

LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

①

INVENTORY

AD-A261 989



DTIC ACCESSION NUMBER



LEVEL

AFOSR-TR-93-0111

DOCUMENT IDENTIFICATION

Dec 92

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

DISTRIBUTION STATEMENT

ACCESSION FOR	
NTIS	GRAB
DTIC	TRAC
UNANNOUNCED	
JUSTIFICATION	
BY	
DISTRIBUTION/	
AVAILABILITY CODES	
DISTRIBUTION	AVAILABILITY AND/OR SPECIAL
A-1	

DISTRIBUTION STAMP

DTIC
ELECTE
MAR 10 1993
S C D

DATE ACCESSIONED



DATE RETURNED

93 3 8 162

~~93 3 4 097~~

DATE RECEIVED IN DTIC

93-04705



REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

H
A
N
D
L
E
W
I
T
H
C
A
R
E

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1992
SUMMER FACULTY RESEARCH PROGRAM
(SFRP) REPORTS

VOLUME 2

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 UPLANDER WAY
CULVER CITY, CA 90230-6608

SUBMITTED TO:

LT. COL. CLAUDE CAVENDER
PROGRAM MANAGER

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE
WASHINGTON, D.C.

DECEMBER 1992

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY

28 Dec 92

Annual 1 Sep 91 - 31 Aug 92

4. TITLE AND SUBTITLE

No 2
1992 Summer Faculty Research Program (SFRP)
Volumes 1 - 16

749620-90-C-0076

5. AUTHOR

Mr Gary Moore

AFOSR-TR- 93 0111

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research & Development Laboratories (RDL)
5800 Uplander Way
Culver City CA 90230-6600

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NI
110 Duncan Ave., Suite B115
Bldg 410
Bolling AFB DC 20332-0001
Lt Col Claude Cavender

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

UNLIMITED

13. ABSTRACT (Maximum 200 words)

The purpose of this program is to develop the basis for continuing research of interest to the Air Force at the institution of the faculty member; to stimulate continuing relations among faculty members and professional peers in the Air Force to enhance the research interests and capabilities of scientific and engineering educators; and to provide follow-on funding for research of particular promise that was started at an Air Force laboratory under the Summer Faculty Research Program.

During the summer of 1992 185 university faculty conducted research at Air Force laboratories for a period of 10 weeks. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS

15. SECURITY CLASSIFICATION OF REPORT

UNCLASSIFIED

16. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

17. SECURITY CLASSIFICATION OF ABSTRACT

UNCLASSIFIED

UL

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1992
SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 2

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Lt. Col. Claude Cavender

Program Manager, RDL
Billy Kelley

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1992

PREFACE

This volume is part of a 16-volume set that summarizes the research accomplishments of faculty, graduate student, and high school participants in the 1992 Air Force Office of Scientific Research (AFOSR) Summer Research Program. The current volume, Volume 2 of 16, presents the final research reports of faculty (SFRP) participants at Armstrong Laboratory.

Reports presented herein are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Research reports in the 16-volume set are organized as follows:

VOLUME	TITLE
1	Program Management Report
2	Summer Faculty Research Program Reports: Armstrong Laboratory
3	Summer Faculty Research Program Reports: Phillips Laboratory
4	Summer Faculty Research Program Reports: Rome Laboratory
5A	Summer Faculty Research Program Reports: Wright Laboratory (part one)
5B	Summer Faculty Research Program Reports: Wright Laboratory (part two)
6	Summer Faculty Research Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory; Frank J. Seiler Research Laboratory; Wilford Hall Medical Center
7	Graduate Student Research Program Reports: Armstrong Laboratory
8	Graduate Student Research Program Reports: Phillips Laboratory
9	Graduate Student Research Program Reports: Rome Laboratory
10	Graduate Student Research Program Reports: Wright Laboratory
11	Graduate Student Research Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory; Frank J. Seiler Research Laboratory; Wilford Hall Medical Center
12	High School Apprenticeship Program Reports: Armstrong Laboratory
13	High School Apprenticeship Program Reports: Phillips Laboratory
14	High School Apprenticeship Program Reports: Rome Laboratory
15	High School Apprenticeship Program Reports: Wright Laboratory
16	High School Apprenticeship Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory

1992 FACULTY RESEARCH REPORTS

Armstrong Laboratory

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
1	Mathematical Modeling of the Human Cardiovascular System Under Acceleration: Analysis of the Arterial Blood Flow using the Thick-Wall Model	Dr. Xavier J. R. Avula
2	GC/MS Protocols for EPA Method 1625	Dr. Stephan B. H. Bach
3	An Intelligent Tutor for Sentence Combining	Dr. Margaret W. Batschelet
4	Articulated Total Body Model Dynamics: Verification and Demonstration Simulations	Dr. Larry A. Beardsley
5	A Study of the Effects of Low Update Rate on Visual Displays	Dr. Jer-Sen Chen
6	Development of a Research Paradigm to Study Collaboration in Multidisciplinary Design Teams	Dr. Maryalice Citera
7	An Approach to On-Line Assessment and Diagnosis of Student Troubleshooting Knowledge	Dr. Nancy J. Cooke
8	Intelligent Decision Making with Qualitative Reasoning	Dr. Aresh Das
9	Effects of Feedback Delay on Tracking	Dr. Kent M. Daum
10	Hand Torque Strength for Small Fasteners	Dr. S. Deivanayagam
11	Design and Modification of a Threat Site/Emitter Laydown Database and Threat Emitter Parameters Database for the Advanced Defensive Mission Planning System and B1-B/B2 Engineering Research Simulators	Dr. John C. Duncan
12	Fractal and Multifractal Aspects of an Electroencephalogram	Dr. John E. Erdei
13	Terminal Arteriolar Density, Total Peripheral Resistance and Optimal Ventriculo-Arterial Coupling	Dr. Daniel L. Ewert
14	TSH and Free T ₄ Testing in Lieu of Traditional "Long" Thyroid Profiles	Dr. Tekum Fonong
15	A Theoretical and Computational Model of Laser Induced Retinal Damage	Dr. Bernard S. Gerstman
16	Estimation of Dioxin Half-Life in the Air Force Health Study	Dr. Pushpa L. Gupta
17	Survival Analyses of Radiated Animals Incorporating Competing Risks and Covariates	Dr. Ramesh C. Gupta
18	Thermal Stress in Seven Types of Chemical Defense Ensembles During Moderate Exercise in Hot Environments	Dr. Richard A. Hengst

Armstrong Laboratory (cont'd)

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
19	The Potential of Electronic Teams for Distance Education	Dr. DeLayne Hudspeth
20	Evaluation of Glucose Monitoring Devices for Use in Hyperbaric Chambers	Dr. Catherine H. Ketchum
21	A New Protocol for Studying Carotid Baroreceptor Function	Dr. Arthur J. Koblasz
22	A Study on Teleoperated System Containing an Exoskeleton	Dr. A. J. Koivo
23	A Composite Person-Factor Measure for Selecting Fighter Pilots	Dr. W. F. Lawless
24	Local Administration of Excitatory Amino Acid Antagonists Attenuates Light-Induced Phase Shifts and C-Fos Expression in the Hamster Suprachiasmatic Nuclei	Dr. L. M. Lutton
25	Collaborative Instructional Development Environment: A Stage for the AIDA	Dr. Robert G. Main
26	A Method for Comparison of Alternative Multiship Aircraft Simulation Systems Utilizing Benefit Estimation	Dr. William C. Moor
27	Vocalizations of Naturally Ranging Groups of the Rhesus Macaque	Dr. B. E. Mulligan
28	Modelling Selective Brain Cooling in Humans	Dr. David A. Nelson
29	Fundamental Skills Training Project: Life Science Tutor	Dr. Carolyn J. Pesthy
30	Model Development for Use in Hyperbaric Oxygenation Therapy Research	Dr. Edward H. Piepmeier, Jr.
31	A Study of the Water Forces Analysis Capability for the ATB Model with Emphasis on Improved Modeling of Added Mass and Wave Damping	Dr. David B. Reynolds
32	A Study of the Effects of Microwave Radiation and Temperature on Amino Acid Metabolism by Mouse Macrophage Cells	Dr. Donald K. Robinson
33	The APT: Psychometric, Measurement, and Predictive Properties	Dr. Mary Rosnowski
34	Analysis of Isocyanates in Spray-Paint Operations	Dr. Walter E. Rudzinski
35	Evaluation of Astronaut Practice Schedules for the International Microgravity Laboratory (IML-2)	Dr. Robert E. Schlegel
36	SEM-EDXA Analyses of Airborne Inorganic Fibers for Qualitative Identification	Dr. Larry R. Sherman
37	Choice Between Mixed and Unmixed Goods in Rats	Dr. Alan Silberberg
38	The Use of Small Groups in Computer-Based Training	Dr. Stanley D. Stephenson

Armstrong Laboratory (cont'd)

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
39	Arterial Compliance and Total Peripheral Resistance for Varying +GZ Forces; Ventriculoarterial Coupling for Maximum Left Ventricular Work; A Latex Tube Model of the Aorta and Nonlinear Pressure Effects in Short Arm Centrifuges	Dr. Richard D. Swope
40	Examination of Response Latencies to Personality Inventory Items	Dr. Lois E. Tetrick
41	The Similarity of Air Force Specialties as Analyzed by Additive Trees, Networks, and Multidimensional Scaling	Dr. Stephen A. Truhon
42	Observations on the Distribution, Abundance, and Bionomics of <i>Aedes albopictus</i> (Skuse) in Southern Texas	Dr. Michael L. Womack

MATHEMATICAL MODELING OF THE HUMAN CARDIOVASCULAR SYSTEM
UNDER ACCELERATION: ANALYSIS OF THE ARTERIAL BLOOD
FLOW USING THE THICK-WALL MODEL

Xavier J. R. Avula
Professor
Department of Mechanical and Aerospace Engineering
and Engineering Mechanics

University of Missouri-Rolla
Rolla, MO 65401

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC 20332

September 1992

MATHEMATICAL MODELING OF THE HUMAN CARDIOVASCULAR SYSTEM
UNDER ACCELERATION: ANALYSIS OF THE ARTERIAL BLOOD
FLOW USING THE THICK-WALL MODEL

Xavier J. R. Avula
Professor

Department of Mechanical and Aerospace Engineering
and Engineering Mechanics
University of Missouri-Rolla

Abstract

Recently developed high performance aircraft would expose the human body to acceleration injury if appropriate life-supporting devices are not incorporated in their design. The cardiovascular system, being central to the maintenance of homeostasis, is adversely affected by sustained accelerations produced during the operation of these aircraft. To aid in the construction of the needed life-support systems, and to understand the response of the cardiovascular system in the adverse environment of high-speed aircraft maneuvers, mathematical models of the cardiovascular system are highly desirable. In this study, the proposed model consists of a thick-walled, highly deformable elastic tube in which the blood flow is described by linearized Navier-Stokes equations. The governing equations, which are coupled nonlinear partial differential equations, were solved by the fourth-order Runge-Kutta numerical scheme. The thick-wall model predicts higher pressures in comparison with the earlier results of the thin-wall model. Further study involving the model of the entire circulation system with realistic tissue properties, and using the finite element method is proposed to develop a comprehensive mathematical model for better understanding.

MATHEMATICAL MODELING OF THE HUMAN CARDIOVASCULAR SYSTEM
UNDER ACCELERATION: ANALYSIS OF THE ARTERIAL BLOOD
FLOW USING THE THICK-WALL MODEL

Xavier J. R. Avula

INTRODUCTION

The human body is well accustomed to the earth's force of gravity. However, recent aerospace developments have occasioned its exposure to the hazards of abnormal gravitational fields which are manifested in the form of vibration, impact, weightlessness, and rectilinear and angular accelerations beyond the levels of human tolerance. Abnormal accelerations on the human body, depending upon their severity and kind, are known to cause a variety of pathophysiologic effects such as headache, abdominal pain, impairment of vision, hemorrhage, fracture, and loss of consciousness [1-4]. All these effects influence the performance of pilots, compromising the high-performance capabilities of the new generation aircraft they operate, let alone the loss of aircraft, and possibly lives, by crash. In the context of space travel, cardiovascular adaptation to microgravity is of concern: shift of blood and tissue fluids from lower to upper body causing headward edema, reduced blood volume, and perhaps altered autonomic control of the circulation are significant issues.

The cardiovascular system, being central to the homeostasis of the organism, is extremely susceptible to the hostilities of the changes in the environmental forces. The design of protective devices and procedures, and medical selection and physiological training protocols which are expected to provide acceleration tolerance and protection for the pilots during aircraft and spacecraft maneuvers must take into consideration the response of the cardiovascular system to the altered gravitational environment. Therefore, a thorough understanding of the cardiovascular system

and its structure-function relationship in abnormal force environments is essential to any effort directed to overcome the gravitational trauma.

The prohibitiveness of actually subjecting the human body to abnormal, altered gravitational forces to gain knowledge of the cardiovascular system's response is obvious. The alternative is to develop a mathematical model and investigate its response. The need for mathematical models and the analysis of model features for prediction of system performance are well recognized in view of the cost and risk involved in testing the original system which, in this case, is the human subject. In this report, an analysis of the arterial blood flow is attempted using the thick-wall model with large deformations. The mathematical formulation is general and could be easily extended to myocardial deformations taking advantage of the topological similarity between the arteries and the heart.

PRESENT STATE OF KNOWLEDGE

There is no dearth of mathematical models of the cardiovascular system in the scientific literature. Womersley [5] and Noordergraaf [6] presented mathematical analyses of some aspects of the cardiovascular system by using lumped parameter models. Taylor [7], and Kenner [8], and Attinger et al. [9] used distributive parameter models to analyze pressure-flow relationships in arteries and veins. Sagawa [10] described the overall circulatory regulation and the influence of the mechanical properties of the cardiovascular system on the control of circulation. Boyers et al. [11] simulated the steady-state response of the human cardiovascular system with normal responses to change of posture, blood loss, transfusion, and autonomic blockage. Porenta et al. [12] have developed a finite element model for blood flow in arteries with taper, branches, and stenoses assuming thin wall and axisymmetry.

Complete models of the cardiovascular system under acceleration stress have been attempted by a few investigators. Avula and

Oestreicher [13] presented a mechanical model consisting of linearized Navier-Stokes and finite elasticity equations to predict blood pooling under acceleration stress assuming a spherical model for the heart and axisymmetric thin tube model for blood vessels. Moore and Jaron [14], and Jaron et al. [15] have developed a non-linear multielement model of the cardiovascular system which can calculate blood pressures and flows at any point in the cardiovascular system. It includes the effects of forces caused by acceleration as well as the effects of several G-protection modes. The problem with multielement models is that the capacitive and resistive elements introduce more uncertainties in the system model as opposed to direct mechanical models.

The stresses and deformations in the myocardium and the arterial wall were of concern to several investigators. Mirsky [16] calculated the shear and bending stresses in the ventricular wall using a thick-walled model of the left ventricle. Demiray [17], and Demiray and Vito [18] calculated the stresses in the arterial and ventricular walls using the theory of finite deformations and experimentally determined strain energy function for soft biological tissue.

In view of the topological similarity between the blood vessels and the heart, the present state of knowledge provides the basis for a mathematical formulation of the blood circulation problem using a thick-wall cylindrical tube which is capable of undergoing large deformations.

MATHEMATICAL FORMULATION

A. Equations of Fluid Motion

The geometry of the artery containing blood in motion is shown in Fig. 1. Let r, θ, z are the cylindrical polar coordinates, and let u, v, w be the velocity components in

the corresponding directions. Assuming axial symmetry in flow and tube deformation, the Navier-Stokes equations for the flow of blood can be written as:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial r} + w \frac{\partial u}{\partial z} = - \frac{1}{\rho_0} \frac{\partial p}{\partial r} + \nu \left[\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial z^2} - \frac{u}{r^2} \right] \quad (1)$$

$$\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial r} + w \frac{\partial w}{\partial z} = - \frac{1}{\rho_0} \frac{\partial p}{\partial z} + \nu \left[\frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \frac{\partial^2 w}{\partial z^2} \right] + g(t) \quad (2)$$

where p is the pressure, ν is the kinematic viscosity, ρ_0 is density of blood, and $g(t)$ is the body force per unit mass caused by the acceleration. The continuity equation is

$$\frac{\partial u}{\partial r} + \frac{u}{r} + \frac{\partial w}{\partial z} = 0 \quad (3)$$

The above equations are non-dimensionalized using a typical length, R_0 , which is the initial (undeformed) radius of the aorta, and U , the average velocity of blood in the aorta. Introducing the new quantities

$$t^* = \frac{tU}{R_0}, \quad r^* = \frac{r}{R_0}, \quad z^* = \frac{z}{R_0}, \quad w^* = \frac{w}{U} \quad (4)$$

$$u^* = \frac{u}{U}, \quad p^* = \frac{p}{\rho_0 U^2}, \quad g^* = \frac{R_0 g}{U^2}, \quad Re = \frac{UR_0}{\nu}$$

After deleting the "stars" for simplicity, Eqs. (1)-(3) in terms of the newly defined variables become

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial r} + w \frac{\partial u}{\partial z} = - \frac{\partial p}{\partial r} + \frac{1}{Re} \left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial z^2} - \frac{u}{r^2} \right) \quad (5)$$

$$\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial r} + w \frac{\partial w}{\partial z} = - \frac{\partial p}{\partial z} + \frac{1}{Re} \left(\frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \frac{\partial^2 w}{\partial z^2} \right) + g(t) \quad (6)$$

$$\frac{\partial u}{\partial r} + \frac{u}{r} + \frac{\partial w}{\partial z} = 0 \quad (7)$$

The boundary and initial conditions are

$$\begin{aligned} u &= \frac{dR_1}{dt} \quad \text{at } r = R_1 \quad t \geq 0 \\ w &= 0 \quad \text{at } r = R_1 \quad t \geq 0 \\ w &= 1 \quad \text{at } w = 0 \quad t \geq 0 \end{aligned} \quad (8)$$

where R_1 is the inside radius of the blood vessel in the deformed state.

B. Equations of Motion for a Thick-Walled Elastic Tube

The theory of large elastic deformations is utilized to describe the time-dependent deformation of the blood vessels. In view of the published results on blood pooling and the consequent cardiac insufficiency, the application of large deformation theory appears necessary. Demiray and Vito [18] have previously used this theory to calculate the deformation of arteries.

The undeformed and deformed cylindrical tubes are shown in Fig. 2. Let r, θ, z represent a point in the wall of the undeformed tube, and R, θ, z in the deformed tube. r_1, r_2 are inside and outside radii, respectively, of the undeformed tube, and R_1, R_2 those at the deformed tube. Axial stretch of the tube is neglected because of tethering caused by the surrounding tissue. Assuming the material of the blood vessels to be homogeneous, incompressible, and isotropic, the stress at any point can be written as [19]:

$$\tau^{ij} = \phi g^{ij} + \psi B^{ij} + P G^{ij} \quad (9)$$

where $\phi = 2(\delta W / \delta I_1)$, $\psi = 2(\delta W / \delta I_2)$, $B^{ij} = I_1 g^{ij} - g^{ij} G$, P is a scalar function which represents a hydrostatic pressure, W is the strain energy function, T_1 and T_2 are the strain invariants, and g^{ij} , g_{ij} , G^{ij} , and G_{ij} are the contravariant and

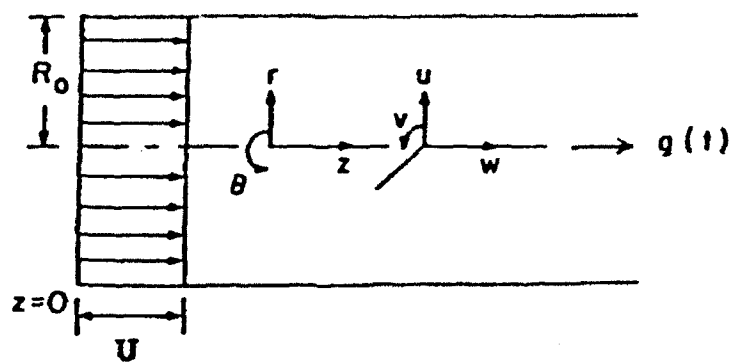


Fig. 1. Blood vessel geometry

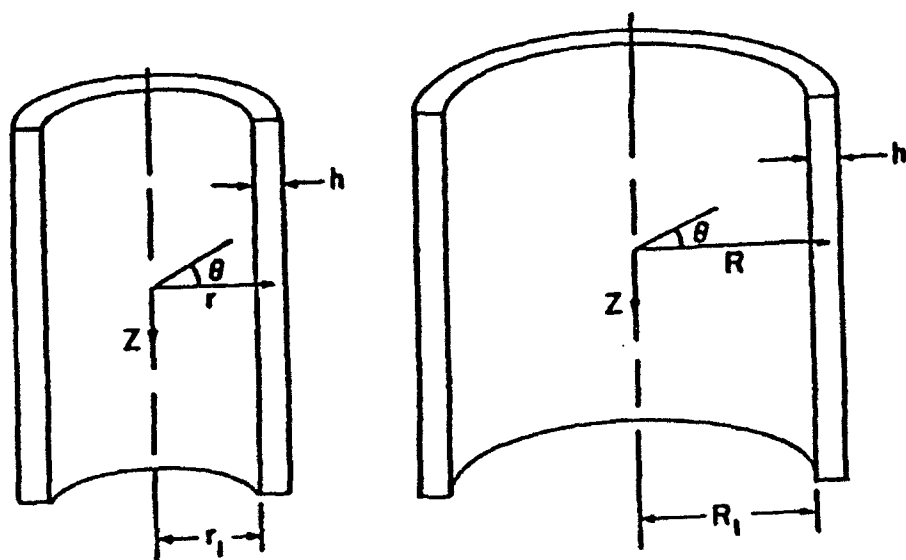


Fig. 2. Undeformed and deformed elastic thick-wall tube

covariant metric tensors. The indices i and j take the values 1, 2, and 3. The equations of motion are given by:

$$\tau^{ij} \parallel_1 + \rho_w F^1 = \rho_w f^1 \quad (10)$$

where \parallel denotes covariant differentiation, ρ_w is the density of the vessel wall, F is the body force, and f is the acceleration. Let us neglect the body force on the vessel wall in comparison to its effect on the fluid flowing in the cylindrical tube. Performing the covariant differentiation on the remaining part of the equation of motion we get

$$\tau_{ri}^{ij} + \Gamma_{jr}^1 \tau^{ij} + \Gamma_{jr}^1 \tau^{jr} = \rho_w f^1 \quad (11)$$

where Γ_{jk} represent the Christoffel symbols of the second kind.

It has been shown that for a biomaterial, a reasonable strain energy function [17,18] is

$$W = \frac{\beta}{2\alpha} [c^{\alpha(I_2 - 3)} - 1] \quad (12)$$

in which α and β are material constants. Defining the circumferential stretch ratio $\lambda = R/r$, the stresses in the r , θ , z directions can be expressed as

$$\tau^{11} = P + \beta \left(1 + \frac{1}{\lambda^2}\right) e^{\alpha(I_2 - 3)} \quad (13)$$

$$R^2 T^{22} = P + \beta (1 + \lambda^2) e^{\alpha(I_2 - 3)} \quad (14)$$

$$\tau^{22} = P + \beta \left(\frac{1}{\lambda^2} + \lambda^2\right) e^{\alpha(I_2 - 3)} \quad (15)$$

Substitution of the above equations and the appropriate Christoffel symbols in Eq. (11) gives the equation of motion in the form

$$\frac{\partial}{\partial R} \left[P + \beta \left(1 + \frac{1}{\lambda^2}\right) e^{\alpha(I_2 - 3)} \right] + \frac{\beta}{R} \left(\frac{1}{\lambda^2} - \lambda^2\right) e^{\alpha(I_2 - 3)} = \rho_w \frac{\delta^2 R}{\delta t^2} \quad (16)$$

The incompressibility condition leads to:

$$R^2 - R_1^2 = r^2 - r_1^2 \quad (17)$$

and differentiating twice yields

$$\frac{\partial^2 R}{\partial t^2} = -\frac{R_1^2}{R^3} \left(\frac{dR_1}{dt} \right)^2 + \frac{1}{R} \left(\frac{dR_1}{dt} \right)^2 + \frac{R_1}{R} \frac{d^2 R_1}{dt^2} \quad (18)$$

With p_1 , p_2 denoting the pressure on the inside and outside wall, respectively, of the blood vessel, the use of the boundary conditions, $\tau^{11} = -p_1(t)$ at $R = R_1$ and $\tau^{11} = -p_2(t)$ at $R = R_2$, substituting Eq. (18) into Eq. (16) and integrating yields

$$p_1(t) - p_2(t) = \rho_w R_1 \frac{d^2 R_1}{dt^2} \ln \frac{R_2}{R_1} + \left(\frac{dR_1}{dt} \right)^2 \rho_w \left[\ln \frac{R_2}{R_1} + \frac{1}{2} \left(\frac{R_2^2}{R_1^2} - 1 \right) \right] - \beta \int_{\lambda_1}^{\lambda_2} \frac{1 + \lambda^{2\alpha} (\lambda^2 + \frac{1}{\lambda^2} - 2)}{\lambda^3} d\lambda \quad (19)$$

The relationship $I_2 = 1 + \lambda^2 + 1/\lambda^2$ has been used to obtain Eq. (19). Equation (19) is nondimensionalized using the variables

$$p^* = \frac{P}{\rho_o U^2}, \quad R_1^* = \frac{R_1}{R_o}, \quad R_2^* = \frac{R_2}{R_o}, \quad t^* = \frac{tu}{R} \quad (20)$$

$$\beta^* = \frac{\beta}{\rho_o U^2}, \quad \rho_w^* = \frac{\rho_w}{\rho_o}$$

The resulting equations will be in terms of the "starred" (non-dimensional) variables. However, by deleting the "stars" for convenience, these equations become

$$p_1(t) - p_2(t) = \rho_w R_1 \frac{d^2 R_1}{dt^2} \ln \frac{R_2}{R_1} + \rho_w \left(\frac{dR_1}{dt} \right)^2 \left[\ln \frac{R_2}{R_1} + \frac{1}{2} \left(\frac{R_2^2}{R_1^2} - 1 \right) \right] - \beta \int_{\lambda_1}^{\lambda_2} \frac{1 + \lambda^{2\alpha}}{\lambda^3} e^{\alpha(\lambda^2 + 1/\lambda^2 - 2)} d\lambda \quad (21)$$

The initial conditions are:

At time $t = t_o$, $R_1 = R_o$, $dR_1/dt = u$, radial velocity of the fluid.

A fourth-order Runge-Kutta numerical scheme was used to solve the linearized version of Eqs. (5) and (6) in conjunction with Eq. (21). The requirement of brevity and the universal familiarity of the numerical method preclude the presentation of computational details here.

RESULTS AND DISCUSSION

As mentioned above, the solution consists of solving Eqs. (5), (6) and (21). The following constants were used in the solution: $2R = 2.94$ cm, $U = 11.9$ cm/s, $h = 0.164$ cm, $\rho = 1.05$ g/cc, $\rho = 1.05$ g/cc, $\nu = 0.038$ Stoke, $\alpha = 0.8$ and $\beta = 113,500$ dynes/sq cm. In this study, the impact deceleration profile shown in Fig. 3 was used to determine the pressure development in a segment of the human aorta.

Figure 4 depicts a comparison of computed pressures in thin-wall (solid line) and thick-wall (dotted line) models. Clearly, the thick-wall model predicts higher pressures in the vessel segment considered as opposed to the thin-wall. Also, the thick-wall solution deviated from the thin-wall with increasing time. The reason for this can be explained as follows: the thick-wall blood vessel, by virtue of its higher stiffness, is deformed less, thus expending less strain energy of deformation. Then a portion of the impact energy is translated into greater fluid pressure. (Such an argument can also be extended to explain hypertension in atherosclerotic patients). Since this is only a preliminary study, a more careful evaluation of the solution needs to be conducted. The deviation of pressure with increasing time is intriguing. A closer look at the constitutive equation of the tissue is necessary.

RECOMMENDATIONS

Most investigators working with mathematical models of the circulatory system have addressed the tissue deformation and blood

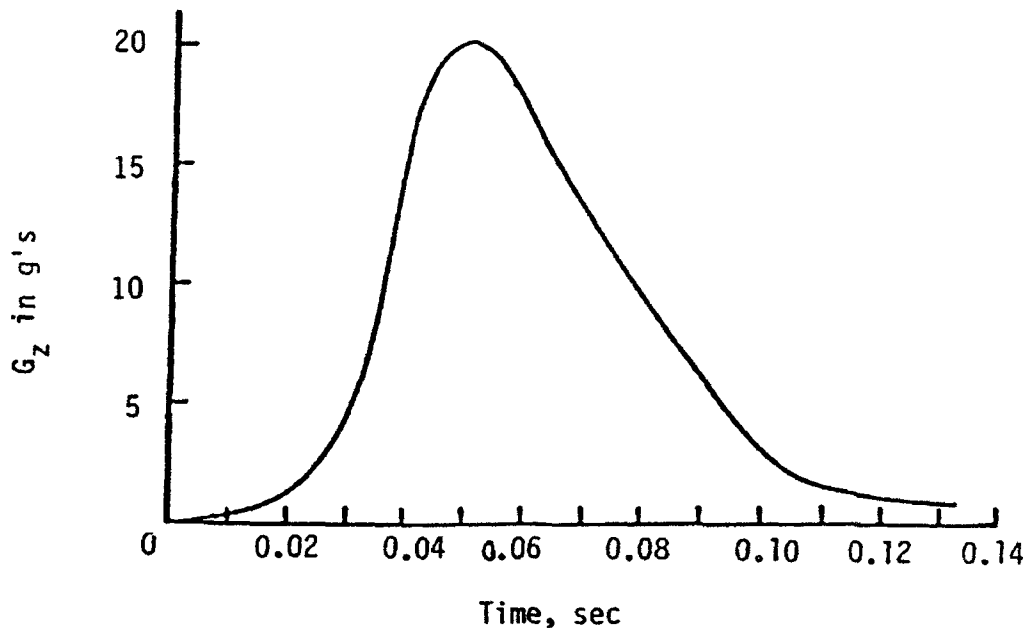


Fig. 3. Experimental deceleration profile (see [13])

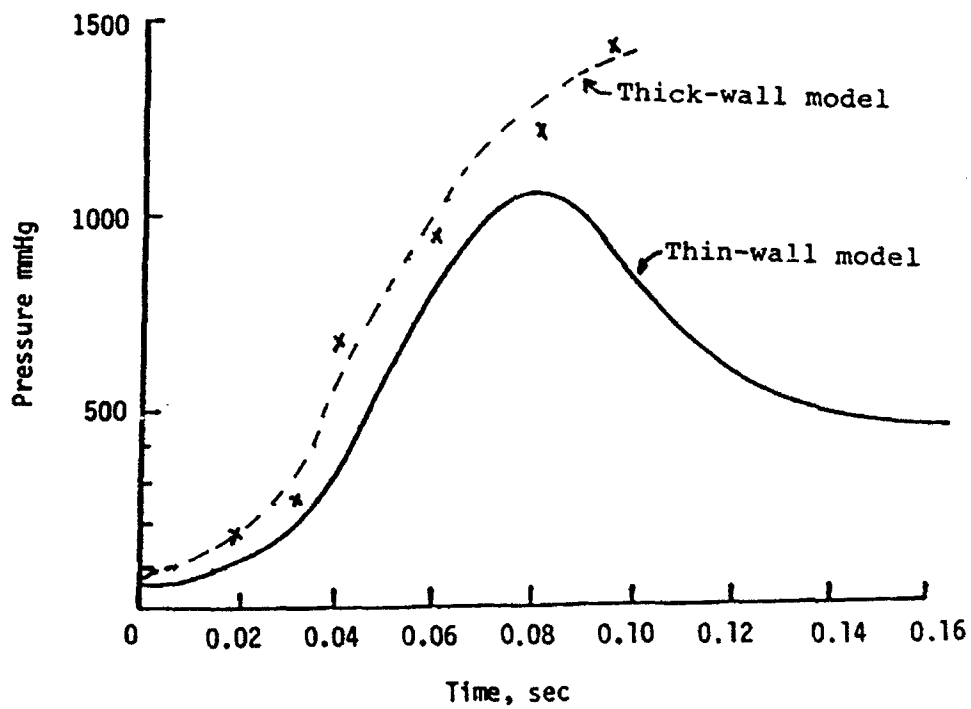


Fig. 4. Pressure in the aorta for the deceleration profile of Fig. 3

flow patterns as separate issues. It is desirable to incorporate into the model realistic tissue properties and address the above issues simultaneously. The heart and blood vessels have a complicated non-symmetric geometry. As the finite element method is generally more suitable for problems with non-regular geometry (see Oden [20]), it is recommended that a comprehensive model of the entire cardiovascular system be constructed using this method with realistic tissue properties. The variable tissue properties can be easily accommodated into the discretized elements resulting in more realistic models. One of the recent issues in acceleration trauma is the combined stress of subclinical cardiovascular disease and acceleration. Particularly, the mitral valve prolapse among the pilots manifests in lower tolerance for gravitational stress with susceptibility to developing mitral regurgitation leading to heart failure [21]. The finite element model of the heart can effectively address this problem if accurate description of the ventricles and the heart valves are available.

ACKNOWLEDGMENT

The author wishes to thank Dr. Sherwood Samn of the Armstrong Laboratory, Brooks Air Force Base, San Antonio, Texas for helpful discussions during this investigation. The Summer Research Fellowship awarded by the Air Force Office of Scientific Research to conduct this study is gratefully acknowledged.

REFERENCES

- [1] Hiatt, E. P., Meehan, J. P., and Galambos, R., Reports on Human Acceleration, Publication 901, National Academy of Sciences - National Research Council, Washington, DC, 1967.
- [2] Space Sciences Board, NAS-NRC, Physiology in the Space Environment, Vol. 1 - Circulation, Publication 1485A, National Academy of Sciences, Washington, DC, 1968.
- [3] Burton, R. R., Leverett, S. D., Jr., and Michaelson, E. D., "Man at High Sustained +Gz Acceleration: A Review", Aerospace Medicine, Vol. 45, pp. 1115-1136, 1974.
- [4] Whinnery, J. E., "Medical Considerations for Human Exposure to Acceleration Induced Loss of Consciousness", Aviation, Space and Environmental Medicine, pp. 618-623, July 1991.

- [5] Womersley, J. R., "Mathematical Analysis of the Arterial Circulation in a State of Oscillatory Motion", Wright Air Development Center, Technical Report WADC-TR-56-164, 1958.
- [6] Noordergraaf, A., "Hemodynamics", in Biological Engineering, Schwan, H., (Ed.), McGraw-Hill, New York, 1969.
- [7] Taylor, M. D., "The Input Impedance of an Assembly of Randomly Branching Elastic Tubes", Biophysical Journal, Vol. 6, pp. 25-51, 1966.
- [8] Kenner, T., "Flow and Pressure in Arteries", in Biomechanics: It's Foundations and Objectives, Fung, Y. C. et al. (Eds.) Prentice-Hall, Englewood Cliffs, New Jersey, 1972.
- [9] Attinger, E. O., Anne, T., Mikami, T., and Sugawara, H., "Modeling of Pressure Flow Relationships in Veins", in Hemorheology, Copley, A., (Ed.), McGraw-Hill, New York, 1964.
- [10] Sagawa, K., "The Circulation and Its Control I: Mechanical Properties of the Cardiovascular System", in Engineering Principles in Physiology, Vol.2, Brown, J. H. V. and Gann, D. S., (Eds.), Academic Press, New York, 1973.
- [11] Boyers, D. G., Cuthbertson, J. G., and Luetscher, J. A., "Simulation of the Human Cardiovascular System: A Model with Normal Responses to Change of Posture, Blood Loss, Transfusion, and Autonomic Blockage", Simulation, Vol. 18, No. 6, pp. 197-206, 1972.
- [12] Porenta, G., Young, D. F., and Rogge, T. R., "A Finite Element Model of Blood Flow in Arteries Including Taper, Branches, and Obstructions", Journal of Biomechanical Engineering, Vol. 108, No. 2, pp.161-167, 1986.
- [13] Avula, X. J. R. and Oestreicher, H. L., "Mathematical Model of the Cardiovascular System Under Acceleration Stress", Aviation, Space, and Environmental Medicine, pp. 279-286, January 1978.
- [14] Moore, T. W., Jaron, D., "Cardiovascular Model for Studying Circulatory Impairment Under Acceleration", IEEE Transactions on Engineering in Medicine and Biology", Vol. 10, No. 1, pp. 37-40, 1991.
- [15] Jaron, D., Moore, T. W., and Bai, J., "Cardiovascular Responses to Acceleration Stress: A Computer Simulation", IEEE Proceedings, Vol. 76, No. 6, pp. 700-707, 1988.
- [16] Mirsky, I., "Left Ventricular Stresses in the Intact Human Heart", Biophysical Journal, Vol. 9, pp. 189-208, 1969.

- [17] Demiray, H., "Stresses in Ventricular Wall", Transactions of the ASME, J. of Applied Mechanics, pp. 194-197, June 1976.
- [18] Demiray, H. and Vito, R. P., "Large Deformation Analysis of Soft Biomaterials", Developments in Theoretical and Applied Mechanics, Vol. 8, Proc. of the 8th Southeastern Conference on Theoretical and Applied Mechanics, pp. 515-522, 1976.
- [19] Green, A. E. and Adkins, J. E., Large Elastic Deformations and Nonlinear Continuum Mechanics, Clarendon Press, Oxford, 1960.
- [20] Oden, J. T., Finite Elements of Nonlinear Continua, McGraw-Hill, Inc., New York, New York, 1972.
- [21] Samn, S., "The Role of Subclinical Cardiovascular Disease in High-G Flying: A Mathematical Modeling Approach", The Physiologist, Vol. 35, No. 1, Suppl., 1992.

GC/MS PROTOCOLS FOR EPA METHOD 1625

Stephan B.H. Bach

**Assistant Professor
Division of Earth and Physical Sciences
University of Texas at San Antonio
San Antonio, Texas 78249-0663**

**Final Report for:
Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

September 1992

GC/MS PROTOCOLS FOR EPA METHOD 1625

Stephan B.H. Bach
Assistant Professor
Division of Earth and Physical Sciences
University of Texas at San Antonio

Abstract

A Finnigan 5100 GC/MS is set-up to operate EPA Method 1625. The tuning parameters for the instrument had to be optimized after restarting the instrument. Application of the quantitation software available with the 5100 to operate EPA Method 1625 is undertaken. A short description for troubleshooting various instrumental problems encountered while making the instrument and the software operational are described.

GC/MS PROTOCOLS FOR EPA METHOD 1625

Stephan B.H. Bach

INTRODUCTION

Today, in Environmental Chemistry there are numerous challenges for the chemist. One area, that is constantly changing and growing, is Environmental regulations. Because the regulations are being constantly updated, compliance with the regulations becomes increasingly difficult. Compliance involves using published EPA Methods for determining the quantities of target compounds (unknowns in the sample) specified by the particular EPA Method in routine samples. Our goal was to develop EPA Method 1625¹ for production runs of waste water samples on a Finnigan 5100 GC/MS instrument. The difficulty in this project was not the chemistry directly, since it is documented in the Code of Federal Regulations (CFR),¹ but adapting these published methods to the requirements of the laboratory and reactivating the 5100.

APPARATUS

These experiments were carried out using a Finnigan 5100 quadrupole mass spectrometer (MS) with a GC interface. The MS has a mass range from 4 to 850 atomic mass units (amu). We swept the quadrupoles at a rate of one scan every 0.5 seconds, and scanned a mass range from 35 to 450 amu. The ionizer was set at 70 eV. Extraction and lens voltages were varied, as necessary, to tune the instrument for EPA Method 1625.

The GC oven attached to the 5100 is a Finnigan model 9611. The GC portion of this work was done using a commercially available DB-5 column (5% diphenyl dimethyl siloxane), with a film thickness of 1 micron, and 30 m long. The injections were done using a modified splitless routine. The injection port

temperature was routinely held at 250 °C. The transfer oven temperature was held at 270 °C. Helium (99.999%) was used as the carrier gas.

TUNING

In order to meet EPA guidelines very stringent instrumental conditions must be maintained. The instrumental parameters are not set by specific instrumental settings, but are determined by the fragmentation pattern of perfluorotributylamine (CF43). The parent peak for this compound is $m/z = 671$. But under the ionization conditions prescribed by the EPA Method 1625, the base peak in the mass spectrum is $m/z = 69$. The CF_3^+ ion is a small fragment of the parent compound which indicates very high ionization energies. The target percentages of the various fragments which are observed when the instrument is tuned properly are given in Table I. The major problem with the tuning procedure is that a large variety of settings will achieve a similar fragment distribution. If the settings are within typical ranges for the instrument (Table II), this does not pose a problem. But when the tune settings fall outside of those settings, problems arise. These problems are due to the electrostatic field lines created within the ionizer by the extractor and lens voltages. Distortions can also occur when the ion program and ion energy voltages are set at extremes. These problems are usually observed as aberrations in the peak shape. Either the resolution of the peak will not meet specifications or the peak will appear as a doublet or multiplet. When these problems are not detected, the mis-settings can result in poor quantitation for the amount of the target compound present or misidentification of the compound.

Once the mass spectrometer is tuned properly with CF43, a GC run is then be made with decafluorotriphenylphosphine (DFTPP) at 20ng/ul in either methanol or methylene chloride. The mass spectrometer then needs to be de-tuned slightly to give the proper fragmentation ratios for the DFTPP. The proper percentages are listed in Table III. The mass spectrometer is now ready to acquire data.

The EPA methods as described in the CFR seem to be straight forward analytical techniques. Unfortunately this seeming simplicity is an illusion.

Applying the EPA protocols to field samples results in considerable complexity. In order to obtain reliable results one must keep the GC/MS within parameters described in the EPA methods by verifying daily that the instrument is properly tuned. This may seem simple at first until one considers the number of variables which must be kept constantly in control. The most obvious variable from sample to sample is how the sample is introduced into the GC/MS. Great care should be taken with the injection technique in order to get reproducible results. Column condition must also be monitored from sample to sample. This can be accomplished by observing the GC peak shape and the retention times of the respective compounds. The primary problem that we encountered with the GC was solvent tailing. This was due to improperly setting the starting temperature of the GC oven. Care must be taken to set the initial oven temperature either above or below the boiling point of the solvent. Otherwise the solvent will tend to boil off slowly during the time that the GC oven is at its initial temperature (usually 4 minutes).

The temperature program for the GC can be a variety of different sequences of settings. The goal is to separate the target compounds sufficiently so that they can be correctly identified from their retention time and their mass spectrum. There are a variety of published GC oven sequences which seem to work equally well. An important point though, is to acquire data until the final peak of the target compounds (benzo(ghi)fluoranthene) is eluted. At this point the acquisition of data can be stopped.

QUANTITATION²

For quantitating the target unknowns several methods are allowable, external standard, internal standard, and isotope. We will be using the internal standard method. This involves spiking each sample with a known amount of a compound. The ideal compounds are deuterated species which do not coelute with any of the target compounds and whose mass spectrum is readily identifiable. Typical compounds used are d10-anthracene, d5-nitrobenzene, d14-terphenyl. The

internal standards should be picked so that they elute through out the GC run and not just in one region of the GC.

Once the standards have been decided upon, known concentrations of the target compounds and internal standards are run. The concentration of the internal standards are kept at 20ng/uL. The target compound concentrations vary from 20 to 200 ng/uL. The concentrations of the target compounds should cover the expected ranges of concentrations of the target compounds that will be found in the field samples. We used concentrations of 20, 50,100, and 200 ng/ul.

Creating a Library....

A new library should be created for each EPA method to be run on the instrument. Be sure to set the proper library before running AUTOQUANT. To change to the proper library, use the SET4 XX command. XX is the name of the library to be used for the particular EPA Method. Each EPA Method to be run on the instrument will require its own Library for quantitation. These libraries are constructed using the chromatograms collected from the runs of target compounds and internal standards where the concentrations of the target compounds are known. The chromatogram is displayed using the CHRO command. The mass spectrum of each chromatographic peak is displayed using the Y';ED command string. The Y' command enhances the mass spectrum of the GC peak by performing a subtraction subroutine to eliminate extraneous peaks and then enters the current library. The ED command then brings up the library editor.

Once in the library editor, the mass list from the mass spectrum is displayed with the relative intensity (1000 = 100%) of each of the masses. All but the five or six most intense peaks are eliminated using the A -j,k command string. The "-" tells the program to eliminate peaks between j and k. If eliminating a single mass, use the same command string , but set k equal to j, otherwise all peaks larger than j will be eliminated. Higher mass peaks should be retained preferentially over smaller masses, because they will be more unique to the compound than the smaller masses. For those compounds with a limited number of peaks, retain as

many as possible even if it means retaining some peaks that are not very intense. These peaks will be better than none at all when the quantitation program goes to calculate fit and intensity.

With the mass table reduced to an acceptable number of masses (5-6), the mass spectrum can now be inserted into the Library for this method. The I command inserts the mass spectrum and its parameters at the end of the current library. The I' command inserts the mass spectrum into the current slot of the library. The software will prompt for answers to several questions. The manual is fairly clear except for several points. The reference compound is entered as a library location in the current library (XX.00). The XX is the two letter code for the current library and 00 is the library entry number. If a mistake is made during this procedure, type N when asked if everything is correct (at the end of the procedure). This will not write the responses to disk. Then use the I' command to insert the mass spectrum into the current library location. This will avoid inserting erroneous duplicate entries into the library which would occur if the I command were used in this situation. After all target compounds and the internal standard have been inserted into the library in this fashion, check the library entries for any mistakes. Once this is done, create the Master Library list.

....Lists

The Master Library list is used by the quantitation program to find the target compounds in the unknown samples. It is created using the LL MASTERLIST command where MASTERLIST is the name given to the master list. Once the master list is created it must be saved using the W command. Two other lists must also be created from the quantitation library, the retention time list (RT) which contains the library location of the internal standard to be used in a limited region of the GC spectrum, and the target compound (TC) list which contains the library locations of the target compounds that the quantitation software will be searching for. The TC lists should contain no more than ten compounds to operate properly. Several RT and TC lists will have to be created for most EPA Methods.

These RT and TC lists are then combined into a Driver List which lists the order that the software will follow in searching for the compounds and the internal standards to be used for those compounds. The manual contains a complete description of how to put these files together.

The next step is to create the 11 Table which at first contains the retention times of the target compounds and the internal standards. After running TCA, it will be filled in by the quantitation software to produce the quantitation report. The command string used to create the 11 Table is TCALIB FILENAME,11NAME, MASTER. FILENAME is the filename of a standard GC run. The 11NAME is the filename for the 11 Table, and MASTER is the filename given to the Master List used for the current analysis. Before the 11 Table is used it should be checked for accurate target compound retention times. If the retention times are good, move the retention times from column five to column three using the UPDATE 11NAME command string. Quantitation can now be attempted.

The TCA FILENAME,11NAME, DRIVER command string is used to start the quantitation. FILENAME is the name of the GC file to be quantitated. For the software to function properly, the GC file name and the calibration table file name should be the same except for the three letter endings. The 11NAME is the file name of the converted 11 Table (correct retention times in column 3). The DRIVER file contains the RT and TC file pairs in the sequence that the software is suppose to perform the reverse search for the compounds in the GC scan. The TCA software uses the reverse search subroutine to match the known peaks in the library with the peaks found in the GC file.

For becoming familiar with the software and how to prepare the various files for Autoquant, it is not necessary to have the instrument perfectly tuned. Once it was apparent that the day to day reproducibility of the 5100 was acceptable, we went ahead and set up the files necessary for the Autoquant software. Day to day reproducibility was satisfactory if autotune could tune the instrument approximately every other day.

TROUBLESHOOTING

Autotune

If the autotune routine is not able to tune the instrument, set the default tuning parameters by typing DF in MSDS. This will reset the tuning parameters in the file used by autotune. Try autotune again to see if the instrument will tune up. If not, several instrumental parameters need to be checked. First, verify that the voltages going to the MS are correct by using the test points on the ionizer board (located in the card cage under the GC). Before checking the voltages, use SCAN and set the scan time to 4 or 5 seconds. This is done so that variable voltages such as those used for the ion program routine can be checked with a standard digital multi-meter (DMM). A short in the ionizer is usually indicated by a zero voltage where there should be a nonzero voltage (voltages are listed in the manual). This can be verified by disconnecting the multi-pin connector at the rear of the MS. If there is a short in the ionizer, the zero voltage should now be at the proper setting. The ionizer can easily be shorted because of the ion-volume slipping in the ionizer. We inserted the ion-volume manually as the last step in putting the instrument back together.

Ionizer Short

Another ionizer problem which effects the operation of the MS, is a short of the heater to the ionizer. This will cause the wrong voltages to be put on the ionizer. The usual indication of this situation, is if all else is functioning properly, but the mass spectrum is very noisy. To check for this situation, turn the heater off by setting the ionizer temperature to a value that is below room temperature. A check should also be made of the pin outs from the ionizer and heater. This is done by checking the multipin connector at the back of the instrument. A diagram of the pin assignments is given in the manual. A short is indicated if the resistance between the two pins is very low (100 ohms or less). The resistance between the two pins should be infinite.

Transfer Oven

The transfer oven temperature is critical for eluting and detecting the compounds towards the end of the GC run. These are usually polyaromatic hydrocarbons (PAC). If the transfer oven temperature is too low, then the PAC's will become "trapped" in the transfer oven and will fail to elute out before the end of the GC run. Indications of this problem are either the absence of the PAC's when they are known to have been in the sample, or very broad PAC peaks in the GC. The transfer oven temperature should be kept at least 20 °C above the maximum temperature that the GC oven reaches during the GC run.

1. EPA Method 1625 Revision B, Title 40 Code of Federal Regulations, part 136, Appendix A, p 504 (July 1, 1989).
2. The current Version 8.0 software will not operate the quantitation programs properly. Version 6.5 is the most recent software release that works properly for using the quantitation software.
3. EPA Method 1625 Revision B, Title 40 Code of Federal Regulations, part 136, Appendix A, p 504 (July 1, 1989); "Reference Compound to Calibrate Ion Abundance Measurement in Gas Chromatography-Mass Spectrometry Systems", J.W. Eichelberger, L.E. Harris, W.L. Budde, *Anal. Chem.*, **47**, 955 (1975).

TABLE I: CF43 Fragmentation Ratios³

Mass	Target Percentage
69	base peak, 100% relative abundance
219	30.0 to 75.0 % of m/z = 69
414	1.40 to 4.00 % of m/z = 69
502	0.80 to 4.00 % of m/z = 69

Isotope Ratio Targets

70/69	0.80 to 1.30 %
220/219	3.50 to 5.20 %
415/414	7.20 to 10.80 %
503/502	8.10 to 12.10 %

TABLE II: Typical Instrument Settings

Resolution (HI)	128 ± 10
Resolution (LO)	128 ± 10
Ion Energy	5 ± 5
Ion Program	10 ± 5
Lens Voltage	15 ± 10
Extractor	10 ± 5

TABLE III: DFTPP Fragmentation Ratios³

51	30.0 to 60.0 % of m/z = 198
68	<2 % of m/z = 69
70	<2 % of m/z = 69
127	40.0 to 60.0 % of m/z = 198
198	base peak, 100% relative abundance
199	5.0 to 9.0 % of m/z = 198
275	10.0 to 30.0 % of m/z = 198
365	>1 % of m/z = 198
441	present, but less than 443
442	>40 % of m/z = 198
443	17.0 to 23.0 % of m/z = 442

AN INTELLIGENT TUTOR FOR SENTENCE COMBINING

Margaret W. Batschelet
Assistant Professor
Division of English, Classics and Philosophy

University of Texas at San Antonio
6900 North Loop 1604 West
San Antonio, Texas 78249-0643

Final Report for:
Air Force Office of Scientific Research
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC

August 1992

AN INTELLIGENT TUTOR FOR SENTENCE COMBINING

Margaret W. Batschelet
Assistant Professor
Division of English, Classics, and Philosophy
University of Texas at San Antonio

Abstract

Sentence combining has become a well established technique for teaching grammar, syntax, and some forms of punctuation. Using sentence combining, students practice producing syntactically complex sentences by adding to, deleting, and rearranging simple "kernel sentences." Research on sentence combining by Frank O'Hare and others has established that students who have sentence combining practice produce more syntactically mature sentences as well as better written compositions. Some work has been done on using computers for sentence combining, but only one sentence combining program has been written: a drill-and-practice program for Apple II computers. The R-WISE tutor, a tutoring system aimed at improving the reading and writing skills of high school freshmen, includes a sentence combining module as part of its editing tools. The sentence combining and reduction module allows students to revise clusters of sentences within their drafts which need to be either combined or broken into more than one sentence. Students can work with their sentences or go through a set of exercises designed to improve their sentence combining and reducing skills. Some tutoring is provided, based in part on the number of sentence clusters or run-ons remaining after the module is completed and on student performance on the exercises.

AN INTELLIGENT TUTOR FOR SENTENCE COMBINING

Margaret W. Batschelet

INTRODUCTION

Sentence combining is a technique used to teach elements of grammar, syntax, and punctuation. Using sentence combining, students practice producing syntactically complex sentences by adding to, deleting, and rearranging simple "kernel sentences." Sometimes students are cued as to the particular combinations that must be made: they are provided with clues that suggest which words are to be added or deleted or they are given patterns to follow in producing the finished sentence. In other cases students are given the kernels and allowed to make whatever combinations seem most appropriate.

For example, a student might be given the following kernels:

All the English students loved their teacher.

The students were *cool*.

The teacher was *charming*. (O'Hare, 90)

Here the student is cued to use only the words *cool* and *charming* from the second and third kernels. The resulting combination would be

All the *cool* English students loved their *charming* teacher.

Or the student might be given the following:

Pattern: One of my favorite movies, *Casablanca*, is on tonight.

The little boy is in front.

The little boy is a redhead.

The little boy is my cousin.

In this case, the student would combine the kernels to produce a sentence that follows the pattern given above. The resulting combination would be

The little boy in front, the redhead, is my cousin.

Sentence combining allows teachers to accomplish a number of purposes within a relatively simple set of exercises. As Erika Lindemann points out:

We can demonstrate how sentences are constructed without beginning with grammatical terms or singling out errors in a particular student's paper. We can avoid terminology altogether or introduce it after students understand how to perform the transformations. We can describe punctuation, mechanics, and conventions of edited American English in the context of writing, not isolated from it. Because students will combine sentences in various ways, we can also point out rhetorical choices, differences in emphasis which might work more effectively in some larger contexts than in others. (137)

Because of these advantages, the members of the design team working on the R-WISE (Reading and Writing in a Supportive Environment) reading and writing tutor decided to incorporate a sentence combining module as part of the tutor's editing section. Although the tutor could not be used for daily practice in sentence combining, placing such a module in the editing section of the program would allow students to use sentence combining as a means of revising sentence structure and improving their syntax.

In this report I will present a short survey of the research concerning sentence combining and of other sentence combining computer programs. Then I will present an overview of the sentence combining module which I helped to design and implement as part of R-WISE.

METHODOLOGY

Sentence Combining Research

The idea of sentence combining derives from Noam Chomsky's theories of generative-transformational grammar published in *Syntactic Structures* in 1957. A variety of studies in the 1950s and 60s had demonstrated that traditional grammar study in the English classroom had little positive effect (see Braddock). After Chomsky's theories became widespread, several researchers attempted to study whether using some form of generative-transformational grammar instruction might produce better results in the classroom. An initial study by Bateman and Zidonis suggested that transformational grammar instruction did just that, but subsequent researchers argued that the size of the sample in the study (twenty students) was too small to draw definitive conclusions and that it was possible that the students improved their syntactic maturity not because of their study of transformational grammar but because of the sentence manipulations that that grammar introduced.

A subsequent study by John Mellon attempted to measure the effect of transformational grammar and sentence combining on a larger student sample. Mellon divided 247 Boston seventh graders into three groups: one group received instruction in transformational grammar with a heavy sentence combining component, a second group received traditional grammar instruction, and a third group studied no grammar but received additional lessons in literature and composition. At the end of a year, the first group showed significant growth in syntactic maturity, while the second group showed much less.

The most significant study of sentence combining as a pedagogical technique was published by Frank O'Hare in 1973. O'Hare wished to test the hypothesis that the increase in syntactic maturity which the students in the earlier studies experienced was the result of their practice in sentence combining rather than their formal instruction in transformational grammar. The Florida seventh grade students in O'Hare's study were given simple instructions and cues for sentence combining, as well as the same sentence combining exercises used by Mellon. They were given no formal instruction in transformational grammar and followed the normal seventh grade curriculum for literature and composition. A control group also followed the seventh grade curriculum, but received no practice or instruction in sentence combining. The students' writing was assessed for syntactic maturity and writing quality in pre- and post-tests. Again the students in the experimental groups showed significant improvement both in syntactic maturity and in the quality of their overall writing; their scores were also superior to those of the control group.

In the years following O'Hare's study, other studies by researchers such as Donald Daiker, Max Morenberg, and Andrew Kerek have confirmed that sentence combining is a powerful tool in teaching both grammar and punctuation. The technique has been widely adapted in textbooks for elementary, secondary, and college students and is seen as "a viable way to help students write cleanly focused, grammatically correct, and thoughtfully worded sentences" (Connors and Glenn, 1992).

Sentence Combining and Computers

Given the widespread acceptance of sentence combining and the growing use of computers in English classrooms, it is hardly surprising to find that some teachers have suggested using computers to teach combining. Several instructors have recommended using word processors as sentence combining tools, although none have run controlled studies of their use. See, for example, the papers by Anne Wright, Joseph Milner, and Gwen Solomon.

In 1989 a Florida study gave fourth graders classroom instruction in sentence combining while allowing them to practice using data files presented in a word processing program (Laframboise). The final results of this study have not yet been published, but the preliminary report was favorable.

In 1984 Ann Humes and her colleagues at the Southwest Regional Laboratory for Educational Research and Development at Los Alamitos, California constructed a simple branching program which allowed elementary and middle school students to practice combining sentence kernels in an electronic workbook setting. The resulting program seemed to work well for the students, but there were some problems with slow response as a result of the branching structure.

Some "style check" programs also give users some feedback on their sentence structure. For example, as part of the analysis of a text, Right Writer will give the user what are termed "Sentence Structure Recommendations" which include the number of sentences with multiple clauses and the number of simple sentences.

Although all of these programs offer some tentative direction in merging sentence combining with computer technology, they represent only a beginning. The Humes and Laframboise programs are limited to workbook exercises, while programs like Right Writer offer no real advice on sentence combining as a revision technique, relying instead on omnibus comments such as "Most sentences contain multiple clauses. Try to use more simple sentences" and "Few compound sentences or subordinate clauses are being used." Although an experienced writer might be able to make use of such advice, it is probable that an inexperienced writer could not.

Thus although the idea of performing sentence combining on a computer has been explored, its full potential has not yet been reached.

AN INTELLIGENT TUTOR FOR SENTENCE COMBINING

R-WISE

R-WISE (Reading and Writing in a Supportive Environment) is an Intelligent Tutoring System which is part of the Fundamental Skills Tutor project. Its aim is to improve the reading and writing skills of high school freshmen. R-WISE consists of a series of individual modules focused on various parts of the reading and writing process; many of these modules are intelligent tutors, collecting data about individual student performance on particular tasks and tutoring on specific skills. The Sentence Combining and Reduction Module of R-WISE is one of these intelligent tools.

The Sentence Combining and Reduction Module was planned as an editing tool, placed in the revision section of the tutor. The purpose of the module is to provide students with a means of analyzing and revising their sentences, and, to a lesser extent, their punctuation and grammar. Thus the module is not really intended to teach sentence combining as O'Hare did, but rather to use it as a means of improving the prose style of a student's draft.

The Sentence Combining Module

The tutor begins by analyzing the student's draft for clusters of two or more "sentence clusters" meeting the following criteria:

- o Length of fewer than twelve words
- o No coordinating conjunctions
- o No relative pronouns
- o No subordinating conjunctions
- o No conjunctive adverbs.

The length criterion was taken from O'Hare's study; the other criteria were added to avoid picking up short sentences which were, despite their length, relatively complex. In practice, these criteria have helped to make the module more selective, centering on sentence clusters which need to be combined. If the student's draft reveals no sentence clusters meeting these criteria, then the student is allowed to pass over the module.

The module itself consists of three major sections: a "workboard", a series of relationship explanations, and a "lesson." After the student's draft is analyzed, each sentence cluster first appears on a "Workboard" space at the lower left of the screen. The student is

also given a histogram representing the sentence lengths of each paragraph so that she can see the patterns of variation she has in her sentence length. The student is then given a choice of options: she can press a Work button which enables her to begin working with the sentence cluster on the Workboard, she can press an Explanation button which brings forward a screen with explanations of seven possible combination types, or she can press a Pass button which allows her to pass on making any changes in the sentence cluster.

The Work option is perhaps the simplest of the three. If the student sees immediately how the sentences can be combined, she presses Work, makes her corrections, and presses a Finished button when they are completed. The revised sentence is then placed in her draft in place of the original sentence cluster, and the next sentence cluster appears on her workboard.

If the student is uncertain how to combine the sentences, or if she is confused about sentence combining in general, she can press the Explanation button. On the Explanation screen she is given a list of possible relationships she might see between sentences (addition, subtraction, alternatives, explanation, illustration, conclusion/causal, and time); these relationships are based on work by Ross Winterowd and Bonnie Meyer. Clicking on any of the words in the relationship list will give her a more extended explanation of the relationship along with examples of combinations. After studying her sentence cluster and looking through the possible relationships, the student can press Work to try one of the combinations on her

sentences. Again, pressing Finished will enter the revised sentence into her draft.

While on the Explanation screen, the student can choose the Lesson option. Pressing this button brings her to a set of sentence combining exercises using the seven relationships explained on the Explanation screen. Each exercise contains a set of kernel sentences and a series of blanks into which the kernels can be entered to make a new, combined sentence; sometimes the kernels are entered in their entirety and sometimes they must be modified to fit a finished combination. These exercises thus represent a kind of cued practice in that the student is given a pattern to follow. The student can work through all of the exercises, or she can press a Quit button at the lower left of her screen to return to the Explanation screen and the Workboard.

If the student feels that the cluster of sentences is effective as is, or if she is simply unwilling to combine them, she can press Pass. She will then be given a message asking if she is certain that the sentences are satisfactory as is; if she presses Yes, the sentence cluster will be highlighted in her draft for future reference. If she presses No, she will be returned to the Workboard to make another choice. At the end of the Sentence Combining session, after all of the student's sentence clusters have been examined, she will be shown the sentence clusters she passed on once again. At this point the student will be given one more chance to work with the sentences, with buttons for deleting them, rewriting them, or, once again, passing.

After all of the sentence clusters in the draft have been either revised or passed, the tutor once again analyzes the draft to see how many sentence clusters remain. The final count of sentence clusters is added to the student model as another indication of the student's progress on sentence combining.

R-WISE also includes a Sentence Reduction module within the same suite of tools, although the two tools are not intended to be used during the same session. Using this tool the tutor analyzes the student's draft for "run-on" sentences, sentences which fit one or more of the following criteria:

- o Length more than five words above the average for the paper
- o More than one coordinating conjunction
- o Commas separated by groups of three or more words

The Sentence Reduction tool is organized following the same pattern as the Sentence Combining tool: the student has options for working, getting an explanation of various techniques for revising run-on sentences, doing a set of exercises for run-on revisions, or passing on the sentence. The Explanation section presents techniques for dealing with three types of run-on sentences: clauses joined without punctuation, clauses joined with more than one coordinating conjunction, and independent clauses joined by a comma (a comma splice). At the end of the session, the tutor analyzes the draft to find the number of run-on sentences remaining; that number then becomes part of the student model.

Intelligence

The student model for the sentence combining module (as for much of R-WISE) is currently under design, thus its exact measurements can only be suggested. However, some tutoring is possible. Simple, event-driven help can be offered for situations such as excessive time without character entry. In addition, if the student has a certain percentage (e.g., more than 50%) of sentence clusters remaining after she has completed the module, she can be directed to read her draft for sentence variety, being aware of the effect that a series of short, simple sentences may have on her reader. Within the Lesson section of the tutor, the student can be tutored on the various relationship types if it seems that she is having difficulty with one type of combination. Moreover, records of difficulty on individual combinations within the lesson can be combined with records from other modules within the tutor to obtain some indication of difficulties with particular organizing principles such as description or causation (see Meyer). The student can then be tutored on a higher cognitive level. For example, if the student is consistently having problems with causation relationships (corresponding to the conclusion/causal relationship in the sentence combining module), then the tutoring on the sentence combining module can focus on the nature of the cause and effect relationship.

CONCLUSIONS

Obviously, the Sentence Combining tool will not teach sentence combining in isolation. It is intended to suggest various forms of

sentence revision to a student; ideally, it will be used in conjunction with classroom instruction in sentence combining. And, considering how widespread sentence combining has become as an instructional tool, it is not unreasonable to assume that most students will have received some classroom experience in combining.

The Sentence Combining and Sentence Reduction tools will be tested starting in September 1992 at MacArthur High School in the Northeast Independent School District in San Antonio. Twenty-four classes of ninth graders will use the various modules of the R-WISE program for a year as part of their basic classwork. These students, as well as control groups, will receive pre- and post-tests in both reading and writing; writing samples from both experimental and control groups will be evaluated, using a standard analytical scale, on the following criteria:

- o Clear and logical organization of ideas
- o Development and support of a central idea
- o Appropriate response to the purpose/audience of a written composition
- o Control of the English language
- o Mature syntactic structures

The final criterion most obviously involves the skills students will ideally acquire from sentence combining, yet the development of a mature style will influence several other criteria as well. Sentence combining can play an integral part in the overall success of the tutor in improving students' writing skill.

Since the Sentence Combining tool is only one of a series of tools within R-WISE, and since it will not be used with the intensity of the sentence combining exercises in O'Hare's study, we do not hope to achieve results that rival O'Hare's. However, we hope that the availability of a sentence combining module will help the student users of R-WISE to work more closely with sentence structure and ultimately to achieve a more mature and complex syntax.

References

- Bateman, Donald, and Frank J. Zidonis. *The Effect of Transformational Grammar on the Writing of 9th and 10th Graders*. Research Report 6. Urbana IL: National Council of Teachers of English, 1966.
- Braddock, Richard, Lloyd-Jones, R., and Schoer, L. *Research in Written Composition*. Urbana, IL: National Council of Teachers of English, 1963.
- Chomsky, Noam. *Syntactic Structures*. The Hague: Mouton, 1957.
- Connors, Robert and Cheryl Glenn. *The St. Martins Guide to Teaching Writing*. Second edition. New York: St. Martins Press, 1992.
- Daiker, Donald. "Sentence-Combining and Syntactic Maturity in Freshman English." *College Composition and Communication*. 29(1978): 36-41.
- Humes, Ann. "Computer Instruction on Sentence Combining." Southwest Regional Laboratory for Educational Research and Development. Los Alamitos, CA. ERIC ED 239 580.
- Kerek, Andrew, Donald A. Daiker, and Max Morenberg. "Sentence-Combining and College Composition." *Perceptual and Motor Skills*. 51(1980): 1059-157.
- Laframboise, Kathryn L. "The Effects of Sentence-Combining Using Word Processing Technology on the Reading Comprehension and Writing Fluency of Low-Achieving Fourth Grade Students." ERIC ED 302 845.
- Lindemann, Erika. *A Rhetoric for Writing Teachers*. Second edition. New York: Oxford University Press, 1987.
- Mellon, John. *Transformational Sentence-Combining: A Method for Enhancing the Development of Syntactic Fluency in English Composition*. Urbana: National Council of Teachers of English, 1969.
- Meyer, Bonnie J.F. "Signaling the Structure of Text." *The Technology of Text: Principles for Structuring, Designing, and Displaying*. vol. 2. Englewood Cliffs, NJ: Educational Technology Publications, 1985. 64-89.
- Milner, Joseph O. "Micro to Main Frame Computers in English Education." ERIC ED 255 915.
- Morenberg, Max, Donald A. Daiker, and Andrew Kerek. "Sentence-Combining at the College Level: An Experimental Study." *Research in the Training of English* 12(1978): 245-56.

O'Hare, Frank. *Sentence-Combining: Improving Student Writing without Formal Grammar Instruction*. Urbana, IL: National Council of Teachers of English, 1973.

Right Writer. *User's Manual*. RightSoft, Incorporated. 1989

Solomon, Gwen. "writing with Computers." *Electronic Learning* 5(Nov.-Dec. 1985): 39-43.

Winterowd, W. Ross. "The Grammar of Coherence." *Contemporary Rhetoric: A Conceptual Background with Readings*. Ed. W. Ross Winterowd. New York: Harcourt, Brace, Jovanovich, Inc.: 1975. 225-33.

Wright, Anne. "Computers as Instructional Tools." ERIC ED 275 000.

Acknowledgments

This research was performed in collaboration with Dr. Linda Woodson of the University of Texas at San Antonio and Dr. Patricia A. Carlson of Armstrong Laboratory, with support from Ms. D'Ann Johnson and Ms. Elaine Hiltzfelder of MacArthur High School of San Antonio. Programmers for the Sentence Combining and Reduction Module were Jennifer K. Morales and George Stallsworth.

**Articulated Total Body Model Dynamics
Verification and Demonstration Simulations**

**Larry A. Beardsley
Assistant Professor
Department of Math, Physics, & Computer Science
Savannah State College
State College Branch
Savannah, GA 31404**

**Final Report for:
Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC**

September, 1992

**Articulated Total Body Model Dynamics
Verification and Demonstration Simulations**

by

Larry A. Beardsley

Abstract

The Armstrong Laboratory at Wright-Patterson Air Force Base, Ohio utilizes a program called DYNAMAN, written by General Engineering and Systems Analysis Company (GESAC) for the purpose of performing predictive simulations of human body dynamics resulting from various external force applications. The DYNAMAN program runs on a PC based computer and uses the Articulated Total Body (ATB) model program code for the dynamics simulations. Although the ATB model code was developed to study human body dynamics, the program can be used to simulate general coupled rigid body dynamics. The present study examined some of the basic mechanisms in the model to validate their response against closed form mathematical solutions. Particular emphasis was placed on studying elastic and energy absorbing impacts and it was found the simulation results agreed very closely with the closed form mathematical solutions.

ACKNOWLEDGMENTS

I am extremely grateful to the Air Force Material Command, the Air Force Office of Scientific Research, and to Research Development Laboratories for the opportunity to participate in the summer research program for faculty and graduate students. This summer proved to be a very rewarding experience professionally.

In particular, I give thanks to Dr. Ints Kaleps, branch chief of AL/CFBV, and the lab focal point, for his direction in the research effort, and to Brian Self for the time he gave in acquainting me with the DYNAMAN program. Also, I am appreciative to Dr. Louise Oberlgefell for her effort in familiarizing me with the ATB model. Many thanks are extended to all of the people at AL/CFBV for their friendliness and encouragement.

I. INTRODUCTION

The Articulated Total Body (ATB) model is used at the Armstrong Laboratory (AL) to study human body biomechanics in various dynamic environments, especially aircraft ejection with windblast exposure. The model was developed by Calspan Corporation (formerly Cornell Aeronautical Laboratory) in the early mid-seventies for the U.S. Department of Transportation. The model is designed primarily to evaluate human body response when modeled as a system of rigid bodies when subjected to a dynamic environment consisting of applied forces and interactive contact forces.

Although the ATB model was originally developed to model the dynamic response of crash dummies and with later modifications, the response of the human, the ATB model is quite general in nature and can be used to simulate a wide range of physical problems that can be, according to Obergefell [3], approximated as a system of connected or free rigid bodies. The model has been used to simulate such widely diverse physical phenomena as human body dynamics, the motion of balls in a billiards game, and the transient response of an MX Missile suspended from cables in a wind tunnel.

The approach used in the ATB model is to consider the body as being segmented into individual rigid bodies called segments. Each segment has a mass of the body between joints or, in the case of single-jointed segments, such as the foot, the mass distal to the joint. An example would be the left upper arm segment which represents the mass of the body between the shoulder joint and the elbow. Segments are assigned mass and moments of inertia and joined at locations representing the physical joints of the human body, such as the shoulder joint or the knee joint.

The model can simulate a system which is made up of one or more segments, connected or free. The system or body is assembled as a chain of individual segments. More complex bodies can take on a tree-like structure, with several chains branching out from several connected segments. The limitation of this approach is that a closed loop for connected segments is not allowed. An example of a closed loop would be the right and left lower arms being connected by a joint at the hands.

A program called DYNAMAN was written by the General Engineering and Systems Analysis Company (GESAC) for the purpose of executing the ATB model on a personal computer.

II. THE DYNAMAN PROGRAM

DYNAMAN is a software package written for a personal computer that allows an analyst to simulate the dynamics of an occupant of a vehicle involved in a collision. The software consists of these modules:

1. A preprocessor that allows an analyst to interactively set up an input file to carry out a simulation.

2. A simulation module which accepts the input file that was created using the preprocessor and produces output files that contain various kinematic variables that describe the three-dimensional motion of the occupant inside the vehicle and the contact force tables.
3. A postprocessor that can be used to view the output of the simulation in pictorial, graphical, and tabular forms.

III. OBJECTIVES OF THE RESEARCH EFFORT

The DYNAMAN program is relatively new with updates to be considered. It is important that the output results be accurate. To determine accuracy of the DYNAMAN program based on the ATB model, a series of simulations were performed to verify the correctness of predictions against closed form mathematical solutions. The effects of inertial, elastic, and viscous forces were examined as well as geometry based calculations. When possible, solid graphics representations of the responses were generated.

IV. THE MATHEMATICS OF THE MODEL

The ATB model utilizes many reference coordinate systems within the program. The primary coordinate systems used in the model are the inertial vehicle local body segment, principal, joint, and contact ellipsoid reference coordinate systems. The specification of each reference coordinate system requires an origin and a direction cosine matrix (usually initially specified by three rotation angles, yaw, pitch, and roll) which relates one reference coordinate system to another. All of the coordinate systems are considered to be orthonormal.

The ATB model assumes that the coordinates of the origin of the inertial reference coordinate system are zero and all other coordinate systems are specified with respect to this system. The user may equate the origin of the inertial reference coordinate system to any convenient point from which his data are referenced. The frame of reference is arbitrary and is partially specified by defining which way is down by the values supplied for the components of the gravity vector. It has been customary to supply (0, 0, g) as the components of the gravity vector to specify that the positive Z axis is pointing downward.

Up to six vehicles with specified motion can be defined. The origin of each of the vehicle coordinate systems is arbitrary, and any convenient reference point may be chosen for the input and output data which would be most meaningful.

Each body segment has a local reference coordinate system, and a mass and principal moments of inertia. The local reference coordinate system has its origin at the segment mass center. The principal moments of inertia are with respect to this origin. The principal axes for the principal moments are specified with respect to the local reference system. The ellipsoidal contact surface origin and orientation are also specified with respect to the local reference system. A body

segment can have up to six degrees-of-freedom. A body segment is given an initial position, orientation, linear and angular velocity and its motion is then computed for the remainder of the simulation subject to any imposed constraints. The motion of the body segment cannot be specified unless the segment is also defined as a vehicle. Contact planes can also be attached to the body segments to provide another way for the body segment to interact with the environment. Although the orientation of the segment local reference coordinate systems can be arbitrarily defined, the convention has been to choose the axes so that when the body is in an upright, standing position, the positive Z axis is downward, the X axis is to the front, and the Y axis is to the body's right.

The principal axes are fixed with respect to the segment local reference axis, and their orientation needs to be specified only once. Other model input and output referring to the segments is in terms of the body segment local reference coordinate system. The segment data is transformed to principal axes, the equations are numerically solved in the inertial system, and the predicted responses are transformed to the local systems for output. When the input description refers to the local body reference, the local and not the principal moment of inertia reference coordinate system is implied. Mathematically what takes place is the following:

The matrix

$$I^{(L)} = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

is a non-zero symmetric matrix whose diagonal elements represent the respective moments of inertia in the local coordinate system. It is diagonalized to obtain the matrix $I^{(P)}$, where

$$I^{(P)} = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix},$$

which is a diagonal matrix whose diagonal elements are the eigenvalues (the solution to $\det(\lambda I - I^{(L)}) = 0$) of the characteristic equation of $I^{(L)}$, and these elements are the principal moments. The conversion from the local reference system to the principal reference system may be represented, where, for example, the Z axis is the Z_L axis in the local system and the Z_P axis is the Z axis in the principal reference system. The origin is at the segment mass center.

The direction cosine matrix, A_{10} , is a 3x3 matrix which converts the components of a vector in one reference system (0) to components in another reference system (1) and is formed in the following manner:

$$A_{10} = \begin{bmatrix} \hat{i}_1 \cdot \hat{i}_0 & \hat{j}_1 \cdot \hat{i}_0 & \hat{k}_1 \cdot \hat{i}_0 \\ \hat{i}_1 \cdot \hat{j}_0 & \hat{j}_1 \cdot \hat{j}_0 & \hat{k}_1 \cdot \hat{j}_0 \\ \hat{i}_1 \cdot \hat{k}_0 & \hat{j}_1 \cdot \hat{k}_0 & \hat{k}_1 \cdot \hat{k}_0 \end{bmatrix},$$

where the dot product of $\hat{i}_1 \cdot \hat{i}_1$, for example, is the cosine of the angle between the two vectors (since the dot product is defined to be the product of the magnitudes of the two vectors times the cosine of the angle between them). The multiplication of $A_{M \times 3}$ and $X_{3 \times 1}$, yields a 3×1 matrix which consists of components of the X vector in the other system. Since the direction cosine matrix is orthogonal, $A A^T = A^T A$ holds.

A direction cosine matrix is assigned to each segment which defines its angular orientation. This matrix is updated during integration by the use of quaternions [2]. The integrator integrates the quaternion equation. The yaw, pitch, and roll angles or Euler angles are also computed at each time step from the direction cosine matrices.

Problem No. 1

The first problem was to model a ball which is dropped from a given height onto a plane with given friction and force-deflection characteristics. Several simulations were run varying these characteristics as well as giving the ball different initial velocities.

The equations used in this model are the following:

The laws of motion where "v" is the velocity at time "t", "v₀" is the initial velocity, "a" is the acceleration (386.1 in/sec²), and "z" is the distance the ball falls:

- (1) $v = v_0 + at$
- (2) $z = v_0 t + \frac{1}{2} gt^2$
- (3) $v^2 = v_0^2 + 2gz$

Other basic equations used are:

$$\Sigma F = ma$$

and

$$W = mg,$$

where F is a force acting on a body or plane, m is the mass, W is weight, and g is gravity, or 386.1 in/sec².

The deflection force, $F(\delta)$ was initially given a value of $F(\delta) = 20\delta$.

To validate calculations of the DYNAMAN simulations, time velocities were compared with equation (8) which was derived through fundamental laws of physics and integration.

Given that $\Sigma F = ma$, then

$$a = \frac{1}{m} \Sigma F$$

$$= \frac{1}{m} [W - F(\delta)] \quad (1)$$

where W is positive since the positive z -axis is pointing downward.

Also, we have that $W = mg$. So,

$$W = mg \Rightarrow m = \frac{w}{g} = \frac{1lb}{386.1 \text{ in/sec}^2}$$

(the weight was given to be 1 lb)

$$\Rightarrow \frac{1}{m} = 386.1 \frac{\text{in}}{\text{lbsec}^2} \quad (2)$$

Substituting this equation in (1), we have:

$$a = 386.1 \text{ in/lbsec}^2 \cdot [W - F(\delta)] lb$$

$$= 386.1 \text{ in/sec}^2 [W - F(\delta)] \quad (3)$$

Substituting the right hand side of the above equation into (1), we have:

$$a_z = 386.1 \frac{\text{in}}{\text{sec}^2} [W - F(\delta)] \quad (4)$$

Since the weight is given to be 1 lb in the problem to be simulated, and the force is 20 lbs, the resulting equation is:

$$a_z = 386.1 \frac{\text{in}}{\text{sec}^2} [1 - 20z'] \quad (5)$$

$$\text{where } Z' = Z - Z_0 = \delta$$

$$\text{Since } a_z = \frac{dv}{dt} \text{ and } dz' = v dt = dt \cdot \frac{dz'}{v},$$

equation (5) is equivalent to:

$$\begin{aligned} \frac{dv}{dz'} &= 386.1 (1 - 20z') \\ \int v dv &= 386.1 \int (1 - 20z') dz' \\ \text{Therefore } \frac{v^2}{2} &= 386.1 (z' - 10z'^2) + C \end{aligned} \quad (6)$$

The velocity at which the ball impacts the plane is found by solving:

$$v_f^2 = 2ay,$$

where $y=7$ (since the ball falls 7 inches) and $a=386.1$

$$v_f = 73.521 \text{ in/sec} \quad (7)$$

This result gives us the initial condition of $v=73.521$ when $Z' = 0$.

Substituting (7) into equation (6), and solving for the constant C, we find $C \approx 2702.7$. So, equation (6) becomes

$$\begin{aligned} \frac{v^2}{2} &= 386.1 (Z' - 10Z'^2) + 2702.7 \\ -v^2 &= 772.2 (Z' - 10Z'^2) + 5405.4 \end{aligned} \quad (8)$$

Of course as the force deflection varies, the polynomial representing $F(\delta)$ may be substituted in equation (5), and will yield a new expression to integrate.

Problem No. 2

A second problem which was modeled was that of a ball(s) given an initial velocity inside a cube and analyzing the motion and forces. The problem was divided into two parts. The first part was (a) to model the perfect gas law and (b) to analyze the data giving the ball(s) different values for force-deflection characteristics and an initial velocity.

(a) To verify the perfect gas law using the DYNAMAN, it may be helpful to review a concise derivation of the perfect gas law. Present [4] offers an elaborate derivation of the law, and Halliday and Resnick [3] arrive at the same result in a more succinct presentation.

Briefly, the perfect gas law may be arrived at in the following way.

Given a cube having edges of length ℓ , let the faces of the cube be normal to the x-axis where the positive x-axis points to the right.

Let S_1 and S_2 be the faces normal to the x-axis of area ℓ^2 . The velocity, V , may be resolved into components V_x , V_y , and V_z for a given molecule. Assuming that each wall is perfectly elastic, a particle will rebound after a collision with S_1 , with its x-component of velocity reversed. Since there will be no effect on the velocity in the y-direction, V_y , or in the z-direction, V_z , the change in the particles momentum will be

$$\Delta p = p_f - p_i = (-mv_x) - (mv_x) = -2mv_x \quad (1)$$

where P_f is the final momentum, P_i is the initial momentum, m is the mass, and V_x is the velocity in the x-direction, normal to S_1 .

Therefore, the total momentum which S_1 receives is $2mv_x$ since the total momentum is conserved.

Now assume the same particle reaches S_2 without striking another particle in the interim. The time required for the particle to transverse the cube is:

$$\frac{l}{v_x}$$

When the particle impacts S_2 , it will have its x-component of the velocity reversed again, and it will return to S_1 . Assuming there are no collisions in between, the round trip will take a time of

$$\frac{2l}{v_x}$$

So, the number of collisions per unit of time the particle imparts to S_1 is given to be $v_x/2l$, and the rate at which momentum is transferred to S_1 is

$$2mv_x \frac{v_x}{2l} = \frac{mv_x^2}{l} \quad (2)$$

The total force, the rate at which momentum is imparted to S_1 by all of the molecules on S_1 , may be found by summing up mv_x^2/l for all of the particles. To find the pressure, we divide this force by the area of S_1 , which is l^2 . Given that m is the mass of each molecule, we have

$$p = \frac{m}{l^3} (v_{x_1}^2 + v_{x_2}^2 + \dots) \quad (3)$$

where v_{x_i} is the velocity of molecule m_i . If N is the total number of particles in the container, (2) can be rewritten as

$$p = \frac{mN}{V} (v_{x_1}^2 + v_{x_2}^2 + v_{x_3}^2 + \dots) \quad (4)$$

However, mN is the total mass of the gas and mN/V is the mass per unit volume or the density ρ . Also, the expression

$$\left(\frac{v_{x_1}^2 + v_{x_2}^2 + \dots}{N} \right)$$

is the average value of $v_{x_i}^2$ for all of the particles in the container.

This may be denoted as \bar{v}_x^2 .

$$\text{For any given particle, } v^2 = v_x^2 + v_y^2 + v_z^2.$$

However, since in general there are many particles moving at random, the average values of v_x^2 , v_y^2 , and v_z^2 are equal, and the value of each is one-third of the average value of v^2 . Also, without loss of generality, the molecules may assume motion along any one of the three axes. Therefore,

$$\begin{aligned} \bar{v}_x^2 &= \frac{1}{3} \bar{v}^2, \text{ and} \\ p &= \rho \bar{v}_x^2 = \frac{1}{3} \rho \bar{v}^2 \end{aligned} \quad (5)$$

Equation (1), (3) and (5) are the three equations which will be used to compare with the simulation.

Methodology for Problem 2

A cube was constructed of six planes whose normals to each plane pointed towards the interior. For the first simulation, a ball (sphere) of weight .1 pound was used with an initially velocity of 451.6 in/sec in the x-direction and -451.6 in/sec in the z-direction with no angular velocity and gravity not a factor. The deflection force was initially given to be $F(\delta) = 100\delta$ pounds where δ is in inches and the coefficient is in lbs/in. Friction was given a value of zero, and the restitution characteristic was initially set to one to indicate that the total momentum is conserved.

For subsequent simulations, the number of balls was increased and/or the velocity was increased. Later, the friction force was varied to verify that linear energy was transferred to angular energy. Since there is a constraint on the maximum number of plane segment contacts, a maximum of five balls was used.

The Mathematics Behind Simulation 2(a).

Since the average force on a given plane in the cube is approximately the same after a large number of impacts, one plane could be considered for analyzing the force. For each ball impact, $F(\delta) = 100\delta$ is the given force in pounds. Now, if the plane is impacted with a force F_1 over a time Δt and F_2 over a time Δt , and so on for the time the ball is deformed, the total force the

segment imparts to the wall is expressed by

$$\sum_{t=0}^{t-t_f} F \Delta t, \quad (1)$$

and for more than one segment, the total force is given by

$$\sum_{\text{Segment 1}}^{\text{Segment n}} \sum_{t=0}^{t-t_f} F \Delta t, \quad (2)$$

where the average force is given by

$$\sum_{\text{Segment 1}}^{\text{Segment n}} \sum_{t=0}^{t-t_f} F \Delta t / t. \quad (3)$$

The pressure P is found by dividing (8) by the area, A, of a given plane.

$$\sum_{\text{Segment 1}}^{\text{Segment n}} \sum_{t=0}^{t-t_f} F \Delta t / tA = P \quad (4)$$

Now, this result of P should compare with the P in (5), page 4-12.

The Mathematics Behind Problem 2(b).

The second part of the problem using the cube is to analyze the results of a ball(s) given an initial velocity with different values of force deflection characteristics.

First, as with problem 1, an expression had to be derived using the fact that

$$a_z = 386.1 \text{ in/sec}^2 [W - F(\delta)] \quad (1)$$

However, we assume gravity does not play a role in this problem. So, we have

$$a_z = 386.1 \text{ in/sec}^2 (-F(\delta)) \text{ where } F(\delta) = 100\delta \quad (2)$$

Integrating as we did in the first problem, we find that

$$\int v dv = 386.1 \int -100x' dx' \text{ where } x' = x - x_0 = \delta \quad (3)$$

$$\Rightarrow \frac{v^2}{2} = 386.1(-50x'^2) + C \quad (4)$$

find

Initially, $v_i = 451.6 \text{ in/sec}$. at $x' = x - x_0$. Therefore, substituting the initial conditions, we that $C = 101971.28$. So,

$$\frac{v^2}{2} = 386.1(-50x'^2) + 101971.28 \quad (5)$$

$$v^2 = -38610x'^2 + 203942.56 \quad (6)$$

which is an expression of velocity in terms of the deflection x' . This expression was used to validate the results of DYNAMAN.

I.V. RESULTS

Data from the first simulation was collected and compared with results one would expect to obtain from the closed form equations derived.

With the first simulation of a free falling ball, different values of the deflections, Z' , for given times were matched with the corresponding velocities using the time history tables generated by DYNAMAN. Using the equation (8) of the first problem, $v^2 = 772.2 (z' - 10z'^2) + 5405.4$, which gives the velocity as a function of deflection, the velocity was calculated for a given deflection value. The calculated value for velocity was then compared to the velocity given in the table. For example, using the closed form equation above, the calculated velocity for $Z' = .4$ is 66.923 in/sec to the nearest thousandth. The corresponding velocity for $Z' = .4$ is at $t = 196$ msec (see table 1) and is 66.939 in/sec (see table 2). Using the error formula

$$\frac{[v_e - v_o]}{[v_e]} - \text{relative error,}$$

where v_e is the expected velocity and v_o is the observed velocity, we find the relative error is approximately .02 percent. Similarly, using a value of $z' = .641$ in the equation, we find the value for the velocity should be 52.226. Using the time history table, the value for the velocity corresponding to $z' = .641$ is 52.260 at $t = 200$ (see tables 1 and 2 again). The value generated by DYNAMAN was again accurate to .06%. This verification was made to ensure that the time history table was yielding accurate values after full deformation but before the ball left the plane.

Other verifications were made, such as the magnitude of the greatest deflection generated by DYNAMAN compared to the value one obtains from the closed form equation. Of course, this verification was made by letting the velocity equal zero in the equation and solving for z' . The table value was accurate to within .1 percent. Validation of the numerical integration was also reinforced by observation of plots generated by DYNAMAN. For example, a parabolic curve indicates the symmetry of the normal force curve representing when the ball impacts the plane and then leaves it (see Figure 2) for the case where the restitution factor is 1.0. Figure 1 represents the velocity curve for the first model. Also, Figures 3 and 4 represent the output for model 1 with a restitution factor of .5.

Several other simulations were run using different force deflection characteristics. For example, using a restitution factor of .5 versus a factor of 1.0 in the first simulation, a loss of energy is expected. The tables indicate this as well as the plots. When higher degree polynomial functions were used for the force of deflection, integration was performed again to derive closed form equations. These equations were again analyzed to validate the results generated by DYNAMAN in the same manner as the first simulation mentioned before. Due to the requested brevity of this paper, these results will be given in a supplemental paper and sent to the laboratory.

The results of the second problem using closed form equation were again very close to the results of DYNAMAN. Due to the limitations of the number the plane segment contacts, the problem could not be simulated with a larger number of balls.

VI. CONCLUSIONS AND RECOMMENDATIONS

Using the two models mentioned in this paper, the DYNAMAN generated results were very close to what one would expect from the closed form mathematical solutions for most cases. However, in some instances where the restitution factor was given a value of a number greater than .9 but less than 1.0, the unloading curve was above the force deflection curve, which is physically not possible. Also, when the restitution factor was given a value less than 1.0, tabular results did not compare as close as one would expect with respect to the loss of energy. These differences may be due to numerical error using a certain method of approximation. This should be analyzed in more depth to see if better results may be obtained, or to see if there may be an error in the algorithm. In any case, a more detailed analysis should be carried out using the same models, and some new ones to analyze all possible physical forces.

Since the DYNAMAN program is based on the ATB model, and the elements of this model are fairly well documented, a more detailed study of the previous simulations, as well as new ones, can be carried out.

It is the researcher's opinion that a more in depth study of the ATB model using DYNAMAN would be very useful in making sure that the accuracy of results is optimized.

TABLE 1

FILE: B:ONE4.008
CONTACT FORCES - SEGMENT PANELS VS. SEGMENTS

TIME (MSEC)	PANEL 1 (ground) VS. SEGMENT 1 (lt)				CONTACT LOCATION (IN)		
	DEFL- ECTION (IN)	NORMAL FORCE (LBS)	FRICTION FORCE (LBS)	RESULTANT FORCE (LBS)	(lt REFERENCE) X	Y	Z
189.000	-0.104	0.00	0.00	0.00	0.000	0.000	0.000
190.000	-0.031	0.00	0.00	0.00	0.000	0.000	0.000
191.000	0.043	0.85	0.00	0.85	0.000	0.000	0.043
192.000	0.116	2.32	0.00	2.32	0.000	0.000	0.116
193.000	0.189	3.79	0.00	3.79	0.000	0.000	0.189
194.000	0.261	5.23	0.00	5.23	0.000	0.000	0.261
195.000	0.332	6.63	0.00	6.63	0.000	0.000	0.332
196.000	0.400	8.00	0.00	8.00	0.000	0.000	0.400
197.000	0.465	9.31	0.00	9.31	0.000	0.000	0.465
198.000	0.528	10.55	0.00	10.55	0.000	0.000	0.528
199.000	0.586	11.73	0.00	11.73	0.000	0.000	0.586
200.000	0.641	12.82	0.00	12.82	0.000	0.000	0.641
201.000	0.691	13.81	0.00	13.81	0.000	0.000	0.691
202.000	0.736	14.71	0.00	14.71	0.000	0.000	0.736
203.000	0.775	15.51	0.00	15.51	0.000	0.000	0.775
204.000	0.809	16.19	0.00	16.19	0.000	0.000	0.809
205.000	0.838	16.75	0.00	16.75	0.000	0.000	0.838
206.000	0.860	17.20	0.00	17.20	0.000	0.000	0.860
207.000	0.876	17.51	0.00	17.51	0.000	0.000	0.876
208.000	0.885	17.70	0.00	17.70	0.000	0.000	0.885
209.000	0.888	17.77	0.00	17.77	0.000	0.000	0.888
210.000	0.885	17.70	0.00	17.70	0.000	0.000	0.885

TABLE 2

FILE: B:ONE4.008
POINT REL. VELOCITY (IN / SEC)

TIME (MSEC)	POINT (0.00, 0.00, 0.00) ON SEGMENT NO. 1 - lt			
	X	Y	Z	RES
189.000	0.000	0.000	72.973	72.973
190.000	0.000	0.000	73.359	73.359
191.000	0.000	0.000	73.660	73.660
192.000	0.000	0.000	73.433	73.433
193.000	0.000	0.000	72.639	72.639
194.000	0.000	0.000	71.284	71.284
195.000	0.000	0.000	69.379	69.379
196.000	0.000	0.000	66.939	66.939
197.000	0.000	0.000	63.983	63.983
198.000	0.000	0.000	60.532	60.532
199.000	0.000	0.000	56.615	56.615
200.000	0.000	0.000	52.260	52.260
201.000	0.000	0.000	47.502	47.502
202.000	0.000	0.000	42.378	42.378
203.000	0.000	0.000	36.927	36.927
204.000	0.000	0.000	31.190	31.190
205.000	0.000	0.000	25.213	25.213
206.000	0.000	0.000	19.042	19.042
207.000	0.000	0.000	12.723	12.723
208.000	0.000	0.000	6.306	6.306
209.000	0.000	0.000	-0.159	0.159
210.000	0.000	0.000	-6.623	6.623

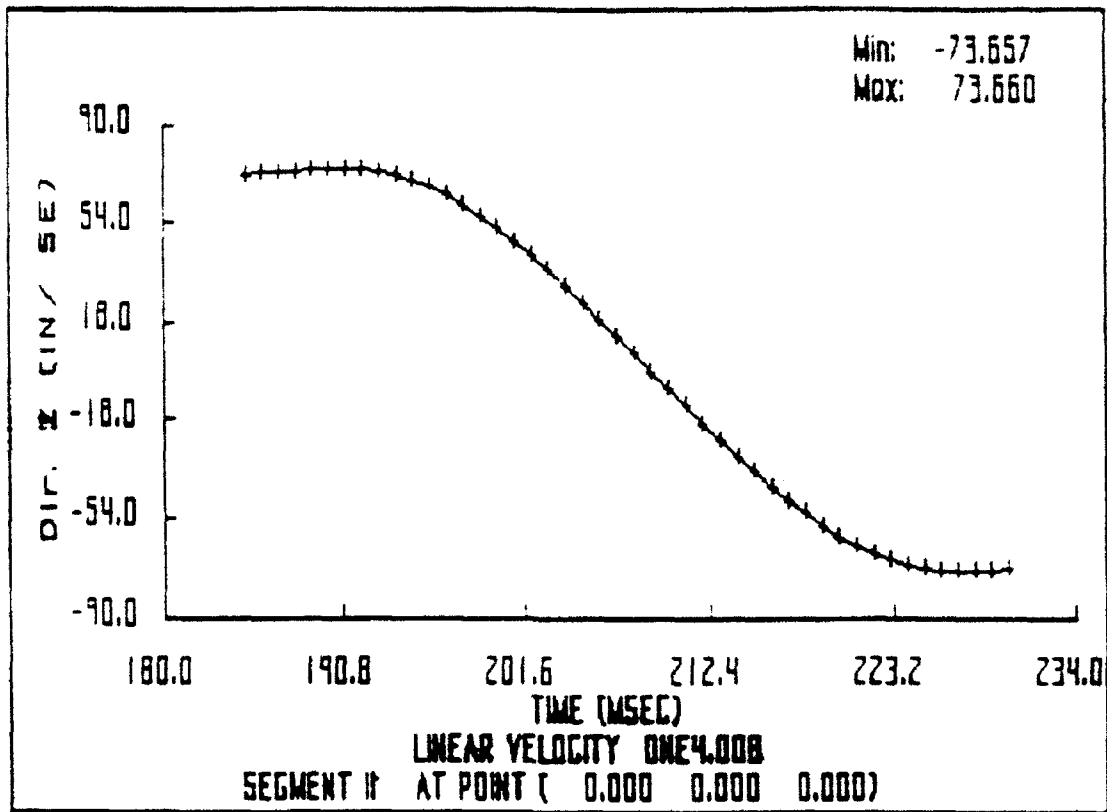


FIGURE 1

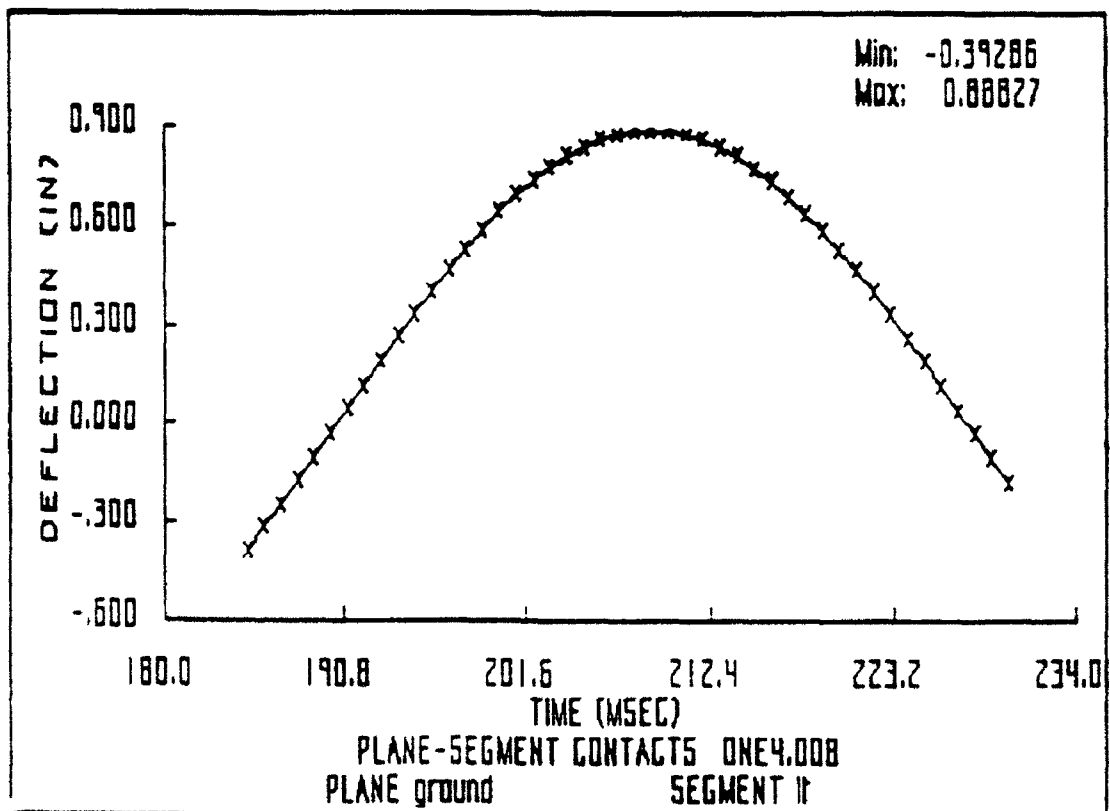


FIGURE 2

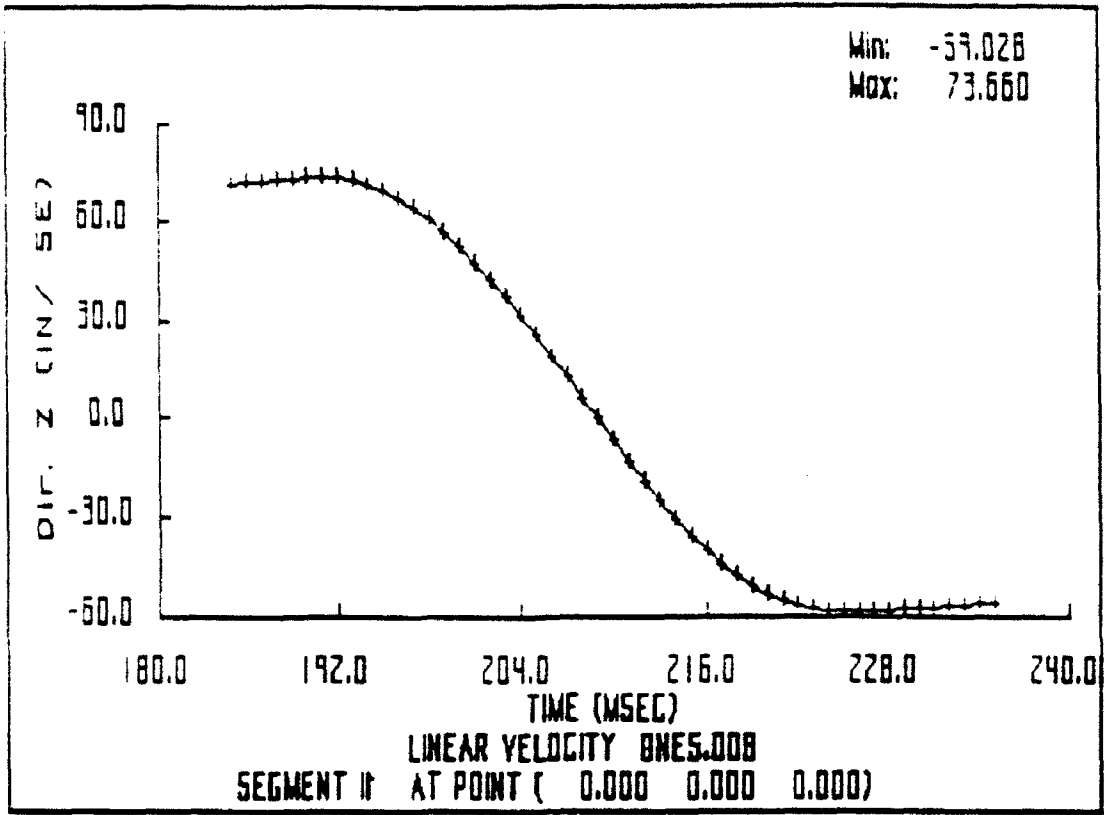


FIGURE 3

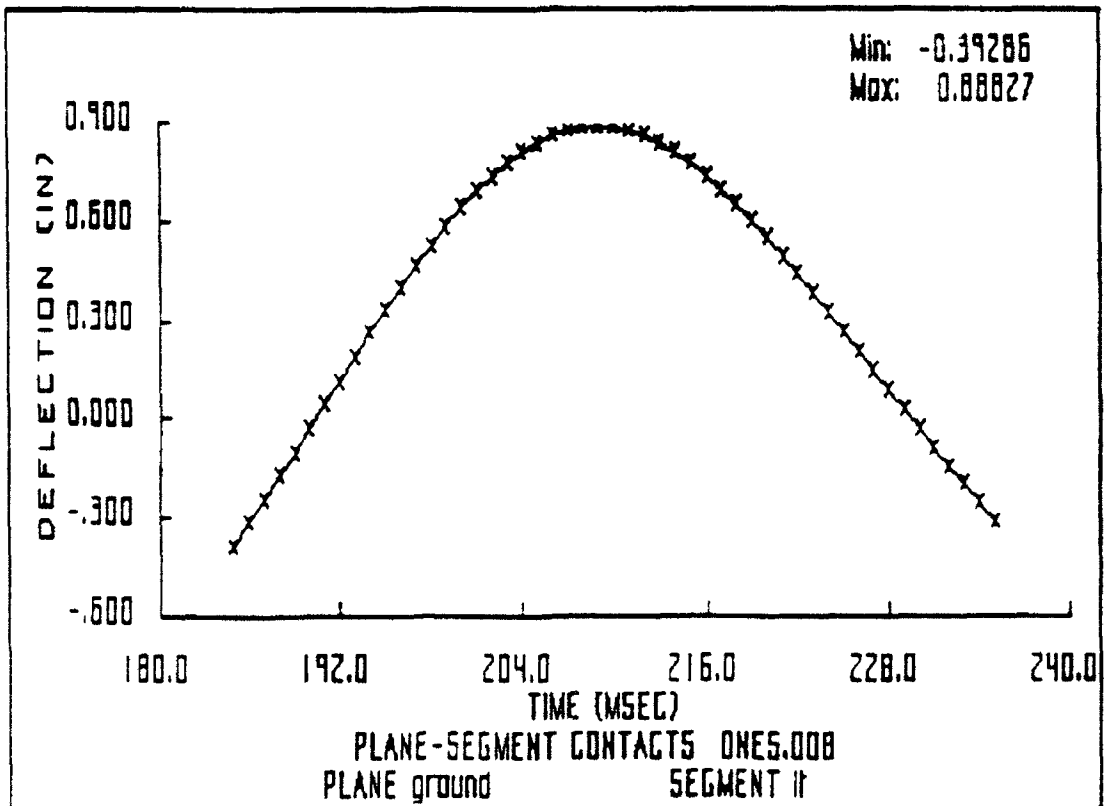


FIGURE 4

REFERENCES

1. Corben, H.C. and Stehle, P.; Classical Mechanics, John Wiley & Sons, Inc., New York, 1950.
2. Fleck, John and Butler, Frank; Report No. Z5-5881-V-1, "Validation of the Crash Victim Simulator, Vol 1," Engineering Manual by Calspan Corporation Advanced Technology Center, Buffalo, New York, December 1981.
3. Halliday, David and Resnick, Robert; Fundamentals of Physics, John Wiley and Sons, Inc., New York, 1974.
4. Obergefell, Gardner, Kaleps, and Fleck; "Articulated Total Body Model Enhancement, Volume 2: User's Guide," January 1988.
5. Present, R.D.; Kinetic Theory of Gases, McGraw Hill Book Co., New York, 1958.
6. Uspensky, J.V.; Theory of Equations, McGraw Hill Book Co., New York, 1948.

A STUDY OF THE EFFECTS OF LOW UPDATE RATE
ON VISUAL DISPLAYS

Jer-Sen Chen
Assistant Professor
Department of Computer Science and Engineering

Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

A STUDY OF THE EFFECTS OF LOW UPDATE RATE
ON VISUAL DISPLAYS

Jer-Sen Chen

Assistant Professor

Department of Computer Science and Engineering

Wright State University

Abstract

This report contains a study on the effect of low and incompatible update rate to a 60 Hz refresh visual display and propose solutions to remove the undesired artifacts. Since the contrast sensitivity of human eyes drops significantly above 60 Hz, visual display devices usually operate at 60 Hz refresh rate. If the refresh rate is low, for instance at 30 Hz, the undesired flickering artifact will be detected by human visual perception. Due to limitation of technology, even though we are exploring very high definition video signals it is just impossible to achieve the 60 Hz requirement. Simple remedies to the problem of lower update rate, such as to display the same frame multiple times or to blank out the non-updated frames will introduce undesired artifacts of multiple images or flickering. This report proposes approaches to alleviate the problem by either an algorithmic scheme to interpolate the missing frames or by a field sequential scheme on binocular displays. A spatially interpolative scheme, though computationally efficient, is not a solution to generating the missing frames. A temporally interpolative scheme should be employed to generate the correct frames though it is computationally extensive and sometimes not feasible for real time applications. The binocular field sequential approach provides an alternative to solving the problem of low update rate, though the temporal integration behavior of binocular perception is yet to be further studied.

A STUDY OF THE EFFECTS OF LOW UPDATE RATE ON VISUAL DISPLAYS

Jer-Sen Chen

1 Introduction

Temporal anti-aliasing has long been studied in the research of computer-generated animation [Max85, Korein83, Potmesil83]. Aliasing is an artifact when the sampling rate is not fast enough to handle the rapid change of signals, either spatially or temporally sampled. The Nyquist sampling rate to avoid aliasing is at least twice the maximum of the signal frequency. In the temporal domain, the aliasing is usually observed in the presence of fast moving objects. A typical anti-aliasing technique involves certain degree of smoothing or averaging, either pre-sampling or post-sampling. And a typical approach for temporal anti-aliasing is using *motion blur* algorithm. When a camera films a fast moving object, motion blur is generated through the integration process during the time period when the shutter is open. The techniques of generating motion blur are often employed to computer generated animation to produce more natural and realistic video.

Since the human eyes are less sensitive at high frequency, in particular beyond 60 Hz [Carterett75], the usual display refreshing rate is set to 60 Hz to avoid artifacts such as flickering. If the refreshing rate is low, for instance at 30 Hz or lower, the undesired flickering artifact will be detected by human eyes. However, due to the limitation of technology, the update rate of the video signals is usually lower than the 60 Hz refresh rate. For instance in a flight simulation computer program, the complexity of the simulated scene may be so complicated that a new frame can only be generated every 1/10 second, namely the update rate is only 10 Hz. When the 10 Hz updated images are displayed on a 60 Hz refresh rate visual display, different artifacts are detected by the human eyes depending on the techniques employed to accommodate the 60 Hz refreshing speed.

Even though very high resolution video displays, such as 1000 scan lines and 1000 pixels per scan line, are becoming more and more popular these days, it is usually impossible to achieve the 60 Hz update rate especially certain processing is involved. With a display resolution of one million pixels

(1000x1000), 60 Hz processing means 60 million pixels of processing per second. Even with the fast advances of computing technology, it is beyond the reach of the capability of most computers with the possible exception of massively parallel computers that have thousands of processors.

So the question is: suppose we have an update rate below 60 Hz, say 30 Hz or 15 Hz, how are we going to present the signal to a 60 Hz refreshing rate display? Motion blur techniques, unfortunately, can not be directly applied to this *inverse* problem. That is, what we have are the images derived from lower update rate output instead of the *object models* and their motion in the real 3-dimensional world. The problem is therefore more image processing rather than computer graphics oriented. A study on the effects of lower update rates on flight simulation visual displays was conducted by Kellogg and Wagner [Kellogg88]. They tried to determine the maximum acceptable display velocities when the display starts to degrade significantly. We shall conduct some experiments and study under the assumption of 30 Hz update rate and 60 Hz refresh rate unless otherwise specified.

We first illustrate the artifacts exhibited by some naive display oriented approaches to present 30 Hz updated signals on 60 Hz refresh rate displays. We then proceed the study of algorithmic approaches which aims at regenerating the missing frames due to low update rate through some interpolation schemes. And finally, a field sequential display oriented approach is studied through binocular displays to investigate the temporal intergation characteristics of human eyes. In summary, this report consists of two different approaches to investigate the removal of artifacts introduced by incompatible update rate and discuss their feasibility. The two approaches are:

- **monocular interpolated scheme:** In this scheme we conduct our experiment in monocular environment. The visual stimuli are presented on a usual monocular display such as the monitor of a computer workstation. We use computer graphics and image processing techniques to investigate the feasibility and visual effect of regenerating the missing frames by interpolative schemes. We shall see simple spatial interpolation, though computationally simple and efficient, is not all a solution to this problem. On the other hand, the correct solution should be obtained by temporal interpolation. Temporal interpolation, however, involves the calculation of optical flow which is very computationally extensive. Feasibility and robustness of temporal interpolation scheme needs further studied and justification.
- **binocular field sequential scheme:** The visual stimuli are presented on binocular displays

that feed separate input to two human eye channels. For instance, we use two computer workstations to generate separate stimuli to left and right eyes through a synchronization mechanism at both electronic scanning and frame generation level. The experiment in this scheme focuses on presenting different stimuli in terms of time delay to left and right eyes. Through the study of the behavior of temporal integration, we aim at reducing the computational cost of exact temporal interpolation while alleviating the artifacts introduced by the low update rate.

All the following experiments are conducted using IBM PC for simple stimuli, and Silicon Graphics IRIS-4D VGX machines for high performance shaded or colored graphics. The binocular field sequential study is conducted through the Helmet-Mounted Displays (HMD).

2 A Simple Experiment

We first conduct a simple experiment illustrating the artifact introduced by simplistic approaches dealing with the cases of lower update rate than display refresh rate. A simple tested graphics is shown in figure 1. the equally spaced short line segments in the middle of the screen mark the translational amount of every $1/60$ of a second. That is, the moving line segment shown above the markers smoothly moves from left of the screen to the right and it moves from one marker to the next at each refresh of the screen. When a 60 Hz update rate is employed, no artifact is detected when the eyes track the moving line segment. We then proceed the experiment to investigate the artifacts introduced by lower update rate.

2.1 Display the Same Frame Twice

At the update rate of 30 Hz, a new frame arrives only at every other refreshing period. For instance, if the first frame arrives at time $0/60$, the second frame arrives at time $1/30$ (or $2/60$ of a second). There is no frame arriving at the time tick of $1/60$ second. So a naive way of displaying a 30 Hz update rate signal would be simply showing the same frame twice. That is, we show the same frame of time $0/60$ at time $1/60$ second again, and show the same frame of time $2/60$ at time $3/60$ again, and so on.

This scheme works for a static scene, that is when there is no motion from either the viewer or the

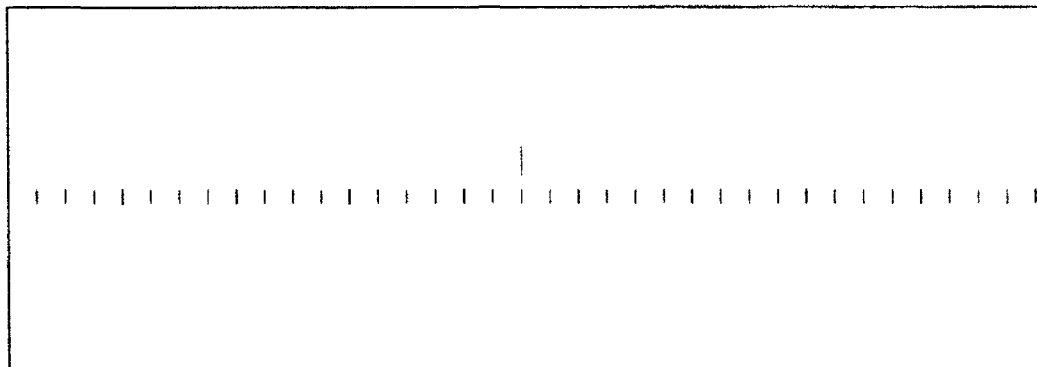


Figure 1: Original display

environment. Since the refreshing rate is 60 Hz, no artifact or flickering is perceived. For a dynamic scene, however, a very undesirable artifact arises. The most significant artifact is that, when the eyes are tracking the moving vertical line, TWO lines are perceived as shown in figure 2. We can also observe that the contrast reduced in both perceived lines. The reason for this artifact can be explained as follows:

When the eyes are tracking the moving line, they anticipate certain amount of motion from the position perceived at time $N/60$ second to time $(N+1)/60$ second. The amount of movement is of course the distance between two adjacent markers. Since we are display the same frame twice, the line in the second frame is not displayed at the position *it is supposed to be*. Instead the second frame displays the line at the same position of the previous frame which create a lag and therefore present a false line to the eyes. The spacing of the two lines is exactly one marker apart.

We then lower the update rate even further, say 20 Hz. For the refresh rate of 60 Hz, we have only one updated frame arrives every 3 refreshed frames. If we display the same frame three times, that is, the moving line segment is displayed at the same location 3 times, we can detect the artifact of three moving line segments instead of one. The perceived contrast of three moving line segments are further reduced.

We also conduct similar experiment using color graphics. For the 30 Hz update rate, we first

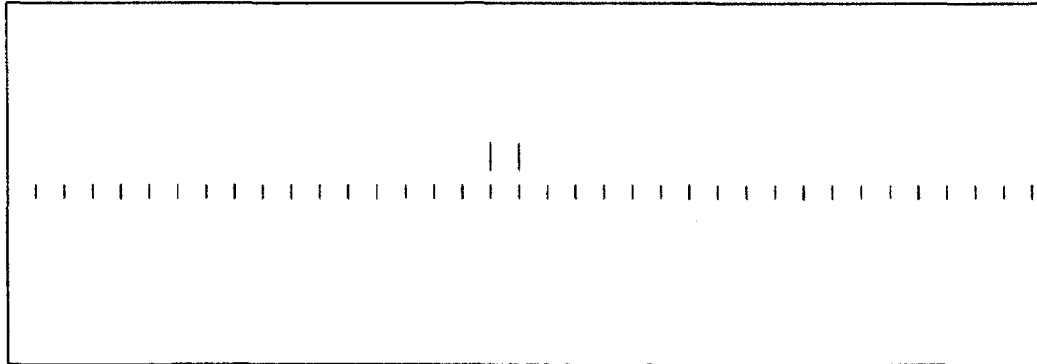


Figure 2: Perceived display when tracking the target

present the line segment in RED color and at next frame refresh we draw the line segment at the same location, but with a different color, say GREEN. When the eyes are not tracking the moving line segment, it is perceived as a single YELLOW (red plus green) line segment, even though only for a very brief moment since it is not tracked. When the moving line segment is tracked, again we observe two moving line segments. In particular, one is RED and the other is GREEN, and the RED one leads the GREEN one by exactly one marker.

2.2 Display One Frame and Blank Out The Other

The other naive approach of displaying a low update rate video is to display the true frame once and display nothing (BLANK or average grayscale) for other frames. For instance, if the update rate is 30 Hz then for 60 Hz refresh rate we display the newly updated frame first and blank out the second frame. If the update rate is 15 Hz, we display every one out of four frames and black out the other three frames.

This approach, although removing the *multiple image* artifact of the previous approach, has several drawbacks. Since the display is now virtually the update rate (15 Hz or 30 Hz) instead of 60 Hz refresh rate, the eyes can easily pick up the low frequency flickering artifact. Furthermore, since the true scene is only displayed every one out of several frames, it also reduces the contrast of the images.

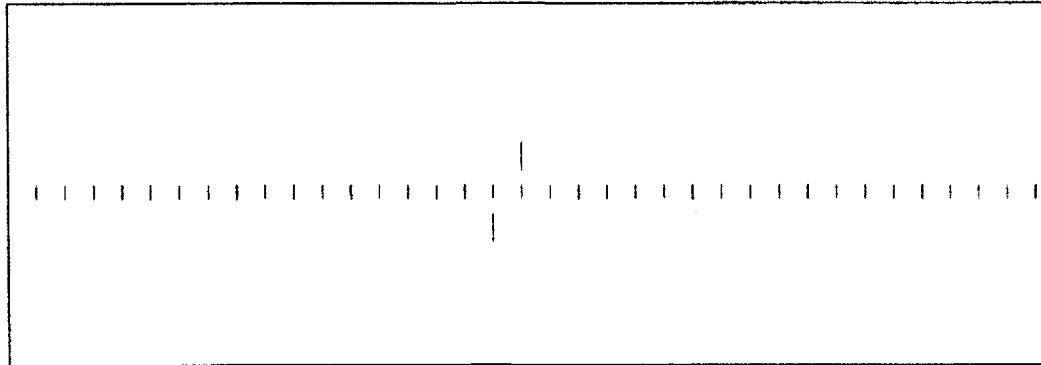


Figure 3: Two flickering moving line segments

We also conduct the following experiment of displaying two moving line segments, one above the markers, the other below. Each one is presented in the same fashion above that we display every other frame and blank out the other. However, the upper line segment is displayed first followed by a blank one; the lower line segment is just the opposite, that is, the first frame is blank, followed by the frame with the line segment. The observed result is shown in figure 3. The upper line segment leads the lower one by exactly one marker. Of course, both moving line segments are flickering and their contrast is reduced since they are both at 30 Hz.

3 Algorithms for Interpolating Missing Frames

The display approaches of bridging the incompatibility between lower update rate and the 60 Hz refresh rate, either displaying the same frame multiple times or displaying one frame and blanking out the rest, introduce various artifacts. An algorithmic approach aims at generating the missing frames for 60 Hz through some interpolative schemes. The algorithmic interpolation schemes, however, involve a great deal of computation to generate the missing frames and are definitely subject to the feasibility consideration in real time applications.

3.1 Spatial Interpolation

To display a 30 Hz updated signals on a 60 Hz refresh rate display, an interpolated artificial frame need to regenerated for every two consecutive real updated frames. Let $F_1(x, y)$ and $F_2(x, y)$ be two consecutive updated frames arriving at time t_1 and t_2 respectively, the time difference $t_2 - t_1$ is $1/30$ of a second for a 30 Hz update rate. In order to accommodate the 60 Hz refreshing rate display, a frame, say $F_{1.5}(x, y)$, at $t_{1.5} = \frac{t_1+t_2}{2}$ is generated by interpolation. In the case of repeating the same frame twice, $F_{1.5}(x, y)$ is set to be exactly $F_1(x, y)$, while in the case of blanking every other frame, $F_{1.5}(x, y)$ is set to be background color.

One naive way on generating $F_{1.5}(x, y)$ is *spatial interpolation*, i.e., every pixel of $F_{1.5}$ is generated by the average value of F_1 and F_2 at the same pixel location. The spatial interpolation scheme, though simple and computational efficient, does not produce true interpolated results. Furthermore, the artifact remains. Consider a simple square moving from left to right. Figures 4(a) shows an one-dimensional tested intensity plot extracted from a thermal image shown in figure 5(a). The total number of points (pixels) is 364, the marker at the bottom of the plot is equally spaced at 20 pixels. We assume that the object in the scene is moving horizontally from left to right at the speed of 20 pixels per $1/30$ second. The true frame at next $1/60$ second is the signal in figure 4(a) translated to the right by half a marker. Suppose that we have an update rate of 30 Hz, so the next updated frame is the signal in figure 4(a) translated to the right by a full marker. A spatial interpolation scheme generates the result of figure 4(b), where the intensity at the left and right boundaries being averaged as well as the span of the signal being extended by one marker.

One noticeable distortion can be observed at both leading and trailing edges of the nonzero intensity region. For the trailing edge on the left, a spatial interpolation scheme averages the pixel values therefore creating a band, which is exactly a marker apart (20 pixels), of averaging values between the signal and the background. A discontinuity from the background to the signal is interpolated as two discontinuities of half of the original differential. This effect is also noticeable in the two-dimensional image of figure 5(b) which shows the result of two-dimensional spatial interpolation. The artifacts are definitely not removed on the visual displays using spatial interpolation scheme since the interpolated image already contains multiple objects.

3.2 Temporal Interpolation

A true interpolated frame should be generated through temporal instead of spatial interpolation. For instance, the trailing edge of the signal in the first frame F_1 is at pixel location 78 and for F_2 is at 98, then the truly interpolated frame $F_{1.5}$ should contain only one single trailing edge at pixel location 88. Temporal interpolation, involving an inverse problem of the computation of optical flow, is computational much more extensive and furthermore the exact solution may sometimes be impossible to derive [Horn86]. The optical flow is represented by a vector image $V(x, y)$ in which the vector at each pixel location identifies the motion of the pixel therefore the resulting location of the pixel at next frame can be computed.

Temporal interpolation starts with the computation of optical flow of the imagery to recover the motion field of the moving objects as well as background. Let $u(x, y)$ and $v(x, y)$ be the x and y components of the optical flow vector at that point, namely $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$. Also let $E(x, y, t)$ be the intensity of the image frame at spatial location (x, y) and temporal instant t , then the *optical flow constraint equation* [Horn86] is

$$E_x u + E_y v + E_t = 0$$

where $E_x = \frac{\partial E}{\partial x}$, $E_y = \frac{\partial E}{\partial y}$, and $E_t = \frac{\partial E}{\partial t}$. It expresses a constraint on the components u and v of the optical flow.

In the case of one-dimensional signal, the optical flow constraint equation reduces to

$$E_x u + E_t = 0$$

The translational velocity u can therefore be solved from

$$u = -\frac{E_t}{E_x}$$

Since the noisy nature of the real signal, a smoothing stage is employed before the computation of the optical flow. There is also limitation in recovering the motion field from optical flow. Inside a homogeneous region of a moving object, no apparent spatial or temporal intensity gradient will be detected, therefore no optical flow is detected. Fortunately, for the purpose of visual display, these points can be trivially set to the same intensity value of the pixel at same location. Once we recover

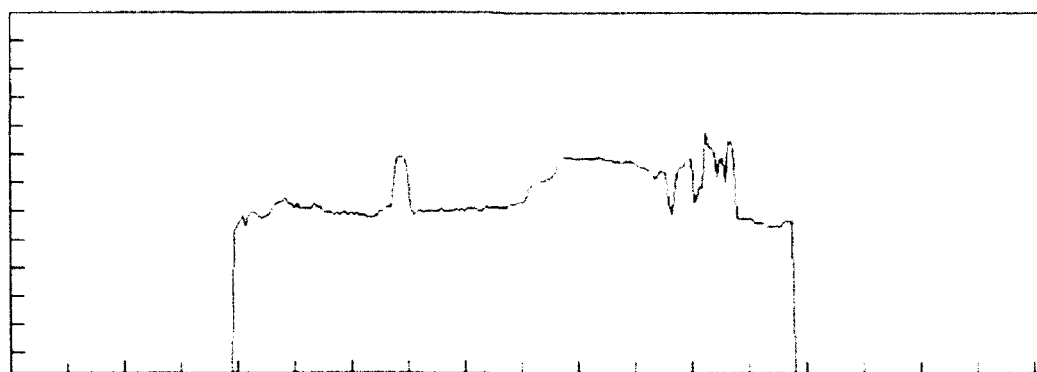
the translational velocity, in terms of pixels per updated frame, of the moving object, we then proceed the temporal interpolation to generate the missing frame. For instance, if the translation of pixel i is 1 ($u = 1$) we then set the grayscale value of pixel i of $F_{1.5}$ to be the average of the pixel $i - 2$ of F_1 and the pixel $i + 2$ of F_2 . Ideally, of course, the pixel $i - 2$ of F_1 and the pixel $i + 2$ of F_2 should have the same grayscale value, but in practice there is always certain amount of error. The error can be reduced through relaxation scheme which adds another degree of computational complexity [Horn86].

The temporally interpolated signal using the same example of figure 4(a) is shown in figure 4(c). We can observe that the temporally interpolated signal is much more resemble to the original signal compared to the spatially interpolated signal. The temporal interpolation algorithm is simply applying an extensive spatial smoothing to the original signals first then followed by a motion estimation, using $u = -\frac{E_t}{E_r}$. We have to point out this simple temporal interpolative scheme works for our example because the motion is only translational in known direction, namely from left to right. In the real two dimensional imagery, even with only translational motion, the recovery of optical flow and motion field is much more tedious. The translational amount as well as the direction has to be recovered. The spatial smoothing stage before the computation of optical flow can not be employed in a radially symmetric fashion, at least not to a large degree of smoothing.

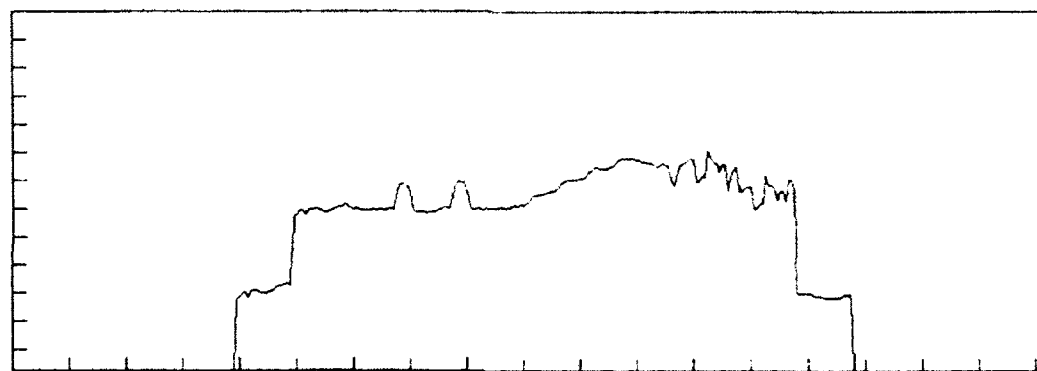
Figure 5 shows the comparison of the application different interpolation schemes to the two-dimensional image. The image is actually processed scanline-by-scanline independently. Figure 5(a) shows the original image, figure 5(b) shows the result of spatial interpolation, where figure 5(c) shows the result of temporal interpolation through a very simple scheme of estimating optical flow. We then playback the interpolated images and compare two interpolated schemes. The spatial interpolation scheme still exhibits the artifact of multiple images. Only the other hand, the artifact of multiple response is very much alleviated in the temporal interpolation scheme except at regions where the recovery of the motion field is incorrect.

4 Field Sequential Approach

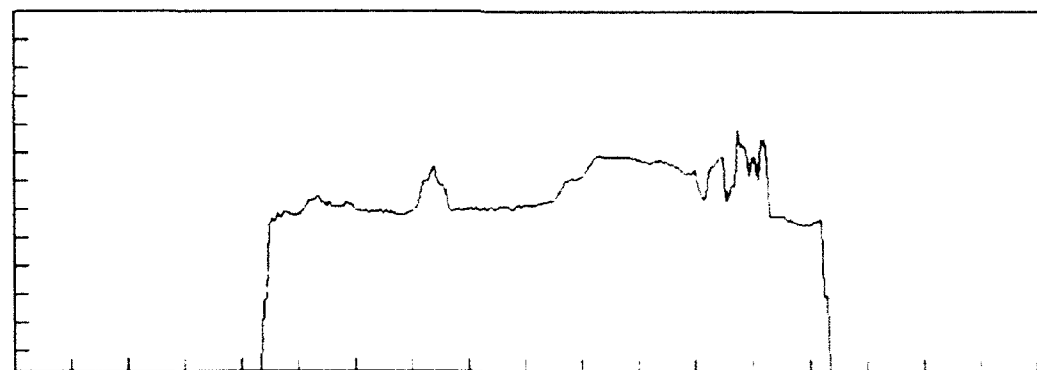
Integration of the visual input from both eyes has long been studied. Spatial integration of human eyes is evident from the random-dot stereogram analysis [Julesz71]. A pair of random dot images which are a stereopair can be fused to yield a square popping out in front of a reference plane.



(a) original signal



(b) spatially interpolated signal



(c) temporally interpolated signal

Figure 4. Comparison of Different Interpolated intensity profiles



(a) Original IR image



(b) Spatial Interpolation



(c) Temporal Interpolation

Figure 5: Comparison of Different Interpolation Schemes

Temporal factor in visual perception can be traced back to the basic law underlying a lot of temporal perceptual phenomena, namely the Bloch's law in 1885 [Carterett75]:

$$I \times t = k, \quad t < \tau$$

where I is the threshold intensity, t is the duration of the test, k is a constants, and τ is the *critical duration*. Visual motion perception which studies physiology of human eyes perception of moving stimuli can also be traced by to the early 60's. Temporal integration, the study of integrating behavior from the two eye channels, however, is much less explored due to the limitation in technology.

4.1 Repeated Monocular Experiments

A field sequential approach can be employed for binocular displays. The stimuli to the left and right eyes can be arranged in a fashion to remove undesired artifacts. The study of field sequential approach was conducted using two IRIS-4D VGX stations to generate the stimuli. The synchronization of the video signal of two channels is done by utilizing the Broadcast Video Options (BVO) provided by Silicon Graphics Inc., at the electronic scanning level package. The synchronization at the frame level is done by a client-server mechanism through logical links among running processes.

We first conduct the same experiments in the previous section of monocular cases to check our binocular configuration. The synchronied 60 Hz updated video was first fed into both eye channels. The result is the smoothed moving object on the helmet-mounted binocular displays. We then proceed to conduct the experiment of 30 Hz update rate video. Again identical and synchronized video was sent to both eyes. When we use the strategy of displaying the same frame twice, we perceive the DOUBLE moving objects as in the case of monocular case. When we display one frame and blank out the other, we see the flickering and contrast-reduced video. Suppose F_1, F_2, F_3, \dots represent the

30 Hz updated frames, the signals are presented as follows for the double display:

time	1/60	2/60	3/60	4/60	5/60	6/60
left	F_1	F_1	F_2	F_2	F_3	F_3
right	F_1	F_1	F_2	F_2	F_3	F_3

where the synchronized display of one frame and blanking out the other is as follows:

time	1/60	2/60	3/60	4/60	5/60	6/60
left	F_1	Blank	F_2	Blank	F_3	Blank
right	F_1	Blank	F_2	Blank	F_3	Blank

4.2 Field Sequential Experiment

We then proceed to conduct a simple test on field sequential approach. That is, to the both eyes, the 30 Hz video is presented with blanking of every other frame and the signals send to both eyes have a time difference of 1/60 of a second. Again, suppose F_1, F_2, F_3, \dots represent the 30 Hz updated frames, the signals are presented as follows:

time	1/60	2/60	3/60	4/60	5/60	6/60
left	Blank	F_1	Blank	F_2	Blank	F_3
right	F_1	Blank	F_2	Blank	F_3	Blank

The result of the binocular field sequential stimuli is that no more doubling artifact is observed. Flickering, as a result of 30 Hz update rate, still remains. The above configuration, however, seems to exhibit a less degree of flickering artifact than the synchronized 30 Hz video with blanking of every other frame.

5 Conclusion and Future Directions

A preliminary study of the artifact introduced by incompatible update rate to the display refresh rate as well as intuitively proposed solution is presented in this report. The artifact, can be in general categorized into:

- **multiple images:** when displaying the same frame multiple times, and

- **flickering**: when displaying the updated frame only once and blanking the missing frames.

The proposed solution in this report as well as for future research direction can be either algorithmic or display oriented. The algorithmic approach is to generate the missing frames due to the low update rate. The display oriented approach, without generating the missing frames, tries to bridge the gap between update rate and refresh rate by directly employing the characteristics of human eye responses to various display configurations.

We have demonstrated that simple repetition of the same frame several times to accommodate the refresh rate will generate the artifact of multiple responses of the single eye-tracked moving object. Another display oriented approach that simply to blank out the missing frames will produce the flickering artifact since the eyes are virtually looking at the update rate video, say 30 Hz, instead of refresh rate video at 60 Hz. Field sequential approach is also display oriented since we are not generating any missing frames. By utilizing binocular display devices, like most of the Head-Mounted Displays, we can present different visual stimulus to two human eye channels. In particular, we have conducted an experiment by simply presenting the 30 Hz updated video to both eyes by introducing an 1/60 second delay from one eye to the other. The result is that the multiple-object artifact is removed but the 30 Hz flickering is still present, although seems to be at a lesser degree.

Algorithm approaches are basically schemes to regenerate the missing frames, for instance, every other frame for 30 Hz update rate and 60 Hz refresh rate. Simple spatial interpolation by averaging two consecutive updated frames to generate the intermediate frame, though computationally simple and efficient, is by no means a solution to remove the artifacts. It actually introduces more artifacts since it creates more inconsistency such as one line becoming two lines.

On the other hand, temporal interpolation provides a better algorithmic approach for removing the artifacts. The computation though, involving at least the optical flow up to the true recovery of the motion field, is very extensive and expensive. We have shown some simple temporal interpolation involving only translation in known direction. The result is satisfactory though still far from robustness. A combination of algorithmic and display oriented approaches provides another promising direction in find solution to removing the artifacts.

Acknowledgement

This report describes part of my work of 1992 Air Force Office of Scientific Research Summer Faculty Research Program conducted at the **Armstrong Laboratory** at Wright Patterson AFB, Dayton, Ohio. I would like to express my sincere gratitude to Dr. Brian H Tsou, my laboratory focal point at AL/CFHV, for his advice and assistance throughout the program. I would also like to thank all the people at AL/CFHV who make this summer research a very fruitful and pleasant experience.

References

- [Carterett75] "Handbook of Perception, Volume V, Seeing," edited by E.C. Carterett and M.P. Friedman, Academic Press, 1975.
- [Horn86] . B.K.P. Horn, "Robot Vision," MIT Press, Cambridge, Massachusetts, 1986.
- [Julesz71] . B. Julesz, "Foundation of Cyclopean Perception," University of Chicago Press, 1971.
- [Kellogg88] G.V. Kellogg and C.A. Wagner, "Effects of Update and Refresh Rates on Flight Simulation Visual Displays," NASA-TM-100415, 1988
- [Korein83] J. Korein and N. Balder, "Temporal Anti-Aliasing in Computer Generated Animation," Computer Graphics, Vol. 17, No. 3, July 1983, pp.377-388.
- [Max85] N.L. Max and D.M. Lerner, "A Two-and-a-Half-D Motion Blur Algorithm," Computer Graphics, Vol. 19, No. 3, July 1985, pp.85-93.
- [Potmesil83] M. Potmesil and I. Chakravarty, "Modeling Motion Blur in Computer-Generated Images," Computer Graphics, Vol. 17, No. 3, July 1983, pp.389-399.

**Development of a Research Paradigm to Study
Collaboration in Multidisciplinary Design Teams**

**Maryalice Citera
Assistant Professor
Department of Psychology
Wright State University**

**Jonathan A. Selvaraj
Graduate Research Assistant
Department of Psychology
Wright State University**

**Final Report for:
AFOSR Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.**

September 1992

**Development of a Research Paradigm to Study
Collaboration in Multidisciplinary Design Teams**

**Maryalice Citra
Assistant Professor
Department of Psychology
Wright State University**

**Jonathan A. Selvaraj
Graduate Research Assistant
Department of Psychology
Wright State University**

Abstract

The purpose of this research project was to develop a research paradigm to investigate the collaborative process of design in multidisciplinary teams. Two phases of task development are described. The initial phase involved identifying a design problem that could be used to create the experimental task. The problem selected was the design of a navigation system for an automobile. The second phase involved collecting knowledge about the problem to make the task as realistic and interesting as possible. Knowledge was collected from design experts using a concept mapping technique. The results highlighted many tradeoffs and design issues that could be integrated into an experimental task. Future steps necessary for producing the design paradigm are described.

Development of a Research Paradigm to Study
Collaboration in Multidisciplinary Design Teams

Maryalice Citera and Jonathan A. Selvaraj

Recently a great deal of enthusiasm has been generated for the multidisciplinary team approach to design (e.g., Sobieski, 1990). This approach is in direct contrast to the conventional approach to design. The conventional approach typically involves designers from different disciplines working independently and sequentially on the same project. Sequential designs can be costly in terms of time and money. The inefficiencies result from viewing the problem from a single, narrow perspective. For example, user requirements and needs may not be fully considered early in the design process if human factors experts become involved only after the design is complete. Their role then becomes one of design evaluation. Unfortunately, at this point, there may be a reluctance to make the recommended modifications or changes. From the user's standpoint, the outcome may be a less than adequate design. Similarly, finished designs are often given to manufacturing engineers to produce. If manufacturing issues were not considered earlier in the design process, the design may be impractical or difficult to mass produce. In addition, when components of a design are designed independently, the parts may not be integrated easily. Thus, the conventional approach to design has several drawbacks and may result in poor designs or costly redesigns.

A multidisciplinary teamwork approach allows for the integration of inputs from various disciplines. Teamwork allows the team to partition the design problem so that each member has a manageable task. The multidisciplinary aspect allows design teams to draw on the variety of expertise that individuals bring to the design process (e.g., human factors, industrial design, engineering, marketing). In multidisciplinary teams, tasks can be distributed so that each member handles the components which best suits his/her expertise. Also, team members can share information across disciplines and hopefully inform each other on the possible limitations and weaknesses in various options. In other words, designs are scrutinized from multiple viewpoints prior to fixing or constraining the design. Also, the multidisciplinary team approach permits members to concurrently work on a design. This allows the team to get the final product "out the door" quicker because tasks can be done simultaneously.

Unfortunately, bringing together multidisciplinary design teams has disadvantages.

Teams do not effectively share distributed knowledge. According to Stasser (1992), groups often focus on commonly shared information and neglect to discuss unique or unshared information. Sharing may be difficult because experts from different disciplines often speak their own languages and are biased by tacit assumptions from their unique perspectives (Boff, 1987). Design experts from a particular discipline may not be aware of the assumptions they make and of the design constraints that they intuitively accept. This makes it difficult for design team members to effectively communicate and convey their expertise to other team members. Even when communicated, domain specific constraints may actually conflict across disciplines without clear resolution. Team members must strategically make tradeoffs to arrive at a mutually acceptable design. Thus, collaboration is analogous to the process of negotiation.

The fact that a design often does not progress in a strictly rational and analytical way may compound the communication problem. Due to time constraints, designers often do not generate and carefully evaluate all possible design alternatives. Instead of starting from scratch, designers tend to use previous cases as starting points for their next designs (Gero, 1990; Klein, 1987). This has been referred to as case-based design. Case-based designs are efficient because features of a previous design case can be incorporated in the new design project. But, unnecessary features of the previous design might also get incorporated. Because these features are spuriously associated with essential ones, they needlessly restrict the design options that are considered. Designers may have trouble differentiating and articulating which features are essential and which are superfluous. The result may be misanalogies. Misanalogies occur when previous learning is applied to a situation where it may be inappropriate or conflict with other aspects of the design. Misanalogies may make it even more difficult for the designers to communicate and share their expertise with other members of the team.

Furthermore, cross-disciplinary team members are often physically separated by distance and their tasks are performed asynchronously. Under these conditions, both communication and coordination will be more difficult. Although a variety of communication media can be used to bridge the distance (e.g., e-mail, fax, telephone), these media can create additional problems such as psychological distance (Wellens, 1986) and complicate the team's communication problems. On the other hand, electronic means of communicating may offer innovative improvements that aid collaboration in design (e.g., electronic group memory,

electronic knowledge elicitation).

Our goal is to design a research paradigm which can be used to study a variety of alternative ways computer technologies (e.g., electronic media, groupware, concept mapping tools) can be used to facilitate collaboration in design teams. The remainder of this paper describes the preliminary steps involved in developing a research paradigm to study collaborative work in design teams.

Methods

Designing a research paradigm to study collaborative design issues involved finding a design problem that could be developed or adapted into an experimental task. The process of finding and developing an experimental task was conducted in two phases. The first was an extensive literature review to find a design problem. The second phase involved acquiring knowledge about the design problem that was selected. The following sections outline the process that was undertaken.

Finding a Design Problem

A literature review was conducted to find a design problem that could be adapted for use in our research paradigm. Engineering, design, and human factors journals were extensively surveyed. In addition, computerized searches of PSYCHLit and ERIC databases were conducted.

To be considered as a task candidate, design problems were judged in relation to five criteria that were established prior to our literature review. The first criterion was that the task had to be based on a "real world" design problem. This "real world" perspective was driven by the desire to produce an experimental context that was relevant, meaningful, and engaging for design teams to participate in. Also, this would enable research to be undertaken in both the field and laboratory settings. In addition, a "real world" based task would increase the external validity of the task. Thus, the generalizability of future research would be enhanced.

The second criterion was that the problem had to be multidisciplinary in nature. This meant that the design problem had to involve considerations from various disciplines. Therefore, the problem had to be sufficiently complex so that knowledge from more than one discipline would be necessary to complete the design. The disciplines that might be involved in such a design would be marketing, engineering (i.e., electrical, computer, human factors, industrial, etc.), psychology (i.e., experimental, social, etc).

Third, the problem had to incorporate or include consideration of human factors issues

(e.g., display design issues, human-computer interaction issues, mental workload issues, ergonomics, anthropometry, etc). This criterion was included because of relevance to the overall purpose of our research team. We were interested in how human factors information is incorporated into collaborative design projects. We hoped this knowledge would ultimately contribute to the development of groupware products that facilitate the retrieval and incorporation of human factors information and expertise into design.

Fourth, the design problem needed to have the potential to be completed in an individual and team context. This would provide flexibility in designing future experiments (e.g., multiple levels of analysis).

Fifth, design professionals, as well as, university students must be able to complete the problem. This meant that the problem should be simple enough for a student to complete. The problem should not overwhelm the student. At the same time it should be complex enough for a real design professional to find it interesting and engaging. Ideally, the problem would have characteristics that were familiar to both university students and professional designers.

Forty-one design problems were selected for consideration using these criteria. The majority of these design problems had been used as instructional projects in college level design and engineering courses. Some were actual on-going "real world" design projects. A table of these problems is available from the authors upon request. From these problems we selected the task of designing a navigation display for an automobile (Dingus, 1990).

This design problem satisfied all of our pre-established criteria. First, designing a navigation system for an automobile was a real design project. Several systems have been designed and some are currently being tested. Second, this design problem required the input of many disciplines (e.g., engineering, psychology, hardware, software, etc.). Third, human factors issues were taken into consideration in previous design endeavors of automobile navigation systems. Fourth, we felt that the design could be completed by both individuals and teams of designers. Fifth, the design problem focused on a system to aid individuals in driving and navigating an automobile. Because these behaviors are familiar to both designers and students, we felt that it had the potential to be an interesting and engaging experimental task.

Acquiring Knowledge about the Design of an Automobile Navigation System

Once a design problem had been chosen, the second phase of task development

began. The second phase involved acquiring knowledge or information about the issues involved in designing a navigation system for an automobile. In particular, we focused on potential tradeoffs involved in designing within a multidisciplinary team context. Concept mapping interviews were conducted with design experts to acquire knowledge and information concerning our design problem.

Concept Mapping. The interview technique employed to elicit knowledge was Concept Mapping (McNeese, Zaff, Peio, Snyder, Duncan, & McFarren, 1990; Novak & Gowin, 1984). Concept mapping is an interactive interview methodology used to elicit information from an expert concerning a particular subject area. During the exchange, an interviewer converts the elicited information into a heterarchical graphical network of concepts nodes (i.e., a concept map). Concept nodes are usually conveyed as actions, events, or objects. The concept nodes are connected by various relational links. The relational links are conveyed as prepositions or verbs (McNeese, et al., 1990). A concept-relation-concept unit is called a concept "triplet" and is a useful unit of analysis.

The result is mapped onto a white board as the expert speaks. The expert is encouraged to interact with the map by suggesting changes, additions, and elaborations. The map becomes an external memory aid indicating to the expert what has been discussed and stimulating the recall of additional information (McNeese, et al., 1990).

Using concept mapping to acquire knowledge has several advantages. First, graphically representing information allows the expert and interviewer to see what has been communicated. Any misunderstandings or misinterpretations that the interviewer may have can be corrected "on-line" by the expert. Second, the external representation of knowledge also helps the expert organize the information that has been communicated. Third, concept mapping allows the interviewer to locate and identify central issues and concepts nodes within a given subject area. This is facilitated by the arrangement of concepts nodes in relation to other concept nodes (McNeese, et al., 1990). These advantages are dependent on how actively the expert guides the construction of their concept map. Usually, as the level of interaction between the map and the expert increases, the quality and quantity of information increases (McNeese et al., 1990).

Subjects. Thirteen design professionals were interviewed using the concept mapping methodology. The design professionals included human factors psychologists and engineers, a software specialist, a display hardware specialist, an electrical engineer, and an industrial

engineer.

Concept Mapping Session Apparatus and Procedure. Sessions were conducted in a conference room that had been modified for the purpose of concept mapping interviews. The room contained seven whiteboards fastened to the walls. The mapper captured the expert's information by drawing the concept map on the whiteboards. A Macintosh computer was located on a conference table in the center of the room. The Macintosh was used to input a computerized version of the expert's map for later analysis. The application program used for this purpose was called the Concept Interpreter. (See Snyder, McNeese, & Zaff, 1991.) In addition, the sessions were audio taped for later review and analysis. Each concept mapping session took approximately one to three hours.

The concept mapping sessions involved the expert, concept mapper, and an additional panel of interviewers. The concept mapper converted the knowledge elicited from the expert into a concept map and had the most interaction with the expert. The additional panel of interviewers served to elicit additional information from the expert as needed. The panel used probe questions to clarify or expand existing concepts.

At the onset of the concept mapping session, experts were seated at the conference table. The concept mapper then gave a brief introduction to concept mapping and its advantages over more traditional forms of interview techniques. The concept mapper, also, described the purpose of the interview. This included conveying information concerning our goal to construct a research paradigm and experimental task. At this point the concept mapping interview started.

A total of 13 concept maps concerning collaboration and design issues involved in designing a navigation system for an automobile were produced. These maps were analyzed using the Concept Interpreter. The Concept Interpreter was developed at Armstrong Laboratories to facilitate the recording and analysis of concept maps (Snyder, et al., 1991). The application sorted the triplets according to categories pre-defined by the researcher. As stated previously, a triplet is a concept-relation-concept unit. The sorting process produced a matrix that indicated whether the category was represented in each expert's map. The Concept Interpreter also produced the list of triplets for each individual's map that reflected a particular category.

The categories were defined using keywords. The keywords for each category were identified in several ways. First, keywords were identified by the researchers based on their

experiences in the concept mapping sessions. Second, keywords were identified based on a review of the audio tapes and the experts' concept maps. Third, the matrix contained an undefined category that was iteratively reviewed by the researchers to identify relevant triplets. After the computer generated the matrix and corresponding lists of triplets, the researchers assessed the relevance of triplets within a particular category. Triplets that were deemed irrelevant were eliminated from further consideration in that category. The following section summarizes the results and conclusions of our analyses.

Results

The concept interpreter matrix is presented in Figure 1. The matrix indicates the number of concepts and links in each of the maps. The number of concepts discussed by each expert ranged from 77 to 191. The number of triplets (i.e., links) ranged from 79 to 207. The matrix, also, indicates whether or not reference to a particular category appeared in each of the 13 concept maps. A dot indicates that an element of that category was mentioned once. A circle and a dot indicate that two or more elements from that category were discussed. As can be seen from Figure 1, there was a good deal of consistency across the experts. The experts covered many of the same issues. The next section summarizes the key issues identified by the experts.

To facilitate the discussion, the categories from the concept interpreter were broken down into two groupings: process type issues and tradeoff issues. Process type issues focused on identifying team members (i.e., who is on the team), the timeline of a "typical" design, and the objective or mission of the design. Tradeoff issues focused primarily on the types of constraints and tradeoffs different team members might face in designing a navigation system for an automobile. Both types of information will be useful for building a task. The process information helps to specify what types of disciplines (e.g., software engineers, marketing analyst, program manager) should be represented in the research paradigm, the steps involved in designing a navigation system for an automobile, and what would be a reasonable task for designers to do in a limited amount of time. The tradeoff issues help identify the structure of the task and provide "real world" information for creating the task materials. Each grouping will be discussed in more detail below.

Process Issues

Team Composition. Six disciplines were mentioned consistently across the maps. Over half of the experts who discussed team composition indicated these disciplines. The

Team	Timeline	Objective	Constraints	General Hardware Issues	Software Constraints	HF Constraints	Marketing	Users	Managing	Format of Display	Error Management	Context	Display Hardware	Navigation System	Computer System	Control Hardware	Car/System Interface	Maintainability	UNDEFINED
B. M. (97 nodes / 107 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
J. H. (165 nodes / 176 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
S. T. (118 nodes / 128 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
D. M. (130 nodes / 147 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
C. M. (147 nodes / 165 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
B. P. (77 nodes / 75 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
G. B. (130 nodes / 146 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
C. A. (132 nodes / 176 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
J. G. (150 nodes / 160 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
T. C. (191 nodes / 207 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
T. H. (97 nodes / 115 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
J. E. (120 nodes / 138 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
E. H. (109 nodes / 121 links)	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙

Concept Map

Figure 1: Concept Interpreter Matrix

prototypical design team included: 1) a human factors psychologist/engineer, 2) a hardware/display engineer, 3) a software engineer/programmer, 4) a manager, 5) a marketing analyst/user representative, and 6) a system/safety specialist.

Timeline Phases. Four phases were discernible from the matrix results. The steps identified by the experts were: 1) marketing definition phase, 2) the assembly of a team, 3) developmental planning phase, 4) full scale development and prototyping phase, and 5) implementation phase. According to the experts, the marketing definition phase would involve the identification of a niche or marketing need for the product. This would include conducting a small feasibility study to assess whether the product could be produced profitably. In the second phase, team members would be identified and assigned responsibilities. Typically the team members would represent the disciplines discussed above. During the developmental planning phase, the team would stipulate the specifications and performance requirements for the product. The full scale development phase would be an iterative process that involves prototyping, identifying problems, correcting problems, and refining the design. This would be the phase where most design tradeoffs and bottlenecks would become apparent. During this phase, beta versions of the product would be created, tested, and evaluated. The final phase, implementation, would involve releasing the final product.

Tradeoff Issues

This section discusses design tradeoffs identified by the experts. The tradeoffs have been summarized by discipline perspective and tradeoff issue. Table 1 presents the results for the hardware, software, and human factors perspectives. Table 2 presents the results for the management, marketing, and system/safety perspectives. Because of space limitations, we discuss two representative tradeoffs to illustrate the type of information summarized in these tables.

Display Type. The experts identified several types of displays that can be used in an automobile. These included liquid crystal displays (LCD), cathode ray displays (CRT), electro-luminescent displays, plasma displays, and projection displays. The consensus among the experts was that the most likely choices were the LCD and the CRT displays. Thus, we focus our discussion on the tradeoffs the experts identified between these two types of displays.

From a hardware perspective, the advantage of an LCD is that it requires less power and puts less of a strain on the car's electrical system than a CRT display. The drawbacks of

Table 1: Design Tradeoffs for Hardware, Software, and Human Factors

Discipline Perspective			
Tradeoff	Hardware	Software	Human Factors
Display Hardware			
Type	<ul style="list-style-type: none"> -size -power supply -cooling requirements -raster vs. stroke reliability -flat panel vs. curved screen -color 	<ul style="list-style-type: none"> -resolution--# of pixels -refresh/update rate -display drawtime 	<ul style="list-style-type: none"> -resolution -brightness -size affects readability -viewing angle -display flicker
LCD vs. CRT	<ul style="list-style-type: none"> -LCD available in limited sizes -LCD has limited colors -LCD less power requirements -LCD requires driver hardware -LCD requires backlighting 	<ul style="list-style-type: none"> -LCD has less resolution -LCD has slower update rate 	<ul style="list-style-type: none"> -LCD has limited brightness -LCD has limited viewing angle
Color vs. Monochrome	<ul style="list-style-type: none"> -LCD has limited color -CRT is available in more colors 	<ul style="list-style-type: none"> -color affects update rate -color requires more lines of code -color can be used to highlight 	<ul style="list-style-type: none"> -color helps identify objects -color aids readability -color reduces workload
Controls	<ul style="list-style-type: none"> -touch screen -switches -buttons -voice commands -bezel switch -trackball 	<ul style="list-style-type: none"> -touch screen increases complexity of program 	<ul style="list-style-type: none"> -reach/anthropometry issues -touch--screen may be too hot -usability -access to display -interface shouldn't distract
Format			
Track-up vs. North-up Orientation		<ul style="list-style-type: none"> -track-up requires rotation -raster/bitmap cannot easily be rotated -vector drawings are more easily rotated 	<ul style="list-style-type: none"> -track-up preferred -vehicle may be difficult to find -north-up requires mental rotation -mental rotation is cognitively demanding

Discipline Perspective			
Tradeoff	Hardware	Software	Human Factors
Ego-centric vs. Earth-centric		<ul style="list-style-type: none"> -easier to animate smaller objects -car is easier to animate -map is harder to scroll -must track vehicle location 	
Menus		<ul style="list-style-type: none"> -requires more complex programming 	<ul style="list-style-type: none"> -should minimize user programming -should be straightforward -shouldn't distract driver -workload shouldn't be increased
Variable Zoom	<ul style="list-style-type: none"> -data storage device -CD-ROM -Hard drive -CPU processing speed & time -Zoom requires more detailed information 	<ul style="list-style-type: none"> -database issues -zoom requires information filtering -auto decluttering needed requires choice of lowest type of road to display (i.e., alley, two lane, etc.) -details in latitude and longitude -raster/bitmaps vs. vector drawings -bitmaps have limited zoom -bitmaps are not easy to filter -vector drawings can be zoomed -zoom affects map dynamics -zoom changes update speed requirements -zoom affects resolution -overlaps/overwrites may occur 	<ul style="list-style-type: none"> -zoom affects clutter -zoom needed in most congested or populated areas -zoom affects workload requirements
Visual vs. Auditory	<ul style="list-style-type: none"> -type of visual display -type of speech generator or voice synthesizer 	<ul style="list-style-type: none"> -requires programming voice synthesizer or speech generator 	<ul style="list-style-type: none"> -workload -distractions -auditory display can serve to remind driver of approaching turn -auditory display does not require looking away from road -visual display requires driver to look away from road--primary task

Discipline Perspective			
Tradeoff	Hardware	Software	Human Factors
Navigation System			
GPS vs. INS	<ul style="list-style-type: none"> -GPS requires antennae & receiver -INS requires compass & motion sensors -Computer must compare position information to map database 	<ul style="list-style-type: none"> -GPS packet information received from satellite at rate of 400 ms -Level of GPS available to public is accurate to 100 ft -Inaccuracy requires error checking -Off-course must allow user to correct 	<ul style="list-style-type: none"> -Errors add to workload
Context			
Weather	<ul style="list-style-type: none"> -Weather may cause corrosion & damage -Excessive temperatures may cause damage (Car temperature can rise to 120 degrees) -Weather may interfere with reception 	<ul style="list-style-type: none"> -how can program deal with interference 	<ul style="list-style-type: none"> -inclement weather increases workload
Safety	<ul style="list-style-type: none"> -shocks -mechanical -electrical 	<ul style="list-style-type: none"> -glitches may cause accidents -only programmable in park 	<ul style="list-style-type: none"> -reach -safety -overheating
Glare	<ul style="list-style-type: none"> -tilt of display -curve & type of display -viewing angle of display -may require glare screen 		<ul style="list-style-type: none"> -readability -brightness -tilt of display
Passengers	<ul style="list-style-type: none"> -may damage equipment- throw something at screen 		<ul style="list-style-type: none"> -increase workload—crying baby
Traffic		<ul style="list-style-type: none"> -integration of data into route selection algorithm 	<ul style="list-style-type: none"> -more traffic increases workload

Table 2: Design Tradeoffs for Management, Marketing, and System/Safety

Discipline Perspective			
Tradeoff	Management	Marketing/Users	System/Safety
Display Hardware			
Type	-Displays vary in cost	-What user is willing to spend determines cost limit	-fit in dash -tilt of display -cooling requirements -power supply -weight -depth of display -screen implosion or chemical leak -safety screen may be needed
LCD vs. CRT	-LCD is 3 times more expensive than CRT -LCD is more reliable -LCD has lower repair cost	-better reliability & maintainability means easier sell	-LCD has less cooling requirements -LCD doesn't respond well to extremely high temperatures -LCD is flat panel--has less depth -LCD weigh's less -CRT requires more space -CRT screen can implode -LCD chemicals can leak
Color vs. Monochrome	-Color is more expensive -Monochrome=\$700 -4 color=\$1500 -16 color=\$3000	-Color is aesthetically pleasing -Color preferred by customers	
Controls	-Touch screens cost more	-User friendly controls are more appealing	-location of controls (e.g., on steering column) -touch screen may be too hot
Format			
Track-up vs. North-up Orientation	-increasing complexity of program may increase costs and delay the schedule	-Track-up preferred -Raster/bitmaps are more realistic--may be preferred by customers	

Discipline Perspective			
Tradeoff	Management	Marketing/Users	System/Safety
Ego-centric vs. Earth-centric			
Menus	-Increasing complexity of programming may increase costs and delay schedule	-Customers prefer simpler interface to a more complex one	
Variable Zoom	-Required detail needed for zoom will be more costly -Filtering information requires more complex programming--may affect cost and schedule	-Storage device must not be too costly	-Storage device must fit in car
Visual vs. Auditory Display	-Auditory display adds to cost	-Realism & understandability of voice	-heat of car limits type of synthesizer that can be used -must fit in car
Navigation System			
GPS vs. INS	-GPS receivers may be expensive -INS devices involve costs	-GPS antenna is ugly -aesthetics	-inaccuracy may cause accidents -fit of equipment and cabling -additional weight
Context			
Weather			-weather can impact vehicle & equipment
Safety	-want to avoid lawsuits	-want to avoid accidents	
Glare	-Glare screen might increase cost		-Changing tilt of screen to avoid glare may conflict with dashboard sizing
Passengers			
Traffic	-Traffic updates increase costs	-traffic updates may appeal to consumers	

the LCDs include the need for backlighting and driver hardware. CRT displays may be more advantageous than LCDs because they are available in a larger variety of sizes and with a wider range of color options for screen display.

From a software perspective, CRTs may be preferred because they have better resolution and a quicker update rate than LCDs. These advantages may facilitate the programmer in creating clear and readable display images.

From a human factors perspective, LCDs may be difficult to read and adversely affect the user because of their limited brightness and viewing angle.

From management's perspective, the disadvantage of an LCD is its cost. LCDs are approximately three times more costly than CRTs. The advantages of an LCD is that it is more reliable and cheaper to repair than a CRT.

From a marketing perspective, CRTs may be easier to sell than LCDs because of readability and increased aesthetic appeal due to color availability.

From a system/safety point of view, an LCD may be preferable to a CRT because it may be easier to fit within the car's structure and layout. Since LCDs are flat panel displays, they are more compact and require less overall space. LCD also requires less cooling than CRTs. But, LCDs are sensitive to extremely high temperatures. Given the high temperatures in a locked car during the middle of the summer (about 120 degrees) this may be a problem. If damaged, CRTs can implode and be hazardous to the driver and passengers. LCDs, on the other hand, if damaged, will leak dangerous chemicals.

These constraints indicate that different perspectives may opt for different types of displays based on their constraints. The "best" choice for any particular perspective may not be the "best" choice for all, given a particular situation.

Orientation of Display Format. Two issues were raised in the orientation of the navigation display. The first issue is whether the display should be track-up or north-up. In a north-up display, the map is always presented with north at the top, south at the bottom, east to the right, and west to the left. If the driver's route runs east-west, the map would display the route running right to left. In a track-up display, the map is always presented with the route running bottom to top regardless of the compass direction. This means that even when the mission is to drive south, the driver's destination will appear at the top of the display. The second issue is whether the display should be ego-centric or earth-centric. In an ego-centric display, the car (i.e., the individual's present location) would remain fixed and the map would

scroll. In an earth-centric display, the map would remain steady and the car would move across the map.

From a hardware perspective, a track-up orientation, because it requires rotation of the entire map as the car makes a turn, may dictate a faster update rate than a north-up orientation. A faster update rate may also be more necessary for an ego-centric orientation than for an earth-centric orientation. More information will need to be refreshed if the map scrolls instead of merely animating the car.

From a software perspective, the way the data is coded in the map database may constrain the orientation choices. The data may be coded as either raster/bitmaps or as vector drawings. Raster/bitmaps will be more difficult to rotate than vector drawings. Orientation may also play a role in the complexity of the program and the difficulty of the programming task. Since smaller objects are easier to animate, earth-centric displays will be somewhat easier to program than ego-centric ones.

From a human factors perspective, preferences for north-up versus track-up vary depending on the task involved. In the flight context, north-up is preferred by navigators and track-up is preferred by pilots. North-up may increase the driver's workload because the driver must mentally rotate the map to figure out whether to make a left- or right-hand turn. In a track-up orientation, turns indicated on the display map directly to the driver's right and left-hand. In addition, the driver's workload may increase under certain combinations of orientations (such as earth-centric, track-up) because it may be difficult to keep track of the car's position.

From a management perspective, the amount of resources (time and money) required to program the different format orientations is a key consideration. From a marketing perspective, user preferences may be important. From a system/safety perspective, there are no obvious tradeoffs based on orientation of the display.

Individuals from different perspectives see the orientation of the display from different vantage points. For the team to arrive at an optimal solution, information from all perspectives will be necessary.

Discussion

The knowledge elicited during the concept mapping sessions proved to be fairly consistent across experts. The concept maps highlighted the key tradeoffs and design issues faced by collaborators designing an automobile navigation system. The information collected

from the design experts represents a useful database from which an enriched design task can be created.

Collecting and summarizing this information, however, was only the first step involved in developing the design paradigm. Further work needs to be done to construct the task. The next step is to create roles and task materials for each of these roles. A role would be created for each of the disciplines identified as key players on the team. The task materials would present information relating to the tradeoffs identified for each perspective. For example, the team member representing the hardware perspective might get technical information (e.g., information concerning the number of pixels per inch, refresh rate, screen phosphors) on different types of displays available. The team member representing the human factors perspective might get information concerning readability under different levels of screen brightness or using different colors. In addition, outcome measures of team performance and team processes would need to be developed. The measures will depend on both the task materials and the research questions to be addressed.

This paper described the preliminary steps involved in developing a research paradigm to study collaboration in design. The first step was the identification of a design problem. Based on five criteria the researchers selected the problem of designing a navigation system for an automobile. The second step was the collection of knowledge from design experts. Using concept mapping, many relevant design issues and tradeoffs were identified. Future steps in the development of this paradigm include creating experimental task materials and developing outcome measures to assess team performance.

Acknowledgements

The authors would like to thank Clifford Brown, Michael McNeese, and Brian Zaff for their generous assistance and guidance with this project. Many of the ideas presented in this paper were developed in collaboration with them. They, also, were instrumental in collecting and analyzing this data. We would also like to thank the design experts from ASC and the Armstrong Lab for participation in the concept mapping interviews. Furthermore, we would like to acknowledge Daniel Snyder for his help with the Concept Interpreter and Margaret Hahn and Michael Nancarrow for their aid in transferring the maps to the computer.

References

- Boff, K. R. (1987). The tower of Babel revisited: On cross-disciplinary chokepoints in system design. In W. B. Rouse and K. R. Boff (Eds.) System design: Behavioral perspectives on designers, tools, and organizations (pp. 83-96). New York: Elsevier Science Publishing.
- Dingus, T. A. (1990). Teaching human factors design skills: Science of engineering doesn't do the trick. Proceedings of the Human Factors Society, 34rd Annual Meeting (pp. 525-528). Santa Monica, CA: The Human Factors Society.
- Gero, J. S. (1990). Design prototypes: A knowledge representation schema for design. AI Magazine, 11(4), 28-36.
- Klein, G. A. (1987). Analytical versus recognitional approaches to design decision making. In W. B. Rouse and K. R. Boff (Eds.) System design: Behavioral perspectives on designers, tools, and organizations (pp. 175-186). New York: Elsevier Science Publishing.
- McNeese, M. D., Zaff, B. S., Peio, K. J., Snyder, D. E., Duncan, J. C., & McFarren, M. R. An Advanced Knowledge and Design Acquisition Methodology: Application for the Pilot's Associate (AAMRL-TR-90-060). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division, Air Force Systems Command.
- Novak, J. D. & Gowin, D. B. (1984). Learning How to Learn. New York: Cambridge University Press.
- Snyder, D. E., McNeese, M. D., & Zaff, B.S. (1991). Identifying design requirements using integrated analysis structures. Proceedings of 1991 National Aerospace and Electronics Conference (NAECON '91), 2, 786-792.
- Sobieski, J. (1990, December). Multidisciplinary design optimization. Aerospace America, p.65.
- Stasser, G. (1992). Pooling of unshared information during group discussions. In S. Worchel, W. Wood, and J. A. Simpson (Eds.) Group Processes and Productivity (pp. 48-67). Newbury Park: Sage.
- Wellens, A. R. (1988). Use of a psychological distancing model to assess differences in telecommunication media. In L. Parker & C. Olgen (Eds.), Teleconferencing and electronic media: Vol. V (pp. 347-361). Madison, WI: Center for Interactive Programs.

**AN APPROACH TO ON-LINE ASSESSMENT AND DIAGNOSIS OF STUDENT
TROUBLESHOOTING KNOWLEDGE**

**Nancy J. Cooke
Assistant Professor
Department of Psychology**

and

**Anna L. Rowe
Graduate Student
Department of Psychology**

**Rice University
P.O. Box 1892
Houston, TX 77251**

**Final Report for:
Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

AN APPROACH TO ON-LINE ASSESSMENT AND DIAGNOSIS OF STUDENT
TROUBLESHOOTING KNOWLEDGE

Nancy J. Cooke
Assistant Professor
and
Anna L. Rowe
Graduate Student
Department of Psychology
Rice University

Abstract

Intelligent tutors have the potential to enhance training in avionics troubleshooting by giving students more experience with specific problems. Part of their success will be associated with their ability to assess and diagnose the students' knowledge in order to direct pedagogical interventions. The goal of the research program described here is to develop a methodology for assessment and diagnosis of student knowledge of fault diagnosis in complex systems. Along with this broad goal, the methodology should: (1) target system knowledge, (2) provide rich representations of this knowledge useful for diagnosis, (3) be appropriate for real-world complex domains like avionics troubleshooting, and (4) enable assessment and diagnosis to be carried out on-line. In order to meet these requirements a general plan for mapping student actions onto system knowledge is proposed and research from one part of this plan is presented. Results from a Pathfinder analysis on action sequences indicate that action patterns can be meaningfully distinguished for high and low performers and that the patterns reveal specific targets for intervention. Short- and long-term contributions of this work are also discussed.

AN APPROACH TO ON-LINE ASSESSMENT AND DIAGNOSIS OF STUDENT TROUBLESHOOTING KNOWLEDGE

Nancy J. Cooke
and
Anna L. Rowe

INTRODUCTION

As tasks become more cognitively complex and demand more specialized skill, training issues are increasingly critical. The domain of avionics troubleshooting is a good example of such a task. The cognitive complexity of this task, combined with the personnel downsizing currently faced by the Air Force, make the role of training even more crucial. Personnel will be required to become skilled quickly, and their skill will be required to span a broad range of equipment. In addition, the automatization of many aspects of the troubleshooting task greatly reduces the amount of time spent manually troubleshooting faults. The difficulties associated with the resulting lack of troubleshooting experience are particularly apparent when the automatization fails and manual troubleshooting becomes essential. Training programs need to address these rare, yet critical, events.

How can training programs meet these requirements? One approach is through the use of computerized intelligent tutoring systems (ITSs). These systems enable individuals to spend time learning a skill in a one-on-one environment in which a computer takes on the role of a human tutor. One goal of ITSs is to incorporate individualized instruction based on a detailed assessment of student knowledge and diagnosis of cognitive strengths and weaknesses. Instructional intervention can then be directed at these strengths and weaknesses. The purpose of the work described in this paper is to develop a methodology for the assessment and diagnosis of student knowledge in the context of ITSs.

The problem of assessment and diagnosis for ITSs has been approached in a number of ways. One approach involves "debugging" a student's knowledge after inferring misconceptions or "mal-rules" from patterns of student errors (e.g., Burton, 1982; Stevens, Collins, & Goldin, 1979). Although this approach has intuitive appeal, there is some evidence that errors are not as systematic as would be implied by underlying misconceptions (Payne & Squibb, 1990). Anderson, Boyle, and Reiser (1985) take a different approach and model student actions in terms of a set of production rules. These rules are then compared to an ideal student model in order to determine student deficiencies. These approaches and other related ones attempt to model the student by mapping either errorful actions or all actions onto misconceptions or deficiencies in the student's knowledge. The approaches are similar in that this mapping is achieved rationally. That is, the ideal model or rules for scoring are constructed through an analysis of domain principles, rather than through an empirical investigation of expert or ideal student behavior. Interestingly, many of the domains studied in ITS research have involved rather abstract, academic subjects such as algebra, geometry, and computer programming. These topics tend to lend themselves to a rational analysis because they are well-specified, well-structured, and typically associated with an

organized and well-documented body of knowledge. Although many of the principles and techniques derived from such studies may generalize to other similar domains, it is not clear how such findings can be extended to more complex and concrete domains such as avionics troubleshooting in which knowledge acquisition is often a prerequisite for tutor development (Psozka, Massey, & Mutter, 1988).

Likewise, in most real-world domains the first question to be addressed in assessment and diagnosis is exactly what knowledge is necessary to perform the task? Hall, Gott, and Pokorny (1991) have developed, PARI, a procedure for analyzing the cognitive requirements of a task for this purpose. The procedure involves a series of structured interviews with subject matter experts (SMEs) in which a specific problem is dissected in terms of its precursors, actions, results, and interpretations (i.e., PARI). For instance, a PARI analysis of the avionics troubleshooting domain has indicated that there are several types of knowledge relevant for successful troubleshooting performance. These types include: (1) system (or how it works) knowledge, (2) strategic (or how-to-decide-what-to-do-and-when) knowledge, and (3) procedural (or how-to-do-it) knowledge (Gott, 1989).

The kind of information obtained from a cognitive task analysis can guide assessment and diagnosis tasks. For example, Pokorny and Gott (1992) have devised an assessment procedure to identify general deficits in the different knowledge types (i.e., system, procedural, and strategic) of airmen tasked with troubleshooting technical electronic equipment. In general, points are deducted from these three different knowledge categories depending on the errors that the student makes. Note that this is similar to the debugging approach, except that general deficiencies are identified, not specific misconceptions. Gitomer (1992) has developed a related procedure that involves mapping student actions onto these same types of deficits. In both of these cases, a cognitive task analysis that involved knowledge elicitation from SMEs was required to determine ideal student behavior. Optimally, assessment and diagnosis in this domain would consist of identifying the specific content of student knowledge of these different types and comparing it to SME knowledge.

The Problem: Eliciting System Knowledge

Evidence exists to suggest that system knowledge may be the most critical of the three types of knowledge in troubleshooting ill-defined problems in complex systems (Gitomer, 1984). Although much can be learned about procedural and strategic knowledge from observing the actions of a problem solver, it is much less clear how system knowledge is revealed. Furthermore, the definition of system knowledge or, what many refer to as a mental model, is not completely clear, or at the least, agreed upon (Rouse & Morris, 1986; Wilson & Rutherford, 1989). Despite the lack of a clear definition, research employing the mental model construct is fairly prolific, with different researchers using their own operationalizations of the construct. The different methods of examining mental models can be classified into four categories: 1) accuracy and time measures, 2) interviews, 3) process tracing/protocol analysis, and 4) structural analysis.

Accuracy and time measures are often taken of problem solving behavior and used to make inferences about mental models. This method is similar to the approach discussed above for

debugging student knowledge, both in terms of methodology and limitations. That is, time and errors do not always map neatly onto a specific mental model or misconception (Cooke & Breedin, 1992). As a consequence, most attempts to measure mental models in the literature have combined this basic measure with one or more of the other relatively richer measures such as interviews or think-aloud verbal reports. Interviews can be more or less structured, with the content and course of the interview being more or less predefined. Unstructured interviews do not follow a prespecified format, whereas structured interviews do and may focus on: (1) a specific system component--e.g., location, purpose, function (Gitomer, 1984), (2) diagrams--e.g., enumerate concepts, show physical and/or functional relations, designate related components (Gitomer, 1984; Hall, Gott & Pokorny, 1991), or (3) a specific example of system behavior (Stevens, Collins & Goldin, 1979). Process tracing/protocol analysis involves asking subjects to "think aloud" as they solve a problem. Subjects are asked to generally describe their thought processes and to state reasons for their actions. The protocol is subsequently analyzed (i.e. protocol analysis) either to generate hypotheses about mental models or to support or reject a proposed model. Although such verbal reports have been criticized for their reliability and accuracy (e.g., Nisbett & Wilson, 1977), others (Ericsson & Simon, 1984) have attempted to define the conditions under which verbal protocols are appropriate. The fourth type of technique, structural analysis, entails collecting pairwise proximity estimates for a set of system-relevant items. These estimates are then submitted to a descriptive multivariate statistical technique (e.g., multidimensional scaling, cluster analysis, or network clustering technique) which reduces the estimates to a simpler form. For example, Kellog and Breen (1990) used the Pathfinder network structural technique (Schvaneveldt, 1990) to derive and compare user's mental models with an idealized system model. One of the strengths of structural analysis is that it is able to convey quantitative, as well as qualitative information about mental models.

In summary, four very different types of measurement methods have been used in research on mental models. The different measurement approaches may each provide different sorts of information, making generalizations across studies difficult, if not impossible. In addition, the different approaches have not been evaluated in terms of their respective validity as measurement instruments. In general, each of the different methods is likely to have advantages and disadvantages (Cooke, 1992a), and no one method of measuring mental models has received universal acceptance. Therefore the selection of a single optimal method for on-line student assessment is an uncertain enterprise at best. In this paper a pragmatic view is taken in which valid methods are minimally assumed to elicit knowledge that is relevant to task performance.

Another difficulty associated with using most of these methods for on-line assessment of student system knowledge is that most involve the collection of "extra" data (e.g., verbal reports, similarity ratings) not typically collected in interactions with the tutor. Thus, the use of these methods would entail interruption of the tutoring process to collect data in a task that would most probably seem artificial to the student. The single exception to this limitation is the collection of time and accuracy measures. Unfortunately, time and accuracy data are impoverished compared to the much richer data obtained from verbal reports and structural analyses. These richer methods go

beyond the student's actions, facilitating the jump from actions to the cognitive underpinnings of those actions. Therefore, what is needed is not only a sound method for measuring system knowledge, but one that can derive rich representations of this knowledge from student actions derived on-line. This is the focus of our project. The goal is to be able to map student actions (both errorful and correct) collected on-line onto a rich representation of student system knowledge. This representation can then be used to assess and diagnose student system knowledge and identify targets for intervention. The domain selected for this project is avionics troubleshooting.

The Plan: Mapping Student Actions onto System Knowledge

Basically, the general problem identified above involves making detailed inferences about a student's system knowledge from that student's actions. One way to dissect this problem is to work backwards from the goal state (system knowledge), to the initial state (student actions). Interviews, process tracing, and structural analytic methods offer rich representations of system knowledge. However, it is necessary to know which of these methods provides the best measure of system knowledge in the domain of avionics troubleshooting (see Figure 1.1). Therefore, the first subgoal in solving the above problem involves identifying a valid method for eliciting and representing system knowledge required for avionics troubleshooting. Assuming that system knowledge is critical for performance, then a valid method of measuring this knowledge should reveal differences among subjects that correspond to performance differences.

Of course, these techniques require data collected off-line. Therefore, the next subgoal involves determining how to derive this type of data from on-line interactions with the tutor. Can we make use of the data already collected on-line to derive representations of system knowledge? In other words can we identify general relationships between student actions and patterns of system knowledge derived off-line, so that later predictions can be made about system knowledge based on student actions? It is generally assumed that actions are, at least partially, the result of knowledge and that certain patterns of actions reflect specific types of troubleshooting knowledge (Pokorny & Gott, 1992). Gott, Bennett, and Gillet (1986, p 43) label the assumption that "thinking is for the purpose of doing" the theory of technical competence. But how do we make sense of all of these actions? What is needed is a means of identifying meaningful patterns or summaries of student actions. A pattern of actions can be thought of as an intermediate representation of student troubleshooting knowledge (see Figure 1.2). Although patterns in student actions are likely to emerge, their meaningfulness is an empirical question. Specifically, do differences revealed in identified action patterns correspond to actual differences in other measures of student performance? Thus, the identification of action patterns and the evaluation of the meaningfulness of these patterns is a second subgoal.

Once meaningful patterns of actions (i.e., troubleshooting knowledge) have been identified, the next subgoal entails mapping these patterns onto patterns of system knowledge (see Figure 1.3). Can we identify patterns of actions that correspond to distinct representations of system knowledge? Of course this step requires the elicitation of both actions and system knowledge from the same subjects. Assuming that the previous subgoals have resulted in meaningful patterns of

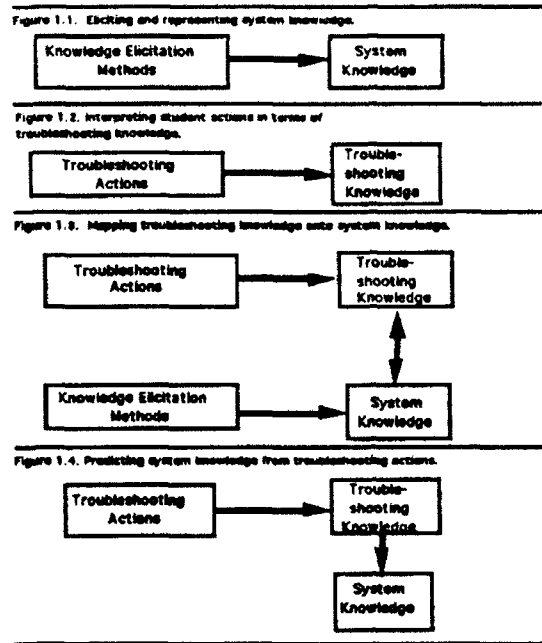


Figure 1. Steps involved in mapping student actions onto system knowledge.

actions and representations of system knowledge and assuming that system knowledge underlies actions (at least partially), then some correspondence should emerge. For instance, students who swap a card before checking the data flow to that card may do so for several reasons. This mapping procedure may indicate that students who demonstrate this action pattern tend not to understand the relationship between data flow and signal flow. Finally, if this correspondence does emerge, then it would be possible to make predictions about system knowledge from troubleshooting actions collected on-line, thereby eliminating the extra data collection step (see Figure 1.4). These predictions could be evaluated by comparing them to predictions made by SMEs or by implementing them in a tutor and evaluating the tutor.

The four subgoals represented in Figure 1 comprise the long-term plan associated with the development of a new approach for assessing and diagnosing student system knowledge. The subgoals represented in Figures 1.1 and 1.2 are prerequisites to the later subgoals, but even in isolation, these preliminary steps make important contributions to the general problem of student assessment and diagnosis. More specifically, identifying optimal methods for eliciting system knowledge is useful for stages of tutor development in which knowledge of this type needs to be elicited from domain experts. In addition, although less efficient than the long-term plan, these techniques could be used to assess student system knowledge off-line. The second subgoal would also contribute by identifying meaningful action patterns, useful in and of themselves in assessing and diagnosing other types of student knowledge (i.e., procedural or strategic knowledge). The remainder of this report focuses on progress made toward the long-term plan, specifically, the subgoal portrayed in Figure 1.2.

Research Progress: Interpreting Student Actions

The goal of this part of the project is to identify meaningful patterns in students'

troubleshooting actions. These patterns are referred to generally as "troubleshooting knowledge," because it is assumed that they are influenced by the three forms of knowledge central to troubleshooting, namely strategic, system, and procedural knowledge. If the resulting action patterns capture troubleshooting knowledge in a meaningful way, then they should minimally be able to differentiate high and low performers.

One way that action patterns can be derived is through the use of the Pathfinder network scaling procedure. The Pathfinder procedure is a descriptive statistical technique that represents pairwise proximities in a graphical form (Schvaneveldt, 1990; Schvaneveldt, Durso, & Dearholt, 1985; 1989). In the graph, concepts or entities are represented as nodes and relations between entities as links between nodes. Each link is associated with a weight that represents the strength of that particular relationship. These weights are based on proximity estimates which can be collected in a number of ways including pairwise relatedness ratings, co-occurrence of items in a sorting task, or event co-occurrence. Pathfinder networks can have directed links given asymmetrical proximity estimates and unconnected nodes if proximity estimates between an item and all other items exceed a maximum criterion set by the experimenter. It should also be noted that although the links represent semantic relations, the algorithm does not identify the specific relation associated with each link. The Pathfinder procedure determines whether or not to add a link between each pair of nodes. Basically, a link is added if the minimum distance between nodes based on all possible paths (i.e., chains of one or more links) is greater than or equal to the distance indicated by the proximity estimate for that pair. Two parameters, r and q , determine how network distance is calculated and affect the density of the network. Dearholt and Schvaneveldt (1990) provide a detailed discussion of Pathfinder.

Pathfinder has several advantages, including the fact that it is not constrained to hierarchical configurations like most cluster analysis routines, and it is able to represent asymmetrical relations (Dearholt & Schvaneveldt, 1990). In addition, results from several studies indicate that Pathfinder network representations are psychologically meaningful in that they are predictive of recall order and judgment time (Cooke, 1992b; Cooke, Durso, & Schvaneveldt, 1986). Pathfinder networks have, in fact, been used to reliably distinguish skilled and unskilled performers in domains such as air-combat flight maneuvers (Schvaneveldt, Durso, Goldsmith, et al., 1985), computer programming (Cooke & Schvaneveldt, 1988), and interface design (Kellog & Breen, 1990). They have also been used to assess student classroom performance (Goldsmith & Johnson, 1990). In this study the similarity between student and instructor networks was highly correlated ($r = .74$) with final class grade.

The Pathfinder procedure has typically been used to represent knowledge in the form of conceptual or declarative relationships (e.g., Cooke & Schvaneveldt, 1988; Schvaneveldt, Durso, Goldsmith, et al., 1985). However, it has also been used in one case to represent action sequences (McDonald & Schvaneveldt, 1988). In this study McDonald and Schvaneveldt collected co-occurrence frequencies of UNIX commands issued by users who interacted with the system. They used Pathfinder to summarize these data in terms of a network of the most frequently occurring action paths. Thus, because of Pathfinder's ability to represent action sequences and deal with the

asymmetrical and nonhierarchical relations typically found in actions, it was selected as a vehicle for interpreting actions in the present study.

Such a representation of actions would be desirable for several reasons beyond the overall goal of mapping actions onto system knowledge. First, on-line assessment in tutors could be achieved by deriving an individual's network from actions executed during problem solving and comparing this network to an "expert" network in terms of structural similarity. Second, the qualitative nature of the network representation allows a more detailed diagnosis of student troubleshooting knowledge. The Pathfinder network analysis could highlight specific actions and action sequences that are not "expert-like," allowing them to be targeted for remediation. Likewise, positive aspects of performance (expert-like actions) could be identified and targeted for positive feedback to the student. Thus, one additional benefit of this methodology is that it is capable of providing both quantitative assessment information at a global level and qualitative information at a more detailed level. Finally, because of the bottom-up nature of this approach, the Pathfinder representations may incidentally reveal specific patterns of actions that distinguish high and low performers, that would not have been recognized or verbalized by the SMEs.

METHOD

Actions taken by subjects completing troubleshooting tests described by Nichols, Pokorny, Jones, Gott and Alley (1989) were used to develop Pathfinder networks. In the Nichols et al. study the effects of an ITS called SHERLOCK were examined by comparing the performance of technicians who received both on the job training (OJT) and SHERLOCK training (experimental group) to the performance of technicians who received only OJT (control group).

Subjects.

The subjects were 37 manual avionics shop technicians stationed at one of two AF bases, Langley AFB or Eglin AFB. Supervisors had identified the subjects as being at a beginning or intermediate skill level (3 or 5) and available for the study duration (1 mo.). Five subjects were later dropped from the study: two subjects were transferred, and three subjects were identified as being more skilled than previously determined, leaving a sample of 32 technicians. The subjects were first matched on the basis of a verbal troubleshooting score and a number of other scores (e.g., mechanical and electrical tests). Then members of each matched pair were randomly assigned to either the experimental or control group. The 30 subjects who completed a specific set of three verbal troubleshooting problems were used in the present analyses.

Individual subjects were classified as either high or low performers on each problem based on the score they received from the scoring worksheet (Pokorny & Gott, 1992), the current assessment method in this domain. This score is derived by subtracting a predetermined number of points for each error that the student makes in troubleshooting. For the pretest problem, high performers were defined as those subjects who received a score of 85 or greater, whereas low performers were defined as those subjects who received a score of 35 or less. These cutoffs were arrived at by identification of natural breaks in the frequency distribution of scores. Four of six high performers and three of eight low performers were in the experimental group. Subjects were later reclassified as high and low performers based on their performance on the posttest problem.

Specifically, subjects were classified as high performers if they received a score of 85 or greater, and subjects who received a score of 55 or lower were classified as low performers. Interestingly, all of the high performers and only one of the low performers were in the experimental group.

Materials and Procedure.

A brief description of the methodology used by Nichols et al. (1989) follows. All subjects participated in a training period in which they received either OJT or OJT and SHERLOCK. The pre- and posttest measures referred to below were administered before and after this training period, respectively. Various measures of aptitude, experience, subjective opinions of the tutor, and troubleshooting performance were collected in the study, however, only problems from the verbal troubleshooting data were analyzed for the present study.

The verbal troubleshooting test is an individually administered structured problem solving test. The test begins with the examiner describing a fault that has occurred. The subject then attempts to isolate the fault and repair the equipment through a series of recursive action-result steps. In each step the subject specifies an action he/she would take and the reason for taking that particular action. The examiner responds by informing the subject of the action's effect on the equipment, and requests the subject's inference concerning equipment operation based on that effect. The cycle continues until the problem is solved, the one hour time limit expires, or the subject gives up. Thus, although subjects are not working on the actual equipment, they have to make use of all of the technical data that they would require if they were troubleshooting real equipment.

Six pretest and four posttest verbal troubleshooting problems were administered by Nichols et al. Only the data from three problems were used in the present analyses, specifically pretest 1, pretest 2, and posttest 1. The complete analysis described below was conducted on data from pretest 2 and posttest 1 because these problems were comparable in terms of type and difficulty. The pretest 1 problem was primarily analyzed to determine the optimal coding scheme.

RESULTS AND DISCUSSION

A coding scheme for students' actions was developed using the data from the pretest 1 problem. This scheme was then applied to and modified slightly for the remaining two problems, referred to herein as pretest and posttest. The purpose of the scheme was to be able to classify discrete actions into meaningful action units that could be represented as nodes in a Pathfinder network. The main categories of actions for both problems included equipment checks, data flow tests, signal flow tests, and swaps. The most abstract action unit was used unless the same action would, in some cases, result in a pass and in others, a fail. In this case, the lower, more specific level of abstraction was used. Using this decision rule, for each problem an action unit was associated with one and only one troubleshooting outcome. The resulting coding schemes consisted of 63 action units/nodes categories for the pretest and 62 action units/nodes for the posttest problem.

Transition probabilities for all pairs of actions (in both directions) were calculated for individual subjects by dividing the frequency with which specific action transitions (e.g., swap UUT followed by check DMM fuse) occurred by the frequency with which the first item in the

sequence occurred. For example, if swap UUT occurred twice and was followed by check DMM fuse on one of those occasions then the transition probability would be 0.5. Note that these are first-order transitions only. Higher order transitions (i.e., the probability of swap UUT followed by check DMM fuse either immediately or with one or more actions intervening) were considered, but not used because the immediate transitions were considered to be the most meaningful. Transition probabilities were also calculated across groups of subjects using frequencies summed across all subjects in the high or low performer groups.

Four matrices of transition probabilities (high and low performers, pre- and posttest) were submitted to the Pathfinder network scaling technique (Schvaneveldt, 1990). Figures 2 and 3 illustrate the pretest problem network representations resulting from the high and low performers' probabilities, respectively. Figures 4 and 5 illustrate the posttest problem network representations resulting from the high and low performers' probabilities respectively. Details of these networks will be discussed below in the section on diagnosis.

Assessment

One of the major questions to be asked of this approach is whether Pathfinder networks of actions can distinguish high and low performers for the purposes of assessment. In this study the subjects' score for each problem derived using the scoring worksheet is assumed to be the "true score" indication of their performance on that problem. Therefore, to answer the above question one can look at the correlation between an assessment measure derived from Pathfinder networks and the score derived from the scoring worksheet procedure. To assess students using Pathfinder, for each problem an ideal or expert network can be compared to the network representation of each nonexpert individual. The C measure (Goldsmith & Davenport, 1990) provides a quantitative index of network similarity that can be used for this purpose. This measure is based on proportion of shared nodes and links in two networks and ranges from 0 (low similarity) to 1 (high similarity). For the pre- and posttest problems, the networks based on the aggregate actions of the six highest performers were used as ideals for those problems. The remaining nonexperts were evaluated in terms of these standards. Note that use of the six highest performers as the ideal greatly restricts the range of the data for the remaining nonexperts on which the correlations were based. This procedure was necessary because there were only incomplete data available for the SMEs, the obvious choice for the ideal. Thus, it should be kept in mind that the correlations reported here may be underestimated due to this constraint.

The correlations between troubleshooting scores and this network similarity measure for the 24 nonexperts in each problem are presented in Table 1. In addition, two other assessment measures that were related to the network similarity measure were calculated and included in the analysis to aid in distinguishing relevant from irrelevant aspects of the Pathfinder-based measure. One of these measures was derived from a correlation of action frequencies (i.e., the frequency with which each action unit occurred) associated with an individual's protocol and action frequencies associated with the aggregate high performer protocol. Thus, this measure should be high to the extent that the nonexpert performed the same actions as the high performers the same number of times. It should overlap with the Pathfinder network similarity measure in that they

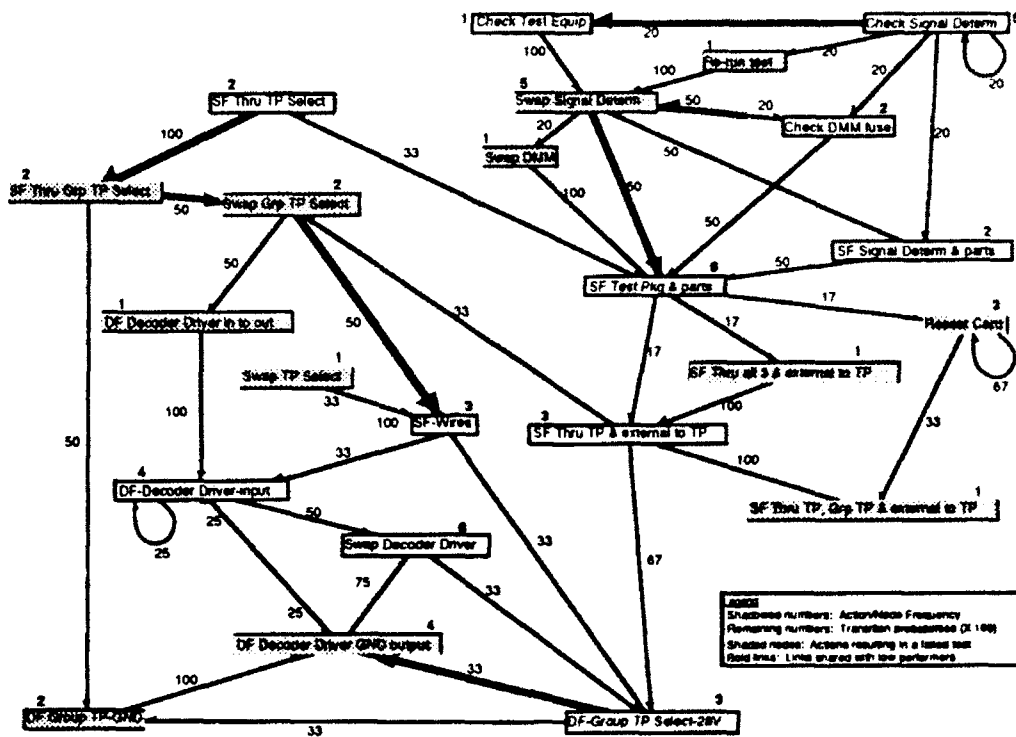


Figure 2. Pretest network based on aggregate transition probabilities of the six high performers.

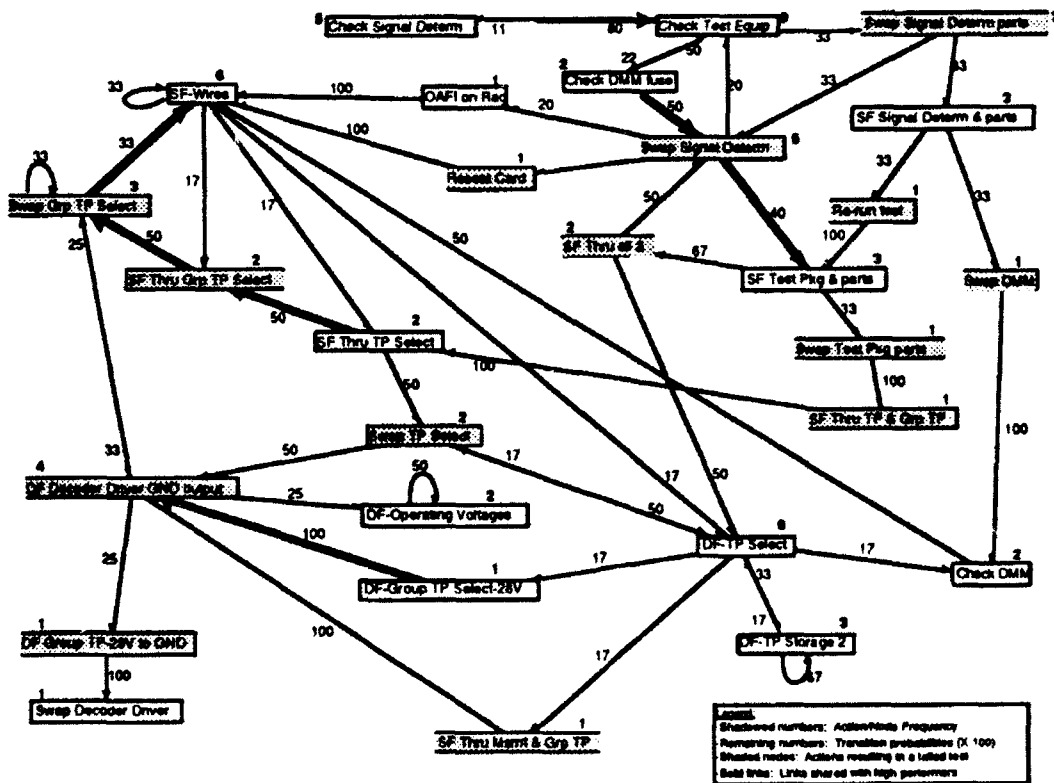


Figure 3. Pretest network based on aggregate transition probabilities of the eight low performers.

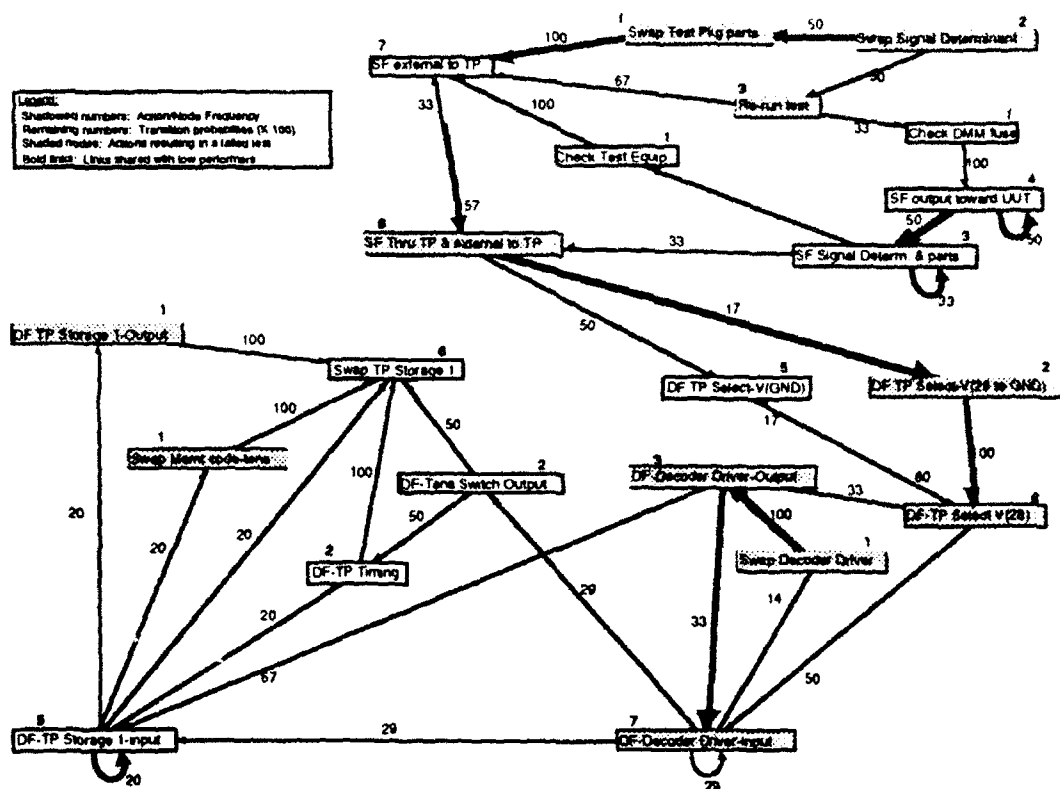


Figure 4. Posttest network based on aggregate transition probabilities of the six high performers.

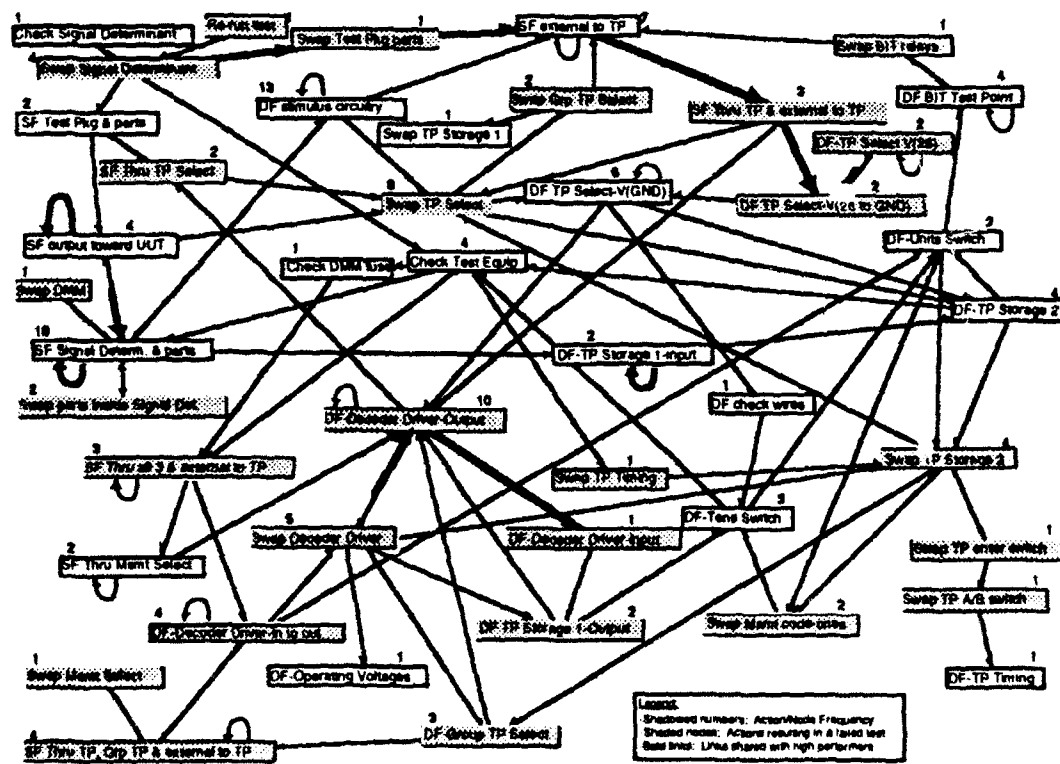


Figure 5. Posttest network based on aggregate transition probabilities of the eight low performers (link weights are omitted due to graph complexity).

both take shared actions into account. However, the Pathfinder measure includes information on action sequences, whereas the action frequency measure includes frequency of individual actions. Finally, the second other measure was the total number of actions that each subject executed (i.e., the number of steps to solution).

Examination of Table 1 indicates that the Pathfinder similarity measure is predictive of troubleshooting scores for the pretest ($r(22) = .57, p < .01$), but not for the posttest ($r(22) = .26$).

Table 1. Intercorrelation matrix of four assessment measures. (VT score = verbal troubleshooting score; PF sim = similarity of Pathfinder network with expert network; ActFreq = correlation of action frequencies with expert action frequencies; No.Act = number of actions)

Table 1a					Table 1b				
Pretest Measures					Posttest Measures				
	1	2	3	4		1	2	3	4
1. VT score	1.0	.57**	.65**	.38	1. VT score	1.0	.26	.76**	-.35
2. PF sim		1.0	.47*	.38	2. PF sim		1.0	.55**	.22
3 ActFreq			1.0	.30	3 ActFreq			1.0	-.17
4. No.Act-				1.0	4. No.Act				1.0

* $p < .05$; ** $p < .01$

However, the action frequency measure is predictive of scores for both the pre- ($r(22) = .65, p < .01$) and the posttest ($r(22) = .76, p < .01$). Other significant correlations indicate that the two measures of Pathfinder similarity and action frequency are highly intercorrelated, as was predicted. However, at least for the pretest, both measures seem to independently account for a portion of the variance. The correlation between the troubleshooting score and the action frequency measure remains significant when the Pathfinder similarity measure is partialled out ($r(21) = .53, p < .01$). Also, the correlation between the troubleshooting score and the Pathfinder similarity measure is marginally significant when the action frequency measure is partialled out ($r(21) = .39, p < .07$).

Another way of looking at these data is to compute change scores for subjects from pretest to posttest and correlate these scores. Because only 20 of the 24 nonexperts were classified as nonexperts for both tests, data were analyzed for only these 20 subjects. As might be expected from the previous analysis, the change in troubleshooting score was highly correlated with both the change in Pathfinder similarity ($r(18) = .51, p < .05$) and the change in action frequency ($r(18) = .55, p < .05$). The only other correlation to reach significance was between change in Pathfinder similarity and change in action frequency ($r(18) = .49, p < .05$).

Taken together, these results suggest that the types of actions subjects perform and the frequency with which they perform them are predictive of both the pre- and posttest scores. In addition, the specific sequence in which actions are executed is predictive of the pretest scores. As will be discussed below, there was a much wider range of actions performed by the low

performers in the posttest compared to the pretest which may have overwhelmed any predictive power of sequential variation.

Finally, the assessment measures can also be compared in terms of their ability to discriminate subjects in the experimental and control groups. The mean scores of experimental and control subjects for the pretest and posttest are presented in Table 2. As should be expected, there were no pretest differences between experimental and control groups. Interestingly, the only significant difference between these two groups at posttest is for the Pathfinder similarity measure ($t(22) = 2.07, p < .05$). Subjects in the experimental condition had networks that were more similar to the ideal network than did subjects in the control condition. The lack of a significant verbal troubleshooting score difference between the two groups is most likely due to the restriction of range that occurred by eliminating the six highest performers on the posttest. The fact that Pathfinder accounts for experimental vs. control differences, but not the action frequency measure, suggests that subjects who were trained on SHERLOCK learned more expert-like action sequences than those who were not.

Table 2. Mean assessment measures for experimental and control groups on pre- and posttests. (VT score = verbal troubleshooting score; PF sim = similarity of Pathfinder network with expert network; ActFreq=correlation of action frequencies with expert action frequencies; No.Act = number of actions)

	Pretest Mean	Posttest Mean
<u>VTscore</u>		
Experimental	42.00	68.00
Control	47.00	59.00
<u>PFsim</u>		
Experimental	.05	.07
Control	.05	.04
<u>ActFreq</u>		
Experimental	.38	.41
Control	.33	.24
<u>No.Act</u>		
Experimental	12.60	15.80
Control	11.50	16.60

In sum, this procedure seems to identify meaningful action patterns. Assessment in avionics troubleshooting is currently carried out using the scoring worksheet (Pokorny & Gott, 1992) and, as demonstrated above, an assessment measure based on Pathfinder action patterns corresponded to that of the scoring worksheet. Although the long-term plan does not entail diagnosis of student knowledge directly from action patterns, one of the purported benefits of the Pathfinder analysis is its ability to offer information beyond the mere assessment of student knowledge. In the following section diagnostic implications of the Pathfinder analyses are discussed, the main question being does Pathfinder highlight specific strengths and weaknesses in students' knowledge that can be targeted for intervention? The analysis that follows entails identifying the strengths and

weaknesses of the low performers as a whole relative to the high performer ideal, although an identical analysis could be performed at an individual level.

Diagnosis and Intervention

The Pathfinder networks for the high and low performers differed both quantitatively and qualitatively. Some of the quantitative differences between individuals and high performers were captured in the network similarity measures described above. General quantitative differences between the two groups can be seen in terms of the number of nodes and links present in the networks of the high and low performers. The high performers' networks had fewer nodes (i.e., actions; pretest=23, posttest=21) than the low performers' networks (pretest=28, posttest=45), especially at posttest (see Figures 2 through 5). In other words, the high performers as a group executed fewer distinct actions than the low performers, indicating a less varied repertoire of actions across all high performers for this problem. High performers seem to agree on the relevant actions for this problem in comparison to low performers. Although the low performers at posttest executed over twice as many distinct actions as the high performers, they shared all but one of the high performer's actions (shared nodes at pretest=17, at posttest=20). Thus, at posttest the low performer's applied a wide repertoire of actions as a group, including actions that were expert-like. These results suggest that the low performers as a group have knowledge about a wide variety of actions by posttest, yet they do not seem to understand when these actions apply. Interestingly, the subjects in the experimental group executed fewer distinct actions (35) than those in the control group (48). Thus, SHERLOCK may be effective in teaching students the conditions under which various actions apply.

What do these differences indicate in terms of diagnosis and intervention? First, the six pretest nodes in the high performers' network that were not contained in the low performers' network consisted of signal flow and data flow tests. In addition, at pretest, low performers executed 11 actions (corresponding to 11 extra nodes) high performers did not, seven of which were data flow and signal flow tests. At posttest, half of the additional actions executed by low performers were data flow and signal flow tests and half were swaps. Thus, these errors of omission and commission indicate that intervention in these particular cases should be targeted at learning the appropriate data flow and signal flow actions. A more detailed target may be derived by a focus on individual nodes.

In addition to having fewer nodes than low performers, high performers' networks also had fewer links than the low performers' at both pre- and posttest (high performers: pretest=44, posttest=38; low performers: pretest=52, posttest=94). This is to be expected given the fact that fewer nodes necessarily implies fewer links. However, the number of links per node does not differ greatly for the four networks. For each node there are approximately 2 links per node (range = 1.8 to 2.1) across the four networks. However, the number of links shared between the high and low performers increased slightly from pre- (7) to posttest (11), suggesting that the low-performers began demonstrating action sequences more like those of the high performers. This pattern is verified by the C measure of similarity between the networks of the low and high performers at pre- ($C = .04$) and posttest ($C = .07$). Although the resulting C values are relatively

low, they do indicate that the low performers became more like the high performers at posttest. For instance, even the low performers at posttest had learned to conduct the signal flow and data flow tests which the high performers used at posttest to pinpoint the fault. Thus, the low performers learned more expert-like sequences of actions, given training.

The networks of the high and low-performers also differed in some more global ways. First, the high performers (both tests) appeared to follow a rule about the general sequence of actions which were taken: (1) general checks outside of the test package, including visual checks, equipment checks, and swaps, (2) signal flow tests inside the test station, (3) data flow tests of components inside the test station and (4) swapping. Low performers, on the other hand, committed violations in this general sequence in both tests. Second, the low performers exhibited several of what may be termed meaningless action sequences on both tests, whereas high performers did not. For example, after completing a signal flow or data flow check which indicated that the component was functional, some low performers chose to swap the component anyway.

CONCLUSIONS

The results obtained from the work completed thus far are promising in that they indicate that meaningful patterns of actions can be identified using the Pathfinder network scaling procedure, thereby achieving the subgoal indicated in Figure 1.2. The network patterns are also meaningful in the sense that they can differentiate high and low performers as defined by the scoring worksheet. In addition, the Pathfinder networks reveal qualitative differences between high and low performers that are suggestive of targets for intervention (e.g., data flow and signal flow tests). Finally, this bottom-up approach to knowledge elicitation resulted in general action patterns that may not have been verbalized in a typical knowledge elicitation interview (e.g., the general sequence executed by high-performers). These results are even more promising when the source of the ideal or expert network used to make these comparisons is considered. Specifically, high-performers were used here as the ideal. An even better ideal would probably result from the use of recognized SMEs. In addition, the use of subjects with more expertise would widen the range of performance, which would likely result in enhanced assessment and diagnostic capabilities. The next step of this project is the evaluation of different measures of system knowledge (the subgoal represented in Figure 1.1). The longer-term goals include the mapping of system knowledge onto action patterns and prediction of system knowledge based on this mapping.

The short term (one year) contributions of this work include: (1) A method of generating network representations of student actions and an evaluation of this method. (2) An alternative to, or extension of, existing methods for assessing student troubleshooting knowledge on-line. (3) A method for targeting specific actions or sequences associated with overall knowledge strengths or deficits. (4) A method or set of methods that have been determined to be optimal for eliciting and representing the system knowledge of students.

The longer-term contributions of this work are: (1) A procedure for on-line assessment and diagnosis of student's system knowledge which involves mapping action patterns onto deficits or proficiencies in system knowledge. (2) A procedure which summarizes actions (errorful and

correct) in terms of a rich representation of student knowledge that lends itself to qualitative analysis useful for diagnosis and intervention. (3) An assessment and diagnosis procedure that targets the complex domain of avionics troubleshooting. (4) A methodology that can be applied to the problem of eliciting knowledge from SMEs for tutor development. (5) A general test of the assumption that system knowledge underlies troubleshooting actions.

REFERENCES

- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. Science, 228, 456-462.
- Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 157-183). London: Academic Press.
- Cooke, N.J. (1992a). A Taxonomy and Evaluation of Knowledge Elicitation Techniques. Lockheed Engineering and Sciences Corporation Technical Report. Houston, TX: Human-Computer Interaction Laboratory
- Cooke, N. J. (1992b). Predicting judgment time from measures of psychological proximity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 640-653.
- Cooke, N. J. & Breedin, S. D. (1992 manuscript under review). Constructing naive theories of motion on-the-fly. Journal of Experimental Psychology: Learning, Memory, and Cognition.
- Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory and Cognition, 12, 538-49.
- Cooke, N. M., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. International Journal of Man-Machine Studies, 29, 407-427.
- Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. Schvaneveldt (Ed.), Pathfinder associative networks: Studies in knowledge organization (pp. 267-277). Norwood, NJ: Ablex.
- Ericsson, K.A., & Simon, H.A. (1984). Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press.
- Gitomer, D.H. (1984). A cognitive analysis of a complex troubleshooting task. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.
- Gitomer, D.H. (1992). Foundations of HYDRIVE. Presentation given at the Basic Job Skills Progress Review, Brooks AFB, San Antonio, April 7-9.
- Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization (pp. 75-87). Norwood, NJ: Ablex.
- Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization (pp. 241-253). Norwood, NJ: Ablex.
- Gott, S. P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E. Z. Rothkopf (Ed.), Review of Research in Education, Vol 15. (pp. 97-169). Washington D.C.: American Educational Research Association.
- Gott, S. P., Bennett, W., & Gillet, A. (1986). Models of technical competence for intelligent tutoring systems. Journal of Computer-Based Instruction, 13, 43-46.
- Greeno, J.G. (1983). Conceptual entities. In D. Gentner & A.L. Stevens (Eds.), Mental Models (pp. 227-252). Hillsdale, NJ: Erlbaum .
- Hall, E.M., Gott, S.P., & Pokorny, R.A. (1991). A Procedural Guide to Cognitive Task Analysis: The PARI Methodology. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

- Kellog, W.A., & Breen, T.J. (1990). Using Pathfinder to evaluate user and system models. In R. W. Schvaneveldt, (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization . Norwood, NJ: Ablex.
- McDonald, J. E., & Schvaneveldt, R. W. (1988). The application of user knowledge to interface design. In R. Guindon (Ed.), Cognitive Science and its Applications for Human-Computer Interaction (pp. 289-338). Hillsdale, NJ: Erlbaum.
- Nichols, P.D., Pokorny, R., Jones, G., Gott, S.P., & Alley, W.E. (1989). Effective Instructional Processes Within an Intelligent Tutoring System. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.
- Payne, S. J., & Squibb, H. R. (1990). Algebra mal-rules and cognitive accounts of error. Cognitive Science, 14, 445-481.
- Pokorny, R., & Gott, S. (1992, in preparation). The Evaluation of a Real-World Instructional System: Using Technical Experts as Raters. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.
- Potka, J., Massey, L. D., & Mutter, S. A. (1988). Knowledge acquisition. In J. Potka, L. D. Massey, and S. A. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 15-19). Hillsdale, NJ: Erlbaum.
- Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.
- Schvaneveldt, R. (Ed.), (1990). Pathfinder Associative Networks: Studies in Knowledge Organization . Norwood, NJ: Ablex.
- Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1985). Pathfinder: Scaling with Network Structures. Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory, New Mexico State University.
- Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 24, pp. 249-284). New York: Academic Press.
- Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & DeMaio, J. C. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.
- Stevens, A., Collins, A., & Goldin, S.E. (1979). Misconceptions in student's understanding. International Journal of Man-Machine Studies, 11, 145-156.
- Wilson, J.R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. Human Factors, 31, 617-634.

INTELLIGENT DECISION MAKING WITH QUALITATIVE REASONING

Ashesh Das

**Department of Statistics and Computer Science
West Virginia University
Morgantown, WV 26506**

Final Report for :

**Summer Faculty Research Program
Armstrong Laboratory : Human Resources Directorate
Logistics Research Division**

**Sponsored by Air Force Office of Scientific Research
Wright-Patterson Air Force Base, Dayton, Ohio.**

July, 1992

INTELLIGENT DECISION MAKING WITH QUALITATIVE REASONING

Ashesh Das
Department of Statistics and Computer Science
West Virginia University
Morgantown, WV 26506

Abstract

A class of common sense reasoning, called qualitative physics has been treated, where, physical systems being subjected to deep reasoning studies on their behavioral pattern, can focus one dimensional qualitative laws free from mathematical intricacies, but sufficient to describe their mainstream behaviors with discrete real numbers. The nature of decision making with tools using qualitative physics in an uncertain environment has been discussed. It is stressed that the notion of generic task structures play a dominant role in design/diagnostics with qualitative reasoning. This idea has been extended to decision making in robot-automated operations, where, planning and decision making can be done by constructing partially ordered bases consisting composite pairs of geometric descriptions of a subject and the corresponding generic task structure. The theory has been illustrated with a simple example.

INTELLIGENT DECISION MAKING WITH QUALITATIVE REASONING

Ashesh Das

1. Introduction:

Critical technologies, as coined by the White House office of Science and Technology Policy, partly cover these items:

- | | |
|---------------------------------|---------------------------------------|
| 1. Simulation and Modeling, | 2. Machine Intelligence and Robotics, |
| 3. Flexible Manufacturing, | 4. Software Engineering, |
| 5. High Performance Computing, | 6. Photonics, |
| 7. Sensitive Radars, | 8. Passive Sensors, |
| 9. Signal and Image Processing, | 10. Data Fusion, |
| 11. Composite Materials, | 12. Computational Fluid Dynamics, |
| 13. Superconductivity, | 14. Biotechnology, |
| 15. Superconductivity, | 16. Semiconductor Materials/Circuits. |

Tools of artificial intelligence (AI) are applied in all of them for intelligent decision, management, understanding and learning of information, and for enhanced automated production, design and diagnostics. Practical problems in each of the cases are enormous. It is, mostly, because of the massive amount of knowledge and data of varying nature associated with complex high-tech machines. They are constantly being updated too.

A major activity in AI is to represent all types of knowledge and data appropriately, and then perform symbolic operations. Knowledge is very often represented with predicate logic, with production rules, with frames, with semantic nets, with tools of nonmonotonic reasoning (like default reasoning, modal logic, fuzzy logic, etc.), with conceptual dependency, with scripts etc. But extracting knowledge (knowledge elicitation/acquisition) from domain experts is a complex and time-consuming process. When knowledge and data are time-varying and incomplete, forming a calculus of intelligent reasoning is even more complex. A class of AI operations called common sense reasoning is a powerful applicable tool in such contexts [1]. It works with "essential knowledge" ignoring detailed data-driven complex processes. Let us consider FIG.1 in this context.

A cylinder is shown with several spherical devices inside. The cylinder is oscillating. Imagine that, suddenly one of the spherical devices is stripped off. Is it possible to say which one has been stripped off, and what may be the immediate possible consequence? A common sense reasoning is applicable now. Let us observe the oscillating cylinder.

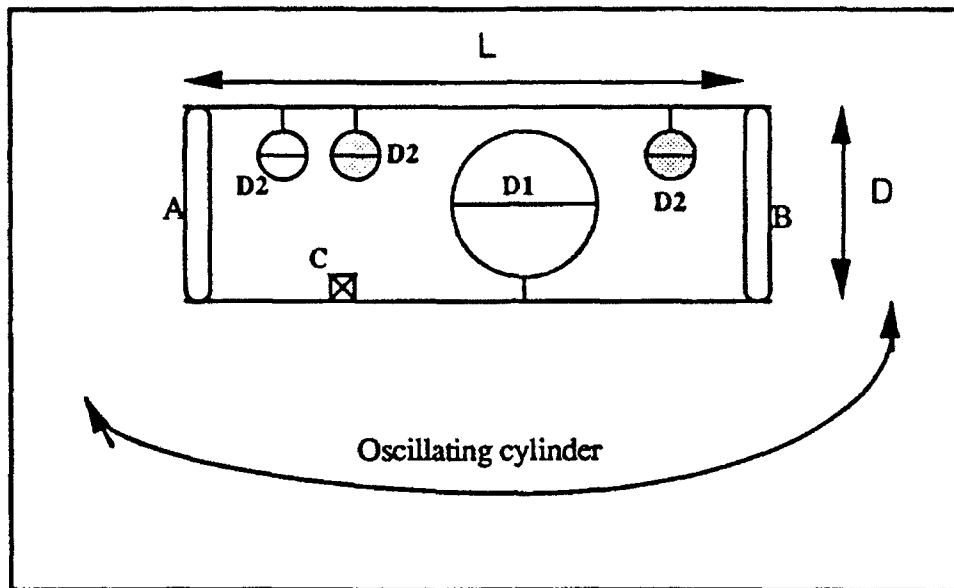


FIG. 1

There is a spherical device with diameter $D1$ inside is known. There are other spherical devices with diameter $D2$ is known too. It is known that $D1$ is much bigger than $D2$. The diameter of the cylinder is D , and its length is L . While the cylinder oscillates, the stripped off spherical device starts hitting side A and side B , periodically. The bigger spherical device with diameter $D1$ will take smaller time to hit sides A and B than smaller devices with diameter $D2$. Common sense quickly calculates that $(L-D2) > (L-D1)$. Thus a prediction can be made after listening to the time of hitting sides A and B . Imagine now that C is a delicate instrument inside the cylinder. If from the time of measurement it is inferred that $D1$ is rolling inside, then there is a bigger chance of the instrument C to be damaged. Common sense immediately decides that the weight of $D1$ is much larger than the weight of C . The problem can be reformulated in this way : it has been reported that as soon as the cylinder AB started oscillating, the instrument C broke down. An immediate common sense inference is that the bigger spherical device $D1$ is stripped off, and is rolling inside. It did hit the instrument C , and broke it. The common sense reasoning counted upon the description of the system, initial conditions, looked for the temporal behavior, and put forth an explanation.

Question: Is computerized simulation of common sense reasoning applicable in all of these critical technology issues mentioned earlier ?

Answer: It lies in the motivation while approaching the problem. To apply common sense reasoning, the domain of application must be precisely stated with a theory for the domain working with valid ontologies. Entities defining a system, relations between them, and rules controlling them constitute what is called ontology. Knowledge must be appropriately represented in the domain, and there must be a technique of drawing inferences. Such paradigms of common sense reasoning have been successfully applied in the physical world. It has been called qualitative physics, or qualitative reasoning [2].

Causal ordering of relations among tasks and variables defining a system play a very important role in qualitative reasoning. Herbert Simon is one of the architects of causal theories [3], and this theory has been applied to develop a theory of qualitative reasoning by Iwasaki and Simon [4]. According to their theory, causal ordering is context sensitive. For example, there may be a set of equations describing the behavior of an oscillating cylinder and its parts inside, as shown in FIG.1, but only those equations are to be taken that reflect correct understanding of physics involved producing correct causal structure. If the causal ordering is established, a precise functional form of equations describing the system may be ignored. de Kleer and Brown [5] have addressed this problem using systems with infinitesimal time intervals. Reiger and Greenberg [6] developed a theory of qualitative reasoning with events, actions, and states connected by causal links that describe how states, actions, and events trigger each other. Chandrasekaran and coworkers [7] developed a theory of qualitative reasoning applicable in engineering diagnostics and engineering design, where, a device is described by a series of states connected by causal links.

Qualitative reasoning very often extracts information by considering the system as closed and characterizable by one-dimensional parameters, where multiple theories are applicable. An electrical circuit, or a simple heat engine or a hydraulic system can be considered as closed systems (isolated from external world, stimulus, input) describable by one dimensional parameters. If there are causally ordered relations between these parameters, the system behavior can be studied with multiple theories: sometimes electronics, sometimes mechanics, or sometimes thermodynamics, sometimes statistical mechanics. Sometimes a competing theory may dominate the whole situation, sometimes several theories may combine to form a single theory. Sometimes a mixing of different environments is done to define

ontologies of a system. Hayes [8] introduced a qualitative reasoning for the physics of liquids isolating them by their containers.

Physical systems, being subjected to deep reasoning studies on their behavioral pattern, can focus some one dimensional qualitative laws free from mathematical intricacies but sufficient to describe their mainstream behavior with discrete real numbers. A study of such laws forms a class of qualitative reasoning, called qualitative physics.

2. Qualitative physics:

There are three major approaches to qualitative physics : component-based approach of de Kleer (5), process-based of Forbus (9), and constraint-based approach of Kuipers (10). In component-based approach, a device is seen as composed of three different types of constituents : materials, components, and conduits. To study the behavior of a device, a qualitative study of constituent behavior is made. Examples of materials are, water, air, electrons. The form of characteristic properties of materials can be changed by components. Qualitative equations are formed which encode rules for determining different types of components. Conduits are physical parts of a device that transport materials from one component to another without altering any information related to the materials. Examples of conduits are pipes, wires, cables etc. In process-based approach the prime issue is to build models of physical phenomena. A physical situation is modeled as collection of objects, their relationships, and processes. Variables are used to represent object properties which are influenced by processes in time. A closed world assumption is made there upon to make appropriate set of constraints controlling process evolution. The constraint-based approach starts with a set of qualitative constraints, and an initial state, and predicts the set of possible futures for the system. This approach has a precise mathematical semantics as an abstraction of differential equations.

In qualitative physics calculations are done with intervals on the real number space instead of actual numbers themselves. For example, the 'sign' + represents the interval $(0, +\infty)$, 0 represents $[0,0]$, and - represents $(-\infty,0)$. Qualitative variables take these values only, which are considered well-behaved in the sense that, they are continuous and differentiable. The notation has a distinct role in qualitative physics. It is used to give the sign of a quantitative variable. Iff the quantitative values of a variable is showing behavior given by variable $x > 0$, then it is stated that $[x] = +$. Similarly, iff $x = 0$, then $[x] = 0$, and iff $x < 0$, then $[x] = -$. Qualitative addition and multiplication tables are shown in FIG.2.

		[x]+[y]					[x] * [y]		
		[x]	-	0			+	[x]	-
[y]	-	-	-	?	-	+	0	-	
	0	-	0	+	0	0	0	0	
	+	?	+	+	+	+	-	0	

FIG.2

It is seen that $[x]+[y]$ is not always defined. $[x]*[y]$ is always defined. Similarly $[x] [y]$ is not always defined [4]. If + stands for pressure increasing, - stands for pressure decreasing, and 0 stands for no change, then: pressure increasing + pressure increasing gives pressure increasing. Similarly, pressure decreasing - pressure decreasing gives pressure decreasing. But, pressure increasing + pressure decreasing is ?, that is undefined. A qualitative equation $mx+ny = 0$, where, m and n are two positive constants, in qualitative version becomes $[x]+[y]=0$. This equation, also known as confluence, will follow the sign convention given above for $[x]+[y]$. A differential equation of the type $dx/dt = c(x+y)$, where, $c=0$, in qualitative form looks like $[dx]=[x]+[y]$. Here $[dx]=[dx/dt]$. The sign of a qualitative variable is very important. $[dx]$ is +, when x increases. $[dx]$ is minus when x decreases, and $[dx]$ is zero, when x is zero. Quantitative equations are manipulated by conventional mathematical rules and axioms, and they are much more flexible than qualitative equations (confluences). Thus qualitative operations very often lead to some inherently built-in ambiguous situation. Such ambiguities lead to multiple behavioral predictions, called interpretations. Among many interpretations, one must correspond to the actual physical situation. A qualitative simulation sometimes must split into two or more paths. To make proper predictions and a comprehensive understanding of the system, all possible qualitative states and transitions must sometimes be networked to form a system, de Kleer and Brown [5] called it envision. Under an appropriate device description qualitatively, envision can predict the behavior of a device in a sequence of all possible causally connected future states. In component-based approach a basic policy is to guarantee that device behavior is describable from individual components and their connections and the global function of the device is not determinable by the local function of its components. This is done to confirm that any prediction made on

the behavior of a component by qualitative reasoning is not already contained in the description of it. Two principles are working in this line : locality principle, and no-function-in-structure principle. Locality principle assures that a component description is always confined to its own topology without any reference to any other component in the system. No-function-in-structure principle guarantees that a component description must not reference in any way the functioning of the whole device.

However, the prime question is : why one should attempt doing qualitative physics ? Qualitative physics works with small number of variables that take small number of values. Consequently, it can function with incomplete data and incomplete models. Very often such models demand provision for causal explanations building for its predictions. Such explanations make subsequent quantitative analysis easier. A comparative summary of qualitative and quantitative approach is shown in FIG.3.

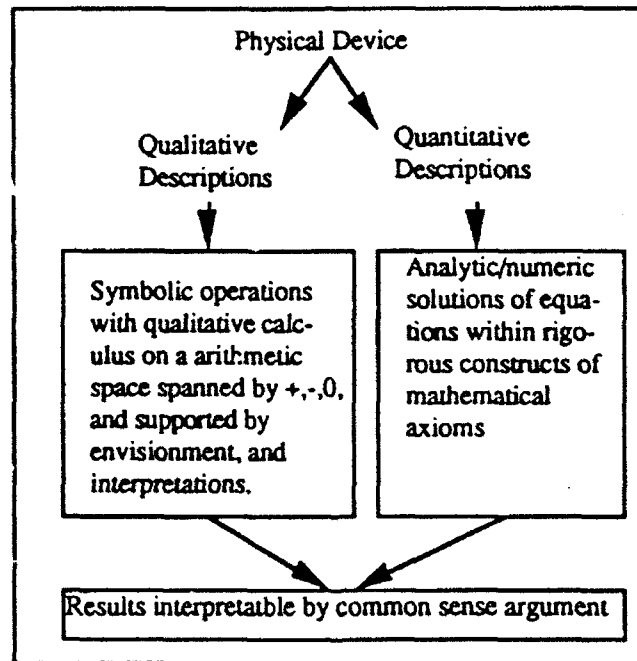


FIG.3

3. Decision Making with Qualitative Physics :

Decision making with qualitative physics is a class of model-based deep reasoning system. This should be contrasted with shallow level rule-based approach as seen in traditional expert systems. The best side of model-based approach is having a knowledge representation that depend on the specific problem, and on the characteristics of the knowledge to be represented. This provides a compatibility of

observation made to the inferencing tool. In shallow level rule-based reasoning such a compatibility is missing, unless there are methods for incorporating causal relations among the observations, as demonstrated by Patil et al [11], and Pople [12]. In a study on causal reasonings by expert physicians Kuipers and Kassler [13] observed that, one important representation for the model of a mechanism consisted of qualitative descriptions of continuous variables, their directions of change, and the constraints among them. This observation is in complete conformity with classic works in qualitative physics by de Kleer [5], Forbus [9] and Kuipers [10]. In diagnostics and design decisions, the role model-based qualitative physics can be described by FIG. 4.

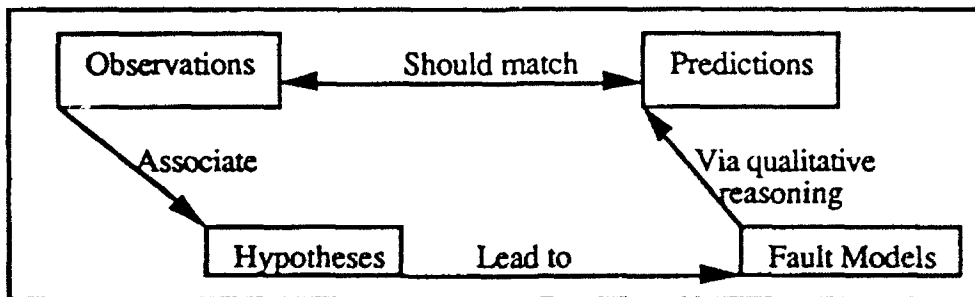


FIG.4

To make decisions with the help of decision trees is a common practice. But most of the time decision trees become very large requiring a massive storage in the computer. Influence diagrams are therefore used [14]. These are much smaller diagrams than the corresponding decision trees which contain three sets of knowledge : causal knowledge about events influencing each other, knowledge about what steps may be taken, and normative knowledge about how good the consequences are. Influence diagrams are signed directed graphs, and their real strength is in representing a state of partial knowledge during model creation. However, a decision tree or an influence diagram is good for open systems. Traditional qualitative physics is a closed system, by definition. A closed system guarantees locality and no-function-in-structure principle. Kuipers [15] pointed out that decision making using qualitative physics actually comes in between the numerical solutions of differential equations and influence diagrams. A major difficulty with differential equations is that functions relating variables must be stated explicitly. Since qualitative physics works with intervals on real numbers instead of the whole number space, it forms its own quantity lattice, as conceived by Simmons [16] where, decision making can be seen as a graph searching problem, or a numeric constraint propagation problem, or

interval arithmetic problem to constrain values of arithmetic expressions, or relational arithmetic problem to constrain values of arithmetic expressions, or constant elimination arithmetic problem to constrain the value of arithmetic expressions. At the same time they provide more inferential power than influence diagrams by applying limit analysis to changing variables, and by having a mathematically precise semantics as an abstraction of differential equations. The discretization of number space of the traditional quantitative analysis gives a strong visibility to qualitative physics. One discrete value changes to another, where semantically important points for the system are changing. Decision making thus boils down to searching a highly truncated search space. This is a major computational power.

4. The question of uncertainty:

If decision making is primarily done by searching, then incorporating heuristic knowledge in ordering and pruning the search is a good practice. de Kleer and Brown [5] used heuristic rules that restricted the types of assumptions introduced during value propagation. But a practical question is, what is the nature of searching or decision making in the presence of uncertainties? Moreover, qualitative physics at the generic phase involve some ambiguities. This comes because of working with only three quantized values +, -, and 0. Incorporating quantitative measures with qualitative calculations, as done by Simmons [16], is one way of reducing uncertainties. Another approach is due to Kuipers [10] the qualitative calculus works with a number of quantized values, as long as the number is finite. Apart from the built-in ambiguity, there are mainly three reasons for uncertainties to enter into qualitative physics: (1) the subject, where, qualitative physics was applied had uncertain information on its structure, performance, and life cycle. (2) Qualitative rules giving the performance of the subject carried uncertainty. (3) Such rules were dealt with ill-structured qualitative or quantitative intervals, causing further imprecision to occur. Whatever may be the source of uncertainty in qualitative physics calculation, it falls under Judea Pearl's classification of approaches to probability: extensional and intentional. If the uncertainty in a qualitative physics calculation is dealt with the techniques of production systems, or rule-based systems then it comes under extensional class. If the uncertainty is dealt with declarative rules dealing belief and possible worlds, then it comes under intentional class. Since declarative rules are model-based, and qualitative physics deals with model-based reasoning, uncertainty in qualitative physics will be dealt with the notion of intentional class. Here probability ultimately takes the role of belief with four relations working with it: likelihood (the pressure in

the device is more likely to increase than the temperature), conditioning (if pressure increases, then the cylinder breaks), relevance (whether the cylinder will break depends on whether the pressure crosses the breaking stress), and causation (having the breaking stress crossed, the cylinder caused it to break). Pearl [14] has demonstrated how these relations lead to a beautiful theory of probabilistic networks with propagating belief, and a normative decision theory develops.

5. Planning and decision making:

There has not been much work in planning with qualitative physics. Within the context of classical temporal planning theory, Hogge [17] constructed domain compilers that takes a qualitative physics domain as input, and produces rules suitable for a temporal planner. The domain compilers, given the description of certain liquid flow as input, produces inference rules describing the situation behind the liquid flow to happen. Such rules with other inference rules, and specifications of the actions an agent may take, help the planner to create plans involving processes as intermediaries, such as filling a kettle by moving it under a faucet and turning it on. However, such a system will work only in a closed world environment, where there is no robot, no machine controller doing design/diagnostic studies to have confluences rewritten. One of the reasons behind that, generating knowledge preconditions for actions and plans require flexibility in confluence-controlled behavior whose nature cannot be a priori assumed. Moreover, incompleteness in knowledge brings uncertainty in plan execution. A planner cannot be absolutely sure that the goal after plan execution is absolutely fulfilled.

For decision making and planning under uncertainty the use of nonmonotonic logic is a common practice. Langlotz and Shortliffe [18] has attempted to establish a connection between decision theory and nonmonotonic logic. They argued that planning using nonmonotonic logic comprises two decision-theoretic concepts: probabilities that represent degree of beliefs in planning systems, and utilities that represent degree of preferences for planning outcomes. In fact, a major part of decision theory is to deal preferences over outcomes. Possible consequences of executing a plan is known as outcome. With outcomes is associated a degree of desirability called utility. The expected utility over m outcomes is computed as: $\langle U(P_i) \rangle = \sum_{j=1}^m p(O_j | P_i) * U(O_j)$, where j is summed from 1 to m , $U(P_i)$ is the expected utility of the i th plan, $p(O_j | P_i)$ is the probability of the j th outcome after executing the i th plan, and $U(O_j)$ is the utility of the j th outcome. The knowledge engineer has a crucial role in planning within decision context. He/she very carefully extracts probabilities signifying the likelihood of potential consequences of actions, and utilities that give the degree to which each

potential consequences satisfies the aimed goal. The plan with highest expected utility is the best choice. To achieve this the planner sometimes adopts a myopic policy for selecting information sources [14]. Fig 5 may be referenced now.

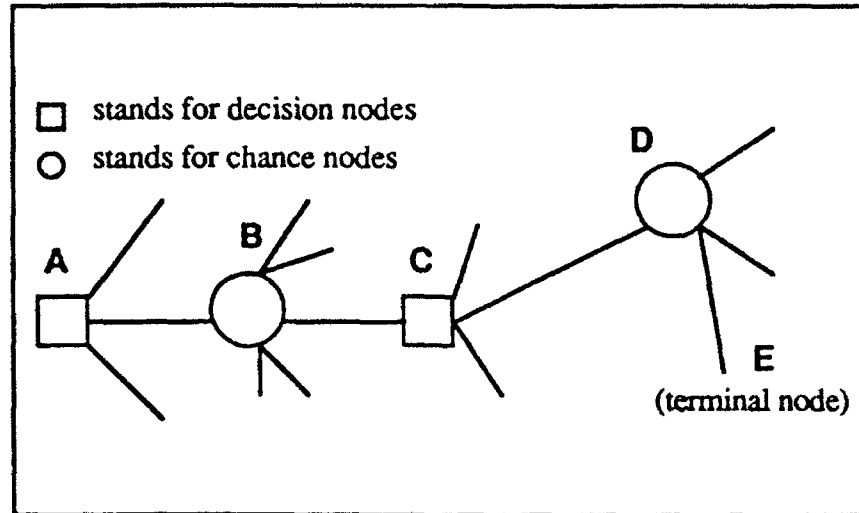


FIG.5

Boxes, i.e., A and B are decision nodes, and correspond to actions over which the decision maker has control. Circles, i.e., B and D are chance nodes, and correspond to situations described with probabilities, and are beyond the decision maker's control. E is the terminal node, which is the final outcome state. The qualitative planner now will have to assign probabilities to the branches of each node. This probability is subjective in nature. Objective data are descriptive of past events, and decision models use its probability to predict future events. Objective data will work as guidance to subjective estimates. A subjective probability will express a degree of belief in the proposition. A qualitative planner may start with some sort of prior belief on the fairness of the outcome of assigning probability to decision tree branches, and then update this belief using Bayes rule when quantitative data becomes available. To the planner the terminal node E (FIG. 5) is the goal. All by itself the goal will not have any feed back to the planner for improving its performance. Thus the planner will reach the goal with varying degree of certainty. Some AI planners apply heuristic evaluation functions to overcome such situations [19]. When the decision tree is an influence diagram, Wellman [20] has given a neat theory of probabilistic network with qualitative influence propagation.

6. Task structure, qualitative physics, and decision making :

Qualitative modeling requires that the model builder has a deep understanding of the subject to be modeled. During the past few years much attention has been drawn in identifying generic tasks on the knowledge base identifiable with certain type of domain specific knowledge and associated control strategies [21]. The generic task has been formulated by (1) task function, (2) knowledge representation/organization, and (iii) control strategy. Task function will enumerate types of problems and kinds of inputs/outputs involved or expected. Then comes the question of selecting primitives for representing the knowledge of tasks. The next step involves inference strategy adoptable to accomplish generic task function. Bylander and Chandrasekaran [22] showed that such task-specific knowledge-based techniques help in knowledge acquisition and explanation generation.

Knowledge compilation compatible to generic tasks must be well-performed. This will reduce the uncertainty associated with ambiguities built-in with qualitative reasoning. Transforming a quantitative description to a qualitative form is one discipline of knowledge compilation. As a result the qualitative form becomes more suitable with admitted assumptions on the signs of variables. The notion of knowledge compilation is involved in the process because breaking a task into a series of generic tasks (and then reformulating them adoptive to qualitative descriptions) directly links to the idea that, types of problem solving and types of knowledge are closely related. Generic task structure guarantees these following features [21].

Multiformity : Each generic task provides one unique way to organize and use knowledge. The knowledge engineer designing the decision making program decides which task to use and when to use.

Modularity: In every qualitative arithmetic space there is one (or a combination of) generic task(s) satisfying local goals.

Knowledge Acquisition: Every generic task structure has its own knowledge acquisition strategy.

Explanation: The tactical control strategy on the qualitative arithmetic space, and data matching therein , may be supported by explanation generation.

Exploiting Inference Between Knowledge and Inference: The compiled knowledge for each model is domain-specific, and therefore application/control specific. This guarantees modularity, multiformity, and tractability of generic task structures.

Design problem solving is a complex activity that can be represented by a 4-tuple: $\langle A, \rightarrow, B, C \rangle$. Here A represents specifications, \rightarrow represents transformations, B represents goal, and C represents constraints to be satisfied. If the above expression is

translated into a world of qualitative physics with enough flexibility to handle three issues discussed by the IJCAI-91 panel on AI and design [23]. These issues are: (1) decision making, (2) representations, and (3) knowledge handling including modeling and simulation.

The first issue, decision making address to questions like: how to handle poorly defined and incomplete design specifications, how to handle multiple interacting goals and constraints, especially in tasks of concurrent design and those involving trade-off decisions, how to choose decompositions in various problem solving contexts, how to handle formation problems, how to handle hierarchical reasoning process in design, and how to handle incremental design and redesign tasks. The second issue, representations address to questions like: how to represent design methods so as to facilitate explanation and design reuse, how to represent design from multiple view points, and how to organize large reusable knowledge bases. The third issue, knowledge handling including modeling and simulation, address to questions like: how to use domain knowledge for a priori guiding the generation of candidate solution rather than (in addition to) a posteriori evaluation of candidate solutions, how to find and exploit useful approximation to main theories/models to enable reasoning from function to structure and to provide computationally tractable evaluations of solution candidates at various levels of solution construction, how to integrate qualitative and quantitative knowledge and domain in the context of design problem solving how to automatically acquire/refine domain knowledge, and how to learn from design experience.

Such a venture on design problem solving with AI reasoning proposes that there may be a set of generic design primitives creating adoptable confluences. It is useful to identify such design primitives as generic task structures, then design can be seen as searching a space of subassemblies [24]. Then a qualitative reasoning for identifying appropriate generic task modules become a graph search searching problem [16], where the task structure for design boils down to propose-critique-modify (PCM) family of methods. According to PCM methods, the qualitative reasoning must be able to support subtasks of proposing partial or complete decision design solutions, verifying the proposed solutions, critiquing the proposals by identifying possible causes of failure, and modifying proposals to satisfy design goals. It is obvious that PCM methods will have to survive with inherent ambiguities in qualitative reasoning, or will have to create qualitative arithmetic spaces with solvable confluences to get rid of ambiguities .

It has been claimed that here are approaches in qualitative physics with notions of generic task built inside. One such approach has been called assembling a device [25].

where local qualitative constraints have been fused to global relations. For some typical design task to be performed by simulation, observation etc., some reference variables are selected. The assembling step produces confluences with these reference variables. Determining nonambiguous variables for a particular assignment of the reference quantities becomes an easy task. As an illustration FIG. 6 is referenced [25].

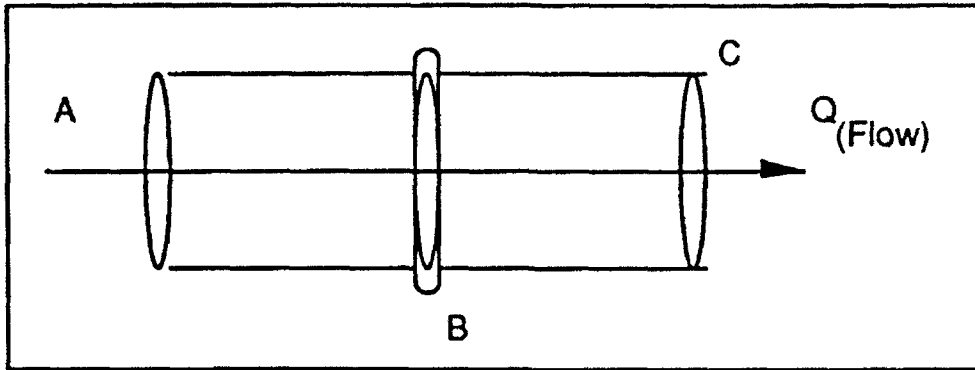


FIGURE 6

Two pipes AB and BC have been connected. The confluences describing this situation are (P is pressure and Q is flow): $[dP_A] - [dP_B] - [dQ] = 0$, and $[dP_B] - [dP_C] - [dQ] = 0$, which gives (dP_B has been eliminated) $[dP_A] - [dP_C] - [dQ] = 0$. Since Gaussian elimination of $[dP_B]$ does not always work for confluences, a typical qualitative resolution rule has to be used, that ensures globality. It is not difficult to identify such resolution rules as methods in the generic task package. Such attempts exemplifying that qualitative reasoning must be able to reason, and make decisions on the relationships between the device geometry and its function in the device assembly. Describing kinematic behavior of a mechanism as possible motions in causal terms, a qualitative design space on configuration space can be created [26]. In this space a causal description is a collection of state diagrams [5, 9]. Each state corresponds to a qualitatively different behavior. For making any bonafide decision, the configuration space is made acceptable under a prescribed set of confluences.

7. Decision making in robot-automated operations

Existing theories of robot navigation with techniques like triangulation, ranging sensors, auto-focus, stereo techniques, dead reckoning, inertial navigations, geo-satellite location, correspondence of map data with the robot's location, and obstacle avoidance techniques have been criticized to be brittle and error accumulating, and are limited by range of active sensors depending on accurate quantitative measurements of distance/directions perceived or travelled. Managing massive

quantitative data for robots are really difficult. Recently, an attempt has been made to reformulate robot navigation by qualitative reasoning [27]. Robots may be able to take decisions on the basis of some land mark values defining transitions of processes, or change in geographic topologies. It is actually searching robot memory in a quick short-cut way and to enhance decision making processes. It appears that with the aid of generic task structure method a planning theory can be designed for robots taking decisions in a task oriented environment like manufacturing plants. A plan is a sequence of time-indexed actions. This implies that when the robot fulfils a plan, a prescribed series of such hierarchically coming time-indexed actions have been executed. Such time-indexed actions can be written as a set of composite pairs [28, 29]:

$$\langle R_1, T_1 \rangle + \langle R_2, T_2 \rangle + \langle R_3, T_3 \rangle + \dots + \langle R_n, T_n \rangle, \quad (\text{Eq.1})$$

where, R stands for device-level qualitative parameters, and T stands for task-level actions (plan fragments) corresponding to R. R gives the topological configuration of the device, i.e., geometric description of the system. T gives the generic task corresponding to such descriptions. R is a set of qualitatively understood values (land marks (see ref.[5])). Let us reference FIG. 6 showing that two pipes have been joined. It is clear that the final confluence $[dP_A] - [dP_C] - [dQ] = 0$ contributes to the composition of the qualitative parameter R. If the plan is to open the joined section and separate the two pipes, then the confluences confirming this stage will be $[dP_A] - [dP_B] - [dQ] = 0$, for the first pipe, and $[dP_B] - [dP_C] - [dQ] = 0$, for the second pipe. After fulfilling the aforesaid part of the plan, say the next step is to join the first pipe A to a third pipe D. Another composite pair $\langle R, T \rangle$ comes into action. The generic task structure is constrained by a confluence of the type $[dP_A] - [dP_D] - [dQ] = 0$. Matching a qualitatively determined geometric parameter R to a task level issue T can be done on a blackboard [29]. A blackboard is a problem solver [30] enriched with highly domain-specific heuristic knowledge. Since the robot does not do one single job, or it may have to remove ambiguities in matching R to T, a comparative diagnostics can be done on the blackboard. Two plan with two different sets of qualitative parameters R1 and R2 can be compared side by side with plan-invoking macros on the blackboard. The blackboard will organize the robot's visual memory about local coordinate systems of landmarks as the primary means of defining locations, maintain memory structures that associate local landmark systems along paths of motion that the robot executed when it decided upon the landmarks [27].

Let us assume that in Eq.(1) $\langle R, T \rangle$ s are taken from a space of partially ordered bases. This gives freedom in forming the plan vector intelligently. The total geometric space configuration assumed in performing a job may then be seen as composed of a

series of subtasks. Each of these subtasks was constructed out of realizable <R,T>s. Thus two plans designed for two totally different jobs may be known as differing by additions, subtractions, modifications of some of the <R,T>s. Let us reference FIG 7.

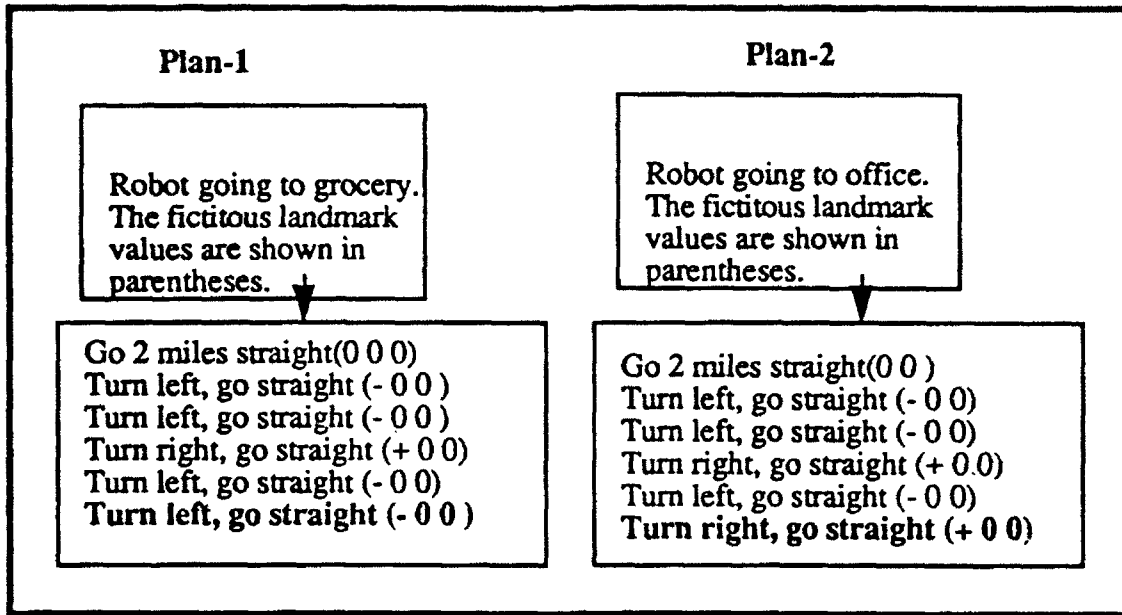


FIGURE 7

In one plan, say the robot is going to the grocery store, and in the other plan, it is going to the office. The <R,T>, where, these two plans differ is the final turn, as shown in the figure. Bases <R,T>s are partially ordered in both the plans. Therefore, if one plan is implemented successfully, and the other is not, a comparative diagnostic measure can be worked out, possibly searching common macros in both the cases.

However, one should be aware of possible troubles in partially ordered <R,T>s. It actually presupposes the existence of a strong table management system admonishing context dependent properties of <R,T>s. A comparative diagnostics of a faulty plan, or, an working plan showing bugs latter, requires all the macros in correct order invoking the two plans. If in the second plan, for example, a fault is observed in the last right turn, then a comparative diagnostics requires an account of the past history in correct order in both the plans. If the robot knows, how to go the grocery store (plan 1), then going to office (plan 2) needs a little modification in the final phase. Tracing the two plan vectors side by side with <R,T>s implementing them, is necessary in case of bugs observed in one or both of the plan execution. Such tracing also requires cataloging and comparing time of occurrences of all subtasks, their duration,etc. Obviously, this is a horrendous task.

8. Concluding remarks:

Qualitative reasoning of intelligent searching, matching./comparing, is a powerful tool for making decisions. The discipline actually forms a new generation of decision technology with machine intelligence. Massive amount of quantitative data already crippled many existing powerful intelligent systems. Since qualitative reasoning is model-based, discrete/selected arithmetic-based (landmark values), and is capable of performing simulation with incomplete knowledge, designing decision making tools with such reasoning is worth perusing. Such tools will require symbolic reasoning with qualitative calculus on arithmetic spaces of discrete values supported by envisionment and interpretations. Even though qualitative modeling admits lack of quantitative knowledge, libraries of prototypes with gradually increasing support from quantitative data defining generic task structures are needed. This makes decision making with machine intelligence more anthropomorphic.

Acknowledgement:

Members at the Logistic Research Branch of the AL/HRG, WPAFB, specially, Capt. Gary Worrall, provided immense help in completing this summer project. It is gratefully acknowledged. I am grateful to Gary Moore, Program Director, RDL, California, and Dan Danishek, RDL's on-site coordinator at the WPAFB, for their sincere interest in this work.

Note:

A detailed and rigorous discussion of the subject treated in this report has been done under the title, **QUALITATIVE PHYSICS FOR INTELLIGENT DECISIONS AND AUTOMATIC TECHNICAL MANUAL GENERATION**, and has been submitted for publication as a technical report to the Armstrong Laboratory, Human Resources Directorate, Logistic Research Division, Wright-Patterson Air Force Base, Dayton, Ohio. The report can be obtained from the division chief Bert Cream.

9. References

- [1] Davis, E. (1990). *Representation of commonsense knowledge*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [2]. Weld, D. and de Kleer, J. (1989). (Eds): *Qualitative physics about physical systems*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [3] Simon, H.A. (1952). Causal ordering and identifiability. In *Studies in Econometric Methods*. New York : Wiley International, 49-70.
- [4] Iwasaki, Y and Simon, H. (1986). Causality in device behavior. *Artificial Intelligence*. 29, 63-72.
- [5] de Kleer, J. and Brown, J.S. (1984). A qualitative physics based on confluences. *Artificial Intelligence..* 24, 7-83.

- [6] Reiger, C. and Grinberg, M. (1977). The declarative representation and procedural simulation of causality in physical mechanism. In the *Proceedings of Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)*, Cambridge, Mass.
- [7] Sembugamoorthy, V. and Chandrasekaran, B. (1986). Functional representation of device and compilation of diagnostic problem solving. In Kolodner, J. L. and Reitsbeck, C.K. (Eds): *Expertence, Memory and Reasoning*. New Jersey : Lawrence Erlbaum Associates. 47-73.
- [8] Hayes, P. (1985). Naive physics I : Ontology for liquids. In Hobbs, R. and Moore,R (Eds): *Formal Theories of Common Sense World*. Norwood : Albex Pub. Inc.
- [9] Forbus, K.D. (1984). Qualitative process theory. *Artificial Intelligence*. 24, 85-168.
- [10] Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence*. 29, 289-338.
- [11] Patil, R.S. Szolovits, P., and Schwartz, W. B. (1981). Causal understanding of patient illness in medical diagnosis. In the *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*. San Mateo : Morgan Kaufmann Pub. Inc.
- [12] Pople, H.E. Jr. (1982). Heuristic method for imposing structure on ill structured problems : the structuring of medical diagnosis. In P. Szolovits (Ed): *Artificial Intelligence in Medicine*. AAAS/Westview Press, Boulder, Colorado.
- [13] Kuipers, B. and Kassirer. (1984). Causal reasoning in medicine : analysis of a protocol. *Cognitive Science*., 8, 363-385.
- [14] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [15] Kuipers, B. (1989). Qualitative reasoning : modeling and simulation with incomplete knowledge. *Automatica*. 25, 571-585.
- [16] Simmons, R. (1986). "Commonsense " arithmetic reasoning. In the *Fifth National Conference on Artificial Intelligence (AAAI-86)*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [17] Hogge, J. (1987). Compiling plan operators from domains expressed in qualitative process theory. In the *Proceedings of the National Conference of Artificial Intelligence(AAAI-87)*. San Mateo(CA): Morgan Kaufmann Pub. Inc. 229-233.

- [18] Langlotz, C. and Shortliffe, E.H. (1989). Logical and decision theoretic methods for planning under uncertainty. *AI Magazine*. 10, 39-47.
- [19] Dean, T. and Wellman, M.P. (1991). *Planning and Control*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [20] Wellman, M.P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*. 44, 257-303.
- [21] Chandrasekaran, B. (1987). Towards a functional architecture for intelligence based on generic information processing tasks. In the *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [22] Bylander, T. and Chandrasekaran, B. (1987). Generic tasks for knowledge-based reasoning : the right level of abstraction for knowledge acquisition. *International Journal of Man-Machine Studies*. 26, 231-243.
- [23] Amarel, S., Despain, A., Nil, P., Steinberg, L., Tenenbaum, M., and Will, P. (1991). Panel on AI Design. In the *International Joint Conference on Artificial Intelligence (IJCAI-91)*. San Mateo (CA): Morgan Kaufmann Pub. Inc.
- [24] Chandrasekaran, B. (1990). Design problem solving : A task analysis. *AI Magazine*. 11, 71-59.
- [25] Dormoy, J-L. and Raiman, O. (1988). Assembling a device. In the *Seventh National Conference on Artificial Intelligence(AAAI-88)*. San Mateo (CA): Morgan Kaufmann Pub. Inc. 330-335.
- [26] Joskowicz, L. and Addanki, S. (1988). From kinematics to shape : an approach to innovative design. n the *Seventh National Conference on Artificial intelligence(AAAI-88)*. San Mateo (CA): Morgan Kaufmann Pub. Inc. 347-352.
- [27] Levitt, T.S. and Lawton, D.T. (1991). Qualitative Navigation for a mobile robot. *Artificial Intelligence*. 44, 305-360.
- [28] Das, A. (1987). Embedding intelligence in robot-automated operations. In the *First International Conference on the Application of Artificial Intelligence and Expert Systems in Engineering (AAI/ES-87)*. Tullahoma, TN. 1083-1088.
- [29] Das, A. (1989). Constructing a mutating plan vector for a learning robot. In the *Proceedings of the IEEE Southeastern Conference (IEEESECON-89)*. Columbus, SC.
- [30] Hayes-Roth, B. (1985). A blackboard architecture for control. *Artificial Intelligence*. 26, 251-321.

EFFECTS OF FEEDBACK DELAY ON TRACKING

Kent M. Daum, O.D., Ph.D.
Associate Professor
School and Department of Optometry

The University of Alabama at Birmingham
1716 University Boulevard
Birmingham, Alabama 35294-0010

Final Report for:
AFSOR Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 16, 1992

EFFECTS OF FEEDBACK DELAY ON TRACKING

Kent M. Daum, O.D., Ph.D.
Associate Professor
School and Department of Optometry
The University of Alabama at Birmingham

ABSTRACT

All flight simulators have delays in both the presentation and the use of information. In the fiber-optic helmet-mounted display (FOHMD), these delays are associated partly with the detection and measurement of eye position. In currently used simulators that measure head and eye position, the presentation delay in response to an eye movement has been measured at 90 to 140 ms. In addition to these delays, the human oculomotor system also has an inherent delay of about 150 ms. A large body of literature has shown that humans can track predictable stimuli with very small errors as long as the stimuli are within certain limits of amplitude, frequency and bandwidth. Most evidence suggests that the ability to track these targets is a result of an element within the feedback control system that predicts target motion and acts to eliminate delays which would otherwise drastically interfere with tracking and task performance. Although there has been substantial study of many of these factors, there has been little work aimed at assessing the effects of alterations of the amount of delay on tracking performance. This lack of information is critical in assessing tracking tasks with an additional delay such as is necessarily found in a simulator. It is also critical in determining simulator fidelity for any task involving eye movements. The objective of this project was to study the effects of feedback delay on tracking performance. We studied five subjects (mean age: 30.4 yrs; range: 21 to 40 yrs.). The subjects tracked a small green dot on a CRT while attempting to keep a laser spot (indicating their eye position) superimposed on the stimulus. The laser spot marked eye position with or without an additional delay. We examined the effects of feedback delays ranging from 0 to 225 ms. on sinusoidal motion (5 frequencies: 0.1 to 1.0 Hz.) and an unpredictable smooth stimulus with a bandwidth of 0.00 to 0.95 Hz. Each waveform and frequency was repeated 3 times. Both the mean number of saccades per second and the mean amplitude of the saccades increased as a function of the frequency being tracked and as a function of delay. Saccades generated while tracking the random smooth pursuit waveform occurred frequently (relative to the other waveform) and, as expected, were not affected by the presence of a feedback delay. The amplitude of saccades for this waveform fell about in the middle of the other frequencies, suggesting that the frequency composition of the random stimulus was well matched to the sinusoidal stimuli. These data suggest that the addition of modest delays (on the order of 75 ms.) result in significantly poorer tracking. They further imply that delays involving eye movements in the construction of flight simulators should be minimized. Further study is necessary to define the capabilities of the oculomotor system in dealing with such delays and in defining the parameters precisely responsible for the detrimental tracking effects.

EFFECTS OF FEEDBACK DELAY ON TRACKING

Kent M. Daum, O.D., Ph.D.

INTRODUCTION

High fidelity simulators play an important part in the training of aviators flying high performance aircraft. All simulators have delays in both the presentation and the use of information. In certain simulators (such as the fiber-optic helmet-mounted display {FOHMD}), these delays are associated in part with the detection and measurement of eye and head position. The response delay in the presentation of an updated stimulus to an eye movement in such a simulator has been measured to be 90 to 140 ms. (Wetzel, 1992) and delays of 200 ms. or more sometimes occur. There are also delays as a result of the transmission and analysis of the resulting information along with considerable image processing and presentation demands. Delays involving visual measurement and presentation are common in simulators. A realistic simulator has feedback delays which closely match those seen in the actual aircraft. Poor simulators typically have delays longer than those seen in the aircraft.

The human oculomotor system also has certain associated delays. The smooth pursuit system has an inherent delay of about 150 ms. (Rashbass, 1961). The saccadic system has a delay of a somewhat greater amount, about 250 ms., depending on the particular stimulus (Findlay, 1981). The fovea has a radius of about 0.5° (Bahill and McDonald, 1983). To obtain the best visual acuity (Barmack, 1970), it is necessary to track a target extremely precisely since both visual acuity and the resulting visual performance fall off steeply as a function of retinal eccentricity. In order to fixate moving targets within a fraction of a degree, the oculomotor system must completely compensate for any delays in its system (Barmack, 1970; Bahill and McDonald, 1983).

A large body of literature has shown that humans can track predictable stimuli with very small errors as long as the stimuli are within certain limits of amplitude, frequency and bandwidth (Bahill and McDonald, 1983; Kowler et al, 1984; Fender and Nye, 1961; Fuchs, 1967; Michaels and Jones, 1966). Although there have been suggestions to the contrary (St-Cyr and Fender, 1969; Lisberger et al, 1981), most evidence suggests that the ability to track these targets is a result of a feedback control element that predicts target motion and acts to eliminate delays which would otherwise drastically interfere with tracking and task performance (Dallos and Jones, 1963; Yasui and Young, 1975; Young, 1977; Stark et al, 1962; McDonald and Bahill, 1980).

There has been substantial study of many of these effects on the gain and phase of the tracking system response: frequency (Fuchs, 1967; St-Cyr and Fender, 1969), amplitude (Lisberger et al, 1981), velocity (Pola and Wyatt, 1980), acceleration (Lisberger et al, 1981), predictability (Michael and Jones, 1966; Kowler et al, 1984), type of feedback (Fender and Nye, 1961), type of cues (Mather and Lackner, 1980) and the nature of the waveform (Michael and Jones, 1966; Sugie, 1971). To date, however, there has been little work aimed at assessing the effects of alterations in the amount of delay on tracking performance (Brown, 1990; St-Cyr and

Fender, 1969). This lack of information is critical in assessing tracking tasks with an additional delay such as is necessarily found in a simulator.

OBJECTIVE(S)

The objective of this project is to study the effects of feedback delay on tracking performance. A corresponding theoretical objective is to study the characteristics of the predictor operator of the oculomotor pursuit system over a range of different waveforms and feedback delays.

SIGNIFICANCE

Since simulators using an area of interest (AOI) approach (such as the FOHMD) necessarily have delays, it is critical to ascertain the effects of such feedback delays on the oculomotor system in order to be certain that the simulator is as realistic as possible. This study will document the range of such delays that allow normal oculomotor performance under these conditions. Since there are a number of significant differences between the experimental arrangement used in this study and commonly used simulators, the study is not designed to provide data that will directly characterize what happens in a simulator with a large AOI display.

Also, the predictor operator is the most essential factor in the pursuit system. Within a certain range of parameters, this operator determines the degree of success or failure of any tracking task. Since the capabilities of this predictor have not been fully described, this study will provide a more complete understanding of this system.

EXPERIMENTAL DESIGN CONSIDERATIONS

We aim to examine waveforms, frequencies and delays that are representative of those seen in the simulator environment. The experimental arrangement will allow only horizontal eye movements and stimuli covering a total range of 20 degrees. Such limitations are not present in the simulator.

The experimental design also neglects the significance of the size of any AOI (typically 25° in extent; Wetzel et al, 1990) that may be in a simulator. It is likely that an AOI has an effect on eye movement characteristics, but, to date, no studies have examined this issue.

METHODS

Inclusion/Exclusion Criteria

We used the following inclusion/exclusion criteria: All subjects were volunteers. In addition, all subject's ages must be between 18 and 40 yrs. The upper age limit was to eliminate subjects who may not have been able to accommodate on the targets for the required length of time. The subjects must have had at least 20/20 visual acuity in the eye to be used for tracking at the distance of the task

(34 cm.) without correction or with contact lenses. This criterion was to eliminate tracking difficulties due to poor acuity due to refractive errors or otherwise. Finally, all subjects must have reported that they did not have any significant visual problems and were free of known ocular disease. The protocol used in this experiment was approved by the U.S. Air Force Armstrong Laboratories Institutional Review Board.

Eye Movement Measurement

Eye movements of the subjects were assessed using an infrared limbus tracker technique that has been previously described (Wetzel, 1988; Young and Sheena, 1975). All measurements were on the left eye; the other eye was occluded during the experiment. The eye was illuminated with infrared light using a low level LED source. A pair of phototransistors balanced about the limbus detected movements of the eye as a result of the imbalance of reflected light (Young and Sheena, 1975; Truong and Feldon, 1987).

This apparatus was able to detect changes in eye position of less than 0.1 degrees; had a maximum sensitivity of 1.7 arc-minutes resolution or more; and, was highly linear over the $\pm 10^\circ$ range used in the experiment (Wetzel, 1988; Young and Sheena, 1975). All eye movements were assessed at 500 Hz. All subjects used a bite bar and head and chin rests throughout the experiment.

Blinks

All records were evaluated before analysis. Movements considered blinks and were edited from the eye movement trace. In the determination of saccadic counts, amplitudes and other basic data concerning eye movement performance, the periods of time where blinks occurred were not used in the analysis. In calculations of gain and phase, portions with blinks will be replaced by a line connecting the eye position at the times just before and after the movement.

Calibration procedures

Initially in the calibration procedure the phototransistors and the infrared source were carefully positioned about 10 mm. from the cornea with a slight upward angle of view (Truong and Feldon, 1987). After the experimenter set the zero and gains, a computer-controlled algorithm was used to assess the gain, slope and intercept relationship of the voltage output from the phototransistors to the angle of gaze over the experimental range using a sample of points on both sides of 0 at 2.5, 5.0, 7.5 and 10 degrees. The best fitting least squares equation was used to determine the appropriate calibration coefficients. We used three criteria for linearity: The first was a correlation coefficient (r) of 0.9993 or greater. At this level, r is significant at the 0.1% level (Diem, 1962). The second was that the stability of the eye must be less than or equal to 5 arc min. at each position. The third was that the calculated residual

errors at each position must be less than 0.50° . The calibration was repeated before each set of data was taken (approximately every 15 to 30 sec of data). The calibration trial lasted 13.50 sec. and took on the order of a minute to complete once the subject was arranged in the apparatus and the initial calibration adjustments were completed. Experimental trials were initiated within seconds after a successful calibration.

Tracking Task

Through a beam splitter, the subject tracked a small green dot moving about the face of a CRT (Tektronix model 602) which was 34 cm away from the corneal pole. This distance was chosen so that the face of the CRT would subtend a total of 20 degrees ($\pm 10^\circ$).

Feedback Techniques

The signal indicating the position of the subject's eye was sent to a horizontal position motor which (via a mirror) controlled the position of a small attenuated laser spot. This spot was projected from the rear onto a translucent screen located 1 m. in front of the subject. With zero delay, the subject saw a small red spot of light corresponding to his/her eye position. When a feedback delay was instituted, the subject saw the red laser spot corresponding to his/her eye position at the present time minus the amount of the delay. That is, the subject saw the spot corresponding to where his/her eyes were 200 ms. ago if the delay was 200 ms. The subject's task was to keep the laser spot indicating eye position superimposed on the green dot which was to be tracked. When this was accomplished, the subject had negated the effects of delay in the system.

Waveforms, Frequencies and Delays

Each subject tracked two different types of waveforms (sinusoid and random smooth pursuit; Table 1). One of the waveforms (sine) had five associated frequencies (0.1, 0.3, 0.5, 0.7 and 1.0 Hz.). The random smooth pursuit waveform was constructed by filtering the output of a random noise generator. Consequently, the random smooth pursuit waveform was composed of frequencies from 0.0 to 0.95 Hz. The spectral composition of this waveform is shown in figure 1.

The feedback presented to the subject for each waveform and frequency had one of five delay steps (an absence of feedback being one). There were three trials for each waveform, frequency and delay. Each subject tracked a total of 90 calibration and 90 experimental trials. At eighteen trials per session, each subject spent five sessions completing the experiment. We used a total of five subjects in the experiment.

Typically, data taking sessions lasted from about 25 to 30 minutes to about an hour. The subjects were constantly reminded of the specific practical objective of the experiment: to keep the laser

spot corresponding to their eye position superimposed on the green dot indicating the target on the CRT. The subjects also were encouraged to take breaks as frequently as necessary to remain fresh over the course of the session. Generally, subjects took a 10 to 15 minute break midway through each data taking session.

Sample Size

We considered changes in eye position of 0.05 to 0.10° (3 to 6 arc min or slightly less) significant. Steady fixation by a good observer is slightly below this level. Since there have not been studies where the feedback tracking delay has been systematically altered, there are no estimates of the variance of the various components that may result. Our criterion for Type I errors was 0.05.

Bias Control/ Masking

We used a Latin square procedure to determine the presentation order of each particular waveform (with its associated frequency and delay). The data portion of the experiment was divided into five sessions where 18 different waveforms (along with the calibration trials) were presented. Each waveform, frequency and delay was presented three times. All of the waveforms, frequencies and delays were presented in each third of the data taking.

The subjects were informed only of the general nature of the experiments. None of the subjects were informed of the type of waveform or the delay before a trial began.

Analysis Techniques

We used a variety of criteria for recognizing a saccadic movement (Table 2). These included trend: The movement must continue in the same direction for a minimum length of time (5 to 16 ms). Another criteria used for all trials was velocity threshold: This varied from 30 to $60^\circ/\text{sec}$. In order to discriminate saccadic movements for the higher frequencies (0.7 and 1 Hz), we used an additional acceleration threshold, $1500^\circ/\text{sec}^2$. We established these criteria after extensive, careful examination of the records by hand. In a small number of instances, portions of the records suggested as being saccadic were examined and edited after saccadic recognition programs were completed. In most cases, these were instances where the velocity/amplitude relationship did not fit the main sequence for saccades.

We analyzed the effects of delay on tracking in several different ways. Our objective was to determine the extent tracking was effected by the delays, and to determine the nature of the differences. Poorer tracking results in a higher number of saccades. In addition, the mean saccadic amplitude should increase when the subject has difficulty tracking.

RESULTS

Five subjects participated in the experiment (three females and two males). The mean age of the subjects was 30.4 yrs. (std. dev. = 9.1 yrs.; range: 21 to 40 yrs.). All of the subjects except one had previously participated in eye tracking experiments.

We recorded a total of 900 trials. Four hundred and fifty of these trials were calibration trials and were not included in the analysis. Each subject completed a total of 90 trials used for analysis. These included 3 trials of 5 frequencies with 5 delay states for each sinusoid. In addition, the subjects tracked the random smooth pursuit for 3 trials with the same 5 delay states. We discarded two trials from one subject and one trial from another because of excessive blinking and/or saturations in the recording. The resultant 447 trials form the basis for this work.

Figure 2 shows three examples of the tracking by one of the subjects. The figure shows tracking of a 0.1, 0.3 and 0.5 Hz. targets with feedback delays of 0, 0 and 150 ms., respectively. In these cases, the subject has tracked both the 0.1 and 0.3 Hz. targets very well: The largest positional errors are less than 1° and there is no phase lag. A significant phase lag is present with the 0.5 Hz. stimulus. The eye is lagging behind the stimulus by about 190 ms. at the beginning of the trace and is lagging about 130 ms. behind at the end of the trace. Such directional asymmetries were relatively common. Directional tracking asymmetry has been previously reported (Meyer et al, 1985).

The effect of delay on mean saccadic amplitude is shown in figure 3. For the sinusoidal waveform, the addition of a delay in the feedback loop resulted in an increased mean saccadic amplitude for each frequency. As frequency increased, the delay had a more significant effect on the amplitude. For the 1 Hz. and 0.5 Hz. frequencies there is evidence of a saturation effect: The mean saccadic amplitude was smaller with a delay of 225 ms. than it was at 150 ms.

For each frequency and waveform, the case where feedback was not provided resulted in a smaller mean saccadic amplitude. For the sine wave, the mean saccadic amplitude was a function of frequency with the faster frequencies having a greater difference between the no feedback and zero (or correct) feedback states.

Delays did not affect the random smooth pursuit waveform the same way as the predictable sinusoidal waves. As the delay increased for the random waveform, the mean saccadic amplitude remained essentially the same. The mean saccadic amplitude for the random waveform was about in the middle of those found with the various sinusoidal waveforms since the random waveform was composed of sinusoidal waveforms spanning the same range.

There was a directional effect to tracking with a range of feedback delays (figure 4). Over the waveforms tracked in this experiment, mean saccadic amplitudes were consistently larger going from left to right (temporal to nasal) than tracking from right to left (nasal to temporal). This trend occurred even for the case in which no feedback was provided. This perspective on the data is useful because it shows that saturation effects occur at a smaller delay than is apparent with the summed data (as shown in figure 3). With each waveform

(and frequency) the mean saccadic amplitude from delays of 75 ms. and up is roughly level or it declines. An exception to this is the random waveform; The addition of a delay has an insignificant effect on saccadic amplitude.

Feedback delay also has an effect on the number of saccades made per unit time (figure 5). There is a definite trend toward more frequent saccades as the feedback delay time is increased. As in the other cases, the random smooth waveform is an exception: The addition of delay has absolutely no effect on the number of saccades made in a given time. Correlated with the difficulty in tracking such an unpredictable waveform, there were considerably more saccades made per unit time with the random smooth waveform than with any of the others.

Figure 6 shows these data in a way which emphasizes the effect of frequency on the amplitude of saccades. The number of saccades per second increased steadily from 0.1 to 0.5 Hz. At 0.7 Hz. the number of saccades decreased considerably before increasing again at 1.0 Hz.

Our subjects were consistently able to achieve extremely high pursuit velocities on the order of 50 to 60°/sec. They were capable of essentially matching the peak velocity of the 1.0 Hz. stimulus (63°/sec.) for extended periods of time (several seconds).

DISCUSSION

These data suggest that the addition of any additional delay to the oculomotor feedback system has a pervasive detrimental effect on oculomotor tracking performance. Additional feedback delay consistently results in more and larger saccades than when there is none or when the delay is appropriate.

With regard to the oculomotor predictor system, this study suggests that the visual system has relatively little capability of immediately dealing with such additional delays and that, at a minimum, the system needs substantial learning time. This study was specifically designed to minimize learning since all trials were randomized and masked. We will be able to examine two aspects of these data which will allow some judgement of the amount of learning/prediction which took place in this study. One method is to compare the positional errors for the trials for the initial vs. the final seconds of the trials. Another method is to compare the positional errors (or some similar aspect such as gain or phase) from the first to the second and third trials. If learning or prediction is taking place to a noticeable extent the errors in both of these methods should diminish over time and trial.

Many oculomotor adaptations involving vergence (a disconjugate movement) such as prism adaptation (Carter, 1963) occur over short time periods (seconds to minutes). Other adaptations involving conjugate eye movements (vertical pursuits) have been shown to take considerably longer, on the order of an hour of exposure or more (Gleason et al, 1991). In order to fully explore the extent that the oculomotor system can learn to compensate for additional delays, protocols allowing substantially longer and more consistent exposure to the delays must be developed, carried out and analyzed.

The results of this paper must be interpreted cautiously with regard to their practical application in simulators. Clearly, in and of itself, substantial delay in the feedback system of a simulator above that which is typical to the actual aircraft is detrimental and should be minimized where possible. This study does not provide any information about the *practical* significance of such delays in a simulator environment.

Essentially, the feedback system used in this experiment mimicked a design with an area of interest that was very small, approaching zero. The addition of an AOI of substantial size should minimize the effects of delay. The relationship of the size of the AOI and the nature of any feedback delay to eye movements made and to the effect on task performance is undetermined.

There are a number of other questions about the effects of feedback delay that remain unanswered. One issue concerns the algorithms used to discriminate saccades. As shown in Table 2, we customized the analysis programs for each waveform to effectively recognize saccades. We did this because of the high pursuit velocities reached in the 0.7 and 1.0 Hz. waveforms. Data shown in figure 6, however, suggest that the number of saccades made per second decreased at 0.7 Hz. below what it was at 0.5 Hz. Presently, we are unaware of any physiological mechanism that might explain this result. To justify the result of the algorithm, it will be necessary to manually analyze the saccades in each record. In addition, statistical analysis of the significance of changes in the number or amplitude of saccades made in these various conditions remains to be completed.

Considerable other analysis of the data in this experiment remains. As noted above, its important that the accuracy of the analytic algorithm which recognizes saccades be confirmed. This is critical to providing confidence that the data is sound. In addition, an analysis of positional errors with each of the delay conditions will allow a judgement of whether tracking was poorer in one case than another. In addition, gain and phase relationships between the target and the eye are standard methods of examining tracking ability. Although these have not been completed at this time, they will be important in understanding the nature of the oculomotor response to the delay because they are a global view of the data. If the visual system is adapting to the delay, then the phase relationships of the eye to the target should be nearly the same as the delay is changed. If the additional delays impair tracking then the gain of the response should decrease. Further analysis of the percentage of time spent in smooth pursuit as a function of waveform and delay also remains. The greater the time spent in pursuit, the better the eye is able to track the target. If the additional delays affect tracking then the time spent in smooth pursuit should decrease.

We also plan to examine the nature of individual differences in tracking in the presence of these delay conditions. Understanding these differences is important because they are important in estimating the significance of the changes in the data as a function of delays. In addition, they will also provide insight into the range of response to additional delay in the tracking ability of the population. This is important because its possible that some people may be able to deal with delays that bother others.

The extremely high pursuit velocities seen in this experiment have been reported previously (Meyer et

al, 1985). These investigators reported that eye velocity was about 90% of that of the target up to target velocities of about 100°/sec. Meyer et al's work assessed pursuit by hand, used an EOG apparatus and unidirectional tracking (for a particular trial). The subjects in this experiment sustained their tracking for a much longer period of time (several seconds vs. about 450 ms.) and the calibration process and the method of differentiating saccades from pursuit was much more rigorous.

REFERENCES

- Bahill AT, McDonald JD. The smooth pursuit eye movement system uses an adaptive controller to track predictable targets. *IEEE*: 274-278, 1981.
- Bahill AT, McDonald JD. Model emulates human smooth pursuit system producing zero-latency target tracking. *Biol Cybernetics* 48: 213-222, 1983.
- Bahill AT, McDonald JD. Smooth pursuit eye movements in response to predictable target motions. *Vision Res* 23(12): 1573-1583, 1983.
- Barmack NH. Dynamic visual acuity as an index of eye movement control. *Vision Res* 10: 1377-1391, 1970.
- Brown C. Gaze controls with interactions and delays. *IEEE Trans on Systems, Man and Cybernetics* 20(1): 518-527, 1990.
- Carter DB. Effect of prolonged wearing of prisms. 40: 265-273, 1963.
- Dallos PJ, Jones RW. Learning behavior of the eye fixation control system. *IEEE Trans on Automatic Control* AC8: 218-227, 1963.
- Deno DC, Crandall WF, Sherman K, Keller EL. Characterization of prediction in the primate visual smooth pursuit system. *Invest Ophthalmol Vision Sci suppl*.31(4): 532, 1990.
- Diem K. *Documenta Geigy Scientific Tables*, 6th ed. Geigy Pharmaceuticals, Ardsley, New York, p. 66.
- Engelkin EJ, Wolfe JW. A modeling approach to the assessment of smooth pursuit eye movement. *Aviat Space Environ Med* 50: 1102-1107, 1979.
- Fender DH, Nye PW. An investigation of the mechanisms of eye movement control. *Kybernetik* 1: 81-88, 1961.
- Findlay JM. Spatial and temporal factors in the predictive generation of saccadic eye movements. *Vision Res* 21: 347-354, 1981.
- Fuchs AF. Periodic eye tracking in the monkey. *J Physiol* 193: 161-171, 1967.
- Gleason GA, Lunn R, Scar CM. Direction-specific and position-specific mechanisms underlie nonconjugate adaptation of vertical pursuits. *Invest Ophthalmol Visual Sci suppl* 32(4): 896, 1991.
- Greene DE, Ward FE. Human eye tracking as a sequential input adaptive process. *Biol Cybernetics* 33: 1-7, 1979.
- Kowler E, Steinman RM. The effect of expectations on slow oculomotor control: III. Guessing unpredictable

target displacements. *Vision Res* 21: 191-203, 1981.

Kowler E, Martins AJ, Pavel M. The effect of expectations on slow oculomotor control: IV. Anticipatory smooth eye movements depend on prior target motions. *Vision Res* 24(3): 197-210, 1984.

Lisberger SG, Evinger C, Johanson GW, Fuchs AF. Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit in man and monkey. *J Neurophysiol* 46(2): 229-249, 1981.

Mather JA, Lackner JR. Multiple sensory and motor cues enhance the accuracy of pursuit eye movements. *Aviat Space Environ Med* __: 856- 859, 1980.

McDonald JD, Bahill AT. Adaptive control model for the human smooth pursuit system. In: Vogt W, Mickle M, eds. *Modeling and Simulation, Proc Eleventh Annual Modeling and Simulation Conference*, Pittsburgh, PA., Instrument Society of America, Pittsburgh, 283-287, 1980.

Meyer CH, Lasker AG, Robinson DA. The upper limit of human smooth pursuit velocity. *Vision Res* 25(4): 561-563, 1985.

Michael JA, Jones GM. Dependence of the visual tracking capability upon stimulus predictability. *Vision Res* 6: 707-716, 1966.

Pola J, Wyatt HJ. Target position and velocity: The stimuli for smooth pursuit eye movements. *Vision Res* 20: 523-534, 1980.

Stark L, Vossius G, Young LR. Predictive control of eye tracking movements. *IEEE Trans Human Factors Electron HFE3*: 52-57, 1962.

St-Cyr GJ, Fender DH. Nonlinearities of the human oculomotor system: Gain. *Vision Res* 9: 1235-1246, 1969.

St-Cyr GJ, Fender DH. Nonlinearities of the human oculomotor system: Time delays. *Vision Res* 9: 1491-1503, 1969.

Sugie N. A model of predictive control in visual target tracking. *IEEE Trans on Systems, Man and Cybernetics SMC1*(1): 2-7, 1971.

Truong DM, Feldon SE. Sources of artifact in infrared recording of eye movement. *Invest Ophthalmol Vis Sci* 28: 1018-1022, 1987.

Wetzel PA. Error reduction strategies in the oculomotor control system. Dissertation University of Illinois at Chicago, Chicago, Illinois, 1983.

Wetzel PA. Personal communication. 1992.

Wetzel PA, Thomal ML, Williams TT. Development and evaluation of eye tracker performance for use with the fiber optic helmet mounted display. *Proc 12th Interservice/Industry Training Systems Conference*, November 6-8, 1990, pp. 273-280.

Yasui S, Young LR. Perceived visual motion as an effective stimulus to pursuit eye movement system. *Science* 190: 906-908, 1975.

Young L. Pursuit eye movements: What is being pursued? In: Baker R, Berthoz A, eds. *Control of Gaze by*

Brain Stem Neurons. Elsevier Press, Amsterdam, 1977.

Young LR, Sheena D. Survey of eye movement recording methods. *Behav Research Methods Instrum* 7(5): 397-429, 1975.

FIGURE LEGENDS

- Figure 1: Fast-fourier analysis of the spectral composition of the random waveform over the bandwidth from 0 to 3.0 Hz.
- Figure 2: Examples of the tracking of one of the subjects. For each of the figures parts, up is to the right and down is to the left. The total amplitude of the stimulus was 20°, only portions of the total amplitude are shown. From the left of the figures to the right is 1278 ms. a. The stimulus was moving at 0.1 Hz. and feedback delay was 0 (i.e. The laser spot indicated actual eye position). The small saccade to the right was 1.16° in amplitude lasting 30 ms. b. The stimulus was moving at 0.3 Hz. again with 0 feedback delay. The cursors to the left are 100 ms. apart. The positional errors at the two cursors are 0.99 and 0.82 degree, respectively. c. The stimulus was moving at 0.5 Hz. and the feedback delay was 150 ms. The saccade to the right had an amplitude of 1.80° lasting 20 ms. There is considerable phase lag with the eye about 190 ms. behind the stimulus on the right and about 130 ms. behind the stimulus on the left.
- Figure 3: Effect of amount of feedback delay on the mean absolute saccadic amplitude as a function of various sinusoidal frequencies and the random waveform (RND) for the combined group of subjects.
- Figure 4: Effect of direction of tracking and the amount of feedback delay on the mean absolute saccadic amplitude as a function of various sinusoidal frequencies and the random waveform (RND) for the combined group of subjects. Since all subjects tracked with their left eye, right to left is temporal to nasal and left to right is nasal to temporal.
- Figure 5: The number of saccades per second as a function of the amount of feedback delay for each of the sinusoidal frequencies and the random waveform (RND).
- Figure 6: The number of saccades per second as a function of the sinusoidal frequency and the random waveform (RND) as a function of the amount of feedback delay,

Table 1: Details of the trials

WAVEFORM	FREQUENCY (Hz) AMPLITUDE (°) AVERAGE/PEAK VELOCITY (°/sec)	DELAY	SECONDS PER TRIAL	NUMBER OF TRIALS
sine	0.1 10 4.0/6.0	none ^a 0 75 150 225	30.00	3 each
sine	0.3 10 12/18.4	none ^a 0 75 150 225	30.00	3 each
sine	0.5 10 20/31	none ^a 0 75 150 225	30.00	3 each
sine	0.7 10 28/43	none ^a 0 75 150 225	14.28	3 each
sine	1.0 10 40/62	none ^a 0 75 150 225	15.00	3 each
random smooth waveform	bandwidth: 0.0 to 0.95 10 varies	none ^a 0 75 150 225	32.12	3 each

^a None means that there was no feedback provided to the subject; The laser feedback spot was turned off.

Table 2: Criteria used for recognizing saccades

FREQUENCY (Hz)	VELOCITY (°/sec)	ACCELERATION (°/sec ²)	MINIMUM DURATION (ms)
<i>0.1</i>	30	-	8
<i>0.3</i>	30	-	10
<i>0.5</i>	40	-	16
<i>0.7</i>	50	1500	10
<i>1.0</i>	60	1500	10
<i>random</i>	30	-	10

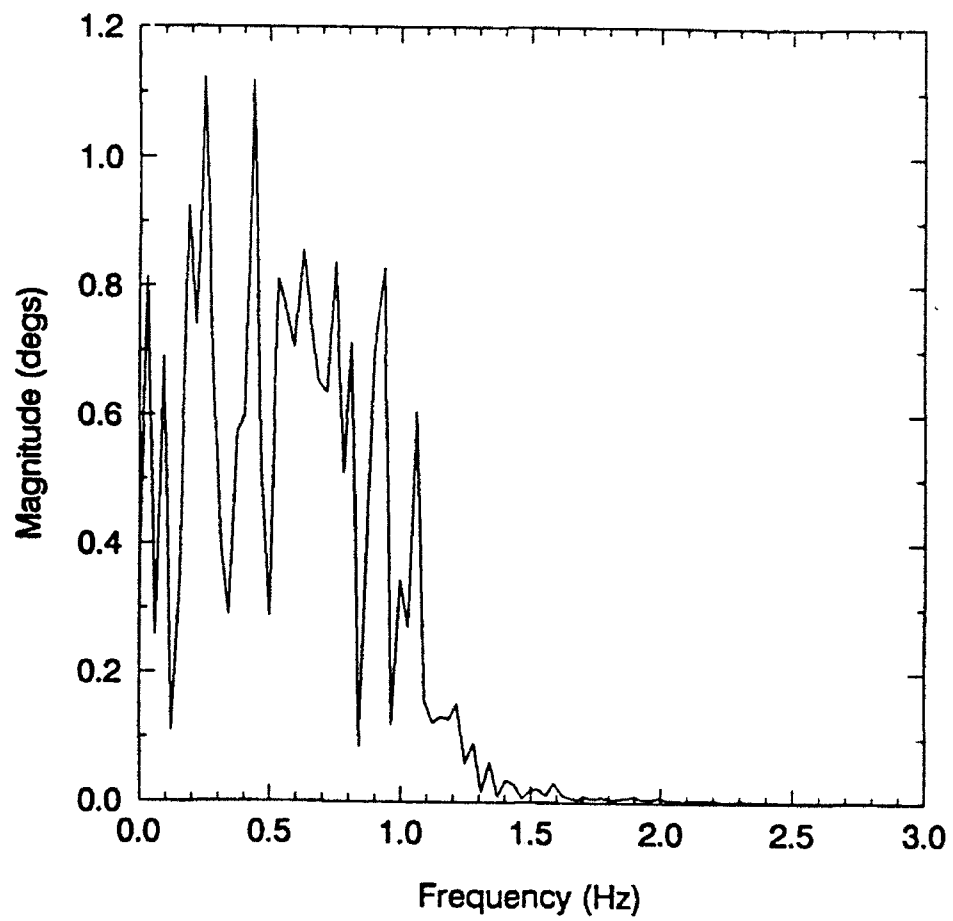
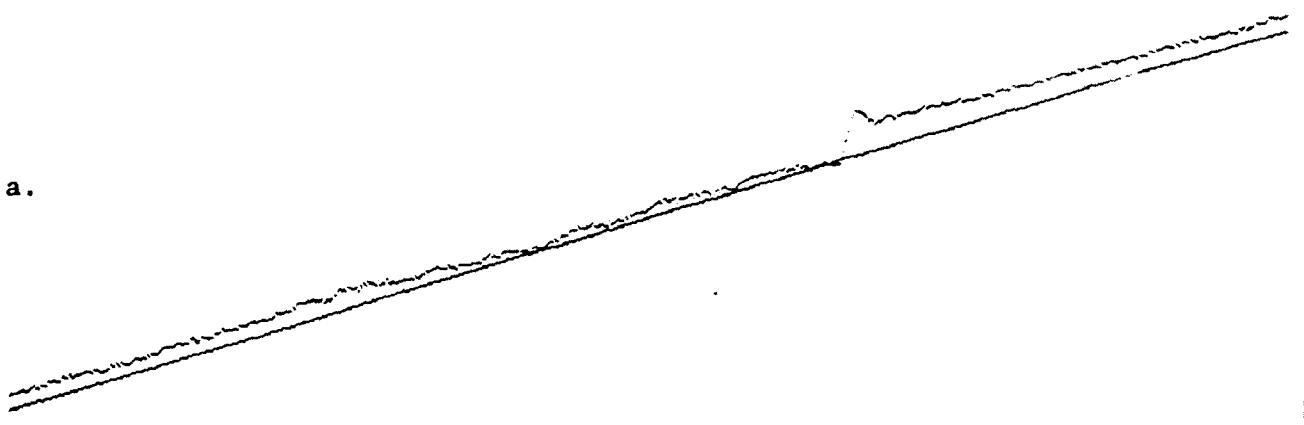
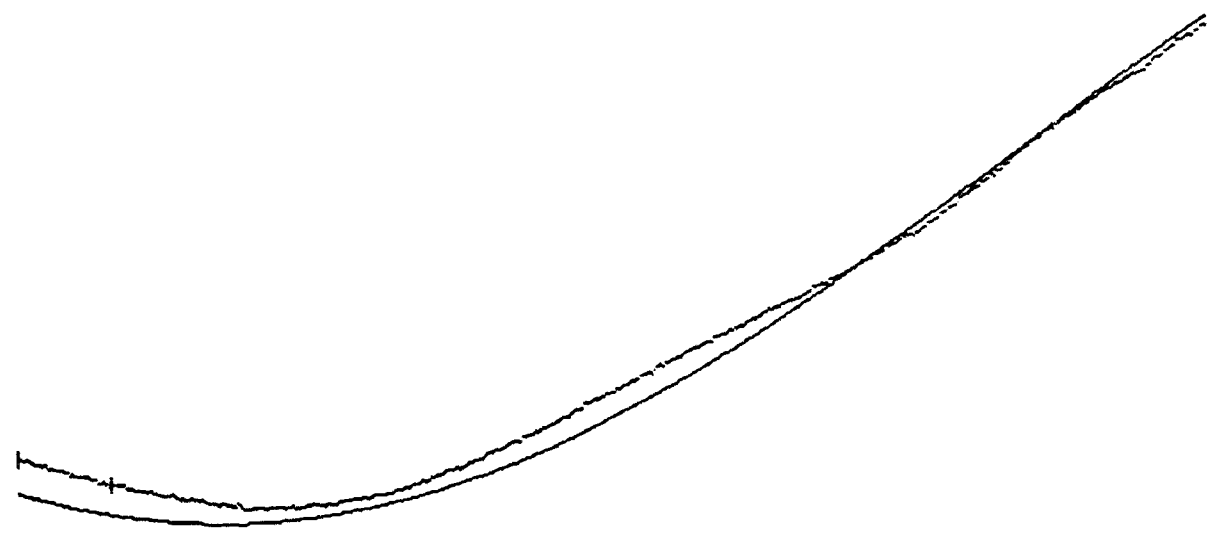


Figure 1

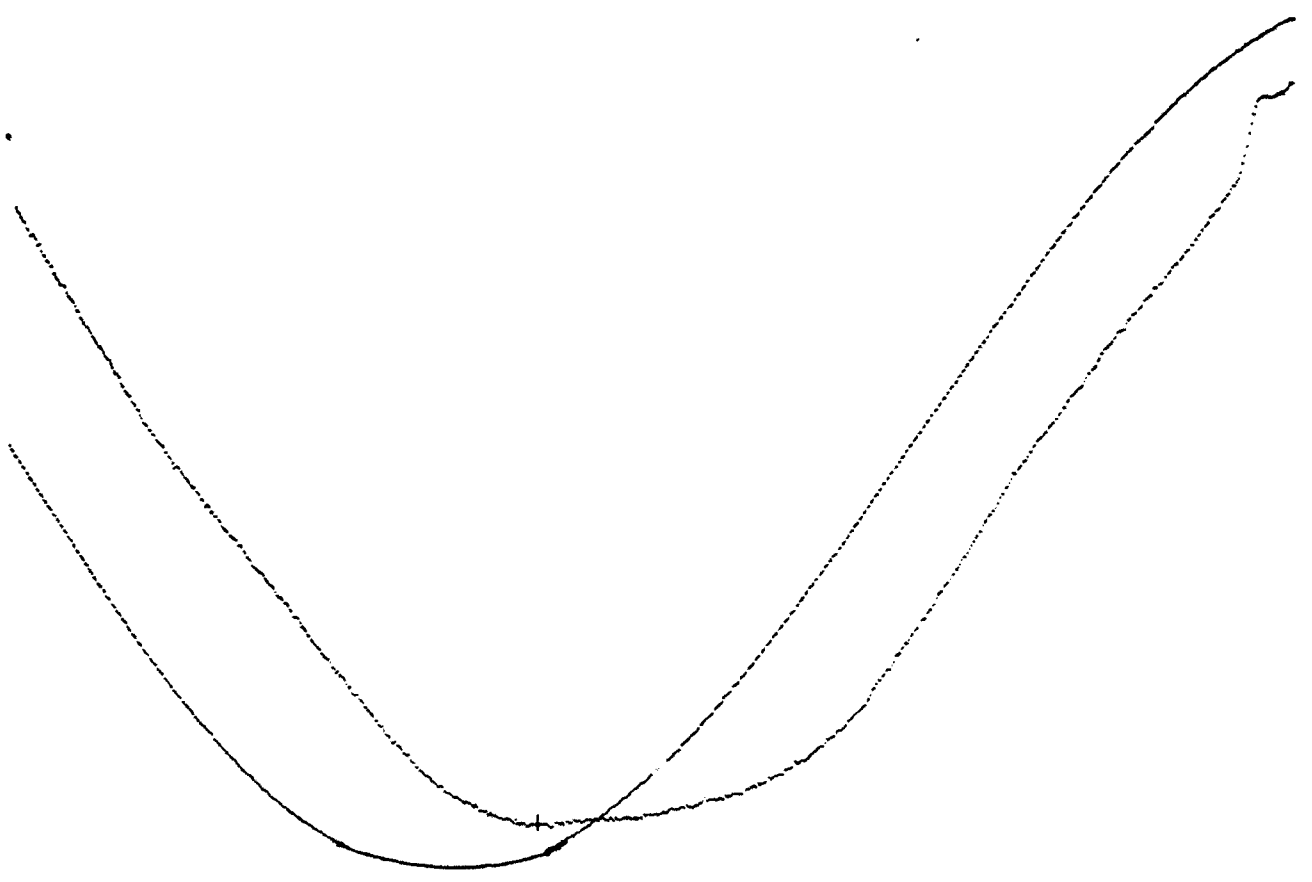
a.



b.



c.



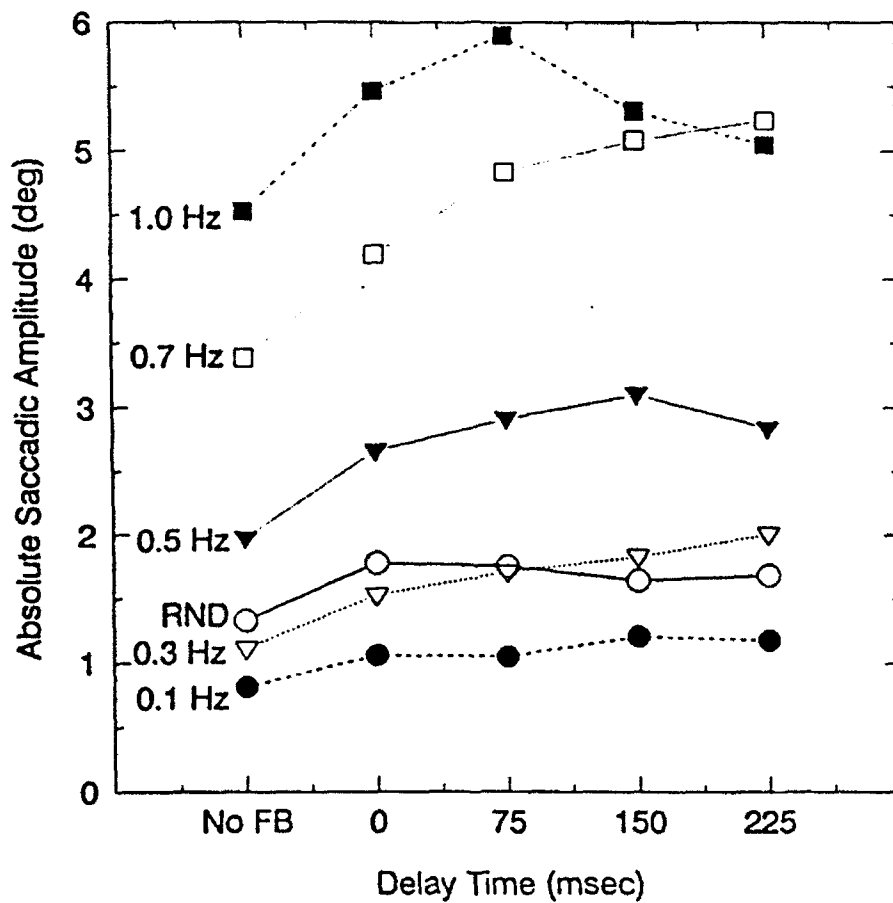


figure 3

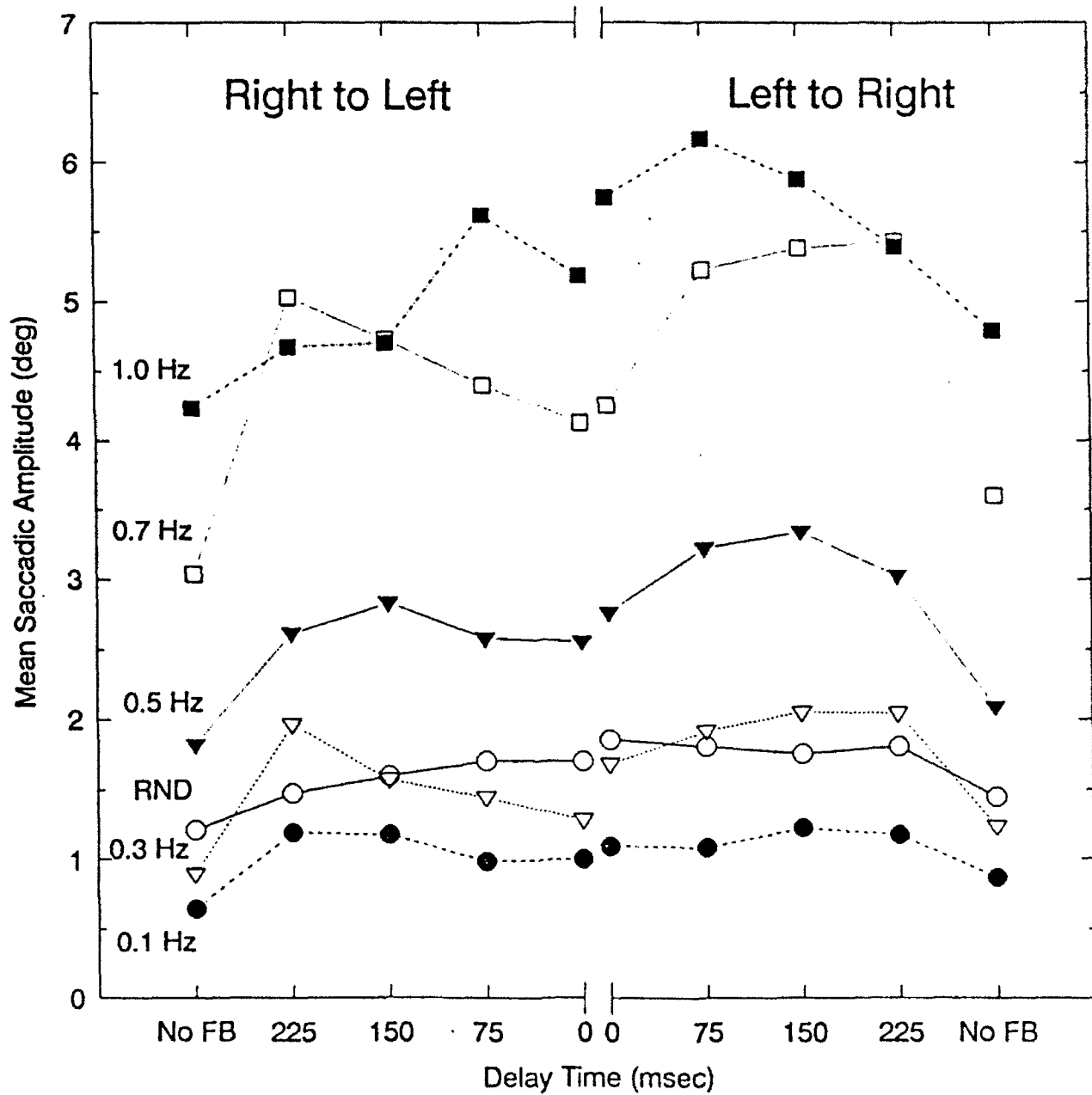


figure 4

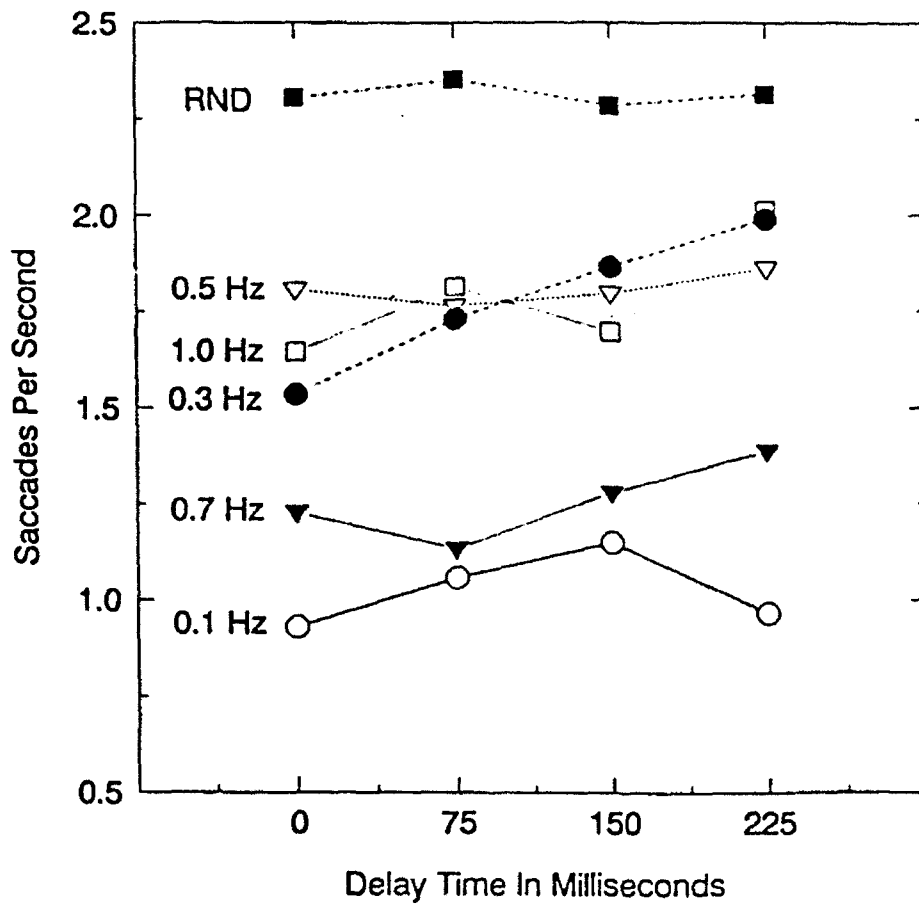


figure 5

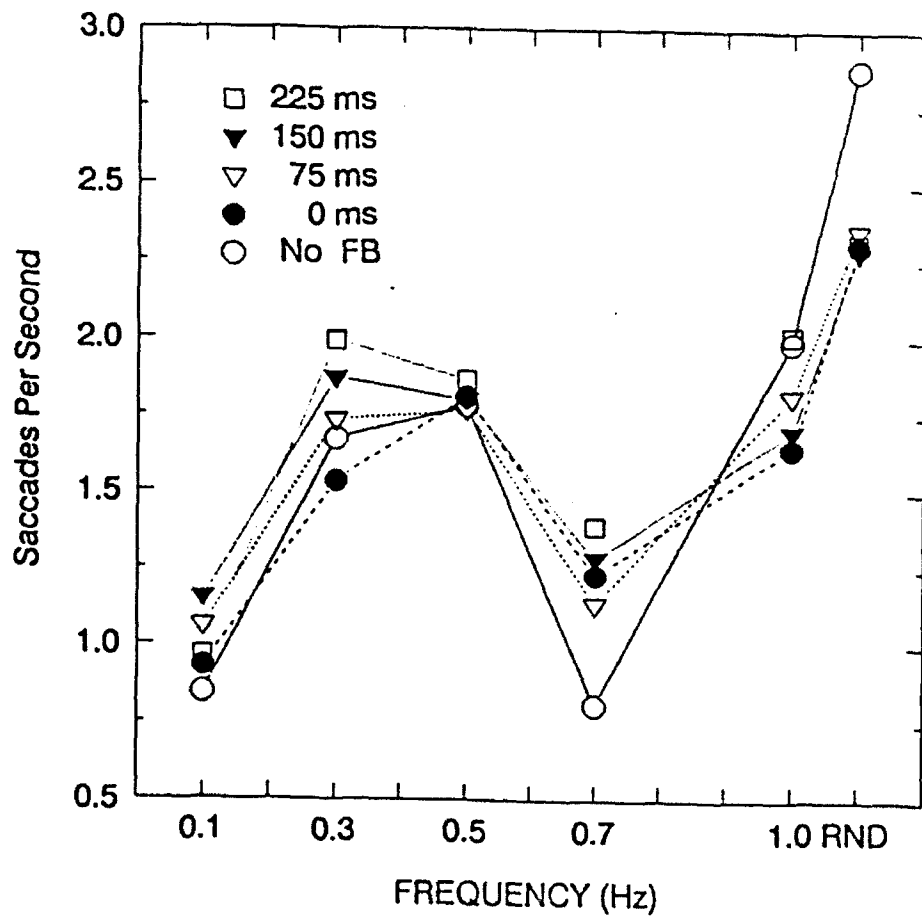


figure 6

HAND TORQUE STRENGTH FOR SMALL FASTENERS

S. Deivanayagam
Professor of Industrial Engineering

Tennessee Technological University
Cookeville, TN 39501

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Boling Air Force Base, Washington, D.C.

August 1992

10-1

Hand Torque Strength for Small Fasteners

S. Deivanayagam
Professor of Industrial Engineering
Tennessee Technological University

Abstract

This report describes an experimental investigation on the hand torque strength capabilities of persons for tightening and loosening small fasteners such as wing-nuts, keys and Allen screws. During the typical maintenance activities of modern systems the technician often is required to perform such tasks. This study used 14 males and 14 females as subjects and each subject performed 72 test exertions under simulated task conditions in the laboratory. The results will be useful as a data base for computerized human models such as CREW CHIEF and DEPTH for modeling ergonomic capabilities of persons.

I. INTRODUCTION

Maintainability is a very important characteristic that influences the system effectiveness and the life cycle cost of any modern system. A system normally is designed and put into operation with the objective of fulfilling certain goals. Such a system must be in operable condition for it to function effectively and meet the goals as desired. When the system is ready to operate on demand, it is said to be "available". The availability of the system is a function of its reliability and maintainability. Simply put, maintainability refers to the ease with which an unavailable (not ready) system can be put back in "ready" condition for operation. A system that is being serviced will not be ready for operation during the time it is being serviced.

In recent times, many modern weapon systems that are highly sophisticated and complex have been found, in practice, to have low maintainability characteristics. The reason for this situation can be attributed to the insufficient consideration of maintainability during the practice, most often, is not any willful negligence on the part of the designer, but the absence of appropriate tools and techniques to include maintainability considerations in the early stages of system design. Realizing this deficiency, the United States Air Force (USAF) initiated several programs during the last decade with the objective of developing design tools to improve the maintainability of its weapon systems. CREW CHIEF is one such program that is being developed by Armstrong Laboratory at Wright-Patterson Air Force Base, Ohio. CREW CHIEF is a computerized man model of the USAF maintenance technician performing aircraft flight line maintenance functions. The CREW CHIEF model, among other things, has extensive data bases on human strength capabilities for torquing with wrenches and materials handling. However, the model, currently does not have strength data on torquing small objects bare hands. The research described in this report was undertaken to gather torque strength data for turning certain types of small fasteners and tools with fingers under conditions representing typical maintenance tasks.

During the installation, removal, replacing and other such maintenance activities the technician often uses his/her hand to tighten or loosen small fasteners such as wing-nuts, Allen screws, or flat key-like devices. The wing-nut and key-like device is usually held in position with fingers and turned clockwise or counter-clockwise in order to effect a tightening or loosening action. The Allen wrench, an L-shaped hexagonal bar, is held with the fingers, mated to the matching recess in the fastener and then turned in order to effect the desired action.

Objective and Scope of the Research

The objective of this study was to determine the torquing capabilities of a person for tightening and loosening small fasteners such as wing-nuts and keys with bare hands and small Allen screws with Allen wrenches. The torque strength capabilities under a variety of simulated task conditions were gathered using human subjects in a laboratory setting.

The results of this study will be of help to the designers of modern complex systems in the selection and location of small fasteners while considering maintainability along with the other usual criteria. The results could also be incorporated, after normalizing for the USAF maintenance technician population, into the data bases for computerized human models (viz., CREW CHIEF and DEPTH) that are being developed for the use of designers.

In many modern hardware systems, small fasteners are used to a significant level to hold and attach small parts, components, and panels to larger units. During the maintenance of the part or component, the technician will be required to remove and replace such fasteners. The fasteners typically require torquing actions involving the muscles and joints of the upper extremity. The small fastener or the small tool (Allen wrench) is usually grasped and held by the fingers while a turning moment is applied on the surface so held. This is typically a major difference in turning a small fastener and a large fastener. In the case of a large fastener the tool (wrench, etc.) is held by the entire hand and the

muscles of upper arm and shoulder provide much of the power. In the case of small fasteners the muscles of hand and fore-arm are responsible for most of the power for the turning action. Further the hand-tool contact area for the large fastener is generally much larger than the finger-fastener contact area for the small fastener or tool. Therefore the force per unit area of the contact surface can be much higher in the case of a small fastener leading to more discomfort and/or pain to the person.

The torquing capability of a person working on a small fastener will depend to a large extent on the type of finger grip, which in turn depends on the size, shape, location, orientation, etc. of the fastener. Therefore it becomes necessary for the designer to be knowledgeable regarding the hand torquing capabilities of the technician population in order to select and locate the ergonomically appropriate/acceptable small fastener. The task of tightening or loosening will then be performed by the technician with lesser effort and fatigue and will ultimately result in better maintainability of the whole system.

II. EXPERIMENTAL STUDY

Fasteners and Tools Selected

For the purpose of this study, the following four different cases of fasteners and tools were selected.

1. Flat rectangular keys (KY)
2. Wing-Nuts (WN)
3. Allen wrench with the long end as handle (AL)
4. Allen wrench with the short end as handle (AS)

Allen wrench was included in the study even though it is not a fastener by itself. However in order to tighten or loosen an small Allen screw this wrench is held in the hand in a manner very similar to holding the other two types of fasteners. The flat key was selected to represent the generic cases of such small fasteners.

The following sizes of the fasteners were used in the study.

1. Three sizes of keys:

Large - 2" long x 3/4" wide x 1/8" thick;

Medium - 1" long x 3/4" wide x 1/8" thick

Small - 1/2" long x 3/4" wide x 1/8" thick

A fourth size, 3" long was used in the study as the experimental fastener for bench mark (BM) torque exertions at the beginning and end of each session of testing.

2. Three sizes of wing-nuts:

Large - 1/2" thread diameter (2.1" max. width X 0.88" max. height)

Medium - 5/16" thread diam. (1.22" max. width X 0.58 " max. height)

Small - #8 size (0.88" max. width X 0.42" max. height)

3. Allen wrench:

The Allen wrench used was 3/16" standard size. The long end was 2.75" long and the short end was 1" long. Torque with the Allen wrench was applied through a 3/16" socket attached to the torque dynamometer

All fasteners were welded to a 3.5" square steel plate suitable for attaching to the torque dynamometer described in the following section.

Torque Dynamometer and Data Acquisition System

The torque dynamometer consisted of a Lebow Model 2192-200 torque sensor (0-200 inch-pounds, capacity) mounted on a wooden base and a metal frame. A 3.5" square aluminum face plate with a circular magnet press-fit in the center was attached to the front of the torque sensor. This face plate also had two 1/4" dowel pins to aid in attaching and holding the steel plates welded with the experimental fasteners. The frame work allowed adjustment of the height of the torque dynamometer. Figure 1 shows the dynamometer and part of the frame work. Figure 2 shows the experimental set up for the large size wing-nut for clockwise torque at 0 degree position and facing orientation.

The computerized data acquisition system consisted of a bridge amplifier (Honeywell make, Accudata model), a 16 bit analog to digital converter board (Tecmar make), a microcomputer (Compaq make) and a printer (Okidata make). The analog signal output from the torque sensor was

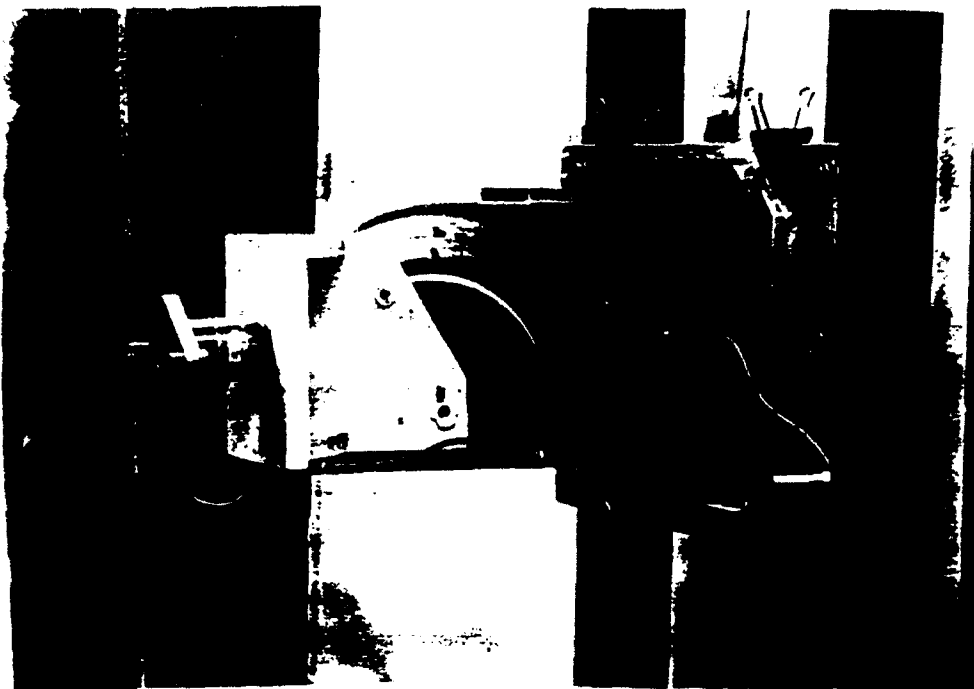


Figure 1. Torque Dynamometer and part of the frame work used in the study

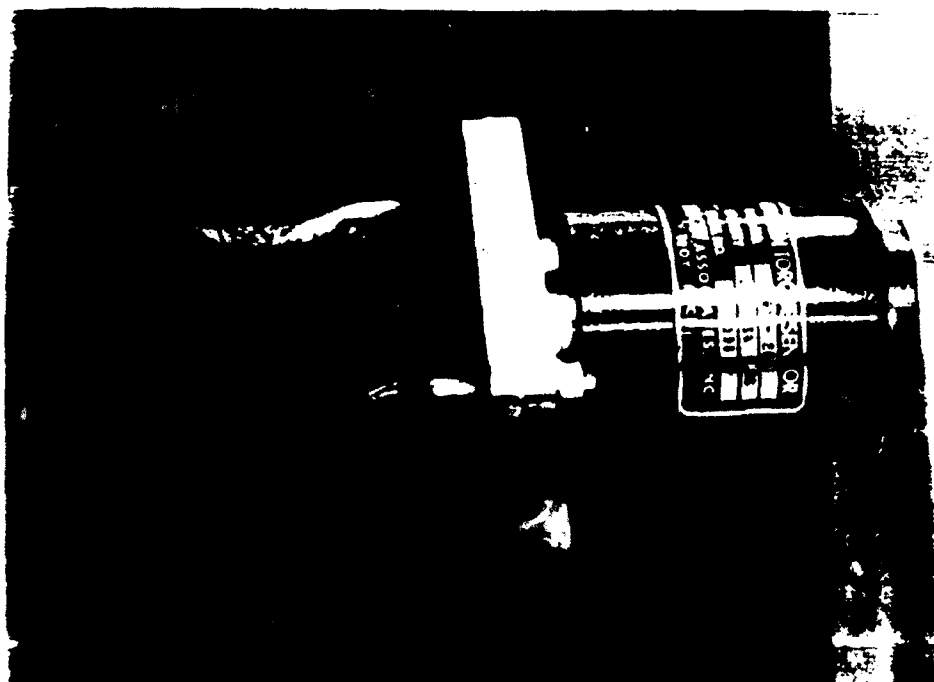


Figure 2. Experimental set up for Wing-Nut

amplified, sampled at the rate of 10 per second, digitized, scaled and printed out. The data, if determined acceptable, was stored on a 5.25" floppy disc for later analysis. Except for the torque sensor the rest of the experimental hardware has been used earlier in the laboratory for similar torque strength studies for CREW CHIEF program (1).

Experimental Variables

The following were selected as the independent variables for this experimental study.

1. Type of fastener/tool: Key, Wing-nut, Allen long, Allen short
2. Size: 2", 1", and 1/2" width for keys
1/2", 5/16" and #8 size for wing-nut
3/16" for Allen long and Allen short
3. Orientation: Facing and Transverse
4. Position: 0 and 90 degrees for key, wing-nut and Allen short
0, 90, 180, and 270 degrees for Allen long
(0 degrees refers to the vertical and upwards positioning of the fastener/tool)
5. Direction: Clockwise and Counter-clockwise

The mean torque strength computed as the arithmetic average of the 30 sample values during the exertion period of 1 to 4 seconds is the dependent variable. The above design resulted in a total of 72 exertions with 24 for key, 24 for wing-nut, and 24 for Allen long and short combined. Therefore each of the above set of 24 exertions were considered as a block in the experimental design and the individual exertions were randomized within the block for each subject. The order of the experimental blocks was also randomized. The 72 exertions were completed in 2 sessions of 36 exertions each. Thus 1 1/2 blocks were completed in one test session by a subject. In addition to the test exertions the subject also performed 2 bench mark exertions each at the beginning and end of each session. The subject was given a rest period of at least two minutes between consecutive torque exertions in order to overcome the fatigue and/or soreness in the muscles of finger, hand and arm. Such a time period was also needed for the experimenter to get the computer printer output for the exertion, make the decision to accept or not, and get ready for the next exertion. The subject was also encouraged to take additional time as break should he or she feel so. In fact several subjects did take longer breaks now and then.

The hand grip to be employed by the subject for each exertion was decided earlier by trying out on three different individuals in the laboratory and the consensus method was standardized. The standardized gripping technique was arrived by taking into consideration the task conditions, the desirability of keeping the hand, wrist, and forearm in neutral positions to avoid excessive stress the joints. The standardized gripping technique for the exertions are shown in Table 1. Minor variations from the standard technique was permitted to suit the individual hand and finger sizes so that the subject did not feel undue discomfort during gripping and turning. The subject was instructed before each exertion on the specific gripping technique for that exertion and was asked to try it without actually applying any large force. All exertions were done with the right hand of the subject.

The object being torqued was placed in two orientations relative to the subject, facing and transverse. In the facing orientation, the axis of rotation was pointing towards the subject. In the transverse orientation the axis of rotation was pointing toward the subject's right.

Subjects

Fourteen males and fourteen females in the age range 18 to 31 years participated as subjects in this study. Most of the subjects were students enrolled in the nearby educational institutions. The subjects were recruited and paid by Logicon Technical Services under a master contract with the Armstrong Laboratory of USAF for such services.

All potential subjects were pre-screened by the recruiting contractor for their health and physical conditions. The selected subjects further answered a written questionnaire first on their coming to participate in the experiment. Anyone with a current or historical condition of back injuries, back pain, hernia or other such musculoskeletal problems was excluded from participation. After this the subject was briefed on the experiment and the procedure to be followed. Then the subject signed the consent forms. The physical fitness questionnaire and consent forms have

Table 1. Instructions for Gripping

Key, Wing-Nut and Allen Wrench (short grip)

Facing at 0 degrees:

CW - Palm facing down to start and Thumb left top

CCW - Palm facing up to start and Thumb right top

Facing at 90 degrees:

CW - Palm facing left to start and Thumb right top

CCW - Palm facing left to start and Thumb left top

Transverse at 0 degrees:

CW - Palm facing plate to start and Thumb near high

CCW - Palm facing down to start and Thumb near low

Transverse at 90 degrees:

CW - Palm facing plate to start and Thumb far top

CCW - Palm facing plate to start and Thumb near top

Allen Wrench (long grip)

Facing at 0 degree:

CW - Palm facing left, grip with fingers, thumb around and over the index finger

CCW - Palm facing up, thumb on end, and index finger on bend

Facing at 90 degrees:

CW - Palm facing down, thumb at the bend and other fingers over the long end

CCW - Palm facing person, thumb at the bend, fingers below the long end

Facing at 180 degrees:

CW - Palm facing left, thumb on top of the bend, and other fingers along the long end

CCW - Palm facing person, thumb at the bend, and other fingers along the long end

Facing at 270 degrees:

CW - Palm facing down, thumb below the end, and other fingers over the handle

CCW - Palm facing up, thumb at the bend, and other fingers around the handle

Transverse at 0 degrees:

CW - Palm facing left, thumb at the end, and other fingers at the bend

CCW - Palm facing right, thumb downwards along the handle towards the bend, and other fingers around the handle

Transverse at 90 degrees:

CW - Thumb at the end, and index finger at the bend

CCW - Base of thumb at the bend, and index finger below the end

Transverse at 180 degrees:

CW - Thumb at the bend, and other fingers around the handle

CCW - Thumb at the end, and index finger at the bend

Transverse at 270 degrees:

CW - Thumb at the bend and index finger at the end

CCW - Thumb over and along the handle, and index finger at the bend

been approved by the Human Use Review Committee of Armstrong Laboratory and have been used in previous similar experimental studies for CREW CHIEF program. Copies of the forms can be found in reference (1). Except for one male and one female subject, who were ambidextrous, all others were right-handed subjects.

Experimental Procedure

As mentioned earlier the entire experiment of 72 test exertions and 8 bench mark exertions were performed by a subject in two sessions separated by at least three days. The first session lasted nearly three hours and the second lasted about two hours. At the start of the first session 12 anthropometric measurements and a battery of five benchmark strength tests were performed. The anthropometric measurements were selected in consultation with the CREW CHIEF research personnel and Dr. John McConville of ARP, Inc., Yellow Springs, Ohio. The strength tests were standardized tests developed and used for CREW CHIEF program. The statistical summary of the anthropometric and strength measurements of the subjects is given in Table 2.

After the anthropometric and strength measurements, the subject performed 36 test exertions in the randomized sequence previously determined. The torque dynamometer was adjusted vertically so that the fastener center was at a height equal to 50% of the subject's vertical reach height. To control the distance from the object being torqued, the subject stood with the toe of the right foot just behind a line on the floor set at 50% of the subject's forward grip reach. The subject was free to place the left foot at any position to the left and rear of the right foot for good balance and comfort. The subject then was instructed on the proper gripping and was asked to try and be familiar with it.

When the subject was ready the experimenter entered the command for the computer to start the data collection and also cued the subject to start applying the torque. The subject had to develop the maximum torque in during the first second and then maintain it for three more seconds.

Table 2. Anthropometric and Strength Statistics Summary

	Males n = 14		Females n = 14	
	Mean	S.D.	Mean	S.D.
AGE - Age, years	22.86	5.16	20.14	1.46
WT - Weight, pounds	165.00	24.33	145.71	17.89
STAT - Stature, mms	1775.00	44.23	1669.79	82.98
GRHT - Grip Height, mms	2157.07	69.02	2014.07	111.56
FGL - Fwd.Grip Length., mms.	541.43	66.67	486.86	51.71
HLN - Hand Length, mms	190.64	7.79	178.21	11.07
HB - Hand Breadth, mms	84.79	5.29	76.36	3.23
D1L - Digit I Length, mms ..	121.00	4.27	111.86	6.33
D2L - Digit II length, mms ..	98.57	6.72	94.21	5.19
WCR - Wrist Circum., mms....	168.07	7.45	148.64	6.06
FCR - Forearm Circ., mms ...	290.21	19.99	249.50	14.05
BCR - Biceps Circum., mms ...	321.29	35.57	276.43	21.74
GST - Grip Strength, pounds.	108.00	12.82	69.76	12.50
TEL - 38cm. Lift, pounds ...	228.50	52.32	146.00	29.53
ELL - Elbow Ht, Lt. pounds..	94.43	18.66	53.57	8.28
PS - One Hand Pull St.pds..	110.50	23.22	72.50	15.11
SIXFTL-6' Lift St. pounds....	107.85	30.43	51.43	10.27

Note: For definitions and procedures of hand size measurements, see (2).
For all others, see (1).

Table 3. Analysis of Variance (ANOVA) Summary

	AL		AS		KY		WN	
	F	M	F	M	F	M	F	M
Orientation			**	**	**	**	**	**
Direction			**	**			**	*
Position	**	**			**	**		
Size	n/a	n/a	n/a	n/a	**	**	**	*
Orient. X Direction	**	**	**					**
Orient. X Position	**	*		*	**	**	**	**
Orient. X Size	n/a	n/a	n/a	n/a	**	**	**	**
Direct. X Position	**	**	**					**
Direct. X Size	n/a	n/a	n/a	n/a	*		**	*
Positi. X Size	n/a	n/a	n/a	n/a	**	**		**
Orie. X Dir. X Posi.	**	**						
Orie. X Dir. X Size	n/a	n/a	n/a	n/a			*	**
Orie. X Pos. X Size	n/a	n/a	n/a	n/a		**	*	
Dire. X Pos. X size	n/a	n/a	n/a	n/a	*			**
Ori.X Dir.X Pos.X Size ..	n/a	n/a	n/a	n/a				*

** indicates significant at 0.05 level
* indicates significant at 0.10 level
n/a indicates not applicable

F = Females
M = Males

The computer kept track of the time and torque and signaled the subject at the end of four seconds to stop. Immediately the computer printed out the sampled values of the applied torque and certain statistics for acceptance. To be accepted the exertion should be such that, (a) the ratio of peak torque during 0 to 1 second to the mean torque during 1 to 4 seconds must be within the range 0.85 to 1.15 and (b) during seconds 1 to 4, no more than 5 samples be found beyond the range of 10% on either side of the mean value. If the above criteria were not met the exertion was repeated. However if the criteria could not be met in three trials, then the best of the three exertions was accepted.

Results

The experimental study resulted in a total of 2016 data points of mean torque values representing the various experimental conditions. The torque strength (mean of the torque applied during the time period 1 to 4 seconds) values were averaged for each combination of the independent variables separately for males and females. The results are presented in graphical form in figures 3 through 6. The data was also grouped into 8 sets for the four types of fasteners and the gender of subject. Then an analysis of variance was performed using the BMD statistical package available on the VAX computer of the Armstrong Laboratory. Table 3. shows the summary of ANOVA analysis and Table 4. shows examples of linear regression models. Discussed below are some important observations.

1. The highest average torque strength value for males (70.96 inch-pounds) and females (51.42 inch-pounds) is noticed while using the Allen wrench with long end in transverse position at 180 degrees and clockwise turning. In this case the torque is primarily due to a pulling action with finger, wrist, elbow, and even shoulder muscles contributing to the action. The longer grip makes possible to exert the force with less discomfort and pain.

2. The weakest exertion overall was with the small (1/2") key in the facing orientation and clockwise direction. The weakest position was 0 degrees for males (5.25 inch-pounds) and 90 degrees for females (4.05 inch-pounds). The position effect may be considered negligible in this case. The small

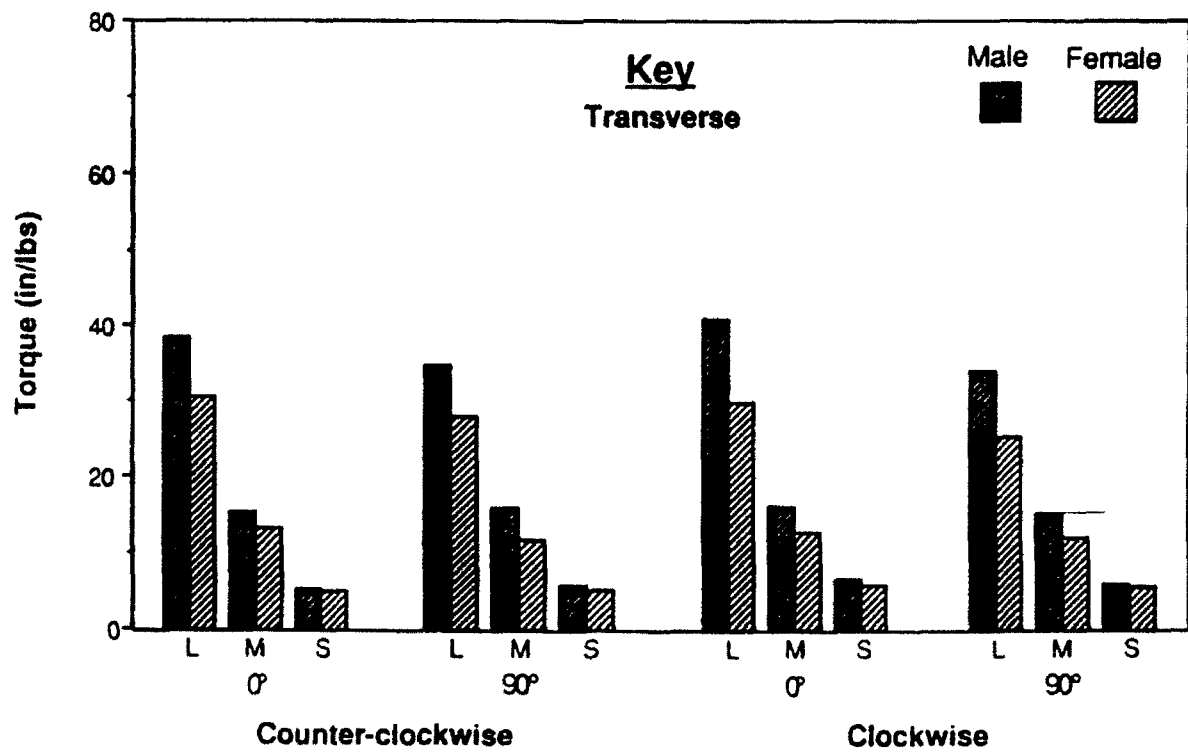
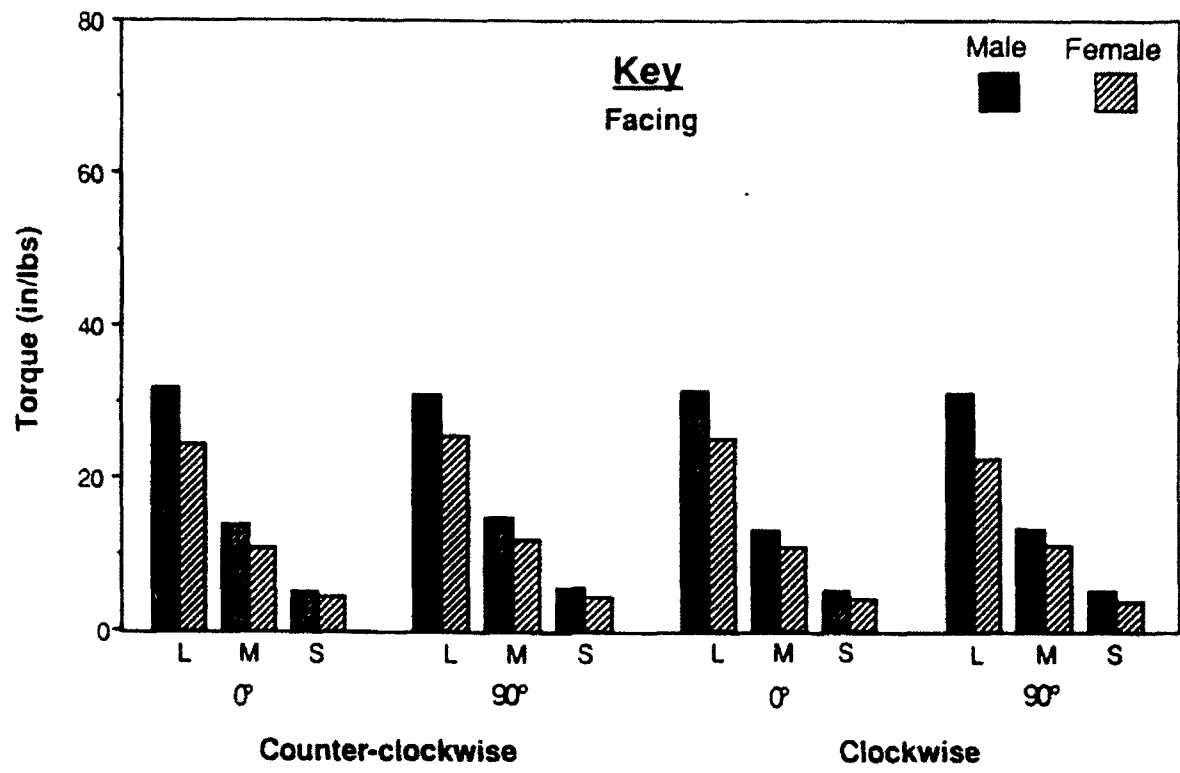


Figure 3. Torque Strength for Key

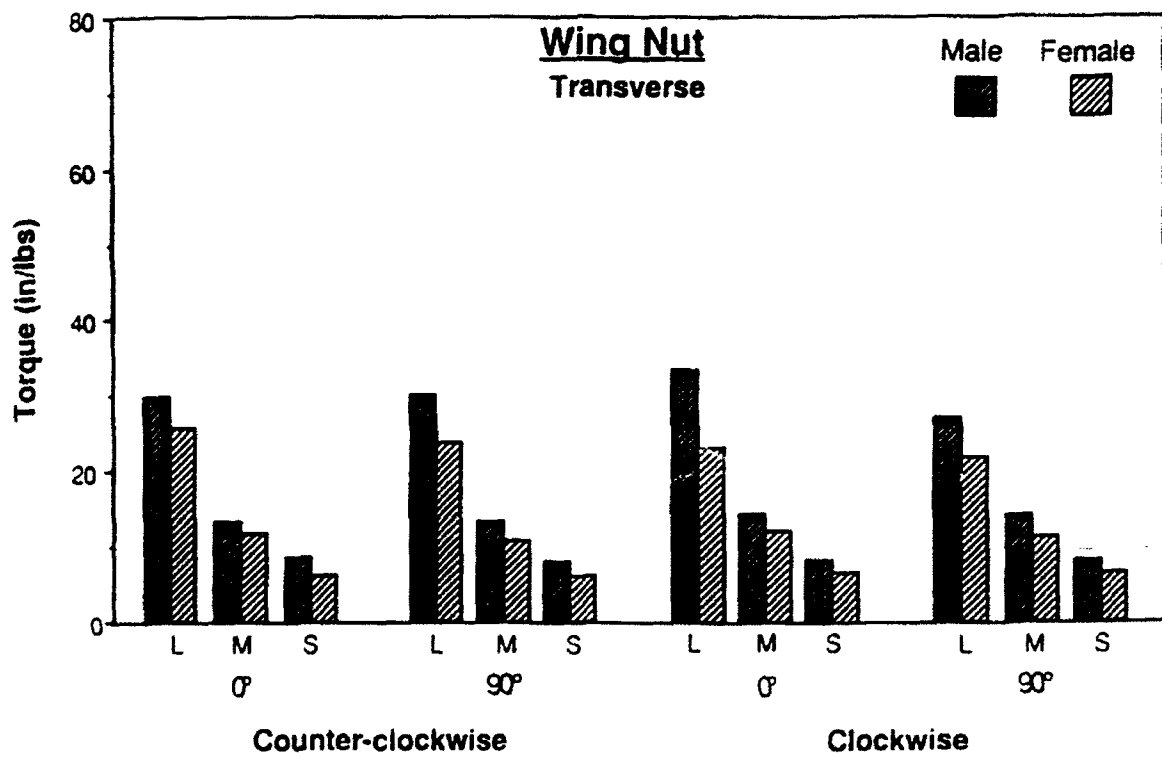
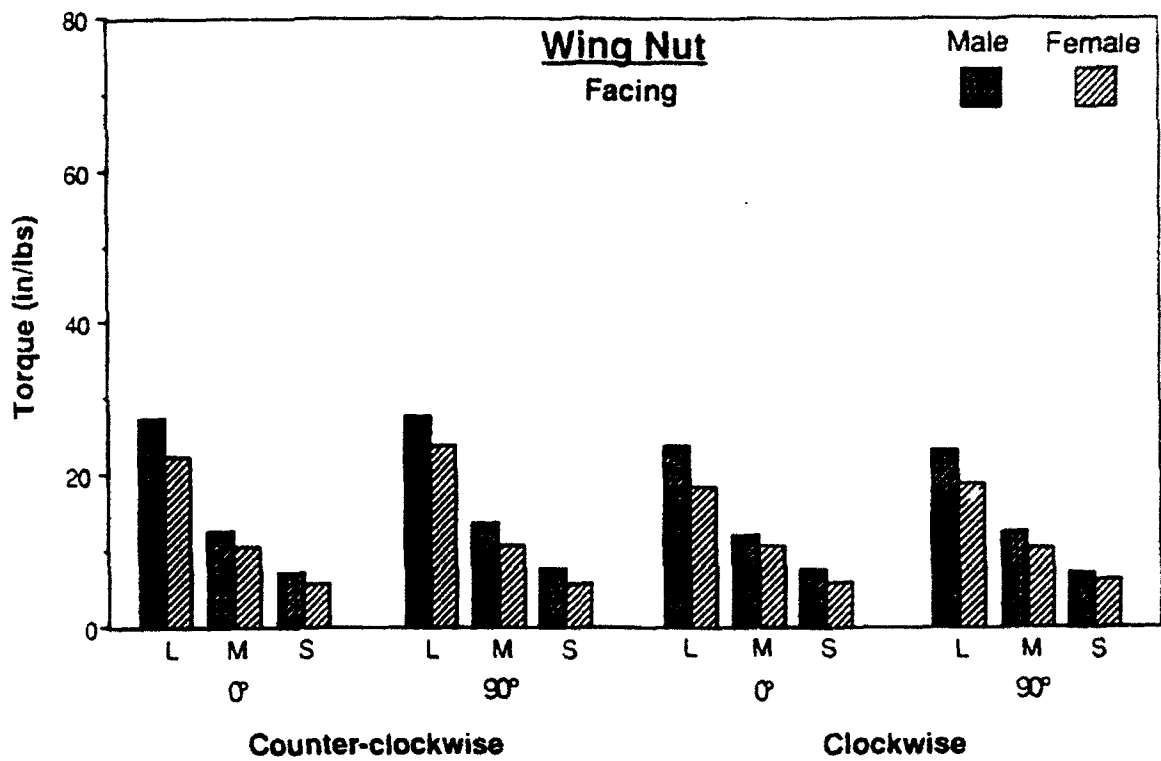


Figure 4. Torque Strength for Wing-Nut

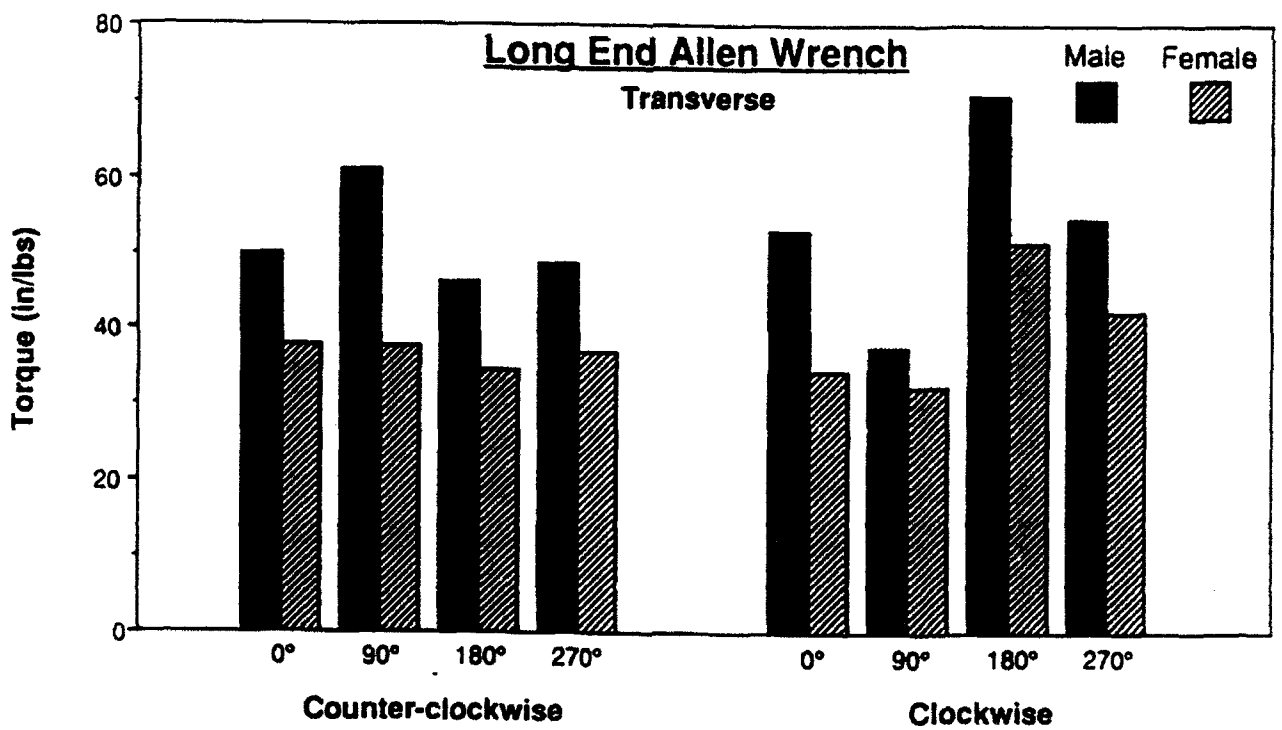
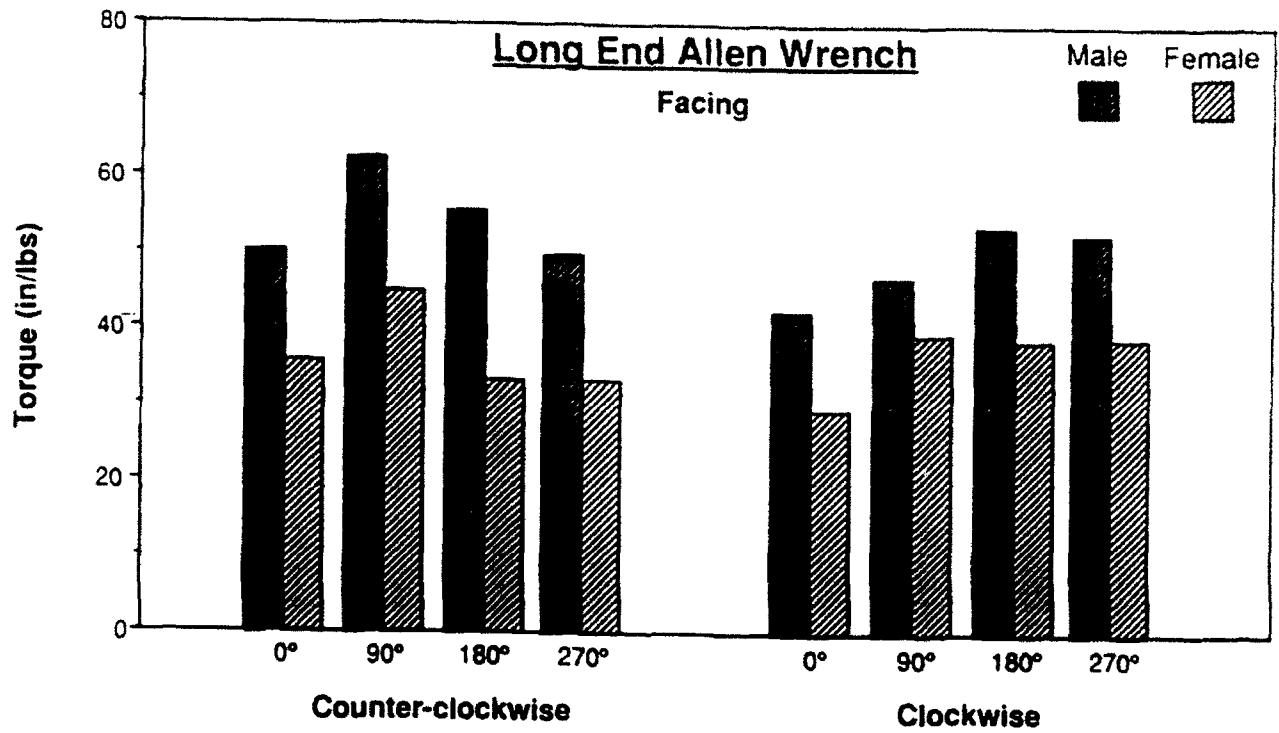


Figure 5. Torque Strength for Long End Allen Wrench

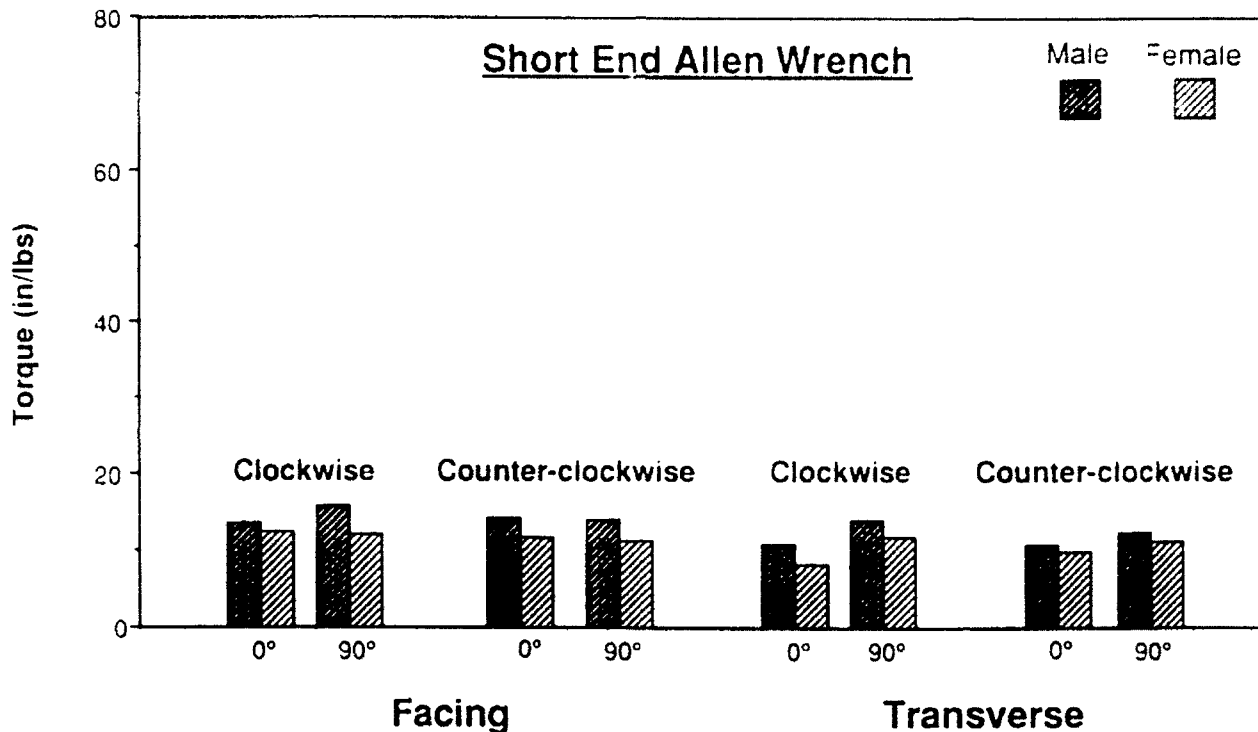


Figure 6. Torque Strength for Short End Allen Wrench

Table 4. Examples of Linear Regression Models

CASE 1: Female, Medium Key, Transverse Orientation, 0-degree Position, Clockwise Torque

$$\text{TORQUE} = 32.22 + 0.097 (\text{TEL}) - 0.361 (\text{WCR}) + 0.217 (\text{GST}) + 0.056 (\text{FCR}) \dots\dots\dots R = 0.90$$

CASE 2: Male, Allen Wrench-Long Grip, Facing, Orientation, 180 degree Position, Counter-clockwise Torque

$$\text{TORQUE} = 118.83 - 1.005 (\text{WCR}) - 1.192 (\text{HLN}) - 1.695 (\text{HB}) + 1.207 (\text{AGE}) - 0.202(\text{SIXFTL}) \dots\dots\dots R = 0.88$$

CASE 3: Female, Allen Wrench -Short Grip, Facing Orientation, 90-degree Position, Clockwise Torque

$$\text{TORQUE} = -13.63 + 0.140 (\text{HLN}) \dots\dots\dots R = 0.66$$

CASE 4: Males, Wing-Nut, Facing Orientation, 0-degree Position, Clockwise Torque

$$\text{TORQUE} = 77.26 + 0.112 (\text{ELL}) - 0.689(\text{HB}) + 0.047(\text{FGL}) + 0.066(\text{TEL}) - 0.302(\text{D1L}) \dots\dots\dots R = 0.90$$

(Note: Torque values in INCH POUNDS ; For others see Table 2)

size results in a grip with the thumb and the lateral side of the index finger, somewhat like the lateral pinch grip causing most of the turning force to come from finger muscles. The pain experienced along the sides of the first and second phalanges of the index finger limited any attempt by the person to make use of stronger muscles of the arm.

3. The gender effect of males with higher strength can be seen clearly. However, the relative difference is much smaller in the present study compared to previous studies on strength. This is probably because even if a person is physically strong, his/her ability to convert that strength to effective torquing action is limited by the size of the fastener and the individual's pain tolerance.

4. The effect of orientation in general has been noticed as transverse with higher torque strength than facing. In the transverse orientation the person is able make use of stronger muscles and joint actions.

5. The size effect, as can be expected, is significant for key and wing-nut. The larger ones result in higher torque values because of larger contact areas and more comfortable and stable grips. Table 5 below presents the torque data relative the effective grip diameter of each type of fastener/tool. The effective grip size the estimated distance in inches. The torque (inch pounds) is averaged over all subjects, orientations, positions, and direction. The 2 1/4" size key is the largest size used only for benchmark torque exertions. This table suggests may be used, with caution, to interpolate the torque strength values for other sizes.

Table 5. Torque Applied Vs. Effective Grip Diameter

	Effective Grip Diameter (inches)								
	Key		Wing-nut			AL	AS		
Size	1/4	5/8	1 1/2	2 1/4	9/16	15/16	1 5/16	2 1/2	3/4
Torque	5.19	13.34	30.39	40.93	7.04	12.19	25.00	45.10	12.31

6. Position effects are noticed in the cases of Allen wrench (long grip) and key. With 4 levels of the position variable such an effect was clearly expected. However, among the other three types of fasteners, each with two levels, only key seems to be influenced by the position. At this time this effect can not be explained.

7. The direction is significant only for wing-nut and Allen wrench, small grip. Again the reason for this is not clear. It is possible that the effect of direction is masked as second order interaction effects with other primary variables for key and Allen wrench, long grip.

8. The second order interactions can be explained by the fact that different combinations of the variables resulted in different grips as indicated in Table 1.

9. The higher order interactions are difficult to explain at this time. It is suspected that many will disappear if the sample size is larger.

The torque strength data was also used to develop linear regression models. Initial trials to develop general models including anthropometric, strength and task variables resulted in models with low multiple -R values. Therefore it was decided to develop individual regression models for each task condition and gender separately with the anthropometric and strength values as the independent variables. This resulted in a total of 144 regression models with R values for most not less than 0.8. (Computer print outs of all the 144 regression models are included in the appendix to this report transmitted to the CREW CHIEF program personnel.) Some of the models ended up with more than eight variable while some others had only one or two variables. Also some models resulted in multiple-R values less than 0.3. Obviously such cases need further to study to include other variables not considered in this study. CREW CHIEF research personnel have been using the 6' feet lift test (SIXFTL) as the primary variable for regression models. In the present case this variable does not appear in several cases. The SIXFTL has been found to be a very good predictor of overall body strength for heavy physical work. In turning small fasteners it was observed that in many cases the limiting factor was the pain tolerance of the person while trying to apply the turning force over a relatively small contact area. Therefore the any indicator of overall strength by itself may not be good predictor of the torque strength capacity of the individual. It is suggested that one or two hand/finger dimensions be included in the regression models along with the strength indicator for modeling torque strength for turning small fasteners and tools with bare hands.

III. REFERENCES

1. Gibbons, L.E., Summary of Ergonomics Research for the CREW CHIEF Model Development (U) AAMRL-TR-90-038, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 45433, December 1989.

2. Griener, Thomas, M., Hand Anthropometry of U.S. Army Personnel, Technical Report Natick/TR-92/011, United States Army Natick Research, Development and Engineering Center, Natick, MA 01760-5000, December, 1991.

IV. ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to Mr. Edward Boyle (AL/HRG), Dr. Joe W. McDaniel (AL/CFHD) and Mr. Nilss Aume (AL/CFHD) for suggesting this research problem and for their encouragement and assistance during the period of the study. The author also wishes to acknowledge the willing assistance provided by the following employees of The University of Dayton Research Institute during various times of the research: Mr. Phil Krauskopf, Mr. Glenn Robbins, Mr. Van Thai, Mr. Don Haddox, Mr. Glenn Severt, Mr. L. E. Gibbons, Mr. David Kancler, Ms. Becky Unger and Ms. Kathy Smith.

DESIGN AND MODIFICATION OF A THREAT SITE/EMITTER LAYDOWN
DATABASE AND THREAT EMITTER PARAMETERS DATABASE
FOR THE ADVANCED DEFENSIVE MISSION PLANNING SYSTEM
AND B1-B/B-2 ENGINEERING RESEARCH SIMULATORS

John C. Duncan
Assistant Professor
Aerospace Engineering Technology

Kent State University
Kent, Ohio 44242

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

DESIGN AND MODIFICATION OF A THREAT SITE/EMITTER LAYDOWN
DATABASE AND THREAT EMITTER PARAMETERS DATABASE
FOR THE ADVANCED DEFENSIVE MISSION PLANNING SYSTEM
AND B1-B/B-2 ENGINEERING RESEARCH SIMULATORS

John C. Duncan
Assistant Professor
Aerospace Engineering Technology
Kent State University

ABSTRACT

The primary goal of the Advanced Defensive Mission Planning System (ADMPS) is to allow rapid Mission Generation - integrating terrain information, routes, and threats - by a non-expert Mission Planner. The ADMPS should provide a realistic and accurate model of threat/site laydown using current doctrine, strategy and practices. At the same time, each threat model should simulate realistic threat emitter characteristics and operation. Together, the digitized terrain database, threat site laydown and emitter simulation should provide a realistic simulated electronic warfare environment that can be used in various Defensive System simulations. Once fully implemented, the ADMPS will be able to easily and quickly provide total mission scenarios - including training and EWO missions - for use in studies in various aircraft simulations.

DESIGN AND MODIFICATION OF A THREAT SITE/EMITTER LAYDOWN
DATABASE AND THREAT EMITTER PARAMETERS DATABASE
FOR THE ADVANCED DEFENSIVE MISSION PLANNING SYSTEM
AND B1-B/B-2 ENGINEERING RESEARCH SIMULATORS

John C. Duncan

INTRODUCTION

The Armstrong Aerospace Medical Research Laboratory is involved with ongoing human factors research and studies concerning the analysis of defensive crew station design for strategic aircraft including the B-52, B-1B, and B-2 bombers. The laboratory uses high fidelity defensive crew station simulators to simulate defensive crew member avionics, including visual displays and aural cues, in order to investigate, analyze and evaluate the decisions and actions performed by a Defensive Systems Officer. One of the primary requirements of the evaluations is the use of a realistic simulation of threat emitters. The threat simulation must provide a realistic and accurate model of threat/site laydown using current doctrine, strategy and practices. Historically, the placement of threats for a mission has been accomplished manually, site by site, requiring a high degree of expertise and knowledge of threat characteristics and tactical/strategic deployment. At the same time, each site should simulate realistic threat emitter characteristics and operation. The purpose of the ADMPS is to allow rapid Mission Planning by a non-expert by automating threat placement during the mission build process and providing highly detailed threat emitter models.

METHODOLOGY

The work completed this summer was a continuation of the tasks initiated at Armstrong Laboratory in the summer of 1991. The tasks identified and methodology followed are described in the previous summer's final report (Duncan, 1991). The tasks identified in the summer 1991 report that were completed this summer are described in this report.

Results, Accomplishments, Comments and Recommendations

Accomplishments

1. Mode sequencing/transition parameters were developed and defined for all emitters based on emitter system (type), beam #, and operational mode. The Emitter Mode Data parameters (see Figure 1) provide a method for simulating realistic operational mode transitions for a threat emitter and includes specific mode transition information. Note: Mode transitioning in the High Fidelity (Hi-Fi) simulation is simple and crude. It does not currently incorporate the parameters and methods described by the Emitter Mode Data parameters.

2. A description of PRI parameters for PRI agile emitters were developed, and the PRI values and sequencing - based on emitter system (type), beam #, and mode - are described in the Emitter PRI Agility Data tables (see Figure 2). The Emitter PRI Agility Data parameters provide a method for simulating the PRI characteristics of an agile PRI emitter, specifically discrete PRI sequences, staggered PRI sequences, PRI jitter levels, and PRI switch-on-scan PRI values. Note: the Hi-Fi simulation does not currently incorporate the PRI agility types described by the Emitter PRI Agility Data parameters.

3. Scan type/motion definitions were developed and a description of the emitter signal scanning characteristics was provided for all emitters - based on emitter system, beam #, and mode - through the Emitter Scan Data parameters (see Figure 3). The emitter's physical scan motion is sufficiently described so that the generation of aural signals, including timing and amplitude, is either explicitly described or can be derived from the parameters given. The scan motion parameters are not based solely on the aural simulation requirements, but provide a description of the physical scanning motion characteristics of each emitter. Note: The Hi-Fi simulation threat database is not fully transportable because of the method of implementing audio response. Some pertinent scan motion parameters are "missing" since the database was "kluged" to make it work for the Hi-Fi simulation: scan parameters that describe the actual scan motion are not

used - only the audio period is provided - and scan timing synchronization between co-emitters of multi-beam threat systems is not simulated.

4. Independent Acquisition radars were separated from associated Fire Control radars for those systems in which the radars were combined into a single system. Several emitter beams were identified as separate Acquisition/Search radars that are not slaved to a Fire Control radar. In both the new systems modeled and existing systems that were incorrectly modeled, the Acquisition radars have been broken out from the SAM/AAA Fire Control radars and incorporated as separate systems/emitters with their own mode transitions. Note: The Acquisition radars are independent from the Fire Control radars and should not transition in mode along with the Fire Control radars. These systems were identified and corrections noted.

5. Several new threats were added to the list of simulated threats. The new threats include some Acquisition Radars that were broken out from Fire Control radars, and a number of new emitter system models that were added to provide more capability and increase realism. In addition, the current ERS threat database was updated to include several threat systems that were modeled in the ADMPS laydown database and were previously absent from the threat database. A list and general description of those threats is included in the classified notebook.

6. Some data concerning existing threats was modified because of additional or updated information, or incorrect modeling of an emitter. The corrections were noted through red lines to the current ETCL listing and/or completion of new emitter parameter sheets (see Figure 4). Included in the modifications were the addition of emitters that were not modeled because they could not be detected by the ALQ-161 (out of band). Note: Many phased array/multi-function emitters/systems beams utilize time sharing techniques to perform different functions. The signal(s) from these types of systems usually appear to an observer as several different signals transmitting in rapid succession. This type of operation is seen by most defensive systems as several unique signals (multi-beam operation) emitted from the same threat. These types of threats/signals were treated like multi-beam systems

in order to simulate the defensive system responses. Changes were made to the threat systems models that use multi-function radars so that a multi-beam simulation has been implemented.

High-Fidelity Simulation Problems

There are several shortcomings in the threat signal audio simulation in the Hi-Fi simulation. Dick Smith is aware of these problems and he and I have discussed possible software solutions to some of them. The problems are noted below.

1. Scan timing problems - scan synchronization, typically for back-back circular scanning emitters, is not provided. The simulation will not provide the proper time interval between signals for those systems with different RFs and/or PRFs that are synchronized with each other.
2. The Hi-Fi simulation uses pseudo-random PRI selection for Stagger levels. Actual emitter PRIs should be provided and the levels and sequencing are described in the Emitter PRI Agility Data tables. (see Figure 2).
3. The Hi-Fi simulation does not correctly model raster scanning emitters. The simulation does not simulate the signal timing and changes in received power level as the emitter main beam moves from bar to bar during the raster scan. The same problem exists to a lesser degree for helical and spiral scanning systems.
4. Signals that cannot be detected (out of band) by the ALQ-161 are not modeled. Ideally, the threat data would completely represent all threat emitter parameters - to maintain portability - and the simulation would screen out those signals which cannot be detected by a particular defensive system.
5. There is no method for simulating intermittent emitters (IFF, CDL, GCA, GCI, communications). The signals are identified in the ADMPS laydown data, but the Hi-Fi simulation does not simulate the random/intermittent

transmission of these signals.

6. RF signal agility and multi-beam signal RF synchronization or separation is not simulated.

7. "Out-of-Sector" sector scanning is not provided by the Hi-Fi simulation for raster scanning, unidirectional sector scanning, bidirectional sector scanning, helical scanning, or spiral scanning systems. The Hi-Fi simulation treats all sector type scanning as if the target aircraft (B1-B) is always centered within the sector, which causes audio signals to appear at regular time intervals and power levels. In actuality, the target aircraft is not likely to appear exactly in the center of the sector much of the time, thus causing the audio signals to appear at irregular intervals with varying power levels. Note: This is a particularly noticeable problem with Airborne Interceptor (AI) threats because their scan sector width is limited and the sector scan is centered/boresighted along the aircraft's longitudinal axis.

Recommended Threat Simulation Enhancements

As described above, there are several changes that should be made to correct existing deficiencies in the ADMPS and Hi-Fi simulation. In addition, there are a few features that could be incorporated that would increase the fidelity and realism of the threat simulation. Suggestions are given below in a prioritized list.

1. Modify the Hi-Fi threat simulation so that realistic mode changes are incorporated into the design. Threats should be allowed to transition through various operational modes with changes to affected emitter parameters in the same way that actual threats change mode. The mode change criteria has been defined for each threat by the Mode Data parameters and should be implemented during real-time by simulation Mode Control software. Software must be developed to control mode changes, and to update the emitter parameters when the emitter changes mode. A high fidelity simulation of mode changes would require that the Mode Controller calculate detection and break-lock conditions

based on received signal strength (based on range, effective radiated power, ground clutter, and aircraft radar cross-sectional area), occulting, and countermeasures effects, for each active beam. It will be difficult to establish the criteria for countermeasures effectiveness (unless EID data is the only criteria used) and the overhead involved in analyzing every emitter will be high. Therefore, I recommend a simple method of controlling mode changes where the Mode Controller ignores countermeasures effects and changes modes based only on target detection, and the range and time requirements described by the Emitter Mode Data parameters. It will be simpler to allow the simulation instructor/operator to manually change threat modes to simulate countermeasure effectiveness.

2. There are a number of significant threat types missing from the threat simulation. In particular, the database does not contain several Red AAA system, has no Red Naval systems, and is missing most Blue/Gray threats. In addition, the current threat database does not include several recent threat systems (listed in the ADMPS laydown database) because of a lack of information concerning those systems. Because of the changing roles of strategic bombers - and the new emphasis on the conventional role - the incorporation of Blue/Gray threats has great importance. The threats are widely deployed, in significant numbers, to third world nations and U.S. allies and a conventional role makes the probability of encountering some of these systems very high. In addition, ground based jammer systems - which are currently ignored, should be added to the threat database.

3. Add "Ambush" versions of the threats which can use optical, electro-optical, or directed tracking against a target. These types of systems operate by keeping the system emitters in a passive state, pointing the system emitters at a target through guidance/direction from external sources, and then only radiate signals when ready to launch. In essence, the system emitters bypass the normal search and acquisition phases and come on the air in a tracking mode. The ambush mode can be easily modeled through the Emitter Mode Data parameters without changing any other emitter characteristics.

4. In addition to the mode related parameters, the software should calculate

received signal strength (at the aircraft) for each emitter based on sector coverage, occulting, effective radiated power, and ground clutter.

5. RF signal agility and multi-beam signal RF synchronization or separation should be simulated.

6. Develop graphical site laydown illustrations for the mission planner. The illustrations should show threat emitter location/laydown - by showing the number of emitters and their range and bearing - and can be presented either online during mission planning or through hard copy illustrations in a planning guide.

7. Implement the simulation of intermittent emitters (IFF, CDL, GCA, GCI, communications). (The signals have already been identified in the ADMPS laydown data.)

8. "Out-of-Sector" sector scanning should be provided by the Hi-Fi simulation for raster scanning, unidirectional sector scanning, bidirectional sector scanning, helical scanning, or spiral scanning systems.

References

Duncan, J.C., 1991. DESIGN OF A THREAT SITE/EMITTER LAYDOWN DATABASE AND THREAT EMITTER PARAMETERS DATABASE FOR THE ADVANCED DEFENSIVE MISSION PLANNING SYSTEM AND B1-B/B-2 ENGINEERING RESEARCH SIMULATORS, AFOSR Summer Faculty Program Final Report, September, 1991.

EMITTER MODE DATA

SYSTEM NAME		EMITTER NAME		# OF MODES	

MODE #	MODE NAME	NEXT HIGHER MODE	NEXT LOWER MODE	MODE RANGE LIMIT	MAXIMUM TIME IN MODE	MODE CHANGE DELAY TIME	BEAM ACTIVE
				km	seconds	seconds	TRUE/FALSE
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

MODE #	MODE NAME	NEXT HIGHER MODE	NEXT LOWER MODE	MODE RANGE LIMIT	MAXIMUM TIME IN MODE	MODE CHANGE DELAY TIME	BEAM ACTIVE
				km	seconds	seconds	TRUE/FALSE
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

MODE #	MODE NAME	NEXT HIGHER MODE	NEXT LOWER MODE	MODE RANGE LIMIT	MAXIMUM TIME IN MODE	MODE CHANGE DELAY TIME	BEAM ACTIVE
				km	seconds	seconds	TRUE/FALSE
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

MODE #	MODE NAME	NEXT HIGHER MODE	NEXT LOWER MODE	MODE RANGE LIMIT	MAXIMUM TIME IN MODE	MODE CHANGE DELAY TIME	BEAM ACTIVE
				km	seconds	seconds	TRUE/FALSE
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

Comments:

Figure 1

EMITTER PRI AGILITY DATA

SYSTEM NAME	EMITTER NAME
-------------	--------------

MODE #	MODE NAME	BEAM #
--------	-----------	--------

DECALTE P/N TABLE #	JITTER LEVEL	WORKSATION RATE	RANDOM P/N LOWER LIMIT	RANDOM P/N UPPER LIMIT	RANDOM P/N CHANGE RATE

# OF STAGGER LEVELS	P/N STAGGER 01	P/N STAGGER 02	P/N STAGGER 03	P/N STAGGER 04	P/N STAGGER 05	P/N STAGGER 06	P/N STAGGER 07	P/N STAGGER 08	P/N STAGGER 09	P/N STAGGER 10

P/N SWITCH-ON-SCAN LEVELS	P/N LEVEL 01	P/N LEVEL 02	P/N LEVEL 03	P/N LEVEL 04	P/N LEVEL 05	P/N LEVEL 06	P/N LEVEL 07	P/N LEVEL 08	P/N LEVEL 09	P/N LEVEL 10

Comments:

Figure 2

EMITTER SCAN DATA

SYSTEM NAME	EMITTER NAME		# OF MODES	# OF BEAMS

MODE #	MODE NAME /	BEAM #

SCAN TYPE	ADMIN SCAN PERIOD	ELEVATION SCAN PERIOD	RETRACE PERIOD	CONICAL SCAN RATE	PARABOLIC SCAN RATE	ADMIN BECTOR WIDTH	ELEVATION BECTOR WIDTH	ELEVATION ANGLE	AI ADMIN BECTOR WIDTH LIMIT	AI/UPPER ELEVATION BECTOR WIDTH LIMIT	AI/LOWER ELEVATION BECTOR WIDTH LIMIT
	seconds	seconds	microseconds	Hz		degrees	degrees	degrees	degrees	degrees	degrees

MODE #	MODE NAME	BEAM #

SCAN TYPE	ADMIN SCAN PERIOD	ELEVATION SCAN PERIOD	RETRACE PERIOD	CONICAL SCAN RATE	PARABOLIC SCAN RATE	ADMIN BECTOR WIDTH	ELEVATION BECTOR WIDTH	ELEVATION ANGLE	AI ADMIN BECTOR WIDTH LIMIT	AI/UPPER ELEVATION BECTOR WIDTH LIMIT	AI/LOWER ELEVATION BECTOR WIDTH LIMIT
	seconds	seconds	microseconds	Hz		degrees	degrees	degrees	degrees	degrees	degrees

Comments:

Figure 3

Emitter Name _____

Mode Name _____

Beam #

() Upper_Freq (Mhz) Lower_Freq (Mhz) Freq_Type Upper_PRI (usec) Lower_PRI (usec)

() PRI_Type Upper_Scan_Rate (sec) Lower_Scan_Rate (sec) Scan_Mod (-dbm) Scan_Type

() Upper_Pulse_Width (s) Lower_Pulse_Width (s) ERP (dbm) MultiBeam_Ind Beam_Width

Beam #

() Upper_Freq (Mhz) Lower_Freq (Mhz) Freq_Type Upper_PRI (usec) Lower_PRI (usec)

() PRI_Type Upper_Scan_Rate (sec) Lower_Scan_Rate (sec) Scan_Mod (-dbm) Scan_Type

() Upper_Pulse_Width (s) Lower_Pulse_Width (s) ERP (dbm) MultiBeam_Ind Beam_Width

Figure 4

Fractal and Multifractal Aspects of an
Electroencephalogram

John E. Erdei
Associate Professor
Department of Physics

and

Elaine M. Brunsman
Department of Physics

University of Dayton
300 College Park
Dayton, Ohio 45469-2314

Final Report for:
Summer Research Program
Armstrong Laboratory/Crew Systems
WPAFB

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

July 1992

Fractal and Multifractal Aspects of an
Electroencephalogram

John E. Erdei
Associate Professor
and
Elaine M. Brunsman
Department of Physics
University of Dayton

ABSTRACT

The tools of fractal analysis have been used in an attempt to better understand the relationship between dynamical systems analysis and cognitive task assessment. The main thrust of the research has been to determine both the fractal and multifractal scaling behavior of an electroencephalogram (EEG), and to determine if this scaling behavior is sensitive to tasks being carried out by a subject. Universal features of the EEG have been examined by comparing the scaling behavior for 5 subjects, each of which performed the same tasks. The results were compared with known results for random noise, in an effort to determine the amount of randomness in the signal. Upon determination of the details of the scaling behavior of the EEG, the sensitivity of this behavior to cognitive task was examined by comparing EEG recorded under 12 different task conditions. Four 30 second EEG segments taken from a single 3 minute record were examined, and the scaling behavior for each segment was averaged to produce a single result for the given task. Scaling exponents and the multifractal spectrum was computed for each subject under all task conditions. Although the determination of the scaling exponents and the multifractal spectrum was successful, association of these characteristics with particular tasks yielded ambiguous results. A great deal of the ambiguity stems from the lack of stationarity of the signal, since the scaling behavior is not uniform from one section of the EEG to another.

Fractal and Multifractal Aspects of an Electroencephalogram

John E. Erdei
Elaine M. Brunzman

INTRODUCTION

The area of dynamical systems analysis has shown that even though the behavior of a non-linear system can be quite complex, there are often general features which can tend to unify the way in which these systems are studied. For example, many behaviors are well described by an evolution equation of the form

$$\frac{\partial x_i}{\partial t} = F_i(x_1, x_2, \dots, x_n, t, c_1, c_2, \dots, c_p) \quad i = 1, 2, 3, \dots, n \quad (1)$$

where the x_i are a set of variables which can be measured experimentally, and the c_j ($j = 1, 2, \dots, p$) are a set of control parameters. The control parameters correspond to any relevant set of quantities which can be fixed by the experimenter. For a physical system, the details of the form of the functions F_i are often either only partially understood or not understood at all. At times, a reasonable guess can be made for the form of the F_i , the number of variables x_i , and the types of control parameters involved in the model. An example can be seen in the study of population dynamics, where the x_i are associated with predators and prey, the functions F_i can be constructed by making reasonable assumptions about the interactions between the x_i , and the c_j are associated with food supplies, living space, etc. In other situations, the form of F_i may not be known, but the control parameters are well defined. This situation can occur in fluid dynamics, where flows are controlled by a set of established numbers, like the Reynolds number, and the variables are often the components of a velocity field.

There are times, however, when the best one can do is make measurements of one or more variables as a function of time, without prior insight into either the form of F_i or the relevant control parameters. The study of electroencephalograms (EEG) can be placed in this last category, since although temporal records of the voltage activity of the brain can be recorded, little is known about the ingredients of the right hand side of equation (1) or even if equations of this type are appropriate.

Fortunately, some help has been provided through the study of chaotic systems. Single probe measurements, by use of the time delay method, can be used to reconstruct the complete phase portrait of the system. From this portrait, an estimation of the number of variables required to reconstruct the dynamics can be obtained through the correlation dimension, and the orbital stability of trajectories in phase space can be determined through the Lyapunov exponent.

In an effort to further understand the dynamics of the EEG, we have focused our attention on the scaling, or self-similar, behavior of the EEG. The main goal is to provide a foundation for the assessment of cognitive task difficulty. The assessment of task difficulty can be important in many areas of interest to the Air Force. From the standpoint of dynamical systems analysis, the assessment of cognitive tasks can be difficult, since the dynamical model is not well defined.

The approach we have taken highlights the fractal nature of the EEG. Typically, a self-similar or statistically self-similar fractal is characterized by the existence of a scaling law defined over some characteristic of the object. This scaling law is the result of a homogeneous function defined over the set. A function is homogeneous if it is defined by the relationship

$$F(\lambda x) = \lambda^p F(x), \quad (2)$$

where λ is some number. This equation is called a scaling law, and p is referred to as a scaling exponent. If this scaling behavior is defined in terms of a measure as a result of counting the number of "boxes" required to cover the set, then the scaling exponent is called the *fractal dimension* of the set, or more appropriately, the *box dimension*. Many other characteristics can also be used to define homogeneous functions for a fractal set. An extension of homogeneity occurs when the scaling behavior must be defined over two quantities. A generalized homogeneous function is defined to be a function for which

$$F(\lambda^a x, \lambda^b y) = \lambda^p F(x, y). \quad (3)$$

Curves which satisfy this condition are called self-affine.

We have tested the self-affine character of the EEG, with the intention being to determine an accurate measurement of the scaling exponent, p . The technique used is well suited for calculations of the statistical self-affinity of a time series and is called rescaled range analysis. The scaling behavior is associated with the statistics of the EEG, with the scaling exponent being known as the Hurst exponent, H ($0 < H < 1$). From H , the fractal dimension can be determined from $d_f = 2 - H$.

The motivation behind this stems from many considerations. First, no clear statement of the value of scaling exponents for an EEG appears to exist in the literature. It is worthwhile, therefore, to clarify the self-similar (or self-affine) nature of the EEG. Second, questions exist regarding the relationship between noise (i.e. randomness) and the EEG signal. Many types of random signals have been tested by Mandelbrot and Wallis¹. Their results show that $H = 0.5$ for random time series generated using many different types of distribution functions. A Hurst exponent in the neighborhood of 0.5 for an EEG would be a strong indication that the signal is dominated by random noise of some type.

Having established the scaling behavior of the EEG, the final task is to relate the scaling to the cognitive task. The working assumption is that the Hurst exponent, H , is a function of the control parameters and that the cognitive task being performed can be used as a control parameter. The hope is that the change in task will be reflected in a change in Hurst exponent, so that this exponent can be used as a relative measure of the task performed.

We have also initiated a study of the multifractal characteristics of an EEG. In this study, the EEG is cut through with a threshold, thereby producing a point set. The use of thresholds can be associated with mimetic techniques in the study of EEG. The resulting point set is treated as a (truncated) Cantor set, and analyzed using multifractal methods. The set is decomposed into fractal subsets (S_α) each with its own characteristic fractal dimension ($f(\alpha)$). The resulting multifractal spectrum gives information about the dynamics behind the information. Our assumption here is the same as with the calculation of the Hurst exponent, i.e., can variation in the multifractal spectrum be used as a measure of the cognitive task performed.

RESOURCES

The EEG used for this study were previously recorded at the Crew Systems Directorate of Armstrong Laboratory at WPAFB. The records were made at 21 sites across the head and under 12 different test conditions. The data collection rate was 128 Hz, and all recordings were made for three minutes. The three minute records were then subdivided into shorter segments for analysis. The conditions were eyes closed no task, eyes open no task, visual Sternberg of 1 and 4 letters, auditory Sternberg of 1 and 4 letters, 1 and 3 dial probability monitoring, two levels of spatial monitoring, an auditory reaction time, and a visual reaction time. At the time of completion of the summer fellowship, most of the study centered around the Pz lead on the head.

RESCALED RANGE ANALYSIS

The first approach to classifying EEG's was to calculate their fractal dimension. A literature search revealed that box counting algorithms had been used in previous research to find the fractal dimension of an EEG. However, the dimensions reported in the literature seemed to be contradictory. Katz² reports a dimension around 1.24, while Arle and Simon³ state a dimension of 1.7. Considering both studies were recorded for a normal awake adult and that fractal dimensions of a curve must be between one and two, these dimensions are on opposite sides of the range. We have examined similar EEG using box counting and found that the final plot did not yield a well-defined linear region, and therefore no well-defined fractal dimension.

Box counting seems inadequate when analyzing EEG's. There are a few reasons for this being the case. Box counting is generally used for curves with the same type of dimension on both axes. A classic example would be a coastline where the dimension would be a length on both sides. An EEG complicates matters by needing a box with a voltage dimension on the vertical edge, and a time dimension on the horizontal edge. Another reason box counting might produce ambiguous results stems from the fact that an EEG is a temporal record. The programmer must decide the temporal length of the record to be analyzed. A short record would be comparable to stretching out a curve to give a dimension close to one. A long record would be comparable to compacting a curve to make it space filling and giving it a dimension close to two.

The inadequacies of box counting generated a search for a better method to compute the fractal dimensions of an EEG. Since the method of calculating fractal dimensions varies depending upon the type of curve being analyzed, it is important to classify the EEG curve type. The EEG signal is a function of two variables, time and voltage. It is possible that the signal may be better described as a generalized homogeneous function, depending on whether or not it follows the appropriate scaling law. It is improbable that a record looks exactly the same on every scale since it is a physical measurement and not a mathematical fractal. It is more likely that the EEG exhibits statistical scaling laws. The Gaussian distribution is one example of statistics that follow the generalized homogeneous scaling law. The distribution is given by

$$p(\xi, \tau) = \frac{1}{\sqrt{4\pi D\tau}} \exp\left(-\frac{\xi^2}{4D\tau}\right). \quad (4)$$

When replacing ξ with $b^{1/2}\xi$ and τ with $b\tau$, the scaling relation for the probability density goes like

$$p(b^{1/2}\xi, b\tau) = b^{-1/2} p(\xi, \tau). \quad (5)$$

Here, the ξ are considered to be a set of random variables, τ is a microscopic time and D is the diffusion coefficient. The exponent of b on the right hand side of equation 5 is considered to be the scaling exponent of the statistics and for this case is one-half. Another change of variables allows the scaling exponent to be a number other than one-half. This change occurs when

$$\xi = \xi\left(\frac{|t-t_0|}{\tau}\right)^H. \quad (6)$$

We will call the new variable ΔX and $t-t_0$, we will call Δt . The probability distribution now becomes

$$p(\Delta X(t)) = \frac{1}{\sqrt{4\pi D \left(\tau \left(\frac{|\Delta t|}{\tau}\right)^{2H}\right)}} \exp\left(-\frac{\Delta X^2}{4D \left(\tau \left(\frac{|\Delta t|}{\tau}\right)^{2H}\right)}\right). \quad (7)$$

Again, replacing ΔX with $b^H \Delta X$ and t with bt , the scaling relation for the probability density goes like

$$p(b^H \Delta X(bt)) = b^{-H} p(\Delta X(t)). \quad (8)$$

The scaling exponent is no longer one-half like it is for the Gaussian distribution, but becomes a scaling exponent called the Hurst exponent (H), which can vary between zero and one. Curves that follow this type of scaling behavior are known as statistically self-affine curves.

One way to discover whether or not EEG's are statistically self-affine is to see if they follow the appropriate type of scaling law. Mandelbrot and Wallis¹ propose a specific scaling law for these curves of the form

$$\left\langle \frac{R(t, \tau)}{S(t, \tau)} \right\rangle \propto \tau^H, \quad (9)$$

where t is the time, τ is the length of the record, R is the range, S is the standard deviation, and H is the scaling or Hurst exponent. $\langle R/S \rangle$ is the average rescaled range, and the method using the above scaling law to interpret curves is called rescaled range analysis. If the EEG turns out to be statistically self-affine, then the fractal dimension of the EEG can be computed by taking $2-H$.

A program was therefore developed to carry out rescaled range analysis. The algorithm was tested first to ensure that it gave the proper results. A cosine curve yielded the expected Hurst exponent of one, and random numbers yielded a Hurst exponent close to the theoretical value of one-half. This brings up another advantage of rescaled range analysis. It has been suggested by Pijn et. al.⁴ that an EEG could either be due to chaos or noise. According to Mandelbrot¹, many types of noise (hyperbolic non-Gaussian noise, log normal random function...) exhibit the same scaling exponent as the Gaussian distribution, namely one-half. Rescaled range analysis on deterministic chaotic models which behave like random signals such as the Lorenz model and the logistic map have given results that do not equal one-half. If the EEG is dominated by noise then rescaled range analysis should give an exponent of one-half. The program was then used to analyze EEG's. The initial

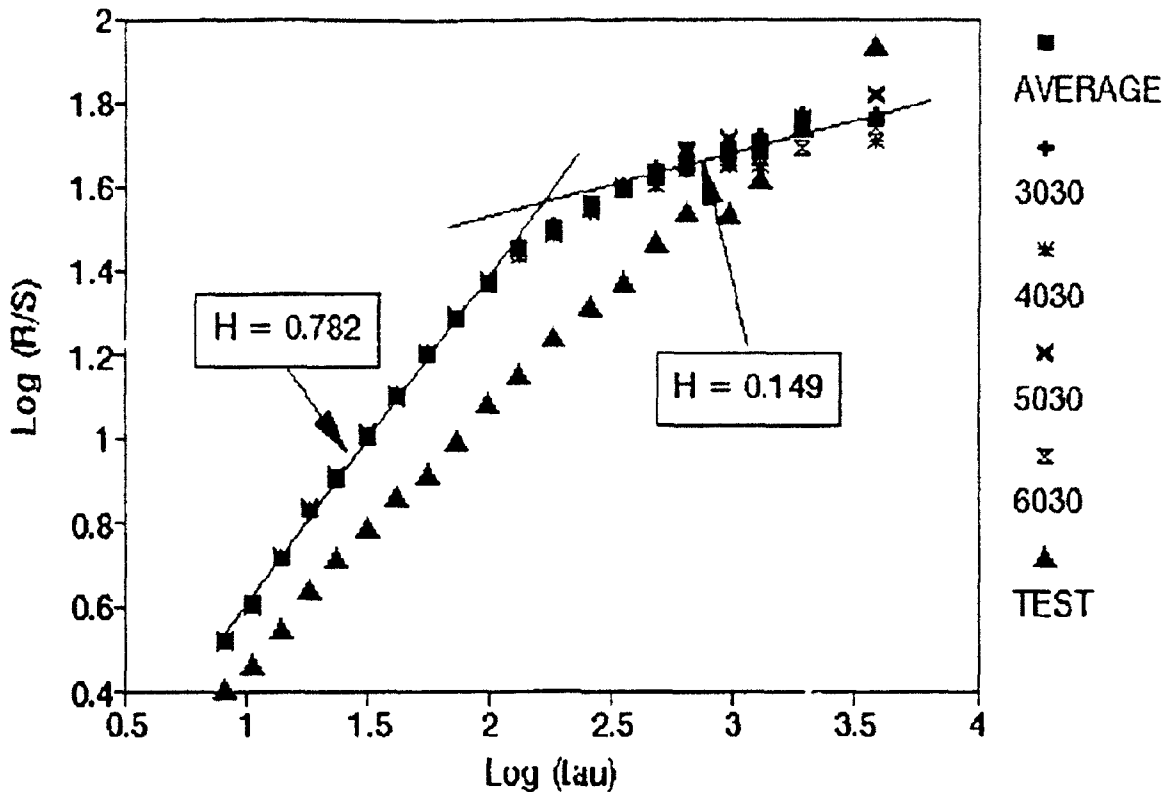


Figure 1. The average result of rescaled range analysis on four 30 second segments of MSV415 are designated by the filled box. The triangle symbol represents the results for random numbers with corresponding slope of one-half. All other symbols show the individual results of each segment of MSV415.

results of rescaled range analysis on an EEG showed two separate scaling regions (Figure 1).

The two regions could have occurred for a variety of reasons. The routine itself may have put the break into the data. As can be seen in figure 1, the test data revealed no such break. It should be noted that when the EEG voltages were put into a random order and then analyzed the break disappeared and the Hurst exponent became one-half as expected. The algorithm probably did not create the break in the results. A second possibility would be that the device(s) used to record the EEG put in the break. Discussions with the people at WPAFB has led us to believe this is not the case. A third reason for the break is that it is characteristic of the EEG. The break occurred in all the EEG's we examined, and it occurred at roughly the same place in each record (Figure 2).

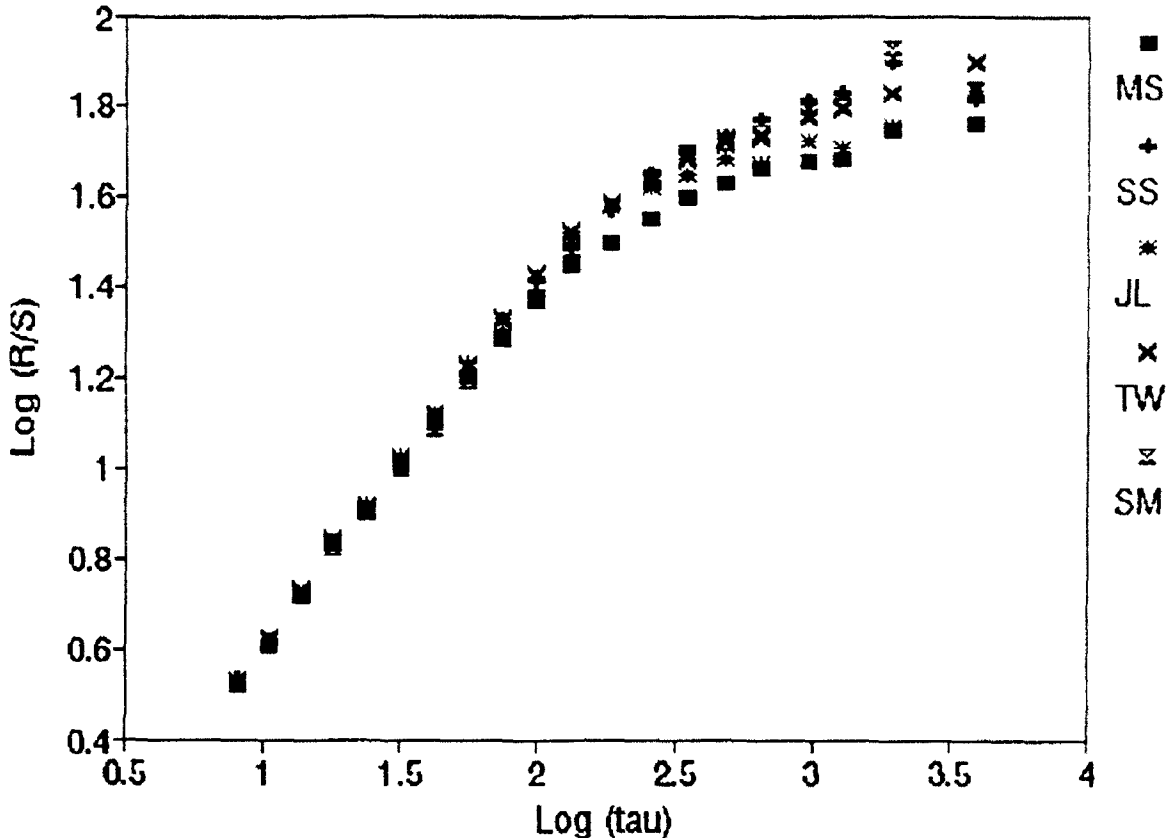


Figure 2. Rescaled range analysis results for a visual Sternberg of 4 letters, on 5 subjects (MS, SS, JL, TW, SM).

It is therefore believed that the break is characteristic of all EEG's.

Our rescaled range analysis results indicate that an EEG is well described as a self-affine function. A signal exhibits two scaling regions, each with its own Hurst exponent or fractal dimension. It is also interesting to note that neither slope is one-half, and therefore the EEG signal is probably not best described as noise. As a diagnostic tool rescaled range analysis breaks the EEG's into two groups. One group has fractal dimensions similar to the visual Sternbergs, the other group has fractal dimensions close to the auditory Sternbergs. Rescaled range analysis falls short of categorizing the EEG's any further for two reasons. First, in examining rescaled range analysis over 30 second subsections of longer EEG, we found that variation in the results were larger than was hoped. Although averages of rescaled range analysis may show variation from one task to another, a single 30 second segment from a particular task may yield ambiguous results when compared with all of the averages (Figure 1). The results also show more variation from person to person than there is

from one task to the next. There is also no discernable trend in the fractal dimension from one task to another. For example, the lower Hurst exponents for subject SS increases from visual Sternberg of one letter to visual Sternberg of 4 letters, while subject JL Hurst exponents decreases over the same interval (Figures 3 and 4).

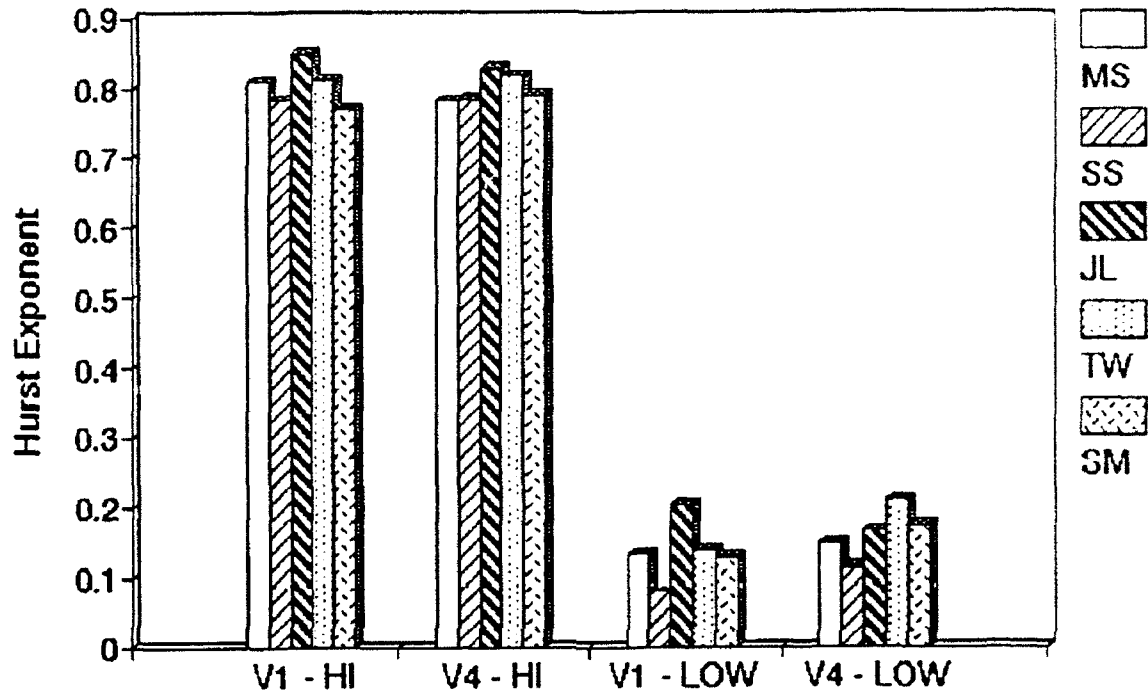


Figure 3. Hurst exponents of visual Sternbergs one and four letters, across five subjects.

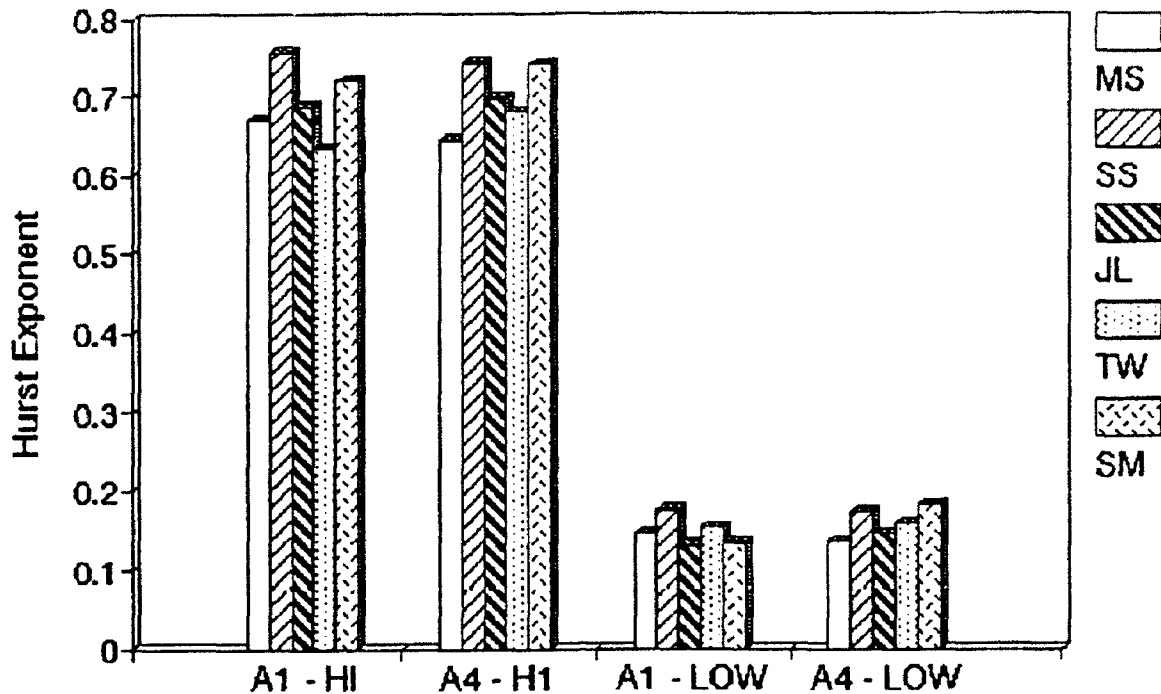


Figure 4. Hurst exponents of auditory Sternbergs of one and four letters, across five subjects.

MULTIFRACTAL ANALYSIS

The second approach to the study of the fractal nature of EEG was to examine the multifractal spectrum of the information. The development of this approach is due to many authors (see, for example Halsey⁵), and has been used to characterize fluid flows⁶, Josephson junctions^{7,8}, and has been applied to mathematical models which exhibit a quasiperiodic transition to chaos such as the circle map⁵. An early application of multifractals can be found in the description of the "curdling" of a Cantor set. In this process, the initiator of the Cantor set is assumed to have a density (mass/length), and that the generator

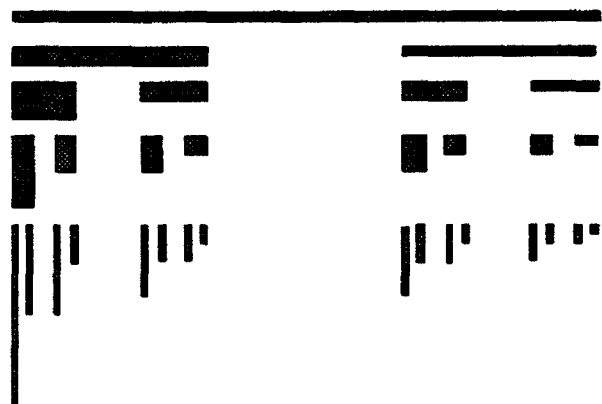


Figure 5. The curdling of the Cantor set.

preserves the total mass of the rod. A particular example of a curdled Cantor set is shown in figure 5. The initiator is a line segment of length 1 and mass m_0 . The generator cuts the line at $L=2/3$ and the "hammers" the longer segment down to a length $L=1/3$. Thus the generator creates two lines of equal length ($1/3$), but unequal densities ($\rho_1=2m_0/L$ and $\rho_2=m_0/L$). This procedure is continued indefinitely. Since the mass in each segment remains finite while the length of each segment goes to zero, the densities of each segment will diverge, but the density of each segment will diverge at a rate characteristic of the segment. It should be pointed out that the figure can be misleading in that the set is considered to always lie along a one dimensional line. The vertical length of each bar actually represents the density within that segment, with a longer vertical line indicating a higher density. Also note that box counting on the resulting set will produce a dimension of 0.631 (the dimension of the Cantor middle third set), but the information contained in the curdled set is much richer than that of the simple Cantor set.

Multifractal analysis takes into account the densities (or specifically the "local" rates of divergence of the densities) by decomposing the set into fractal subsets. Each of these subsets have points which are characterized by densities which have the same rate of divergence. Each subset S_α (where iso- α sets diverge at the same rate) has a fractal dimension $f(\alpha)$. A plot of $f(\alpha)$ vs α is called the *multifractal spectrum*, or the *spectrum of singularities*.

The relation to EEG stems from the use of threshold cuts which can be used to reduce the EEG to a point set, which can be studied as if it were a truncated Cantor set. The process is illustrated in figure 6. The EEG is shown on the top, while the intersection of the EEG with the voltage=0 line is shown in the middle. The variation in local density of the point set is clearly evident. The point set is divided into (in this example) 100 bins, and a density is computed by adding the number of points which appear in each of the bins. These densities are plotted in the lower portion of the figure.

The behavior of the densities as the number of bins increases determines the iso- α subsets. Typically, the fractal dimension of each of the subsets lies between 0 and 1, since the set is a point set. A general characteristic of the $f(\alpha)$ spectrum is that

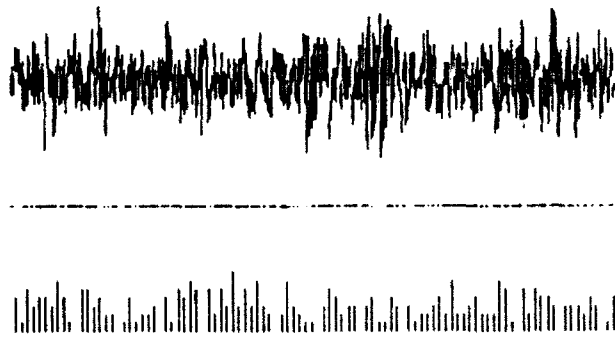


Figure 6. The top represents one of the EEG used in the study. The middle figure depicts the location of zero-voltage threshold crossings. The lower portion shows the density of points contained in 100 bins.

$$\frac{\partial^2 f(\alpha)}{\partial \alpha^2} < 0, \quad (10)$$

and that $f(\alpha)$ has a single maximum, so that the $f(\alpha)$ vs α plot behaves roughly as an inverted parabola, with a maximum at some value α_0 . The details of the shape of this curve for specific experiments have been compared with those of known mathematical models in an effort to model very complex behaviors⁶.

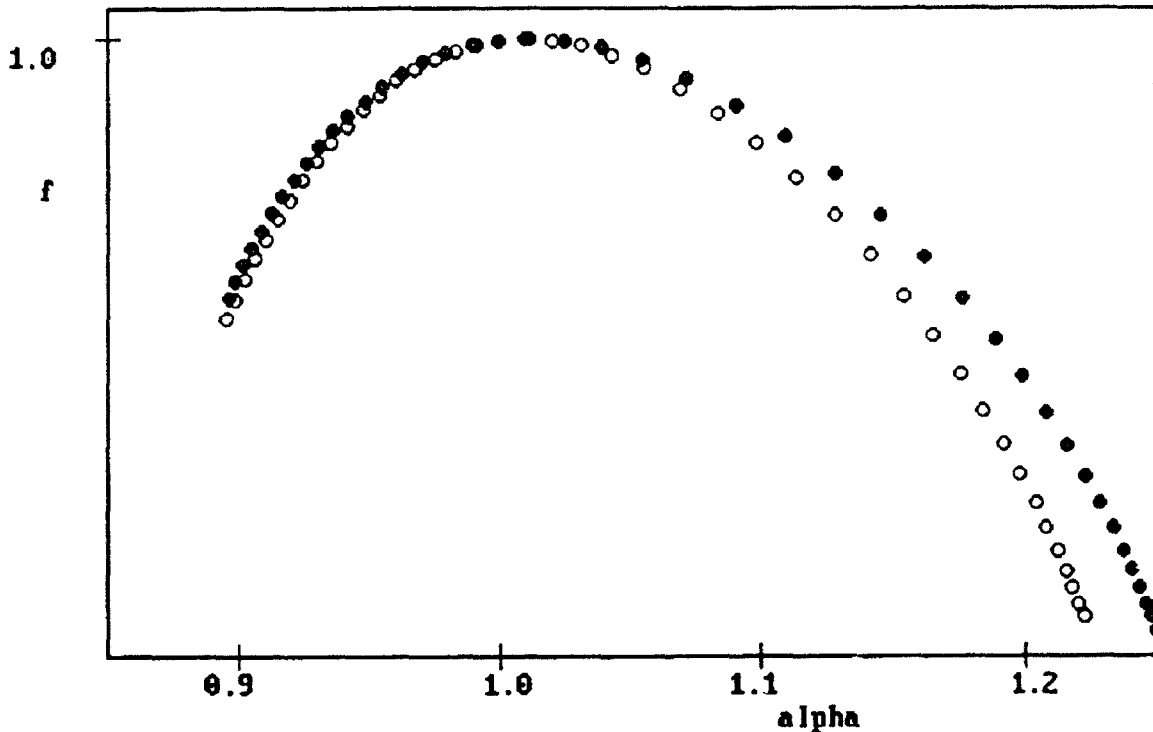


Figure 7. Multifractal spectrum for EEG recorded under two task conditions. The open circles represent a spatial monitoring task of relatively low difficulty, while the filled circles represent a spatial monitoring task of higher difficulty.

A typical result for an EEG is shown in figure 7. Two separate records are represented in this figure. The two plots are averages taken over four 30 second subsections of the EEG. The spectrum exhibits the typical "inverted parabola" shape of the $f(\alpha)$ vs α curve, with some variation evident between the two tasks. In order to parameterize the shape of this curve, a Taylor expansion for $f(\alpha)$ should be carried out,

$$f(\alpha) = a_0 + a_1(\alpha - \alpha_0) + a_2(\alpha - \alpha_0)^2 + \dots, \quad (11)$$

with the coefficients a_i acting as numerical comparisons from one task to the next. The higher order terms are required since the curve is not exactly parabolic.

Although the spectrum of the EEG studied all exhibit the same universal shape, stationarity of the signal still poses a significant problem in the analysis. Once again, although there is some task dependence in the averaged results, individual subrecords can yield ambiguous spectra for comparison.

CONCLUSIONS

We have been able to identify what would appear to be universal features in the scaling laws associated with EEG. It has been possible to determine the scaling behavior of EEG across various subjects and tasks. Results of rescaled range analysis show two typical scaling behaviors, each associated with its own Hurst exponent. Although the dynamical model of the EEG is not known, the EEG is not dominated by random noise since there are no indications of an $H = 0.5$ behavior. In general, situations where the primary inputs are auditory can be distinguished from situations where the primary inputs are visual by the values of H . It is unclear whether or not the task(s) are a control parameter, or if the scaling behavior of the EEG is only a weak function of task. Specific results associated with cognitive task were ambiguous. Similar results were produced for the multifractal analysis. Characteristic spectra of universal shape were produced for all subjects studied, but ambiguities in the results make it difficult to assess the utility of treating cognitive tasks as control parameters.

Further research could be performed to determine if task is a control parameter. During the course of the summer, simple signals with varying signal to noise ratios were studied. The goal was to determine the effect of increasing amounts of noise on the scaling behavior of the statistics of the signal. Well defined signals (non-chaotic) with no noise exhibit a major break in its scaling behavior at record length which corresponds to the longest periodicity in the data. With no noise, the Hurst exponent for record lengths longer than this periodicity is approximately zero. As the signal to noise ratio is increased, the Hurst exponent for records of length longer than the longest periodicity grows, until, once the signal is dominated by noise, the exponent approaches the random noise value of 0.5.

Our present interpretation of the EEG over all subjects and task conditions studied is that the scaling behavior of short records (approximately 2-3 seconds) is dominated by a deterministic signal (since $H \neq 0.5$). The change over in scaling behavior which occurs at records longer than 2-3 seconds indicate some basic periodicity which we do not yet understand. The values of H for these longer records (generally greater than zero but less than 0.3) may result from the signal to noise ratio of the deterministic signal and background noise. This could be determined by examining the EEG of a subject who has been placed in a situation which could introduce a larger amount of randomness in the electrical activity. One possibility would be to force the subject to perform task(s) that it was unable to perform. This could produce an extreme case where the

cognitive task control parameter is driven to an "extreme" value. If no significant alteration is produced in the scaling behaviors, there would be little hope of using scaling behaviors for assessing cognitive tasks. We have no such information currently, but it would be worthwhile in the future to examine such EEG if the data can be obtained.

Finally, there was not enough time during the summer research period to examine all 21 leads for all subjects available. It is possible that information collected from only one lead is insufficient to properly characterize variations in the scaling behaviors. Further research should be carried out to assess the spatial dependence of scaling in the EEG across the head.

ACKNOWLEDGMENTS

The authors would like to thank both RPL and AFOSR for their support during the summer. The authors would also like to thank all of the members of the Crew Systems Division of Armstrong Laboratory at WPAFB for their kindness and hospitality during the summer. We would especially like to thank Maj. Al Badeau for all of the help he provided during the research period. Without his help, we would not have had the opportunity to conduct this research.

APPENDIX 1: A brief listing of the programs used in calculating scaling behaviors for EEG.

Rescaled Range Analysis

EEGBATCH.BAS : creates EEGBAT.BAT, rerun when change subject or tasks
RRA2BAT.BAS : creates RRA2.BAT, rerun when change subject or tasks
EEGBINFO.BAS : creates input files for EEGBAT.BAT, rerun when change subject or tasks
RRABIN2.BAS : creates input files for RRA2.BAT, rerun when change subject or tasks
EMB.BAT : runs EEGBAT.BAT followed by RRA2.BAT
EEGBAT.BAT : runs EEG.EXE which creates data files of different EEG's
RRA2.BAT : runs RRA3.EXE on 4 segments of each data file created by EEGBAT.BAT
to get rescaled range analysis on each segment of each EEG file
AVERS.BAS : averages the 4 segment results of RRA3 for each separate EEG
TAURS.BAS : gives best fit line for the average RRA3 result

Multifractals

MFBATCH.BAS : creates MF.BAT for 14 tasks of one subject and one lead
MFABATCH.BAS : creates MFA.BAT for 14 tasks of one subject and one lead
BATCHDIR.BAS : creates input files for MF.BAT
BATCHDA.BAS : creates input files for MFA.BAT
THRESH3.BAS : makes threshold cuts on each of 4 segments for each EEG to be used in
MFBOX.FOR
MF.BAT : runs MFBOX.FOR, D14R1.FOR, and D3R7.FOR for each segment and each
EEG
MFBOX.FOR : calculates

$$\text{Log} \left(\sum_{i=1}^n \mu_i^q \right) \text{ vs } \text{Log } \delta$$

MFLOT.BAS : plots the results of MFBOX.FOR
D14R1.FOR : calculates the slopes of the above plots to get $\tau(q)$ versus q
D3R7.FOR : converts τ and q into f and α
MFAVE.BAS : calculates the average of the data files created by MFBOX.FOR and stores
averages to be used in MFA.BAT
MFA.BAT : runs D14R1.FOR and D3R7.FOR on average mfbox data file

REFERENCES

1. Mandelbrot B. B., and Wallis J. R., "Robustness of the Rescaled Range R/S in the Measurement of Non Cyclic Long Run Statistical Dependence", *Water Resources Research*, 5 No. 5, 967 (1969).
2. Katz M. J., "Fractals an the Analysis of Waveforms", *Computer Biological Medical*, 18 No. 3, 145 (1988).
3. Arle J.E., and Simon R. H., "An Application of Fractal Dimension to the Detection of Transients in the Electroencephalogram", *Electroencephalography and Clinical Neurophysiology*, 75, 296 (1990).
4. Pijn J. P., Van Neerven J., Noest A., and Lopes Da Silva F. H., "Chaos or Noise in EEG Signals; Dependence on State and Brain Site", *Electroencephalography and Clinical Neurophysiology*, 79, 371 (1991).
5. Halsey T. C., Jensen M. H., Kadanoff L. P., Procaccia I., and Shraiman B. I., "Fractal Measures and Their Singularities: The Characterization of Strange Sets", *Phys. Rev. A*, 33, 1141 (1986).
6. Jensen M. H., Kadanoff L. P., Libchaber A., Procaccia I., and Stavans J., "Global Universality at the Onset of Chaos: Results of a Forced Rayleigh-Benard Experiment", *Phys. Rev. Lett.*, 55, 2798 (1985).
7. Jensen M. H., Bak P., and Bohr T., "Transition to Chaos by Interaction of Resonances in Dissipative Systems. I. Circle Maps", *Phys. Rev. A*, 30 No. 4, 1960 (1984).
8. Bohr T., Bak P., and Jensen M. H., "Transition to Chaos by Interaction of Resonances in Dissipative Systems. II. Josephson Junction, Charge Density Waves, and Standard Maps", *Phys. Rev. A*, 30 No. 4, 1970 (1984).

**TERMINAL ARTERIOLAR DENSITY, TOTAL PERIPHERAL RESISTANCE
AND OPTIMAL VENTRICULO-ARTERIAL COUPLING**

**Daniel L. Ewert
Assistant Professor
Department of Electrical Engineering**

**North Dakota State University
Fargo, ND 58105**

**Final Report for:
Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

TERMINAL ARTERIOLAR DENSITY, TOTAL PERIPHERAL RESISTANCE AND OPTIMAL VENTRICULO-ARTERIAL COUPLING

Daniel L. Ewert
Assistant Professor
Department of Electrical Engineering
North Dakota State University

Abstract

Three main areas were studied. These were 1) a proof-of-concept experiment to characterize the coronary artery pressure and flow relation as a function of occluded terminal arterioles in goat myocardium; 2) development of a mathematical technique to estimate total peripheral resistance (TPR) during unsteady conditions in a high $+G_z$ state and 3) a mathematical description to determine optimal ventriculo-arterial coupling based on the principle of maximum external work transfer.

The proof-of-concept experiment was performed and approximately 42 million, nonradioactive 15 μm microspheres were injected into the left ventricle myocardium. The relationship between coronary pressure and flow was a decreasing nonlinear function of occluded terminal arterioles.

Two equations were developed which predicted arterial capacitance and total peripheral resistance during transient $+G_z$ episodes. In addition it is possible to estimate the volume of blood that pools in the venous system during high $+G_z$ states.

An equation was developed which estimated the optimum ratio of arterial elastance to left ventricle elastance for maximum external work transfer from heart (left ventricle) to the arterial system.

TERMINAL ARTERIOLAR DENSITY, TOTAL PERIPHERAL RESISTANCE AND OPTIMAL VENTRICULO-ARTERIAL COUPLING

Daniel L. Ewert

Three main topics were researched: 1) A proof-of-concept experiment to characterize the coronary artery pressure and flow relation for goat myocardium. 2) Development of a new mathematical technique to estimate total peripheral resistance (TPR) during unsteady conditions in a high $+G_z$ state, and 3) A mathematical description to determine optimal ventriculo-arterial coupling based on the principle of maximum external work transfer. Each topic will be discussed in a separate section.

TERMINAL ARTERIOLAR DENSITY

Introduction:

Coronary artery disease is the most prevalent condition causing medical disqualifications among USAF aviators (Whilton, 1984). It was estimated that 10-20% of USAF aviators have some degree of coronary artery disease (Pettyjohn and Meckin, 1975). In 1980 the cost associated with personnel afflicted with coronary vascular disease was estimated to be 50 million dollars per year (DeHart; 1980).

Mild coronary artery disease may be exacerbated during $+G_z$ stress (Hickman and McGranahan, 1985). This exacerbation is most probably due to reduced coronary flow reserve. Hoffman (1984) defined coronary flow reserve as the difference between maximum vasodilated flow and autoregulated flow at a particular pressure.

The slope of the maximum vasodilated curve is the maximum conductance of the coronary circulation. The reciprocal of this conductance is resistance. If coronary artery disease is present, the resistance should increase (slope decreases).

As increased coronary flow is required, (increased metabolic state) the autoregulated curve moves upward to approach the maximum vasodilated curve. This is accomplished by vasodilating the arterial side and by recruitment of additional flow passages by opening terminal arterioles. Thus, the theoretical maximum autoregulated flow will eventually approach the maximum vasodilated flow. In practice this is not exactly true (Belloni 1979, Canty and Klocke 1985).

What is not well understood is how the opening and closing of terminal arterioles affect the transformation of autoregulated flow to maximally dilated flow. The research aim was to characterize the relation between the coronary pressure, flow and the number of occluded terminal arterioles.

It was expected that the relation between the number of closed terminal arterioles and the steady coronary flow would be a nonlinear, monotonically decreasing function due to the parallelism and connectivity of the coronary circulation.

Methods:

An isolated heart preparation was assembled as shown in Figure 1. A temperature-controlled bath was maintained at 37°C. A modified Tyrode's solution (115.0 mM NaCl, 15 mM KCl, 1.2 mM MgCl₂ · 6H₂O, 2 mM NaH₂PO₄ · H₂O, 4.0

mM Na-Lactate, 1.2 mM Na SO₄, 27 mM NaHCO₃, 1.5 mM CaCl₂, 11.1 mM glucose, 95% O₂ - 5% CO₂) was used. Adenosine (A-9251, Sigma, St. Louis, MO) was added to the solution (100 μM) to dilate the coronary arterial system. The elevated potassium concentration caused the heart to arrest in diastole.

A goat heart was obtained by an open chest technique according to an animal use committee approved protocol. The pericardium was opened and chilled Tyrode's solution was poured over the heart until the heart rate dropped. Then the heart was cut free and placed in a beaker of chilled Tyrodes.

Next, the heart was hung on the assembly in Figure 1 and perfused with warm, oxygenated Tyrodes. An electromagnetic flowmeter was placed on the left main coronary artery.

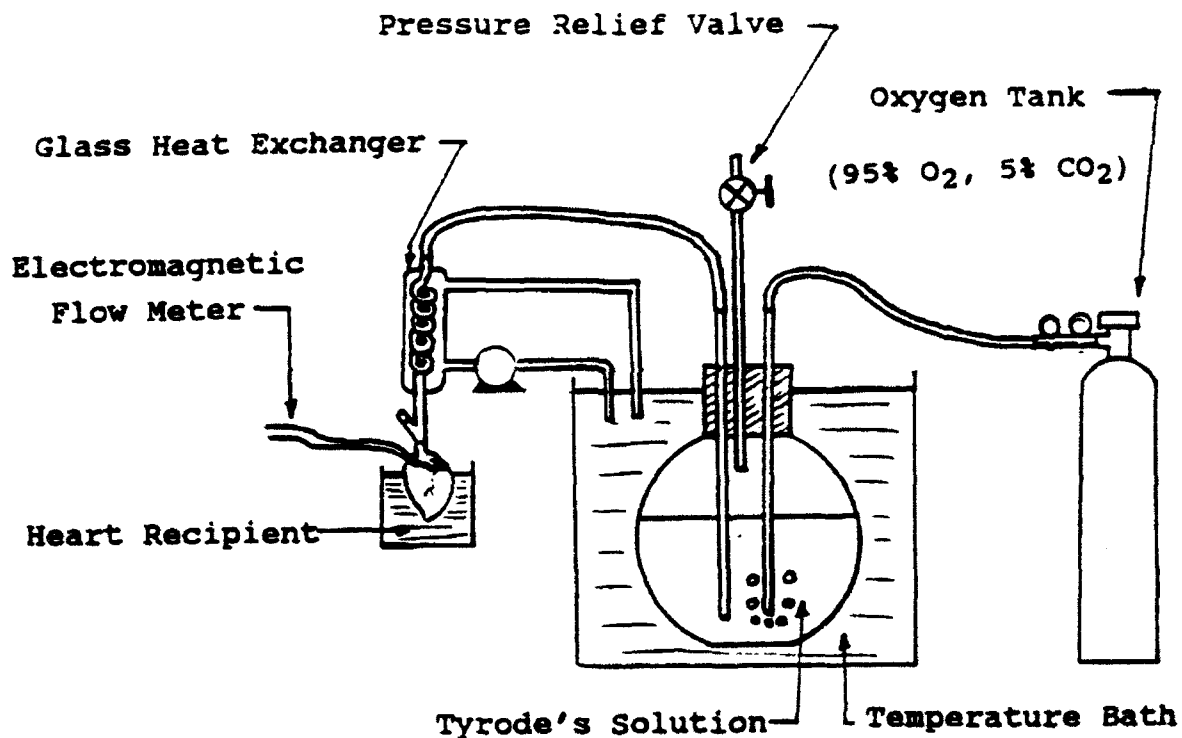


Figure 1: Isolated Heart Preparation

After the pressures and flows equilibrated, 15 μm nonradioactive microspheres of different colors were injected to occlude the myocardium's terminal arterioles. Pressures and flows were recorded for each injection. After the experiment the heart was placed in formaldehyde and 5 serial, 15 μm thick histological sections were taken in the left circumflex, left anterior descending, and septal regions of the left ventricle.

Results and Discussion:

The outcome of the experiment is shown in Figure 2. As expected, the coronary flow decreased as more microspheres were injected and the higher the coronary artery perfusion pressure, the larger the coronary flow. The flow appears to remain relatively constant at large microsphere numbers. This may be due to collateral flow or perhaps the microspheres do not totally occlude the terminal arterioles. Also it may be that more than one microsphere may occlude a terminal arteriole. Therefore, 5 serial, 15 μm thick histological sections were taken to estimate the "packing factor" - the average number of microspheres occluding a terminal arteriole. At the time of this report, the slides were unavailable. Because the aim of the study was to perform a proof-of-concept experiment, only one heart was done.

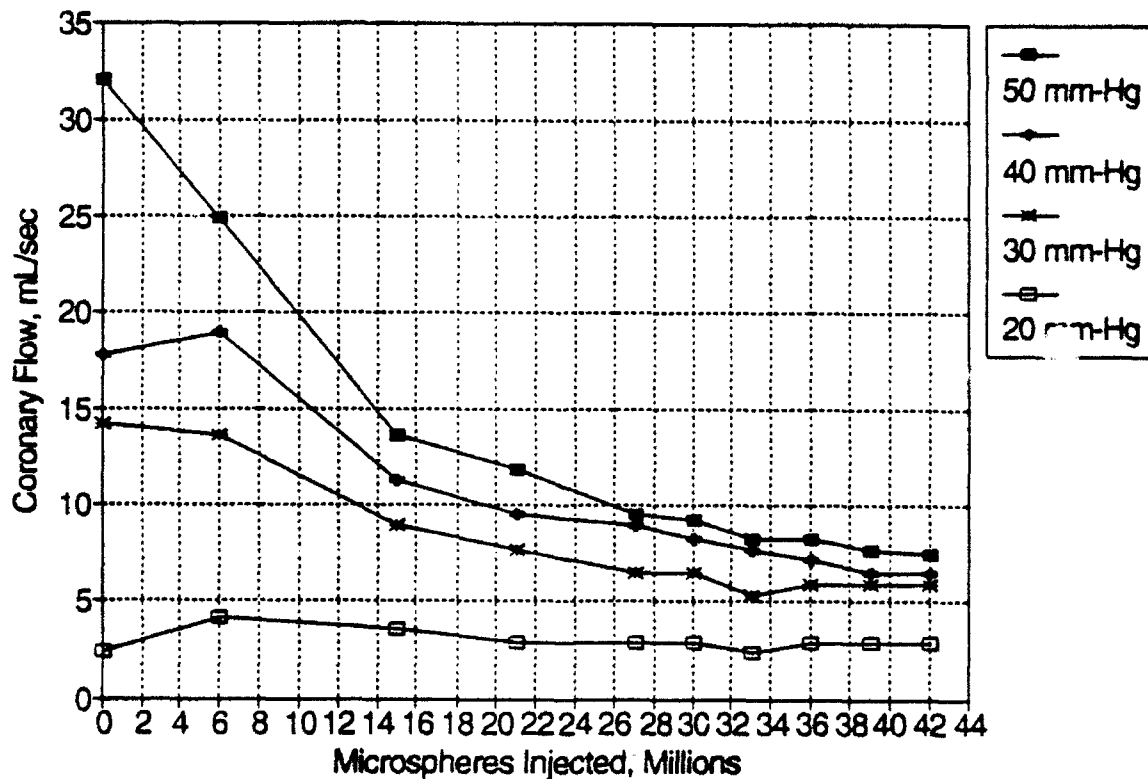


Figure 2: Pressure, Flow and Microspheres Injected Relation

TOTAL PERIPHERAL RESISTANCE

Introduction:

As the cardiovascular system is subjected to transient $+G_z$, the usually steady nature of the aortic pressure and flow waveforms becomes highly unsteady. When the waveforms become unsteady the traditional approach for calculating total peripheral resistance (TPR), (mean aortic pressure/mean aortic flow), yields inaccurate TPR estimates. Furthermore, for the high $+G_z$ environment, the blood volume tends to pool in the venous system - especially without G_z pants. Thus, the objective of this study was to find a method to estimate TPR and blood volume increases in the venous circulation.

Other investigators have developed methods to estimate TPR. Perhaps Toorop et.al (1987) best characterized unsteady TPR by using a Newton fitting procedure to find optimal values of TPR and arterial compliance. However, blood volume changes in the venous circulation were not addressed. Liu et. al (1986) developed a method for estimating TPR that uses an aortic pressure waveform and stroke volume (SV). In addition, information about the diastolic portion of the curve must be known. Furthermore, in their calculations, to obtain arterial compliance, TPR must be calculated and its value is based on the mean aortic pressure/mean aortic flow rendering this approach ineffective to unsteady beat analysis. The aim of this study was to develop a model and equations that would estimate TPR, arterial compliance and venous blood volumes.

Methods:

Shown in Figure 3 is the electrical analog model used to develop the mathematical formulations.

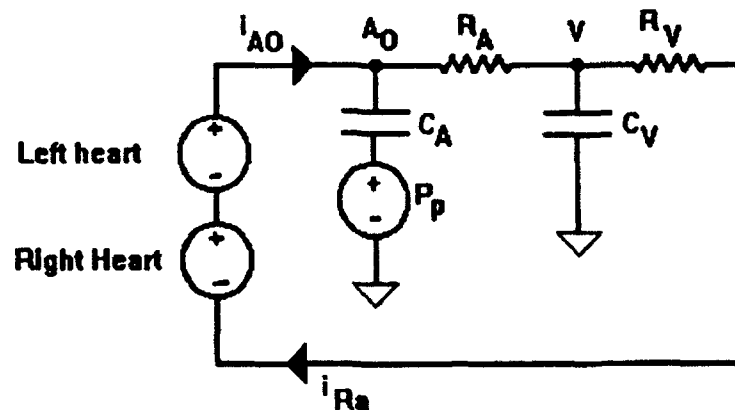


Figure 3: Circuit Analog of Systemic Circulation

- Definitions:
- i_{AO} = aortic flow
 - C_A = "Lumped" arterial capacitance
 - R_A = "Lumped" arterial resistance
 - C_v = "Lumped" venous capacitance
 - R_v = "Lumped" venous resistance
 - i_{RA} = Right heart flow
 - P_p = Pleural pressure

Assumptions:

- (1) C_A, R_A, C_V, R_V , are constant over one beat
- (2) $R_A \gg R_V$ ($PR_A \approx P_V$)
- (3) Total blood volume = constant

Applying Kirchoff's current law at node A_O yields the following equation.

$$i_{ao} = C_A \frac{d(P_{ao} - P_p)}{dt} + \frac{P_{ao} - P_v}{R_A}$$

i_{Ao}, P_{ao}, P_p and P_v are experimentally measured.

C_A and R_A are unknowns.

Applying Kirchoff's current law at node V yields the following equation.

$$i_{RA} = \frac{P_{ao} - P_v}{R_A} - C_V d \frac{P_v}{dt}$$

P_{ao}, P_v, P_p and i_{Ra} are experimentally measured

C_V and R_A are unknowns

Next, integrate both equations over two intervals

$$\int_{t_0}^{t_1} i_{ao} dt = C_A [(P_{ao}(t_1) - P_p(t_1)) - (P_{ao}(t_0) - P_p(t_0))] + \frac{1}{R_A} \int_{t_0}^{t_1} (P_{ao} - P_v) dt$$

$$\int_{t_1}^{t_2} i_{Ao} dt = C_A [(P_{ao}(t_2) - P_p(t_2)) - (P_{ao}(t_1) - P_p(t_1))] + \frac{1}{R_A} \int_{t_1}^{t_2} (P_{ao} - P_v) dt$$

$$\int_{t_0}^{t_1} i_{Ra} dt = \frac{1}{R_A} \int_{t_0}^{t_1} (P_{ao} - P_v) dt - C_V [P_v(t_1) - P_v(t_0)]$$

$$\int_{t_1}^{t_2} i_{Ra} dt = \frac{1}{R_A} \int_{t_1}^{t_2} (P_{ao} - P_v) dt - C_V [P_v(t_2) - P_v(t_1)]$$

Thus for arterial side there are two equations and two unknowns and for the venous side one obtains two equations and two unknowns.

Results and Discussion:

A short duration $+G_z$ run was conducted on fully instrumented baboons. Shown below are the results of TPR analysis. Due to instrumentation problems (pulmonary flow) the venous system equation could not be checked. The results are shown in Figures 4, 5 and 6. In all cases (3,6, and 8 $+G_z$) the TPR fell at onset of $+G_z$. The fall is thought to be due to the larger hydrostatic pressures increasing the diameter of the resistance vessels. The greater the $+G_z$, the greater the drop in resistance. After about 4-5 secs the resistance begins to increase when, it is thought that, the baroreceptor reflex is activated. After the high $+G_z$ episode the resistance begins to decrease, probably due to combined effects of a lower hydrostatic pressure and reduced baroreceptor reflex activity.

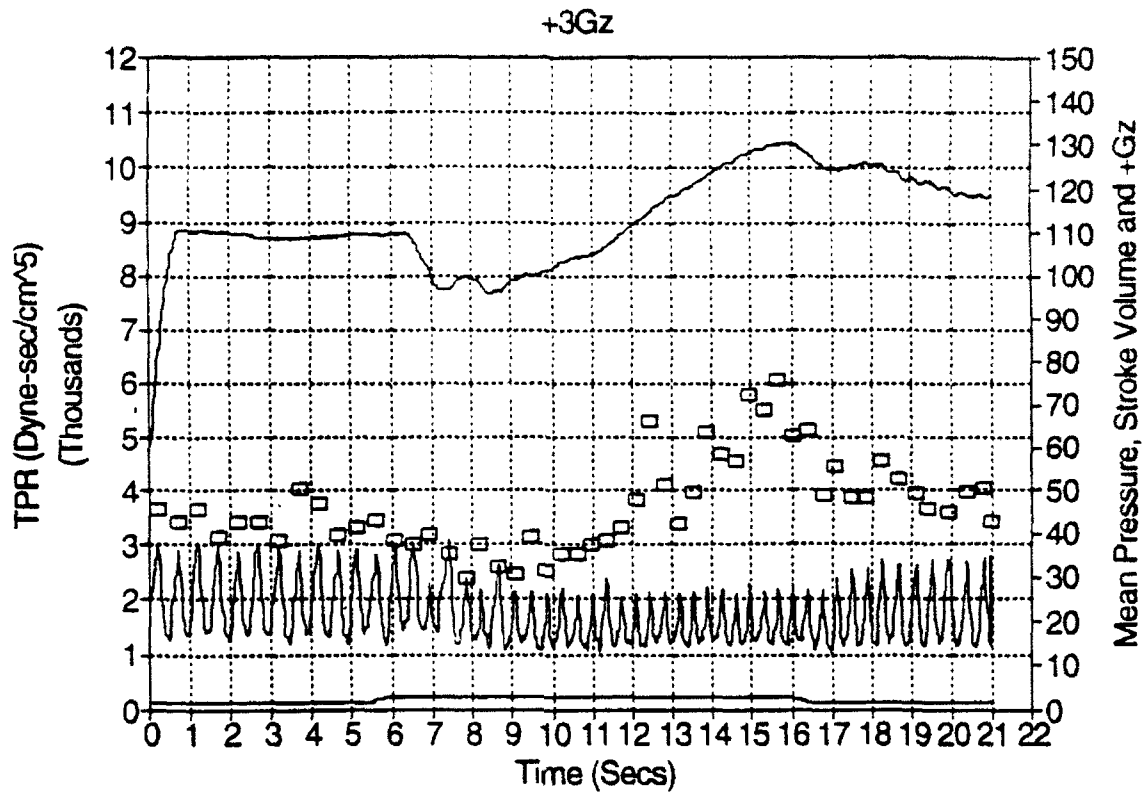


Figure 4: 3G_z Total Peripheral Resistance

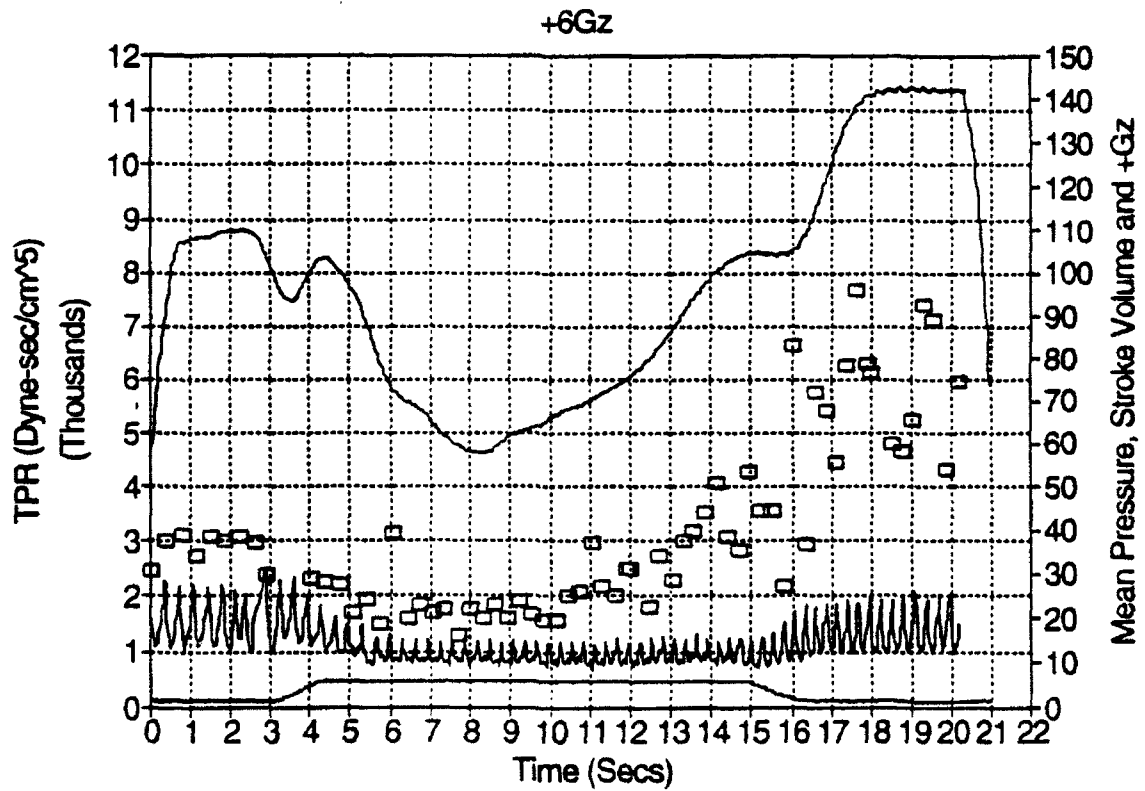


Figure 5: +6G_z Total Peripheral Resistance

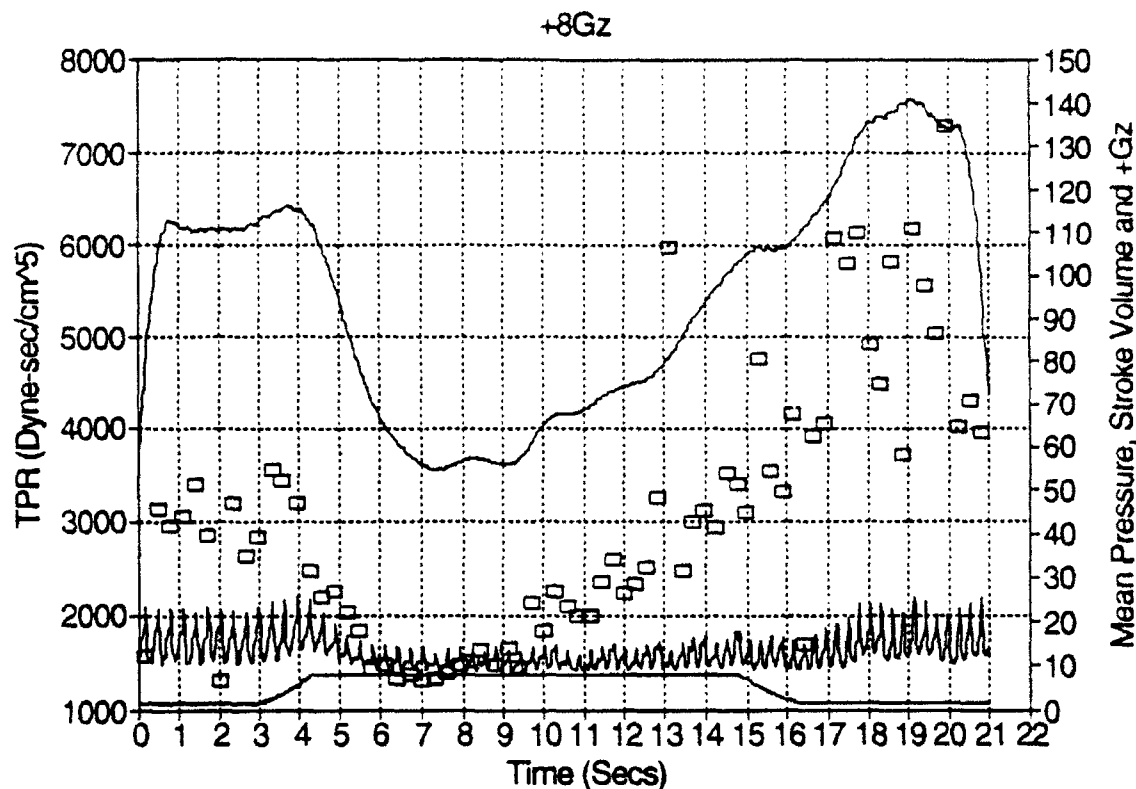


Figure 6: +8G_z Total Peripheral Resistance

Choosing the optimal intervals has not been fully investigated due to time constraints. However, it appears that the interval of time t_0 to t_2 can be taken as one cardiac cycle. The interval t_0 to t_1 should extend from the start of aortic pressure increase to maximum aortic pressure. The interval t_1 to t_2 should range from maximum aortic pressure to the end of the cycle.

VENTRICULO-ARTERIAL COUPLING

Introduction:

Two mechanisms, maximum external work (EW) transfer and maximum mechanical efficiency, have been proposed to explain optimal coupling between the left ventricle and arterial circulation (Elzinga and Westerhof, 1980, Sunagawa et. al., 1983). This study focuses on the maximum external work transfer mechanism, whose foundation was proposed by Kubota et. al.(1992). Their analysis showed that to maximize EW transfer, the end-systolic elastance of the left ventricle should equal the "effective" arterial elastance. It should be noted that this "effective" arterial elastance does not represent the lumped physiological arterial elastance (Sunagawa et. al., 1983).

Much scientific literature has been generated that makes use of the unphysiological "effective" arterial elastance which limits the usefulness and interpretations of the method.

The objective of this study is to define the physiological arterial elastance for maximum EW to be transferred from the left ventricle to the arterial system.

Method:

Shown in Figure 7 is the electrical analog model used to develop the mathematical formulations.

Model:

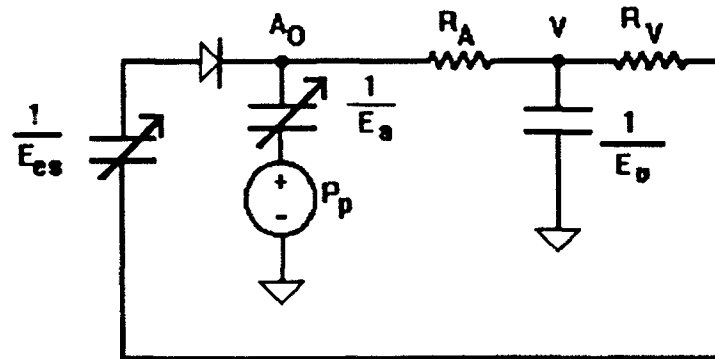


Figure 7: Optimal Coupling Circuit Analog

Definitions:

- E_{es} = "Lumped" elastance of left ventricle
- E_a = "Lumped" elastance of arterial system
- P_p = pleural pressure
- P_{ao} = aortic pressure
- R_a = Peripheral resistance
- E_v = venous elastance
- R_v = venous resistance

Assumptions:

- $R_A \gg R_v$
- E_a is a constant over one beat
- Neglect E_v
- R_A is a constant over one beat
- E_{es} is a constant over ejection
- P_v is a constant
- P_p is a constant
- Neglect passive properties of left ventricle
- Ideal diode

Shown in figures 8 and 9 are graphical depictions of left ventricle elastance and arterial elastance.

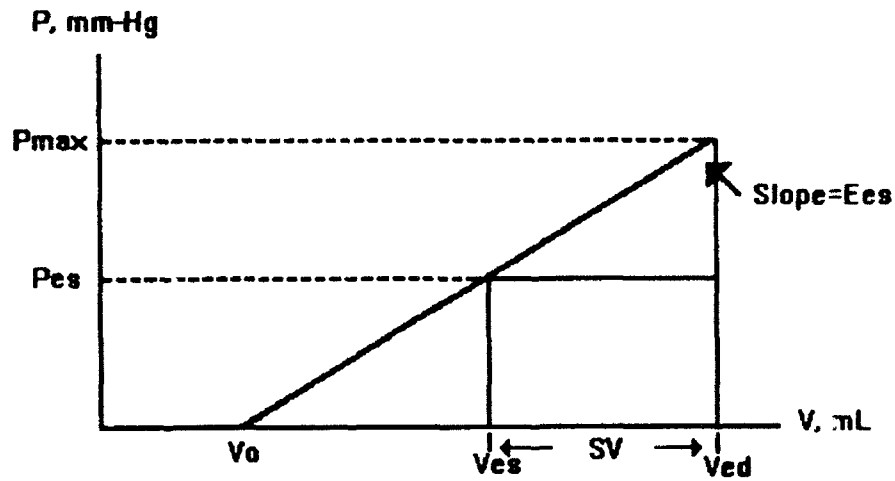


Figure 8: Left Ventricle Elastance

where:

P_{max} is the isovolumic maximum pressure

P_{es} is the end systolic pressure

V_{ed} is the end diastolic volume

V_{es} is the end systolic volume

$SV = V_{ed} - V_{es}$; stroke volume

V_0 = volume in left ventricle when pressure is zero.

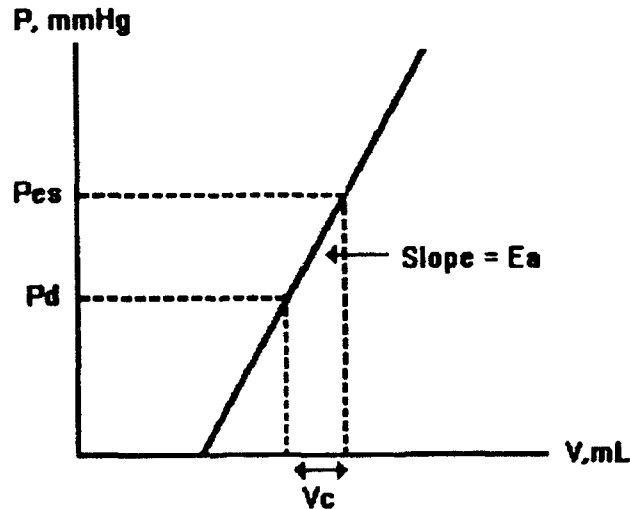


Figure 9: Arterial Elastance

where:

P_d is the aortic pressure at end diastole.

V_c is the volume of blood injected into the arterial capacitance.

The stroke volume ejected from left ventricle must either be stored in the arterial capacitor or be passed on to the venous system.

$$SV = V_c + V_R$$

where:

$$V_c = \frac{1}{E_a} (P_{es} - P_d)$$

$$V_R = \frac{1}{R_A} \int_{Tej} (P_{ao} - P_v) dt$$

Tej = ejection time interval

Shown below are the elastance formulations for left ventricle and arterial system.

$$E_{es} = \frac{P_{es}}{V_{ed} - SV - V_o}$$

$$E_a = \frac{P_{es} - P_d}{V_c}$$

Substituting the above equations into $EW = P_{es} \cdot SV$ and maximizing EW with respect to E_a , differentiating EW and setting it to zero, it can be shown that the optimal elastance ratio for maximum external work transfer is:

$$\frac{E_a}{E_{es}} = \frac{\left(\frac{V_R}{V_{\max}} - 1 \right) + \frac{P_d}{P_{\max}} \left[3 - 2 \left(\frac{V_R}{V_{\max}} + \frac{P_d}{P_{\max}} \right) \right]}{\left(1 - 2 \frac{V_R}{V_{\max}} \right) \left(\frac{P_d}{P_{\max}} + \frac{V_R}{V_{\max}} - 1 \right)}$$

where $V_{\max} = V_{ed} - V_o$

If $V_R \ll V_{\max}$ then

$$\frac{E_a}{E_{es}} = \frac{1 - 2 \frac{P_d}{P_{\max}}}{1 - \frac{T_{ej} P_a E_{es}}{R_a (P_{\max} - P_d)}}$$

Where $\overline{P_a}$ is the average aortic pressure over a beat.

Results and Discussion:

Shown in Figure 10-15 are graphical representations of the long and short equations for some typical physiological situations.

Note that only for a few specific conditions will the optimum ratio of E_a to E_{es} be unity as suggested by Kubota et.al. (1992). This is because this analysis uses a "lumped" physiological arterial elastance rather than a non-physiological "effective" arterial elastance. The optimum ratio depends on how much volume flows through the peripheral resistance during ejection. The volume that flows through the peripheral resistance depends on the value of resistance. Thus as TPR decreases the elastance will increase - less volume will be stored in the arterial elastance. Figures 10-12 demonstrate the effect of higher aortic diastolic pressures - the higher the diastolic pressure the less rigid the arterial elastance. Figures 13-15 demonstrate the effect of larger end diastolic volumes. As the end diastolic volume increases the smaller the elastance at lower TPR values.

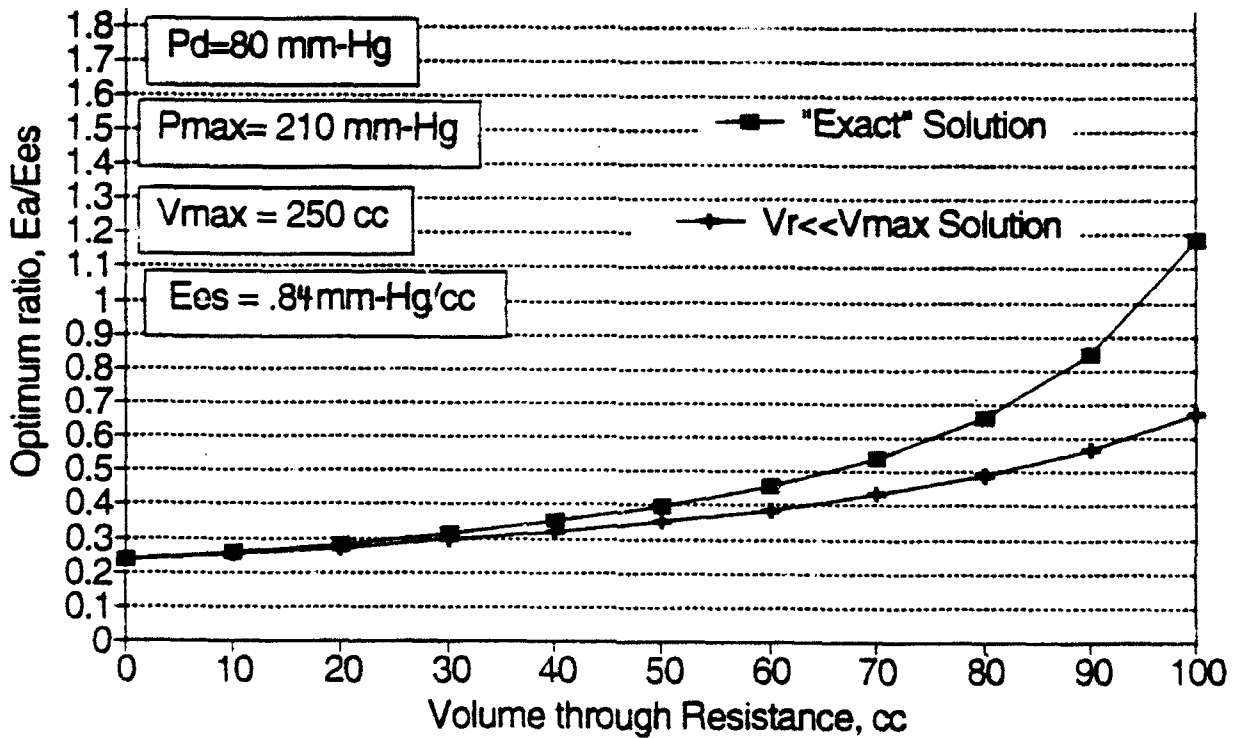


Figure 10: Optimum Ratio with $P_d = 80$ mm Hg

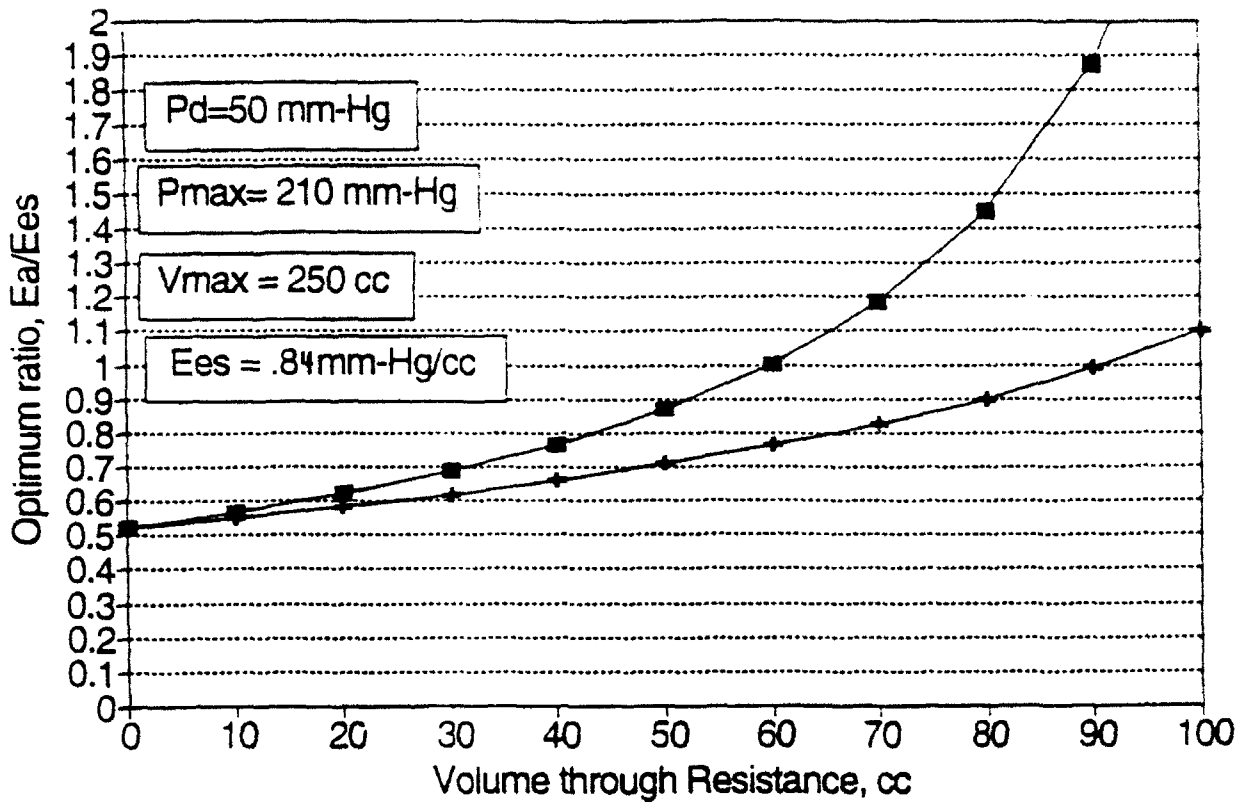


Figure 11: Optimum Ratio with $P_d = 50$ mm Hg

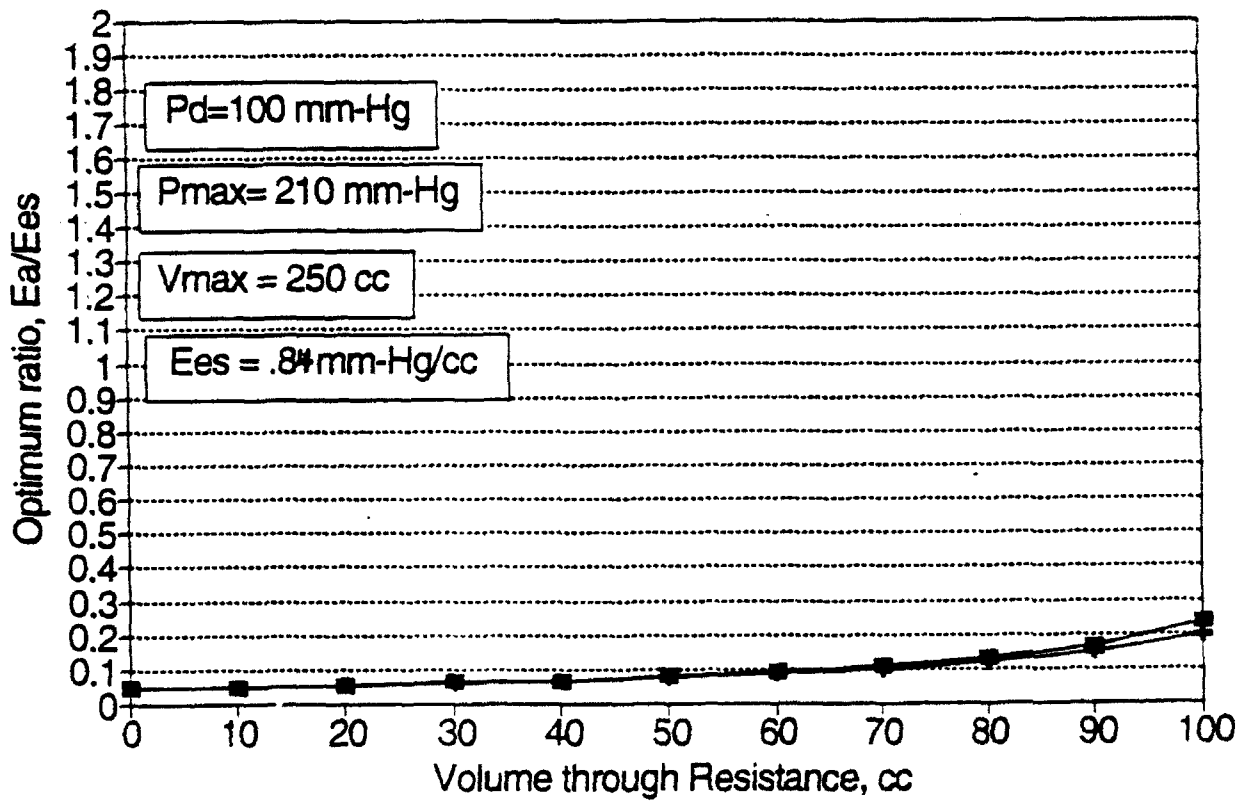


Figure 12: Optimum Ratio with $P_d = 100$ mm - Hg

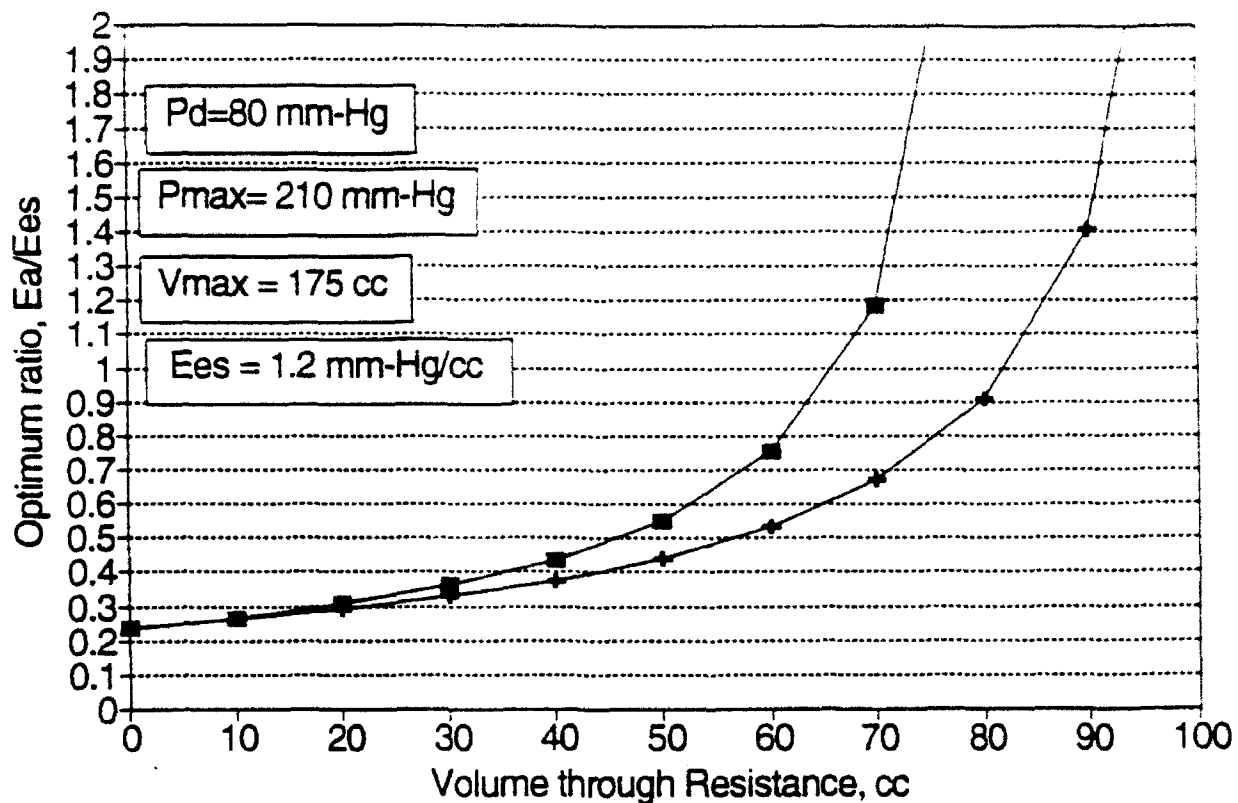


Figure 13: Optimum Ratio with Small End Diastolic Volume

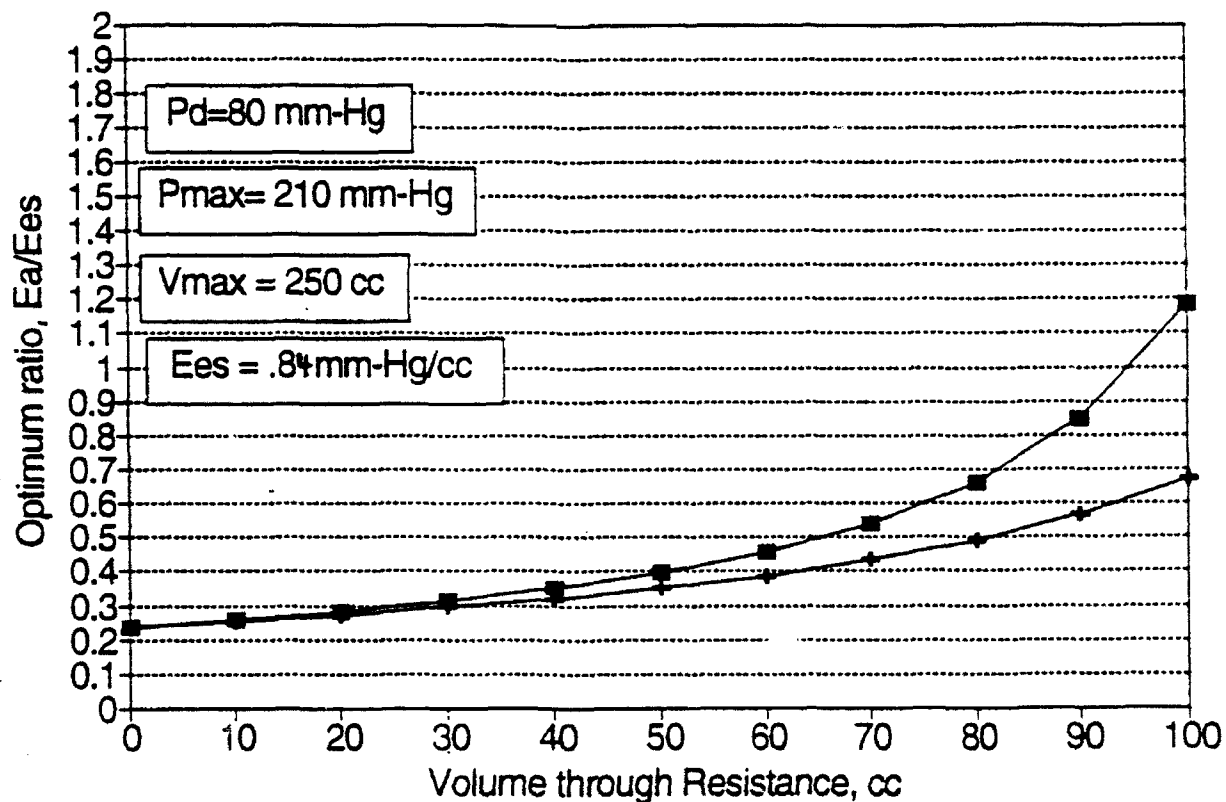


Figure 14: Optimum Ratio with Normal End Diastolic Volume

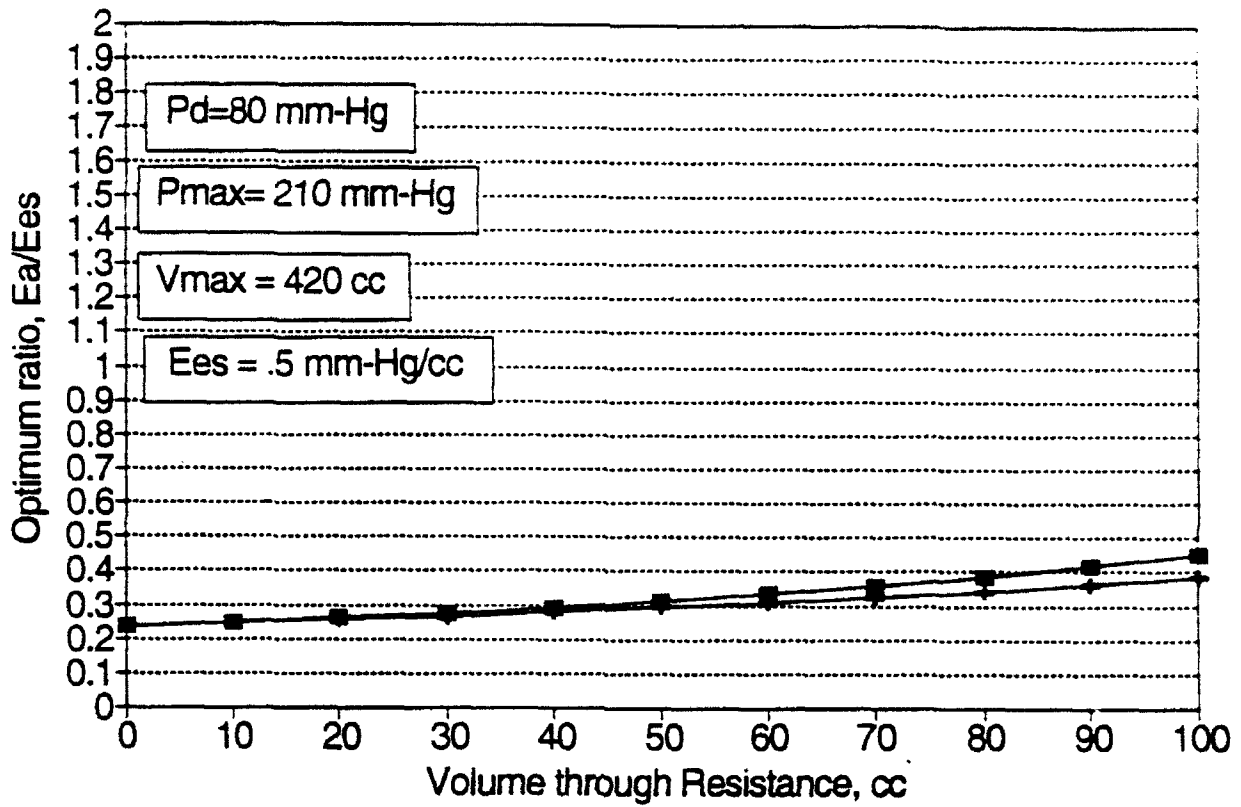


Figure 15: Optimum Ratio with Large End Diastolic Volume

Acknowledgments:

This author would like to thank the team from the Laboratory for Aerospace Cardiovascular Research for their support. Furthermore, the major intellectual contributions of Dr. R.D. Swope, Capt. D.A. Self, and Dr. R. Crisman were greatly appreciated.

REFERENCES

- Belloni, F. L., "The Local Control of Coronary Blood Flow", *Cardiovascular Research*, 13, 83-85.
- Canty J. M. and F. J. Klocke, "Reduced Regional Myocardial Perfusion in the Presence of Pharmacological Vasodilator Reserve", *Circulation*, 71, 370-7.
- DeHart, R. M., "Coronary Heart Disease: An Expensive Air Force Problem", *Aviation, Space and Environmental medicine*, 51(9), 1057-1063.
- Elzinga, G. and N. Westerhof, "Pump Function of the Feline Left Heart: Changes with Heart Rate and its Bearing on the Energy Balance", *Cardiovascular Research*, 14, 81-92, 1980.
- Hickman, J. R. and G. M. McGranahan, Jr., Fundamentals of Aerospace Medicine, Chapter 14, Edited by DeHart, Lea and Febiger, Philadelphia, 1985.
- Hoffman, J. I. E., "Maximal Coronary Flow and the Concept of Coronary Vascular Reserve", *Circulation*, 70(2), 153-9, 1984.
- Kubota, T., Alexander, Jr., J. Itaya, K., Todaka, K., Sugimachi, M., Sunagawa, K., Nose, Y. and Akira Takeshita, "Dynamic Effects of Carotid Sinus Baroreflex on Ventriculoarterial Coupling Studied in Anesthetized Dogs", *Circulation Research*, 70, 1044-1053, 1992.
- Liu, Z., Brin, K. P., and F. C. P. Yin, "Estimation of Total Arterial Compliance: an Improved Method and Evaluation of Current Methods", *American Journal of Physiology*, 251 (Heart, Circ. Physiol. 20), H588-H600, 1986.
- Pettyjohn, F. S., and R. R. Meekin, "Joint Committee on Aviation Pathology: XVI Coronary Artery Disease and Preventive Cardiology in Aviation Medicine", *Aviation, Space, and Environmental Medicine*, 46, 1299-1304, 1975.
- Sunagawa, K., Moughan, W. L., Burkoff, D., and K. Sagawa, "Left Ventricular Interaction with Arterial Load Studied in Isolated Canine Ventricle", *American Journal of Physiology (Heart, Circ. Physiol. 14)*, H733-H780, 1983.
- Toorup, G. P., Westerhof, N., and G. Elzinga, "Beat-to-Beat Estimation of Peripheral Resistance and Arterial Compliance during Pressure Transients", *American Journal of Physiology*, 252, (Heart, Circ. Physiol. 21), H1275-H1283, 1987.
- Whitton, R. C., "Medical Disqualification in USAF Pilots and Navigators", *Aviation, Space, and Environmental Medicine*, 55(4), 332-6, 1984.

TSH AND FREE T₄ TESTING IN LIEU OF TRADITIONAL
"LONG" THYROID PROFILES

Tekum Fonong
Associate Professor
Department of Chemistry

Hampton University
Hampton, Virginia 23668

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

TSH AND FREE T₄ TESTING IN LIEU OF TRADITIONAL "LONG" THYROID PROFILES

Tekum Fonong
Associate Professor
Department of Chemistry
Hampton University

Abstract

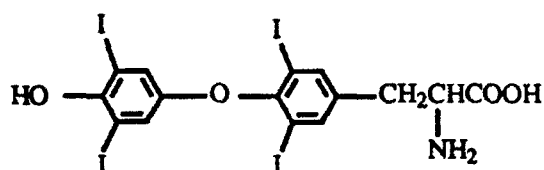
A laboratory test protocol consisting of the measurement of *sensitive* TSH (sTSH) and free thyroxine (FT₄) in human serum was evaluated as a possible replacement for the traditional "long" thyroid profile for assessing thyroid dysfunctional states. The traditional "long" thyroid profile consists of the measurement of serum TSH, total thyroxine (TT₄), triiodothyronine uptake (T₃U), and a calculated free thyroxine index (FTI). The traditional "long" thyroid profile is presently offered to clients by the Epidemiology Research Division of Armstrong Laboratory, Brooks Air Force Base, San Antonio Texas. In this study, sTSH and FT₄ were measured in serum specimens from 89 patients previously evaluated for possible thyroid disorder at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX. The sTSH and FT₄ test results were found to be in support of the clinical diagnoses in the patient groups studied and they were in good agreement with results from the "long" thyroid profile. Two recommendations are made from these findings: (i) that the sTSH and FT₄ combination be provided as the first line test protocol for the laboratory diagnosis of thyroid dysfunction and (ii) that the Epidemiology Research Division of Armstrong Laboratory offer the sTSH and FT₄ test protocol to clients in place of the "long" thyroid profile. There are benefits to the laboratory for implementing a sTSH and FT₄ protocol - the sTSH and FT₄ combination provides increased diagnostic sensitivity compared to conventional RIA methods; the sTSH assay evaluated in this study can clearly distinguish euthyroid from hyperthyroid patients; the chemiluminometric endpoint of the sTSH and FT₄ assays avoids the need for radioisotopes with their potential health and disposal problems; and the reduced number of initial tests using the sTSH and FT₄ protocol is a definite cost-effective strategy for the laboratory.

TSH AND FREE T₄ TESTING IN LIEU OF TRADITIONAL "LONG" THYROID PROFILES

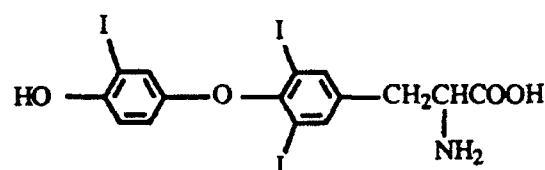
Tekum Fonong

INTRODUCTION

The thyroid gland produces three hormones - thyroxine (T₄), triiodothyronine (T₃), and calcitonine. T₄ and T₃ are synthesized by the gland following entrapment of iodide, conversion to iodine, and coupling of two iodinated tyrosine molecules (1). The T₄ and T₃ so formed are attached to thyroglobulin for storage and are released, as needed, as protease splits them from the thyroglobulin. Daily requirement for iodine is about 200 µg. Iodine is fairly completely absorbed from the upper portion of the gastrointestinal tract, and the excess is excreted in the urine. The structures of T₄ and T₃ are presented below.



Thyroxine (T₄)



Triiodothyronine (T₃)

At moderate concentration levels thyroid hormones cause increased synthesis of RNA and protein, followed by an increased basal metabolic rate (increased caloric production measured by increased oxygen consumption compared with normal). They stimulate oxidative enzyme systems, generally, and enhance release of free fatty acids from adipose tissue. At high concentrations they decrease protein synthesis and uncouple oxidative phosphorylation, energy being dissipated as heat rather than being harnessed for production of adenosine triphosphate (ATP). Thyroid hormones cause increased intestinal absorption of glucose.

CONTROL OF THYROID HORMONE PRODUCTION

Thyrotropin (thyroid stimulating hormone; [TSH]) from the anterior pituitary gland stimulates the production of thyroid hormone by promoting iodide uptake, conversion to iodine, coupling of iodinated tyrosine molecules, and splitting of the hormones from thyroglobulin. The T_3 and T_4 provide feedback to the pituitary gland and, probably, to the hypothalamus. Low levels of thyroid hormones stimulate TSH production and high levels inhibit it. Also, hypothalamic-factor output is stimulated by stress and, in turn, increases TSH output. On the other hand, somatostatin inhibits resting and stimulated secretion, not only of growth hormone, but also of thyrotropin-releasing-factor (TRF) and TSH. Somatostatin inhibits secretion of prolactin as well.

TRANSPORT OF CIRCULATING THYROID HORMONES

Both T_4 and T_3 circulate in plasma bound largely to thyroxine binding globulin (TBG). They are bound to thyroxine binding prealbumin (TBPA) to a lesser extent and, only when in excess, the hormones circulate bound to albumin. The usual concentration ratio of T_4 to T_3 is 9:1, but T_3 has considerably greater physiologic activity. It is the small free fraction of the thyroid hormones (less than 0.1% of the total) that is physiologically active and determines the clinical thyroid status of the patient - hyperthyroid, euthyroid, or hypothyroid.

Primary alterations of TBG are fairly common. An increase is regularly seen in pregnant women, as well as in those following estrogen therapy, or in patients on oral contraceptives. Primary decreases are commonly seen in patients with massive protein losses, such as the nephrotic syndrome. Levels of TBG can also be altered by nonthyroidal illness (NTI). Very rarely, a decrease may be the result of a congenital deficiency. Certain drugs, for example, diphenylhydantoin (Dilantin), diazepam, salicylate, tolbutamide, and chlorpropamide, are able to compete with T_4 binding sites on TBG.

The amount of free hormones (FT_4 and FT_3) present in the plasma depends on the amount of TT_4 and TT_3 and TBG present. Free hormone is increased whenever the ratio of hormone to TBG is increased, whatever the absolute levels. Conversely, it is decreased whenever the ratio of

hormone to TBG is decreased, whatever the absolute level.

In the past, FT₄ determination was tedious and expensive. FT₄ has been estimated by multiplying the results of the TT₄ test by those of the T₃U (2). Nowadays, there are commercially available kits, with equal or better sensitivity than RIA kits, designed for the direct determination of FT₄ in human serum. Some of the kits utilize nonisotopic signals for detection. Because the free hormone is the active form, it is possible that a state of euthyroidism may exist in spite of increased or decreased TT₄ due to abnormal levels of TBG.

HYPERTHYROIDISM

Hyperthyroidism is a syndrome caused by excess production of thyroid hormone, usually T₄, but on occasion, T₃. Hyperthyroidism can have different etiologies, but by far the commonest is Graves' disease, an autoimmune disorder in which serum autoantibodies bind to TSH receptors. Once bound, the antibodies stimulate the production and release of thyroid hormone. These antibodies belong to the IgG fraction generically termed thyroid-stimulating immunoglobulins (TSIg) (2). They cause thyroid hyperactivity with consequent suppression of TSH. With Graves' disease the gland is diffusely enlarged, and the patient may have exophthalmos. Another cause of hyperthyroidism is the presence of functioning thyroid adenomas.

The signs and symptoms of hyperthyroidism include weight loss, fine tremors, tachycardia, cardiac arrhythmias, fatigue, heat intolerance, sweating, and diarrhea. Patients are often nervous and irritable. In women scant menses are common. In hyperthyroid patients (Graves' disease or functioning thyroid adenomas) TSH is low. This observation is also true in patients with hyperthyroidism secondary to an autonomously hyperactive thyroid adenoma. A condition known as thyroid storm may be triggered by stress. It consists of an acute exacerbation of signs and symptoms, including fever, marked tachycardia, heart failure, shock, and mental disturbances (1).

The diagnosis of Graves' disease is based on the clinical presentation along with laboratory evaluation that includes serum TSH, T₄, T₃, T₃U, and calculated FTI. In problem cases, the definitive tests of serum FT₄, thyrotrophin-releasing hormone (TRH) stimulation, and TSIg determinations may be necessary.

HYPOTHYROIDISM

Hypothyroidism (myxedema) is a syndrome caused by insufficient thyroid hormone. Signs and symptoms include lethargy; weight gain; cool, pale, dry skin with puffiness around the eyes and a puffy appearance of the face (secondary to mucinous infiltration of the skin); sparse hair with marked loss of the outer portions of the eyebrows; low body temperature; cold intolerance; fatigue; dizziness; and voice hoarsening. Patients may be anemic and depressed, with loss of libido. When children are afflicted they have growth impairment. Cretinism is the result of thyroid deficiency of the fetus. Primary hypothyroidism may result from thyroiditis, surgical or isotopic ablation of the thyroid, ingestion of thyroid inhibitors, or congenital thyroid enzyme deficiency (1).

OBJECTIVES OF THIS STUDY

- (1) To measure sTSH and FT₄ in human serum using assay kits from Amersham Corporation and to compare the results with those obtained using kits from Bio-Rad Corporation and Tosoh Corporation.
- (2) To evaluate the utility of a sTSH and FT₄ test combination as a possible replacement for the traditional "long" thyroid profile (consisting of TSH, TT₄, T₃U, and FTI) currently offered to clients by the Epidemiology Research Division of Armstrong Laboratory, Brooks Air Force Base, San Antonio Texas.
- (3) To develop an implementation plan for the routine use of sTSH and FT₄ assays at the Epidemiology Research Division of Armstrong Laboratory, Brooks Air Force Base, San Antonio Texas.

MATERIALS AND METHODS

The Amerlite sTSH and FT₄ kits used in this study were purchased from Amersham Corporation, Arlington Heights, IL. The sTSH assay utilizes an immunometric technique and the FT₄ assay utilizes a competitive immunoassay technique. Both assays are based on enhanced luminescence for the measurement of the respective hormones in human serum. The kits were developed for the Amerlite automated Analyzer.

The sTSH assay uses two mouse monoclonal antibodies which bind different epitopes of TSH. One antibody (anti- β subunit of TSH) is coated on wells and the other is conjugated to horseradish-peroxidase. In the FT₄ assay, there is competition between FT₄ from the sample with horseradish-peroxidase-labeled T₄ for a limited number of binding sites on a T₄-specific sheep antibody. The antibody simultaneously reacts with donkey anti-sheep antibody coated onto wells. Following completion of the procedures according to the manufacturer's directions, the light signals are read in the Amerlite Analyzer and related to the concentration of hormones.

RESULTS

Table I shows the TSH and FT₄ results from patient sera analyzed using the Tosoh, Bio-Rad, and Amersham kits. Table IIA shows the data for TSH, TT₄, T₃U, FTI, and FT₄ using the Bio-Rad and Amersham kits for patients with known thyroid disease. Table IIB shows the of TSH, TT₄, T₃U (with calculated FTI), and FT₄ using the Bio-Rad and Amersham kits for patients evaluated at the Endocrinology Department with no known of thyroid disease.

*Table I. Comparison of TSH and FT₄ data from the Tosoh, Bio-Rad, and Amersham kits.

Patient Code	Tosoh TSH	Bio-Rad TSH (μIU/mL)	Amersham TSH	Bio-Rad FT ₄	Amersham FT ₄ (ng/dL)	Tosoh FT ₄
42	0.8	0.7	0.47	1.2	1.36	0.8
44	0.9	0.8	0.74	0.7	1.12	0.8
47	0.9	1.1	0.74	1.1	1.20	0.9
52	1.4	1.3	0.94	1.3	1.60	1.2
66	6.7	6.8	4.93	0.9	0.92	0.8
79	10.2	9.3	7.65	0.7	1.06	0.8

*Results from the Bio-Rad and Tosoh kits were determined by other workers at earlier evaluation of patients at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX.

Table I cont'd. Comparison of TSH and FT₄ data from the Tosoh, Bio-Rad, and Amersham kits.

Patient Code	Tosoh TSH	Bio-Rad TSH (μIU/mL)	Amersham TSH	Bio-Rad FT ₄	Amersham FT ₄ (ng/dL)	Tosoh FT ₄
36	0.7	0.6	0.63	1.4	1.65	1.2
54	1.9	1.6	1.75	1.1	0.97	0.8
45	1.4	0.9	0.34	1.3	1.45	1.1
46	0.9	1.0	0.97	1.3	1.30	1.3
58	2.9	2.4	2.49	0.7	0.92	0.7
53	1.0	0.8	0.57	1.3	1.52	1.1
57	2.1	2.2	1.84	1.1	1.16	0.8
16	0.09	0.06	0.06	0.8	1.04	0.8
41	0.8	0.7	0.72	1.8	1.68	1.4
71	7.8	7.4	4.73	1.1	1.23	1.0
19 ^a	0	0.06	0.05	1.7	1.95	1.6
55	1.6	1.6	1.20	1.4	1.53	1.3
20	0.1	0.07	0.10	1.6	1.63	1.2
7 ^a	0	0	0.05	1.0	1.19	0.9
28	0.2	0.2	0.09	1.8	1.72	1.3
51	1.7	1.3	1.15	1.5	1.82	1.2
15 ^a	0	0.05	0.07	2.3	2.23	1.7
56	1.7	1.7	1.37	1.1	1.10	0.9
74	9.1	8.1	6.57	1.4	1.37	1.1
88	15.2	13.8	10.6	1.3	1.44	1.3

*Results from the Bio-Rad and Tosoh kits were determined by other workers at earlier evaluation of patients at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX.

^aData entries of "0" are actually below the detection limit (DL) of the instrument and should be reported as <DL.

Table I cont'd. Comparison of TSH and FT₄ data from the Tosoh, Bio-Rad, and Amersham kits.

Patient Code	Tosoh TSH	Bio-Rad TSH (μIU/mL)	Amersham TSH	Bio-Rad FT ₄	Amersham FT ₄ (ng/dL)	Tosoh FT ₄
31	0.3	0.3	0.16	1.2	1.46	1.1
81	12.1	9.9	8.73	0.8	0.96	0.8
17 ^a	0	0.06	0.05	1.7	1.82	1.4
5 ^a	0	0	0.02	2.5	2.52	2.0
6 ^a	0.1	0	0.06	1.2	1.42	1.0
10 ^a	0	0	0.07	2.0	1.74	1.5
21	0.2	0.1	0.10	1.8	1.82	1.5
30	0.3	0.3	0.15	1.7	1.89	1.5
33	0.4	0.4	0.35	1.7	1.86	1.6
59	2.7	2.8	2.82	1.7	1.58	1.3
76	9.0	8.3	6.24	1.7	1.61	1.5
73	7.7	7.5	5.41	1.4	1.34	1.1
75	8.7	8.1	5.05	0.9	1.24	1.1
77	8.3	8.3	6.28	0.7	0.82	0.6
83	7.0	6.6	4.58	1.3	1.31	1.0
84	13.3	12.0	9.44	1.1	1.24	1.0
90	14.8	14.2	11.4	0.9	0.97	0.8
67	6.9	7.0	4.58	0.5	0.67	0.5
64	7.9	6.7	5.37	1.2	1.30	1.0
85	13.6	12.2	11.0	1.6	1.49	1.2
80	10.3	9.7	7.83	0.5	0.7	0.6

*Results from the Bio-Rad and Tosoh kits were determined by other workers at earlier evaluation of patients at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX.

^aData entries of "0" are actually below the detection limit (DL) of the instrument and should be reported as: <DL.

Table I cont'd. Comparison of TSH and FT₄ data from the Tosoh, Bio-Rad, and Amersham kits.

Patient Code	Tosoh TSH	Bio-Rad TSH (μIU/mL)	Amersham TSH	Bio-Rad FT ₄	Amersham FT ₄ (ng/dL)	Tosoh FT ₄
63	7.0	6.6	4.58	1.3	1.31	1.0
4 ^a	0	0	0	2.6	2.62	1.9
24	0.2	0.1	0.11	1.9	1.99	1.5
02 ^a	0.2	0	0.04	1.8	1.66	1.5
03 ^a	0	0	0.04	2.7	2.52	2.4
04 ^a	0	0	0	2.0	1.84	1.6
05 ^a	0	0	0	1.8	2.27	1.5
07	0.2	0.2	0.11	2.0	1.89	1.5
09	0.3	0.3	0.23	2.8	2.44	2.0
10	0.2	0.2	0.11	1.7	1.63	1.3
15	0.1	0.06	0.08	1.7	1.78	1.5
17	0.2	0.1	0.08	2.1	2.29	1.7
18 ^a	0	0	0.02	1.8	2.35	1.8
19 ^a	0	0	0.03	1.6	1.67	1.1
21	0.2	0.2	0.14	1.4	1.50	1.1
24 ^a	0	0	0.02	1.4	1.46	1.1
26 ^a	0	0	0	2.0	2.12	1.9
29 ^a	0	0	0	3.9	4.91	4.1
30	0.2	0.2	0.12	2.9	3.41	3.0
32	1.3	1.2	1.05	1.4	1.44	0.9
36	1.2	1.2	1.08	1.4	1.55	1.1

*Results from the Bio-Rad and Tosoh kits were determined by other workers at earlier evaluation of patients at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX.

^aData entries of "0" are actually below the detection limit (DL) of the instrument and should be reported as <DL.

Table I cont'd. Comparison of TSH and FT₄ data from the Tosoh, Bio-Rad, and Amersham kits.

Patient Code	Tosoh TSH	Bio-Rad TSH (μIU/mL)	Amersham TSH	Bio-Rad FT ₄	Amersham FT ₄ (ng/dL)	Tosoh FT ₄
40	0.9	0.7	0.68	1.1	1.23	0.9
42	0.6	0.5	0.25	0.9	1.36	0.8
43	0.6	0.6	0.42	1.3	1.49	1.0
45	1.5	1.4	1.05	1.2	1.45	1.0
46	0.7	0.5	0.59	1.7	1.81	1.5
48	0.8	0.7	0.59	1.8	1.32	1.0
50	2.5	0.7	0.81	1.1	1.14	0.9
51	1.4	1.3	1.20	1.8	1.73	1.6
52	3.2	3.2	2.35	0.7	0.95	0.6
60	6.7	6.7	4.66	1.1	1.27	1.1
67	13.0	12.4	7.89	0.9	1.18	0.9
69	11.6	10.3	7.20	1.1	1.15	0.9
70	8.0	7.1	5.43	1.5	1.59	1.3
71	6.6	7.1	4.93	0.8	0.76	0.6
74	7.0	7.2	4.57	1.4	1.37	1.2
75	7.4	3.9	5.16	0.9	0.92	0.6
77	14.7	14.0	11.5	1.3	1.10	0.9
83	13.3	13.2	10.3	1.1	1.10	0.8
84	7.6	7.4	5.39	1.1	1.06	0.8
85	7.5	6.6	5.84	1.5	1.39	1.0
89	9.6	9.3	6.71	0.7	0.91	0.6

*Results from the Bio-Rad and Tosoh kits were determined by other workers at earlier evaluation of patients at the Department of Endocrinology, Wilford Hall USAF Medical Center, Lackland AFB, San Antonio, TX.

Table IIA. Amersham test kit results for patient group on treatment for known thyroid disease.

Patient Code	TSH (μIU/mL)	TT ₄ (μg/dL)	T ₃ U (%)	FT ₄ (ng/dL)	FTI	Clinical Diagnosis
15	0.07	11.0	34.6	2.23	3.8	<i>Total Thyroidectomy</i>
56	1.37	9.8	25.4	1.10	2.5	"
74	6.57	7.2	31.9	1.37	2.3	"
88	10.6	12.4	26.2	1.44	1.9	"

31	0.16	9.6	27.1	1.46	2.5	<i>Partial Thyroidectomy</i>
81	8.73	8.7	21.7	0.96	2.0	"
17	0.05	11.0	31.2	1.82	3.4	"

5	0.02	13.1	35.2	2.52	4.6	<i>1° Hypo - Over-replaced</i>
6	0.06	7.8	31.9	1.42	2.5	"
10	0.07	10.3	28.5	1.74	2.9	"
21	0.10	9.5	31.4	1.82	3.0	"

30	0.15	14.9	26.1	1.89	3.9	<i>1° Hypo - Normal</i>
33	0.35	11.7	28.0	1.86	3.3	"
59	2.82	11.5	28.4	1.58	3.3	"
76	6.24	15.1	29.0	1.61	4.4	"

73	5.41	11.2	25.9	1.34	2.9	<i>1° Hypo - Under-replaced</i>
75	5.05	8.0	28.2	1.24	2.2	"
77	6.28	6.5	26.3	0.82	1.7	"

Table IIA cont'd. Amersham test kit results for patient group on treatment for known thyroid disease.

Patient Code	TSH (μIU/mL)	TT ₄ (μg/dL)	T ₃ U (%)	FT ₄ (ng/dL)	FTI	Clinical Diagnosis
83	4.58	7.1	30.3	1.31	1.9	<i>1^o Hypo - Under-replaced</i>
84	9.44	10.2	25.2	1.24	2.6	"
90	11.4	6.5	28.5	0.97	1.3	"

67	4.58	4.9	28.2	0.67	1.4	<i>1^o Hypo - Dx < 12 mo.</i>
64	5.37	7.9	31.6	1.30	2.5	"
85	11.0	9.0	29.7	1.49	2.7	"
80	7.83	4.6	28.2	0.70	1.9	"
63	4.58	7.1	30.3	1.31	2.2	"

4 ^a	0	11.1	41.4	2.62	4.6	<i>Goiter - T₄ Suppression Tx</i>
24	0.11	7.5	37.7	1.99	2.8	"
19	0.05	10.5	36.8	1.95	3.9	<i>Thyroid Nodule, No Rx</i>
55	1.20	9.5	29.3	1.53	2.8	"
20	0.10	7.9	34.6	1.63	2.7	<i>"Hot" Nodule</i>

7	0.05	11.5	23.5	1.19	2.7	<i>Graves' Disease - <12 mo Dx</i>
28	0.09	10.0	30.4	1.72	3.0	"
51	1.15	9.5	32.1	1.82	3.0	<i>Hyperthyroidism</i>

^aData entries of "0" are actually below the detection limit (DL) of the instrument and should be reported as <DL.

Table IIB. Amersham test kit results for patient group with no known thyroid disease.

Patient Code	TSH (μIU/mL)	TT ₄ (μg/dL)	T ₃ U (%)	FT ₄ (ng/dL)	FTI	Clinical Diagnosis
42	0.47	5.2	34.7	1.36	1.8	<i>Hyperlipidemia</i>
44	0.74	4.3	34.9	1.12	1.5	"
47	0.74	5.8	32.5	1.20	1.9	"
52	0.94	7.1	34.1	1.60	2.4	"
66	4.93	7.1	26.1	0.92	1.9	"
79	7.65	5.6	32.4	1.06	1.8	"

36	0.63	8.3	33.4	1.65	2.8	<i>CNS</i>

54	1.75	6.2	30.1	0.97	1.9	<i>Pituitary Disease</i>

45	0.34	9.5	28.1	1.45	2.7	<i>Psyche</i>
46	0.97	13.5	26.4	1.30	3.6	"

58	2.49	3.5	35.0	0.92	1.2	<i>Myopathy</i>

53	0.57	11.2	25.4	1.52	2.8	<i>Cardiac Symptoms</i>
57	1.84	6.2	30.3	1.16	1.9	"

16	0.06	5.1	40.4	1.04	2.1	<i>Euthyroid Sick</i>
41	0.72	9.8	31.7	1.68	3.1	"

71	4.73	7.6	28.8	1.23	1.4	<i>Dental Abscess</i>

DISCUSSION OF RESULTS

The three test kits evaluated in this study are based on different detection systems. The Bio-Rad procedure is a conventional RIA assay and the Tosoh procedure is an immunoenzymometric assay (IEMA). The Amerlite immunoassays from Amersham utilize enzyme-based, enhanced luminescence technology in the qualitative and quantitative determination of antigens found in blood serum, plasma and other biological fluids. The enzyme horseradish peroxidase is used as the label in the immunoassay reaction. The enzyme acts as a catalyst in the oxidation reaction of a luminescent component, whereby luminescence is produced. The light would normally be of low intensity and short lived, if it was not for the enhancer which amplifies and sustains the light. The light is measured at any time during a period of between two and twenty minutes after the signal reagent is added. By virtue of their high specific activity and the possibility of precise rapid quantitation of photon emissions, chemiluminescent labels probably offer the greatest sensitivity for immunoassays (3).

Linear Least squares regression plots of the Bio-Rad (x-axis) vs. Amersham (y-axis) data for FT_4 gave a straight line of equation $Y = 0.991x + 0.121$ with a correlation coefficient (r) of 0.925. A plot of FT_4 data for Bio-Rad vs. Tosoh gave excellent linearity with a correlation coefficient approaching unity. Similar results were obtained from plots of TSH data. Such results indicate that the Bio-Rad thyroid test procedures based on RIA can be replaced with either the Tosoh or Amersham nonisotopic procedures. However, the laboratory must determine new reference ranges for each analyte being measured by the substitute assay. The reference ranges for sTSH, TT_4 , T_3U , FTI, and FT_4 (Amersham kits) at the Epidemiology Research Division of Armstrong Laboratory are (0.15 - 3.20) $\mu IU/mL$, (5.1 - 12.4), $\mu g/dL$, (23.4 - 38.5) %, (1.6 - 3.9), and (0.7 - 1.61) ng/dL , respectively. The sensitivity of the sTSH assay is 0.04 $\mu IU/mL$.

Objectives of Thyroid Testing

- (1) to define the level of thyroid function - hyperthyroid, euthyroid, or hypothyroid;
- (2) to identify the nature and location of the disorder, and
- (3) to monitor the effectiveness of treatment.

The diagnosis of thyroid dysfunction is based on clinical examination supported by the appropriate laboratory tests. The clinical expression of thyroid disorders is so varied that the laboratory plays a central role in their assessment and an understanding of laboratory tests is essential to effective treatment.

Of the 89 patient sera analyzed in this study, only 50 patient charts were readily available for correlation of the laboratory studies with the clinical diagnoses. Of the 50 patients, thirty-four patients (68%) had earlier been diagnosed with thyroid disease and the majority of these were on treatment for their condition at the time of this study (Table IIA). Sixteen patients (32%) had no known history of thyroid disease although they had been diagnosed with a variety of nonthyroidal illnesses (Table IIB).

The information from the charts of the 34 patients was as follows: three patients were not on any form of treatment; three patients were being evaluated less than twelve months from initiation of treatment for hyperthyroidism; four patients had total thyroidectomy; three patients had partial thyroidectomy; fourteen patients were on replacement therapy for primary hypothyroidism; five patients were on treatment for primary hypothyroidism diagnosed in the last 12 months before the study; and two patients were on suppression doses of T_4 for goiter.

Careful evaluation of the sTSH and FT_4 laboratory data presented in Table IIA shows the data to be consistent with the clinical diagnoses of the patient groups studied and the results are also in good agreement with the "long" thyroid profile results using FTI as an estimate of FT_4 . Although the measurement of sTSH and FT_4 in this study was carried out principally for the purpose of monitoring the effectiveness of treatment of previously diagnosed thyroid disorders, sTSH and FT_4 measurements should be the tests of choice for the initial assessment of thyroid function (see Figure 1, page 14-19).

Measuring TSH in serum provides useful information about both thyroid and pituitary function. Immunoassays for TSH have become extraordinarily specific for TSH through the application of monoclonal antibodies. Current trends in TSH assay development have emphasized the lower end of the normal range in order to distinguish small amounts of TSH ($<0.1 \mu\text{U/mL}$ from true 0 levels) for detection of the depression of TSH, which occurs in hyperthyroidism and secondary hypothyroidism (4, 5). The TSH assay has its greatest value in

distinguishing primary thyroid deficiency from hypothyroidism secondary to pituitary dysfunction. Levels of TSH also help distinguish true hyperthyroidism from functionally euthyroid conditions with low thyroid hormone levels.

TSH values are subnormal in all patients with hyperthyroidism except for the rare cases of a pituitary TSH-secreting adenoma (4). The present availability of sTSH assays using monoclonal antibodies has greatly increased the analytic sensitivity of TSH assays. The increased specificity makes it possible to distinguish subnormal TSH levels due to hyperthyroidism. Such TSH tests are now routinely used for confirming the diagnosis of hyperthyroidism (TSH <0.1 μ IU/mL).

Serum TSH measurements are also useful in determining the response to exogenous TRH in the investigation of hypo- and hyperthyroidism in differentiating between pituitary and hypothalamic causes of hypothyroidism. Measurement of TSH is also useful in monitoring thyroxine replacement therapy in hypothyroidism and in following patients who have undergone thyroid surgery or developed hypothyroidism following radiation (6). When treatment is appropriate, generally there is a return to normal.

CONCLUSIONS

The sTSH and FT₄ test results using kits from Amersham Corporation were found to be in support of the clinical diagnoses in the patient groups studied. The results were also in good agreement with results from the "long" thyroid profile (consisting of TSH, TT₄, T₃U, and FTI). This means (i) that the sTSH and FT₄ can be provided as the first line test protocol for the laboratory diagnosis of thyroid dysfunction and (ii) that the Epidemiology Research Division of Armstrong Laboratory can offer the sTSH and FT₄ test protocol to clients in place of the "long" thyroid profile. There are benefits to the laboratory for implementing the sTSH and FT₄ protocol: the sTSH and FT₄ combination provides increased diagnostic sensitivity than conventional RIA methods; the sTSH assay evaluated in this study can clearly distinguish euthyroid from hyperthyroid patients; the chemiluminometric endpoint of the sTSH and FT₄ assays avoids the need for radioisotopes with their potential health and disposal problems; and the reduced number of initial tests using the sTSH and FT₄ protocol is a definite cost-effective strategy for the laboratory.

RECOMMENDATIONS FOR IMPLEMENTATION OF sTSH AND FT₄ TESTING AT THE
EPIDEMIOLOGY RESEARCH DIVISION OF ARMSTRONG LABORATORY,
BROOKS AIR FORCE BASE, SAN ANTONIO TEXAS.

According to the American Thyroid Association Guidelines (7, 8) a *sensitive* TSH assay is one which has sufficient sensitivity to discriminate between abnormally low values and the low range of normal. This criterion is clearly met by the Amerlite TSH assay from Amersham Corporation which was evaluated in the study reported here. Accordingly, the following are suggestions for implementing a sTSH and FT₄ test combination in the clinical laboratory.

- (1) The reference range used in the laboratory for thyroid testing must be established locally since it is influenced by diet and analytic methodology.
- (2) The sTSH assay should be the primary test for evaluating thyroid function as well as for monitoring therapy in thyroid disease.
- (3) When a measured sTSH is low the laboratory should automatically determine FT₄.
- (4) For patients on thyroxine replacement therapy, to identify patients who are overtreated, FT₄ should automatically be measured when sTSH is low.
- (5) When a measured FT₄ is within the borderline low range value with
 - (a) a slightly elevated TSH, this may indicate early hypothyroidism. In this case, the laboratory should recommend measurement of circulating thyroid antibodies to assess the presence of autoimmune thyroiditis.
 - (b) a normal TSH value, this suggests either that there is competition by a drug for binding sites on TBG, or that hypothyroidism is secondary to hypothalamic disease (9). In this case, the laboratory should reassess the drug history of the patient. If secondary hypothyroidism is suspected (communication with the patient's primary care physician), the laboratory should suggest to the clinician to proceed to TRH testing. A slightly low plasma T₄ with a normal TSH level may also be due to severe nonthyroidal disease.

- (6) Laboratory test slips should request information about patient medication history.
- (7) The laboratory is an important partner in the delivery of efficient and cost-effective care to the patient. As such, the laboratory *must* make an effort to keep in regular communication with the patients' primary care physician (especially in problem cases) in an effort to achieve this goal. This will also insure that laboratory personnel are kept abreast of the latest developments in the delivery of quality patient care.

A strategy for the investigation of thyroid function in ambulatory patients is presented in Figure 1 below.

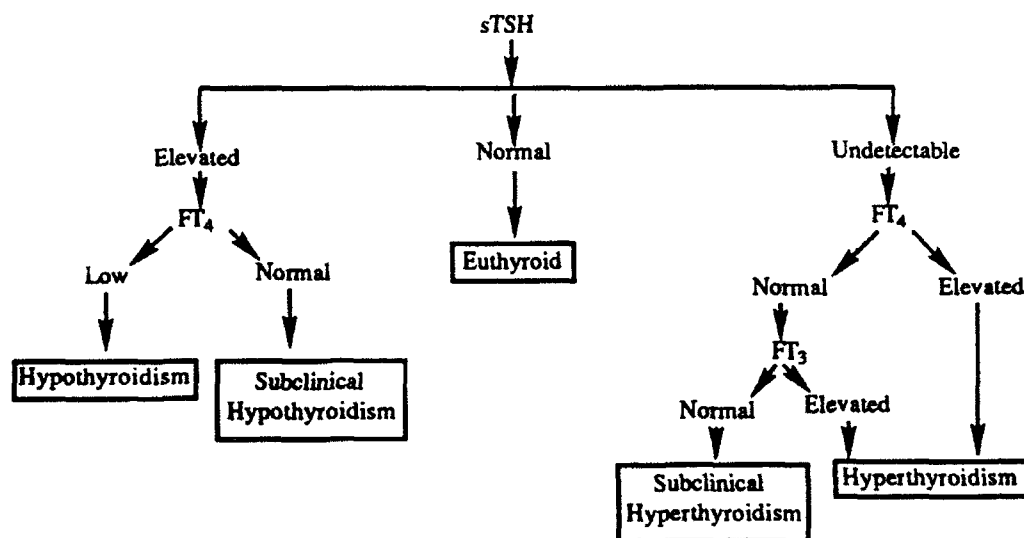


Figure 1. Strategy for Investigation of Thyroid Function

In evaluating thyroid status in relation to serum thyroid hormone levels, it should be noted from Figure 1 that in the case of T_3 toxicosis, a form of hyperthyroidism, serum FT_4 alone is not adequate to indicate the patient's status. In general, FT_4 is within normal limits and FT_3 is elevated (10). Also, although the biologically active hormone is T_3 (3 to 4 times more active than T_4), serum FT_4 is a better index of thyroid function than FT_3 in relation to the need for follow-up treatment.

REFERENCES

1. Beeler MF. Interpretations in Clinical Chemistry, 1st ed. Chicago: American Society of Clinical Pathologists, 1981.
2. Gornall AG. Applied Biochemistry of Clinical Disorders, 2nd ed. Philadelphia: Lippincott Co., 1986.
3. Weeks I, Sturgess ML, Woodhead JS. Chemiluminescence immunoassay: an overview. Clin Sci 1986;70:403-08.
4. Bayer MF. Effective Laboratory Evaluation of Thyroid Status. Medical Clinics of North America; 1991;75:1-27.
5. Nicoloff JT, Spencer CA. The Use and Misuse of the Sensitive Thyrotropin Assays. J Clin Endocrinol Metab 1990;71:553-58.
6. Watts NB. Use of a Sensitive Thyroxine Assay for Monitoring Treatment With Levothyroxine. Arch Intern Med 1989;149:309-12.
7. Surks MI, Chopra IJ, Mariash CN, Nicoloff JT, Solomon DH. American Thyroid Association Guidelines for Use of Laboratory Tests in Thyroid Disorders. J Am Med Assoc 1990;63:1529-32.
8. Hay ID. et al. American Thyroid Association Assessment of Current Free Thyroid Hormone and Thyrotropin Measurements and Guidelines for Future Clinical Assays. Clin Chem 1991;37:2002-08.
9. Zilva ZF, Pannell PR, Mayne PD. Clinical chemistry in Diagnosis and Treatment, 5th ed. Chicago: Year Book Medical Publishers Inc., 1988.
10. Caldwell G. et al. A New Strategy for Thyroid Function Testing. Lancet 1985;May 18:1117-19.

**A THEORETICAL AND COMPUTATIONAL MODEL OF
LASER INDUCED RETINAL DAMAGE**

Bernard S. Gerstman
Associate Professor
Department of Physics

Florida International University
University Park
Miami, FL 33199

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, San Antonio, TX

September 1992

A THEORETICAL AND COMPUTATIONAL MODEL OF
LASER INDUCED RETINAL DAMAGE

Bernard S. Gerstman
Associate Professor
Department of Physics
Florida International University

Abstract

Laser induced retinal damage was studied from a theoretical standpoint. A computational model was developed to simulate and understand the underlying physical mechanisms leading to damage. The model permits the varying of the density of the melanin granules which are responsible for the high absorptivity of visible radiation in the retinal pigment epithelium, the location of the primary absorption site leading to damage. The model also permits the varying of the input laser pulse characteristics, thus allowing a systematic computational investigation of the effects of different laser characteristics and the subjects' retinal anatomical differences on damage mechanisms. A retinal thermal profile as a function of position and time is computed for a given laser pulse and melanin density. The laser induced temperature rise can then be used in a preliminary thermodynamic model of the damage process that was also developed. This model takes into account both damage and repair mechanisms and supplements the single term Arrhenius damage integral, which is shown to be inadequate. In, addition the "nanosecond discrepancy" is shown not to exist.

A THEORETICAL AND COMPUTATIONAL MODEL OF LASER INDUCED RETINAL DAMAGE

Bernard S. Gerstman

I. INTRODUCTION

An important ramification of the continual progress in laser development is the increased danger of injury, especially to the visual system. This danger of injury is present in military and civilian use, and can be either accidental or intentional. This subject continues to gain in importance as a result of the steady advancement in laser technology (increased energy and irradiance, shortening of pulse lengths into the femto-second regime, etc.). Though damage can be caused by electromagnetic radiation from non laser sources, laser sources are the most dangerous due to their extreme values of the above mentioned beam properties, and therefore lasers were considered as the radiation sources for this research. The following report is equally valid for non laser sources, i.e. the specific characteristic of coherence of laser radiation is not the property that causes damage. Therefore, when reference is made to the "beam", the source can be either laser or non laser.

Experimental tests of laser induced retinal damage have been carried out under a variety of controlled and uncontrolled circumstances. We have developed a computational model of the primary steps of laser induced damage in the retina, as well as a preliminary theoretical thermodynamic model of the process. Together, these will allow increased understanding of the underlying damage mechanisms and analysis of how the damage depends on practical considerations of properties such as those of the retina and the laser beam. This information can then be used to make protective eyewear more effective.

II. OVERVIEW OF DAMAGE PROCESS

The sequence of steps leading to visual system damage is the following:

- 1) beam propagation through the visual system
- 2) Absorption of radiation by elements of the visual system
- 3) Initial biological damage mechanisms resulting from absorption
- 4) Functional damage

Additionally, the location of the initial absorption and damage can be separated into two general classes:

- 1) The parts of the eye normally involved only in optical transmission (cornea, lens, humours)
- 2) The retina and its associated layers

A full picture of radiation induced damage to the visual system requires an understanding of all aspects of all the categories listed above and their coupling. The subject of this report is the absorption and initial damage produced in the retina. The value of this work can only be realized when coupled to the

other aspects. For example, the significance of damage to cells of the retina can only be decided based upon the loss of visual function as expressed in a decrease of visual task performance. (Functional loss was not investigated during the period covered by this research and will not be discussed in this report. A basic question which should be investigated in the future, relevant to functional loss due to beam absorption is, how many photoreceptor cells need to be inactivated in the fovea before functional loss occurs?)

In order for radiation to cause damage in the retinal region, the beam must enter the eye and be transmitted to the retina. While being transmitted through the ocular media, the beam properties can be changed. However, for the present study, the only input information necessary in our modeling of retinal damage are the beam characteristics when it hits the retina:

- 1) Total energy deposited
- 2) Beam size and profile
- 3) Pulse length
- 4) Wavelength shifts

The ocular media (cornea, lens, humours) therefore, are relevant to this study only in that they may affect the beam properties compared to those incident at the cornea. For example, energy reaching the retina will be less than incident energy due to absorption; the beam profile is affected by geometric focusing by the cornea and lens, scattering, or self-focusing at high intensities; pulse length can be modified by plasma generation; wavelength shifts can result from Raman or Brillouin processes. Since these beam characteristics serve as input data to modeling retinal damage, the most straightforward way to get this information would be by subjecting eyes, or synthetic systems designed to simulate the optical media of the eye, to beam pulses of all ranges of the above properties and measuring the properties of the pulses after traveling through the optical media. This can be accomplished by putting a beam detector at the back end of the optical media, where the retina would be in an intact eye. Clearly this is an over-simplification of the measurements involved, and the experimentalists at the Laser Laboratory at Brooks Air Force Base should be consulted concerning the aspects of these measurements. I have of course, ignored the important phenomena that can occur due to interaction and absorption of radiation by parts of the eye that are designed only for transmission. These phenomena, such as cataracts, can lead to severe functional loss and, for high power levels, new phenomena can occur such as plasma generation, the functional damage effects of which still require investigation.

Having briefly explained how retinal damage is coupled to other aspects of the visual process, we now concentrate on retinal damage induced by absorption of electromagnetic radiation. The actual photon absorption causes electronic excitation of bio-molecules, but the resulting biochemical damage can be attributed to different underlying physical causes following the photon absorption:

Thermal: Electron-phonon coupling converts the electronic excitation energy into heat. On a molecular level, the heat is excitation of vibrational normal modes (temperature increase) which can result in protein denaturation.

Chemical: The electronic excitation leads to changes in chemical bonds producing noxious chemical reactions.

Photo-toxic: Electronic excitations lead to electronic bond breaking which makes the molecule dysfunctional (denaturation by electronic means rather than thermal).

Acousto-Mechanical waves: Electronic excitation energy is converted to mechanical energy which results in stresses that disrupt structures such as membranes.

Electrical: Extremely high intensities may have electric fields strong enough to electrically disrupt membranes.

From the standpoint of functional impairment to the visual system affecting task performance, damage to the photoreceptors and their associated nerve transmission pathways is the critical consideration. In order to determine which of these damage mechanisms are relevant, we must look at the characteristics of the beam that hits the retina, as well as the molecular absorption sites in the retinal region.

In investigating the retina, only radiation of $\lambda > 400$ nm need be considered since radiation of shorter wavelength is almost completely absorbed by the cornea and lens (removal of the lens can lead to retinal damage due to increased transmission of UVA radiation). Thus, initial damage will occur at sites capable of absorption of visible and infra-red radiation. Since we are concerned with damage to the photoreceptors, and since the photoreceptors absorb visible light, damage to photoreceptors due to their own absorption is an obvious possibility. However, the photoreceptors only absorb ~5% of the incident light, and the visual system is specifically designed to channel this energy away through chemical reactions, preventing heat buildup. It is possible that damage can result from absorption by the photoreceptors through photochemical and phototoxic mechanisms (e.g. permanent bleaching), but this requires long pulses ($\tau > 1$ sec) along with short wavelength and high energies (5 J/cm² at 325 nm, 20 J/cm² at 442 nm, 3000 J/cm² at 632 nm)[1]. These energies are all well above thresholds for damage, and therefore other absorption sites besides the photoreceptors themselves must be responsible for threshold damage.

The absorption site that is most important for retinal damage is the monolayer of cells called the retinal pigment epithelium (RPE). As shown in Fig. 1, this layer of cells is in direct contact with the photoreceptors, allowing strong coupling of any phenomena resulting from initial photon absorption in the RPE. The

importance of the RPE as the initial damage site is due to its high absorption (~50% of incident visible radiation) by the melanin located in the RPE cells. Evidence that near threshold damage by visible light is centered in the RPE comes from experimental observations by Gueneau, et. al. [2]. They made observations using a beam with $\lambda=593$ nm (Rhodamine 6G), $\tau=600$ nsec, and beam diameter $d=570$ μ m. At 24 hours post exposure, their observations on the test animal were the following

Radiance on Retina

Observed Damage

12 mJ/cm²

RPE cells damaged, slight damage to photoreceptor outer segments.

16 mJ/cm²

RPE badly damaged, O.S. disrupted.

45 mJ/cm²

RPE destroyed, photoreceptors partially destroyed.

118 mJ/cm²

All retinal layers damaged. RPE, photoreceptor outer and inner segments destroyed but Bruch's membrane not broken.

These observations that the near threshold damage is centered in the RPE implies that the initial absorption site for threshold damage is the melanin that is the dominant absorber in RPE cells. The possible damage mechanisms resulting from photon absorption by the melanin granules are the following:

- 1) Direct damage to melanin granules. This is unlikely because melanin granules are extremely robust, e.g. they show no anatomical changes after being heated to 350 C or boiled in concentrated acid (sulfuric acid, hydrochloric acid, alcoholic KOH) [3].
- 2) Thermal damage due to heat conducted away from melanin granules, leading to protein denaturation in the RPE and photoreceptors (denaturation can start at temperature increases of only 10 C)
- 3) Acoustic transients leading to disruption of cellular components. Membranes would be most susceptible due to stress and shear along their extended length.

COMPUTATIONAL MODEL FOR DAMAGE

The dominant damage mechanisms are likely to be from thermal effects and acoustic transients (2 and 3 above). We have developed a computational model to investigate these effects. The important physical quantities to be calculated are the temperature rise and pressure gradients resulting from the photon absorption in the RPE. This is done by having photon absorption occurring only by melanin granules in the RPE cells. The RPE cellular volume is divided into grid points, and heat is transferred by conduction between grid points at each time increment.

BIOLOGICAL STRUCTURE OF ABSORPTION REGION

The initial absorption occurs in melanin granules that are dispersed in the cells of the close packed RPE monolayer. The RPE cells are in direct contact with the photoreceptors and have extrusions that surround part of the photoreceptor outer segments. The dimensions of the relevant components can be seen in Fig. 1. The melanin granules are ellipsoids with a semi-major axis of

approximately $1.5 \mu\text{m}$ and an axis ratio of 1.5:1 (close to spherical). The melanin granules are not randomly distributed, but instead are more densely packed in the region of the RPE cell closer to the photoreceptor (explaining why damage to photoreceptors occurs at lower beam radiance than damage to the Bruchs membrane which is on the opposite side of the RPE cell).

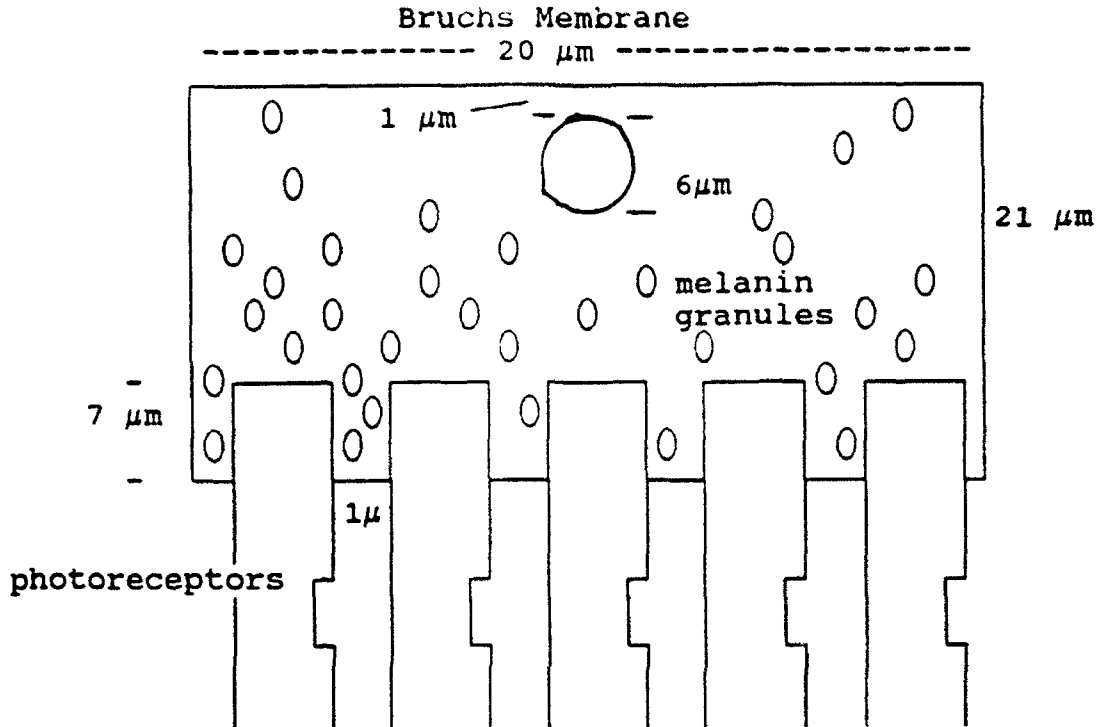


Figure 1. Schematic diagram of a single RPE cell. A sample distribution of melanin granules is shown with increased density near photoreceptor outer segments. Nucleus is treated as having same water-like thermal properties of rest of cellular fluid, but is explicitly included because it is a region of cell without melanin.

MELANIN DISTRIBUTION

In our computer calculations for increases in temperature and pressure, biological variability in melanin granule distribution is taken into account by making the number of melanin granules per cell an input variable in the program. This allows us to systematically determine the effect of differing melanin density among individuals or species, on damage mechanisms. Another variable takes into account the observation that the density of melanin granules is low in the part of the cell closer to the Bruchs membrane. This is incorporated in the program with a variable that allows the choice of what percentage of melanin granules are within $7 \mu\text{m}$ of the Bruchs membrane boundary.

SIZE OF CELLULAR VOLUME MODELLED

The number of RPE cells, and hence the amount of cellular material used in the program is chosen so that the final, uniform

equilibrium temperature increase ($t \rightarrow \infty$) resulting from the laser energy deposited will be small enough to be assured of causing no ongoing thermal damage. Enough cells must be used so that when the total laser energy absorbed by the RPE is uniformly equilibrated, the ΔT is less than the 10 C at which proteins begin to denature. The greater the laser energy absorbed, the greater the volume of cellular material required. However, the more cellular material incorporated into the program, the more grid points needed, and the longer the computation will take. A starting volume can be determined for the amount of cellular material to be used by selecting the appropriate volume (with water-like thermal properties) such that the uniform equilibrium ΔT will be 2 C. If the program runs too slowly for each input laser pulse modelled, then the equilibrium ΔT can be raised to 4 C, which will allow a decrease in the total amount of cellular material by half, thus shortening the computation time.

III. CALCULATION OF ACOUSTIC PRESSURE TRANSIENTS

Gradients in pressure cause mechanical stress or strain which can tear bonds in biological structures. The weakest bonds in biomolecules are hydrogen bonds, ~ 6 kcal/mole. Using a Morse potential analysis, this equates to a breaking force of 0.33×10^{-4} dyne/bond. We will now describe our approach to modelling laser induced pressure gradients and show that the resulting forces are on the order to break enough bonds to cause damage in RPE cells.

The simple expression relating forces to pressure gradients is $F = \nabla P dV$. In order to calculate ∇P we need the spatial and time dependence of $P(r, t)$. The standard thermo-mechanical expression is

$$P(r, t) = B\beta\theta(r, t) - B\nabla \cdot \xi(r, t) \quad (1)$$

$\theta(r, t)$: temperature rise, ΔT

$\xi(r, t)$: linear displacement of volume element from equilibrium r

B : Bulk modulus

β : thermal coefficient of expansion

Thus, the calculation of $P(r, t)$ only requires determination of $\theta(r, t)$ and $\xi(r, t)$. Equation (1) can be further simplified with the assumption that $\theta(\text{cell}) = 0$ during the short times of acoustic wave propagation. This is justified because the speed of sound is orders of magnitude higher than the speed of thermal conduction in cellular material

$$\frac{v(\text{thermal})}{u(\text{sound})} = \frac{\frac{\kappa}{\rho c L}}{\sqrt{B/\rho}} = \frac{.01 \text{ m/s}}{1500 \text{ m/s}} < 10^{-5} \quad (2)$$

where we have used the values of water: $\kappa = .419 \text{ J/m} \cdot \text{s} \cdot ^\circ\text{C}$, $c = 4.19 \times 10^3 \text{ J/Kg} \cdot ^\circ\text{C}$, $\rho = 10^3 \text{ Kg/m}^3$, $B = 2.1 \times 10^9 \text{ Kg/m} \cdot \text{s}^2$ and a distance $L \sim 10 \mu\text{m}$ (size of one cell). With the assumption of $\theta = 0$, the calculation of P reduces to

$$P = -B\nabla \cdot \xi \quad (3)$$

An equation for ξ is derivable from an equation expressing the flow of momentum density $p = \rho v$ for any volume element located at r at a time t ,

$$p(r+vdt, t+dt) - p(r, t) = Fdt - \nabla P dt + f(\eta)dt \quad (4)$$

with

$$v = d\xi/dt$$

F : external force/volume

∇P : pressure gradient (force/volume)

$f(\xi)$: viscous forces/volume

We make the assumptions that $v \cdot \partial p / \partial r$ and $f(\eta)$ are both small compared to external forces F and ∇P , and that the cellular fluid is incompressible ($dp/dt = 0$). Equation (4) then simplifies to

$$\rho \partial v / \partial t = \rho \partial^2 \xi / \partial t^2 = F - \nabla P \quad (5)$$

Since $u(\text{sound}) \gg v(\text{thermal})$, the motion is occurring under adiabatic conditions in which $P = -B\nabla \cdot \xi$, which allows Eq. (5) to be written as

$$B\nabla(\nabla \cdot \xi) - \rho \partial^2 \xi / \partial t^2 = -F \quad (6)$$

To solve this equation for $\xi(r, t)$ (and thus to get the pressure via Eq. (3)) requires an explicit expression for the external force F . This force is generated by the laser energy absorbed by the melanin granules. This absorption of energy causes sudden heating of the cellular fluid localized around the melanin granule leading to a force due to the gradient of the thermal expansion

$$F = -B\nabla \theta \quad (7)$$

Equation (6) now gives us a wave equation for acoustic pressure transients

$$\nabla(\nabla \cdot \xi) - \rho/B \partial^2 \xi / \partial t^2 = \nabla \theta \quad (8)$$

with the velocity of acoustic wave propagation given by $u = \sqrt{B/\rho}$ as used in Eq. (2).

Equation (8) is an inhomogeneous, three-dimensional, thermally uncoupled equation relating the electromagnetic energy induced θ to thermal elastic waves. It is thermally uncoupled because thermal effects, θ , only appear in the driving force, not in the dynamic response of the system. Solution of this equation only requires boundary conditions for ξ and the initial localized $\theta(r, t)$.

Equations (7) and (8) allow an analysis showing that shorter laser pulse duration τ , and hence shorter energy deposition time in melanin granules, increases the strength of the resulting pressure waves. The magnitude of the response depends on the strength of the driving force. This driving force is proportional to the

gradient of the temperature increase, $\nabla\theta$, not merely θ itself. The temperature increase of the cellular fluid immediately surrounding a melanin granule is proportional to the total energy absorbed by the granule. If the energy is deposited over pulse times long compared to thermal conduction times of the surrounding fluid, then significant amounts of the laser energy deposited can be conducted away during the time of deposition. Thus, during the time of energy deposition the energy conducted away allows warming of fluid not in immediate contact with the melanin granule, which also leaves less heat to warm the fluid layer that does immediately surround the granule. Both of these effects result in a smaller $\nabla\theta$ for the fluid layer in immediate contact with a melanin granule and thus reduced amplitudes in the pressure waves. Maximum driving force occurs for maximum $\nabla\theta$, which occurs if the laser pulse τ is short compared to thermal conduction times. With a $v(\text{thermal}) \sim .01$ m/s, and relevant distances for a cellular fluid layer immediately surrounding a melanin granule of $l \sim 1 \mu\text{m}$, we get relevant thermal conduction times of .1 msec. Thus, for laser pulses with $\tau > 10^{-4}$ s, we can expect pressure induced damage to be negligible compared to a laser pulse with the same spatial energy profile, but with $\tau < 10^{-4}$ seconds.

Though damage resulting from laser induced acoustic transients is small compared to thermal damage, there is both experimental evidence and theoretical justification for its importance. Histological observations by Marshall [4] show evidence of outer segment membrane damage which were interpreted as due to pressure waves. Actual pressure measurements by Cleary [5] have implied $P \sim 10,000$ atmospheres near the edge of an absorbed laser pulse. This number can be used, along with some assumptions, to make a quick calculation of a possible underlying molecular biological mechanism for damage. The pressure waves are observed to have a frequency of $\nu \sim 10^6$ hertz, which with $u \sim 1500$ m/s, gives a wavelength of $\lambda \sim 1.5 \times 10^{-3}$ m over which this wave of amplitude of 10,000 atmospheres = 10^6 dyne/cm² goes from peak to trough. This gives a maximum force of the traveling wave of approximately

$$F = \nabla P dV = V_{\text{cell}} \Delta P / \lambda \approx 10^2 \text{ dyne} \quad (9)$$

The number of H-bonds that can be broken by this force is $10^2 \text{ dyne} / (.33 \times 10^{-4} \text{ dyne/H-bond}) \approx 3 \times 10^6$ H-bonds. If we picture membranes as ladders held together by rungs composed of H-bonds, and these H-bond rungs are spaced 1 Å apart, then the length of membranes that can be disrupted are

$$L \approx 3 \times 10^6 \text{ bonds} \times 10^{-10} \text{ m/bond} \approx 3 \times 10^{-4} \text{ m} \sim 15 \text{ RPE cells} \quad (10)$$

This can lead to immediate tearing of membranes associated with the RPE or photoreceptors, as well as damage resulting from the separation of the RPE membrane from the photoreceptor membranes which prevents the RPE from transferring nutrients and waste products between the photoreceptors and the choroid. This will also lead to death of photoreceptor cells. Thus, the possibility

of acoustic pressure transient damage is significant. As shown in Eqs. (3) and (8), this requires the computational modelling of the laser induced $\theta(r,t)$, which also leads to an examination of the more important damage mechanism due to temperature rises in the cell.

IV. THERMAL MODEL COMPUTATIONS

The most important mechanism for near threshold laser damage is thermal. We have set-up a detailed computer model in order to investigate the importance of various laser and biological characteristics on the damage caused by heating of the cell (as well as pressure effects described above). The model incorporates the following assumptions and protocol:

- 1) Photons are absorbed only by melanin granules in the RPE.
- 2) The biological volume, including the melanin granules, is divided into computational cells whose size is determined by computational time limitations.
- 3) A finite difference algorithm is employed to determine the heat entering each cell from adjacent cells which may be at different temperatures.
- 4) Time evolves through finite increments. The resolution of the time increments is determined by computational speed limitations and may vary during the course of a simulation of a single laser pulse (i.e. finer scale immediately after laser deposition, coarser scale as system approaches equilibrium).
- 5) At each time step, the heat flowing into or out of each cell is determined and the new temperature for each cell is computed.
- 6) Damage occurs when ΔT in a cell stays above a critical value for a critical length of time. The values of these parameters can either be chosen based upon an assumed underlying damage mechanism (i.e. protein denaturation for $\Delta T > 10$ C) or can be fitted based upon experimental data on threshold damage measurements.
- 7) Time stepping continues until spatially homogeneous ΔT occurs.

Using this finite difference model, the following output information will be generated:

- 1) θ (initial) for pressure calculations.
- 2) Largest ΔT .
- 3) Largest ΔT that occurs specifically in the photo-receptor region.
- 4) Spatial and temporal evolution of ΔT , especially at damage inducing levels (> 10 C).

Of crucial importance in this study is that this output information will be calculated as a function of experimentally variable input variables: laser pulse characteristics, and melanin granule density.

The underlying damage mechanism, consideration (6) above, will now be discussed.

V. THEORETICAL ANALYSIS OF DAMAGE MECHANISM

The above section describes the computational model that will be used to calculate ΔT of the biological material under a variety

of input conditions. This ΔT is the primary cause of the damage. The underlying mechanism that causes the damage to the biological tissue is of primary concern from a scientific viewpoint, as well as having importance in terms of practical considerations such as protective measures. In a complicated biological structure such as the retina, there are many coupled processes that are responsible for the biological functioning. This leads to great difficulty in determining which microscopic molecular process is the one that suffers the disruption and leads to the observed damage. However, anything less than the pinpointing of the underlying, primary molecular mechanism leaves incomplete understanding of the damage process.

At this time, understanding of the biochemical processes in cells is not at a level of pinpointing a molecular damage site. However, it is possible to model the kinetics of the process and thus calculate model independent parameters such as reaction rates and activation energies. These parameters can then be used as clues in the formation of a theoretical model for the molecular processes involved.

The connection to the computer Thermal Model is through ΔT . At present, interpretation of experimental results is based upon an Arrhenius damage integral formalism which has been used by a variety of researchers. It is based upon an analysis by Henriques [6] who admittedly used this analysis as merely an initial description of the process, and not as a detailed theory. As I will show below, this Arrhenius damage integral has recently been used in a detailed way that is unjustified on underlying theoretical grounds, and is of limited use in analyzing experimental results. The Arrhenius damage integral assumes that the damage mechanism follows a standard reaction rate expression:

$$\frac{d\Omega}{dt} = P e^{-\Delta E/R(T_0 + \Delta T)} \quad (11)$$

Attempts to use this expression in the rigorous manner intended for chemical reactions has led to problems when trying to adopt it directly to model laser induced damage. The parameter ΔE has a concise meaning in chemical reactions in terms of the activation energy for a reaction leading directly to the observed product. This assumes knowledge of a well defined chemical reaction pathway and it is not clear that this parameter is straightforwardly applicable for the complicated processes involved in laser induced retinal damage. Next, in using this equation to model the damage process, more attention needs to be paid to the 12D P. Again, the numerical value of this parameter determined from fitting the experimental data should be related to physical considerations of ΔS and attempt frequencies of the underlying processes. In addition, this equation implies that even at normal temperatures T_0 , $d\Omega/dt \neq 0$, yet buildup of damage does not occur at normal temperatures. Thus, simple use of Eq. (11) ignores the kinetics of a repair process.

Finally, another problem concerning Eq. (11) is its lack of

dependence on the characteristics of the laser pulse. The only laser depend quantity in Eq. (11) is ΔT which implies that all laser pulses inducing the same $\Delta T(t)$ will result in the same $d\Omega/dt$. For example, since thermal relaxation times are less than a μsec , any laser pulse with a duration τ_p shorter than this deposits its energy like $\delta(t)$. Thus, for pulse widths shorter than a μsec , the actual pulse duration is not important and the only important characteristic is the total energy of the pulse in determining $\Delta T(t)$. This is of course found to be true from experimental measurements of ED_{50} 's which remain approximately constant for pulse durations less than one μsec (the nanosecond "discrepancy" data are artifacts, and the ED_{50} curve should remain flat through the nanosecond time range, as will be shown later in this report). So, Eq. (11) at least seems to be successful in modeling the experimentally observed constancy of ED_{50} below a μsec . However, this success is of small value because constant ED_{50} for $\tau_p < \mu\text{sec}$ does not necessarily depend on the temperature profile following an exponential decay in time; any $f(\Delta T)$ in Eq. (11) would give the same prediction.

Given the problems just discussed, there are nevertheless reasons to maintain the Arrhenius damage integral formalism. Most importantly, the underlying mechanism causing the damage is some type of biological process and an exponential dependence on temperature can be expected, though in a more complicated fashion than expressed in Eq. (11). Also, there is a curious coincidence that when an expression as simple as Eq. (11) is used to fit the data, the value of ΔE is about 150 Kcal/mole which is just within the upper limit for energies for protein denaturation, a very possible cause for the observed damage. Thus, it is worthwhile to develop a theoretical model that is a more realistic version of Eq. (11) and which will incorporate features to correct some of the faults pointed out above.

This newer model assumes that the visible damage represented by Ω does depend on ΔT via intermediate biochemical reactions on some actual molecular system S . We now develop a theoretical treatment that incorporates both damage and repair mechanisms and thus leads to a more realistic and predictive understanding of the damage process. We will not be able to pinpoint specific biochemical molecules and reactions responsible for damage, but we will be able to better quantify the reaction parameters.

The following parameters are defined as:

S_D : Concentration of damaged molecular systems

S_U : Concentration of undamaged systems

Clearly, the observed damage Ω is a function of the concentration of damaged molecular systems; $\Omega = \Omega(S_D)$ and additionally $S_D = S_D(\Delta T)$. We will focus on the theoretical modeling of S_D as a step towards understanding the molecular cause of Ω .

There is a clear boundary condition on S_D as a function of temperature; when T is at normal temperatures T_0 , no molecular systems are damaged (This assumes that the only damage mechanism is thermal in nature. Pressure or chemical-toxic effects can be added.):

$$\frac{dS_D(T_0)}{dt} = 0 = \frac{dQ}{dt} \Big|_{T_0} = 0 \quad (12)$$

If we assume that the damage mechanism is a thermally activated process which speeds up with increasing temperature, such as protein denaturation, then there must be also be a repair or renewal process to balance out the damage that occurs slowly even at T_0 . The general thermodynamic picture of thermal damage and repair is

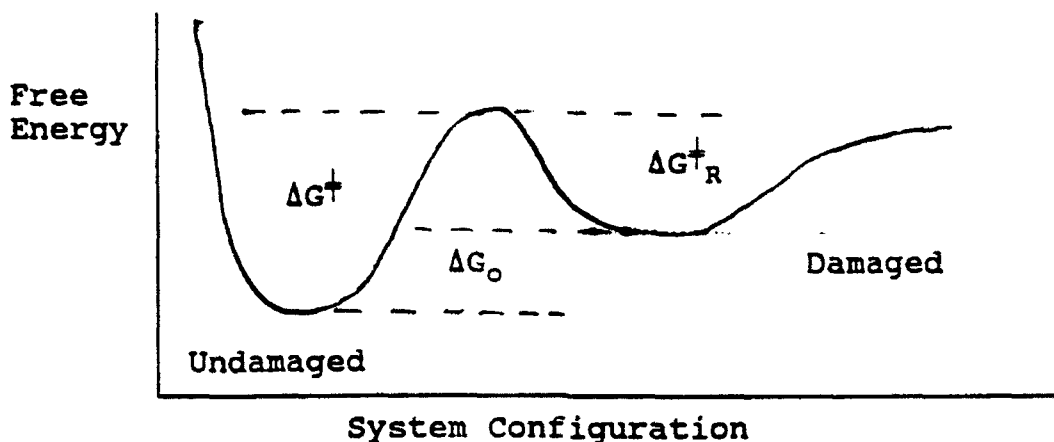


Figure 2. Free energies in two state transition model.

In the diagram, the various free energy differences have their usual significance: ΔG^\ddagger controls the rate at which undamaged systems are damaged, ΔG^\ddagger_R controls determines the rate of repair of damaged systems, and ΔG_0 controls the equilibrium concentration

$$K^{Eq} \equiv \frac{S_D}{S_U} = e^{-\Delta G_0/RT} \quad (13)$$

As in any transition state process, the rate of configuration interchange depends on several factors:

1) ν , barrier attempt frequency of a molecule due to vibrations of active normal modes. In thermal equilibrium, populated vibrational degrees of freedom have $\nu \approx kT/h \sim 10^{13} \text{sec}^{-1}$

2) $e^{-\Delta G^\ddagger/RT}$, the probability per attempt that the molecular system has the necessary energy to get to the top of the barrier

3) A , probability of going to other side if system gets to the top. (In an adiabatic process in which the initial and final electronic and nuclear states of the system are strongly coupled by many degrees of freedom, $A \sim 1$.)

The rate constant for a single process is $K(\text{sec}^{-1}) = \nu A e^{-\Delta G/RT}$. For the diagram above, we need to include both the damage process and the repair process

$$\frac{dS_U}{dt} = -v e^{-\Delta G^\ddagger/RT} S_U + v_R e^{-\Delta G_R^\ddagger/RT} S_D \quad (14)$$

Additional conditions are $S_U(t) + S_D(t) = S$ and, in equilibrium, $S_D/S_U = e^{\Delta G_0/RT}$. Relationships between various parameters in Eq. (14) can be determined for various conditions using the equality

$$\Delta G_R^\ddagger + \Delta G_0 = \Delta G^\ddagger \quad (15)$$

If the molecular systems have had time to equilibrate at a temperature T we have

$$\frac{dS_U}{dt} = -v e^{-\Delta G^\ddagger/RT} S_U + v_R e^{-\Delta G_R^\ddagger/RT} e^{-\Delta G_0/RT} S_U \quad (16)$$

At T_0 (normal retinal temperature) we know that in a healthy individual $dS_U/dt = 0$, then Eqs. (15) and (16) give the relationship

$$v(T_0) = v_R(T_0) \quad (17)$$

Interestingly, it would be expected that $v > v_R$ since almost any normal mode can lead to damage, but only a few are likely to lead to repair. Therefore, for Eq. 17 to hold true it implies that other factors besides intramolecular vibrations are assisting in repair, such as enzyme catalyzed processes, in order to raise v_R to v at T_0 .

Eq. (16) allows a theoretical analysis of the conditions leading to damage. For example, even though $v(T_0) = v_R(T_0)$ it is not necessary to have $v(T) = v_R(T)$ for temperatures other than T_0 . A repair involves a decrease of entropy of the molecular system which therefore requires coupling to peripheral reactions. As T increases, the coupled reactions are likely to be less effectively coupled and coordinated which will prevent v_R from increasing with T as quickly as v (in fact v_R may actually decrease with increasing T). Thus, Eq. (16) predicts damage occurring for $T > T_0$ because of $v(T) > v_R(T)$. An important question for theoretical investigation is to understand on a molecular mechanism how v and v_R vary with temperature, and for experimental measurements to be made that can focus on these dependences.

Another source of net damage in Eq. (16) is the substitution in Eq. (14) of

$$S_D = S_U e^{-\Delta G_0/RT} \quad (18)$$

to get the repair rate contribution in Eq. (16). However, Eq. (18) is only valid if the ensemble of systems S have attained configurational equilibrium at the new temperature T . For times shorter than this equilibration time, $t \ll \tau_{Eq}$, the number of damaged systems will still be stuck at the equilibrium concentration of the initial temperature $S_D \approx S_D(T_0)$ which is less than the concentration

of damaged systems once equilibrium has been attained at a higher T, i.e.

$$\text{for } t \ll \tau_{\text{Eq}}, \quad S_D \approx S_D(T_0) < S_D^{\text{Eq}}(T > T_0) \quad (19)$$

Therefore, damage builds up because too few S_D are available to be repaired compared to the correct equilibrium concentration at the new T.

We can therefore expect damage build up from Eq. (16) at elevated temperatures because of both $v > v_R$ and Eq. (19). It is worthwhile to note that these different underlying physical mechanisms may be discernible experimentally because of the different time scales in which they dominate in causing damage:

1) Ultrashort ($t \leq 10^{-13}$ sec), $S_D < S_D^{\text{Eq}}(T)$ dominates. At these short times there has not been enough time for either damage or repair processes to make many attempts at crossing the barrier so the difference between v and v_R is unimportant. Therefore

$$\frac{dS_D}{dt} \approx v (-e^{-\Delta G^*/RT} S_D + S_D(T_0) e^{-\Delta G_R^*/RT}) \quad (20)$$

and using $S_D(T_0) = S_U(T_0) \exp(-\Delta G_0/RT_0)$ we have

$$\frac{dS_D}{dt} \approx v (-e^{-\Delta G^*/RT} + e^{-\Delta G_0/RT_0} e^{-\Delta G_R^*/RT}) S_D < 0 \quad (\text{for } T_0 < T) \quad (21)$$

2) $t >$ inter-configurational equilibrium time (e.g. protein folding times of ~ 1 second) damage is caused by $v(T) > v_R(T)$. For these longer times $S_D = S_D^{\text{Eq}}(T) = S_U \exp(-\Delta G_0/RT)$ and

$$\frac{dS_D}{dt} \approx S_U (-v e^{-\Delta G^*/RT} + v_R e^{-\Delta G_0/RT} e^{-\Delta G_R^*/RT}) = S_U e^{-\Delta G^*/RT} (-v + v_R) \quad (22)$$

where Eq. (15) has been used. With $v(T) > v_R(T)$, Eq. (22) shows that $dS_D/dt < 0$ and therefore $d\Omega/dt > 0$.

3) Middle times (10^{-13} sec $< t < 10^0$ sec): both of the above can contribute.

The parameters of interest for understanding the underlying physical damage mechanism(s) are the $v(T)$'s, and ΔG 's of Eq. (14) just discussed. To get these values requires experimental measurements that track $d\Omega/dt$ and dS_D/dt as functions of ΔT and time. These measurements are straightforward for times greater than seconds, but are obviously much more difficult for shorter times. A major experimental challenge will be the tracking of damage at ultrashort time scales. It is likely that tracking damage development at these ultrashort times will rely on spectroscopic techniques such as absorptivity, reflectivity, and circular dichroism changes, since it is not obvious that chemical assay techniques can discriminate over these short time scales.

Assuming that experimentalists can figure out a way of making the relevant measurements, the theoretical analysis would continue

as follows. We rewrite Eq. (14) and substitute in $K=S_D/S_U$ where the ratio K depends on T and t as explained above.

$$\frac{dS_U}{dt} = -v e^{-\Delta G^*/RT} S_U + v_R K e^{-\Delta G_R^*/RT} S_U \quad (23)$$

This leads to

$$S_U(t) = S_U(T_0) e^{\int_0^t (-v e^{-\Delta G^*/RT} + v_R K e^{-\Delta G_R^*/RT}) dt} \quad (24)$$

We emphasize that the two terms in Eq. (24) is the minimum number that must be present to model the damage and repair mechanisms. We also reiterate that only two terms have been used as an initial step towards a full model and that inherently multi-reaction repair mechanisms have been compressed (and oversimplified) into the parameters $v_R(T)$ and ΔG_R^* . The next step in developing the model is to determine if v 's and ΔG 's can be found to reproduce the experimentally observed time course of $S_U(t)$. If this turns out to be not possible then it will be necessary to add additional terms to Eqs. (23) and (24) to reflect the complexity of the damage and repair mechanisms.

To make the connection to experimental measurements also requires the determination of the functional dependence of the observable damage Ω to the molecular damage $S_D=S-S_U$. It is possible with standard benchtop biochemistry to take retinal segments and expose them to elevated temperatures and measure molecular damage $S_D(T,t)$. The relationship of this *in vitro* molecular damage to observable lesions *in vivo* is at present unknown. The observable damage may scale linearly with the molecular damage

$$\Omega(t) - S_D(t) = S - S_U(t) \quad (25)$$

or, there may be a critical concentration of molecular damage S_D^c necessary before observable damage Ω appears:

$$\Omega(t) - S_D - S_D^c = (S - S_D^c) - S_U(t) \quad (26)$$

where

$$S_D^c = S - S_U^c = S - S_U(T_0) \exp_0 \int^{t_c} (-v e^{-\Delta G^*/RT} + v_R K e^{-\Delta G_R^*/RT}) dt \quad (27)$$

This equation presents a possible approach of using experimental data to get underlying physical parameters like v 's and ΔG 's. We can assume that the critical molecular damage $S_D^c=S-S_U^c$ at which Ω first is observable is a property of the retina and independent of the characteristics of the laser pulse, i.e. the observable retinal damage has the same experimental endpoint. In this situation, the time t_c , the time for a minimum visible lesion (MVL) to build up, depends on the time course of the temperature dependence $T(t)$. The

computational thermal model described earlier in this report can then be used to get the $T(t)$ profile for any given laser characteristics that have been found experimentally to produce an MVL. Since we assume that all MVL have the same Ω when damage is first observed at the time t_c (which will vary depending on the laser characteristics), then S_D^c will be the same for all MVL and we can therefore say that

$$\int^{t_c} (-v e^{-\Delta G^*/RT} + v_R K e^{-\Delta G_R^*/RT}) dt = \text{Constant} \quad (28)$$

It is important to note that the Constant in Eq. (28) has no physical meaning. Its importance lies in the fact that it is assumed to be the same for all laser pulses. Inserting into Eq. (28) various functional forms of $T(t)$ determined for different MVL laser pulses from computations using the thermal computational model, the integral in Eq. (28) can be carried out analytically to give the constant in terms of t_c . Experimental measurements, if possible, of t_c 's for these laser pulses might then be used to get values of v 's and ΔG 's.

VI. RESOLUTION OF THE "NANOSECOND DISCREPANCY"

Finally, this report will end with an explanation that resolves the question of the apparent "nanosecond discrepancy". Experimental observations by Sliney [1] appeared to show that the ED_{50} for pulse lengths of approximately 100 nsec were an order of magnitude higher than the ED_{50} 's for pulse lengths shorter or longer. If this were true, it would clearly be very important from the practical standpoint and would also be a valuable source of information for understanding the interplay of the basic, underlying physical mechanisms that cause damage. However, Allen et. al. [7] saw no experimental evidence for an increase in the ED_{50} in this time range, and they postulated that the results of Sliney's group were due to the fact that the 100 nsec pulses used by Sliney were actually composed of multiple shorter pulses. As will now be shown, computational analysis strongly supports this multiple short pulse proposal and therefore the nanosecond discrepancy is likely not to exist.

We will now show that when the anomalously high Sliney ED_{50} in the 100 nsec range is treated as a multiple pulse energy, then each of the constituent shorter pulses has an ED_{50} of approximately the same value as measured for other pulses of duration less than a μ sec. The basis for this analysis is the empirical relationship that has been discovered that relates a single pulse ED_{50} with that of N multiple pulses of the same duration, ED_{50}^N

$$ED_{50}^N = \frac{ED_{50}}{N^{1/4}} \quad (29)$$

Sliney's group used a Q-switched ND-YAG laser that likely had an undetected model locked pulse train of 83 psec pulses with the

time between pulses approximately 20 times longer than the duration of the 83 psec pulses themselves. Therefore, in a time of 100 nsec there is room for approximately 50 of the shorter 83 psec pulses along with their concomitant delay time in between. The total exposure, reported as the single pulse $ED_{50}(100 \text{ nsec})$, was $5 \times 10^{-4} \text{ J/cm}^2$. Each of the fifty 83 psec pulses in this 100 nsec exposure would have approximately $1.0 \times 10^{-5} \text{ J/cm}^2$. Thus the 100 nsec exposure gives an $ED_{50}^N(83 \text{ psec})$ of $1.0 \times 10^{-5} \text{ J/cm}^2$ for $N \sim 50$. We now compare this to the ED_{50}^N predicted from Eq. (25). The ED_{50} in the 100 psec range has been measured to be approximately $5 \times 10^{-5} \text{ J/cm}^2$. Using Eq. (25), we therefore would expect an $N=50$ multiple pulse train to have $ED_{50}^N = 5 \times 10^{-5} \text{ J/cm}^2 / \sqrt{50} = .71 \times 10^{-5} \text{ J/cm}^2$ which is extremely close to the energy per pulse if the 100 nsec exposure actually was composed of fifty 83 psec pulses.

The extremely close agreement between the two values, i.e. $1.0 \times 10^{-5} \text{ J/cm}^2$ vs. $.71 \times 10^{-5} \text{ J/cm}^2$, is strong supporting evidence that the anomalous 100 nsec exposure was actually composed of multiple exposures and that there is therefore no nanosecond discrepancy.

VII. SUMMARY

A detailed computer model has been developed of the primary absorption site of visible radiation leading to retinal damage. This site is the retinal pigment epithelium. The computer program derived from the model allows the density of the highly absorbing pigment granules to be varied. Also incorporated as a user supplied input variable are the laser pulse characteristics: duration, energy, width. This computational model allows for a systematic investigation of the functional dependence of the damage on the retina's properties and the laser properties. Additional, variables can be included as deemed important. Upon the choice of melanin density and laser characteristics, the computer modeling will produce a temperature rise profile as a function of time and location within the retinal region.

The temperature profile calculations can then be used with a thermodynamic analysis of the damage process in order to understand and predict damage under given conditions. A preliminary thermodynamic model is presented to supplant the simpler Arrhenius damage integral, whose limitations are discussed.

Finally, the "nanosecond discrepancy" that appeared to show an ED_{50} for nanosecond pulses that was an order of magnitude higher than for pulses of both shorter or longer duration has been shown to most likely not exist. It is instead probably due to the picosecond multiple pulse makeup of the nanosecond pulses used in the experiment that reported the anomalously high ED_{50} .

Acknowledgements

This work was done in collaboration with Lt. Colonel M. Rogers, who also arranged for my visit, and Captain R. Thompson, both of Armstrong Laboratory, Brooks Air Force Base. This work was also aided by helpful discussions with Captain P. Roach and Major F. Cheney of Armstrong Laboratory, Brooks Air Force Base, and Dr.

R. Glickman and Dr. J. Zuchlich of Krug Sciences, San Antonio, Texas. The work was funded by an Air Force Office of Scientific Research Summer Faculty Research Appointment.

REFERENCES

- 1) D. H. Sliney, First International Symposium on Laser Biological Effects and Exposure Limits, Ed. by L. Court, A. Duchene, and D. Courant, Paris, Nov. 1986.
- 2) G. Gueneau, V. Baille, D. Courant, M. Dubos, and L. Court, First International Symposium on Laser Biological Effects and Exposure Limits, Ed. by L. Court, A. Duchene, and D. Courant, Paris, Nov. 1986.
- 3) Hayes and Wolbarsht, *Aerospace Medicine*, 39, 474, 1968.
- 4) J. Marshall, *Invest. Ophthalmol.*, 9, 97, 1970.
- 5) S. F. Cleary, *Laser Applications in Medicine and Biology*, Edited by M. Wolbarsht, Plenum Press, NY, 1977.
- 6) F. C. Henriques, *Archives of Pathology*, 43, 489, 1947.
- 7) R. G. Allen, S. J. Thomas, R. F. Harrison, J. A. Zuchlich and M. F. Blankenstein, *Health Physics*, 49, 685, 1985.

Estimation of Dioxin Half-Life in the Air Force Health Study

Pushpa L. Gupta
Professor
Department of Mathematics

University of Maine
321 Neville Hall
Orono, Maine 04469

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

**ESTIMATION OF DIOXIN HALF-LIFE IN THE
AIR FORCE HEALTH STUDY**

Pushpa L. Gupta

Professor

Department of Mathematics

University of Maine

ABSTRACT

In this project the estimation of half-life of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) in humans is studied by examining the mixed effects model with repeated measurements studied by Michalek et. al. (1992). In the case of two measurements and a single covariate, analytical expression for the restricted maximum likelihood estimators of the parameters and their variance-covariance matrix are derived. The results are then extended to the general case with arbitrary number of measurements and covariates. The purpose of this investigation is to compare the efficiencies of the estimator of half-life thus obtained and the estimator of half-life obtained by using the unconditional model (full model) introduced by Gupta (1990, 1991). This will enable the Air Force scientists to select an appropriate method for estimating the half-life.

ESTIMATION OF DIOXIN HALF-LIFE IN THE AIR FORCE HEALTH STUDY

Pushpa L. Gupta

1. INTRODUCTION

The estimation of half-life of TCDD in Ranch Hand Vietnam veterans has been of considerable interest to the U.S. Air Force scientists. For the past several years, epidemiology division of the USAF School of Aerospace Medicine at Brooks Air Force Base has been conducting studies to estimate the median half-life of the TCDD concentration level in the Ranch Hands. Estimation of TCDD half-life has been made possible due to advances in chemistry which have allowed the scientists to measure TCDD concentration in human serum. The scarcity of exposed subjects with saved serum has led to half-life studies based only on two measurements per subject. These measurements were taken 15 to 20 years after exposure in Vietnam.

There are five procedures available for estimating the half-life of TCDD in Ranch Hand Vietnam veterans: (1) a non-parametric estimate of Phillips (1989), (2) a non-parametric estimate of Pirkle et. al. (1989), (3) A parametric estimate of Michalek et. al. (1989), (4) an estimate based on mixed-linear model with covariates, Tripathi (1989, 1990) and Michalek et. al. (1992), (5) a parametric estimate based on conditional and unconditional model (full model) and on extensions of the full model accommodating multiple measurements and covariates, Gupta (1990, 1991).

For the Ranch Hand data, the half-life estimate, based on mixed-effects linear model, was obtained using SAS GLM procedure, as the

analytical expressions for the variance and covariances of the parameter estimators were not yet available. Thus there was a need to derive such analytical expressions which could be applied to any other data set and which would enable us to compare these estimators with the estimators obtained by other methods.

Section 2 contains the review of the full model along with the analytical expression for the variance of the estimator of half-life. In section 3, we derive the variance-covariance matrix of the parameters of the mixed-effects linear model with two measurements and one covariate. These results are extended, in section 4, for multiple measurements and covariates. Some conclusions and recommendations are given in section 5.

2. UNCONDITIONAL MODEL (FULL MODEL)

In this section we will consider the unconditional (full model) and derive the maximum likelihood estimators (MLE's) for the parameters which in turn will yield estimator for the median half-life, and an estimator of its standard error. We assume that the first order kinetics model

$$C_t = C_0 e^{-\lambda t} \quad (2.1)$$

holds, where C_t is the TCDD concentration t years after exposure, C_0 is the initial exposure (unknown) and λ is the unknown decay rate which we are assuming to be random. Then the half-life is given by $Y = \frac{\log 2}{\lambda}$, where \log indicates the natural logarithm. Based on two TCDD measurements C_1 and C_2 taken t_1 and t_2 years after exposure, Y can be written as

$$Y = \Delta t \log 2 / \log(C_1/C_2) = d/X,$$

where $\Delta t = t_2 - t_1 > 0$, $X = \log(C_1/C_2)$, and $d = \Delta t \log 2$. Since

$\log C_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$ (see Wolfe et. al. 1988),

$\log(C_1/C_2) \sim N(\mu, \sigma^2)$, where $\mu = \mu_1 - \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$,

$\sigma_{12} = \text{Cov}(\log C_1, \log C_2)$.

We shall denote the distribution function and the density function of a standard normal by $\Phi(\cdot)$ and $\phi(\cdot)$ respectively. Then the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{d}{y^2} e^{-(1/2\sigma^2)((d/y)-\mu)^2}, & y \neq 0 \\ 0, & y = 0 \end{cases} \quad (2.2)$$

The median half-life M is given by

$$M = \frac{d}{\mu + \sigma \Phi^{-1}(a + \phi(-\mu/\sigma))}, \quad (2.3)$$

where $a = \pm .5$ according as $M \gtrless 0$.

We denote the nonconstant part of the full log-likelihood as ℓ_F , where

$$\ell_F(\mu, \sigma^2 | y_1, i=1, 2, \dots, N) = -N \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^N \left(\frac{d_i}{y_i} - \mu \right)^2 \quad (2.4)$$

where $d_i = \Delta t_i \log 2$, $i = 1, 2, \dots, N$.

The maximum likelihood estimators of μ and σ together with their expectations, variances and covariances are given by the following:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N d_i / y_i, \quad E(\hat{\mu}) = \mu, \quad \text{Var}(\hat{\mu}) = \sigma^2 / N$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N ((d_i/y_i) - \hat{\mu})^2, \quad E(\hat{\sigma}^2) = \left(\frac{N-1}{N}\right)\sigma^2, \quad \text{Var}(\hat{\sigma}^2) = \frac{2(N-1)\sigma^4}{N^2},$$

$$\text{Cov}(\hat{\mu}, \hat{\sigma}^2) = 0.$$

Thus the estimator of the median half-life M is

$$\hat{M} = \frac{d}{\hat{\mu} + \hat{\sigma}\phi^{-1}(a + \phi(-\hat{\mu}/\hat{\sigma}))} = g(\hat{\mu}, \hat{\sigma}^2) \quad (2.5)$$

and

$$\begin{aligned} \text{Var}(\hat{M}) &= \text{Var}(g(\hat{\mu}, \hat{\sigma}^2)) = \left(\frac{\partial g}{\partial \hat{\mu}}\right)^2 \text{Var}(\hat{\mu}) + \left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)^2 \text{Var}(\hat{\sigma}^2) + 2\left(\frac{\partial g}{\partial \hat{\mu}}\right)\left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)\text{Cov}(\hat{\mu}, \hat{\sigma}^2) \\ &= \left(\frac{\partial g}{\partial \hat{\mu}}\right)^2 \left(\frac{\sigma^2}{N}\right) + \left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)^2 \left(\frac{2(N-1)\sigma^4}{N^2}\right), \end{aligned} \quad (2.6)$$

see Elandt-Johnson and Johnson (1980), where the partial derivatives are evaluated at $(\mu, (N-1)\sigma^2/N)$. The partials of $g(\cdot)$ are given by

$$\frac{\partial g}{\partial \hat{\mu}} = -\frac{g^2}{d} \left[1 - \frac{\phi(-\hat{\mu}/\hat{\sigma})}{\phi\left(\frac{d - g\hat{\mu}}{g\hat{\sigma}}\right)} \right]$$

$$\frac{\partial g}{\partial \hat{\sigma}^2} = -\frac{g^2}{2d\hat{\sigma}^2} \left[\frac{d}{g} - \hat{\mu} \left(1 - \frac{\phi(-\hat{\mu}/\hat{\sigma})}{\phi\left(\frac{d - g\hat{\mu}}{g\hat{\sigma}}\right)} \right) \right]$$

$$= \frac{-g}{2\hat{\sigma}^2} \left[1 + \frac{\hat{\mu}}{g} \frac{\partial g}{\partial \hat{\mu}} \right], \quad (2.7)$$

where g is given by (2.5). For more information, see Gupta (1990).

3. VARIANCE-COVARIANCE MATRIX OF THE ESTIMATORS OF THE PARAMETERS (SINGLE COVARIATE)

In this section, we shall obtain the variance-covariance matrix of the estimators of the parameters in the mixed-effects linear model, which is introduced below.

Let the j th measurement on the i th individual be given by

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 t_{ij} + \epsilon_{ij}, \quad j = 1, 2.$$

Then

$$\begin{aligned} Y_i &= \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} \\ &= X_i \beta + Z_i \alpha_i + \epsilon_i, \quad i = 1, 2, \dots, N, \end{aligned} \quad (3.1)$$

where X_i and Z_i are known 2×2 and 2×1 design matrices, β is a vector of 2 fixed population parameters to be estimated, α_i is random effect, and ϵ_i is a random error vector associated with the i th individual, such that

$$\epsilon_i \sim N(0_{2 \times 1}, \sigma^2 I_2)$$

$$\alpha_i \sim N(0, \sigma_\alpha^2 = \sigma^2 \theta).$$

It follows that

$$Y_i | \alpha_i \sim N(X_i \beta + Z_i \alpha_i, \sigma^2 I_2)$$

and

$$Y_i \sim N(X_i \beta, \Sigma_i),$$

where $\Sigma_i = \sigma^2 V_i(\theta)$, $V_i(\theta) = I_2 + \theta \mathbf{1}_2 \mathbf{1}_2'$ for all i . I_2 denotes 2×2 identity matrix and $\mathbf{1}_2$ is a 2×2 matrix of ones. Y_i and Y_j are independent for $i \neq j$.

Now if we let $Y' = [Y'_1, Y'_2, \dots, Y'_N]$, $X' = [X'_1, X'_2, \dots, X'_N]$,
 $Z = \text{diag}(Z_1, Z_2, \dots, Z_N)$, $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_N)$,
 $\epsilon' = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_N)$, $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_N)$, then we can write
the full model as

$$Y = X\beta + Z\alpha + \epsilon, \quad (3.2)$$

where

$$Y|\alpha \sim N(X\beta + Z\alpha, \sigma^2 I_{2N})$$

and

$$Y \sim N(X\beta, \Sigma),$$

where

$$\Sigma = \sigma^2 I_{2N} + \theta I_{2N}.$$

Since $Y_i \sim N(X_i\beta, \sigma^2 V_i)$, $i = 1, 2, \dots, N$, and independent, the full-likelihood, without the constant term, is given by

$$\ell_F = -N \log \sigma^2 - \frac{1}{2} \sum_{i=1}^N \log |V_i(\theta)| - \frac{1}{2\sigma^2} \sum_{i=1}^N r_i^T V_i^{-1}(\theta) r_i, \quad (3.3)$$

where $r_i = y_i - X_i\beta$. Since $V_i(\theta)$ is dependent on θ , in order to save space we will drop the argument θ and write V_i instead of $V_i(\theta)$.

A general criticism of the ML estimators for the variance components σ and θ , obtained from (3.2), is that they are biased downward because they do not take into account the loss of degrees of freedom from the estimation of β . The restricted maximum likelihood (RML) method corrects for this by defining estimates of the variance components as the maximizers of the log-likelihood based on $2N - 2$ linearly independent contrasts, where $2N$ is the total number of observations from all individuals. Thus our estimates of θ , σ and β

are based on restricted log-likelihood (see Lindstrom and Bates, 1988), ℓ_R (ignoring the constant term) given below,

$$\begin{aligned} \ell_R(\beta, \sigma, \theta | y) &= -\frac{1}{2} \log \left| \sigma^{-2} \sum_{i=1}^N X_i^T V_i^{-1} X_i \right| + \ell_F(\beta, \sigma, \theta | y) \\ &= -(N-1) \log \sigma^2 - \frac{1}{2} \log \left| \sum_{i=1}^N X_i^T V_i^{-1} X_i \right| - \frac{1}{2\sigma^2} \sum_{i=1}^N r_i^T V_i^{-1} r_i \\ &\quad - \frac{1}{2} \sum_{i=1}^N \log |V_i|. \end{aligned} \quad (3.4)$$

To simplify calculations we obtain first the RML estimate of σ^2 , which is

$$\hat{\sigma}_{RML}^2 = \frac{\sum_{i=1}^N r_i^T V_i^{-1} r_i}{2N-2} \quad (3.5)$$

and substitute $\hat{\sigma}_{RML}^2$ in equation (3.4) to get the profile log-likelihood of β and θ , as

$$p_R(\beta, \theta | y) = -(N-1) \log \sum_{i=1}^N r_i^T V_i^{-1} r_i - \frac{1}{2} \log \left| \sum_{i=1}^N X_i^T V_i^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^N \log |V_i|. \quad (3.6)$$

The derivatives of the profile log-likelihood with respect to β and θ are given below as:

$$\frac{\partial p_R(\beta, \theta | y)}{\partial \beta} = 2(N-1) \frac{\sum_{i=1}^N X_i^T V_i^{-1} r_i}{\sum_{i=1}^N r_i^T V_i^{-1} r_i} = 2(N-1) \frac{H_2(V^{-1})}{H_1(V^{-1})}. \quad (3.7)$$

$$\frac{\partial p_R(\beta, \theta | y)}{\partial \theta} = (N-1) \frac{\sum_{i=1}^N r_i^T A r_i}{\sum_{i=1}^N r_i^T V_i^{-1} r_i} + \frac{1}{2} \text{trace} \left(\left(\sum_{i=1}^N X_i^T V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i^T A X_i \right) \right)$$

$$\begin{aligned}
& - \frac{1}{2} \sum_{i=1}^N \text{trace} \left(v_i^{-1} \frac{\partial v_i}{\partial \theta} \right) \\
& = \frac{(N-1)}{(2\theta+1)^2} \frac{\sum_{i=1}^N r_i^T (\perp_2) r_i}{\sum_{i=1}^N r_i^T v_i^{-1} r_i} + \frac{1}{2} \text{trace} \left(\left(\sum_{i=1}^N X_i^T v_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i^T (\perp_2) X_i \right) \right) / (2\theta+1)^2 \\
& - \frac{N}{2} \text{trace} \left(\frac{1}{(2\theta+1)} (\perp_2) \right) \\
& = \frac{N-1}{(2\theta+1)^2} \frac{H_1(\perp_2)}{H_1(v^{-1})} + \frac{1}{2(2\theta+1)^2} \text{trace} \left((H(v^{-1}))^{-1} H(\perp_2) \right) - \frac{N}{2\theta+1} \quad (3.8)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta} & = 2(N-1) \left[- \frac{\sum_{i=1}^N X_i^T v_i^{-1} X_i}{\sum_{i=1}^N r_i^T v_i^{-1} r_i} + \frac{2 \left(\sum_{i=1}^N X_i^T v_i^{-1} r_i \right) \left(\sum_{i=1}^N X_i^T v_i^{-1} r_i \right)^T}{\left(\sum_{i=1}^N r_i^T v_i^{-1} r_i \right)^2} \right] \\
& = 2(N-1) \left[- \frac{H(v^{-1})}{H_1(v^{-1})} + \frac{2H_2(v^{-1})(H_2(v^{-1}))^T}{(H_1(v^{-1}))^2} \right] \quad (3.9)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2} & = \frac{(N-1)}{(2\theta+1)^3} \left[- \frac{4H_1(\perp_2)}{H_1(v^{-1})} + \frac{1}{(2\theta+1)} \left(\frac{H_1(\perp_2)}{H_1(v^{-1})} \right)^2 \right] \\
& + \frac{1}{2(2\theta+1)^3} \text{trace} \left((H(v^{-1}))^{-1} H(\perp_2) (H(v^{-1})) H(\perp_2) - 4(H(v^{-1}))^{-1} H(\perp_2) \right) \\
& + \frac{2N}{(2\theta+1)^2} \quad (3.10)
\end{aligned}$$

$$\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta \partial \beta} = \frac{2(N-1)}{(2\theta+1)^2} \left[- \frac{H_2(\perp_2)}{H_1(v^{-1})} + \frac{H_2(v^{-1}) H_1(\perp_2)}{(H_1(v^{-1}))^2} \right] \quad (3.11)$$

where $v_1 = v = \begin{bmatrix} \theta+1 & \theta \\ \theta & \theta+1 \end{bmatrix}$, $|v_1| = 2\theta+1$, $v_1^{-1} = \frac{1}{2\theta+1} \begin{bmatrix} \theta+1 & -\theta \\ -\theta & \theta+1 \end{bmatrix}$,

$$\perp_2 = \begin{bmatrix} 11 \\ 11 \end{bmatrix}, \quad A = v_1^{-1} \frac{\partial v_1}{\partial \theta} v_1^{-1} = \frac{\perp_2}{(2\theta+1)^2}, \quad v_1^{-1} \frac{\partial v_1}{\partial \theta} = \frac{\perp_2}{(2\theta+1)},$$

$$\text{and } H(S) = \sum_{i=1}^N X_1^T S X_1, \quad H_1(S) = \sum_{i=1}^N r_1^T S r_1, \quad \text{and } H_2(S) = \sum_{i=1}^N X_1^T S r_1,$$

for details, see appendix I.

Since $\sigma^{-2} H_1(V^{-1}) \sim \chi_{2N}^2$ and $H_2(V^{-1}) \sim N(0_{2 \times 1}, \sigma^2 H(V^{-1}))$ (see Appendix II),

$$\begin{aligned} E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta}\right) &= 2(N-1) \left[-H(V^{-1}) E\left(\frac{1}{H_1(V^{-1})}\right) + 2E\left(\frac{H_2(V^{-1})(H_2(V^{-1}))^T}{(H_1(V^{-1}))^2}\right) \right] \\ &= 2(N-1) \left[-\frac{H(V^{-1})}{2(N-1)\sigma^2} + 2E\left(\frac{H_2(V^{-1})(H_2(V^{-1}))^T}{(H_1(V^{-1}))^2}\right) \right] \\ &= 2(N-1) \left[-\frac{H(V^{-1})}{2(N-1)\sigma^2} + 2 \frac{E(H_2(V^{-1})(H_2(V^{-1}))^T)}{E((H_1(V^{-1}))^2)} \right] \\ &= 2(N-1) \left[-\frac{H(V^{-1})}{2(N-1)\sigma^2} + 2 \frac{\sigma^2 H(V^{-1})}{\sigma^4 [2(2N) + (2N)^2]} \right] \\ &= -\left(\frac{N^2+1}{N^2+N}\right) \frac{H(V^{-1})}{\sigma^2} \\ &= -\left(\frac{N^2+1}{N^2+N}\right) \frac{1}{(2\theta+1)\sigma^2} \begin{bmatrix} 2N & \sum_{i=1}^N (t_{11} + t_{12}) \\ \sum_{i=1}^N (t_{11} + t_{12}) & \sum_{i=1}^N (\theta+1)(t_{11} - t_{12})^2 + 2t_{11}t_{12} \end{bmatrix}. \end{aligned} \tag{3.12}$$

Since $\frac{\sigma^{-2}}{2(2\theta+1)} H_1(\pm_2) \sim \chi_N^2$ and $\sigma^{-2} H_1(V^{-1}) \sim \chi_{2N}^2$,

$$\frac{\sigma^{-2} H_1(\pm_2)}{2(2\theta+1)\sigma^{-2} H_1(V^{-1})} = \frac{1}{2(2\theta+1)} \frac{H_1(\pm_2)}{H_1(V^{-1})} \sim \text{Beta}(N/2, N/2) \text{ (appendix II).}$$

Therefore,

$$F\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2}\right) = \frac{(N-1)}{(2\theta+1)^3} \left[-4E\left(\frac{H_1(\pm_2)}{H_1(V^{-1})}\right) + \frac{1}{2\theta+1} E\left(\frac{H_1(\pm_2)}{H_1(V^{-1})}\right)^2 \right]$$

$$\begin{aligned}
& + \frac{2N}{(2\theta + 1)^2} + k \\
& = \frac{N-1}{(2\theta + 1)^3} \left[-4(2\theta + 1) + 4(2\theta + 1) \frac{(N+2)}{4(N+1)} \right] + \frac{2N}{(2\theta + 1)^2} + k \\
& = -\frac{(N^2 - 3N - 2)}{(N+1)(2\theta + 1)^2} + k, \tag{3.13}
\end{aligned}$$

where

$$k = \frac{1}{2(2\theta + 1)^3} \text{trace} \left(\frac{(H(V^{-1}))^{-1}H(\perp_2)(H(V^{-1}))^{-1}H(\perp_2)}{2\theta + 1} - 4(H(V^{-1}))^{-1}H(\perp_2) \right)$$

$$\begin{aligned}
E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta \partial \beta}\right) &= \frac{2(N-1)}{(2\theta + 1)^2} \left[-E\left(\frac{H_2(\perp_2)}{H_1(V^{-1})}\right) + E\left(\frac{H_2(V^{-1})H_1(\perp_2)}{(H_1(V^{-1}))^2}\right) \right] \\
&= \frac{2(N-1)}{(2\theta + 1)^2} \left[-\frac{E(H_2(\perp_2))}{E(H_1(V^{-1}))} + \frac{E(H_2(V^{-1})H_1(\perp_2))}{E(H_1(V^{-1}))^2} \right] \\
&= \frac{2(N-1)}{(2\theta + 1)^2} \left[0_{2 \times 1} + 0_{2 \times 1} \right] = 0_{2 \times 1},
\end{aligned}$$

since $E(H_2(\perp_2)) = 0_{2 \times 1}$ and $E(H_2(V^{-1})H_1(\perp_2)) = 0_{2 \times 1}$ (for more explanation, see appendix II). (3.14)

Therefore, the information matrix I of $\hat{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$ is of the form

$$I = \begin{bmatrix} a & 0 & 0 \\ 0 & b & d \\ 0 & d & c \end{bmatrix},$$

where

$$\begin{aligned}
a &= -E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2}\right) \\
&= \frac{N^2 - 3N - 2}{(N+1)(2\theta + 1)^2} - \frac{1}{2(2\theta + 1)^3} \text{trace} \left(\frac{(H(V^{-1}))^{-1}H(\perp_2)(H(V^{-1}))^{-1}H(\perp_2)}{2\theta + 1} \right)
\end{aligned}$$

$$+ \frac{1}{2(2\theta + 1)^3} \text{trace} (4(H(V^{-1}))^{-1}H(\pm_2)),$$

$$b = - E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta_0^2} \right) = \frac{2N(N^2 + 1)}{\sigma^2(2\theta + 1)(N^2 + N)}, \text{ see (3.12)}$$

$$d = - E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta_0 \partial \beta_1} \right) = \frac{\left(\sum_{i=1}^N t_{i1} + t_{i2} \right) (N^2 + 1)}{\sigma^2(2\theta + 1)(N^2 + N)},$$

$$c = - E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta_1^2} \right) = \frac{\sum_{i=1}^N \{ (\theta + 1)(t_{i1} - t_{i2})^2 + 2t_{i1}t_{i2} \} (N^2 + 1)}{\sigma^2(2\theta + 1)(N^2 + N)}.$$

Hence the variance-covariance matrix of $\hat{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$ is

$$I^{-1} = \begin{bmatrix} \frac{1}{a} & 0 & 0 \\ 0 & \frac{c}{bc-d^2} & \frac{-d}{bc-d^2} \\ 0 & \frac{-d}{bc-d^2} & \frac{b}{bc-d^2} \end{bmatrix}. \quad (3.15)$$

$$\text{Hence } \text{Var}(\hat{\theta}) = \frac{1}{a}, \text{Var}(\hat{\beta}_0) = \frac{c}{bc-d^2}, \text{Var}(\hat{\beta}_1) = \frac{b}{bc-d^2},$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-d}{bc-d^2} \text{ and } \text{Cov}(\hat{\theta}, \hat{\beta}_0) = 0 = \text{Cov}(\hat{\theta}, \hat{\beta}_1).$$

The half-life of the population for the model (3.1) is given by

$$t_{1/2} = \frac{\ln 2}{\lambda} = - \frac{\ln 2}{\beta_1}.$$

We are interested in the $\text{Var}(t_{1/2})$ and hence in the $\text{Var}(\hat{\beta}_1)$, which is given by

$$\text{Var}(\hat{\beta}_1) = \frac{b}{bc-d^2}$$

$$= \frac{\frac{2N(N^2 + 1)}{\sigma^2(2\theta + 1)(N^2 + N)}}{\left(\frac{N^2 + 1}{\sigma^2(2\theta + 1)(N^2 + N)} \right)^2 \left\{ 2N \left(\sum_{i=1}^N (\theta + 1)(t_{i1} - t_{i2})^2 + 2t_{i1}t_{i2} \right) - \left(\sum_{i=1}^N (t_{i1} + t_{i2}) \right)^2 \right\}}$$

$$= \frac{2N(N^2 + N)\sigma^2(2\theta + 1)}{(N^2 + 1) \left\{ N \left\{ \sum_{i=1}^N (\theta + 1)(t_{i1} - t_{i2})^2 + 2t_{i1}t_{i2} \right\} - \left(\sum_{i=1}^N t_{i1} + t_{i2} \right)^2 \right\}} \quad (3.16)$$

4. VARIANCE-COVARIANCE MATRIX OF THE ESTIMATORS OF THE PARAMETERS OF THE GENERAL MODEL

In this section, we shall obtain the variance-covariance matrix of the estimators of the parameters in the mixed effects model for the repeated measures. Here we will consider N subjects, $(p + 1)$ covariates, and J measurements on each subject whereas in section 3, $p = 1$, $J = 2$. Let the j th measurement on the i th individual be given by

$$Y_{ij} = \beta_0 + \alpha_i + \sum_{\ell=1}^p \beta_{\ell} x_{ij\ell} + \epsilon_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, J$$

Then

$$Y_i = X_i \beta + \alpha_i Z_i + \epsilon_i, \quad i = 1, 2, \dots, N, \quad (4.1)$$

where X_i and Z_i are $J \times (p + 1)$ and $J \times 1$ design matrices, β is a $(p + 1) \times 1$ vector of fixed population parameters to be estimated, α_i and ϵ_i are random such that

$$\epsilon_i \sim N\left(0_{J \times 1}, \sigma^2 I_J\right)$$

$$\alpha_i \sim N\left(0, \sigma_{\alpha}^2 = \theta \sigma^2\right).$$

It follows that $Y_i | \alpha_i \sim N(X_i \beta + Z_i \alpha_i, \sigma^2 I_J)$ and

$$Y_i \sim N(X_i \beta, \Sigma_i),$$

where $\Sigma_i = \sigma^2 V_i(\theta) = I_J + \theta Z_i Z_i^T = I_J + \theta 1_J$ and Y_i is independent of Y_j , $i \neq j$. As in section 3, we can represent the full model as

$$Y = X\beta + Z\alpha + \epsilon, \quad (4.2)$$

where $Y' = (Y_1', Y_2', \dots, Y_N')$, $X' = [X_1', X_2', \dots, X_N']$
 $Z = \text{diag}(Z_1, Z_2, \dots, Z_N)$, $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_N)$, $\epsilon' = (\epsilon_1', \epsilon_2', \dots, \epsilon_N')$
 and

$$Y|\alpha \sim N(X\beta + Z\alpha, \sigma^2 I_{jN})$$

and

$$Y \sim N(X\beta, \Sigma),$$

where $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_N) = \sigma^2 V$, $V = I_{jN} + \theta ZZ' = I_{jN} + \theta 1_{jN}' 1_{jN}$
 is a matrix of ones.

Since $Y_i \sim N(X_i \beta, \sigma^2 V_i(\theta))$, $i = 1, 2, \dots, N$ and independent,
 the non-constant part of the full log-likelihood is given by

$$\ell_F = -\frac{NJ}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^N \log |V_i(\theta)| - \frac{1}{2\sigma^2} \sum_{i=1}^N r_i^T V_i^{-1}(\theta) r_i, \quad (4.3)$$

where $r_i = y_i - X_i \beta$. In order to save space $V_i(\theta)$ is written as V_i .
 Since the ML estimates are biased downward, our estimates of θ , σ , and β are based on the restricted log-likelihood given below

$$\begin{aligned} \ell_R(\beta, \theta, \sigma | y) = & -\frac{[NJ - (p+1)]}{2} \log \sigma^2 - \frac{1}{2} \log \left| \sum_{i=1}^N X_i^T V_i^{-1} X_i \right| \\ & - \frac{1}{2\sigma^2} \sum_{i=1}^N r_i^T V_i^{-1} r_i - \frac{1}{2} \sum_{i=1}^N \log |V_i|. \end{aligned} \quad (4.4)$$

The RML estimate of σ^2 is given by

$$\hat{\sigma}_{RML}^2 = \frac{\sum_{i=1}^N r_i^T V_i^{-1} r_i}{NJ - (p+1)} \quad (4.5)$$

Substituting (4.5) in equation (4.4), we get the restricted profile log-likelihood of β and θ as follows:

$$p_R(\beta, \theta | y) = -\frac{NJ - (p+1)}{2} \log \sum_{i=1}^N r_i^T V_i^{-1} r_i - \frac{1}{2} \log \left| \sum_{i=1}^N X_i^T V_i^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^N \log |V_i|. \quad (4.6)$$

The derivatives of the restricted profile log-likelihood with respect to β and θ , after simplification, are given as

$$\frac{\partial p_R(\beta, \theta | y)}{\partial \beta} = [NJ - (p+1)] \frac{H_2(V^{-1})}{H_1(V^{-1})} \quad (4.7)$$

$$\begin{aligned} \frac{\partial p_R(\beta, \theta | y)}{\partial \theta} &= \left[\frac{NJ - (p+1)}{2(J\theta + 1)^2} \right] \frac{H_1(\perp_J)}{H_1(V^{-1})} + \frac{1}{2(J\theta + 1)^2} \text{trace} \left((H(V^{-1}))^{-1} H(\perp_J) \right) \\ &- \frac{NJ}{2(J\theta + 1)}. \end{aligned} \quad (4.8)$$

$$\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta} = [NJ - (p+1)] \left[\frac{H(V^{-1})}{H_1(V^{-1})} + 2 \frac{H_2(V^{-1}) (H_2(V^{-1}))^2}{(H_1(V^{-1}))^2} \right]. \quad (4.9)$$

$$\begin{aligned} \frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2} &= \frac{NJ - (p+1)}{2(J\theta + 1)^3} \left[-\frac{2JH_1(\perp_J)}{H_1(V^{-1})} + \frac{1}{(J\theta + 1)} \left[\frac{H_1(\perp_J)}{H_1(V^{-1})} \right]^2 \right] \\ &+ \frac{1}{2(J\theta + 1)^4} \text{trace} \left[(H(V^{-1}))^{-1} H(\perp_J) \times (H(V^{-1}))^{-1} H(I_J) \right] \\ &- \frac{J}{(J\theta + 1)^3} \text{trace} \left[(H(V^{-1}))^{-1} H(\perp_J) \right] + \frac{N(J)}{2(J\theta + 1)^2}. \end{aligned} \quad (4.10)$$

$$\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta \partial \beta} = \frac{NJ - (p+1)}{(J\theta + 1)^2} \left[-\frac{H_2(\perp_J)}{H_1(V^{-1})} + \frac{H_2(V^{-1}) H_1(\perp_J)}{(H_1(V^{-1}))^2} \right], \quad (4.11)$$

where $V_1(\theta) = V$ has $1+\theta$ in the diagonal and θ off diagonal, $|V_1| = J\theta + 1$,
 $V_1^{-1} = [v_{i,jk}]$, $v_{i,jk} = \frac{-\theta}{1+J\theta}$, $j \neq k$, $v_{i,jj} = 1 - \frac{\theta}{1+J\theta}$, $A = V_1^{-1} \frac{\partial V_1}{\partial \theta} V_1^{-1} = \frac{1}{(J\theta+1)^2}$,

$V_1^{-1} \frac{\partial V_1}{\partial \theta} = \frac{1}{J\theta+1}$ and $H(S) = \sum_{i=1}^N X_i^T S X_i$, $H_1(S) = \sum_{i=1}^N r_i^T S r_i$, and $H_2(S) = \sum_{i=1}^N X_i^T S r_i$.

Since $\frac{H_1(V^{-1})}{\sigma^2} \sim \chi_{JN}^2$ and $H_2(V^{-1}) \sim N(0_{J \times 1}, \sigma^2 H(V^{-1}))$,

$$\begin{aligned} E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta}\right) &= [NJ - (p+1)] \left[-H(V^{-1}) E\left(\frac{1}{H_1(V^{-1})}\right) + 2E\left(\frac{H_2(V^{-1})H_2(V^{-1})^T}{[H_1(V^{-1})]^2}\right) \right] \\ &= [NJ - (p+1)] \left[\frac{-H(V^{-1})}{(NJ-2)\sigma^2} + 2 \frac{\sigma^2 H(V^{-1})}{E(H_1(V^{-1}))^2} \right] \\ &= [NJ - (p+1)] \left[\frac{-H(V^{-1})}{(NJ-2)\sigma^2} + \frac{2\sigma^2 H(V^{-1})}{\sigma^4 [2(JN) + (JN)^2]} \right] \\ &= - \frac{[NJ - (p+1)] [(JN)^2 + 4] H(V^{-1})}{[(JN)^2 - 4JN] \sigma^2}. \end{aligned} \quad (4.12)$$

Since

$$\frac{H_1(1_J)}{J(J\theta+1)\sigma^2} \sim \chi_N^2, \quad \frac{H_1(V^{-1})}{\sigma^2} \sim \chi_{JN}^2, \quad \frac{1}{J(J\theta+1)} \frac{H_1(1_J)}{H_1(V^{-1})} \sim \text{Beta}(N/2, (J-1)N/2)$$

(see appendix II).

$$\begin{aligned} E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2}\right) &= \frac{NJ - (p+1)}{2(J\theta+1)^3} \left[-2JE\left(\frac{H_1(1_J)}{H_1(V^{-1})}\right) + \frac{1}{J\theta+1} E\left(\frac{H_1(1_J)}{H_1(V^{-1})}\right)^2 \right] \\ &\quad + \frac{1}{2(J\theta+1)^4} \text{trace}\left((H(V^{-1}))^{-1} H(1_J) \times (H(V^{-1}))^{-1} H(1_J)\right) \\ &\quad - \frac{J}{(J\theta+1)^3} \text{trace}\left((H(V^{-1}))^{-1} H(1_J)\right) + \frac{N}{2} \left(\frac{J}{J\theta+1}\right)^2 \\ &= \frac{[NJ - (p+1)]J}{2(J\theta+1)^2} \left[-2 + \frac{N+2}{NJ+2} \right] + \frac{N}{2} \left(\frac{J}{J\theta+1}\right)^2 + k \end{aligned}$$

$$= \frac{J}{2(J\theta + 1)^2} \left[p + 1 - \frac{[NJ - (p+1)][NJ - NJ]}{NJ + 2} \right] + k,$$

where $k = \frac{1}{2(J\theta + 1)^4} \text{trace} \left([H(V^{-1})]^{-1} H(\perp_J) \times [H(V^{-1})]^{-1} H(\perp_J) \right)$

$$- \frac{J}{(J\theta + 1)^3} \text{trace} \left((H(V^{-1}))^{-1} H(\perp_J) \right). \quad (4.13)$$

$$E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta \partial \beta} \right) = \frac{[NJ - (p+1)]}{(J\theta + 1)^2} \left[- \frac{E(H_2(\perp_J))}{E(H_1(V^{-1}))} + \frac{E(H_2(V^{-1})H_1(\perp_J))}{E((H_1(V^{-1}))^2)} \right]$$

$$= \frac{[NJ - (p+1)]}{(J\theta + 1)^2} [0_{J \times 1} + 0_{J \times 1}]$$

$$= 0_{J \times 1},$$

since $E(H_2(\perp_J)) = 0_{2 \times 1}$ and $E(H_2(V^{-1})H_1(\perp_J)) = 0_{J \times 1}$ (see appendix II).

$$(4.14)$$

Therefore the information matrix I , of $\hat{\theta}$, $\hat{\beta}$ is of the form

$$I = - \left[\begin{array}{c|c} E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2} \right) & 0_{1 \times (p+1)} \\ \hline 0_{(p+1) \times 1} & E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta} \right) \end{array} \right], \text{ where } E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \theta^2} \right) \text{ and}$$

$E \left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial \beta^T \partial \beta} \right)$ are given by (4.13) and (4.12). Variance-covariance

matrix of $\hat{\theta}$ and $\hat{\beta}$ is given by the inverse of the information matrix I given above.

5. CONCLUSIONS AND RECOMMENDATIONS

The aim of this project was to compare the relative efficiency of the estimators of the half-life given by the full model and the mixed-effects linear model. In sections 2 and 3 we have derived the analytic expressions for the variances of the estimators given by two models. In addition, the analytical expression for variance-covariance matrix for the parameters of the general mixed-effects linear model has been obtained. These expressions will enable us to select an appropriate model for the estimation of half-life.

Due to the shortage of time, the comparisons could not be done. These comparisons will be taken in a future study if funding is available.

ACKNOWLEDGEMENTS

I would like to thank the Air Force Systems Command and the Air Force Office of Scientific Research for sponsoring this research. I would also like to thank the RDL for their efficiency in administrative aspect of the program.

My special thanks are due to Dr. Joel Michalek for providing me some insight into this problem. I cannot thank Mr. Tom White enough for his help in computational work.

REFERENCES

- Elandt-Johnson, R. C. and Johnson, N. L. (1980). Survival Models and Data Analysis. John Wiley and Sons.
- Gupta, P. L. (1989). Dioxin Half-life Estimation in Veterans of Project Ranch Hand. UES Report.

_____ (1990). An investigation of Dioxin Half-life Estimation in Veterans of Project Ranch Hand. UES Mini Grant Report.

Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. Journal of the American Statistical Association, 83, 1014-1022.

Michalek, J. E., Mihalko, D., White, R., and Patterson, Jr., D. G. (1989). Maximum Likelihood Estimation of Half-Life Based on Two Measurements per Subject (unpublished manuscript).

Michalek, J. E., Tripathi, R. C., Caudill, S. P. and Pirkle, J. L. (1992). Investigation of TCDD Half-Life Heterogeneity in Veterans of Operation Ranch Hand. Journal of Toxicology and Environmental Health, 35, 29-38.

Phillips, D. L. (1989). Propagation of Error and Bias in Half-Life Estimates Based on Two Measurements. Archives of Environmental Toxicology, 18, 508-514.

Pirkle, J. L., Wolfe, W. H., Patterson, D. C. Needham, L. L., Michalek, J. E., Miner, J. C., Peterson, M. R., and Phillips, D. G. (1989). Estimates of the Half-Life of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin in Vietnam Veterans of Operation Ranch Hand. Journal of Toxicology and Environmental Health, 27, 165-171.

Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology. Brooks/Cole.

Tripathi, R. C. (1989). An Investigation of Dioxin Half-Life Heterogeneity in Humans Based on Two Measurements per Subject. UES Report.

_____ (1990). In Investigation of Dioxin Half-Life Estimation in Humans Based on Two or More Measurements per Subject. UES Report.

Wolfe, W. H., Michalek, J. E., Miner, J. C., Peterson, M. R., Pirkle, J. L., Patterson Jr., D. G., and Needham, L. L. (1988). Serum 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin levels in Air Force Health Study Participants-Preliminary Report. Morbidity and Mortality Weekly Report, 37, 309-311.

SAS Institute, Inc. (1990). SAS/STAT Users Guide, Version 6, 4th ed., Vol. 2, Cary, N. C.: SAS Institute, Inc.

**SURVIVAL ANALYSES OF RADIATED ANIMALS
INCORPORATING COMPETING RISKS AND COVARIATES**

Ramesh C. Gupta
Professor
Department of Mathematics
Room 320

University of Maine
5752 Neville Hall
Orono, ME 04469-5752

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

SURVIVAL ANALYSES OF RADIATED ANIMALS
INCORPORATING COMPETING RISKS AND COVARIATES

Ramesh C. Gupta
Professor
Department of Mathematics
University of Maine

Abstract

The effects of radiation, taking into account the cause of death (cancer or heart disease) along with the covariates such as sex, age, type of exposure, dose, are examined. A general log linear hazard model approach is studied. The model estimates the cause specific hazard rates, assuming piecewise exponential distribution and exhibits the survival function for each of the covariate groups and the probability of death due to each cause. A data set called "Delayed Bio-Effects Colony", of radiated animals, is analyzed and some conclusions are drawn.

SURVIVAL ANALYSES OF RADIATED ANIMALS INCORPORATING COMPETING RISKS AND COVARIATES

Ramesh C. Gupta

I. Introduction

The United States Air Force has been interested in studying the effect of different types of radiation encountered by its personnel in space. A number of balloon and unmanned satellite flights had provided some idea of the types of radiation which would be encountered in space. These were (1) electromagnetic radiation, (2) electrons, (3) protons, and (4) nuclei of elements of higher Z numbers. For physical details, see Dalrymple et. al. (1991).

In order to study the radiation hazards of space, a decision was made to use the rhesus monkeys as the primary biological subjects. The USAF School of Aerospace at Brooks Air Force Base undertook this study as they had appropriate facilities for holding the animals together with very excellent veterinary service. The animals in this study were, therefore, rhesus monkeys of both sexes caught in the wild that were given single, high-dose rate, total body surface exposures of monoenergetic protons as part of a series of experiments to determine the relative biological effectiveness of protons in the manifestation of the acute radiation syndrome. The estimated age of the subjects ranged from 18-36 months at the time of irradiation, based on body fat and dentition.

301 monkeys survived a postirradiation observation period of 100 days and were considered suitable candidates for extended observation. These animals together with 57 age-matched nonirradiated monkeys, were retained for longitudinal survey of delayed effects. For further details of the data set, see Wood (1991).

A preliminary study on the above data set has been described in Wood (1991). The purpose of the present project is to study, statistically in detail the effect of radiation taking into

account the cause of death (cancer or heart-disease) along with covariates such as sex, age, type of exposure, dose, etc.

The problem, therefore, lies in the frame work of competing risk data with covariates and right censored observations. Covariate methods have been examined quite extensively in the context of survival (i.e. single risk) studies, see for example Fiegel and Zelen (1965), Zippin and Armitage (1966), Glasser (1967), Cox (1972), Prentice (1973), Peto and Lee (1973) and Breslow (1974). Lagakos (1978) have extended these methods to an exponential competing risks model which incorporates covariate variables and right censored observations.

In this project we follow the general log linear hazard model approach described in Larson (1984). The model estimates the cause specific hazard rates assuming piece wise exponential distribution and incorporates the different covariates.

Section 2 contains the notations and definitions while section 3 describes the log-linear hazard model and the inference procedures. The data set, mentioned in the introduction, is described in section 4 and analyzed using the machinery in section 3. Finally, some discussions and conclusions are presented in section 5.

2. Notations and Definitions

Let us consider a population in which J causes of death are operating and each cause of death has its own failure time, for example a person can have heart-disease, kidney disease, lung disease etc. and all these causes are competing for the person's life. The methods of analyzing survival data in such a population are called competing risk analyses. We also have a fundamental assumption that each death can take place due to specific cause.

Suppose the J causes of death are time dependent and that the data consists of the pair (X, T) where T is the time of death and X is the cause of death taking the values $1, 2, 3, \dots, J$. We shall use $X = 0$ to denote that the data is right censored. In addition to (X, T) , suppose one records a p dimensional vector $Z =$

(Z_1, Z_2, \dots, Z_p) of covariates.

The j^{th} cause specific hazard function $\lambda(j, t; z)$ is the instantaneous rate from cause j in the presence of other risks and is defined as

$$\lambda(j, t; z) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t, X = j | T > t)}{\Delta t} \quad (2.1)$$

The survival function, the probability of surviving a specified time, is given by

$$Q(a; z) = \exp \left\{ - \int_0^a \lambda(\cdot, t; z) dt \right\} \quad (2.2)$$

$$\text{where } \lambda(\cdot, t; z) = \sum_{j=1}^J \lambda(j, t; z) \quad (2.3)$$

is the combined failure rate from all causes. It may be noted that (2.3) may be satisfied even without the assumption of independence between the competing risks.

We shall divide the time interval into K disjoint intervals with end points $0 = w_0 < w_1 < \dots < w_K$. The cause specific hazard function is approximated by a step function in these intervals and is given by

$$\lambda(j, t; z) = \lambda(j, k; z), \quad w_{k-1} < t \leq w_k \quad (2.4)$$

The survival function (2.2) can be written as

$$\begin{aligned} Q(a; z) &= \exp \left\{ - \left[\sum_{r=1}^{k-1} \int_{w_{r-1}}^{w_r} \lambda(\cdot, t; z) dt + \int_{w_{k-1}}^a \lambda(\cdot, t; z) dt \right] \right\} \\ &= \exp \{ - \lambda(\cdot, k; z) (a - w_{k-1}) \} \prod_{r=1}^{k-1} \exp \{ - \lambda(\cdot, r; z) (w_r - w_{r-1}) \} \end{aligned} \quad (2.5)$$

for $w_{k-1} < a \leq w_k$.

The conditional probability of failure due to cause j given that the failure occurs in the k^{th} interval is given by

$$P_r(j; z | k) = \frac{\lambda(j, k; z)}{\lambda(\cdot, k; z)} \quad (2.6)$$

Thus the conditional probabilities due to different causes follow

a multinomial distribution. The unconditional probability of failure due to cause j is, therefore, given by

$$\begin{aligned} \Pr(j; z) &= \sum_{k=1}^k \Pr(j; z | k) P(\text{failure occurs in the } k^{\text{th}} \text{ interval}) \\ &= \sum_{k=1}^k \{ \lambda(j, k; z) / \lambda(\cdot, k; z) \} \{ Q(w_{k-1}; z) - Q(w_k; z) \} \end{aligned} \quad (2.7)$$

3. Log Linear Hazard Model

3.1 Description of the Model

For the sake of convenience, we shall confine our attention to only one discrete covariate with L categories. The extension to more than one covariate is straightforward.

The data can be classified into $J \times K \times L$ table where J denotes the number of competing risks, K denotes the number of time intervals and L denotes the number of covariate categories. Thus there are $J K L$ failure rates to estimate which are modeled as follows:

$$\begin{aligned} \ln \lambda(j, k; \ell) &= \alpha + \alpha_x(j) + \alpha_T(k) + \alpha_z(\ell) + \alpha_{xT}(j, k) \\ &\quad + \alpha_{xz}(j, \ell) + \alpha_{TZ}(k, \ell) + \alpha_{xTZ}(j, k, \ell) \end{aligned} \quad (3.1)$$

for $j = 1, 2, \dots, J$; $k = 1, 2, \dots, k$; $\ell = 1, 2, \dots, L$. Here X denotes the failure type, T denotes the time interval and Z denotes the covariate, $j = 1, 2, \dots, J$; $k = 1, 2, \dots, k$ and $\ell = 1, 2, \dots, L$. The subscripted terms must be zero-sum across each factor. Thus the constraints are

$$\begin{aligned} 0 &= \sum_{j=1}^J \alpha_x(j) = \sum_{k=1}^K \alpha_T(k) = \sum_{\ell=1}^L \alpha_z(\ell) \\ 0 &= \sum_{j=1}^J \alpha_{xT}(j, k) = \sum_{k=1}^K \alpha_{xT}(j, k) \\ 0 &= \sum_{j=1}^J \alpha_{xz}(j, \ell) = \sum_{\ell=1}^L \alpha_{xz}(j, \ell) \\ 0 &= \sum_{k=1}^k \alpha_{TZ}(k, \ell) = \sum_{\ell=1}^L \alpha_{TZ}(k, \ell) \\ 0 &= \sum_{j=1}^J \alpha_{xTZ}(j, k, \ell) = \sum_{k=1}^K \alpha_{xTZ}(j, k, \ell) = \sum_{\ell=1}^L \alpha_{xTZ}(j, k, \ell) \end{aligned} \quad (3.2)$$

Thus α_x has $J - 1$ independent parameters, α_{xT} has $(J - 1)(K - 1)$ independent parameters and so on. Looking at this from analysis

of variance point of view, here the grand mean is α , the main effects are α_X , α_T and α_Z ; the two way interactions are α_{XT} , α_{XZ} and α_{TZ} ; and the three way interaction is α_{XTZ} . The main effects can be written as

$$\begin{aligned}\alpha_X(j) &= \frac{1}{KL} \sum_{k=1}^K \sum_{\ell=1}^L \ln \lambda(j, k; \ell) - \alpha, \quad j = 1, 2, \dots, J \\ \alpha_T(k) &= \frac{1}{JL} \sum_{j=1}^J \sum_{\ell=1}^L \ln \lambda(j, k; \ell) - \alpha, \quad k = 1, 2, \dots, K. \\ \alpha_Z(\ell) &= \frac{1}{JK} \sum_{k=1}^K \sum_{j=1}^J \ln \lambda(j, k; \ell) - \alpha, \quad \ell = 1, 2, \dots, L.\end{aligned}\quad (3.3)$$

Similarly the two way and three way interaction terms can be written.

3.2 The Likelihood Function

Consider data on N subjects. The observe data on the n^{th} subject consists of the triplet (X_n, T_n, Z_n) . Its contribution to the likelihood function (LF) can be obtained as follows:

Suppose the observation terminates in the k^{th} interval and the covariate value is ℓ , so that $Z_n = \ell$. Let

$b_{r,n}$ = subject's exposure time in the r^{th} interval

= $\{\text{Min}(T_n, W_r) - W_{r-1}\} \delta(T_n > W_{r-1})$ where $\delta(\cdot)$ denotes the indicator function.

It is clear that $b_{r,n} = 0$ if $T_n \leq W_{r-1}$. Using (2.5), the contribution of this individual to the LF can be written as

$$\prod_{j=1}^J \{\lambda(j, k; \ell)\}^{\delta(x_n=j)} \prod_{r=1}^k \exp\{-\lambda(\cdot, r; \ell) b_{rn}\} \quad (3.4)$$

The overall LF from N subjects is, therefore,

$$\prod_{n=1}^N \left[\prod_{j=1}^J \{\lambda(j, k; \ell)\}^{\delta(x_n=j)} \prod_{r=1}^k \exp\{-\lambda(\cdot, r; \ell) b_{rn}\} \right] \quad (3.5)$$

$$= \left[\prod_{n=1}^N \prod_{j=1}^J \{\lambda(j, k; \ell)\}^{\delta(x_n=j)} \right] \left[\prod_{n=1}^N \prod_{r=1}^k \exp\{-\lambda(\cdot, r; \ell) b_{rn}\} \right]$$

The log likelihood is, therefore

$$\begin{aligned}
\theta &= \sum_{n=1}^N \ln \left(\prod_{j=1}^J \{ \lambda(j, k; \ell) \}^{\delta(x_n=j)} \right) + \sum_{n=1}^N \ln \left(\prod_{r=1}^k \exp \{ -\lambda(\cdot, r; \ell) b_{rn} \} \right) \\
&= \sum_{j=1}^J \sum_{n=1}^N \delta(x_n=j) \ln \lambda(j, k; \ell) - \sum_{n=1}^N \sum_{r=1}^k \lambda(\cdot, r; \ell) b_{rn} \\
&= \sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L [\ln \lambda(j, k; \ell) D_{jkt} - \lambda(j, k; \ell) U_{kt}] \quad (3.6)
\end{aligned}$$

where $\lambda(j, k; \ell)$ is given by (3.1) and

$$D_{jkt} = \sum_{n=1}^N \delta(x_n=j) \delta(W_{k-1} < T_n \leq W_k) \delta(Z_n = \ell)$$

= # of those subjects who die with cause j in the interval k and covariate value ℓ , and

U_{kt} = total exposure time of all subjects with covariate ℓ during the interval k . D_{jkt} and U_{kt} form the sufficient statistics for our model, see Bishop, Fienberg and Holland (1975).

Now consider independent random variables D_{jkt} having Poisson distribution with mean $\lambda(j, k; \ell) U_{kt}$. The LF for this sample is

$$\prod_{j,k,t} [e^{-\lambda(j,k;\ell)U_{kt}} \{ \lambda(j,k;\ell)U_{kt} \}^{D_{jkt}} / D_{jkt}!]$$

The log LF is, therefore

$$\sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L [D_{jkt} \ln \{ \lambda(j, k; \ell) U_{kt} \} - \lambda(j, k; \ell) U_{kt} - D_{jkt}!] \quad (3.7)$$

Thus the two likelihood functions (3.6) and (3.7) have the same kernel and, therefore, the techniques employed to analyze Poisson count data can be used to analyze competing risk data using log linear models.

3.3 Fitting of the model and testing for significance.

The maximum likelihood estimates of the parameters of our model are obtained by setting the partial derivatives equal to 0. We shall use the statistical package BMDP, 4F, for fitting an appropriate model. The program gives the values of the estimators and their standard errors so that necessary hypotheses can be tested. The expected values \hat{D}_{jkt} corresponding to the

cell counts D_{jkt} , are also obtained by using the program. Having obtained \hat{D}_{jkt} , the estimates of $\lambda(j,k;\ell)$ are given by

$$\hat{\lambda}(j,k;\ell) = \hat{D}_{j,k,\ell} / U_{k\ell} \quad (3.8)$$

The goodness of fit for each model can be tested by using the likelihood ratio statistic

$$G^2 = 2 \sum_{j=1}^J \sum_{k=1}^K \sum_{\ell=1}^L D_{jkt} \ln (D_{jkt} / \hat{D}_{jkt}).$$

G^2 has, asymptotically, a chi-square distribution with ν degrees of freedom where ν is the difference between the number of non zero estimates and the number of independent parameters of the model. Testing for specific parameters is based on the difference in G^2 statistic for the two nested models, one of which does not contain the specific parameter to be tested. The difference in G^2 statistics will have a chi-square distribution with degrees of freedom equal to the degrees of freedom between the two models.

4. Analyses of the 'Delayed Bi-Effects Colony' Data.

The data consists of 358 subjects (rhesus monkeys), 207 males and 151 females. 301 of these subjects were exposed to different kinds of radiations such as 2.0 MeV X ray, 32.0 MeV Proton, 55.0 MeV Proton, 138 MeV Proton, 400 MeV Proton, 2.3 GeV Proton, 1.6 MeV Electron, 2.0 MeV Electron and Solar Flare while 57 subjects formed the control group. The exposure dosage ranged from 0 (control) to 1500 rads. These subjects were followed over a period of 338 months. The competing causes of death are (1) infections (2) organ degeneration (3) tumors (tumors of the bone skin, hematopoietic system, tumors of nerve tissue and tumors of viscera), (4) endometriosis and others. 289 subjects died of different causes and 69 were still alive (censored). At the time of death, time of death and cause of death are recorded. For the censored observations, only the time of their leaving the study is observed.

Since endometriosis occurs only in females, three different

analyses were performed, one for males only, one for females only and one for males and females combined.

4.1 Analyses for Male subjects

In order to perform the log-linear hazard model approach on the 207 male subjects, we considered three competing risks (1) tumors (2) infection and (3) others, and exposure dosage as the only covariate partitioned into three categories. (1) No radiation (controls), (2) between 1 and 500 rads (3) more than 500 rads.

The time was divided into four intervals with $W_0 = 0$, $W_1 = 100$, $W_2 = 200$, $W_3 = 300$ and $W_4 = 400$. Thus $X = 1$ (tumor), $X = 2$ (infection) and $X = 3$ (other causes); $Z = 0$ (no exposure), $Z = 1$ (between 1 and 500 rads) and $Z = 2$ (> 500 rads). The observed number of death D_{jk} , are given in Table 1.

<u>INTERVALS</u>						
<u>Z</u>	<u>X</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>TOTAL</u>
0	1	0	0	2	0	2
	2	1	2	0	0	3
	3	2	3	9	2	16
1	1	4	5	8	1	18
	2	5	9	6	0	20
	3	2	6	23	8	39
2	1	9	8	2	0	19
	2	7	6	3	0	16
	3	7	10	5	1	23

TABLE 1

The exposure times $U_{k\ell}$ are

<u>EXPOSURE TIMES</u>				
<u>INTERVALS</u>				
<u>Z</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
0	3279	2781	1890	163
1	10534	8007	6169	589
2	5108	2812	1024	93

TABLE 2

The best fitted model by the program is XZ, XT, TZ, i.e.

$$\ln \lambda(j,k;\ell) = \alpha + \alpha_X(j) + \alpha_T k) + \alpha_Z(\ell) + \alpha_{XT}(j,k) + \alpha_{XZ}(j,\ell) +$$

$\alpha_{Tz}(k, \ell)$. This model gave $G^2 = 9.08$ with 10 degrees of freedom.

The expected values $\hat{D}_{j,k}$ using the above model are given in Table

3.

INTERVALS

Z	X	1	2	3	4	TOTAL
0	1	0.5	0.6	0.9	0.0	2.0
	2	0.9	1.2	0.9	0.0	3.0
	3	1.6	3.2	9.2	2.0	16.0
1	1	3.5	5.1	8.6	0.9	18.0
	2	4.8	8.4	6.9	0.0	20.0
	3	2.7	6.6	21.6	8.1	39.0
2	1	9.0	7.4	2.6	0.1	19.0
	2	7.4	7.4	1.2	0.0	16.0
	3	6.7	9.2	6.2	0.9	23.0

TABLE 3

The estimates, $\hat{\lambda}(j, k; \ell)$, of the failure rate are then obtained by

$\hat{\lambda}(j, k; \ell) = \hat{D}(j, k, \ell) / U_{k\ell}$ and are given in Table 4.

Dose	Cause	Interval 1	Interval 2	Interval 3	Interval 4
Controls	Tumors	1.524855138761818D-04	2.157497303128371D-04	4.761904761904762D-04	0.00000000000000E+00
Controls	Infection	2.744739249771272D-04	4.314994606256742D-04	4.761904761904762D-04	0.00000000000000E+00
Controls	Other	4.879536444037816D-04	1.150665228335131D-03	4.867724867724867D-03	1.226993865030675D-02
<=500	Tumors	3.322574520599962D-04	6.369426751592356D-04	1.39406710974226D-03	1.528013582342954D-03
<=500	Infection	4.556673628251377D-04	1.049082053203447D-03	1.118495704328092D-03	0.00000000000000E+00
<=500	Other	2.563128915891399D-04	8.242787560884226D-04	3.501377857027071D-03	1.375212224108659D-02
>500	Tumors	1.761942051683634D-03	2.631578947368421D-03	2.53906250000000E-03	1.075268817204301D-03
>500	Infection	1.448707909162099D-03	2.631578947368421D-03	1.17187500000000E-03	0.00000000000000E+00
>500	Other	1.311667971808927D-03	3.271692745376956D-03	6.05468750000000E-03	9.67741935483871D-03

TABLE 4

The overall failure rates (without taking into consideration the competing risks are given in Table 5.

Dose	Interval 1	Interval 2	Interval 3	Interval 4
Controls	9.149131E-04	1.797915E-03	5.820106E-03	1.226994E-02
<=500	1.044238E-03	2.510303E-03	6.013941E-03	1.528014E-02
>500	4.522318E-03	8.534851E-03	9.765625E-03	1.075269E-02

TABLE 5

We observe that for the control group as well as for the low dose group, the failure rates are increasing except for the last interval because in some of the cases no animal died in that

interval. For the high dose group, both for tumors and infections the failure rate increases from the first interval to the second and then it decreases in the third and fourth interval. This suggests that for the high dose group, the effect of radiation is initially very high and then it starts diluting with time.

Using Table 4 and 5, equation (2.5) was employed to calculate the survival function. The survival function suggests that there is not a significant difference between the control group and the low dose group but the survival probabilities are significantly low in the high dose group. The difference is more alarming in the 100-300 range, see Figure 1.

The probabilities of death due to different causes, were calculated by using equation (2.7). The results are given in Table 6.

	Tumors	Infection	Other
Controls	0.06011492	0.08979256	0.7251982
<=500	0.1857949	0.1857565	0.5451115
>500	0.3049938	0.2494232	0.4107624

TABLE 6

Note that these probabilities do not add up to exactly 1. More accurate results can be obtained by taking larger number of intervals.

4.2 Analyses for Female subjects

For analyzing the data on 151 female subjects, we considered four competing risks (1) tumors (2) endometrosis (3) infection and (4) others, and exposure dosage as the only covariate partitioned as before. The time was also divided into four

intervals as before. Thus in this case $X = 1$ (tumor), $X = 2$ (endometrosis), $X = 3$ (infection) and $X = 4$ (other causes); $Z = 0$ (no exposure) $Z = 1$ (between 1 and 500 rads) and $Z = 2$ (> 500 rads). The observed number of deaths $D_{j,k,l}$ are given in Table 7.

		<u>Intervals</u>				
<u>Z</u>	<u>X</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>Total</u>
0	1	0	0	3	1	4
	2	1	1	0	0	2
	3	1	0	1	0	2
	4	2	1	4	1	8
1	1	0	0	6	0	6
	2	4	24	1	0	29
	3	4	3	6	0	13
	4	6	5	7	9	27
2	1	1	2	1	0	4
	2	4	6	1	0	11
	3	4	3	1	0	8
	4	4	6	9	0	19

TABLE 7

The exposure times U_k are

		<u>Exposure Times</u>			
		<u>Intervals</u>			
<u>Z</u>		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
0		2168	1773	939	14
1		7805	5260	2418	257
2		4173	2193	856	0

TABLE 8

The best fitted model by the program is XZ, XT, TZ as before. This model gave $G^2 = 17.94$ with 13 degrees of freedom. The expected values $\hat{D}(j,k;l)$ using the model are given in Table 9.

Intervals

Z	X	1	2	3	4	Total
0	1	0.3	0.2	3.2	0.3	4.0
	2	0.9	0.9	0.2	0.0	2.0
	3	0.9	0.2	0.9	0.0	2.0
	4	1.9	0.7	3.7	1.7	8.0
1	1	0.3	0.9	4.1	0.7	6.0
	2	5.4	22.1	1.5	0.0	29.0
	3	4.3	3.6	5.1	0.0	13.0
	4	4.0	5.4	9.4	8.3	27.0
2	1	0.4	0.9	2.7	0.0	4.0
	2	2.7	7.9	0.3	0.0	11.0
	3	3.7	2.2	2.1	0.0	8.0
	4	6.1	6.0	6.9	0.0	19.0

TABLE 9

The estimates $\hat{\lambda}(j,k;l)$ of the failure rates are then obtained by using $\hat{\lambda}(j,k;l) = \hat{D}(j,k;l)/U_{kj}$ and are given in Table 10.

Dose	Cause	Interval 1	Interval 2	Interval 3	Interval 4
Controls	Tumors	1.383763837638376D-04	1.128031584884377D-04	3.407880724174654D-03	2.142857142857143D-02
Controls	Endo	4.151291512915129D-04	5.076142131979696D-04	2.129925452609159D-04	0.000000000000000E+00
Controls	Infection	4.151291512915129D-04	1.128031584884377D-04	9.584664536741214D-04	0.000000000000000E+00
Controls	Other	8.763837638376383D-04	3.948110547095318D-04	3.940362087326944D-03	1.214285714285710E-01
<=500	Tumors	3.843689942344651D-05	1.711026615969582D-04	1.695616211745244D-03	2.723735408560311D-03
<=500	Endo	6.918641896220372D-04	4.201520912547529D-03	6.203473945409429D-04	0.000000000000000E+00
<=500	Infection	5.509288917360666D-04	6.844106463878327D-04	2.109181141439206D-03	0.000000000000000E+00
<=500	Other	5.124919923126201D-04	1.026615969581749D-03	3.887510339123242D-03	3.229571984435798D-02
>500	Tumors	9.58543014617781D-05	4.103967168262654D-04	3.154205607476636D-03	0.000000000000000E+00
>500	Endo	6.470165348670022D-04	3.602371181030552D-03	3.504672897196261D-04	0.000000000000000E+00
>500	Infection	8.866522885214474D-04	1.003191974464204D-03	2.453271028037383D-03	0.000000000000000E+00
>500	Other	1.461778097292116D-03	2.735978112175103D-03	8.060747663551403D-03	0.000000000000000E+00

TABLE 10

The overall failure rates (without taking into considerations the competing risks) are given in Table 11.

Dose	Interval 1	Interval 2	Interval 3	Interval 4
Controls	1.845018E-03	1.128032E-03	8.519702E-03	1.428571E-01
<=500	1.793722E-03	6.083651E-03	8.312655E-03	3.501945E-02
>500	3.091301E-03	7.751937E-03	1.401869E-02	0.000000E+00

TABLE 11

We observe that for the control group as well as for the high dose group, the tumor rates are initially high. They go

down from the first to the second interval, but then they go up in the third interval. However, for the low dose group, the tumor rates always increase. In the case of endometriosis, for all groups, the rates first increase from the first to the second interval and then they go down after 200 months. For infections, the exposed groups failure rates always increase whereas for the control group, they first decrease and then go up in the third interval. Notice that zero failure rates indicate that no deaths occurred during those intervals. The overall failure rates go up for the exposed groups; while for the controls the rates go down in the second interval and then they start going up.

Using Tables 10 and 11, equation (2.5) was employed to calculate the survival function. The survival function suggests that initially there is no difference between the controls and the low dose group but the difference is significant between 100 and 300 months time period. The difference between the controls and the high dose group is little in the beginning but then it becomes high between 100 and 300 months time period. The difference between low dose and high dose seems to stay uniform up to 300 months. After 300 months, the situation is very interesting, see Figure 1.

The probabilities of death due to different causes were calculated using equation (2.7). The results are given in Table 12.

	Tumors	Endo	Infection	Other
Controls	0.2394166	0.08847239	0.09469794	0.5774127
<=500	0.08155257	0.3453729	0.1584405	0.4084637
>500	0.08656117	0.2460318	0.1721196	0.4120613

TABLE 12

4.3 Analyses for All Subjects (Combined)

In order to analyze data on all 358 subjects (males and females combined), we considered only two competing risks (1) tumors, and (2) others, and exposure dosage as the only covariate as before. The time was also classified into four intervals as before. Thus in this case $X = 1$ (tumor) and $X = 2$ (other causes); $Z = 0, 1, 2$ as before. The observed number of deaths D_{jk} are given in Table 13

Z	X	<u>Intervals</u>				Total
		1	2	3	4	
0	1	0	0	5	1	6
	2	7	7	14	3	31
1	1	4	5	14	1	24
	2	21	47	43	17	128
2	1	10	10	3	0	23
	2	26	31	19	1	77

TABLE 13

The exposure time $U_{k\ell}$ are

<u>EXPOSURE TIMES</u>				
<u>INTERVALS</u>				
<u>Z</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
0	5447	4554	2829	177
1	18339	13267	8587	846
2	9281	5005	1880	93

TABLE 14

The best fitted model by the program is XT, TZ and $G^2 = 13.46$ with 8 degrees of freedom. The expected values $\hat{D}(j,k;\ell)$ using the model are given in Table 15.

INTERVALS

Z	X	1	2	3	4	Total
0	1	1.4	1.0	4.3	0.3	7.1
	2	5.6	5.9	14.7	3.7	29.9
1	1	5.1	7.8	12.8	1.6	27.3
	2	19.9	44.2	44.2	16.4	124.7
2	1	7.4	6.2	4.9	0.1	18.6
	2	28.6	34.8	17.1	0.9	81.4

TABLE 15

The estimates $\hat{\lambda}(j, k; \ell)$ of the failure rates are then obtained by using $\hat{\lambda}(j, k; \ell) = \hat{D}(j, k; \ell) / U_{jk}$, and are given in Table 16.

Dose	Cause	Interval 1	Interval 2	Interval 3	Interval 4
Controls	Tumors	2.570222140627870-04	2.195871761089150-04	1.519971721456350-03	1.694915254237290-03
Controls	Other	1.028088856251150-03	1.29556433904260-03	5.196182396606580-03	2.090395480275990-02
<=500	Tumors	2.780958612792410-04	5.879249265093840-04	1.490625363922210-03	1.891252955082740-03
<=500	Other	1.085119144991550-03	3.331574583553180-03	5.147315709793880-03	1.938534278959810-02
>500	Tumors	7.973278741514920-04	1.238761238761240-03	2.60638297972340-03	1.07526881720430-03
>500	Other	3.08156448658550-03	6.953046953046950-03	9.395744680851060-03	9.677419354838710-03

TABLE 16

The overall failure rates (without taking into considerations the competing risks are given in Table 17.

Dose	Interval 1	Interval 2	Interval 3	Interval 4
Controls	0.001285	0.001515	0.006716	0.022599
<=500	0.001363	0.003919	0.006638	0.021277
>500	0.003879	0.008192	0.011702	0.010753

TABLE 17

We observe that for the control group, the tumor rates decrease from the first to the second interval and then they go down while for low dose group, the rates always go up. For the high dose group, the tumor rates go up to 300 months and then decrease perhaps because there are very few animals left to die in the last interval. The overall failure rates go up for the control as well as for the low dose group while for the high dose group, the rates go up to 300 months and then decrease as in the case of tumor.

Using Tables 16 and 17 equation (2.5) was employed to

calculate the survival function. The survival function indicates that initially there is insignificant difference between the controls and the low dose group but the difference is significant between 100 and 300 months. The difference between the controls and the high dose group is markedly high and tends to decrease after a long period of time. The difference between the low dose and high dose is also very high, see Figure 1.

The probabilities of death due to different causes were calculated using equation (2.7). The results are given in Table 18.

	Tumors	Other
Controls	0.151633	0.808072
<=500	0.156441	0.807396
>500	0.175519	0.792816

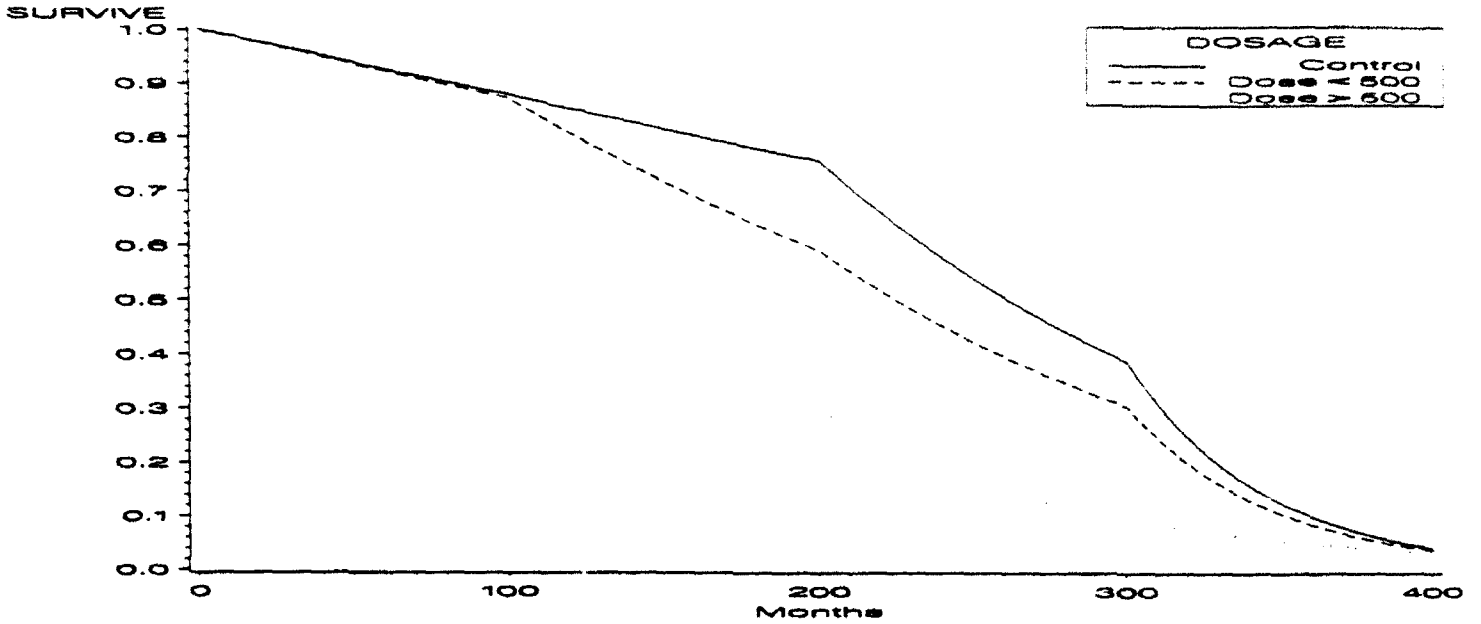
TABLE 18

5. Conclusions and further analyses

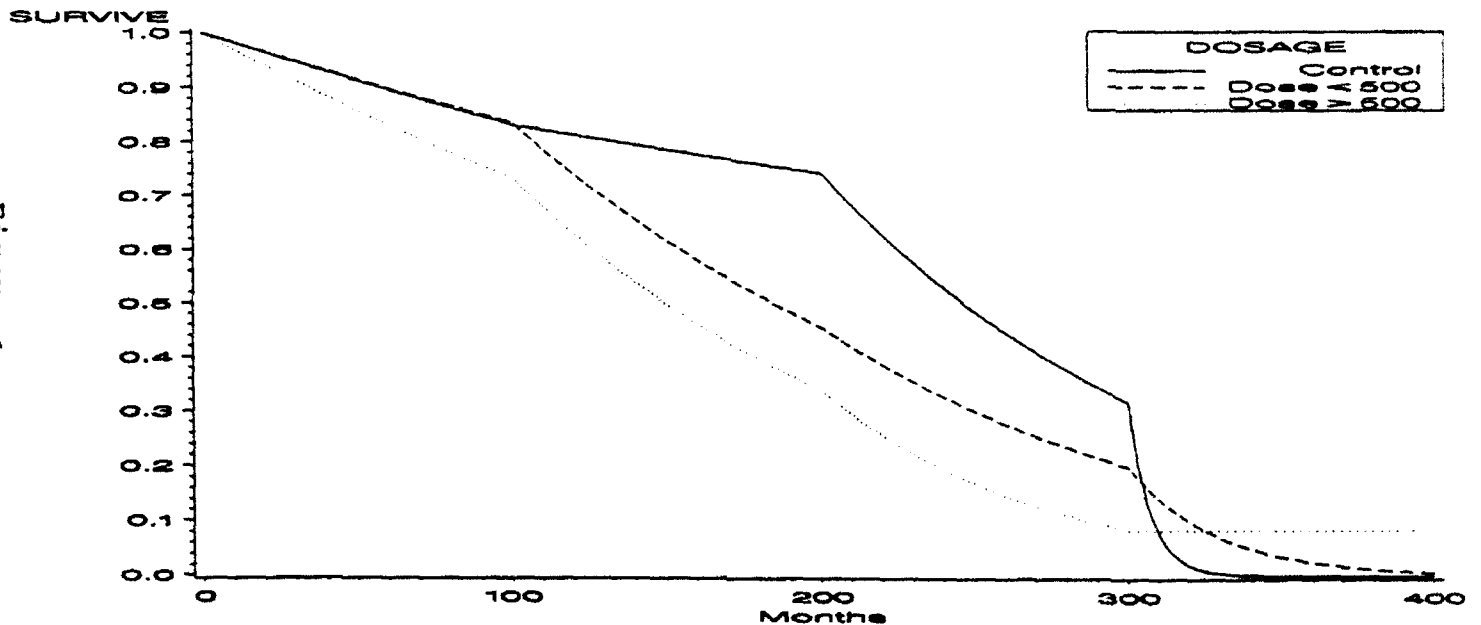
The three analyses indicate that the probability of developing a tumor is .06 in the case of males and .24 in the case of females so that overall probability (males and females combined) is .15. The probability of infection is almost the same, about 9% in both the sexes. The survival curves suggest that the males do not react to low dose of radiation as much as the females. The high doses have profound effect in both the groups.

Overall the study has brought out the effect of high and low doses of radiation on both the male and female groups. The probabilities of developing tumor and other diseases obtained by the model will be useful in determining the effect of radiation (high and low doses) in the two groups (males and females).

All Subjects



Females



Males

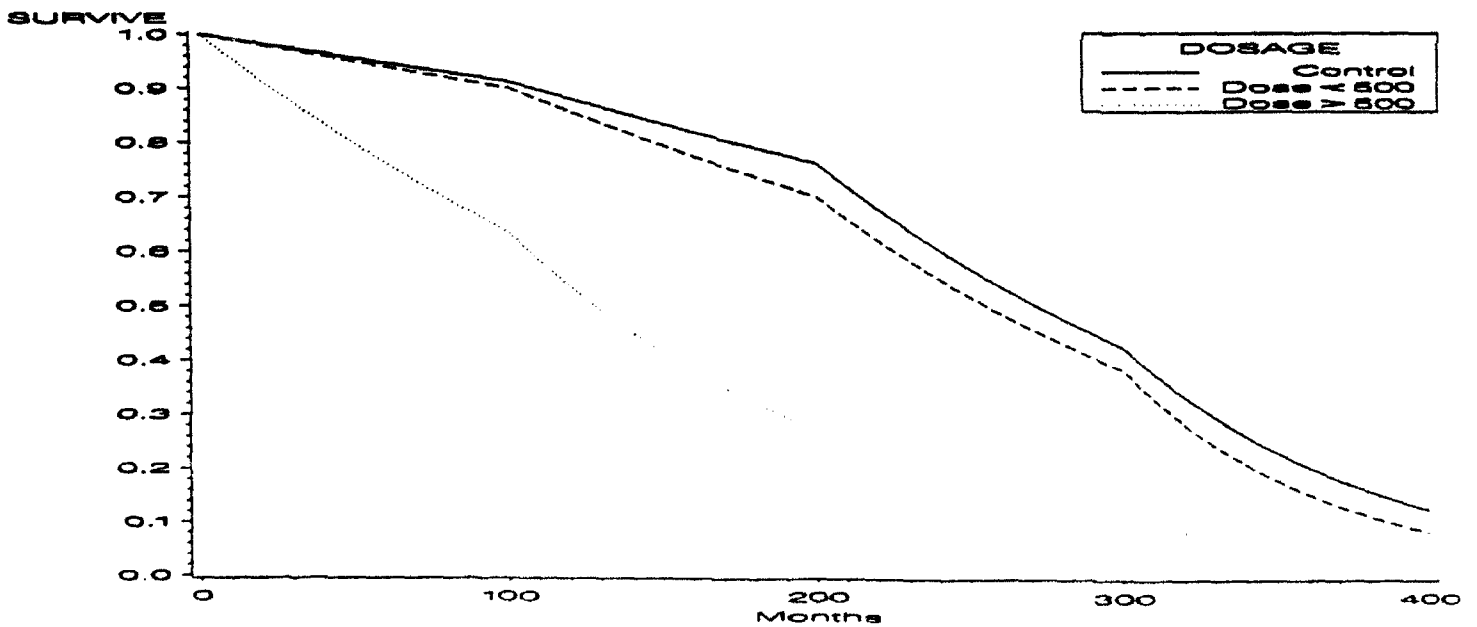


Figure 1

Acknowledgement

I am thankful to Dr. Richard Albanese for supporting this project and to the Radiation Science Division of the Armstrong Laboratory for providing the necessary facilities.

REFERENCES

- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Massachusetts.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89-100.
- Cox, D.R. (1972). Regression models and life tables. *Jour. Royal Statist. Soc. B*, 34, 187-220.
- Dalrymple, G.V., Lindsay, I.R., Mitchell, J. and Hardy, K.A. (1991). A review of the USAF/NASA proton bieffects project: rationale and acute effects. *Radiation Research* 126, 117-119.
- Fiegel, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* 21, 826-838.
- Glasser, M. (1967). Exponential survival with covariance. *Journ. Amer. Statist. Association* 62, 561-568.
- Lagakos, S.W. (1978). A covariate model for partially censored data subject to competing causes of failure. *Appl. Statist.* 27(3), 235-241.
- Larson, M.G. (1984). Covariate analysis of competing risks data with log-linear models. *Biometrics* 40, 459-469.
- Peto, R. and Lee, P. (1973). Weibull distributions for continuous-carcinogenesis experiments. *Biometrics* 29, 457-470.
- Prentice, R.L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, 60, 279-288.
- Wood, D.H. (1991). Long-term mortality and cancer risk in irradiated rhesus monkeys. *Radiation Research* 126, 132-140.
- Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information of an exponential survival parameter. *Biometrics* 22, 665-672.

THERMAL STRESS
IN SEVEN TYPES OF CHEMICAL DEFENSE ENSEMBLES
DURING MODERATE EXERCISE
IN HOT ENVIRONMENTS

Richard A. Hengst
Associate Professor
Department of Biological Sciences

Purdue University North Central
1401 South U.S. 421
Westville, IN 46391

Final Report for :
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research

September 1992

THERMAL STRESS
IN SEVEN TYPES OF CHEMICAL DEFENSE ENSEMBLES
DURING MODERATE EXERCISE
IN HOT ENVIRONMENTS

Richard A. Hengst
Associate Professor
Department of Biological Sciences
Purdue University North Central

Abstract

Data collected from experiments conducted in 1991 were organized into a technical report format. This report covers exposure tolerance and physical effects of thermal stress on subjects performing mild exercise in desert conditions when wearing seven different chemical defense ensembles. Subjects walked at 3 mi h^{-1} in ambient conditions of 40°C and 20% RH until rectal temperature increased 1.5°C above starting values. Ensembles tested included the BDO with BDU, BDO without BDU, UK with BDU, Gore-Tex with PJ-7, MLFS, CWU-77P, and PJ-7. The performances of these ensembles were compared with BDU results. Three garments, the PJ-7, MLFS, and CWU-77P performed within 94% of the trial exposure of the BDU and all three performed significantly better than other garments tested. CDEs with longer trial lengths had lower mean skin temperatures, lower heart rates, lower heat storage rates and slower rates of rectal temperature increase. The significantly longer times and lower thermal stress characteristic of better performing ensembles correlate well with higher percentage sweat evaporation values with consequent improvement in heat dissipation.

**THERMAL STRESS
IN SEVEN TYPES OF CHEMICAL DEFENSE ENSEMBLES
DURING MODERATE EXERCISE
IN HOT ENVIRONMENTS**

INTRODUCTION

The US Air Force has a continuing interest in the thermal properties of chemical defense ensembles (CDEs). This is because heat stress proves to be the overriding physiological problem for individuals wearing CDEs, particularly in temperate or hot climates (Nunnely, 1983). Even moderate exercise accelerates body temperature increases which can cause deterioration in task performance and possibly physical damage or even death. Preparation for chemical defense scenarios dictates the need for testing thermal properties of CDEs in a wide variety of environments. This testing includes present concerns for understanding garment performance and possible effects in desert conditions.

It is important to examine thermal tolerance and physiological response in new CDE designs in desert conditions. This testing answers concerns for personnel safety and performance under field conditions in protective uniforms. The CDE presently worn by most US Air Force personnel functions well in the cooler environments for which it was designed, but alternatives must be examined and verified for hot environments. For instance, flight crew have unique thermal and garment design needs. Similarly, improvements in CDE heat dissipating properties are of extreme importance to ground crew working in hot environments.

Solutions proposed for the thermal stress problem include improvement in fabric and garment designs, alternation of rest and work periods, and various methods of cooling suits. Artificial cooling methods are useful in some instances but are not feasible for personnel whose tasks place them at any distance from cooling equipment. Alternation of rest with work is a method commonly employed at the present to slow heat accumulation and thermal stresses on the body. This method works better in some environments than others but is generally enhanced by CDE designs that effectively dissipate heat. The most promising solution appears to be improvement in fabric and garment design.

The purposes of this study are to consolidate and analyze data from seven CDE studies for thermal stress effects and to prepare a follow-up report discussing this analysis.

MATERIALS AND METHODS

Subject Pool

Eleven USAF volunteer subjects were used in this study. Subjects were informed of potential risks and signed consent forms previously approved by the local human use committee before participating in any experiments. This group is considered representative of USAF personnel now serving on active duty status. All subjects participated in all trials except the BDU trial which was completed by ten of the subjects. Average subject characteristics were as follows: Age = 34 yr \pm 5.6; Weight = 79.6 kg \pm 8.5; Height = 178.0 cm \pm 3.6.

Experimental Design

These experiments were designed to determine how long normal individuals wearing different CDE designs could continue moderate exercise in a desert environment. A second goal was to document physiological accommodations to the thermal burden imposed by CDEs during this exercise. Environmental, exercise, and non-garment factors were held constant so that differences between trials were attributable to garment differences. It was not necessary to exercise subjects until collapse or dangerously core temperatures occurred; it was sufficient to establish the pattern of temperature increase over a lower, safer, range and extrapolate responses to higher, more critical, body temperatures. Thus, the 1.5°C T_{re} increase was chosen for these experiments.

The environmental conditions selected for these tests are representative of average daily midday temperature and humidity in Dhahran, Saudi Arabia, as determined by the U.S. Army during Operation Desert Shield. Ambient temperature was held at 40°C (104°F) and relative humidity was controlled at 20%. Overhead radiant heat was used to simulate solar effects on body heat storage. Experimental conditions did not vary significantly from a dry bulb temperature (T_{db}) of 40°C (104°F), a wet bulb temperature (T_{wb}) of 27°C (80.6°F), and a black globe temperature (T_{bg}) of 45°C (113°F).

Constant work loads were achieved continuous treadmill walking at 3 miles/hour up a 5% grade. For purposes of USAF application, the selected activity level corresponds to moderate exercise performed by active flight line ground crews during combat turns. Metabolic heat produced by this effort was approximately 450 kcal/h and is comparable to the ground-crew activity levels targeted. Activity was terminated when (T_{re}) rose 1.5°C (2.7°F) above starting values. Trial length of the subject pool was considered to be the median time to reach target temperature.

The major variable in these thermal experiments was the chemical defense suit worn during the trial. However, certain aspects of clothing were held constant in all trials. These constants included a butyl rubber hood, the MCU/2P face mask and respirator, rubber gloves with cotton liners, and athletic shoes. Standard Mission Oriented Protective Posture (MOPP) IV clothing involves this garb with the obvious exception of athletic shoes. Athletic shoes rather than standard overboots were worn as a concession to subject safety and subject comfort when walking on

the treadmill.

Eight suits were tested in these trials. Each garment is briefly described in turn.

BDU: A summer weight BDU was included for purposes of comparing thermal effects of a normal duty uniform worn under desert conditions to those of the various CDEs. This ensemble has no chemical defense technology applied in design or manufacture.

BDO+BDU: A two-piece, trouser and jacket, standard issue CDE using charcoal-foam sandwiched within cloth layers. The outer fabric layer is treated to repel liquids. This CDE is now used by the USAF and is normally worn over the BDU in cooler climates.

BDOnoBDU: The standard issue CDE just mentioned worn without the BDU in hot climates.

UK+BDU: A one-piece coverall worn under flight suits or BDU. Charcoal adsorbent is latex bonded to liquid repellent viscose nylon and requires long underwear when worn.

Gore-Tex+PJ-7: A two-piece, trouser and jacket, liquid and aerosol repellent shell of Gore-Tex™ worn with a PJ-7 as an undergarment and vapor adsorbent layer using Bluecher carbon sphere technology.

PJ-7: A one-piece, vapor adsorbent garment without liquid repellent cloth treatment. As mentioned earlier, this garment uses Bluecher carbon spheres as the vapor adsorbent material. It was designed to be used as an undergarment but could be used alone in hot climates.

MLFS: A two-piece, trouser and jacket, over garment that may be worn alone in hot climates. Jacket design has an integral hood as part of the design. Vapor adsorbent layer uses Bluecher carbon spheres.

CWU-77P: A one-piece coverall worn over underwear with the outer fabric treated to repel liquid agents. Vapor adsorbent layer uses Bluecher carbon sphere technology.

Physiological Measurements

Electrocardiogram (ECG) and body temperatures were monitored continuously throughout the experiments. The ECGs were continuously displayed and monitored via Transkinetics ECG transmitters to ensure subject safety. These signals were processed to determine heart rate (HR) on a beat-to-beat basis. Temperatures were measured from thermistors placed at specific body and clothing sites including rectal (T_{re}) (10 cm insertion depth), forearm (T_{fa}), chest (T_{ch}), thigh (T_{th}), and calf (T_{ca}). Clothing temperatures were recorded from sites immediately external to the placement of skin thermistors. Clothing thermistors measured forearm clothing (T_{clfa}), chest clothing (T_{clch}), thigh clothing (T_{clth}) and Calf clothing (T_{clca}) temperatures. From skin or clothing temperatures, it was possible to calculate mean temperature of skin (T_{msk}) and clothing (T_{mcl}). Mean skin temperature was calculated by weighted averages of leg, torso, and arm temperatures. Total body heat storage was calculated from body weight, rectal and mean skin

temperatures. Temperatures, HR, and calculations were recorded every 30 seconds by a computer data acquisition system.

It was important that data be gathered on sweat production, sweat rate, evaporation, and percent evaporation since sweat evaporation is known to be an important cooling factor in CDE clothing. These data were determined from the differences between clothed weight and between nude weight of subjects before and after each experiment.

Subjective Measures

Subjective evaluation of work effort and of thermal comfort were taken intermittently to gauge subject reactions to garments in a quantitative fashion. Initial readings were taken 5 min after walking began and every five minutes for the duration of the experiment. Ratings of exertion and thermal comfort were manually entered into the computer data acquisition system.

Statistical Analysis of Data

Data were analyzed from samples taken every 5 min of a trial. Values used in analysis were median values from the period occurring 1 min before, at, and after each 5 min interval. Analysis of parameter change over time was by three-way repeated measures analysis of variance (ANOVA) (Suit x Subject x Time). Separate analyses of conditions at selected times (beginning, 30 min, and final) were conducted by two-way ANOVA (Suit x Subject) by variable. These three points were chosen to compare the initial physical condition of subjects, status after thermal stresses in various suits had time to develop effects on subjects, and condition at the end of experiments. The last analysis was to determine if differences present at 30 min persisted to the target temperature. Variables analyzed included trial length, HR, T_{re} , T_{msk} , heat storage, thermal comfort (TC) ratings, and rated perceived exertion (RPE) scores, sweat production, sweat evaporation, percent evaporation and sweat rate. Differences were considered significant if $P < 0.05$.

RESULTS

Our primary goals of this investigation were: (1) to determine the length of time that subjects wearing CDEs could be expected to perform moderate work in a desert environment; and (2) to document biological accommodations of personnel to work in these conditions. A trial's length was determined by how long it took an individual's rectal temperature T_{re} to rise 1.5°C (2.7 °F) above the starting value. Physiological responses to this temperature change were used to establish predictable patterns that characterized an ensemble's thermal stress. The selected temperature increase allowed for a realistic sampling of responses over a thermally stressful range without unduly forcing subjects to medically critical temperature extremes. Given constant work loads and conditions, physiological measures are expected to change at the rate established over the experimental core temperature range. Therefore, responses to core temperatures moderately above

those of the experimental range are predictable from data obtained over a lower, and safer, temperature range.

Ensemble Performance

Mean times for each ensemble's performance are shown in Table 1. It is clear that the data are clustered into a long tolerance group and a short tolerance group. A two way ANOVA (Suit by Time) confirmed that significant differences ($P < 0.001$) existed between data means. Duncan's multiple range test established the two groupings differed significantly from each other but garments within each group were not significantly different. An especially great similarity in performance occurred within the long tolerance grouping in which the three CDE's differed by a total of 36 seconds. Garments in the shorter performance grouping varied more than the long but not enough to be significantly close to any of the better performing CDE's.

LENGTH OF ENSEMBLE PERFORMANCE TRIALS

TABLE 1

Ensemble	Mean Time (min)	Percent of BDU Time
Long Tolerance Times		
BDU	55.9	100
PJ-7	53.4	96
MLFS	53.2	95
CWU-77P	52.8	94
Short Tolerance Time		
UK+BDU	38.9	70
BDO no BDU	37.1	66
G-Tex + PJ-7	36.2	65
BDO + BDU	32.2	58

Sweat Production and Sweat Evaporation

Cooling through sweat evaporation is an important mechanism for resisting thermal stress in a hot environment. Table 2 shows the results of sweat produced and evaporated in each garment trial. The strong relationship between sweat evaporated and trial length may be seen in Figure 1. Figure 1 is a plot of trial duration as a function of percent evaporation. A positive correlation exists between percent sweat evaporated and tolerance times whereas that between sweat rate and tolerance time is a negative one (see Figure 2).

Table 2
Sweat Produced and Evaporated

SUIT	Time (min)	Sweat Total (g)	Percent Evap.	Evaporation Total (g)	Sweat Rate (g min ⁻¹)
BDO+BDU	32.2	1423.6	31.4	433.2	31.0
G-Tex + PJ-7	36.2	1480.0	32.2	461.4	29.2
UK+BDU	38.9	1588.2	37.1	546.4	30.0
BDO no BDU	37.1	1476.5	36.5	551.0	28.9
MLFS	53.2	1736.5	45.7	786.1	26.1
CWU-77P	52.8	1900.9	47.0	867.3	28.6
PJ-7	53.4	1876.8	48.4	892.3	27.6
BDU	55.9	1801.0	53.2	945.0	25.6

Physiological State At the Beginning of Experiments

This study was conducted over an 8-month period. It was important to verify that subjects were physiologically similar at the start of all trials to facilitate direct comparisons of responses. Data comparing responses are shown in Table 3. Analysis showed heart rate in the Gore-Tex+PJ-7 (98.0 Beat min⁻¹) was the only variable identified as being significantly different from the rest (109 beats min⁻¹). Within 5 min this difference in HR was no longer present (HR = 126 beats min⁻¹; mean at 5 minutes = 125 beats min⁻¹). All subjects were seen as physiologically similar with the exception of this small variation.

Physiological State After Thirty Minutes

A detailed three-way ANOVA (Suit x Subject x Variable) was performed and variables compared on data sampled from the 30-min interval of experiments. The 30-min point represented the last 5-min interval for which data were available for all ensembles and for which sufficient time had passed to develop differences in response to thermal stress of each suit. Table 4 shows results of mean physiological and psychological measures at 30 min. As expected, physiological responses generally paralleled performance times. The MLFS, CWU-77P, PJ-7 and BDU all had lower mean heart rates, lower T_{re} , lower T_{msk} , lower heat storage, somewhat lower RPE values, and lower TC values than the BDO+BDU, BDO no BDU, Gore-Tex+PJ-7, and UK+BDU. The first group with the lower physiological rates or measures also had longer performance times. This group also had lower subjective ratings for RPE and TC.

Initial Physiological Measurements

TABLE 3

Suit	HR (Beats/min)	T _{re} (°C)	T _{msk} (°C)
BDU	109.2	37.1	36.5
PJ-7	113.7	37.0	36.5
CWU-77P	108.3	37.1	36.5
MLFS	111.5	37.1	36.5
UK+BDU	106.6	37.1	36.6
BDO _{no} BDU	111.5	37.1	36.6
GTEX+PJ-7	98.0*	37.1	36.7
BDO+BDU	120.1	37.2	36.6

* Significant difference $P < 0.05$

Mean Responses After 30 Minutes

TABLE 4

SUIT	HR (Beats/min)	T _{re} (°C)	T _{msk} (°C)	H _{lstr} (kcal/kg)	RPE	TC
BDU only	147.6	37.8	36.4	35.7	12.5	5.6
PJ-7	143.6	37.8	37.0	52.7	12.6	5.5
MLFS	148.3	37.9	36.8	47.1	12.8	5.8
CWU-77P	145.0	37.8	36.8	45.5	13.1	5.9
UK+BDU	156.6	38.2	37.3	81.8	14.0	6.2
BDO no BDU	159.3	38.2	37.5	74.5	14.2	6.3
GTex + PJ-7	168.2	38.2	37.9	80.1	13.7	6.7
BDO+BDU	169.3	38.4	37.9	88.9	13.9	6.2

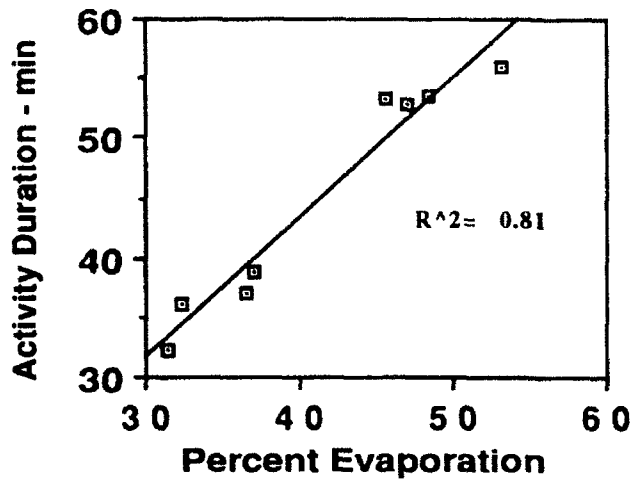


Figure 1

Trial Length as a Function of Percent Evaporation

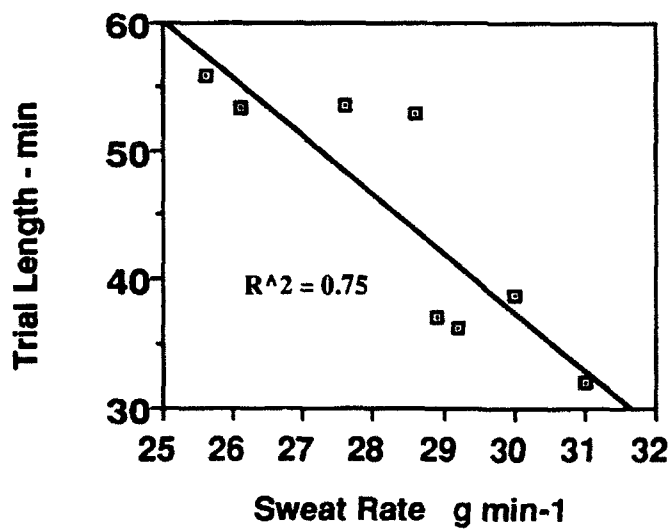


Figure 2

Trial Length as a Function of Sweat Rate

Physiological State At The End Of Experiments

Mean responses of subjects at the end of experiments are shown in Table 5. A three-way ANOVA (Suit x Variable x Time) performed on these results identified significant differences in T_{msk} and heat storage. Application of Duncan's multiple range test to these two parameters showed subjects had significantly lower T_{msk} and heat storage when wearing the CWU-77P, MLFS, PJ-7 or BDU than when wearing the BDO+BDU, BDOnoBDU, UK+BDU, or Gore-Tex+PJ-7. Final T_{msk} varied from 36.4 to 37.2 °C with a group mean of 36.8°C for the long-exposure group, while T_{msk} varied from 37.4 to 38.0 °C with a mean of 37.7 °C for the short-trial group. Heat storage averaged 74.3 kcal kg⁻¹ (range = 69.2 to 76.4 kcal kg⁻¹) for the long exposure group and 86.0 kcal kg⁻¹ (range = 78.9 to 95.3 kcal kg⁻¹) for the short-exposure group.

Responses of Subjects at the End of Experiments

TABLE 6

SUIT	TIME (min)	HEART RATE (b min ⁻¹)	T_{re} (°C)	T_{msk} (°C)	HtStr (KCal/kg)	RPE	TC
BDU	55.9	160.9	38.5	36.7	80.1	14.4	6.1
PJ-7	53.4	164.1	38.4	37.0	86.0	15.3	6.6
MLFS	53.2	160.5	38.5	37.0	83.0	11.3	6.3
CWU-77	52.8	156.8	38.5	37.1	84.4	14.4	6.4
UK+BDU	38.9	162.0	38.5	37.6	91.3	14.9	6.5
BDO no BDU	37.1	164.5	38.5	37.6	88.3	13.9	6.7
GTex + PJ-7	36.2	164.1	38.4	38.1	96.2	11.2	6.5
BDO+BDU	32.2	170.4	38.5	38.0	103.2	13.9	6.2

Changes Over Time

When the data for most variables are viewed over the time course of the experiments, the obvious groupings and the changes to the long- and short-duration groups are easily seen. Figure 3 shows T_{msk} over time. Garments from the short duration group show T_{msk} s that increase rapidly and remain at a higher value than for subjects wearing suits with longer tolerance times. Rectal temperatures increase more slowly in the four suits of the long-exposure group but eventually the same point is reached as is demonstrated in Figure 4.

DISCUSSION

Maintenance of consistent environmental conditions and work load limited experimental variables to measures of physiological stress imposed by various chemical defense garments. Choices of a 40°C ambient temperature, a radiant overhead heat load, and a 20% relative humidity place a premium on evaporation as the prime biophysical mechanism for dissipating

metabolic heat. Steady walking rates were controlled through a treadmill and resulted in constant rates of metabolic heat production. Consistent external and internal thermal stresses in combination with subject pool containing the same individuals in all experiments minimized differences between experimental trials except for suit effects.

Fabric manufacturers and garment designers often strive to minimize thermal stress on the basis of theory and prior experience. Simple theoretical models have identified a number of different clothing factors that interact to affect thermal stress (Parsons, 1988, Whistler 1990). Prominent variables include the area of the body covered by a garment, the insulative properties of that garment, and the evaporation occurring from that garment (Parsons, 1988). However, new designs require both laboratory and field testing since the interactions of fabric, design, fit, and numerous other factors are too complex to be accurately modeled from fabric test data alone. For most of the skin surface, insulation and evaporation properties of garments combine to cause differential rates of heat storage and thermal stresses on the body.

Two of the garments are not always used in the test configurations as chemical warfare ensembles. The BDU was evaluated by itself in these experiments to answer two questions. First, how long can an individual wearing a normal duty uniform be expected to perform moderate work in desert conditions before reaching a criterion body temperature? Secondly, what physiological accommodations are made to desert conditions while performing work in this environment? Both of these questions help to establish an operational standard to which chemical defense ensembles can be realistically compared. In this trial the rest of the ensemble (respirator mask, hood, and gloves with liners) was donned as in the standard MOPP IV posture.* Use of the PJ-7 as a separate trial also warrants discussion. This suit was designed as an undergarment to be used under repellent or impermeable outer shells although it could be worn in a hot environment. In these experiments, the PJ-7 was the vapor adsorbing layer for a Gore-Tex shell. Therefore, it was necessary to test the thermal properties of the PJ-7 alone to distinguish Gore-Tex effects from those of the PJ-7.

In the Results section, it was established that data tended to fall naturally into two groups by exposure tolerance time, each group being internally similar, but differing significantly from one another. An examination of similarities and differences in suit design may serve to explain the data clusters observed.

Suits in the long-exposure group included the BDU, the PJ-7, the MLFS and the CWU-77P. The last three CDEs have distinctly different designs: the PJ-7 is a close fitting one-piece garment; the MLFS is a loose fitting, two-piece trouser and jacket with integral hood; and the CWU-77P is a

* An exception is noted for the MLFS which has an integral cloth hood in its design.

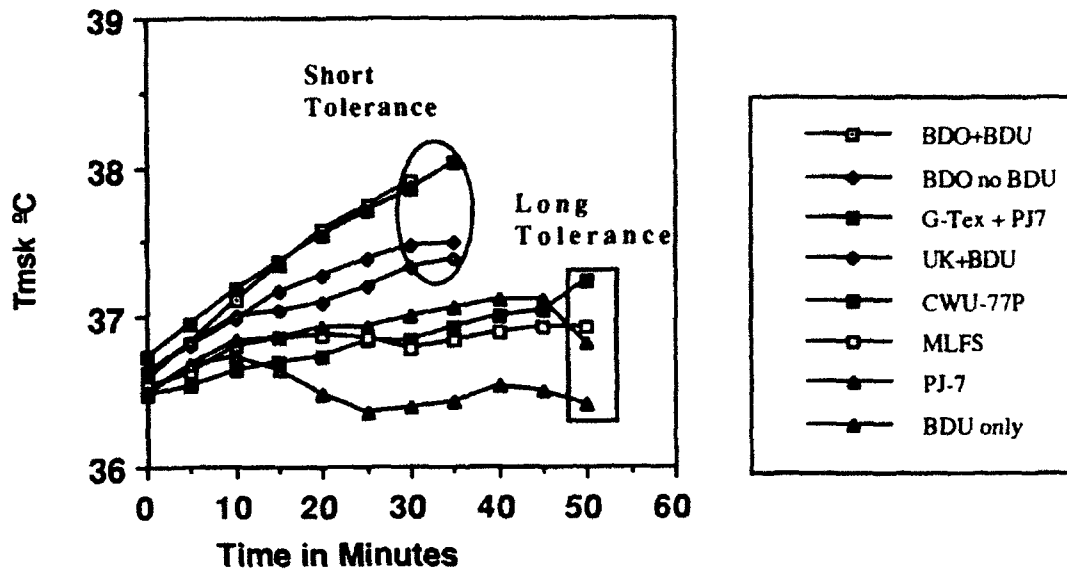


Figure 3

Mean Skin Temperature Over Time

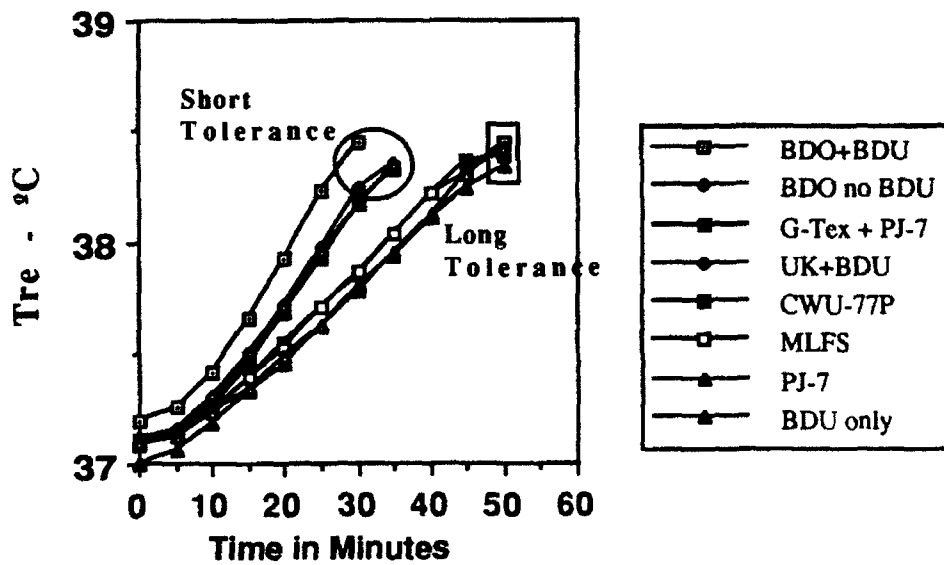


Figure 4

Rectal Temperature during Experiments

one-piece coverall usually worn over underwear. Despite obvious differences in appearance, all three suits have several construction features in common. These features include a fabric with a relatively porous weave, the use of Bluecher process carbon spheres, and a design that allows them to be worn as a single clothing layer. Separate tests have established that all three suits can provide adequate levels of protection as single clothing layer garments. These CDEs proved so surprisingly similar to the BDU's thermal performance that further improvement in heat dissipation for CD garments might only bring about modest payoffs in work times in these environments.

In contrast to the single clothing layer/Bluecher Sphere technology characteristic of the long-exposure group, the short-exposure time suits had designs using both single and multiple clothing layers (three of four garment designs) and poor vapor permeability. Multiple layers have greater resistance to heat transfer, although this may not always be detrimental in hot environments. Since multiple layers have more air gap insulation, work time is expected to decrease. When the BDU is worn with the BDO, trial time decreased by 17.3%. The fact that performance time of a BDO alone was not significantly above the variability of the BDO+BDU, indicates that the BDO itself retains most of the heat of the ensemble.

Short trial suits also had relatively poor vapor transfer characteristics. Short-duration garments consistently had lower percent evaporation making heat transfer more difficult. This evaporation is a property of the cloth and design used in the manufacture of these garments. In contrast, all three ensembles with long trial times use Bluecher carbon spheres laminated between two plies of relatively porous cloth. Bluecher spheres are highly adsorbent for CW vapors but allow sufficient sweat evaporation through the fabric/carbon bead combination to presumably cool the suit environment. Reduced interference with physiological cooling enhances heat dissipation and increases working time.

Data presented earlier support a relationship between evaporation and performance. The good fit ($R = 0.81$) of this data to its regression line supports the contention that effective suit evaporation extends work times.

A relationship also appears to exist between sweat rate and trial length. In general, higher sweat rates were associated with briefer work periods. It is possible that high rates of sweat production are a response to high heat storage rates or to higher mean skin temperatures occurring in these suits. Both of these parameters develop significant differences which continue throughout the entire trial. This relationship of sweat rate to heat storage or T_{msk} indicates that the thermal stress imposed by a suit influences the sweat rate evoked.

A case has been established for a relationship between percent evaporation and thermal tolerance times. Heat storage occurs more slowly in suits with greater evaporative percentages.

Since heat storage is calculated from core and skin temperatures, it is logical to examine the data for relationships between evaporation and body temperatures. It is also logical to examine possible effects on HR, RPE, and TC since each response is related to temperature and/or heat stored.

As ambient temperature increases, the significance of evaporative cooling also increases. It is a basic assumption that increased evaporation through a suit translates to increased cooling of skin surfaces (Parsons, 1988). The data support this assumption in that T_{msk} s are consistently lower in suits with higher evaporation percentages and longer trial times. Figure 4 shows T_{msk} over time for each suit. Skin temperatures are lower in suits with long-exposure times. Effective evaporation should keep T_{msk} s lower and the data are in agreement with this expectation. Thus, suit evaporation effectively cools the skin. In this study, garments constructed from single clothing layers made from porous weave cloth and that used Bluecher carbon spheres as an adsorbent showed consistently cooler T_{msk} s and longer work times.

Rectal temperatures increased at distinctly different rates for long- and for short-exposure suits. Short duration suits had more rapid T_{re} increases than long duration suits. These increases are similar to the rates at which T_{msk} rose in the two groups with T_{msk} increasing more quickly in the short-trial suits. Better cooling properties were likely the source of difference in T_{re} increase since percent evaporation was higher for long-exposure suits. Differences in the rates that T_{re} rose between long- and short-exposure groups can only occur if cooling mechanisms are more effective in one group than the other. Since T_{re} rose at a much slower rate in suits with higher evaporative rates, it is concluded that thermal stress is more effectively compensated in suits with higher evaporative percentages than in those with lower percentages. Not surprisingly, evaporation properties of a suit directly affect core temperature increases. This effect may be extended to differences in heat storage rates and reflects differences in thermal stress imposed by various suits and the degree to which these suits help or hinder physiological compensation mechanisms.

Heart rate was the major physiological measurement other than skin and rectal temperatures in these experiments. Increased body temperatures result in cardiac adjustments beyond those necessary to support physical efforts. Since work load was a controlled experimental factor, changes in HR are related to changes in heat storage associated with thermal stresses. Suits with better evaporative percentages took longer for HR to increase by an equal amount. There is no question that subjects wearing the BDU, PJ-7, MLFS, or CWU-77P took significantly longer times to raise heart rates to the values seen in the short-duration garments. Consequently, the cardiac work load is reduced in suits with lower thermal stress.

It was expected that no differences would be present at the conclusion of experiments since core temperatures increase the same 1.5°C for all subjects and environmental and work stresses were the same for every trial. However, analysis demonstrated that significantly lower T_{msk} s and heat

storage were present in subjects with longer performance time. Lower heat storage should result in more rapid recovery from thermal stress, an important consideration in personnel engaged in continuing field operations. The lower T_{msk} s associated with the better performing suits are significant in that the first step in recovering is normally cooling the skin. The already lower T_{msk} of individuals in Bluecher-type suits should help in this cooling and recovery process. Although thermal stresses cannot be completely eliminated from chemical defense garments, these garments can be designed to maximize compensation through normal thermoregulatory mechanisms.

Subjective measures are important in estimating performance, can give some indication of physiological status, and certainly can indicate factors affecting motivation in thermally stressful circumstances. RPE, a score of how hard the subjects feel they are working during an activity, may correlate with HR. Thermal comfort is commonly used as a relative indicator of T_{msk} . Analysis of these data showed only weak correlations between physiological and psychological data. For these reasons, subjective measures are not good indicators of physiological performance, at least under the environmental conditions used in this study.

Conclusions

It is clear that newer garment designs utilizing Von Blucher carbon sphere technology provide significantly less thermal burden in a hot environment than alternative technologies. Thermal tolerance times are significantly prolonged and physiological strains are significantly reduced for subjects in all of the garments tested that use this protective system. The advantage of this carbon sphere bead system appears to reside in the permeability of suits to sweat evaporation through the fabric thus cooling the wearer. Percent evaporation appears to be a good predictor of physiological response with fairly high correlations between evaporation percentages and HR, heat storage, T_{msk} , and thermal tolerance. Subjective measures were not good predictors of physiological status. It is speculated that evaporative qualities of CDEs may improve physiological recovery in addition to prolonging thermal tolerance to work.

References

- (1) Constable, S. H., Alleviation of thermal stress in ground crew supporting air operations in a chemical warfare scenario. Prepublication article. , ca. 1991.
- (2) Craig, F.N., Moffitt, J.J., Efficiency of evaporative cooling from wet clothing. *J Appl Physiol* 1974; 36: 313-18.
- (3) Fox, R.H., Goldsmith, R., Kidd, D.J., and H.E. Lewis. Acclimatization to Heat in Man by Controlled Elevation of Body Temperature. *J. Physiol.* (1963), 166, 530-547.
- (4) Gonzalez, R.R., Santee, W.R., and Enstruck, T.L., Physiological and biophysical properties of a semipermeable attached hood to a chemical protective garment. In: *Performance of Protective clothing: Fourth Volume*, ASTM STP1133.
- (5) Nunnally, S.A., Design and Evaluation of Clothing for Protection from Heat Stress: An Overview. *Acta Physiol. Scand.* 1983. Suppl 1; 87-98.
- (6) Pandolf, K.B., and Goldman, R.F. Convergence of Skin and Rectal Temperatures as a Criterion for Heat Tolerance. *Avia. Space Environ. Med.* 49:1095-1101, 1978.
- (7) Parsons, K.C. Protective clothing: Heat exchange and physiological objectives. *Ergonomics*, 1988 vol 31, 991-1007.
- (8) Wissler, E. Mathematical Simulation of Human thermal behavior using whole-body models. In: Shitzer, A., ed. *Heat Transfer in Biology and Medicine.*, New York, NY: Plenum, 1985: 325-74.

The Potential of Electronic
Teams for Distance Education

DeLayne Hudspeth
Associate Professor
Area of Instructional Technology
College of Education
The University of Texas
Austin, Texas

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

July, 1992

The Potential of Electronic
Teams for Distance Education

DeLayne Hudspeth
Associate Professor
Area of Instructional Technology
College of Education
The University of Texas
Austin, Texas

Abstract

The potential use of collaborative electronic networks was investigated as a means to involve and motivate distant learners, provide a cooperative base of learning and interaction and provide an environment for learning in a society where the half life of knowledge shortens and the distinction between work and learning blurs. Literature from a number of disciplines was reviewed and suggests considerable potential for electronic support of geographically separate persons who might function as electronically supported learning teams (E/L-teams). It is recommended that costs and benefits of existing services such as telephone bridges and groupware be explored to empirically determine the pro's and con's of electronically supported teams.

The Potential of Electronic
Teams for Distance Education

DeLayne Hudspeth

BACKGROUND

It was my intent when coming to the Armstrong Human Resources Laboratory to explore various issues related to testing and assessing achievement of distant learners. It is my opinion, as documented in my paper presented at the Distant Learning conference, April 21-23, Brook AFB, Texas that we currently test poorly, at low cognitive levels, and that such testing is a disservice to both the student and the field of distance education. I made a good many assertions in this paper and felt that a review of the literature, a research framework and the development of research priorities during the summer of 1992 would provide a useful line of inquiry for a RIP proposal in the fall of 1992.

I also wanted to explore whether selected aspects of the testing process could be automated (perhaps building on basic tenets of AIDA) and to consider whether the constructs found in Artificial Intelligence and Intelligent Tutoring Systems could be used to increase the power and potential of systems of testing. It is my current feeling that there are many useful answers to be gleaned from HRTC activities, such as AIDA. The basis for this need is that few Subject Matter Experts and faculty receive training in systematic assessment and testing.

However, at the HRTC Distance Education Conference the broad issue of electronic network communication was discussed and a number of issues surfaced which broadened my field of inquiry.

Specifically, the issue of electronically supported collaborative learning teams were discussed as a means to involve and motivate distant learners, provide a cooperative base of learning and interaction, and to create an environment of learning for an age wherein the distinction between continuing our education and the world of work is blurred. The kind of environment that I see for the future is referred to as "Just-in-Time Education" found in the June, 1992 issue of Educational Technology.

LEARNING TEAMS

Discussion with personnel at HRTC supported inquiry into electronic collaborative learning networks although a possible conflict with the Army was seen, given their charge to study work teams and groups. Therefore, an acronym that precluded the specific use of the words 'team' or 'group' was considered such as: "Collaborative Electronic Network Technology - CENT"; or, "Collaborative Network Techniques - CNT"; or, "Virtual Work Force - VWF"; etc.

However, a review of literature found the greatest yield for 'team' as the basic term. We decided to be straight forward in current discussion and worry about turf at a later date. It may even be possible that the effort proposed below could function as an inter-service project. Hence, the final term used in this study was 'E-teams' (Electronically supported work teams) for general discussion and the more specific, 'E/L-teams' (Electronically supported learning teams) when discussing groups of students assigned a group learning task.

Rationale for Investigation

Although dyads, triads, small and large group assignments are typically under used instructional strategies, their use in distance education seems to be unknown. Using a variety of machine search sources (ERIC, Wilson, PsychLit, etc.) not a single research report was uncovered on the topic of using remote teams for distance learning. This is curious when considering that various forms of technology support have existed for several decades (such as audio conference calls) and that business, the military and other sectors routinely create and use teams of persons who are geographically separate. (Similar groups have been defined as a task-force, study group, blue-ribbon committee, etc.)

The lack of research on how to best structure learning teams of geographically separate individuals is all the more startling when considering that new communication technology will radically broaden our contact with different people, and that we will increasingly depend on others in different ways. Surely with distant learners the issue of cultural interactions, work and learning styles, and research to suggest how we can capitalize on might learn with and about these differences, is important.

It also seems clear that the declining costs of computing and of computer-based communication is a significant factor for distance education. A much wider range of instructional strategies is now available that were not previously possible because of costs. Hence, data concerned with electronically supported instructional strategies such as grouping techniques could be very useful in the design of some courses.

COMPUTER SUPPORTED INTERACTION

That human interaction is possible with computer based communication was demonstrated more than a decade ago when ARPANET made available an electronic mail (e-mail) function. Once e-mail was available large numbers of computer scientists around the country initiated questions, shared data and began to exchange ideas, rapidly and casually. Soon, a large electronic community was formed of friends who had not seen each other, and collaborators who had not met in person.

During the last decade a quiet revolution has been in progress such that any person with computer, modem and phone line can develop their own private/public communication systems. The generic Bulletin Board System (BBS) can provide a discussion forum, share either public or the most private mail, store files ranging from poetry to music compositions (MIDI files), provide synchronous or asynchronous communication and other forms of communication.

The variety of messages and casualness with which persons communicate electronically will eventually provide a second order effect in terms of expectations and investment in communications which are little studied at the moment. From a social/learning perspective, the most significant role the computer may play is acting as a communication tool within and among data and people. Of these access points it is the communication with people that has the most potential to revolutionize how we learn.

Another example of how rapidly telecommunications is moving is the Texas Education Network (TENET). This network does not

provide the full features of a sophisticated BBS but does provide E-mail, conferences, full File Transfer Protocol (FTP, limited to Kermit) and access to Internet, a communication protocol such that message exchange between networks such as Tenet, Bitnet and other systems is possible. In short, every teacher and eventually every student can communicate with the world for less than \$.02 a day registration fee.

The point of this discussion is not to push BBS or any other communication system but instead to note that the full impact of any relatively new technology is hard to foresee. The unanticipated consequences usually have less to do with efficiency and more with social/organizational effects. The social/organizational effects often emerge slowly as people negotiate new roles and relationships. It is an assumption of this research that as distance education becomes more interactive a number of roles will change especially the roles traditionally defined as instructor and student. Further, the presence or "place" created by electronic communications typically reduces the difference between rank and years in service and relies instead on the intrinsic logic of the message. While crossing organization and rank barriers will be readily accepted by some, it may not be accepted by others. Research which supports the power and utility of E-Teams and E/L-teams-Teams is needed to help determine if increased productivity is worth the price it may extract.

The last reason offered in this brief report as a need to research the design, use and effect of electronically support learning teams (E/L-teams) is that in most sectors of the world

we live in, knowledge work is largely collaborative. This concept was beautifully supported, as this report was being prepared, with a walk past a dozen offices in the Armstrong Research Center. In three of these offices, clusters of three and four persons were huddled around a CRT while one person acted to key in the concepts being reviewed and captured with word processing and spreadsheet software. Geographically, these persons could have been anywhere and with modem shared one "host" computer. Participants could have discussed this work over the telephone as they viewed their CRT. In fact, this work probably would have gone faster if each person had been in their office with a direct view of the CRT and a headset with which to communicate, rather than awkwardly leaning over shoulders and trying to point to key words on the CRT.

PRELIMINARY FRAMEWORK

Although a number of conceptual frameworks could drive inquiry into electronically supported learning teams (E/L-Teams) one broadly accepted framework is described given the **time** and **place** of the meeting. These are:

Same Time-Same Place (Persons meeting face-to-face in a conference room) Research for this meeting condition goes back 20 years using mainframe computer terminals in front of each participant for Delphi forecasting, anonymous voting, capture of ideas while others are speaking, etc. Today, similar systems are used by professionals who specialize in leading meetings. The limited research available suggests that the combined skill of the meeting convener and the technology invariably improve the

effectiveness of participants (number and quality of ideas generated, quality of problem solutions, etc.). However, there is little evidence that these meeting support systems have been used for formal education and no studies were found using this technology for the purpose of supporting a distance education course.

Same Time-Different Place (Persons simultaneously meeting in various locations) Support for this condition also goes back several decades through such services as audio phone bridges and special mainframe software that allow various persons to key a message, simultaneously, for an on-line communication. Today, Bulletin Board Systems offer "chat" (usually limited to a handful of persons). Also, video conferences are being used either full motion or through compressed and decompressed video (CODEC). Usually a limited number of locations can be served.

In practice, a well designed video conference room can effectively allow several locations and a dozen folk to meet and exchange ideas much as they would within a conference room. However, no research was obtained to indicate the effect of these systems for distance education.

Different Time-Same Place (Persons who rotate through some training or work facility such as a nuclear plant team; or, foreman rotating through multiple work shifts.) Support for this condition has received relatively little attention except with traditional office support systems which now include E-mail, word processing, process control displays, etc. Since the effort of my study is to involve distant learners, the condition of Different Time-Same Place was down played although there is some

research in this area related to Total Quality Control and similar management control systems.

Different Time-Different Place (Persons who have no meeting schedule and who are geographically separate.) Many special purpose teams such as a task force or a blue ribbon committee, will do most of their business in an asynchronous fashion and there are a good many systems to support these teams. Included are random phone calls, mail and more recently the fax, the phone answering machine, E-mail, BBS conferences and forums and increasingly phone mail.

Interestingly, two studies noted that if a team began by using electronic text support systems, such as a typical BBS might provide, then little use was made of voice system such as the telephone, phone answering devices and voice mail; and, vice versa. (Since each medium has both strengths and strong limitations, it would be interesting to determine why a single medium is used.) In corporations that are supported by a good E-mail system, it has been noted that the 3-office rule seems to prevail... if a colleague is located more than 3 doors away, then it is easier and more productive to use E-mail than to play telephone tag.

GENERAL RESEARCH FINDINGS

On the whole, research with tele-communications technology indicates a positive contribution to teams and group work. The few experimental studies that exist typically support the effect of the intervention. (Although, how many unsuccessful tries go not reported is anyone's guess!). There are more qualitative

studies than quantitative; and, a good many thought pieces! Generally, research indicates that E-Teams provide greater contributions, turn out better products, are better coordinated, hold fewer meetings (but communicate more frequently among members) and cross more organizational barriers than do teams who rely on face-to-face meetings and traditional communication.

RESEARCH ISSUES

Unfortunately, the generalizations described above are drawn from a variety of disciplines and are not easily studied within a theoretical framework such as organizational behavior, communication or telecommunications theory. Communications Development has some procedures that may be useful but the relevant research elements have typically has been used to described communications in communities or large complex organizations rather than a smaller team or group. Therefore, our initial review of the literature would suggest that an important research issue is that the study of electronically supported learning teams is a multi-disciplinary task made complicated by multiple terms to describe similar effects.

Nonetheless, it is possible to put some perspective on the plethora of issues by seeing E/L-teams in the context of a system having input, process, output and existing within some supra-structure. A tentative list of issues within this framework includes:

Input

- Optimum team member selection criteria (what kind of people work best with others; how do personalities interact with different types of tasks.

- Orientation? If so, what kind? How long? Must it be face-to-face? Can trust be developed outside of face-to-face?
- Minimum input skills? Keyboarding (for chat)?
- Can previous team skills transfer to new teams? E/L-teams?

Process

- What is the optimum size of a team?
- How best should an E/L-team add and drop members?
- How are roles defined on an E/L-team? What are the optimum number of roles?
- How do different roles interact with the specific task facing the team?)
- What are the characteristics of good supporting team software? How does a team navigate?
- Who, how, when, is feedback provided? How is team feedback different from individual feedback?
- Are there definable stages of a E/L-team (as there seem to be with traditional teams).

Output

- What is the best method of debriefing or closing down a team?
- How is a E/L-team (qua team) best evaluated?
- How are records kept and archived for a E/L-team?

Supra-Structure

- For what kind of learning outcomes are E/L-teams the most appropriate instructional strategy?
- Can or should hardware and software be standardized?
- Is there an optimal network configuration?

- How can quality be maintained with all aspects of E/L-teams?-

- What kinds of institutional research are appropriate?

ALTERNATE RESEARCH FRAMEWORKS

As with much research, circumstance may dictate resources such as funds, personnel, subjects and subject taught, etc. Subsequently, an ongoing continuing education course with a stable curriculum, multiple instructors and administration that would support different treatments would be ideal. Further, the odds are high that the circumstances of the course would suggest a place to start in terms of specific variables such as medium used, major attitudes and skills that could be the target for improvement. Local conditions are also an important consideration for the type of technology used both to originate and for reception.

Another approach might be to explore various delivery systems to determine costs vs. yield. Or, stated in other terms, are there selected instructional telecommunication technologies that seem stable and powerful enough such that they can become the building blocks around which important process questions can be explored. For example, various uses of the telephone can be documented which range from hotline support to its use to replace a live instructor. At the other extreme mega-funds support tele-business ventures such as CENET used by IBM for corporate training with 11 uplinks and 38 receive sites.

In the context of research with learning teams, one starting point for 'appropriate technology' research could be the use of

phone circuitry for two purposes: 1) small group audio conversation using an 800 number and "Meet-me" bridges (persons calling a given number are automatically connected together for full duplex operation); and, 2) groupware which could support, when appropriate, multiple persons working on the same computer program. This could be word processing, spreadsheet or other software appropriate for course objectives. This might mean two lines (although many individuals who have modems now have two lines) or special communications software such as distributed by Optel which multiplexes modem and voice frequencies on one line. Access to a computer also promises BBS and other computer mediated instruction for one-to-one or small group interaction.

With this relatively inexpensive technology traditional means of instruction could be maintained and various combinations of small groups supported with voice and computer mediated instruction could be used to test the effects of training, delivery, group character, achievement and problem solving skills, etc. Clearly, the issue is not equipment nor communication systems. Instead, an institution that supports on-going distance education is needed in order to find the supporting supra-structure within which research can occur.

NEXT STEPS

A RIP proposal will be prepared for September submission to the University of Texas at Austin. Unfortunately, the overhead rate of this institution now exceeds 50% which leaves little energy for research especially where technology is involved. However, every effort will be made to conceptualize a research project for the estimated \$9,000 available to investigators after

overhead is taken by UT officials. In the past notification of these awards was in December which means that research could begin early in 1993. Data could be collected during the Spring and a second summer faculty grant would be requested to finish data collection and describe the results. However, acceptance of a possible summer grant will be contingent upon knowledge of whether a graduate student will be assigned to this project. If this is not possible then it is highly unlikely that a second grant will be sought.

Note: Special thanks goes to Mr. George Houtman, graduate student in Instructional Technology of the University of Texas, who agreed to work on this project with neither pay nor expense money. He had primary responsibility for literature review and produced several excellent bibliographies (available from the author).

EVALUATION OF GLUCOSE MONITORING DEVICES FOR USE
IN HYPERBARIC CHAMBERS

Catherine H. Ketchum, Ph.D., NRCC
Instructor
Department of Pathology
Division of Laboratory Medicine

University of Alabama at Birmingham
West Pavilion, P-230
618 South 18th Street
Birmingham, AL 35233-7331

Final Report for:
Air Force Office of Scientific Research
Summer Research Program
Davis Hyperbaric Laboratory
Armstrong Laboratory
Brooks Air Force Base, TX

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

EVALUATION OF GLUCOSE MONITORING DEVICES FOR USE
IN HYPERBARIC CHAMBERS

Catherine H. Ketchum, Ph.D., NRCC
Instructor
Department of Pathology
Division of Laboratory Medicine
University of Alabama at Birmingham

Abstract

Significant decreases in whole blood glucose (WBG) levels have been noted in diabetic patients undergoing hyperbaric oxygen therapy thus making accurate monitoring of WBG's within a hyperbaric chamber critical to patient care. Five WBG monitoring instruments were evaluated to determine what effects, if any, increasing the partial pressure of oxygen by increasing chamber pressure from 1.0 to 2.4 atmospheres absolute (ATA) would have on instrument performance. Reagent strips exposed to chamber conditions for 30 dives showed no adverse effects in precision or stability. Using whole blood samples ranging from 25 to 250 mg glucose/dl, we found that the hyperbaric environment did have a significant effect on accuracy and did affect the precision of one instrument. All glucose results obtained with the Glucometer M+ were increased significantly, as were those less than 100 mg/dl with the ExacTech pen. All results were decreased significantly with the HemoCue and the Companion 2. Although significant decreases were noted with concentrations greater than 150 mg/dl, the OneTouch II was not affected in the hypoglycemic range.

EVALUATION OF GLUCOSE MONITORING DEVICES FOR USE IN HYPERBARIC CHAMBERS

Catherine H. Ketchum, Ph.D., NRCC

INTRODUCTION

Diabetic patients who have wound healing complications refractory to standard medical and surgical therapy must often be treated by more invasive interventions, i.e., limb amputation. Hyperbaric oxygen (HBO) therapy is a useful adjunct in the treatment of these patients resulting in an improvement in wound healing or a reduction in the amount of tissue amputated (1,2). HBO acts to improve wound healing by elevating tissue oxygen tension which results in increased collagen formation, enhanced fibroblast proliferation, and improved leukocyte function (1).

HBO therapy in diabetic patients, however, is not without risks as rapid and substantial drops in whole blood glucose (WBG) levels have been noted during and following therapy. Springer reported an average WBG decrease of 51 mg/dl in a study of 25 insulin dependent diabetics pre- and post-therapy (3). Patients experiencing decreases in blood glucose levels during HBO therapy may exhibit seizure-like activity which mimics the neurological manifestations of oxygen toxicity. Treatment for hypoglycemia and oxygen toxicity differ; therefore, it is important for hyperbaric personnel to be able to accurately monitor the glucose levels of these patients while within the hyperbaric chamber.

Some of the WBG monitoring devices designed for near-patient testing have been used in hyperbaric chambers but there have been questions regarding the effects of the hyperbaric environment, particularly the increase in oxygen availability, upon the device utilized. Most monitoring devices measure glucose via modified glucose oxidase based methods. Glucose oxidase impregnated onto reagent strips catalyzes the reaction between glucose and oxygen to gluconic acid and hydrogen peroxide. This reaction is either coupled to a second reaction which can be monitored photometrically or monitored electrochemically. Conditions of increased or decreased oxygen partial pressures have

been reported to interfere with the glucose oxidase reaction and the electrochemical reaction (4-9).

We evaluated five whole blood glucose monitoring devices to determine what effects increasing the partial pressure of oxygen by increasing the chamber pressure from 1.0 ATA to 2.4 ATA would have on performance. Our evaluation included four glucose oxidase based devices and one glucose dehydrogenase based device.

MATERIALS AND METHODS

Instruments: The glucose oxidase based instruments included the ExacTech pen (MediSense, Inc, Cambridge, MA), the Companion 2 (MediSense, Inc, Cambridge, MA), the OneTouch II (Lifescan, Inc, Milpitas, CA), and the Glucometer M+ (Miles Inc, Elkhart, IN). The glucose dehydrogenase based instrument was the HemoCue (Mission Viejo, CA). The Vision glucose assay (Abbott Laboratories, Abbott Park, IL), a hexokinase based assay, was used outside the chamber as a reference method.

Glucose samples: A pool of remnant (Na₂)EDTA anticoagulated whole blood was allowed to degrade to a glucose level of less than 20 mg/dl. A 10% dextrose solution was used to prepare aliquots with glucose concentrations of 25, 50, 100, 150, and 250 mg/dl. The glucose concentrations were verified using the Vision pre- and post-dives. Control solutions provided by the manufacturers were used also.

Evaluation protocol: Hyperbaric medicine personnel were trained to use specific instruments. The evaluation was conducted in two phases (dives) to avoid excessive exposure of support personnel to HBO. Reagent strips for the ExacTech pen, the OneTouch II, and the Glucometer M+ were divided into three groups of 100 strips each. Group 1 strips were stored per the manufacturer's instructions within the laboratory facility. Group 2 strips were stored per instructions in the hyperbaric chamber. Group 3 strips were sealed hermetically within metal containers prior to storage in the chamber. The

Companion 2 reagent strips and the HemoCue cuvettes were included only in Group 1 because of strip availability and storage restrictions respectively.

Dive 1. Using Group 1 reagent strips and cuvettes, all instruments were calibrated and controls verified. First, the precision of each instrument was determined by assaying low and high control solutions in replicates of ten. Each aliquot was assayed in replicates of ten at 1.0 ATA. Personnel and equipment were transferred to the hyperbaric chamber and pressurized to 2.4 ATA with a chamber environmental oxygen concentration of 21%. The aliquots were reassayed in replicates of ten. Dive 2: Fresh glucose samples were prepared and assayed as described in Dive 1 using Group 2 and Group 3 strips.

RESULTS

The precision studies performed at 1.0 ATA using control solutions are summarized in Table 1. All instruments were found to perform within the manufacturer's claims under these conditions. As can be seen in Table 2, the changes in precision were not significant when testing the WBG samples at 1.0 ATA and at 2.4 ATA using the Companion 2, the ExacTech pen, the OneTouch II, and the HemoCue. There was a significant ($p < 0.05$) change in precision at 1.0 ATA vs 2.4 ATA when assaying the 25, 50, and 100 mg/dl samples using the Glucometer M+.

We were unable to measure the 25 mg/dl WBG sample using the ExacTech pen. At 2.4 ATA, we observed a significant ($p < 0.05$) increase at the 50 and 100 mg/dl concentrations and a nonsignificant decrease at the 150 and 250 mg/dl concentrations. We found the concentrations greater than 100 mg/dl increasing 18.5 to 104.9% at 2.4 ATA using the Glucometer M+. There was a decrease at 2.4 ATA with all aliquots tested with the Companion 2 and the HemoCue. The decrease was significant statistically ($p < 0.01$) at all concentrations for the HemoCue and at the 25, 100, and 150 mg/dl concentrations for the Companion 2. We found a significant decrease at 2.4 ATA in the

results obtained with the OneTouch II at the 150 and 250 mg/dl concentrations.

The regression analyses comparing each instrument with the Vision are as follows: Companion 2, $y = 1.02x - 0.89$, $r = 0.99$; Glucometer M+, $y = 0.96x + 12.83$, $r = 0.97$; ExacTech pen, $y = 0.85x - 2.29$, $r = 0.99$; OneTouch II, $y = 0.81x + 1.09$, $r = 0.98$; HemoCue, $y = 0.93x + 5.43$, $r = 0.99$.

DISCUSSION

At 1.0 ATA all of the instruments evaluated were found to meet manufacturer's claims with respect to precision using control solutions. Additionally, these initial studies allowed operators adequate time to become familiar with and demonstrate proficiency with their assigned instruments. The HemoCue and the OneTouch II were the most precise when testing the WBG samples at 1.0 ATA.

At 1.0 ATA, we found the instruments to correlate with the Vision. Since the Vision actually separates the erythrocytes and uses the plasma filtrate as the sample, the values were expected to be 10 to 15 % higher than those obtained with the WBG devices.

By testing the WBG samples first at 1.0 ATA and then at 2.4 ATA, we were able to determine that increasing the chamber pressure from 1.0 ATA to 2.4 ATA and thus the partial pressure of oxygen did effect the performance of the instruments evaluated. Although the chamber air was maintained at 21% oxygen, as the partial pressure of oxygen was increased from 159.6 mmHg to 383 mmHg, more oxygen was available to react with glucose in the glucose oxidase catalyzed reaction or to interfere with the electrochemical reaction.

Accuracy was decreased significantly with the Glucometer M+, but precision was decreased at the lower concentrations. This may be due, in part, to a matrix effect when the blood was wiped

from the reagent strip prior to reading. Some investigators have shown the wipe technique to decrease the glucose determination by as much as 50% (10), but, the precision obtained with the control solutions suggests that operator technique was not a contributing factor. However, removal of the blood did increase strip exposure to the oxygen enriched environment. The ExacTech pen did not have the analytical sensitivity necessary to monitor hypoglycemic events. Additionally, the lower concentrations were falsely increased which is also unacceptable when monitoring hypoglycemic events. This is most likely due to interference of the increased environmental oxygen with the electrochemical reaction. Glucose values decreased at 2.4 ATA with the Companion 2 and the HemoCue. While the decreases were significant statistically, they were not significant clinically. The effect observed with the HemoCue cannot be related to the increase in oxygen as oxygen does not play a role in the glucose dehydrogenase reaction, but may be an effect of the increased pressure on the electronics of the instrument. We observed an increase in the reaction time at 2.4 ATA which would support this theory. While the decrease in the normoglycemic to hyperglycemic glucose values obtained with the OneTouch II were statistically significant the decrease was not clinically significant. This instrument performed the best with respect to precision.

The finding that the reagent strips could be stored in the chamber was expected since the strips are not exposed in the chamber to any of the conditions known to increase deterioration, i.e., high humidity and temperatures less than 0 C or higher than 30 C.

In summary, we found the hyperbaric environment to affect all of the instruments evaluated. The instruments from least to most effected were the OneTouch II, HemoCue, Companion 2, ExacTech pen, and the Glucometer M+.

REFERENCES

1. Davis JC. The use of adjuvant hyperbaric oxygen therapy in treatment of the diabetic foot. *Clin Podiatr Med Surg* 1987 4:429-437.
2. Baroni G, Porro T, Faglia E, et al. Hyperbaric oxygen in diabetic gangrene treatment. *Diabetes Care* 10:81-86 (1987).
3. Springer T. The importance of glucometer testing of diabetic patients pre and post dive. *Undersea Biomed Res* 18 (suppl):20 (1991).
4. Zel G. Inaccuracies of blood sugar determinations in the hyperbaric environment using Chemstrips. *Undersea Biomed Res* 16 (suppl):20 (1989).
5. Gregory M, Tyan F, Barnett JC, Youtz T. Altitude and relative humidity influence results produced by glucose meters using dry reagent strips. *Clin Chem* 34:1312 (1988).
6. Barnett C, Ryan F, Ballonoff L. Effect of altitude on the self monitoring of blood glucose (SMBG). *Diabetes* 35 (suppl): 117A (1987).
7. Shafner MR. The effect of increased atmospheric pressure on glucose strip accuracy. *Milit Med* 157: 162-165 (1992).
8. Khraisha S. Comparative study of serum insulin, glucose, growth hormone and cortisol of students at 794.7 mmHg (Dead Sea level) and 697.5 mmHg (Amman) barometric pressures. *Aviat Space Environ Med* 61:145-147 (1990).
9. Giordano BP, Hodges C, Thrash W, et al. Performance of seven blood glucose testing systems at high altitude. *Diabetes Educat* 15: 444-448 (1989).
10. Jones B and Bachner P. Bedside glucose monitoring data analysis and critique. Q-Probes, College of American Pathologists Quality Assurance Service, 91-09A, 1992.

Table 1. Precision Studies at 1.0 ATA Using Control Solutions

	<u>Companion</u>		<u>Glucometer</u>		<u>ExacTech</u>		<u>OneTouch</u>		<u>HemoCue</u>	
n	10	10	10	10	10	10	10	10	10	10
x(mg/dl)	46	317	36	390	101	391	54	337	34	450
sd(mg/dl)	3.5	15.7	1.5	12.2	5.7	13.6	1.3	4.5	1.7	23.7
cv(%)	7.7	4.9	4.2	3.1	5.6	3.5	2.5	1.3	5.0	5.3

Table 2. Summary of Whole Blood Glucose Data (1.0 ATA/2.0 ATA)

COMPANION 2

n	10/10	10/10	10/10	10/8	10/5
x(mg/dl)	34/27	65/61	116/102	138/117	255/257
sd(mg/dl)	4.6/3.5	4.6/5.2	8.0/9.6	6.9/12.3	12.7/0.8
cv(%)	13.7/13.2	7.0/8.5	6.9/9.4	5.0/10.6	4.9/0.3
[(x1-x2)/x1]100	-20.6	-6.2	-12.1	-15.2	+1.0

Regression equation (1.0 ATA) $y = 1.02x - 0.89$, $r = 0.99$

GLUCOMETER M+

n	10/10	10/10	10/10	10/6	10/8
x(mg/dl)	30/28	72/73	121/141	187/273	232/>500
sd(mg/dl)	1.6/0.5	15/8	4.7/15.9	32/40	21/-
cv(%)	5.4/1.7	20.7/11.1	3.9/11.3	17/14.7	9.1/-
[(x1-x2)/x1]100	-6.6	+1.4	+16.5	+46.0	+116

Regression equation (1.0 ATA) $y = 0.96x + 12.83$, $r = 0.97$

EXACTECH

n	10/10	10/10	10/10	10/10	10/10
x(mg/dl)	L/L	58/65	91/99	116/110	210/202
sd(mg/dl)	-/-	7.8/5.1	6.7/5.1	5.8/7.7	9.4/13.2
cv(%)	-/-	13.4/7.8	7.3/6.9	5.1/7.0	4.5/6.6
[(x1-x2)/x1]100	-	+12.1	+8.8	-5.2	-3.8

Regression equation (1.0 ATA) $y = 0.85x - 2.29$, $r = 0.99$

HEMOCUE

n	10/10	10/10	10/10	10/10	10/10
x(mg/dl)	29/23	58/56	93/85	125/112	233/214
sd(mg/dl)	2.0/1.6	1.9/1.0	3.2/2.4	2.2/0.9	2.9/3.4
cv(%)	6.8/7.0	3.3/1.7	3.4/2.8	1.7/0.8	1.2/1.6
[(x1-x2)/x1]100	-20.6	-3.4	-8.6	-10.4	-8.2

Regression equation (1.0 ATA) $y = 0.93x + 5.43$, $r = 0.99$

ONETOUCH II

n	10/10	10/10	10/10	10/6	10/8
x(mg/dl)	31/30	59/60	83/82	111/105	209/185
sd(mg/dl)	1.3/1.1	3.2/2.2	2.1/3.2	2.3/2.1	2.8/3.6
cv(%)	4.3/3.6	5.4/3.6	2.5/3.9	2.0/2.1	1.3/1.9
[(x1-x2)/x1]100	-3.2	+1.7	-1.2	-5.4	-11.5

Regression equation (1.0 ATA) $y = 0.81x + 1.09$, $r = 0.98$

A NEW PROTOCOL FOR STUDYING CAROTID BARORECEPTOR FUNCTION

Arthur J. Koblasz
Associate Professor
School of Civil Engineering

Georgia Institute of Technology
CE 0355
Atlanta, GA 30332

Final Report for
Summer Research Program
Armstrong Laboratory

Sponsored by
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

A NEW PROTOCOL FOR STUDYING CAROTID BARORECEPTOR FUNCTION

Arthur J. Koblasz
Associate Professor
School of Civil Engineering
Georgia Institute of Technology

Abstract

A new protocol was evaluated for characterizing carotid baroreceptor function in baboons. Three male baboons were anesthetized with Ketamine (30mg IM) followed by periodic doses of α -Chloralose (50mg/kg IV, plus maintenance doses of 20mg/hr). Pancuronium Bromide (.1mg/kg IV) was also given to each animal to reduce muscle activity. The right carotid sinus was isolated by ligating all incoming and outgoing vessels approximately 1 cm above and below the sinus. A small catheter was inserted into the carotid sinus, along with a 2-French Millar (single-tipped) pressure transducer. A Skinner three-way hydraulic valve (Type B14) was used to switch the catheter between two reservoirs of warmed physiological saline solution, which were positioned at different heights above the carotid sinus.

This apparatus allowed the pressure in the carotid sinus to be quickly shifted between two different levels, creating both pulse and step changes in pressure. Since the carotid sinus was completely isolated, the pressure shifts occurred nearly instantaneously with no significant flow of saline into the sinus.

A NEW PROTOCOL FOR STUDYING CAROTID BARORECEPTOR FUNCTION

Arthur J. Koblasz

INTRODUCTION

Baboons have been previously used in chronic centrifuge studies at Armstrong Laboratories and are typically instrumented with the following transducers:

1. Doppler ultrasound crystal positioned near the outlet of the aortic valve,
2. Konigsberg pressure transducer sutured into the aorta approximately 1 cm downstream from the doppler crystal,
3. 3-French Millar (single-tipped) pressure transducer inserted through an implanted catheter and hemostatic valve into the Right Atrium.

The baboons are exposed to different g-accelerations using the Brooks Centrifuge, and the resulting cardiovascular changes are continuously monitored. The protocol described below is recommended at the time of euthanasia to provide in vitro data which can be compared with the in vivo centrifuge data.

APPARATUS

The pulse and step changes in pressure were produced by switching between two reservoirs suspended at different heights above the carotid sinus. A Skinner three-way hydraulic valve (Type B14) was used to switch between the two reservoirs. The circuit diagram shown in Figure 1 produces periodic pulses and step signals which were used to drive the solenoid mechanism attached to the Skinner valve.

At the conclusion of each experiment, we also inserted a small gauge EMG Electrode (Teca MG-37) into the lumen of the sinus, and 10 msec pulses of current (amplitude less than 1 mAmp) were used to electrically stimulate the baroreceptors. The pulses of electrical stimuli were presented in two different patterns-- periodic pulses occurring once every 11 seconds and random interval pulses occurring at a minimum interval of 110 msec. Figure 1 also shows the circuit diagram for creating these electrical stimuli.

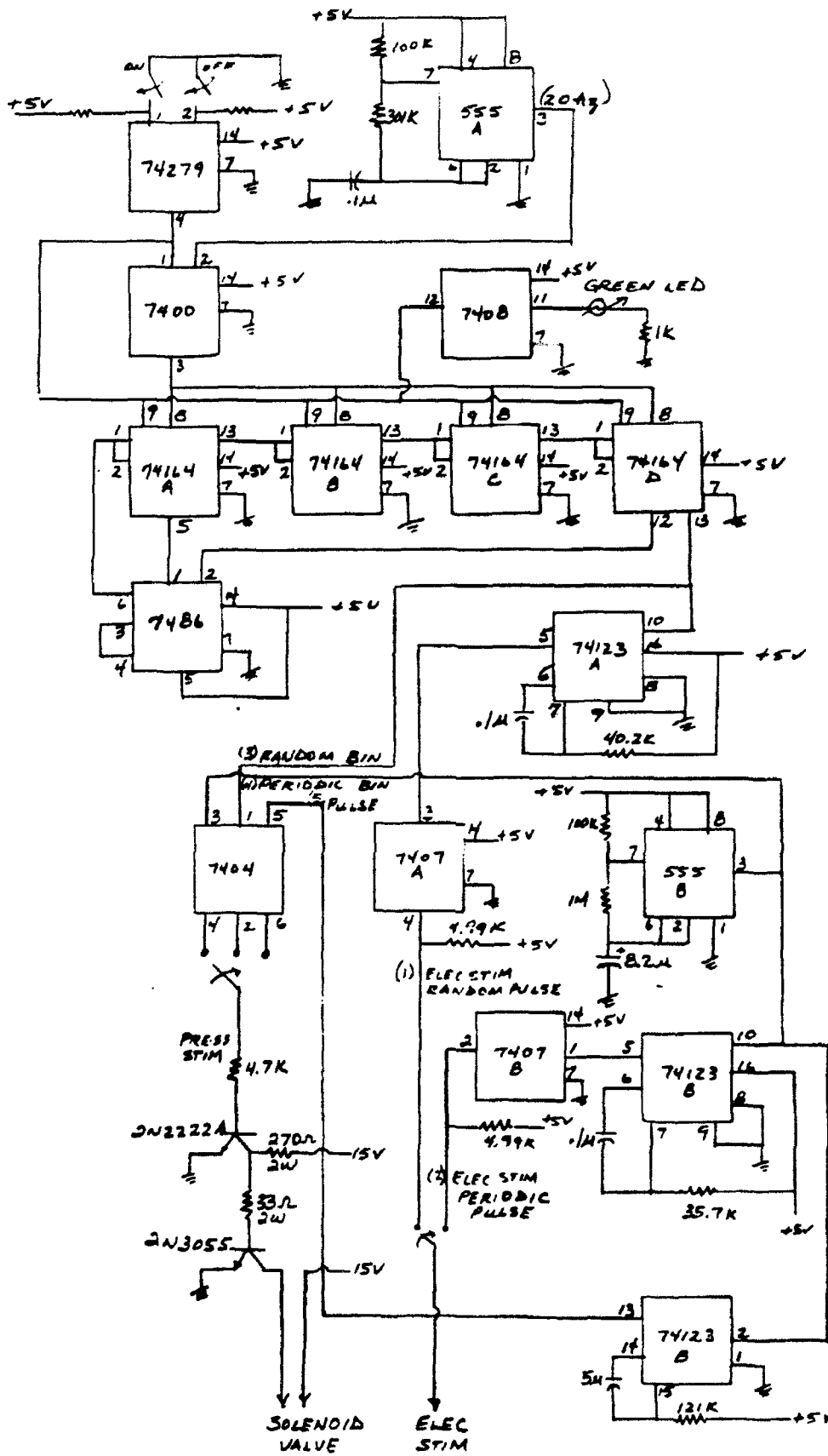


Figure 1
Electrical Circuit

PROTOCOL

Pressure Stimuli. In the first test condition, a small catheter was inserted into the right carotid sinus. The sinus was isolated by ligating all incoming and outgoing vessels. The height of the "normally-on" reservoir was then adjusted to provide a steady carotid pressure of 100mm Hg. The level of the second reservoir was adjusted to provide square-wave and pulse changes in pressure as indicated in Figure 2. Each stimulus condition was sustained for a period of three minutes (see Figure 2). In the second and third test conditions, the same square-wave and pulse modulations were repeated at different mean levels of 70mm Hg and 40mm Hg.

In an attempt to quantify amplitude-dependent nonlinearities of the baroreceptor response, the second test condition (i.e. mean carotid pressure of 70mm Hg) was repeated with larger amplitudes of square-wave and pulse modulations (i.e. 50mm Hg changes in carotid pressure).

Electrical Stimuli. After completing the battery of different pressure stimuli, a new protocol was evaluated for eliciting an electrical response. The Millar transducer was removed from the carotid sinus, and the normally-on reservoir was readjusted to provide a constant carotid pressure of 100mm Hg. A stainless steel EMG epidermal electrode was inserted into the lumen of the carotid sinus, and a stainless steel wire electrode was wrapped around the outside of the sinus. A ground electrode was clipped to the right ear lobe. Pulses of electrical current (i.e. 5 mSec duration and <1 mAmp amplitude) were used as electrical stimuli, as shown in Figure 3.

The pressure levels in the aorta and right atrium were measured continuously using the implanted transducers described above. In addition, the pressure in the right femoral artery was measured using a 3-French Millar transducer. All three pressure measurements were recorded on a Racal Instrumentation Recorder (V-Store, 24 Channel), along with the EKG and the pressure/electrical stimuli. The protocol described above was repeated on three adult male baboons.

DISCUSSION

The six channels of recorded data were each low-pass filtered (100 Hz corner frequency) and played into a 12 Bit A/D converter (Data Translation, DT2801A) at a sampling rate of 200 Hz. The data was then transferred via Ethernet to the Hydra Computer (Sequent S81) at Georgia Tech. A total of 16 Megabytes of data were collected for the different pressure and electrical stimuli. We are now trying to recruit a graduate student at Georgia Tech who will help us analyze these data.

The flow rate transducers implanted in the aorta of each baboon were not producing reliable signals; therefore we did not transcribe this data. During the autopsy of each baboon, the implanted ultrasound crystals were examined under a microscope. The acrylic coating over each doppler crystal had decomposed, leaving a scaly surface over the crystal. A new design for the crystal housing was suggested, which is depicted in Figure 4.

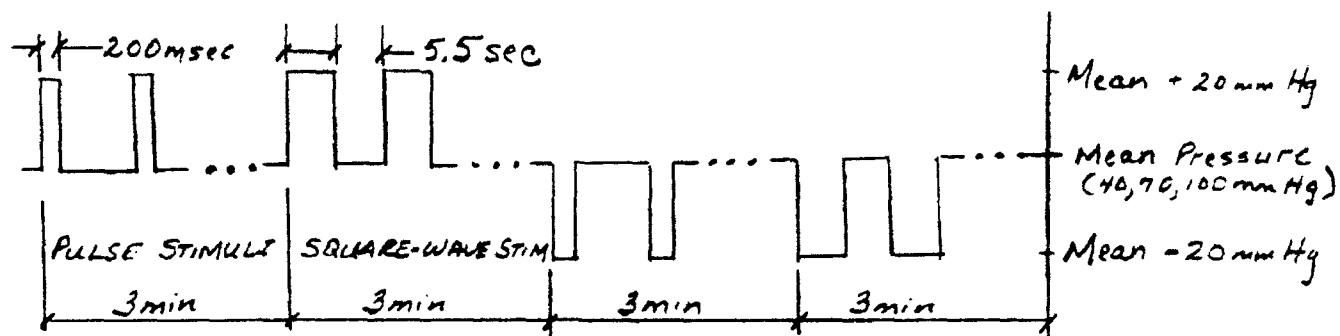


Figure 2
Pressure Stimuli

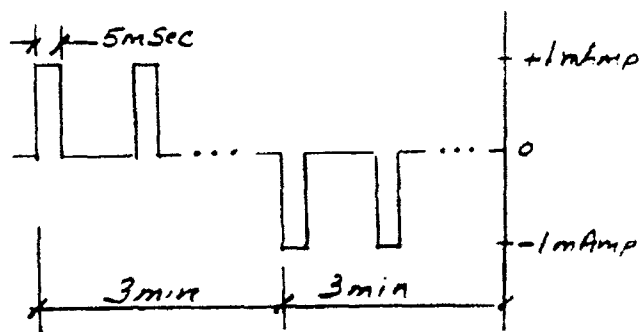


Figure 3
Electrical Stimuli

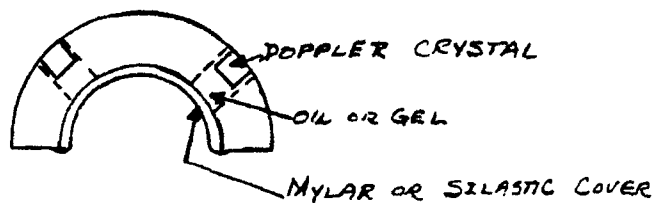


Figure 4
Ultrasound Crystal Housing

A STUDY ON TELEOPERATED SYSTEM CONTAINING AN EXOSKELETON

**A.J. Koivo
Professor
School of Electrical Engineering**

**Purdue University
1285 Electrical Engineering Building
West Lafayette, IN 47907-1285**

**Final Report for:
AFOSR Summer Research Program
AL/CFBS, Armstrong Laboratory
Wright Patterson AFB, OH 45433
RESEARCH & DEVELOPMENT LABORATORIES**

June 3 - August 19, 1992

A STUDY ON TELEOPERATED SYSTEM CONTAINING AN EXOSKELETON

A.J. Koivo
Professor
School of Electrical Engineering
Purdue University

Abstract

A teleoperated system consisting of a human operator, an exoskeleton (leader-master), a robotic manipulator (follower-slave) and environment is studied. The kinematics of the exoskeleton are first studied by assigning the coordinate frames by following the Denavit-Hartenberg (D-H) guidelines. Then the kinematic relations are developed, including the forward kinematic equations and the Jacobian matrix, the transpose of which relates the joint torques to the generalized force (force and torque) acting on the end-effector. Since the joint shafts are driven by a cable (wire) system, the torques of the motors are next calculated in terms of the joint torques. Thus, the kinematic relations for the exoskeleton are determined.

Then, the entire teleoperated system is considered. This system with the exoskeleton is currently being designed and constructed in the Laboratory on Human Sensors, Feedback and Telerobotics. The use of two-port networks to study the behavior of teleoperated systems was reviewed. The ideal response of the teleoperated system is defined as the steady-state response in which (i) the velocities of the leader (master) and follower (slave) are the same and (ii) the (generalized) forces exerted by the follower on the environment and by the leader (master) on the exoskeleton are equal. The conditions for the ideal response are presented in terms of the two-port network model which is used to model the behavior of the teleoperated system.

The effect of the time delay in the system is discussed. A recommendation for the control architecture of the teleoperated system is then presented.

A STUDY ON TELEOPERATED SYSTEM CONTAINING AN EXOSKELETON

A.J. Koivo

1. INTRODUCTION

An exoskeleton is a unique structure that complements a human arm. It is attached to the human arm (Figure 1), and it moves when the arm is moved; thus, it can provide the representation of the motion of the human arm. The motion of the exoskeleton can be represented, for example, by means of the readings of the encoders attached to the joints of the exoskeleton. These measurements can be transmitted to a remote site in order to duplicate the motion of the exoskeleton and thus that of the human arm. Therefore, the exoskeleton makes the leader-follower (master-slave) teleoperation possible; that is, the other (robotic) arm (called the follower or slave) at a distant location is made to perform the same motion as that of the human arm.

In order to make the follower (slave) perform tasks which require dexterity, the system should be equipped with sensory feedback, for example, the (wrist) force-torque sensors, and possibly tactile sensors. These devices can provide the leader (master) with force-torque information which is essential in the completion of tasks requiring dexterity, for example, in assembly. Another possible sensory modality can be furnished to the leader (master) by camera images.

An exoskeleton designed and manufactured by Odetics, Inc., has been delivered to the U.S. Air Force Armstrong Laboratory. Our study will focus on the establishment of basic characteristics for this exoskeleton and for the teleoperated system of which the exoskeleton is a part.

2. KINEMATIC RELATIONS IN EXOSKELETON

2.1. Positional Relations for Exoskeleton

A schematic drawing of the exoskeleton is shown in Figure 2. The exoskeleton has seven revolute joints (links). The axes of the rotations are marked in Figure 2. To establish the kinematic relations for the exoskeleton, the coordinate frame $O_i-x_iy_iz_i$, $i = 0,1,\dots,6$ with the origin at O_i is assigned for each link. It is accomplished by following the Denavit-Hartenberg (D-H) guidelines. It should be noted that the origin O_0 of the base coordinate frame, which is somewhat arbitrary, is chosen on the rotational axis (z_0) of the first link (joint) and placed to the point where the rotational axis (z_1) of the second link (joint) intersects that (z_0) of the first link (joint). The remaining coordinate frames are determined by the standard D-H procedure.

Since the rotational axis of the handle (the end-effector) is placed away from the rotational axis (z_6) of the seventh joint (link), a gripper coordinate frame $O_G-x_Gy_Gz_G$ is defined so that the z_G -axis coincides with the vertical axis of the gripper. The x_G -axis is chosen so that it is in the plane of the z_6 - and z_G -axes. The y_G -axis completes the Cartesian coordinate system. The origin O_G of the gripper coordinate frame $x_Gy_Gz_G$ is placed at the center point of the handle

(palm). Thus, the assignments of the coordinate frames for the exoskeleton are completed.

A vector p_i in the i th coordinate frame, $i = 1, \dots, 6, G$, can be expressed in the $(i-1)$ st coordinate frame by using the following relation:

$$p_{i-1} = A_{i-1}^i p_i \quad (1)$$

where A_{i-1}^i is a homogeneous transformation matrix of dimension (4×4) .

To determine matrix A_{i-1}^i , the structural kinematic parameters (a_i, α_i, θ_i and d_i) of the exoskeleton are determined by following the (D-H) guidelines. The application of these guidelines to Figure 2 gives the results displayed in Table 1.

TABLE 1

i	a_i	α_i	θ_i	d_i
1	$a_1 = 0$	$\alpha_1 = 90^\circ$	θ_1	$d_1 = 0$
2	$a_2 = 0$	$\alpha_2 = -20^\circ$	θ_2	$d_2 = 0$
3	$a_3 = -1.969$	$\alpha_3 = 120^\circ$	θ_3	$d_3 = 14.640$
4	$a_4 = 1.969$	$\alpha_4 = -70^\circ$	θ_4	$d_4 = 0$
5	$a_5 = 0$	$\alpha_5 = 70^\circ$	θ_5	$d_5 = 11.770$
6	$a_6 = 0$	$\alpha_6 = 90^\circ$	θ_6	$d_6 = 0$
$7 = G$	$a_7 = -1.625$	$\alpha_7 = 0^\circ$	$\theta_7 = 0^\circ$	$d_7 = \text{const.}$

The units of the numbers are in inches. Note: $d_5 = \overline{C_4 O_5}$ in Figure 2.

Using the values of the structural kinematic parameters given in Table 1, the homogeneous transformation matrix A_{i-1}^i , $i = 1, 2, \dots, 6, G$ can be obtained. Its general expression is as follows:

$$A_{i-1}^i = \begin{bmatrix} \cos\theta_i & -\cos\alpha_i \sin\theta_i & \sin\alpha_i \sin\theta_i & a_i \cos\theta_i \\ \sin\theta_i & \cos\alpha_i \cos\theta_i & -\sin\alpha_i \cos\theta_i & a_i \sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The substitution of the values from Table 1 to the expression in equation (2) gives

$$A_0^1 = \begin{bmatrix} c_1 & 0 & s_1 & 0 \\ s_1 & 0 & -c_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} ; \quad A_1^2 = \begin{bmatrix} c_2 & \frac{1}{2}s_2 & \frac{\sqrt{3}}{2}s_2 & 0 \\ s_2 & -\frac{1}{2}c_2 & -\frac{\sqrt{3}}{2}c_2 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$A_2^3 = \begin{bmatrix} c_3 & \frac{1}{2}s_3 & \frac{\sqrt{3}}{2}s_3 & -1.969c_3 \\ s_3 & -\frac{1}{2}c_3 & -\frac{\sqrt{3}}{2}c_3 & -1.969s_3 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 14.64 \\ 0 & 0 & 0 & 1 \end{bmatrix} ; \quad A_3^4 = \begin{bmatrix} c_4 & -\rho_1 s_4 & -\rho_2 s_4 & 1.969c_4 \\ s_4 & \rho_1 c_4 & \rho_2 c_4 & 1.969s_4 \\ 0 & -\rho_2 & \rho_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$A_4^5 = \begin{bmatrix} c_5 & -\rho_1 s_5 & \rho_2 & 0 \\ s_5 & \rho_1 c_5 & -\rho_2 c_5 & 0 \\ 0 & \rho_2 & \rho_1 & 11.77 \\ 0 & 0 & 0 & 1 \end{bmatrix} ; \quad A_5^6 = \begin{bmatrix} c_6 & 0 & s_6 & 0 \\ s_6 & 0 & -c_6 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} ; \quad A_6^G = \begin{bmatrix} c_7 & -s_7 & 0 & a_7 c_7 \\ s_7 & c_7 & 0 & a_7 s_7 \\ 0 & 0 & 1 & d_7 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

where $\rho_1 = \cos 70^\circ$ and $\rho_2 = \sin 70^\circ$.

As an example on the application of equation (1), the center point O_G of the gripper can be expressed in the base coordinate system $O_0-x_0y_0z_0$. Indeed, the coordinates of O_G in the gripper coordinate frame are $p_{GO} = [0 \ 0 \ 0 \ 1]^T$. The coordinates of this same point in the base (zerth) coordinate system are

$$p_{0G} = A_0^1 A_1^2 A_2^3 A_3^4 A_4^5 A_5^6 A_6^G p_G = A_0^G p_{GO} \quad (5)$$

where matrix $A_0^G = A_0^1 A_1^2 A_2^3 A_3^4 A_4^5 A_5^6 A_6^G$ can be computed by means of expressions in equations (3) and (4).

2.2. Velocity Relations for Exoskeleton

In the teleoperation, the follower (slave) is often required to track the velocity profile of the end-effector of the leader (master). The operational velocity V of the end-effector expressed in the fixed base coordinate system $O_0-x_0y_0z_0$ is determined by the rotational velocities $\dot{\theta}_i$, $i = 1, \dots, 6, G$ of the joints, as is well-known:

$$V = J_G(\theta)\dot{\theta} \quad (6)$$

where the generalized velocity vector $V = [v^T \ w^T]^T$ consists of the translational (v) and rotational (w) velocity vectors, and the superscript "T" refers to transposition, and the joint velocity vector $\dot{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_6 \ 0]^T$. The Jacobian (6x7) matrix $J_G(\theta)$ can be expressed in a general form as follows:

$$J_G(\theta) = \begin{bmatrix} \vec{k}_{z_0} \times (\vec{p}_{0G} - \vec{p}_{00}) & \vec{k}_{z_1} \times (\vec{p}_{0G} - \vec{p}_{01}) & \dots & \vec{k}_{z_6} \times (\vec{p}_{0G} - \vec{p}_{06}) \\ \vec{k}_{z_0} & \vec{k}_{z_1} & \dots & \vec{k}_{z_6} \end{bmatrix} \quad (7)$$

where the "x" denotes a cross-product, \vec{k}_{z_i} is the unit vector along the positive z_i -axis, vector p_{0i} , $i = 1, 2, \dots, 6, G$, is the coordinate of the origin of the i th frame expressed in the zeroth (base) coordinate system. One recalls that vector p_{0i} appears as the fourth column of the homogeneous transformation matrix A_0^i . It should be noted that all vectors in expression (7) must be expressed in the same coordinate system, for example, in the zeroth coordinate system.

Since \vec{p}_{0G} and \vec{p}_{0i} , $i = 0, 1, \dots, 6$ are expressed in the base coordinate system, the unit vector \vec{k}_{z_i} in the $(i-1)$ column in equation (7) is converted to the base coordinate system by means of the rotation matrix $(A_0^i)_R$ of the homogeneous transformation matrix $(A_0^i)_R$. Thus,

$$(A_0^i)_R k_{z_i} = [k_{z_i}^{0x} \ k_{z_i}^{0y} \ k_{z_i}^{0z}]^T \quad (8)$$

where $k_{z_i} = [0 \ 0 \ 1]$ and the superscript of the variables on the right side specify the direction of the component in the zeroth coordinate system.

For computational efficiency, the cross-products in equation (7) may be converted to a matrix-vector product as follows:

$$\vec{k}_{z_i} \times (\vec{p}_{0G} - \vec{p}_{0i}) \rightarrow \begin{bmatrix} 0 & -k_{z_i}^{0z} & k_{z_i}^{0y} \\ k_{z_i}^{0z} & 0 & -k_{z_i}^{0x} \\ -k_{z_i}^{0y} & k_{z_i}^{0x} & 0 \end{bmatrix} \begin{bmatrix} p_{G_i}^x \\ p_{G_i}^y \\ p_{G_i}^z \end{bmatrix} \quad (9)$$

where $[p_{G_i}^x \ p_{G_i}^y \ p_{G_i}^z]^T = p_{0G} - p_{0i}$ and $i = 0, 1, \dots, 6$.

Equation (7) can be calculated for any given configuration of the exoskeleton. Hence, the relationship between the (rotational) joint velocities and the Cartesian generalized velocity is determined.

3. FORCE RELATIONS IN EXOSKELETON

3.1. Joint Torques as Functions of Force/Torque Measurements

An external (static) force f_e on the center point of the gripper (handle) and an external (static) moment n_e acting about the axis of the gripper coordinate frame which in the sequel will be called concisely the generalized external force, F_G , can be related to the torques acting on the shafts of the rotational joints by means of the Jacobian matrix, i.e.,

$$\tau = J_G^T(\theta)F_G \quad (10)$$

where the joint torque vector $\tau = [\tau_1 \ \tau_2 \ \dots \ \tau_6]^T$ and $\tau_i, i = 1, \dots, 6$ acts on the shaft of joint i , and vector $F_G = [f_e^T \ n_e^T]^T$ consists of the force vector $f_e^T = [f_{ex} \ f_{ey} \ f_{ez}]$ and the torque vector $n_e^T = [n_{ex} \ n_{ey} \ n_{ez}]$ with the subscript indicating the direction of the particular component.

In the exoskeleton of Odetics, the wrist force/torque sensor is not located on the center point of the handle. Instead, the center point of the force/torque sensor is on the z_6 -axis at the point $(0,0,l_s)$ relative to the sixth coordinate frame. It is convenient to define a coordinate frame $O_s-x_s y_s z_s$ for the wrist force/torque sensor in such a way that the origin is in the center point of the sensor and the axes are parallel with the respective axes of the $O_6-x_6 y_6 z_6$ coordinate frame. The wrist force/torque sensor measures force $f_s^T = [f_{sx} \ f_{sy} \ f_{sz}]$ acting on point O_s and moment $n_s^T = [n_{sx} \ n_{sy} \ n_{sz}]$ exerting a torque about the coordinate axes of $O_s-x_s y_s z_s$ and the subscript indicates the direction of the component. These measured values represent external force and torque components which cause translational and rotational motions of the exoskeleton.

The problem is now to determine an equivalent force f_G acting on point G and an equivalent moment n_G exerting a torque about the axes of the $O_G-x_G y_G z_G$ system such that the response of the system to $[f_G^T \ n_G^T] = F_G^T$ is the same as the system response to $F_s^T = [f_s^T \ n_s^T]$. The solution is provided by the force transmission matrix L_S^G which defines the equivalence of the generated forces F_s and F_G :

$$F_s = L_S^G F_G \quad (11)$$

Matrix L_S^G can be decomposed as follows:

$$L_S^G = \begin{bmatrix} I & 0 \\ E_S^G & I \end{bmatrix} \quad (12)$$

where I represents a (3×3) identity matrix. Submatrix E_S^G can be written on the basis that the generated forces F_s and F_G have the same dynamic effect on the motion. It follows that

$$E_S^G = \begin{bmatrix} 0 & -d_7 & 0 \\ d_7 & 0 & -a_7 \\ 0 & a_7 & 0 \end{bmatrix} \quad (13)$$

It may be noticed that the equivalence relation between F_s and F_G can also be written as

$$F_G = L_G^S F_S \quad (14)$$

where

$$L_G^S = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & | & & \\ 0 & 1 & 0 & | & & \underline{0} \\ 0 & 0 & 1 & | & & \\ \hline 0 & d_7 & 0 & | & 1 & 0 & 0 \\ -d_7 & 0 & a_7 & | & 0 & 1 & 0 \\ 0 & -a_7 & 0 & | & 0 & 0 & 1 \end{array} \right] \quad (15)$$

It may be readily checked that ${}^G_S L_G^S = I$.

The joint torques can now be determined in terms of the measured values. Indeed, the substitution of equation (14) into equation (10) yields:

$$\tau = J_G^T(\theta) L_G^S F_S \quad (14)$$

Equation (16) gives the torques on the joint shafts when the measurement F_S is known.

The joint torques are generated by the motors (actuators) using cable wires. Thus, the next task is to relate the components of the joint torque vector τ to the torque vector τ_m of the motors (actuators).

3.2. Joint Torques as Functions of Motor Torques

Each joint is driven by an actuator (motor) via a system that consists of a gear train installed in the cage of the motor and a series of pulleys driven by cable wires [1]. If it is assumed that the gear ratio is 1:N, then the motor dynamics obey the following equation:

$$I_m \ddot{\theta}_m + B_m \dot{\theta}_m = \tau_m - N(\tau_f + \tau_c) \quad (17)$$

where θ_m is the motor shaft position relative to a reference frame, I_m is the (effective) inertia on the motor shaft, B_m is the coefficient of viscous friction (reflected to the motor axis), τ_m is the torque generated by the motor, τ_f is the frictional torque (sensed on the motor shaft), and τ_c is the reaction torque exerted on the output shaft from the cable wire system.

If the radius of the pulley attached to the motor shaft is r_m , then it follows [1] that

$$\tau_c = r_m (T_1 - T_1') \quad (18)$$

where T_1 and T_1' are the tension forces acting on the opposite sides of the pulley. Let the compliance coefficient of the cable (wire) between the pulleys number $(i-1)$ and (i) be denoted by c_i on one side and c_i' on the other side of the pulley. Moreover, let x_i and x_i' be the cable displacements at the pulley i , and t_i and t_i' are the initial tensions of the cable wires on the opposite sides of the pulley i . Thus, the following equations apply to the cable wire on the motor pulley:

$$Nr_m \theta_m - x_1 = c_1(T_1 - t_1) \quad (19)$$

$$x'_1 - Nr_m \theta_m = c'_1(T'_1 - t'_1) \quad (20)$$

If T_i and T'_i represent the tensions on the cable wire at the i th pulley, then

$$x_{i-1} - x_i = c_i(T_i - t_i) \quad (21)$$

$$x'_i - x'_{i-1} = c'_i(T'_i - t'_i) \quad (22)$$

where $i = 2, 3, \dots, n-1$. The n th pulley drives the n th joint shaft. If the radius of the n th pulley is r_n , then it follows that

$$x_n - r_n \theta_n = c_{n+1}(T_{n+1} - t_{n+1}) \quad (23)$$

$$r_n \theta_n - x'_n = c'_{n+1}(T'_{n+1} - t'_{n+1}) \quad (24)$$

Furthermore, if the joint shaft has a moment of inertia I_n , and its rotational angle is denoted by θ_n , then

$$I_n \ddot{\theta}_n = \tau_n - \tau_{fn} \quad (25)$$

where $\tau_n = r_n(T_{n+1} - T'_{n+1})$ is the moment causing the rotation of the joint shaft and τ_{fn} is the frictional moment opposing the motion. Equations (14) through (25) describe the force transmission through the cable wires, if the friction was negligible.

Suppose that the friction on the opposite sides at the i th pulley is denoted as f_{ci} and f'_{ci} . Then, the tensions (forces) for the cable wire between the i th and $(i+1)$ st pulleys are:

$$T_i - f_{ci} = T_{i+1} \quad (26)$$

$$T'_{i+1} + T'_i + f'_{ci} \quad (27)$$

where $i = 1, 2, \dots, n$. It will be assumed in the sequel that $f_{ci} = f'_{ci}$ and $c_i = c'_i$ for convenience.

In order to express τ_n in equation (25) in terms of motor torque τ_m and frictional forces, equations (26), (27) and (14) through (24) can be manipulated to obtain for the joint torque τ_n :

$$\begin{aligned} \tau_n &= r_n(T_{n+1} - T'_{n+1}) \\ &= 2r_n[Nr_m \theta_m - \sum_{k=1}^n C^k f_{ck} - r_n \theta_n]/C^{n+1} \end{aligned} \quad (28)$$

where $C^k = \sum_{j=1}^k c_j$ and $\theta_m = [\tau_m - N(\tau_f + \tau_c)]/a(s)$ with $a(s) = I_m s^2 + B_m s$ (s being the Laplacian operator). Thus, the n th joint torque τ_n can be explicitly written [1] as

$$\tau_n = 2r_n \{Nr_m [\tau_m - N(\tau_f + \tau_c)]/a(s) - \sum_{k=1}^n C^k f_{ck} - r_n \theta_n\}/C^{n+1} \quad (29)$$

Equation (29) relates the motor torque τ_m to the joint torque τ_n . If τ_f , τ_c and f_{ck} are negligible,

then

$$\tau_n = 2r_n [Nr_m \tau_m / a(s) - r_n \theta_n] / C^{n+1} \quad (30)$$

Equations (29) and (30) reveal that to determine the torque τ_n on the shaft of the nth joint, it is necessary to specify the transfer function of the motor (i.e., I_m and B_m , the ratio N of the motor gear, the radius (r_m) of the motor pulley and that (r_n) of the pulley on the nth joint, and the compliance coefficient ($c_i, c'_i, i = 1, 2, \dots, n+1$) of the cable (rope) wire, and friction (f_{cn}, f'_{cn}) between the pulley and the cable wire, if the friction cannot be neglected. It should be noted that the calculation of C^{n+1} requires that the number of pulleys between the motor shaft and the nth joint shaft must be known; the derivation of equations (29) and (30) assumes that each joint has one pulley and that there are no additional pulleys between any two adjacent joints.

The parameters appearing in equation (30) for the exoskeleton (Odeucs) are listed in Table 2. The parameters I_m, B_m, N and r_m for the motor driving joint i are also given.

TABLE 2*

Joint shaft number $n =$	1	2	3	4	5	6
Number m of pulleys betw. motor/joint n	n_1	n_2	n_3	n_4	n_5	n_6
Compliance coefficient c_n	c_1	c_2	c_3	c_4	c_5	c_6
Radius r_n of pulley n	r_1	r_2	r_3	r_4	r_5	r_6
Friction f_{cn}	f_{c1}	f_{c2}	f_{c3}	f_{c4}	f_{c5}	f_{c6}
Effective inertial I_m of motor for joint n	I_{m1}	I_{m2}	I_{m3}	I_{m4}	I_{m5}	I_{m6}
Effective frictional coeff. B_m of motor for joint n	B_{m1}	B_{m2}	B_{m3}	B_{m4}	B_{m5}	B_{m6}
Gear ratio N_m of motor for joint n	N_{m1}	N_{m2}	N_{m3}	N_{m4}	N_{m5}	N_{m6}
Pulley radius of motor for joint n	r_{m1}	r_{m2}	r_{m3}	r_{m4}	r_{m5}	r_{m6}

If the number of pulleys is different from the number of the joints, then the number (m) of the pulleys between the motor shaft and joint n must be substituted in equation (29) for the upper limit of the sum and C^{n+1} is to be replaced by C^{m+1} .

*This table is to be filled with numerical values when the final report of Odeucs on the exoskeleton will become available.

4. TELEOPERATED MANIPULATOR SYSTEM

A typical teleoperated manipulator system consists of a human operator who moves the exoskeleton, i.e., the leader (master) manipulator. This manipulator can be attached (mechanically) to the arm of the operator as an exoskeleton. The motion of the leader arm is transmitted through a communication network to a remotely located follower (slave) manipulator which is made to mimic the motion of the human arm, and to interact with the environment. The communication network transmits the motion information (for example, from the joint sensors) through the transmission medium to the remote site to the follower (slave) arm which is to perform a specified task.

The block diagram of a typical teleoperated manipulator system is shown in Figure 4. It is common that the velocity (position) command from the motion of the leader (master) is fed forward to the follower (slave) manipulator. The operation of the system ideally is such that (i) the position and velocity of the follower are the same as those of the human arm and that (ii) the leader senses the same forces as the follower whose end-effector touches the environment. Therefore, the human operator must receive sensory feedback from the follower (slave) which is interacting with the environment. For example, the forces/torques acting on the end-effector (gripper) of the follower arm when it manipulates objects or touches the environment are fed back to the leader (master) manipulator and thus, to the human operator.

4.1. A Two-Port Network Models a Teleoperated System

For the analysis, it has become common to model the subsystem consisting of the leader, the communication link and the follower (see Figure 4) as a two-port network. The input-output relations can then be expressed in terms of the velocities v_1, v_2 and generalized (forces and torques) forces F_1, F_2 where the subscript "1" refers to the leader manipulator and "2" to the follower manipulator. Thus, the following models are obtained:

$$\begin{bmatrix} F_1 \\ v_1 \end{bmatrix} = C \begin{bmatrix} F_2 \\ v_2 \end{bmatrix} \quad C = \text{chain matrix} \quad ; \quad \begin{bmatrix} F_1 \\ v_2 \end{bmatrix} = H \begin{bmatrix} v_1 \\ F_2 \end{bmatrix} \quad H = \text{hybrid matrix} ;$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = Y \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \quad Y = \text{admittance matrix} \quad ; \quad \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = Z \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad Z = \text{impedance matrix} ;$$

$$F - v = S(F + v) \quad S = \text{scattering matrix}$$

where $F = [F_1^T \ F_2^T]^T$ and $v = [v_1^T \ v_2^T]^T$. Thus, various choices are used to select the set of independent and dependent variables. For the linear time-invariant models, matrices C, H, Y, Z and S can be determined by the standard methods of the circuit theory.

It has become a common practice in the literature that the analysis is performed for the (scalar) force and velocity only. Thus, in the sequel, F_1, F_2, v_1 and v_2 are scalars. Therefore, matrices C, H, Y, Z and S are (2x2)-matrices. Since many teleoperated systems have manipulators with six DOF (degrees of freedom), the extension of the analysis to the six-dimensional

cases will render the foregoing matrices to be of size (12×12).

Example 1.

It is common in teleoperated systems to have velocity v_1 fed forward (forward flow) from the leader (master) to the follower (slave) and the force F_2 measured at the end-effector of the follower fed backwards to the leader [2]. A natural two-port model in this case is the one involving the hybrid matrix H:

$$\begin{bmatrix} F_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ F_2 \end{bmatrix} \quad (31)$$

where h_{ij} , $i,j = 1,2$ represents the (ij) element of H. For n joint manipulators, then h_{ij} represents a submatrix of dimension $n \times n$.

By assuming that the two-port model is described by a linear time invariant relation of equation (31), the elements h_{ij} of H, $i,j = 1,2$ can be determined as follows: $h_{11} = F_1/v_1 |_{F_2=0}$, $h_{12} = F_1/F_2 |_{v_1=0}$, $h_{21} = v_2/v_1 |_{F_2=0}$, and $h_{22} = v_2/F_2 |_{v_1=0}$. It should be noted that when $v_2 = v_2(v_1, F_2)$ and $F_1 = F_1(v_1, F_2)$ are nonlinear relations, then the quotients specifying h_{ij} should be replaced by appropriate partial derivatives.

By utilizing the mechanical impedance $Z = \text{force/velocity}$, equation (31) can be expressed as

$$\begin{bmatrix} F_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} Z_{in} & \text{scaling} \\ \text{scaling} & \text{factor} \\ \text{factor} & 1/Z_{out} \end{bmatrix} \begin{bmatrix} v_1 \\ F_2 \end{bmatrix} \quad (32)$$

where Z_{in} and Z_{out} are the mechanical input and output impedances of the two-port model, respectively.

A teleoperated system should behave in such a way that (i) the speed of the follower is directly proportional to the velocity of the leader, and (ii) the generalized force on the human arm is directly proportional to the generalized force sensed by the end-effector of the follower, i.e., $v_2 = k_1 v_1$ and $F_1 = k_2 F_2$ where k_1 and k_2 are constants. In the ideal case, $k_1 = 1$ and $k_2 = -1$. Thus, $v_1 = v_2$ and $F_2 = -F_1$, and the velocity and the generalized force are transmitted without any loss of information and delays through the communication media.

4.2. Common Architectures of Teleoperated Systems

Several architectures for controlling a teleoperated manipulator system have been proposed in the literature. As an example, Figure 5a shows schematically the velocity forward loop and the generalized force feedback.

The realization of the architecture in Figure 5a for multi-joint manipulators is shown in Figure 5b. It is noticed that the joint velocity vector of the leader is converted first to the Cartesian velocity vector by means of the master Jacobian matrix J_m . Then, the Cartesian velocity is

transmitted to the follower (slave) site, where it is converted to the joint velocity of the follower (slave) by using the inverse Jacobian matrix J_s^{-1} of the follower. Then, the actuators of the joints generate this additional velocity component. To feed back the force sensed by the end-effector of the follower, the measured generalized force ($e_{34} = F_2$) is first expressed in the wrist coordinate frame by means of matrix T_w , and then the result is transformed to the Cartesian base coordinate system by means of the rotation submatrix of the homogeneous transformation matrix $(A_6^0)^{-1}$. The result is communicated to the leader and converted to the joint torques by the transposed Jacobian matrix J_m^T . Thus, the leader will "feel" the remotely sensed generalized force.

Experimental studies on the teleoperated system shown in Figure 5b have demonstrated that the system is extremely sensitive to (even small) time delays. On the basis of the experiments, local control loops have been added to each (leader and follower) side [2]. The local control law which is built around an actuator on both sides enforces a desired impedance and circumvents the problems caused by the time delay.

4.3. Effect of Time Delay in Teleoperation

The signals transmitted through a communication media between the leader and follower usually exhibit time delays. Moreover, the processing of information in a computer and human reaction time can also result in additional time delays. The time delays may vary in the range of a few milliseconds to 6-10 seconds (or even more in space applications). A time delay in a control system such as in the teleoperated system is known to have a destabilizing effect in the system response, which is also demonstrated experimentally by laboratory experiments.

In order to compensate for the destabilizing effect of the time delay in the force reflection (feedback) loop, a local feedback loop at the follower's end [4,5] has been introduced. Another approach is to determine the control law based on the scattering theory that makes the communication link with the delay appear as a passive transmission line [5]. The resulting system is stable for passive loads at the leader and follower ports.

One of the most attractive approaches to handle problems caused by time delays is called shared compliant control [7]. In the shared compliant control, the control task is shared by the human operator (direct manual control) and the automatic compliant control of the follower manipulator. The compliant component is implemented in software as a damper-spring model for the contact. The force feedback loop resides entirely in the follower robot side, and the communication delay does not cause stability problems, because the force feedback is done locally. The motion of the end-effector of the follower will result in a soft touch of walls and objects; thus, it also facilitates the telemanipulation. It is this approach that is recommended for the implementation.

References

- [1] N. Imamura, M. Kaneko, K. Yokoi, and K. Tanie, "Development of a Two-fingered Robot Hand with Compliance Adjustment Capability," *1990 Japan-U.S.A. Symposium on Flexible Automation*, Kyoto, Japan, 1990.
- [2] B. Hannaford, "A Design Framework for Teleoperators with Kinesthetic Feedback," *IEEE Trans. on Robotics and Automation*, Vol. 5, No. 4, August 1989, pp. 426-434.
- [3] Y. Strassberg, A.A. Goldenberg and J.K. Mills, "A New Control Scheme for Bilateral Teleoperating Systems: Performance Evaluation and Comparison," *Proc. of the 1992 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, Raleigh, NC, July 1992, pp. 865-872.
- [4] S. Lee and H.S. Lee, "An Advanced Teleoperator Control System: Design and Evaluation," *Proc. of the 1992 IEEE Intl. Conference on Robotics and Automation*, Nice, France, May 1992, pp. 859-864.
- [5] R.J. Anderson and M.W. Spong, "Bilateral Control of Teleoperators with Time Delay," *IEEE Trans. on Automatic Control*, Vol. 34, No. 5, May 1989, pp. 494-501.
- [6] Y. Yokokohji and T. Yoshikawa, "Bilateral Control of Master-Slave Manipulators for Ideal Kinesthetic Coupling," *Proc. of the 1992 IEEE Intl. Conference on Robotics and Automation*, Nice, France, May 1992, pp. 849-858.
- [7] W.S. Kim, B. Hannaford, and A.K. Bejczy, "Shared Compliance Control for Time-Delayed Telemanipulation," *Proc. of the 1991 IEEE Intl. Conference on Robotics and Automation*, Cincinnati, OH, May, 1990.

A composite person-factor measure for selecting fighter pilots

W.F. LAWLESS
Assistant Professor
Departments of Mathematics and Psychology

Paine College
1235 15th Street
Augusta, GA 30910

Final Report For:
SUMMER FACULTY RESEARCH PROGRAM
Armstrong Laboratory/HRA

Sponsored by:
Human Resources Directorate, Aircrew Training Research Division
Williams AFB, AZ 85240-6457

Air Force Systems Command, Brooks AFB, TX

August 1992

A composite person-factor measure for selecting fighter pilots

W.F. LAWLESS
Assistant Professor
Departments of Mathematics and Psychology
Paine College

Abstract

The research activities completed this summer were designed to find the factors that would identify the top fighter pilot candidates for the United States Air Force. A comprehensive literature search was completed which led to the initial person factors of expertise and history. In parallel, working with Mike Houck, UDRI, and Wayne Waag, HRA, scales to measure fighter pilot situation awareness were crafted with a format redesigned for optical scanning data collection. Other scales written by the author included a wingman interdependence scale, a revision of a background scale, assistance with a peer evaluation scale, and a post-SA SimTest scale. A theory was constructed based on the literature review and the author's dissertation, the SAAC-engagements data base analyzed for the combined selection person factors, a chapter for a report written, an outline of a journal article planned, a proposal for the 1992 Fall SA SimTest engagements written, and a paper written to present to the Aerospace Medical Association. Additionally, initial plans were made to analyze a second data base for other articles, and later this Fall for the analysis of a third data base for the construction of an ANN enemy pilot.

A composite person-factor measure for selecting fighter pilots

W.F. LAWLESS

INTRODUCTION

One of the more research activities today in the human factors area with the Air Force is the attempt to find factors that would identify the top fighter pilot candidates. Selecting the best candidate is important for many reasons, not the least of which are savings in Air Force flight school training costs, operational fighter pilot proficiency training costs (check rides and tactical range sorties), aircraft safety and accident losses, and national defense. Choosing the wrong USAF candidates for basic flight training could account in part for its current 20-22% attrition rate; each 1% reduction in the flight training attrition rate would save \$1 million annually (references are cited in the forthcoming report chapter; a copy is available from the author). Graduating the wrong candidates could account in part for increased operational proficiency costs or for the Air Force accident rate; with 60% of all Air Force aircraft accidents due to pilot error, for fighter aircraft alone, these losses totaled more than \$1.125 billion between 1980-89 (comparable losses and pilot error rates were found in Naval aviation). And employing the wrong graduates in combat could cause a higher loss rate of aircraft and trained pilots and could be construed as a weakened national defense; e.g., the USAF fighter pilot kill ratio of 10:1 in Korea fell to 2.5:1 in Vietnam. If a composite measure of ACM performance established a profile to select the best fighter pilot candidates, as thought possible by Youngling and his colleagues (1977), training costs should be reduced,

safety rates improved, and national defense strengthened. In addition, selection and training theory could be advanced by becoming disentangled from the concept of universal pilot training, that is the belief that training skills can overwhelm person factors (e.g., from leadership studies, leadership training skills have not been found to predict leadership outcomes). Instead, scarce resources could be concentrated on training the most capable fighter pilot candidates.

Theory and results from the SAAC engagement data were drafted into a report chapter for publication. A paper based on the chapter has been submitted to the Aerospace Medical Association for presentation next year. And a journal article is planned for completion later this Fall. The following discussion provides details of some of the activities completed this summer.

Activities Review:

Activities initially proposed for this summer included the finalization of the Situational Awareness (SA) and Pilot Peer Review (PR) questionnaire forms. Both of these instruments had been tasked to the University of Dayton, principally M.R. Houck. Because of my research background in dynamic behavior, especially designing surveys and testing them in a dynamic environment, I assisted Houck in the design of the forms, plans for data collection and reduction, and inference analyses, the latter with linear and non-linear techniques. Where possible, data reduction will include a 4-dimensional presentation of the results (three parameter dimensions, and a fourth performance outcome dimension in color).

Summary Accomplishments:

1. Overall, the SAAC engagements data were analyzed to test for combined selection person factors that might identify the best fighter pilots for the USAF. Based on the an extensive literature search and the data results, a report chapter was written, a paper drafted for presentation at the ASM meeting next Spring, and a journal article planned for the journal Aviation, Space, and Environmental Medicine. In addition, as the next stage in this research program, for a set of simulated engagements (the SA SimTest) later this Fall, surveys were crafted and revised, and an analyses plan completed.
2. SA survey forms: the SA forms were field tested at Luke, AFB, revised and converted to an OPSCAN format. A trial data run with the opscan forms was performed at ASU and the data transmitted to the VAX at Williams AFB.
3. PR survey form: the Peer Evaluation form was completed and prepared for OPSCAN data collection.
4. Background data form: the background data form was completed and prepared for OPSCAN data collection.
5. Archival data, held by MPC, was studied as a source of information.
6. Arozona State University: Len Mitchel, University Testing Services. A contact was written with ASU to custom design OPSCAN data collection forms for AL/HRA. A data test run transmitted between ASU, HRL0T1, and Virginia Tech.

I. Summer Proposal: Finalize the SA global variable survey form.

Accomplishments:

1. The shape of the predictor-criterion parameters are converging to the use of four measures: BAT (a psychophysiological or psychomotor battery, comparable to performance measures on the WAIS test; one factor), AFOQT (cognitive measures, similar to the SAT and IQ tests, without psychomotor measures; five subfactors), SA, and PR. These four measures will be four of the predictor variables in a path analysis to predict the criterion of simulator outcomes. (BAT/AFOQT data collection and analyses were coordinated with Bill Alley, Maj. Dave Perry, and Rick Siem at 8-240-2244, at Brooks.)
2. A potential addition to the path analyses is individual pilot check ride grades and range scores. (Data collection information obtained from Bernie Edwards, Williams, AFB, for ATC; and with LtCol. Perry Mize at ACC.)

I. A. Summer Proposal: Include an item scale that would permit non-linear item-response theory analyses (e.g., LOGIST). Thus, Likert numbering to facilitate IRT should be even numbered, viz., either 6 or 8 points wide.

Accomplishments:

1. SA redesigned to a 6-unit Likert response field. PR and other forms coded for IRT analyses.

I. B. Summer Proposal: Data was to be collected by OPSCAN methods (i.e., optical sensitive scoring). Len Mitchel, Arizona State University, was to contract with HRAT to collect the data via optical scanning, and transmit the data to HRAT via e-mail (i.e., internet/bitnet).

Accomplishments:

1. One ream of 500 sheets of general NCS form #16482 (10 wide response field) OPSCAN forms ordered from NCS at 1-800-533-0518; FAX 507-451-1560.

2. Negotiated with Len Mitchell, Computer Support Supervisor, Arizona State University (602-965-7146; e-mail: ICLAM@ASUVM.INRE.ASU.EDU), to produce a custom designed optical sensitive form to meet HRAT specifications. Rates approximately \$35/hr to produce form, to collect data from subject completed forms, or to transmit the results electronically to the HRAT VAX account at Williams AFB. Reviewed with Wayne Waag and Mike Houck.

3. After the Luke AFB initial field test of the SA survey packet, it was decided to immediately produce the custom designed OpSCAN forms, and to skip the general NCS forms even during development. Negotiations begun and finalized with Mitchell at ASU on these forms. He estimated about 1 hr per page, or about 6 hrs for the initial run (approximately \$210).

4. Tested the ASU opscan designed forms on July 31, 1992. Test data given to ASU on opscans, scanned by ASU, and e-mailed to Lawless%HRLOT1.DECNET@EIS.BROOKS.AF.MIL. Reply e-mailed back to ASU received.

I. C. Summer Proposal: As possible, if a sufficiently sized sample pool was available ($n > 100$), scale design will use exploratory factor analyses (EFA) with oblique rotations, and should result in a reliability ALPHA of 0.80 or higher, preferably above 0.90; final scale confirmed with confirmatory factor analyses (e.g., LISREL or Proc Calis). Hierarchical multiple regressions and dummy variables were to be used to assess the continuous, bipolar variable SA scale as a predictor of pilot behavioral outcomes in simulators/actual flight. Results were to be displayed in 3-dimensions (SA by situations by time, with performance behavior in color as a fourth dimension).

Accomplishments:

1. Wayne Waag requested early field testing and data collection. The initial field test was held at Luke AFB on June 19, 1992 (cf. memo to M. Houck, 6/19/92, copy attached). Overall, the pilot subjects were engaged by the SA scale and found it to be a meaningful exercise; the pilots (three: Dave Greschke; Jet Jackson; and Bob Liotta; Bart Respotnik, Mike Houck, and Bill Lawless also attended) spent 4 hrs providing extensive feedback and commentary that led to a redesign of the survey forms and preparations to collect data.

II. Summer Proposal: Finalize the Pilot Peer Review form.

Accomplishments:

1. Pilot peer evaluation survey report (PR) completed and submitted to ASU for publication in an OPSCAN format.

II. A. Summer Proposal: Peer review should include all pilots in a squadron roster. Each evaluator will evaluate the squadron pilots and the pilots of his flight team.

Accomplishments:

1. First field test at Luke AFB 6/19/92 (cf. memo to M. Houck, 6/19/92, copy attached).

II. B. Summer Proposal: Decide whether to use an anonymous (evaluator not identified; evaluations available to command and to evaluated individuals) or confidential (evaluator identified; evaluations available to command and to evaluated individuals) evaluation approach, or a combination approach that would protect the evaluators and those evaluated.

Accomplishments:

1. Peer evaluations will be anonymous. IP SA evaluations will be identified and confidential to protect privacy.
2. All squadron pilots will also complete confidential self assessments on themselves with a background survey, SA, wingman relationships survey, and a post-SA SimTest survey.

II. C. Summer Proposal: Include in the peer evaluation items the evaluator's preferences to fly/not fly with each pilot evaluated, with the scale reference point set at mission readiness. From the literature, preferences would make the scale more dynamic (Snyder & Ickes, 1985, Handbook of Social Psychology).

Accomplishments:

1. Each pilot will peer evaluate all pilots in his squadron by rank ordering, and assess each pilot's general fighter pilot ability and SA ability.

III. Summer Proposal: The overarching goal was to assist in the derivation of global static predictors of pilot dynamic performance outcomes. To this end, BAT and psychomotor data were to be assessed along with SA and Peer Rating data to determine the best static measures of dynamic performance for simulated and actual cockpit outcomes.

Accomplishments:

1. Plans were made to include BATS and AFOQT data.
2. It was considered important to obtain pilots with air-to-air credits from Desert Storm/Desert Shield in the data as part of a baseline. A contact letter to be used by MPC at Randolph AFB was drafted and has been provided to Wayne Waag.
3. SAAC data was analyzed and the results were significant for the two implicit person factors of expertise and history. Theory development for the person factors was begun, a report chapter written, a paper proposed for presentation at the annual ASM meeting, and a journal article for ASEM was proposed. Follow-on work will include data analyses of the Fall, 1992, SA SimTest data, 30 engagement data from SAAC, and 500 human engagement opponent data.

Partial Reading List

Articles:

17. Blower, D.J. & Dolgin, D.L. (1992). Articles on Navy predictors of flight performance (8 references). Blower, D.J. & Dolgin, D.L. (1990). An evaluation of performance based tests designed to predict success in primary flight training. Proceedings of the Human Factors Society 34th Annual Meeting. pp. 949-53. Jensen, R.S. (1989). Aviation psychology. Brookfield, VT: Gower Publ. Dolgin, D.L., Shull, R.N., & Gibb, G.D. (1987). Risk assessment and the prediction of student pilot performance. Proceedings of the 4th International Symposium on Aviation Psychology, pp. 480-5. Hilton, T.F. & Dolgin, D.L. (1991). Pilot selection in the military of the free world. In R. Gal & A.D. Mangelsdorff (Eds.), Handbook of Military Psychology, pp. 81-101. Sussex, England: Wiley. Blower, D.J. & Dolgin, D.L. (1991). An evaluation of performance-based tests designed to improve naval aviation selection. NAMRL-1363. Pensacola, FL: Naval Aerospace Medical Research Laboratory. Shull, R.N., Dolgin, D.L., & Gibb, G.D. (1988). The relationship between flight training performance, a risk assessment test, and the Jenkins Activity Survey. NAMRL-1339. Pensacola, FL: Naval Aerospace Medical Research Laboratory. Dolgin, D.L. & Gibb, G.D. (1988). A review of personality measurement in aircrew selection. Pensacola, FL: Naval Aerospace Medical Research Laboratory.
19. Salas, Ed. (1992). Navy cockpit resource management (23 articles).

Stout, R.J., Cannon-Bowers, J.A., Salas, E., Morgan, Jr., B.B. (1990). Does crew coordination behavior impact performance? Paper presented at the annual meeting of the Human Factors Society, Orlando, FL. Franz, T.M., McCallum, G.A., Lewis, M.D., Prince, C., & Salas, E. (1990). Pilot briefings and aircrew coordination. Paper presented at the 12th Annual Department of Defense Symposium, Colorado Springs, CO. Franz, T.M., Prince, C., Cannon-Bowers, J.A., & Salas, E. (1990). The identification of aircrew. Paper presented at the 12th Annual Department of Defense Symposium, Colorado Springs, CO. Prince, C., Salas, E., & Franz, T. (1990). Aircrew coordination performance and skill development. Paper presented at the Symposium on Independent Research/Independent Exploratory Development, Baltimore, MD. Cannon-Bowers, J.A., Prince, C., Owens, J.M., Morgan, Jr., B.B., & Gonos, G.H. (1989). Determining aircrew coordination training effectiveness. Paper presented at the 11th annual meeting of the Interservice/Industry Training Systems Conference, Ft. Worth, TX. Hays, R.T., Jacobs, J.W., Prince, C., & Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. Military Psychology, 4, 67-74. Oser, R.L., Prince, C., & Morgan, Jr., B.B. (1990). Differences in aircrew communication content as a function of flight requirement: Implications for operational aircrew training. Paper presented at the 1990 meeting of the Human Factors Society, Orlando, FL. Lassiter, D.L., Vaughn, J.S., Smaltz, V.E., Morgan, Jr., B.B., & Salas, E. (1990). Paper presented at the 1990 Meeting of the Human Factors Society, Orlando, FL. Oser, R.L., Prince, C., Morgan, B.B., & Simpson, S.S. (1991). An analysis of aircrew communication patterns and content. Technical Report 90-009. Naval Training Systems Center, Human Factors Division,

Orlando, FL. Smith, K.A. & Salas, E. (1991). Training assertiveness: The importance of active participation. Paper presented at SEPA, New Orleans, LA. Prince, C. & Salas, E. (1991). The utility of low fidelity simulation for training aircrew coordination skills. Paper presented at the 2nd annual International Training Equipment Conference, Maastricht, Netherlands. Bowers, C.A. & Salas, E. (1991). The assessment of coordination demand for helicopter flight requirements. Paper presented at the 6th International Symposium on Aviation Psychology, Columbus, OH. Prince, C., Oser, R., Salas, E., & Shrestha, L. (1992). Paper presented at the International Training Equipment Conference, Luxembourg. Baker, D.P., Salas, E., & Prince, C. (1991). Team task importance: Implications for conducting team task analysis. Paper presented at the 6th meeting Society for Industrial and Organizational Psychology, St. Louis, MO. Baker, D.P., Bauman, M., & Zalesny, M.D. (1991). Development of aircrew coordination exercises to facilitate training transfer. Paper presented at the 6th International Symposium on Aviation Psychology, Columbus, OH. Urban, J.M., Franz, T.M., & Morgan, Jr., B.B. (1991). A comparison of behavioral observations and subjective evaluations in the assessment of aircrew communication. Paper presented at SEPA, New Orleans, LA. Hartel, C.E.J., Smith, K., Prince, C. (1991). Defining aircrew coordination: Searching mishaps for meaning. Paper presented at the 6th International Symposium on Aviation Psychology, Columbus, OH. Baker, D.P. & Salas, E. (in press). Principles for measuring teamwork skills. Human Factors. Morgan, Jr., B.B., Herschler, D.A., Wiener, E.L., & Salas, E. (1992). Implications of automation technology for aircrew coordination and performance. In W.B. Rouse (Ed.), Advances in Man-Machine Systems Research,

6. Bowers, C., Salas, E., Prince, C., & Brannick, M. (in press). Games teams play: A methodology for investigating team coordination and performance. Behavior Research Methods, Instruments and Computers. Swezey, R.W., Llaneras, R.E., Sterling, V., Prince, C., & Salas, E. (1991). Instructional strategy for aircrew coordination training. Paper presented at the 6th Biannual Symposium of Aviation Psychology, Columbus, OH. Prince, C., Cannon-Bowers, J., Salas, E., Owens, J., & Morgan, Jr., B.B. (1989). Aircrew coordination training: Requirements and challenges. An unpublished manuscript. Prince, C. & Salas, E. (1989). Aircrew performance coordination and skill development (Tech Report 89-009). Prince, C., Chidester, T.R., Cannon-Bowers, J., & Bowers, C. (1992). Aircrew coordination--achieving teamwork in the cockpit. In R. Swezey & E. Salas (Eds.), Teams: Their training and performance. New York: Ablex.
20. Freeman, J. LtCol. (1992). Safety reports. HQ AFSA/SEL, Norton AFB, CA, 92409-7001.

Books:

1. Jackson, R. (1986). Flying modern jet fighters. Wellingborough: Patrick Stevens.
2. Hall, G. (1987). Top gun. The Navy's fighter weapons school. Novato, CA. Presidio Press.
3. Trotti, J. (1984). Phantom over Vietnam. Fighter pilot, USMC.

Novato, CA. Presidio Press.

4. Clausewitz, C.V. (1976). On war. M. Howard & P. Paret (Eds./Trs.).
Princeton: Princeton University Press.

LOCAL ADMINISTRATION OF EXCITATORY AMINO ACID ANTAGONISTS
ATTENUATES LIGHT-INDUCED PHASE SHIFTS AND C-FOS EXPRESSION
IN THE HAMSTER SUPRACHIASMATIC NUCLEI

L. M. Lutton(SFRP Fellow)*, B. Buckley*, and M. A. Rea**

*Associate Professor
Department of Biology
Mercyhurst College
Erie, PA 16546

**Circadian Neurobiology Research Group
Armstrong Laboratory (AL/CFTO)
Brooks AFB, TX 78235

Final Report for:
AFOSR Summer Research Program
Armstrong Laboratory
Brooks AFB, San Antonio, TX

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

LOCAL ADMINISTRATION OF EXCITATORY AMINO ACID ANTAGONISTS
ATTENUATES LIGHT-INDUCED PHASE SHIFTS AND C-FOS EXPRESSION
IN THE HAMSTER SUPRACHIASMATIC NUCLEI

L. M. Lutton*, B. Buckley and M. A. Rea
Associate Professor
Department of Biology
Mercyhurst College

Abstract

The effects of local administration of excitatory amino acid antagonists on light-induced phase shifts of the circadian activity rhythm and on light-induced c-fos expression in the SCN were determined. Syrian hamsters, stereotaxically fitted with guide cannulas, received 300 nl injections of CNQX, MK-801 or vehicle in the SCN area 5 min prior to a 10 min, 20 lux light exposure. Injection of either 1mM CNQX or MK-801 resulted in a 77% reversible inhibition of light-induced phase advances (CNQX=12±7 min; MK801=12±5 min; veh=52±9 min; p<0.05). The effect of CNQX was dose related, while MK-801 was equally effective at the lowest dose tested (0.01mM). 1 mM MK-801 at CT13 resulted in a 71% inhibition of light-induced phase delays (vehicle=-51±6 min; MK-801=-15±5 min; p<0.05), but 1 mM CNQX failed to significantly do so (-39±10 min). Injection of MK-801 alone was without effect at CT18 or CT13, but CNQX alone at CT13, but not CT18, appeared to cause a significant phase delay (-20±6 min).

Light exposure at CT18 resulted in a characteristic c-fos expression in the SCN. Local administration of 1 mM CNQX and MK-801 reduced the number of Fos-immunoreactive cells by about 32% and 44%, respectively: (CNQX=643±135; MK-801=533±143; vehicle=951±79; p<0.05). These results indicate that both NMDA and non-NMDA receptors in the SCN mediate light-induced phase shifts and Fos expression and they provide evidence that this Fos expression represents a necessary event in the photic entrainment of the SCN circadian pacemaker.

LOCAL ADMINISTRATION OF EXCITATORY AMINO ACID ANTAGONISTS
ATTENUATES LIGHT-INDUCED PHASE SHIFTS AND *C-FOS* EXPRESSION
IN THE HAMSTER SUPRACHIASMATIC NUCLEI

L. M. Lutton*, B. Buckley and M. A. Rea

Introduction

The suprachiasmatic nuclei (SCN) are the site of the light entrainable circadian pacemaker responsible for the generation of circadian rhythmicity in mammals (Rusak and Zucker, 1979; Meijer and Reitveld, 1989). SCN driven rhythms, which persist under constant conditions, are normally entrained to the environmental light-dark (LD) cycle. Potentially entraining photic information is conveyed to the SCN by at least two pathways a monosynaptic projection from retinal ganglion cells (Moore and Lenn, 1971; Pickard, 1982; Johnson et al., 1988b), the retinohypothalamic tract (RHT), and an indirect retinal projection through the ventral lateral geniculate nucleus/ intergeniculate leaflet, the geniculo-hypothalamic tract (GHT) (Card and Moore, 1982). While the GHT appears to modulate the response of the SCN to light (Harrington, 1986; Pickard et al., 1987), the RHT projection is necessary and sufficient for photic entrainment of the SCN pacemaker (Johnson et al., 1988a).

Photic entrainment of mammalian circadian rhythms occurs as a consequence of the phase-specific effects of environmental light on the circadian pacemaker as defined by the phase response curve to light "pulses" presented to organisms in constant darkness (DeCoursey, 1964; Takahashi et al., 1983). Light exposure during the early subjective night causes phase delays in pacemaker driven rhythms, and phase advances in the later subjective night. Light pulses during the subjective day do not cause phase shifts. Light exposure during the subjective night also induces the expression of *c-fos* related (*Fos*) genes in a population of SCN cells (Rea, 1989; Aronin et al., 1990; Kornhauser

et al., 1990; Rusak et al., 1990; Earnest et al., 1990). Furthermore, light does not stimulate *c-fos* expression during the subjective day (Kornhauser et al., 1990; Rea, 1992), suggesting that *c-fos* expression may represent an early event in the light-induced phase alterations of the pacemaker.

Peripheral and intracerebroventricular (icv) injection of excitatory amino acid (EAA) antagonists has been reported to attenuate light-induced phase shifts of the free-running activity rhythm (Colwell et al., 1990; 1991) and to suppress Fos expression in the SCN (Abe et al., 1991; 1992; Ebling et al., 1991). However, these modes of administration result in the widespread distribution of drugs, precluding the precise localization of their effects. Therefore, we determined the effects of direct application of EAA antagonists into the suprachiasmatic hypothalamus on light-induced phase advances and delays of the free-running activity rhythm in hamsters, and light-induced Fos expression in the hamster SCN.

Methods

Animals. Male, Syrian hamsters (*Mesocricetus auratus*; Charles River, Wilmington, MA) were housed in groups of 6 and maintained under LD 14:10 for at least 2 weeks prior to experimentation. Under LD 14:10, cage level illuminance was approximately 200 lux. Animals received food and water *ad libitum*.

Surgical Procedure. Hamsters weighed 125 - 150 g at the time of surgery. Under pentobarbital (90 mg/kg), 26 gauge cannula guides containing 33 gauge stylets (Plastic Products, Roanoke, VA) were stereotaxically implanted to a depth of 2.9 mm below dura and fixed to the skull with fine machine screws and dental cement. Infusion cannulas (33 gauge) delivered EAA antagonists in the area of the SCN. Stereotaxic coordinates were 1.0 mm rostral and 1.5 mm late-

ral to bregma at a 10° angle from vertical; upper incisor bar at 0. Fascia and skin were sutured and the animals were allowed to recover for 1 week under LD 14:10. Animals received post-operative analgesia (0.05 ml buprenex, i.m.) at 2 and 12 hr after surgery and were carefully monitored during recovery. *Activity Rhythms.* After 1 week in LD 14:10, cannulated hamsters were transferred to individual cages equipped with 14 inch running wheels (Columbus Instruments, Columbus, OH) and maintained in constant darkness (DD). Wheel running activity was monitored continuously using a Zenith 248 computer with Dataquest III data acquisition software (Minimitter Co., Inc, Sunriver, OR). Wheel running records (actograms) were generated and analyzed using *Circadia* software (Behavioral Cybernetics, Cambridge, MA) with a MacIntosh IICx computer.

The onset of wheel turning activity, designated as circadian time (CT) 12, was used as a phase reference point for the timing of photic stimulation (Daan and Pittendrigh, 1976). Activity onset was defined as the first 6 minute interval that: (a) exceeded 10% of the maximum rate for the day, (b) was preceded by 4 hours of inactivity, and (c) was followed by 1 hour of sustained activity. The period of the free running rhythm (τ) was calculated as the average amount of time between activity onsets. CT12 on the day of stimulation was predicted by the extrapolation of the least squares line of the activity onsets of the five days preceding that day.

Drug Injection Procedure. A 33 gauge infusion cannula attached to a 1 μ l Hamilton syringe (Albers et al., 1984) was used for local administration of EAA antagonists (300 nl). The cannula extended 4.4 mm beyond the tip of the cannula guide to a position just dorsal to the SCN (Fig 1). Drugs (0.01-1.0 mM) were prepared in a bicarbonate-buffered physiological saline (aCSF): NaCl, 122mM; KCl, 3.8mM; MgSO₄, 1.2mM; KH₂PO₄, 1.2mM; NaHCO₃, 25mM; and CaCl₂, 2.5mM.

All drug solutions were prepared the day of injection.

Drug Diffusion Experiments. The autoradiographic procedure of Talman et al. (1991) was used to determine the diffusion pattern of [^3H] CNQX after injection into the SCN region. Using our standard procedure 300 nl of 1 mM CNQX was injected. Hamsters were sacrificed 5 min after injection. Brains were rapidly removed, cut in the coronal plane on either side of the hypothalamus and frozen on dry ice. Frozen, 20 μm coronal sections were cut on a cryostat, pressed onto cold (-20°C) subbed microscope slides, and stored in a desiccator at -20°C overnight. Dried sections were fixed for 15 min in formaldehyde vapor, placed in film cassettes against ^3H sensitive film (Hyperfilm- ^3H ; Amersham, Arlington Heights, IL) and stored at 4°C for 5 days. Film was developed and dimensions of the exposed regions were measured with a micrometer.

Light-Induced Phase Shifts. After at least 10 days in DD, hamsters received injections followed by light stimulation at CT13.5 or CT18.5. These 300 nl injections of vehicle, CNQX, or MK-801 were performed under dim red illumination (< 0.5 lux). Animals were gently restrained for approximately 30 seconds during the injection, and the infusion cannula remained in place for 10 - 15 seconds after the injection. Five min after injection, hamsters were transferred to the light stimulation chamber, a 12 cm x 20 cm white metal cylinder covered with a sheet of frosted glass. Each animal received 10 minutes of white light at an average illuminance of 20 lux. The light stimulation apparatus was described previously (Rea, 1989). Light intensity was determined with a Tektronix J16 digital photometer and a J6511 illuminance probe. After stimulation animals were returned to their wheel cages in DD for 8-10 days.

Quantification of Phase Shifts. Phase shifts were calculated using *Circadia* according to the procedures of Daan and Pittendrigh (1976). After data col-

lection, animals were sacrificed, and the brains were removed and fixed in 10% buffered formalin. The location of the injection site was verified histologically by examining 50 μ m vibratome sections of the SCN.

Light-Induced Fos Expression. Cannulated hamsters were housed in LD 14:10 for at least 7 days after surgery. On the day before treatment, hamsters were transferred to constant darkness. After 30-32 hrs in constant darkness (mid-subjective night), they received 300 nl injections of, 1mM CNQX, or 1 mM MK-801 into the SCN area. Five min after injection, each hamster was exposed to 20 lux of white light as described above and then returned to darkness.

Fos Immunocytochemistry. Hamsters received 90 mg/kg pentobarbital i.p. 2 hrs after light stimulation and were perfused transcardially with 100 ml heparinized phosphate buffered saline (pH 7.4), followed by 100 ml of 4% paraformaldehyde in Na_3PO_4 buffer (pH 7.4). Brains were removed and postfixed in 4% paraformaldehyde overnight (4°C), followed by 0.1 M Na_3PO_4 buffer (pH 7.4; 4°C; 24 hrs). Frontal sections (70 μ m) cut on a vibratome were incubated at 4°C overnight in Fos antiserum (1:5000; provided by Dr M.Iadarola, NIDR). The antiserum was raised against synthetic peptide corresponding to residues 128-153 of the M-peptide region of rat c-fos (Curran et al.,1987). It recognizes several Fos-related antigens (Iadarola et al.,1989). Fos immunoreactivity (Fos-ir) was detected with a Vectastain ABC kit (Vector Labs, Burlingame, CA) (Rea,1989).

Statistical Analysis. Statistical significance between means was determined using the two-tailed Student's "t" test.

Materials. Paraformaldehyde and all salts were obtained ACS grade from Sigma Chemical Co., (St. Louis, MO), 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and (5R,10S)-(+)-5-methyl-10,11-dihydro-5H-dibenzo[a,b]cyclohepten-5,10-imine hydrogen maleate (MK-801) from Research Biochemicals, (Natick, MA) and [^3H]-CNQX (17 Ci/mmol) from NEN-Dupont, (Wilmington, DE).

Results

Drug Diffusion. Autoradiographic examination of the distribution of [^3H]-CNQX revealed that 5 minutes after injection, the drug occupied a spheroid of approximately 1.2 ± 0.2 mm in diameter ($n=3$). Evidence of diffusion of the drug into the lateral ventricle was apparent in 1 animal. Based on these observations, only data from animals with injection sites located 0.5 mm from the anatomical center of the SCN and without disruption of the ependymal cell layer were accepted.



Fig 1. Ideal injection site. Cresyl violet stained frontal section of hamster brain at the level of the SCN. Injection tract is visible descending from the upper right to a site just dorsal to the SCN. (25X)

Light-Induced Phase Shifts of the Free-Running Activity Rhythm. Light exposure at CT18 consistently resulted in stable phase advances (81 ± 8 min) of the free-running activity rhythms (Fig 2,4). Injection of vehicle 5 min prior to light exposure inhibited the phase response by 36% (52 ± 9 min; $p < 0.05$). Light exposure at CT13 resulted in stable phase delays of approximately -41 ± 3 min (Fig 5,6). Vehicle injections were without effect (-51 ± 6 min).

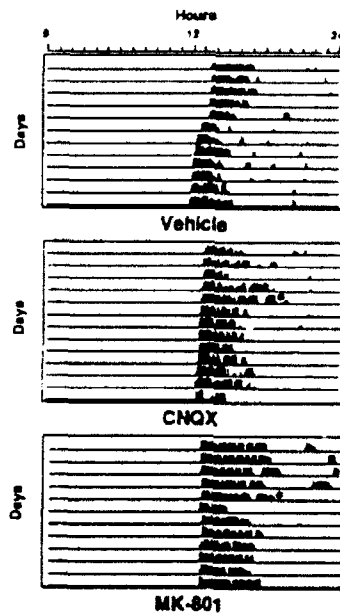


Fig. 2. EAA antagonists attenuate light-induced phase advances. Representative actograms demonstrating the effect of 300 nl of vehicle (top panel), 1 mM CNQX (middle panel) or 1 mM MK-801 into the SCN area prior to a 10 min light exposure at the times indicated by the asterisks (*).

Effects of EAA Antagonists on Light-Induced Phase Advances. Local administration of CNQX and MK-801 blocked light-induced phase advances of the free-running activity rhythm (Fig 2). The blockage was reversible (Fig 3), indicating that the effects of local injection of the drugs were not due to RHT damage. Injection of 1mM CNQX resulted in a 77% inhibition of these phase advances (12 ± 7 min; $p < 0.05$), was dose related and, alone, did not alter the phase of the activity rhythm (Fig 4). Injection of 1 mM MK-801 also inhibited these phase advances (12 ± 5 min). It was effective at a concentration as low as 0.01 mM and, alone, did not alter the phase of the activity rhythm (Fig 4).

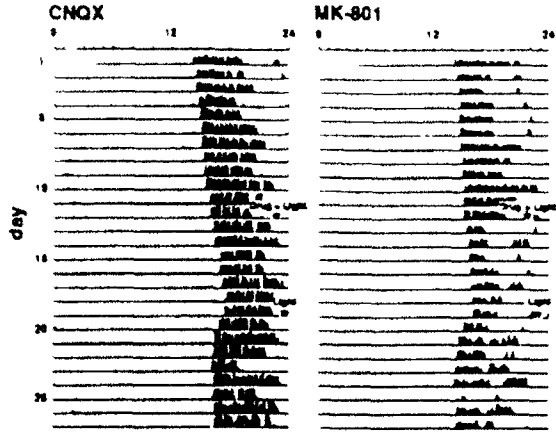


Fig. 3. Effects of EAA antagonists are reversible. Actograms show that injection of 1 mM CNQX (left panel) or 1 mM MK-801 (right panel) five minutes prior to a 10 min light exposure at the time of the asterisk (*) on day 12 prevented the phase advance, while a subsequent exposure to 10 min of light at the same CT on day 19 resulted in a phase advance.

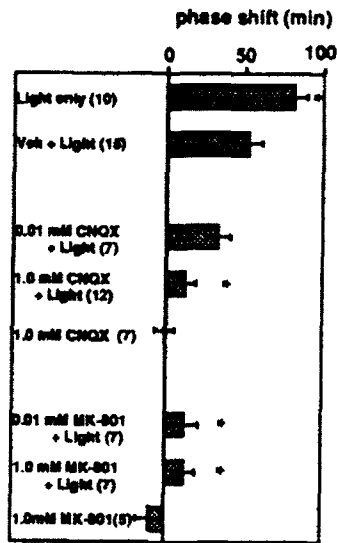


Fig. 4. Effect of local administration of EAA antagonists on light-induced phase advances. Data represent the MEAN \pm SD. The number of animals indicated in parentheses. Asterisks (*) indicate statistically significant differences ($p < 0.05$) relative to the Veh + light group.

Effects of EAA Antagonists on Light-Induced Phase Delays. Local administration of 1 mM MK-801 resulted in a 71% inhibition of light-induced phase delays (vehicle = -51 ± 6 min; MK-801 = -15 ± 5 min; $p < 0.05$; Fig 5,6). Alone it was without effect. 1 mM CNQX failed to significantly inhibit these phase delays (-39 ± 10 min; Fig 5,6) but, alone, caused a significant delay (-20 ± 6 min).

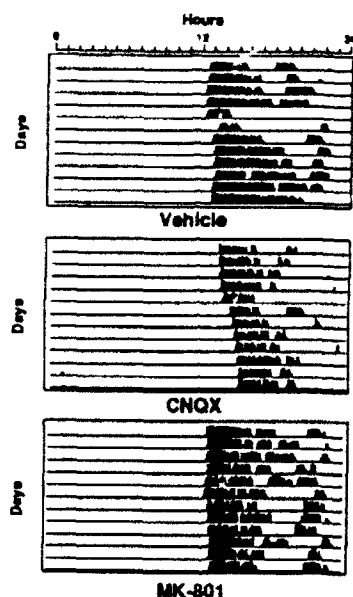


Fig. 5. MK-801 attenuates light-induced phase delays. Representative actograms demonstrating the effect of 300 nl of vehicle (top panel), 1 mM CNQX (middle panel), or 1 mM MK-801 into the SCN region 5 min prior to a 10 min light exposure at the times indicated by the asterisks (*).

Effects of EAA Antagonists on Light-Induced Fos Expression. Light exposure at CT18 resulted in a characteristic pattern of Fos expression among cells in the suprachiasmatic hypothalamus (Fig 7). Fos-ir cells were observed within the SCN, predominately localized to the ventrolateral third of the nucleus. In addition, many Fos-ir cells were observed in the region dorsal to the SCN, including the periventricular area.

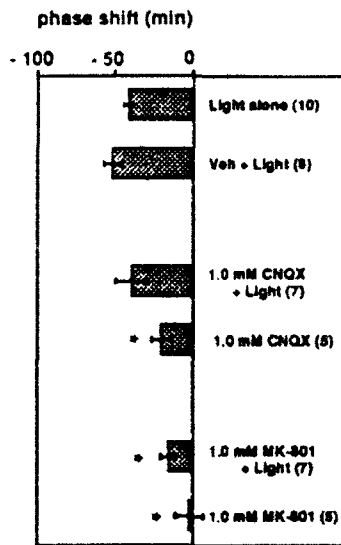


Fig. 6. Effect of local administration of EAA antagonists on light-induced phase delays. Data represent the MEAN \pm SD. The number of animals indicated in parentheses. Asterisks (*) indicate statistically significant differences ($p < 0.05$) relative to the aCSF + light group.

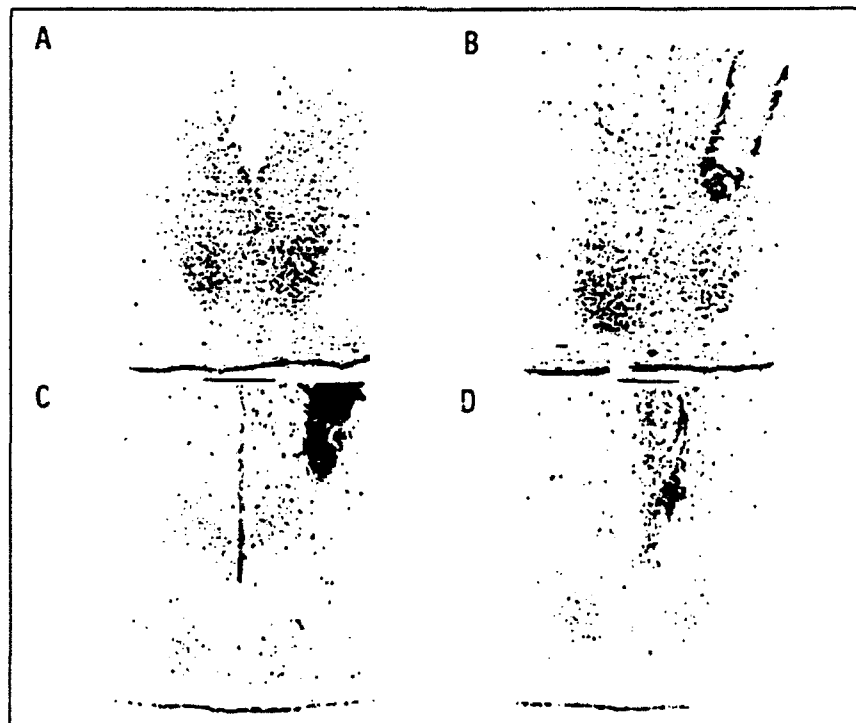


Fig. 7. Photomicrographs of representative frontal sections from the caudal third of SCN immunostained for Fos proteins. (A): light alone. (B): vehicle. (C): 1 mM CNQX. (D): 1 mM MK-801. (25X)

Injection of aCSF did not alter the pattern (Fig 7) or density (Fig 8) of Fos-ir in the SCN area (light = 974 ± 99 cell nuclei; vehicle = 951 ± 79 cell nuclei) except along the injection tract. Introduction of 1 mM CNQX or MK-801 into the SCN region reduced the intensity of immunostaining (Fig 7) and the number of Fos-ir cells present after light exposure (Fig 8). CNQX and MK-801 reduced Fos-ir cell nuclei relative to vehicle injected controls by about 32% and 44%, respectively (CNQX = 643 ± 135 , $p < 0.05$; MK-801 = 533 ± 143 , $p < 0.05$).

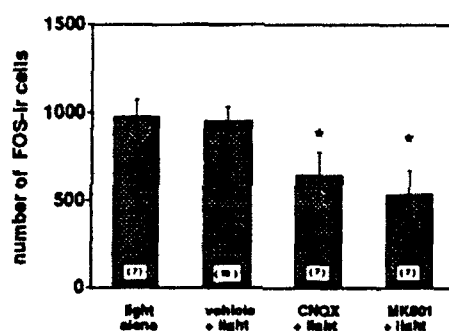


Fig. 8. EAA antagonists attenuate light-induced Fos expression in the SCN. Data represent the MEAN \pm SD. The number of subjects indicated in parentheses. Asterisks (*) indicate statistically significant differences ($p < 0.05$) relative to the vehicle + light group.

Discussion

Considerable data suggest that RHT neurotransmission is mediated by the release of EAAs. EAA antagonists dose-dependently attenuate optic nerve stimulated excitatory postsynaptic potentials (Kim and Dudek, 1991) and field potential responses (Shibata et al, 1986; Cahill and Menaker, 1986; 1987) in the SCN in vitro. Furthermore, Liou and coworkers (1986) reported that elec-

trical stimulation of the optic nerve resulted in EAA release from a hypothalamic slice preparation. Finally, neurophysiological evidence suggests that EAAs mediate fast excitatory neurotransmission at the retinofugal synapses in the lateral geniculate nucleus (Kemp and Sillito, 1982; Crunelli et al., 1987).

Consistent with this hypothesis, several laboratories have reported that peripheral or icv administration of EAA antagonists attenuates light-induced phase shifts of the free-running activity rhythm (Colwell et al., 1990; 1991) and c-fos expression (Abe et al., 1991, 1992; Ebling et al., 1991) in the hamster SCN. However, these modes of administration preclude the localization of drug effects to the SCN. In the present study, care was taken to insure that EAA antagonist administration was localized to the region of the SCN. The diffusion of [³H] CNQX 5 min after injection into the SCN (time to light onset in phase shift experiments) was determined using autoradiographic techniques. These injections resulted in a pattern of drug distribution limited to an area approximately the size of the SCN. Only animals with injection sites within 0.5 mm of the anatomical center of the SCN and without evidence of damage to the third ventricle were accepted for data analysis. As a result, we were able to demonstrate that local administration of EAA antagonists attenuates both light-induced phase shifts and Fos expression in the SCN.

Both CNQX and MK-801 reversibly blocked light-induced phase advances (Figs 2,3) of free-running activity rhythms. Inhibition by CNQX was dose related (Fig 4) while MK-801 was effective at the lowest dose tested (0.01 mM). Neither drug advanced the pacemaker when applied alone. These results are consistent with previously published data involving systemic or icv injections of EAA antagonists (Colwell et al., 1990; 1991) and suggest that both NMDA- and non-NMDA receptors mediate the phase advancing effects of light on the SCN.

The pharmacology of light-induced phase delays appears to differ from that of the advances. While local infusion of MK-801 significantly attenuated light-induced phase delays (Figs 5,6), CNQX appeared to be ineffective. This finding disagrees with Colwell et al.(1991) that icv injection of DNQX, a non-NMDA antagonist, DNQX, blocked both light-induced phase advances and delays. The reason for this discrepancy is unclear, but may be methodological. The inhibitory effect of DNQX on phase delays observed by Colwell et al.(1991) may involve a site of action outside the SCN. EAAs appear to be the principal fast excitatory neurotransmitter in the hypothalamus (van den Pol and Dudek, 1991) and electrical stimulation of the surrounding hypothalamus has been shown to elicit EAA-dependent EPSPs in SCN neurons (Kim and Dudek, 1991). Alternatively, the apparent failure of CNQX to block light-induced phase delays in the present study may be due to a phase delaying effect seen when injected alone (Fig 6). An appealing interpretation of the current results is that neurochemical mechanisms underlying light-induced advances and delays are different. Ralph and Menaker (1985;1986) report that GABA system drugs differentially alter light-induced phase advances and delays of hamster activity rhythms. The existence of distinct mechanisms underlying delays and advances is also supported by the observation that different populations of SCN cells express Fos in association with phase delays and advances (Rea, 1992).

In the present study, brief light exposure during mid-subjective night resulted in a characteristic pattern of Fos-ir in the hamster SCN (Rusak, et al., 1990; Ebling, et al., 1991; Rea, 1992). Injection of CNQX or MK-801 5 min prior to light exposure significantly reduced the number and intensity of Fos-ir cell nuclei (Figs 7,8). This observation is in agreement with the recent report of Abe et al (1992) that icv injection of NMDA and non-NMDA anta-

gonists attenuated light-induced Fos expression in the SCN. Moreover, these investigators observed that a small population of Fos-ir cells in the dorsolateral SCN were resistant to EAA antagonists, suggesting that other neurotransmitter systems might mediate light-induced Fos expression in these cells. Similarly, Ebling et al. (1991) reported that systemic MK-801 failed to inhibit Fos expression in this area of the SCN in response to icv injections of NMDA. In the present study, we also observed light-induced Fos expression in the dorsolateral SCN that was resistant to EAA antagonists. Unfortunately, variability in the pattern of diffusion of drugs after intra-SCN administration did not permit a precise localization of this cell population.

The observation that doses of CNQX and MK-801, which inhibited light-induced phase advances, diminished Fos expression throughout the ventrolateral region of the SCN is consistent with the suggestion that NMDA- and non-NMDA EAA receptors may reside on the same cells. (Abe et al., 1992). Indeed, Kim and Dudek (1991) have reported that optic nerve stimulation-induced EPSPs in SCN neurons contain both NMDA- and non-NMDA receptor-mediated components. Furthermore, the observation that light-induced phase shifts and Fos expression show similar pharmacology provides further evidence that Fos expression plays an intimate role in the biochemical mechanisms responsible for the photic entrainment of the SCN circadian oscillator.

References

- Abe, H.; Rusak, B.; Robertson, H.A. Photic induction of Fos protein in the suprachiasmatic nucleus is inhibited by the NMDA receptor antagonist MK-801. *Neurosci. Lett.* 127: 9-12; 1991.
- Abe, H; Rusak, B; Robertson, H.A. NMDA and non-NMDA receptor antagonists

- inhibit photic induction of Fos protein in the hamster suprachiasmatic nucleus. *Brain Res. Bull.* 28:831-835; 1992
- Albers, H.E.; Ferris, C.F.; Leeman, S.E.; Goldman, B.D.; Avian pancreatic polypeptide phase shifts hamster circadian rhythms when microinjected into the suprachiasmatic region. *Science* 223: 833-835; 1984.
- Aronin, N.; Sagar, S.M.; Sharp, F.R.; Schwartz, W.J. Light regulates expression of a Fos-related protein in rat suprachiasmatic nuclei. *Proc. Nat. Acad. Sci. USA* 87:5959-5962; 1990.
- Card, J.P.; Moore, R.Y. Ventral lateral geniculate nucleus efferents to the rat suprachiasmatic nucleus exhibit avian pancreatic polypeptide-like immunoreactivity. *J. Comp. Physiol.* 206: 390-396; 1982.
- Cahill, G.M.; Menaker, M. Kynurenic acid blocks suprachiasmatic nucleus responses to optic nerve stimulation. *Brain Res.* 410: 125-129; 1987.
- Cahill, G.M.; Menaker, M. Effects of excitatory amino acid receptor antagonists and agonists on suprachiasmatic nucleus responses to retinohypothalamic tract volleys. *Brain Res.* 479: 76-82; 1989.
- Colwell, C.S.; Ralph, M.R.; Menaker, M. Do NMDA receptors mediate the effects of light on circadian behavior? *Brain Res.* 523:117-120, 1990.
- Colwell, C.S.; Foster, R.G.; Menaker, M. NMDA receptor antagonists block effects of light on circadian behavior in the mouse. *Brain Res* 554:105-110;1991
- Crunelli, V.; Kelly, J.S.; Leresche, N.; Pirchio, M. On the excitatory postsynaptic potential evoked by stimulation of the optic nerve in the rat lateral geniculate nucleus. *J. Physiol.* 384: 603-618; 1987.
- Curran, T.; Gordon, M.B.; Rubino, K.L.; Sambucetti, L.C. Isolation and characterization of the c-fos(rat) cDNA and analysis of post-translational modification in vitro. *Oncogene* 2:79-84; 1987.

- Daan S.; Pittendrigh, C.S. A functional analysis of circadian pacemakers in nocturnal rodents: II. The variability of phase response curves. *J. Comp. Physiol.* 106:253-266; 1976.
- DeCoursey, P.J. Function of a light response rhythm in hamsters. *J. Cell. Comp. Physiol.* 63: 189-196; 1964.
- Earnest, D.J.; Iadarola, M.; Yeh, H.H.; Olschowka, J.A. Photic regulation of c-fos expression in neuronal components governing the entrainment of circadian rhythms. *Exp. Neurol.* 109: 353-361; 1990.
- Ebling, F.J.P.; Maywood, E.S.; Staley, K.; Humby, T.; Hancock, D.C.; Waters, C.M.; Evan, G.I.; Hastings, M.H. The role of N-methyl-D-aspartate-type glutamatergic neurotransmission in the photic induction of immediate-early gene expression in the suprachiasmatic nuclei of the Syrian hamster. *J. Neurophysiol.* 3:641-652, 1991.
- Harrington, M.; Rusak, B. Lesions of the thalamic intergeniculate leaflet alter hamster circadian rhythms. *J. Biol. Rhythms* 1: 309-325; 1986.
- Iadarola, M.J.; Yeung, C.I.; Draisci, G.; Ruda, M.A. Prolonged elevation of multiple Fos immunoreactive proteins in rat spinal cord: Western blot and immunocytochemical analysis. *Soc. Neurosci. Abstr.* 15:468; 1989.
- Johnson, R.F.; Moore, R.Y.; Morin, L.P. Loss of entrainment and anatomical plasticity after lesions of the hamster retinohypothalamic tract. *Brain Res.* 460:297-313; 1988a.
- Johnson, R.F.; Morin, L.P.; Moore, R.Y. Retinohypothalamic projections in the hamster and rat demonstrated using cholera toxin. *Brain Res* 462:301-312;1988b
- Kemp, J.A.; Sillito, A.M. The nature of the excitatory transmitter mediating X and Y cell inputs to the cat dorsal lateral geniculate nucleus. *J. Physiol.* 323: 377-391; 1982.

- Kim, Y.I.; Dudek, F.E. Intracellular electrophysiological study of suprachiasmatic nucleus in rodents: Excitatory synaptic mechanisms. *J. Physiol.* 444: 269-287; 1991.
- Kornhauser, J.M.; Nelson, D.E.; Mayo, K.E.; Takahashi, J.S. Photic and circadian regulation of c-fos gene expression in the hamster suprachiasmatic nucleus. *Neuron* 5:127-134; 1990.
- Liou, S.Y.; Shibata, S.; Iwasaki, K.; Ueki, S. Optic nerve stimulation-induced increase of release of ^3H -glutamate and ^3H -aspartate but not ^3H -GABA from the suprachiasmatic nucleus in slices of rat hypothalamus. *Brain Res. Bull.* 16: 527-531; 1986.
- Meijer, J. H.; Rietveld, W. J. Neurophysiology of the suprachiasmatic circadian pacemaker in rodents. *Physiol. Rev.* 69:671-707; 1989.
- Moore, R.Y.; Lenn, N. J. A retinohypothalamic projection in the rat. *J. Comp. Neurol.* 146:1-14; 1971.
- Pickard, G.E. The afferent connections of the suprachiasmatic nucleus of the golden hamster with emphasis on the retinohypothalamic projection. *J. Comp. Neurol.* 211: 65-83; 1982.
- Pickard, G.E.; Ralph, M.R.; Menaker, M. The intergeniculate leaflet mediates effects of light on circadian rhythms. *J. Biol. Rhythms* 2: 35-56; 1987.
- Ralph, M.R.; Menaker, M. Bicuculline blocks circadian phase delays but not advances. *Brain Res.* 325: 362-365; 1985.
- Ralph, M.R.; Menaker, M. Effects of diazepam on circadian phase advances and delays. *Brain Res.* 372: 405-408; 1986.
- Rea, M. Light increases Fos-related protein immunoreactivity in the rat suprachiasmatic nuclei. *Brain Res. Bull.* 23:577-581; 1989.

- Rea, M. Different populations of cells in the suprachiasmatic nuclei express c-fos in association with light-induced phase delays and advances of the free-running activity rhythm in hamsters. *Brain Res.* 579:107-112; 1992.
- Rusak, B.; Robertson, H.A.; Wisden, W.; Hunt, S.P. Light pulses that shift rhythms induced gene expression in the suprachiasmatic nucleus. *Science* 248:1237-1240; 1990.
- Rusak, B.; Zucker, I. Neural regulation of circadian rhythms. *Physiol. Rev.* 59:449-526; 1979.
- Shibata, S.; Liou, S.Y.; Ueki, S. Influence of excitatory amino acid receptor antagonists and of baclofen on synaptic transmission in the optic nerve to the suprachiasmatic nucleus in slices of rat hypothalamus. *Neuropharmacology* 25: 403-409; 1986.
- Takahashi, J. S.; DeCoursey, P. J.; Bauman, L.; Menaker M. Spectral sensitivity of a novel photoreceptive system mediating entrainment of mammalian circadian rhythms. *Nature* 308:186-188; 1983.
- Talman, W.T.; Colling, J.E.K.; Robertson, S.C. Glycine microinjected into nucleus tractus solitarii of rat acts through cholinergic mechanisms. *Amer. J. Physiol.* 260: H1326-H1331; 1991.
- van den Pol, A.N.; Wuarin, J.P.; Dudek, F.E. Glutamate, the dominant excitatory transmitter in neuroendocrine regulation. *Science* 250: 1276-1278; 1990.

Acknowledgements

The authors gratefully acknowledge the skilled technical assistance of Ms Anna Marie Michel. We are indebted to Dr Michael Iadarola for providing the Fos antiserum. Finally, we thank to Dr. Robert L. Gannon for a constructive critique of the manuscript. This work was supported by AFOSR 2312W6 (MAR).

COLLABORATIVE INSTRUCTIONAL DEVELOPMENT ENVIRONMENT:
A STAGE FOR THE AIDA

Robert G. Main, Ph.D.
and
Andrew S. Wilson, M.A.
Department of Communication Design

California State University, Chico
Chico, CA 95929-0504

Final Report for:
Summer Research Program
Armstrong Laboratory: HRTC

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

Collaborative Instructional Development Environment:
A Stage for the AIDA

Robert G. Main, Ph.D.
and
Andrew S. Wilson, M.A.
Department of Communication Design
California State University, Chico

Abstract

Computer-based media production tools have matured sufficiently to enable the Air Force to readily provide very powerful curriculum materials development tools based on the existing Workstation III or IV. However, providing instructional designers and developers with a multimedia development workstation is not equivalent to providing them the power to use them well. While an Automated Instructional Design Advisor will certainly aid designers and developers in choosing appropriate media to solve instructional problems, provision of such powerful media production tools will require a commitment within the Air Force to provide technical and creative support. Only this will ensure effective, motivational media design.

An examination of matured computer-based media production technology was undertaken and a group of ISD experts was impaneled to discern which available tools hold the most promise and value for instructional design. This study presents the findings of the Delphi panel as well as considering the impact of providing such tools to designers and developers. We make the recommendation that implementation of a computer-mediated communication system be concurrent with the emplacement of computer-based media production tools to create a collaborative instructional development environment that will improve media creativity and dynamism especially with respect to computer-based training. In addition, such a system will provide for centralized archiving of reusable and repurposed media, effective formative and summative evaluation, increased collaboration between instructional designers, developers, subject matter experts and media production experts, thus increasing instructional quality, employee productivity and job satisfaction.

Collaborative Instructional Development Environment: A Stage for the AIDA

Robert G. Main & Andrew S. Wilson

INTRODUCTION

The Air Force has identified a goal concomitant with the development of the Advanced Instructional Design Advisor (AIDA) to create a multi-use instructional design workstation that will provide designers/developers the power to locally produce instructional materials ranging from graphics to video to computer-based interactive training. The Collaborative Instructional Development Environment (CIDE) Workstation is a set of functional specifications for hardware, software and network communications to operate on the Air Force Workstation III and IV. The specifications were developed from the responses of a Delphi committee asked to evaluate potential components of such a system in terms of their most common tasks and development efforts. The purpose of this research is to ascertain not only the computer-based media development tools required for effective instructional design by the Air Force, but to explore the type of collaborative support environment that will make available to instructional developers the expertise necessary to produce curriculum materials that fully exploit the power of media to motivate as well as teach.

DISCUSSION OF THE PROBLEM

Well-developed, appropriate media enhances instructional quality (Johnston, 1987). However, its use and effectiveness are hampered by:

- High cost
- Inability to determine appropriate media
- Lack of production expertise
- Substantial lead time for production
- Communications problems

While an Advanced Instructional Design Advisor (AIDA) can aid in selecting the appropriate media for an instructional activity, the problems listed represent serious impediments to the instructional developer in incorporating that media into the lesson. An experienced live instructor can often compensate for problems with instructional media, but the trend toward electronic delivery of instruction demands highly effective stand-alone media that communicates and motivates (Winn, 1987). Thus, there is a pressing need to empower instructional designers/developers (IDD's) with the tools to create powerful and dynamic instructional aids both for standup instruction and for computer-based instruction and distance learning.

The software and hardware tools now exist to create a multimedia production workstation that will allow development of a wide array of instructional media from documents to animation to interactive digital video. However, it is unrealistic to believe that instructional developers, often subject matter experts in a particular area of training, will also have the skills to make the best use of the powerful tools such a workstation will provide. Therefore, how to empower IDD's to produce and manipulate a variety of instructional media without placing on them the burden of requisite expertise in what are typically specialist areas has become a primary concern. For example, providing a graphics software package that is easy to use does not endow the user with

the talent to generate visuals that are creative and powerful in their ability to facilitate the transfer of knowledge and also grab the learner's attention and hold his or her interest.

AIDA offers a partial solution. AIDA can assist an IDD in selecting an appropriate medium to accomplish specific instructional objectives. It will guide developers through many complex design decisions and help them clarify goals and directions. But AIDA will not be able to review the aesthetics of media development. It will not be able to examine a proposed media solution for clarity of purpose and execution. Though it will be able to point out potential errors, it will not be able to answer specific questions regarding the easiest way to solve a particular communication problem; nor can AIDA provide years of media production experience to a neophyte designer.

So while AIDA provides less experienced IDD's important guidance and assistance in instructional design decisions, it cannot give comprehensive direction and evaluation in media production. They need both the valuable guidance of AIDA and a support infrastructure for technical and creative assistance in making the best use of the increasingly complex tools at their disposal and for producing the media that will most effectively support the instructional objectives.

THE COLLABORATIVE INSTRUCTIONAL DEVELOPMENT ENVIRONMENT (CIDE)

The increasing demand for multimedia instructional materials has created the need for a collaborative instructional development environment that includes not only instructional design and subject matter expertise, but media development specialists as well. The addition of a communications network to the computer platform on which designers/developers will be running both AIDA and the designer-specified hardware and software for media production can create a collaborative open network for sharing expertise in mentor/apprentice relationships between developers and special interest advice conferences led by media specialists. Providing on-demand assistance should solve most media production problems and provide a logical path for the instructional developer to travel from the recommendations of the AIDA to the finished lesson material required by the instructional design.

The inherent versatility of an open network brings a number of added benefits to Air Force instructional design organizations including:

- Centralized Multimedia Education and Training Archive (META) of reusable instructional media for easy updates, adaptation and integration.
- Ability to search local and remote archives for appropriate existing media.
- Organizational gateway for send/receive faxing and E-mail.
- Centralized electronic media publishing.
- Automated document flow for formative and summative evaluation.
- Improved task management and group coordination.

This study will address the following questions pertinent to a collaborative instructional development environment:

1. Is electronic collaboration effective?
2. What are the instructional media needs of Air Force instructional developers and how are they currently being met?

3. Can a collaborative instructional development environment be created with current commercial-off-the-shelf technology?

ASSUMPTIONS

That electronic media provide effective enhancement to instructional strategies in terms of improved cognition, information organization and integration, and learner motivation has been generally accepted since the 1968 research of Chu and Schramm (1979). Therefore, this study does not address the use of electronic media in improving the effectiveness of instruction. The term "effective" implies a judgment relative to a standard. This has customarily meant comparison of some electronic medium relative to face-to-face instruction without electronic media. "Medium" is defined by the American Heritage Dictionary as "an agency, such as a person, object or quality, by means of which something is accomplished, conveyed or transferred." The electronic media are simply carriers of information for presentation to the learner. The potential of a medium for transfer of knowledge and skills has more to do with how the information is packaged and accessed by the learner than it does with the characteristics of the medium itself (Johnston, 1987).

Information encoded as print differs very little whether it is presented by video, CRT or in a handout. The spoken word may be carried by the instructor's voice in person, on audio tape or by video. The assumption is made that learning does occur with electronic media and that the way the information is packaged, accessed and presented can have great effect on the learning process. Given this assumption, the problem becomes how to provide the tools and the expertise to permit the instructional developer to have choices in the selection of an appropriate medium and the most effective packaging of that instruction for optimal learning. An important consideration in the design and development process is the "mindware", a term coined by Salomon (1985) that refers to the mindset a learner brings to the instructional process. It includes the learner's propensities and associations with media. Main (1992) has attempted to systematize these motivational factors by generating a model of instructional design which integrates the affective domain into the curriculum development process. Although empirical data is sparse, the attractiveness of electronic media is evident in the amount of leisure time spent with television and electronic games by both children and adults. Though there is some ongoing discussion in the literature over how much and what types of media are appropriate to various tasks (see Friedman, Polson and Spector, 1991), we will not attempt to explore these issues here since they are more the province of an AIDA, SME's and IDD's than they are dependent on a CIDE.

The very strength of a collaborative instructional development environment is the versatility to produce any and all types of media as deemed suitable to a particular task by the cadre of professionals employed for ISD. Whether the media need is a published document, graphic slide, computer-based animation, or edited videotape, the personal computer has matured sufficiently as a platform to produce professional quality products. We state here that the PC is sufficiently robust to simultaneously support a panoply of media production tools, an AIDA, and a collaborative network to form electronic work groups.

Electronic work groups as discussed here are the result of computer-mediated communication (CMC). Analog communication networks are not considered because of their relatively high cost when used for media exchange. Computer-based communication systems range from simple electronic mail to voice mail, interactive chat forums, desktop video conferencing, document transfer and shared screen editing. Basic CMC systems have gained enormous popularity over the last ten years as

exemplified by the rise in usage of public network services such as MCI Mail, Prodigy and America On-Line, and the world-wide research collaboration taking place on Internet, NSFnet and Bitnet. For millions of people, checking their e-mail messages or logging into an interactive chat conference has become a daily ritual. Internet, a consortium of universities and research institutions, expects to have more than two million participants by 1995 (*Communications Week*, July 6, 1992, p. 1).

These services are active collaborative communities. Over one thousand special interest forums on the Internet, for example, allow asynchronous discussion of social, technological and scholarly issues ranging from animal psychology to quantum physics. The participants represent a huge knowledge base. By actively sharing information they multiply their individual skills and abilities.

IS ELECTRONIC COLLABORATION EFFECTIVE?

The literature on collaborative work is rich with examples of increased employee productivity. Though one study of Air Force cadets found that highly competitive people perform best when given individual rewards (Porter, Bird, & Wunder, 1990), the majority of researchers have reported improvement in performance on complex tasks by collaborative groups (Bassin, 1988; Blaye, Light, Joiner, & Sheldon 1991; Johnson, Maruyama, Johnson, Nelson & Skon, 1981; Katz, Kochan & Weber, 1985). In their excellent literature review, Tjosvold & Tsao (1989) state:

Considerable research, including field experiments, indicate that people in cooperation compared to those in competition exchange resources, assist each other, and manage conflicts constructively so that they are all successful (p. 189).

Citing the findings of Johnson et. al. (1981) they continue, "as they work cooperatively, employees explore issues and make successful decisions, and are more productive especially on complex tasks that benefit from sharing information" (p. 189).

According to Bassin (1988), "It's not the gifted individuals who make peak performance possible as much as the dynamics of belief, collaboration and support" (p. 64). Bassin believes cooperative work groups are effective because of the resources of individual members, diversity of ideas, emotional support, mutual motivation and increased job satisfaction. He feels that isolated employees are at a fundamental disadvantage, unable to grasp how their work output fits into the overall performance of the organization. Collaborative teams solve these problems. According to Tjosvold et. al. (1989):

In cooperation, people believe their goals are positively linked; one's goal attainment helps others reach their goals. Alternatively, mistrust, individual tasks, and win/lose rewards induce competition. Competitors believe their goals are negatively correlated so that one's goal attainment makes it more difficult for others to attain their goals (p. 189).

Finally, members of cooperative work groups report increased job satisfaction and organizational loyalty (Finholt, Sproull & Kiesler, 1990; Tjosvold et. al., 1989; Bassin, 1988; Sproull and Kiesler, 1986; Tjosvold, Andrews and Jones, 1983; Johnson et. al., 1981). A reduced sense of isolation, greater understanding of organizational objectives, emotional support and social interaction all seem to play important roles. Sproull and Kiesler (1986) point out that cooperative work groups often use electronic mail to provide a productive outlet for natural desires for sociability and organizational attachment. "People like to be sociable at work. A technology that makes it easy to be sociable—be it a water fountain, coffee pot, telephone, or EMS [electronic messaging system]—will be used for sociability" (p. 1151). Likewise, Tjosvold, et. al. (1983) suggest that cooperative interaction strengthens morale, commitment to the organization and productivity. The positive experiences of working

together lead employees to believe they have gained a great deal from the employer, and teamwork binds them to each other and to the organization.

It is generally felt that members of a cooperative work group benefit from the strengths of the talented individuals of whom it is comprised. (Bassin, 1988). Technical expertise and design experience more readily cross the organizational lines in an ad hoc cooperative group, providing just-in-time support for mission-critical objectives (Finholt, et. al., 1990). Employees participate in organizational goals and enjoy increased productivity and greater job satisfaction. But traditional methods of forming and maintaining ad hoc work groups such as face-to-face meetings may cost an organization a great deal in terms of travel, time to distribute materials, and time required to meet and to schedule more meetings, especially amongst geographically remote participants (Finholt et. al., 1990).

The literature provides sufficient evidence that electronic collaboration is effective. There is a note of caution. Changing communication patterns and protocols also changes organizational culture. These issues are not addressed in this paper but considerable empirical evidence is available and should be examined before establishing capabilities (see for example Lea and Spears, 1991; Dubrovsky, Kiesler and Sethna, 1991; Smilowitz, Compton and Flint, 1989; Sproull et. al., 1986).

WHAT ARE THE INSTRUCTIONAL MEDIA NEEDS OF AIR FORCE IDD'S AND HOW ARE THEY NOW BEING MET?

A Delphi group of expert Air Force IDD's was used to determine the instructional media needs of the Air Force and what tools would most enrich the collaborative instructional development environment.

The methodology for this study involved a review of state-of-the-art technologies in the field of personal computers and desktop workstations, media production tools and communication software. The evaluation was limited to commercial-off-the-shelf applications or products being beta tested for commercial release. An examination of the trade publications in personal computing, desktop publishing, digital photography, graphic design, video production and communication networking was used to establish a taxonomy of available products and services that appeared to have value for instructional development. A trip was made to the InfoMart in Dallas to see some of the candidate technologies demonstrated.

The information from the technology review was used to generate a list of 42 product categories divided into three areas of instructional design/development: 1) Instructional materials development, 2) Management and 3) Collaboration. To provide a rational method for evaluating the importance of the functions represented by these product categories, a combination of the Delphi methodology and the Kepner-Tregoe rational decision model (1965) was used.

The Kepner-Tregoe rational decision model uses the technique of determining what are the essential outcomes and what are the desirable outcomes for any decision situation. It is widely used in the evaluation of competing systems because it provides a quantifiable method for comparing products with a variety of disparate features. The "must" category of features must be met by all candidate systems or they are dropped from further consideration. Those features deemed desirable but not essential are labelled "wants" and are assigned weights (usually by a panel of users). The evaluation is made by experts who test the system's ability on each item.

For our panel of experts we elected to use Air Force IDD's from a number of organizations that would reflect a variety of instructional development needs from standup courses to computer-based technical training. We opted for this approach over using consultants or academics because of the importance of user involvement in system design.

Systems development theory (Kling, 1991; Conner, 1985; Boar, 1984) and practical field experience (Kyng, 1991; Perin, 1991) both indicate that potential users of a system must be involved during the early stages of design. Kyng (1991) advocates a doctrine of "mutual learning" where designers teach users about the technological possibilities while users instruct designers in the task specifics of their work. Perin (1991) discusses the problems created when systems are mandated for unwilling users. Computer systems that extend the abilities of subordinates, and especially those that may create informal social fields among them, may threaten managers. "The challenge is to create computer support that acknowledges, if not incorporates these realities, rather than presuming the technology will by itself reform or obliterate them" (p. 81).

Therefore, while expert consultants might easily specify an extremely competent design system in terms of the prevalent ISD models and perceived needs of IDD's, there is no certainty that such a system will be readily adopted by Air Force IDD's. For these reasons, we assembled a Delphi panel of experienced users to assist in developing the functional design requirements for the CIDE. We were assisted in identifying expert IDD's by Lt. Sheila Robinson HQ ATC/TTDD and by Maj. Richard O'Neal HQ ATC/XPCR at Randolph Air Force Base.

Delphi is a technique developed by the RAND Corporation to be used in technical forecasting or to achieve consensus amongst a group of experts without undue influence (halo effect) by prestigious individuals (Tersine and Riggs, 1976). For this study, a Delphi group of ten experienced Air Force IDD's was selected. Their combined experience totals 108 years in curriculum development. The participants are expert practitioners rather than a representative sample of Air Force instructional developers. A survey by Walsh, Yee, Grozier, Gibson & Young (1992) of 256 Air Force personnel involved in developing computer-based instruction (CBI) found the average experience of the IDD's to be just 20 months. Participants were selected for this panel because of their knowledge and experience with Air Force instructional design and development, not because they represented typical IDD's.

Eight members of the panel of experts are male and two are female. Seven are civilian employees of the Air Force and three are career military personnel. The level of sophistication of the group is quite high. One participant is a manager of an instructional development group. Although not involved in the actual design of instruction at this time, he has more than seven years of prior experience in training development. He is included because his managerial responsibilities include the design and development of all types of instruction from traditional classroom to CBI. Eight of the members are experienced in designing and developing CBI and three of them do this exclusively. Five panel members are involved with the design of standup training using static media aids and six develop dynamic media for their instructional programs. Seven of the designers have at least some experience in multimedia CBI development. This level of expertise and experience in the field makes this group well qualified to offer expert evaluations concerning the functional requirements and desired features for a collaborative workstation to improve both the productivity and quality of Air Force instructional design and development.

INSTRUMENTATION AND EVALUATION

A structured questionnaire was distributed to the panelists in which they were asked to specify the percentages of instruction created using different types of media both within their organization and Air Force-wide. We then asked them to tell us how much instruction using each type of media they thought would be most appropriate for use by their organization and the Air Force.

They were presented with the list of 42 candidate technologies developed from the trade journals, literature review and vendor presentations. They were told their expertise was being solicited to assist in determining the design features of a collaborative instructional development workstation. They were asked to evaluate each technology category to determine if they felt it was essential (a "must") to quality curriculum development. If the technology was judged not to be essential, the panelists were asked to place a value on its worth (0=valueless to 20=nearly essential) to an instructional developer.

Finally, they were asked about media they can and cannot presently develop in-house and their collaborative relationships with other designers and subject matter experts.

We sought to answer five basic questions that bear directly on the functional specifications of the CIDE:

- What types of instructional media are presently being developed?
- What types of instructional media would IDD's prefer to develop if they had more resources?
- By whom is various media now developed (IDD's, non-training agencies, contractors)?
- What kind of collaboration is necessary to the development of effective media for instruction?
- What technologies are perceived as essential to IDD; which are desirable, and which are unnecessary?

The data gathered was averaged and used to rank order potential technologies that could be included in the CIDE. Using the Kepner-Tregoe (1965) decisioning system, we were able to determine which technologies constitute the necessities of the system and which the niceties. Using this "rational" decisioning system helps forestall the desire to add every available technology under the assumption that if we provide it to designers they will learn to want it and use it—the "Field of Dreams" approach.

THE TECHNOLOGIES

Our only constraint (self-imposed) was that all software and hardware technologies specified for the design of the CIDE should be compatible with the Air Force Workstation III and IV. Our intent in this is not only to reduce eventual development costs and to work with a computing platform that has already been approved and implemented by the Air Force, but also to ensure compatibility with the Advanced Instructional Design Advisor being developed for Air Force ISD (Hickey, Spector and Muraida, 1992).

Technological feasibility was determined through review of computer trade publications and an on-site visit to Dallas' InfoMart. While specific software and hardware selections will require further study and additional input from potential users, there will be a discussion below of critical technologies that match the user requirements determined by the Delphi panel. The research and development paradigm is to establish a rational ordering of functional requirements and assess the status of commercial-off-the-shelf (COTS) tools available to meet those requirements.

FINDINGS

Our hypothesis holds that Air Force IDD's are probably designing more standup instruction and more instruction with static media than they would prefer. If true, we would suspect it is because of an inability to design more dynamic curriculum materials stemming from a variety of reasons ranging from lack of skills to lack of equipment to insufficient time. To test this theory, we asked the Delphi panel for their best estimate of the quantities of instructional media of various types being produced by them, their organizations, and their best estimate of the media type's use Air Force-wide. Summaries of their responses are contained in figures 1-8.

As we postulated, individual Air Force IDD's generally feel they are developing more instruction without media, or instruction that is dominated by static media than they would prefer (Figure 1). The mean portion of curriculum hours developed as standup instruction with no media was estimated by our panel to be 40 percent within their own organizations and 32 percent overall for the Air Force. They believed a more suitable amount of this type of instruction would be about 25 percent.

The participants also indicated they would prefer to see less instruction supported by static media such as slides, overhead transparencies, etc. (Figure 2). They estimated instruction with static media accounted for almost 50 percent of the hours of instruction produced in their organization and nearly 60 percent Air Force-wide. Their preference was that approximately one-third of the instructional hours be standup instruction supported by static media.

The use of dynamic media for instruction shows an opposite result, i.e., the participants would like to use dynamic media more than it is being used now (Figure 3). Participants would like to increase their organization's use of dynamic media from 30 to nearly 40 percent of instruction designed, and would like to see it account for one-third of total Air Force instruction.

As we suspected, print-based media is still the most widely used medium (Figure 4). More than 90 percent of instructional hours are supported by some printed materials in the form of student handouts and workbooks. The effective penetration of desktop publishing and familiarity of nearly all instructional designers with paper-based production certainly facilitates its ubiquity. Nevertheless, panelists feel that the amount could be reduced somewhat without damage to the instructional process.

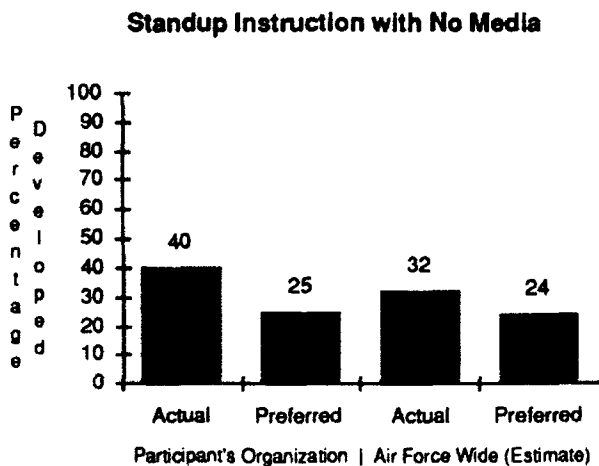


Figure 1

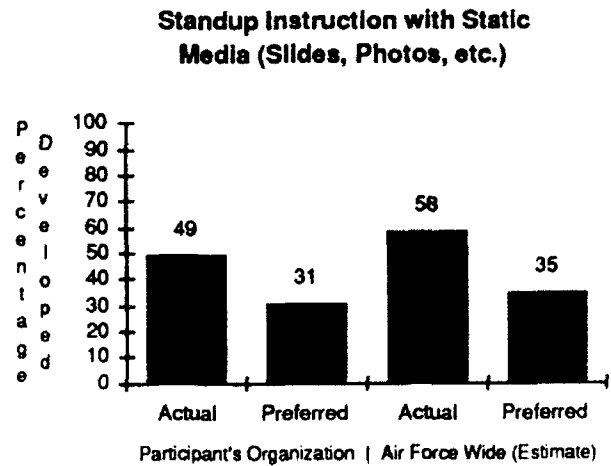


Figure 2

Standup Instruction with Dynamic Media (Slide-Tape, Video etc.)

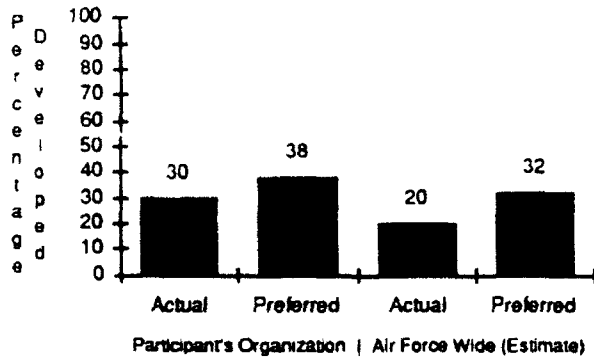


Figure 3

Instruction with Student Handouts and/or Workbooks

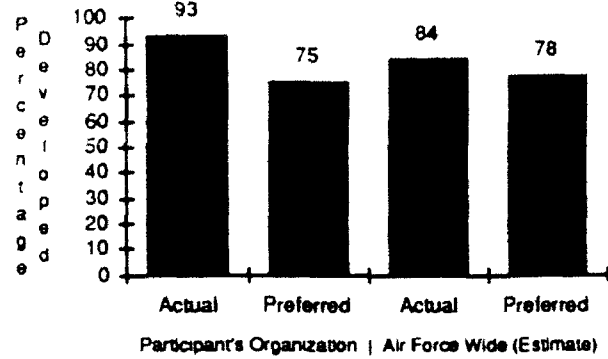


Figure 4

While the percentage of instruction presently developed for computer-based delivery is low (18 percent within our panelist's organizations and less than 10 percent estimated Air Force-wide), most participants would like to see CBI use increased greatly (Figure 5). Although one panelist charged exclusively with CBI development feels that CBI should only account for 10 percent of all instruction, other participants felt the amount of CBI desired should be nearly one-third of Air Force-wide instruction.

Every respondent wants to see more multimedia CBI developed for Air Force instruction (Figure 6). At present, one-third of all CBI being developed within participants' organizations consists of dynamic multimedia, along with an estimated 25 percent of such CBI Air Force-wide. But panelists believe that the amount should be pushed above 50 percent, that is, more than half of all computer-based instruction should be multimedia. This suggests a clear need for interactive dynamic multimedia. However, as participant comments allude, many of the tools required to develop motivational multimedia are presently unavailable in the field.

Air Force IDD's also indicate an interest in developing more instruction for distance learning applications (Figure 7). By their estimate a scant two percent of Air Force instruction now constitutes distance learning. However, they believe as much as

Curriculum Presented as Computer-Based Instruction

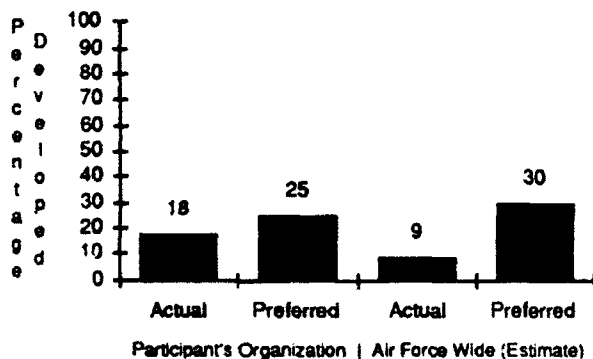


Figure 5

Computer-Based Instruction with Multimedia Presentation

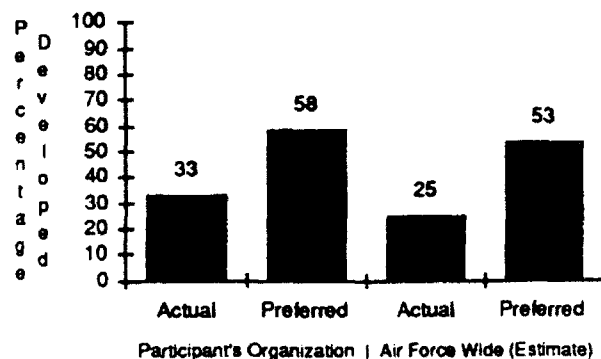


Figure 6

27 percent of their instruction has distance learning applications, and see a potential for 22 percent of total Air Force instruction to be delivered remotely.

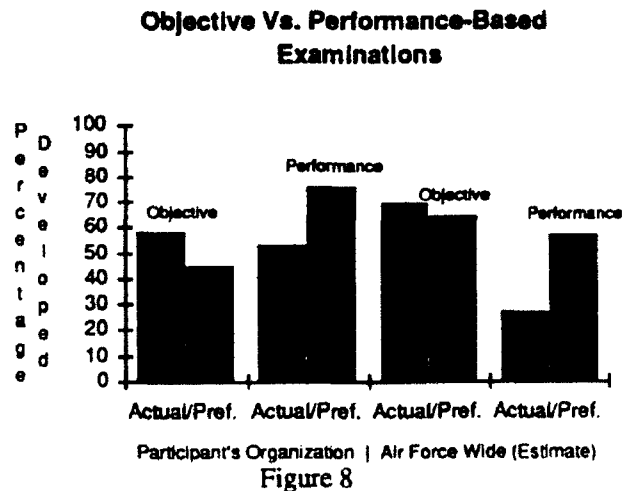
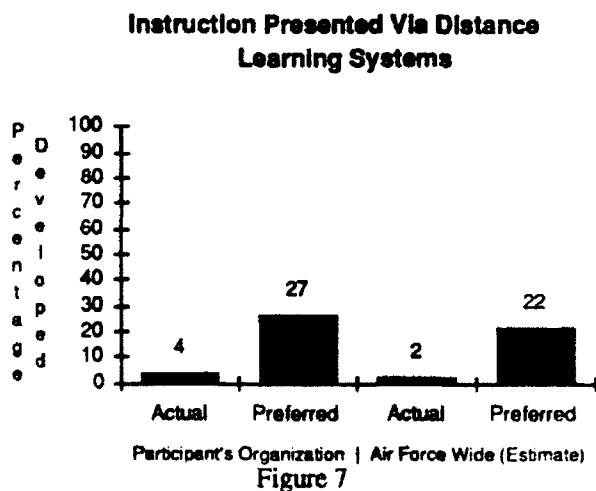
Finally, panelists feel that too great an emphasis is now placed on objective examinations. They show a clear preference for developing performance-based evaluations both within their own organizations and Air Force-wide (Figure 8). The data suggests most of the experts want performance-based evaluations to supplement rather than totally replace objective exams. Given the nature of many of the tasks for which Air Force IDD's develop instructional systems, it seems likely that objective measures may often be insufficient. Many skills-based tasks can only be effectively evaluated by proficient performance. Objective exams are generally easier to develop, administer and evaluate than performance-based tests, suggesting that IDD's may benefit from tools that help them develop more innovative evaluation measures.

In summary, the experienced Delphi panel would like to develop less standup instruction that is unsupported by media or has only static media. Although they would like to reduce their reliance on traditional text-oriented student handouts and workbooks, they still want print support for three-fourths of their instruction. They would like to increase the use of dynamic media as a support for standup instruction and they would like to increase CBI and in particular the use of CBI that includes multimedia presentation. They would also like to increase the amount of courses offered through distance learning systems and they would like to implement more performance-based evaluations.

MEDIA PRODUCTION METHODS

Most of our Delphi panelists (seven of nine responding) contract the final development of graphics. Reasons cited for this range from not violating a base-negotiated contract by developing graphics in-house, to a respondent who cites excessive time and effort spent to produce non-professional looking graphics. Only two of nine IDD's report they develop their own final graphics. Two report they are starting to develop more internally. One of those cited slow turnaround by contractors.

All respondents' organizations contract for printing services, though six of nine provide camera-ready copy, indicating the penetration of desktop publishing. One panelist reports that their organization is beginning to desktop publish and hope to soon provide camera-ready copy; one panelist reports that contractor turnaround is, "not very fast". Two panelists report that printing services are handled by contractors, but do not specify who provides camera-ready copy.



Five of seven respondents use on-base contractors for all photography. Two of seven provide their own photos. Similarly, five of nine respondents use on-base contractors for all video footage, while two of nine produce their own internally. Of these two respondents, one is tasked with CBI development, the other is a manager whose organization develops primarily textual media and team training. Two of nine share the task of video development with contractors. Three of seven respondents contract for slide/tape program production, two produce in-house and two report no slide/tape productions. Three of eight panelists report audio production is provided by on-base contract while two develop in-house. Three panelists state audio production services are not available (even though computer-based audio production tools are highly developed and inexpensive).

ESSENTIAL AND DESIRED DEVELOPMENT TOOLS

Respondents were asked to determine with a yes or no vote whether a variety of computer-based media development tools were essential to ISD. Where they voted no, they were asked to determine the usefulness of the tool for ISD on a scale of 0-20. Each "yes" vote is valued at 30 points, while each "no" vote is valued at its given weight. These data are totaled, divided by 30 and used to create the Kepner-Tregoe decision tree shown in Figure 9. Tools that receive 60 percent of possible points (180 points of the 300 points possible) are considered to be essential to ISD and, therefore, to the CIDE workstation; those totaling 50 percent (or 150 points) are deemed highly desirable; others below 50 percent are considered useful in proportion to their weights. Note that no tool received a weight below 30 percent (90 points), and at least two panelists considered any given tool absolutely essential (yes votes). Thus, all of these tools should probably be available as add-on features to the workstation to support the task needs of particular designers.

Clearly Air Force instructional designers and developers would like to be producing more dynamic and motivational media for standup instruction, computer-based instruction and for distance learning applications. A variety of reasons for the present lack of media are suggested in the panelists' comments on their organizations' current arrangements for final media production.

One respondent states that the existing base-negotiated graphics contract legally prevents them from developing graphics in-house. One is simply lacking sufficient equipment. Three others cite training and poor final quality of in-house work due to "seldom used but technically difficult skills". Four of the respondents bemoan long turnaround time for most contracted media, while three others cite low quality in contractor-developed media due to poor communication or insufficient familiarity with the subject matter. Despite these problems, all parties indicated the need to use more dynamic and motivational instructional media.

The literature cited earlier suggests that many of these problems could be remedied by the installation of a collaborative network. Creativity and expertise hurdles can be surmounted by special interest groups and just-in-time technical support. Communications problems with contractors can be circumvented with more timely collaborative sessions. With the appropriate tools, more preparatory development work can take place in-house even if prior agreements stipulate that contractors must produce final media. And with the right tools, designers will have the freedom to explore more creative, dynamic and motivational media solutions to instructional problems.

From the Kepner-Tregoe decisioning tree it is fairly obvious that IDD's themselves recognize this. Of course there was unanimous agreement for word processors, authorware, flowcharts and test development tools—the staples of the trade. But not

Kepler-Tregoe Decisioning Tree:
 The Value of Media Production Technologies
 for Instructional Development in Percentages

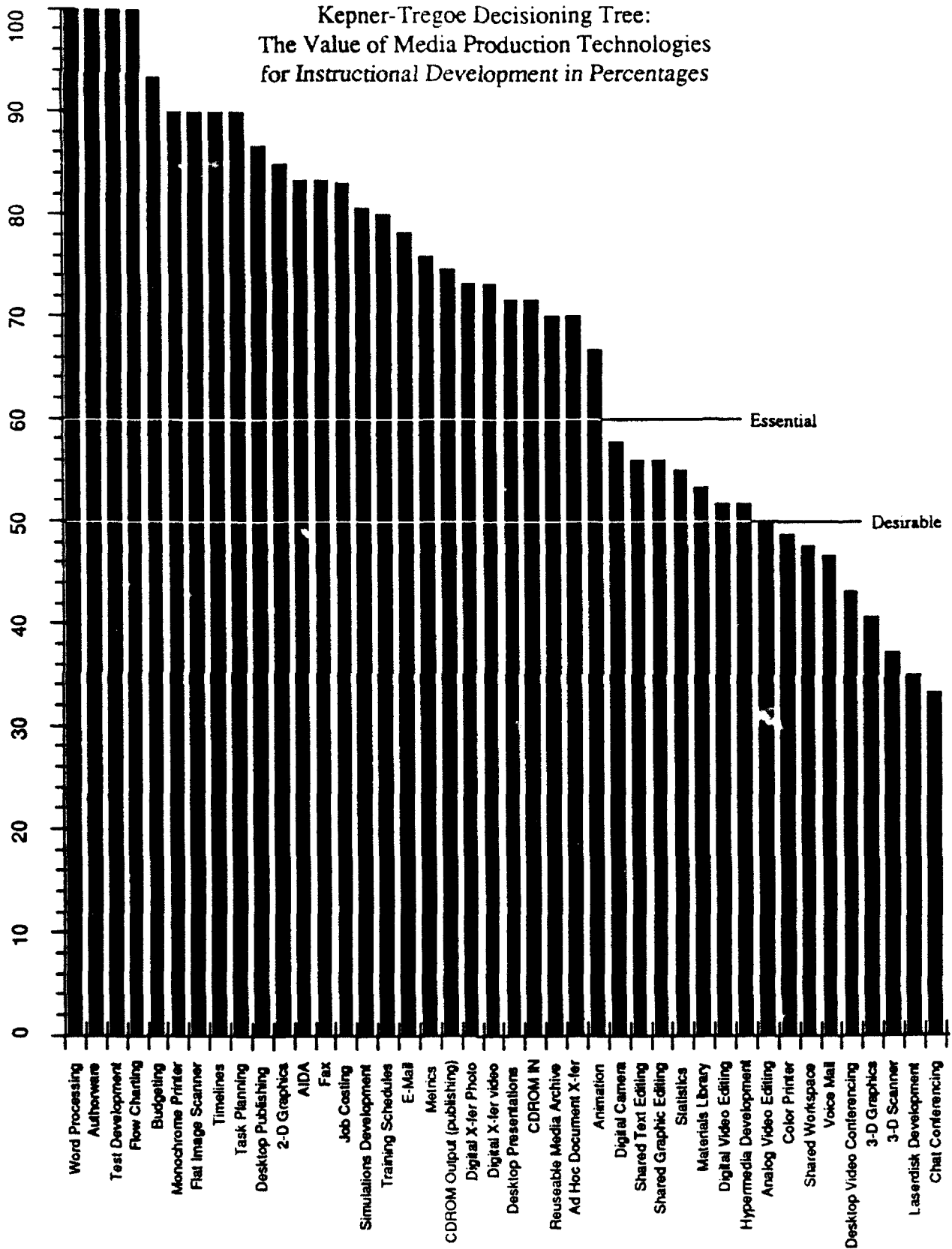


Figure 9
 25-14

surprisingly, there was extremely strong interest in desktop publishing, two-dimensional graphics, image scanners, simulations development, digital and analog video editing, video and photography transfer, and CDROM input and publishing. These are tools to create a media-rich instructional environment. They are not simple to use, requiring technical proficiency and creativity, but IDD's understand their value to the development of highly motivational, dynamic media.

The data indicate, as well, that they understand the value of collaboration with peers and specialists to quality ISD. Every respondent reported routine collaboration with a subject matter expert. Four participants agreed with the panelist who stated, "[it is]...impossible to develop any kind of quality training without an SME". In addition, four of nine cite frequent collaboration with other IDD's for ideas, and evaluation; five of nine report collaboration with graphics specialists; and three of nine report contact with technical experts and media specialists: "without this link, our product would not get out".

Presently, the bulk of this collaboration is conducted face-to-face or over the telephone. Where media production and graphic design specialists are concerned, considerably more work is conducted face-to-face than by any other method. Yet, most of these experts cite time and communication factors as primary impediments to more extensive use of motivational media. Clearly a well-implemented collaborative network that connects IDD's, SME's and media production specialists—including contractors—would solve most of the aforementioned problems, streamline collaborative processes by eliminating many face-to-face meetings, and encourage the creative development of more effective, dynamic and motivational media.

Finally, the experience of the Navy cited by Cantor (1988) suggests that centralized archival of reusable, adaptable media would save time and money, while encouraging IDD's to make use of the best stock media available. This type of networked archive would make media produced by the most highly skilled and talented producers readily available to IDD's throughout the Air Force's widely distributed ISD agencies.

CAN A COLLABORATIVE INSTRUCTIONAL DESIGN ENVIRONMENT BE CREATED WITH CURRENT COMMERCIAL-OFF-THE-SHELF TECHNOLOGY?

The answer to this questions is a qualified yes. Every item on the list of 42 functions identified as essential or desirable for at least some IDD's are available right now. The open architecture and communication networking capabilities are present. What cannot be answered by this study is what integration software, degree of data interchange standards and communication data speeds are necessary and available for implementation of every function. To answer this question definitively will require additional study, prototype development and beta testing which is strongly recommended. A beta test would create a field laboratory environment that could be useful in answering a variety of research questions regarding process and task procedures for optimal use of the collaborative instructional development environment. The constraining factors for such a system are not hardware and software.

SUMMARY

As indicated throughout the study, the mere provision of development tools to IDD's does not empower them individually to design efficiently or effectively. Instructional Systems Design (ISD) is a complex task requiring variously the input of IDD's, SME's, graphic designers, video production specialists, technical writers and CBI programmers. While increased attention to the user interface has created a generation of media production software much more friendly to the average user, true mastery of

any such tool is the result of both practice with the software and a thorough understanding of the knowledge domain to be represented and the traditions and techniques native to the media being developed. The proliferation of desktop publishing provides an illustrative example. While user-friendly programs like PageMaker and Ventura Publisher brought computer-based publishing to everyone, they did not communicate the traditions of the typographer's art. As a result, the average quality of typeset materials now varies a great deal. For expert typesetters, the software tools were a productivity boon, allowing them to create high-quality typeset pages more quickly and less expensively than ever before. However, in the hands of the non-typesetter, they allowed only the quick production of readable but inelegant pages devoid of ligatures, gendered quotes, and properly kerned letter pairs. To quote Bob Krejci, "There is no one-person authoring tool that can produce the kind of product that an experienced staff of designers, subject matter experts, artists, and programmers can develop... There is no 'Van Gogh in a spray can' product" (1992).

This is not to suggest that there is not merit in providing IDD's with the wide variety of tools required to develop affective instructional media. On the contrary, while questions remain as to the amount and design of potent media for various instructional objectives and within specific domains of knowledge, there is no doubt that visual and auditory media can be an appropriate and highly effective means of organizing and presenting some information (Gildea, Miller and Wurtenberg, 1990). Analysis of specific appropriate forms and applications of graphics have been begun with respect to the development of an Automated Instructional Design Advisor (Friedman, et. al., 1991). While an AIDA may well include guidance towards appropriate applications and designs for graphic instructional media, the goal of a development workstation should be to provide robust fully-featured design tools. Whether a screw or nail is appropriate is the decision of the carpenter and architect. A good toolbox contains both driver and hammer. For example, the graphic design and imaging segments of the PC market have matured sufficiently that there are a variety of extremely competent software packages presently available for the development of fine art (e.g., Fractal Design Painter), line art (e.g., Corel Draw, Adobe Illustrator) CAD (e.g., AutoCad, Eazy Cad), solid model rendering (e.g., AutoCad, Renderman) and photo image processing (e.g., Image-In, Photo Finish). While the Air Force may choose to standardize on one or more of these packages, the CIDE workstation will be designed to accommodate one or all of them in ad hoc arrangements to support the task at hand. However, any such integration of new tools requires an equal commitment to technical and creative support services to aid in the transfer of sufficient expertise to make the tools useful. The creation of ad hoc collaborative work groups can address the problem of providing technical development assistance to less experienced IDD's and those without sufficient expertise in development of specific types of media. Simultaneously, collaboration should improve worker efficiency and job satisfaction as well as creating a community of expertise and a professional growth environment for the motivated IDD.

The superimposition of a collaborative open network on the instructional development environment will permit ad hoc working arrangements for mentor/apprentice relationships, creative and technical consulting support and multiple problem solving perspectives for individual IDD's as well as bringing other benefits to improve both ISD and IDD performance. A collaborative open network among instructional developers may include technologies as commonplace as fax, e-mail and voice mail or those as esoteric as ISDN-based digital document transfer and two-way video conferencing. Fundamentally, it constitutes the creation of open communication channels that engender the formation of ad hoc work groups and technical interest

forums, allow the transfer of documents and resource materials, and provide an easily accessed, non-threatening means to seek technical help and creative assistance.

Building an open network will allow both synchronous (live chat forums or video conferences) and asynchronous (e-mail, fax) communications amongst IDD's. But perhaps just as valuable, it will enable document transfer for evaluation purposes, scheduling and coordination, and sharing of valuable resources such as adaptable existing media and a centralized archive for some materials. Perez (1992) in an exploration of traits of the expert training developer discovered that senior designers developing instruction through a team approach, "...developed formal conventions and guidelines to insure the uniform execution of the instructional design" (p. 13). CMC can support this kind of control and coordination as well as increasing efficiency in the group design process. Cantor (1988) reported that an automated curriculum design environment developed for the Navy that included archiving of boilerplate text and graphics as a shared resource reduced time spent on repetitive work by allowing incorporation and adaptation of existing materials. He cited an aggregate reduction in ISD time-on-task from between 45 and 66 percent.

The study by Walsh et. al. (1992) of Air Force CBT developers reinforces the value a collaborative instructional development environment could provide. They found 78 percent of CBI development team members were inexperienced in CBI design and development. More than one in four of the development team members felt the team's activities were not well coordinated and that communication between team members was unclear and ineffective.

A formal job/task analysis was not performed in one of every three CBI development efforts. About one in four CBI instructional developers relied on previous course materials or analyses. Forty percent used learning objectives modified from previous lessons. Only one-half of developers indicated a media analysis was performed as part of the CBI design process, and when an analysis was performed a subject matter expert was used over two-thirds of the time. Of the 253 CBI designers surveyed, not one indicated a media specialist was involved in the media analysis process. Of development team members, just over ten percent were described as media experts and they were all graphic artists. For the CBI projects, 95 percent contained some graphic components (icons, charts, tables, diagrams, maps, equipment, human figures, even animation); 48 percent had audio content (bells, beeps, tunes as rewards, signals, music, engines and verbal commands, questions, etc.); and 45 percent included some still or motion video (for identification of equipment, body parts, panels, etc., procedures and interpersonal and communication skills). The most commonly cited reasons for not using multimedia were lack of capability, not enough time and not being trained for development. There was no mention at all of interactive media applications.

Walsh's survey of practicing computer-based instructional developers strongly indicates that Air Force CBI development could benefit substantially from the availability of a collaborative instructional development environment as outlined in this study. The data from the Delphi group indicates non-CBI instructional development needs a similar capability.

Providing computer-based media development tools in a collaborative environment should streamline many work processes and stimulate interactive evaluation of instructional components. Tessmer and Wedman (1992) discovered that the most common reason cited by professional IDD's for skipping an ISD activity was not lack of money or experience but lack of

time. They state, "A means of 'cutting corners while controlling risk' in ID/D [ISD] project(s) needs to be developed" (p. 16). We believe the CTIDE can assist in cutting those corners.

CONCLUSIONS

The Air Force should explore implementation of a collaborative instructional development environment workstation. All the technologies involved are readily available in COTS packages. Many of the individual media development technologies such as flat bed scanning, graphic design, desktop publishing, authorware and laser printing are already in place in some ISD organizations. Issues to be resolved include:

- 1) the bandwidth and best method for ad hoc networking and multi-network management (e.g., simple ethernet, ISDN, FDDI, ATM, SNMP, etc.);
- 2) whether members of the collaborative community can benefit from broadband communications technologies such as desktop video teleconferencing, voice mail and groupware (shared graphic and text editing);
- 3) what kinds of materials should be centrally stored in a Multimedia Education and Training Archive (META) and whether those archives should be maintained within each ISD organization or by on-base visual information (VI) agencies such as Combat Camera, or both.
- 4) strategies for perusing, indexing and previewing contents of the META, including Boolean search techniques and indexed multimedia information retrieval.
- 5) determining any potential adverse consequences of installing a CMC network and developing strategies to offset them.

With the development of AIDA, the firm adoption of a standardized computer platform, and the maturation of computer-based media production tools ranging from graphic design to desktop digital video editing, the time is ripe for implementation of an instructional development environment with standard tools, central archiving, and a collaborative network for exchange of creative and technical support, reusable media, and formative evaluation. Changes to organizational culture, while predictable, will remain minor and manageable. In reality they will most likely contribute to increased employee satisfaction and organizational loyalty. Without question, such a system will contribute to the development of more effective, motivational instruction.

Bibliography

- Bassin, M. (1988). Teamwork at General Foods: New and improved. *Personnel Journal*, 67, 5 62-70.
- Blaye, A., Light, P., Joiner, R. and Sheldon, S. (1991). Collaboration as a facilitator of planning and problem solving on a computer-based task. *British Journal of Developmental Psychology*, 9, 471-483.
- Boar, B. H. (1984). *Application Prototyping*. New York: John Wiley & Sons, Inc.
- Cantor, J. A. (1988). An automated curriculum development process for Navy technical training. *Journal of Instructional Development*, 11, 4, 3-11.
- Chu, G. C., and Schramm, W. (1979). *Learning from Television—What the Research Says*, 4th Ed. (first edition released in 1968), Washington, D.C.: National Association of Educational Broadcasters.
- Communications Week (1992). Intel board to vote on IP fix. *Communications Week*, #410, July 6, 1992, p. 1.
- Conner, D (1985). *Information system specification and design road map*. Englewood Cliffs, NJ, Prentice-Hall.
- Dubrovsky, V.J., Kiesler, S. and Sethna, B.N. (1991) The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human-Computer Interaction*, 6, 119-146.
- Finholt, T., Sproull, L., and Kiesler, S. (1990) Communication and performance in ad hoc task groups. In, Galegher, J., Kraut, R. E., & Egido, C (Eds), *Intellectual Teamwork, Social and Technological Foundations of Cooperative Work*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Friedman, A., Polson, M. C., and Spector, J. M. (1991). Designing an advanced instructional design advisor: Incorporating visual materials and other research issues. USAF Document # AL-TP-1991-0017-Vol-4 HRTC, Armstrong Laboratories, Brooks, AFB, TX.
- Gildea, P. M., Miller, G. A. and Wurtenberg, C. L. (1990). Contextual enrichment by videodisc. In: *Cognition, Education, Multimedia*, Hillsdale, NJ, Lawrence Erlbaum Associates, Inc.
- Hickey, A. E., Spector, J. M. and Muraida, D. J. (1992). Design specifications for an Advanced Instructional Design Advisor (AIDA). USAF Document AL-TR-1991-0085-Vol-2 HRTC, Armstrong Laboratories, Brooks, AFB, TX.
- Hiltz, S.R. and Turoff, M. (1978). *The Network Nation: Human Communication via Computer*. Reading, MA: Addison-Wesley.
- Johnson, D. W., Maruyama, G., Johnson, R. T., Nelson, D. and Skon, S. (1981). 'Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis', *Psychological Bulletin*, 89, 47-62.
- Johnston, J. (1987). *Electronic Learning from Audiotape to Videodisc*. Hillsdale, NJ.: Lawrence Erlbaum Associates.
- Katz, H. C., Kochan, T. A. and Weber, M. R. (1985). 'Assessing the effects of industrial relations systems and efforts to improve the quality of working life on organizational effectiveness', *Academy of Management Journal*, 28, 509-526.
- Kepner, C. H. and Tregoe, B. B. (1965). *The Rational Manager*. New York, McGraw Hill.
- Kling, R. (1991). Cooperation, coordination and control in computer-supported work. *Communications of the ACM*, 34, 12, 83-88.
- Krejci, B. (1992) Response to a letter to the editor. *Instruction Delivery Systems*, 6, 4, p.5.

- Kyng, M. (1991). Designing for cooperation: Cooperating in design. *Communications of the ACM*, 34, 12, 65-75.
- Lea, M. and Spears, R. (1991). Computer-mediated communication, de-individualization and group decision-making. *International Journal of Man-Machine Studies*, 34, 283-301.
- Linde, C. (1988). The quantitative study of communicative success: Politeness and accidents in aviation discourse. *Language and Society*, 17, 375-399.
- Main, R. (1992). Integrating the affective domain into the instructional design process. Brooks A. F. B., Texas, Armstrong Laboratories Report AL-TP-1992-0004, HRTC, Armstrong Laboratories, Brooks, AFB, TX.
- Monge, P. R. and Kirste, K.K. (1980). Measuring proximity in human organization. *Social Psychology Quarterly*, 43, 1, 110-115.
- Perez, R. S. (In Press). Modelling the expert training developer. In *Advanced Training Technologies Applied to Training Design*. (Eds.) R.J. Seidel & P. Chatelier, Plenum Press.
- Perin, C. (1991). Electronic Social Fields in Bureaucracies. *Communications of the ACM*, 34, 12, 75-82.
- Porter, D. B., Bird, M. E., & Wunder, A. (1990). Competition, cooperation, satisfaction and the performance of complex tasks among Air Force cadets. *Current Psychology: Research & Reviews*, 9, 4, 347-354.
- Ridgeway, C.L., Berger, J. and Smith, L., (1985). Nonverbal cues and status: An expectation states approach. *American Journal of Sociology*, 90, 5, 955-979.
- Salomon, G. (1985). Information technologies: What you see is not (always) what you get. *Educational Psychologist*, 20, 4, 207-216.
- Smilowitz, M., Compton, D. C., & Flint, L. (1989). The effects of Computer Mediated Communication on an individual's judgment: A study based on the methods of Asch's social influence experiment. *Computers In Human Behavior*, 4, 311-321.
- Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science*, 32, 1492-1512.
- Tersine, R. J., and Riggs, W. E. (1976). The Delphi technique: A long-range planning tool. *Business Horizons*, April, 51-56.
- Tessmer, M., & Wedman, J. (1992). The practice of instructional design: a survey of what designers do, don't do and why they don't do it. San Francisco, CA: paper presented at the annual meeting of the American Educational Research Association.
- Tjosvold, D. & Tsao, Y. (1989). Productive organizational collaboration: The role of values and cooperation. *Journal of Organizational Behavior*, 10, 189-195.
- Tjosvold, D., Andrews I. R. and Jones, H. (1983). 'Cooperative and competitive relationships between superiors and subordinates', *Human Relations*, 36, 1111-1124.
- Walsh, W. J., Yee, P. J., Grozier, S. A., Gibson, E. G., and Young, S. A. (1992). *A survey of Air Force computer-based training (CBT) planning, selection, and implementation issues*. AL-TP-1991-0059, HRTC, Armstrong Laboratories, Brooks, AFB, TX.
- Winn, W. (1987). Instructional design and intelligent systems: Shifts in the designer's decision-making role. *Instructional Science*, 16, 59-77.

A METHOD FOR COMPARISON OF ALTERNATIVE MULTISHIP AIRCRAFT
SIMULATION SYSTEMS UTILIZING BENEFIT ESTIMATION

William C. Moor
Associate Professor
Department of Industrial and Management Systems Engineering

Arizona State University
Tempe, AZ 85287

Final Report for:
Summer Research Program
Armstrong Laboratory
Aircrew Training Division
Williams Air Force Base, AZ

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

September 1992

A METHOD FOR COMPARISON OF ALTERNATIVE MULTISHIP AIRCRAFT
SIMULATION SYSTEMS UTILIZING BENEFIT ESTIMATION

William C. Moor
Associate Professor
Department of Industrial and Management Systems Engineering
Arizona State University

Abstract

This study improves and refines a benefits model developed in previous work. It creates the operational procedures necessary to acquire all data required for estimating benefits. In addition, the study completes an operational test of these procedures demonstrating their feasibility.

This research focuses on the benefit component of this analysis as it is elements of the benefits computation that require the most refinement. The overall thrust of the model building remains the same as for the original effort. It is the desire of the author to make the computation model as clear as possible to the potential user and to build it in a form that facilitates use. All computational work is placed in LOTUS 1-2-3 spreadsheets that are annotated for data entry and use. The model is built in reference to a specific, operational aircraft (F-15) but is easily modified to allow comparison to any air superiority jet fighter for which multiship simulators would be developed.

A METHOD FOR COMPARISON OF ALTERNATIVE MULTISHIP AIRCRAFT
SIMULATION SYSTEMS UTILIZING BENEFIT ESTIMATION

William C. Moor

INTRODUCTION

The USAF desires, in so far as possible, that proposed capital expenditures be based on a benefit-cost comparison among all competing alternatives. The Aircrew Training Division of the Armstrong Laboratories (formerly the Human Resources Laboratory (HRL)) is actively engaged in research on the development of aircrew training simulators but does not pursue the manufacture or distribution of such simulators to operational units. A difficulty exists in that no widely accepted method of evaluating the benefit-cost impacts of these devices is in use. Since these simulators represent significant capital expenditures, a method of evaluating their benefit-cost relationships is desired to help evaluate their usefulness from both a management and a research perspective.

These factors lead to the objective of developing a method of applying benefit-cost analysis to simulators which are designed for implementation at the operational (squadron or wing) level. The simulators to be evaluated are those appropriate for multiship activity and training.

BACKGROUND

In 1990, W. C. Moor, working in conjunction with personnel at HRL, developed a preliminary model for the benefit-cost evaluation of multiship simulator alternatives (Moor, 1991a; Moor, 1991b; Moor and Andrews, 1992). This model, while it shows promise of meeting the objective stated above, is in need of refinement and application testing. Since much of the work presented in this report is an effort to improve the definitions and use of some of the variables required by this model, the names and definitions of these variables are shown in Table 1.

The original model included the capacity to evaluate and compare multiple simulation environments as an explicit element. This study focuses on refining the method of benefit determination for a single simulation environment assuming that this method can be generalized for multiple environments.

Table 1
Variable Identification and Definition
(Extracted from Moor, 1991a)

Performance Area: An operational activity which would be required by a combat pilot and would behaviorally complex enough that training emphasizing its acquisition and maintenance is appropriate. The Performance Area is identified as PA(i) where i refers to a specific performance area.

Emulation Capability of the Simulation Environment: The degree to which the simulation environment represents the actual environment experienced in the aircraft for the specific performance area. The Emulation capability of the Simulation environment is identified as ESIM(i).

Continuation Use of the Simulator: The degree to which a simulator would be used to train in a performance area after initial skill training had been accomplished. The Continuation USE is identified as CUSE(i)

Necessity of Use of the Simulator: The degree to which a simulator must be used to train in a performance area (usually because of extreme hazard/danger or legality of operation). The Necessity of USE is identified as NUSE(i).

This research focuses on the benefit component of this analysis as it is elements of the benefits computation that require the most refinement. The general computational model for benefits determination has not been altered and is shown in Figure 1.

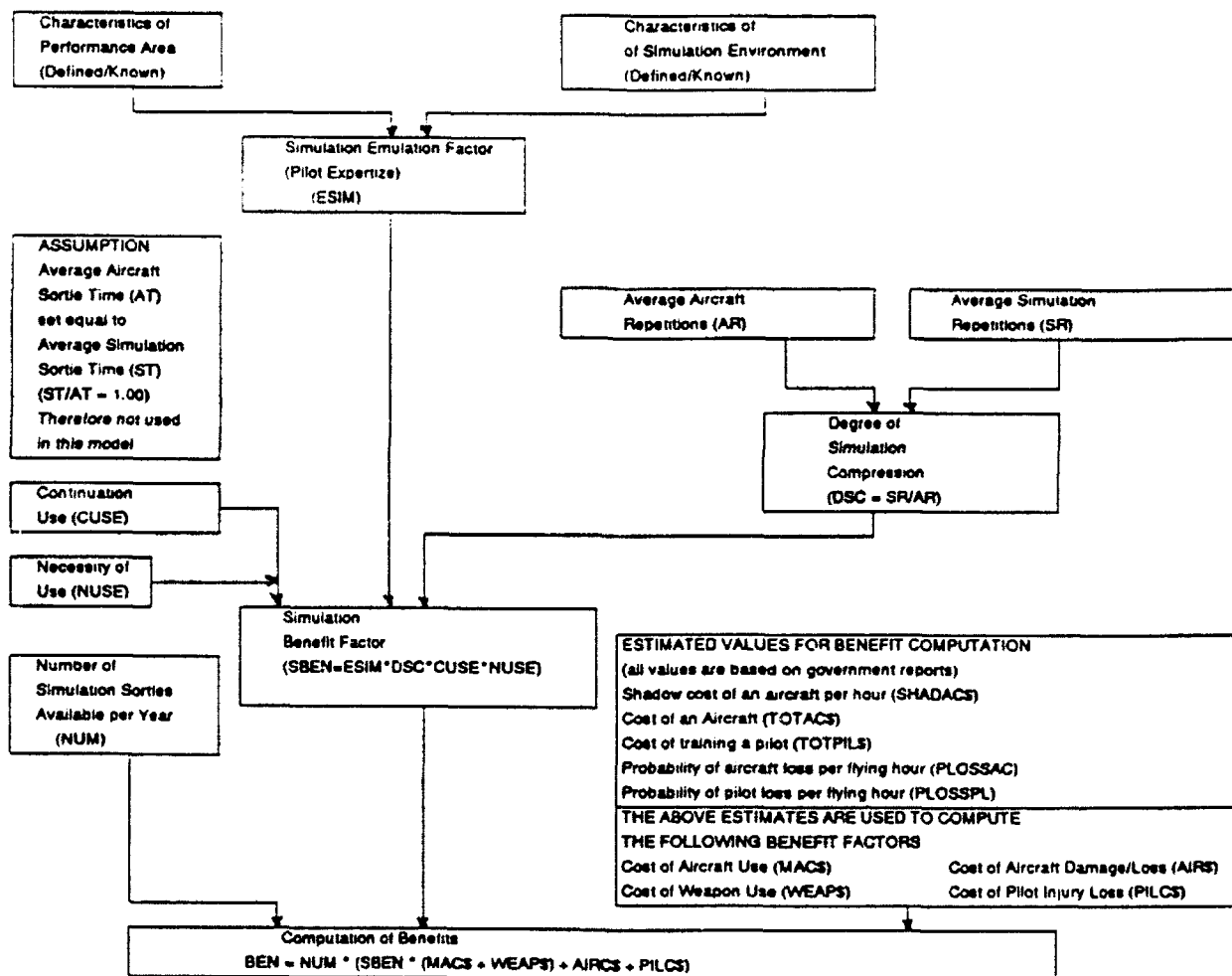
The overall thrust of the model building remains the same as for the original effort. It is the desire of the author to make the computation model as clear as possible to the potential user and to build it in a form that facilitates use. All computational work is placed in LOTUS 1-2-3 spreadsheets that are annotated for data entry and use. The model is built in reference to a specific, operational aircraft (F-15) but is easily modified to allow comparison to any air superiority jet fighter for which multiship simulators would be developed.

BASIS FOR BENEFITS COMPUTATION

The model uses the shadow price of aircraft operation as the basis for benefits computation. It has been argued that this implies a direct trade of aircraft hours flown in exchange for

hours spent in the simulator using the two variable economics utility trade-off curve. This trade-off was never intended, implicitly or explicitly, since the simulators to be evaluated are not presently in use, it can not be assumed that their use would replace the aircraft.

Figure 1
Benefits Computation Flowchart
(Extracted from Moor, 1991b)



Schmid provides a good perspective of this issue (Schmid, 1989). Arguably, in Schmid's discussion of the methods of computing benefit values, the method used here would appear to be the "Cost Saving Method" (Schmid, 1989, p. 66) however, the following argument shows that the use of the shadow price of the

aircraft to form the basis of dollar estimation of benefits is more appropriately seen as the "Intermediate Good Method" (Schmid, 1989, p. 62) and this estimate is a minimum value for the comparison of two different simulators.

For purposes of this comparison, utility may be defined as the degree of "combat readiness". Figure 2 can be seen as presenting, at the squadron level, this utility in terms of the hours of training received, which reflects the syllabus describing the mixture of tasks being trained. The iso-utility curve shown describes the "trade-off" between simulator hours of the simulators currently in use and aircraft hours.

However, for a given budget level, it is reasonable to assume that the best trained pilots possible are being produced (the training syllabus is as good as possible in its mix of tasks trained and flown, for that budget level.). Therefore there is no good reason to "trade-off" aircraft hours for simulator hours.

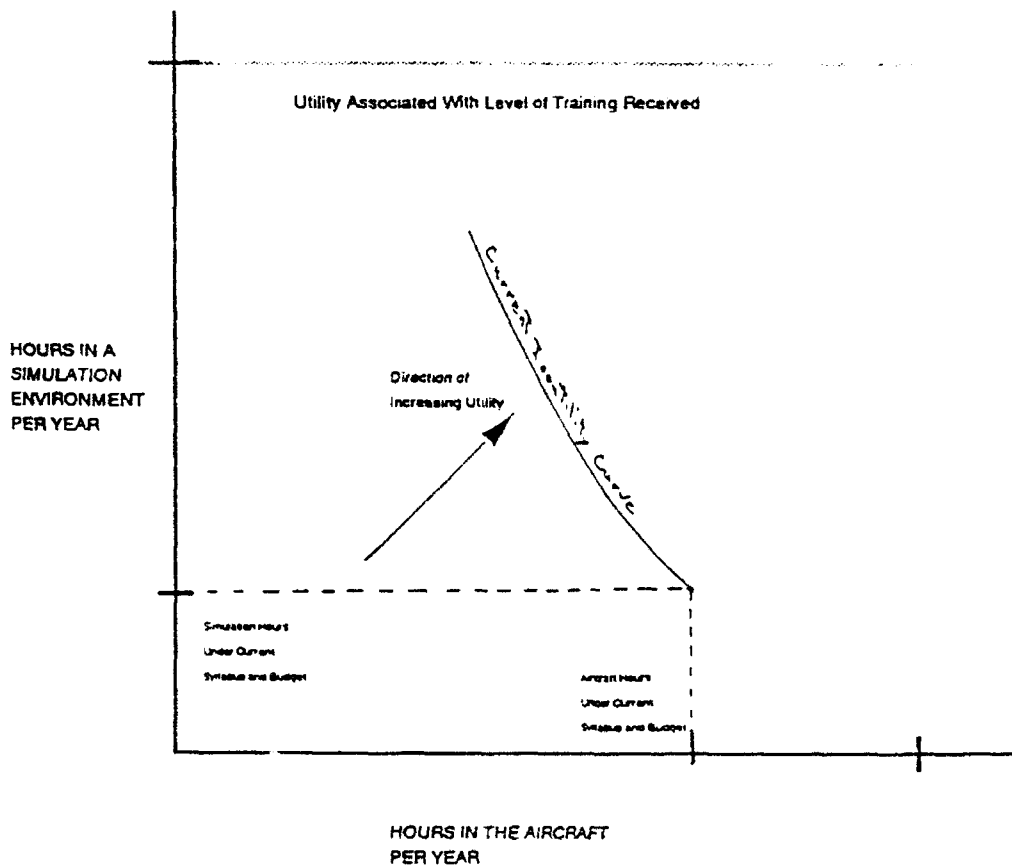
If, with improved models of simulators and for a given level of aircraft flight hours, an apparent increase may be made in the hours of simulator time available (and corresponding changes to the training syllabus made) the iso-utility curve is being changed (upward) yielding a better trained pilot. This "new" iso-utility curve is one that is defined by an increased number of aircraft hours, even though no additional hours are authorized under the budget. This is illustrated in Figure 3.

Therefore, using the current budgeted average cost per aircraft hour as a starting point for benefit computation is justifiable. Cost per aircraft hour is a "savings" (benefit) at the next "higher" utility curve associated with the improved simulators (or simulator system) that allowed the curve. However, if the next higher utility curve actually applied, the cost per aircraft hour would be less due to economies of scale. It appears reasonable, therefore, to start with the current cost and decrease it incrementally to complete a parametric or sensitivity analysis.

PERFORMANCE AREAS

Performance areas were originally defined as identical to the sortie training purposes presented in AFM 51-50 (Dept. of the Air Force, 1982). This was done in an effort to provide mission ready pilots with a reference that would be meaningful to them for comparison to simulators. The original evaluation conducted (interviews with pilots) showed that this was not successful. The pilots interviewed did not see the simulators as total sortie training devices (as they probably should not be) and they could not make the comparisons.

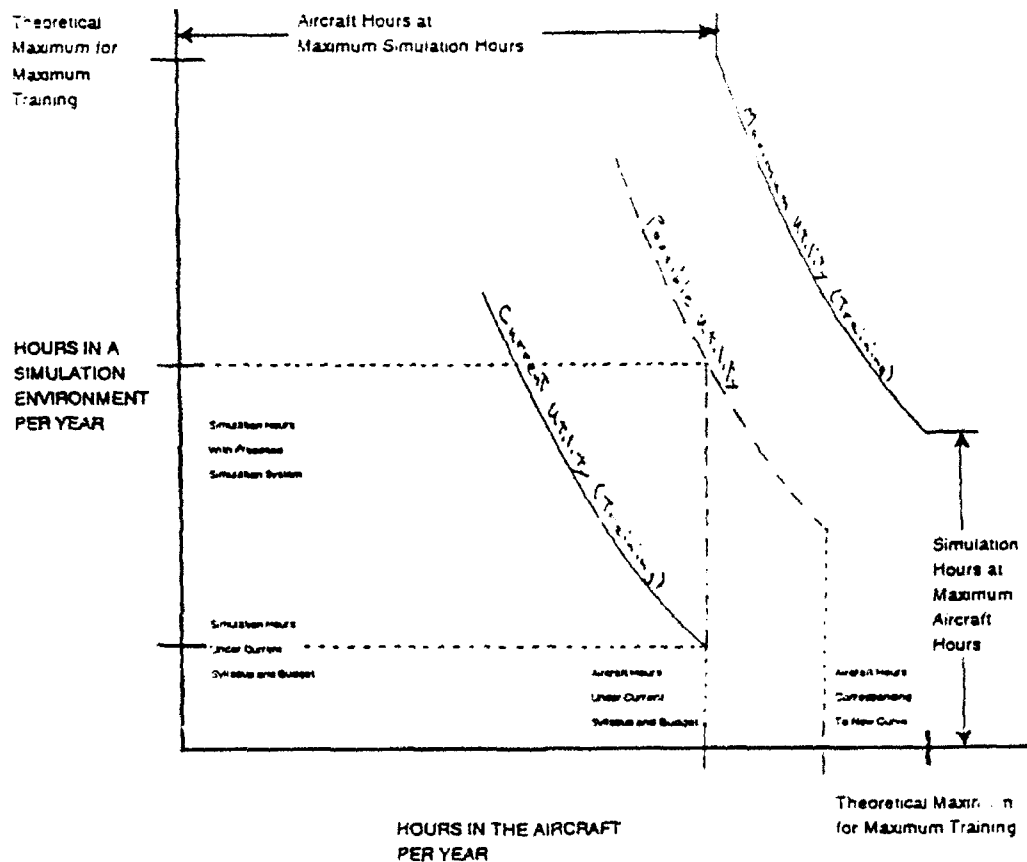
Figure 2
Iso-Utility Curve Illustrating Pilot Training
Under Current Training Conditions



In addition, if the "new" simulators allowed the pilot to be further trained in skills which are currently not in the syllabus or which can be trained only rarely (for example, air combat

skills currently practiced in major exercises available to the pilot once every year or two), then even further changes in the utility curve could be argued. The benefit-cost model, based on the shadow price of aircraft hours, provides a way to choose between competing proposed simulators for the provision of these enhanced skills.

Figure 3
 Iso-Utility Curve Illustrating Pilot Training
 Under "New Simulator" Training Conditions



The purpose of this research emphasizes the operational use of simulators, therefore a comparison that is not meaningful to pilots is not useful. Extensive research conducted using the F-15 simulators at McAir (Houck and Thomas, 1989; Houck, Thomas and Bell, 1991), has yielded an empirically derived set of "tasks"

based on interviews with pilots that are far more practical. These tasks form the current basis for Performance Areas. A list of the names of these performance areas is provided in Table 2. Their definitions may be found in the original sources.

Table 2
Definitions of Performance Areas

Tactics/Mission Planning and Briefing:	Mission Debriefing:
Escort Tactics:	Visual Low Level:
Night Tactics:	Low Altitude Tactics:
Visual Lookout:	Radar Lookout:
Tactical Formations:	Two-Ship Tactics:
Four-Ship Tactics:	Beyond-Visual-Range (BVR) Employment:
All-Aspect Defense:	All-Weather Employment:
Communications Jamming:	Tactical Electronic Warfare System (TEWS) Assessment:
Chaff/Flare Employment:	Reaction to Surface-to-Air Missiles (SAMs):
Reaction to Antiaircraft Artillery (AAA):	Reaction to Air Interceptors (AIs):
Radar Employment/Sorting:	Visual Identification (VID):
Electronic Identification (EID):	Tactical Intercept:
Multibogey, Four or More:	Intraflight Communications:
Electronic Countermeasure/Electronic Counter-Countermeasure (ECM/ECCM) Employment:	

These "new" performance areas have the advantage of being meaningful to pilots who can compare the simulated environment to the environment they experience in the aircraft. In addition, these areas include some that are rarely encountered and not part of the training manual.

The set of performance areas is probably too large (27 areas) to grasp for the purposes of making ESIM comparisons. It is proposed that, when a simulator is to be evaluated, the developers of the simulator select those areas which are best presented by their system. The number created would probably be considerably smaller and would be most advantageous to that particular simulator. A questionnaire to facilitate this selection has been developed using the instructions and measurement scale shown in Table 3.

Once the performance areas are specified by the developer of the simulator, training experts from the USAF would be asked to determine the relative desirability of each of these areas for training. No performance area would be eliminated at this step, but a relative weighting would be obtained which would be used to determine the allocation of training time in the simulator. This

allocation would be computed by multiplying the relative weighting (scaled to a summed value of 1.0) by the number of hours the simulator could be used if it were operational. The relative weightings would be assumed to be maintained for the proportion of scheduled sorties in computing the benefits values. A questionnaire has been developed to facilitate this weighting. The instructions and measuring scale for this questionnaire are shown in Table 4.

Table 3
Selection of Performance Areas
Instructions and Measurement Scale

Considering each of the following air combat tasks, please evaluate the total simulation environment (cockpit, visual, audio, etc.) in terms of its capability to represent the task from the pilots perspective in the F - 15 aircraft.

Please use the following scale when evaluating each task.

- | | |
|--------------------------|---|
| Unacceptable: | The simulator is totally inappropriate for the task, it is possible that negative training could occur. |
| Marginal: | Significant deficiencies exist which require correction before widespread use of the system. |
| Adequate: | System is usable, but could/should be greatly improved. |
| Acceptable: | Only minor deficiencies are noted. |
| Fully Acceptable: | No improvements are required. |

Circle, or place a check mark on the evaluation scale for each task.

DETERMINATION OF NUse AND CUse VALUES

Once the performance areas are selected for each simulator to be evaluated and the relative weights of these areas are evaluated, experts in the training needs for the USAF would be asked to provide an evaluation of the timing of this training for mission ready pilots. Currently there are 4 categories to evaluate each performance area. A questionnaire has been developed to ask the USAF experts how each performance area could be trained in the aircraft. This evaluation is used to provide the basis for numeric estimates for CUSE and NUSE and also a basis for the number of times a particular performance area could be repeated in a simulation sortie versus an aircraft sortie.

Table 4

Selection of Relative Weights of Performance Areas
Questionnaire Instructions and Measuring Scale

EVALUATION OF THE RELATIVE TIME SPENT TRAINING

Assuming the simulator were to be implemented for training at the squadron level and that the following set of tasks were the only ones to be trained using the simulator.

(List of performance areas derived for this simulator)

Show the relative amount of time that should be spent training Task A versus Task B.

Use the following scale:

1. Equal amount of time
2. Barely more time
3. Weakly more time
4. Moderately more time
5. Strongly more time
6. Very strongly more time
7. Critically more time
8. Absolutely more time

If the relative time should be reversed between any two tasks indicate by showing reversing arrows, i. e.

TASK A	TASK B	
TASK A	Relative Time	TASK B
_____	_____	_____

(Tasks shown on a paired comparison basis)

The values for CUSE and NUSE are both set to 1.00 by default, unless and until there is a clear reason to set them to some other value. The following decision rules have been developed to assign these values.

CUSE should be set at some value less than 1.00 when it is known that a particular simulator is only useful to train the initial phases of the performance area learning curve. CUSE would be set at the value of 0.50 if the simulator would not be used to train the performance area at the squadron level. Normally this would occur when the performance area was routinely practiced during any sortie.

NUSE should be set at a value of 2.00 if the performance area cannot be trained in the aircraft at the squadron level (due to legal, technological or other restrictions) although the aircraft is fully capable of performing its portion of the task. The only way this performance area may be trained is through some form of

special exercise or in a non-flying simulator. NUSE would be set to a value of 0.00 if the performance area cannot, and should not, be practiced in the simulator.

These evaluations are made with respect to the performance areas, not with respect to the specific simulator. The questionnaire asks the expert to rate each performance area according to the categories shown in Table 5. The appropriate values for CUSE and NUSE are shown.

These definitions and numeric assignments of values have been developed during the course of this study. They have not yet been validated as the most appropriate ones for the purpose of computing benefits values.

DETERMINATION OF ESIM VALUES

The conceptual definition of the emulation capability of the simulator has not been changed during the course of this study. However, the operational definition has been extended through the development of a measuring scale and a questionnaire to acquire pilot values for ESIM.

An empirical test of the questionnaire revealed the fact that all pilots questioned ($N = 3$) wanted to reserve the right to evaluate the capability of the simulator to represent the performance area in two dimensions. One of these was the dimension of initially acquiring the skills necessary to perform. The second dimension was the maintenance of the skills necessary to perform. Therefore, two questionnaires were developed for this measure. The exact method of pooling this data has not yet been determined (other than an arithmetic average). Clearly this data would also be useful as a supplement to the evaluation of the CUSE values.

The questionnaire instructions and measuring scale for the acquisition of skills are shown in TABLE 6. The instructions and measuring scale for the maintenance questionnaire are identical except for necessary wording changes ("maintenance" replaces "acquisition").

Table 5
Performance Area Categories
Repetition of Task, CUSE and NUSE Values

1. CONTINUOUS - Performance areas which are continually practiced (or ready to use on an intermittent basis) during the course of any sortie. No sortie is designed, necessarily, to practice these performance areas but they are performed (practiced) as needed. e. g. Intraflight Communications. CUSE would be assigned a value of 0.5, NUSE would be assigned a value of 1.0.
2. DISCRETE CLASS I - Performance areas which are performed a finite number (N) of times during the course of a sortie but, logically, would never be performed more than N times during any sortie of approximately equal duration. e. g. Night Tactics, Briefing/Debriefing. CUSE would be assigned a value of 1.0, NUSE would be assigned a value of 1.0.
3. DISCRETE CLASS II - Performance areas which are performed a finite number (N) of times during the course of a sortie planned for their practice. N repetitions are all that can be performed due to physical constraints on the environment, equipment and/or pilot. e. g. Radar Employment/Sorting. There would be a multiplier effect if the simulator, during the course of a sortie, could fully repeat the task N' times. N/N' is the repetition (REP) correction to apply to benefits determination. This must be determined by analysts familiar with the simulator and trainers familiar with the aircraft. CUSE would be assigned a value of 1.0, NUSE would be assigned a value of 1.0.
4. DISCRETE CLASS III - These are performance areas which can be performed during war or under extremely unusual and rarely occurring conditions (special exercises or locations for the sortie). Since this performance area is designed to be performed in the aircraft, a sortie may be specified and planned but it may not be actually practiced in the aircraft. Therefore the sortie design is understood analytically. Strictly speaking there is no multiplier effect accruing due to the use of the simulator (since $N = 0$), but for purposes of comparing different simulators N may be arbitrarily set equal to 1. Then, if this performance area may be practiced one (or more) time per simulated sortie, the benefit factor would be "corrected" by a factor of 1/1, any further practice in the simulator yields a REP number enhancing benefit. CUSE would be assigned a value of 1.0, NUSE would be assigned a value of 2.0.

COMPUTING NUMBER OF SIMULATION SORTIES

The computation of benefits for a particular simulation system is directly dependent on the number of simulation sorties possible. Obviously, this depends on the number of simulators expected to be in use and the operating schedule for that use. It is proposed that, when simulators are to be compared, operating conditions as nearly equal as possible be used (hours of operation, number of simulation cockpits, etc.). The only differences allowed would those technologically intrinsic to the simulator (required maintenance downtime, reliability, etc.).

The number of simulation sorties available may be easily computed based on the organizational configuration being examined through the use of a spreadsheet developed for this project. A portion of the spreadsheet is presented in Figure 4.

The value computed by this spreadsheet (NUM) is used as a basis for comparison and also as a multiplier for the relative

weightings determined for each performance area, therefore its importance cannot be overemphasized. Every effort must be made that completely valid figures be used when preparing the inputs for the computation of NUM.

ESTIMATION OF BENEFIT CONVERSION FACTORS

The establishment of benefits accruing to any simulation alternative is based on a comparison to a specific aircraft. Most of the modeling presented to this point has been as generic as possible. The final conversion of modeled variables to benefit values requires the specification of the aircraft. The original test of this model referenced the F-15 and this reference will be maintained.

1. SHADOW COST OF THE AIRCRAFT - The "shadow cost" of the aircraft (marginal operating cost) is based on USAF budget figures (Dept. of the Air Force, 1988a, 1991).

SHADOW COST OF THE AIRCRAFT = \$5,000

2. WEAPONS USE BY PERFORMANCE AREA - The model includes the capacity to treat as a benefit the cost savings accruing to weapon deployment in the simulator that does not represent the actual consumption of the weapon. This factor is in the model and is available for further refinement but at present it is not used.

WEAPONS COST (PERFORMANCE AREA) = \$0

3. PILOT COST - The possibility of a pilot flying a sortie incurring an accident which leads to death or injury is very real. Therefore, a benefit is computed corresponding to the cost of training the pilot multiplied by the probability of death and/or injury per flying hour. Currently, this benefit factor uses only the cost of training a pilot (General Accounting Office, 1987) corrected by a USAF inflation correction factor (Directorate of Engineering and Services, 1988) multiplied by a rough estimate of pilot death while flying (Dept. of the Air

Force, 1992). The cost of injury could be implemented but is not part of the model.

$$\begin{aligned} \text{PILOT COST} &= (\$7,504,281 \times 1.18028) \times (0.0000205) \\ &= \$182 \text{ per flying hour} \end{aligned}$$

Table 6
Acquisition of ESIM Values

EVALUATION OF SKILL ACQUISITION USING THE SIMULATOR

Consider each of the following Air Combat Tasks and using the following scale, rate the capability of the simulator to train the initial acquisition of the skills necessary to perform the task at the squadron level.

Scale of Measurement

Comparison of Learning to Perform a Task in the Simulator To Learning to Perform a Task in the F-15.

Measurement	Definition
0.00	Absolutely no training/learning potential in the simulator. The task must be trained/learned entirely in the F-15.
5.00	The task can be partially learned in the simulator but must be practiced in the F-15 to be fully learned.
10.00	Perfect training/learning environment in the simulator. The task never needs to be practiced in the F-15. Expectation is that, the first time the task is performed in the F-15, it will be performed correctly.

4. AIRCRAFT COST - The cost accruing due to loss of the aircraft is treated in a similar manner and for a similar reason as the cost of the pilot. The benefit factor allows for the use of cost of damage but this value is not yet implemented. Currently, this benefit factor uses only the cost of replacing the aircraft (drawn from USAF budget figures (Dept. of the Air Force, 1988a, 1988b)) multiplied by a rough estimate of total loss of the aircraft (Dept. of the Air Force, 1992). Since the values are drawn for the F-15 (not F-15E), this must be seen as an arbitrary estimate. The aircraft is no longer being

manufactured and would not be identically replaced if lost.

$$\begin{aligned} \text{AIRCRAFT COST} &= (\$40,000,000) \times (0.0000308) \\ &= \$1232 \text{ per flying hour} \end{aligned}$$

Figure 4
Data Spreadsheet to Compute Number of Simulation Sorties

NUMCALC
FORM FOR THE COMPUTATION OF THE TOTAL NUMBER OF SIMULATION
SORTIES THAT COULD BE FLOWN. SORTIES ARE COMPUTED BY UNIT
LEVEL AT WHICH THE SIMULATOR WOULD BE USED.
THESE CALCULATIONS ARE FOR A SINGLE SPECIFIC TYPE OF SIMULATOR.

		UNIT LEVEL FOR COMPUTATIONS					
		SQUADRON VARIABLE VALUE	RANGE NAME	WING VARIABLE VALUE	RANGE NAME	REGIONAL VARIABLE VALUE	RANGE NAME
ORGANIZATIONAL CHARACTERISTICS:							
	NUMBER OF SIMULATORS PER UNIT	1	NSIMSQ	1	NSIMWG	1	NSIMRE
	NUMBER OF UNITS	1	NUMSQ	1	NUMWG	1	NUMRE
OPERATING CHARACTERISTICS PER UNIT:							
	NUMBER OF HOURS OPERATED PER DAY	8	NHRDYSQ	10	NHRDYWG	12	NHRDYRE
	NUMBER OF DAYS OPERATED PER WEEK	5	NDYWKSQ	6	NDYWKWG	6	NDYWKRE
	NUMBER OF WEEKS OPERATED PER YEAR	52	NWKYRSQ	50	NWKYRWG	52	NWKYRRE
	SORTIE DURATION	1.5	SRTMSQ	1.2	SRTMWG	1.5	SRTMRE
	AVERAGE BRIEFING TIME PER SORTIE (HOURS)	1	BRFTMSQ	1	BRFTMWG	1.2	BRFTMRE
	AVERAGE DEBRIEFING TIME PER SORTIE (HOURS)	1.5	DBRFTMSQ	1.1	DBRFTMWG	1.3	DBRFTMRE
	UTILIZATION RATE (AVAILABILITY PER DAY) (stated as a decimal between 0.00 and 1.00)	80.00%	UTILSQ	90.00%	UTILWG	0.98	UTILRE
OUTPUT VALUES ->							
	NUMBER OF SORTIES PER DAY PER SIMULATOR	2.9333	NSORTSQ	5.925	NSORTWG	6.08	NSORTRE
	INTEGER NUMBER OF SORTIES PER DAY PER SIMULATOR	2	NNSORTSQ	5	NNSORTWG	6	NNSORTRE
	TOTAL NUMBER OF SORTIES PER YEAR FOR THIS UNIT LEVEL (this is used as TOTNUM(I) in the benefits computation)	520	TOTNUMSQ	1500	TOTNUMWG	1872	TOTNUMRE

BENEFITS COMPUTATION MODEL

The total benefits computation model that was originally developed has been modified to include all the factors that are described in this report. The actual computation equations are shown in Figure 1. In addition, editing comments to assist in data entry have been placed on the spreadsheets which are the operational representation of the model. These spreadsheets, due to space constraints, cannot be presented as Figures. The files, names shown and in LOTUS 1-2-3 format, are available to anyone who wishes to examine them.

The benefits spreadsheet allows for up to 8 different performance areas to be named and entered. More performance

areas may be used but the spreadsheet must be modified to accommodate this increase by adding rows to the spreadsheet and duplicating the computational equations as necessary.

The spreadsheet model includes the capability of simultaneously evaluating up to four distinctly different (or four variants) simulator environments. These are labeled with a column in each matrix corresponding to a different environment.

TEST OF THE BENEFITS COMPUTATION PROCEDURE

A preliminary test of the procedure and benefits computation model was conducted as part of this study. The simulator used as a focus was the Air Intercept Trainer (AIT) which is currently in use with several Air National Guard F-15 squadrons and which was developed by HRL. Several of the HRL personnel, who had helped develop and implement the AIT, assisted as "experts" for questionnaire-based input and three pilots, mission ready and experienced with the AIT (although not necessarily F-15 qualified), served as "experts" for pilot input. The procedural steps are described below.

1. The initial step was to administer a questionnaire to two of the HRL experts asking them to select the performance areas, from the list of 27, that the AIT was designed to perform and their impressions of the quality of this performance. Those performance areas which received a vote of "Acceptable" or better from both experts were selected as the evaluation performance areas. These yielded a set of 8 performance areas, named in Table 7, as a basis for the remaining benefits estimation.

Table 7

Performance Areas Selected for AIT Evaluation

Radar Lookout:	Tactical Formation:
Two-Ship Tactics:	Four-Ship Tactics:
Beyond-Visual-Range (BVR) Tactics:	Radar Search/Sorting:
Tactical Intercept:	Multibogey, Two or More:

2. The same two experts then were asked to perform the paired comparison of these 8 areas to determine the relative time that

should be devoted to each of them in allocating simulation use. These proportions were input to the benefits matrix prior to determining the total number of simulation sorties that could be flown.

3. The decision was made to evaluate one pair of AIT simulators implemented into one Air National Guard squadron. The HRL experts, who were familiar with the operating character of this squadron, provided input for the computation of the total number of simulation sorties available.

4. A different HRL manager, who was knowledgeable about the AIT and who had helped develop the initial list of performance areas, completed the questionnaire providing preliminary values for CUSE and NUSE.

5. The three pilots then evaluated each of the 8 areas with respect to the capability of the AIT to provide emulation capability with respect to the aircraft. (It was at this point that the distinction between "acquisition of ability" and "maintenance of ability" was requested and accepted for current purposes.) The averages for each performance area and each use plus a combined average were maintained, yielding a total of three simulation environments to be compared.

6. All experts were polled to determine the expected number of repetitions of each performance area. This same poll yielded average duration of aircraft sortie and simulation sortie. This information was also used to complete the evaluation of the CUSE and NUSE variables.

7. All benefits computation values presented earlier were the ones used.

8. All data was input into the benefits computation matrix for examination.

The final results of this preliminary test are available for examination but are not included in this report due to space constraints. This was not considered a true benefits evaluation of the AIT and it would be misleading to present numerical values that are tentative at best.

CONCLUSIONS AND RECOMMENDATIONS

This study improved and refined the benefits model developed in previous work. It created the operational procedures necessary to acquire all data required for estimating benefits. In addition, the study completed an operational test of these procedures demonstrating their feasibility.

The study did not address additional refinement of the original cost model or the model used to present benefit cost summary information. These models are available as originally developed in 1990.

Obvious areas requiring future research are listed in the study. These include improvements and refinements in the methods of: 1) acquiring and using the CUSE and NUSE values; 2) validating the ESIM values; 3) validating the use of the master list of performance areas; 4) justifying the proportionality values for the number of simulation sorties per performance area; and, 5) determining the general usability of the operational procedures. Additional areas for further development are those cited above.

REFERENCES

Department of the Air Force, "Justification of Estimate for Fiscal Years 1992/1993: Operation and Maintenance, Air Force, Volume II", Biennial Budget Estimates, Submitted to Congress, Washington D.C., February, 1991, AD A236189

Department of the Air Force, Safety Center, Summary Accident Statistics for Fiscal Years 1990 and 1991, Personal Communication, 6/10/92.

Department of the Air Force, F-15 Aircrew Training, Volume VII, Flying Training, TAC/AAC/PACAF/USAFEM Manual 51-50, Volume VII, March, 1982

Department of the Air Force, "Justification of Amended Fiscal Years 1988/1989: Operation and Maintenance, Air Force, Volume 1", Biennial Budget Estimates, Submitted to Congress, Washington D.C., February, 1988a, AD A198257

Department of the Air Force, "Justification of Amended Fiscal Years 1988/1989: Aircraft Procurement, Air Force, Volume 1", Biennial Budget Estimates, Submitted to Congress, Washington D.C., February, 1988b, AD A198259

Directorate of Engineering and Services, "Annual Construction Pricing Guide for FY 90 Program", Department of the Air Force, Headquarters United States Air Force, Washington, D.C., (HQ USAF/LEECD, Pentagon, Room 5D483), April, 1988

General Accounting Office, "Air Force Pilots: Developing and Sustaining a Stable, Combat-Ready Force", Briefing Report, GAO/NSIAD-88-49BR, December, 1987, Washington, D.C.

Moor, W. C., and D. H. Andrews, "Benefit-Cost Model for the Evaluation of Simulator-Based Multiship Training Alternatives", Armstrong Laboratory, Human Resources Directorate, Aircrew Training Research Division, Williams AFB, AZ, AL-TP-1992-0023, Pp. 1-54, July 1992.

Moor, W. C., "Benefit-Cost Evaluation of Simulator Based Multiship Training Alternatives", in United States Air Force Summer Faculty Research Program: Program Technical Report, Rodney Darrah, editor, Universal Energy Systems, Inc., Dayton, OH, 1991a (AD Number: A244517)

Moor, W. C., "Development of Key Variables for Multiship Simulation Benefit/Cost Analysis", Final Report, Report #CRR-92036, Engineering Research Center, Arizona State University, Universal Energy Systems, Inc., Aircrew Training Division, Armstrong Laboratory, USAF, Contract No. F49620-88-C-0053/SB5881-0378, Pp. 1- 95, December, 1991b.

Schmid, A. Allan, Benefit-Cost Analysis: A Political Economy Approach, Westview Press, Boulder, Colorado, 1989.

VOCALIZATIONS OF NATURALLY RANGING
GROUPS OF THE RHESUS MACAQUE

B. E. Mulligan
Professor
Department of Psychology

University of Georgia
Athens, GA 30602

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

VOCALIZATIONS OF NATURALLY RANGING
GROUPS OF THE RHESUS MACAQUE

B. E. Mulligan
Professor
Department of Psychology
University of Georgia
Athens, GA 30602

Abstract

Acoustical recordings and analyses of vocalizations and associated behavior of naturally ranging groups of *Macaca mulatta* were carried out in order to establish a natural basis for understanding the vocal behavior of these animals under captive, laboratory conditions. This was important because vocal behavior can be a strong indicator of emotional condition which, in turn, is reflective of the psychological wellbeing of laboratory-housed research animals, especially non-human primates. Knowledge of normally occurring vocal behavior was an essential first step in the development of an acoustical approach to the assessment of emotionality of rhesus macaques. It was found that adult rhesus utilize seven acoustical categories of calls in communicating vocally under natural conditions. Acoustically different calls also appear to differ in behavioral context and functional significance. Specific types of information transmitted in calls appears to include individual location and identity, distress, solicitations of aid, predator alarm, threat of aggression, defense, submission, and the caller's condition of arousal.

VOCALIZATIONS OF NATURALLY RANGING
GROUPS OF THE RHESUS MACAQUE

B. E. Mulligan

INTRODUCTION

Vocal communication among animals has been for many years a subject of considerable scientific interest, especially vocal behavior of non-human primates (e.g., Cheney and Seyfarth, 1992; Snowdon, Brown, and Petersen, 1982). Our interest in rhesus monkeys, *Macaca mulatta*, is both scientific and pragmatic. We want to establish the vocal repertoire of *M. mulatta*, as it occurs under natural conditions, to serve as a foundation for the evaluation of their "normal" vocal behavior under captive, laboratory conditions. This would provide a non-invasive, objective approach to the assessment of emotionality and, by extension, the psychological health of rhesus macaques employed as laboratory research subjects.

The first acoustical studies of rhesus vocalizations were conducted by Rowell (1962) and Rowell and Hinde (1962) on a captive colony of 22 animals. They distinguished between two general vocal classes, agonistic "noises" and mostly non-agonistic "clear calls". Thirteen years later, Erwin (1975) cited this work as "...the best description available..." of a highly variable vocal system that is extremely difficult to classify on

either acoustical or functional grounds. In fact, Rowell and Hinde (1962) took the position that their acoustical analyses did not separate rhesus vocalizations into exclusive call categories and they presented some evidence of sounds they said to be "transitional" between call categories. Erwin (1975) likewise noted that vocalizations of rhesus may contain mixtures of tonal clear calls with atonal noises. Indeed, he suggested that rhesus vocalizations function primarily to signal the caller's emotive condition.

Rhesus agonistic "screams" were extensively studied by Gouzoules, Gouzoules, and Marler (1984) in a large, free-ranging colony on Cayo Santiago. Contrary to previous findings, Gouzoules *et al.* found that sounds emitted during agonistic encounters were acoustically divisible into five discrete call categories each of which functions as a representational signal. Calls were found to be associated with status of the opponent (dominance rank or matrilineal relatedness) and level of aggression. Because they did not observe that either scream type or probability of subsequent behavior of the receiver varies with the caller's level of arousal level, they concluded that arousal information is not a primary informational component of rhesus screams.

Most recently, Hauser (1992) reported a study on the rhesus "coo" vocalization, a tonal, or clear call that is said to function as a "social contact" signal that contributes to group cohesiveness. Recordings were made from the colony on Cayo Santiago. The call acoustic structure of the call was reported to

consist of closely spaced (about 400 Hz) harmonically related narrow bands below about 10 kHz. Call duration was found to vary between 100 and 650 msec. Remarkably, the members of one matriline were found to produce coos that were consistently more "nasal" than coos emitted by other animals. Nasality was defined acoustically as the level of noise between bands in the call. While it is interesting that nasality appears to be limited to a single matrilineal line, it was not determined whether monkeys respond differentially to this condition.

It is implicit in the above discussion that the vocal repertoire of *M. mulatta* is not yet established. Our own cursory observations in 1991 of vocalizations of laboratory-housed rhesus suggested some differences between their calls and those reported in the literature. It was not possible for us to assess whether these differences were an effect of laboratory housing, or perhaps a result of the improved acoustical instrumentation that we were using. In either case, a careful field study of free-ranging rhesus macaques using modern instrumentation was needed to establish an base of "normality" for comparison with the vocal behavior of laboratory animals.

METHODOLOGY

All recordings of *M. mulatta* vocalizations were made on Morgan Island (operated by LABS, Inc.) off the coast of South Carolina. The area of the island is about 500 acres and contains abundant natural vegetation (large live oaks, pines, cabbage

paimetto, etc.). The colony of free-ranging rhesus on the island is relatively large, approximately 4000 animals in about 32 social groups, although animal density appears to be relatively low. Stations are located around the island where animals are fed daily and provided continuous access to fresh water.

Vocalizations were recorded with a Sennheiser MKH 416 P48 supercardioid/lobe condenser microphone (frequency response 40-20000 Hz) feeding a Sony TCD-D10 Pro II digital audio recorder sampling at 48 kHz. The microphone was held in a shock mount and, for most recordings, it was enclosed in a Sennheiser MZW 60 blimp windscreen. Acoustical analyses were performed using a Sony PCM-2700 recorder to read the prerecorded digital tapes into a Kay CSL model 4300 unit operating at a rate of 51,200 samples per second within a 80486 33 MHz computer with SVGA display. Data was spooled directly from the PCM-2700 via CSL to 90 Mbyte Bernoulli disks and stored as files for analysis. This system permitted temporal, spectral, and three dimensional graphic and numeric analyses of very large vocal data sets.

Data collection was begun in the Spring of 1992 and continued through early Fall. Acoustical recordings were supplemented by video-sound recordings for behavioral analyses and by recorded commentaries of primatologists experienced in field studies of rhesus behavior. Categories of rhesus calls were established entirely on the basis of distinctive acoustical characteristics that were found to be common to groups of vocalizations. Behavioral interpretations of each call category

were derived from field observations made both at the time of recording and later after acoustical segmentation of vocalizations into categories of calls.

RESULTS

Vocal Repertoire of Adult *Macaca mulatta*

1. **Low Bark.** These relatively quiet, low-pitched vocalizations are emitted by females and males when group behavior appears to be relatively stable, but with some small potential for danger, or conflict. In this context, *low barks* often are produced by only one individual at a time in a group, without apparent focus, possibly serving as acoustic markers of the caller's location and condition of mild arousal. Slightly more intense *low barks* that are focused toward a specific source of arousal, e.g., an intruder or an aggressor against a coalition member, may be emitted by one or more individuals.

The low pitch and flutter-like quality of this call results from both its spectral and pulsatile structure. *Low barks* are composed of 2-5 pulses (mean 3.2) each and the pulses are generated at audible rates between 16 and 24 pps (mean 19.4 pps). Individual pulse durations vary between 31 and 54 msec (mean 45 msec). The effect is a low guttural flutter. Low pitch is reinforced by pulse spectral composition; energy is primarily concentrated in two bands, one very low band below about 890 Hz on average and one intermediate band that averages between about

2.5 and 5.9 kHz. Excited callers may expand this band to frequencies as high as 10 kHz. Conversely, the call may be emitted with energy no higher than about 4 kHz. Call durations also appear to covary with intensity, ranging between brief low level calls of about 84 msec and more intense calls of about 165 msec (mean 166.6 msec).

Rate and intensity of vocalization appear to reflect the caller's arousal level and whether the call is specifically directed or not. *Low Barks* may be directed as mild threats toward a particular recipient, or generally broadcast without specific focus. Under apparently mild arousal conditions (e.g., group foraging on ground), *low barks* may be emitted non-directionally in clusters of several calls each with variable silent periods intervening between clusters. In this context, the non-directed *low bark* may signal proximity or spacing and may contribute to maintenance of group cohesion.

In the case of *low barks* directed toward a specific recipient, these calls may be produced at sustained rates of 2 or 3 per second over periods of several minutes by an individual responding focally to the presence of a non-aggressing intruder. Effectively, the calls mark the continued presence of the intruder and may act as group alerting, or warning, signals.. If the caller is suddenly further aroused by aggressive action of the intruder, *low barks* are likely to be replaced by either *threat barks* or *alarm barks*, depending on characteristics of the intruder (e.g., size, sex, rank, group membership, etc.) and the

caller's resulting emotional valence (anger-defense vs fear-flight). If the caller is attacked, its vocalizations are likely to become *screams*.

2. *Gruff Bark*: More intense and temporally compact than *low barks*, these low-pitched, harsh vocalizations are emitted while facing toward the focus of aggression. At their lowest levels *gruff barks* exceed *low barks* in intensity by 10 to 15 dB and the intensities at which *gruff barks* are produced may be varied by as much as 30 dB, a relatively large dynamic range for one type of call. At lowest levels, *gruff barks* may be emitted singly, or at irregular intervals, and may serve as a prelude to imminent physical aggression, especially when head-bobbing, piloerection, and tail elevation are evident in the threatening animal. The more intense *gruff barks* tend to be emitted at rates averaging between about 3 and 5 calls per second and are nearly always associated with physical attack by the caller. Conflicts between adults frequently involve simultaneous production of aggressive *gruff barks* and defensive (or, possibly submissive) *screams* alternating with aid-soliciting *shrill cries*.

The *gruff bark* is a simple, but acoustically distinctive, variation of the *low bark*. The slow pulses of the *low bark* are temporally fused into continuous bands of energy, one band below about 700 Hz on average, and a second broad band which averages between about 2.5 and 13 kHz. This spectral structure differs from that of the *low bark* primarily in the greater average width

of the mid-frequency band, a result simply of the greater average intensity of threat calls; spectra of *gruff barks* emitted at low intensities match pulse spectra of the more intense *low barks*.

Durations of *gruff barks* average about 157 msec. This is about the same as the average of the distribution of *low bark* durations, but there is significantly less variability in the durations of *gruff barks* than *low barks* (standard deviations of 32 and 59 msec, respectively). The amplitude pattern of *gruff barks*, though variable, tends toward sharp rise times followed by slower decays; rise times average about 38 msec (range 21-75 msec), less than a third of the call duration. This asymmetrical onset/offset pattern of relatively stable duration probably reflects vocal limits in the production of high intensity calls.

3. ***Echoic Bark***: Most acoustically intricate of rhesus atonal calls, the structure of *echoic barks* appears to mimic reflected sounds. These intense calls seem to reverberate through the canopy as if reflected by distant and scattered surfaces, perhaps creating an illusion of calls originating from multiple locations and distances greater than actually exist. The effect might be acoustic camouflage, i.e., disguise of acoustic cues to the caller's spatial location, a particularly advantageous feature for a call that signals alarm.

The echoic semblance of this call is probably due to amplitude differences and time intervals between sharply-onset pulses. The call contains several pulses emitted in close

succession at increasing amplitudes. The first pulses (usually two) are brief (mean 36 msec; range 12-42 msec) and precede a more intense pulse of longer duration (mean 71 msec; range 33-101 msec). In many calls, this main pulse offsets directly into a less intense band of noise of about the same duration (mean 88 msec; range 58-143 msec) which effectively extends the main pulse of the bark, but at a lower energetic level. Durations of *echoic barks* (mean 199 msec; range 138-321 msec) exceed on average durations of *low* and *gruff barks*.

Intervals between successive pulses of *echoic barks* are approximately the same, averaging about 28 msec. If these intervals actually were reverberation times, they would correspond to increases in propagation distances of about 32 ft between first and second pulses, and twice this distance between first and third (main) pulses. Unlike the slow and audible pulse rates in *low barks*, pulse rates in *echoic barks* (mean 37 pps; range 34-43 pps) are auditorially fused, but not temporally fused as in *gruff barks*. Fusion of temporally delayed segments in *echoic barks* would contribute to the perception of spatial displacement of these calls. The illusion of reverberation might be reinforced by the substantial increments in pulse amplitudes (10 dB mean increment between first and second pulses; 14 dB mean increment between second and main pulses) and by the associated increments in upper limit of pulse bandwidths (first pulse 0.09-4.84 kHz; second pulse 0.16-13.62 kHz; main pulse 0.09-22.00 kHz; offset noise 1.3-16.3 kHz) which may be heard as an increasingly

loud, rising pitch.

Echoic barks function as alarm signals. They are emitted by males and females preparatory to, and during, escape from frightening and dangerous situations, e.g., approach of a potential predator. They seem to be emitted when the source of danger (rarely a conspecific) is within sight of the caller and they nearly always appear to be associated with escape movements. *Echoic barks* from one animal may excite other group members to orient toward the source of danger and to begin emitting the same call, especially if the group is arboreal. Animals on ground may take to the trees in response to *echoic barks* from the canopy and those in trees may move to higher branches, or the group may flee. Emission of these barks during flights on ground through vegetation appear to be infrequent.

Aroused individuals facing a moderately threatening potential predator may equivocate between aggressive displays and escape movements while alternately emitting *gruff* and *echoic barks*. If distance between the source of danger and the caller is abruptly shortened, equivocation gives way to escape and alarm signals. If the predator does not pursue and if escape succeeds in re-establishing a safe distance, the caller may again resume equivocating between aggression and escape.

4. **Scream:** As its name may suggest, the scream is an atonal, noisy expulsion of acoustic energy over a broad band of frequencies. Elicited by threat of aggression or actual attack,

both the energy content and the spectral structure of *screams* appear to depend on the severity of the aggression. A threatening gesture received from a higher ranking individual may elicit low level *screams* from the recipient. If the gesture is followed by further aggression (lunging toward or chasing), the energy levels and spectral distributions of the ensuing *screams* expand. The defensive individual may emit a series of *screams* interspersed with *shrill cries* (and "looks" for coalitionary support), the number and intensity of its calls apparently depending on the effectiveness of its attempts to escape. If the aggression involves actual physical contact (e.g., slapping, biting), the defensive *scream* usually becomes a super intense acoustic *blast* directed toward the aggressor's face.

Screams may consist of single vocalizations, but usually they are composed of several successive bursts of broadband energy. Occasionally, *screams* may be preceded by *clicks*, perhaps warning of very high stress levels. Tonal *shrill cries* soliciting coalitionary support also are often associated with *screams*, either preceding or interspersed among them in series of calls. *Screams* usually are broadcast directly toward the antagonist while *shrill cries* appear to be emitted with the caller's head turned away, probably in the direction of animals who might provide aid. Alternatively, rather than serving as a counter-aggression, defensive expressions, screams might also signal some degree of submission.

The spectral structure of *screams* is built around a wide

band (2 to 8 kHz bandwidth) of intense energy often displaying irregular frequency modulation within its more intense regions. The lower edge of the high energy band is in the region 2.4-3.2 kHz with little energy below about 1.7 kHz in moderate level *screams*. As *scream* intensity increases, the width of the high energy band expands to higher frequencies (an increase in call intensity by 25 dB corresponds to an expansion in upper frequency of the band from about 5.4 kHz to about 10.6 kHz). Although the upper region of this band usually does not exceed about 12 kHz, lower level noise extends to beyond 22 kHz. In the most intense forms of this call, *blasts*, it appears that the high energy band expands to encompass the entire spectrum which is nearly flat to beyond 22 kHz.

Although the abruptness of *scream* onsets appears to increase as intensity increases (from about 93 to 26 msec), call duration is not correlated with intensity. Calls range in duration between 150 msec to 1.2 sec (mean 436 msec), the length of the *scream* seeming to depend on the severity and persistence of aggression. The nature of the attack also is evident in irregular patterns of amplitude modulation that may be present in calls from an animal experiencing sustained aggression from which it cannot immediately escape. In this case, individual *screams* may consist of 2-5 contiguous ragged pulses of energy. Times between successive *screams*, or *screams* and *shrill cries*, of animals under attack range between about 20-48 msec, a nearly continuous cacophony of intense sound. Such rapid sequences of calls may be

difficult for human listeners to resolve without practice, especially if these calls are mixed with intense *gruff barks* from an aggressing animal.

5. *Shrill Cry*. This piercing tonal call may be emitted as brief chirps or intensely warbling whistles by males or females soliciting coalitionary support while receiving threats or responding directly to physical aggression. In addition to the apparent recruitment-of-aid function of the *shrill cry*, its intense and acoustically complex nature probably makes it an ideal signal for the transmission of location information, as well as the caller's individual identity and condition of distress. The tonal nature of *shrill cries* is due to the banded structure of their energy spectra which we usually hear as having pitch. Though adult *shrill cries* may occur independently of other calls, often they are emitted in conjunction with atonal *screams*. In these defensive (and/or submissive) sequences, it appears that the caller faces its aggressor as it *screams*, but looks away, probably toward supporters, as it *shrill cries*.

One of the more acoustically variable of the rhesus calls, the *shrill cry* exhibits complex modulation in both amplitude and frequency domains. Tight distributions of energy in narrow bands (range 0.2-3.0 kHz; mean 900 Hz) are spaced roughly at harmonic intervals that co-vary in center frequency and amplitude over the variable durations of the *shrill cry* giving this type vocal expression a sound quality that may range from brief, intense

chirps to more protracted and sharply pitch-modulated whistle sounds.

Patterns of frequency modulation may involve combinations of such pitch-varying effects as abrupt up or down sweeps in bands, harmonic warbling, and variations in sharpness of tuning (width) of bands. The extent of frequency modulation in *shrill cries* exceeds that of other rhesus calls, ranging between 0.7 and 5.5 kHz (mean modulation range 3.1 kHz). Amplitude also may undergo appreciable modulation (e.g., 7-19 dB) during *shrill cry* emission depending apparently on momentary events to which the caller is reacting, as well as its arousal level. Irregular amplitude patterns composed of multiple pulses of different shapes and durations appear to be associated with high levels of distress.

Shrill cries usually contain at least three detectable energy bands separated by intervals of 1.4-5.2 kHz (mean 3.9 kHz). The lowest and most intense band tends to be in the range 1.6-7.7 kHz (mean 5.1 kHz) and averages 17 dB greater SPL than its first harmonic (it is probably this lowest band that determines the pitch heard by humans). The average difference in SPL between the first and second harmonic is only 4 dB, which appears to represent the rate at which the intensity of the higher harmonics declines. However, even if *shrill cries* contain more than the first three bands, some of the higher bands may be missing, or buried in noise. In the case of relatively intense calls, the upper range of the highest band may extend beyond 20 kHz.

6. *Coo*: A clear, tonal call that is emitted at low-to-moderate intensities usually by several individuals in a group under conditions that appear to be behaviorally reinforcing, e.g., conditions involving either receipt of food or removal of threat. For example, sight of food and events anticipatory of food presentation elicit from a group in trees near a familiar feeding place a chorus of excited *coos*. Similarly, if withdrawal of a fear-inducing situation (aggressing intruder on the ground; monkeys in trees) results in cessation of alarm barking, group *coo*-calling tends to follow the apparent reduction of group anxiety. It is not inconsistent with this association between group *coo*-calling and reinforcing conditions to hear occasional isolated *coos* from single individuals in low states of arousal, e.g., while reclining on a shady limb in mid-afternoon heat.

General features of *coo* spectral and temporal structures are relative stable. Energy is distributed throughout the spectrum in very narrow bands (bandwidth: 100-200 Hz) at approximately regularly spaced intervals that range between 400 and 900 Hz, depending on the frequency of the fundamental. However, one or more energy bands may be absent, or at undetectably low amplitudes, in the spectra of some calls. Bands usually are not found at frequencies above about 11 kHz.

The pattern of pitch variation in most *coos* is visible in the patterns of frequency modulation that spectral bands undergo. Pitch gradually climbs to about the midpoint of the call and then declines, the direction and range of modulation increasing

towards the middle of the call. The range of frequency modulation of energy bands increases with frequency, though not with harmonic regularity. Modulation may be negligible in low frequency bands near the fundamental, as much as 700 Hz in mid spectrum bands, and approach 1.6 kHz in upper spectral bands. However, the pattern of frequency modulation in upper bands often is not present in lower bands, and upper band structure may vary in calls successively emitted by the same animal. These differential vocal effects may constitute recognizable individual differences or possibly contextually significant information. In any case, we have firm evidence that individuals may alter the upper spectral structure of successively emitted *coos*. Spectral structure is not fixed.

Call durations, relative amplitudes, and amplitude patterns appear to covary and may be related to differences in general features of social contexts. Although the range of relative amplitudes of *coos* appears narrow (4-6 dB) in comparison to other calls, amplitudes of calls produced by excited animals in anticipation of food tend to be greater than those produced by animals relieved of tension. Likewise, while amplitude patterns of low level *coos* tend to be flat with gradual, symmetrical rise/decay times of about 64 msec and moderate durations (mean 192 msec), more intense *coos* exhibit sharper rise times (about 45 msec) and longer durations (mean 328 msec). The apparent contextual dependency of these acoustic differences may not be valid if, indeed, the differences turn out to be reliable.

7. *Click*: The briefest of rhesus vocalizations, this sound is a single, intense pulse ranging in duration from 12 to 40 msec (mean 26 msec). *Clicks* may be emitted singly or in irregularly distributed pulse trains at average rates of about 44 pps for several seconds at a time. Although most energy is concentrated in a band between about 2.6 and 12 kHz, the spectra of more intense *clicks* are relatively flat with energy extending to beyond 22 kHz. *Clicks* may occur independently of any other calls, or in conjunction with *screams* and *shrill cries*. The short durations and broad spectra of *clicks* would appear to make these sounds readily localizable, especially when emitted as pulse trains. It is possible that *clicks* signal a high level of arousal, perhaps predictive of likely aggressive or defensive reactions if arousal-inducing conditions persist. However, *click* pulse structure offers little room for individual variation and probably would not be a strong signal of caller identity.

Cheney, D.L., and Seyfarth, R.M. (1992). Precis of how monkeys see the world. *Behavioral and Brain Sciences*, 15(1):135-182.

Erwin, J. (1975). Rhesus monkey vocal sounds. In: G.H. Bourne (ed.) *The Rhesus Monkey*, Vol.I. New York:Academic Press, pp. 365-380.

Gouzoules, S., Gouzoules, H. and Marler, P. (1948). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behavior*, 32:182-193.

Hauser, M.D. (1992). Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production? *Journal of the Acoustical Society of America*, 91(4):2175-2179.

MODELLING SELECTIVE BRAIN COOLING IN HUMANS

David A. Nelson
Associate Professor
Department of Mechanical Engineering - Engineering Mechanics

Michigan Technological University
1400 Townsend Drive
Houghton, MI 49931

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September, 1992

MODELLING SELECTIVE BRAIN COOLING IN HUMANS

David A. Nelson
Associate Professor
Department of Mechanical Engineering - Engineering Mechanics
Michigan Technological University

Abstract

Models of heat exchange in the human cavernous sinus between the carotid artery and the jugular vein were developed. The models were used to test existing theories of Selective Brain Cooling (SBC) in humans, as proposed by Cabanac and Caputa (1979). Results show that the heat exchange in the cavernous sinus is insignificant, and that in the carotid/jugular arterio-venous pair is, at most, a secondary means of cooling the brain.

A simple finite-difference steady state thermal analysis of the brain suggests that only slight temperature increases are anticipated in the brain, even under conditions which would be expected to produce severe heat stress. These results tend to cast doubt on the appropriateness of invoking SBC models in humans.

MODELLING SELECTIVE BRAIN COOLING IN HUMANS

David A. Nelson

Introduction

The human head is an important site for heat exchange in both hot and cold environments. It is particularly sensitive to high temperatures, which can produce heat stroke and extensive tissue damage. Some authors (e.g., Hayward and Baker, 1968; Taylor, 1969) have identified specialized anatomical structures for brain thermoregulation and cooling in various species. Specifically, some higher mammals (sheep, e.g.) possess a carotid rete which is, in effect, a heat exchanger which may substantially cool cerebral arterial blood and helping to avoid overheating of the brain in those animals.

Humans, in contrast, do not have a carotid rete. This raises the question of whether there are analogous specialized structures in humans which serve to maintain the brain temperatures within a normal physiologic range. Various authors have devised models of "Selective Brain Cooling" (SBC) in humans which rely on counter-current arterio-venous heat exchange for cooling the brain (e.g., Cabanac and Caputa, 1979; Cabanac, 1986; McCaffrey, et al., 1975; Nielsen, 1988). However, there are no quantitative estimates of the heat exchange rates associated with SBC. Such estimates are essential to establishing or denying the validity of SBC models.

This work described here attempted to model the arterio-venous heat exchange processes associated with SBC to determine whether physiologically significant rates of heat transfer are feasible with this mechanism. Additionally, a finite-difference thermal model of the brain, lacking specialized cooling, was developed and used to determine whether it is necessary to invoke SBC to describe brain thermoregulation.

Methodology

Selective Brain Cooling in humans presupposes significant heat transfer rates can be achieved between arterial and venous blood in the head and neck. Two sites proposed for this heat exchange are the cavernous sinus, where the internal carotid artery transfer some heat to the venous blood in the sinus; and the juxtaposed lengths of the carotid artery (including the common and internal carotid) and the jugular vein. Estimates of the heat transfer rate achievable in both these structures are developed here, along with some parametric analysis of heat exchange in the carotid/jugular pair.

Carotid/Jugular Heat Exchange

The model of SBC proposed by Cabanac and Caputa (1979) relies upon the existence of a mechanism for transferring heat from the warm carotid arterial blood to venous blood draining from the face. The face, which is cooled by evaporation of sweat and by convection from the facial prominences, is perfused primarily via the external carotid artery. Blood which is cooled in flowing over the face would drain through the jugular vein, which runs parallel to and close to the internal carotid artery, which provides a portion of the arterial supply to the brain. Under this model, the jugular vein and the carotid artery are, in effect, a counter-flow heat exchanger the warm carotid blood is cooled by the venous blood which has been chilled in the face. The internal carotid blood which perfuses the brain is thus "pre-cooled" and depresses brain temperature. Absent this mechanism, there is no substantial thermal communication between the cooled regions of the face and the metabolically active brain. Obviously, this mechanism is vital to the viability of the SBC theory.

Estimates of the rate of heat transfer between the carotid artery (internal and common) and the jugular vein were obtained by treating the complex as a dual-pipe, counter-flow heat exchanger. (See Figure 1.)

The rate of heat transfer q between the two fluids of such a heat exchanger may be described in terms of the effectiveness, ϵ :

$$q = \epsilon C_{\min} (T_{h,i} - T_{c,i}) \quad (\text{EQ 1})$$

where C_{\min} is the lesser product of mass flow rate and specific heat $\dot{m}C_p$ for the two fluids. The effectiveness is defined as the heat transfer rate between the two fluids in the heat exchanger, divided by the heat transfer rate in a similar heat exchanger of infinite length. The effectiveness thus represents the ratio of actual to "ideal", or maximum potential heat transfer rates.

For a counter-flow heat exchanger, the effectiveness ϵ is determined according to:

$$\epsilon = \frac{1 - \exp(-NTU(1 - C_r))}{1 - C_r \exp(-NTU(1 - C_r))} \quad (\text{EQ 2})$$

where $C_r = C_{\min}/C_{\max}$ and NTU , the "number of transfer units" is a dimensionless quantity which is defined in terms of the overall heat transfer coefficient U between the two fluids:

$$NTU = (UA) / C_{\min} \quad (\text{EQ 3})$$

where A is the surface area of one of two vessels and U is the overall heat transfer coefficient between, based on the same vessel area (Incropera & DeWitt, 1990); i.e., UA is the inverse of the total thermal resistance between the two fluids (arterial and venous blood). That total thermal resistance is the sum of the resistance associated with convection heat transfer between the two fluids and the respective vessel walls, plus the resistance associated with conduction heat transfer through the vessel walls and the intervening tissue between the vessels:

$$\frac{1}{UA} = \frac{1}{h_v A_v} + \frac{1}{h_a A_a} + R_{\text{cond}} \quad (\text{EQ 4})$$

where h_v and h_a are the venous and arterial convection coefficients and A_v and A_a are the respective vessel surface areas (based on vessel inside diameter). The conduction thermal resistance R_{cond} represents the heat flux impedance associated with conduction through the vessel walls and the intervening tissue. See Figure 1.)

The conduction resistance R_{cond} between the two vessels can be estimated from the known solution for conduction between circular cylinders of arbitrary diameter in an infinite medium (Lemons, et al., 1987). This yields a an expression for the thermal resistance of the form:

$$R_{\text{cond}} = \frac{2\pi kL}{\text{acosh} \left(\frac{4w^2 - D_1^2 - D_2^2}{2D_1 D_2} \right)} \quad (\text{EQ 5})$$

where k is the thermal conductivity of the medium separating the two fluids (i.e., the vessel walls and intervening tissue), D_1 and D_2 are the diameters of the two vessels and w is distance separating the centers of the vessels.

The convection coefficients on the arterial and venous sides (h_a and h_v in Eq 4) were estimated for various flow rates and vessel lengths by employing the Seider-Tate correlation for the Nusselt number as a func-

tion of the Reynolds number (Re_D) and the Prandtl number (Pr) for thermal and momentum entry-length problems in tube flows: (Incropera & DeWitt, p. xx)

$$\overline{Nu}_D = 1.86 \left(\frac{Re_D \cdot Pr}{L/D} \right)^{0.333} \left(\frac{\mu}{\mu_s} \right)^{0.14} \quad (\text{EQ 6})$$

where the Nusselt number $\overline{Nu}_D = \bar{h}D/k$ and L and D are the vessel length and diameter, respectively. The quantity μ/μ_s is the ratio of fluid (i.e., blood) viscosity at the bulk temperature to the viscosity at the vessel surface temperature. This ratio is an adjustment factor to account for fluid property variations with temperature. For the small temperature differences encountered in this application (no more than a few degrees Celsius), this factor is, effectively, unity. Thus, Eq. 6 can be simplified to :

$$\overline{Nu}_D = 1.86 \left(\frac{Re \cdot Pr}{L/D} \right)^{0.333} \quad (\text{EQ 7})$$

A parametric study of the estimated rate of arterio-venous heat transfer was performed by evaluating Eqs. (1) through (5) and Eq. (7). Estimates were based upon assumed flow rates in the two vessels, estimated vessel lumen diameter, wall thickness and separation distance. Tissue and blood properties (density, viscosity, thermal conductivity) were estimated from values in the literature (Bowman, et al, 1975; Olsen, 1985). A FORTRAN code was developed to evaluate the model equations from the input data.

Cavernous sinus

Calculations were performed to estimate potential heat transfer rates between interanal carotid arterial blood and the venous blood in the cavernous sinus, consistent with the methodology outline above for the carotid/jugular heat exchange.

Discrete model of the brain.

Models of thermoregulation which incorporate SBC are generally based on the premise that, absent such cooling mechanisms, the human brain is prone to overheating. To test the validity of that assumption, a simple steady-state, two-dimensional finite-difference thermal model of the brain was developed. The model was based on the Pennes (1948) bio-heat equation for perfused tissues. The model consisted of four hemispherical layers: deep brain (gray matter), cerebral cortex (white matter), the skull and an outer, dermal layer incorporating the underlying fat and fascia. By invoking symmetry, the domain of the problem was reduced to quarter of a sphere. The discretized equations were solved using Gauss-Seidel iteration to obtain the temperature estimates at the selected sites. Temperatures were calculated at eight locations, as shown in Fig. 2.

A constant-temperature boundary condition was applied along the lower surface as indicated. A convection boundary condition was applied along the spherical surface, where the local convection coefficient was one of the parameters of interest. The range of coefficients examined was based on the results of Mount (1979). Additional surface heat loss in the form of latent heat could be included by specifying the sweat evaporation rate.

Within each layer, perfusion was assumed uniform. Blood flow per unit volume for each of the four tissue layers, as well as metabolic heat production rates and tissue properties, were estimated from the literature.

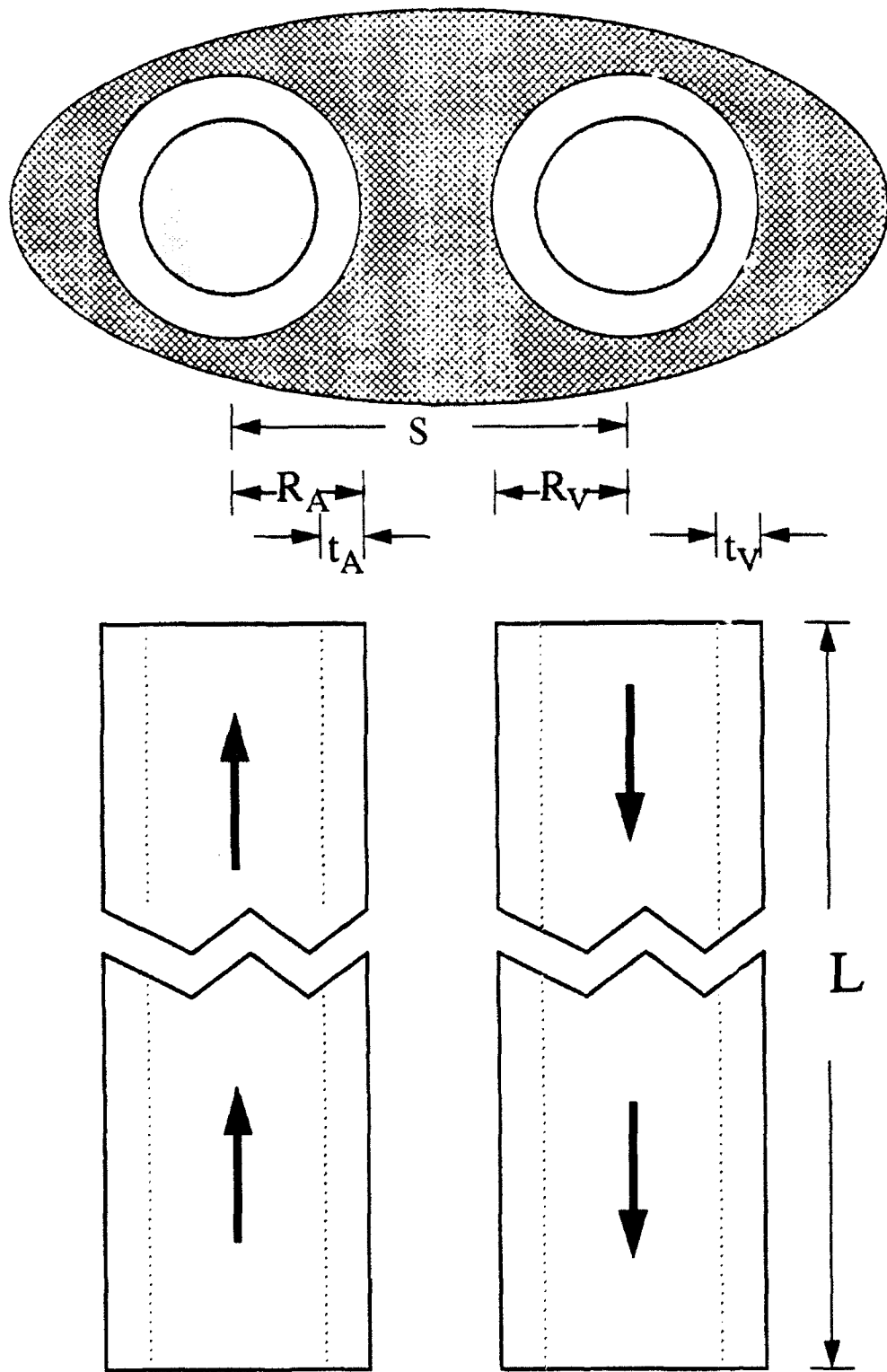


Figure 1. This illustrates the geometry used to model carotid-jugular heat exchange.

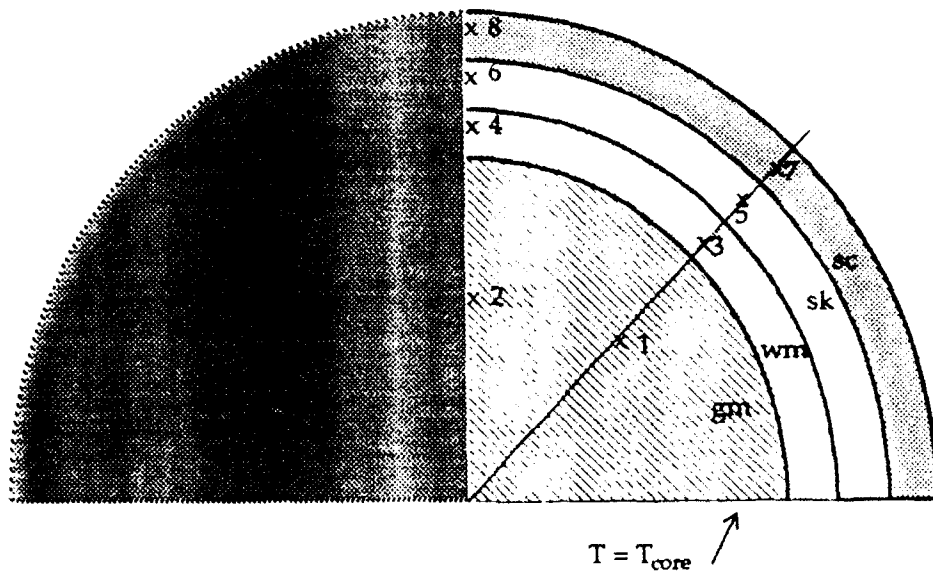


Figure 2. The brain is modelled as a 4-layer hemisphere, consisting of the gray matter (gm), white matter (wm), skull (sk) and scalp/fascia (sc). Symmetry reduces the domain to one-quarter of a sphere. Data are reported for the numbered locations. (Figure is not drawn to scale.)

Results

Heat Exchange in the Cavernous Sinus

Calculations of the heat exchange rates between arterial and venous blood in the cavernous sinus suggest only very limited heat transfer could occur in that region. Based on an estimated sinus length of 2.0 cm and an inside diameter of the internal carotid artery of 0.4 cm, the best estimate of the effectiveness of the heat exchanger is 2.6×10^{-3} ; i.e., the estimated heat transfer rate is less than 0.3% of the potential rate. For an arterial flow rate of 4 ml/s and a temperature difference of 3°C between the two sides, this effectiveness implies an arterio-venous heat transfer rate in the cavernous sinus of approximately .1 W. This is physiologically insignificant.

Carotid-Jugular Heat Exchange

Figure 3 shows a plot of the dimensionless arterial temperature drop $(T_{ao} - T_{ai}) / (T_{vi} - T_{ai})$, as a function of "Heat Exchanger Length" (i.e., the length over which the internal/common carotid artery and the jugular vein are juxtaposed). The quantities T_{ao} , T_{ai} are the temperatures corresponding to the "outlet" and "inlet" on the arterial side of the heat exchanger and T_{vi} is the "inlet" temperature on venous side. The quantity $(T_{vi} - T_{ai})$ thus represents the amount by which the venous blood temperature has been depressed, due to cooling of some portion of that blood by convection and evaporation from the face and scalp. The dimensionless quantity displayed, then, demonstrates the extent to which carotid-jugular heat exchange allows this facial cooling mechanism to affect arterial temperature in the brain. The results shown are based on vessel diameters of 0.4 cm (both vessels), with the center separation ("S" in Figure 1) of 0.6 cm (e.g., the vessel walls have a thickness of 0.1 cm, with the outer surfaces in direct contact with each other). Larger separation distances drastically reduce the heat exchange rate.

The three curves in Fig. 3 correspond to volumetric flow rates of 10 ml/s, 4 ml/s and 1 ml/s in both the carotid artery and jugular vein. Thus, at a flow rate of 10 ml/s in both 20-cm long vessels, the arterial blood

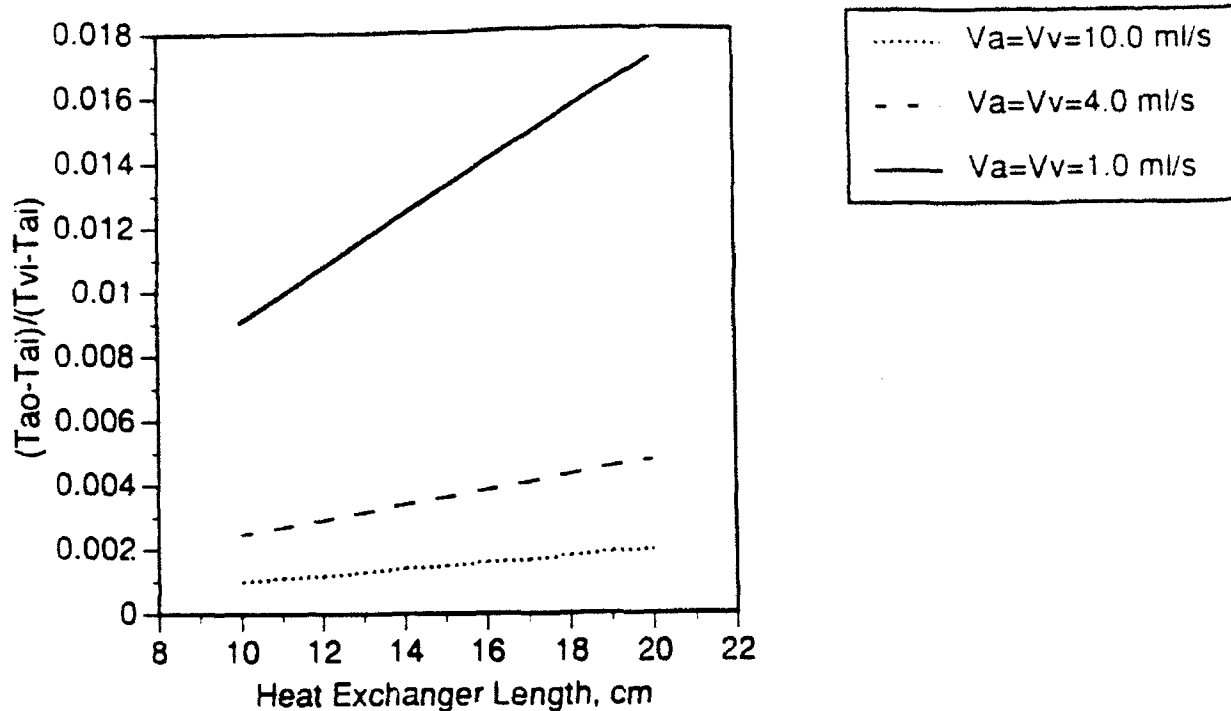


Figure 3. The dimensionless carotid arterial temperature decrease due to heat transfer to the venous blood is estimated as a function of length of the juxtaposed carotid/jugular pair.

will be cooled by 0.018°C for every degree C of temperature difference between the arterial inlet and venous inlet. Other conditions examined demonstrate less of an effect on arterial temperature. (See Fig. 3.)

The results depicted in Fig. 3 show the greatest temperature depression of the arterial blood occurs at low volumetric flow rates. However, the rate of heat transfer between the two fluids shows a very different relationship to flow rate, as illustrated in Fig. 4. The results shown in Fig. 4(a) are for the same geometry as for Fig. 3 ($R_A=R_V=0.2$ cm; $S=0.6$ cm), although for a slightly narrower range of volumetric flow rates (2 - 8 ml/s). The results of Fig. 4(b) are for a similar geometry (vessel inside diameter of 0.4 cm), but with a separation distance $S=0.8$ cm (i.e., $t_A=t_V=0.1$ cm, with 0.2 cm of tissue between the adjacent vessels). The heat transfer rates shown are based on a 1°C temperature difference between arterial and jugular blood.

Estimates of Brain Temperatures

Figure 5 shows representative temperature estimates as a function of convection coefficient for the eight locations of interest. These data are for the case of a normothermic individual (i.e., $T_{\text{core}} = T_{\text{arterial}} = 37^{\circ}\text{C}$) with a sweat evaporation rate of zero in an environment which is at 40°C . The data show that, although there may be a moderate temperature elevation in the superficial regions of the head, brain temperature is only slightly affected under the rather extreme conditions of an individual exposed to a high, hot, saturated wind.

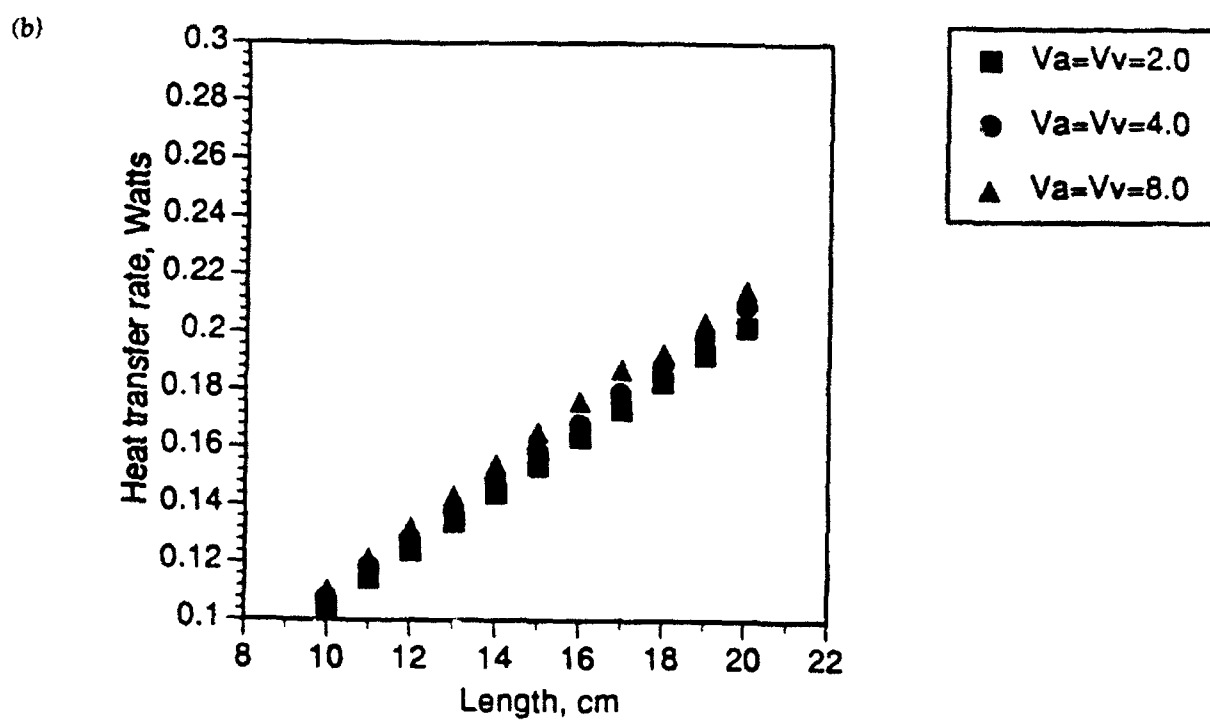
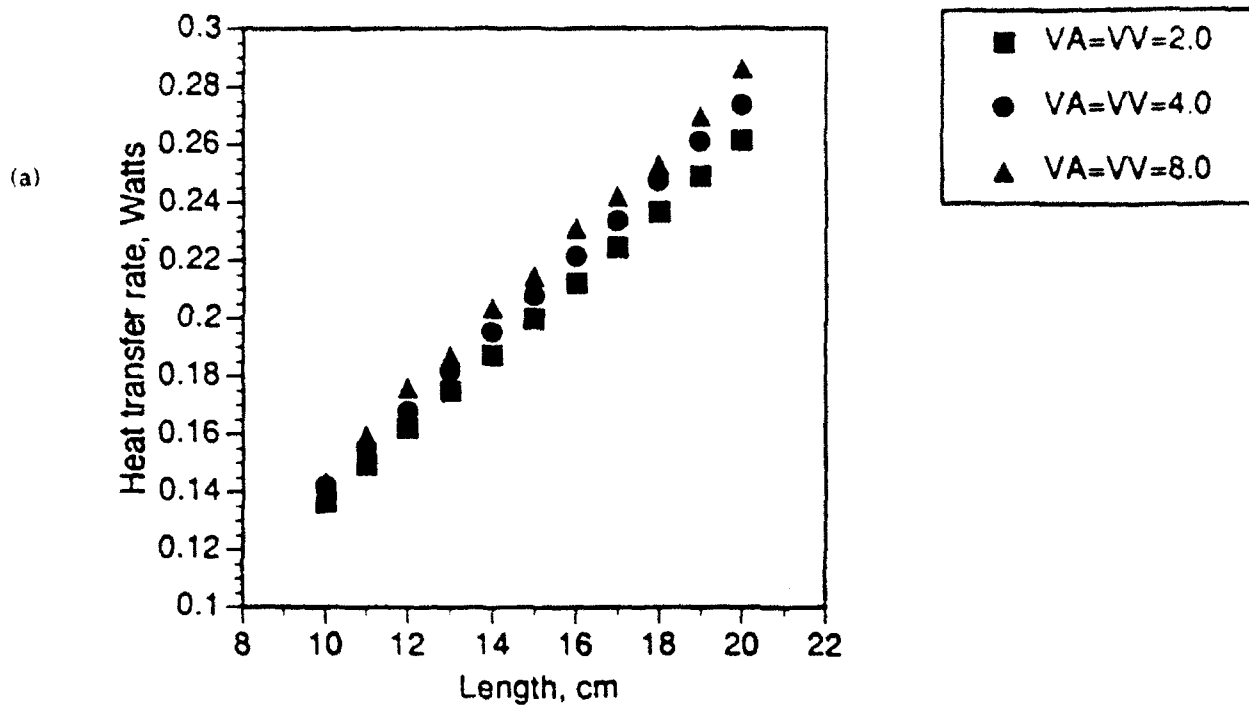


Figure 4. The estimated rate of arterio-venous heat transfer is shown as a function of length of the juxtaposed carotid/jugular pair. In both cases, vessel diameters of 0.4 cm are assumed. In (a), the separation between vessel centers (dimension "S" in Fig. 1) is 0.6 cm. In (b), the separation is 0.8 cm. The volumetric flow rates given are in ml/s.

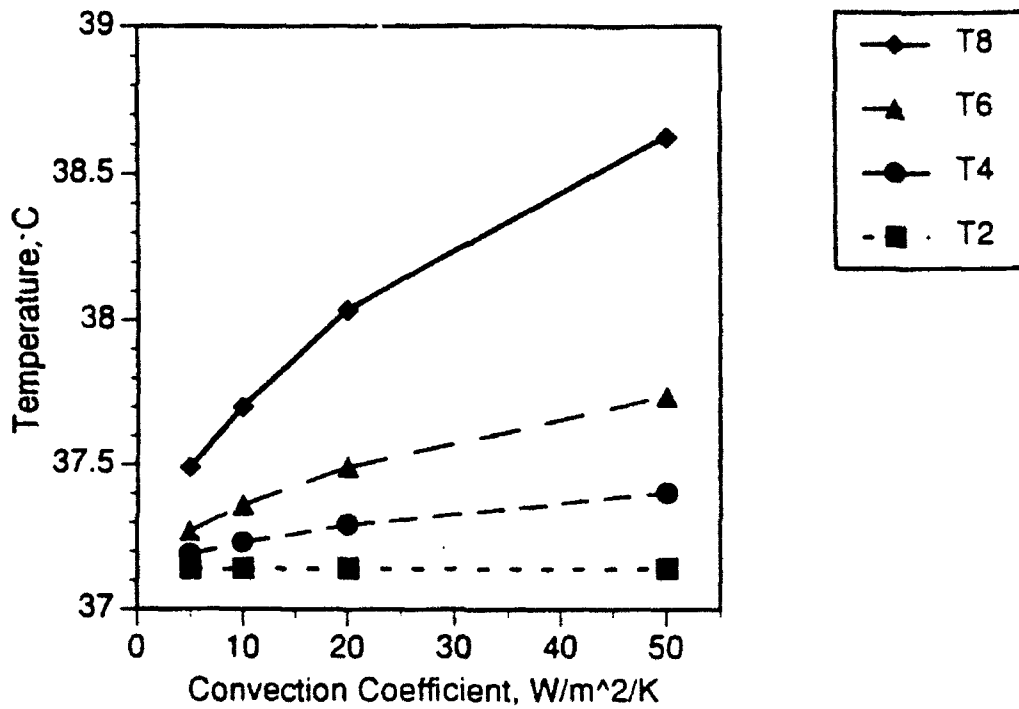
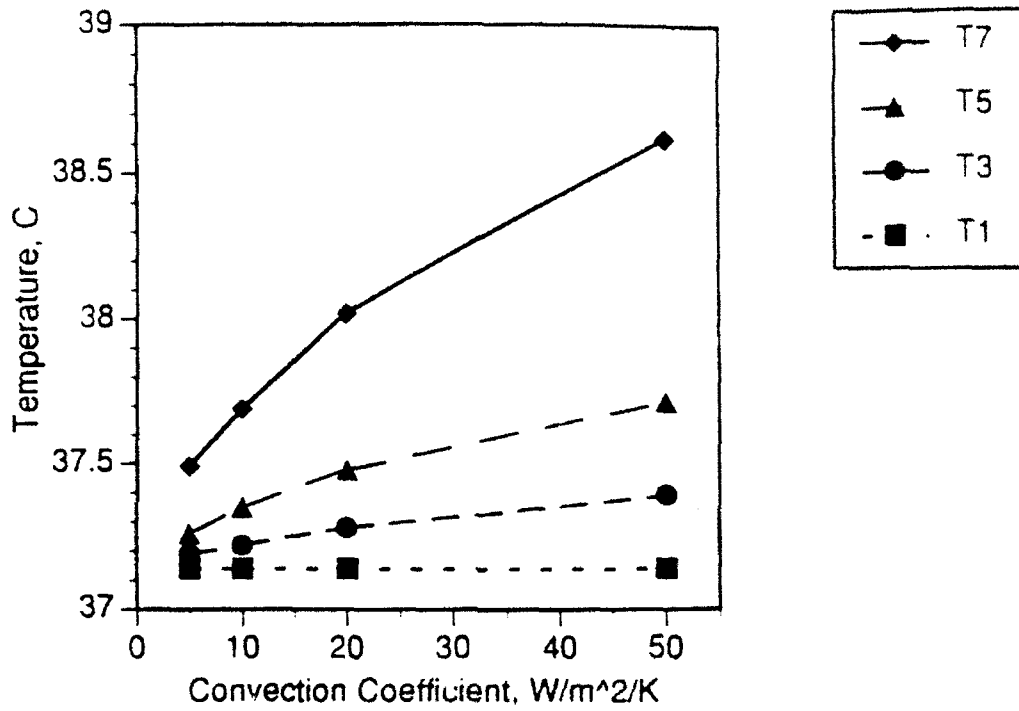


Figure 5. Estimated tissue temperature are shown as a function of the convection coefficient for the eight sites indicated in Figure 2. Results are for a normothermic individual (arterial and core temperatures of 37°C) in a hot environment (40°C) with no sweating.

Discussion

Arterio-Venous Heat Exchange

The results depicted in Fig. 3 suggest a very small cooling effect and low rate of heat exchange between the juxtaposed sections of the carotid artery and jugular vein. Even if all jugular venous blood were cooled by 3°C below arterial temperature, the most it could affect internal carotid temperature would be to decrease it by 0.05°C . As internal carotid flow represents only a portion of the brain's blood supply (the vertebral arterial flow being unaffected by this cooling mechanism), it is questionable whether any measurable effect on brain temperature could be demonstrated.

The rate of heat exchange is also small, even under highly favorable assumptions. (See Fig. 4.) Note that the rate of heat transfer increases only marginally as the flow rates through the vessels are increased from 2.0 ml/s to 4.0 ml/s , and again from 4.0 ml/s to 8.0 ml/s . This indicates the flow rates through the vessels do not significantly limit the heat exchange. The small surface areas available for heat transfer undoubtedly pose a more substantial limitation on the rate of heat transfer. Additionally, comparison of Figs. 3(a) and (b) demonstrates the heat transfer rate is quite sensitive to the separation distance between the vessels.

Given that the results of Fig. 4 are for a single arterio-venous pair and assume a temperature difference of 1°C between arterial and venous inlets, the rate of heat exchange for both vessel pairs (right and left sides) may be in the range $1 - 2\text{ W}$. This is small in comparison to the normal metabolic heat load from the brain, which is estimated at $15 - 20\text{ W}$ according to the results of Tabaddor, et al. (1972). While arterio-venous heat exchange in the neck might be an auxiliary mechanism for cooling the brain, it does not appear that it is a primary means for heat removal.

The extremely low rates of heat transfer from the internal carotid artery to the venous pool of the cavernous indicate this structure does not have a credible role in brain cooling.

Brain Temperature Estimates

The results obtained from the discretized thermal model suggest that, at least in a steady thermal environment, it is not necessary to invoke SBC to obtain adequate protection from hyperthermia in the brain. Even in the absence of any evaporative cooling from the scalp, tissue temperatures in the cortex and deep brain do not appear to be drastically elevated. Whether there is potential danger of large transient elevations of brain temperature is not addressed with this model and will be a point for future investigations.

It should be noted that there is very little variation in temperature in the azimuthal dimension within a tissue layer. This suggests that future analyses may be appropriately performed using either a one-dimensional or a lumped thermal model of the brain.

References

- Bowman, H.F., Carvalho, E.G., Woods, M. "Theory, measurement and application of thermal properties of biomaterial" *Ann. Rev. Biophys. and Bioeng.* 4:43-80 (1975).
- Cabanac, M. "Keeping a cool head" *News in Physiol. Sci.* 1:41-44 (1986).
- Cabanac, M. and Caputa, M., "Natural selective cooling of the human brain: evidence of its occurrence and magnitude" *J. Physiol.* 286:255-64 (1979a)
- Guyton, A.C. *Textbook of Medical Physiology*, Fifth ed. Philadelphia: W.B. Saunders, p. 416 (1976).
- Hayward, J.N. and Baker, M.A. "A comparative study of the role of the cerebral arterial blood in the regulation of brain temperature in five mammals" *Brain Res.* 16:417-440 (1969).
- Incropera, F.P. and DeWitt, D.P. *Introduction to Heat Transfer*, 2ed. New York: J. Wiley & Sons, pp.617-8.
- Lemons, D.E., Chien, S., Crawshaw, L.I., Weinbaum, S. and Jiji, L.M. "Significance of vessel size and type in vascular heat exchange" *Am. J. Physiology* 253:R128-35 (1987).
- McCaffrey, T.V., McCook, R.D. and Wurster, R.D. "Effect of head skin temperature on tympanic and oral temperature in man" *J. Appl. Physiol.* 39:114-118 (1975).
- Mount, L.E. *Adaptation to Thermal Environment*. Baltimore: University Park Press, pp. 70-72 (1979).
- Nielsen, B. "Natural cooling of the brain during outdoor bicycling?" *Pfluegers Arch.* 411:456-461 (1988).
- Olsen, R.W. *Temperature Distributions During Induced Deep Hypothermia and Subsequent Circulatory Arrest: An Experimental and Numerical Study*. Ph.D. dissertation, U. Texas Health Science Center, Dallas (1985).
- Pennes, H.H. "Analysis of tissue and arterial blood temperatures in the resting human forearm" *J. Appl. Physiol.* 1(2): 93-122 (1948).
- Tabaddor, K., Gardner, T.J. and Walker, A.E. "Cerebral circulation and metabolism at deep hypothermia" *Neurology* 22:1065-1070 (1972).

FUNDAMENTAL SKILLS TRAINING PROJECT:
LIFE SCIENCE TUTOR

Carolyn J. Pesthy
Science Department

Douglas Mac Arthur High School
2923 Bitters
San Antonio, TX 78217

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

October 1992

FUNDAMENTAL SKILLS TRAINING PROJECT:
LIFE SCIENCE TUTOR

Carolyn J. Pesthy
Science Department
Douglas Mac Arthur High School

Abstract

Given that we believe that children are this country's most valuable resource, and therefore, that education must be a priority in this country to prepare today's children to be successful in tomorrow's world, the federal government, through the Fundamental Skills Training (FST) project, will be enhancing technological and academic skills in developing literate, life-long learners.

Presently, there are no standards that have been developed and documented on or for science on a national scale. Documentation has been reviewed from different states and national organizations for guidance in this quest. From this, a directory or menu of essential skills and tasks have been developed that incorporate performances which enhance students in becoming successful adults.

The purpose of this paper is to briefly outline the current deficits in American science education and to describe the potential value of advanced educational technologies in remedying these deficits. One particular instructional scenario is described in detail to illustrate this potential.

INTRODUCTION

If the United States is to maintain its competitive edge and technological leadership in the next century, all segments of the population must receive quality education and training in science and mathematics (Rakow, 1984). In addition, because we face an astounding shortage of qualified scientists and engineers especially among minorities and women, we must find new and more successful ways to attract all segments of the population to pursue careers in science, mathematics and technical fields. Because we live in an age of exponentially-increasing technology, the desirability for professionals in technical fields related to science and math has become of paramount importance and a matter of immediate urgency (Burns, 1982). Society cannot be content with supporting mediocrity and the status quo.

The educational goals put forth by the National Science Teachers Association (1982) include a focus in developing scientifically literate individuals who understand how science, technology, and society influence one another, and who are able to use this knowledge in the everyday decision-making process. NSTA sees science as an important element in accomplishing demands of everyday life. Ideally, all segments of our population should have sufficient knowledge based on facts, concepts, conceptual frameworks, and process skills which will enable them to continue to learn and think logically and, ultimately, to value science and technology.

This demand especially for scientific skills, attitudes, and knowledge touches on an important issue. A technologically-

oriented, democratic society cannot develop to its potential with a large section of its population ignorant of science and technology. Attitudes, skills, reasoning abilities and knowledge are requisite to regain a sense of control over the destiny and future of this country.

Because of this deficit in scientific knowledge and reasoning capabilities within our population, U.S. business and industry is increasingly concerned about its ability to compete effectively in the international marketplace (SCANS, 1990). The competitive foundation of any business enterprise is dependent on the productivity and quality of its workforce and consequently, this nation's economic security is at risk when the competitiveness of the economy is directly linked to the quality of this workforce. Therefore, the need to graduate young people who are prepared to function effectively in the workplace of the future has become one of the most alarming concerns of this Nation. Employers find it increasingly difficult to hire literate workers and consequently, corporations are spending more on remedial education in the basic skills. Today's workers must be literate enough to read complex manuals, analyze data and make judgements based on sound, logical, problem-solving basics (Barton & Kirsch, 1990). As the role of advanced technology increases in all types of U.S. industries, so does the workforce's need to support and competitively utilize this technology.

The effort to remediate the existing workforce and maintain critical skills and competencies will require the development of

new technologies and the ultimate transfer of these technologies to needed sectors both in education and business. These technologies can be used to address both the immediate needs of the present workforce but more importantly perhaps, to address the needs of the future workforce... those individuals who are represented in our present-day classrooms.

Many recent studies are critical of the poor performance of American students in basic skills such as reading, mathematics and sciences, including skill requirements to make them comfortable with technology (SCANS). In response to this outcry, the Armstrong Laboratory's Human Resource Directorate has initiated a long-term research project to study the efficacy of applying state-of-the-art intelligent tutoring technology to public education in the hopes of correcting some of these deficiencies.

This project, called the Fundamental Skills Training (FST) project, may hold significant promise in providing teachers with a tool that can effectively support them in providing meaningful education for students lacking in the skills, concepts and basic factual knowledge. The primary goal of the FST project is to research, develop, and transfer three prototype intelligent tutoring systems (ITSS) to the public education sector. ITSS deliver individualized instruction in specified subject matters by applying artificial intelligence techniques. These techniques, unlike most educational software, judge what the student knows and how well the student progresses. In addition, students interact with the ITS rather than merely responding to

it in a specified and passive way. This active participation allows the student to learn by doing. Instruction is then tailored to the student's needs, automatically, with the "expert" instructor coaching as the student engages in the activities. This requires the representation of a domain expert's knowledge, an instructor's pedagogy and a dynamic blueprint of the student who is being taught. With this form of facilitation, the teacher can now provide initial or additional instruction elsewhere in the classroom to individuals who have a greater need.

The first system, initially prototyped in 1990-1991, was the pre-algebra word problem tutor. This tutor was designed to teach students how to solve word problems of all types typical in pre-algebra curricula and to enhance general problem-solving strategies that could be applied to other curriculum areas.

The second tutor, for early high school reading and writing, facilitates development of an individual's writing skills by guiding them through prewriting, drafting, revision and editing phases of the writing process. This tutor uses strategies and tools that structure the students' writing activities and allows them to develop the cognitive skills used in writing. Included are significant reading strategies such as finding key words and making connections, ultimately to move towards summarization and paraphrasing, a skill used in all academic areas.

The third tutor in this project is the science tutor. The goal of the FST science tutor is to enhance the abilities of students who successfully complete their science education in becoming competent in science and having an appreciation of its

endeavors. By "competent in science", the interdisciplinary project team means that each student will be literate, functional, critical and appreciative. To be "literate" denotes the ability to read, comprehend, and write science-related concepts and principles. To be "functional", refers to the ability to use and apply scientific methods and technologies. To be "critical" denotes abilities to evaluate and judge the soundness of scientific approaches and outcomes. To be "appreciative" means to be fully aware of the value and significance of science and technology.

OVERVIEW OF THE FUNDAMENTAL SKILLS TUTOR IN SCIENCE

Virtually all industrialized countries value scientific literacy as a major component in competing successfully in the world market. For the past few decades, it has been a primary concern in the United States that the quality of students entering the workforce continues to decline. Specifically, scientific knowledge and abilities have been comparatively low. Many young adults do not have the background to understand a newspaper article that is scientifically based. Asking a recent high school graduate to analyze a set of data or to make a decision about a specialized scientific procedure would most likely end in frustration.

Mullis and Jenkins (1988) summarized the results of the National Assessment of Educational Progress (NAEP) dealing with science in a report entitled The Science Report Card. They concluded that most students aged 13 and 17 years were highly

proficient in knowledge of everyday science facts and in understanding simple scientific principles. However, a dramatic decline in proficiency occurred when these students were asked to apply basic scientific information, to analyze scientific procedures and data, or to integrate specialized scientific information. As we progress into the twenty-first century, the chief concern of our society is to implement a way to increase scientific competency to a level that is competitive on a global scale.

Apart from the knowledge base that is required for a citizen to be scientifically competent, certain skills are also needed. A sampling of the science curricula of several states reveals numerous processes that form the core of science education in the United States. Although the curriculum guides from California (1990), Florida (1984), Georgia (1986), Pennsylvania (1987), Texas (1991), and Washington (1985) organize these processes in different paradigms, there is a commonality among them. All of them distinguish between the content of science and the process of science. In addition, the processes of science that figure in the different frameworks are fairly uniform, although they may be variously labeled (skills / processes / outcomes / competencies) and variously grouped (domains / goals / criteria / standards). McCormack and Yager (1989) suggest a taxonomy for science education that expands the usual dichotomy of content and process of science by adding three other components: creativity, attitudes, and applications / connections. Thus, from the myriad items that are proposed in the literature as to which skills and

attributes constitute a scientifically competent citizenry, we have constructed an archetype for the Fundamental Skills Tutor (FST) in science.

Goals of the Fundamental Skills Tutor: Science

The goal of the FST science tutor is for students who complete their science education to be competent in science. By "competent in science," the project team means that each student will be literate, functional, and critical in the domain of science. To be literate denotes the abilities to obtain, comprehend, and communicate scientific material. To be functional refers to the ability to utilize the methods, principles, and technologies that pertain to science. To be critical indicates the abilities to assess the soundness of scientific approaches and outcomes and to judge the significance of science and technology in society. (A detailed listing of specific skills is presented in Appendix A.)

Abilities of a scientifically literate citizen

The ability to obtain information is a fundamental skill in science that can be achieved in several ways. Students can gain information from empirical sources through their own senses of sight, hearing, touch, smell, and taste. Empirical sources include observations of natural events, actions of others, or activities that students undertake themselves. They can also gain information from the works of others not directly observed.

These works are documented in books, periodicals, and video and audio recordings. A subset of skills includes being able to

1. utilize printed and non-printed resources,
2. discern similarities and differences,
3. recognize patterns and inconsistencies, and
4. correlate relationships between objects/events.

The ability to comprehend information is an essential skill in science that exists at various levels in the cognitive process. Comprehension begins with understanding the words which convey the thoughts that are conceptualized, organized, and integrated with prior knowledge. A subset of skills includes being able to

1. construct meaning from observations and readings,
2. make inferences and deductions,
3. attain concepts, and
4. link new knowledge to prior knowledge.

The ability to communicate information is a basic skill in science that can take various forms. Most scientific information is relayed in writing, but non-written and non-verbal formats have become increasingly important. A subset of skills includes being able to

1. produce written items such as reports, manuals, graphs, charts, and related documents; and
2. create non-verbal items such as drawings, diagrams, pictures, and related illustrations.

Abilities of a scientifically functional citizen

The ability to utilize science entails both the content and the process. It is a skill that is as relevant to the home and workplace as it is to the classroom and laboratory. A subset of skills includes being able to

1. apply scientific principles to solve problems in everyday life,
2. use scientific methods to conduct standard experimental investigations, and
3. implement scientific technology.

Abilities of a scientifically critical citizen

The ability to assess the soundness of scientific approaches and outcomes is demonstrable in several ways. It involves both selecting pertinent information and evaluating it. A subset of skills includes being able to

1. discriminate between fact and opinion,
2. distinguish science and non-science, and
3. determine the statistical probability of results.

The ability to judge the significance of science and technology in society is one that involves attitudinal components. Human values and feelings, as well as decision-making skills, impact science and society. A subset of skills includes being able to

1. recognize the strengths and limitations of science,
2. weigh the harmful and beneficial applications of scientific knowledge, and

3. justify the conclusions reached or actions taken.

EXEMPLAR MODULE

This exemplar module is a very small part of a much larger simulation based system. The simulation represents a closed, underwater environment. A very small part of this underwater world is the Aqualab. Suppose that within this underwater research station of the future, a jealous fellow scientist has broken into your lab, and made a shambles of everything. As you enter this lab, it has become apparent that this was no mistake, and intentionally done. You look around the lab for clues that would help identify the "culprit". For scientists, learning how to play detective is essential.

In order to play detective, students need some fundamental skill training on obtaining evidence or making observations. This particular module is built around science skill I.A.2: obtaining information from empirical and indirect sources, discern similarities and differences

Goal of module: Skills in acquiring data through the senses

The goal of this module is to have students practice multiple methods of observation. A correct understanding of scientific explanation requires the ability to distinguish what is observable. What flows from these observations are abstract deductions.

Interface Description

The interface consists of 4 sections: information presentation, comparison development for similarities and differences with normal pictures, comparison development with abstract pictures, and cartoon analysis. The information presentation section has windows that students read an introduction and instruction on how they are to proceed. Students will be given pictures or icons of animals, plants, or cartoons. Students should make observations based on similarities or differences in color, size, shape, texture, composition, dimension or method of movement. These pictures may be animated. The following are six example scenarios:

1. Two goldfish will be presented with five differences.

Students will list the five differences. For example, goldfish that are presented should be a different color, ie. one should be a solid gold color and the second should be a white background with orange speckles. The second will also have huge "bugeyes" and fins that are large and "feathery". Students will also be asked to list a certain number of similarities.

If students have problems, alternative presentations will be given to the students. For example, two sharks will be presented. One shark will be a Leopard shark and will have a distinctive caudal (rear) fin. Another shark, a Mako, will also be presented with its caudal fin (forked). Other differences and similarities will be focused upon. In another example, two leaves will be represented. They may

have different venation, margins, shape, color, pattern etc. Pairs such as, porpoises and dolphins can also be used.

2. Four foxes will be presented (or four dragons) with a greater number of differences. One dragon could produce smoke while some of the others "breathe" fire, while similarities and/or differences are being distinguished.
3. Students will be given cartoons and asked to find some number of differences in the cartoons or pictures.
4. Students will be given cartoons and asked to find what items do not belong in the picture.
5. A series of unrelated pictures or cartoons are given. Then a series of separate items are presented. Students are asked to place the items into the picture in which they correctly belong.
6. Unfamiliar items will be given, such as graphs.
Similarities and differences will be expected here also.

Instructional Module

A series of questions will be asked to the student through lines of reasoning and conclusions. Most questions require responses in words, symbols, or numbers chosen by the student. The computer might recognize a verbal response by searching for specified combinations of key words or phrases. Graphic displays are used throughout; objects move across the screen; lightbulbs light with various degrees of brightness; the student uses the built-in pointer to point to places in the cartoons or pictures in answer to some of the questions.

Analysis of verbal responses is prepared for in the writing and editing of the dialog. Real person-to-person dialog is used to anticipate the general course of the dialog or responses: additional unanticipated answers are collected and edited into the program through trial runs. A correct answer allows the students to continue in the main sequence; an incorrect answer is dealt with in a remedial sequence. The instructional module will check for the correct responses against possible student responses either through a drawing template or word template. At the end of the sequence, a test will be given where students are asked to study a picture for a certain length of time. The picture will then be removed from the screen and detailed questions will be asked about the particulars of the picture.

Information Presentation

Science is somewhat like playing "Clue" or playing Sherlock Holmes in a detective mystery. A scientist, like a detective, is constantly gathering evidence (facts) either directly from what he/she can pick up from the scene of the crime (observations) or indirectly, perhaps from the lab. Curiosity is typical of a scientist. Curiosity is a type of inquisitiveness which scrutinizes everything, including the commonplace and familiar, with the questioning attitude of why or how.

Testing and Evaluation

The number of hints and type of hints should be determined by the tutor development team. A hint would be constructed so as

to offer guidance, rather than the answer. An example is: "You have correctly identified 3 out of 5 possible differences of similarities. Search more carefully for the remaining two." Another possible hint is: "Would you like to continue searching on your own, or would you like some help?" If hints are given, because the student has failed to make an observation about a particular portion of the anatomy, the tutor could show a close-up or use a circle to pinpoint the spot where the student should be looking. If another hint is required, a list of tail characteristics could be presented so student could choose from the list. At the same time, positive incentives should be built in such as, as more differences are correctly chosen, fish could swim away.

In each of the basic skills provided the student, an example will be given, perhaps by having the icon flashing. The student will then be asked to perform the same skill. (A timed element could be introduced if students are catching on quickly). If students are having trouble, the teacher might suggest a "logical" approach to the student's observational process; perhaps a pattern needs to be developed to help the student be more successful. Heuristics, such as seeking from top to bottom, left to right, or working from the outside of the page to the center of the page, could be introduced.

Technology Applications

Any system developed is only as good as its weakest link. Teachers need to make new technology work in their classroom.

The quality of any computer software may be directly coupled to the incorporation and application of the software to the curriculum offered by the teacher. For example, in the 1920's, education was introduced to a different type of educational plan designed to individualize learning called the Morrison Plan and the Winnetka Plan. Because of this innovative strategy, learning machines gradually took the student, via table-top mechanical devices, down a diversified pathway to learning. These mechanical devices have included television, photocopying machines, film and slide projectors, and the VCR. More recently, the computer, which is a technology that has yet to fulfill its full educational potential, has widened the educational technology spectrum. These adjunct pieces of machinery potentially enhance education and enable the teacher and the students to have access to learning resources previously not imagined. However, these tools are only as good as the teacher chooses to make them. The desire to use these newest technologies is offset by the fear of the unknown. To be used effectively, teachers must be trained, not only on the hardware but the software as well. Often forgotten is the training and practice of strategies and methods involved in integrating this "know-how" with the curriculum. For this technology to truly become theirs, teachers need to totally incorporate the technology to the point of being creative and synthesizing new lessons of their own.

BIBLIOGRAPHY

- Barton, P., & Kirsch, I. (1990). Workplace Competencies: The need to improve literacy and employment readiness. Policy Perspectives. (Report No. 433J47900887). Washington, DC: U.S. Department of Education.
- Burns, M., Gerace, W., Maestre, J. & Robinson, H. (1982). The current status of Hispanic technical professionals: How can we improve recruitment and retention. Integrated Education, 20(1-2), 49-55.
- McCormack, A.J. & Yager, R.E. (1989). Towards a Taxonomy for Science Education. The Science Teacher.
- Mullis, I.V.S. & Jenkins, L.B. (1988). The 1986 Science Report Card: Elements of Risk and Discovery. National Assessment of Educational Progress, Educational Testing Service. Princeton, N.J.: Education Testing Service.
- Rakow, S. (1984). What's happening in elementary science: A national assessment. Science and Children, 22(2), 39-40.
- Recommended Secondary Science Competency Continuum for Pennsylvania Schools. (1987). Pennsylvania Department of Education. Harrisburg, Pennsylvania.
- Science Curriculum Guidelines Framework. Guidelines for Science Curriculum in Washington Schools. State Board of Education. Olympia, Washington.
- Science Framework. (1990). California State Board of Education. Sacramento, Calif.
- Staff. (1990, October) Skilled, trained, and educated workers...Our best hope for tomorrow. Texas Business Today. pp.6-11
- Texas Essential Elements. Texas Education Agency. State Board of Education. Austin, Texas.
- Vogel, D. (1991). Examining Ethics. The Science Teacher, pp. 38-42.
- What work requires of schools (1991). A SCANS Report for America 2000. Washington, DC: U.S. Department of Labor.
- Whitney, R. & Urquhart, N. (1990, March) Microcomputers in the mathematical sciences: Effects on courses, students, and instructors. Academic Computing.

APPENDIX A

- I. A scientifically literate citizen will be able to do the following things effectively.
 - A. Obtain information from empirical and indirect sources.
 1. Utilize printed and non-printed sources.
 - a. Deal with the features of a textbook and their organization.
 - b. Access periodical literature from indices.
 - c. Select printed references appropriate for the task.
 - d. Conduct an informative interview.
 2. Discern similarities and differences.
 - a. List likenesses of objects or events.
 - b. Describe differences of objects or events.
 3. Recognize patterns and inconsistencies.
 - a. Organize a hierarchy by rank.
 - b. Arrange a series in order.
 4. Correlate relationships between objects/events.
 - a. Describe direct and inverse relationships.
 - b. Establish a cause-and-effect relationship.
 - B. Comprehend information in verbal and non-verbal form.
 1. Construct meaning from observations and readings.
 - a. Ascertain meaning from context clues.
 - b. Deduce meaning from root words and affixes.
 2. Make inferences and deductions.
 - a. Syllogize conclusions.
 - b. Interpret texts and diagrams.
 3. Attain concepts.
 - a. State generalizations of texts or diagrams.
 - b. Formulate consistent models.
 4. Link new knowledge with prior knowledge.
 - a. Create metaphors.
 - b. Write analogies.
 - C. Communicate information in verbal and non-verbal modes.
 1. Produce written documents.
 2. Create non-verbal items.
- II. A scientifically functional citizen will be able to do the following things effectively.
 - A. Apply scientific principles in everyday life.
 1. Troubleshoot problems.
 - a. Identify the nature of the problem.
 - b. Identify the cause of the problem.
 2. Generate potential solutions.
 - a. Brainstorm the possibilities.
 - b. Select the most effective one.

- B. Use standard scientific methods.
 - 1. Execute confirmation investigations.
 - a. Follow explicit instructions.
 - b. Produce expected results.
 - 2. Conduct experimental inquiries.
 - a. Ask valid questions.
 - b. Form a testable hypothesis.
 - c. Design a controlled experiment.
 - d. Manipulate variables.
 - e. Gather, record, and present metric data.
 - f. Analyze results.
 - g. Draw conclusions.
 - h. Accept or reject hypothesis.
- C. Implement scientific technology.
 - 1. Choose technology appropriate for the task.
 - 2. Employ tools and techniques correctly.

III. A scientifically critical citizen will be able to do the following things effectively.

- A. Assess the soundness of scientific approaches and outcomes.
 - 1. Discriminate between fact and opinion.
 - 2. Distinguish science from non-science.
 - 3. Apply statistical procedures to data.
 - a. Interpret statistical data from different types of graphs.
 - b. Determine the use of mean, median, and mode.
 - c. Understand statistical significance.
 - d. Distinguish between possibility and probability.
 - e. Discriminate between validity and reliability.
 - f. Explain the role of sampling in statistics.
- B. Judge the significance of science and technology in society.
 - 1. Recognize the strengths and limitations of science.
 - 2. Weigh the harmful and beneficial applications of scientific knowledge.
 - 3. Justify the conclusions reached or actions taken.
 - 4. Apply ethics in science and society.
 - a. Record and report all scientific findings honestly.
 - b. Present and interpret scientific issues that may be socially controversial in an unbiased manner and within the range of scientific acceptance.

MODEL DEVELOPMENT FOR USE IN
HYPERBARIC OXYGENATION THERAPY RESEARCH

Edward H. Piepmeier, Jr.
Assistant Professor
College of Pharmacy

University of South Carolina
Columbia, SC 29208

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington D. C.

August 1992

MODEL DEVELOPMENT FOR USE IN
HYPERBARIC OXYGENATION THERAPY RESEARCH

Edward H. Ptepmeter, Jr.
Assistant Professor
College of Pharmacy
University of South Carolina

Abstract

One in vivo model and one in vitro model were developed. A rabbit model was developed for use in determining the effectiveness of hyperbaric oxygenation on muscle regeneration following beta-streptococcal induced myonecrosis. Experimental results indicate that a rabbit model has been defined which presents beta-streptococcal induced myonecrosis similar to that exhibited in two cases of streptococcal infection presented at Wilford Hall Medical Center. This model will be used in future experiments to determine the efficacy of hyperbaric oxygenation upon beta-streptococcal induced myonecrosis, and to identify any detrimental effects of co-administration of ibuprofen during beta-streptococcal induced myonecrosis. An in vitro model was developed for determination of the interaction of hyperbaric oxygenation with glucose analyzers' ability to measure blood glucose levels both with and without hyperbaric oxygenation. Five glucose analyzers which measure glucose based upon a glucose oxidase reaction were examined in the initial study. There was a significant difference between the readings each glucose analyzer gave both under normobaric and hyperbaric conditions for all five blood concentrations tested ($p < 0.01$). There was a significant difference between the measurements given by each glucose analyzer under normobaric and hyperbaric conditions for three of the blood glucose concentrations tested ($p < 0.01$).

MODEL DEVELOPMENT FOR USE IN HYPERBARIC OXYGENATION THERAPY RESEARCH

Edward H. Piepmeier, Jr.

INTRODUCTION

Models can be developed to (1) describe complex dynamic processes in biomedical research, (2) test hypotheses, (3) improve predictive abilities, (4) extrapolate experimental observations to untested conditions, and (5) minimize the use of experimental animals in an effective manner. The beta-streptococcus rabbit model addresses all of these goals and will assist in the development of new treatment strategies. To date, studies involving the effects of hyperbaric oxygenation (HBO) upon muscle regeneration have been limited but valuable. It is hypothesized that HBO will help the muscle regenerate quicker than it would under non-treated conditions. The glucose analyzer model identifies possible sources of inaccuracies in the use glucose meters which are presently on the market. Hyperbaric medicine will provide improvement in diabetic therapy if the need for glucose monitoring under hyperbaric conditions can be met. Most portable glucose analyzers currently on the market are based upon the action of the enzyme glucose oxidase. Glucose oxidase catalyzes the oxidation of glucose in the blood in the presence of oxygen in the atmosphere, producing gluconic acid and hydrogen peroxide. The rate of the reaction may then be determined by measuring the amount of evolution of hydrogen peroxide. The rate of the glucose oxidase reaction is dependent upon the concentration of glucose if the presence of oxygen in the atmosphere is known and remains constant. Therefore, erroneous blood glucose measurements may be obtained under hyperbaric therapy.

METHODOLOGY

Animal Model - Beta-streptococci was isolated from two patients who were presented at Wilford Hall Medical Center with massive necrotizing fasciitis and myonecrosis. The virulent strain was cultured for use in developing a rabbit model. Two controls were used in the development of this model. One rabbit was injected with twenty milliliters of the broth used to

culture the bacteria containing no bacteria or endotoxin. The second rabbit was injected with twenty milliliters of the broth in which bacteria was originally present but later removed by filtration. This second sample contained endotoxin produced by the bacteria. Rabbits were then injected with broth containing 10^3 , 10^5 , 10^6 , and 10^8 bacteria. Daily blood samples were used to perform chemistry screens, and blood cultures. In addition, the rabbits were evaluated for their overall condition in terms of ability to apply weight to the affected limb, body temperature and respiratory rate. The size of the necrotic area was measured and photographed. Upon termination of the animal a necropsy was performed upon the affected tissue. Based upon these observations, the rabbits which received 10^5 and 10^6 bacteria exhibited similar time courses of infection as did the patients admitted to Wilford Hall Medical Center.

Glucose Analyzer Model - Five glucose meters; Hemoque®, Exactech®, One Touch®, Glucometer®, and the Companion® were used to determine blood glucose levels in normobaric conditions, and in hyperbaric conditions. The strips used by the glucose meters were separated into three treatment groups. Group 1 served as the control group and the strips were maintained in the original containers at ground level pressure. The second group of strips was maintained in the original containers at hyperbaric pressure. Group 3 was maintained in the original containers which were placed into a metal container before being placed at hyperbaric pressure. All groups were maintained in their respective treatment groups for a period of one month. Each glucose analyzer was calibrated with high, low, and normal glucose controls prior to testing of the blood samples. Glucose was added to five aliquots of blood to obtain approximate glucose concentrations of 250, 150, 100, 50, and 25 mg/dL. Each blood sample was measured ten times each under normobaric and hyperbaric conditions by each of the five glucose analyzers. This analysis was performed for each of the three treatment groups of strips described above.

RESULTS

Animal Model - At the time of this writing only subjective data was available indicating the

10^5 and 10^6 beta-streptococci bacteria injected into a rabbits hind limb produced a suitable model. It is anticipated that tests currently being performed will provide objective data necessary to further define this model.

Glucose Analyzer Model - There was no significant difference between the glucose analyzers' measurement of glucose when using each of the three treatment groups of strips. In addition there was no significant difference between normobaric and hyperbaric readings of all glucose analyzers tested for the blood samples containing approximately 100 and 50 mg/dL. There were significant differences ($p < 0.01$) between the values obtained from each of the five glucose analyzers tested at each blood glucose concentration tested under normobaric and hyperbaric conditions. There were also significant differences ($p < 0.01$) between glucose analyzer readings under normobaric and hyperbaric conditions for blood samples containing approximately 250, 150, and 25 mg/dL.

CONCLUSIONS

Animal Model - The development of animal models is necessary for use in increasing our knowledge regarding physiology and drug disposition and activity during hyperbaric oxygenation. The results from this study will be used not only in the development of a protocol for the treatment of a virulent beta-streptococcal infection, but also for other studies of the efficacy of hyperbaric oxygenation as an adjunct treatment for infection.

Glucose Analyzer Model - Further preliminary studies were performed upon completion of the study outlined in this report. These studies compared the ability of the glucose analyzers to provide consistent readings of blood glucose in oxygenated and deoxygenated blood. Preliminary results indicate that higher oxygen concentrations in the blood drive the glucose oxidase reaction and result in a higher glucose reading. Further studies were then performed to determine if blood concentrations of 0.1 mM nitroglycerin or 0.1 mM nitroprusside had any effect upon the blood oxygen levels thus resulting in altered readings of blood glucose concentrations by the glucose analyzers. Preliminary data collected in this study did not show a significant effect of the organic nitrates upon blood glucose analysis in oxygenated or deoxygenated blood.

A STUDY OF THE WATER FORCES ANALYSIS
CAPABILITY FOR THE ATB MODEL WITH EMPHASIS
ON IMPROVED MODELING OF ADDED MASS AND WAVE DAMPING

DAVID B. REYNOLDS
Associate Professor
Department of Biomedical and Human Factors Engineering

Wright State University
3050 Colonel Glenn Highway
Dayton, OH 45435

Final Report for:
Summer Research Program
Armstrong Laboratory (AL/CFBV)

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D.C.

September 1992

A STUDY OF THE WATER FORCES ANALYSIS
CAPABILITY FOR THE ATB MODEL WITH EMPHASIS
ON IMPROVED MODELING OF ADDED MASS AND WAVE DAMPING

David B. Reynolds

Associate Professor

Department of Biomedical and Human Factors Engineering
Wright State University

Abstract

The Water Forces Analysis Capability (WAFAC) module for the Articulated Total Body (ATB) model currently does not account for wave damping of a body moving near a free surface. We developed a simplified dimensionless relation between the wave damping coefficient for heave $b_{w,h}$ and the draft of the body, namely: $b_{w,h}/(\rho A_{proj}^{5/4} g^{1/2}) = a_h (1 - \exp[-c_h H])$ where ρ is fluid mass density, A_{proj} is the projected area of the body on a plane normal to its relative velocity, g is the gravitational acceleration, $H \equiv a/d$ is the ratio of body half length to draft, and a_h, c_h are empirically determined constants. From theoretical calculations of heave damping in the literature, we found that $a_h=0.24$ and $c_h=0.53$ provided a good fit for damping averaged over $2.4 \leq \lambda/a \leq 31.4$ for H from 0 to ∞ , where λ is the wavelength. Using this result, we validated the WAFAC/ATB model for experimental forced oscillation of a heaving hemisphere, indicating not only the importance of considering wave damping, but also the accuracy of the above expression. Parametric studies of the motion of a sphere with half the density of water released from the water surface show that (1) the damped natural frequency agrees with theory, (2) wave damping is an important component of the overall damping, and (3) the motion of the sphere in waves demonstrates broad band resonance. Although it appears from others' computations that our dimensionless damping relation is approximately valid for heaving ellipsoids of revolution, this should be verified by future experiments. Also, dimensionless damping coefficients for a surging sphere at the same draft are 1.5 to 9 times larger than for heave, indicating that a single damping expression may not be valid for more general motion than was considered here. This suggests that surging/swaying experiments be conducted to validate the model for other motions.

A STUDY OF THE WATER FORCES ANALYSIS
CAPABILITY FOR THE ATB MODEL WITH EMPHASIS
ON IMPROVED MODELING OF ADDED MASS AND WAVE DAMPING

INTRODUCTION

The Articulated Total Body (ATB) model is a general purpose computer model for predicting the motion of multiple segment linked bodies of general form [4,11]. Recently, in response to the need of the Coast Guard, an additional program having the capability to calculate forces on a body near a free surface was added [3]. This program, termed the Water Forces Analysis Capability or WAFAC, was developed so that the ATB model can be used to analyze the performance of Personal Floatation Devices (PFD).

More generally, three objectives were envisioned by the developers: (1) to develop a method to predict human body response in water, especially with waves, with the body fully or partially submerged and with or without attached PFDs; (2) to develop algorithms and code for applying water forces and movements to body segments within the framework of the ATB model, and (3) to conduct preliminary validations of the WAFAC model through simple simulations. The work described here mainly deals with a continuation of the third objective, although in doing so, we found it necessary to add a simplified method of computing wave damping, which is a force proportional to the relative velocity of a body near a free surface and arises from the generation of waves due to the body's motion. We found that by including this force, the model is validated for the motion of a simple object based on previously obtained experimental data. Once validated, we turned to parametric studies, particularly looking at the effects of frictional drag, wave damping, added mass, and wavelength (or wave frequency) on the motion of a sphere released at a finite distance from its equilibrium position.

BACKGROUND

In their final report, the WAFAC developers, General Engineering and Systems Analysis Company, Inc. (GESAC) recommend several enhancements to

improve the water forces model [3]. Topmost on their list is accounting more accurately for added mass and damping in heave, surge and sway. GESAC had included added mass coefficients, and assumed that they were equal in the three translational directions. Although viscous drag was modeled, the damping due to creation of waves by the motion of a body on or near the free surface was discussed in their report but not included, probably due to its complexity and the perceived need for modifying the governing system of equations in the ATB model. The problem we focused on during the summer research period was how to include wave-damping effects in the model without modifying the system equations and yet be reasonably accurate in predicting the motion of simple objects undergoing oscillations on a calm surface and in waves. We also reviewed added mass, especially for spheres and ellipsoids.

Added Mass and Damping Coefficients

From theoretical as well as experimental studies, the added mass and damping are strong functions of the incident or forced wave frequency (or equivalently, wavelength), but in the linearized theory are not functions of wave amplitude. This is especially true in the low frequency region from 0 up through resonant frequency. It should be noted that added mass coefficients for a body in an infinite fluid are in general different from those for the body at or near a fluid surface. Damping coefficients as defined here only result from motion on or near the surface. For example, the damping during heave motion is a result of energy dissipated due to "wave-making" as the body moves vertically in proximity to the water line. Kim [7] shows that the heave damping coefficient attains a maximum value when the draft is zero. However, since the theories which result in the computation of added mass and damping coefficients of a body floating near the surface do not consider viscosity, it is very possible that these dissipative forces may not be negligible especially as the draft of the body increases. This is why it may be necessary to have an effective damping coefficient which is some combination of wave-making damping and "drag" damping.

As mentioned above, added mass coefficients are different for a body moving in an infinite medium as compared to a floating body moving at or near a free surface. The former are a function of the body geometry and orientation with the direction of motion. The latter in addition depends in a

complicated manner upon wave frequency and draft of the body. There are no analogous damping coefficients for a body immersed in an infinite fluid since these require a free surface and waves.

Several investigators have computed these coefficients for floating bodies with different simple geometries, including spheroids [7], several types of cylinders, and two other ship-like shapes [12]. For our purposes, it is probably most interesting to review the data from spheroids.

Bodies Immersed in an Infinite Medium

Newman [10] presents graphical results of added mass coefficients for ellipsoids of revolution as functions of the major to minor axis ratio (a/b). For the degenerate case of a sphere, all translational added mass coefficients are equal to $1/2$, indicating that the added-mass is half the displaced mass. For a general ellipsoid, added-mass varies in a complex way with a/b and no equations fitting the graphical results are given. Added moments of inertia for spheroids are also given by Newman.

As was stated above, a damping coefficient for a body in an infinite medium analogous to that for a body floating on or near a free surface does not exist because a body in an infinite medium cannot create surface waves. The question arises: How far below a free surface must a body be to have negligible wave-making damping? Kim [7] studied these coefficients for simple geometries as a function of mode of motion, frequency, and draft. His results for heaving spheroids indicate that wave-damping for spheroids submerged deeper than having the water surface tangent to the top have very small heave damping. Interestingly, for heaving spheres and ellipsoids, Kim found that added mass and damping coefficients decreased with increasing draft whereas for surging or swaying spheres and surging ellipsoids, the coefficients increased with draft, at least to the point where the body was just fully submerged. These computations were made over a range of dimensionless frequency ($a\omega^2/g = ka = 2\pi a/\lambda$, where a is the body half length, ω is the angular frequency, g is gravitational acceleration, k is the wave number, and λ is the wavelength) between 0 and 4, corresponding to wavelength-to-body-length ratios of infinite to about 1, which covers the range of practical interest. Possibly more striking than the effect of draft on Kim's results, however, is the effect of frequency. Most of the frequency dependence of the

several added-mass and damping coefficients appears to be in the range of dimensionless frequency from 0 to 4. His results also seem to be an accurate extension of those of previous investigators, as evidenced by good agreement with the latter. However, one should note that in nondimensionalizing the damping coefficient, the latter is divided by a term involving frequency, creating the possibility of an artificial variation with frequency. In subsequent work described in the methodology section, we derive a different dimensionless damping coefficient which is independent of frequency.

Drag on Floating Bodies in Waves

Newman [10] notes two important cases where viscous forces significantly affect the motions of floating bodies in waves. The first case occurs when the body shape is such that inertial forces are small in comparison to viscous forces, as for the streamlined bodies or flow along flat surfaces. The second case concerns cross-flow drag over long slender bodies which have a velocity relative to the local wave velocity.

Examples of the first case, i.e., predominant viscous shear forces, include angular motion of a body of revolution about its axis, such as yaw moment of an axisymmetric buoy, roll motion of bodies with nearly circular cross-section, or heave motion of a slender buoy with large draft. In these cases, viscous damping is particularly important at the resonant frequency, which occurs where inertial and hydrostatic restoring forces cancel each other. The second case occurs for slender bodies with lengths about equal to the wavelength (length is the dimension of the body in the surge direction) and transverse dimensions comparable to the wave amplitude. A rigid body will then experience motion relative to the waves and cross-flow drag, similar to that for bodies in an infinite medium, can be estimated if the drag coefficient, relative velocity, and projected area of the body are known. Specific examples of floating bodies in waves with significant cross-flow drag include the horizontal force on a slender spar buoy with large draft or the vertical force on the hull of a thin ship whose length is on the same order as the wavelength and whose "thinness" is comparable to the wave amplitude. In the latter situation, the craft will not follow the wave motion, in contrast to an object much smaller than the wavelength and amplitude, such as a cork bobbing in waves.

Added-Mass and Damping Coefficients for Floating Bodies

In Newman's words, "if a body moves in an infinite ideal fluid, hydrodynamic pressure forces and moments will result which can be expressed most simply in terms of the added-mass coefficients m_{ij} . A more complicated and less general situation follows for a body moving on or near a free surface" [10]. The analogous added-mass coefficient in the latter case results in complex coefficients, with both the real and imaginary parts depending on frequency. The real part is known as the added-mass coefficient a_{ij} because it represents the force component proportional to the acceleration. Newman points out that the latter coefficient will differ from that for a body in the absence of a free surface m_{ij} and one must refrain from assuming that all properties of m_{ij} apply to a_{ij} . The imaginary part gives rise to a force which is proportional to the body velocity, and the corresponding coefficient b_{ij} is known as the damping coefficient. This force results from the generation of waves on the free surface due to the body motion.

Wehausen [13] has reviewed the various aspects of predicting the motion of floating bodies, including theoretical determination of added-mass and damping coefficients and comparison with experimental data. As mentioned above, most interesting from our standpoint is the work of Kim [7], who computed a_{ij} and b_{ij} for spheroids as a function of body motion, frequency, and draft. His results show that a_{ij} and b_{ij} are strong functions of frequency at least up to a dimensionless frequency of 4.

Unfortunately, the added-mass and damping results of Kim and others are numerical computations presented in graphical form. No empirical fitting of these results has been done as far as we are aware. What we describe in the next section is the approach we have taken to simplify the inclusion of wave damping in the WAFAC module of the ATB model.

METHODOLOGY

The framework for including wave-damping in the WAFAC module without altering the governing system of equations in ATB is to compute a total drag

force which is assumed to be the sum of that due to viscous drag (as is currently done in WAFAC) and wave-making drag, i.e.,

$$\begin{aligned} F_{TOT} &= F_v + F_w \\ &= \frac{1}{2} C_D A_{proj} \rho |V_{rel}| V_{rel} + b_w V_{rel} \end{aligned} \quad (1)$$

where C_D is the drag coefficient, A_{proj} is the projected area of the body on a plane normal to the direction of the relative velocity, V_{rel} , ρ is the fluid mass density, and b_w is the wave-damping coefficient. What we will describe in this section is our method of reducing the data for b_w , which in general is a function of body shape and size, oscillation frequency, direction of motion, and draft.

Data Reduction for Frequency

Many investigators report damping coefficients made dimensionless by dividing by $\rho a^3 \omega$, or in general, mass density times a volume scale times angular frequency. Some authors use displaced volume instead of a^3 [10]. Dimensionless damping coefficients are plotted as a function of the dimensionless frequency ka discussed above. To eliminate the artificial variation of dimensionless damping with frequency, we compute another dimensionless damping coefficient: $b_w / \rho a^{5/2} g^{1/2}$. The latter dimensionless damping coefficient can easily be shown to be the product of $b_w / \rho a^3 \omega$ and the square root of dimensionless frequency, $\sqrt{ka} = \omega \sqrt{a/g}$.

In this way, damping coefficient computations can be plotted in terms of a dimensionless coefficient which does not include the variable of frequency. Although the revised damping coefficient does show a dependence on frequency, as it must since damping has been theoretically shown to go to zero at $\omega=0$ and $\omega=\infty$ [10], we then average the dimensionless damping coefficient for a given set of data over the range $0.2 \leq ka \leq 2.6$. This is equivalent to wavelengths greater than 2.4 times the radius but less than 31.4 times the radius or half length of the body. This range of frequency was chosen because it lies within the computations of several investigators and covers a practical range of interest.

Data Reduction for Body Shape and Size

For a given body shape, dimensional analysis scales the results for size. It may not be so helpful in reducing data for different shapes. Because we are mainly interested in damping for spheres and ellipsoids which have much geometric similarity, we propose that body length scaling be proportional to A_{proj} as defined in the drag equation. One of the convenient aspects about using A_{proj} is that it is a variable currently computed by WAFAC. The dimensionless damping coefficient now becomes: $b_w / \rho A_{proj}^{5/4} g^{1/2}$.

Data Reduction for Draft

Most of the damping coefficients computed in this study were taken from the results of Kim [7], who presented dimensionless damping for a particular geometry (mostly spheroids, i.e., ellipsoids of revolution) and motion as a function of dimensionless frequency for up to five values of dimensionless draft, H . The latter was defined as the ratio of the body half length to submerged depth. Thus, for a sphere half-submerged, $H = 1$, for one just totally submerged, $H = 1/2$, while for one floating on the very top of the free surface, $H = \infty$. The effect of draft was accounted for by empirically fitting average dimensionless damping coefficients as a function of H . It should be noted that WAFAC also computes the variables required to calculate H , namely BET and BTE in subroutine DRAGCHK. BET is the distance from the centroid of an ellipsoid to the free surface, positive if the centroid is out of the water and negative if below the water line. BTE is the distance from the centroid to the lowest point on the ellipsoid which is also on a plane parallel to the water surface tangent plane. Thus, $H = BTE / (BTE - BET)$.

Model Validation and Computer Simulations

The paucity of experimental data [13] for sphere motions on a free surface makes validation of the model difficult. The only experimental measurements we have found for added mass and damping coefficients are of a heaving sphere, which do agree with theory [2]. In these experiments, the sphere was forced to oscillate with constant small amplitude in calm water. Magnitude and phase of the oscillatory force and surface pressure distribution

were measured. To simulate this with WAFAC, we specified the forcing function at a given frequency to match experimental data and compared the predicted displacement amplitude with that used experimentally. This was repeated at each frequency of interest to generate a dimensionless driving force-frequency curve to compare with the experimental data and theory.

The WAFAC module acting in conjunction with ATB was run for a 5 inch radius sphere having half the density of water, thus half-submerged at equilibrium. Runs were made in calm water and in waves. In all cases reported here, the sphere was initially dropped from a point above the water surface for which the sphere bottom was just touching the surface. We systematically studied the effect of damping and added mass by varying the coefficients between 0 and a nominal value found in this study or previously used. We rewrote the equation computing drag force to include wave damping as given in Eq. 1. Wave heights (= twice the wave amplitude) of 6 and 2 in were studied. For the 2 in wave, we studied wavelengths of 21, 31.5, 45, 60, 75, 90 and 100 in. Only the latter was studied for the 6-inch wave. Output variables included x, y, z components and resultant of acceleration, velocity, displacement, wave excitation force and drag force. Other outputs are available but were not plotted for all cases.

RESULTS and DISCUSSION

Added Mass Coefficients - Spheres

Although the added mass coefficient for a sphere moving in an infinite medium is $\frac{1}{2}$, i.e., the added mass is $\frac{1}{2}$ the mass of the displaced fluid, the latter value is also a reasonable frequency-averaged value for the added mass coefficient of surging/swaying spheroids as well. The latter statement is not obvious from a cursory look at Kim's computations of added mass coefficient as a function of ka for several values of H [7], but if his coefficient is recast in terms of the mass of *displaced* fluid instead of total fluid mass then the curves for the several values of H are nearly the same. On the other hand, added mass coefficients for heaving spheroids demonstrate very different behavior, with coefficients increasing with decreasing draft, i.e., the largest added mass occurs when the sphere is floating on top of the fluid [7]. Nondimensionalizing by displaced fluid mass rather than total fluid mass only

separates the curves further. However, for a heaving sphere half submerged, the added mass approaches $\frac{1}{2}$ the displaced mass as $\omega \rightarrow \infty$, and $\frac{1}{2}$ also appears to be a reasonable average value for $0 \leq ka \leq 3$ [10]. Based on frequency-averages of both surge/sway added mass at all levels of draft and heave added mass also averaged over $0 \leq H \leq 2$ (i.e., from very deeply submerged to 84% of the sphere volume out of the water), a "grand" average value of $\frac{1}{2}$ for the added mass coefficient appears reasonable.

Added Mass Coefficients - Ellipsoids

Kim [7] also gives results for a heaving and surging (motion in the longitudinal direction) ellipsoid of revolution with major to minor axis ratio (a/b) of 4. Added mass coefficients for this particular surging ellipsoid are 4 to 5 times smaller than for a surging sphere with the same proportion of displaced volume. This is similar to the data for added-mass coefficients of ellipsoids in an infinite medium [10]. For $a/b \geq 0.8$, added mass coefficients in the longitudinal and lateral directions in an infinite medium are approximately given by

$$m'_{11} \equiv m_{11} / (\frac{4}{3} \pi \rho a b^2) \approx 0.5(b/a) \quad (2a)$$

and

$$m'_{22} \equiv m_{22} / (\frac{4}{3} \pi \rho a b^2) \approx 1 - 0.5(b/a), \quad (2b)$$

respectively. These equations are exact for a sphere. Added mass for a heaving ellipsoid is larger than that for a surging one at the same draft [7]. These data are similar to differences noted above for surging and heaving added mass for spheres. Thus, it appears reasonable to accept Eq 2 a,b as approximating added mass coefficients for ellipsoids moving near a free surface. For a general ellipsoid with three planes of symmetry, graphical data may be found in Kochin, Kibel, and Roze [8]. If the two minor axes are not too much different, Eq 2 a,b may be used to approximate added mass coefficients, for which b would be taken as the mean of the two minor axes.

Damping Coefficients - Sphere

Wave damping for a heaving sphere was initially studied by Havelock [5]. Later, Barakat [1] also studied a heaving sphere and found somewhat different

results for damping than Havelock at low frequency. Later, Wehausen [13] extensively reviewed the literature and noted that Barakat's work "is unfortunately marred by errors in the numerical calculation." Other investigators have also noted Barakat's error [6]. Kim's [7] results agree with and extend Havelock's in terms of studying the effect of draft on heave damping.

For a heaving sphere the frequency-averaged damping coefficients are given in Table 1. The results are computed mainly from Kim's graphs [7], but half-submerged sphere ($H=1$) data was based upon Hulme's [6] results. Averaging was done over $0.2 \leq ka \leq 2.6$, and coefficients of variation, CV% (standard deviation as percentage of the mean) are also given.

Dimensionless Draft, $H = a/d$	Heave Damping Coefficient $b_{w,h}/(\rho A_{proj}^{5/4} g^{1/2}) \pm CV\%$
1/2	0.040 \pm 61.9%
1	0.103 \pm 24.9%
2	0.162 \pm 20.6%
4	0.203 \pm 23.5%
∞	0.238 \pm 27.2%

Table 1. Frequency-Averaged Heave Damping Based on Calculations by Kim [7] and Hulme [6].

The heave damping decreases with increasing submerged depth (note, H is proportional to the reciprocal of the draft). The largest damping occurs when the sphere is floating on the top of the fluid, $H=\infty$. The variability over frequency is largest for the smallest damping ($H=1/2$), and the CV of the larger values averages about 24%. This indicates that averaging over this frequency range gives reasonably accurate values for heave damping of a sphere.

The data in Table 1 can be very well fit by an exponential model, yielding

$$b_{w,h}/(\rho A_{proj}^{5/4} g^{1/2}) = a_h (1 - \exp[-c_h H]) \quad (3)$$

with $a_h = 0.24$ and $c_h = 0.53$. This nonlinear curve fit was done on SigmaPlot® (Jandel Scientific, Corte Madera, CA), and yielded CV% for a_h and c_h of 4.1% and 11.0%, respectively, indicating a good fit using this model.

Wave damping of surging/swaying spheres (i.e., $b_{w,s}$, the coefficient with dimensions) demonstrates similar frequency variation as for heave but quite different variation with draft [7]. Largest $b_{w,s}$ are for a sphere just

totally submerged beneath the surface ($H=1/2$) with progressively smaller values as the sphere floats higher in the fluid. Variation with H is reduced drastically with our choice of dimensionless $b_{w,s}$, i.e., $b_{w,s}/(\rho A_{proj}^{5/4} g^{1/2})$. This is due to the factor A_{proj} , which decreases from πa^2 for $H=1/2$ to $0.227 a^2$ for $H=4$, i.e., a sphere having a draft of $1/4$ its radius. These frequency-averaged surge/sway damping coefficients are given in Table 2.

Dimensionless Draft, $H = a/d$	Surge/Sway Damping Coefficient $b_{w,s}/(\rho A_{proj}^{5/4} g^{1/2}) \pm CV\%$
1/2	0.369 \pm 45.8%
1	0.416 \pm 51.8%
2	0.410 \pm 61.6%
4	0.347 \pm 65.5%

Table 2. Frequency-Averaged Surge/Sway Damping Based on Calculations by Kim [7].

One immediately notices that: (1) these coefficients are larger than computed for heave at the same draft; (2) they don't have the dependence on draft that is seen in heave; and (3) they are more variable over the range of dimensionless frequency studied. As for any wave damping coefficient, surge/sway damping must go to zero as draft gets large, i.e., as $H \rightarrow 0$.

Damping Coefficients - Ellipsoids

Kim's ellipsoid damping coefficients for heave and surge are quite a bit smaller than those for a sphere. This appears to be mainly a result of the nondimensionalization he chose to use. If the characteristic length in his dimensionless damping, frequency, and draft variables is chosen as the radius of a circle with an area equal to ellipse area, ellipsoid and sphere data for heave damping virtually coincide. Kim only computes ellipsoid surge damping for $H=4$ and 8 , and although damping normalized as suggested above compared with that for a surging sphere demonstrates less agreement than found for heave, the mean values for the two cases of H compare favorably with that for a sphere. This suggests that our dimensionless damping results for spheres may be applied with reasonable confidence to ellipsoids.

A Note on the Range of Validity of Parameters

The WAFAC model was developed assuming a linearized free-surface condition [3]. What this means in terms of allowable wave amplitude A is that A/λ should be ≤ 0.01 for the kinetic energy to be $\leq 3\%$ of the potential energy on the free surface of a wave [9]. Now, our frequency (or wavelength) averaged damping coefficient is valid over the range $2.4 \leq \lambda/a \leq 31.4$. Thus, A/a (wave amplitude/body half-length) should be ≤ 0.314 . Thus, results of simulation for wave heights much larger than $1/3$ the body length must be viewed with some suspicion.

Relative Importance of Wave Damping and Drag Forces

If the wave damping and drag forces in Eq. 1 are equated and Eq. 3 substituted for wave damping, the relative velocity at which the two forces are equal is given by

$$|V_{rel}| = 0.48 (1 - \exp[-0.53H]) A_{proj}^{1/2} g^{1/2} / C_D \quad (4)$$

For example, for a 5 in. radius sphere half-submerged (i.e., $H=1$), $C_D = 1/2$ (approximately valid for a sphere for $700 \leq Re \leq 3 \times 10^5$, where $Re = 2 V_{rel} a/\nu$ is the Reynolds number, ν the kinematic viscosity), $|V_{rel}|$ is 23.1 in/sec for heave. Taking the $H=1$ value for surge damping from Table 2 results in $|V_{rel}| = 40.9$ in/sec. Because these velocities are in the range expected for model validity, these sample calculations indicate that wave damping forces are significant with respect to frictional drag forces for objects moving in proximity to a free surface.

Model Validation

The only experimental data the authors are aware of for motions of a sphere on a free surface area by Cumming [2]. He forced a hemisphere to oscillate with constant small ($\%$ in) amplitude in calm water. He measured force and pressure distribution at several oscillatory frequencies. From these data he computed wave damping and added mass and compared the calculations with theory, obtaining good agreement. We simulated Cumming's

experiment by prescribing a sinusoidal forcing function and obtaining the displacement amplitude as output of the model. For the first series, we retained both $C_D = 1$ and wave damping. The forcing functions and output displacements are given in Table 3. Displacement amplitude was determined once steady-state conditions were reached. The amplitudes should be compared to $\frac{1}{8}$ in.

With wave damping included, we obtained good agreement with Cumming's experiment. At the higher frequencies, we underpredicted the amplitude slightly. This may be due to the increasing influence of the $C_D = 1$ term, as the velocity amplitude was about 3 in/sec with the higher frequencies studied. Steady state was reached within 5 sec for each forcing frequency studied.

Forcing Function (lb)	Displacement Amp., in Wave Damping, $C_D=1$	Displacement Amp., in No Wave Damping, $C_D=1$
0.774 sin 3.59t	0.24	*
0.490 sin 5.67t	0.24	*
0.328 sin 6.95t	0.28	*
0.160 sin 8.22t	0.24	0.63
0.449 sin 9.825t	0.21	*
1.030 sin 11.76t	0.22	*
1.378 sin 12.68t	0.22	*

*In the cases without wave damping, these displacement results were erratic and did not appear to reach steady state, unlike the cases with wave damping. Therefore, we did not calculate an amplitude.

Table 3. Model Validation Results for Forced Oscillation of a Half-Submerged 6 in Radius Sphere.

In the second series, we eliminated the wave damping term and only included frictional drag. The displacement amplitude was expected to be considerably larger, and this was observed. Except for the resonant case, the displacement was highly variable with cycle and did not appear to reach a steady state. These results indicate that wave damping is an important term to include in the model, for without it, we could not validate the model with Cumming's experimental data.

Computer Simulations

Initially, we ran the WAFAC/ATB without considering wave damping, but studied the effect of frictional drag and added mass on the motion of a 5 in radius sphere in calm water and with 6 in amplitude, 100 in wavelength waves. Later, we added wave damping as given by Eq. 3 and made the same runs as well as runs with a reduced wave amplitude of 2 in. With the latter amplitude, we investigated the effect of wavelength on the motion, studying additional wavelengths of 21, 31.5, 45, 60, 75, and 90 in.

Calm Water. In all the calm water studies reported here, the sphere was released from a point for which the bottom was just touching the water surface. Data are probably best reduced by computing damping coefficient and damped natural frequency. Table 4 shows this data for the several combinations of drag coefficient, added mass coefficient, and wave damping. The data were averaged over several cycles after initial transients had died out.

Drag C_D Coeff, C_D	Add. Mass Coeff.	Wave Damping	Damped Nat. Freq., f_d (Hz)	Damping Coeff., ζ
0	0	No	1.481 \pm 0.72%	0
1	0	No	1.709 \pm 0.85%	0.0353 \pm 103%
1	0.5	No	1.393 \pm 0.68%	0.0366 \pm 86.8%
*0	0.5	No	1.107 \pm 0.50%	0
0	0	Yes	1.700 \pm 1.03%	0.0812 \pm 1.62%
1	0	Yes	1.702 \pm 0.45%	0.0969 \pm 15.3%
1	0.5	Yes	1.393 \pm 0.53%	0.0791 \pm 17.5%
0	0.5	Yes	1.385 \pm 1.35%	0.0664 \pm 4.1%

*In this run, the sphere came out of the water for every cycle and acceleration was also particularly asymmetric.

Table 4. Data Reduction Results for Parametric Studies of a Sphere Released above the Equilibrium Position in Calm Water.

The decrease in natural frequency with added mass by a factor of $\sqrt{1.5}$ is expected from "strip" theory [9], which was observed. With wave damping present, higher damping was found for the cases with no added mass and unity drag coefficient. The decrease in damping with added mass by a factor of $\sqrt{1.5}$ also agrees with theory. For the cases with $C_D = 1$, we noted that damping coefficient decreased with cycle, and approached the damping with $C_D = 0$ for

the final cycles (runs were carried out for 10 sec). This is due to the fact that the drag force term, proportional to velocity squared, becomes relatively insignificant. With wave damping and constant added mass, no difference in the natural frequency with variation of C_D was observed. This is also consistent with the strip theory, in which the oscillating sphere behaves like a mechanical oscillator and the natural frequency is proportional to the square root of restoring to inertial forces. However, with wave damping, the natural frequencies reported here are about 40% higher than estimated from that theory. This is probably because in the strip theory, the oscillations are assumed small and the body assumed "wall-sided", i.e. no curvature at the water line. Neither of these assumptions is valid for the computer runs in calm water. The natural frequency does agree with calculations based upon "exact" theory, as given by Cumming [2].

Waves. Because of the restrictions on wave amplitude and frequency discussed above, we only present results for the 2 in amplitude wave. Displacement in the z direction (heave) is shown for the several wavelengths in Fig. 1. The most striking feature of these results is that for wavelengths of 21, 31.5 and 45 in (corresponding to frequencies of 1.71, 1.40 and 1.17 Hz) the oscillation is independent of wavelength and has a period of about 830 ms ($f=1.20$ Hz). At the higher wavelengths of 60, 75 and 100 in (corresponding to frequencies of 1.01, 0.905 and 0.784 Hz) there appears to be an initial higher frequency oscillation transient followed by an oscillation nearly matching the frequency of the incident wave. In the higher wavelength cases, the resonance of the sphere's natural oscillation appears to die out and is followed by motion at the wave frequency. At the lower wavelengths, the wave frequencies surround the damped natural frequency of the sphere determined previously (~ 1.39 Hz) and a resonance phenomenon appears to be occurring. This broad band resonance is typical of "blunt" objects such as a sphere, whereas more pencil-like objects, such as a vertical spar buoy, experience a larger, more highly tuned resonant response [10].

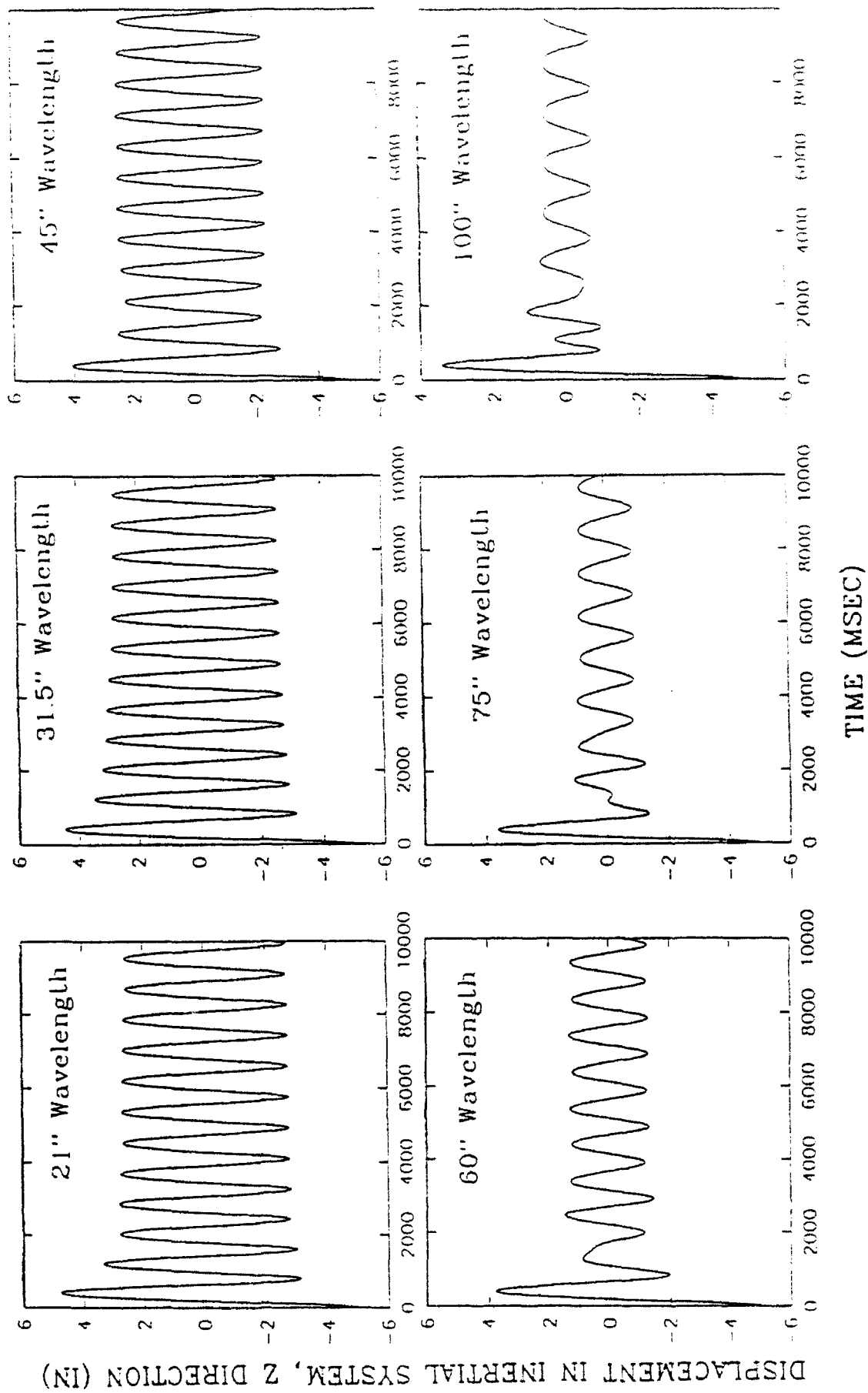


Figure 1. Heave displacement for a 5 in radius sphere having half the density of water, released on water surface, 2 in amplitude waves, $C_D=1$, $C_{AM}=0.5$, with wave damping.

CONCLUSIONS AND RECOMMENDATIONS

Wave damping for a sphere moving near a free surface is an important force in determining its motion via use of the WAFAC/ATB model. For heave and surge, the frequency dependence of the damping coefficient can be minimized by averaging over a frequency range of interest. The resulting simplified equation inserted in the model has been shown to accurately predict the motion of a hemisphere undergoing forced heaving oscillations. These results are not predicted when only frictional drag is considered. Parametric studies of the effect of wavelength on the heave motion of a sphere show a broad band resonance phenomenon, which is predicted for a sphere, lending additional credence to the use of the wave damping expression we have developed.

We recommend that experiments be conducted on ellipsoids in heave, surge and sway to validate the model for this geometry and more general motion. Once the model has been validated for single ellipsoidal objects in heave, surge and sway, more complex objects, such as multi-segment human models with attached PFD's, can be studied.

ACKNOWLEDGEMENTS

The author would like to thank Drs. Ints Kaleps and Buford Shipley of the Armstrong Lab, Vulnerability Assessment Branch for their helpful discussion of this work. Also many thanks to Dr. Shipley for his technical assistance, without which we could not have generated so much useful data.

REFERENCES

1. Barakat, R. Vertical motion of a floating sphere in a sine-wave sea, J. Fluid Mech. 13:540-56, 1962.
2. Cumming, R.A. The experimental determination of forces and pressures acting on a hemisphere oscillating on a free surface, Univ. Calif. Inst. Engr. Res., Contract No. Nonr-222(30), 1963.

3. Development of Water Forces Analysis Capability (WAFAC) for the Articulated Total Body Model, Final Report, General Engineering and Systems Analysis Co., Inc., Kearneysville, WV.
4. Fleck, J.T. and F.E. Butler, Validation of the Crash Victim Simulator, Final Report, U.S. Dept. Transport. Natl. Highway Traffic Safety Admin., 1981.
5. Havelock, T.H. Waves due to a floating sphere making periodic heaving oscillations, Proc. Roy. Soc. A, 231:1-7, 1955.
6. Hulme, A. The wave forces acting on a floating hemisphere undergoing forced periodic oscillations, J. Fluid Mech. 121:443-63, 1982.
7. Kim, W.D. On the harmonic oscillations of a rigid body on a free surface, J. Fluid Mech. 21:427-51, 1965.
8. Kochin, N.E., I.A. Kibel, and N.V. Roze, Theoretical Hydrodynamics, English translation of 5th Russian ed., Wiley, New York, 1964.
9. McCormick, M.E. Ocean Engineering Wave Mechanics, Wiley, New York, 1973.
10. Newman, J.N. Marine Hydrodynamics, MIT Press, Cambridge, MA, 1977.
11. Obergefell, L.A., T.R. Gardner, I. Kaleps, J.T. Fleck, Articulated Total Body Model Enhancements, Final Report, AAMRL-TR-88-043, 1988.
12. Vugts, J.H. The hydrodynamic coefficients for swaying, heaving and rolling cylinders in a free surface, Int. Shipbuilding Progr., 15:251-76, 1968.
13. Wehausen, J.V. The motion of floating bodies, Ann. Rev. Fluid Mech. 3:237-68, 1971.

A STUDY OF THE EFFECTS OF MICROWAVE RADIATION
AND TEMPERATURE ON AMINO ACID METABOLISM BY
MOUSE MACROPHAGE CELLS

Donald K. Robinson
Associate Professor
Department of Chemistry

Xavier University of LA
New Orleans LA 70125

Final Report for:
Summer Research Program
Armstrong Laboratory/OEDR

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

August 1992

A STUDY OF THE EFFECTS OF MICROWAVE RADIATION AND
TEMPERATURE ON AMINO ACID METABOLISM BY
MOUSE MACROPHAGE CELLS

Donald K. Robinson
Associate Professor
Department of Chemistry
Xavier University

Abstract

In this study we sought to determine if the changes in amino acid metabolism by mouse macrophage cells exposed to microwave radiation (2450 MHz, 30 minutes) are caused by "hot spots" or heat effects. For this study we exposed the cells to 2450 MHz, 103 w/kg, for 30 minutes. Amino acid profiles of the cell culture media were determined at 24, 48, and 72 hours. Treatments consisted of microwave radiation at 37° and sham treatments at 37, 38, and 39°C. In addition, all of these studies were performed using two cell densities; one cell density was approximately 10 fold less dense than the other. The changes in amino acid profiles were determined by thin-layer chromatography. Changes in amino acid profiles were not observed to be more pronounced at 48 hours and 72 hours post-treatment. The changes in amino acid metabolism by mouse macrophage cells were more apparent in the dense cell cultures. The amino acid bands of the sham treated cells at 38 and 39°C were less intense than the samples microwaved and sham treated at 37°C. This indicates that the previously found changes in amino acid metabolism by mouse macrophage cells are due to the bioeffects of microwave radiation rather than to temperature effects. If the changes were due to temperature effects, the band intensities of the microwaved samples would have been considerably diminished.

A STUDY OF THE EFFECTS OF MICROWAVE RADIATION
AND TEMPERATURE ON AMINO ACID METABOLISM BY
MOUSE MACROPHAGE CELLS

Donald K. Robinson

INTRODUCTION

The use of radar and radio transmitters in the civilian and in the military sectors of our society continues to increase. Such microwave and radiofrequency radiation has increased the interest in its possible health effects. This led to a wide range of studies on the bioeffects of microwave radiation although many of the studies have been negative (1,2,3). Even though microwave exposure has been reported to cause changes in chromosome number and structure (4,5), to cause cataracts in humans (6), to promote malignant tumor formation in rats (7), and to increase tumor production and leukemias in humans (8), their effects are less certain if temperature and other variables are carefully controlled. Failure to adequately control temperature can cause thermal effects to be mistaken for bioeffects of microwave radiation. The aim of this study was to determine if small changes in temperature can cause changes in amino acid metabolism of mouse macrophage cells. During the summer of 1990 we found significant changes in amino acids in cells exposed to microwave radiation compared to temperature regulated sham controls. Most of the changes occurred at 48 hours post exposure, although significant changes in several amino acids were seen at 24 hours (9). Even though the studies were controlled for temperature, it is still possible that the microwave radiation might have increased the temperature one or two degrees and this might have been enough to produce the changes in amino acid metabolism. In addition we sought to determine if microwave radiation and temperature effects are more apparent in less dense cell cultures than in cell cultures with a higher density. High density samples may be more susceptible to the added stress of microwave radiation or temperature increases. For this study we expanded our time studies to include evaluations at 24, 48, and 72 hours. These time studies will be of value in determining if microwave radiation results in increased uptake of amino acids by the cells or if it results in decreased release of amino acids into the cell culture media. This kind of information is very important because it will help define the mechanism by which microwave radiation exerts its effect. For our studies we used thin-layer chromatography (TLC) to ascertain amino acid profiles of the test samples. The TLC experiments are easy to perform, inexpensive, fast, and changes in amino acid profiles can be easily seen and interpreted. Aliquots of the samples were saved for quantitative amino acid analyses by a high performance amino acid analyzer.

OBJECTIVES

1. Determine if the previously found changes in amino acid metabolism of mouse macrophage cells at 37°C are due to microwave radiation or to temperature (heat effects). For the study, amino acid profiles of cells exposed to microwave radiation at 37°C, 30 minutes, will be compared to those non-microwaved or sham treated at 37, 38, and 39°C for 30 minutes.
2. Determine the changes in amino acid profiles of the cell culture supernatants at 24, 48, and 72 hours post-exposure. This determination will be made by comparing the amino acid profiles by thin-layer chromatographic analyses of cell culture supernates.
3. Determine if microwave and temperature effects are more apparent in less dense cell cultures than in more dense cell cultures. This determination will be made by using two separate cell cultures that are approximately 10 fold different in cell densities.

METHODOLOGY

The study protocol utilized RAW 264.7 mouse macrophage cells at cell densities ranging from 5.6 to 7.3×10^5 cells per ml in RPMI mammalian culture media (Table 1). Four milliliter aliquots of the cells were transferred to sterile culture tubes and inserted into the temperature controlled microwave chamber for microwave exposure or sham treatments. Microwave exposure was at 2450 MHz, continuous wave, with a mean specific absorption rate of 103.5 ± 4.2 W/Kg for 30 minutes. After microwave exposure or sham treatment, each culture tube was vortex-mixed for 5 seconds. A 0.5 ml aliquot of cell suspension was removed for counting with a Coulter cell counter. The remaining 3.5 ml of cell suspensions were allowed to grow in an incubator at 37°C. At 24, 48, and 72 hours, 1ml aliquots were removed. One-half of each aliquot was used for counting and the other 0.5 ml was centrifuged to remove the cells. The supernatant phase was removed for amino acid analysis. All of the supernates were stored in the freezer until analyzed. The same experiments were repeated but with a cell density of 4.4 to 5.2×10^4 cells per ml.

Thin-Layer Chromatography (TLC)

Seven microliters of each supernate were applied to a pre-coated, thin-layer chromatographic plate (Sigma cell Type 100 cellulose on polyester, 20 x 20 cm., Sigma T-6890). The plates were sprayed with 0.2% ninhydrin in ethanol to visualize changes in amino acid profiles.

Cell Density

Cell densities were determined using a Model ZBL Coulter cell counter, Hialeah Florida.

RESULTS

Effects of microwave radiation and temperature (37-39°C) on amino acid metabolism at 24 hours post-exposure.

The chromatograms of the microwaved samples and the sham treated samples, 37 - 39°C, displayed very similar banding patterns. There were three distinct amino acid banding regions. The band intensities of the 38 and 39°C sham treated samples were significantly less intense than the band intensities of the 37°C microwaved and sham treated samples.

Effects of microwave radiation and temperature (37-39°C) on amino acid metabolism at 48 hours post-exposure.

The chromatograms of the microwaved and sham treated samples at 37°C displayed very similar banding patterns. The band intensities of the 38 and 39°C sham treated samples were significantly less intense than the band intensities of the 37°C microwaved and sham treated samples.

Effects of microwave radiation and temperature (37-39°C) on amino acid metabolism at 72 hours post-exposure.

The chromatograms of the microwaved and the sham treated samples at 37°C again displayed very similar banding patterns at 72 hours post-exposure. The band intensities of the 38 and 39°C sham treated samples were significantly less intense than the band intensities of the 37°C microwaved and sham treated samples.

The effects of microwave radiation and temperature (37-39°C) on amino acid metabolism by cells that differ 10-fold in densities.

With the less dense cell culture, the differences in amino acid band intensities of the sham treated samples at 38 and 39°C were less pronounced than those observed for the microwaved and sham treated samples at 37°C.

DISCUSSION

The chromatograms obtained from the cell culture supernates subjected to sham treatment at 38 and 39°C 24 hours post exposure, differed significantly from the chromatograms obtained from samples that were exposed to microwave radiation or sham treatment at 37°C. The band intensities of the sham treated samples decreased with increase in temperature from 37 to 39°C. This suggests that the differences in amino acid concentrations between microwaved and sham treated samples previously found after high performance amino acid analysis were not due to "hot spots", or temperature effects. If temperature

effects were responsible for these differences in amino acid concentrations, the amino acid band intensities of the microwaved samples at 37 degrees Celsius would parallel or be less intense than the band intensities of the 38 and 39°C sham treated samples. Thin-layer chromatography (TLC) analysis of the cell supernates at 48 hours and 72 hours post treatment gave results that paralleled the banding patterns found at 24 hours post treatment. Namely, the band intensities of the samples sham treated at 38 and 39°C were significantly less intense than the samples microwaved and sham treated at 37°C. It appears as though the changes in amino acid metabolism are not more pronounced 48 hours and 72 hours post-treatment by microwave radiation or increases in temperature. However, high performance amino acid analysis of all sample supernates is required to quantitate the amino acid concentrations and thus validate our qualitative analyses. The changes in amino acid metabolism by mouse macrophage cells were not amplified by using less dense cell suspension samples. This was probably due to the net decrease in metabolism of the culture media by each cell suspension sample.

TABLE I RAW 264.7 Mouse Macrophage Cells

Number of Cells/ml

<u>SAMPLE</u>	<u>Treatment(°C)</u> <u>Sham(s)/Microwaved(m)</u>	<u>BEFORE</u> <u>INCUBATION</u> (x10 ⁵)	<u>AFTER</u> <u>24HRS</u> (x10 ⁵)	<u>AFTER</u> <u>48HRS</u> (x10 ⁵)	<u>AFTER</u> <u>72HRS</u> (x10 ⁵)
1.	39(S)	5.6	4.0	3.5	7.3
2.	38(S)	5.7	4.1	3.9	7.9
3.	37(S)	5.7	4.3	4.2	7.5
4.	37(M)	6.3	4.6	8.3	8.5
5.	37(M)	7.3	4.6	7.0	9.4
6.	37(S)	6.3	4.1	4.8	7.7
7.	38(S)	6.2	4.5	8.6	8.8
8.	39(S)	6.3	4.5	5.6	8.2
9.	39(S)	5.8	4.3	5.9	9.8
10.	38(S)	5.8	4.0	5.0	8.8
11.	37(S)	6.4	5.2	5.2	8.7
12.	37(M)	5.9	4.2	5.7	9.8
<u>Diluted Samples</u>		<u>(x10⁴)</u>	<u>(x10⁴)</u>	<u>(x10⁴)</u>	<u>(x10⁴)</u>
1.	39(S)	5.2	3.9	4.2	6.7
2.	38(S)	4.3	3.3	3.6	4.8
3.	37(S)	4.6	3.4	3.4	4.7
4.	37(M)	4.9	3.5	4.5	5.4
5.	37(M)	5.1	3.4	4.7	6.5
6.	37(S)	4.5	3.5	3.8	5.0

TABLE I RAW 264.7 Mouse Macrophage Cells (Cont'd)

<u>SAMPLE</u>	Treatment(°C) <u>Sham(s)/Microwaved(m)</u>	Number of Cells/ml			
		<u>BEFORE</u> <u>INCUBATION</u> (x10 ⁵)	<u>AFTER</u> <u>24HRS</u> (x10 ⁵)	<u>AFTER</u> <u>48HRS</u> (x10 ⁵)	<u>AFTER</u> <u>72HRS</u> (x10 ⁵)
7.	38(S)	4.7	4.2	4.4	6.0
8.	39(S)	4.3	3.0	3.6	5.6
9.	39(S)	4.6	3.4	3.7	5.1
10.	38(S)	4.4	3.7	3.6	5.2
11.	37(S)	5.0	3.6	4.4	6.3
12.	37(M)	4.4	2.9	4.1	5.5

REFERENCES

1. Kiel, J.L., Wong, L.S., and Erwin, D.N., Physiological Chemistry and Physics and Medical NMR 18: 1986, pp 181-187.
2. Rama Rao, G., Cain, C.A., and Tompkins, W.A., Effects of microwave exposure on the hamster immune system. III. Macrophage resistance to Vesicular stomatitis Virus infections., Bioelectromagnetics 5:
3. Parker, J.E., Kiel, J.L., and Winters, W.D., Physiological Chemistry and Physics and Medical NMR, 20: 1988, pp 129-134.
4. Yao, K., J. of Heredity 73: 1982, pp 133-138.
5. Manikowska, E., Luciani, J.M., Servanti, B., Czerski, P., Obrenovitch, J. and Stahl, A., Experiments 35: 1979, pp 388-390.
6. Hirsch, F.C., and Parker, J.T., Arch. Indust. Hyg. 6: 1952, pp 512-517.
7. Kuntz, L.L., Johnson, B.S., Thompson, D., Crowley, J., Chou, C-K., and Guy, A.W., U.S. Air Force School of Aerospace Medicine document TF-85-11 1985, pp 1-66.
8. Szmigielski, S., Szudzinski, A., Pietraszek, A., Bielec, M., Janiak, M., and Wrembel, J., Bioelectromagnetics 3: 1982, pp 179-191.
9. Schwertner, H.A., Robinson, D.K., Kiel, J.L., Dunham, R.G., Proceedings (in press), 1992.

ACKNOWLEDGEMENTS

I wish to thank Air Force Systems Command, and the office of Scientific Research for sponsorship of this research.

I am deeply indebted to many individuals here at Brooks AFB for a very productive, refreshing, and stimulating research experience. Ms Marina Jaramillo and Mr Eugene Vega were extremely cooperative and proved to be invaluable in the preliminary and final preparation of this final report. SSgts John Alls and Richard Weber gave freely of their expert technical assistance in the laboratory. I am grateful to Dr Jill Parker and Major Eric Holwitt for their kindness, cooperation and consultations. I am thankful to Dr John Bruno for his gift of RAW 264.7 mouse macrophage cells, his instructions on the use of the wave guide, and his congenial comraderie. The kindness, cooperation, encouragement, direction, and support of Dr Johnathan Kiel contributed greatly to the success of this summer research project.

THE REPORT:
PSYCHOMETRIC, MEASUREMENT, AND PREDICTIVE PROPERTIES

Mary Roznowski
Assistant Professor
Department of Psychology

David Dickter
Graduate Associate
Department of Psychology

Ohio State University
1885 Neil Avenue
Columbus, OH 43210-1222

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, Texas

September, 1992

THE APT:
PSYCHOMETRIC, MEASUREMENT, AND PREDICTIVE PROPERTIES

Mary Roznowski
Assistant Professor

David Dickter
Graduate Associate

Department of Psychology
Ohio State University

Abstract

This report describes the results of a research project carried out to investigate aspects of the measurement quality and construct validity of a battery of cognitive information processing measures (the APT: Automated Personnel Testing). Two samples were used to investigate these issues. Criteria included scores on the Armed Services Vocational Aptitude Battery (ASVAB) and performance measures from computerized, intelligent tutors. The data indicate that the APT battery does an excellent job predicting performance on these various criteria and that these tests could be used to predict individuals' abilities to learn knowledge and skills relevant to many Air Force occupations. Other results indicate acceptable to good psychometric properties of the individual tests comprising the APT. Suggestions are given to improve the psychometric and predictive properties of the measures.

THE APT: PSYCHOMETRIC, MEASUREMENT, AND PREDICTIVE PROPERTIES

Mary Roznowski, David Dickter

Introduction

This report describes the results of a research project carried out to investigate aspects of the measurement and validity of a battery of cognitive information processing measures. The measures constitute a subset of the tests designed by the Learning Abilities Measurement Program (LAMP) which is a program of basic research sponsored by the Air Force Office of Scientific Research and the Air Force's Armstrong Laboratory. A primary goal of this research program is ultimately to improve the selection and classification system used by the U.S. Air Force as well as to learn more about the nature of human abilities and to model human information processing and learning. One hope is that the new types of cognitive measures will improve the predictive validity of the current tests being used as well as increase the coverage of the potential domain of cognitive and intellectual behaviors. The Automated Personnel Testing (APT) battery, which is a subset of the LAMP tests, consists of the tests that are ultimately to be made operational. The APT battery used computerized administration of cognitive tasks in order to measure abilities that include speed of information processing and working memory capacity. These tests will be investigated in this report along with additional measures from the LAMP's CAM (Cognitive Abilities Measurement) battery (Kyllonen et al., 1990).

In order to validate these experimental tests, intelligent microcomputer tutors have been designed which involve up to 30 hours of extensive subject training and testing (Shute, 1992a, 1992b). These microprocessor administered tutors constitute excellent criteria with which to validate the cognitive measures and assess a wide range of verbal, spatial and quantitative skills (see Region & Shute, 1992 for information on automated instruction). These tutors assess the acquisition of knowledge and skills that are relevant to many Air Force specialities.

General Approach

This paper contains a discussion of several types of results. Firstly, item level analyses and results are reported. These analyses will include investigations of item difficulties, item discrimination indices and coefficient alphas. Classical test theory definitions will primarily be used here. Correlation matrices will also be presented in order to examine overlap between these tests and other available cognitive measures such as ASVAB scores. Finally, results will be presented reporting

regression results for the APT tests predicting outcome criteria from the computerized tutors.

This study consisted of the analysis of two separate data sets. Data for both studies were obtained from a data set collected in a long-term series of studies of aptitude-treatment interactions and learning environments (Shute, 1992a)¹. The analyses will be reported in a single section of the paper, however. This is done because of the fact that both data sets were used to answer the same general types of questions.

Method

For data set 1, subjects were 344 temporary service workers who volunteered to be tested on the battery of cognitive tests and the intelligent tutor. Individuals were paid for their participation. 81% were males and 15% were females. These subjects completed the electricity concepts tutor. Subjects completed the ASVAB, the battery of CAM tests and the computerized tutor. Both tutors were self-paced which allowed subjects to control their own rate of progressing through the tutor.

For data set 2, subjects were 402 temporary service workers. 75% were males and 25% were females. These subjects completed the flight engineering tutor. Table 1 contains a summary of demographic characteristics for the two samples and the combined sample. (All tables follow at the end of the report.)

The cognitive ability battery (APT) consists of the following types of measures: processing speed, working memory, fact learning, skill learning, induction and general knowledge. Each of these types of measures or "factors" was assessed with three different content domains: verbal, quantitative, and spatial. Within processing speed, for instance, verbal, spatial and quantitative content domains were assessed. Item level data were available for all of these measures except skill learning. A total of 19 tests were analyzed as part of the APT battery. In addition, analyses were carried out which investigated the remaining tests in the CAM battery (a total of 58 tests). These analyses were done to determine whether tests not chosen for the original APT battery would be suitable for inclusion in a future version of the battery.

The primary criteria used in this research included performance scores on the intelligent microcomputer tutors. One tutor taught students about Ohm's law and other basic electricity principles. The second tutor taught subjects basic skills, knowledge, and abilities about flight engineering.

Subjects mastered the tutor by correctly solving several consecutive problems on a given principle or skill (Shute, 1992). Thus, subjects' time to complete

the tutor could vary greatly and would serve as an excellent indicator of general ability to learn complex skills and perform in a novel environment. Following completion, subjects were given a test of the various principles and knowledge taught in the tutor. A final score, based on overall accuracy on this test, was used as the second criterion. Both provided excellent indicators of individuals' performance and ability to learn and use novel material.

Results

Several types of summary data were available. For each trial within each test type, both correct/incorrect and response time scores were available. Further, both mean and median response time across trials within a test were scored. These various scores were examined and it was decided that median response time for all items (CTA) and percent correct for all items (PCA) would be used.

Table 2 contains item statistics for binary item response data from data set 1 (item data were available for 317 subjects). To get a general idea of the overall quality of the tests in the battery, various item statistics are presented. Included in the table are item difficulties (mean, range), item intercorrelations (mean, range), test-sans-item item-total correlations (range) and coefficient alpha for the latency data. (Test-sans-item item-total correlations are the correlations between a given item and the total score comprising the *remaining* items. Thus, the item is not included in the total for the correlation computed for that item.)

Item difficulties are quite variable with processing speed difficulties being the highest and most values averaging near .90. Some of the fact learning item difficulties were quite low as were several of the induction tests. Item intercorrelations were at reasonable levels with the exception of some negative correlations which might indicate deficiencies with a subset of the items in the various tests. Item total correlations (TSI) were quite good overall. Again, however, some small negative correlations indicate that some items might be deleted from future batteries or perhaps rewritten. The working memory tests look excellent overall with high item total correlations, high alphas and reasonably wide ranges of item difficulties.

Table 3 contains item statistics for item response latency data from the same data set. Included in this table are item intercorrelations (mean, range), test-sans-item item-total correlations (range) and coefficient alpha. The few blanks in the table are a result of not being able to compute the given statistics due to computer memory limitations. Item statistics for the response latencies look quite good overall. Average item intercorrelations are fairly high with very few negative correlations appearing.

Item total correlations are good overall as well. Some of the items with lower discriminations might be examined at and considered for rewriting or removal.

Presented in Table 4 are correlations between the APT tests and AFQT scores. Correlations for both percent correct (PCA) and latency average scores (CTA) are given. The two data sets were combined here. The PCA scores show consistently high correlations with AFQT scores. Note especially the working memory percent correct score correlations. The latency score composites fair less well. The processing speed and working memory latency scores seem to do quite well here, however. These validities are likely to improve somewhat with some psychometric attention paid at the item/trial level.

Table 5 contains correlations between the APT tests and the criterion scores from the electricity tutor (data set 1). Both types of scores are presented (percent correct and latency average scores). PCA correlations look very good overall, especially for the accuracy criterion. It is impressive that such large correlations were obtained given that many of the individual tests were composed of less than 20 items/trials. The working memory tests again perform quite well (both types of scores). Percent correct scores for the induction tests perform very well also. The tutor scores appear to be excellent measures of general cognitive ability. Indeed, the tutors required substantial manipulation of verbal, quantitative and spatial content.

Table 6 contains correlations between speed accuracy scores and the tutor criteria for the electricity tutor. These scores were created for the processing speed and working memory tests only. These tests were used because the latencies showed considerable promise in the previous analyses. Speed accuracy scores were computed using both accuracy and latency scores. The scoring formula used number rights, number wrongs, mean for latencies and a correction for number of options. Correlations of similar magnitude as the original percent correct scores were obtained.

Table 7 contains regression results for the first sample for the electricity concepts tutor. Both accuracy and total time criteria were presented here. First, in the left column, PCA is used to predict the accuracy criterion. Next, the CTA composite of scores is entered along with PCAs and finally AFQTs are entered with the PCAs and CTAs. In the second set, CTA and PCA are reversed where CTA scores are entered first. It is clear that PCAs do an excellent job predicting the accuracy criterion ($R = .75$, $R^2 = .56$). CTAs alone do a reasonable job predicting accuracy though not as well as percent correct scores. Similar results can be seen for the time criterion, with the latency scores fairing a bit better in terms of predictability.

Regression analysis for the second sample for the flight engineering tutor was also carried out. (Results will not be presented in table format to conserve space.) Here, post-test accuracy was predicted from PCAs, PCAs along with CTAs and finally all possible predictors (PCAs, CTAs, AFQT). Similar results to the previous analyses for the electricity tutor were found. Results indicated that both types of scores (PCAs, CTAs) do an excellent job in the prediction of the accuracy criterion. It was also evident that percent correct does a far superior job to latency scores predicting the criteria ($R = .87$, $R^2 = .76$ compared to $R = .56$, $R^2 = .31$).

Table 8 contains regression results for the speed-accuracy scores for the processing speed and working memory tests along with percent correct and response latency scores. The criteria here are the accuracy and total time spent from the electricity tutor. The first set of results show the multiple R, multiple R^2 , differences (increments to R^2) and the F value and significance of the increment. These regression analyses were hierarchical where AFQT was entered, then a composite of PCA scores, a composite of CTA scores and finally the speed accuracy scores. Thus, the final row (marked SP-AC) represent the full model which contains AFQT, PCAs, CTAs and speed accuracy scores. The second set of analyses include the model composed of AFQT and speed accuracy scores alone. These two sets of analyses allow examination of the usefulness of the additional scores in the prediction of the criteria of interest.

The results indicate excellent predictability of the criteria with the various predictors. In addition, it appears that the speed accuracy scores do contribute to prediction, but that the other types of scores are indeed necessary and do contribute unique variance to prediction beyond speed accuracy.

Table 9 presents the regression analyses carried out with either PCAs or SP-AC scores entered first. These analyses were carried out to investigate whether the SP-AC scores could take the place of the individual percent correct and response latency scores. One result of note is that PCAs and CTAs alone do an excellent job in the prediction of the criterion ($R = .67$ for accuracy; $R = .74$ for total time). Further, the speed accuracy scores do an inferior job in prediction compared to either PCAs alone or PCA and response times in combination.

Discussion

The results presented in this report indicate very high quality for the APT battery at both the psychometric and predictive levels. Using two data sets, both with excellent external criteria, scores from the APT tests were found to do an excellent job predicting performance on two complex computerized tutors. Multiple Rs were very high. Further, scores from APT tests added predictability beyond the ASVAB tests.

Similar results were found in work by Christal (1991) for a different set of LAMP tests. Especially promising are the scores reflecting error rates on the computerized cognitive measures. Latencies were also useful and were predictive of the various criteria. Speed-accuracy scores were predictive as well, although did not outperform error rate scores used alone.

Psychometric properties of the measures were good overall. Some improvements in overall quality are very likely with careful attention paid at the item level, such as eliminating items with low or negative item-total correlations. Examining items with negative item intercorrelations would be beneficial as well. Finally, close examination of item difficulties and distributions of item difficulties would be essential.

It is important to point out here that latencies were considered as assessments of a unitary construct. However, this assumption was very likely inappropriate. That is, some of the latencies were from tests that emphasized speed while others stressed accuracy. It is not meaningful to equate these types of latency scores and assume they are equally precise or are equivalent assessments of an underlying construct. Future research needs to treat latencies in a more appropriate fashion.

Authors' Note

The authors are indebted to Dr. Valerie Shute for her extensive work on the tutors and for her efforts in data collection, programming and scoring for both data sets.

References

- Christal, R.E. (1991). Comparative validities of ASVAB and LAMP tests for logic gates learning. (AL-TP-1991-0031). Brooks Air Force Base, Tx: Manpower and Personnel Division, AFHRL.
- Kyllonen, P.C., Christal, R.E., Woltz, D.J., Shute, V.J., Tirre, W.C. and Chaiken, S. (1990). Cognitive Abilities Measurement (CAM) Test Battery, Version 4.0. Unpublished test battery, Armstrong Laboratory, Brooks Air Force Base, Tx.
- Regian, W.J., & Shute, V.J. (1992). *Cognitive Approaches to Automated Instruction*. Hillsdale, NJ: LEA.
- Shute, V.J. (1992a). A Macroadaptive approach to tutoring. *Journal of Artificial Intelligence and Education*.
- Shute, V.J. (1992b). A comparison of learning environments: All that glitters... In S.P. Lajoie and S.J. Derry (Eds.), *Computers as cognitive tools*. Hillsdale, NJ: Erlbaum.

Table 1

Demographic Characteristics of Data Sets: Individual and Combined

	<u>Data Set 1</u> N = 344	<u>Data Set 2</u> N = 402	<u>Combined Data</u> N = 746
Females	50 15%	99 25%	149 20%
Males	279 81%	301 75%	580 78%
Hispanic	147 43%	188 47%	335 45%
Black	33 10%	56 14%	89 12%
White	133 39%	139 35%	272 36%
High School	133 39%	119 30%	252 34%
GED	48 14%	45 11%	93 12%
Some College	129 39%	195 49%	324 43%
<i>English Main</i>			
Language	290 84%	351 87%	641 86%

Table 2
Item Statistics for Binary Item Response Data (Data Set 1)

<u>Percent Correct Statistics</u>				
	Item Difficulty	Item Intercorrelations	TSI	Alpha
	\bar{x} range	\bar{x} range	range	
<i>Processing Speed</i>				
Quantitative 4	.86 (.69 to .97)	.09 (-.12 to .49)	-.06 to .56	.89
Verbal 4	.92 (.83 to .96)	.04 (-.11 to .35)	-.02 to .41	.73
Spatial 4	.92 (.78 to .99)	.06 (-.11 to .48)	-.02 to .56	.79
<i>Working Memory</i>				
Quantitative 4	.62 (.43 to .87)	.41 (.21 to .63)	.40 to .72	.92
Verbal 4	.56 (.34 to .75)	.39 (.19 to .58)	.43 to .69	.91
Spatial 4	.33 (.19 to .58)	.45 (.32 to .60)	.57 to .74	.93
<i>Induction</i>				
Quantitative 1	.47 (.39 to .61)	.03 (-.10 to .22)	-.01 to .23	.23
Quantitative 2	.50 (.22 to .89)	.19 (.03 to .38)	.16 to .56	.70
Quantitative 3	.55 (.17 to .79)	.25 (.01 to .50)	.16 to .63	.73
Verbal 1	.57 (.46 to .68)	.21 (.04 to .42)	.22 to .52	.73
Verbal 2	.36 (.13 to .88)	.11 (-.02 to .29)	.03 to .36	.56
Verbal 3	.34 (.02 to .56)	.28 (-.01 to .58)	.11 to .69	.82
Spatial 1	.60 (.22 to .92)	.09 (-.11 to .33)	-.08 to .40	.50
Spatial 2	.79 (.46 to .92)	.21 (.00 to .42)	.26 to .44	.71
Spatial 3	.50 (.19 to .84)	.21 (-.07 to .45)	-.02 to .57	.71

Note: N = 318.

Table 2 (cont.)

		<u>Fact Learning</u>			
Quantitative 1	.44 (.12 to .78)	.21 (-.00 to .41)	.22 to .54	.87	
Quantitative 2	.74 (.49 to .90)	.07 (-.19 to .48)	.04 to .48	.80	
Verbal 1	.14 (.01 to .44)	.18 (-.11 to .89)	.11 to .21	.27	
Verbal 2	.75 (.62 to .82)	.36 (.15 to .53)	.40 to .68	.92	
Spatial 1	.30 (.07 to .47)	.29 (-.13 to .51)	.02 to .60	.84	
Spatial 2	.77 (.42 to .93)	.12 (-.13 to .46)	.13 to .55	.87	
		<u>General Knowledge</u>			
Verbal 1	.58 (.17 to .86)	.23 (.02 to .44)	.18 to .59	.88	
Verbal 2	.78 (.50 to .93)	.05 (-.14 to .34)	-.00 to .29	.71	
Spatial	.73 (.51 to .86)	.14 (-.19 to .52)	.09 to .46	.83	

Note: N = 318.

Table 3
Item Statistics for Response Latency Data (Data Set 1)

<u>Response Latency Statistics</u>			
	Item Intercorrelations	TSI	Alpha
	\bar{x} range	range	
<i><u>Working Memory</u></i>			
Quantitative 4	.38 (.17 to .57)	.46 to .67	.91
Verbal 4	.31 (.15 to .45)	.43 to .60	.87
Spatial 4	.32 (.16 to .45)	.44 to .60	.88
<i><u>Processing Speed</u></i>			
Quantitative 4	.28 (.00 to .98)	---	.95
Verbal 4	.27 (.00 to .93)	---	.93
Spatial 4	.30 (-.05 to .98)	---	.94
<i><u>Induction</u></i>			
Quantitative 1	.44 (.34 to .52)	.56 to .65	.89
Quantitative 2	.27 (.12 to .45)	.32 to .55	.80
Verbal 1	.32 (.18 to .46)	.43 to .59	.82
Verbal 2	.27 (.01 to .41)	.16 to .57	.79
Verbal 3	.32 (.15 to .48)	.41 to .62	.82
Spatial 1	.28 (.15 to .43)	.35 to .59	.79
Spatial 2	.28 (.12 to .42)	.36 to .56	.78
Spatial 3	.36 (.18 to .58)	.43 to .67	.83

Note: N = 318.

Table 3 (cont.)

<u>Response Latency Statistics</u>			
	Item Intercorrelations	TSI	Alpha
	\bar{x} range	range	
		<i>Fact Learning</i>	
Verbal 1	.33 (.24 to .46)	.42 to .53	.71
Verbal 2	.28 (.14 to .46)	.37 to .58	.88
Spatial 1	.26 (.05 to .54)	.29 to .61	.71
Spatial 2	.30 (.03 to .55)	.26 to .65	.95
Quantitative 1	.24 (.03 to .44)	.24 to .57	.88
Quantitative 2	.27 (.00 to .59)	.28 to .65	.93
		<i>General Knowledge</i>	
Verbal 1	.42 (.16 to .73)	.45 to .79	.93
Verbal 2	.39 (.08 to .66)	---	.96
Quantitative 1	.41 (.29 to .53)	.43 to .61	.80
Quantitative 2	.54 (.39 to .70)	.55 to .73	.86
Quantitative 3	.46 (.36 to .55)	.54 to .65	.83
Quantitative 4	.53 (.40 to .64)	.58 to .76	.92
Spatial 2	.18 (-.05 to .52)	.13 to .59	.85

Table 4

Correlations between APT Tests and AFQT Scores for Both Types of Scores (Combined Data)

<u>Score type</u>	<u>AFQT</u>	
	<u>PCA</u>	<u>CTA</u>
PSV4	.08	-.39
PSQ2	.45	-.21
PSS1	.47	-.12
WMV1	.63	-.31
WMQ4	.59	-.46
WMS3	.63	-.20
FLV3	.36	-.38
FLQ2	.40	-.04*
FLS1	.57	-.12*
SLV2	.46	-.21
SLQ3	.59	-.02*
INV1	.56	.03*
INQ3	.54	.02*
INS2	.60	-.18
GKQ1	.15	-.10
GKQ2	.16	.01*
GKQ3	.56	-.02*
GKQ4	.33	-.17
GKS2	.64	.06

Note: N = 615. All significant at .01 except as marked (*).

Table 5

Correlations between APT Tests and Electricity Tutor
Criterion Scores (Data Set 1)

<u>Score type</u>	<u>Post Accuracy</u>		<u>Total Time</u>	
	<u>CTA</u>	<u>PCA</u>	<u>CTA</u>	<u>PCA</u>
PSV4	-.33	.10*	.36	-.09*
PSQ2	-.06*	.35	.12*	-.40
PSS1	.01*	.48	.07*	-.47
WMV1	-.24	.49	.36	-.56
WMQ4	-.30	.45	.43	-.45
WMS3	-.12*	.51	.18	-.55
FLV3	-.34	.35	.35	-.26
FLQ3	.04*	.31	.10*	-.25
FLS1	-.04*	.55	.17	-.44
SLV2	-.03*	.40	.13*	-.41
INV1	.08*	.43	-.02*	-.39
INQ3	.13*	.47	-.05*	-.40
INS2	-.05*	.52	.10*	-.55
GKQ1	-.04*	.04*	.09*	.04*
GKQ2	.04*	.07*	.04*	-.14
GKQ3	.02*	.49	.01*	-.43
GKQ4	-.06*	.19	.13*	-.13*
GKS2	.05*	.53	.04*	-.54

Note: N = 287. All significant at .01 except as marked.

Table 6

Correlations for Speed Accuracy Scores for PS and WM tests (Data Set 1)

	<u>Accuracy</u>	<u>Total Time</u>
PSV4	.38	-.42
PSQ2	-.07*	.03*
PSS1	.28	-.36
PSQ4	.26	-.36
PSS4	.19	-.22
WMV1	.37	-.43
WMQ4	.47	-.49
WMS3	.51	-.54
WMV4	.49	-.48
WMS4	.52	-.46

Note: N = 318.

Table 7

Regression Results for the Electricity Concepts Tutor Criteria

(Data Set 1)

	<u>Post Accuracy</u>		<u>Total Time</u>	
PCA	R = .75		PCA R = .73	
	R ² = .56		R ² = .53	
+ CTA	R = .75	D = .00	+ CTA R = .77	D = .06
	R ² = .57	F = .44 (ns)	R ² = .59	F = 3.6 (.00)
+ AFQT	R = .78	D = .04	+ AFQT R = .77	D = .01
	R ² = .60	F = 24.2 (.00)	R ² = .60	F = 9.7 (.00)
CTA	R = .50		CTA R = .57	
	R ² = .25		R ² = .32	
+ PCA	R = .75	D = .32	+ PCA R = .77	D = .26
	R ² = .57	F = .12 (.00)	R ² = .59	F = 9.9 (.00)
+ AFQT	R = .78	D = .04	+ AFQT R = .78	D = .01
	R ² = .60	F = .24 (.00)	R ² = .60	F = 9.7 (.00)

Note: N = 300.

Table 8

Regression Results for AFQT, PCA, CTA and Speed Accuracy
Scores for PS and WM Tests

	<u>Post Accuracy</u>	<u>Total Time</u>
AFQT	R = .72 R ² = .52	R = .68 R ² = .47
+ PCA	R = .75 D = .04 R ² = .56 F = 2.59 (.00)	R = .73 D = .07 R ² = .53 F = 4.21 (.00)
+ CTA	R = .76 D = .01 R ² = .58 F = .78 (ns)	R = .77 D = .06 R ² = .59 F = 4.06 (.00)
+ SP-AC	R = .77 D = .01 R ² = .59 F = .68 (ns)	R = .77 D = .01 R ² = .60 F = .38 (ns)

AFQT	R = .72 R ² = .52	R = .68 R ² = .47
+ SP-ACCR	R = .75 D = .03 R ² = .56 F = 2.00 (.03)	R = .73 D = .07 R ² = .54 F = 4.56 (.00)

Note: N = 300.

Table 9

Regression Results for Speed Accuracy for PS and WM Tests

	<u>Post Accuracy</u>	<u>Total Time</u>
PCA	R = .66	R = .68
	R = .43	R ² = .47
+ CTA	R = .67 D = .02	R = .74 D = .08
	R = .45 F = 1.10 (ns)	R ² = .55 F = 5.10 (.00)
+ AFQT	R = .76 D = .12	R = .77 D = .04
	R = .58 F = 80. (.00)	R ² = .59 F = 30.7 (.00)

SP-ACC	R = .62	R = .66
	R = .39	R ² = .43
+ AFQT	R = .75 D = .16	R = .73 D = .11
	R = .56 F = 106. (.00)	R ² = .54 F = 68. (.00)

Note: N = 300.

ANALYSIS OF ISOCYANATES
IN SPRAY-PAINT OPERATIONS

Walter E. Rudzinski
Professor
Department of Chemistry

Southwest Texas State University
San Marcos TX, 78610

Final Report for:
Summer Research Program
Armstrong Laboratory
Analytical Services Division

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

ANALYSIS OF ISOCYANATES
IN SPRAY-PAINT OPERATIONS

Walter E. Rudzinski
Professor
Department of Chemistry

Southwest Texas State University
San Marcos TX, 78610

Abstract

NIOSH Method 5521 was used for the analysis of toluene diisocyanate (TDI), 4,4'-diphenylmethane diisocyanate (MDI), and hexamethylene diisocyanate (HDI) monomers, Des N-75 polymer, as well as for the kinetic analysis of the reactivity of Des N-75 in a paint formulation. Preliminary results were also obtained for the chromatographic separation of Des N-3300 and Des Z-4370 prepolymers, and isophorone monomer. The NIOSH method was modified for the analysis of polyisocyanate by the use of standards prepared from the bulk prepolymers.

Field studies were also conducted at several spray painting operations, and NIOSH Method 5521 was compared directly with OSHA Method 42 for the analysis of HDI and Des N-75. At high concentrations of particulate, NIOSH Method 5521 gave higher HDI results than those obtained from OSHA Method 42. Total nuisance dust and particle size distributions were evaluated in one field study, and the Des N-75 polyisocyanate results were reasonable when compared with the ratio of polyisocyanate expected in the sample based on the MSDS data sheets and the paint formulation.

ANALYSIS OF ISOCYANATES IN SPRAY-PAINT OPERATIONS

Walter E. Rudzinski

Introduction

Methods have been developed for the analysis of various isocyanates: toluene diisocyanate (TDI), 4,4'-diphenylmethane diisocyanate (MDI) and hexamethylene diisocyanate (HDI). These methods depend on sample collection with either an impinger or filter, derivitization of the reactive isocyanate, then separation using reversed phase HPLC with either uv, fluorescent or electrochemical (ec) detection. The entire area has been the subject of various reviews [1,2].

Currently there are three accepted procedures which are employed for the analysis of isocyanates:

Mobay Method 1.4.3 [3] which involves sample collection with an impinger, derivitization of the reactive isocyanate with N-4-nitrobenzyl-propylamine, and separation using reversed phase HPLC with uv detection. The method was developed for the analysis of TDI, HDI, MDI and 4,4' methylene bis(4-cyclohexyl) isocyanates (monomer and polyisocyanates).

NIOSH Method 5521 [4] which analyzes for TDI, MDI and HDI using an impinger, 1-(2-methoxyphenyl)-piperazine in toluene as the derivitizing reagent, a C-8 column for HPLC separation and uv and ec detectors for eluent quantitation. The method is adequate for TDI, MDI and HDI monomers but gives low recoveries for MDI and HDI prepolymers.

OSHA Method 42 [5] which collects samples by drawing air through a fiber filter coated with 1-(2-pyridyl)-piperazine. The isocyanatoureas are then desorbed, and analyzed using HPLC with uv or fluorescence detection. The derivatives have higher molar absorptivities than those formed with nitro reagents. The method has been validated for TDI and HDI (not MDI), although because of laboratory limitations, controlled test atmospheres of diisocyanates were not generated. Humidity dramatically affected the sample recovery, and phenol which is present in inhibited isocyanate is an interferent.

Of the three methods of analysis, all appear to be adequate for

isocyanate monomer, but only Mobay Method 1.4.3 and NIOSH Method 5521 appear to be promising for the analysis of isocyanate oligomer. Recent comparative studies of isocyanate sampling protocols show that the impinger collection method is superior to filter collection whenever the particulate concentration is high [6], or when the particle size is large [7]. Furthermore of the two methods which employ an impinger, NIOSH Method 5521 appears to be more promising since the limits of detection for the method are 0.2 µg/L .

In the following report, NIOSH Method 5521 was used for the analysis of HDI, TDI and MDI monomers, Des N-75 polymer, as well as for the kinetic analysis of the reactivity of Des N-75 in a paint formulation. Preliminary results were also obtained for the chromatographic separation of Des N-3300 and Des Z-4370 prepolymers, and isophorone monomer.

Field studies were also conducted at several spray painting operations, and NIOSH Method 5521 was compared directly with OSHA Method 42 for the analysis of HDI and Des N-75.

Experimental

Materials and Methods

HDI, TDI and MDI urea standards were prepared as described in the OSHA Method 42 and NIOSH Method 5521 protocols. An isophorone urea standard was prepared by reacting a measured amount of isophorone diisocyanate with excess 1-(2-methoxyphenyl) piperazine (Aldrich) in toluene, and then letting the solution sit overnight. Ten µL of acetic anhydride were then added, the solution evaporated, and then reconstituted in 5 mL of methanol.

Des N-75 standards were prepared from bulk prepolymer, Des N-100 (Miles, Pittsburgh, PA), which is 35-40% biuret trimer of HDI and 35% polyisocyanate (MSDS data sheet and private communication). Des N-3300 and Des Z-4370 standards were prepared from the corresponding bulk prepolymer solutions (Miles, Pittsburgh, PA). Stock solutions were prepared by dissolving c.50 mg of the bulk prepolymer in 50 mL of toluene or methylene chloride. Aliquots were then added to 1.5 mL of 430 mg/L 1-(2-methoxyphenyl) piperazine (MOPIP) and the solution brought up to 10-15 mL with toluene. The excess derivitizing reagent was acetylated with 10 µL of acetic anhydride, and the toluene evaporated at 55 °C under nitrogen. The polyisocyanate urea derivatives formed were then reconstituted in 2-5 mL of methanol.

All isocyanates were analyzed according to NIOSH Method 5521 protocols, except that the concentration of organic modifier in the mobile phase was sometimes increased to either: 50% methanol: 50% methanolic acetate buffer or 45% acetonitrile: 55% methanolic acetate buffer.

The polyisocyanate in polyurethane paints based on Des N-75 was quantitated using Des N-75 standard solutions. For other polyurethane paint formulations polyisocyanate was only estimated since genuine standards were not available. In these cases, polyisocyanate was estimated based on the Des N-75 electrochemical/uv response ratio of Des N-75 standard solutions.

A second ratio method was also evaluated for the quantitation of polyisocyanate. After air sampling for total particulate, the result is then multiplied by the ratio of % isocyanate/% solids in the paint formulation. The ratio was compared directly with the results obtained from NIOSH Method 5521.

Instrumentation

A Hewlett-Packard Series II/L 1090 high performance liquid chromatography (HPLC) equipped with an autosampler, diode array uv-vis detector and/or 1049A programmable electrochemical detector, and HP 3396 series II integrators was used to analyze for all isocyanates. The electrochemical detector was equipped with a glassy carbon electrode operated in the amperometric mode at +0.8 V, while the diode array detector was operated at 242 nm. Either one, two or three Hypersil ODS 5 μ columns in series were used for the analyses. Three columns were required to completely resolve HDI in the presence of TDI.

Sampling

The sampling methods employed were:

1. NIOSH Method 5521 which employs an impinger containing 1-(2-methoxyphenyl) piperazine (MOPIP) reagent in toluene. The MOPIP converts isocyanate into a urea derivative.
2. OSHA Method 42 which employs a 37 mm cassette containing a glass fiber filter coated with 0.1 mg of 1-(2-pyridyl) piperazine (PPIP). The OSHA Method was modified so that 2.0 mg of (PPIP) reagent were used.
3. NIOSH Method 0500 which measures the total nuisance dust. Area samples were collected for the duration of the spray-painting.

Spray Paint Operations

(Langley AFB)

Five different spray paint operations were evaluated. High volume low pressure (HVLP) spray guns were used throughout. Several methods were used concurrently to collect samples in the area of overspray. Air sample inlets were 3-7 ft downdraft from the equipment being painted. Sample inlets were 3 1/2 - 4 1/2 ft above the floor. For both OSHA method 42 and NIOSH Method 0500, samples were collected open-faced. For all sampling the air pump flow rate was 1 L/ minute.

In four of the spray paint operations, breathing zone impinger samples were collected. In two of these four, operations #4 and #5, a personal cascade impactor was also used to evaluate the particle size distribution in the breathing zone.

Operations #1 and #2 involved spray-painting an aircraft towing vehicle (USAF MB-4) and a small generator respectively in corrosion control operations using a Deft polyurethane paint (olive drab pigment: N-75 hardener in a 1:1 ratio).

Operation #3 involved the spray-painting of a quarter panel of a vehicle with PPG Star (pigment:hardener in a 4:1 ratio). The pigment was blue, and the hardener was a mixture of Des N-3300 and Des Z-4370 which contain, predominately, the isocyanurate of HDI and a trimer of isophorone diisocyanate respectively. The area sampler was positioned 5 feet downdraft at right angles to the overspray. No personal samples were taken.

Operation #4 involved spray painting the lower half of a generator with Dupont Imron Polyenamell and Imron 192S Hardener in a 4:1 ratio. Imron 192S consists of a mixture of Des N-75 and Des Z-4370 which contain HDI, aliphatic polyisocyanates and a trimer of isophorone.

Operation #5 involved spray-painting the upper half of a generator with Deft, glossy black, polyurethane paint (black pigment: N-75 hardener in a 1:1 ratio). The area sampler was positioned 4 feet downdraft.

(Edwards AFB)

Two different spray paint operations involving both orange and white polyurethane paints were evaluated. The protocols in NIOSH Method 5521 were used for all area and personal sampling. Area impingers were located in the

vicinity of an F-16 fighter plane, but away from the tail which was being painted. Personal samples were taken in the breathing zone of the painters. The polyurethane paint hardener was Des N-75 .

(Travis AFB)

Five different spray paint operations were sampled in a paint booth over a period of one week. The sampling methods were those outlined in NIOSH Method 5521 and OSHA Method 42. Samples were taken at both exhaust and recirculation ducts as well as two personal samples for each painter during the spray paint operation.

(Wright-Patterson AFB)

Two different spray paint operations were sampled over a period of two weeks. The sampling methods were those outlined in NIOSH Method 5521.

Results and Discussion

HDI Standards

Figures 1 and 2 plot the ec and uv responses obtained for five HDI standards prepared according to NIOSH Method 5521.

Figure 1. EC Response for HDI (NIOSH Method 5521)

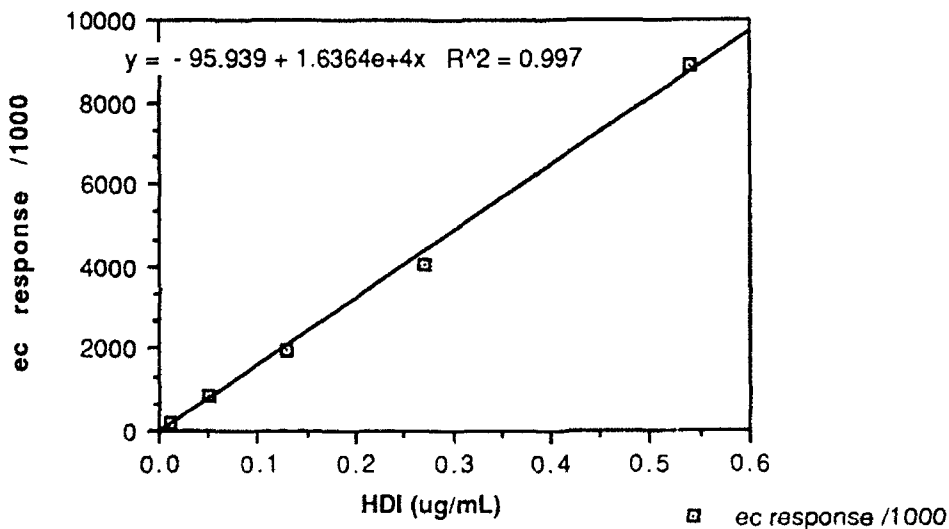
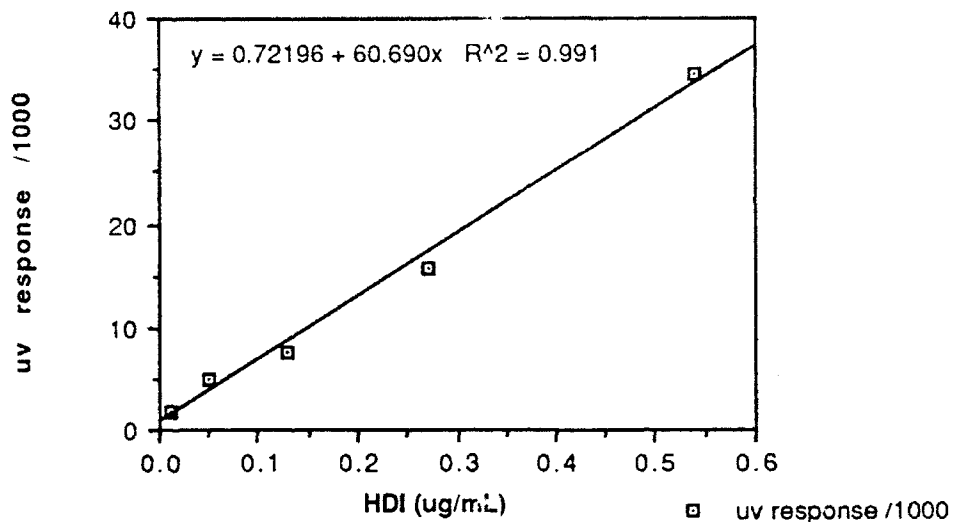


Figure 2. UV Response for HDI (NIOSH Method 5521)

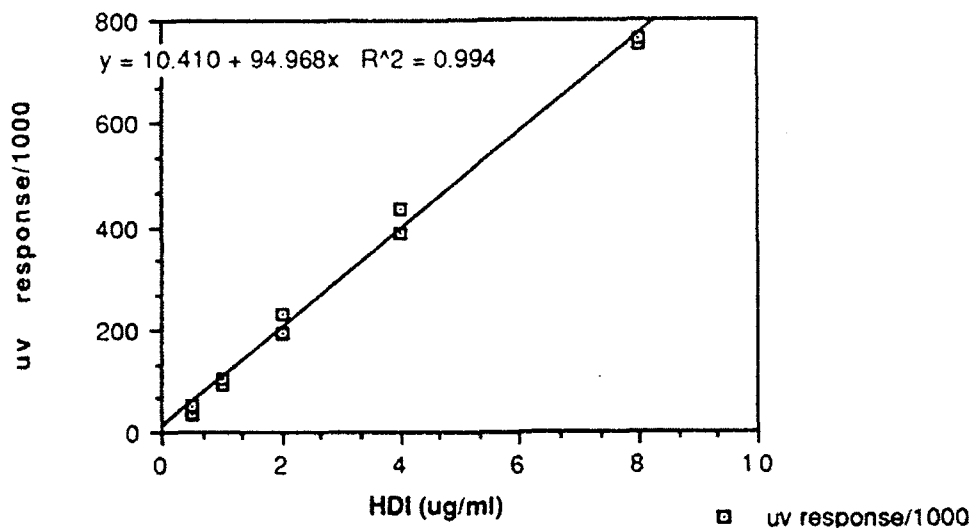


The coefficient of correlation is greater than 99% for each working curve. The linear range of the ec detector extends from the limit of quantitation which is below 0.01 $\mu\text{g}/\text{mL}$ to at least 0.54 $\mu\text{g}/\text{mL}$, while the linear range of the uv detector begins at 0.2 $\mu\text{g}/\text{mL}$ and extends to beyond 0.54 $\mu\text{g}/\text{mL}$ when injecting 10 μL of sample.

The slope associated with the ec response is a measure of the analytical sensitivity and is 16.4×10^6 counts/ ($\mu\text{g}/\text{mL}$), while the slope of the uv response is 60.7×10^3 counts / ($\mu\text{g}/\text{mL}$). The relative analytical sensitivity then, as measured by the ratio of slopes or the ec/uv response ratio, is 269 which means that the sensitivity of the ec detector is 269X as large as that of the uv detector. This relative sensitivity is variable and depends primarily on the physical state of the glassy carbon electrode of the ec detector. For this data set (obtained on 6-8-92) the electrode had just been polished, and so the relative ec/uv ratio was high. With continuous use the analytical sensitivity of the ec detector diminishes and then remains relatively constant at about 6×10^6 counts/ ($\mu\text{g}/\text{mL}$). The analytical sensitivity for the uv detector remains at about 60×10^3 counts / ($\mu\text{g}/\text{mL}$), and is not as susceptible to long term drift. Ordinarily the ec/uv ratio is about 100 for HDI. The relative standard deviation from the mean for 7 working standards at 0.6 $\mu\text{g}/\text{mL}$ was 4% for the uv detector and 2% for the ec detector.

Hexamethylene diisocyanate urea, HDIU, derivatives were also prepared by the reaction of 1-(2-pyridyl)-piperazine, PPIP, with hexamethylene diisocyanate, HDI, according to OSHA Method 42. Figure 3 plots the uv response obtained for five HDI standard solutions.

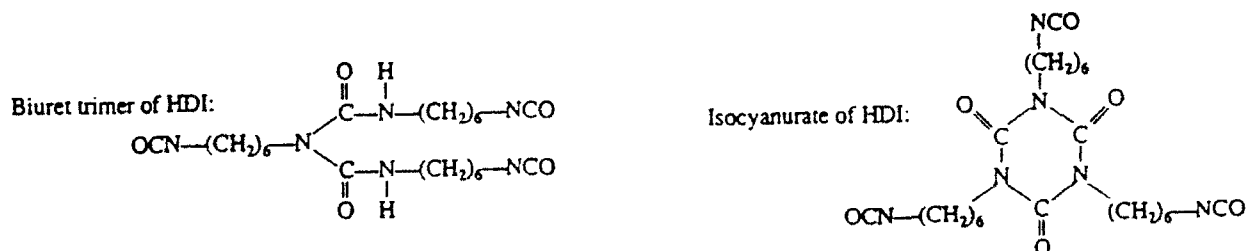
Figure 3. UV Response for HDI (OSHA Method 42)



The response is linear with a slope of 95×10^3 counts/($\mu\text{g}/\text{mL}$). This value is close to that reported in OSHA Method 42 [5].

Polyisocyanate Standards

Standard solutions were prepared from the following Miles' hardeners: Des N-75, Des N-3300 and Des Z-4370. Des N-75 contains HDI, the biuret trimer of HDI, and other higher molecular weight oligomers of HDI (3-5 ring structures). Des N-3300 contains mostly the isocyanurate of HDI. Des Z-4370 contains isophorone and the trimer of isophorone. The biuret and isocyanurate structures are depicted below:



Chromatograms were obtained for all of the polyisocyanates under identical conditions, and Figure 4 represents the chromatogram for Des N-75.

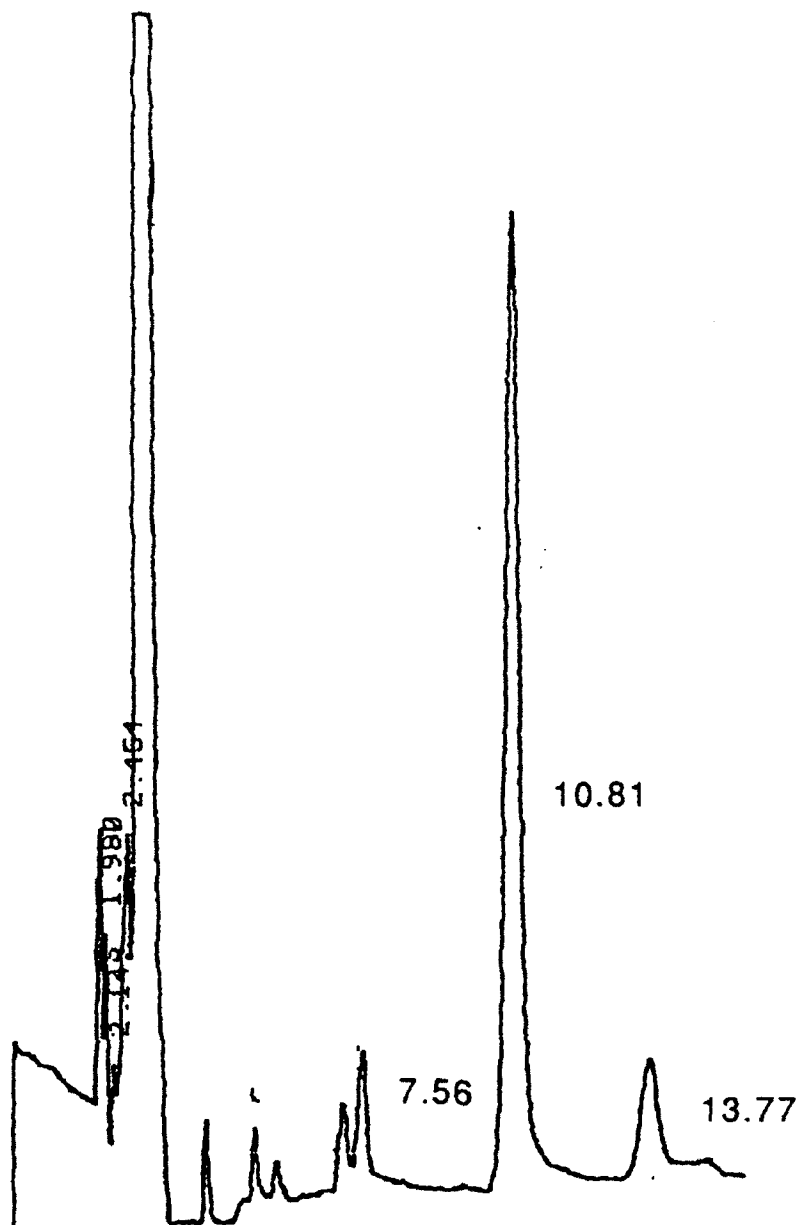


Figure 4. Chromatogram of the separation of Des N-75. Columns, two HP Hypersil ODS 5 μ , 100 mm X 4.6 mm; mobile phase, acetonitrile: methanolic acetate buffer, pH =6 (45:55); flow rate = 1.0 mL/min.; detection, uv at 242 nm.

Des N-75 has 3 characteristic peaks. The largest peak represents 64-80% of the total area (excluding the solvent and reagent blank peaks) and is attributed to the biuret of HDI; the peak appears after 10.8 minutes (k' = 4.4).

This peak is flanked by 2 smaller peaks: the first peak, which has about 10% of the area of the largest peak, appears after 7.6 minutes ($k' = 2.8$); the second peak, with about 20% of the area of the largest peak, appears after 13.8 minutes ($k' = 5.9$) and is attributed to the isocyanurate of HDI (based on a comparison with Des N-3300 which is predominately this species and has a chromatographic peak at 13.8 minutes). Des N-75 also has additional peaks which are sometimes observed after 36 and 40 minutes. These latter two peaks can be attributed to higher molecular weight oligomers.

Des N-3300 has two characteristic peaks: a large peak which represents about 66-74% of the total area, is attributed to isocyanurate and appears after 13.8 minutes ($k' = 5.9$); a second smaller peak which represents about 25 % of the area of the largest peak and appears after 8 minutes ($k' = 3.5$)

Des Z-4370 has two characteristic peaks with very long retention times: a larger peak which represents about 42% of the total area, and appears after 27 minutes ($k' = 14$), and a second peak at 16.5 minutes ($k' = 8$).

Under the same chromatographic conditions, HDI, IPDI and MDI appear at 4.06, 6.23 and 6.45 minutes respectively. The two isomers of TDI appear at 4.06 and 4.30 minutes respectively, and the first isomer of TDI cannot be resolved from HDI.

Figures 5 and 6 plot the ec and uv responses for the largest peak in the chromatogram (assumed to be biuret) for six N-75 standards.

Figure 5. EC Response for N-75

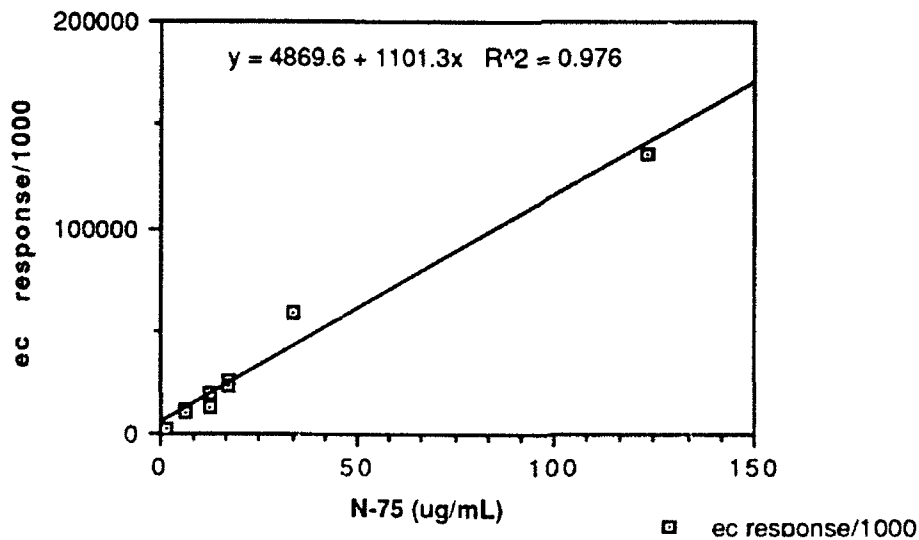
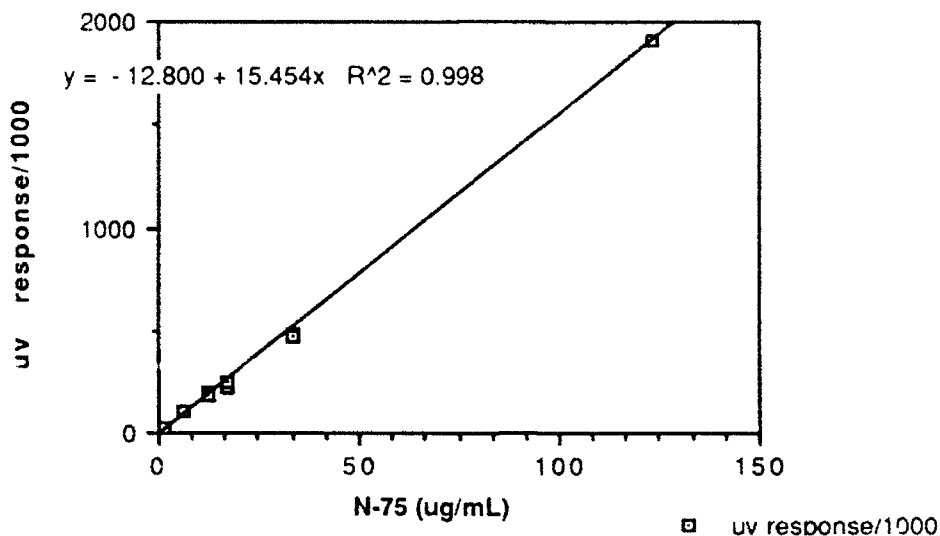


Figure 6. UV Response for N-75



The chromatographic response is linear for N-75 with a coefficient of correlation greater than 97% for each working curve. The linear range of both the ec and uv detectors extends from the limit of quantitation which is near 1 $\mu\text{g/mL}$ to at least 123 $\mu\text{g/mL}$ when injecting 10 μL of sample.

The slope of the ec response is 1×10^6 counts/ $(\mu\text{g/mL})$, while the slope of the uv response is 15.5×10^3 counts/ $(\mu\text{g/mL})$. The ec/uv ratio is on the order of 100. The predicted relative response for the biuret of HDI with respect to HDIU is 0.21 based on an equivalent weight of 159 grams for the biuret (compare with the equivalent weight of HDI, i.e., 84 g) and based on a concentration of 40% biuret in N-100. The actual relative response was 0.25 for the uv detector, and 0.16 for the ec detector. Both of these responses bracket the expected value of 0.21 and lend credence to the assumption that the largest peak is indeed the biuret of HDI.

Field Data

Table I on the next page presents the results for the analysis of HDI for 5 field operations at Langley AFB. All concentrations (conc) are expressed in $\mu\text{g/mL}$.

TABLE I. Langley AFB, HDI Analysis (6-23-92)

op	ID	type	uv resp	ec resp	ec/uv	conc	mg/m3
1	584	personal	96000	12564000	131	1.68	0.12
	585	personal	0	0		0	
	586	personal	102000	6207000	61	1.74	0.14
	587	area	253000	17199000	68	4.80	0.14
2	588	area	266000	17960000	68	2.55	0.27
	588	area	277000	11047000	40	3.00	0.32
	589	personal	38000	1987000	52	0.54	0.12
	590	personal	30100	1982000	66	0.54	0.12
3	593	area	0	0		0	
4	594	personal	59000	9253000	157	2.19	0.08
	594	personal	60000	7815000	130	2.19	0.08
	595	area	9400	658000	70	0.15	0.005
5	591	blank	0	0		0	
	592	blank	0	0		0	
	596	personal	16000	1780000	111	0.27	0.02
	597	area	26400	1262000	48	0.33	0.02
	598	blank	0		0	0	
					ave ec/uv	83.4	
				std. dev.	38		
	HDI	standard		2295000		0.53	
	HDI	standard	34600	2232000	65	0.53	
	HDI	standard	31900	2215000	69	0.53	
	HDI	standard	16400	1112000	68	0.26	
	HDI	standard	17700	1108000	63	0.26	
	HDI	standard	8400	577000	69	0.13	
	HDI	standard	8058	577000	72	0.13	
				ave ec/uv	67.4		
				std. dev.	3		

The ec/uv response ratio for field samples is highly variable, but the true value, 67, lies within the 95% confidence interval of the mean of the sample data set (83 ± 24). Individual sample ec/uv values are often outside the range of 0.75 to 1.5 times the average ec/uv ratio for the HDI standard solutions, but the chromatographic retention time for HDI is diagnostic, and can be used to readily identify HDI.

Table I also shows that detectable HDI concentrations range from a low of 0.005 to a high of 0.32 mg/m3 depending upon the type of spray paint operation. The TLV value is 0.034 mg/m3 so that personal protective equipment is needed, especially for the spray painting of large vehicles and objects (operations #1 and #2). Spot spray painting did not seem to generate as much HDI (operations #4 and #5). Sampling during operation #3 was at right angles to the overspray, and little HDI was observed as expected.

TABLE II. Langley AFB, N-75 Analysis (6-23)

op	ID	type	uv resp	ec response	ec/uv	conc	mg/m3
1	584	personal	1167000	36130000	31	88	6.3
	585	personal	27200	0			0
	586	personal	1148000	71577000	62	87	6.7
	587	area	2882000	85377000	30	228	6.7
2	588	area	3849000	108145000	28	154	16.2
	588	area	1948000	55764000	29	152	16.0
	589	personal	558000	16955000	30	38	8.5
	590	personal	798000	19804000	25	58	12.9
5	591	blank	124000	4002000	32	3	
	592	blank	0	0		0	
	596	personal	1008000	33430000	33	75	5.0
	597	area	2243000	59480000	27	176	13.0
	598	blank	10700	287000	27	0	
				ave ec/uv	32		
				std. dev.	3		
	N-75	standard	558000	22346000	40	39	
	N-75	standard	571000	24355000	43	39	
	N-75	standard	276000	11513000	42	16	
			ave ec/uv	41			
			std. dev.	1			

Standard deviation of sample set excludes #586

Table II presents the results for the analysis of N-75 taken on the same day and on the same data set as in Table I. The ec/uv ratio for a series of samples and standards ranges from a low of 25 to a high of 43 (the ec/uv value = 62 is presumed to be an outlier). The ec/uv ratios are consistently lower than any ec/uv value for a genuine HDI monomer standard (these range from 63-72). These results call into question the premise postulated in MDHS Method 25 [9,10] that the isocyanate peaks may be identified and quantified on the basis of their ec/uv response ratio and standard solutions based on isocyanate monomer. The results also agree with a recent report prepared by NIOSH staffers [11] which indicates that monomer solutions can not be used as standards for the quantitation of polyisocyanate derivatives. All the polyisocyanate concentrations in air were therefore analyzed on the basis of Des N-75 standard solutions. All hardeners used at Langley AFB were based on Des N-75, except for field operation #4 which used Des N-3300.

Table II shows that the N-75 concentration in air during spray paint operations ranges from 5.0-16.2 mg/m3. At present there are no OSHA permissible exposure limits (PELs) or short-term exposure limits (STELs), no NIOSH recommended exposure limits (RELs) , and no American Conference of Governmental Industrial Hygienists (ACGIH) threshold limit values (TLVs) for

these compounds. The manufacturer (Miles) recommends 1 mg/m³ as an exposure limit, while the Swedish standard lists a 5-min STEL of 0.2 mg/m³ and a TLV-TWA (threshold limit value-time weighted average) of 0.09 mg/m³ [12]. The Oregon OSHA standard has a ceiling standard of 1 mg/m³ and a 8-hr TWA of 0.5 mg/m³ [13]. The spray paint operations at Langley AFB exceed the ceiling standard of 1 mg/m³ by a factor of at least 5 consistently. These results are very typical. In a 10 year Oregon study HDI polyisocyanates exceeded the ceiling value 42% of the time, and reached up to 29.5 mg/m³.

In order to verify that the N-75 samples did indeed exceed the ceiling values, the samples were diluted by a factor of 10 and rerun after 2 weeks. In every case, the recalculated value was above the ceiling limit, within the working range of the N-75 standards, and within $\pm 25\%$ of the original value. The results suggest that the N-75 concentration is reproducible, and that the samples are stable for two weeks after sample preparation when stored at 4 C.

TABLE III. Langley AFB, Isocyanate Analysis

op	type	HDI		N-75		T.P.	MMAD
		NIOSH	OSHA	(6-23)	(7-7)		
1	personal	0.12		6.3	5.1		
	personal	0.00					
	personal	0.14		6.7	7.2		
	area	0.14	0.067	6.7	5.1	8.4	31.43
2	area	0.30	0.119	16.1	13.9	12.4	46.85
	personal	0.12		8.5	9.2		
	personal	0.12		12.9	9.7		
3	area	0.00	< 0.016			0.0	0.03
4	personal	0.08		11.1	4.7		34
	area	0.00	< 0.008		4.9	2.1	15.27
5	personal	0.02		5.0	5.2	10.4	28
	area	0.02	< 0.016	13.0	11.1	5.6	15.05

op = operation

theor. = expected concentration of N-75 polyisocyanate based on weight percent in paint formulation

T.P. = total particulates, for personal samples these were obtained using a midget cascade impactor.

MMAD = median mass aerodynamic diameter in microns.

Table III summarizes the data obtained at Langley AFB for the analysis of HDI and HDI polyisocyanate. In a direct comparison of the filter (OSHA) and impinger (NIOSH) methods for HDI, both were equivalent at a particulate concentration of 15 mg/m³, but impinger sampling is superior when the particulate concentration exceeds 30 mg/m³.

Total particulates ranged from 15-46 mg/m³ for all Langley AFB operations. If the expected concentration of N-75 polyisocyanate is calculated based on the weight percent in the paint formulation, a theoretical concentration of HDI polyisocyanate can be determined (Theor). This value was ordinarily within a factor of 2 of the true value determined experimentally. In accordance with the suggestion in the AFOEHL Newsletter[14], the amount of polyisocyanate may be estimated by calculation based on the total particulates so long as the accuracy desired is only within a factor of two.

For two personal samples a midget cascade impactor was used for sampling. The total particulates were estimated and the mass median aerodynamic diameter (MMAD) in microns was calculated. The values determined 4.1 ± 1.7 and $6.9 \pm 2.5 \mu$ are well within the range previously observed in spray paint operations[15]. Further the particle size distribution is such that collection by an impinger should be reasonably efficient; impingers work well so long as the particle size is greater than 2μ [16].

In order to assess the general applicability of the modified NIOSH Method 5521 employed under the controlled sampling conditions at Langley AFB, we evaluated samples from other airbases.

In field trials involving orange poly operations at Edwards AFB, CA workers were exposed to a low level (0.16 mg/ m³) of N-75 polyisocyanate (TLV = 1.00 mg/ m³) and between 0.003-0.011 mg/ m³ of HDI (TLV = 0.034 mg/ m³). Impingers located in the vicinity of the F-16, but located away from the tail which was being painted showed non-detectible levels of both HDI and N-75.

In field trials involving white poly operations at Edwards AFB, CA, the results showed that the three workers were all exposed to low levels of HDI (between 2 and 6 μ g/m³) and variable levels of N-75 polyisocyanate. Worker #1 was exposed to 20-50 mg/m³ of polyisocyanate, while workers #2 and #3 were exposed to 150-210 mg/m³ and 330 mg/m³ of polyisocyanate respectively.

None of the workers at Edwards AFB were exposed to concentrations of HDI or N-75 which exceeded the threshold limit value of 1000 μ g/m³ (Moby).

TABLE IV. Travis AFB, HDI Analysis (7-13-92)

op	#	ID	ec/1000	uv/1000	ec/uv	ug/m3
1	337	exhaust duct		10.6		16.1
	338	recirc. duct		12.1		17.5
	339	personal	28728	197.0	146	269.8
	340	personal				0.0
2	373	personal				0.0
	374	personal	34542	36.8	939	42.2
	375	recirc. duct	42011	23.1	1819	29.9
	376	exhaust duct				
3	298	exhaust duct		6.5		8.5
	299	recirc. duct				0.0
	300	personal		13.7		20
	301	personal				0.0
4	453	field blank				0.0
	454	recirc. duct				0.0
	455	exhaust duct	10270	15.2	676	23.1
	456	personal				< LOQ
	457	personal		14.3		20.1
5	413	personal				0.0
	414	personal				0.0
	415	recirc. duct				0.0
	416	exhaust duct				0.0
	417	exhaust duct				0.0
op	Name	ID	ec/1000	uv/1000	ec/uv	ug/mL
	HDI	standard	4724	58.4	81	0.75
	HDI	standard	4367	54.5	80	0.75
	HDI	standard	4184	57.3	73	0.75
	HDI	standard	4237	42.7	99	0.75
	HDI	standard	4769	48.9	98	0.75
	HDI	standard	4459	65.9	68	0.75
	HDI	average	4457	54.5	83	
	HDI	std. dev	245	8.0	13	
	HDI	standard	1506	8.5		0.15

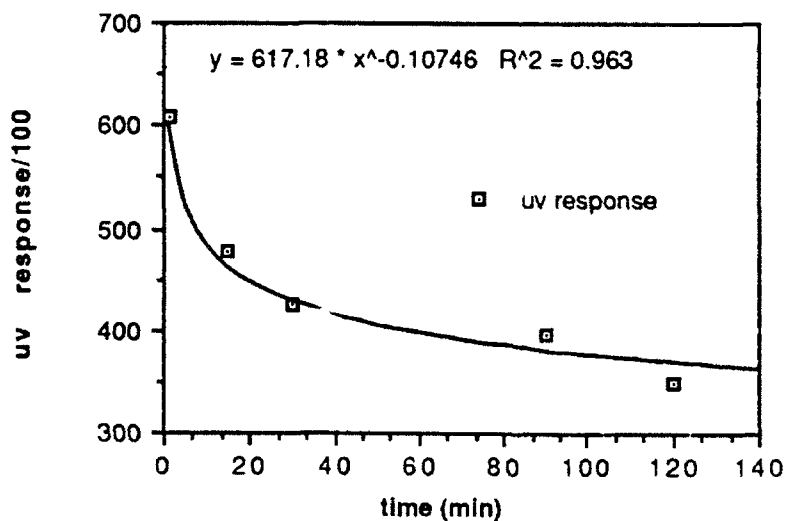
In field trials at Travis AFB, over a period of one week, five different spray paint operations were sampled. Samples collected at the exhaust and recirculation ducts in a spray paint booth showed detectable HDI monomer in 5 cases. Personal samples in 4 of 11 cases showed definite exposure to HDI monomer, and in 6 of 11 cases exposure to oligomer. Two personal samples showed HDI levels which exceeded the TLV value of 34 µg/m³. Oligomer peaks, although observable in the chromatograms, could not be quantitated since the type of hardener employed was not known. Based on chromatographic retention times, the hardeners may have been Des N-100 and/or Des N-3300.

In field trials at Wright-Patterson AFB, two different spray operations were sampled. On June 30, 1992, the collected sample showed 0.19 mg/m³ of HDI and 9.0 mg/m³ of polyisocyanate (both of these values exceed the threshold limit value by about a factor of 6-9); whereas, in a separate field trial exactly two weeks later, the sampling indicated that 0.002 mg/m³ of HDI and 0.3 mg/m³ of polyisocyanate were present. It is not clear why the second sample gave lower results, but factors such as the type of paint, the type and adjustment of the air spray gun or the type of spray paint operation may all have been factors.

Kinetic Analysis of the Reactivity of Des N-75

In order to assess whether the Des N-75 concentration observed might vary with time, a series of experiments were performed to evaluate the chromatographic response for HDI as a function of time. Hardener was mixed with pigment in a 1:1 ratio, and sample aliquots were removed at periodic intervals and injected into an impinger solution. Figure 7 shows the chromatographic response as a function of time.

Figure 7. N-75 Kinetic Response



A plot of chromatographic response as a function of the logarithm of time yielded a linear relationship, thus indicating a first order decrease in the amount of HDI. The half life was calculated to be 6 hours. The results indicate that the time interval between mixing and spray painting may be a key parameter which affects the ultimate amount of isocyanate found in air.

Conclusions

Based on the evaluation of standards in the laboratory, and field data from Langley AFB, as well as Edwards, Travis and Wright-Patterson AFB's, we have observed the following:

1. NIOSH Method 5521 for the analysis of HDI appears to give higher results and has a lower limit of detection when compared with OSHA Method 42.
2. NIOSH Method 5521 should be used for the analysis of polyisocyanates; however, the method should be modified so that prepolymers are used as standards instead of relying on the ec/uv ratio and the response based on a monomer standard. For Des N-75, the polyisocyanate had a lower ec/uv ratio than the monomer, and therefore polyisocyanate concentrations based on monomer standards would have been low.
3. When evaluating field data, polyisocyanate concentrations should be reported as well as monomer concentrations, even though there are no current, regulatory requirements for polyisocyanate. Field studies show that concentrations of polyisocyanate in excess of 1 mg/m³ are common and therefore may pose a health hazard to spray paint workers.

References

1. Melcher, R.G.: Industrial Hygiene. Anal. Chem. 55: 40R-56R (1983).
2. Purnell C.J. , R. Walker: Methods for the Determination of Atmospheric Organic Isocyanates. A Review. Analyst. 110 893 (1985).
3. Miles Corporate Industrial Hygiene Laboratory: Method 1.4.3 : Determination of Isocyanates in Spray Mist Environments by Sampling in an Impinger with N-4-nitrobenzyl-N-n-propylamine in Toluene and Analysis by High Performance Liquid Chromatography. in Methods for Sampling and Analysis of Airborne Isocyanates pp. 1-16, Miles Inc., Pittsburgh, PA, May (1992).
4. National Institute for Occupational Safety and Health: NIOSH Analytical Method #5521: Isocyanates. in Manual of Analytical Methods. 3d ed. Vol. 2. U.S. Department of Health and Human Services, Publication No. 84-100, Washington D.C. (1989).
5. Occupational Safety and Health Administration: OSHA Method 42: Diisocyanates. in OSHA Methods Manual pp 42-1:42-39 (1983).
6. Czarnecki, B.: Polyisocyanate Aerosol Sampling Using Coated-Filter vs. Impinger Collection Systems. Poster presented at the American Industrial

- Hygiene Conference. Boston, MA, U.S.A. June (1992).
7. Dharmarajan, V., P.M. Ohrin, D.R. Hackathorn, P.D. Ziegler: Evaluation of Air-purifying respirators in simulated isocyanate based paint spraying operation. Private Communication (1992).
 8. National Institute for Occupational Safety and Health: NIOSH Analytical Method #0500: Total Nuisance Dust. in NIOSH Manual of Analytical Methods. 3d ed. Vol. 2. pp. 1-3. U.S. Department of Health and Human Services, Publication No. 84-100, Washington D.C. (1989)
 9. Health and Safety Executive: Methods for the Determination of Hazardous Substances. Organic Isocyanates in Air. MDHS 25. HSE, London (1983).
 10. Bagon, D.A., C.J. Warwick, R.H. Brown: Evaluation of Total Isocyanate-in-Air Method Using 1-(2-Methoxyphenyl)piperazine and HPLC. Am. Ind. Hyg. Assoc. J. 45 : 39-43 (1984).
 11. Streicher, R.P., J.E. Arnold, C.V. Cooper, and T.J. Fischbach: Investigation of the Ability of MDHS Method 25 to Determine Urethane-Bound Isocyanate Groups. Paper 313 presented at the American Industrial Hygiene Conference. Boston, MA, U.S.A., June (1992).
 12. Tornling, G., R. Alexandersson, G. Hedenstierna, and N. Plato: Decreased Lung Function and Exposure to Diisocyanate (HDI) and (HDI-BT) in Car Repair Painters: Observations on Re-examination 6 Years after Initial Study. Am. J. Ind. Med. 17: 299-300 (1990).
 13. Janko, M., K. McCarthy, M. Fajer, J. van Raalte: Occupational Exposure to 1,6-Hexamethylene Diisocyanate-Based Polyisocyanates in the State of Oregon, 1980-1990. Am. Ind. Hyg. Assoc. J. 53(5) :331-338 (1992).
 14. Poitras, B.J., D. Carpenter. Sample Collection, Analysis and Respirator Use with Isocyanate Paints. AFOEHL Report 90-030CA00156BPA. pp 1-5. AF Occupational and Environmental Health Laboratory (AFSC) Human Systems Division, Brooks AFB, Texas 87235-5501, February (1990).
 15. D'Arcy, J. B. and Tai L. Chan: Chemical Distribution in High-Solids Paint Overspray Aerosols, Am. Ind. Hyg. Assoc. J. 51(3) :132-138 (1990).
 16. Dharmarajan, V., R.D. Lingg, K.S. Booth D. R. Hackathorn: Recent Developments in the Sampling and Analysis of Isocyanates in Air, in Sampling and Calibration for Atmospheric Measurements, ASTM STP 957, J.K. Taylor, Ed. ASTM, Philadelphia 190-202 (1987).

**EVALUATION OF ASTRONAUT PRACTICE SCHEDULES FOR
THE INTERNATIONAL MICROGRAVITY LABORATORY (IML-2)**

Robert E. Schlegel
Associate Professor

Randa L. Shehab
Graduate Student

School of Industrial Engineering
The University of Oklahoma
Norman, Oklahoma 73019

Final Report for:
Summer Research Program
Armstrong Laboratory
Crew Systems Directorate
Sustained Operations Branch
Brooks Air Force Base, Texas

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

EVALUATION OF ASTRONAUT PRACTICE SCHEDULES FOR THE INTERNATIONAL MICROGRAVITY LABORATORY (IML-2)

Robert E. Schlegel, Associate Professor
Randa L. Shehab, Graduate Student
School of Industrial Engineering
The University of Oklahoma
Norman, Oklahoma 73019

Abstract

The National Aeronautics and Space Administration (NASA) is currently conducting a series of space shuttle launches to enable scientists to study the effects of microgravity on a variety of factors. Included in the second International Microgravity Laboratory mission (IML-2) will be an extensive study of the effects of microgravity on astronaut cognitive performance ability. The Sustained Operations Branch of the USAF Armstrong Laboratory (AL/CFTO) has primary responsibility for this effort. This large collaborative study will include the training and testing of astronauts on a battery of human cognitive performance tests prior to launch, periodically during the space mission, and after the flight.

To permit an accurate identification of performance decrements caused by microgravity in space, it is essential to collect stable pre-flight baseline data. A preliminary investigation was conducted to determine the impact on baseline performance stability of less than optimal practice schedules and testing lapses due to such factors as launch delays.

A total of 21 subjects were trained on the NASA Performance Assessment Battery and then assigned to one of five practice schedules. Two groups practiced each day for 15 consecutive days. Two other groups followed a schedule of 5 days testing, 2 days off, 5 days testing, 3 days off, 5 days testing. The fifth group followed a schedule of 2 days testing, 5 days off, 2 days testing, 5 days off, 2 days testing. Then, either three days or five days after the last practice session, subjects returned for five days of retesting to represent mission days.

The study confirmed the overriding importance of providing an adequate number of practice sessions to achieve performance stability. By comparison, occasional missed sessions (i.e., the 5-on, 2-off schedules) had little impact on ultimate performance at the end of practice. The data indicated a possible performance difference between those subjects with only a 3-day gap between practice and "mission days" vs. those with a 5-day gap. High levels of differential stability and reliability were observed for at least one measure on all tests but the Critical Tracking test. Excellent software reliability was demonstrated by less than 0.02% missed data collection points.

Acknowledgments

The authors would like to thank the Air Force Office of Scientific Research, and Research and Development Laboratories for the opportunity to share a productive and enjoyable summer with the Sustained Operations Branch, Crew Systems Directorate of Armstrong Laboratory at Brooks AFB, Texas. The authors especially wish to express their great appreciation to all members of the branch, including those associated with NTI, Inc., Systems Research Laboratories, Inc. (SRL), and KRUG Life Sciences, all of whom provided substantial support and an excellent working atmosphere.

In terms of specific individuals, many thanks are due Dr. Samuel G. Schiflett (AL/CFTO) and Dr. Douglas R. Eddy (NTI), the Principal Investigator and Associate Investigator, respectively, for the overall program on "Microgravity Effects on Standardized Cognitive Performance Measures". Both individuals provided their full technical and personal support for the research. The scientific discussions were highly enlightening and enjoyable. Special thanks are extended to Dr. William F. Storm for his friendship and genuine interest in the project.

The authors wish to acknowledge the skilled and timely programming contributions of Kathy M. Winter, Sam J. LaCour, Kathy Raynsford, and Sam "Major" Moise. At several points during the research period, these individuals worked under tight deadlines to provide software changes needed to conduct the study.

Last, but certainly not least, enormous thanks are offered to those who really made the study possible - the experimental subjects. Their contribution of time and steadfast adherence to testing schedules (including weekends) provided the valuable data that enabled the project to be successful. Their involvement also made it easier for the summer researchers to become acquainted with the many people that comprise the Sustained Operations Branch at Brooks AFB.

EVALUATION OF ASTRONAUT PRACTICE SCHEDULES FOR THE INTERNATIONAL MICROGRAVITY LABORATORY (IML-2)

Robert E. Schlegel
Randa L. Shehab

INTRODUCTION

The United States National Aeronautics and Space Administration (NASA) is currently conducting a special series of space shuttle launches to enable scientists to study the effects of microgravity (weightlessness or near weightlessness) on a variety of factors. These missions have been named International Microgravity Laboratory (IML) spacelab missions. Included in the second IML mission (IML-2) with a scheduled launch date of July 1994 will be an extensive study of the effects of microgravity on astronaut cognitive performance ability. The Sustained Operations Branch of the USAF Armstrong Laboratory (AL/CFTO) has primary responsibility for this effort. This large collaborative study will include the training and testing of astronauts on a battery of human performance tests prior to launch, periodically during the space mission, and after the flight. In preparation for this study, considerable research must be conducted to establish the exact nature of the training and testing protocol and the degree to which the testing protocol will be robust to such factors as launch delays, disruptions in the testing schedule, and various test parameters.

To permit an accurate identification of performance decrements caused by microgravity in space, it is essential to collect stable pre-flight baseline data. However, the astronaut's pre-flight schedule is extremely demanding. Only a limited amount of time is available for pre-flight training on any one project. Thus, there is a major need to identify the optimal practice schedule and test parameters and measures in order to achieve stable baseline performance data prior to the IML-2 spaceflight mission. It is also necessary to determine the impact of less than optimal practice schedules on the stability of the baseline performance data. Another objective of this study is to determine the influence of testing lapses (due to such factors as launch delays) on baseline stability. It is essential to understand the influence of such delays on baseline performance so as to determine when and how much additional practice may be needed after such delays to re-establish adequate baseline performance. A preliminary examination of these issues was conducted during Summer 1992 under the AFOSR Faculty and Graduate Student Summer Research Programs. This pilot study, which involved 21 male and female subjects with ages ranging from 23 to 52 years, provided a tentative answer to the

question of optimal practice schedules and the impact of launch delays. This report presents the background and methodology for the study along with preliminary conclusions based on the data analyses completed at the time of report writing.

METHODOLOGY

In summary, this AFOSR summer study served as a precursor to a larger research project, the objectives of which are to:

- (1) determine the optimal practice schedule for acquiring stable performance on a number of basic human performance tests,
- (2) determine whether time lapses in practice on these tests affect the retention of performance capability,
- (3) evaluate, improve, and verify the reliability of the testing software and the stability of the test measures,
- (4) evaluate and establish task parameter levels for those tests not achieving stability, and
- (5) provide a normative database for classifying astronaut performance.

Test Battery

Several factors were considered in selecting tests for the NASA IML-2 experiment. One of the most important of these is the restrictive time available during flight for performance assessment. Another critical factor is the specific information processing skills necessary for mission success. The final and most relevant issue is the information provided by a specific test that could aid in identifying the cognitive processes or information processing stages affected by microgravity. These and other factors were taken into account in reviewing a large number of human performance task batteries. As a result, six performance tests and two subjective scales from these various batteries were selected for inclusion in the NASA Performance Assessment Battery (NASA-PAB).

Brief descriptions of the performance tests and subjective rating scales selected for the IML-2 experiment follow in the order in which they are presented in the battery:

Mood Scale II - consists of 36 questions and involves pressing a numbered key to indicate the level of agreement with a descriptive adjective. The test takes 1 minute.

Critical Tracking - involves tracking an unstable object on the display using a trackball for 2 minutes.

Spatial Matrix - involves indicating whether a matrix of squares is the same as one previously presented. The test lasts 1.5 minutes.

Sternberg Memory Search - involves indicating whether a letter is the same as one of those in a previously memorized set. The test lasts 2 minutes.

Continuous Recognition Memory - involves pressing a key to indicate whether a number is the same as one previously memorized. The test lasts 2 minutes.

Switching Task - involves responding to 1 of 2 tasks presented simultaneously on each screen display. In the Manikin task, the subject presses a key to indicate which hand of a manikin holds a matching symbol. In the Mathematical Processing task, the subject presses a key to indicate whether a sum of three numbers is greater or less than 5. The test lasts 4 minutes.

Dual Task - involves performing the Sternberg Memory Search while Tracking and lasts 3 minutes.

Fatigue Scale - involves pressing a key to indicate which statement best matches the subject's fatigue state and takes less than 15 seconds.

More detailed descriptions of the various tasks are available from the publications listed in the References section.

Equipment

The NASA-PAB is presented on a small portable microcomputer called the Performance Assessment Workstation (PAWS). The PAWS consists of a GRiD 1530 Computer and MSI Trackball. The GRiD 1530 Computer is a laptop portable computer that is IBM AT compatible. It uses GRiD MS-DOS (Microsoft Disk Operating System) and has a built-in flat screen EL display, a built-in keyboard, a 32-bit 80C386 central processor, 4 Mb of RAM, an 80887 math coprocessor, an internal 30 Mb hard drive, an internal 1.44 Mb floppy drive, and an RS-232 interface along with other features. It has a special ROM BIOS chip to allow MS-DOS to support the EL display screen.

Subjects

A total of 22 volunteer subjects, recruited from branch personnel available during the summer, were trained on the PAWS. One subject was unavailable for further testing after the training period. Of the 21 remaining subjects, there were 6 women and 15 men. Subject age ranged from 23 to 52 with a mean of 31.5 and a standard deviation of 10.2. Subjects were screened for self-reported normal (or corrected-to-normal) vision, normal hearing, and the absence of any central nervous system stimulant or depressant medication. Approval for the use of human subjects was obtained from the Advisory Committee on Human Experimentation (ACHE).

PROCEDURES

The overwhelming majority of testing sessions were conducted at subject workstations located in Room 24E of Bldg. 170 at Brooks AFB. Weekend testing sessions were conducted at subjects' residences. Data were collected from each subject over a period of two months. An orientation session provided subjects with background information about the project and instructions on the individual tasks. All subjects completed four one-hour training sessions over a two-day period (one each morning and afternoon). All subsequent practice and test sessions lasted 20 minutes per day.

After the initial training routine, subjects were assigned to one of five practice groups, distinguished by practice schedule and length of delay prior to "mission" testing. Two groups (A and B) practiced each day for 15 consecutive days. Two other groups (C and D) followed a schedule of five days testing, two days off, five days testing, three days off, five days testing. The fifth group (E) followed a schedule of two days testing, five days off, two days testing, five days off, two days testing. Then, either three days (Groups A and C) or five days (Groups B, D, and E) after the last practice session, subjects returned for five days of retesting used to represent mission days. Total testing time per subject for the entire study ranged from 12 to 14 hours.

In each session, subjects performed the six tests and completed the two subjective scales using the PAWS for a total of 20 minutes. The tasks involved viewing a computer display screen, responding by pressing keys on the keyboard, and moving a trackball. The PAWS has been designed to allow a subject to perform the tests independently without

experimenter assistance. It is automated to minimize the time required for a well-practiced subject to perform the tests. It automatically performs all housekeeping functions, such as subject identification, file naming, test sequencing, and data backup. This feature was important for testing over weekends where subjects administered the tests themselves without experimenter supervision. This approach was very effective and demonstrates the reliability and ease of use of the software.

In the initial training sessions, the interval between tests was subject-determined, that is, the tests did not start automatically. Subjects needed to press a key to start the next task. This allowed an opportunity for the subjects to ask questions and receive feedback. Summary feedback was provided at the end of each test during all sessions.

RESULTS

The data were summarized to examine practice group differences in performance means, and analyzed with respect to differential stability and reliability. Due to the large number of performance measures, graphs for a sample of tasks are presented below to illustrate typical patterns of performance during training, practice, and baseline. Figures 1 and 2 present the training performance, averaged across all 22 subjects for the Mathematical Processing component of the Switching Task and for the Unstable Tracking component of the Dual Task. The task learning illustrated by the faster response times and higher accuracy, and the decrease in control losses over time is representative of the performance trends during training. In general, performance improved rapidly over the first three to five trials. The rate of improvement leveled off by the eighth trial.

Figures 3 and 4 summarize the practice performance by Practice Group for the Critical Tracking lambda (index of task difficulty) and Continuous Recognition throughput measures. These graphs clearly illustrate that average performance continues to improve throughout the fifteen practice trials, with the rate of improvement dependent on the particular task and measure. It is also apparent that the rate of improvement was approximately the same for each of the five groups. In general terms, there was little to distinguish Groups A, B, C, and D from each other. All four groups had the same number of practice trials (15), although groups C and D had occasional two-day or three-day lapses. On the other hand, Group E subjects demonstrated a similar rate of improvement on the tasks, but could not, in general, achieve the same level of performance by the end of their relatively short six-session practice period.

Switching - Math Processing

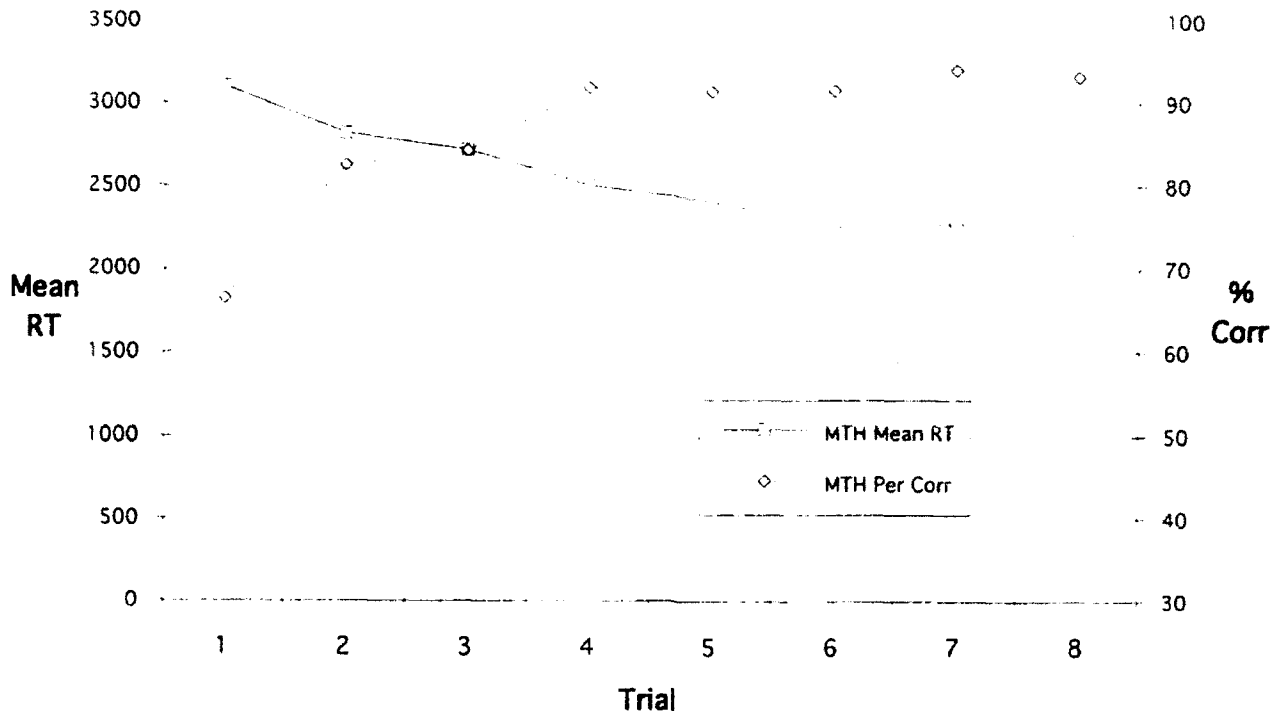


Figure 1. Mean RT and Accuracy for Mathematical Processing Training.

Dual - Tracking

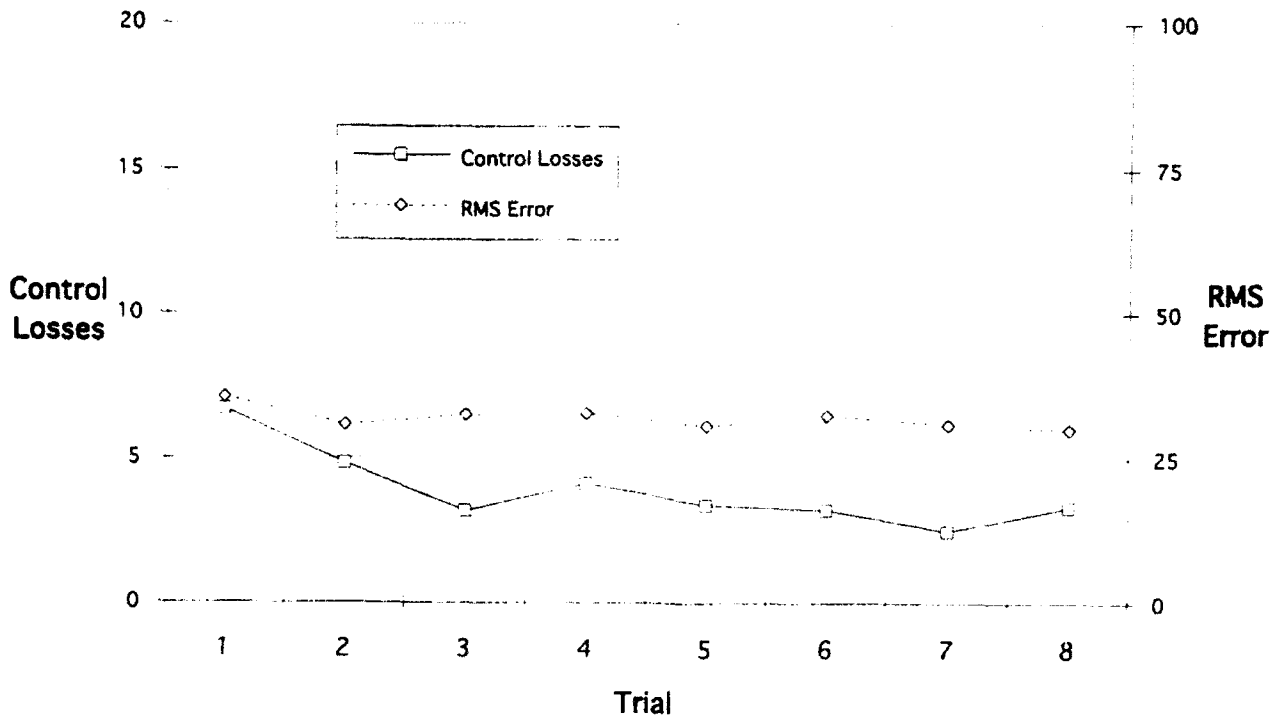


Figure 2. Mean Control Losses and RMS Error for Dual Tracking Training.

Critical Tracking

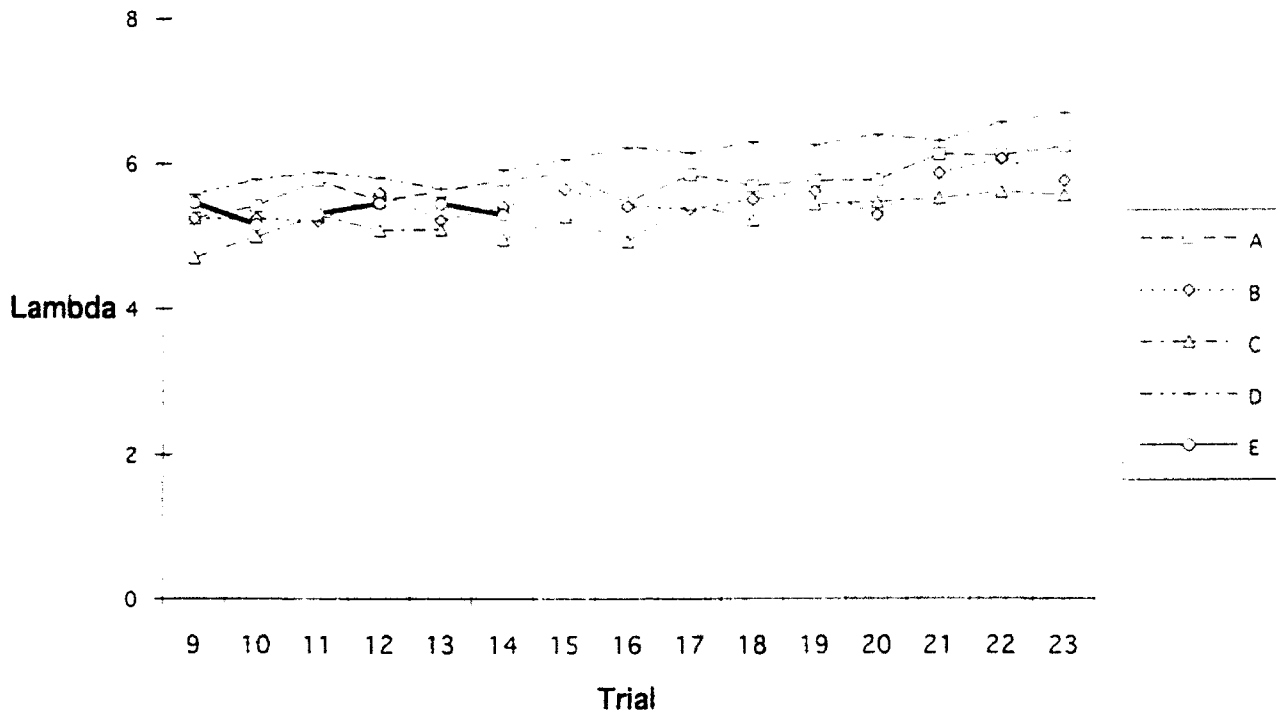


Figure 3. Mean Lambda Score for Critical Tracking Practice.

Continuous Recognition

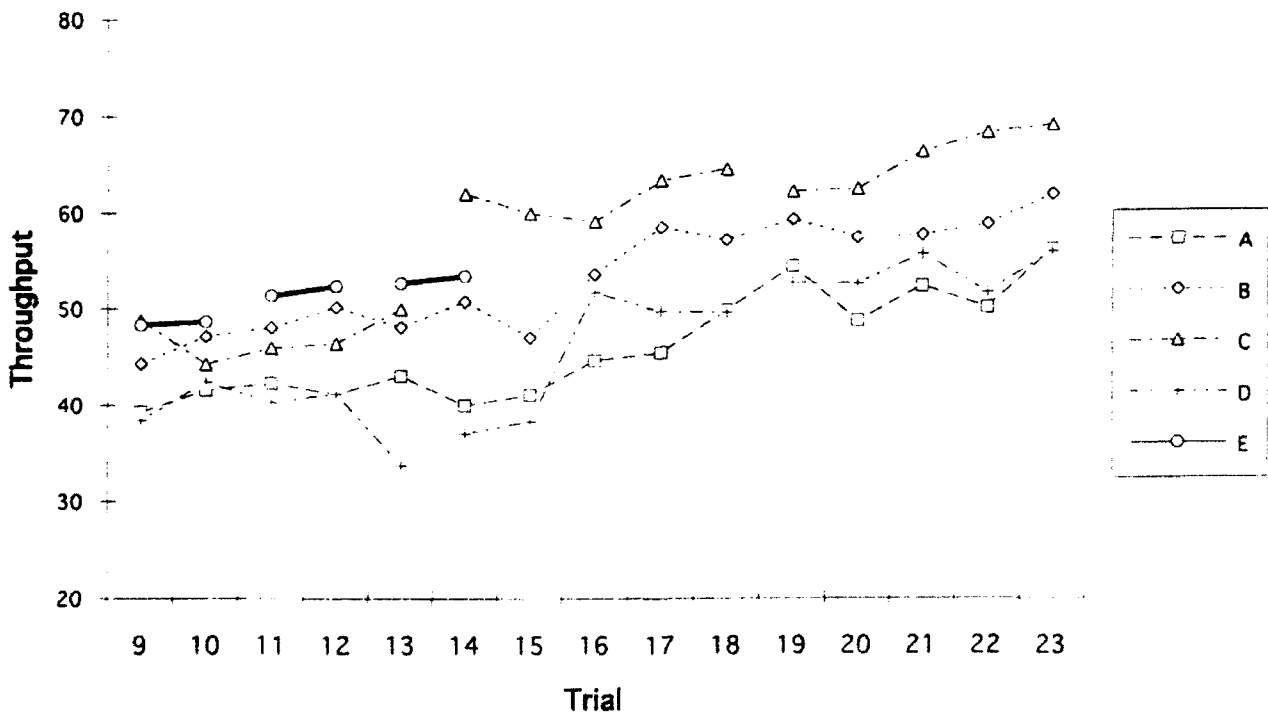


Figure 4. Mean Throughput for Continuous Recognition Practice.

Figures 5 through 8 present the data for four of the task measures and allow a comparison of how the five subject groups performed at the end of Training, the start and end of Practice, and during the five analogous Mission Days (MD-1 through MD-5). The represented tasks include Critical Tracking, the Matrix Task, Memory Search, and the Manikin Task component of the Switching Task. Data for most of the other task measures are similar. In general, the pattern of results may be described as follows.

In terms of absolute performance, there were differences among the group means that were dependent upon the particular task and measure. These differences first became apparent during training, continued throughout the practice trials, and persisted during the "mission" testing. However, the performance improvement from the start to the end of practice was approximately the same for the four groups that completed 15 trials of practice (Groups A, B, C, and D). As mentioned previously, the improvement for Group E was much lower over the same period of time due to the limited number of practice sessions (only 6), although the rate of improvement appeared to be equal to the overall rate for the other groups. The limited number of practice sessions placed the Group E subjects at a distinct disadvantage throughout the practice period and during mission testing, so much so that subjects in this group continued to improve on the tasks at a fairly high rate throughout the final five test sessions. In summary, subjects in this group did not achieve a stable level of performance that would allow an accurate comparison with mission day performance. The mission testing data would lead one to the erroneous conclusion that performance improved as a function of the duration of the space flight. Although this pattern did not hold for all task measures, often due to ceiling effects of the measures themselves, it occurred with a sufficient number of tasks to warn against the use of such little data to establish a performance baseline prior to launch.

As an example, performance of the Group E subjects on the Manikin Task (Figure 8) placed this group in the middle compared with the other groups at the end of Training and the beginning of Practice. By the end of Practice, Group E had the longest mean RT. Throughout the last five mission sessions, the Group E mean RT continued to improve while the mean RT values for Groups A through D were essentially constant across sessions. For many other tasks, Group E lost its intermediate performance ranking and fell to last place as a result of insufficient practice.

In contrast, there was little to distinguish the other four groups, either during Practice or Mission testing. The two-day and three-day breaks taken by Groups C and D (indicated by

Critical Tracking

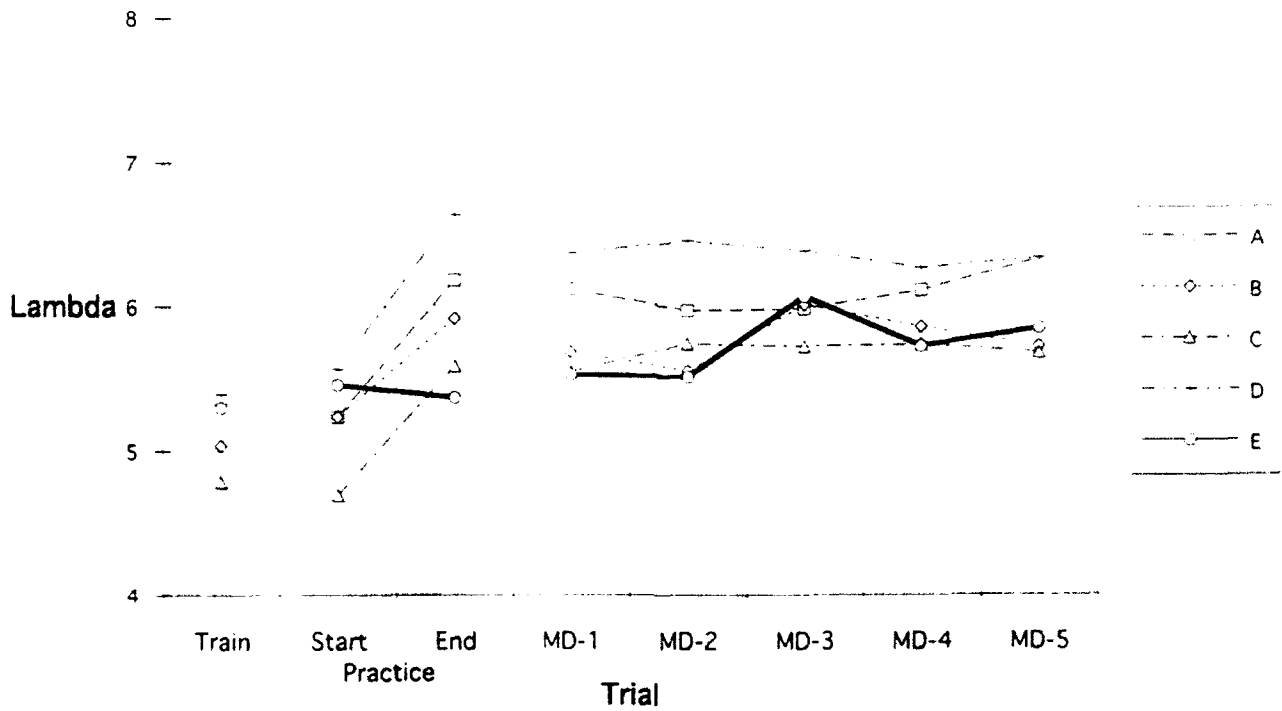


Figure 5. Mean Lambda Score for Critical Tracking Mission Days (MD).

Matrix

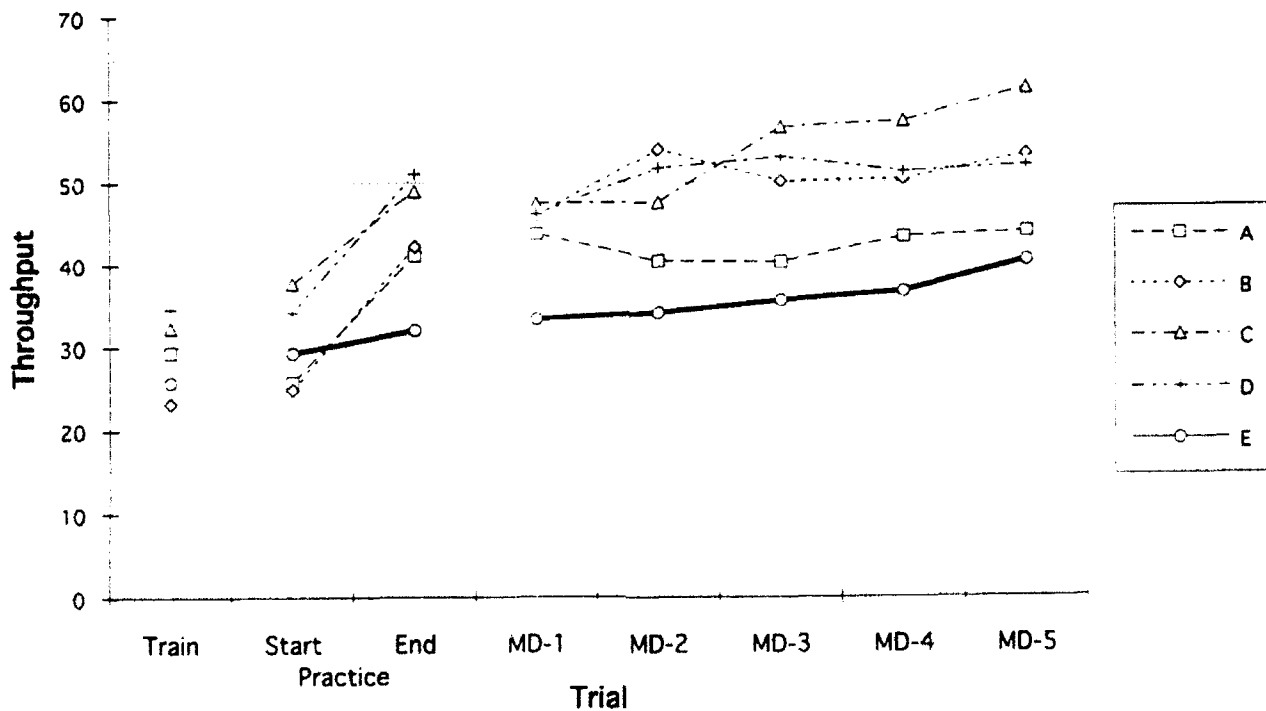


Figure 6. Mean Throughput for Matrix Task Mission Days (MD).

Memory Search

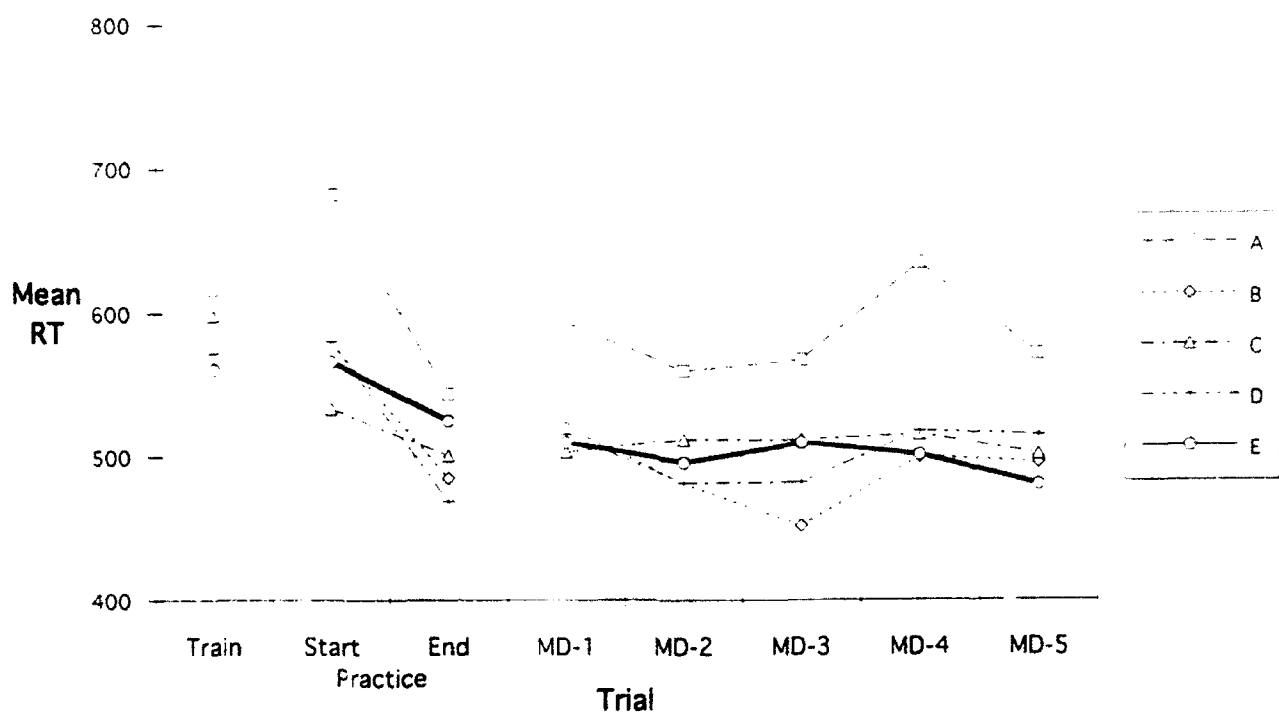


Figure 7. Mean Response Time for Memory Search Mission Days (MD).

Switching - Manikin Task

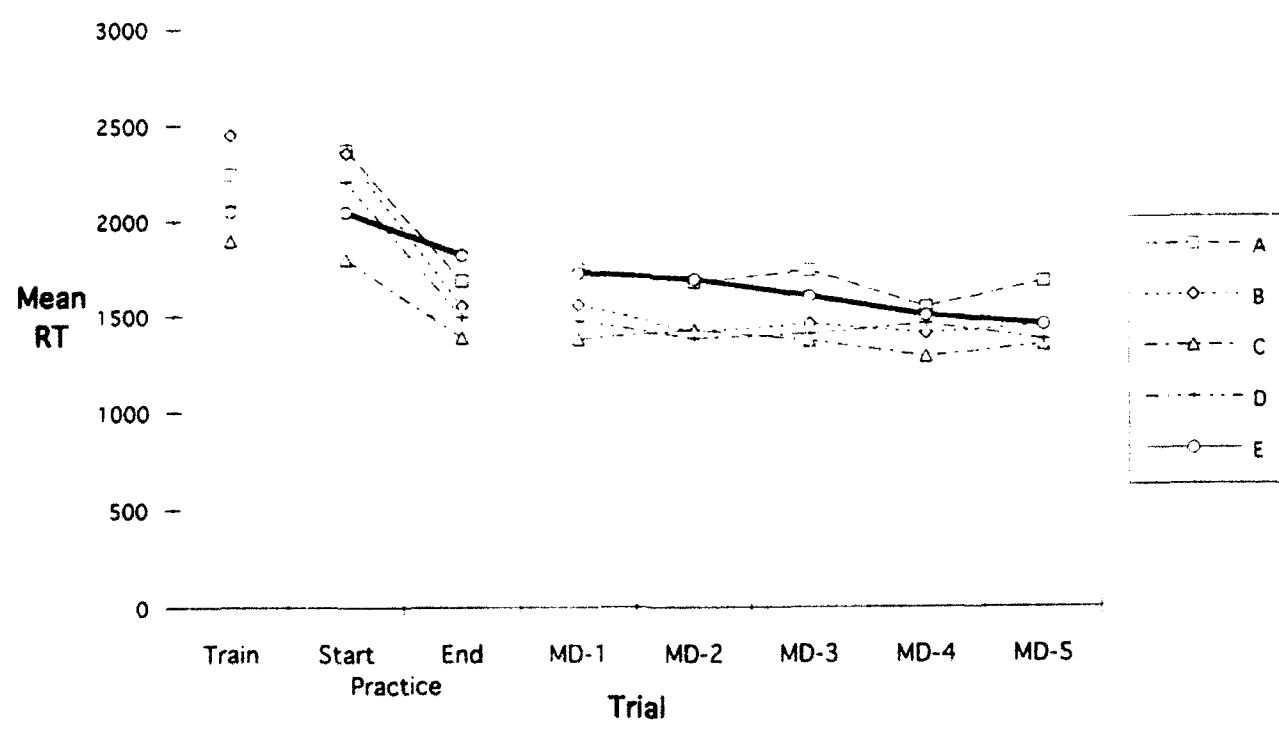


Figure 8. Mean RT for Switching - Manikin Task Mission Days (MD).

the break between Trials 13 and 14 and between Trials 18 and 19 in Figures 3 and 4) did not appear to interrupt the rate of learning or current level of performance for most tasks. In other words, there was little performance regression following the breaks.

Data comparing a three-day vs. a five-day gap between practice/baseline performance and mission testing were inconclusive but suggest that a decline in performance may occur for some tasks following a five-day gap. This performance drop, if it exists, disappears by the second trial following the gap.

Intertrial correlations and Lord and Novick (1968) reliabilities computed for the last five training trials and the five mission testing sessions are presented in Table 1. With the exception of the Critical Tracking task, at least one measure from each task was able to provide an average intertrial correlation above 0.82, with many tasks yielding values in excess of 0.95. Lord and Novick (1968) reliabilities, computed as the ratio of (Between-subject Variability minus Within-subject Variability) divided by Between-subject Variability, were of similar magnitude and paralleled the intertrial correlations.

CONCLUSIONS

The major conclusions of the study can be summarized as follows:

- (1) The total number of practice sessions is critical in terms of providing subjects with ample opportunity to improve and to achieve a stable level of performance. For the majority of tasks, subjects in Group E (2-on, 5-off) demonstrated poorer performance during the "mission" days than the other groups. On many of these tasks, these subjects were performing at or near the top at the start of the practice period. However, the six practice sessions spread over sixteen days resulted in substantially less improvement by the end of the practice period compared with the other groups. More importantly, Group E subjects continued to improve their performance during the actual "mission" days, thus demonstrating their lack of performance stability.
- (2) The distribution of practice sessions appears to be less important than the total number, provided the sessions are distributed somewhat evenly across the practice period, and a continuous series of sessions is obtained immediately prior to the pre-launch blackout. There were no apparent differences in the end-of-practice performance levels between those subjects who practiced 15 days straight and those who were on a 5-on, 2-off schedule.

Table 1. Intertrial Correlations and Lord & Novick Reliabilities

Task -- Measure	Training		Mission Days	
	Correlation	Reliability	Correlation	Reliability
Tracking -- Lambda	0.69	0.66	0.69	0.70
Tracking -- RMS Error	0.88	0.87	0.78	0.77
Matrix -- Mean RT	0.93	0.90	0.97	0.95
Matrix -- Percent Correct	0.28	0.25	0.68	0.64
Matrix -- Throughput	0.92	0.85	0.96	0.94
Memory -- Mean RT	0.77	0.65	0.84	0.79
Memory -- Percent Correct	0.41	0.19	0.53	0.42
Memory -- Throughput	0.70	0.59	0.79	0.74
Con. Rec. -- Mean RT	0.91	0.80	0.91	0.90
Con. Rec. -- Percent Correct	0.80	0.80	0.85	0.85
Con. Rec. -- Throughput	0.90	0.86	0.94	0.93
Manikin -- Mean RT	0.82	0.71	0.90	0.86
Manikin -- Percent Correct	0.84	0.46	0.56	0.55
Math -- Mean RT	0.90	0.79	0.93	0.92
Math -- Percent Correct	0.60	0.54	0.65	0.60
Dual -- Control Losses	0.93	0.91	0.69	0.63
Dual -- RMS Error	0.90	0.91	0.82	0.82
Dual MS -- Mean RT	0.74	0.73	0.91	0.91
Dual MS -- Percent Correct	0.18	0.18	0.52	0.44
Dual MS -- Throughput	0.79	0.74	0.88	0.89

- (3) Although not demonstrated statistically at this time, a possible performance difference on the first "mission day" existed between those subjects with only a three-day gap between practice and mission days and those with a five-day gap. On a number of tasks, the observed decline in performance between end of practice and mission day 1 (MD-1) was either lower or non-existent for the three-day groups compared with the five-day groups.
- (4) Very good to excellent levels of differential stability and reliability (typically on the order of 0.90 or above) were obtained for at least one measure on all of the performance tests

except the Critical Tracking test (0.69 for the Lambda measure and 0.82 for RMS Error). This indicates a need to examine this test to identify alternative scoring metrics. For the remaining tests, this provides solid justification for their inclusion in the NASA-PAB.

- (5) Excellent software reliability was demonstrated by fewer than 10 missed tests out of 4344 tests administered during 543 sessions.

Recommendations for Further Research

Based on the results of this study, the following suggestions for the next phase of this research are provided:

- (1) Additional data based on a substantially larger subject sample are required to confirm the tentative results of this preliminary study.
- (2) Additional investigation of performance declines following a simulated launch delay is warranted.
- (3) It may be of importance to examine other alternative practice schedules such as 3-on, 2-off, 3-on, 2-off, 5-on or an intensified schedule of two practice sessions per day over seven days.
- (4) An alternative scoring metric for the Critical Tracking task (such as a mean or median lambda value for each session) should be explored.
- (5) The influence on performance of the actual stimulus sequence which is based on the session number) should be examined.
- (6) A large normative database must be provided to allow comparison of the performance ranges of the specific Mission Specialist astronauts to a larger population.

References

- Benson, A.J. and Gedye, J.L. (1963). Logical processes in the resolution of orienting conflict. *RAF Rpt. 259*, Farnborough, UK, Royal Air Force Institute of Aviation Medicine.
- Damos, D.L. and I.yall, E.A. (1984). The effect of asymmetric transfer on dual-task assessment of voice technology. *Proceedings of the 28th Annual Meeting of the Human Factors Society*, Santa Monica, CA: Human Factors Society.
- Damos, D.L. and Wickens, C.D. (1980). The identification and transfer of timesharing skills. *Acta Psychologica*, **46**, 15-39.
- Damos, D.L. (1991). *Multiple Task Performance*. London, UK, Taylor & Francis, LTD.

- Hunter, D.R. (1975). Development of an enlisted psychomotor and perceptual test battery. *AFHRL-TR-75-60*, Brooks AFB, TX: Air Force Human Resources Laboratory.
- Lord, F.M. and Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McRuer, D.T. and Jex, H.R. (1967). A review of quasi-linear pilot models. *IEEE Transactions on Human Factors in Electronics*, **8**, 231-249.
- Miller, J.C., Takamoto, G.M., Bartel, G.M., and Brown, M.D. (1985). Psychophysiological correlates of long-term attention to complex tasks. *Behavior Research Methods, Instruments, and Computers*, **17**(2), 186-190.
- Nesthus, T.E., Schiflett, S.G., Eddy, D.R., and Whitmore, J.N. (1991). Comparative effects of antihistamines on aircrew performance of simple and complex tasks under sustained operations. *AL-TR-91-104*, Brooks AFB, TX: Armstrong Laboratory, Crew Technology Division.
- O'Donnell, R.D. (1991). *Scientific validation of the Novascan (tm) tests: Theoretical basis and initial validation studies*. NTI Report to Nova Technology, Inc., 19460 Shenango Drive, Tarzana, CA: NTI, Incorporated.
- O'Donnell, R.D. and Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff et al. (eds), *Handbook of Perception and Human Performance*, Vol II, New York: Wiley, 42-1 to 42-49.
- Perez, W.A., Masline, P.J., Ramsey, F.R., and Urban, K.E., (1987). Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology. *AAMRL-TR-87*, Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.
- Schlegel, R.E. and Storm, W.F. (1983). Speed-accuracy tradeoffs in spatial orientation information processing. *Proceedings of the 27th Annual Meeting of the Human Factors Society*, Santa Monica, CA: Human Factors Society.
- Reader, D.C., Benel, R.A., and Rahe, A.J. (1981). Evaluation of a manikin psychomotor task. *USAFSAM-TR-81-10*, Brooks AFB, TX: USAF School of Aerospace Medicine.
- Shingledecker, C.A. (1984). A task battery for applied human performance assessment research. *AFAMRL-TR-84-071*, Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratories.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.

SEM-EDXA ANALYSES OF AIRBORNE INORGANIC FIBERS
FOR QUALITATIVE IDENTIFICATION

LARRY R. SHERMAN
University of Scranton
Scranton, Pennsylvania 18510-4286

ABSTRACT

The health hazard posed by airborne asbestos fibers is well documented, as well as numerous analytical methods for the identification and characterization of the fibers. Analytical methods for other lung abatement or innocuous fibers are not as well defined and this project was initiated to determine if the deficiency could be overcome using equipment available in the Occupational Environmental Analyses section of the Armstrong Laboratory.

The author and his graduate student used a number of fiber standards to establish a library for fiber identification using morphology and the elemental X-ray analyses obtained with the Amray 1820 Scanning Electron Microscope and Tracor Northern X-ray Analyzer. From the elemental analyses a series of linear equations were developed which are unique for the qualitative and quantitative analysis of each fiber type. These equations along with the average elemental fiber composition were put into computer programs for analyzing air fibers submitted to the AL-OEAO. The success rate for analyzing fibers on standard cellulose acetate filters was approximately 88% and would have been greater if a larger library were available.

INTRODUCTION

Although asbestos is currently believed to pose the greatest microscopic fiber health hazard to workers in either the fabricating or users' industries, other inorganic fibers may also cause lung problems (1). When exposed to high temperature, man-made refractory fibers (RCF) can be converted into cristobalite, a crystalline silica, which is listed as a possible carcinogen by the Internal Agency for Research on Cancer (2) and OSHA has established a permissible exposure limit (PEL) of 0.05 mg/m³ for those fibers with a length to aspect ratio of 3:1. Workers are exposed at this level if they wear no breathing protection equipment when removing and repairing certain high temperature equipment. The health hazards posed by other inorganic fibers, like fiber glass or ceramics, is much less known and methods for their identification and characterization in lung abatement has not been well studied (3).

Although there are no current PEL for most inorganic fibers, many industrial hygienists have adopted the OSHA PEL for asbestos, i.e. 0.2 f/cc for an eight hour weighed average and 0.01 f/cc for a 30 minute exposure, and recommend monitoring all inorganic fibers until further health hazard data is available (4).

The Air Force normal fiber identification procedure uses Phase Contrast Microscopy, NIOSH Method 7400, but this procedure can not distinguish non-asbestos inorganic fibers from asbestos fibers. One of the major problems has been separating acicular minerals, fibrous minerals and asbestiform minerals. Furthermore, the NIOSH method is unable to distinguish organic from inorganic

fibers, thus many high fiber count samples actually pose no health hazard because the fibers are not lung abatement materials. Furthermore, on the sub-micrometer scale, cleavage fragments can appear as fibrous material which add further ambiguity when TEM is used. With TEM there is more difficulty in this type of separation than SEM. On the other hand, assuming that all fibers are inorganic or even upgrading the fiber count to asbestos fibers can be an uneconomical evaluation and require much more effort than justified if a complete fiber identification were detailed. Nevertheless, the cost is too high for analyzing every fiber; the current project was designed to develop a scanning electron microscopy-energy dispersed X-ray analysis method for rapidly analyzing a representative fraction of the fibers in air samples to help evaluate the environment where the sampling occurred.

METHODOLOGY

SEM-EDXA

Standard inorganic fiber samples were mounted on carbon coated aluminum SEM studs and overlaid with a Au/Pd coating using the Anatech Ltd. Hummer VI as previously described in the literature (5). The samples were placed in the vacuum chamber of the Amray 1820 scanning electron microscope (SEM) and searched for suitable fibers. The selected fibers were photographed at a magnification which illustrated their morphology. Morphology is extremely critical, e.g. chrysotile is a serpentine fiber consisting of sheets where the SiO_4 tetrahedral all point in the same direction; other serpentine fiber may have the same chemical composition as chrysotile but do not have asbestos morphology

(6). A characteristic portion of the fiber was analyzed using the Energy Dispersed X-ray (EDXA) emitted by the impinging electrons on the surface of the fiber. The photons were analyzed using a Tracor Northern X-ray Analyzer (TN) and corrected for absorption, interference and atomic number. Using "canned" programs supplied by TN, the TN computer calculated the raw data as either atomic percentage or weight percentage. Weight percentage was easiest to handle in this study. A minimum of five fibers were analyzed for statistical purposes. All methodology has been previously published (5,7,8) (see appendix A for an example of the data sheet produced from this experimental work) (9).

A 4 mm triangle was cut from each unknown cellulose acetate air sample submitted to the Armstrong Laboratory. The 4 mm sample was mounted on an aluminum stud covered with wet carbon paint. When dry, the sample was coated with a 11 nm Au/Pd coating (5,7,8), then placed in the SEM vacuum chamber. Suitable fibers, those with correct morphology, were analyzed for elemental composition. Since there was no preconceived knowledge of the fiber composition, eleven elements were assayed (Si, Mg, Fe, Ca, Mn, Na, S, P, Cl, K and Al). The X-ray assay data was transferred to a Basic computer program on the Zenith 100 microcomputer for data reduction.

STANDARDS

All seven types of asbestos and twenty non-asbestos man-made and natural fibers were analyzed on the SEM (see Table 1). Sixty other mineral samples have been prepared but have not yet been analyzed. The EDXA data was processed to determine the

average % composition and standard deviation and to establish a 2,3,4, or 5 element linear equation which mathematically describes the sample (MATRIX.BAS program). The average % composition, standard deviation and the coefficient for the linear equation along with representative factors constitute the library for determining the unknown fibers.

MATRIX.BAS

The MATRIX.BAS is a computer program, which accepts X-ray data from the standard samples. It is used for deriving a linear equation along with the average composition and standard deviation for describing the fiber's chemical characteristics. Up to five representative elements can be entered into the matrix. Although zero is a valid number for the matrix solution, it should be avoided unless the standard deviation for the average composition overlaps zero. The computer program initially computes the mean and standard deviation but allows substitution of the mean for a value which may appear to skew the values in a positive or negative manner. The determinate is diagonalized and the coefficients for the equation are printed. If the coefficients are greater than +10, especially for minor constituents, they can severely interfere with the ultimate fiber selection process. Table 1 contains the fibers analyzed along with their average composition, standard deviations and the coefficient for each element found in the library.

ASBTOS.DAT and FIBER.DAT

These programs construct the fiber libraries for asbestos and non-asbestos fibers. Both programs are basically the same for constructing DATA.ASB and DATA.FIB data files for analyzing

unknown fibers. The fiber name, representative elements, linear equation coefficients, average composition, standard deviation and representative percentages (factors) are put into a sequential file for use in the IDEN.BAS or ANAL.1 programs

Although both programs are designed to accept input from the keyboard to expand the files, they currently have all processed data stored in a data file starting on line 5000.

IDEN.BAS

IDEN.BAS is a computer program for searching all fibers (both asbestos and non asbestos) in the library and matching the X-ray data with the library data. No prior editing is required of the data. IDEN.BAS compares the elemental data for each fiber in the library with the normalized elemental values from the unknown assay. When a match occurs, the similarity and variance is printed on the line printer. After checking each element a linear equation is constructed from the matched elements, if the product of the unknown equation and product of the standard equation in the library are within $\pm 20\%$, the equation is printed on the line printer.

Fortuitous results are not deleted in the data processing. The final fiber selection must be made by the analyst. Better selection is possible using the ANAL.1 program.

ASBT1.BAS

When a fiber appears to be an asbestos fiber, the Si, Mg, Fe, Ca, Mn and Na composition can be entered into the ASBT1.BAS program and the fiber delineated as to asbestos type. The use of this fiber analysis method has previously been published (5,7).

Documentation and print out of the program is available from Brooks AFB or the P.I. The data analysis program calculates the Mg/Si, Fe/Si, Ca/Si Mn/Si and Na/Si ratios for the fiber; from the ratios, it attempts to determine the fiber, then computes the cation/anion factor for further identification. If a fiber with the correct morphology has been selected, a high degree of reliability can be given to the printout.

A report printing sequence has been added to the basic program. It is used when all the fibers are believed to be asbestos. The printer sequence has no provision for printing non-asbestos fibers, these should be processed with the REPORT.ASB program. The reports are prepared to assist the air base environmental officer in making decisions for handling high fiber counts.

ANAL.1

ANAL.1 program performs the functions as IDEN.FIB except that it further searches the fiber data to eliminate obvious errors, i.e. fibers which are fortuitous identified because some major element was eliminated during the construction of the linear equation. This program can delete fibers which will appear with the IDEN.FIB program.

REPORT.ASB

The program is used to prepare a report which can be sent to a base (see appendix C).

RESULTS.

On July 30, 1992, OEAO began to use the analytical method developed in this report to prepare sample analyses sheets for the air bases submitting air samples which exceeded the action

limit. Table 2 presents statistics concerning the reports mailed as of August 20, 1992. Appendix B illustrates a sample report and Appendix C a copy of the survey letter which was sent along with the sample report. As of August 20, one survey sheet has been returned; the results were very favorable.

Since the linear equation for amosite in IDEN.FIB did not appear to be correct, a standard amosite sample was obtained from the Bulk Asbestos Section. When the fiber was analyzed, it did not fit into the analysis scheme. It reacted poorly in a magnetic field and had the wrong refractive index. A phone call to the supplier revealed that he had never fully characterized the sample. Further checking of the asbestos fiber (ASBT1.BAS), identified the fiber as anthophyllite. A FAX of the fiber analysis was sent to the supplier in order to correct the labeling error.

PREPARING A REPORT FOR AN AIR BASE.

To save time in identifying a fiber, the analyst should inspect the Mg/Si, Na/Si and Ca/Si ratios. If the Mg/Si ratio is less than 0.15 or greater than 1.0, the sample can not be asbestos. Furthermore, if the Na/Si ratio is greater than 0.45 or the Ca/Si ratio is greater than 0.6, the sample can not be asbestos. If the above mentioned ratios indicate that the sample could be asbestos, the ASBT1.BAS program should first be executed to determine if the fiber is asbestos. If the latter is correct, the asbestos type is printed and an opportunity is presented for preparing a base report. In a previous study, 94% of the fibers were correctly identified with the first generation ASBT1.BAS data processing program (5,7).

When the element ratios are incorrect or if the ASBT1.BAS program indicate the fiber is not asbestos, the IDEN.FIB or ANAL.1 program are used to search the library for a fiber match. The computer processes the elemental data to determine if there is a match between the normalized elemental composition and the library data. The first intersection is based upon average composition and the standard deviation. If a match occurs, the elemental match is printed. Non-matches appear on the video terminal. The operator may stop the processing at anytime (F5) to inspect the results (see program documentation). After comparing the elemental composition the program constructs a linear equation for the fiber from the elements which intersect in both sets (unknown and library). The product of the equation in the library (nominally 100) and the equation for the unknown fiber are compared; if the intersection is within $\pm 20\%$, the data is either printed or stored for future evaluation.

A final selection must be made by the analyst who must make the best comparison between the library fibers and unknown fibers. Since sodium is a ubiquitous contaminate in all real samples, it is useless in the final appraisal unless its composition is greater than 10%. Calcium is a critical element. Calcium is found in most microscopic minerals and its match with known materials is very difficult. It is found in most complex silicates. Both the IDEN.FIB and ANAL.1 programs will attempt a match the calcium very closely.

INSTRUMENTATION

AL-OEA possesses an Amray 1820 Scanning Electron Micro-

scope with a Tracor Northern X-ray Analyzer. The instrument is capable of viewing and analyzing fibers with a diameter or length of 0.3 μm or greater. The use of the instrument in fiber analyses has been previously documented (5,7,8). All data processing was performed on a Zenith 100 microcomputer.

FUTURE WORK

Since calcium is critical to the final determination of a fiber composition, a data reduction system should be set up which uses calcium as the prime element for the decision making process. This would include normal calcium bearing minerals as well as tremolite, actinolite and ferroactinolite, the major calcium bearing asbestos. The data processing would be very specific but the library could be very large requiring more memory than the Zenith 100 possess.

Seven specific computer programs have been written for analyzing fibers during the summer fellowship. To make the system more user friendly, all these program should be merged and put into a tree which is assessable through the keyboard. This work will requires considerable programmer time and may not be possible with the Zenith 100. Furthermore, the Zenith 100 is too slow. A more modern computer, like a Dynal 586, is needed for processing the laboratory information. The P.I. is capable of performing this future work in his home institution and will request a mini grant for this purpose.

Glossary

Action level-more than 30 fibers/100 fields and 0.01 fibers/cc of air.

Asbestiform-fibrous material with an asbestos morphology.

Acicular-needle like crystals

Fibrous-minerals which crystallize in elongated needle like grains.

Man-made fibrous-man synthesized materials.

Normalized elemental data-conversion of selected elemental data used in a single fiber analysis to 100%.

ACKNOWLEDGEMENTS.

The P.I. wishes to thank Messieurs T. Thomas, R. Richardson and K. Roberson for their assistance and encouragements in this project. He also wishes to thank Ms D. Tessmer for providing asbestos samples/standards needed and Mr. R. Diskin for performing most the SEM-EDXA technician work. He wishes to thank to the Air Force Office of Scientific Research for the opportunity to perform this work.

References

1. N.A. Esmen, Y.Y. Hammad "Recent Studies of the environment in ceramic Fiber production. In: Biological Effects of Man-made Mineral Fibers, Vol. 1, pp 222-231. WHO/IARC Publication Lyon France (1984).
2. Int. Agency Res. Cancer "Silica and Some Silicates. IARC Monographs, Vol. 42 Lyon France (1987).
3. J.R. Kramer "Fibrous and Asbestiform Minerals" in Proc. Workshop on Asbestos: Definitions and Measurement Methods" ed. by C.C. Gravatt, P.D. LaFleur & K.F.J. Heinrich, Nat. Bureau of Stand. U.S. Govt Printing Office, Washington, D.C. (1978) p.19-33.
4. R.T. Cheng, H.J. McDermott, G.M. Gia, T.L.G. Cover & M.M. Dude "Exposures to Refractory Ceramic Fibers in Refineries and Chemical Plants" Appl. Occup. Environ. Hyg. 7, 361-7 (1992).
5. L.R. Sherman, K.T. Roberson & T.C. Thomas "Qualitative Analysis of Asbestos Fibers in Air, Water and Bulk Samples by SEM-EDXA" J.PA Acad. of Sci. 63, 28-33 (1989).
6. J. Zussman, "Crystal Structures of Amphiboles and Serpentine Minerals," P.D. Lafleur & K.J. Heinrich ed., Nat. Bur. of Stand. Washington, 20234 (1978).
7. K.T. Roberson, T.C. Thomas & L.R. Sherman, "Comparison of Asbestos Air Samples by SEM-EDXA and TEM-EDXA" Ann. Occup. Hyg. 36, 265-9 (1992).
8. L.R. Sherman "Final Report Summer Fellowship," US Air Force of Office of scientific Research (1988)
9. R. Diskin "Final Report Summer Fellowship" Inorganic Fiber Analysis by SEM-EDXA.

TABLE 1
Library data for Man-Made and Natural Fibers

Fiber	Element	Ave % Composition	Sigma	Coefficient
ASBESTOS FIBERS				
Amosite	Si	43.84	0.77	0.3511
	Mg	8.89	0.98	14.8002
	Fe	41.15	0.64	-1.1401
Anthophollite	Si	58.32	0.45	0.7446
	Mg	28.20	0.66	1.7616
	Fe	11.06	0.91	0.4903
Crocidolite	Na	17.37	1.02	2.5274
	Si	40.95	2.58	0.7662
	Fe	35.50	1.17	0.6964
Chrysotile	Na	1.20	0.40	1.5908
	Si	49.92	2.39	0.8076
	Mg	43.92	1.56	1.1281
	Fe	1.85	0.26	4.2994
	Al	1.58	0.42	0.2499
Tremolite	Si	56.34	0.76	1.1565
	Mg	22.58	0.58	0.7884
	Fe	0.39	0.16	3.8940
	Ca	17.56	0.89	0.7889
	Mn	0.16	0.10	-1.5329
Actinolite	Si	55.05	0.77	2.2558
	Mg	20.68	0.79	-0.5440
	Ca	15.50	0.70	-0.9777
Ferroactinolite	Si	54.08	1.31	1.1084
	Mg	19.90	0.96	0.5924
	Fe	7.77	0.63	1.4703
	Ca	16.64	1.20	0.9433
	Mn	0.23	0.10	0.4873
NON-ASBESTOS FIBERS				
Alumina Saffil	Si	6.00	0.20	8.8802
	Al	93.04	1.00	-0.5026
Alumina	Si	4.425	0.67	0.1444
	Al	94.165	0.09	1.0552
Bulk 6000	Si	54.96	0.85	1.1461
	Al	43.24	1.52	0.8549
	Fe	1.11	0.32	-0.01767
Ca Na Phosphate	Ca	29.77	0.92	0.4783
	Na	8.50	0.38	-0.6232
	P	60.03	0.57	1.4986
Calcium Silicate	Ca	58.23	1.00	1.02
	Si	41.77	1.00	0.99
Calcium Sulfate	Ca	55.00	2.00	1.0
	S	45.00	2.00	1.32
Cerawool	Si	51.96	0.19	1.1081
	P	15.37	0.13	3.0724
	Al	30.80	0.22	-0.2086

Cooperknit	Si	97.11	0.27	1.0192
	Ca	0.12	0.06	-2.9376
	K	0.42	0.10	0.6266
Fiberfrax	Si	33.58	0.59	0.9795
	Mg	0.25	0.02	13.6011
	Al	65.31	0.53	0.9742
Fiberfrax 7000	Si	57.47	0.24	-16.0688
	Mg	0.12	0.04	-118.435
	Al	41.69	0.14	24.9009
Fiberglass O-C	Si	63.73	0.46	1.1259
	Ca	12.95	0.50	0.7228
	Na	17.46	0.43	0.9231
	Al	2.28	0.20	0.9913
Graphite P55	Na	4.31	0.83	0.7818
	Si	5.29	1.88	-3.0205
	S	83.83	4.63	0.5238
	Fe	6.41	0.84	10.4628
	K	3.79	1.00	-0.3927
Graphite T300	Na	56.18	4.76	0.9365
	Si	20.77	5.01	-1.2376
	Mg	9.97	5.33	-0.4726
	Fe	17.34	12.82	0.0
	K	10.89	0.93	7.1433
Koawool	Si	56.33	1.02	0.8313
	Al	42.64	0.10	1.2471
Metallized Plas	Si	49.33	2.11	1.2876
	Mg	19.00	1.58	0.5274
	Na	14.65	2.31	-1.0156
	Al	14.89	1.88	2.4439
Mineral Wool A	Si	27.93	6.51	0.1477
	Ca	57.74	7.39	1.5237
	Mg	5.62	2.40	2.0618
	Fe	5.75	1.19	-0.8916
	Si	26.52	0.77	0.7170
Rockwool A	Mg	5.21	0.27	-0.1535
	Fe	4.21	0.15	3.0585
	Ca	53.20	0.98	0.7618
	Al	6.00	0.20	4.3249
	Si	99.09	0.50	1.0071
Sandtex	Al	0.24	0.06	-0.1843
	Si	67.88	0.65	0.4118
Talc 11	Mg	30.15	0.32	2.2039
	K	0.57	0.09	9.8008
	Si	57.17	1.45	1.2138
Tritan Kaowool	Mg	0.25	0.10	-5.9538
	Fe	1.45	0.23	3.5505
	K	0.63	0.10	-1.0439
	Al	40.25	0.76	0.6934

Table 2
 Statistical Analyses of Reports Sent to Bases

AL-OEA Sample Number	Total Fibers Assayed	Number of Fibers Identified
45727	5	4
45795	5	4
45796	5	2
45381	5	5
45382	5	5
45915	5	3
45917	5	4
46781	5	5
47944	5	5
47945	5	5
47986	5	5
47987	5	5
47089	5	5
47990	5	4
Ward Sci.	4	4
Total	74	65 88%

APPENDIX A

SQ: ASBTOS3
 ?31?31 .3100E+02
 LABEL:CHRYSOTILE STANDARD
 FILE NO. (1-4095)=?31
 DRIVE NO. (0- 4)=?3
 LABEL: CHRYSOTILE

CHRYSOTILE STANDARD
 Standardless Analysis
 20.0 KV 38.5 Degrees

Chi-sqd = 0.71

Element	Rel. K-ratio	Net Counts
Si-K	0.47277 +/- 0.04923	413 +/- 43
Mg-K	0.41973 +/- 0.04002	336 +/- 32
Fe-K	0.10423 +/- 0.04790	30 +/- 14
Ca-K	0.00000 +/- 0.00000	0 +/- 0
Mn-K	0.00328 +/- 0.03792	1 +/- 13
Na-K	0.00000 +/- 0.00000	0 +/- 0

PT.NO. 1 FILE 31
 F Correction 20.00 kV 38.50 deg
 No.of Iterations = 5

Element	K-ratio	Z	A	F	Atom%	Wt%
Si-K	0.330	0.992	1.634	1.000	52.70	53.58
Mg-K	0.294	0.992	1.314	0.991	43.09	37.92
Fe-K	0.073	1.114	1.013	1.000	4.07	8.23
Ca-K	0.000	1.015	1.163	0.997	0.00	0.00
Mn-K	0.002	1.135	1.021	1.000	0.13	0.27
Na-K	0.000	1.017	1.609	0.985	0.00	0.00
					Total=	100.00%

SQ:

APPENDIX B

AIR FORCE
OCCUPATIONAL AND ENVIRONMENTAL HEALTH DIRECTORATE
Brooks Air Force Base, Texas 78235-5000
Report of Fiber Analysis as Determined by
Scanning Electron Microscopy

The date is 08-27-1992 Time is 07:30:08

The results of this report are based on the composition and mass ratios of the elements characterized as asbestos, ceramic fibers, graphite or fiberglass. These mass ratios allow this laboratory to make an educated estimation of the type of fibers found on filters with fiber counts greater than 30 fibers/100 fields. It must be made clear, however, that the results of this report are only the opinion of the laboratory analyst.

The number of fibers analyzed on the filter was 5

OEHL Sample No.	BASE Sample No.	Results
92 47986	SZ920412	AMOSITE This fiber is an amphibole
92 47986	SZ920412	AMOSITE This fiber is an amphibole
92 47986	SZ920412	NON-ASBESTOS COMPLEX SILICATE
92 47986	SZ920412	NON-ASBESTOS FIBERGLASS
92 47986	SZ920412	NON-ASBESTOS CALCIUM SILICATE

62 Medical Squadron
McCord AFB WA
98573-5000

KENNETH T. ROBERSON
Chief of Air Asbestos and
Particle Analysis

APPENDIX C

SURVEY FIBER IDENTIFICATION

Standard air samples which are submitted to the Armstrong Laboratory for asbestos air counting may in the very near future be identified as to the exact composition of the fibers on the cellulose acetate filter, e.g. chrysotile, amosite, ferroactinolite, ceramic fibers, graphite or Fiberglas. Since this qualitative analysis is time consuming and expensive, only samples which have sufficiently large counts (>30 fibers per 100 fields with a level of 0.01 fibers/cc or greater) justify the additional analyst time. These representative fibers should give the environmental health officer a statistical representation of what is on the filter in the area where the samples were collected.

Please help us to provide better service by answering the following questions regarding the sample(s) you recently submitted to the Armstrong Laboratory.

1. Would you like to have this service included with your reports when the fiber count justifies the added expense? Yes ___ No ___

2. If this service were provided, could you use this information in further environmental evaluations. HOW? Yes ___ No ___

3. How often would you request this service. Once a year ___ Now and then ___ with every sample ___ never ___

4. Was all pertinent information provided on the report correct? i.e. Base sample number, your address, units of measures) Yes ___ No ___.

5. Is the report format easy to read/understand? Yes ___ No ___.

6. Would a report like the enclosed report be satisfactory? Yes ___ No ___

7. If you marked no in question 4, what would you like included\excluded?

8. What is your overall satisfaction with the service you receive from the Analytical Service Division of AL? Excellent ___ Good ___ Poor ___ Not Applicable ___.

9. Which of the following do you consider the most important in this area (1 being priority)

- ___ Timeliness of the service
- ___ Quality and accuracy of the service
- ___ Quick response to resolving the issue

Point of Contact:
K.T. Roberson
Autovon 240-3626

A. Richardson III
Chief
AL/OEAO
Brooks AFB TX 78235

CHOICE BETWEEN MIXED AND UNMIXED GOODS IN RATS

**Alan Silberberg
Professor
Department of Psychology**

**John Widholm
Graduate Student
Department of Psychology**

**The American University
Washington, DC 20016-8062**

**Final Report for:
Summer Research Program
Armstrong Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, Texas**

August 1992

CHOICE BETWEEN MIXED AND UNMIXED GOODS IN RATS

Alan Silberberg
Professor
Department of Psychology
The American University

John Widholm
Graduate Student
Department of Psychology
The American University

Abstract

Twelve food- and water-deprived rats chose between two levers. A multiple fixed-ratio 49 fixed-ratio 1 schedule was associated with one lever and a multiple fixed-ratio 25 fixed-ratio 25 was associated with the other. In Phase 1 for both levers, one of the two components defining the multiple schedule delivered access to 0.1-cc of water while the other component delivered a single 45-mg food pellet. The order of food and water presentations was counterbalanced across subjects. To prevent absolute preference for an alternative from developing, the values of each multiple schedule were adjusted according to a titration schedule: If a multiple schedule was selected four times in succession, its ratio values were incremented. In Phase 2, half the rats were exposed to only water reinforcement, while the other half received only food reinforcement. In all other ways, the experiment was unchanged from Phase-1 conditions. There was no reliable change in preference, an outcome incompatible with the economic notion that organisms prefer mixtures of goods over unitary presentations of a good.

CHOICE BETWEEN MIXED AND UNMIXED GOODS IN RATS

Alan Silberberg
John Widholm

Herrnstein (1961) demonstrated that when pigeons chose between two keys each associated with a variable-interval (VI) schedule of reinforcement, the ratio of choices to one of the two alternatives equalled the ratio of reinforcers delivered. For example, if one key provided 20 reinforcers/hour and the other provided 40/hour, pigeons might respond 1000 times to the former key and 2000 times to the latter key in a one-hour session.

This phenomenon of matching has been used to scale reinforcer values (Baum & Rachlin, 1969). The rationale for this use can be illustrated in terms of the example offered above: Receipt of 40 reinforcers in an hour is, by definition, twice as valuable as receipt of 20. That a pigeon reflects this value ratio in terms of a 2:1 behavior ratio means that value ratios are perfectly mapped by behavior ratios.

The occurrence of matching on concurrent VI VI schedules introduces the possibility of scaling the value ratios of different reinforcers. Miller (1976) exposed pigeons to two keys, each associated with a VI 2-minute schedule. The reinforcers (hemp, wheat, corn, millet) differed between keys. When wheat and corn served as the reinforcers, all birds preferred wheat.

Miller argued that this outcome established not only the ordinal value of these reinforcers (i.e., wheat is preferred to corn), but also the value ratio of these goods (defined by each subject's choice ratio).

While Miller's conclusions are compatible with his data, his goal of scaling reinforcers that differ in kind is, from the perspective of an economist, unrealizable. The problem is that the value of each good can depend on the kind of good offered as its alternative. Where there is a dependency, goods are said to be complements. An extreme example would be to offer a human a choice between a tennis racquet and a tennis ball. Since these two goods are used together, their value individually is low, but together is high. In terms of foods, the value of gin is enhanced when it is offered with tonic, and is diminished when each of these goods is offered separately. In this case, the value of one good depends on the availability of another. Absent independence in a reinforcer's value, it is not possible to scale reinforcers in terms of value ratios as Miller attempted to do.

The present experiment attempts to demonstrate a complementary reinforcer interaction in rats by using food and water as complementary goods in choice. The idea is to give rats a choice between complementary goods that are presented closely together in time vs the same goods spaced apart in time. Presumably, rats would prefer the mixture of these goods over

their separate presentation in the same way that a human would prefer his tonic with his gin rather than separate presentations of each.

Method

Subjects. Twelve, experimentally naive, male white rats at approximately three months of age served as subjects.

Apparatus. A 25 X 30 X 30 cm Coulbourn chamber served as the experimental space. The chamber was composed of Plexiglas except for a response panel and a metal-rod floor. Two 4-cm wide levers were located 5 cm off the floor and 2.5 cm from each side wall. Three small lights were located over each lever. A two-sided feeding cup was located one cm off the floor in the middle of the response panel. Water reinforcement consisted of a single presentation of a 0.1-cc dipper. Food reinforcement consisted of a single 45-mg BioServ food pellet.

Procedure. After subjects were trained by the method of successive approximations to respond to both levers in the chamber, they were placed in the first phase of the experiment. For half the rats in Phase 1 (Group 1), both levers provided first food and then water. One lever provided these reinforcers according to a multiple fixed-ratio (FR) 49 (food) FR 1 (water) schedule while the other lever provided these reinforcers according to a multiple FR 25 (food) FR 25 (water) schedule. For the remaining half of the

subjects (Group 2), the order of food and water was reversed--i.e., rats were exposed to a concurrent multiple FR 49 (water) FR 1 (food) multiple FR 25 (water) FR 25 (food) schedule. Lever availability was cued by illumination of the lights above each lever. Both levers were concurrently available until the fifth response to a lever. At that point, the light above the other lever was extinguished and responses to that lever had no scheduled consequences. Following completion of the operative multiple schedule, there was a 30-second interval during which all lights were extinguished before the next trial began.

The value of the multiple schedules was adjusted automatically based on the pattern of choice during the prior four trials. If during the prior four trials a subject exclusively chose one alternative, the values of its component ratios were altered. If the preference was for the multiple FR 25 FR 25, those values were incremented by 5. If, on the other hand, the exclusive preference was for the multiple FR 49 FR 1 schedule, its value was increased to multiple FR 54 FR 1. Before any schedule adjustment occurred, the values of the alternate schedule's ratio values were determined. If they had been increased previously by exclusive preference to that alternative, they were decreased by a like amount before the lever with the exclusive preference was incremented.

Daily sessions ended after completing the second component of the multiple schedule in operation when 90 minutes of session time had elapsed. Following each session, rats were given unrestricted access to food and water for 15 minutes in their home cages. At that point, all food and water were removed. For some subjects, Phase 1 ended when, in the judgment of the experimenter, choice ratios had been stable for five successive days. Other subjects were switched to the Phase-2 contingencies despite the instability of their choice ratios because the experiment had to be completed before the summer ended.

In Phase 2, half of the subjects in Groups 1 and 2 had their food presentation replaced with a water presentation. Thus, these animals received only water in choice from both levers. The other half of the subjects from Groups 1 and 2 received only food from both levers. Table 1 presents the subject, order of experimental phases, and the number of sessions in each phase.

Results

Figure 1 presents the proportion of all choices that were to the multiple FR 49 FR 1 schedule over the last five sessions of Phase 1 (left-side panels) and Phase 2 (right-side panels). The horizontal line parallel to the X-axis defines the mean performance during the five sessions presented in a panel.

Table 1

Subject number, phase number, experimental condition, and number of sessions.

Subject #	Phase #	Experimental Condition	# of Sessions
1	1	Food - Water	22
	2	Food - Food	19
3	1	Food - Water	28
	2	Food - Food	13
4	1	Food - Water	28
	2	Food - Food	13
5	1	Food - Water	28
	2	Water - Water	13
7	1	Food - Water	22
	2	Water - Water	19
8	1	Food - Water	28
	2	Water - Water	13

(Table 1 continued)

9	1	Water - Food	23
	2	Food - Food	18
10	1	Water - Food	28
	2	Food - Food	13
11	1	Water - Food	22
	2	Food - Food	19
12	1	Water - Food	22
	2	Water - Water	19
13	1	Water - Food	22
	2	Water - Water	22
14	1	Water - Food	22
	2	Water - Water	19

The subject number is presented above each panel. In comparing Phase 1 with Phase 2, no consistent, across-subject changes in relative response rate can be discerned.

Discussion

Economics predicts that all other things being equal, animals will prefer mixtures of goods over their unitary presentation. This prediction follows from the fact that the value of additional amounts of a good diminishes as a function of the amount of that good already in hand (diminishing marginal utility). For example, consumption of the first cookie in store of cookies is more highly valued than consumption of the second and so on. It is improbable that consumption of, say, a fifth cookie would have any utility. If you have many cookies, any given cookie would have little or no value to you. But what of an apple? If you have no apples, and are given the opportunity to trade cookies for apples, you will surely do so because the cookies, by virtue of their abundance, are of low value, while the apples, by virtue of their scarcity, are of high value. Once you have traded cookies for apples, you have a mixture of them, and your action defines your preference for mixtures of goods over any single good.

The present experiment was intended to reproduce these circumstances. For one choice (multiple FR 49 FR 1), two goods (food and water) came in

close proximity. For the other choice (multiple FR 25 FR 25), these goods were spaced further apart. If proximity is an analogue of a mixture, animals should have preferred the multiple FR 49 FR 1 schedule over the multiple FR 25 FR 25 schedule--an outcome that would have been reflected in a decreased relative response rate in Phase 2 compared to Phase 1. As Figure 1 makes apparent, this result did not obtain.

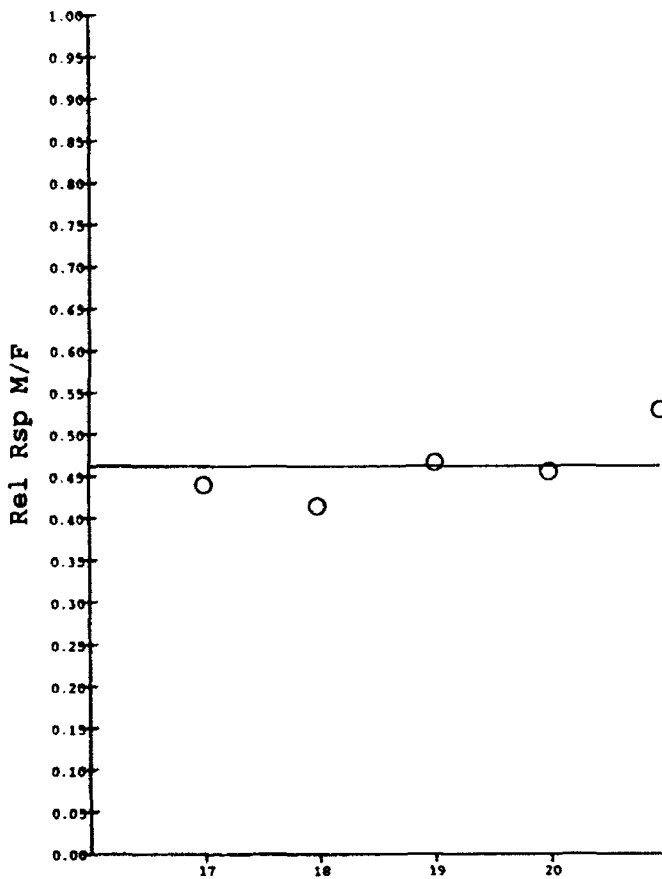
Two explanations seem possible for the failure to demonstrate an experimental effect. First, serial presentation of two goods may not be viewed as a mixture by rats because their capacity to integrate the delivery of these goods over time is too limited. By such an account, neither of the alternatives was viewed as a mixture. Alternatively, if rats integrate events over much longer time periods, the failure to find a preference for mixtures may reflect that fact that both schedules were viewed as providing mixtures. In subsequent work we will try to remedy the failure of this experiment by presenting the "mixed" goods synchronously and by spacing the "unmixed" goods further apart.

References

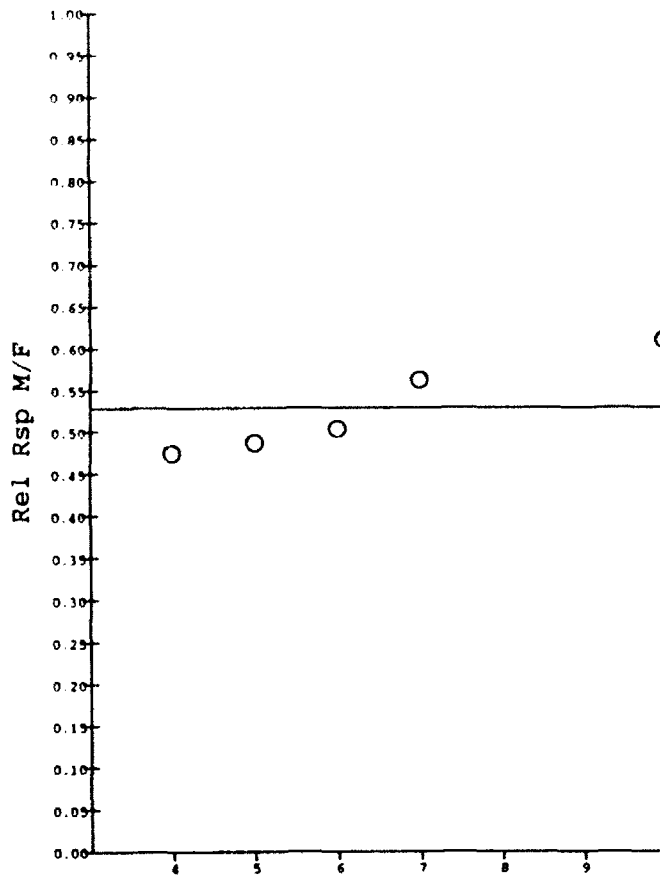
- Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. Journal of the Experimental Analysis of Behavior, 12, 861-874.
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. Journal of the Experimental Analysis of Behavior, 4, 267-272.
- Miller, H. L., Jr. (1976). Matching-based hedonic scaling in the pigeon. Journal of the Experimental Analysis of Behavior, 26, 335-348.

FW1:21-JUL-92:FW:RS=3637

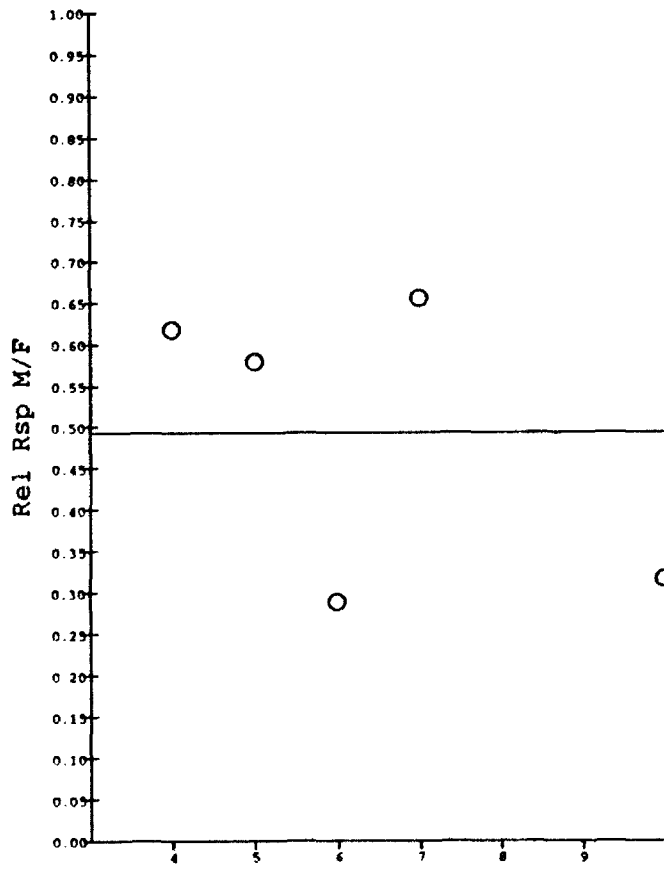
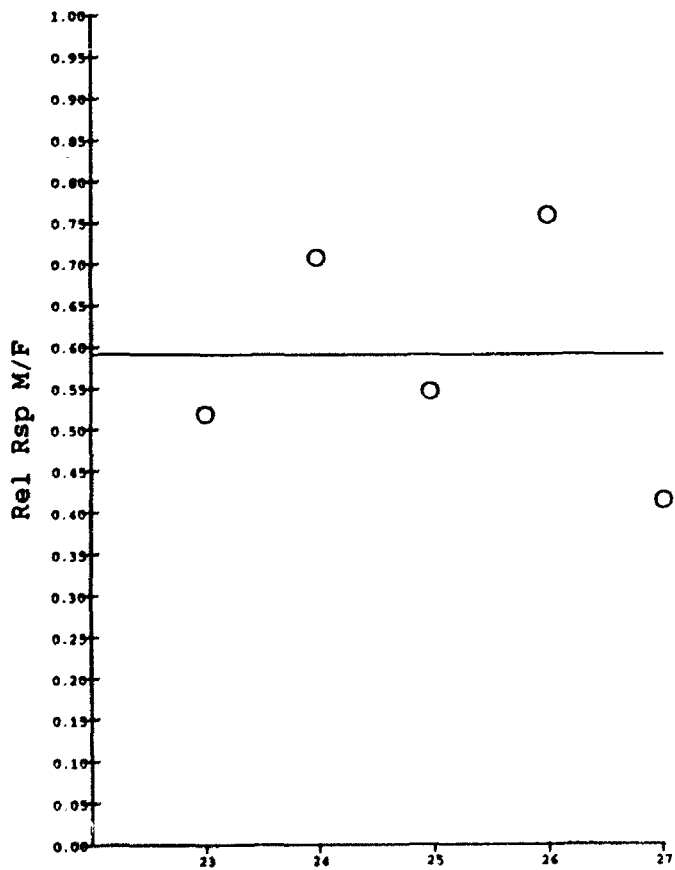
MIX1:10-AUG-92:FF:RS=4774



FW3:27-JUL-92:FW:RS=1627

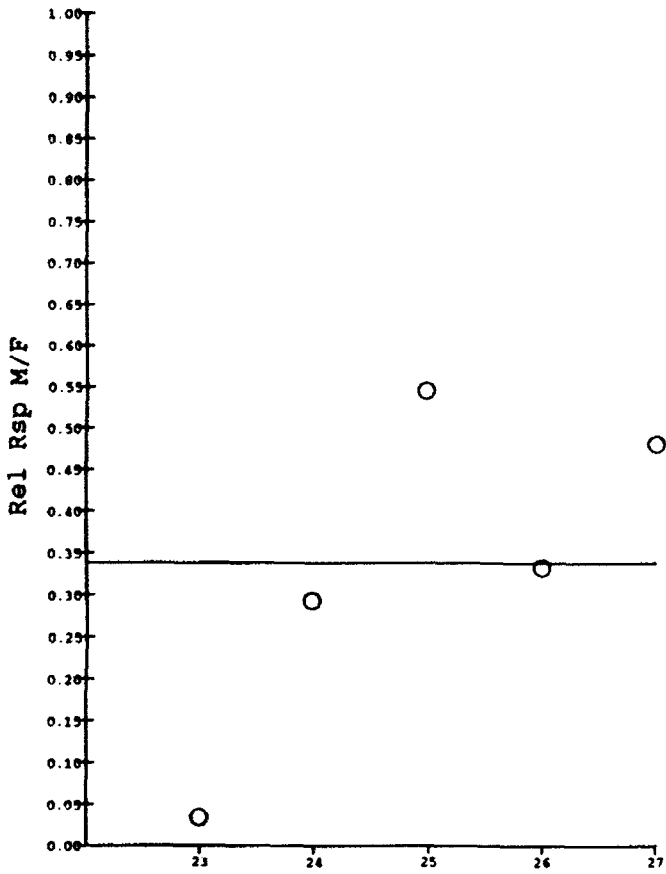


MIX3:10-AUG-92:FF:RS=3005

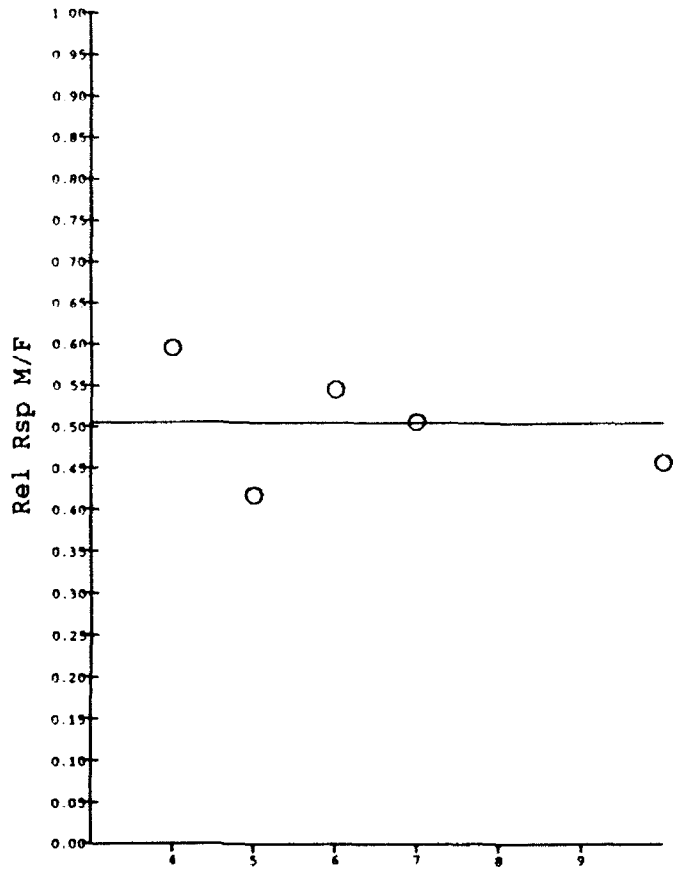


FW4:27-JUL-92:FW:RS=197

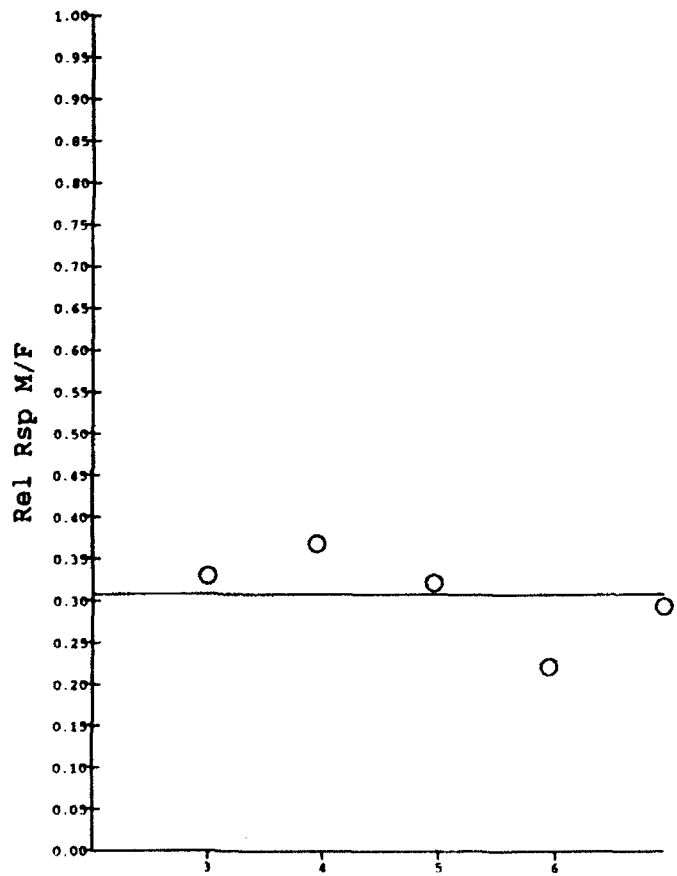
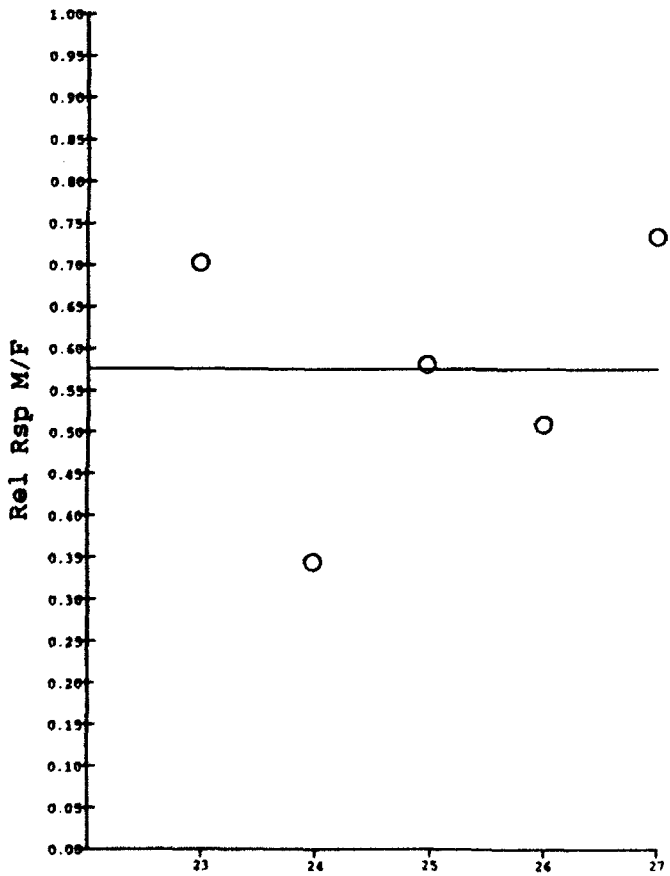
MIX4:10-AUG-92:FF:RS=2747



FW5:27-JUL-92:FW:RS=2086

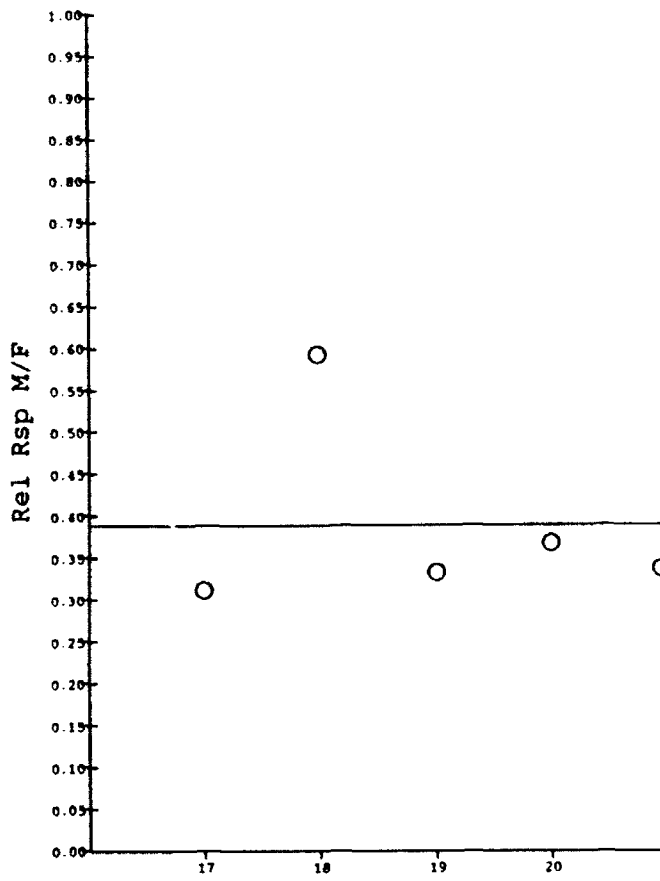


MIX5:07-AUG-92:WW:RS=685

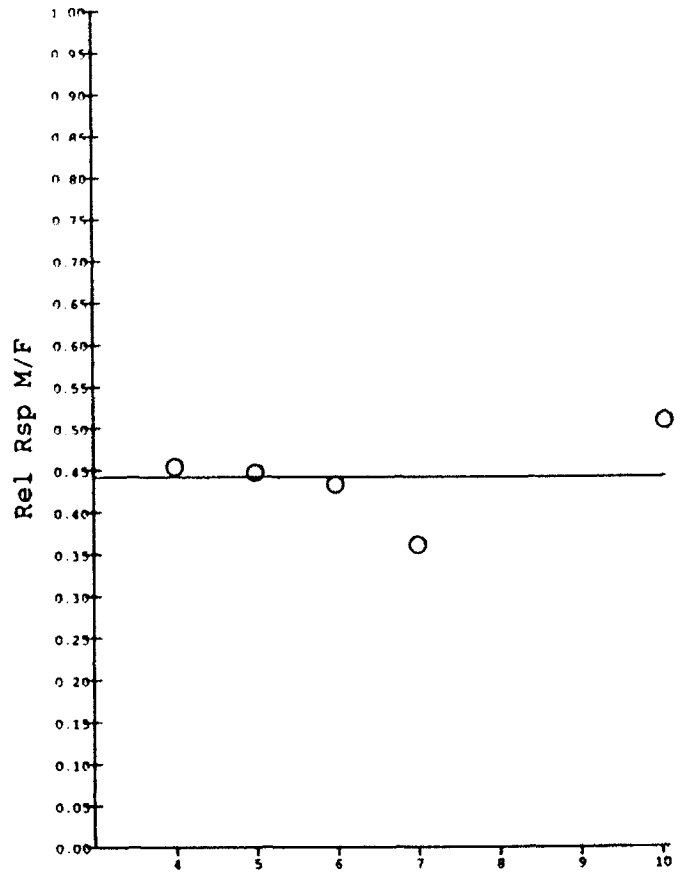


FW7:21-JUL-92:FW:RS=4591

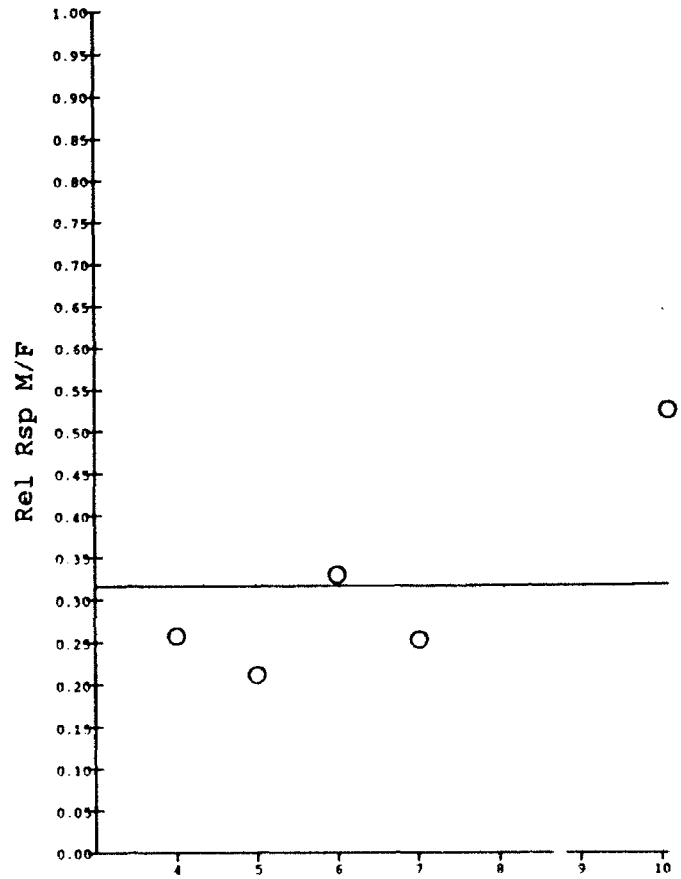
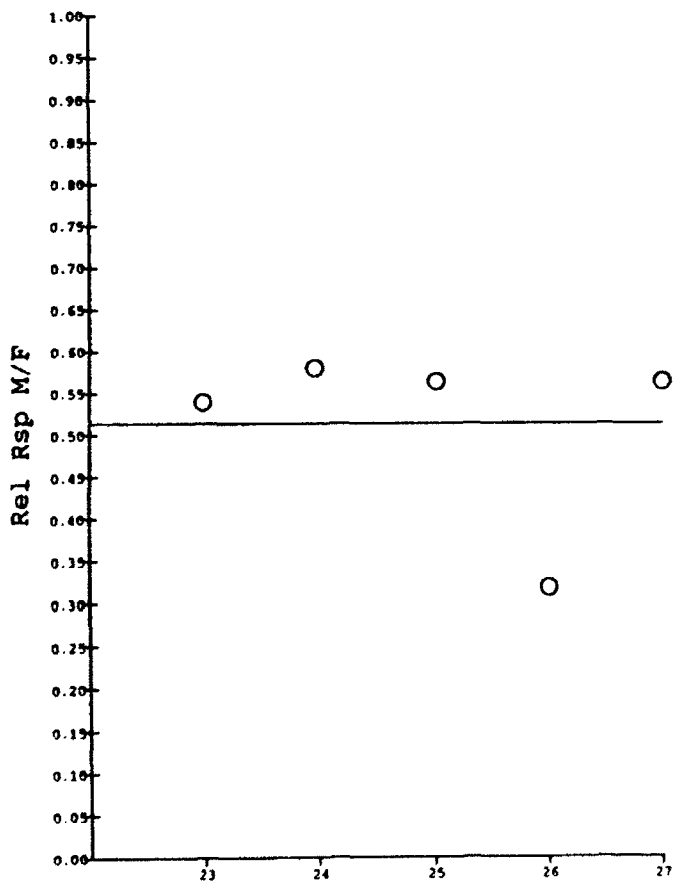
MIX7:10-AUG-92:WW:RS=330



FW8:27-JUL-92:FW:RS=3716

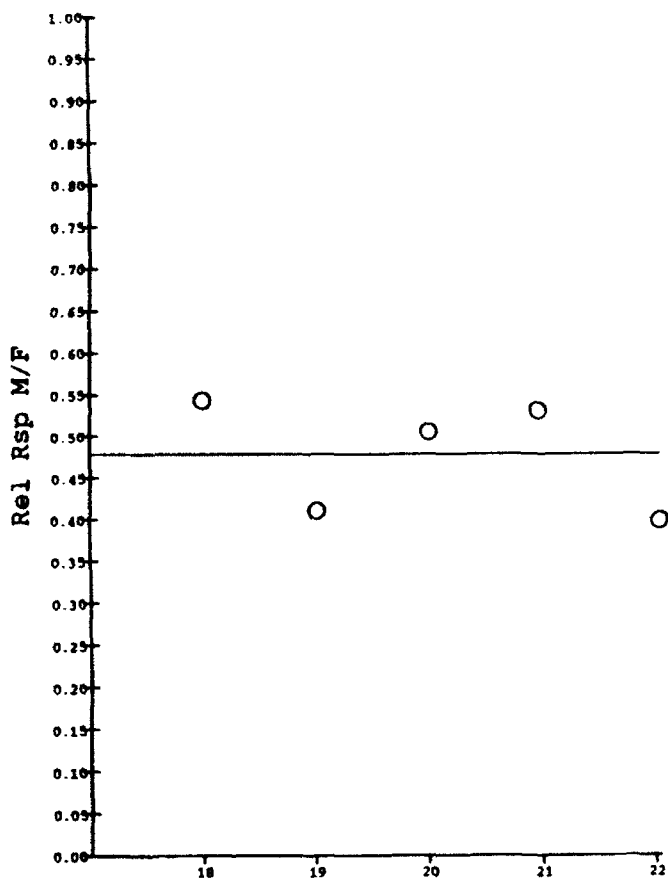


MIX8:10-AUG-92:WW:RS=105

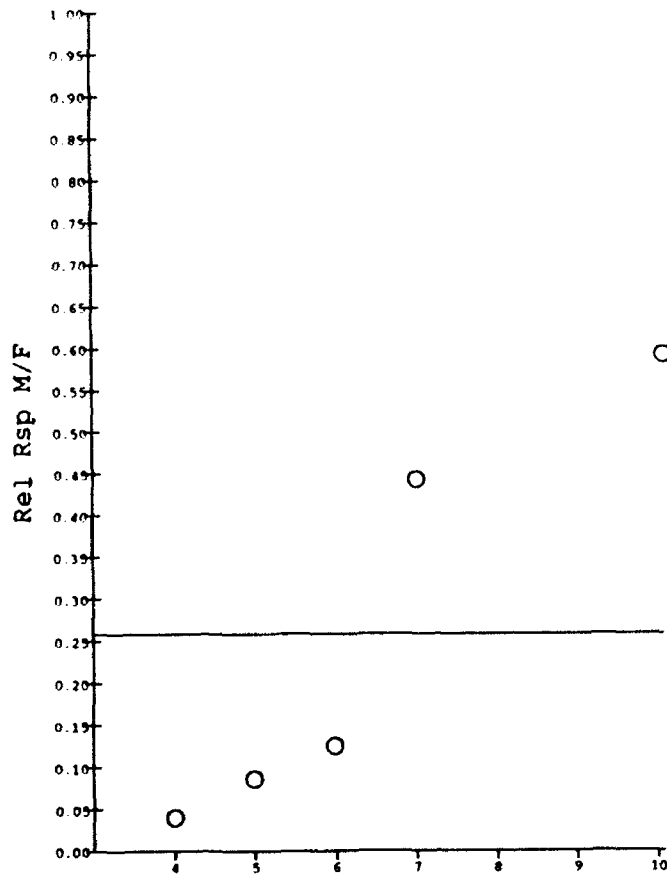


FW9:22-JUL-92:WF:RS=5390

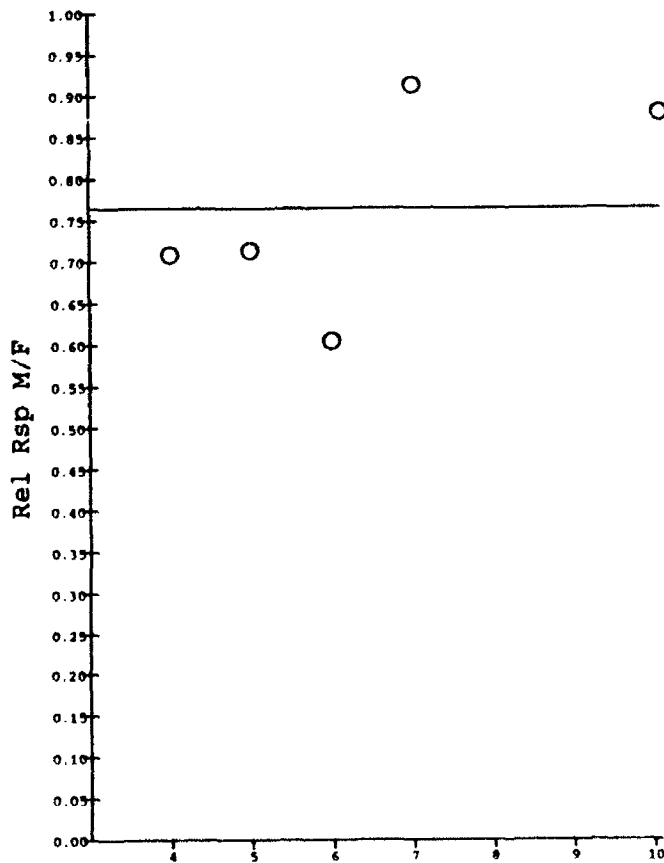
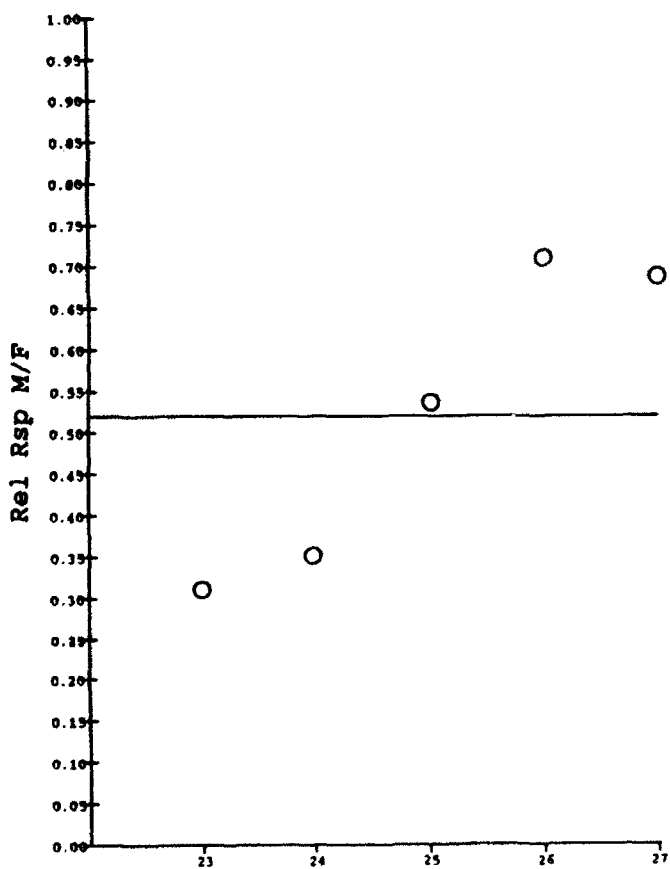
MIX9:10-AUG-92:FF:RS=5570



FW10:27-JUL-92:WF:RS=879

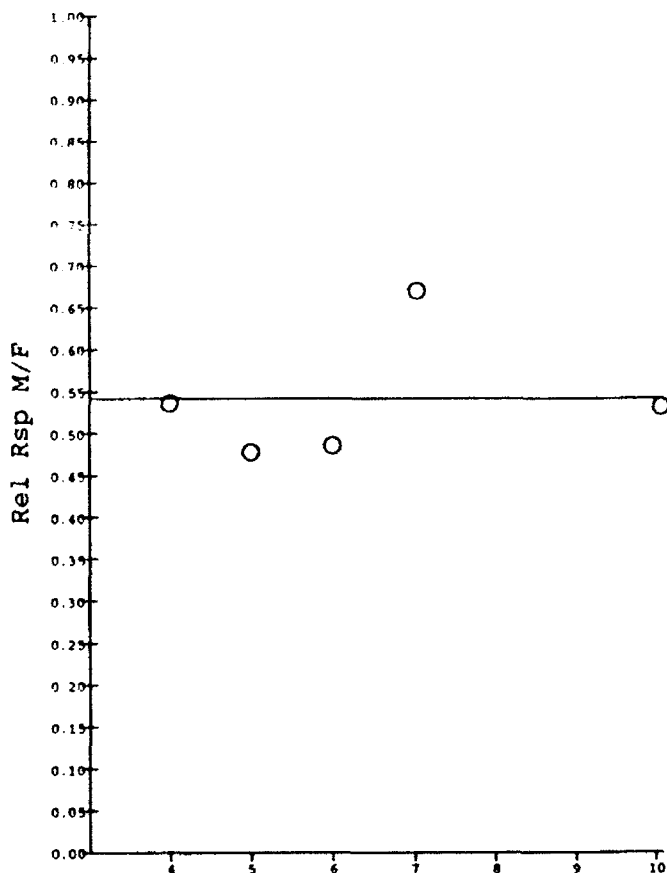
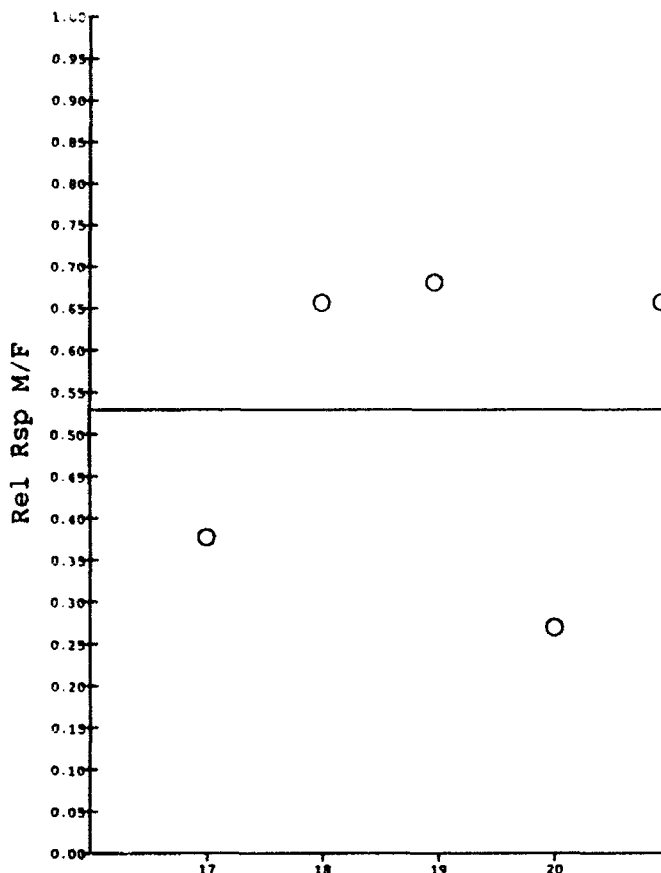


MIX10:10-AUG-92:FF:RS=2977



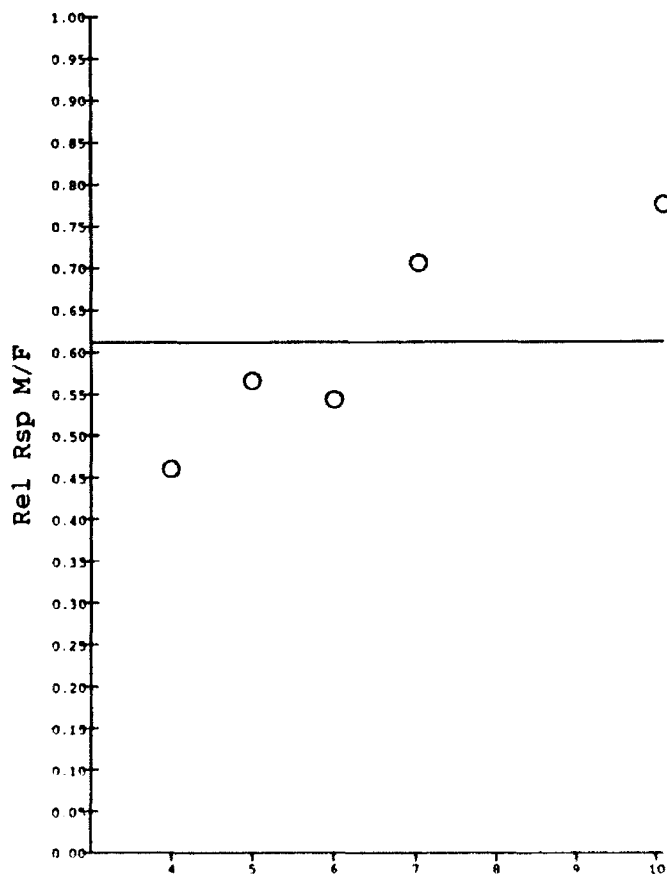
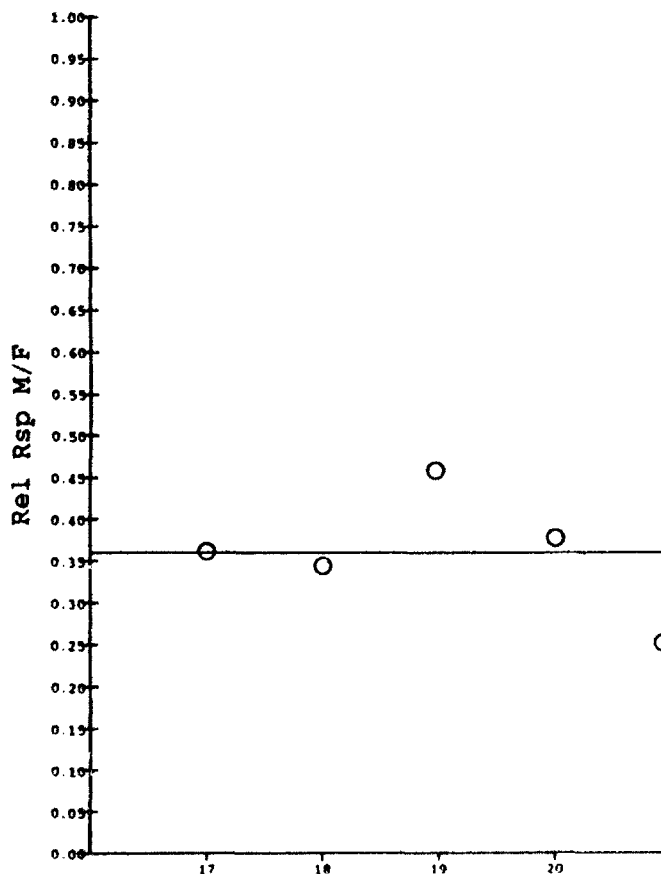
FW11:21-JUL-92:WF:RS=5764

MIX11:10-AUG-92:FF:RS=5895



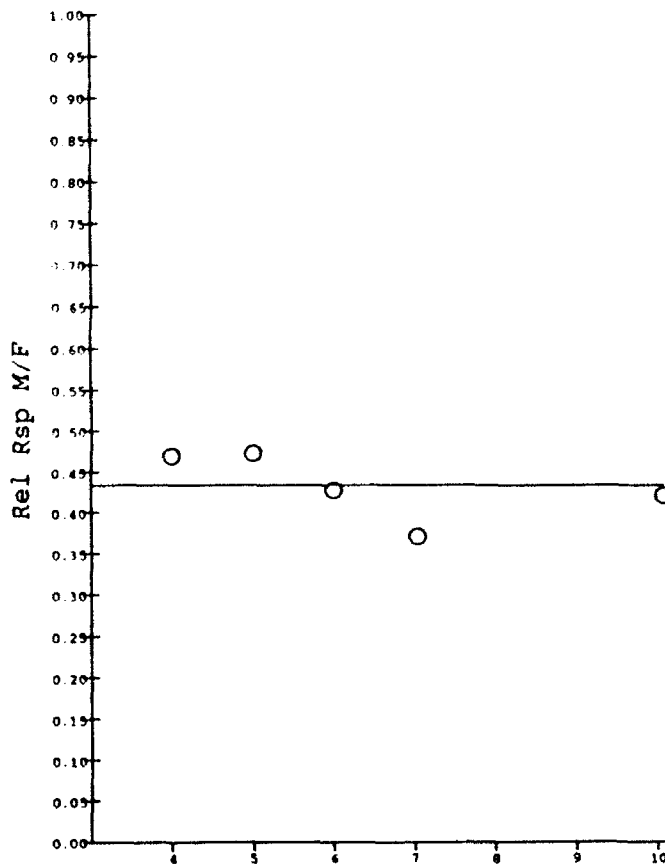
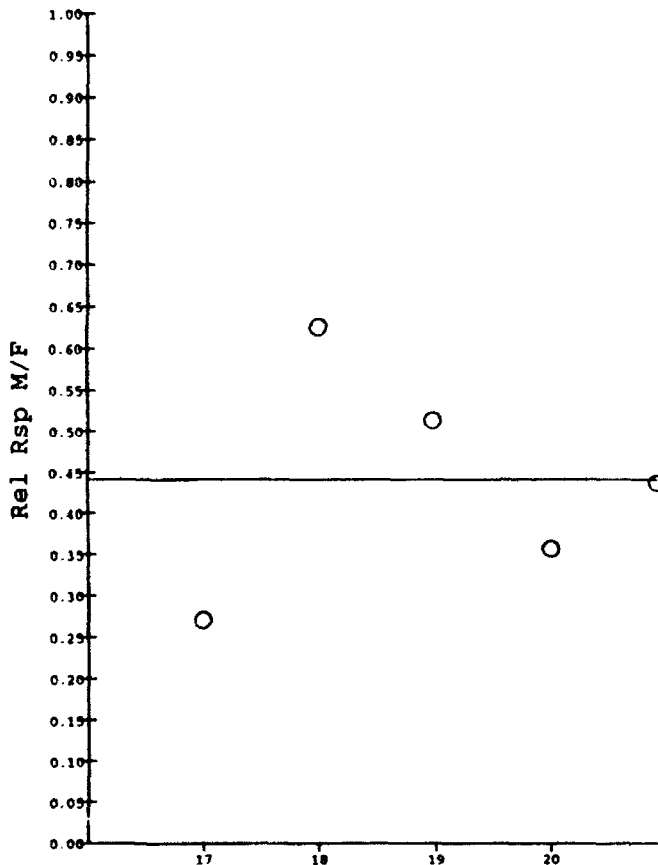
FW12:21-JUL-92:WF:RS=1991

MIX12:10-AUG-92:WW:RS=452



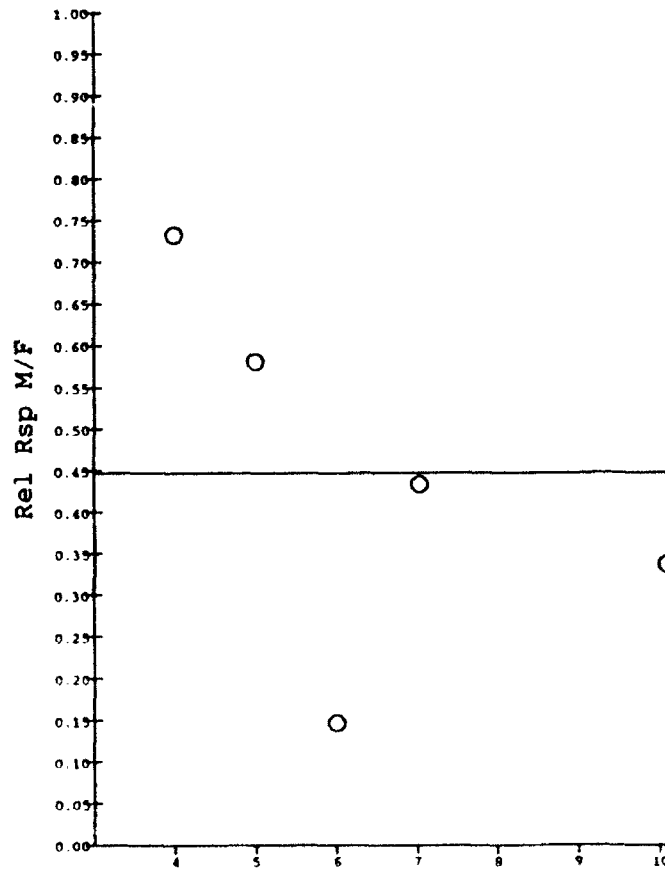
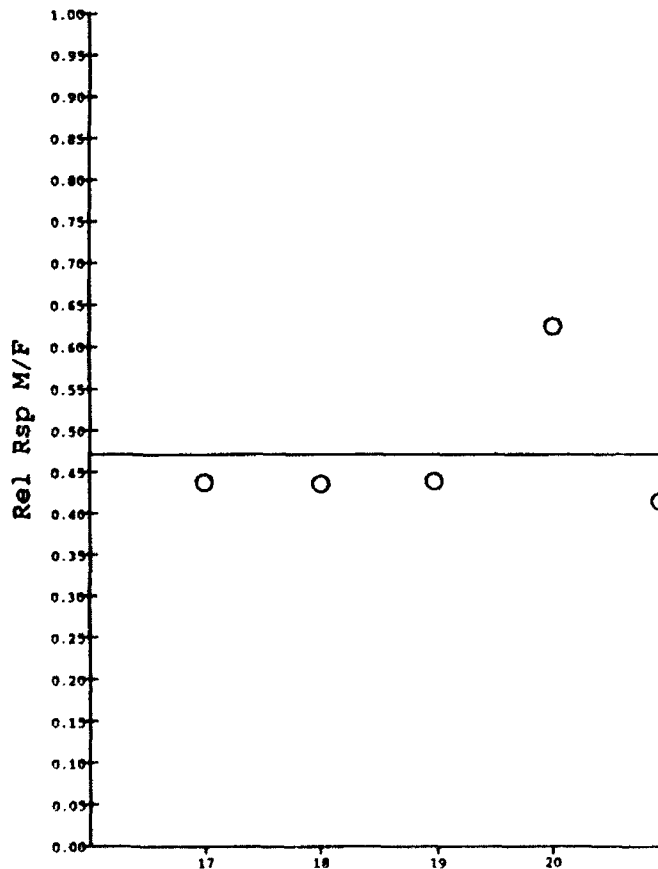
FW13:21-JUL-92:WF:RS=4561

MIX13:10-AUG-92:WW:RS=239



FW14:21-JUL-92:WF:RS=4861

MIX14:10-AUG-92:WW:RS=151



THE USE OF SMALL GROUPS IN COMPUTER-BASED TRAINING

Stanley D. Stephenson
Associate Professor
Department of Computer Information Systems
and Administrative Sciences

Southwest Texas State University
San Marcos, TX 78666

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

THE USE OF SMALL GROUPS IN COMPUTER-BASED TRAINING

Stanley D. Stephenson
Associate Professor
Department of Computer Information Systems
and Administrative Sciences
Southwest Texas State University

Abstract

In the majority of studies which have investigated individual versus small group achievement within the CBT framework, there were no significant differences between the two experimental groups. Some studies produced significantly positive results, but no study produced significantly negative results. However, one would expect groups to outperform individuals. After reviewing the small group CBT literature, this paper suggests that in past studies the behavior of the small group members has not been appropriately structured. Based on related traditional instruction research, it appears that guiding students' discussions following CBT presentation may increase achievement. A reciprocal peer-questioning model is proposed to provide this type of guidance. This model is briefly described and research is suggested. Implications of this model for distance learning are also provided.

THE USE OF SMALL GROUPS IN COMPUTER-BASED TRAINING

Stanley D. Stephenson

INTRODUCTION

Recent computer-based training (CBT) research has indicated that achievement is at least as great when students work in small groups versus working individually. Moreover, small group CBT achievement is never significantly less than individual CBT. This is a significant finding because every review and meta-analysis on the overall effects of CBT versus traditional instruction (TI) has typically reported at least one significantly negative result.

Consequently, the general consensus on small group CBT is that it does not decrease achievement and that, given financial considerations, it is probably more cost effective than having students work CBT individually. However, the notion that small CBT groups might increase achievement has not been seriously suggested even though research from seemingly related areas (e.g., cooperative learning) would suggest that using more than one student on the computer might produce higher performance.

This paper will discuss the use of small groups in CBT with the emphasis on achievement versus, for example, attitudes, interests, etc. First, an overview of CBT research will be given. This overview will rely on recent literature reviews and meta-analyses. Second, the existing small group CBT literature will be reviewed. This review will include all small group CBT studies which have been published and which were available during

the summer of 1992. It is acknowledged that there are a small number of studies which were unavailable and which therefore were not reviewed. Third, small group CBT factors which should be researched will be discussed, and a small group CBT model will be offered. Part three also contain comments on the applicability of the findings in this paper for distance learning.

OVERVIEW OF CBT RESEARCH

Several comprehensive CBT reviews and meta-analyses have been conducted in recent years. These reviews suggest that CBT can produce an Effect Size of about 0.30 along with a reduction in course length of about 33 percent when compared to TI. Also, there is general consensus that the CBT studies which produce the greatest gains are (a) three months or less in length, (b) in the lower grades, and (c) on relatively less complex tasks and less difficult courses.

However, there are several cautions which must be discussed. First, most if not all of the studies reviewed were one time only studies; i.e., we do not have results of the impact of CBT over repeated offerings. Second, some CBT studies do not contain enough information to make definitive statements (see Hmelo, 1989). Third, most of the articles report the efforts of CBT proponents; i.e., researchers who are interested in the CBT concept. We do not know the efficacy of CBT when it is conducted by some one removed from the original excitement of the new technique. Fourth, there has been a consistent discrepancy between published and unpublished results over the effects of

CBT; unpublished reports have typically produced lower scores. Related to this issue is the lack of information about failed CBT projects. And fifth, it appears that about 25 percent of all CBT studies produce results favoring TI, although most are not statistically significant.

THE USE OF SMALL GROUPS IN CBT

A review was conducted of all retrievable studies (n=30) which compared achievement between individual CBT and small group CBT. This paper will offer conclusions and representative references from that review; an expanded paper with a complete bibliography is available from the author.

A major result from these 30 studies is that in no case did individual CBT produce statistically higher achievement than small group CBT. Small group CBT was at least equal to individual CBT and statistically better in 11 (36 percent) studies. This result alone is evidence that small group CBT does not harm achievement. However, the fact that many studies produced insignificant results warrants examination.

In many of the no-difference studies, the length of the experimental session was either one session of less than 60 minutes or several sessions which totaled less than 120 minutes. For instance, in the Carrier and Sales article (1987), which is one of the most frequently cited no-difference studies, the experimental session lasted 26 minutes.

In most small group CBT studies, the dependent variable was measured with either a brief paper-and-pencil or computer-based

multiple choice exam. For instance, in Makuch et al. (1991) the dependent variable was measured with a 16 item post test on which 54 percent of the individual CBT students made a perfect score. It would be difficult for any learning setting to improve on those results. These results suggest that if small group CBT does increase achievement, short multiple choice tests may not have the power to detect the increase, especially if the experimental session is also short. Some other dependent measure, such as task performance, long term retention, etc., may also be necessary.

Several studies reinforce the suggestion that a single multiple choice test may not have the power to detect what was learned in a small group setting. Dick (1963) had individual and paired students cover 28 units of CBT over a 10 week period. A short multiple choice post test produced no immediate difference, but a year later, the paired students recalled significantly more material than did the individual students. A similar result was reported by Shlechter (1990). Stephenson (1992) had individual and paired students work a CBT spreadsheet tutorial in three, 70 minutes sessions and measured two dependent variables: number of spreadsheet commands used and performance on a spreadsheet exercise. There was no difference in number of spreadsheet commands used, but the paired students scored higher on the exercise. These results suggest that paired CBT produces a 'deeper' level of learning, a level which can not be detected by a short exam.

Paired students may spend longer in CBT. Carrier and Sales (1987) and Makuch et al. (1991) reported that even though there were no achievement differences between individual and paired students, paired students took longer to complete the training. In both of these studies, the experimental session was less than 30 minutes. Okey and Majer (1976) found that pairs took longer than individuals but that groups of three or four took less time. Dick (1963) reported that paired students spend an average of 3.7 minutes more on each unit. However, this author found that paired students spent longer on a spreadsheet tutorial during the first of three sessions but spent about the same amount of time during the third session. And Dossett and Hulvershorn (1984) found that over a week long class, paired students required less time than did individual students, a result also reported by Shlechter (1990).

It should be pointed out that there are studies which are methodological sound and which did not produce significant difference between individual and paired student. For instance, Dossett and Hulvershorn (1983) had USAF recruits in an electronic principles course spend a week of training using CBT. They found no difference on an end-of-week written examination, but individual students did not outperform paired students.

Overall, most of the small group CBT studies show higher, but not statistically higher, performance for pairing. Many of the studies have one or more inadequacies, such as short experimental sessions and short multiple choice dependent

measures. Finally, there is no evidence that pairing students lowers achievement.

Specific Small Group CBT Factors

Group aptitude composition is a frequently manipulated small group CBT variables. Naturally, if groups are homogeneous in aptitude, higher aptitude groups will outperform lower aptitude groups (Stephenson, 1992). For heterogeneously mixed ability groups, the results are less clear. Hooper and Hannafin (1988) reported that mixed ability groups produced improved achievement for low ability students without a seriously detrimental impact on high ability students. However, Yeuh and Alessi (1988) found that ability level in mixed-aptitude groups had no effect on achievement, and Hooper et al. (1989) found that heterogenous groups suffered with regard to achievement.

Overall, the literature on aptitude composition of CBT groups is inconclusive. Related to the aptitude question are those studies which report that individual CBT seems to help high ability students disproportionately more than low ability students (Stephenson, 1992). Such results would question whether mixed ability teams would work efficiently, given that in isolation the high ability team member will benefit more from the CBT program than the low ability member. The aptitude grouping issue is probably highly interactive with subject material, software focus, aptitude differential, and course length.

Another important group composition factor is gender. The little work that has been done suggests that small groups (at

least pairs) should be like-gender. Dalton (1990) found that heterogeneous groups of males and females scored lower than homogeneous teams, a result also reported by Underwood et al. (1990). Carrier and Sales (1987) reported that mixed gender teams engaged in more off-task behavior than same gender teams. This author found that mixed gender teams of college students were more socially oriented during CBT sessions. Bellows (1987) reported that when one child was dominant in a mixed-sex group, five out of six times it was the male. Common sense would suggest that age may play a role. Mixed groups of younger students (e.g., K-8) will interact much differently than older students (e.g., college students).

Heterogenous grouping can be based on dimensions other than aptitude or gender. For example, students with similar aptitudes but with different experience levels on a computer could be paired. Or, students of like gender but with different motivational levels (e.g., grade point averages) could be paired. In fact, it would be difficult if not impossible to pair two students who are not similar on at least one dimension critical to CBT achievement.

Group size has not been thoroughly studied; most studies have used two students per team. Cox and Berger (1985) studied group sizes of one, two, three, and five. The highest mean achievement and lowest standard deviation were found in the dyads. Trowbridge and Durnin (1984) studied groups of one, two, three, and four; pairs and triads seemed to be the best

combinations. They also stated that "Quads, however, seemed to be too large, in general, for all four members to maintain high levels of interactivity with either the program or with other members of the group (p. 12)". Bellows (1987) reported that four of 11 groups of three students each produced the "odd man out" syndrome. Okey and Durnin (1984) found that groups of three or four had lower interactions than groups of two. Webb's review (1989) of non-CBT small group learning studies also offers support for a group size of two. Overall, the limited amount of research in this area suggests that no more than three students can work effectively as a team in CBT and that two students may be the ideal number. If group size is two, much of the concern about group aptitude and gender composition is reduced.

In summary, the small group CBT literature is in agreement on the following:

(1) While small group CBT has not shown the same achievement gain and reduction in training time witnessed in individual CBT, small group CBT achievement is never significantly less than individual CBT.

(2) To permit group effects to occur and to properly measure these effects, experimental sessions should be of sufficient length, and at least one performance-based measure should be taken.

(3) Small groups should have either two or three students.

(4) Groups should probably be like-gender. The literature is inconclusive on group aptitude composition.

(5) If group size is two, group composition decisions are not as influential as when group size is larger than two.

SMALL GROUP CBT VARIABLES WHICH SHOULD BE RESEARCHED

Before discussing variables which could possibly affect achievement in small group CBT, several points should be discussed. First, a safe assumption is that none of the paired CBT studies had software written especially for small groups. Also, rarely do we know the aptitude focus of the CBT software used; e.g., we do not know whether the software was targeted for high ability students, low ability students, etc. Second, the physical arrangement of the paired students was not defined in the articles. A natural assumption would be that two or three students simply sat around a terminal configured for one operator and made do. Third, in most of the studies there were no instructions reported as to how the pairs were told to interact, were they told; who should use the keyboard, who should sit where, etc. And fourth, the role of the instructor was not operationally defined. Rarely does it appear that an attempt was made to re-configure the experimental setting to take advantage of whatever benefits small groups might have. In theory, attending to relevant small group variables should only increase the probability that small group CBT achievement will be higher than individual CBT achievement.

Another background issue concerns the adoption of a small group CBT model. There is a strong desire to simply adopt the work done on small group TI cooperative learning. However, small

group cooperative TI and small group CBT are not interchangeable for several reasons; e.g., the group size limitations imposed by the typical CBT setting, the addition of the computer to the learning setting, the emphasis in cooperative learning on reward structure and group goals, etc. Moreover, the cooperative TI model is not as behavioral as might be desired. Although TI cooperative learning has much to offer CBT, that body of research can not be transposed to CBT in wholesale fashion.

With the above as background, what are the small group CBT variables which could be manipulated to influence achievement? One obvious variable is the aptitude focus taken by the CBT software. Related to the software issue is the question of how do multiple students interact with the computer?

With regard to the first two research questions, it should be noted that students like to touch the monitor while working CBT in pairs. Based on observations by this author, it is common to see one student working the keyboard while the other student points to and touches the screen; i.e., students seem to naturally adopt a peer-tutoring pattern of behavior. If both students are required to keyboard responses to the CBT program, this pattern of behavior may be eliminated, perhaps to the detriment of achievement.

The physical arrangement of the small group CBT setting should be studied, as should group composition. However, the importance of these two dimensions may be relatively small if group size is limited to two students.

The role of the small group CBT instructor should be defined. Johnson and Johnson (1988), major advocates of the cooperative TI learning model, have suggested behaviors for the small group TI instructor; these suggestions may be appropriate for the CBT setting also.

Perhaps the most significant research issue for improving achievement in small group CBT comes from recent TI work by King (1990; 1991a,b) Pressley et al. (1988), Webb (1989) and others. These authors have chosen to focus on those small group member behaviors which seem to increase achievement.

For instance, Webb (1989) found that a student's achievement was related to the level of elaborative help that a student gives to other members of a small group. That is, elaborative help (e.g., how to solve a problem) promoted achievement in the help-giver, but non-elaborative help (e.g., providing facts) did not increase achievement.

Pressley et al. (1988) found that when students answered "why" questions (e.g., "Why do women have more surgeries than men?") about to-be-learned facts, they learned the facts better than students who did not answer why questions. Evidently, the why question guided the student's attention to what needed to be learned. In a sense, students were offered an elaboration before-the-fact and then asked to acquire specific support for the elaboration from material to be presented.

King (1990, 1991a,b) trained students to create very specific elaborative questions during review sessions. Students

would then both offer their questions to other members of the small group to answer and also help answer similar questions created by other group members. Students who used this technique had higher achievement than did students who just discussed the material in an unstructured manner. King proposed that the process of elaborating on and explaining material in a social context of reciprocal peer-questioning caused the material to be individually constructed; this knowledge individualization led to higher achievement.

King also proposed that the critical factor in the reciprocal peer-questioning process is the stem of the question; i.e., the proper question stem will promote peer responses which are both highly elaborated and effective. Students trained to ask these types of guided questions outperformed unguided questioners. Guided questioners both gave and received more explanations than unguided questioners and also gave fewer low-level responses (e.g., giving pat answers or facts). Examples of this type of questions are (King, 1991b):

"How are ... and ... alike?

"What is the main idea of ...?"

"What are some possible solutions for the problem of ...?"

The ramifications of the reciprocal peer-questioning model for CBT are enormous, simply because it suggests how small group CBT members should behave to maximize achievement. Within the CBT framework, this type of questioning could be build into (and initially taught by) the computer, as well as being made part of

the students' behavior. For instance, at the end of the first training sessions, the software could offer a guided question stem for one of the students to complete and then present to another student to answer. Question developer and question answerer roles could rotate. As the students progress through the lessons, the creation of appropriate questions could be turned completely over to the students. The computer could record both questions and answers, thereby providing the instructor with data to monitor learning.

This model incorporates two dimensions which seem to affect achievement. First, the model shows the individual small group CBT student how to behave to maximize achievement: develop and answer elaborative questions. Second, the model requires socialization. In a sense, the model assumes that human socialization is an important aspect of learning and then structures that socialization into a form most conducive to learning in the small group CBT format.

Consequently, the most important small group CBT research issue is how to incorporate the reciprocal peer-questioning model into both CBT software and CBT student behavior. This model also provides an umbrella which covers several of the other research questions proposed earlier. For instance, the role of the instructor, group size, and the focus of the software are more easily addressed once the reciprocal peer-questioning model is adopted.

Distance Learning

At the present time, there is no accepted definition of distance learning (DL) except that the student be physically separated from the instructor. CBT could certainly fit or be configured to fit this definition. If CBT is to be treated as DL, a basic distinction must be addressed. Does the physical separation inherent to DL mean that the students are together but at a location separate from the instructor? Or does it imply that the students are also physically separated from each other? If the former is the case, then much of this paper's discussion on small group CBT applies to DL also. However, if the students are themselves separated from each other, then some of this paper's discussion on small group CBT may not apply to DL.

It would appear logical to assume that two students working at a computer in the back of the classroom (a typical small group CBT environment) is very similar to two students working at a computer in another classroom, as in distance learning. The main difference would be the potential for face-to-face student-instructor interaction. In the first case, the students would know that there is a possibility for face-to-face interaction with the instructor, while in the second case face-to-face interaction between the instructor and the student is not possible. Because there is no way to have the instructor be physically located at the DL site, this difference can not be eliminated. Therefore, the reciprocal peer-questioning model should be tested in both settings: instructor present and

instructor absent.

However, the applicability of the peer-questioning model for distance learning is less certain for the situation in which team members are physically separated from each other. For instance, one requirement for cooperative TI learning seems to be considerable face-to-face interaction between the group members (Johnson & Johnson, 1988). If group members are physically separated from each other, electronic interaction may or may not equate to face-to-face interaction. Also, physically separated group members necessitates that both have keyboards, monitors, etc; such an arrangement eliminates the possibility of one member pointing to the screen while the other member's keyboards, although voice and visual interactions could still occur through appropriate links. Consequently, reciprocal-peer questioning could still take place, but it would not be face-to-face questioning. The proposed model should obviously be tested in the physically-separate distance setting also.

In summary, the reciprocal peer-questioning model holds promise for increasing achievement in the small group CBT setting. I plan to submit a 1993 Research Initiation Program (RIP) proposal to initiate this line of research.

REFERENCES

- Bellows, G. P. (1987). What makes a team? The composition of small groups for C.A.I. A paper presented at the annual meeting of the American Educational Research Association, April 20-24, Washington, D. C.
- Carrier, C. A., & Sales, G. C. (1987). Paired versus individual work on the acquisition of concepts in a computer-based instructional lesson. Journal of Computer-Based Instruction, 14, 11-17.
- Cox, D. A., & Berger, C. F. (1985). The importance of group size in the use of problem-solving skills on a microcomputer. Journal of Educational Computing Research, 1(14), 459-468.
- Dalton, D. W. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. Journal of Computer-Based Instruction, 17(1), 8-16.
- Dick, W. (1963). Retention as a function of paired and individual use of programmed instruction. Journal of Programmed Instruction, 11(3), 17-23.
- Dossett, D. L., & Hulvershorn, P. (1983). Increased technical training efficiency: Peer training via computer-assisted instruction. Journal of Applied Psychology, 68(4), 552-558.
- Hmelo, C. E. (1989/1990). Computer-assisted instruction in health professions education: A review of the literature. Journal of Educational Technology Systems, 18(2), 83-101.
- Hooper, S., & Hannafin, M. J. (1988). Cooperative CBI: The effects of heterogeneous versus homogeneous grouping on

- the learning of progressively more complex concepts. Journal of Educational Computing Research, 4, 413-424.
- Hooper, S., Ward, T. J., Hannafin, M. J., & Clark, H. T. (1989). The effects of aptitude composition on achievement during small group learning. Journal of Computer-Based Instruction, 16(3), 102-109.
- Johnson, D. W., & Johnson, R. T. (1988). Cooperative learning and the computer. In J. Ellis (ed.), AETS Yearbook: Information Technology and Science Education (pp 163-172). Columbus, OH: Association for the Education of Teachers in Science.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. American Educational Research Journal, 27, 664-687.
- King, A. (1991a). Effects of training in strategic questioning on children's problem-solving performance. Journal of Educational Psychology, 83(3), 307-317.
- King, A. (1991b). Improving lecture comprehensive effects of a metacognitive strategy. Applied Cognitive Psychology, 5(4), 331-346.
- Makuch, J. R., Robillard, P. D., & Yoder, E. P. (1991/1992). Effects of individual versus paired/cooperative computer-assisted instruction on the effectiveness and efficiency of an in-service training lesson. Journal of Educational Technology Systems, 20(3), 199-208.
- Okey, J. R., & Majer, K. (1976). Individual and small-group learning with computer-assisted instruction. Audiovisual

Communication Review, 24(1), 79-86.

Pressley, M., Symons, S., McDaniel, M. A., Snyder B. L., &

Turnure, J. E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. Journal of Educational Psychology, 80(3), 268-278.

Shlechter, T. M. (1990). The relative instructional efficiency of group versus individual computer-based training. Journal of Educational Computing Research, 6(3), 329-341.

Stephenson, S. D. (1992). The effects of student-instructor interaction and paired/individual study on achievement in computer-based training (CBT). Journal of Computer-Based Instruction, 19(1), 22-26.

Trowbridge, D. & Durnin, R. (1984). Results for an investigation of groups working at the computer. Irvine, CA: University of California, Educational Technology Center. (ED 238 724).

Underwood, G., McCaffrey, M., & Underwood, J. (1990). Gender differences in a cooperative computer-based language task. Educational Research, 32(1), 44-49.

Webb, N. M. (1989). Peer interaction and learning in small groups. International Journal of Educational Research, 13, 21-39.

Yeuh, J. & Alessi, S. M. (1988). The effects of reward structure and group ability composition on cooperative computer-assisted instruction. Journal of Computer-Based Instruction, 15, 18-22.

ARTERIAL COMPLIANCE AND TOTAL PERIPHERAL
RESISTANCE FOR VARYING +GZ FORCES;
VENTRICULOARTERIAL COUPLING FOR MAXIMUM LEFT
VENTRICULAR WORK; A LATEX TUBE MODEL OF THE AORTA AND
NONLINEAR PRESSURE EFFECTS IN SHORT ARM CENTRIFUGES

Richard D. Swope
Professor
Department of Engineering Science

Trinity University
715 Stadium Drive
San Antonio, Texas 78212

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base, San Antonio, Texas

September 1992

ARTERIAL COMPLIANCE AND TOTAL PERIPHERAL
RESISTANCE FOR VARYING +GZ FORCES;
VENTRICULOARTERIAL COUPLING FOR MAXIMUM LEFT
VENTRICULAR WORK; A LATEX TUBE MODEL OF THE AORTA AND
NONLINEAR PRESSURE EFFECTS IN SHORT ARM CENTRIFUGES

Richard D. Swope
Professor
Department of Engineering Science
Trinity University

Abstract

Four projects are considered. (1) A method for computing beat to beat arterial compliance and total peripheral resistance for nonsteady conditions is presented. Data for a baboon-centrifuge run for +GZ ranging from 1.4 to 8g are analyzed and discussed. Total peripheral resistance is found to first decrease with the onset of +GZ and then increase after about 4 to 6 seconds. (2) An expression for the ratio of arterial elastance to end-systolic elastance which maximizes left ventricular external work is given. This optimum ratio depends significantly on diastolic pressure and the fraction of stroke volume which flows through the peripheral resistance during ejection. It decreases with increasing diastolic pressure and increases with increasing peripheral flow. (3) Capillary tube bundles are examined as possible candidates for use as terminal resistors in two element windkessel to be used in conjunction with a branching latex tube model of the aorta. When used with a sudden contraction orifice plate ($d/D = 0.2$), the resistance is a linear increasing function of flow rate but with $d/D = 0.4$ the resistance is nearly independent of flow rate in the physiological range. (4) In centrifuges the "hydrostatic" pressure distribution is a non linear function of distance. In short arm centrifuges ($R = 5$ ft) the nonlinear contribution to the pressure is from 20 to 26% of the linear part for +GZ ranging from 3 to 9g. In a long arm (20 ft) centrifuge the contribution ranges from 5 to 6%.

ARTERIAL COMPLIANCE AND TOTAL PERIPHERAL
RESISTANCE FOR VARYING +GZ FORCES;
VENTRICULOARTERIAL COUPLING FOR MAXIMUM LEFT
VENTRICULAR WORK; A LATEX TUBE MODEL OF THE AORTA AND
NONLINEAR PRESSURE EFFECTS IN SHORT ARM CENTRIFUGES

Richard D. Swope

INTRODUCTION

Four studies are presented: 1) Determining arterial compliance (C_{AO}) and total peripheral resistance (TPR) for non-steady cardiac outputs; 2) Determining the optimum value of the ratio of arterial elastance (E_a) to end-systolic elastance (E_{ES}) for maximizing the external work done by the left ventricle; 3) Characterization and design of a latex tube impedance module model for the aorta, and 4) Evaluation of the "hydrostatic" pressure distribution in short arm centrifuges. For the readers convenience these studies are presented in sequence starting with a complete discussion of 1 followed in order by discussions of 2, 3 and 4. Projects 1 and 2 are also discussed by Dan Ewert (coinvestigator) elsewhere in this publication (see pages 13-1 to 13-17).

1. Introduction

When the left ventricle contracts blood enters the aorta. Some of this blood effectively goes into expanding the volume of the arterial system (mostly the aorta) and some passes through the peripheral vasculature. The aorta acts as a capacitor storing energy during systole and releasing it during diastole. Under steady state conditions the net volume change over any single beat is zero and TPR is classically determined by:

$$TPR = \frac{P_{AO}}{Q_{AO}}$$

where P_{AO} is average aortic pressure and Q_{AO} is average cardiac output.

C_{ao} is then determined by a variety of methods (see, for example, Burattini, et.al. 1989, Campbell, et.al. 1990, Stordahl, et.al., 1990) When the aortic pressure and stroke volume are non steady these classical approaches do not work

(Toorop, et.al. 1987). The primary difficulty lies in the fact that in these cases the net volume change of the aorta over any given beat is not equal to zero. While the aorta is unloading, the average rate of flow of blood through the peripheral resistance is greater than the flow into the aorta. While the aorta is loading the reverse is true. Exposure to a sudden increase in +GZ forces and a sudden return to normal +GZ causes both of these events to occur in sequence.

The analog model shown in figure 1 is adequate for developing a method for determining C_{ao} and TPR. Using Kirchoff's current law (KCL) at node 1 we get

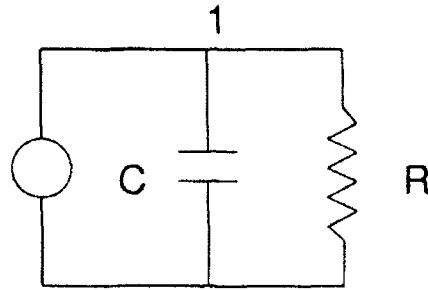


Figure 1

$$i = C \frac{dv}{dt} + \frac{V}{R} \quad (1)$$

or in cardiovascular terms

$$I_{ao} = C_{ao} d(P_{ao} - P_{pleural}) / dt + (P_{ao} - P_{ra}) / TPR \quad (2)$$

where I_{ao} is the entering aortic flow rate, P_{ao} aortic pressure, $P_{pleural}$ - pleural pressure, and P_{ra} - right atrial pressure. $C_{ao} d(P_{ao} - P_{pleural}) / dt$ represents that portion of aortic flow which initially goes into charging the aortic capacitance and $(P_{ao} - P_{ra}) / TPR$ is the peripheral resistive flow. $(P_{ao} - P_{ra})$ then, is the drop

in pressure occurring across the entire lumped peripheral resistance. P_{ra} is equivalent to P_{vein} in the model, corrected for hydrostatic offset. In equation 2, C_{ao} and TPR are unknowns. If we make the assumption that C_{ao} and TPR are constant over a cardiac cycle, then we can divide each beat into two different intervals and integrate equation 2 over each, giving two equations in two unknowns:

$$\int_{t_0}^{t_1} I_{ao} dt = C_{ao} \left[\int_{t_0}^{t_1} dP_{ao} - \int_{t_0}^{t_1} dP_{pleural} \right] + 1/TPR \left[\int_{t_0}^{t_1} (P_{ao} - P_{ra}) dt \right] \quad (3)$$

and

$$\int_{t_1}^{t_2} I_{ao} dt = C_{ao} \left[\int_{t_1}^{t_2} dP_{ao} - \int_{t_1}^{t_2} dP_{pleural} \right] + 1/TPR \left[\int_{t_1}^{t_2} (P_{ao} - P_{ra}) dt \right] \quad (4)$$

1. DISCUSSION OF RESULTS

Figures 2-A, B and C show the TPR results for baboon/centrifuge runs with 10 second exposures to 3, 6 and 8g respectively. TPR is represented by the square symbols and is plotted on the left y-axis. Mean aortic pressure (P_{AO}) stroke volume (SV) and +GZ level are represented by solid lines from top to bottom respectively and plotted on the right y-axes. The mean pressure signals are obtained by low pass filtering the time varying aortic pressure signal and the fall-offs in (P_{AO}) seen at the start and end of the runs are due to data analysis procedures. Figure 3-A shows +GZ rising from 1.4 to 3g at 6 seconds and remaining constant until 16 seconds when it returns to 1.4g. Prior to the onset of +GZ TPR has a mean value of about 3500 Dyne-sec/cm⁵ and is steady. The scatter about this mean includes experimental and computational errors. When the +GZ forces are applied to the blood there is a volume shift from the upper body to the lower body and a rise in the lower body distending pressure. A falling

pressure is experienced by the upper body baroreceptors but the response time of the peripheral resistance to baroreceptor input is from 4 to 6 seconds. The net result of the distending pressure increase and this time delay is an initial reduction in TPR beginning with +GZ onset. Baroreceptor influence is seen to begin at about 10 seconds (4 seconds after +GZ onset) and TPR increases. It reaches its "prior to +GZ" level at about 12 seconds at which time the baroreceptors are exposed to normal pressures. TPR however, continues to increase for about 4 seconds before it begins its descent to normal values. The overshoot is due to the 4 to 6 second time delay of the baroreceptor-system response. Note that the mean aortic pressure follows the same time courses while the stroke volume does not. The stroke volume begins to drop about 1 to 2 seconds after the onset to +GZ and remains low until the +GZ is returned to 1.4g. The fact that TPR and mean aortic pressure begin to rise well before stroke volume rises strongly suggests that hydrostatic pressure effects on TPR and baroreceptor responses are the main causes of TPR and mean aortic pressure variations. Figures 2 B and C show similar results for 6g and 8g.

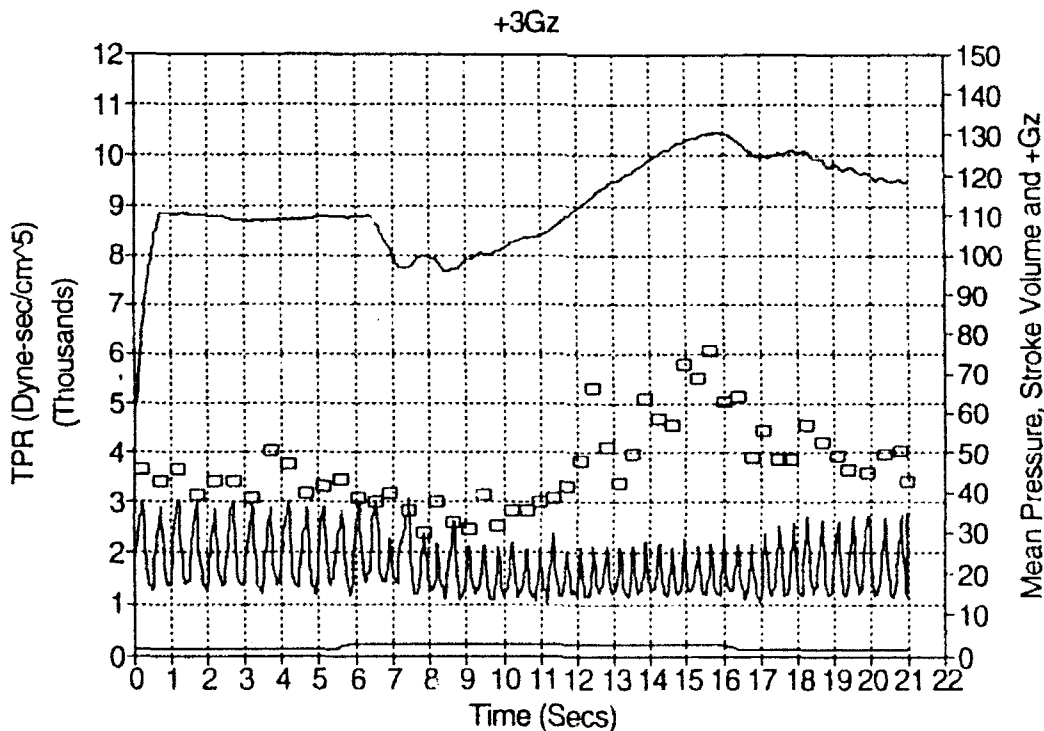


Figure 2-A

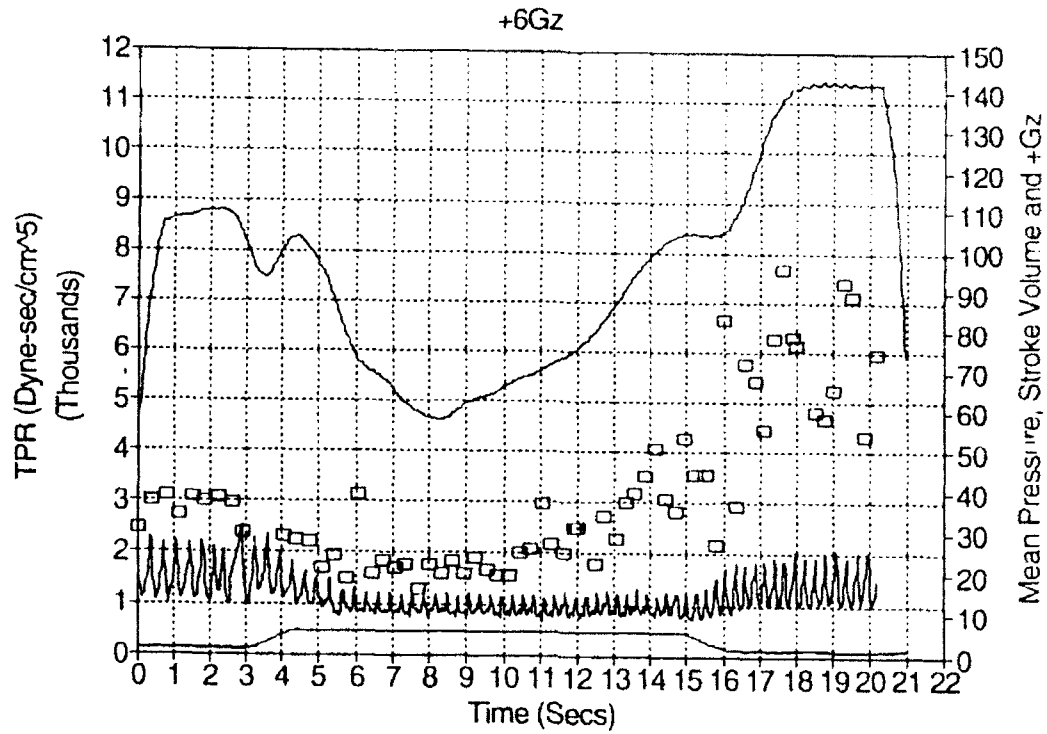


Figure 2-B

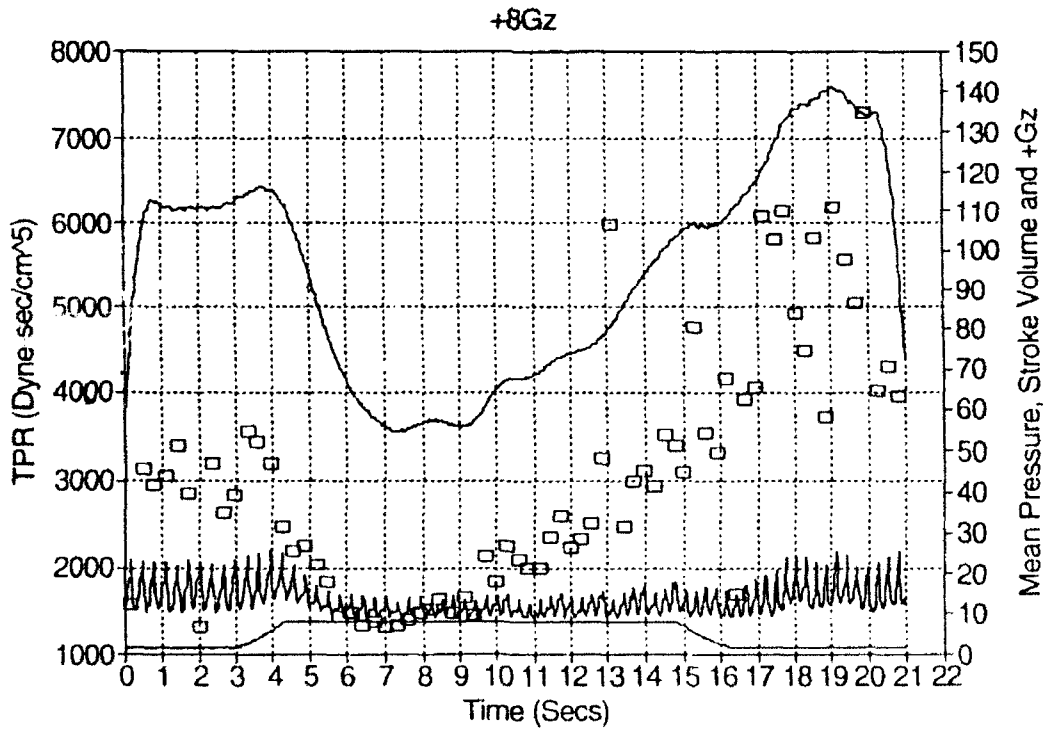


Figure 2-C

For the higher +GZ levels the drop in mean aortic pressure is larger as is the drop in TPR. This is most likely due to the correspondingly higher distending pressure and hence larger peripheral vessel lumina. Four to six seconds after the onset of +GZ the baroreceptor influence is seen and both TPR and mean aortic pressure begins to increase. As in the +3GZ case there is an overshoot, lasting about 4 seconds.

Figures 3 A, B and C show the results for the arterial compliance. There is considerable scatter in the C_{ao} results so it is more difficult to discern what is going on. For the first and second cases C_{ao} begins at about 0.3 cc/mm-Hg for +GZ=1.4. With the onset of increased +GZ there is too much scatter to determine a trend but there does appear to be a slight decrease in C_{ao} when the mean aortic pressure overshoots its 1.4g level. At 8g the scatter is still significant but there does appear to be a slight decrease in C_{ao} at this high +GZ level. This decrease may be due to the longitudinal stretching of the ascending aorta as the heart and blood volume are pulled toward the feet.

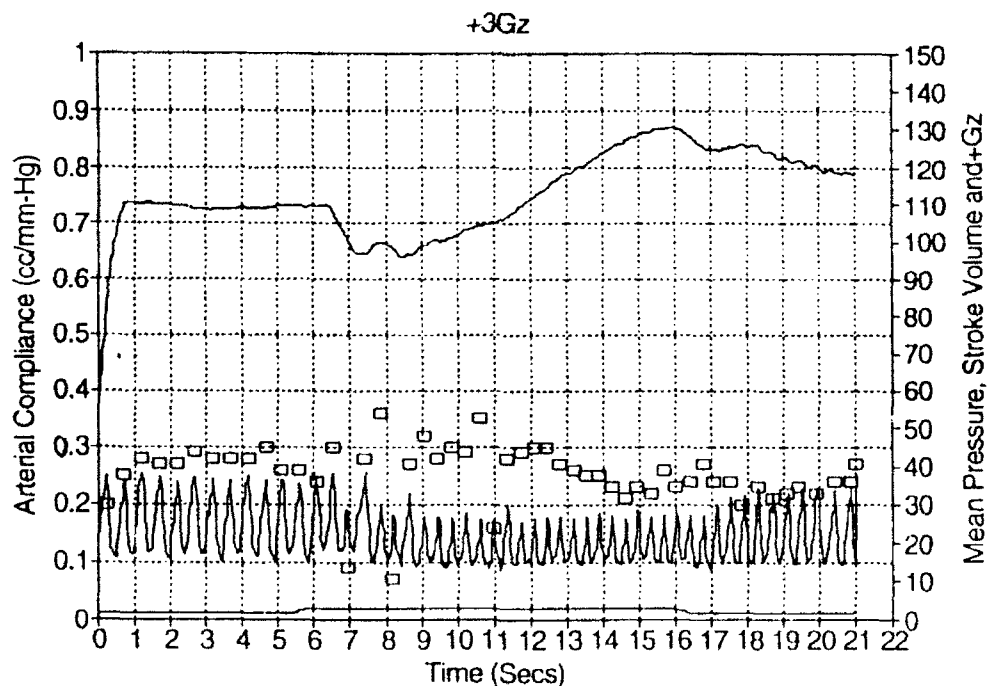


Figure 3-A

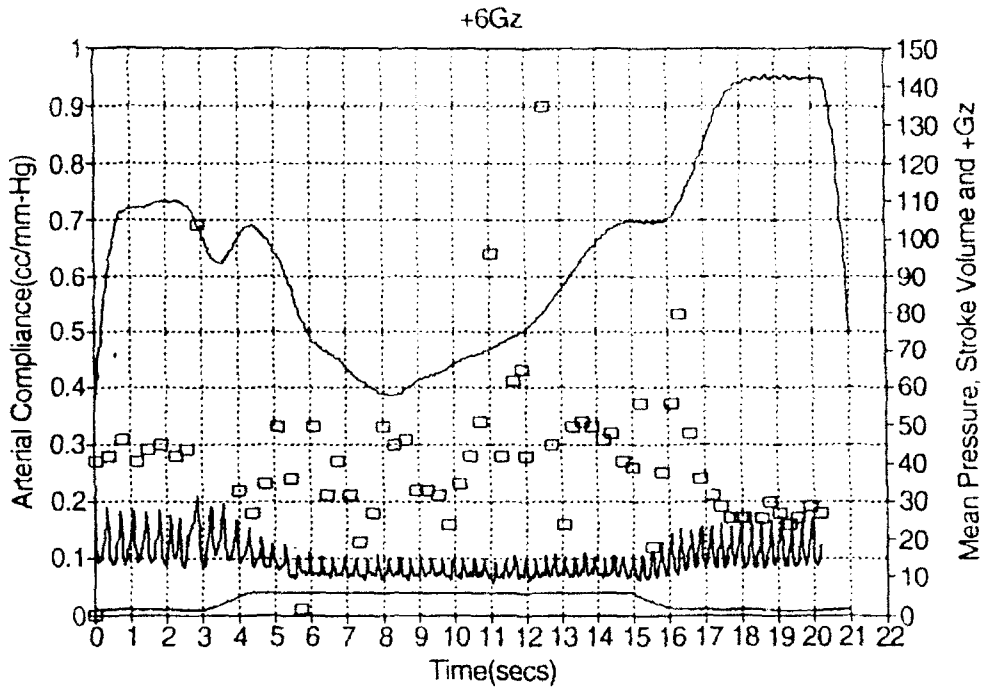


Figure 3-B

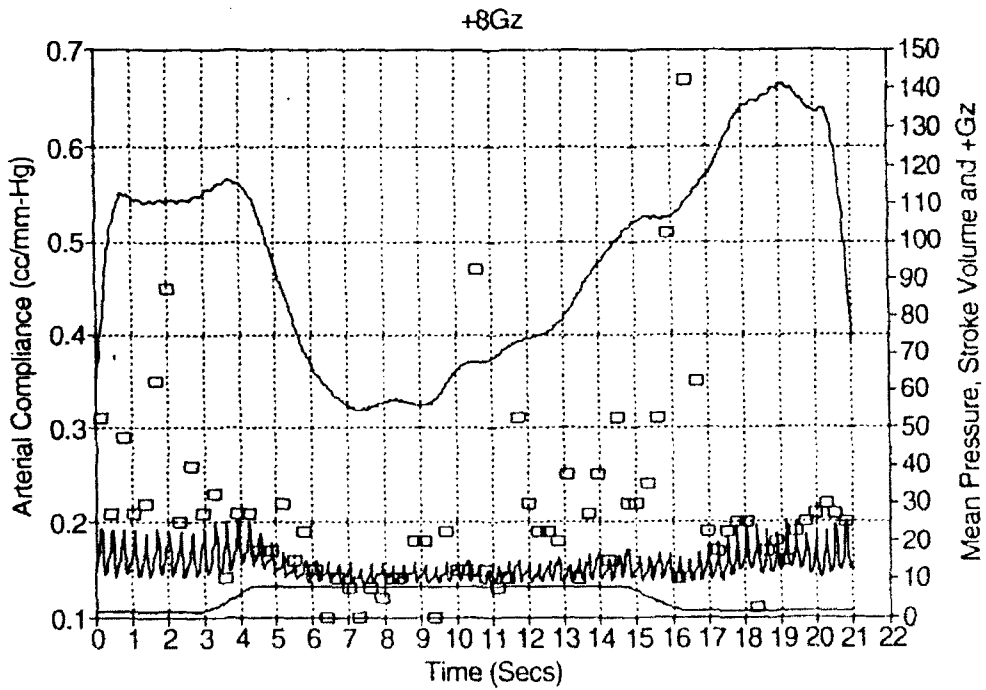


Figure 3-C

C_{ao} values determined by our method are sensitive to interval $t_0 - t_2$ and $t_1 - t_2$ choices while TPR values are not. More work is required to determine a better way to calculate C_{ao} under these harsh non-steady conditions.

2. INTRODUCTION

The left ventricle of the heart is the source of flow energy which propels blood through the systemic circulatory system. Its efficiency depends on the contractility properties of both the ventricle and the arterial system. End-systolic elastance (E_{es}) and arterial elastance (E_a) represent measures of the contractility state of the ventricle and arterial system respectively. E_{es} is defined as the slope of the P_{max} line (see figure 4) and E_a as

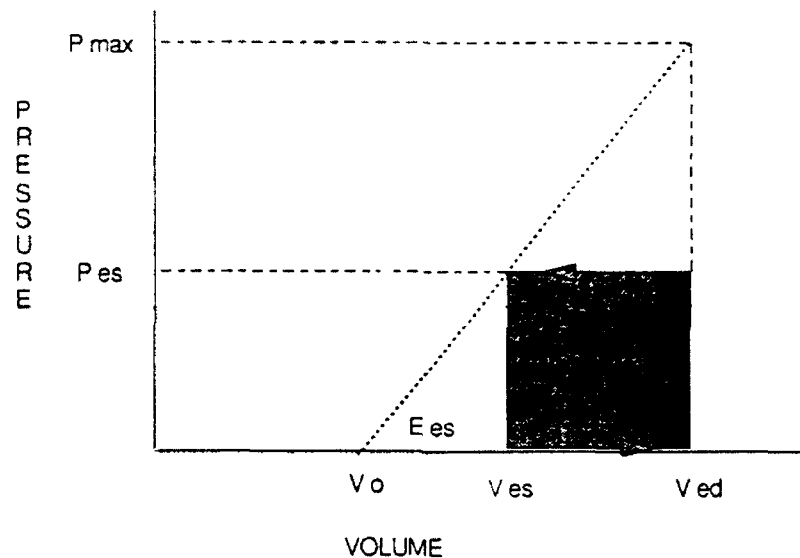


Figure 4

$$E_a = \frac{P_{es} - P_d}{V_c} \quad (5)$$

where P_{es} is end systolic pressure, P_a is diastolic pressure and V_c is the volume change in the arterial system during ejection. It should be noted at this point that other authors (Sunagawa, et.al. 1983; Sunagawa, et.al. 1985 and Kubota, et.al. 1992) use an effective arterial elastance in considering

ventriculoarterial coupling which is different from the one used above. The reader is referred to Sunagawa, 1983 for a detailed discussion of effective arterial elastance which is not the physical elastance of the arterial system. Methods for determining Ees and Pmax can be found in Takeudu, et.al. 1991.

2. Model

The shaded area in the flow loop of figure 4 represents the energy transferred from the left ventricle to the blood and is equal to the external work. Assuming isobaric contraction at Pes this work is given by

$$EW = P_{es} * SV \quad (6)$$

where SV is the stroke volume ($V_{ed} - V_{es}$).

Now Pes and SV are functions of Ees and Ea:

$$P_{es} = E_{es}(V_{e_d} - SV - V_0) \quad (7)$$

$$SV = \frac{1}{E_a} (P_{es} - P_d) + T_{ej} \frac{\bar{P}_a}{TPR} \quad (8)$$

where V_0 is defined in figure 4, T_{ej} is the time for ejection, and \bar{P}_a is mean aortic pressure during ejection minus venous pressure. These relationships are substituted into equation 6 giving external work as a function of the contractility properties E_a and E_{es} and measurable pressures and volumes. Adequate cardiac output for peripheral perfusion can be achieved by multiple combinations of E_{es} , E_a and loading conditions, but it is likely that the cardiovascular control system operates in a way which optimizes ventricular performance. The question addressed in this study: Is there a value of E_a/E_{es} which maximizes external work as given above? Setting $\alpha = E_a/E_{es}$ and setting the first derivative of the external work with respect to α equal to zero yields

$$\alpha_{opt} = \frac{\left(\frac{V_R}{V_{max}} - 1\right) + \frac{Pd}{P_{max}} \left[3 - 2\left(\frac{V_R}{V_{max}} + \frac{Pd}{P_{max}}\right)\right]}{\left(1 - 2\frac{V_R}{V_{max}}\right) \left(\frac{Pd}{P_{max}} + \frac{V_R}{V_{max}} - 1\right)} \quad (9)$$

where $V_{max} = V_e - V_o$ and

$$V_R = \frac{Tej\bar{P}a}{TPR} \quad (10)$$

If $V_R \ll V_{max}$ a simplified version of the optimum value of α (α_{opt}) can be found as:

$$\alpha_{optS} = \frac{1 - 2Pd/P_{max}}{1 - VREes/(P_{max} - Pd)} \quad (11)$$

2. Discussion of Results

Figs. 5 A, B, and C show the optimum values of E_a/E_{es} as a function of the volume of blood which flows through the peripheral resistance during the ejection period for various values of E_{es} . The value of P_{es} is 120 MM-Hg and the stroke volume is taken as 100cc. For $E_{es} = 0.5$ mm-Hg/cc the exact and $V_R \ll V_{max}$ solutions are very close for all values of V_R . However, for larger values of E_{es} the two solutions diverge as V_R increases. These results are consistent with the $V_R \ll V_{max}$ assumption. At relatively low values of V_R i.e. $V_R < 30\% SV$, the optimum ratio is approximately a slowly increasing linear function of V_R ranging from about 0.25 at $V_R = 0$ to about 0.35 at $V_R = 30$ cc. At higher V_R flows the optimum ratio is very sensitive to changes in V_R especially for relatively noncompliant (high E_{es}) systems.

Figure 5D (compared with 5B) shows the effect of increasing diastolic pressure. All other parameters are the same for both figures. α_{opt} is very much influenced by P_d especially at higher values of V_R . At $P_d = 100$ mm-Hg α_{opt} is significantly less than for $P_d = 80$ mm-Hg over the entire range of V_R .

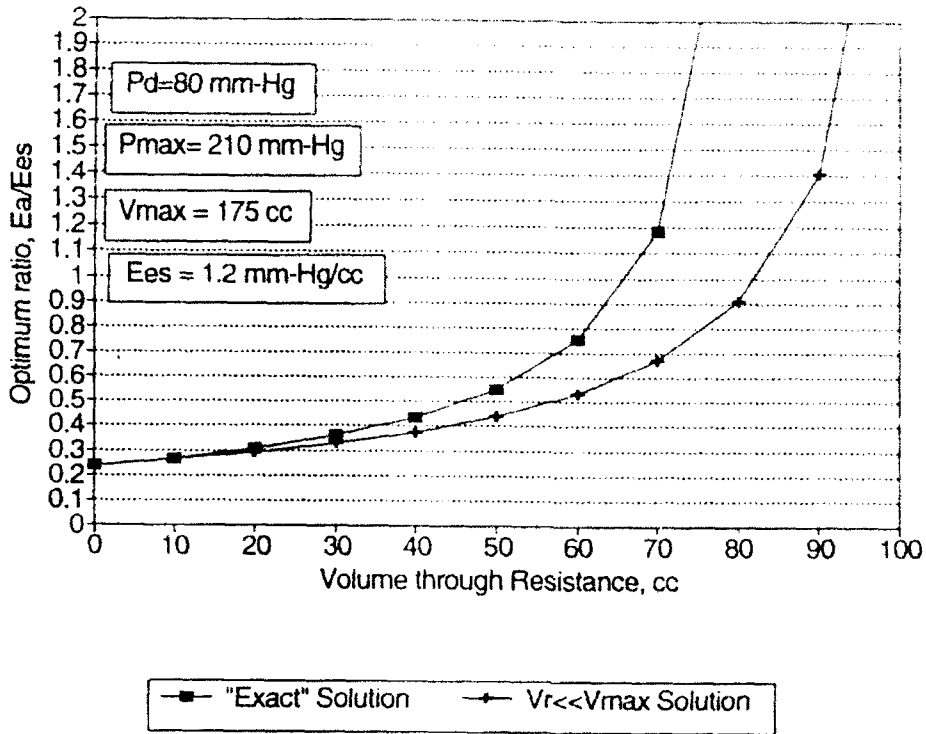


Figure 5-A

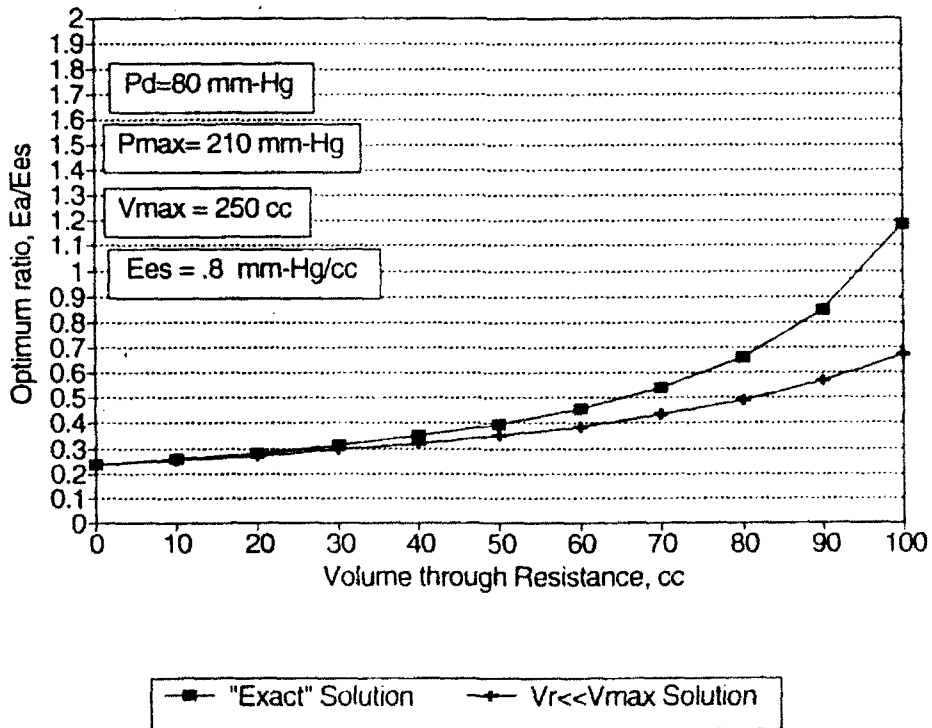


Figure 5-B

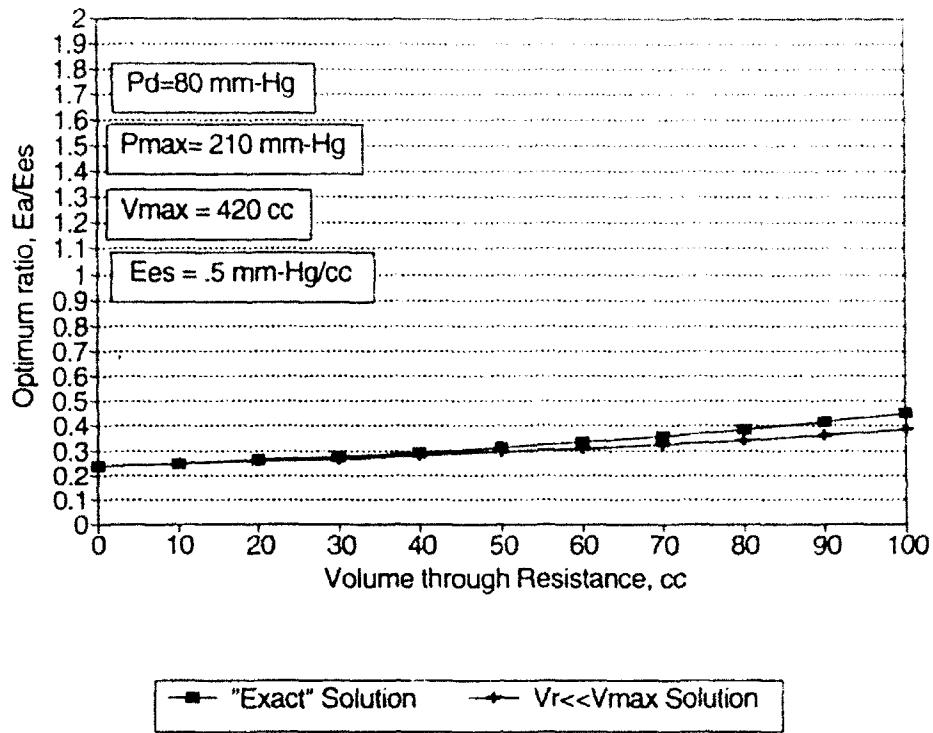


Figure 5-C

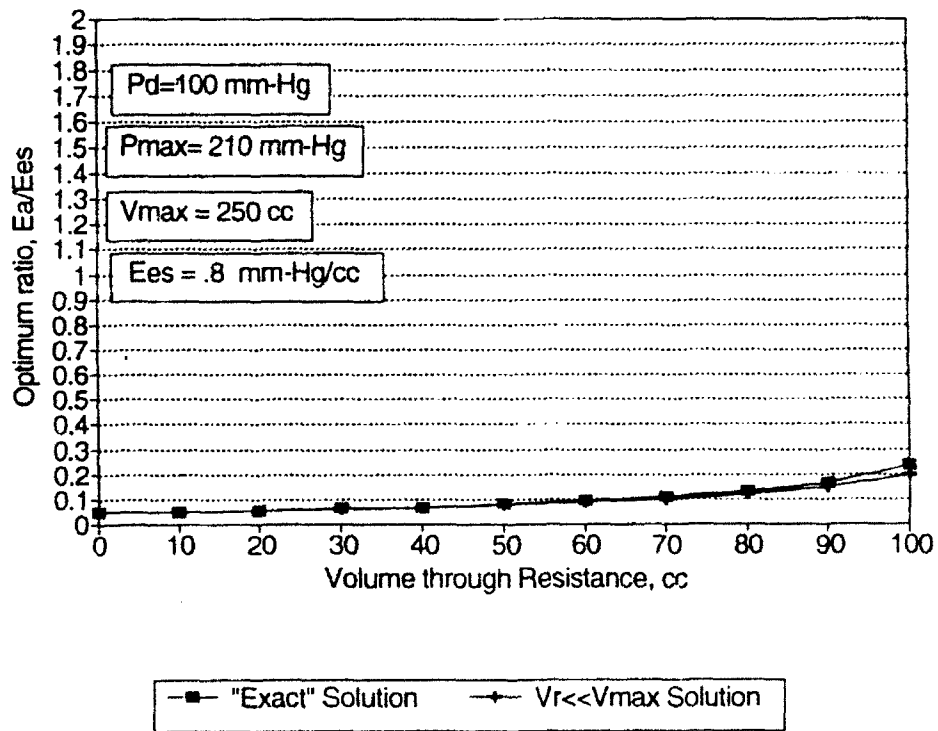


Figure 5-D

3. INTRODUCTION

A latex tube hydraulic model of the aorta is useful in several ways. It can be used as a controllable verification test bed for mathematical/computer models of the aorta as well as for the development and testing of a variety of transducers to measure displacements, velocities and pressures in the aorta. Figure 6 shows a schematic of the latex model used in this study. It consists of a tube with branches representing flows to the head, kidney and legs. Each branch is terminated with a two element windkessel whose input impedance is matched to the anatomical part it represents. The windkessels are made of plexiglass and contain a variable volume air chamber as well as an adjustable resistor. Thus, both the capacitance and resistance are independently adjustable.

3. MODEL

The first step in the development of this system is the characterization of the components. The resistance of the resistors is measured using the constant head flow source system shown in figure 7. A given resistor is placed in the test section and a steady flow rate established. Pressure drop across the resistor is measured and the corresponding resistance is calculated.

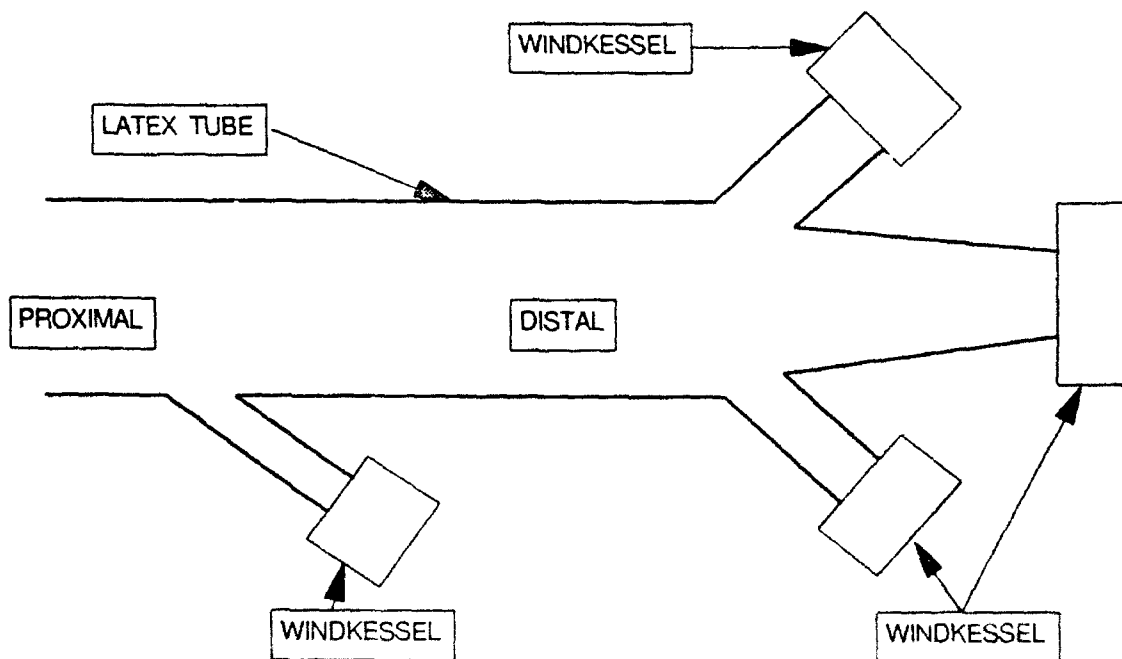


Figure 6

The contractility characteristics of the latex tube are determined by closing off the exit ports of the windkessels and measuring the inside pressure as known increments of water volume are added.

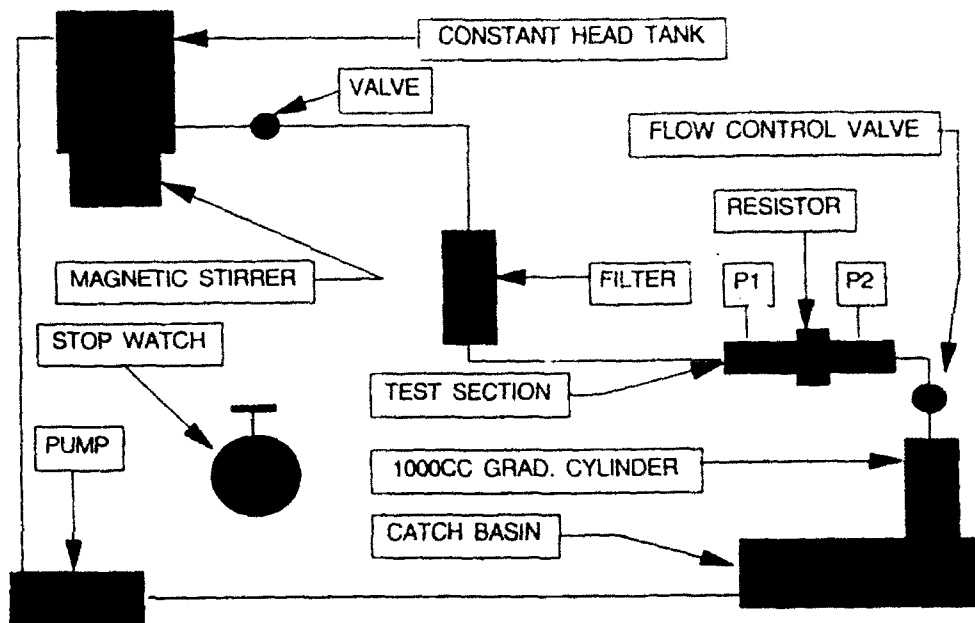


Figure 7

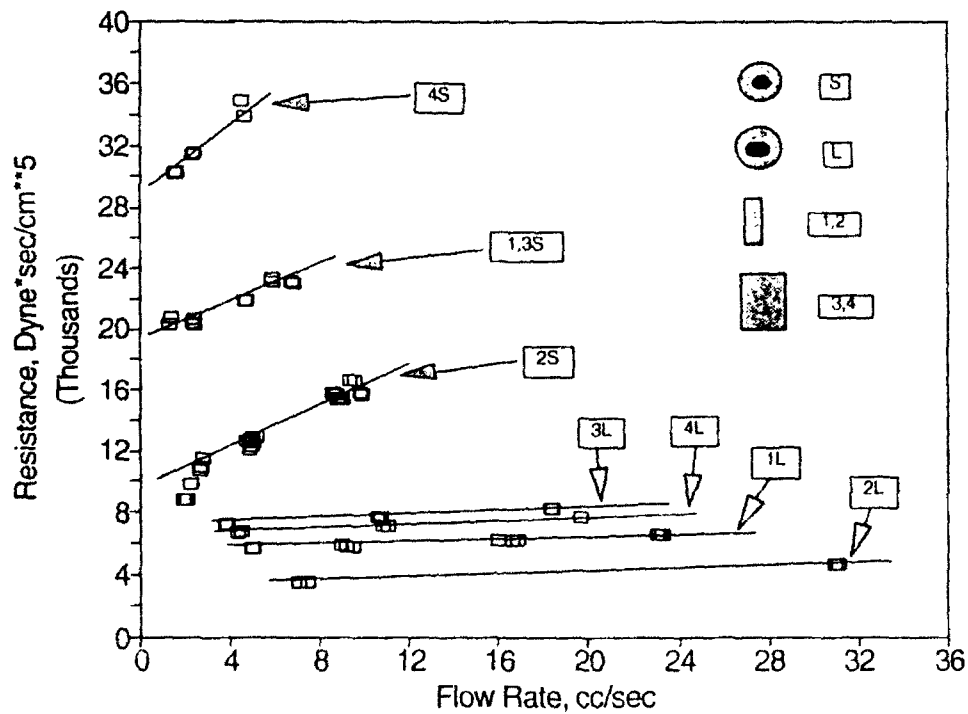


Figure 8

3. Discussion of Results

Figure 8 shows the results of the resistor tests. The resistors are composed of two parts: a short length of approximately 30 micron diameter capillary tubes bundled in parallel and a nylon washer. Resistance is adjusted by placing the washers with different size openings against the open face of the tube bundles thereby allowing more or fewer tubes to transport the fluid. The length of the bundles can also be selected at the time of manufacture. Three bundle lengths are considered: Resistor number 2 is the shortest at 0.192 inches long. Resistor 1 is 0.291, 3 is 0.311 and 4 is 0.391 inches. The washer opening sizes are designated as small, S, and large, L.

With the small hole washer in place the resistance is a linearly increasing function of flow rate, but with the large one the resistance is nearly independent of flow rate. These results are consistent with the flow patterns associated with small and large opening orifice plates (Munson et. al. 1990). It is unlikely that the small hole washer is suitable for our use because it causes the resistance to vary substantially with flow. New capillary bundles with fewer open tubes are being considered as substitutes for future work.

Tube compliance results are shown in Figure 9. There is a hysteresis effect, most likely due to latex creep, in the pressure volume curves. Figure 9D shows the pressure loss over time for a closed tube with various initial filling volumes and pressures. After about 20 seconds there is a very slight linear drop in pressure but during the initial period the drop is significant and non linear. For the entire tube supported by an open cell foam the compliance is about 1.1 cc/mm-Hg (see figure 9A). For the proximal end, figure 9B, (up to just beyond the first branch) the compliance is about 0.6 cc/mm-Hg and the distal value, figure 9C, is about 1.0. This particular latex tube (#8/10/92) is relatively thin and future studies will be conducted with larger tube wall thicknesses of about 1 mm.

4. INTRODUCTION:

Physical deconditioning of astronauts in outer space (0g) environments requires the introduction of some protocol to prevent problems upon return to earth (1g). It has been proposed (Burton, et. al. 1992) that an on board short arm centrifuge could be used for periodic exposure to +GZ levels in flight and with appropriate schedules it might be plausible to prevent some or most of the deconditioning. While this approach might be successful, there are some questions

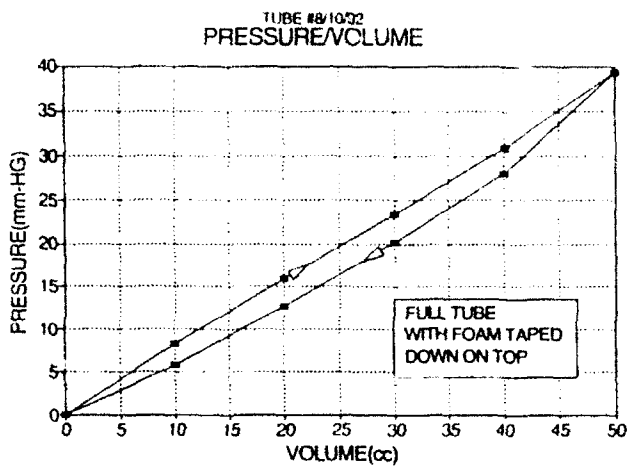


Figure 9-A

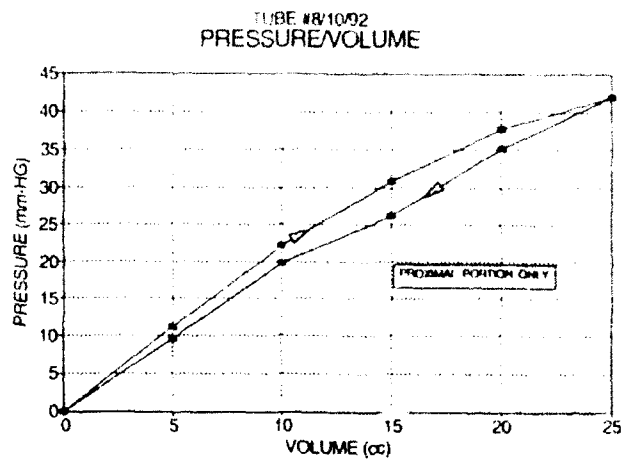


Figure 9-B

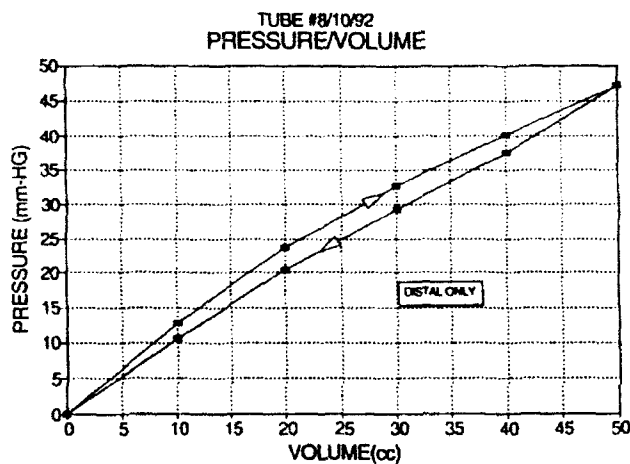


Figure 9-C

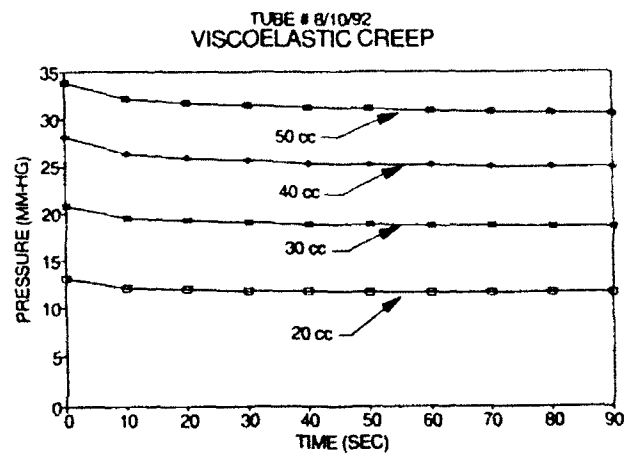


Figure 9-D

that should be answered before proceeding with the construction and testing of a flyable centrifuge.

4. DISCUSSION

On earth, the hydrostatic pressure field established by gravity is linear from head to foot in the standing position. In a centrifuge the +GZ field is produced by the rotational motion and it can be shown (Hansen, 1967) that this in turn causes a non linear pressure field. The nonlinear contribution is small for a 20 foot radius centrifuge, ranging from 4 to 5 % for a +GZ range from 3 to 9g but it is in the 20 to 26% range for a 5 ft radius machine over the same range. Also the proposed seating position on the short arm machine (lying on back with knees drawn up to chest) does not allow the full benefit of rotationally induced +GZ to be realized because the major vascular beds in the legs are very nearly at the same level as the gut. Thus, the hydrostatic "column" is "Z"

shaped and the pressure at knee level is very close to that in the descending aorta. This condition is very different from that of standing in a +GZ field. More work needs to be done to determine if the proposed short arm centrifuge would be beneficial.

REFERENCES

- Burattine, R. and Cambell, K.B., Modified Asymmetric T-Tube Model to Infer Arterial Wave Reflection at the Aortic Root, *IEEE Transactions on Biomedical Engineering*, Vol. 36, No. 8, pp. 805-813.
- Burton, R.R. and Meeker L.J. Physio'ogic Validation of a Short Arm Centrifuge for Space Application, *Aviation, Space and Environmental Medicine*. Vol. 63, No. 6, pp. 476-81.
- Campbell, K.B., Burattine, R., Bell, D.L., Kirkpatrick, R.D., and Knowlen, G.G., (1990), Time-Domain Formulation of Asymmetric T-Tube Model of Arterial System, *American Journal Physiology*, 258, H1776-H1774.
- Hansen, A.G., (1967), Fluid Mechanics, John Wiley & Sons, Inc. New York, New York.
- Kubota, T., Alexander, J., Staya, R., Todaka, K., Sugrmachi, M., Sunagawa, K., Nose, Y., and Takeshita, A., (1992), Dynamic Effects of Carotid Sinus Baroreflex on Ventriculoarterial Coupling Studied in Anesthetized Dogs, *Circulation Research*, 70, pp. 1044-1053.
- Munson, B.R., Young, D.F., and Okiishi, T.H., (1990) Fundamentals of Fluid Mechanics, John Wiley & Sons, Inc. New York, New York.
- Stordahl, S.A., Piene, H., Solbakken, J.E., Rossvoll, O., Samstad, S.O., and Angelson, B.A.J., (1990) Estimation of Arterial Compliance in Aortic Regurgitation: Three Methods Evaluated in Pigs, *Medical Biological Engineering and Computing*, 28, pp. 293-299.
- Sunagawa, K., Maughan, W.L., and Sagawa, K., (1985), Optimal Arterial Resistance for the Maximum Stroke Work Studied in Isolated Canine Left Ventricle, *Circulation Research*, 56, pp. 586-595.
- Sunagawa, K., Maughan, W.L., Burkhoff, D., and Saguawa, K. (1993), Left Ventricular Interaction With Arterial Load Studies in Isolated Canine Ventricle, *American Journal of Physiology*, 245, H773-H780.
- Takeuchi, M., Igarashi, Y., Tomimoto, S., Odake, M., Hayashi, T., Tsukamoto, T., Hata, K., Takaoka, H., and Fukuzaki, H., (1991), Single Beat Estimation of the Slope of the End-Systolic Pressure-Volume Relation in the Human Left Ventricle, *Circulation*, 83, pp 202-212.
- Toorop, G.P., Westerhof, N., and Elyinga, G., (1987) Beat to Beat Estimation of Peripheral Resistance and Arterial Compliance During Pressure Transients, *American Journal of Physiology*, 252, H1275-H1283.

Examination of Response Latencies
to Personality Inventory Items

Lois E. Tetrick
Associate Professor
Department of Psychology

Wayne State University
71 W. Warren
Detroit, MI 48202

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August, 1992

Examination of Response Latencies
to Personality Inventory Items

Lois E. Tetrick
Associate Professor
Department of Psychology
Wayne State University

Abstract

The relations among response latencies and scores on several measures of cognitive information processing tasks and personality measures were examined. The pattern of relations across these measures indicated that response latencies are effected by many aspects of the stimuli presented to individuals. There was no support for a claim that relative response latencies were stable across tasks that might signify such individual differences as impulsivity and intelligence, but rather, they seemed to reflect differential aspects of the task itself, and within subject variability.

Examination of Response Latencies
to Personality Inventory Items

Lois E. Tetrick

Introduction

The ability to capture response latencies in computerized administrations of personnel selection tests has led researchers to question the meaning and usefulness of such data. It has been suggested in the literature that response latencies are indicative of attitude accessibility (Fazio, 1989; Krosnick, 1989), attitude intensity (Markus, 1981), impression management (George & Skinner, 1990), and impulsivity (Dickman & Meyer, 1988), as well as speed of information processing and intelligence (Vernon, 1987). Part of this seeming inconsistency and confusion may be the result of insufficient theoretical and conceptual development of the process involved in responding to different test items ranging from simple reaction time tasks to personality inventories. Use of response latencies to personality items, without appropriate adjustments, may contain variance attributable to factors other than personality thus confounding the measures and obscuring the potential usefulness of this information.

Research investigating response latencies to personality inventories has made little effort to partition the variance into identifiable components of the process of responding to these stimuli. Tetrick (1989) did attempt to control for reading speed and certain item characteristics, but did not have data available that would permit inclusion of such individual characteristics as speed of information processing. It is interesting to note that the measure of reading speed used in this study accounted for less than one percent of the variance in response latency for approximately 40 percent of the participants. Holden and colleagues (1991) also have recognized the need to adjust the raw latencies to remove contamination from individual and item characteristics. They proposed a double standardization procedure that purports to control for both extraneous item and individual characteristics. According to their description of the double standardization procedure, standardization of the latencies within subject across items controls for confounding person variables such as gender and reading speed while the standardization across subjects within item adjusts for the confounding effects of test-item properties such as item length, vocabulary level, and negativity. However, there is no empirical support that this procedure does indeed yield an appropriate adjusted measure of response latency, and it therefore remains controversial.

The purpose of this research project was to examine the effect of various item and personal characteristics on response latencies to personality inventory items.

Methodology

Subjects

The sample consisted of 509 trainees in the USAF Undergraduate Pilot Training (UPT) program. Not all participants had training outcome data or there were missing data for some of the tests; therefore, the sample sizes varied depending on the analyses.

Procedure and measures

Participants took the Basic Attributes Battery (BAT) which is an automated test battery for pilot selection. Item responses and latencies for a subset of the tests included in the battery were merged with UPT outcome and verbal and quantitative AFOQT scores. The specific tests used were the Activities Interest Inventory, Encoding Task, Item Recognition Task, Mental Rotation Task, Time-sharing Task, and the Personality Profiler. A complete explanation of these tasks can be found in AFHRL-TR-87-9.

Activities Interest Inventory. This test required individuals to indicate which activity from each of 81 pairs of activities that they would prefer. Subjects are told to assume that they have the necessary ability to perform each activity. The activities differ with respect to the degree of threat to physical survival. Scores are obtained based on the number of high-risk activities selected by the individual.

Encoding Task. Subjects are presented simultaneously with two letters and are required to make a same-different judgment about the letter pair based on three rules: physical identity (AA versus (Aa), name identity (AA versus AH), and category identity (Vowels versus consonants - AE versus AH). The total task includes 96 trials with 32 trials in each of the rule conditions. The latency of the encoding judgment provides a measure of the speed of the cognitive encoding process. Differences in mean latency for the three rules indicates speed of recoding.

Item Recognition Task. This test assesses individuals' short-term memory. Subjects are presented with a string of one to six digits on the computer screen. The string is then removed and followed, after a brief delay, by a single digit. Subjects respond whether the single digit was one of those presented in the initial string. Speed and accuracy of response are recorded on each of the 48 trials.

Mental Rotation Task. This task assesses spatial transformation and classification ability. Pairs of letters are presented sequentially and subjects are required to make a same-different judgment. The letters in each pair may be either identical or mirror images as well as being in the same orientation or rotated in space with respect to each other. Speed and accuracy of response are recorded on each of the 72 trials.

Time-sharing Task. This test examines an individual's higher-order tracking ability, learning rate, and time-sharing ability as a function of

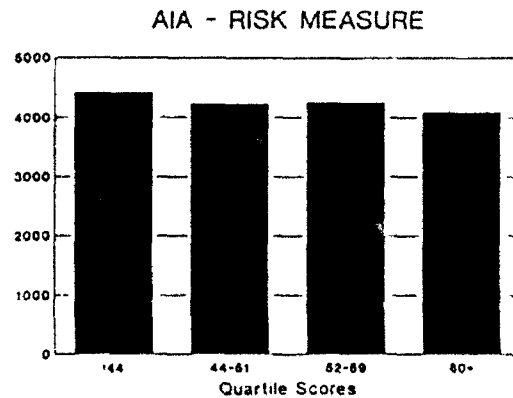
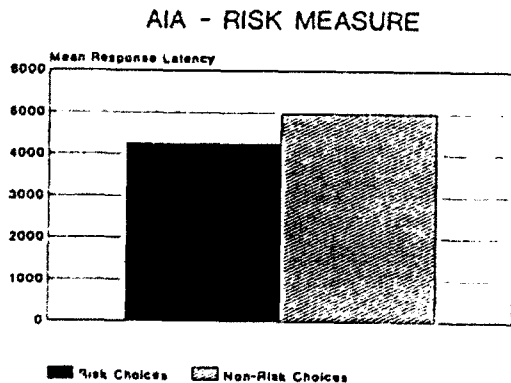
differential task load. Subjects learn a compensatory tracking task which involves anticipating the movement of a marker on a screen and operating a control stick to counteract that movement in order to keep the marker aligned with a fixed central point. Task difficulty is adjusted during the test, depending on the subject's performance on the test. After these "tracking only" trials, the subject is required to track while canceling digits that appear at random intervals and locations on the screen. The information processing load gradually increases during these trials. These dual-task trials are followed by a final block of "tracking only" trials. This study included data from the dual-track tasks only.

Personality Profiler. This test included 202 items mostly drawn from the MMPI. Included were 87 items specifically measuring five dimensions of personality (Costa, Zonderman, McCrae, & Williams, 1985). The five factors were psychoticism, neuroticism, extraversion, inadequacy, and cynicism. There were 13 items on the psychoticism scale, 24 for neuroticism, 12 for extraversion, 13 for inadequacy, and 25 for cynicism. Endorsement/rejection and response latency were recorded for each item.

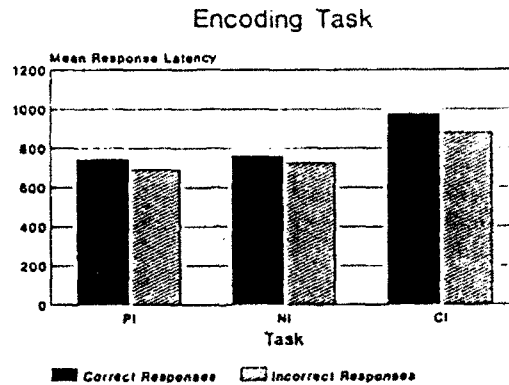
Results

Before examining the relations of response latencies across the tests, response latencies were examined for the dimensions or factors of each test.

Activities Interest Inventory. Overall, subjects chose an average of 50.39 high-risk activities (s.d.=10.75). A comparison of the response latencies for higher-risk choices versus lower-risk choices indicated that respondents took significantly longer to choose lower-risk activities than they did to select the higher-risk activities (means=4986.25 and 4261.59, respectively; $F_{1,503}=305.44, p<.01$) as shown below. In addition, subjects were divided into quartiles based on their total scores and a test of the average response latencies indicated that the highest risk group responded the most quickly and the lowest risk group responded the most slowly as shown below.



Encoding Task. As would be expected, subjects took longer to make judgements under the Category Identity condition than the Name Identity condition with responses being quickest for the Physical Identity rule ($F_{2,718}=112.24, p<.05$). In addition, incorrect responses were quicker than correct responses across all three conditions suggesting possible anticipatory responding ($F_{1,359}=21.89, p<.05$).

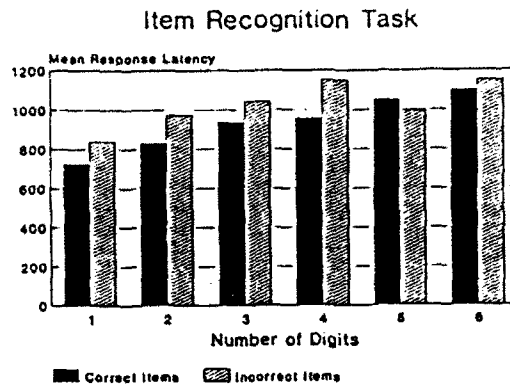


Subjects were grouped based on the number of incorrect responses in the Physical Identity condition. Individuals who missed the most items in the Physical Identity condition were quicker in responding across all conditions than those who missed fewer than 4 items in the Physical Identity condition. Results basically were the same when the groups were formed based on the number of incorrect responses in the Name Identity and Category Identity conditions. However, in the Name Identity condition, the individuals who did not miss any items were slower than the other groups rather than the individuals who missed the most being the quickest.

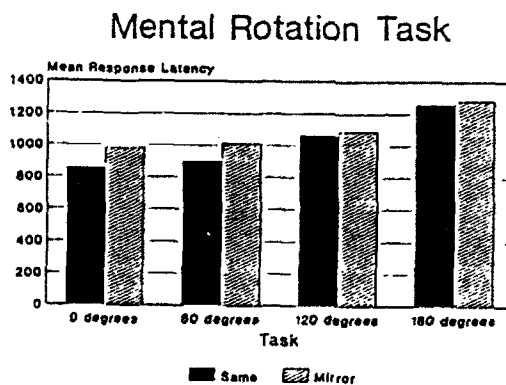
Table 1
Mean Latencies by Accuracy of Response

Number of Incorrect Responses	<u>n</u>	Mean Response Latency		
		<u>PI</u>	<u>NI</u>	<u>CI</u>
0	79	793	802	1024
1	109	765	782	1017
2	98	746	775	992
3	90	753	771	993
4+	124	691	708	987

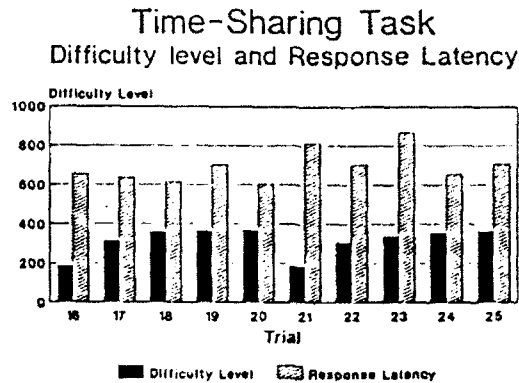
Item Recognition Task. The average response latencies increased as the number of digits in the initial string increased as did the number of incorrect responses ($F_{5,491}=446.19, p<.05$; $F_{5,491}=12.64, p<.05$). Correct responses were made more quickly across all conditions with the exception that when the initial string included 5 digits, incorrect responses were actually quicker.



Mental Rotation Task. Response latencies differed depending on whether the letters were the same or mirror images ($F_{1,491}=130.45, p<.05$) as well as the extent of rotation ($F_{3,491}=552.09, p<.05$). There was a significant interaction as can be seen below ($F_{3,491}=42.74, p<.05$). The source of the interaction was that for rotations of 120 and 180 degrees there was no difference between same and mirror images where for rotations of 0 and 60 degrees, mirror images took longer for the subjects to respond.



Time-sharing Task. The response latencies for digit cancellation with associated difficulty level across trials 16 through 25 are shown below ($F_{13,6409}=187.62, p<.05$). The time to respond differed significantly across trials even when difficulty level was covaried out ($F_{13,6408}=187.05, p<.05$). Longer response latencies were associated with the latter trials.

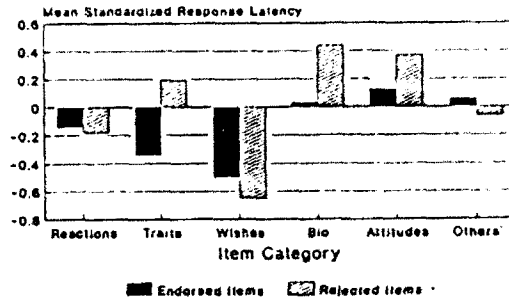


Personality Profiler. Prior to analyzing the response latencies for the personality items, the number of words for each question was computed. For each subject, a regression analysis was performed in which latency was regressed on the number of words in the question to control for reading speed. The standardized residuals for each item based on these analyses were used in all subsequent analyses.

Personality items differ one from another on a lot of dimensions. Angleitner, John, and Lohr (1986) proposed a system for categorizing personality items. Their system includes six major categories: descriptions of reactions, trait attributions, wishes and interests, biographical facts, attitudes and beliefs, and others' reactions to the person. Using their definitions, 6 graduate students sorted all 202 personality items into the categories. The mean response latencies for endorsed and rejected items in each of these categories are shown below. There were significant differences across categories of items ($F_{9,2367}=99.79, p<.05$), significant differences between endorsed and rejected items ($F_{1,263}=77.14, p<.05$), and a significant interaction between categories and endorsement ($F_{9,2637}=63.58, p<.05$). Items in the wishes and intentions category, referring to the intention to engage in particular behaviors or the desire for something, had the shortest response latencies with rejected items having shorter response times than endorsed items. Items describing reactions of the individual showed a similar pattern of response latencies between endorsed and rejected items as those in the wishes category, but with a smaller difference in response latency. Items describing reactions of others toward the person showed this same pattern. However, items assessing traits, biographical facts and attitudes had a

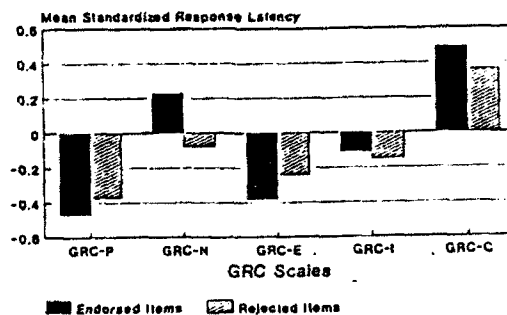
reverse pattern of latencies. For items in these three categories, response latencies were faster for endorsed items than rejected items.

Categories of Personality Items Response Latencies



Additionally, response latencies were examined for those 87 items measuring the five personality factors of Psychoticism, Neuroticism, Extraversion, Inadequacy, and Cynicism. There was an insufficient number of items within each scale to also include the category designation discussed above. The full MANOVA indicated a significant interaction between endorsement/rejection and personality factor. The mean standardized response latencies for each scale and endorsement/rejection responses are shown below. In general, subjects responded more quickly to the psychoticism and extraversion items and most slowly to the cynicism items. Endorsed items were responded to more quickly for the psychoticism and extraversion scales while rejected items were responded to more quickly for the neuroticism and cynicism scales. There was virtually no difference in response latencies for endorsed and rejected items on the inadequacy scale.

GRC SCALES Standardized Response Latencies



Scored if "True Response"

Since it has been postulated that response latencies to personalities are related to the strength or level of a particular personality measure, the correlations among the scale scores and between the scale scores and the

response latencies were examined. Psychoticism scores were correlated with both neuroticism, inadequacy, and cynicism (see Table 2 below). Extraversion was negatively related with inadequacy. Since there were differences in response latencies depending on whether individuals endorsed or rejected various items, Table 3 shows the relations between the scale scores based on the total number of items, endorsed items only, and rejected items only. When the total scale score was considered, only two correlations were significant. Response latencies on the cynicism scale were positively related to scores on the psychoticism scale ($r=.13$) and response latencies on the inadequacy scale were positively related to the total score on the inadequacy scale ($r=.18$).

Table 2
Zero-Order Correlations Among the Five Personality Measure

	GRC-P	GRC-N	GRC-E	GRC-I	GRC-C
	1	2	3	4	5
ST Items Only					
GRC-P	1.0000	.4923**	-.0112	.3137**	.4444**
GRC-N	.4923**	1.0000	-.0186	.3501**	.5767**
GRC-E	-.0112	-.0186	1.0000	-.2142**	.0306
GRC-I	.3137**	.3501**	-.2142**	1.0000	.2319**
GRC-C	.4444**	.5767**	.0306	.2319**	1.0000

N of cases: 509 1-tailed Signif: * - .01 ** - .001

GRC-P=Psychoticism; GRC-N=Neuroticism; GRC-E=Extraversion; GRC-I=Inadequacy; GRC-C=Cynicism

ST=Scale Scores based on items that are scored if response is 'true'.

A different pattern of relations were obtained when scores for endorsed items that were scored for the trait when endorsed was used. For endorsed items, the largest correlations were between the respective scale scores and the associated response latencies. The correlations were all negative indicating that faster responses were related to higher scale scores with the exception of the psychoticism scale. However, there were some off-diagonal correlations that were significant. For example, response latencies to the Neuroticism scale (endorsed items only) also were significantly correlated with the scale score on cynicism ($r=-.22$). Extraversion response latencies were significantly related to the inadequacy scale scores ($r=.16$), and response latencies were significantly related with scales scores on psychoticism and neuroticism ($r=.19$ and $-.23$, respectively).

A somewhat different pattern of relations was obtained when the response latencies for rejected items were examined. Again the diagonal correlations (scale scores with response latencies) were generally the largest correlations

with the exception that neuroticism scale scores were not significantly related to response latencies on the neuroticism items. However, these correlations were positive rather than negative indicating that higher scores were associated with longer responses. The off-diagonal correlations differed as well. For example, response latencies to rejected items on the cynicism scale were positively related to the psychoticism scale and the inadequacy scale ($r=.15$ and $.14$, respectively). Given the pattern of similarities and differences, there is some evidence to support response latencies as indicative of the strength of the trait; however, there is less support for this interpretation for rejected items.

Table 3
Correlations Among Five Personality Scale Scores and Response Latencies

Response Latencies	GRC-P 1	GRC-N 2	GRC-E 3	GRC-I 4	GRC-C 5
Total Scale					
GRC-P	.1087	.0858	.0313	.0944	.1321*
GRC-N	.0861	.0347	-.0479	.0308	-.0377
GRC-E	.0032	.0191	.0199	-.0111	-.0519
GRC-I	.0577	.0701	-.0386	.1782**	.0646
GRC-C	.0146	.0067	-.0941	-.0220	.0461
ST-RT Items Only					
GRC-P	.2985**	.1555*	.0147	.0437	.1923**
GRC-N	-.0988	-.2254**	-.1000	-.0733	-.2294**
GRC-E	.0235	-.0928	-.2651**	.0848	-.0823
GRC-I	-.0939	-.0698	.1561*	-.1778**	-.0521
GRC-C	-.0832	-.2175**	.0424	-.1115	-.2510**
ST-RF Items Only					
GRC-P	.1396*	.1167	.0846	.0642	.1456*
GRC-N	.0660	-.0613	.0764	.0037	.0221
GRC-E	.0494	.0877	.2758**	-.0387	.0169
GRC-I	.1412*	.1271*	-.1729**	.2468**	.1403*
GRC-C	.0655	.1201	-.1545*	.0314	.1727**

GRC-P=Psychoticism; GRC-N=Neuroticism; GRC-E=Extraversion; GRC-I=Inadequacy; GRC-C=Cynicism

Note: TOT=total response latency for scale; STRT=response latencies for endorsed items (response=true) scored 'true'; STRF=response latencies for nonendorsed items (response=false) scored 'true'.

N of cases: 358 1-tailed Signif: * - .01 ** - .001

Next the relations between the scores on the personality scales and the scores on the other tests were examined and are reported in Table 4 below. Scores on the Activities Interest Inventory were significantly related to

Table 4
CORRELATIONS BETWEEN SCORES ON RISK, ENCODING, ITEM RECOGNITION,
MENTAL ROTATION, AND TIME-SHARING TASKS AND FIVE PERSONALITY MEASURES
SCORES ON PERSONALITY SCALES

	GRC-P	GRC-N	GRC-E	GRC-I	GRC-C
ACTIVITIES INTEREST					
RISK	-.0747	-.1942**	.1661**	-.0872	-.0982
ENCODING TASK					
PIRIGHT	-.0568	-.0842	.0428	-.0806	-.0276
NIRIGHT	.0170	-.0461	.0209	-.0156	-.0471
CIRIGHT	-.0797	-.0659	.0237	-.0594	-.0452
ITEM RECOGNITION TASK					
RIGHT1	-.0915	-.0648	-.0265	-.0331	-.0919
RIGHT2	-.1391**	-.0034	.0532	-.0203	-.0469
RIGHT3	-.0758	-.0115	-.0444	.0195	.0015
RIGHT4	-.0798	-.0351	-.0308	-.0386	-.0552
RIGHT5	-.0235	.0703	-.0118	-.0047	-.0091
RIGHT6	-.0136	.0098	.0047	.0003	-.0566
MENTAL ROTATION TASK					
RS0	.0130	.0192	-.0521	.0092	.0400
RS60	.0338	-.0082	-.0348	-.0169	.0340
RS120	.0219	.0482	-.0261	.0598	.0364
RS180	-.0195	.0525	.0092	-.0003	.0405
RM0	-.0142	.0276	-.0251	-.0020	-.0254
RM60	-.0222	-.0254	-.0054	.0204	-.0915
RM120	-.0197	-.0070	-.0142	-.0131	-.0869
RM180	-.0223	-.0053	-.0599	.0467	-.1072*
TIME-SHARING TASK					
DIFF21	-.0095	.0096	.0332	.0338	.0211
DIFF22	-.0304	-.0521	.0009	.0330	-.0919
DIFF23	-.0516	-.0564	.0145	.0192	-.1278*
DIFF24	-.0692	-.0407	.0185	.0209	-.1177*
DIFF25	-.0788	-.0786	.0638	.0187	-.1251*

Note. RISK=Number of risk choices; PIRIGHT=Physical Identity Correct, NI=Name Identity, CI=Category Identity; RIGHT1-6=correct responses digits 1-6; RS0-RS180=Correct responses to Same with corresponding degrees of rotation; RM0-RM180=Correct responses to Mirror with corresponding degrees of rotation; DIFF21-25=difficulty level. Minimum pairwise N of cases: 494 1-tailed Signif: * - .01 ** - .001

scores on neuroticism and extraversion but in opposite directions ($r = -.19$ and $.17$, respectively). None of the Encoding Task scores were significantly related to the personality measures. Only 1 of the 30 correlations among scores on the item recognition task and personality measures was significant and only 1 of the 40 correlations among scores on the mental rotation task and personality measures was significant providing little support for any relation between these two tasks and the five personality measures included in this study. The correlations between scores on the time-sharing task and the personality measures were not much stronger. However the cynicism scale was significantly related to difficulty level for three of the five difficulty levels.

Similarly, the correlations between scores on the tasks and response latencies for endorsed items on the five personality measures were examined and are reported in Table 5. Only the scores on the Activities Interest Inventory were significantly related to any of the personality measures, and then only for the neuroticism scale. Therefore, it appears that the five personality measures and their associated response latencies are distinct from the other tasks included in this study.

Next, the response latencies on the Activities Interest Inventory, the Encoding Task, Item Recognition Task, Mental Rotation Task, and Time-sharing Task were correlated with the scale scores on the personality measures and their associated response latencies (see Tables 6 and 7 below). Response times (both for higher-risk choices and lower-risk choices) to the Activities Interest Inventory were significantly, negatively related to three of the personality measures (psychoticism, neuroticism, and cynicism). This represents a different pattern of relations than was found between scores on the Activities Interest Inventory and the personality measures. Further, the response latencies to the Activities Interest Inventory were unrelated to the response latencies to the personality. Therefore, these data do not support the interpretation of response latencies as measures of the strength of the trait. The Encoding Task and the Item Recognition Task response latencies were unrelated to either the scores on the personality measures or the response latencies on the personality measures which is consistent with the actual scores on the Encoding Task and Item Recognition Tasks with the personality measures. The response latencies on the Mental Rotation Task were significantly related to the cynicism scale scores only, but were not related to the response latencies on the cynicism scale. Response times on two of the five trials of the Time-sharing Task were significantly related to scores on the inadequacy scale whereas the accuracy scores on the task were significantly related to the cynicism scale score. None of the remaining correlations with the Time-sharing Task and the personality measures were significant.

Table 5
CORRELATIONS BETWEEN SCORES ON RISK, ENCODING, ITEM RECOGNITION,
MENTAL ROTATION, AND TIME-SHARING TASKS AND
RESPONSE LATENCIES ON FIVE PERSONALITY MEASURES

	RESPONSE LATENCIES ON PERSONALITY MEASURES				
	GRC-P	GRC-N	GRC-E	GRC-I	GRC-C
ACTIVITIES INTEREST TASK					
RISK	.0859	.1366*	-.0050	.0023	.0530
ENCODING TASK					
PIRIGHT	-.0492	-.0671	-.0970	.0745	.0472
NIRIGHT	-.0292	.0073	-.0458	.0412	.0067
CIRIGHT	-.0581	-.0426	.0378	.0122	.0409
ITEM RECOGNITION TASK					
RIGHT1	-.0158	-.0684	.0300	.0020	.0520
RIGHT2	-.0070	-.0090	-.0576	.0561	.0494
RIGHT3	-.0539	-.0171	.0454	-.0555	.0494
RIGHT4	-.0639	-.0155	-.0367	.0643	.0373
RIGHT5	-.0235	-.0502	.0080	.0704	.0614
RIGHT6	.0507	-.0333	.0390	.0549	.0656
MENTAL ROTATION TASK					
RS0	.0704	.0328	.0082	.0116	.0332
RS60	.0413	-.0184	.0358	.0238	.0365
RS120	.0404	-.0253	-.0086	.0033	.0480
RS180	-.0062	.0058	-.0065	-.0178	-.0021
RM0	-.0519	-.0367	-.0285	.0259	.0529
RM60	-.0415	-.0003	-.0577	.0462	.0418
RM120	-.0242	-.0411	-.0094	.0330	.0359
RM180	-.0840	-.0840	.0114	-.0446	.0421
TIME-SHARING TASK					
DIFF21	-.0287	.0110	.0743	.1139	-.0120
DIFF22	-.0279	.0881	.0379	.0508	-.0199
DIFF23	-.0697	.0916	.0025	.0931	.0170
DIFF24	-.0534	.0734	-.0171	.0833	.0377
DIFF25	-.0366	.0612	-.0333	.0822	.0341

Note. RISK=Number of risk choices; PIRIGHT=Physical Identity Correct, NI=Name Identity, CI=Category Identity; RIGHT1-6=correct responses digits 1-6; RS0-RS180=Correct responses to Same with corresponding degrees of rotation; RM0-RM180=Correct responses to Mirror with corresponding degrees of rotation; DIFF21-25=difficulty level. Minimum pairwise N of cases: 405 1-tailed Signif: * -.01 ** - .001

Table 6
CORRELATIONS BETWEEN RESPONSE LATENCIES ON RISK, ENCODING,
ITEM RECOGNITION, MENTAL ROTATION, AND TIME-SHARING TASKS AND
SCORES ON FIVE PERSONALITY MEASURES

	GRC-P	GRC-N	GRC-E	GRC-I	GRC-C
ACTIVITIES INTEREST TASK					
RISKART	-.1302*	-.1405**	-.0674	-.0530	-.1194*
NRSKART	-.1484**	-.1711**	.0102	-.0899	-.1580**
ENCODING TASK					
PIRAVE	.0588	.0426	-.0040	.0672	.0672
NIRAVE	.0287	.0069	-.0344	.0007	.0352
CIRAVE	-.0082	-.0174	-.0363	.0107	.0188
ITEM RECOGNITION TASK					
ITMR1	.0399	.0752	-.0124	.0293	.0674
ITMR2	.0356	.0775	-.0416	.0509	.0647
ITMR3	-.0023	.0475	-.0570	.0491	.0366
ITMR4	.0200	.0810	-.0371	.0390	.0796
ITMR5	-.0002	.0433	-.0342	.0487	.0205
ITMR6	.0167	.0521	-.0148	.0412	.0427
MENTAL ROTATION TASK					
MNRS0	.0105	.0482	-.0364	.0117	.1197*
MNRS60	-.0075	.0421	-.0023	.0054	.1017
MNRS120	-.0252	.0588	-.0119	.0161	.1145*
MNRS180	.0102	.0197	-.0591	.0104	.0796
MNRM0	.0140	.0950	.0097	-.0006	.1420**
MNRM60	.0318	.0667	-.0220	-.0001	.1292*
MNRM120	-.0089	.0166	-.0625	-.0194	.0841
MNRM180	-.0235	.0089	-.0683	-.0137	.0815
TIME-SHARING TASK					
RESPT21	-.0258	-.0575	-.0512	-.0421	-.0734
RESPT22	-.0852	-.0774	.0384	-.1606**	-.0172
RESPT23	-.0171	-.0710	-.0294	-.0720	-.0578
RESPT24	.0113	-.0773	-.0371	-.1624**	-.0176
RESPT25	-.0436	-.0556	-.0382	-.0980	-.0961

Note. RISKART=mean RT to risk choices; NRSKART=mean RT to non-risk choices; PIRAVE=mean RT Physical Identity Correct, NI=Name Identity, CI=Category Identity; ITM1-6=mean RT correct responses digits 1-6; MNRS0-MNRS180=mean RT correct responses to Same with corresponding degrees of rotation; MNRM0-MNRM180=Mean RT correct responses to Mirror with corresponding degrees of rotation; DIFF21-25=difficulty level.

Minimum pairwise N of cases: 494 1-tailed Signif: * - .01 ** - .001

Table 7

CORRELATIONS BETWEEN RESPONSE LATENCIES ON RISK, ENCODING,
ITEM RECOGNITION, MENTAL ROTATION, AND TIME-SHARING TASKS AND
RESPONSE LATENCIES ON FIVE PERSONALITY MEASURES

	GRC-P	GRC-N	GRC-E	GRC-I	GRC-C
ACTIVITIES INTEREST TASK					
RISKART	-.0181	-.0013	.0420	-.1009	-.0242
NRSKART	-.0008	.0467	.0380	-.0722	.0181
ENCODING TASK					
PIRAVE	-.0273	-.0157	-.0890	-.0108	.0051
NIRAVE	-.0467	-.0406	-.0583	-.0009	.0189
CIRAVE	-.0715	-.0315	-.0234	.0609	.0010
ITEM RECOGNITION TASK					
ITMR1	.0069	-.0356	.0064	.0065	-.0577
ITMR2	-.0277	-.0191	-.0129	.0090	-.0883
ITMR3	-.0529	-.0227	-.0214	.0095	-.0618
ITMR4	-.0214	-.0025	.0037	-.0140	-.0887
ITMR5	-.0462	-.0070	.0208	.0032	-.0910
ITMR6	-.0284	-.0102	-.0136	-.0495	-.0619
MENTAL ROTATION TASK					
MNRS0	-.0162	.0530	.0085	-.0282	-.0479
MNRS60	.0020	-.0083	-.0569	-.0421	-.0735
MNRS120	-.0076	-.0119	-.0317	-.0544	-.0218
MNRS180	.0260	.0481	-.0317	-.0120	-.0671
MNRMO	.0058	-.0549	-.0471	.0168	-.1038
MNRM60	.0282	-.0375	-.0481	-.0053	-.0903
MNRM120	-.0060	-.0250	-.0183	-.0021	-.0886
MNRM180	-.0024	-.0166	.0141	-.0178	-.0835
TIME-SHARING TASK					
RESPT21	.0323	.0022	.0318	-.0640	.0626
RESPT22	.0039	-.0179	-.0011	-.0605	.0273
RESPT23	-.0021	.0003	.0297	-.0485	.0967
RESPT24	-.0207	-.0320	.0443	-.0301	.0502
RESPT25	.0233	.0136	.0133	-.0870	.0911

Note. RISKART=mean RT to risk choices; NRSKART=mean RT to non-risk choices; PIRAVE=mean RT Physical Identity Correct, NI=Name Identity, CI=Category Identity; ITM1-6=mean RT correct responses digits 1-6; MNRS0-MNRS180=mean RT correct responses to Same with corresponding degrees of rotation; MNRMO-MNRM180=Mean RT correct responses to Mirror with corresponding degrees of rotation; DIFF21-25=difficulty level. Minimum pairwise N of cases: 405

The last set of analyses examined the relations between scale scores and latencies on the personality measures and three criteria (UPT, AFOQT-verbal and AFOQT-quantitative). These correlations are shown in Table 8. Scores on the psychoticism and cynicism scales were significantly, negatively related to UPT but none of the personality measure scores were related to the AFOQT scores. Response latencies were significantly related to UPT for the extraversion and cynicism scales only and the response latencies for the cynicism scales were significantly related to scores on the AFOQT-verbal.

Table 8
CORRELATIONS BETWEEN PERSONALITY MEASURES AND CRITERIA
AFOQT SCORES

	UPT	VERB1	QUAN1
TOTAL SCORES			
GRC-P	-.1321*	.0688	.0317
GRC-N	-.0787	.0874	.0326
GRC-E	.0536	-.0409	.0087
GRC-I	-.0974	-.0204	.0219
GRC-C	-.1395*	.0074	-.0259
SCORED TRUE ITEMS ONLY			
GRC-P	-.1142*	.0603	.0171
GRC-N	-.0757	.0953	.0358
GRC-E	.0457	-.0565	-.0008
GRC-I	-.0828	-.0314	.0102
GRC-C	-.1395*	.0074	-.0259
RESPONSE LATENCIES ON SCORED TRUE ITEMS ONLY			
GRC-P	-.0943	.0038	.0261
GRC-N	-.0212	.0013	-.0299
GRC-E	-.1100*	.0027	-.0149
GRC-I	.0140	.0490	-.0011
GRC-C	.1263*	-.1104*	-.0069

GRC-P=Psychoticism; GRC-N=Neuroticism; GRC-E=Extraversion; GRC-I=Inadequacy; GRC-C=Cynicism Scale

Minimum pairwise N of cases: 265 1-tailed Signif: * - .01 ** - .001

Discussion and Summary

This study examined the relations among response latencies and scores on several measures of cognitive information processing and personality measures. The pattern of relations across these various measures indicates that response latencies are extremely variable and appear to be effected by many aspects of the stimuli presented to individuals. There was no support for a claim that relative response latencies are stable across tasks, but rather, they seem to

reflect differential aspects of the task, individual differences across participants, and within subject variability.

There was some weak, limited evidence to suggest that response latencies may be useful for increasing the validity of personality and attitude measures. Quicker response times were associated with scores on most of the personality measures, at least when only those items which were scored if endorsed were used. However, on other tasks as well as reversed scored items, the data suggested that quicker responses may reflect anticipatory responding and less valid responses.

What is evident from the analyses conducted in this project is that response latencies need to be further explored, but in controlled situations. The response latencies across all tasks were not normally distributed. Typically, they were positively skewed. Many researchers have addressed this by truncating latencies above three standard deviations or by setting all latencies above three standard deviations to three standard deviations. However, based on the results of this study, it is necessary to address "too quick" as well as "too slow" responses. Further, since standard deviations are a function of the measure of central tendency used (typically, the mean), these procedures may not make appropriate adjustments. Therefore, further work needs to be done as to the best way to address the distributional adjustments.

This study also highlights the need to examine the structure of personality items and measures. Not only were the effects of scoring direction and response (endorsement versus nonendorsement) of Tetrick (1989) replicated in these data, there also was evidence that the type of item (e.g., trait, attitude, etc.) influenced the time it took for individuals to respond. Before response latencies are used to improve the "validity" of personality measures such as is advocated by Holden, et al. (1991), systematic examination of item characteristics that effect response latencies needs to be conducted.

References

- Angleitner, A., John, O. P., & Lohr, F. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), Personality Assessment via Questionnaires. Springer-Verlag: New York. pp. 61-107.
- Costa, P. T., Jr., Zonderman, A. B., McCrae, R. R., & Williams, R. B., Jr. (1985). Content and comprehensiveness in the MMPI: An item factor analysis in a normal adult sample. Journal of Personality and Social Psychology, 48, 925-933.
- Dickman, S. & Meyer, D. E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. Journal of Personality and Social Psychology, 54, 274-290.
- Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude accessibility. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function. Hillsdale, NJ: Erlbaum. pp. 153-179.
- George, M. S., & Skinner, H. A. (1990). Using response latency to detect inaccurate responses in a computerized lifestyle assessment. Computers in Human Behavior, 6, 167-175.
- Holden, R. R., Fekken, G. C., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test-item response latencies. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3, 111-118.
- Krosnick, J. A. (1989). Attitude importance and attitude accessibility. Personality and Social Psychology Bulletin, 15, 297-308.
- Markus, H. & Smith, J. (1981). The influence of self-schemata on the perception of others. In N. Cantor & J. F. Kihlstrom (Eds.), Personality, Cognition, and Social Interaction. Hillsdale, NJ: Erlbaum. pp. 233-262.
- Tetrick, L. E. (1989). An exploratory investigation of response latency in computerized administrations of the Marlowe-Crowne social desirability scale. Personality and Individual Differences, 10, 1281-1287.
- Vernon, P. A. (Ed.) (1987). Speed of Information-Processing and Intelligence. Norwood, NJ: Ablex.

**THE SIMILARITY OF AIR FORCE SPECIALTIES
AS ANALYZED BY ADDITIVE TREES, NETWORKS, AND
MULTIDIMENSIONAL SCALING**

Stephen A. Truhon

Associate Professor

Department of Social Sciences

**Winston-Salem State University
601 Martin Luther King Jr. Drive
Winston-Salem, NC 27110**

Final Report for:

Summer Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

September 1992

**THE SIMILARITY OF AIR FORCE SPECIALTIES
AS ANALYZED BY ADDITIVE TREES, NETWORKS, AND
MULTIDIMENSIONAL SCALING**

Stephen A. Truhon

Associate Professor

Department of Social Sciences

Winston-Salem State University

ABSTRACT

Air Force specialties which had been studied in the General Work Inventory and the Ease of Movement studies were reanalyzed for similarity through the use of additive trees, networks, and multidimensional scaling. The clustering through the use of additive trees suggested support for the existing taxonomy with exceptions for the clustering of various communication specialties and the clustering together of mechanical and electrical specialties. Extended trees did not much improve the clusters resulting from the additive trees. Network analysis suggested that some specialty areas such as air traffic control, communication, and the use of simulators show a good deal of interchangeability within those areas. In addition, specialties that are described as difficult are good preparation for other specialties, while easy specialties show ease of transfer from other specialties. Multidimensional scaling analyses revealed one dimension common to both studies involving clerical vs. craft skills. Also important in the Ease of Movement study was a dimension involving the amount of technical knowledge required for the specialty. Recommendations for combining specialties and for vocational counseling are made based on these results.

**THE SIMILARITY OF AIR FORCE SPECIALTIES
AS ANALYZED BY ADDITIVE TREES, NETWORKS, AND
MULTIDIMENSIONAL SCALING**

Stephen A. Truhon

INTRODUCTION

The clustering of individuals into jobs and similar jobs into higher levels of job families has long been a concern of personnel psychologists (Harvey, 1986). Such clustering can be helpful in such areas as evaluation of personnel and of jobs, career planning, and test validation. The Air Force has also been concerned with this issue (e.g., Christal, 1974), with increasing importance as the number of Air Force personnel decline and specialties need to be combined.

Two particular studies were of interest in examining the similarity of Air Force specialties: one from the General Work Inventory (GWI) (Ballantine & Cunningham, 1981) and the other from the Ease of Movement (EOM) study (e.g., Mayfield & Lance, 1988). As part of the GWI, seven or eight experts rated the similarity of 40 specialties (table not included) on a five-point scale. As part of the EOM study, personnel in 43 specialties, ranging in number from five to 60 per specialty (table not included)¹, were asked to rate on a nine-point scale the amount of retraining necessary to go from their specialties into other specialties (TO) and into their specialties from other specialties (FROM).

The purpose of the study was to examine the similarity of Air Force specialties from these two data sets with four statistical techniques: 1) clustering through the use of the ADDTREE program (Cortier, 1982; Sattath & Tversky, 1977); 2) clustering through the use of the EXTREE program, an extension of the ADDTREE program

¹ In the interests of space, not all tables and figures are included. An extended version of this report is available from the author.

(Corter & Tversky, 1986); 3) deriving network structures from the PATHFINDER program (Schvaneveldt, Durso, & Dearholt, 1985); and 4) multidimensional scaling through the MDS program (Kruskal, 1964a; 1964b).

ADDITIVE TREES

Sattath & Tversky (1977) devised their ADDTREE program as an improvement to traditional forms of clustering. In traditional clustering programs the two elements of a cluster are always closer to each other than to elements outside the cluster (the ultrametric inequality) and all nodes are equally distant from the root of the tree. In contrast, additive trees allow for inequality between elements outside a cluster and elements within a cluster by allowing the nodes to vary in length to better fit the data.

The ADDTREE program was applied to the similarity matrix from the GWI. The resulting tree structure is shown in Figure 1. Stress was .0461, indicating good fit between the original data and the resulting tree.

Examination of the resulting tree in Figure 1 reveals clustering that, for the most part, follows the current Air Force taxonomy. Medical (90250C, 90252, 91350, and 91450), audiovisual (23150, 23152, 23250), administrative (70250C, 70550, 73250, and 75150), construction (55250, 55251, 55254, and 55255), mechanical (47250, 47251A, 47251D, and 47252), maintenance (43150C, 43150D, 43151, 43152C, and 43152E), control systems (32151G, 32152A, and 32152Q), and simulator specialties (34152, 34153, 34154, 34155, and 34156) all form their own clusters. The lone exceptions involve the avionic systems with the communication and the air traffic control specialties. Note that Avionic Communications Specialist (32850) clusters with the communications specialties (30450 and 30454), while Aviation Navigation Systems Specialist (32851) and Airborne Warning & Control Radar Specialist (32852) cluster with air traffic control (30351, 30352, and 30353) specialties.

The ADDTREE program was also applied to the data from the EOM study. Two asymmetric matrices were derived from the study (TO and FROM). When ADDTREE analyzes an asymmetric matrix, it averages the corresponding values.

The results from the ADDTREE analysis of the TO data are presented in Figure 2. Stress was .0777, indicating reasonably good fit between the data and the resulting tree. Examination of this tree reveals confirmation of the existing taxonomy with some interesting exceptions. The security (811x0 and 811x2) specialties cluster

together as do the inventory, contracting, and logistics (645x0, 645x1, 645x2, 651x0, and 661x0), administrative (702x0, 732x0, and 791x0), medical (902x0, 906x0, and 981x0), vehicle (603x0 and 631x0), and construction (551x0, 551x1, and 553x0) specialties, although it should be noted that vehicle and construction specialties cluster together as well. The weather specialty (251x0) clusters with the airfield specialties (271x1, 272x0, and 276x0). The munitions (461x0 and 462x0) specialties also cluster together.

Among the exceptions to the Air Force taxonomy are that fire protection (571x0) clusters with safety (241x0 and 242x0) specialties, all of which deal with emergency situations. The fitness and recreation specialty (741x1) is off by itself. Morse systems operator (207x1), aircraft control and warning radar specialty (303x2), the two communications specialties (304x0 and 304x4), precision measurement equipment laboratory specialty (324x0), and the avionic communications specialty (455x2), all of which deal with communications, cluster together.

Among the maintenance and mechanical specialties some interesting clusters formed. Flight engineer (113x0C), aircraft loadmaster (114x0), tactical aircraft maintenance (452x4) and strategic aircraft maintenance (457x0) cluster together. The aircraft electrical systems specialty (423x0), aerospace ground equipment mechanic (454x1), special vehicle mechanic (472x1), electrician (542x0), and refrigeration and air conditioning specialty (545x0), all of which involve repair skills, form a cluster.

The additive tree was also derived from the EOM FROM data is not shown. Again good fit was found between the tree and the original data (stress= .0671).

Examination of the resulting tree once again shows some support for the existing taxonomy. Morse systems operator (207x1), weather specialist (251x0), airfield management specialist (271x1), air traffic control operator (272x0), and aerospace control & warning radar specialist (276x0) all cluster together. Munitions system (461x0) and aircraft armament systems (462x0) specialists form a cluster. The security (811x0 and 811x1) specialties cluster together, as do inventory management (645x0), materiel storage & distribution (645x1), supply systems analysis (645x2), contracting (651x0), and logistics plans (661x0) specialists.

Among the exceptions is the cluster formed from pavements maintenance specialist (551x0), construction equipment operator (551x1), engineering assistant specialist (553x0), vehicle operator dispatcher (603x0), and fuel specialist (631x0). Avionic communications specialist (455x2) clusters together with other communications and radar (303x2, 304x0, 304x4, and 324x0) specialists. Flight engineer (113x0C) and aircraft loadmaster (114x0)

cluster with tactical aircraft maintenance (452x4) and strategic aircraft maintenance (457x0) specialists. A cluster formed from mostly electrical specialties: aircraft electrical systems specialist (423x0), aerospace ground equipment mechanic (454x1), special vehicle mechanic (472x1), electrician (542x0), and refrigeration and air conditioning specialist (545x0). Fire protection (571x0) clusters with two safety (241x0 and 242x0) specialties. Finally, the administrative (702x0, 732x0, 741x1, and 791x1) and medical (902x0, 906x0, and 981x0) specialties appear to cluster together.

COMPARISON OF ADDITIVE TREES

While few of the same specialties are included in the GWI and EOM studies, it still is possible to make some comparisons between their additive trees. Medical, administrative, construction, mechanical, and air traffic control clusters form in all three analyses, although it should be noted that, in the EOM study, mechanical specialties group together with electrical specialties. Also, in all three analyses clusters involving air traffic control include specialties that are not directly linked in the current taxonomy.

The TO and FROM additive trees from the EOM study show a great deal of comparability. When the tree structure derived from the TO data was applied to the FROM data by means of the EXTREE program (Corter & Tversky, 1986), good fit occurred which was slightly less good than the tree originally derived for the FROM data (stress = .0707). The reverse also held true. In addition to the similar clusters noted above, security, inventory, vehicle, munitions, and safety clusters form that are identical for both TO and FROM. The few differences may reveal the asymmetry in the data, and not a lack of reliability. For example, it may be easier to retrain from fitness and recreation specialist (741x1) into other administrative specialties than it is to retrain from other administrative specialties into fitness and recreation specialist.

EXTENDED SIMILARITY TREES

When elements in a tree structure are not nested, a lack of fit results between the original data and the additive tree. Extended similarity trees (Corter & Tversky, 1986) compensate for this failure in additive trees by

introducing marked segments of the tree. These marked segments are parts of the arcs of a tree that appear in more than one place. EXTREE does not compute the length of the marked segment in measuring the distance between elements with these common features. As a result, EXTREE improves on ADDTREE's fit between the data and the tree.

Because the existing fit between the data and the additive trees was already quite good, the EXTREE analysis of the GWI and EOM revealed only slight improvements. As a result, those analyses will not be discussed.

PATHFINDER NETWORKS

In networks each concept is represented as a node. Nodes are connected to each other by links, with each link weighted to indicate the proximity between the nodes. Thus two nodes may be connected by a single link or by a path, consisting of a sequence of nodes and links. With symmetrical data, the links are considered undirected (i.e., travel between two nodes can occur in either direction). PATHFINDER attempts to construct a network of nodes and links so that the number of links are minimized while maximizing the fit to the original data (Schvanefeldt et al., 1985).

The PATHFINDER program was applied to the similarity matrix from the GWI. The resulting network is shown in Figure 4. There is reasonably good fit between the data and the network ($r = .6304$; if logarithms are taken, $r = .7429$).²

The network in Figure 4 reveals patterns similar to that shown in Figure 1. Some interesting differences occur. Note links among the following specialties: Avionic Communications (32850), Analog Navigation/Tactics Training Devices (34155) and Digital Navigation/Tactics Training Devices (34156); and among Air Traffic Control Radar (30351), Aircraft Control & Warning Radar (30352), Automatic Tracking Radar (30353), Avionic Navigation Systems (32851), and Airborne Warning & Control Radar (32852). In both cases there are cycles and cliques³ which suggest that relations among these concepts are not hierarchical.

² Esposito (1990) recommends counting the links separating the nodes rather than adding the weights in computing goodness of fit.

³ According to graph theory (Schvanefeldt et al., 1988), a cycle is a non-repeated path of links which returns to a starting node and a clique is a group of three or more nodes, each of which is connected to the other nodes.

Some specialties are more central while others are more isolated. Note that the medical (90250C, 90252, 91350, and 91450), administrative (70250C, 73250, and 75150), construction (55250, 55251, 55254, and 55255), and mechanical (47250, 47251A, 47251D, and 47252) specialties are relatively isolated from other specialties, while air traffic control (30351, 30352, and 30353), communication (30450, 30454 and 32850), and simulator (34152, 34153, 34154, 34155, and 34156) specialties are at the center of the network with many specialties densely linked to one another. This centrality suggests an interchangeability among these specialties.

The PATHFINDER program was also applied to the data from the EOM study. In contrast to ADDTREE, PATHFINDER treats the data from the two halves of an asymmetric matrix separately. As a result, a path between two nodes may be unidirectional or bidirectional. Networks that then result can be quite complex.

The PATHFINDER network for the EOM TO data is shown in Figure 5. Good fit exists between the data and the network ($r = .7288$; if logarithms are used $r = .7906$).

At first glance, Figure 5 is quite confusing. So many links appear that it is difficult to make sense of them. Certain specialties are almost obscured because of the many links to them. They include: Vehicle Operator/Dispatcher (603x0), Fuel Specialist (631x0), Fitness & Recreation Specialist (741x1), and Public Affairs Specialist (791x0). These specialties seem to be regarded as easy specialties. In contrast, some specialties have multiple links leading away from them (e.g., Flight Engineer [113x0C] and Air Traffic Control Operator [272x0]). These require a good deal of specialized training and may be regarded as preparation for other specialties.

The PATHFINDER network for the EOM FROM data was also derived (figure not shown). Reasonably good fit exists between the data and the network ($r = .6420$; if logarithms are taken, $r = .7357$).

The position of the nodes and links between them again produced results comparable to the ADDTREE with the same basic groupings. Some specialties have multiple links from other specialties, producing a pattern that looks like a star. Star patterns are associated with basic level concepts (Schvaneveldt et al., 1985). Specialties such as Vehicle Operator/Dispatcher (603x0), Inventory Management Specialist (645x0), Materiel Storage & Distribution Specialist (645x1), and Administration Specialist (702x0) display this star pattern and may be regarded as relatively easy to retrain to from many specialties.

COMPARISON OF NETWORKS

Despite the many differences between the GWI and EOM data, similarities between their networks exist. Technical specialties are relatively isolated and less technical specialties are central in the networks.

In comparing the TO and FROM networks, frequently the same specialties appear with many links (i.e., easy) and with few links (i.e., difficult). However, there are more links and greater complexity in the TO network. As a result the similarity of the networks is low ($c = .347$)⁴. This finding suggests that personnel believe that it is easier to move to others' specialties than it is for others to move to their specialty.

MULTIDIMENSIONAL SCALING

Multidimensional scaling is a set of techniques for taking measures of similarity between elements and mapping them to points in a multidimensional space, where the distances between points correspond to the similarity between the elements. Because stress, or the lack of fit between the distances and the similarities, decreases as the number of dimensions increases, it is up to the researcher to decide on the number of dimensions and interpret them (Kruskal & Wish, 1978).

The MDS program was applied to the GWI data. Three dimensions seem acceptable (stress = .1762). The resulting three dimensional solution is not shown. The first dimension appears to be a clerical vs. craft or white collar vs. blue collar dimension. Specialties such as Administration (70250C), Legal Services (70550), and Medical Service (90250C) specialists are at one extreme while Airlift/Bombardment Aircraft Maintenance Specialist (43152E), Special Vehicle Mechanic (47251A), and Carpentry Specialist (55250) are at the other extreme.

The second dimension seems to involve whether the specialty deals with operating an object or repairing an object. Specialties such as Helicopter (43150C and 43150D) and Special Vehicle (47251D) mechanics are at one extreme, while Digital Navigation/Tactics Training Devices (34156), Masonry (55251), and Protective Coating (55254) specialists are at the other.

⁴ c is a measure of similarity that ranges from 0 (complementary networks) to 1 (identical networks). It is roughly the ratio of the number of links in common in the networks to the number of links in either network (Goldsmith & Davenport, 1990).

The third dimension appears to deal with whether the type of contact that exists with other personnel is educational or not. Defensive System Trainer (34152), Digital Navigation/Tactics Training Devices (34156), and Education (75150) specialists are at one extreme, while Avionic Navigation Systems (32851), Personnel (73250), and Surgical Service (90252) specialists are at the other extreme.

The MDS program was applied to the EOM TO data. Two dimensions seem interpretable (stress= .3646). The first dimension appears to represent the clerical vs craft dimension. One finds that specialties such as Personnel (732x0), Fitness & Recreation (741x1), and Medical Administrative (906x0) specialists are at one extreme, while Strategic Aircraft Maintenance (457x0) and Aircraft Armament Systems (462x0) specialists are at the other.

The second dimension appears to deal with the technical knowledge requirements of the specialty. Aircraft Control & Warning Radar (303x2) and Medical Service (902x0) specialists are at one extreme, while Pavement Maintenance Specialist (551x0) and Vehicle Operator/Dispatcher (603x0) are at the other extreme.

Multidimensional scaling was also applied to the EOM FROM data. Two dimensions were interpretable (stress= .3242). The resulting two-dimensional solution is not shown. The first dimension appears to be a clerical vs. craft dimension. Administrative (732x0, 741x1, and 791x0) and medical (902x0 and 906x0) specialties are at one extreme, while Munitions Systems Specialist (461x0), Refrigeration and Air Conditioning Specialist (545x0), and Construction Equipment Operator (551x1) are at the other.

The second dimension seems to involve the amount of technical training involved. Inventory Management (645x0) and Security (811x0) specialists are at one extreme, while Flight Engineer (113x0C) and Electrician (542x0) are at the other extreme.

COMPARISON OF SCALING SOLUTIONS

One clear difference between the scaling solutions from the GWI study and the EOM study is better fit in the GWI study. This may be due to the fact that fewer points exist in the GWI study (40 vs. 43) and that the data are symmetrical in the GWI study.⁵ Nevertheless, the first dimension to appear in all three scaling solutions is the

⁵ Kruskal and Wish (1978) note that multidimensional scaling solutions are robust when the similarity matrix is roughly symmetrical.

clerical vs. craft dimension. The lack of comparability between the GWI and EOM studies on the remaining dimensions may indicate the difference in specialties included in the respective studies.

CONCLUSIONS

It would be incorrect to conclude that one particular statistical technique is preferable to another in analyzing the results from the GWI and EOM studies. Each of the methods provides information about the data which is supplemented by the other methods.

The multidimensional scaling analyses provide a global analysis of the data sets. Given the poorer fit between the scaling solutions and the data, attention should be paid to the dimensions that result, rather than where a particular point is located. The fact that the clerical vs. craft dimension shows up in all three analyses suggests that this is a robust distinction. The degree of technical knowledge required which appears in the EOM data analyses also seems to be an important dimension (Mayfield and Lance, 1988).

The additive tree analyses provide support for the existing taxonomy with some interesting exceptions: 1) avionic systems cluster with communication and air traffic control specialties; 2) electrical and mechanical specialties seem closely related; 3) vehicle and construction specialties cluster together; 4) Fire Protection Specialist (571x0) clusters with other safety specialties; 5) Flight Engineer (113x0C) and Aircraft Loadmaster (114x0) cluster with other maintenance specialties; and 6) the distinction between administrative and medical specialties may not be as great as currently suggested.

The networks provide further information about the relationship between these specialties. In the GWI study, the networks suggest that the relations among the communications and radar specialties are not hierarchical may, in fact be interchangeable. In the EOM study, the direction of movement between specialties permits the characterization of some specialties as easy and other specialties as difficult.

DISCUSSION

The findings from the above analyses can be applied in at least two ways: in combining specialties and in providing vocational counseling. The multidimensional scaling solutions will probably not be helpful in making decisions about which specialties to combine, because of the relatively high stress.

The additive trees provide more useful information in this regard by clustering similar specialties. If it is necessary to combine specialties, it would be advisable that they cluster together. The networks provide further help in suggesting which specialties are interchangeable: by cycles and cliques in the GWI and by directed paths in both directions in the EOM. These findings should be used with other sources of information (e.g., similarity of skills required) in making any decisions about combining specialties.

These statistical techniques can be useful in advising airmen about the specialties open to them. While abilities play an important part in the specialty an airman can choose, interests are also important. In the civilian world a popular model of vocational structure based upon interests has been proposed by Holland (1985). While its hexagonal structure has been criticized in favor of a tree structure (Gati, 1979), it has been used to advise people which kind of job environment to seek.

It is possible for the Air Force to have a similar kind of system of vocational counseling which would make use of findings from multidimensional scaling and additive trees. One could determine whether airmen would be interested in a clerical or craft specialty and in the degree of technical knowledge required. When the question came to several possible specialties, the hierarchical structure would provide groups of similar specialties for airmen to choose from. PATHFINDER networks could provide additional help.

REFERENCES

- Ballantine, R.D. & Cunningham, J.W. (1981). Development of the general work inventory. Proceedings of the 23rd Annual Conference of the Military Testing Association, 1, 125-133.
- Christal, R.E. (1974). The United States Air Force Occupational Research Project (AFHRL-TR-73-75). Lackland AFB, TX: Air Force Human Resources Laboratory, Occupational Research Division.
- Corter, J.E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. Behavior Research Methods & Instrumentation, 14, 353-354.
- Corter, J.E., & Tversky, A. (1986). Extended similarity trees. Psychometrika, 51, 429-451.
- Esposito, C. (1990). A graph-theoretic approach to concept clustering. In R.W. Schvaneveldt (ed.), Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex, pp. 89-99.
- Gati, I. (1979). A hierarchical model for the structure of vocational interests. Journal of Vocational Behavior, 15, 96-103.
- Goldsmith, T.E., & Davenport, D.M. (1990). Assessing structural similarity of graphs. In R.W. Schvaneveldt (ed.), Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex, pp. 75-87.
- Harvey, R.J. (1986). Quantitative approaches to job classification: A review and critique. Personnel Psychology, 39, 267-289.
- Holland, J.L. (1985). Making vocational choices (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-28.

Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. Psychometrika, 29, 115-129.

Kruskal, J.B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills, CA: Sage (Number 07-011).

Mayfield, D.L. & Lance, C.E. (1988). Development of a candidate task taxonomy for Air Force enlisted specialties. Unpublished paper.

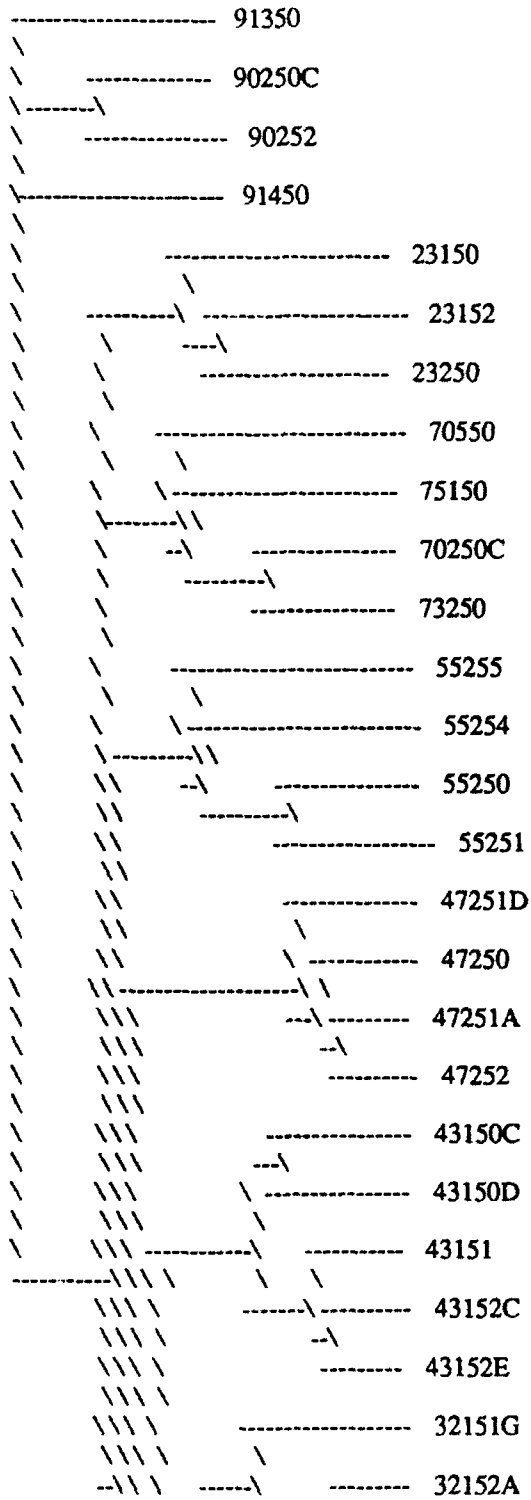
Sattath, S. & Tversky, A. (1977). Additive similarity trees. Psychometrika, 42, 319-345.

Schvaneveldt, R.W., Dearholt, D.W., & Durso, F.T. (1988). Graph theoretic foundations of Pathfinder networks. Computers & Mathematics with Applications, 15, 337-345.

Schvaneveldt, R.W., Durso, F.T., & Dearholt, D.W. (1985). Pathfinder: Scaling with network structures. Memorandum in Computer and Cognitive Science (MCCS-85-9). Las Cruces: New Mexico State University Computing Research Laboratory.

FIGURE 1

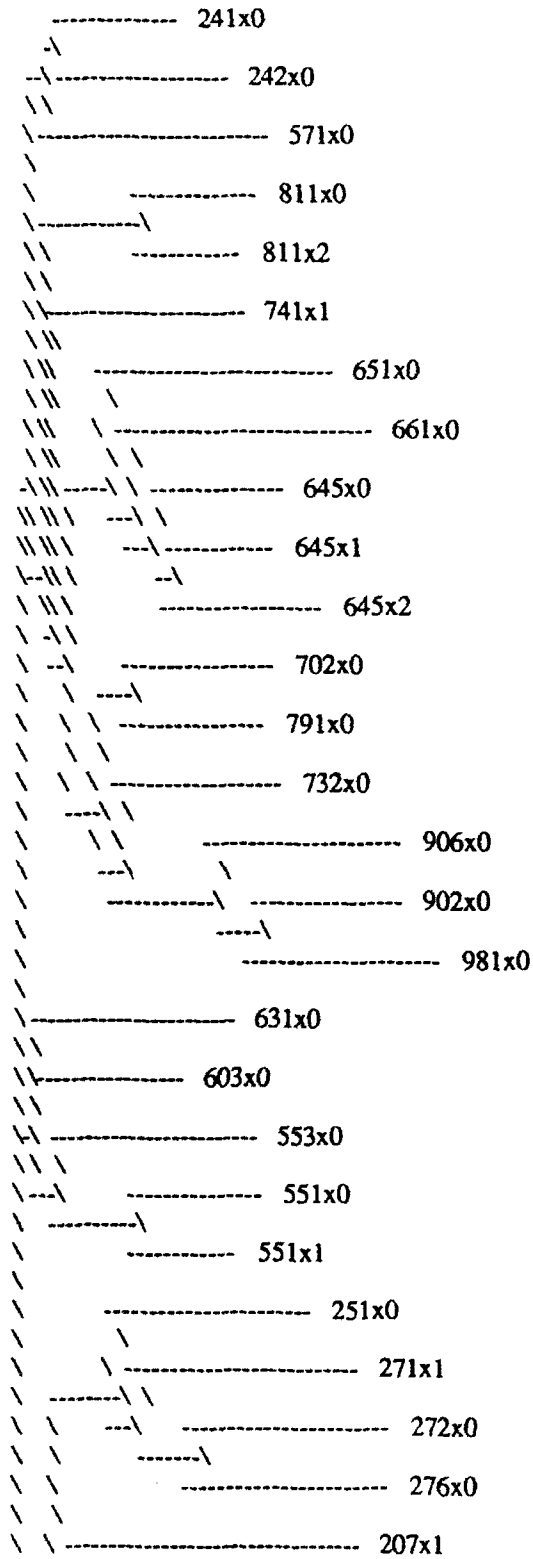
ADDITIVE TREE DERIVED FROM GWI

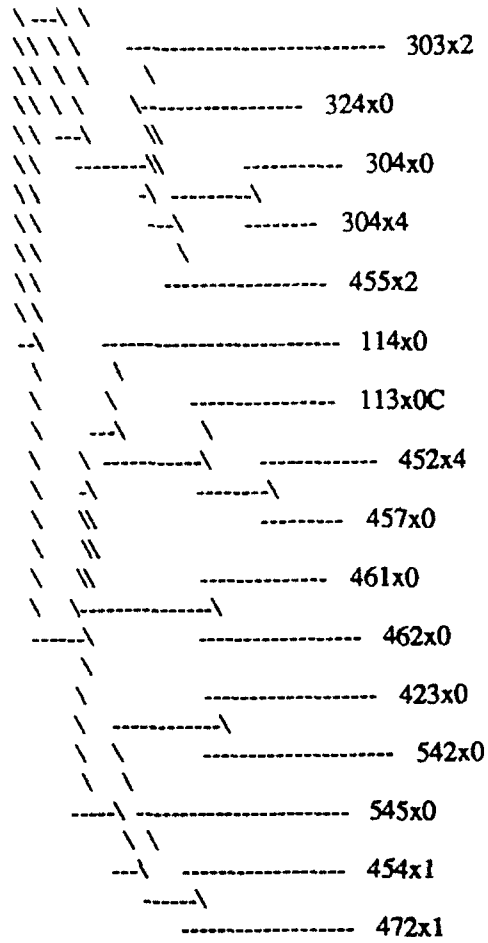


///	/	-----\	
///	/	-----	32152Q
///	/	-----	
--//	/	-----	34155
///	/	-----	
///	/	-----	34156
///	/	-----	
--//	/	-----	34152
///	//	-----	
///	//	-----	34153
///	//	-----	
///	//	-----	34154
-----//		-----	
///		-----	30450
///		-----	
--//		-----	30454
///	//	-----	
///	//	-----	32850
///	//	-----	
///	//	-----	32851
--//		-----	
///		-----	30353
///		-----	
--//		-----	30351
-----\		-----	
-----\		-----	30352
-----\		-----	
-----		-----	32852

FIGURE 2

ADDITIVE TREE DERIVED FROM EOM TO





OBSERVATIONS ON THE DISTRIBUTION, ABUNDANCE,
AND BIONOMICS OF Aedes albopictus (Skuse)
IN SOUTHERN TEXAS

Michael L. Womack
Associate Professor
Division of Natural Sciences and Math

Macon College
College Station Road
Macon, GA 31297

Final Report for:
Summer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

OBSERVATIONS ON THE DISTRIBUTION, ABUNDANCE,
AND BIONOMICS OF Aedes albopictus (SKUSE)
IN SOUTHERN TEXAS

Michael L. Womack
Associate Professor
Division of Natural Sciences and Math
Macon College

Abstract

The distribution and bionomics of the imported Asian tiger mosquito, Aedes albopictus (Skuse), in southern Texas was investigated. Thirty-five new county records were added to its distribution in this state. Additionally, 8 new county records were established for another container breeding mosquito, Aedes aegypti. Aedes albopictus, the Asian mosquito had a wide distribution in southern Texas, but its abundance decreased in the counties bordering the Rio Grande River. An investigation of the bionomics of Ae. albopictus in the San Antonio metropolitan area indicated that this mosquito was utilizing natural tree holes as a larval habitat.

OBSERVATIONS ON THE DISTRIBUTION, ABUNDANCE,
AND BIONOMICS OF AEDES ALBOPICTUS (SKUSE)
IN SOUTHERN TEXAS

Michael L. Womack

INTRODUCTION

The establishment of the imported Asian tiger mosquito, Aedes albopictus (Skuse) in the United States has produced concern as this species will serve as a vector for many endemic and exotic viruses. Laboratory investigations have demonstrated the capability of this vector to transmit Venezuelan Equine Encephalomyelitis (Beaman & Turell 1991), Chikungunya Virus (Turell et al. 1992), La Crosse Virus (Grimstad et al. 1989), and Rift Valley Fever (Turell et al. 1988). This mosquito is also capable of transmitting the viruses for Japanese encephalitis, yellow fever, West Nile, Ross River, 4 serotypes of dengue, St. Louis encephalitis, and western equine encephalitis (Shroyer 1986). In 1991, strains of Eastern Equine Encephalitis virus were isolated from adult Ae. albopictus mosquitoes collected in Florida (MMWR 1992).

A single adult was reported in 1983 from Memphis, Tenn. (Craven et al. 1988), but a breeding population of Aedes albopictus was not discovered until 1985 from collections in Harris County, Texas (Sprenger & Wuthiranyagool 1986). Used tire importation from northern Asia is now accepted as the origin of this insect in the U.S. (Hawley et al. 1987). Since its introduction, this species has increased its range throughout the southeastern and midwestern states. It is the principal nuisance mosquito in many metropolitan areas. Also, in most of its range

it has rapidly replaced another mosquito, Aedes aegypti (Linn.) which also shares the same artificial container larval habitat as Ae. albopictus. In many parts of the southeast, Ae. aegypti is difficult to recover if Ae. albopictus is abundant (Francy et al. 1990).

In 1986, a program was initiated by the U. S. Centers for Disease Control (CDC) to identify the distribution of this potential vector. State health authorities surveyed the Rio Grande Valley area during that time and have continued monitoring up to the present. Several finds of Ae. albopictus were reported in 1988 and a resurvey in 1989 established that Ae. albopictus could spread south and overwinter in the Texas border communities. Matamoros, Mexico reported finding the mosquito in 1988, but their health authorities have failed to detect its continued presence (Francy et al. 1990).

Continued monitoring by health authorities has failed to confirm the establishment of Ae. albopictus in south Texas, and counties within the Rio Grande Valley have not been resurveyed for the Asian tiger mosquito since the original 1986 effort. As this mosquito is a vector of many diseases, it is medically important to affirm or negate the establishment of this potential disease vector in the heavily populated area contiguous to Mexico where active cases of dengue and yellow fever have been reported.

This report describes the results of a resurvey accomplished from June to August 1992 in the Rio Grande Valley area of Texas northward to the boundary of the Edwards Plateau.

Distributional records from the U.S. Air Force ovitrapping program indicates that the Asian tiger mosquito has increased its frequency of occurrence on 4 military installations in San Antonio and suggests that this species is now the dominant container breeding mosquito in the city (McHugh 1992). Consequently, the bionomics and distribution of Ae. albopictus in San Antonio and its immediate environs were also investigated.

Beginning in 1991, the author has been investigating the bionomics and distribution of Ae. albopictus in Georgia. Data collected from this Texas survey will be used to compare the bionomics of western and eastern strains of this mosquito.

METHODS

Southern Texas County Surveys

Tire dealers, auto and truck repair shops, cemeteries, and plant nurseries were the major sites selected for sampling in this study because they offer numerous breeding sites for the maintenance of Ae. albopictus populations. Records from the Centers for Disease Control, Division of Vector-Borne Infectious diseases and the scientific (published) literature were used to determine which southern Texas counties had not reported Ae. albopictus prior to this project (Fournier et al. 1989). New county records for Ae. albopictus were reported to the CDC. Landing/biting adults were captured using vacuum aspirators. Larvae were collected with a turkey baster and were reared on liver powder at 25 C and a photoperiod of 14/10 hours.

San Antonio Metropolitan Area Bionomic Investigations

Recreational parks and cemeteries in widely spaced geographic locations were surveyed for possible habitats for adults and larvae of Ae. albopictus. When present, tree holes and artificial containers were sampled for mosquito larvae. Adults were identified upon emergence.

RESULTS

Southern Texas County Surveys

Thirty-five new county records were added to the distribution of Ae. albopictus in Texas (Table 1). In some instances Ae. aegypti was collected alone or in association with Ae. albopictus. This survey also produced new county records in Texas for Ae. aegypti in Brooks, Comal, Dimmit, Duval, Frio, Jim Hogg, McMullen, Zapata, and Zavala counties (Fournier et. al 1989). Figures 1 and 2 present indicate the county distribution for Ae. albopictus in Texas before and following this project.

Major human population centers are found in the Rio Grande counties of Hidalgo and Cameron. A resurvey for Ae. albopictus in these counties yielded positive sites only in Hidalgo, San Benito, and Olmito. These same areas had been identified in 1988 (CDC data 1992 and Francy et al. 1990). Table 2 lists the sampling points within the Rio Grande Valley populated areas. Brownsville reported no rain during July and no active breeding sites were found. However, San Pedro and Olmito are adjacent to the city limits of Brownsville. The major container breeding mosquito in the valley counties was Ae. aegypti.

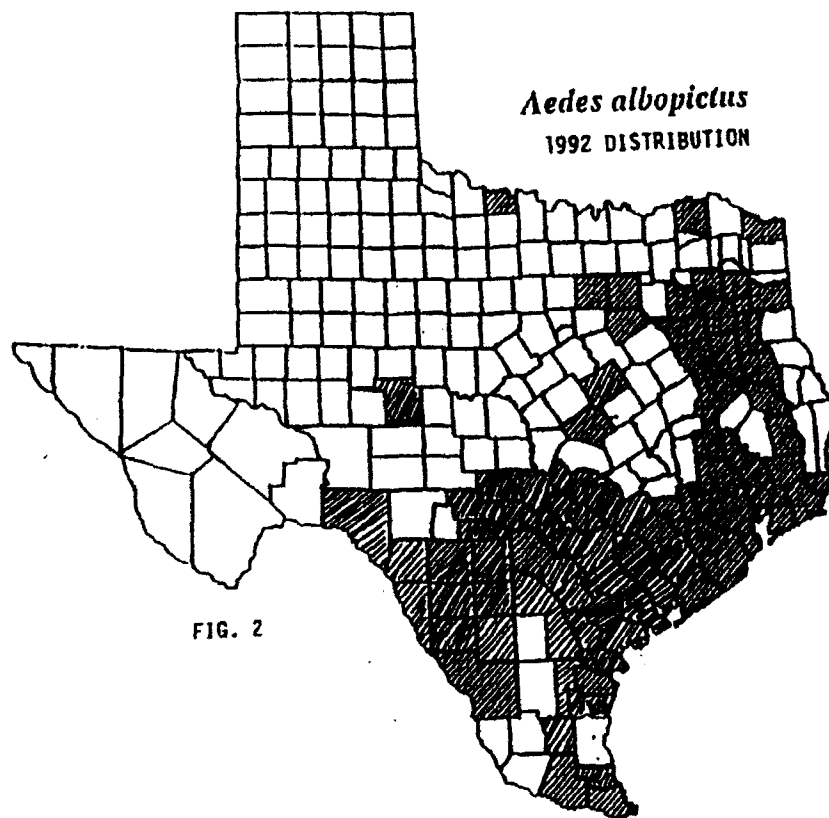
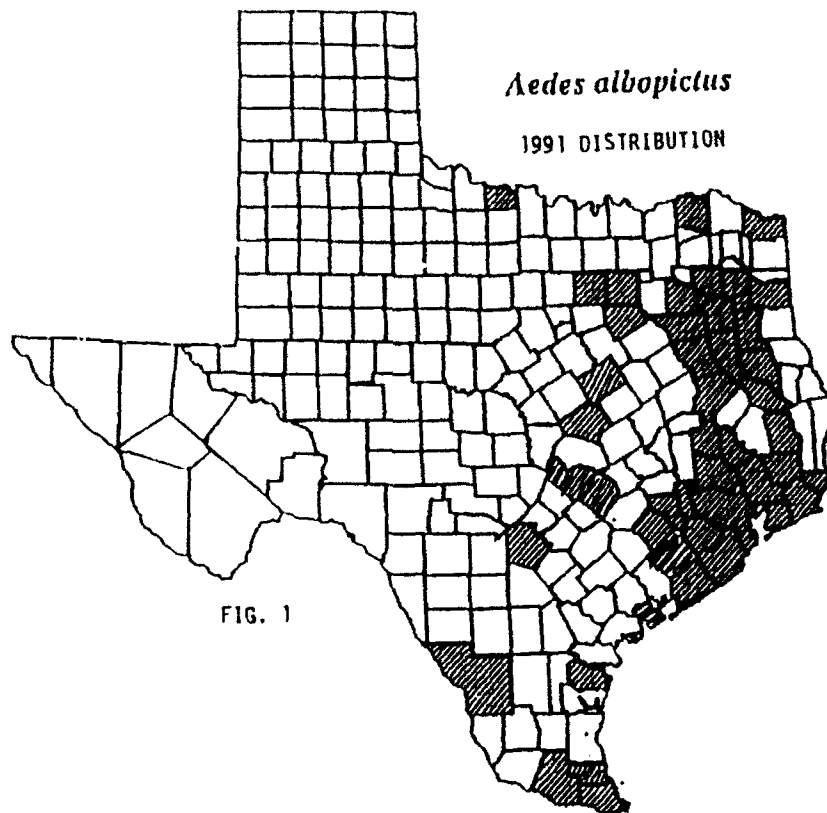


Table 1. Distribution of Ae. aegypti and Ae. albopictus in counties surveyed in southern Texas

COUNTY	CITY/TOWN	ACTIVITY	RESULT
Atascosa	Pleasanton	truck repair	2
Bandera	Medina	stream/vegetation	2
Bee	Beeville	tire dealership	2
Blanco	Johnson City	service station	2
Brooks	Falfurrias	service station	1/2*
Comal	New Braunfels	tire dealership	1/2*
Crawford	Luling	tire dealership	2
Dewitt	Yorktown	service station	2
Dimmit	Carrizo Springs	service station	1/2*
Duval	Freer	service station	1*
Fayette	Schulenburg	tractor repair	2
Frio	Pearsall	tire dealership	1/2*
Gillespie	Fredericksburg	tire dealership	2
Goliad	Goliad	tire dealership	2
Gonzales	Waelder	nursery/tires	2
Guadalupe	Seguin	truck repair	2
Hays	Wimberly	tire dealership	2
Jackson	Edna	tire dealership	2
Jim Hogg	Hebbronville	service station	1*
Jim Wells	Alice	used car lot/tires	2
Karnes	Karnes City	service station	2
Kendall	Boerne	tire dealership	2
Kerr	Kerrville	tire dealership	2
Kinney	Brackettville	recreational area	2
Kleberg	Kingsville	tire dealership	2
LaSalle	Cotulla	service station	1/2
Lavaca	Yoakum	used car lot	2
Live Oak	George West	service station	2
McMullen	Tilden	service station	1*
Maverick	Quemado	auto garage	2
Medina	Natalia	service station	2
Refugio	Refugio	tire dealership	2
San Patricio	Sinton	truck repair	2
Starr	Rio Grande City	tire dealership	1
Uvalde	Uvalde	tire dealership	1/2
Val Verde	Del Rio	truck repair	2
Victoria	Victoria	tire dealership	2
Wilson	Floresville	truck repair	2
Zapata	Zapata	tire dealership	1*
Zavala	Batesville	service station	2
Zavala	La Pryor	tire dealership	1*

1=Ae. aegypti 2=Ae. albopictus 1/2=both
 *new county record for Ae. aegypti

Table 2. Sampling sites in Cameron and Hidalgo counties

COUNTY	CITY	ACTIVITY	RESULT
Cameron	Harlingen	tire dealership	1
	La Feria	used tires	1
	La Feria	used tires	1
	San Benito	tire dealership	1
	San Benito	cemetery	1/2
	San Pedro	roadside dump	1
	Olmito	tire dealership	1/2
	Las Rusios	used tires	1
Hidalgo	Mercedes	used tires	1
	Donna	used tires	1
	Weslaco	used tires	1
	Hidalgo	tire importer	1/2
	Pharr	used cars	1
	McAllen	tire dealership	1
	McAllen	tire importer	1
	McAllen	service station	1
	La Joya	used tires	1
	Edinberg	used tires	1
	McAllen	nursery	1
McAllen	tire dealership	1	

1=Ae. aegypti 2=Ae. albopictus 1/2=mixed

The communities positive for Ae. albopictus were the same as those reported in the 1988 CDC survey. In this investigation, the Asian tiger mosquito demonstrated establishment but limited movement into other areas within the Rio Grande Valley.

San Antonio Metropolitan Area Bionomic Investigations

In contrast, surveys accomplished within the Greater San Antonio metropolitan area revealed an abundance of Ae. albopictus. Feeding adults were common along streams and other shaded areas at Brackenridge, McAllister, Commanche, East Southside Lions, and OP Schnabel parks. San Jose Burial Park,

the nature trail at mission San Juan Capistrano, and the San Antonio Botanical Gardens also were positive. Larvae of this mosquito were recovered in tree holes in association with other mosquitoes at McAllister, East Southside Lions, and Brackenridge parks. Treeholes were common in live oak, Quercus virginia Mill.; cedar elm, Ulmus crassifolia Nutt.; hackberry, Celtis laevigata Willd.; and weeping willow, Salix babylonica L..

DISCUSSION

Environmental factors probably play a major role in defining the distribution of Ae. albopictus in southern Texas. In the South Texas Plains area, total rainfall ranges from 16 to 35 inches annually and generally decreases from east to west. Highest precipitation usually occurs in May followed by a mid-summer decrease in rainfall. Hot summer days with high evaporation rates are common. In September, precipitation often begins to increase. Winters in the Rio Grande Valley are often frost free. Large acreages of cultivated land exist, but much of the area is still occupied by large cattle ranches. Historically, the climax vegetation was grassland or savannah type vegetation. However, grazing and other factors have produced brush vegetation composed of mesquite, post oak, live oak, cacti, and acacias.

The rainfall pattern in the Edwards Plateau region is similar to the South Texas Plains. These regions are predominately rangeland. Farming is restricted to the valley bottoms and around towns. Many species of grasses exist along

with live oak, shinnery oak, junipers, and mesquite. Brush vegetation is not as dominant as the southern plains (Gould 1975).

This survey also included portions of the Gulf and Blackland Prairies regions of Texas. Eastern portions of these regions receive greater and more uniformly distributed rainfall throughout the year. Much of this land is under cultivation. Climax vegetation for these zones are grasses but trees and brush have invaded much of these areas (Gould 1975).

The Asian tiger mosquito has a wide range in southern Texas with greater densities in major cities such as San Antonio and Corpus Christi. The movement of used tires is probably one of the most important factors which accounts for the distribution of Ae. albopictus. Westward toward the Rio Grande River, Ae. albopictus became increasingly uncommon or even absent. Rainfall also decreases in this same direction. The frequency of occurrence of Ae. aegypti was higher in areas which had the least rainfall in contrast to Ae. albopictus.

Cameron and Hidalgo counties, like other Rio Grande River counties had Ae. aegypti as the common container breeding mosquito. This observation was confirmed by discussions with environmental coordinators or vector control personnel in these counties. However, Ae. albopictus was recovered within the communities of Hidalgo, San Benito, and Olmito. This is unusual in that this area of South Texas has a dense human population suggesting that the Asian tiger mosquito should be more common.

However, the combination of high temperature, and low humidity and rainfall have probably contributed to the failure of Ae. albopictus to rapidly establish populations as it has in most of the southeastern United States. Regional weather data showed 1.14 cm of rain in June for McAllen, 2.67 cm for Weslaco, and 3.81 cm for Brownsville. In July, Weslaco reported the highest rainfall with 3.99 cm. Other recording stations reported less than 1.57 cm to no rain for July. Typical temperatures during the survey period ranged from highs of 35.5 C to lows of 22.2 C. These rainfall and temperature patterns are typical for this region during summer.

In conclusion, even though climatic factors may limit the distribution of Ae. albopictus, detailed surveys by local health authorities may yield additional populations of this species in the Rio Grande Valley. Surveys during winter may show the presence of populations of this mosquito as summer temperature and rainfall patterns in the valley region could be restricting the survival of this mosquito during the summer months.

Vast tracts of ranchland, limited vegetation, and the geographic location and west to east orientation of the Guadalupe, San Antonio, and Nueces River subsystems may also be enhancing the natural dispersal of Ae. albopictus westward but limiting dispersal into south Texas. Future surveys should be conducted along river systems to validate the importance of shade, humidity, and treeholes to this insect's natural dispersion requirements. In San Antonio, Ae. albopictus was

commonly collected along wooded, shady stream banks. Finally, the close proximity of Ae. albopictus populations in U.S. cities along the Rio Grande River suggests that public health authorities in Mexico should undertake a container breeding mosquito surveillance program in their border communities. The Asian tiger mosquito may eventually adapt to the environment, and climatic patterns may allow for the expanded naturalization of this disease vector throughout southern Texas.

REFERENCES CITED

- Beaman, J. R., and M. J. Turell. 1991. Transmission of Venezuelan Equine Encephalomyelitis virus by strains of Aedes albopictus (Diptera: Culicidae) collected in North and South America. J. Am. Mosq. Control Assoc. 28:161-164.
- Centers for Disease Control. 1992. Eastern equine encephalitis virus associated with Aedes albopictus - Florida, 1991. MMWR 41:115-121.
- Craven, R. B., D. A. Eliason, D. B. Franczy, P. Reiter, E. G. Campos, W. L. Jakob, G. C. Smith, C. J. Bozzi, C. G. Moore, G. C. Maupin and T. P. Monath. 1988. Importation of Aedes albopictus and other exotic mosquito species into the United States in used tires from Asia. J. Am. Mosq. Control Assoc. 4:138-142.
- Fournier, P. V., G. J. Teltow & J. L. Snyder. 1989. Distribution records of Texas mosquitoes. Texas Department of Health, Bureau of Laboratories.
- Franczy, D. B., C. G. Moore and D. A. Eliason. 1990. Past, present and future of Aedes albopictus in the United States. J. Am. Mosq. Control Assoc. 6:127-132.
- Gould, F. W. Texas plants a checklist and ecological summary. 1975. The Texas A&M University System Agricultural Experiment Station.

- Grimstad, P. R., J. F. Kobayashi, M. Zhang, and G. B. Craig. 1989. Recently introduced Aedes Albopictus in the United States: potential vector of La Crosse virus (bunyaviridae: California serogroup). J. Am. Mosq. Control Assoc. 5:422-427.
- Hawley, W. A., P. Reiter, R. S. Copeland and C. B. Pumpuni. 1987. Aedes albopictus in North America: probable introduction in used tires from northern Asia. Science 236:1114-1115.
- McHugh, C. 1992. Distributional records from the U.S. Air Force ovitrapping program -1991. J. Am. Mosq. Control Assoc. 8:198-199.
- Shroyer, D. A. 1986. Aedes albopictus and arborviruses: a concise review of the literature. J. Am. Mosq. Control Assoc. 2:424-428.
- Sprenger, D. and T. Wuithiranyagool. 1986. The discovery and distribution of Aedes albopictus in Harris County, Texas. J. Am. Mosq. Control Assoc. 2:217-219.
- Turell, M. J., C. L. Bailey, & J. R. Beaman. 1988. Vector competence of a Houston, Texas strain of Aedes albopictus for Rift Valley fever virus. J. Am. Mosq. Control Assoc. 4: 94-96.
- Turell, M. J., J. R. Beaman, & R. F. Tammariello. 1992. Susceptibility of selected strains of Aedes. aegypti and Aedes albopictus (Diptera: Culicidae) to Chikungunya Virus. J. Am. Mosq. Control Assoc. 29: 49-53.