LOAN DOCUMENT
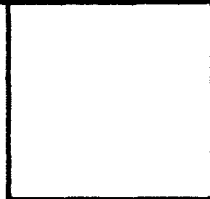
AD-A261 971

DTIC ACCESSION NUMBER

LEVEL

INVENTORY

AEOSR-TR-93-0117

DOCUMENT IDENTIFICATION

Dec 92

DISTRIBUTION STATEMENT
Approved for public release
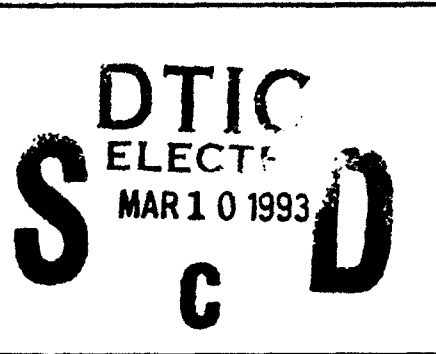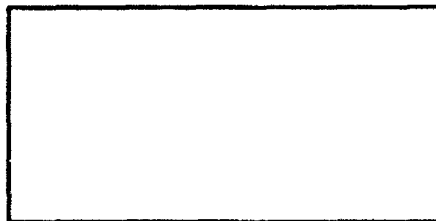Distribution Unlimited

DISTRIBUTION STATEMENT

H
A
N
D
L
E

W
I
T
H

C
A
R
E

ACCESSION FOR
NTIS   GRA&I
DTIC   TRAC
UNANNOUNCED
JUSTIFICATION

BY
DISTRIBUTION/
AVAILABILITY CODES

DISTRIBUTION	AVAILABILITY AND/OR SPECIAL
A-1	

DISTRIBUTION STAMP

DTIC
ELECT
S   D
MAR 1 0 1993
C

DATE ACCESSIONED

DATE RETURNED

DTIC QUALITY INSPECTED 1

93   3 10 011

93   3 4 071

DATE RECEIVED IN DTIC

93-05129

REGISTERED OR CERTIFIED NUMBER

DTIC FORM 70A
JUN 90

DOCUMENT PROCESSING SHEET

LOAN DOCUMENT

PREVIOUS EDITIONS MAY BE USED UNTIL
STOCK IS EXHAUSTED.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	28 Dec 92	Annual 1 Sep 91 - 31 Aug 92

4. TITLE AND SUBTITLE

1992 Summer Faculty Research Program (SFRP)
Volumes 1 - 16 ν-7

F49620-90-C-0076

5. AUTHOR(S)

Mr Gary Moore

AFOSR-TR· 3 0117

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

Research & Development Laboratoreis (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 cintinuing research of
interest to the Air Force at the institution of the faculty member; to stiumlate
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

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1992

GRADUATE STUDENT RESEARCH PROGRAM (GSRP) REPORTS

VOLUME 7

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 7 of 16, presents the final research reports of graduate student (GSRP) 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 GRADUATE RESEARCH REPORTS

Armstrong Laboratory

Report Number	Report Title	Author
1	Feature Conjunction Search in Three-Dimensional Space	Jennifer L. Blume
2	Fractal and Multifractal Aspects of an Electroencephalogram	Elaine M. Brunsman
3	The Relationship Between Serum 2, 3, 7, 8-Tetrachlorodibenzo-Para-Dioxin Blood Levels and Psychological Problems: Social Class, Exposure to Toxins and Mental Distress	Russell P. D. Burton
4	In Utero Taste Aversion Conditioning in the Rat Fetus	Jed S. Cohen
5	Animal Emotionality and Rhesus Macaque (*Macaca Mulatta*) Vocalizations	James W. Collins
6	The Predictive Validity of Automated Personnel Tests and ASVAB Tests for Performance on Air Force Tutors	David Dickter
7	Inorganic Fiber Analysis by SEM-EDXA	Robert F. Diskin
8	Components of Spatial Awareness: Effects of Air Force Fighter Pilot Training and Experience	Itiel E. Dror
9	A Preliminary Investigation of the Effects of a Dynamic Graphical Model During Practice of a Console Operation Skill	John D. Farquhar
10	Validation of the Articulated Total Body Model Regarding Rollover Crashes	Jennifer M. Ferst
11	Transmission Delays and Bursting in Statistical Mechanical Neural Network Models	Michelle A. Fitzurka
12	Assessing the Implementation of a Multivariate Risk Screening Procedure in the Diagnosing of Down Syndrome Associated Pregnancy (DSAP)	William R. Fletcher, Jr.
13	(Report not received)	
14	Probability and Cue Type Manipulations in a Visual Attention Task	Lawrence R. Gottlob
15	The Effects of Two Doses of Exogenous Melatonin on Temperature and Subjective Fatigue	Rod J. Hughes
16	Assisting Air Force Instructional Designers	Tricia Jones
17	The Assessment of Biological Variability	Robert Craig Kundich
18	Multiple Physiological Measures of Performance on a Multiple Task Battery	Scott H. Mills

Armstrong Laboratory (cont'd)

FEATURE CONJUNCTION SEARCH IN
THREE-DIMENSIONAL SPACE

Jennifer L. Blume
Department of Psychology

Texas Tech University
Lubbock, Texas 79409

FEATURE CONJUCTION SEARCH IN THREE-DIMENSIONAL SPACE

Jennifer L. Blume
Department of Psychology
Texas Tech University

Abstract

Assumptions posed by Previc's model (1990) that the upper visual field (UVF) and lower visual field (LVF) are specialized for far and near vision, respectively, were investigated. A further assumption that the left brain hemisphere, and thus the right visual field, as well as the UVF are characterized by focal processing was also tested. Detection speed and accuracy of random-dot stereogram targets in a feature conjunction task were measured as a function of target position (lateral half-field, vertical half-field, eccentricity, and depth). Targets appearing in tne UVF and the right half-field did result in faster responding than targets in the LVF or left half-field. The results are discussed in support of Previc's theoretical model of visual processing.

FEATURE CONJUCTION SEARCH IN THREE-DIMENSIONAL SPACE

Jennifer L. Blume

INTRODUCTION

A recent theoretical paper (Previc 1990) outlined a model of visual processing in three dimensional space which proposes functional asymmetries of the vertical and lateral visual half-fields.

Along the vertical axis, the model defines a distinction between the functions and characteristics of the upper visual field (UVF) and the lower visual field (LVF). Specifically, Previc proposes that the UVF is specialized for far visual space and utilizes a focal mode of processing while the LVF is specialized for near space and is characterized by global processing. One line of evidence supporting these relationships is that as the eyes move up into the UVF, or down into the LVF, they reflexively diverge or converge, corresponding to the eye movements necessary when viewing far or near object (Heuer, 1988). The relationship of far vision to the UVF also is consistent with the fact that as objects move

further in depth, they move higher in the visual
field (i.e., the monocular pictorial depth cue,
"height in field").

Functional distinctions between the left and
right visual fields seem to be due to brain func-
tioning. Several studies have shown that the left
hemisphere, and thus the right visual field, seems
to be specialized for visual search, object and
shape detection, and object recognition (Hellige
1976, Polich 1982, & Yund, 1990). Some evidence
has also shown that global, holistic processing is
a right hemispheric function while focal processing
is a left hemispheric function (Delis, 1986). The
model's propositions regarding left/right asym-
metries define the right visual field as being
specialized for focal (far) processing and the left
visual field for global (near) processing (Previc
1990).

The ecological relevance of these asymmetries
is evidenced by several observations. Most visuo-
motor activity occurring in near space (and the
LVF) is optically degraded due to movement, extreme
diplopia (causing blurred vision), and misaccomoda-
tion which in turn would necessitate a non-focal,

global processing of such information (Previc 1990). Conversely, objects in far vision (and the UVF) tend to be smaller, clearer, and slower allowing focal processing which has been shown to be necessary for object recognition (Treisman & Schmidt, 1982).

Visual search has been defined by some researchers as involving two separate processes. Treisman (1982) and Treisman & Gelade (1980) have proposed that searching for a target that differs from background distractors by one dimension or feature (a square among circles) is a parallel process. Yet, when the target differs by conjunction of two features (a small square among small circles and large squares), the search is a serial process involving focal processes (Treisman, Sykes, & Gelade 1977).

Assuming that feature conjunction tasks do involve focal attention and that focal attention is a characteristic of the UVF (the far visual system) and the right visual field (the left hemisphere), Previc's model would predict that feature conjunction search is better performed in the upper right visual field and poorest in the lower left

quadrant.

In the present study, speed and accuracy of detection, recognition, and localization in a feature conjunction task were measured as a function of depth, eccentricity, vertical visual field, and lateral visual field. According to Previc's model, visual performance should be better when the target appears in the UVF relative to the LVF, and the right visual field relative to the left visual field. Since Previc's model states that the LVF is functionally specialized for near vision and the UVF for far vision, we investigated the relationship between visual field and depth of target. The model predicted that performance should be better for far targets presented in the UVF and near targets presented in the LVF. Further, performance is expected to should decrease as eccentricity increases.

METHOD

Subjects

Ten employees of the USAF Armstrong Laboratory with a mean age of 25.4 participated in the study. All subjects had uncorrected or corrected (with contact lenses only) vision of 20/30 or

better, normal stereoscopic depth perception as measured by the Armed Forces Vision Testing Apparatus, and balanced crossed and uncrossed disparity. All subjects were also right handed, right footed, and right-eye dominant as determined by the Miles ABC test (Miles, 1929).

Apparatus and Stimuli

The stimuli were generated on a Silicon Graphics 3130 Iris computer system and were presented on a Hitachi 60-hz color video monitor. To collect reaction time (RT) data, the Iris computer was linked to a Zenith Datasystems Z-248 computer which provided a clock with a 1-ms resolution.

Subjects viewed the display with their forehead and chin positioned in a standard opthal-mological support brace; eyes were centered vertically on the screen. Stimuli were random dot stereogram (RDS) stimuli, which were created using the red-green anaglyphic technique. A red filter (Kodak Wratten #29) covered the right eye and a green filter (Wratten #58) covered the left eye. Response information was relayed through buttons on two mouses, a "location mouse" and a "search

mouse", resting on the table in front of the subject.

RDS stimuli consisted of diamonds and squares, each in two sizes. For one third of the total trials, each target was presented at a near depth (475-mm apparent distance) a coplaner depth (500 mm distance to screen), and a far depth (522-mm apparent distance). Virtual separation of near and far targets from the fixation plane was 2.3-mm, or 19' arc disparity. This distance was kept constant by sliding the CRT forward or backward according their in interpupillary distance. The areas of the large and small coplaner stimuli were 10.34-mm2 and 5.17mm2, respectively. To create equal projected sizes of all targets in virtual 3-dimensional space, the areas of the large and small near targets were adjusted to 10.68-mm2 and 5.52-mm2, and the large and small far targets to 10.00-mm2 and 4.83-mm2 (see Gulick & Lawson, 1976, chap.10).

Tasks

Search Task

Each trial began with a fixation cross in the center of the screen for 300-ms followed by a 200-ms presentation of one of the target stimuli, which

was then followed by a search field. The search
field was divided into four quadrants by two lines
bisecting the screen horizontally and vertically
present during the search phase only. Nine target
positions were defined in each quadrant for a total
of 36. One target and 35 distractors (the other
three shapes) were presented in three circular
patterns of increasing eccentricity. For each
quadrant, two shapes appeared at the closest eccen-
tricity (20-mm from the fixation point), three at
the middle (40-mm from the fixation point), and
four at the farthest eccentricity from the center
(60-mm from the fixation point). Each of the 4
targets (small square, small diamond, large square,
large diamond) was represented in each of the 3
eccentricities, the 3 depths, and the 4 quadrants
only once per block giving a total ^f 144 trials
per block. Location of targets and distractors was
randomly determined throughout the block of trials.

Each subject was first to search for the
target and respond as soon as possible when the
target was found by pressing the left button on the
"search" mouse with one hand. Upon key press, a
square divided into four quadrants appeared on the

screen. The subject was then to move a cursor to the quadrant corresponding to the location where the target appeared by moving the "location" mouse and clicking the center button with the other hand. The subject was allowed two seconds to make this second response. If the target was not found after tnree seconds, the program scored the trial as "no response" and moved on to the next trial.

Discrimination Task

A discrimination task was also presented as a control to ascertain whether asymmetries shown in the search task were caused by discriminability differences rather than the asymmetries in the search process.

As in the search task, a target shape appeared for 200-ms, but was followed by a 200-ms presentation of only one target or distractor which randomly appeared at one of the 36 locations and three depths. The subject was instructed to determine whether the shape in the search field appeared the same in size and shape as the target previously presented. Responses were indicated with the "search" mouse. On half of the trials, the shape was the same as the target; on the other

half a distractor shape was presented.

Procedure

Each subject participated in a one hour session for seven days. During the first five sessions, the subjects completed the search task (three blocks per session). The discrimination task was completed during the last two sessions (two blocks per session). Each block consisted of 144 trials and took approximately 15 minutes to complete. Intertrial duration was self-paced. All experiments were conducted in a dark laboratory. Instructions for both of the tasks emphasized both speed and accuracy.

In the first session, subjects were familiarized with the search task with a gradual exposure to an increasing number of distractors in the search field. In the second session subjects completed three blocks of trials with the full thirty-six objects in the search field. Data was gathered during the third session. The fourth and fifth sessions were the same as the second and third except the subjects switched response hands to control for hand dominance effects either in detection response or localization response.

For each of the last two sessions, the discrimination task, subjects completed two blocks of practice and two blocks of actual tests. Hand of response was switched between the sixth and seventh sessions.

For both the search and the discrimination tasks, the order of response hand was counter-balanced across subjects.

Design

The experimental design included four main factors; vertical half-fields (2), lateral half-fields (2), depth (3), and eccentricity (3).

RESULTS

Data were analyzed by means of a 5 factor completely repeated measures analysis of variance (ANOVA). Separate ANOVAs were performed for RT and percent correct. Only RT of correct responses were analyzed.

Search Task

Reaction Time

For the search task RT data, all 4 main effects were statistically significant. Results indicated a main effect of vertical half-field $[F(1,9)=17.46; p<.0024]$, lateral half-field

[F(1,9)=5.09; p<.0506], depth [F(2,18)=5.81;

p<.0113], and eccentricity [F(2,18)=133.08;

p<.0001]. When targets were located in the UVF,

average RT (1109.67-ms) was significantly shorter

than those in the LVF (1256.96-ms); and those in

the right visual field yielded faster RT (1155.32-

ms) than the left visual field (1211.31-ms). The

Duncan's Multiple Range post hoc test on the depth

manipulation revealed that the RTs for near

(1209.36-ms) and far (1227.00-ms) targets were not

significantly different, but the coplaner target

RTs (1113.60-ms) were significantly faster than

both the near and far targets. A post hoc test on

the eccentricity effect showed a difference between

all three levels. Targets appearing in the center,

the middle, and the outer eccentricities resulted

in average RTs of 924.80-ms, 1204.01-ms, 1421.13-

ms, respectively.

A two-way interaction between depth and

eccentricity was also significant [F(4,36)=3.82;

p=.0110]. The near and far conditions both gave a

longer RT as eccentricity increased than the

coplaner conditions. The near targets presented on

the center eccentricity resulted in longer RT than

the far targets, but for targets presented on the middle and outer eccentricities, the far targets resulted in a slower response than the near targets. This interaction which shows a cross over between the near and far targets may simply reiterate the lack of a significant difference between the near and far targets.

Percent Correct

The accuracy data for the search task resulted in three of the four main effects reaching significance. Results indicated a main effect of lateral half-field [$F(1,9)=9.22$; $p<.0141$], depth [$F(2,18)=16.8$; $p<.0001$], and eccentricity [$F(2,18)=9.88$; $p<.0013$]. Subjects were significantly more accurate on targets presented in the right visual field (91.296%) than the left (89.097%). The lack of a significant difference in accuracy between UVF and LVF [$F(1,9)=1.24$; $p<.2942$] suggests that there is no speed/accuracy trade-off involved in this kind of task. Post hoc tests on eccentricity showed that accuracy for the targets placed on the outer ring was significantly lower (87.465%) than for the middle ring (90.694%) and the inner ring (92.431%) which did not signifi-

cantly differ from each other. Analysis of the

depth manipulation showed the near stimuli

resulting in significantly lower accuracy (86.354%)

than the coplaner (92.743%) and the far targets

(91.493%) which were not significantly different

from each other.

Discrimination Task

<u>Reaction Time</u>

Results indicated one main effect of

eccentricity [$F(2,18)=12.18$; $p<.0005$]. Post hoc

tests showed that the outer ring was associated

with significantly slower RT (602.631-ms) than the

middle (588.651-ms) and inner ring (582.436-ms)

which were not significantly different from each

other.

<u>Percent correct</u>

Discrimination accuracy data gave no main

effects but did show one three-way interaction

between depth, vertical half-fields, and

eccentricity [$F(4,36)=3.97$; $p<.0090$]. This small

effect seems to be a chance result of little

interpretable meaning.

DISCUSSION

These findings are consistent with Previc's

model of visual processing indicating that the UVF
is specialized for visual search involving feature
conjunction (and thus focal processing), and the
right visual field is specialized for focal
processing. Specifically, results indicated that
detection of feature conjunction targets which
utilize focal processing mechanisms were detected
faster in the UVF with respect to the LVF, and in
the right visual field with respect to the left
visual field.

The depth manipulation did not give the
predicted results that near targets would be
detected faster and more accurately in the LVF and
far targets in the UVF as there was not a
significant interaction between depth and vertical
half-field. But the coplaner depth was did result
in significantly faster and more accurate
performance than the others. This may be due
to the initial presentation of the target appearing
at the coplaner depth, i.e. subjects were primed
for the coplaner depth resulting in faster RT for
coplaner targets.

The effect of eccentricity was consistent with
predictions; specifically, RT was slower and less

accurate for targets presented farther away from
the fixation point than for those presented close
to the fixation point.

Previc's model predicts visual processing
asymmetries as resulting from processes involved in
visual search. The discrimination task's lack of
significant main effects indicates that differences
in visual field processing are in fact due to
effects of search rather than discriminability
processes.

In conclusion, these results support Previc's
model defining distinct functional asymmetries
of processing along the vertical axis (UVF and LVF)
and the lateral axis (right and left visual
fields).

made my summer enjoyable and rewarding.

REFERENCES

Delis, D. C., Robertson, L. C. & Efron, R. (1986). Hemispheric specialization of memory for visual heirarchial stimuli. Neuropsychologia, 24, 205-214.

Gulick, W. L. & Lawson, R. B. (1976). Human Stereopsis: A psychophysical analysis. New York: Oxford University Press.

Hellige, J. B. & Cox, P. (1976) Effects of concurrent verbal memory on recognition of stimuli from the left and right visual fields. Journal of Experimental Psychology: Human Perception and Performance, 2, 210-221.

Heuer, H., Dunkel-Abels, G., Bruwer, M., Kroger, H., Romer, T. & Wischmeier, E. (1988) The effects of sustained vertical gaze deviation on the resting state of the vergence system. Vision Research, 28, 1337-1344.

Miles, W. R. (1929) Ocular dominance demonstrated by unconscious sighting. Journal of Experimental Psychology, 12, 113-126.

Polich, J. M. (1982) Hemispheric Differences for Visual Search: Serial vs. Parallel Processing

Revisited. Neuropsychologia, 20(3), 297-307.

Previc, F. (1990) Functional specialization in the
     lower and upper visual fields in humans: It's
     ecological origins and neurophysiological
     implications. Behavioral and Brain Sciences,
     13, 519-575.

Treisman, A. (1982) Perceptual Grouping and
     Attention in Visual Search for Features and for
     Objects. Journal of Experimental Psychology:
     Human Perception and Performance, 8, 194-214.

Treisman, A., & Gelade, G. (1980) A feature
     integration theory of attention. Cognitive
     Psychology, 12, 97-136.

Treisman, A. M. & Schmidt, H. (1982) Illusory
     conjunctions in the perception of objects.
     Cognitive Psychology, 14, 107-141.

Treisman, A., Sykes, M., & Gelade, G. Selective
     attention and stimulus integration. In S.
     Dornic (Ed.), Attention and Performance, VI.
     Hillsdale, N. J.:Erlbaum, 1977.

Yund, E. W., Efron, R., & Nichols, D., R. (1990)
     Detectability Gradients as a Function of Target
     Location. Brain and Cognition, 12, 1-16.

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

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.

The Relationship Between Serum
2,3,7,8-tetrachlorodibenzo-para-dioxin Blood Levels and
Psychological Problems: Social Class, Exposure to Toxins and
Mental Distress*.

Russell P.D. Burton
Department of Sociology
Kent State University
Kent, Ohio 44240

The Relationship Between Serum 2,3,7,8-tetrachlorodibenzo-para-dioxin Blood Levels and Psychological Problems: Social Class, Exposure to Toxins and Mental Distress*.

Russell P.D. Burton
Department of Sociology
Kent State University
Kent, Ohio 44240

ABSTRACT

This paper has expands on the work of the 1991 AFHS and explores, with newly constructed dependent measures, with updated assumptions regarding consistency across measures, and with more fully specified models, the association between dioxin and mental distress. Results differ considerably from the 1991 AFHS report. While, by no means conclusive, results produced from these analysis strongly suggest that the body burden of dioxin is related to mental distress.

INTRODUCTION

Little is known about the short or long term psychosocial effects of acute or chronic exposure to 2,3,7,8-tetrachlorodibenzo-para-dioxin (dioxin) in humans. The Air Force Health Study (AFHS) is the only study which allows the examination of the long term psychosocial effects of chronic exposure to dioxin. The AFHS is an on going 20 year longitudinal study of the current health of Air Force veterans of Operation Ranch Hand. This unit was responsible for the aerial spraying of herbicides in Vietnam, one such herbicide, of a general class of herbicides which contained trace contaminants of dioxin, was "Agent Orange." These men ether flew or maintained the aircraft or directly handled these contaminated herbicides for a period of one to three years. Thus, the AFHS which began in 1982, some 12 years after the discontinued use of these herbicides, presents an excellent opportunity to examine the long term effects of exposure to these herbicides and, in particular, their toxic contaminant dioxin.

The AFHS is an extremely large undertaking and in the past 10 years the researchers involved in this project have published and presented several papers and technical reports regarding the possible effects of exposure to these contaminated herbicides (e.g. Wolfe et al. 1990; Michalek et al. 1990; Michalek et al. 1992). This project was directed at the examination of the effects of dioxin in 11 general outcome areas with over 250 individual dependent measures.

The last technical report represented a marked improvement in the study of exposure to dioxin (Roegner et al. 1987). Prior to this report no direct measure of exposure to dioxin was available. Thus, research was limited to the classic epidemiological method of comparing those individuals thought to be exposed (i.e. Ranch Hands) to control individuals (referred to as comparisons) who were not occupationally exposed to herbicides or dioxin in Vietnam. However, between 1985 and 1987 a technique was introduced which allowed the direct measurement of exposure through the examination of individual blood samples, for levels of dioxin. Because dioxin has a rather long half life in the body (about 7 years) current levels and original exposure levels could be obtained, with reasonable accuracy. Thus, an important methodological improvement was made and for the first time true dose response estimates could be made.

While this improvement in measurement has contributed immensely to the understanding of the effects of dioxin, a careful review of the psychological assessment section of the 1991 AFHS report suggests three other areas where additional improvement could be made. First, dependent measures, which were designed to examine psychological problems, should be constructed consistent with their theoretical development. Second, outdated assumptions regarding consistencies across three different types of psychological instruments should be dropped. Finally, research models should be constructed based on current theoretical foundations in the sociology of mental distress. The purpose of

this paper is to show that after these changes are implemented the relationship between dioxin on psychological problems will no longer be "inconsistent and isolated", as stated by Roegner, et al. (1991: 1_9), but consistent and prevalent.

This paper is two-fold in its focus. First, research models based on current research regarding mental distress are constructed. Second, the 1987 AFHS data is reanalyzed. The results are discussed in light of the previous theoretical arguments and hypothesis.

RESEARCH MODELS

While most models are misspecified to a certain extent, misspecification of research models leads to biased estimates and should be avoided wherever possible (Pedhazur, 1982). In the psychological section of the AFHS 1991 technical report, only five independent variables were considered for candidates in the construction of their research models: age, race, education, current alcohol consumption and history of alcohol consumption. A brief review of the literature of the epidemiology and sociology of mental distress (see below) suggests that other, common but critical, independent variables should have been included (e.g. individual income).

SOCIAL STATUS AND MENTAL DISTRESS

One of the strongest and most consistent relationships noted in the epidemiology of mental distress is that social status is

negatively related to mental distress (Kessler, 1982)[1]. Two general arguments have been developed as possible explanations for this observed relationship, they are termed social drift and social causation.

The drift or selection argument suggests that psychological disorders result in a drifting of mentally distressed individuals to downward social status positions or the climbing up in social status of non mentally distressed individuals. In both cases the result is a residue of mentally distressed individuals at the lower social status positions. Alternatively, however, the social causation argument suggests that there is something inherent in the lower social positions, such as traumatic life events or intellectual flexibility, which causes mental distress. Again, the lower social statuses have higher levels of mental distress. While both arguments are compelling, data and theoretical details required to sort out such discussions are extremely complex. However, the 1982 work of Kessler has, to some extent, helped clarify this relationship.

Kessler utilized eight broad based normal population surveys in his examination of the effects of differing aspects of social status on mental distress (Kessler 1982). Considering individual income, family income, education and occupational prestige Kessler noted that for males the strongest and essentially the

---

[1] From the outset we should distinguish social status from social class; the latter refers to one's relationship to the means of production and by the former we limit our discussion to education, income and occupational status.

only social status predictor of mental distress was that of individual income (Kessler, 1982). For, females however, the strongest effect was demonstrated by education. Kessler concludes his paper by linking these results to the classic drift vs. social causation debate.

Kessler notes that because the social selection argument emphasizes achieved social status, such as individual income or occupational prestige, the results obtained for men, in his eight surveys, are more constant with the social selection argument than with the social causation argument. That is, one would anticipate, if achieved social status is related to mental distress, that individuals move beyond their more "ascribed" social status (i.e. education) and in doing so one would then observe that income becomes a stronger predictor of mental distress then education.

Alternatively, because several aspects of the social causation argument emphasize socialization processes which occur relatively early in the life course, such as, intellectual flexibility or coping responses, Kessler suggests that the social selection argument may be more consistent with his results noted for women (Kessler, 1982). That is, if there is something inherent about being in the lower social statuses such as receiving socialization which results in low intellectual flexibility, and this intellectual flexibility has an impact on how one responds to stress, and is this intellectual flexibility is relatively fixed for life, one would then expect that

education would be a stronger a predictor of mental distress than income.

Thus, because the data for this paper employ only men we would anticipate, whether one accepts Kessler's social selection and causation explanation or not, that income would be the strongest and only social status predictor of mental distress.

SOCIAL ROLES AND MENTAL DISTRESS

One of the more recent areas of investigation in the epidemiology of mental distress is that of social roles. Work by Thoits (1983, 1986), Menaghan (1989), Burton, Rushing and Ritter (1992) and others have demonstrated that, in general, social roles such as being married and employed, are negatively related to mental distress. While there are some differences by race and gender, the overall results suggest that the social roles of being married and employed are beneficial to mental health (Reskin and Coverman 1985; Mirowsky and Ross 1989a). However, the effects of being a parent and having the children living in the household seems to be positively related to mental distress for women but unrelated to mental distress for men (Mirowsky and Ross 1989a).

Thus, in the AFHS one would expect that the social roles of employee (i.e. working), and spouse (i.e. being married) to be strongly and negatively related to mental distress. One would also anticipate that being a parent and having the child in the home would not be related to mental distress or would be slightly positively related to mental distress.

ALCOHOL CONSUMPTION AND MENTAL DISTRESS

The relationship between alcohol consumption and mental distress is both theoretically and empirically unclear. Theoretically individuals who are depressed, stressed or anxious may be motivated to drink alcohol to alleviate these unwanted feelings (Williams 1970). Thus, current drinking and particular problem drinking may be associated with depression and anxiety. Recent research indicates that at least in the immediate short term (i.e. while under intoxication), in particular situations, at particular times of the week and the day, some stress reduction may occur (Orcutt 1991). However, long term drinking particularly at higher uses may result in a exacerbation of mental distress rather than a reduction (Biener, 1987).

While somewhat unclear, for these data, give the crossectional nature and the time frame, we would expect that both current drinking and drinking history should be positively related to mental distress.

HYPOTHESES

The following list of anticipated relationships should be demonstrated with these data.

1.) Income will not only be negatively related to mental distress but it will be the only aspect of social status which is related to mental distress.

2.) Conversely education will have no effect on mental distress.

3.) Two social roles, that is, being an employee (i.e. working)

and being a spouse (i.e. married), will be negatively related to mental distress.

4.) The social role of being a parent (i.e. having a child) will not be related to mental distress or will have a marginal positive effect.

5.) Both current drinking and history of drinking will be positively related to mental distress.

6.) Dioxin will have a positive effect on mental distress.

SAMPLE

During the Vietnam war a group of individuals were involved in a military operation, code named Operation Ranch Hand. The overall responsibility of these individuals, from January 1962 to October 1971, was the aerial spraying of defoliant herbicides. The primary herbicide, of a general class of herbicides, which contained trace contaminates of dioxin, was code named Agent Orange. The AFHS is an ongoing epidemiological investigation to determine the adverse effects of exposure to these herbicides for the Ranch Hand group. The AFHS was designed as a 20 year longitudinal study starting in 1982 with 5 follow ups to occur at 3 years (1985), 5 years (1987), 10 years (1992), 15 years (1997), and 20 years (2002). What follows is a brief explanation of the population and sampling procedures employed, for more information see Roegner, et al. 1991).

All Ranch Hands were matched to at lest one Comparison but up to 10 Comparisons, based on age, race, and military occupation. The Comparison subjects were drawn from a group of

3-10

Air Force veterans involved in C-130 aircraft missions in Southeast Asia. It was determined that these individuals were suitable as Comparisons because they received similar training, served in Southeast Asia, had sufficient numbers and conducted similar work except that they were not occupationally exposed to herbicides.

RESULTS

Tables 1 through 3 present results from the analysis of the SCL-90-R. Note that all regression results presented through out this paper, except those noted, are in their unstandardized format. Therefore, direct comparison may be made on the same variable across different populations with the same dependent variable, Comparisons may not be made however, across different variables within the same or different populations irrespective of the dependent variable. Table 1 displays the number of valid responses, the means and standard deviations on all variables for all Ranch Hands and all Comparisons. These two groups were matched on age, race and education and we find no significant variation in these variables across the two groups. In general all variables, except drinking history, combat exposure and dioxin levels, are not significantly different across the two groups.

Table 2 Panel A introduces the multivariate analysis of the SCL-90-R for the Comparison group and Table 2 Panel B displays the multivariate equations for all Ranch Hands. With in each cell the superscripts a-d represent the probability values, in

descending order (a = .001; b = .01; c = .05; d = .1), for
testing the hypothesis that the slope for the variable in that
cell is equal to zero. When the superscript e is found in a cell
that indicates the slope for that variable is significantly
different (p < .1) for Ranch Hands and Comparisons; that is,
their exists an interaction between that variable and the two
groups (i.e. Ranch Hands and Comparisons). Additionally, the R^2
for every equation is reported at the far right of each row.

Recalling that these two groups were matched on age, race
and level of education, the most important information to be
drawn from Table 2 is the overall differences between the
equations in Panel A and Panel B. There are two general levels
of differences across these two panels. First, there are the
differences which are statistically significantly across the two
groups (i.e. the cells with e in them). Second are the non-
significant differences which demonstrate constant patterns of
differences across the two groups. While the first level of
differences are clearly what we should rely on to draw our
conclusions, at this early stage in the analysis, it is some
times also helpful to note constant trends in the data. Both at
the level of significance and the trend level we note striking
differences in these two sets of equations, i.e. Panel A and
Panel B.

By far the most obvious difference across these two matched
groups is the differential effects of education and income, with
eleven of the twelve possible interaction terms for education

being significant and seven of the ten possible interaction terms for income being significant. Panel A demonstrates the hypothesized relationships, for a group of males, between individual income, education and mental distress. That is, drawn from the work of Kessler (1982) we anticipated that individual income would be, if not the only, the major social status predictor of mental distress (see hypothesis 1 and 2 above). However, we note that the relationship is completely reversed for the Ranch Hands. That is, for the Ranch Hands, education is, in most cases the only, and in all cases the strongest social status predictor of mental distress (standardized coefficients not shown). We begin to note the first in a series of differences between these two groups and thus the potential effects of exposure to Agent Orange on the Ranch Hands.

Directing the focus to the effects of social roles, it is clear that the next significant difference between these two groups of equations is the extremely strong, positive and consistent effect of being a parent on mental distress observed in the Ranch Hand group but not in the Comparison group. In fact, of the possible ten interaction terms for the role parent, across these two equations, six are significantly different. Thus, it seems, once again that the hypothesized effects of the independent variables in the model (hypothesis 3 and 4) are supported for the Comparison group, but are not supported, for the parent and working roles, in the Ranch Hand group.

In reference to drinking history the effects in both groups are as anticipated. Current drinking has little effect on mental distress in both groups, but the effects are reversed by group. For the Comparison group current drinking is positively related with mental distress, as anticipated, but for the Ranch Hands the effects are negatively related to mental distress. It is noted again that support is found in the Comparison group but not in the Ranch Hand group, for the current drinking hypotheses (i.e. hypothesis 5).

To review the result thus far, it is clear that the Ranch Hand and Comparison groups, although matched on age, military occupation and race, produce completely different equations regarding the prediction of mental distress. Significant differential effects, by group, on mental distress for education, income and the parent role, as well as a strong but non significant trend in the differential effect of the employment role, have been noted. Because a critical covariate which was not adjusted for across these two groups is dioxin levels, assumed to be the result of exposure to Agent Orange, it is possible, but certainly not conclusive, that exposure to Agent Orange and the resulting dioxin levels may be driving these differential effects. The effects of dioxin are considered next.

Table 3 presents these equations. The most interesting aspect of these equations is the fact that dioxin has a significant positive effect for almost all mental distress outcomes. In fact dioxin is a marginal predictor in 3 of the 12

(p < 0.1) equations but is a strong predictor in 6 of the remaining 9 equations (p < 0.05). Such effects are beyond any limits set by random variation with several outcome variables and thus we find clear support for hypothesis 6 regarding the positive effects of dioxin on mental distress. Current drinking is negatively related to mental distress. Additionally drinking history is positively related to mental distress. A marginal and negative effect of education and marital status is also noted.

Thus, a clear dose response between mental distress and dioxin appears to exist at 15 PPTR and greater. At this point it is clear that dioxin body burdens of greater than 15 have a significant effect on mental distress.

CONCLUSION

In conclusion this paper has expand on the work of the 1991 AFHS and explored, with newly constructed dependent measures, with updated assumptions regarding consistency across measures, and with more fully specified models, the association between dioxin and mental distress. Results differ considerably from the 1991 AFHS report (see Roegner et al 1991). While, by no means conclusive, results produced from these analysis strongly suggest that the body burden of dioxin is related to mental distress.

REFERENCES

Burton, Russell P.D., Beth Rushing and Christian Ritter. 1992. "Roles, Race and Subjective Well-Being: A Longitudinal Analysis of Elderly Men." Forthcoming in Social Indicators Research.

Hollingshead, August B. and Frederick C. Redlich. 1953. "Social Stratification and Psychiatric Disorders." American Sociological Review 18:163-169.

Kessler Ronald C. 1982. "A Disaggregation of the Relationship Between Socioeconomic Status and Psychological Distress." American Sociological Review. 47:752-764.

Menaghan, Elizabeth G. 1989. "Role Changes and Psychological Well-Being: Variations in Effects by Gender and Role Repertoire." Social Forces. 67:693-714.

Michalek, Joel E., Ram C. Tripathi, Samual P. Caudill and James L. Pirkle. 1992. "Investigation of TCDD Half-Life Heterogeneity in Veterans of Operation Ranch Hand." Journal of Toxicology and Environmental Health. 35:29-38.

Michalek, Joel E., William H. Wolfe and Judson C. Miner. 1990. "Health Status of Air Force Veterans Occupationally Exposed to Herbicides in Vietnam II. Mortality." Journal of the American Medical Association. 264:1832-1836.

Mirowsky, John and Catherine E. Ross. 1989a. Social Causes of Psychological Distress. Aldine de Gruyter, New York.

Mirowsky, John and Catherine E. Ross. 1989b. "Psychiatric Diagnoses as Reified Measurement." Journal of Health and Social Behavior. 30:11-25.

Reskin B. and S. Coverman. 1985. "Sex and Race in the Determinants of Psychological Distress: A Reappraisal of the Sex-Role Hypothesis." Social Forces. 63:1038-57.

Roegner, Russell H., Willian D. Grubbs Michael B. Lustik, Amy S. Brockman Scott C. Henderson, David E. Willians, William H. Wolfe, Joel E. Michalek and Judson C. Miner. 1991. Air Force Health Study: An Epidemiologic Investigation of Health Effects in Air Force Personnel Following Exposure to Herbicides. Report Number AL-TR-91-0009.

Stellman, Jeanne M., Steven D. Stellman, and John F. Sommer, Jr. 1988. "Social and Behavioral Consequences of the Vietnam Experience Among American Legionnaires." Environmental Research. 47:129-149.

Thoits, Peggy A. 1983. "Multiple Identities and Psychological Well-Being: A Reformulation and Test of the Social Isolation Hypothesis." American Sociological Review. 48:174-187.

Wolfe, William H., Joel E. Michalek, Judson C. Miner, Alton Rahe, John Silva, Wanda F. Thomas, William D. Grubbs, Michael B. Lustik, Theodore G. Karrison, Russell H. Rogner and David E. Williams. 1990. "Health Status of Air Force Veterans Occupationally Exposed to Herbicides in Vietnam I. Physical Health." Journal of the American Medical Association. 264:1824-1831.

TABLE 1

DESCRIPTIVES FOR ALL VARIABLES

FOR ALL RANCH HANDS AND COMPARISONS

	RANCH HANDS			COMPARISONS		
	N	$\bar{x}$	SD	N	$\bar{x}$	SD
SOMAT	854	828.899	470.615	783	808.203	464.449
OBSES	854	828.296	468.450	783	808.061	462.697
INTERP	854	815.391	448.333	783	822.937	459.310
DEP	854	815.708	461.500	783	822.590	221.153
ANX	854	820.937	440.809	783	816.804	440.089
HOST	854	826.197	435.092	783	811.150	433.401
PHOB	854	816.383	343.322	783	821.854	358.330
PAR	854	809.994	417.098	783	828.822	433.060
PSYC	854	814.846	416.376	783	823.531	428.853
GSI	854	817.227	471.208	783	820.934	471.776
PSDI	854	819.901	463.231	783	818.017	456.972
PST	854	817.369	468.264	783	820.779	475.382
EDUC	848	.482	.500	778	.500	.500
AGE	854	43.796	7.538	783	44.027	7.956
BLACK	854	.052	.221	783	.060	.238
DRINK Y	846	2.236*	1.950	781	2.438	1.817
DRINK D	850	-1.010	1.181	783	-.920	1.209
INC	854	7.054	4.174	783	6.913	3.980
COM	852	4.937*	2.311	772	3.812	2.446
WORK	854	.871	.335	783	.870	.337
MAR	854	.840	.367	782	.857	.351
PAR	854	.453	.498	783	.469	.499
LOGPPT	854	3.868*	1.637	783	1.820	1.337

* = Significant Differences Across Groups (P<.1)

TABLE 2 - PANEL A

SLC-90-R MEASURES OF MENTAL DISTRESS

REGRESSED ON ALL SIGNIFICANT COVARIATES

FOR COMPARISONS ONLY (N = 764)

SCL-90 SCALES	EDUC	AGE	BLACK	DRINK Y	DRINK D	INC	COM	WORK	MAR	PAR	R^2
SOMAT	•	•	•	•	•	-8.708[ae]	8.165[c]	-64.776[b]	•	•	.053
OBSES	•	•	•	11.864[be]	•	-4.466[c]	•	-63.349[b]	•	•	.026
INTERP	•	•	•	•	15.050[ce]	-7.181[ae]	•	•	-52.495[c]	•	.030
DEP	•	•	•	8.298[d]	e	•	•	-71.209[b]	-53.997[c]	e	.022
ANX	•	•	•	10.004[c]	•	-7.273[ae]	•	•	-58.326[b]	e	.035
HOST	•	•	•	9.979[c]	•	•	•	•	•	e	.006
PHOB	•	•	•	•	•	-3.616[c]	•	-32.105[d]	•	32.735[b]	.014
PAR	•	•	•	7.711[d]	•	-5.719[be]	•	•	-50.246[c]	•	.023
PSYC	•	•	•	•	•	-5.731[be]	•	-38.849[d]	-50.382[c]	e	.024
GSI	•	•	•	8.848[c]	•	-7.494[ae]	•	-48.542[c]	-42.086[d]	e	.053
PSDI	•	•	•	•	•	-4.302[c]	•	-43.452[d]	•	•	.009
PST	•	•	•	10.242[c]	•	-7.332[ae]	•	-41.335[d]	-43.851[d]	e	.033

a = P≤.001
b = P≤.01
c = P≤.05
d = P≤.1
e = Interaction term is significant with Ranch Hands (P≤.1)
• = Variable was excluded from the equation (P>.1)

TABLE 2 - PANEL B

SCL-90-R MEASURES OF MENTAL DISTRESSS

REGRESSED ON ALL SIGNIFICANT COVARIATES

FOR RANCH HANDS ONLY (N = 838)

SCL-90 SCALES	EDUC	AGE	BLACK	DRINK Y	DRINK D	INC	COM	WORK	MAR	PAR	R^2
SOMAT	-117.613[ae]	.	.	.	.	.[e]	7.691[c]	.	.	.	.053
OBSES	-89.144[ae]	.	.	.[e]	.	.	.	.	.	28.818[d]	.035
INTERP	-53.382[ae]	.	.	.	.	.[e]	.	.	-49.176[c]	35.446[c]	.021
DEP	-74.994[ae]	.	.	12.211[c]	-14.830[d]	6.575[d]	.	.	-71.835[b]	49.852[be]	.047
ANX	-74.080[ae]	.	.	10.565[b]	.	.[e]	.	.	-41.801[d]	42.691[be]	.042
HOST	-65.181[ae]	.	.	6.994[d]	.	.	6.738[d]	.	.	54.399[ae]	.033
PHOB	-44.372[a]	.	.	.	.	-2.836[d]	.	.	.	23.211[d]	.026
PAR	-66.065[ae]	.	.	.	.	.[e]	.	.	-63.434[b]	39.276[b]	.038
PSYC	-64.306[ae]	.	.	.	.	.[e]	.	.	-70.746[a]	37.058[ce]	.039
GSI	-94.671[be]	.	.	.	.	-4.171[ce]	7.688[c]	.	-47.122[c]	37.873[ce]	.053
PSDI	-81.388[ae]	.	.	.	.	.	8.218[c]	.	.	.	.027
PST	-90.473[ae]	.	.	.	.	-4.232[ce]	6.504[d]	.	-50.657[f]	42.148[ce]	.027

a = P≤.001
b = P≤.01
c = P≤.05
d = P≤.1
e = Interaction term significant with Comparisons (P≤.1)
- = Variable was excluded from the equation (P>.1)

TABLE 3

SCL-90-R MEASURES OF MENTAL DISTRESS REGRESSED ON ALL SIGNIFICANT COVARIATES

AND LOG DIOXIN PARTS PER TRILLION (PPTR) FOR RANCH HANDS WITH PPTR >15 (N = 370)

3-20

SCL-90 SCALES	LOGPPTR	EDUC	AGE	BLACK	DRINK Y	DRINK D	INC	COM	WORK	MAR	PAR	TIME	R^2
SOMAT	5.377	-	-	-	5.730[d]	-17.351[b]	-	-	-	-	-	-	.017
OBSES	10.803[c]	26.058[ce]	-	-	6.517[c]	-20.472[ae]	e	-	e	-	-	-	.051
INTERP	13.853[b]	-	-	-	-	-12.457[be]	e	-	-	-34.730[c]	-	-	.045
DEP	7.029	-	-	-	9.673[b]	78.075[be]	-	-	-	-25.760[d]	-	-	.033
ANX	9.144[d]	-	-	-	-	-	e	-	-	-	-	-	.006
HOST	8.046	-19.584[d]	-	-	-	-	-	-	-	-	21.338[ce]	-	.022
PHOB	7.016[d]	-	-	-	-	-	-	-	-	-	-	-	.006
PAR	10.840[c]	-18.135[d]	-	-	-	-	e	-	-	-28.648[c]	-	-	.028
PSYC	10.468[c]	-	-	-	6.659[c]	-16.552[be]	e	-	-	-38.778[b]	-	-	.050
GSI	11.537[c]	-20.233[d]	-	-	8.734[b]	-19.394[be]	e	-	-	-	-	-	.042
PSDI	8.308	-	-	-	9.491[b]	-21.988[ae]	-	4.991[c]	-	-	-	-	.046
PST	11.875[c]	-21.712[de]	-	-	7.158[c]	-17.498[be]	e	-	-	-27.545[d]	-	-	.045

a = P≤.001
b = P≤.01
c = P≤.05
d = P≤.1
e = Interaction term significant with Comparisons (P≤.1)
- = Variable was excluded from the equation (P>.1)

In Utero Taste Aversion Conditioning in the Rat Fetus

Jed S. Cohen
Summer Research Assistant
Department of Biology

Trinity University
715 Stadium Drive
San Antonio, Texas 78212

Summer Research Program
The Armstrong Laboratory
Brooks AFB, TX

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

September 1992

In Utero Taste Aversion Conditioning
in the Rat Fetus

Jed S. Cohen
Summer Research Assistant
Department of Biology
Trinity University

Abstract

The ability to teach rat fetuses a taste aversion *in utero* was examined. Fetuses were conditioned by exposing them to a .15% saccharin solution and to a .8% lithium chloride solution. Initial experimental results indicate that saccharin consumption in all groups was low, possibly due to neophobia. In later test weeks, however, the saccharin and lithium chloride treatment group was found to consume greater quantities of saccharin. Although, the aforementioned group consumed much less saccharin than the other groups, they consumed far more than expected for animals with a conditioned taste aversion. If the protocol did lead to a successful taste aversion, its presence was very subtle. As of this writing, the protocol is being adjusted to ensure the presence of a taste aversion. Future experiments will focus on the ability to transfer the taste aversion from a trained fetus to a naive adult male rat in order to refine neural transplantation techniques.

INTRODUCTION

It has been shown that newborn rat pups are capable of odor and taste aversive learning *in utero*. Stickrod, Kimble, and Smotherman (1982) have shown that aversive learning abilities develop prenatally. They have reported that rat fetuses will form a taste/odor aversion when exposed to apple juice *in utero* followed by lithium chloride. Stickrod further demonstrated that fetal rats can acquire and retain a conditioned taste aversion. He showed that rat fetuses at 20 days of gestation are capable of associative learning which can be demonstrated more than two weeks after birth. The purpose of this experiment is to confirm whether or not a taste aversion can be learned *in utero* and to confirm its effectiveness. We are expanding this original paradigm by initiating conditioning on the 18th day of gestation, testing more than 20 day after birth, and relying on different concentrations of the conditioned stimulus and unconditioned stimulus to determine their effects. Subsequent, more ambitious experiments will examine the feasibility of using neural transplant techniques to transfer a simple memory (the taste aversion) from a trained fetus to a naive adult male rat in order to elucidate how the brain encodes memory, refine neural transplantation techniques, and allow for a description of the neural components of a tissue containing a specific memory.

METHOD

This experiment was conducted to confirm that a taste aversion could be learned *in utero*. In order to expose the fetuses to the CS (conditioning stimulus; saccharin) and the UCS (aversive stimulus; lithium chloride), surgery was performed on pregnant, time-mated, Sprague Dawley rats on Day 18 of gestation. The pregnant female was anesthetized with sodium pentobarbital (60mg/kg). Metophane was administered via nose cone as needed to maintain surgical anesthesia. The abdomen was shaved and treated with betadine scrub and solution. Both horns of the uterus were exposed through a midline laprotomy under aseptic

conditions. A 32-gauge needle was used to inject 10 microliters of the .15% saccharin (CS) solution or distilled water (control) into the amniotic sac adjacent to the nose-mouth area of each fetus. The fetuses were then intraperitoneally injected with 5 microliters of .81% lithium chloride (UCS; 8.1 g LiCl per 1000 ml distilled water) or 0.09 % saline (control). This dose of lithium chloride is known to cause gastrointestinal difficulty in the fetus and is therefore used as the aversive stimulus in establishing the taste aversion. All injections were done under visual guidance. The uterus was replaced after the last injection and maneuvered as close as possible to its original U-shape to prevent any delivery complications at birth. The incision was sutured and treated with Lidocaine, a topical analgesic, to reduce immediate, post-surgical pain. The fetuses were allowed to come to term. Conditioned male rats were weaned and placed in their own cages 25 days after birth. At this time, the female dam and the conditioned females were euthanized using carbon dioxide gas.

Four different treatment groups (TGs) were established. TG1 received saccharin into the amniotic fluid followed by lithium chloride injected intraperitoneally (SAC/LiCl). TG2 received distilled water into the amniotic fluid followed by saline (H2O/SAL). TG3 received saccharin into the amniotic fluid and saline (SAC/SAL). TG4 received distilled water into the amniotic fluid followed by lithium chloride (H2O/LiCl).

The taste aversion testing of the conditioned males was begun immediately after weaning. The rats were allowed to eat and drink ad lib for one week after weaning and before being placed on a timed drinking schedule. After being placed on the drinking cycle, their water bottles were removed at 2 P.M. every Monday. Then, water was available for only 30 minutes per day at the same time every day for three days (Tuesday, Wednesday, and Thursday). This scheduling was required to ensure the rats would drink when liquids were offered to them on the test day. On day #4 of the cycle (Friday: Test Day #1), the rats were offered a choice between

distilled water and the saccarin solution described earlier. The 2 bottles containing the two liquids were placed at opposite ends of the cage top. The drinking solutions were placed in 50 ml centrifuge tubes (Nalgene); each tube was then stoppered with a rubber test tube stopper containing a sipper tube. The left and right bottle positions were switched three times during the drinking period (at 30 seconds, 2.5 minutes, and 3.75 minutes) in order to eliminate side preferences and fixations. The amount of each liquid consumed was recorded at the end of the 30 minute drinking period.

RESULTS

The data collected after each test day were organized and then entered into RS/1, a scientific applications spreadsheet. The data were organized along several lines: the rats were indentified by number, the type of injections they received, the date the test occurred, the amount of water and saccharin each rat consumed, and the percentage of water and saccharin consumed by each of the rats. The most important indicator of a successful taste aversion (TA) was the percentage of saccharin consumed by each rat during the test day. Data interpretation relied heavily on this figure.

Regardless of the type of injections the rats received while *in utero* their initial saccharin consumption in all treatment groups is low. (See graph #TA_EXP1). Lt. Col. Mickley attributed this phenomenon to neophobia -- although the rats have been exposed repeatedly to water prior to the test day, the test day is the first opportunity the rats have had to consume saccharin at will. Thus, they tend to consume the more familiar substance (water) in greater quantity early in the experiment. Interestingly enough, this phenomenon is also seen in the control animals which have received saccharin *in utero* with saline. Why would these animals demonstrate a neophobia to saccharin when they have had some experiences with saccharin? Although we can not explain it directly, this evidence does lend some credence to the neophobia argument.

The data specifically show a neophobia for the saccharin/saline (SAC/SAL) treatment group (control). During test week #1, for example, the SAC/SAL control group had a saccharin consumption rate of 16% that was suprisingly close to that of the H2O/SAL control group's consumption rate of 13%. (The H2O/SAL group was expected to show the lowest saccharin consumption rate because it has never been exposed to saccharin; this assumption should hold true if one gives credence to the notion of neophobia. If the taste aversion conditioning were working, the SAC/LiCl group should show the lowest saccharin consumption.)

Also unexpected was our discovery that the rats which received the lithium chloride (LiCl) injections were consuming more saccharin than their counterparts who received the saline control injections (see graph #TA_EXP1). During test week #1, the SAC/LiCl group consumed 25.9% saccharin while the H2O/SAL group -- the group expected to have the highest saccharin consumption -- had a 13.4% saccharin consumption rate. During test week #2, the SAC/LiCl group consumed 46% saccharin while the H2O/SAL group consumed only 22.3% saccharin. The H2O/SAL group almost manages to almost tie the SAC/LiCl group during test week #5 (64% vs. 64.7%) and actually manages to consume more saccharin during test week #7 (71.9% vs. 70.2%). See graph #TA_STAT for details. Based on this evidence, it is possible that LiCl, administered to cause gastrointestinal distress, may actually have an amnesic effect on the rats.

The clearest separation between the saccharin and the control groups (H2O/SAL and SAC/SAL) occurs on test week #3 (see graph #TA_EXP1). Notice that the SAC /LiCl group has the greatest saccharin consumption rate (67.7%) compared with the H2O/SAL (36.1%) and the SAC/SAL (47.2%) groups. Thus, the group that should demonstrate the taste aversion consumed more saccharin than the control groups.

The only finding consistent with the original TA hypothesis was discovered by Lt. Col. Mickley when he plotted the slopes and intercepts of each of the lines in the #TA_EXP1 graph on RS/1. The SAC/LiCl group exhibits an early saccharin consumption peak followed by a decrease in consumption. The slope for this line is more shallow than those of the control groups or for the H2O/LiCl group. The shallow slope implies that the SAC/LiCl rats do not consume as much saccharin in the long term as do the other rats because the consumption rate remain relatively stable for the duration of the eight week test cycle. Saccharin consumption rates for the other three treatment groups are more sporadic, but generally show a small initial saccharin consumption followed by an increased saccharin consumption later in the test cycle.

Because the initial results were not as promising as expected, the protocol was adjusted slightly with the saccharin and lithium chloride solutions being doubled so that their effects could be more easily seen. Early data suggest that the double concentration groups (which contained only a few rats) consume very low levels of saccharin. The consumption rates are so low, in fact, that differentiation between the experimental and control groups was virtually impossible.

CONCLUSION

If a taste aversion was indeed developed, its presence was very subtle. The only evidence we found to support the original hypothesis was that the SAC/LiCl treatment showed an early saccharin consumption peak followed by a decrease in consumption. Nevertheless, the SAC/LiCl group consumed a great deal of saccharin for animals that should have a taste aversion and, therefore, consume little, if any, saccharin at all.

The experimental protocol is being modified so that a taste aversion can be better demonstrated as in Smotherman's studies.

Taste aversion conditioning is now being done on the 20th day of gestation instead of the 18th. Lower saccharin concentrations (0.1%) are being used for training and testing purposes. Apple juice is also being used in lieu of the saccharin in some cases, an idea borrowed from Smotherman, et. al. Next, the *in utero* saccharin and lithium chloride injections are being separated in time. Initially, each fetus was injected with saccharin in the mouth area followed immediately by lithium chloride in the gut. Now, three fetuses are injected with saccharin and then the lithium chloride injections are administered. Additional planned changes include: weaning the pups earlier so that taste aversion testing can commence sooner, a one bottle (saccharin only) test which may reveal effects obscured by neophobia, and, finally, testing the animals more times during the week.

References

Alleva, E., and Calamandrei, G. Odor-aversion learning and retention span in neonatal mice. <u>Behavioral and Neural Biology</u>, 46 (1986) 348-357.

Smotherman, W.P. Odor aversion learning by the rat fetus. <u>Physiology and Behavior</u>, 29 (1982), 769-771.

Smotherman, W.P., and Robinson S.R. The rat fetus in its environment: Behavioral adjustments to novel, familiar, aversive, and conditioned stimuli presented in utero. <u>Behavioral Neuroscience</u>, 99 (3) (1985) 521-530.

Stickrod G., Kimble, D.P., and Smotherman W.P. In Utero taste/odor aversion conditioning in the rat. <u>Physiology and Behavior</u>, 28 (1982) 5-7.

Stickrod, G., Kimble, D.P., and Smotherman, W.P. Met-enkephalin effects on associations formed in utero. <u>Peptides</u>, 3 (6) (1982), 881-883.

Acknowledgements

Taste Aversion Data from Experiment 1

Taste Aversion Data from Experiment 1

#TA_FIT

Taste Aversion Test

●	SAC+LiCl	
O	H2O+Sal	
■	Sac+Sal	
□	H2O+LiCl	
─·─	0.063595*X	+0.337322
─··─	0.104438*X	+0.055829
───	0.10371*X	+0.12212
━━━	0.093149*X	+0.223744

4-12
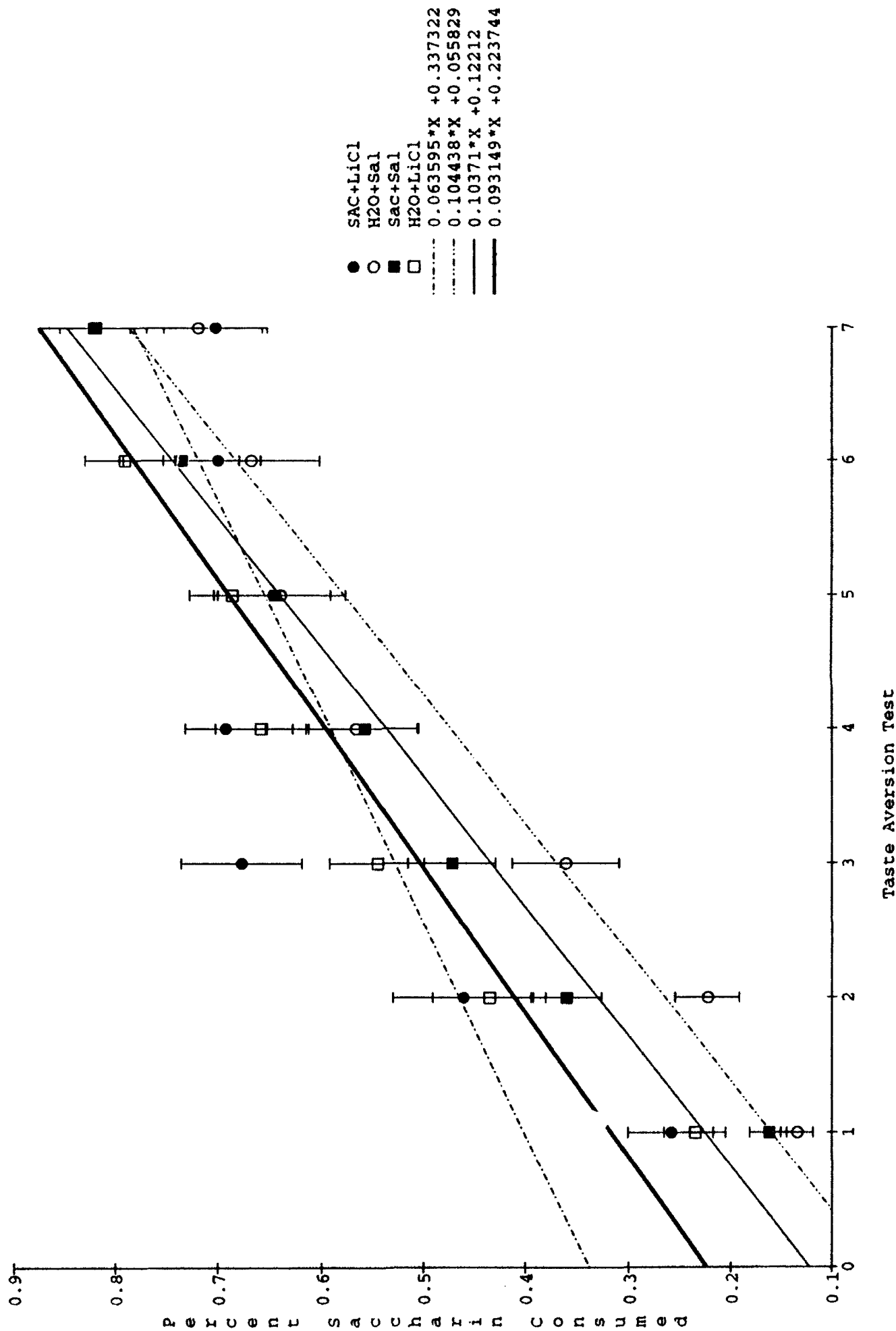
Summary of Taste Aversion Data

0 Rat No	1 Treatment	2 %SAC_T1	3 %SAC_T2	4 %SAC_T3	5 %SAC_T4	6 %SAC_T5	7 %SAC_T6
1 MEAN	1	0.258929	0.460945	0.677562	0.693538	0.647635	0.700657
2 SEM		0.041523	0.068912	0.058358	0.038676	0.056137	0.041601
3 +SEM		0.300452	0.529856	0.735920	0.732214	0.703772	0.742259
4 -SEM		0.217406	0.392033	0.619205	0.654861	0.591498	0.659056
5							
6 MEAN	2	0.134616	0.223457	0.360902	0.567650	0.640472	0.668475
7 SEM		0.016061	0.031629	0.051438	0.061023	0.063779	0.066369
8 +SEM		0.150677	0.255086	0.412340	0.628673	0.704251	0.734844
9 -SEM		0.118555	0.191828	0.309465	0.506628	0.576692	0.602107
10							
11 MEAN	3	0.162959	0.360249	0.472073	0.559031	0.645937	0.736037
12 SEM		0.018454	0.034077	0.043311	0.053906	0.054435	0.056543
13 +SEM		0.181412	0.394326	0.515384	0.612937	0.700372	0.792580
14 -SEM		0.144505	0.326173	0.428762	0.505125	0.591503	0.679494
15							
16 MEAN	4	0.235453	0.435434	0.545553	0.659269	0.686635	0.791764
17 SEM		0.030108	0.055480	0.046606	0.043510	0.041067	0.038410
18 +SEM		0.265561	0.490913	0.592159	0.702779	0.727702	0.830174
19 -SEM		0.205346	0.379954	0.498947	0.615759	0.645568	0.753354
20							
21 MEAN	5	0.183036					
22 SEM		0.031250					
23 +SEM		0.214286					
24 -SEM		0.151786					
25							
26 MEAN	6	0.135376					
27 SEM		0.028954					
28 +SEM		0.164331					
29 -SEM		0.106422					

Treatment: 1=Sac+LiCl, 2=H2O+Sal, 3=Sac+Sal,
4=H2O+LiCl, 5=Sac+LiCl (2X Concentration) 6=Sac+Sal (2X conc)
7=H2O+LiCl (2X conc) 8=AppleJ.+LiCl (2X conc)

Summary of Taste Aversion Data

0 Rat No 8 %SAC_T7

```
 1 MEAN     0.702652
 2 SEM      0.050195
 3 +SEM     0.752847
 4 -SEM     0.652457
 5
 6 MEAN     0.719506
 7 SEM      0.062061
 8 +SEM     0.781567
 9 -SEM     0.657445
10
11 MEAN     0.822441
12 SEM      0.053001
13 +SEM     0.875442
14 -SEM     0.769440
15
16 MEAN     0.820259
17 SEM      0.034832
18 +SEM     0.855091
19 -SEM     0.785427
20
21 MEAN
22 SEM
23 +SEM
24 -SEM
25
26 MEAN
27 SEM
28 +SEM
29 -SEM
```

Treatment: 1=Sac+LiCl, 2=H2O+Sal, 3=Sac+Sal,
4=H2O+LiCl, 5=Sac+LiCl (2X Concentration) 6=Sac+Sal (2X conc)
7=H2O+LiCl (2X conc) 8=AppleJ.+LiCl (2X conc)

ANIMAL EMOTIONALITY AND RHESUS MACAQUE (Macaca mulatta)
VOCALIZATIONS

James W. Collins, M.S.

Department of Psychology
University of Georgia
Athens, GA 30605

ANIMAL EMOTIONALITY AND RHESUS MACAQUE (Macaca mulatta) VOCALIZATIONS

James W. Collins, M.S.

Department of Psychology
University of Georgia
Athens, GA 30605

Abstract

A state-of-the-art computer-based bioacoustics laboratory was established for the recording, digitization and analysis of rhesus macaque (Macaca mulatta) calls, their visual display, and playback. Two trips were taken to record vocalizations under naturalistic conditions from a free-ranging rhesus colony (Morgan Island, South Carolina) maintained by L.A.B.S., in Yemassee, South Carolina. Also, recordings of rhesus monkey calls under differing experimental and control conditions were made from animals at Brooks AFB. The literature on animal emotionality was reviewed and is reported in the form of an annotated bibliography. The great promise that research into the expression of emotionality by animals via their vocalizations holds for detecting the emotional state of animals, especially primates, has not yet been tapped. It is our aim to achieve a higher level of understanding of the emotional component of animal vocalizations. Meeting this goal will significantly expand our understanding of the internal states of animals as well as enable us to comprehend future assessments of research/captive animal wellbeing.

ANIMAL EMOTIONALITY AND RHESUS MACAQUE (Macaca mulatta) VOCALIZATIONS

James W. Collins, M.S.

Department of Psychology
University of Georgia
Athens, GA 30605

During my 12 week tenure at the Armstrong Laboratory at Brooks Air Force Base in San Antonio, Texas, I participated in research projects within the Performance Extrapolation Function Division directed by Dr. Michael R. Murphy. We were interested in measuring rhesus macaque (Macaca mulatta) vocalizations, establishing a model vocal repertoire for this species, and ascertaining how these vocalizations could be used as an index of "emotionality" for rhesus monkeys.

The first task was the establishment of a state-of-the-art computer-based bioacoustics laboratory for the recording, digitization and analysis of macaque calls, their visual display, and playback. This involved the purchase, delivery, and set-up of various individual components into a single, functionally-integrated system. Components of this laboratory are: a portable digital audio recorder with microphone and power pack for field and lab vocalization recording, a digital cassette recorder (DAT) for playback and analysis of recorded audio, a Kay CSL Model 4300 computerized speech lab, an IBM compatible 486 PC AT computer with a superVGA display monitor, and a laser printer for hard copy of visually-represented calls.

Two trips were taken to record vocalizations under naturalistic conditions from a free-ranging rhesus colony (Morgan Island, South Carolina) maintained by L.A.B.S., in Yemassee, South Carolina. Also, recordings of rhesus monkey calls under differing experimental and control conditions were made from animals at Brooks AFB during my tenure. The resulting audio recordings were brought back to the lab and dissected, analyzed, grouped into acoustic and functional categories, and visually presented. This work is still in progress.

The literature on animal emotionality was reviewed (see Appendix A). Subjects examined in these studies include rodents, domestic animals, and nonhuman primates. As a topic of study since the 1930's, there are many published studies of animal emotionality. Most, however, are concerned only peripherally with the "emotionality" exhibited by the species under investigation. Rodent studies usually use an established method of measuring

animal emotionality as the dependent variable in an examination
of drug effects or the effects of brain region ablations.
Studies of emotionality in domestic animals are concerned with
the effects of emotionality in the comparative fitness of the
species for domestication. Primate studies of emotionality often
only indirectly contemplate the emotionality of the subjects
studied, focusing instead on vocalizations produced in a variety
of circumstances. Animal emotionality, a complex subject area,
has not yet received focused attention in published literature,
but may soon with the emergence of an international focus on
animal wellbeing. Supplemental bibliographies are provided for
reference.

For as long as there has been an interest in animal behavior,
there has been a correspondent interest in the subject of "animal
emotionality". Consideration of animal emotionality as a subject
of scientific study with the field of psychology began in the
1930's, with an examination of emotionality in the rat (Hall,
1934). Further studies have sought to devise techniques for the
measurement of emotionality in rodents (Frances, 1988) as well as
consider their theoretical importance (Broadhurst, 1957).
Primate emotionality has not been as extensively studied,
although there has recently been increased interest in this area
as a result of considerations of research/captive primate
wellbeing (Barnard & Hou, 1988; Bayne et al., 1992; Chamove,
1989). The emotional components of primate vocalizations have
been addressed by some studies (Masataka, 1989; Robinson, 1967),
while others have used other indeces of so-called "emotional
responsiveness" in nonhumn primates (Line et al., 1989; Martin et
al., 1988; Troisi, et al., 1991).

An annotated bibliography of important, and sometimes
controversial, research concerned with animal emotionality was
prepared (see Appendix B). The time span of research considered
is from 1934-1992, but more emphasis is placed on recent (1980's
and 1990's) studies. Subject species of these many papers
include mice and rats of various strains, pigtail (Macaca
nemestrina), rhesus (Macaca mulatta), and Japanese macaques
(Macaca fuscata), and the domestic dog and pig. The Roman high-
avoidance, Roman low-avoidance, Maudsley Reactive, and Maudsley
Non-Reactive rat strains are of special interest here because
they have been selectively bred for differences in behavioral
responses that have been linked to a basic animal "emotionality"
(Broadhurst, 1975). Emotionality is not always the central focus
of these papers, but is a significant component of either the
data collected or the theoretical orientation of the work.

Despite being the most widely-used nonhuman primate research
subject, the rhesus macaque (Macaca mulatta) vocal repertoire has
been relatively unexplored and has certainly not been described
to any great level of detail (Rowell, 1962). We believe that the
prime means for the expression of "emotionality" by the rhesus is
it's vocalizations. Early research (Robinson, 1967) has shown

emotionality to be a prime component of the vocal expressions of this species. Likewise, the emotional component of the vocalizations of other species has also been long known (Noirot & Pye, 1969).

The great promise that research into the expression of emotionality by animals via their vocalizations holds for detecting the emotional state of animals, especially primates, has not yet been tapped. It is our aim to achieve a higher level of understanding of the emotional component of animal vocalizations. Meeting this goal will significantly expand our understanding of the internal states of animals as well as enable us to comprehend future assessments of research/captive animal wellbeing.

THE PREDICTIVE VALIDITY OF AUTOMATED PERSONNEL TESTS AND ASVAB
TESTS FOR PERFORMANCE ON AIR FORCE TUTORS

David Dickter   Mary Roznowski
Graduate Associate   Assistant Professor
          Department of Psychology

The Ohio State University
1825 Neil Avenue Mall
Columbus, OH   43210

Final Report for:
Summer Research Program
Armstrong Laboratory

THE PREDICTIVE VALIDITY OF AUTOMATED PERSONNEL TESTS AND ASVAB TESTS FOR PERFORMANCE ON AIR FORCE TUTORS

David Dickter
Graduate Associate

Mary Roznowski
Assistant Professor

Department of Psychology

Abstract

In two studies, regression analyses provided evidence that the computerized Automated Personnel Tests (APTs) are useful supplements to the paper-and-pencil Armed Services Vocational Aptitude Battery (ASVAB) tests for predicting ability to learn from Air Force tutors. Further, the APTs may even outperform ASVAB tests in predicting the criteria when used alone.

THE PREDICTIVE VALIDITY OF AUTOMATED PERSONNEL TESTS AND ASVAB TESTS FOR PERFORMANCE ON AIR FORCE TUTORS

David Dickter and Mary Roznowski

INTRODUCTION

Computers possess unique capabilities for measuring human cognitive information processing skills, including the ability to administer complex tests of working memory and present graphical representations of spatial relations. These features are not available with the traditional paper-and-pencil mode of testing. Indeed, it would be difficult to overestimate the potential utility of computerized measures in conducting research on human cognitive abilities. Computerized tests may not only supplement or replace paper-and-pencil tests, but may be used in research that sheds new light on the nature and measurement of cognitive information processing and intelligence (Fleishman, 1988; Hunt, 1982; Kyllonen & Christal, 1989; Sternberg, 1986).

An ongoing research effort that has been developing novel ways of measuring human cognitive abilities is the Learning Abilities Measurement Program (LAMP), located at the Air Force's Armstrong Laboratory. LAMP utilizes computer administration of cognitive tasks in order to measure abilities that include speed of information processing and working memory capacity. The program maintains the dual goals of modeling cognitive learning and developing improved tests for selection and classification (Kyllonen & Christal, 1989).

The purpose of this study was to measure the predictive

validities of scores from the Automated Personnel Testing (ATP) program, a subset of measures from the Cognitive Abilities Measurement (CAM) test battery (Kyllonen et al., 1990), as well as to compare them to those from paper-and-pencil Armed Services Vocational Aptitude Battery (ASVAB). Studies were conducted with two samples. Performance scores from two computerized Air Force tutors, electricity concepts (in Study 1) and flight engineering knowledge and skills (in Study 2), served as criteria. These tutors measure the acquisition of knowledge and skills that are relevant to many Air Force specialties and are used to study conceptual learning and skill learning (Campbell, 1990; Christal, 1991; Shute, 1992).

STUDY 1

In the first study, the APT and the Air Force Qualification Test (AFQT; a composite based on a subset of tests from the ASVAB) were used to predict performance on the Electricity tutor. Data were obtained from a database of scores on computerized tests administered as part of a study of learning environments (Shute, 1992). Because APT tests were timed examinations, two summary scores were used: percent correct on all test items (PC) and median response time on all items (RT). The total time subjects took to complete the tutor (T-TUT) was recorded. Mastery was defined as correctly solving several consecutive problems on a given principle (Shute, 1992). Following the tutor, subjects were given a test on these principles, and accuracy on the test (POSTACC) was measured. T-TUT and POSTACC

6-4

constituted the criteria used in this study.

METHODOLOGY

Data were obtained for 344 subjects who completed the ASVAB, a battery of CAM tests, and a computerized tutor designed to teach subjects Ohm's law and other principles of electricity. The tutor was self-paced, allowing subjects as much time as needed in order to master the material. Subjects were employees at a local temporary service agency and were paid for their participation. Approximately 15% were women, 39% were White, 43% were Hispanic and 10% were Black.

Regression analyses were performed in order to examine the validities of the APT tests, as well as the unique contributions of the two types of APT measures over the AFQT. The dependent variables were the time spent taking the tutor (T-TUT), and accuracy on a test of knowledge acquired from the tutor (POSTACC). The first criterion measures how long subjects took to acquire an understanding of the principles of electricity. The second criterion measures the number of correct solutions, addressing their actual percentage correct rather than competence alone. In order to keep the number of variables in the regression equation to a minimum, AFQT score used to represent the ASVAB scores. Given that the AFQT is used for selection decisions in the Air Force, and that AFQT scores in the dataset correlated .94 with the sum of the ASVAB scores, this decision seemed to be well justified.

Two sets of regression analyses were conducted. Due to the large number of tests in the APT battery (19), the number of variables used in the regression equation was minimized as much as possible. In the first set, the APT tests were grouped into composites based on six factors measured in the battery: Processing Speed, Working Memory, Fact Learning, Skill Learning, Induction, and General Knowledge. Composites were formed by adding together scores from the quantitative, verbal and spatial tests within each factor. All APT tests were used except two General Knowledge quantitative tests, for which data were not available at the time of the study. Thus, six composites were formed from the mean percent correct scores (PC). These composites were entered simultaneously into the regression equations.

Because subjects were encouraged to answer individual test items quickly and had time limits, another set of composites used the median response time scores (RT). Composites were formed based on the six factors, in the same manner as the PC composites. The RT composites were entered simultaneously into the regression equations.

RESULTS

Tables 1 through 3 present the results of regression analyses using PC composites, RT composites, and AFQT to predict POSTACC. Tables 1 and 2 report validities for these scores for POSTACC, as well as the incremental validities obtained when an

APT score is added to another APT score or to AFQT. Table 3 reports the incremental validities for the three types of scores when each is added to the remaining two scores. In order to provide a complete report of the data, the regression tables include the R-squares adjusted for the differing numbers of variables used for PC, RT and AFQT.

The PC composite has the largest R-square for predicting POSTACC (54.6% of the variance), followed by the AFQT (52.7%) and RT (14.8%). Percent Correct adds unique variance to both RT and AFQT (40.3% and 5.5%, respectively). Moreover, PC makes a unique contribution beyond both of these combined (4.5%). Clearly, Percent Correct scores result in a very useful measure, both on their own and as an addition to AFQT. It is noteworthy, however, that the AFQT also shows incremental validity over the APT, adding unique variance over PC and RT combined (3.6%). Thus, both the computerized and paper-and-pencil tests are useful. On the other hand, response time on APT tests is not useful for predicting variance over and above that accounted for by accuracy scores (PC or AFQT). (See Table 3.)

Table 1. Validity and incremental validity of composites of PC and RT. Criterion: Accuracy on test following the tutor (POSTACC)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.739	.384	.742	.742
R Square	.546	.148	.551	.551
R Square Change			.403*	.005
F Change	56.380	8.121	41.208	.510
Adjusted R Square	.537	.130	.532	.532

* Indicates significance at p<.01. Other incremental F tests are not

Table 2. Validity of AFQT and incremental validity of PC and RT.
Criterion: Accuracy on test following the tutor (POSTACC)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.726	.763	.737
R Square	.527	.582	.542
R Square Change		.055*	.016
F Change	318.278	6.173	1.604
Adjusted R Square	.525	.572	.531

Table 3. Incremental validity for AFQT, PC and RT.
Criterion: Accuracy on test following the tutor (POSTACC)

	AFQT added to PC,RT	PC added to RT,AFQT	RT added to PC,AFQT
Multiple R	.766	.766	.766
R Square	.587	.587	.587
R Square Change	.036*	.045*	.005
F Change	23.969	4.970	.591
Adjusted R Square	.568	.568	.568

Similar analyses were carried out for the time criterion.
Tables 4 to 6 present the results of regression analyses using
subjects' total time to complete the tutor as the dependent
variable.

Table 4. Validity and incremental validity of composites of PC
and RT. Criterion: Total time spent on the tutor (T-TUT)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.702	.503	.743	.743
R Square	.494	.253	.552	.552
R Square Change			.299*	.059*
F Change	46.118	16.067	30.930	6.084
Adjusted R Square	.483	.238	.533	.533

TABLE 5. Validity of AFQT and incremental validity of PC and RT.
Criterion: Total time spent on the tutor (T-TUT)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.680	.723	.715
R Square	.462	.522	.511
R Square Change		.060*	.049*
F Change	248.129	5.970	4.746
Adjusted R Square	.460	.511	.499

Table 6. Incremental validity for AFQT, PC and RT.
Criterion: Total time spent on the tutor (T-TUT)

	AFQT added to APT	PC added to RT,AFQT	RT added to PC,AFQT
Multiple R	.752	.752	.752
R Square	.566	.566	.566
R Square Change	.014*	.055*	.044*
F Change	8.743	5.834	4.636
Adjusted R Square	.546	.546	.546

As observed for the accuracy criterion (POSTACC), the PC composite has the largest R-square. Percent Correct explains 49.4% of the variance in time spent using the tutor. AFQT explains 46.2% of the variance in T-TUT, and RT explains 25.3%. Given that RT is a latency measure, one might expect that it would account for unique variance in T-TUT, even though it did not add unique variance for the accuracy criterion. Indeed, RT added 4.4% unique variance to PC and AFQT combined. On the other hand, Percent Correct adds the most unique variance. (See Table 6.)

As in the analyses with the POSTACC criterion, AFQT also

adds unique variance to PC and RT combined in predicting the time
criterion, though the variance added is small (1%). Thus, in
both analyses, the APT tests add more unique variance to AFQT
than AFQT adds to them, and Percent Correct on the computerized
measures is the best overall predictor.

Another way of keeping the number of variables in the
regression equation at a reasonable level was to select the best
of the APT tests for analysis. The tests were selected based on
their zero-order correlations with the criteria. Although this
might appear to inflate the Multiple R for APT tests falsely, the
analyses reported below will show that this was not the case.
The results are similar to those using composites of APT scores.
Moreover, any such method of selecting the best tests would be
imperfect. If the APT tests correlating highly with AFQT were
chosen, for instance, then the ability to detect incremental
validity of APT beyond AFQT would be jeopardized. The present
method of selecting tests is preferable, since it would be better
to risk committing a Type I error than to risk a Type II error
and possibly overlook incremental validity.

Tests with Percent Correct scores that had moderate to high
positive correlations with the accuracy criterion (POSTACC) and
moderate to high negative correlations with the time criterion
(T-TUT) were identified. Therefore, tests were retained only if
high PC scores corresponded to high POSTACC scores and fast
learning on the tutor.

RT measures with moderately or highly negative correlations

with POSTACC and positive correlations with T-TUT were also identified. Therefore, for RT, tests were retained only if slow reaction time corresponded to poor accuracy and slow learning on the tutor. Tables 7 to 9 present the results of regression analyses with POSTACC as the dependent variable.

Table 7. PC and RT tests correlating most highly with criteria.
Criterion: Accuracy on test following the tutor (POSTACC)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.750	.497	.752	.752
R Square	.562	.247	.566	.566
R Square Change			.320*	.004
F Change			12.244	.445
Adjusted R Square	.536	.231	.530	.530

Table 8. Validity of AFQT and incremental validity of PC and RT.
Criterion: Accuracy on test following the tutor (POSTACC)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.725	.775	.735
R Square	.525	.600	.540
R Square Change		.075*	.015
F Change		3.178	1.535
Adjusted R Square	.52367	.575	.529

Table 9. Incremental validity for AFQT, PC and RT.
Criterion: Accuracy on test following the tutor (POSTACC)

	AFQT added to APT	PC added to RT,AFQT	RT added to PC,AFQT
Multiple R	.776	.776	.776
R Square	.602	.602	.602
R Square Change	.036*	.062*	.002
F Change	24.199	2.586	.237
Adjusted R Square	.568	.568	.568

Although the R-squares are of greater magnitude for the selected PC and RT scores, most increases are less than 2%. The exception is RT, which performs much better using the selection method than using composites (.25 versus .15). In addition, although the R-squares for PC, RT and AFQT are somewhat larger when the best tests are used, the trends among them are comparable to those reported for the composites. Thus, the two methods used for summarizing the data for regression analysis may reflect a range of possible R-square values which depends on which APT tests are used, but they also demonstrate underlying trends among PC, RT and AFQT.

Regression analyses using the selected PC and RT scores were also carried out for the T-TUT criterion. Tables 10 to 12 present the results of these analyses.

Table 10. PC and RT tests correlating most highly with criteria. Criterion: Total time spent on the tutor (T-TUT)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.730	.570	.767	.767
R Square	.533	.325	.589	.589
R Square Change			.264*	.056*
F Change	18.292	13.472	9.915	3.582
Adjusted R Square	.503	.301	.546	.546

Table 11. Validity of AFQT and incremental validity of PC and RT. Criterion: Total time spent on the tutor (T-TUT)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.680	.748	.722
R Square	.462	.559	.521
R Square Change		.098*	.059*

F Change	248.129	3.543	2.432
Adjusted R Square	.460	.530	.502

Table 12. Incremental validity for AFQT, PC and RT.
Criterion: Total time spent on the tutor (T-TUT)

	AFQT added to APT	PC added to RT,AFQT	RT added to PC,AFQT
Multiple R	.777	.777	.777
R Square	.603	.603	.603
R Square Change	.015*	.082*	.044*
F Change	9.728	3.20	2.891
Adjusted R Square	.561	.561	.561

The improvements are similar to those for POSTACC. The tests showed an increase in R-square for T-TUT, usually of 4% or less, when the selection method was used instead of composites. Thus, the same general trends were observed for the selection and composite methods for T-TUT. PC had the largest R-square, followed by AFQT and RT. Both PC and RT added variance over other scores, and AFQT added a small amount of unique variance over PC and RT combined (1.5%, $p < .01$).

STUDY 2

METHODOLOGY

In study 2, 402 subjects completed the ASVAB, a battery of CAM tests, and a computerized tutor designed to teach flight engineering knowledge and skills. As in the first study, subjects were temporary service employees. Approximately 25% were women, 35% were White, 47% were Hispanic and 14% were Black.

Regression analyses were conducted, using the amount of

information learned from the Flight Engineering tutor as the dependent variable. The criterion, POSTFAC, was a measure of the percent correct on a test given following the tutor.

RESULTS

Tables 13 to 15 summarize the results of regression analyses using composites of PC and RT scores. The same tests used for the composites in Study 1 were used for these analyses.

Table 13. Validity, incremental validity of PC and RT composites.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.870	.561	.880	.880
R Square	.757	.315	.774	.774
R Square Change			.459*	.017*
F Change	147.871	21.813	94.492	3.521
Adjusted R Square	.752	.300	.764	.764

Table 14. Validity of AFQT and incremental validity of PC and RT.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.849	.905	.868
R Square	.720	.818	.753
R Square Change		.098*	.033*
F Change	746.110	25.563	6.284
Adjusted R Square	.719	.814	.747

Table 15. Incremental validity for AFQT, PC and RT.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	AFQT added to PC,RT	PC added to AFQT,RT	RT added to PC,AFQT
Multiple R	.913	.913	.913
R Square	.834	.834	.834
R Square Change	.060*	.081*	.016*
F Change	100.613	22.653	4.408
Adjusted R Square	.826	.826	.826

In this analysis, the AFQT and APT tests combined account for more variance than they did in Study 1 (83.4% of POSTFAC in Study 2, versus 58.7% of Accuracy and 56.6% of Total Time in Study 1). This may be due to the very high Multiple R between AFQT and the Study 1 criteria (e.g., .73 between AFQT and POSTACC). On the other hand, the unique variance accounted for by PC and RT over AFQT is nearly twice the amount found in Study 1.

Although there is a difference in the magnitude of the R-squares, the incremental R-square results indicate similar trends to those observed for the composites in Study 1. In both cases, PC has the highest R-square, followed by AFQT and RT. Both PC and RT show incremental validity over each other and over AFQT, as in the analysis with the total time criterion in the first study. The variance added by RT over PC and AFQT combined is small but significant (1.6%, p<.01). PC adds unique variance over RT and AFQT combined (approximately 8%), and AFQT adds unique variance over PC and RT combined (6%).

As in the first study, analyses were performed using a subset of the tests. Tables 16 to 18 present the results of regression analyses using only the tests that were moderately or highly correlated with the criterion, POSTFAC.

Table 16. PC and RT tests correlating most highly with criteria.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	PC	RT	PC added to RT	RT added to PC
Multiple R	.876	.536	.878	.878
R Square	.767	.287	.770	.770
R Square Change	.767	.298	.483	.003
F Change	49.846	14.247	30.941*	.497
Adjusted R Square	.751	.267	.748	.748

Table 17. Validity of AFQT and incremental validity of PC and RT.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	AFQT	PC added to AFQT	RT added to AFQT
Multiple R	.849	.909	.864
R Square	.720	.826	.747
R Square Change	.720	.106*	.027*
F Change	746.110	9.196	3.735
Adjusted R Square	.719	.814	.739

Table 18. Incremental validity for AFQT, PC and RT.
CRITERION: POSTFAC (Accuracy on test of flight engineering)

	AFQT added to PC,RT	PC added to AFQT,RT	RT added to PC,AFQT
Multiple R	.913	.913	.913
R Square	.834	.834	.834
R Square Change	.064*	.087*	.008
F Change	100.867	7.653	1.528
Adjusted R Square	.817	.817	.817

Selecting tests for the regression analysis produces the same results as using composites. Like the analyses in Study 1, selecting the best tests did not substantially raise R-squares in this study. In fact, the one notable difference is that RT drops from an R-square of .315 when composites are used, to one of .287 when the "best" RT scores are used, and the unique variance accounted for by RT over PC is no longer significant. Thus,

selecting tests correlated highly with the criterion is not always effective in increasing R-square. In general, however, it makes little difference whether composites or a select group of tests are used in the regression equation.

DISCUSSION AND CONCLUSION

The results of both studies indicate that scores based on percent correct and reaction time are useful for predicting learning and performance on the complex computerized tutors. Furthermore, these variables add unique variance beyond that contributed by a standardized measure of general aptitude (AFQT). In both studies, the regression analyses indicate that Percent Correct accounts for slightly more variance in criterion scores than that accounted for by AFQT. The analyses also show that the new measures have incremental validity beyond the AFQT in predicting performance on the tutor. Response latency scores add unique variance, but scores reflecting error rates are the most useful predictors, whether they are used alone or in addition to other variables.

One other study has compared the validities of a battery of experimental cognitive tasks and the ASVAB tests in predicting performance on an Air Force tutor. Christal (1991) conducted a study using military recruits, and predicted performance on a computerized tutor designed to teach knowledge and skills on logic gates. His study used a different criterion and a different set of computerized tests. He found that the

computerized measures added as much as 22% unique variance over the ASVAB scores. The preر ent study found not more than 10% unique variance of APT over AFQT, but there was less variance remaining after AFQT had been entered into the regression than there was in Christal's study after the ASVAB had been entered. The correlations between the standardized aptitude measures and the measures from the computer battery were much lower in Christal's study (e.g., .48, vs. .85 in the present study; See Table 8). Thus, it is especially noteworthy that incremental validities were obtained for the percentage correct scores, in spite of the high validities for the AFQT tests.

In sum, two essential points merit repeating:

1) The analyses indicated that percentage correct scores on the cognitive test battery had higher predictive validities than the standardized aptitude measures for the tutor criteria.

2) Incremental validities were obtained for the computerized cognitive information processing scores, over and above the validities of the standardized measures. Note, however, that AFQT scores also added incremental validity over the computerized measures, particularly for the POSTFAC criterion.

The results of this research are encouraging, but a few caveats concerning limitations are warranted. One limitation is the suبject population, which was different in make-up from a typical population of Air Force applicants. The subject population contained a higher proportion of minorities than would usually be found in an Air Force sample. In addition, the

question of motivation arises, as it would in other studies using non-applicant samples. Subjects were encouraged to perform to the best of their ability, but the degree to which their level of motivation would have differed from that of an applicant sample was uncertain. Further empirical research investigating the nature and measurement of subjects' "motivational states" is necessary (Arvey, Strickland, Drauden, & Martin, 1990). A third issue is restriction of range. One might argue that the variance of the data sample should have been corrected to that of the general population. However, since Air Force applicants are a select group with above-average ASVAB scores, and the sample used in the study represented a reasonable match to the population for which the tests were constructed, this procedure was not carried out. As stated in the <u>Principles for the Validation and Use of Personnel Selection Procedures</u> (1987), validation should use a sample that is "reasonably representative of the populations of people and jobs to which the results are to be generalized" (p.7).

The future holds exciting prospects for the development of new computerized tests and their use for the understanding of cognitive information processing and intelligence. Further studies will no doubt bring novel methods for measuring and understanding human cognition.

References

Arvey, R.D., Strickland, W., Drauden, G., & Martin, C. (1990).
Motivational components of test-taking. Personnel Psychology,
43, 695-716.

Campbell, J.P. (1990). An overview of the Army selection and
classification project (Project A). Personnel Psychology, 43,
231-239.

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, Air Force
Human Resources Laboratory.

Fleishman, E.A. (1988). Some new frontiers in personnel selection
research. Personnel Psychology, 41, 679-701.

Hunt, E. (1982). Toward new ways of assessing intelligence.
Intelligence, 6, 231-240.

Kyllonen, P.C., Christal, R.E. (1989). In R.F. Dillon & J.W.
Pellegrino (Eds.), Testing: theoretical and applied
perspectives. (pp.146-173). New York: Praeger.

Kyllonen, P.C., Woltz, D.J., Christal, R.E., Tirre, W.C., Shute,
V.J., & Chaiken, S. (1990). CAM-4: Computerized battery of
cognitive ability tests. Unpublished computer program, Brooks
Air Force Base, Texas.

Society for Industrial and Organizational Psychology Inc. (1987)
Principles for the validation and use of personnel selection
procedures: Third edition. College Park, MD.

Shute, V.J. (1992). A comparison of learning environments: All
that glitters... (AL-TP-1991-0042). Brooks Air Force Base, TX:
Manpower and Personnel Division, Air Force Human Resources
Laboratory.

Sternberg, R.J. (1986). The future of intelligence testing.
Educational Measurement: Issues and Practice, 5, 19-22.

INORGANIC FIBER ANALYSIS
BY SEM-EDXA

Robert F. Diskin
Department of Chemistry
University of Scranton
Scranton, Pennsylvania 18520-4626

Final Report for:
Summer Research Program
Armstrong Laboratory

September 1992

INORGANIC FIBER ANALYSIS
BY SEM-EDXA

Robert F. Diskin
University of Scranton
Scranton, Pennsylvania 18520-4626

Abstract

The air asbestos analysis laboratory at Brooks Air Force Base often needs to identify fibrous material that may or may not be asbestos. In order to help improve the laboratory's ability to identify such fibers, the author prepared and analyzed a number of inorganic fiber standards and several air filters submitted to Brooks with an Amray 1820 Scanning Electron Microscope and Tracor Northern X-ray Analyzer. All seven types of asbestos and twenty non-asbestos man-made and natural fibers were analyzed on the SEM. The objective of the project was to supply the author's professor with the necessary data to design a system for fiber identification using morphology and the elemental analyses obtained.

Introduction

Asbestos is a well-documented health hazard. Exposure to airborne asbestos fibers can lead to diseases such as asbestosis, lung cancer, and mesothelioma (1). It is important therefore that samples suspected to contain asbestos be analyzed for these fibers.

The air asbestos analysis sect..on of the Analytical Services Division of the Occupational and Environmental Health Laboratory analyzes several thousand filters annually for their fiber content according to NIOSH method 7400. This method calls for using phase contrast microscopy to count all fibers with a length to aspect ratio of at least 3:1.

The NIOSH method 7400 receives much criticism since it calls for counting fibers on air filters without distinguishing them as asbestos or non-asbestos fibers. Furthermore, the NIOSH method is unable to distinguish organic from inorganic fibers. Situations arise where this laboratory needs to identify the fibers on the filters as asbestos or some other fibrous material as a service to its constituency. Since the NIOSH method has the ability to exaggerate the potential health hazard, the project objective was to assist the environmental officer in making decisions for handling high fiber counts. One potential way of providing this service is to supplement the fiber count analysis with scanning electron microscopy using an SEM which is equipped with an x-ray analyzer. The combination of an electron micrograph with an elemental composition as determined by energy dispersive x-ray analysis can frequently identify fibers as asbestos or non-asbestos, and often yield a reasonable answer as to the true identity of the fiber. Since it is costly to analyze every fiber, the author's professor designed a method for rapidly analyzing a representative fraction of the fibers in air samples to help evaluate the environment where the sampling occurred (5).

7-3

The project outlined in this paper details the SEM-EDXA technical work that was required for the professor's goal.

Instrumentation

The scanning electron microscope used in this study was an Amray Model 1820 equipped with a data entry keyboard and an air-cooled turbomolecular vacuum pump. It was an automatic imaging SEM, and included digital image storage and processing. It was capable of dot mapping of elemental locations. Based on previous work, all fibers were processed at a tilt angle of 38.5 degrees in order to produce the most intense x-ray pattern sensed by the x-ray detector (4). A detector slide position of 45 mm and a working distance of 12 mm using 20 kev at 90 ma proved most advantageous in obtaining a useful EDXA spectrum in less than two minutes of analysis time (2).

The x-ray analysis unit was a Tracor Northern Series II X-ray Analyzer (TN-II) equipped with a standard beryllium window detector. The TN-II SQ x-ray data acquisition and analyzer software program was used to acquire the X-ray spectrum. The SQ software is capable of performing quantitative analysis on X-ray Spectra using a library of references stored on disk. The program uses Multiple Least Squares Analysis and ZAF or Phi-Rho-Z matrix correction procedures to calculate elemental concentration results. ZAF corrects for atomic number (Z), absorption (A) and fluorescence (F) effects. It is a theoretical correction procedure based on excitation conditions, geometry of the sample and intensity ratios constructed from pure element samples. The package also makes non-linear background corrections on a channel by channel basis and is dependent upon the sample to detector geometry (6).

An Anatech LTD Hummer VI Sputtering System was used to metal coat the

specimens before introduction into the scanning electron microscope. A gold-palladium alloy was used as the coating metal. The coating was about 85% gold by weight.

Methodology

The standard inorganic fiber samples were cut into pieces with a pair of scissors. The pieces were placed on a carbon painted aluminum stud. The studs were painted twice and before the second coat was dry, the fibers were applied which allowed sufficient adherence of the fibers to the studs. The studs were placed in the coating chamber of the Anatech LTD Hummer VI and the system was evacuated to at least 25 millitorr. Argon was introduced into the system to a pressure of 80 millitorr. The high voltage was turned on such that a 10 milliamp current flow was established. A discharge was created which sputtered the Au/Pd alloy onto the specimens. Coating was allowed to continue until a layer 10 nanometers thick was deposited.

The samples were placed in the vacuum chamber of the Amray 1820 scanning electron microscope and searched for suitable fibers. The selected fibers were photographed at a magnification which illustrated their morphology. After the visual inspection , the magnification was increased until the SEM partial field (3cm x 2cm display on the video monitor of the SEM) fitted entirely within the fiber image and the EDXA was obtained for the fiber. Under these experimental conditions the background was minimized but continuous monitoring and adjustments were usually required during EDXA data acquisition to compensate for drift of the particle field, especially at high magnification.

A portion of the selected fiber was analyzed using the Tracor Northern X-ray Analyzer. At least six energy dispersive x-ray analyses were done for each

standard. For each standard, usually three fibers received one analysis while one fiber received three analysis at different locations along the fiber. Analysis was done for a range 0-10 kev for 90 seconds. The quantitative analysis program SQ of the Tracor-Northern system was activated to yield a quantitative analysis of those elements selected. Elemental selection was dependent on the known composition of each standard. The computer gave the data as either atomic percentage or weight percentage. All analyses printouts, the photographs and a printout of the spectra obtained were filed in a binder and used by the author's professor for his research.

For the air samples, a 4mm triangle was cut from each unknown cellulose acetate air filter 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, then placed in the SEM vacu hamber. Suitable fibers, those with correct morphology (length to aspect ratio of at least 3:1), 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). A minimum of five fibers were analyzed for each air filter and selection of fibers was as random as possible meaning no fiber analyzed was in close proximity to any other fiber analyzed.

Results and Discussion

The project was composed of two phases; the collection and analysis of known inorganic fiber standards followed by the analysis of unknown fibers on several air samples submitted to the laboratory. All seven types of asbestos and twenty non-asbestos man-made and natural fibers were analyzed on the SEM (see Table 1). Because this laboratory primarily analyzes samples for asbestos, non-

asbestos fibers that frequently are mistaken for asbestos because of their morphology were selected.

As noted earlier all printouts and photographs were assembled into a small library and submitted to the author's professor. Analysis of air filters are best obtained at around 2000x. At this magnification, fibrous talc and other crystalline materials which commonly yield EDXA data comparable to that obtained for asbestos could be distinguished from asbestos and asbestos-like inorganic fibers (2).

Not infrequently some peak identifications given by the x-ray analysis system were faulty. The most common misidentifications were Nb instead of Au, Tl instead of Cl, Rb and Au instead of Si, and Cl instead of Pd when Cl was also present (4).

A scanning electron microscope equipped with an energy dispersive x-ray analyzer presents a practical method of identifying inorganic fibrous material commonly found in atmospheric environments. When using this equipment, a number of procedures need be followed in order to assure the greatest accuracy. Appropriate magnification of a fiber is that which reveals adequate morphology while being sufficient to completely cover the partial field on the SEM video monitor. Several EDXA spectra should be obtained in order to establish the elemental composition of the standards and fiber makeup of air samples with reasonable accuracy.

Acknowledgements

TABLE 1

NON-ASBESTOS FIBERS		ASBESTOS FIBERS
Alumina Saffil	Fiberglass O-C	Amosite
Alumina	Graphite P-55	Anthophollite
Bulk 6000	Graphite T300	Crocidolite
Ca Na Phosphate	Kaowool	Chrysotile
Calcium Silicate	Metallized Plastic	Tremolite
Calcium Sulfate	Mineral Wool A	Actinolite
Cerawool	Rockwool A	Ferroactinolite
Cooperknit	Sandtex	
Fiberfrax	Talc 11	
Fiberfrax 7000	Tritan Kaowool	

References

1. Pravin P. Parekh, Richard J. Janulis, James S. Webber, and Thomas M. Semkow "Quantitation of Asbestos in Synthetic Mixtures Using Instrumental Neutron Activation Analysis" Analytical Chemistry 64, 320-325 (1992).

2. 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).

3. 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).

4. Edmund C. Shearer "Final Report Summer Fellowship" U.S. Air Force Office of Scientific Research (1989).

5. L.R. Sherman "Final Report Summer Fellowship" U.S. Air Force Office of Scientific Research (1988).

6. R. Accurate "Quantitative Trace Analysis with X-ray Fluorescence in the Scanning Electron Microscopy" Acta (Wien) 1, 193-200 (1987).

COMPONENTS OF SPATIAL AWARENESS:
EFFECTS OF AIR FORCE FIGHTER PILOT TRAINING AND EXPERIENCE

Itiel E. Dror
Department of Psychology
Harvard University

Final Report for
Summer Research Program
Armstrong Laboratory

August 1992

Send correspondence to: Itiel E. Dror
Harvard University
33 Kirkland Street
Cambridge, MA 02138

COMPONENTS OF SPATIAL AWARENESS:
EFFECTS OF AIR FORCE FIGHTER PILOT TRAINING AND EXPERIENCE

Itiel E. Dror
Department of Psychology
Harvard University

Abstract

Air Force pilots with varying experience and training participated in two experiments which assess different components involved in spatial awareness. One experiment examined the ability to track the motion of multiple moving objects; and the other experiment examined the ability to extrapolate motion. We found that motion tracking skills were affected by a "recency of flying" affect --pilots who recently flew were more accurate than pilots who did not. No such affect was found for motion extrapolating skills. We also found that although experienced fighter pilots and novice non-fighter pilots had comparable performance on both tasks on the easiest conditions, the experience fighter pilots were better at both tasks on the most difficult conditions. However, novice pilots were better at the intermediate levels.

COMPONENTS OF SPATIAL AWARENESS:
AFFECTS OF AIR FORCE FIGHTER PILOT TRAINING AND EXPERIENCE

Itiel E. Dror

Pilots rely on spatial awareness for the successful accomplishment of missions as well as for safety. The importance of spatial awareness is further emphasized in the tactical aviation environment in which highly agile fighter aircraft are designed to operate across the spectrum of flight altitudes and orientations. Spatial awareness encompasses being aware of one's own spatial position along with the spatial relations to various elements in the surrounding environment. The high speed and maneuverability of fighter aircraft results in constant and rapid changes in the spatial positioning of the aircraft. The spatial relations to stationary ground targets as well as to other rapidly moving aircraft constantly changes. Subsequently, pilots must always assess and update their situation awareness. This is especially demanding when pilots attend to cockpit displays and are thus disconnected from direct visual contact to their surroundings. Failure to maintain spatial awareness can have disastrous consequences as evidenced by the large number of accidents which have been attributed to loss of situation awareness and spatial disorientation. Furthermore, maintaining spatial awareness during air combat is considered to be an essential element of success.

The goal of this study is to understand and characterize some of the subsystems involved in spatial awareness and the factors that influence them. The study also examines whether spatial awareness abilities are affected by Air Force fighter training and experience, or by a "recency flying affect" -- whether or not the pilots have recently flown, or by some combination of these factors .

Two tasks were designed to tap into the underpinnings of motion extrapolation and motion tracking. In each task we manipulated the number of moving objects and the duration of tracking and extrapolation processes. If these manipulations effect the underling processes then they would force more processing and thus effect accuracy rates and response times. The increased response time and error rate as a function of the manipulations can be calculated as a slope. By comparing such slopes one can examine the

specific components *per se-* independent of the processes involved in encoding the stimulus and generating the response itself (which are reflected by the intercept of the function). This method allowed us to determine the relative contributions of specific types of processing (Sternberg, 1969).

This study is a continuation of a previous study where pilots were found to be better than control non-pilots at certain components of visual cognition but not at others (for a detailed account of the study see Dror, 1991; Dror, 1992; and Dror, Kosslyn, and Waag, 1992).

GENERAL METHOD

The subjects were tested individually on each task in one testing session, which lasted on the average forty-five minutes. Half the subjects were first tested on the motion tracking task and then on the motion extrapolation task; the other half of the subjects were tested in the reverse order. The subjects were given verbal instructions and then were asked to paraphrase them, and any misconceptions were corrected. The subjects began each task with a set of practice trials. During the practice trials, the computer gave feedback by beeping when the subject made an incorrect response, and the subjects were encouraged to ask questions. During the test trials, no feedback was provided and no talking was allowed. The tasks required the subjects to respond by pressing keys marked "yes" (the "b" key) and "no" (the "n" key) on the computer's keyboard. The subjects used two fingers of their dominant hand to press the keys. The tasks were administered on a Macintosh II ci computer with a high resolution video display card (8.24 video card). The computer was connected to a color 13 inch multiscan trinitron super fine pitch Sony monitor. The tasks were administered by a computer program that used the Shell and Macglib libraries of Micro M.L. Inc. All subjects sat so that their heads were approximately 50 centimeters from the computer screen. The subjects were asked to respond as quickly as possible while remaining as accurate as possible.

Subjects

Thirty-four pilots were tested on both experiments. The pilots formed two distinct groups: the first group of 20 pilots were very experienced and highly trained fighter pilots; the second group of 14 pilots were novice pilots who had never flown fighter aircraft. The mean flying hours on fighter

aircraft (F-15 and F-16) of the first group was 1650 (range 1000-3100). Their overall mean flying hours --including non-fighter aircraft-- was 2300 (range 1400-3400). Their mean age was 34.8 (range 29-41). The second group of pilots included pilots who had never flown a fighter aircraft, but were assigned to be fighter pilots. They all had recently finished the undergraduate training of the Air Force and were waiting for their basic fighter aircraft training to began. The pilots in this group had approximately 300 flying hours on non-fighter aircraft. Their mean age was 26.1 (range 24-29). All the pilots had completed at least college education. The pilots were recruited and tested at Luke Air Force Base, AZ.

EXPERIMENT 1: VISUAL EXTRAPOLATION

Visual imagery enables us to manipulate in our mind the visual stimuli we receive from the outside world. We can manipulate and transform images in numerous ways. One such manipulation encompasses shifting and changing the spatial position of objects within a complex image (e.g., one can imagine rearranging items on the desk --shifting the phone and the computer to different locations). Indeed, such manipulations are often used in reasoning (Hayes, 1981). Motion can be accomplished in mental imagery by constantly shifting the spatial positions of items within the image. Motion extrapolation requires such changes, whereby the imagined motion is based on previously seen motion. Correct timing and spatial perception are also key factors that enable the imagined motion to be in the same speed and trajectory as the previously seen motion. Thus, motion extrapolation requires the use of imagery as well as perception of time and space. Pilots ability to extrapolate motion is particularly interesting as previous studies show that pilots have unique abilities in some processes of visual imagery --visual mental rotation (Dror, 1992)-- but not in others --scanning visual images (Dror, Kosslyn, & Waag, 1992).

We used a visual extrapolation task that was a variant of one devised by Dror (1991; see also Dror, Kosslyn, and Waag, 1992). In the previous study, subjects were required to extrapolate the motion of one object that was moving in a circular trajectory. The study revealed that the components involved in motion extrapolation work harder when the trajectory had to be projected a greater distance. In the present study subjects were required to extrapolate the motion of multiple objects moving in various straight

trajectories. It was assumed that it is also more demanding when the motion of more objects had to be extrapolated. A group of moving balls were presented, and the subjects attended to a subgroup of them that were flashing. All balls were then removed from the display, and the screen remained blank. After a delay the balls reappeared, as if they have been in constant motion while the screen was blank. One ball was presented as a probe ball and the subjects were required to judge whether or not this ball was one of the balls that was flashing earlier.

Method

Materials. The stimuli were round disks (the "balls") 18 pixels in diameter – corresponding to 0.668 cm and 0.765 degrees of visual angle-- which moved in straight trajectories on the screen. The balls' initial trajectories were random. The balls bounced off the walls of the screen but moved through each other. Twelve black balls were presented in motion on the screen while a subset of them (1, 2, or 3) flashed --changed colors from black to red and back every 60 ms. The speed of the motion was 4.33 cm per second --corresponding to 4.96 degrees of visual angle. The motion was created by displaying a new screen every 60 ms with each ball advancing 7 pixels in its trajectory --corresponding to 0.260 cm and 0.298 degrees of visual angle. The computer generated the new position for each of the balls throughout the trial. While the screen was blank the computer continued to generate ball motion, however the balls were not displayed on the computer screen

A total of 96 trials were prepared. The trials were constructed in 8 blocks of 12 trials each. Half the trials in each block should have been evaluated as "yes" trials --presenting a ball probe that had flashed earlier-- and the other half were "no" trials --presenting a ball that had not flashed earlier. For each of the 6 "yes" and "no" trials in every block, half had a short time delay and half a long time delay. We increased the short and long time delays used in our previous study (Dror, 1991; see also Dror et al., 1992) as well as increasing the time difference between them. The short time delay was now 2 seconds and the long one 3.5 seconds. For each 3 trials in every block that had the same time delay and the same correct response, one required to extrapolate the motion of 1 ball, one of 2 balls, and one of 3 balls. Thus, each block had 12 trials that included all possible combinations of variables. The trials within

each block were randomly ordered with the constraint that the same number of balls, time delay, or response could not appear more than three times in succession. An additional 12 trials were prepared as practice trials. The block of practice trials had the same structure as the testing blocks and included all variable combinations. The performance of the subjects was examined based on the manipulation of the time delay and the number of balls that were extrapolated.

Procedure. A trial began with an exclamation mark. When the subject was ready, he pressed the space bar, and then 12 moving balls appeared on the computer screen. The subject was told to track the motion of the flashing balls, and to keep tracking them after the balls disappeared as if they were still moving on the computer screen. When the balls re-appeared, one of the 12 original balls was flashing. The subject was to respond "yes" if the flashing ball was one of the balls that was flashing earlier, and "no" if it was not. Immediately after the response, another exclamation mark appeared and a new trial began.

Results

The data were analyzed using an analyses of variance. Separate analyses were performed for response times and error rates. The data included 0.7% responses that were outliers --greater than 2.5 times the mean of the remaining scores in that cell-- and were replaced by the mean of the cell. Incorrect responses were excluded from mean response times computations. Each analysis included two between subject variables: fighter/non-fighter pilots and flow recently/did not fly recently (pilots who had more than 10 flying hours in the past 2 weeks and over 15 in the past 4 weeks were considered as pilots who flow recently, and pilots who did not fly at all in the past 2 weeks and less than 4 hours in the past 4 weeks were considered as pilots who had not flown recently); and two within subject variables: the number of trajectory that were extrapolated (the number of balls) and the duration of extrapolation.

The pilots were able to extrapolate motion with approximately 75% accuracy. They were even able to extrapolate the motion of 3 balls at the same time with an accuracy level of approximately 64%. We were interested to see which factors effected their performance, and whether or not these factors interacted. We were especially interested to see if any of the within variables

(that tap into the extrapolation process *per se*) interacted with any of the between variable (that represent different training and experience of pilots). Such an interaction would show whether training and experience affects components involved in motion extrapolation.

We found that the number of moving balls that were extrapolated effected both response time and error rate, $F(2,60)= 33.19$, $p< .01$ for response time (with means of 1223, 1450, and 1583 ms for 1, 2, and 3 balls, respectively), and $F(2,60)= 53.00$, $p< .01$, for error rates (with means of 12.22, 27.38, and 36.48 percent error for 1, 2, and 3 balls, respectively). Indeed, linear contrasts revealed that response times increased linearly with the number of trajectories that were extrapolated, $F(1,60)= 47.67$, $p< .01$, as did error rates, $F(1,60)= 73.87$, $p< .01$. The duration of the extrapolation did not effect response time, $F(1,30)= 1.01$, $p> 2.5$, but did effect error rate, $F(1,30)= 32.48$, $p< .01$ (with means of 21.26 and 29.47 percent error for short and long time delays, respectively). It is important to realize that the two dependent measures, response time and error rate, are inter-related (see Luce, 1986). Thus, a manipulation can affect performance in either response time, error rate, or both.

After establishing that the manipulation of the number of trajectories that were extrapolated and the duration of the extrapolation did effect performance, we proceeded to see if either of the between variables --being a fighter pilot and recency of flying-- made a difference in how the manipulations affected performance. As illustrated in Figure 1, we found an interaction in error rate between fighter pilot grouping (fighter pilots vs. non-fighter pilots) and the number of trajectories that were extrapolated, $F(2,60)= 3.66$, $p= .03$. This interaction reflected that even though both groups of pilots had similar error rates at the easiest condition when only one trajectory was extrapolated, the fighter pilots were more accurate at the hardest condition when three trajectories had to be extrapolated at once. However, the non-fighter pilots were more accurate in the intermediate condition when only two trajectories were extrapolated. No other interactions were observed, all $p> .2$.

Fighter pilots and non-fighter pilots had comparable overall performance, $F <1$ for both response time and error rate. Finally, we found that pilots who had flown recently and pilots who had not flown recently had comparable performance as well, $F< 1$ for both response times and

comparable error rates.

Insert Figure 1 About Here

Discussion

Replicating the finding of Dror, Kosslyn, and Waag (1991) we found that the longer time delays effected performance. We also found that the increasing number of trajectories that were extrapolated caused linear increases in response time and error rate. Air Force fighter pilot training and experience effected the pilots ability to extrapolate motion, however this effect was limited to the number of trajectories that were extrapolated and did not effect the ability to extrapolate over different time durations.

Although fighter pilots had superior performance in the most difficult conditions, their performance was not better across all conditions.

EXPERIMENT 2: MOTION TRACKING

Tracking a single ball that is moving at high speed is not an easy task. Professional baseball batters need to track a ball moving at speeds of up to 100 mph --producing angular velocities greater than 500 degrees per second. Indeed, professional baseball batters have superior skills at tracking such a high speed moving ball (Bahill and LaRitz, 1984). Tracking multiple objects is even more demanding as the attention system needs to cope with numerous objects moving in different direction at the same time. Several theories have been suggested on how the attention system accomplishes such a task (for a good review see Yantis, 1992). We set out to test pilots on a demanding task which required them to track up to 6 moving balls at the same time.

The motion tracking task was a variant of one used by Intriligator, Nakayama, & Cavanagh (1991), which was based on a task devised by Pylyshyn & Strom (1988). A group of moving balls were presented on the computer screen. The subject was asked to track a flashing subset of the balls, and to continue to do so after they stopped flashing (and thus became indistinguishable from the other balls on the screen). The balls continued to move on the computer screen after the flashing was stopped; after a time delay one of the balls was designated as a probe ball, and the subject had to judge whether or not it was one of the balls that had flashed earlier.

8-9

Method

Materials. The same materials used in the extrapolation task were used here, except that 4, 5, or 6 balls were flashed and the balls did not disappear after the flashing was concluded. The balls remained on the screen and the pilots had to track the balls that flashed earlier. To force actual tracking and not enable the pilots to rely on extrapolation, the trajectory of each ball was changed so the motion was random —every 60 ms the balls changed trajectory by a random degree shift of up to 30 degrees. To avoid confusing between colliding balls, the balls now bounced off each other when they collided.

Procedure. The same procedure used in the extrapolation task was used here.

Results

The data was analyzed as in experiment 1, 0.6% of the data were considered outliers. The pilots were able to track motion of multiple objects with approximately 78% accuracy. Again, we were interested in examining which factors affected their performance. The number of balls affected both response time, $F(2,60)= 5.40$, $p< .01$ (with mean response time of 1164, 1213, and 1220 ms for 4, 5, and 6 balls, respectively), and error rates, $F(2,60)= 10.06$, $p< .01$ (with mean error rates of 18.67, 20.00, and 27.94 % for 4, 5, and 6 balls, respectively). Indeed, linear contrasts showed that response time and error rate increased linearly when the motion of more balls had to be tracked, $F(1,60)= 4.97$, $p= .03$ for response time, and $F(1,60)= 13.79$, $p< .01$ for error rates. The duration of the tracking did not affect response time, $F(1,30)= 1.03$, $p> .2$, but did effect accuracy, $F(1,30)= 8.29$, $p< .01$ (with mean error rate of 21.26 and 26.47 for short and long tracking durations).

Most interestingly, we found that recency in flying had an overall effect on accuracy, $F(1,30)= 3.94$, $p= .05$ (with mean error rate of 26.97 and 22.57 % for pilots who had not flown recently and pilots who had, respectively). No such differences were found in response time, $F< 1$. No overall differences were found between fighter and non-fighter pilots, $F< 1$ for both response time and error rate.

We observed an interaction in error rate between the number of balls

tracked and the pilot grouping (fighter vs. non-fighter pilots), $F_{(2,60)}= 4.01$, $p= .02$. As illustrated in Figure 2, this interaction reflected that even though both pilots groups had similar error rates at the easiest condition when only 4 balls were tracked, the fighter pilots were more accurate at the most difficult condition when 6 balls were tracked at once. However, the non-fighter pilots were more accurate in the intermediate condition when five balls were tracked.

Insert Figure 2 About Here

We also found a trend for a four-way interaction in error rates; $F_{(2,60)}= 2.90$, $p= .06$, for the interaction between the number of balls tracked, the duration of the tracking, the pilot grouping (fighter vs. non-fighter pilots), and recency of flying. As reflected by linear-by-linear contrasts, and illustrated in Figure 3 (top left), looking only at the pilots who had flown recently, the fighter pilots were much better at the most difficult condition relative to the intermediate condition, $F_{(1,60)}= 7.95$, $p< .01$; this is consistent with the interaction we found between pilot grouping and number of balls. However, we found that looking only at the pilots who had not flown recently, as illustrated in Figure 3 (top right), the error rate was especially high for fighter pilots at the intermediate condition and for non-fighter pilots at the easiest condition. Further evidence for this trend was revealed by a linear-by-linear contrast, and is illustrated in Figure 3 (bottom left), showing that for the easiest condition the non-fighter pilots who did not fly recently were much more effected by the duration of the tracking process than those who had flown recently, $F_{(1,60)}= 5.47$, $p= .02$; in contrast no such affect was observed in the fighter pilots group, $F_{(1,60)}= 1.66$, $p> .2$. However, in the intermediate condition, illustrated in Figure 3 (bottom center), the fighter pilots showed such an effect; fighter pilots who had not flown recently were much more effected by the duration of the tracking process than the ones who had flown, $F_{(1,60)}= 13.89$, $p< .01$ for the linear-by-linear contrasts. No such effect was observed for the non-fighter pilots, $F< 1$, for the corresponding linear-by-linear contrast. For the most difficult condition, as illustrated in Figure 3 (bottom right) no such differences were observed, $F< 1$, for both cases.

Insert Figure 3 About Here

Discussion

The most interesting results pertains to the "recency flying effect"
observed in the motion tracking task; this is especially intriguing, given that
we did not find such an effect in the motion extrapolation task. Otherwise,
the general pattern of results was very similar; we found an interaction in
error rate between the number of balls and the pilot grouping (fighter vs.
non-fighter). In both experiments we found that both groups had comparable
error rate at the easiest condition, but the fighter pilots were better at the most
difficulty condition. However, the non-fighter pilots were better at the
intermediate condition.

A hint for a possible explanation to this observation might come from
the trend for the four-way interaction that was found in this task. The pilots
that had not flown recently --fighter and non-fighter-- showed declining
ability in tracking multi objects. The enhanced ability to track objects by pilots
probably does not decrease all at once. Flying probably primes some of the
underlying subsystems, and as time passes some of the processing abilities
decrease. Thus, it is possible that different subsystems decline in different
ways and at different rates; indeed, pilots who did not fly recently showed an
atypical pattern of errors, as illustrated in Figure 3 (top right). Such an
explanation accounts for the interaction data presented in Figure 2, as well as
Figure 1 --because it shows that the pilots who had not flow recently
accounted for much of the error pattern (high error in the intermediate
condition by fighter pilots, and high error in the easy condition by non-fighter
pilots). However, as reflected in Figure 3 (top right), this does not completely
account for the data. There might have also been an affect of age; the fighter
pilots were older than the non-fighter pilots (8.7 years older on the average).
Even though this is not a large difference, the fighter pilots mean age was in
the 30s and the non-fighter pilots was in the 20s. Visual perceptual skills are
at their pick in the 20s and decline in the 30s. Thus, it is possible that the
performance of the fighter pilots was over-clouded in the intermediate
condition by decreased perceptual abilities. In the most difficult condition --

when 6 balls were used-- their superior tracking ability outweighed their decreased perceptual skills and they performed better. Still another possible explanation for this pattern of results is that the groups used different strategies to perform the tasks. If both groups of pilots used the same strategies under the different conditions, then either both groups would perform the same or one group would perform better. Our results show a disassociation of performance across the different conditions. Thus, if different pilots use different strategies to track objects, then such results are possible. Our present research does not allow us to determine what caused this pattern of result. Interestingly, this pattern occurred in both tasks.

GENERAL DISCUSSION

We set out to investigate some system components involved in spatial awareness. We found that Airforce training and experience effects both motion tracking and motion extrapolation abilities. Fighter pilots did perform better in both tasks in the most difficult conditions. However, this superior performance was not observed across all conditions. Although both tasks produced the same pattern of results, the ability to track motion was enhanced by a "recency flying affect", yet the motion extrapolation was not.

Our research explores the abilities to track and extrapolate motion, which seem to be important skills involved in spatial awareness. These skills are especially critical in the domain of piloting in which motion is a key factor. Spatial awareness is a complex ability and probably involves many additional components. One such component concerns the transformation of the viewer centered retinotopic visual input into a spatiaotopic object centered representation (for more detail see Kosslyn et al, 1990; Kosslyn & Koenig, 1992). Another component that is probably involved in spatial the awareness of pilots is the ability to integrate different sources of knowledge. This includes cross modality inputs --pilots receive auditory information regarding spatial positioning of elements in their surrounding via communication as well as directly from their environment. Pilots also need to integrate the spatial information they obtain via the aircraft instruments as well as maps and aerial photographs.

In order to correctly measure spatial awareness ability, one needs to first understand which component subsystems are involved in this process, and how they interact. Then, after having such a model, one needs to explore

which factors affect each component and use these factors as a tool for quantifying the efficiency of the component subsystem. In this study we take a first step in this direction --we explore two such subsystems and speculate on two additional subsystems that may be involved in spatial awareness.

FUTURE ANALYSIS

Additional analysis of the data from this study may provide more insight about motion tracking and extrapolation, and how Air Force training and experience might affect these abilities. Four additional analysis are recommended:

1. To further examine the claim that the two tasks do indeed tap onto different components we need to performed a correlation analysis between the two tasks. It would further establish that motion tracking and motion extrapolation are indeed distinct processes.

2. In the present initial report the mean response time and error rate for each subject was calculated for each cell of trails that had the same conditions (number of balls and duration). We can substitute mean scores of cells by slopes of increase time and error as a function of difficulty. The response of the subjects in each trial can be evaluated by considering how probable it was to make a mistake --and thus how difficult it was to give the correct response. The closer the probe ball was to other balls the more difficult the trial was. Specifically, trials would be more difficult for the "off" trials when there was a small disparity between the location of the probe and an actual position of a ball compared to those that had a great disparity; and similarly for the "on" trials, it would be difficult if a distractor ball --one that did not flash earlier-- was near the probe ball, than when it was far away from it. Thus, instead of using mean response time and error rate per cell, we can calculate a regression slope that correlates the errors and response times to the "likelihood of making a mistake" .

3. Additional analysis can further explore the abilities of the fighter pilots themselves. The group of 20 fighter pilots who were tested on the tasks were composed of 14 pilots who flew F-16 and F-15E and were thus primarily trained on air-to-ground missions, and 6 pilots who flew F-15 and were thus primarily trained on air-to-air combat. It would be interesting to explore if these two types of pilots differed in skills of motion tracking or extrapolation.

4. Test non-pilots control and compare their performance to the pilots (to the

overall performance of all the pilots and to the specific subgroups of pilots). This data can be compared to the data obtained in the previous study where pilots and non-pilots were tested on a variety of visual-spatial tasks (Dror, Kosslyn, & Waag, 1992).

FOOTNOTES

REFERENCES

Bahill, A.T. & LaRitz, T. (1984). Why can't batters keep their eyes on the balls?. *American Scientist*, 72, 249-253.

Dror, I.E. (1991). Visual cognitive profile for pilots. AFOSR, Research and Development Laboratories, TR-91-362.

Dror, I.E. (1992). Visual mental rotation: Different processes used by pilots. *Proceedings of the 36th Annual Meeting of the Human Factor Society*.

Dror, I.E., Kosslyn, S.M., and Waag, W. (1992). Visual-spatial abilities of pilots. Harvard University manuscript.

Hayes, J.R. (1981). *The Complete Problem Solver*. Hillsdale, NJ: Erlbaum.

Intriligator. J., Nakayama, K., & Cavanagh, P. (1991). Attentive tracking of multiple moving objects at different scales. *ARVO Annual Meeting Abstract Issue*, 1040.

Kosslyn, S.M., Flynn, R.A., Amsterdam, J.B., & Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition*, 34, 203-277.

Kosslyn, S.M., & Koenig, O. (1992). *Wet Mind*. New-York: The Free Press.

Pylyshyn, Z.W. & Strom, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179-197.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Doners' method. *Acta Psychologica*, 30, 276-315.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24, 295-340.

CAPTIONS

Figure 1: The results from the motion extrapolation task.

Figure 2: The interaction between pilot grouping (fighter vs. non-fighter) and the number of balls in the motion tracking task.

Figure 3: The four-way interaction in the motion tracking task; top panels: error rate as a function of experience, number of balls, and time delay (left panel- pilots who flew recently, right panel- pilots who did not fly recently); Bottom panels: error rate as a function of experience, recency of flying, and time delay (left panel- easy condition; center panel- intermediate condition; right panel-difficulty condition).

Figure 1

Figure 2

Pilots who did fly recently — Mean Error Rate (%) vs Number of Balls

Pilots who did not recently fly — Mean Error Rate (%) vs Number of Balls

Tracking 4 balls — Mean Error Rate (%) vs Time delay

Tracking 5 balls — Mean Error Rate (%) vs Time delay

Tracking 6 balls — Mean Error Rate (%) vs Time delay
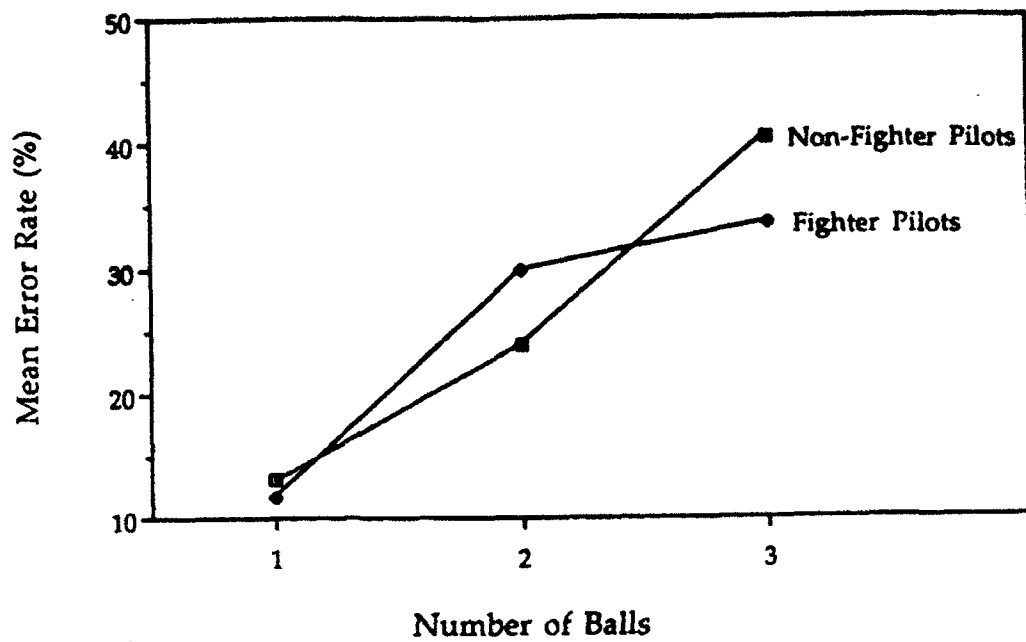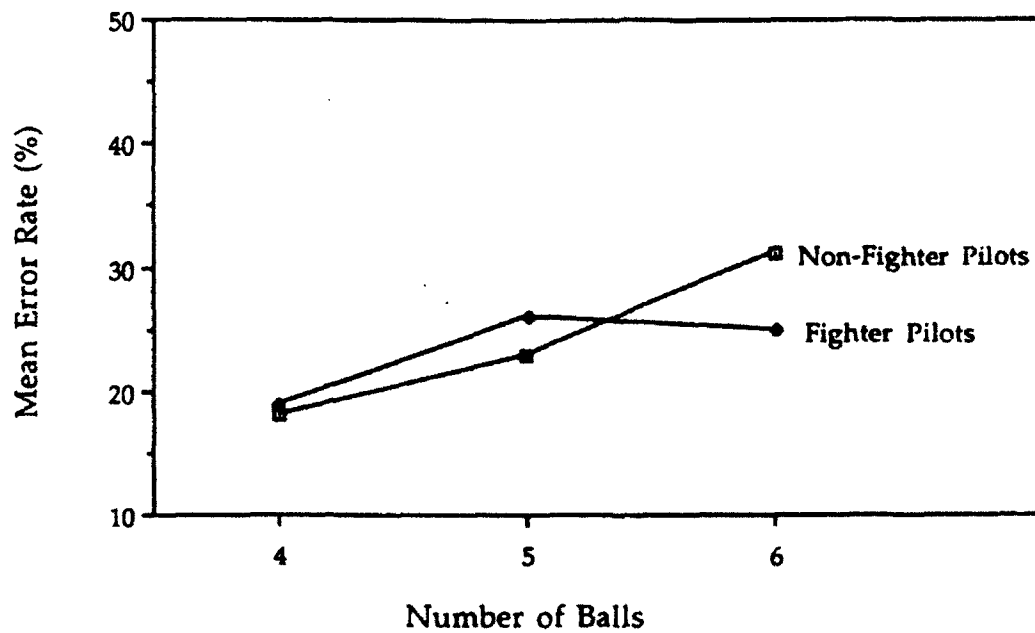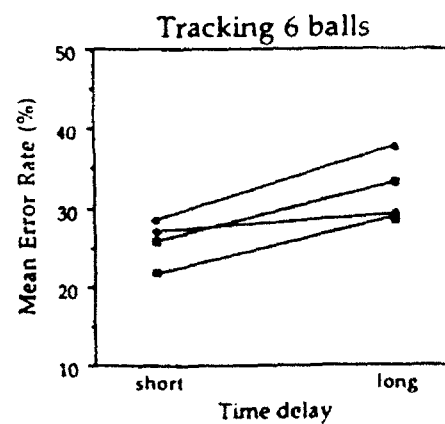
Figure 3

A PRELIMINARY INVESTIGATION OF THE EFFECTS OF A DYNAMIC GRAPHICAL MODEL DURING PRACTICE OF A CONSOLE OPERATION SKILL

John D. Farquhar
Graduate Student
Department of Instructional Technology

University of Georgia
607 Aderhold Hall
Athens, GA 30602

A PRELIMINARY INVESTIGATION OF THE EFFECTS OF A DYNAMIC GRAPHICAL MODEL DURING PRACTICE OF A CONSOLE OPERATION SKILL

John D. Farquhar
Graduate Student
Department of Instructional Technology
University of Georgia

Abstract

A preliminary investigation was conducted to observe the effects of a dynamic model during acquisition of a console operations task. The console operations task studied was a simulation of a procedure involving the operation of a remote crane control arm. One group of subjects had access to a dynamic graphical model during practice of the procedure. The dynamic model appeared as a meaningful displ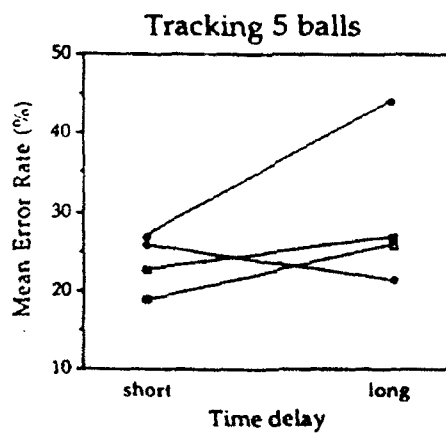ay that responded to console operations. While one group of subjects received the dynamic model during acquisition, all subjects were tested on their knowledge of the procedure without access to the model. Results indicated that during acquisition, time-on-task and errors were roughly equivalent for the model and no-model groups. During testing, however, the model group was able to perform the procedure faster and more accurate. Future studies on the effects of a dynamic graphical model during procedural acquisition are being planned.
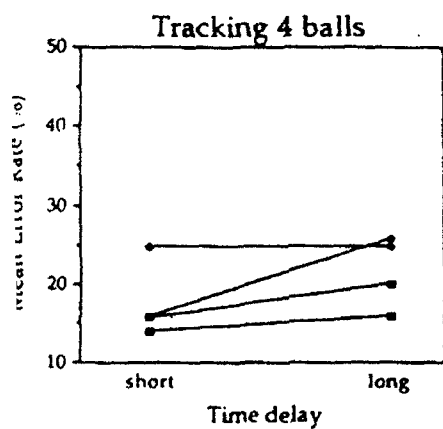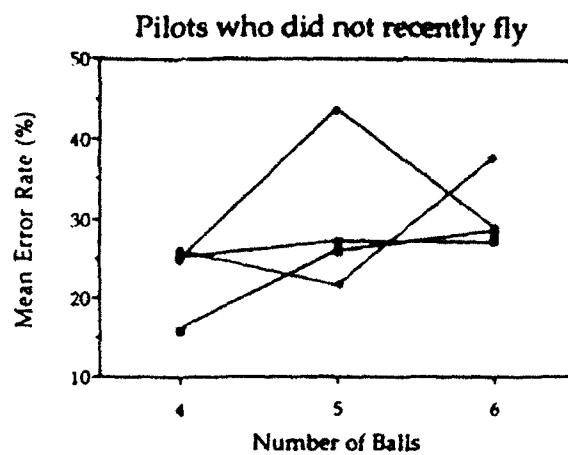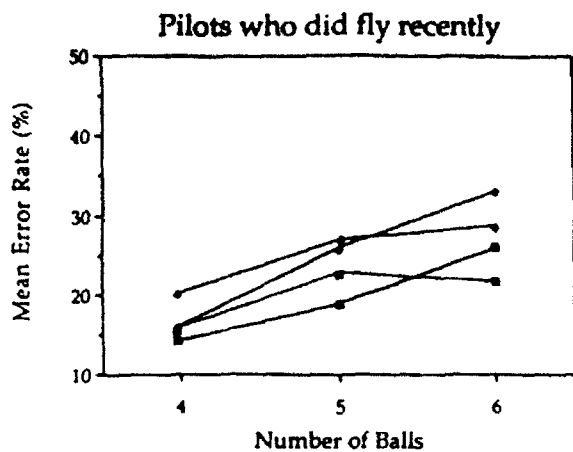
A PRELIMINARY INVESTIGATION OF THE EFFECTS OF A DYNAMIC GRAPHICAL MODEL DURING PRACTICE OF A CONSOLE OPERATION SKILL

John D. Farquhar

INTRODUCTION

Notable theories of cognition and learning have asserted a fundamental distinction between declarative and procedural knowledge types (Anderson, 1983; Kyllonen & Shute, 1989). Using device operation as an example, declarative knowledge is the fact-like (what-it-is, what-it-does, and how-it-works) knowledge of a device. Procedural knowledge is the rule-like (how-its-done) of device operation. Other forms of knowledge, such as qualitative process or mental models describe our understanding of the dynamic causal nature of various devices (Anderson, 1988). Knowledge of a skill, such as device operation, is arguably constructed of various knowledge types or forms of knowledge.

Procedural knowledge is clearly essential for device operation, but the other kinds of knowledge useful in this skill is not as clear. For example, Kieras & Bovair (1984) found that a description of how a device worked, or a device model, affected the acquisition of device operation skills by increasing the speed of acquisition, accuracy of retention, and speed of performance. Reder, Charney, and Morgan (1986), found that certain types of instructional elaborations facilitated the acquisition of procedural computer skills while others did not. Syntactic elaborations (i.e. how-its-done knowledge) improved learning outcomes, whereas conceptual elaborations (i.e. what-it-is knowledge) produced little effect.

This study was simply concerned with the effects of providing a meaningful model of device operation within the context of interactive practice sessions. Little guidance was given on how to interpret the dynamic model, and no instruction directed the subjects to

actively use the model in forming meaning from the device procedure. These deficiencies, however, did not prevent subjects from using the model advantageously.

METHOD

Design: Independent variables for the experiment included the availability of a dynamic graphical model (present or absent from the practice trials), six practice trials, and three test trials. Subjects were randomly assigned to one of two treatments. Each treatment included three phases: an instruction phase, a practice phase, and a test phase. Dependant variables were completion time and number of errors for both sets of practice and test trials.

Subjects: Twenty-nine subjects supplied through a temporary employment agency were paid for their participation in the study. Selection was limited to high-school graduates between the ages of 18 and 27 years. One subject was excluded from failure to complete the final test trial. Subjects completing the experiment early were given additional assignments.

Materials: A console operations task was designed to loosely represent a real-world task in remote crane control arm operation. The task involves the operation of a console to lift, transport, and load canisters from storage bins into rail cars. While skilled console operators follow an interacting set of proscribed procedures, a single 51-step procedure was selected for this experiment.

To provide instruction, practice, and testing for the task, a computer simulation was developed. The simulation displays a complete control panel consisting of a set of 29 interrelated buttons, knobs, and switches. The controls respond to clicks of a mouse with numerous switch and knob settings, meter readings, indicator lights, and an occasional beep. All panel components are appropriately labeled and organized by function on the lower half of the computer screen.

The upper half of the computer screen is reserved for on-screen messages, additional user options (i.e. "Assistance"), and the dynamic graphical display. On-screen messages include the practice or test trial number, a description of the step to perform (e.g. "Set Crane Control to 'Enable'"), and feedback messages for incorrect performance (e.g. "Incorrect. Click as indicated by the arrow.").

Selecting the "Assistance" option reveals a large red arrow indicating the location of the next button or switch, or knob position. An incorrect action in the procedure would automatically select the assistance option and display the red arrow.

The dynamic graphical display, centered at the top of the screen, depicts a representation of the loading task which dynamically responds to control panel actions. The display provides a view of the task as if looking down onto the crane/loading area. Low-fidelity animated sequences demonstrate realistic responses of crane-console interactions.

The simulation software was developed using the ToolBook programming environment and was delivered via a laboratory of 22 individual computer stations. The stations were equipped with Compaq 486/33L microprocessors, VGA-quality monitors, keyboards, and right-handed, three-button mice.

Procedure: Subjects were brought into the laboratory in groups of 15 individuals. One of two treatments was randomly selected for the group. After a brief orientation in use of the equipment, the subjects were instructed to proceed through the program at their own pace. The program included three phases: an instructional phase, a practice phase, and a testing phase.

The instructional phase provided an introduction to the task of operating the crane control arm. This brief tutorial introduced and explained the task of crane control

operations through a dynamic graphical model or display. The graphical model illustrated how the actions of the crane might appear as if the situation was viewed through a window. The instruction did not provide any specific instructions in crane operation, nor did it show or mention the related control panel. Finally, this phase concluded with directions of how the user should interact with the system during the practice and test phases.

During the practice phase, subjects completed 6 trials of the crane control procedure. A text prompt appeared on the screen to direct the subject in performing the individual steps. An on-screen "Assistance" option was available that when selected indicated the location of the next button or switch with a large red arrow.

During practice, one treatment group had available an on-screen display identical to the graphical model presented in the instruction phase. This model represented crane arm-control panel interactions through dynamic changes of the display. The dynamic model was not present for the control group.

Following six practice trials, subjects entered the test phase. In this phase, subjects made three attempts at performing the crane control procedure without the aid of text prompts or assistance. Neither treatment group had available the dynamic graphical model during testing trials. If an error was made during testing, the red assistance arrow would direct the subject to the next step, forcing ultimate correct performance.

RESULTS

Mean completion times and errors for nine trials (six practice, three test) are shown in Figures 1 and 2 respectively. Nearly equivalent completion times and errors were recorded for each of the six practice trials. Test trials yielded significant effects between groups for both completion times, [$F(1,26) = 12.96$, $p < .001$] and errors [$F(1,26) = 15.59$,

9-6

p < .001].

These results indicate that while the presence of a dynamic model during practice of this procedural skill did not alter time-on-task, the model did aid skill performance in testing of the procedure without the model. Further investigations into how dynamic models may facilitate the acquisition, retention, and transfer of procedural skills may yield significant recommendations for the training of device operation skills.
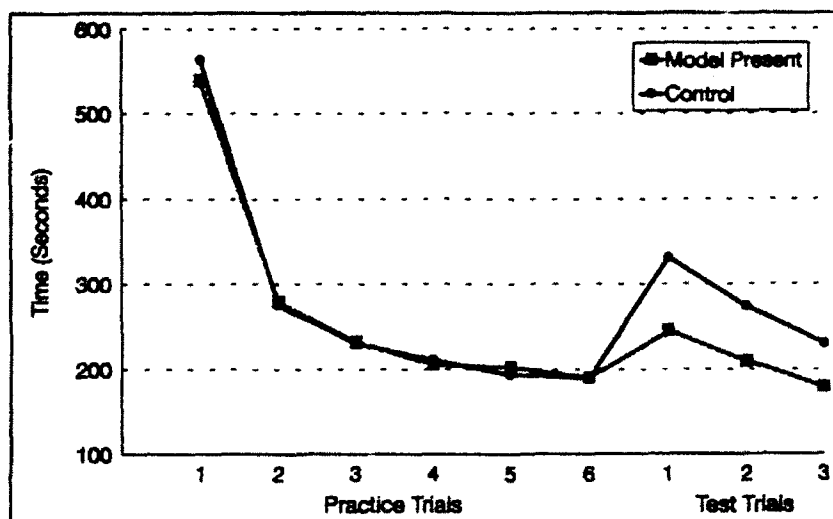
Figure 1: Completion Time by Trial
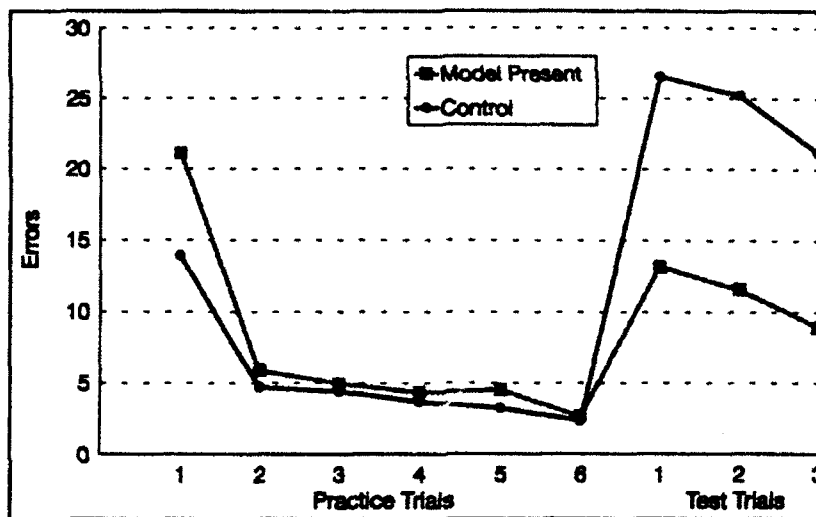
Figure 2: Errors by Trial

REFERENCES

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1988). The expert module. In Polson, M.C., Richardson J.J., and Soloway, E. (Eds.), Foundations of Intelligent Tutoring Systems. Hillsdale, New Jersey: Lawrence Erlbaum.

Kieras, D.E., and Bovair, S. (1984). The role of a mental model in learning to operate a device. Cognitive Science, 8(3), 255-273

Kyllonen, P.C., and Shute, V.J. (1989). A taxonomy of learning skills. In Acherman, P.L., Sternberg, R.J., and Glaser, R. (Eds.), Learning and Individual Differences: Advances in Theory and Research. New York: Freeman & Co.

Reder, L.M., Charney D.H., and Morgan K.I. (1986). The role of elaborations in learning a skill from an instructional text. Memory and Cognition, 14(1) 64-78

VALIDATION OF THE ARTICULATED TOTAL BODY MODEL REGARDING ROLLOVER CRASHES

Jennifer M. Ferst
Graduate Student
Department of Engineering Science and Mechanics

Virginia Polytechnic Institute and State University
Blacksburg, VA 24060

VALIDATION OF THE ARTICULATED TOTAL BODY MODEL
REGARDING ROLLOVER CRASHES

Jennifer M. Ferst
Graduate Student
Department of Engineering Science and Mechanics
Virginia Polytechnic Institute and State University

ABSTRACT

It is known that occupants of vehicle rollover accidents experience complicated and often violent motion. In order to heighten standards for occupant safety during rollover crashes, additional research of occupant behavior in rollover crashes is necessary. Research using actual vehicles would be extremely expensive and time consuming. Alternatively, research on rollover accidents is being done with a human body gross motion simulation program, called the Articulated Total Body (ATB) model. In order to produce reliable data, this model must be validated. Validation is accomplished by comparing actual test results to a computer simulation. In particular, the verification of the ATB model was attempted by comparing the motion of the simulated vehicle and occupant to that recorded on high speed film during an actual 30 mph crash, in which the vehicle rolled over one time and the occupant was a Hybrid III dummy. The vehicle simulation had an acceptable comparison to the test, but the occupant simulation did not compare as well. The difficulty in obtaining a reasonable occupant simulation was due to the unrealistic initial conditions of the test, not to the capabilities of the ATB model. Further work in simulating other rollover tests is necessary to fully validate the predicted motion of an occupant that is produced by the ATB model.

VALIDATION OF THE ARTICULATED TOTAL BODY MODEL
REGARDING ROLLOVER CRASHES

Jennifer M. Perst

INTRODUCTION

Occupant injuries that occur in rollover crashes tend to be very severe and damaging. There were over 9600 persons killed in rollover accidents of passenger cars, pickup trucks, passenger vans, and utility vehicles in 1989. To improve the survivability of an occupant and the crashworthiness of vehicles in rollover crashes, research is necessary. One form of research is to use a computer model to simulate this type of crash. The model that has been used and validated to predict frontal crash results is the Articulated Total Body (ATB) model.[1][2] This model has not been validated in rollover crash situations. Rollover crashes involve three dimensions where frontal crashes involve primarily two. There are many advantages in validating the ATB model. One, if the model can accurately predict the behavior of an occupant involved in a rollover crash, initial conditions of the vehicle motion can be inserted into the model and results can be sought without the physical testing of vehicles. This would save money and time. Two, presently there are no vehicle standards for the automobile industry regarding the protection of occupants if involved in rollover crashes. If the ATB model is validated in predicting occupant behavior, parametric studies could be performed to aid in the development of standards for the automobile industry. A parametric study can include variations such as the study of the effects of roof crush or the study of how different safety devices in a vehicle could prevent injuries. Assuming this were possible, the general safety and welfare of everyone who was an occupant in an automobile could be improved.

DISCUSSION OF PROBLEM

Since the motion of a vehicle and its occupant is extremely violent and complicated in a rollover crash, the Vulnerability Assessment Branch of Armstrong Laboratory has been given the task of validating the ATB model with

respect to this type of automobile accident. This work has been commissioned by the National Highway Traffic Safety Administration (NHTSA). The goal is for the ATB model to be able to predict the dynamics of occupant motion in a rollover crash. The model is being validated by matching the occupant kinematics calculated by computer simulations to tests that have been performed by the Transportation Research Center (TRC) of Ohio. These tests have been conducted with different vehicles including Dodge Caravans and Nissan pickup trucks. A 1989 Nissan pickup truck was evaluated while in the summer research program.

<u>METHODOLOGY</u>

For several years, the ATB model has been used at Armstrong Laboratory for simulating and predicting gross human body dynamics. The model has aided in understanding what might happen in different situations such as aircraft ejection, frontal impacts, automobile panic breaking, and sustained accelerations.[3][4] Recently, the ATB model has been used in the research of automobile rollover crashes. Because of its proven predictive capabilities, the ATB model has been identified and tested as a potential tool to predict the motion of occupants of rollover crashes. However, the model requires validation so future reliance can be attributed to its results.

The model is based on rigid body dynamics. This allows each part of the body to be represented by a rigid segment. The segments are connected in a tree structure by joints. The Hybrid III dummy is represented by a total of seventeen segments coupled by sixteen joints. Each segment has mass, moments of inertia, and a contact ellipsoid associated with it. An example of the occupant, represented by
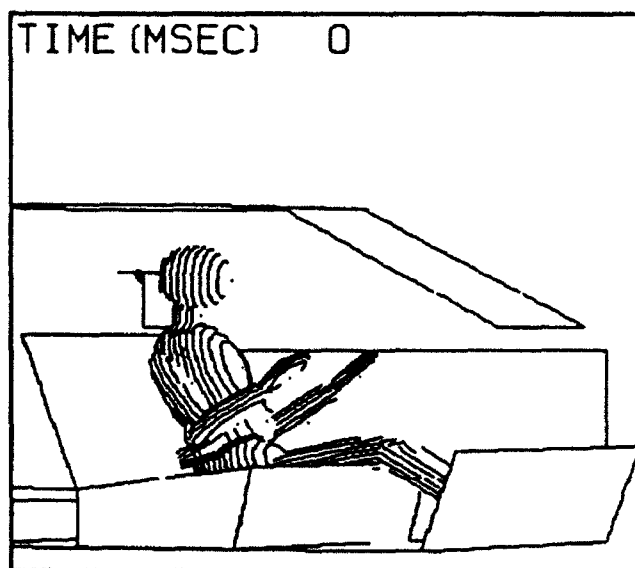
Figure 1

ellipsoids, can be seen in figure 1. He is sitting in the 1989 Nissan pickup truck, which is stationary. The input data file of the ATB model requires definition of the segments, the joint characteristics, the surrounding environment, which is described with planes, the motion of the vehicle, the contacts of the segments with either planes or other segments, the harness belt definition and contact points, and the gravity forces. The output of the model can include time history data of linear and angular accelerations, velocities, displacements, segment-segment contacts, segment-plane contacts, joint forces and torques, and other force or contact information. There is also a visual output available. It graphically depicts the body position with respect to its environment at certain time intervals. This graphical output is available with a graphical display program developed by the Air Force Institute of Technology for a Silicon Graphics computer.

Most of the standard measurements required for the ATB model are based upon the inertial coordinate system. The origin of this coordinate system is defined where test data begins to be collected. The positive x direction is in the direction of motion. The positive y direction is to the right, and the positive z direction is downward. Some information, such as the interior planes of the vehicle, is based upon the vehicle coordinate system. In this coordinate system, the origin is the center of gravity of the vehicle. The positive x direction is toward the front of the vehicle. The positive y direction is to the right side of the vehicle, and the positive z direction is downward. Yaw, pitch, and roll angles are significant factors in the ATB model. Yaw is the rotation about the z axis. Pitch is the rotation about the y axis, roll is the rotation about the x axis.

In order to validate the ATB model, two simulations were performed separately. These were the vehicle simulation and occupant simulation. The vehicle simulation was done first, so that the vehicle kinematics obtained could be used in the occupant simulation. In this case, the vehicle portion simulated the motion of a 1989 Nissan pickup truck, which was released from a NHTSA rollover cart at 30 mph, rolled one full rotation, and came to rest 91.5

feet from its release.[5] The occupant portion simulated the motion of a three-point unibelt restrained and instrumented Part 572E dummy.

Initial conditions were a major and crucial aspect of the vehicle simulation input file. Other information, including the geometry of the vehicle and force deflection characteristics for each possible contact that the vehicle might have were also significant input parameters. These were found in the test report and the data accumulated from the test. The initial conditions that were needed include yaw, pitch, and roll angles, x, y, and z displacements, linear velocities, and x, y, and z angular velocities. Once these initial conditions were calculated, they were entered into the ATB model. The ATB model then simulated four seconds of motion. At this point, the actual test results and simulation results were compared.

For the test, the vehicle was placed on a NHTSA rollover cart situated such that the vehicle's right side would have initial contact with the ground. The vehicle and cart were towed as a unit until they reached a constant speed of 30 mph. The cart and vehicle rolled over a switch which caused hydraulic cylinders to actuate. These cylinders pushed the truck from the cart, and the cart decelerated. The vehicle was instrumented with a three-axis accelerometer and a three-axis angular rate gyro. The rollover cart was instrumented with a three-axis accelerometer. The test was filmed by six high-speed motion picture cameras that recorded at a speed of approximately 500 frames per second. One real time panning camera was also used. The initial roll and pitch at rest were calculated from measurements of the truck on the cart. It was found that the initial roll angle was 29.2 degrees and the initial pitch was .58 degrees. An assumption was made that the initial yaw was -90.0 degrees.

In the ATB model, there would be a difficulty in modeling the truck while it was still on the cart. So, the vehicle simulation began once the vehicle was completely separated from the cart. The time at which the truck has been completely freed from the cart was referred to as the vehicle simulation start time. This time was determined by examining two plots in the

10-6

test report labelled "vehicle/roll
cart separation time - upper switch"
and "vehicle/roll cart separation
time - lower switch."[6]   These
graphs are illustrated in figures two
and three.   The switches that the
plot referred to were those that were
mounted to each side of the truck and
rollover cart.   When the vehicle
broke free from the cart, the
switches were broken, which in turn
sent a signal for each switch.   From
the plots, it appeared that the
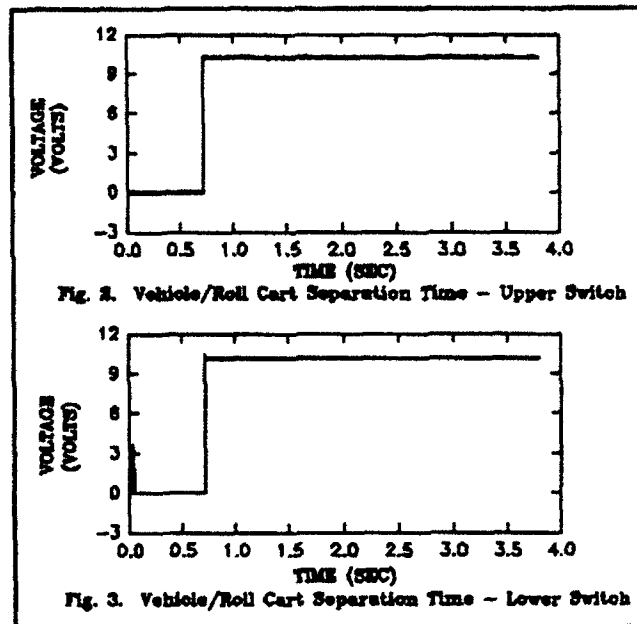
Fig. 2. Vehicle/Roll Cart Separation Time - Upper Switch

Fig. 3. Vehicle/Roll Cart Separation Time - Lower Switch

vehicle was fully separated from the cart at .710 seconds. From further

investigation, mainly viewing the film, it was determined that the truck was

only partially free from the cart at .710 seconds. The vehicle was actually

completely separated from the cart at 1.0 second. This time was chosen to be

time zero for the vehicle simulation.

Since the test start time and simulation start time were different, the

roll, pitch, and yaw angles had to be calculated for the vehicle start time.

This was done by integrating the angular velocity test data in all three

dimensions from 0.0 to 1.0 second.   The integration was calculated using

SigmaPlot.   The integrated number at 1.0 second was the angular position at

that time, which was the beginning of the vehicle simulation.   Therefore, the

roll actually was 87.4 degrees, the pitch was 1.61 degrees, and the yaw was -

92.9 degrees.

The linear displacements were calculated differently for each dimension.

In the z direction, the displacement was the vertical distance between

the ground and the vehicle's center of gravity at one second.   This height was

obtained from using the vehicle's and cart's measurements, photographs of the

truck on the cart, and the roll angle, which was a previous calculation.   This

10-7

distance was found to be -63.12 inches. The linear displacement in the y
direction was assumed to be negligible since the yaw was estimated from
photographs, zero. The x linear displacement was the distance the truck
travelled from the beginning of the test to 1.0 second. The cart acceleration
test data was used to calculate this. The cart data could be used in this
dimension because the vehicle and cart travelled at the same rate for one
second. The calculation involved several integrations, conversions, and an
assumption that the cart reached a constant speed of 30 mph. The vehicle
travelled 428.7 inches in one second. These displacements described the
position of the pick-up at the vehicle simulation start time.

The linear velocity was calculated in the x direction only. In the y
and z dimensions, the linear velocity
was assumed to be zero. These
assumptions were based on the fact
that the roll angle of the vehicle
was approximately 90 degrees.
Therefore, the tangential velocity
had a component only in the x
direction. The linear velocity in
the x direction was calculated using
the cart's acceleration data. The
data was integrated from 0 to 1.0
second, and the rate of 528 inches
per second (30 mph) was added to the
integrated accelerations. So, the
linear velocity in the x dimension
was 429.01 inches/second. This was
the rate at which the truck was
travelling in the x direction at the
vehicle simulation start time.

The angular velocity was

Fig. 4. Vehicle Roll Rate

Fig. 5. Vehicle Pitch Rate

Fig. 6. Vehicle Yaw Rate

10-8
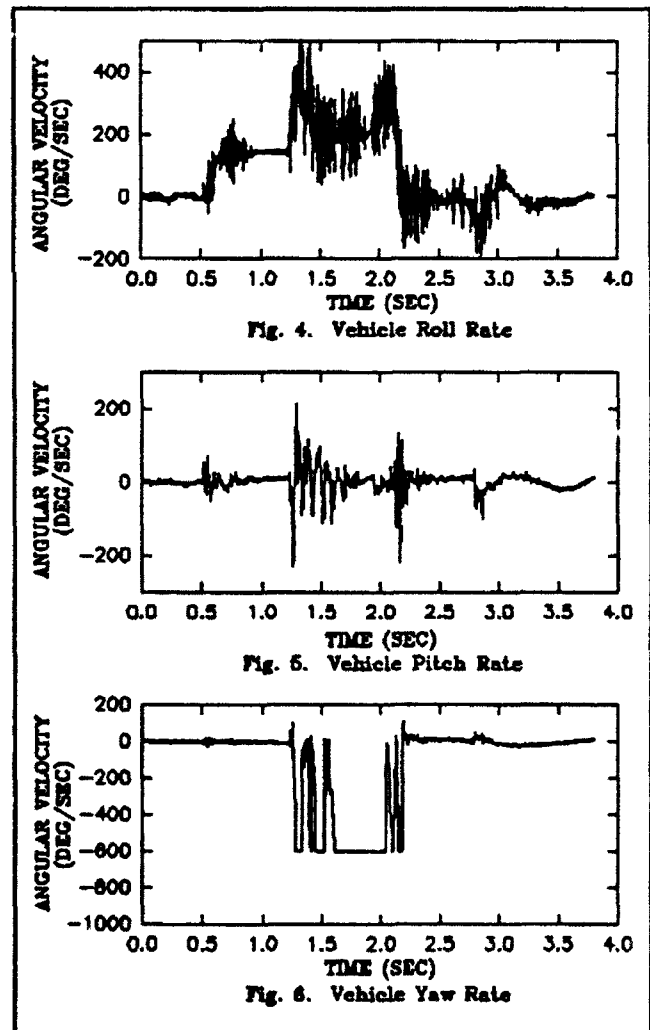
calculated in the x, y, and z directions. These plots are figures 4, 5, and
6. These were determined from the test's angular velocity data. In the test
report, angular velocity was plotted versus time in all three dimensions.
From studying these graphs, the vehicle appeared to have a constant angular
velocity between 1.0 and 1.25 seconds. The data was averaged within this time
interval. In the x direction, the average angular velocity was 145.65
degrees/second. In the y direction, the average was 7.93 degrees/second, and
the average was -7.21 degrees/second in the z direction. These figures
prescribed the rolling motion at the vehicle simulation start time.

Once the vehicle simulation was acceptable, the occupant simulation
input file could be established. First, timing information had to be
acquired. From observing the film, it was evident that the occupant began to
move before the vehicle simulation start time. The occupant simulation should
begin when the occupant is at rest. Generally, the occupant was at rest
before the hydraulic cylinders on the cart began to push the vehicle. So, to
determine the exact time that the
cylinders began to move, plots were
studied. These were in the test
report labelled "roll cart left
cylinder displacement" and "roll cart
right cylinder displacement."[7]
These graphs are illustrated as
figures 7 and 8. The time that the
cylinders began to move was 0.5
seconds. This time was chosen to be
the occupant simulation start time.
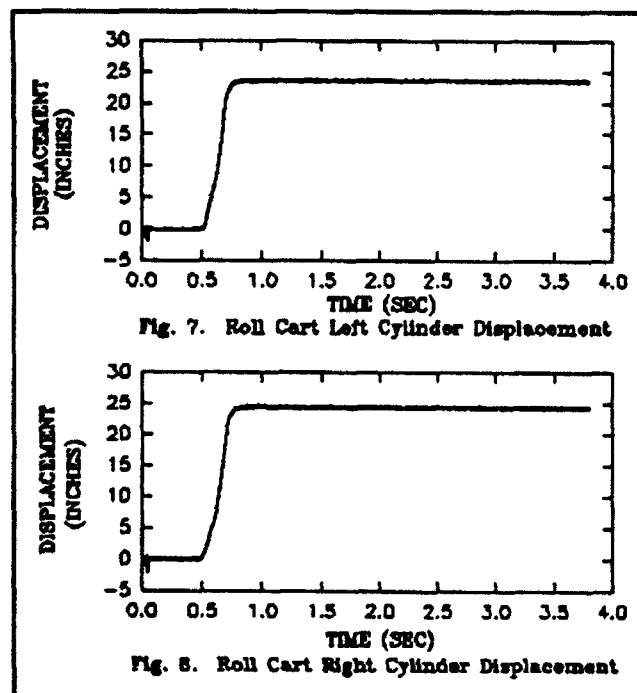
Additional data required for
the occupant simulation were the time
histories of vehicle displacements in
all directions and the vehicle yaw,

Fig. 7. Roll Cart Left Cylinder Displacement

Fig. 8. Roll Cart Right Cylinder Displacement

pitch, and roll angles. Most of these data came from the output of the

vehicle simulation, but some data had to be calculated because of the differences between the occupant simulation start time and the vehicle simulation start time. As stated earlier, the occupant simulation start time was at 0.5 seconds of the test, and the vehicle simulation start time was at 1.0 second of the test. The vehicle displacements and angular orientations had to be calculated between 0.5 and 1.0 second of the test in order to have all of the data needed to run the occupant simulation. These calculations were based on the vehicle test data.

In order to eliminate anomalies in the data, the test data was filtered at 22 Hz. To calculate the angular orientations and vehicle displacements, the same procedure was performed as before when calculating the yaw, pitch, and roll angles, and x and z linear displacements for the vehicle simulation start time. After the calculations had been completed, points were chosen every 40 milliseconds starting at 0.5 seconds of the test. These points were chosen up to 1.0 second of the test. The data from 1.0 to 5.0 seconds of the test was obtained from the time history output files produced by the vehicle simulation.

RESULTS

After the initial conditions were determined for the vehicle simulation, they were inserted into the ATB model input file, and the vehicle motion was simulated. The simulation was compared with the film. The timing of the initial hit was not correct. The initial roll angle did not compare well, and the vehicle's linear displacement in the x direction was not large enough. Adjustments were made to these initial conditions in order for the simulation to be correct. The initial roll angle actually used in the vehicle simulation, in order to match the film, was 95.4 degrees. This change in roll angle altered the calculations of the linear displacements in the x and z directions. The corrected value of the x linear displacement was 434.18 inches, and the z displacement was -59.4 inches. However, additional adjustments were needed for the linear displacements. The vehicle needed to travel further in the x direction in a shorter amount of time. To achieve a

larger x linear displacement, the linear velocity had to be increased. Initially, the assumption was made that the cart reached a constant velocity of 30 mph. This assumption was incorrect. From observing the film, it appeared to be more correct to assume that the cart reached a constant velocity of 31 mph. Therefore, the linear velocity in the x direction was changed to 447.0 inches/second. This new assumption altered the x displacement to 493.7 inches. The z linear displacement had to be adjusted to -49.12 inches. This change allowed the vehicle to fall to the ground in a shorter amount of time. The need for this adjustment could be due to the many assumptions that were taken while originally calculating the z linear displacement, such as scaling cart dimensions from a photograph. After all changes were made, the vehicle simulation was acceptable and matched the vehicle motion on the film fairly well.

After the necessary calculations, all of the information needed to run the occupant simulation was entered into the occupant simulation input file. The simulation was run, and the simulated occupant initially did not behave similarly to the test's occupant. The origination of the problem appeared to be in the x and z displacement data from the occupant simulation start time to 0.5 seconds of the simulation. This was the data that was calculated, not from the vehicle simulation output data. To visually see the data the x and z displacements were graphed versus time. In order to eliminate data points that might cause problems in the simulations, curve fits were performed for the graphs of the x and z displacements between the times of 0.0 to 0.7 seconds of the simulation. The curve fits generated new data points. These were entered into the input data file, and the simulation was run again. The occupant's motion followed the film for the first 400 milliseconds, but then travelled in the opposite direction from what he should have. Every known source of error was accounted for, and the simulated occupant still would not follow the motion of the film's occupant. He moved too far across the seat toward the passenger's side. The simulation was considered complete.

10-11

CONCLUSION

The problems that occurred in the occupant simulation seemed to originate in the calculated vehicle displacement data. This is thought to happen because when the occupant simulation began, the vehicle was still on the rollover cart. Since presently the cart was not modelled, the data calculated to account for that motion was an educated guess. If the vehicle simulation start time and occupant simulation start time were identical, these problems theoretically would not have occurred. Therefore, the ATB model still has the potential to predict the motion of an occupant who is involved in a rollover crash, but further work is required. This work would involve being able to model the vehicle motion while on the cart so the occupant simulation start time and vehicle simulation start time were the same. More tests of the Nissan pickup truck will be modelled to clearly identify if this is a valid problem.

REFERENCES

1. Obergefell, L.A., I. Kaleps, and S. Steele. "Part 572 and Hybrid III Dummy Comparison Sled test Simulations." SAE Paper No. 880639, Feb. 1988.

2. Obergefell, L.A., T.R. Gardner, I. Kaleps, and J.T. Fleck. "Articulated Total Body Model Enahncements; User's Guide." Volume 2, AAMRL-TR-88-043. Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division, Air Force Systems Command, Wright Patterson Air Force Base, OH, 1988.

3. Kaleps, I. "Prediction of Whole-Body Response to Impact Forces in Flight Environments." AGARD Conference Proceedings No. 253, Paris, France, November 1978.

4. Kaleps, I. and J.H. Marcus. "Predictions of Child Motion During Panic Breaking and Impact." Proc. 26th Stapp Car Crash Conference, SAE Paper No. 821166, October 1982.

5. "Vehicle and Dummy Kinematics in a Controlled Rollover Crash; 1989 Nissan Pickup Truck Test Report." The Transportation Research Center of Ohio, East Liberty, Ohio, November-December 1989. (Non-Published)

6. "Vehicle and Dummy Kinematics in a Controlled Rollover Crash; 1989 Nissan Pickup Truck Test Report." The Transportation Research Center of Ohio, East Liberty, Ohio, November-December 1989. (Non-Published)

7. "Vehicle and Dummy Kinematics in a Controlled Rollover Crash; 1989 Nissan Pickup Truck Test Report." The Transportation Research Center of Ohio, East Liberty, Ohio, November-December 1989. (Non-Published)

Transmission Delays and Bursting in Statistical
Mechanical Neural Network Models

Michelle A. Fitzurka
Doctoral Candidate
Department of Physics

The Catholic University of America
620 Michigan Avenue, N.E.
Washington D.C. 20064

Transmission Delays and Bursting in Statistical Mechanical Neural Network Models

Michelle A. Fitzurka
Doctoral Candidate
Department of Physics
The Catholic University of America

Abstract

The ultimate aim of this study is to further our current understanding of the electroencephalogram (EEG), or electrical recording of the activity of the human brain, in its resting state so as to better comprehend the effect of external fields applied to it. To this end, we invoke a neural network model that employs the language and formalism of statistical mechanics to order to explain the behavior of neurons in the brain on average, speaking of the *probability* of a neuron firing, rather than predicting the individual firing pattern of a specific neuron at a specific time. While there is much to be gained by this novel and rich approach, caution must be utilized in the mathematical development and application of any such model to bioloical systems. We intend to investigate the current statistical mechanical neural network models and suggest ways in which to impose more stringent, biological accuracy on models of brain functioning while retaining mathematical tractability. Eventually, we see a need to summon the techniques of non-equilibrium statistical mechanics, a promising direction which to date has not been wholly explored.

Table of Contents

Transmission Delays and Bursting in Statistical Mechanical Neural Network Models

M. A. Fitzurka

Introduction:

Overview of the Literature and Impetus for Further Investigation

The history of the mathematical modeling of complex dynamical systems is both long and vast, as is the legend of the modeling of biological systems. The record of the mathematical modeling of biological systems *as* complex dynamical systems, however, is not as rich and has been limited to those resolute mathemeticians who have attempted to remain faithful to biological realism. While there exists a wealth, a veritable explosion, of information on the utility of *biologically-motivated* mathematical models, strict adherence to biological integrity has been neither widespread nor rigorous. This failing has been well noted by Dr. Judith Dayhoff, [1] in her textbook *Neural Network Architectures: An Introduction*, wherein she provides a detailed delineatation of the various differences between current artificial neural network (ANN) models and true biological neural networks (BNN). See Table 1 in Section 2.2. There is a striking urgency for further rigor in addressing the precise restrictions of biology.

Historically, coherence to biological realism has been avoided due to the greatly complicated, and sometimes intractable, mathematics that is involved as a result of levying strict biological requirements. Also, in many instances, extraordinary practical applications have overshadowed the necessity for imposing biological accuracy. The broad range of successful, though biologically imprecise models has forced the demands of biology to therefore be selectively ignored or postponed. Consequently, for reasons of mathematical complexity and practical efficacy, biologically-correct neural network models have not been fully developed though the need for them is profound. It is our hope that this trend be reversed.

---

[1] Dr. Dayhoff has a PhD in biophysics from the University of Pennsylvania and works at the Naval Surface Warfare Center in Md. She is on the board of the International Neural Network Society (*INNS*) and has expressed interest in my work, as has the President of *INNS*, Dr. Harold Szu. I am grateful that I will be able to rely on their knowledge and experience for insight into my research.

Of the many and interesting problems that could conceivably be enlightened by more stringent studies of biologically-precise mathematical models, those questions dealing with the health and safety of human beings would inarguably be of highest priority. Taking precedence, therefore, would be an investigation for a precise model of the resting (EEG) state of the human brain, which as of yet, has eluded conclusive explanation. This most basic yet fundamentally crucial element in the understanding of human functioning remains unclear despite our current and extensive knowledge of man, so its decoding would be of great benefit. This defines our motivation.

We proceed by expounding a model of J. D. Cowan. In 1970, Cowan resurrected and remarkably refined the mathematics of a model of the statistical dynamics of population kinetics introduced initially by E. H. Kerner in 1957. His model has enormous potential and represents the starting ground for our analysis. A detailed description of this model is in order at this point. Afterwards, there will be a discussion of our intention to improve this scheme, followed by an indication of what we forsee as the future possibilities and advantages of pursuing this perspective, most especially with regard to the safety and integrity of man.

1. <u>Equilibrium Models</u>:

The model of Cowan belongs within a category of models which evoke equilibrium statistical mechanics, and therefore we term this section *equilibrum models*. It will also be seen that as an extension to Cowan's work, more advanced statistical mechanics will eventually be required. Accordingly, there is a section to follow covering these more sophisticated topics.

1.1 - Overall Justification of a Statistical Mechanical Approach

In order to warrant the employment of a statistical mechanical approach in explaining the functioning of the brain, it is necessary to view the brain's average behavior as an aggregate of the behavior of the individual neurons that comprise it. As there are roughly 10^{11} neurons [1] each with 10^5 axonal connections [2] to other neurons; there are therefore a minimum of 10^{16} totally interconnected pathways within the brain, a number that approaches Avagadro's number, $N \sim 10^{23}$. Certainly, it is fitting and proper for us to consequently consider the "motion" (firing rates) of each of these "particles" (neurons) on average, thus invoking the entire language and formalism of statistical mechanics.

Many other important averages and quantities of ultimate interest may be determined by taking advantage of the machinery of statistical mechanics. The average firing rates of neurons, or *activity* in the brain is considered to be the formal analogue of "energy" and a subsequent Hamiltonian mechanics proceeds. The "temperature," corresponding to the *fluctuation* in this activity, admits a Gibbs Ensemble theory for these variations. Obtaining these and other desired values requires the introduction of some of the vocabulary of neurophysiology, neural networks, physics, and biophysics.

1.2 - Background of Cowan's Work on Neuronal Models

The model developed by Cowan borrows heavily from physics and cultivates several of the fundamental concepts and ideals used in physical modeling. The first of these pertinent analogies enlists the aid of the terminology of the classical "lumped equivalent-circuit" of elementary circuit theory in order to describe the functioning of the somato-dendritic neural membrane. Understanding of the mechanisms dictating the workings of this membrane is key in developing a cellular or neuronal model justifying this as a valid starting point. Figure 1, taken directly from Cowan's article (1970), depicts the equivalent analogous circuitry representing neural membrane behavior. To first approximation, all of the relevant information on the linear depolarization of the membrane can be represented by the following first-order differential equation

$$\left[\tau \frac{d}{dt} + 1\right] v(t) = \frac{\tau}{C_m}\left[I_m + \frac{(-E_m)}{R_m} + \frac{(E_i - E_m)}{R_i}\right] \tag{1}$$

where τ is the *time constant* of the membrane given by

$$\frac{1}{\tau} = \frac{1}{C_m}\left[\frac{1}{R_m} + \frac{1}{R_e} + \frac{1}{R_i}\right] \tag{2}$$

reducing to $\tau = R_m C_m$ with open-circuited synapses. These equations and the associated Figure 1 provide a reasonable basis for initiating a mathematical description of biological phenomena.

In this scheme, the neural membrane is a linear system with I_m as an *applied membrane current*. C_m is the *membrane capacitance*. R_e and R_i are the *excitatory* and *inhibitory synaptic resistences* combined in parallel. The g_e and g_i terms represent *exictatory* and *inhibitory conductivities* while g_m is the *membrane conductivity*. E_i and E_m stand respectively for the *inhibitory membrane* and *resting membrane potentials* while $v(t)$ represents the membrane potential measured relative to the resting potential, in this case $v_{inside} - v_{outside} - E$. Resting potentials at-

tributable to the cell or neuron's membrane while it is quiescent are typically −70 mv as indicated in the figure. The IPSP and EPSP symbols represent the *inhibitory* and *excitatory post-synaptic potentials*, labels given to the responses to the pre-synaptic excitation [3].

outside

inside
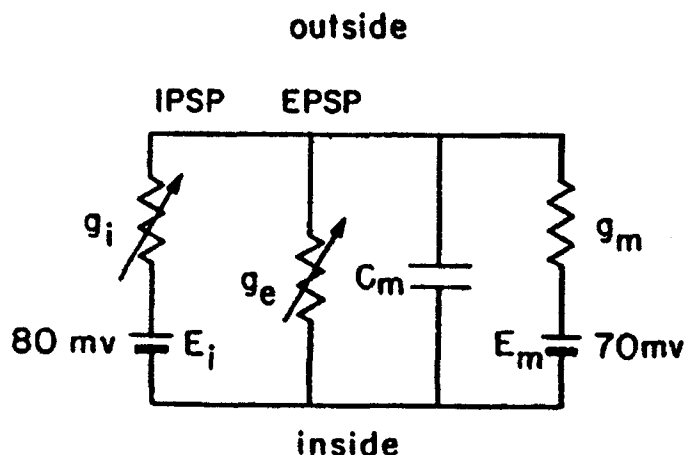
Figure 1: Circuit Model for the Somato-Dendritic Neural Membrane. (Source: Cowan, J.D. 1970, *Some Mathematical Questions in Biology in the Life Sciences*, p. 48.)

The cell membrane, $\approx 75\overset{\circ}{A}$ in thickness, is a highly ordered, bimolecular lipoprotein layer which separates the fluids between and within the cells, *interstitial* and *intracellular fluids* respectively. The composition of these fluids is mostly water with comparable numbers of particles per unit volume dissolved within, sodium $Na+$ and chloride $Cl-$, in the interstitial fluid and potassium $K+$ in the intercellular fluid. Variations in the ionic concentrations in and outside the cell contributes to the electric potential difference between these two fluids, which is relieved or lessened only by the *passive diffusion* or *active transport* of these ions through *channels* or *gates* in the membrane layer. This unidirectional *flux*, (inward at an excitatory synapse; outward at an inhibitory synapse), is opposite to the *concentration gradient* and continues until the net flux is zero and *ionic equilibrium* is reached [4].

In its active state, the neuron is termed either *depolarized*, when the reduction in the magnitude of the potential across the membrane toward 0 mv [5] has the effect of decreasing the negativity of the interior of the cell with respect to the exterior, or *hyperpolarized*, when there is an increase in the potential across the cell's membrane, the inside becoming more negative, approaching −80 mv [5]. Depolarization or hyperpolarization, therefore refers to the net outward or inward ionic movement or flux of ions across the neuronal membrane at *inhibitory* and *exci-*

tatory synapses. These synapses, or junctions, are the connection points between the terminal axon of the antecedent cell to the somato-dendritic membrane of the latter cell.

It is the *impulse,* or *action potential* that is responsible for initially depolarizing the cell until it reaches a characteristic *threshold* value, at which time its reaction is to initiate an impulse which is then transferred to the consecutive neuron via the *synaptic gap.* The neuron becomes hyperpolarized and a *refractory period* of time occurs before it returns to its resting state, where it is able once again to fire. It is the specific nature of this impulsive response of a cell's membrane from its resting state, its action potential, that is a matter of fundamental importance and marks our point of departure from Cowan's work. While Cowan allows for the transmission of single impulses in the mechanics he derives, he does not treat refractory times, the all-or-nothing aspect of the propagated action potential, the finite speed of the action potential, or neuronal bursting.

For the action potential, Cowan assumes the *Heaviside-Dirac* impluse function, to which the membrane responds as

$$h_{ij}(t) = u_{-1}(t) [\delta v_{ij}] \exp(-t/\tau), \tag{3}$$

u_{-1} being the Heaviside step function and δv_{ij} being the deviation of the membrane potential due to the arrival at the jth cell of an impulse from the ith cell. He choses this first-order linear equation, which proceeds directly from EQN(1), since it describes well the *all-or-nothing* character of neurons. This *all-or-none* behavior refers to response of the neuron being fixed in size, shape, duration, and conduction speed [6]. There is also included a decay term which accounts for the drop-off in the effect of the impulse in time. The accumulated effect, or "built-up" potential in the membrane, is consequently represented with the following convolution containing EQN(3)

$$v_i(t) = \sum_{j=1}^{N} \int_0^t h_{ij}(t - \tau) f_j(\tau) d\tau \tag{4}$$

with the introduction of the term $f_j(\tau)$ as the mean frequency of the arrival of impulses from the jth primary cell. This equation carries implicitly the linearity of the system, incorporating the aggregate nature of the neuron by accouting for its linear response to multiple inputs.

Finally, with all the constants written explicitly and reduced by proper approximations, Cowan derives an expression for the transmembrane potential resulting from a large number of inputs

$$v_i(t) = (\sigma/g_m) \left(\sum_j \delta g_{ij} E_j f_j(t) \right) [1 - \exp(-t/\tau)].$$ (5)

The effective operating period for the transmission of excitation from primary to secondary cell is σ, the membrane and corresponding change in membrane conductance are g_m and δg_{ij}, and the change in membrane potential caused by the action of the impulse from the jth cell is E_j. This expression is the result of the reduction of a linear differential system, EQN(1), to a non-linear function of time, EQN(5), and provides a straightforward means of calculating the response of a membrane to its many inputs. With these terms defined accordingly, we are now in a position to show how Cowan manipulates these quantities to develop a statistical mechanical formulation for the activity of a large ensemble of neurons.

1.3 - Cowan's Statistical Mechanical Neuronal Model

The basic equations of Cowan are extensions of the work begun by Kerner in 1957 as previously indicated. The equations of Kerner in turn find their origin and in the monumental treatise of Volterra on population dynamics in 1931 most probably instigated by the suggestion of A.J. Lotka, who in 1925 remarked that "... what is needed is an analysis...that shall envisage the units of a biological population as the established statistical mechanics envisage molecules, atoms, and electrons; that shall deal with such average effects as population density, population pressure and the like after the manner in which thermodynamics deals with the average effects of gas concentration, gas pressures, ..." [7].

While Volterra proposed a system of differential equations to describe the variation in time of the populations N_r of interacting species in a biological association, it was Kerner who noticed first that this system admits a Liouville's theorem, when the log N_r are taken to be the variables, and second that a universal "integral of motion" ensues [8]. Cowan's contribution to this unique development, as we shall see, is that he calculates explicitly the statistical mechanical averages according to Kerner's prescription and shows unambiguously its relation to the specific problem of neural activity.

The focal point around which the dynamics of Cowan is centered is his *neural equation* associating the input $f_i(t)$ and output $f_j(t)$ firing rates. It is from this equation that the *neural Hamiltonian* and the subsequent statistics succeed. In EQN(5), we already have the connection

between the output firing rates and the membrane potential. By recognizing the following basic relationship

$$\delta v_{ij} = \delta Q_{ij}/C_m = i\delta t/C_m, \qquad (6)$$

we can convert the change in membrane potential $\delta v_i(t)$ to the corresponding current $i\delta t/C_m$ produced. Taking appropriate time averages in EQN(5) and solving for $i(t)$ in EQN(6) gives us the current flowing in the cell membrane as a function of time

$$\langle i_i(t) \rangle = \left(\frac{c\sigma}{C_m} \right) \sum_j \delta g_{ij} E_j f_j \langle \exp(-t/c\tau_m) \rangle \qquad (7)$$

or

$$\langle i_i(t) \rangle = \left(\frac{c\sigma\alpha_i}{C_m} \right) \sum_j \delta g_{ij} E_j f_j \qquad (8)$$

with $\langle \exp(-t/c\tau) \rangle$ reducing to $\alpha_i/c\tau_m$. Now, we have the output firing rates also related to the current.

Next, Cowan choses to employ a *logistic function* to represent the nonlinear input-output relation of the cell. This function is a logical choice due to its inherently statistical nature and has some support in the biological literature; however, there is no reason to see *a priori* the distinct advantages that follow from its use. It is chosen in part because the logistic function is one of the easier sigmoidal curves to work with [9]. Thus, according to Cowan, the mean rate at which the ith cell emits impulses as a function of membrane current is

$$f(t) = \left[r \left(1 + \exp \left[-\beta \left(\left[\frac{i(t)}{i_{th}} \right] - 1 \right) \right] \right) \right]^{-1} \qquad (9)$$

This equation can be manipulated further by substituting EQN(8) into EQN(9) to yield

$$f_i(t) = \left[r \left(1 + \exp \left[- \left(\frac{\beta c\sigma\alpha_i}{C_m i_{th}} \right) \sum_j \delta g_{ij} E_j f_j(t) + \beta \right] \right) \right]^{-1} . \qquad (10)$$

Here β is simply

$$\beta = \left(1 - \frac{i_0}{i_{th}} \right)^{-1} \ln \left(\frac{1 - rf_0}{rf_0} \right) \qquad (11)$$

for reasons which become more clear when β is understood to be the equivalent of $\frac{1}{kT}$, k being the standard Boltzmann's constant of statistical mechanics equal to $1.38 \times 10^{-23} \frac{J}{deg}$.

11–10

Finally, we have an expression relating the output firing rates directly to the input firing rates. We can further rearrange this equation, by letting

$$x(t) = 1 - rf(t) = 1 - r/\tau \tag{12}$$

$$\delta g_{ij} E_j = \alpha_{ij} \tag{13}$$

$$\left(\frac{\beta c \sigma \alpha_i}{C_m i_{th} r}\right) = \frac{1}{\beta_i} \tag{14}$$

$$\epsilon_i = \beta - \frac{1}{\beta_i} \sum_j \alpha_{ij}, \tag{15}$$

to reexpress EQN(9) as

$$\ln \frac{x_i(t)}{1 - x_i(t)} = \epsilon_i + \frac{1}{\beta_i} \sum_j \alpha_{ij} x_j(t). \tag{16}$$

The x variable is called the *sensitivity* or *neural sensitivity*. With the neural equation in this form, Cowan makes one final assumption, and one final transformation, before extracting the neural Hamiltonian, G.

EQN(16) is an equation that is valid for *quasi-stationary* inputs. To have a dynamical equation for neural responses that have *non-stationary* inputs, Cowan uses the following ordinary non-linear differential equation in the x variable,

$$\left(\tau \frac{d}{dt} + 1\right) \ln \left(\frac{x_i}{1 - x_i}\right) = \epsilon_i + \frac{1}{\beta_i} \sum_j \alpha_{ij} x_j(t). \tag{17}$$

This equation, which is a dynamical equation for neural responses, can also be written, with the subsidiary conditions that $\alpha_{ij} + \alpha_{ji} = 0$ and $\alpha_{ii} = 0$, as

$$\tau \frac{dx_i}{dt} = \left(b_i u_i + \frac{1}{\beta_i} \sum_j \alpha_{ij} x_j\right) x_i(1 - x_i) \tag{18}$$

where $b_i u_i = \epsilon_i$. In obtaining EQN(18), Cowan neglects the second term on the left-hand side of EQN(17), which he refers to as the *damping term*. Now, we are in a position to consider easily the *stationary states* of the system. As is evident from EQN(18), there are three such stationary states: $x = 0$, $x = 1$ and

$$q_j = \sum_j A'_{ij} \beta_i b_i u_i. \tag{19}$$

Eventually, it is shown that the final ensemble average $\overline{x_i}$ and final time average $\langle x_i \rangle$ of the sensitivity, x_i, are both equal to the stationary state, q_i, defined in EQN(19). This verifies the *property of ergodicity* for this system: $\overline{x_i} = \langle x_i \rangle$. In fact, all ensemble averages may be replaced by time averages [10].

With the following transformation equations

$$v_i = \ln\left(\frac{x_i/q_i}{1 - x_i}\right) \tag{20}$$

$$T = t/\tau \tag{21}$$

the Hamiltonian, G, finally emerges

$$\frac{dv_i}{dT} = \sum_j \left(\frac{\alpha_{ij}}{\beta_i \beta_j}\right) \frac{dG}{dv_j} \tag{22}$$

$$G = \sum_i \beta_i \left[\ln(1 + q_i \exp v_i) - q_i v_i\right]. \tag{23}$$

The "labor" is essentially completed with the acquisition of the system Hamiltonian, the subsequent statistical mechanics is straightforward. If the corresponding *canonical density* is defined as

$$\phi = a \exp\left(-b \sum_{i=1}^{2k} G_i\right), \tag{24}$$

then appropriate ensemble averages for the system may be considered. Whereas Cowan calculates these explicitly, we will merely reference a few of the significant results for demonstrative purposes.

One important parameter to derive is the temperature, defined by Kerner in a previous work as: a measure, in one number common to all the species specifying the mean square deviations of the N_r (population number) from their average values [11]. Given, that

$$\overline{\frac{\partial G}{\partial v_i}} = 0, \tag{25}$$

and

$$\frac{\partial G}{\partial v_i} = \beta_i (x_i - q_i), \tag{26}$$

it can be shown that

$$\overline{v_i \frac{\partial G}{\partial v_i}} = \frac{1}{b}, \tag{27}$$

11-12

and

$$\overline{\beta_i^2(x_i - q_i)^2} = \overline{\left(\frac{\partial G}{\partial v_i}\right)^2} = \frac{(\frac{1}{b})\beta_i q_i(1 - q_i)}{(1 + (\frac{1}{b})\beta_i)}. \tag{28}$$

Solving for the quantity $(\frac{1}{b})$, which he sets equal to θ, Cowan gets the *temperature* or *kinetic energy* of the net, which is *equipartitioned*, or distributed evenly, throughout the net. Thus,

$$\left(\frac{1}{b}\right) = \theta = \left(\beta_i \frac{\overline{(x_i - q_i)^2}}{q_i(1 - q_i)}\right)\left(1 - \frac{\overline{\beta_i^2(x_i - q_i)^2}}{q_i(1 - q_i)}\right) \tag{29}$$

If the Hamiltonian, G, is seen as the *activity* of the network, then this temperature term is considered to be the *amplitude of fluctuation* of the activity. To complete the interpretation of the canonical density, we note that, by definition

$$a = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{(-G/\theta)}dv_1 dv_2\cdots dv_{2k}$$
$$= \prod_1^{2k}\int_{-\infty}^{+\infty} e^{(-G_i/\theta)}dv_i \tag{30}$$
$$= \prod_1^{2k}(q_i^{\beta_i q_i/\theta})\mathcal{B}\left(\frac{\beta_i q_i}{\theta}, \frac{\beta_i(1 - q_i)}{\theta}\right)$$

with $\mathcal{B}(p, q)$ representing Euler's β-function.

From this canonical density, Cowan proceeds to derive the moments of the β-density, the mean, the variance, the time spent above and below the mean, the average *crossing-rate* at which neural fluctuations occur, and the *cross-correlation functions* of the joint activity of two or more cells. Finally, he gives some practical applications with regard to the *thalamo-cortical interaction*, which he sites in proof of his work.

A couple of comments require mention at this point. The mathematics that follows from the dynamics of Cowan, as outlined here, carry three implicit assumptions which deserve attention. First, as introduced prior, there is the matter of the single or impulsive *vs.* packets-of-pulses or bursting nature of the action potential, as seen by referencing EQN(3). In addition to the supposition of a single-pulse character for action potentials, Cowan makes another assumption not cleanly supported by biological knowledge. There is no accounting in Cowan's kinematics for variable delay times as he considers the transmission of signal from neuron to neuron to be instantaneous. Finally, c.f. EQN(18), Cowan models *tonic* cells, that give maintained responses to constant stimuli and not *phasic* cells, which respond only to changing stimuli [12].

11-13

The latter of these issues was to be addressed by Cowan subsequently and will not be the focus of this intended investigation. The former two concerns, that of "bursting" and "transmission delays," as we shall call the two phenomena, will be pursued presently.

2. Non-Equilibrium Models:

As indicated previously, alterations that we make to Cowan's model lead eventually to an analysis that requires more sophistication in the mathematical techniques employed. Although our initial analysis remains at the level of *equilibrium* statistical mechanics, we includ۰ it in this section to emphasize the direction that our mathematics will ultimately take. Fortunately, many of the tools available in non-equilibrium statistical mechanical theory equip us well for the mathematical intricacy that evolves from our approach.

2.1 - Our Corrections to Cowan's SMN Model

Remediation of Cowan's assumption of Heaviside-Dirac pulses and neuron-neuron communication, together with his omission of variable transmission delay time between neurons provides us a working goal. Toward this end, we will make our own set of postulates that we feel lead to more biologically accurate models. Of these, we feel our most critical alteration is the idea that the introduction of a logistic function into the dynamics be postponed until after the definition of the operational kinematics of the individual neurons. This makes the entire derivation less heuristic and therefore more exact. To see this quantitatively, let us compare the derivation of Cowan, and in particular EQN(9) and EQN(17), with the following, which we propose.
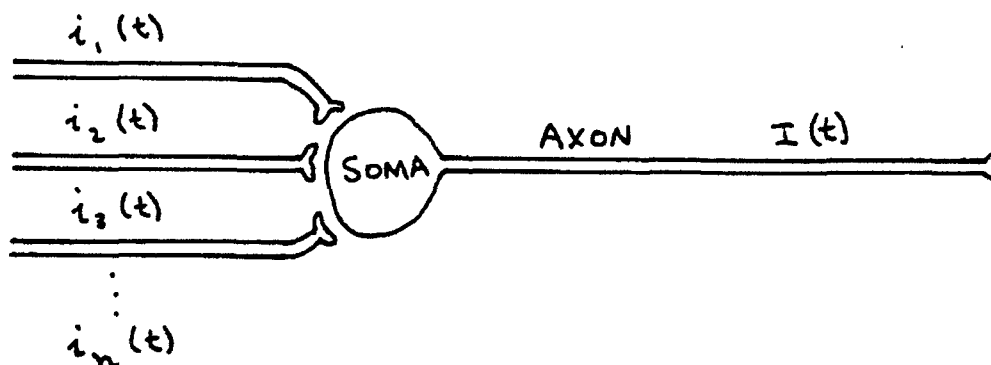
Figure 2: Model for the Neuron as a Summator.

Beginning, as did Cowan, with a model of the neuron as, essentially, a summator: (see Figure 2) taking the collection of combined inhibitory and excitatory signals that impinge upon its cell body, or *soma*, from all antecedent cells (a process known as *convergence*), adding them together by *integration*, and transmitting the singular result through its axon to all sequential cells (*divergence*) [13], we perceive of a situation where the details of the manner in which integration and cell thresholding occurs is critical.

Considering the cell body membrane system to be linear, the membrane potential comparable to EQN(4) in our scheme is

$$v_m(t) = \int_{-\infty}^{+\infty} g(t - \tau) \sum_{j=1}^{n} i_j(\tau) d\tau = \sum_{j=1}^{n} \int_{-\infty}^{+\infty} g(t - \tau) i_j(\tau) d\tau \qquad (31)$$

where for analytical work

$$g(t - \tau) = \exp{-l(t - \tau)} \qquad (32)$$

Here we are representing actual action potential integration rather than following Cowan who used a firing rate relationship based on the assumption of closely spaced Heaviside-Dirac pulses.

The cell body releases the action potential once the membrane potential reaches a critical value, v_c, changing the cell's conductivity to ions which respond by moving through the prescribed *gates* in the membrane and as a result change the potential. If we let $C(t)$ be the concentration of ions in the cell at a given time

$$\frac{dC(t)}{dt} = -\{k_0 + k_1 \mathcal{H}[v_m(t) - v_c]\}[C(t) - C_E] - h[C(t) - C_R], \qquad (33)$$

with k_0, k_1, and h being positive coefficients and with $\mathcal{H}$ being the Heaviside Step Function, estimated for analytical work by

$$\mathcal{H}[v_m(t) - v_c] = \frac{1}{1 + \exp{-\beta[v_m(t) - v_c]}}. \qquad (34)$$

Whenever $v_m(t)$ is less than or greater than v_c, $C(t)$ becomes C_R or C_E respectively. With this prescription, action potential-like behavior is obtained since the change in concentration of the ions in the cell, $\frac{dC(t)}{dt}$ is seen to be the membrane current, $I_m(t)$. The following equations

$$\frac{dv_m}{dt} + lv_m = \sum_{j=1}^{n} i_j(t) \qquad (35)$$

11-15

$$\frac{dC}{dt} + \{k_0 + k_1(1 + \exp-\beta[v_m(t) - v_c])^{-1}\}[C - C_E] + h[C - C_R] = 0 \tag{36}$$

$$I_m(t) = \frac{dC(t)}{dt} \tag{37}$$

define a system of equations resulting from the preceding set of assumptions. These may be manipulated further realizing that for various neurons: β, v_c, k_0, k_1, h, C_E, and C_R may vary statistically.

As can be seen, this approach is more profitable than Cowan's in that the foundation that had been developed heuristically, EQN(1), has now been made more rigorous. Our EQN(33) is not *ad hoc* and specifies more precisely the response of a realistic neuronal membrane. Now, Cowan's hypothesis of pulse summation at the neuron and logistic rate relation can be tested from more fundamental starting points.

2.2 - Future Directions

There remains the as of yet unexplained phenomena of "bursting" that demands to be addresssed. The all-or-nothing description of the emission of pulses is understood to apply to the emission of single pulses and is therefore an oversimplification of the true process. In reality, neurons emit pulses continually, and it is the frequency of pulse emission that increases significantly when the neuron becomes activated. It is this sequence of single pulses (or bursts) in quick succesion, or rather the *frequency* of these bursts that is the unit of interest [14].

To quantify this situation, we look to the discovery of Weber who noticed that the smallest perceptible difference in intensity, ΔI, is a constant, called the *Weber Constant*, $\frac{\Delta I}{I}$, equal to $\frac{1}{30}$. To this contribution, Fechner added a relationship between "stimulus" and "sensation" and derived the *Weber-Fechner law*

$$Sensation = K \log I + C \tag{38}$$

stating that discriminable units are increments of sensation. Thus, there is a logarithmic encoding of intensities being performed in cells [15]. An important distinction becomes notable when extending these models of tonic cells to include phasic cells, another important question for a future study. Whereas for tonic receptors this law, EQN(38), is a fair approximation, for phasic receptors it does not hold. The discharge is limited by prompt adaptation [16].

The inclusion of phasic cells in the statistical mechanical model is but one of the potential areas opened up to investigation by a more complete understanding of our work. Table 1 below, from *Neural Network Architectures* by Judith Dayhoff, lists several more places where further study is needed.

Biological Neural Networks	Artificial Neural Networks
Synapses complex	Synapses simple
Fixed gross wiring structure plus variation in detailed structure	Usually fully interconnected slabs
Pulse transmission	Activity value and connection strengths
Topological mappings	Kohonen feature map
Distributed representations and processing	Distributed representations and processing
Redundancy	Redundancy
Feature detectors	Feature detectors
Learning as fast as one pass	Slow to converge
100 billion neurons	Usually up to hundreds or thousands of neurons
Estimated 10,000 interconnections per neuron	Usually 10 – 10,000 interconnects per neuron
Continuous or asynchronous updating	Generally synchronous updating

Table 1: A Comparison of Biological and Artificial Neural Networks. (Source: Dayhoff, J.E. 1990, *Neural Network Architectures: An Introduction*, Van Nostrand Reinhold: New York, p. 158.)

This table collects a few of the differeneces between artificial and biological neural networks and provides great incentive for those attempting to increase the biological acuracy of current neural network models. The difficulty lies in retaining mathematical tractability with the increase in mathematical complex, but the reward is invaluable.

Another important direction to be pursued includes the explanation of various features of the active state of the brain, the active EEG. The *Visual* (VER) and *Auditory Evoked Responses* (AER) are phenomena that occur in reaction to external stimuli independent of the resting state and are prime candidates for necessitating the introduction of non-equilibrium statistical

mechanics. Both small and large perturbations from equilibrium could be explored, invoking *Kubo Theory, Fokker-Planck* formulations to indicate how the system responds to input and how it handles input-ouput relations. In short, the VER and AFR are problems in non-equilibrium statistical mechanics of neural nets [17].

2.3 - Applications to Health and Safety

In the endeavor to explain the resting EEG, the development of neural network models that are more biologically-accurate is crucial. If we can train these "biophysical neural networks," to correctly predict the electric fields coming from cells in the brain while it is in its resting state, then we will eventually be able to decipher more easily the electric field emissions from the brain in its active state. This knowledge will assist with medical diagnostic EEG analysis and should permit consideration of the effects of applied external fields on the resting and active EEG. Knowing with certainty the expected electric fields of the resting or active brain independent of external effects is vital in isolating the effects of external fields since the external fields may be simply superimposed over endrogenous membrane potentials.

The level of danger presented to human beings by low-level stray electric fields in the environment is unknown at present. Ultimately, showing and explaining the interaction of the brain's electric field with applied fields is our interest. To determine to what degree these artificial environmental fields effect damage to learning and the degradation of memory is an important occupational and environmental medical pursuit, one that our proposed work is designed to advance.

Conclusion:

Nearly 7 decades has elapsed since the introduction of the suggestion by Lotka (1925) to encorporate statistical mechanics rigorously into our theories of biological and neural modeling, and yet there remains a multitude of unpursued avenues and unanswered questions with regard to this pursuit. The work of Volterra (1931, 1937), Rashevsky (1938), Kerner (1957, 1959, 1961, 1964), and Cowan (1970) has laid the groundwork. We are in a postion to more fully answer questions about the human brain from an important perspective. It is our ultimate hope to thereby illuminate some of the mystery surrounding the risk to man posed by external factors in our environment.

References:

Balescu, R. 1975, *Equilibrium and Non-Equilibrium Statistical Mechanics*, Wiley Interscience Publication: New York.

Cowan, J.D. 1970, "A Statistical Mechanics of Nervous Activity," *Some Mathematical Questions in Biology in the Life Sciences*, **2**, pp. 1-57.

Dayhoff, J.E. 1990, *Neural Network Architectures: An Introduction*, Van Nostrand Reinhold: New York.

Kerner, E.H. 1957, "A Statistical Mechanics of Interacting Species," *Bulletin of Mathematical Biophysics*, **19**, pp. 121-146.

Kuffler, S.W., Nicholls, J.G., and Martin, A.R. 1984, *From Neuron to Brain*, Sinauer Associates Inc.: Massachusetts.

Lotka, A.J. 1925, *Elements of Physical Biology*, Williams and Wilkins Co.: Baltimore.

Penrose, R. 1989, *The Emporer's New Mind: Concerning Computer, Minds and the Laws of Physics*, Oxford University Press: New York.

Rashevsky, N. 1938, *Mathematical Biophysics: Physico-Mathematical Foundations of Biology*, Vols. 1 & 2, Dover Publications: New York.

Ruch, T.C., and Patton, H.D. 1965, *Physiology and Biophysics*, W.B. Saunders Company: Philadelphia.

Volterra, V. 1937, "Principes de Biologie Mathématique," *Acta Biotheoretica*, **III**, pp. 1-36.

Endnotes:

[1] Kuffler, et.al., pp. 7 & 12.

[2] Kuffler, et.al., p. 13.

[3] Cowan, p. 48.

[4] Ruch and Patton, pp. 2, 6 & 822.

[5] Kuffler, et.al., p. 17.

[6] Ruch and Patton, p. 29.

[7] Kerner, p. 121-2.

[8] Kerner, p. 121.

[9] Cowan, p. 6.

[10] Cowan, p. 20.

[11] Kerner, p. 121.

[12] Cowan, p. 7.

[13] Kuffler, et.al., p. 17.

[14] Penrose, p. 395-396.

[15] Ruch and Patton, p. 338.

[16] Ruch and Patton, p. 106 & 109.

[17] Balescu.

ASSESSING THE IMPLEMENTATION OF A
MULTIVARIATE RISK SCREENING PROCEDURE
IN THE DIAGNOSING OF DOWN SYNDROME ASSOCIATED PREGNANCY (DSAP)

William R. Fletcher, Jr.
Graduate Student
Department of Mathematics

Howard University
2400 6th St. N.W.
Washington, D.C. 20059

ASSESSING THE IMPLEMENTATION OF A
MULTIVARIATE RISK SCREENING PROCEDURE
IN THE DIAGNOSING OF DOWN SYNDROME ASSOCIATED PREGNANCY (DSAP)

William R. Fletcher, Jr.
Graduate Student
Department of Mathematics
Howard University

Abstract

The mathematical basis of the multivariate risk screening procedure is that, though various nonrelated data originate from the same source and add no more to the modification of the established odds than a single datum from the same source, these multiple independent parameters do, however, improve success. Currently, a quantitative biochemical Alpha-fetoprotein procedure is used. However, according to the Foundation for Blood Research (FBR), it is far less accurate than the proposed AFP diagnostic profile procedure, which not only tests quantities of Maternal Serum Alpha-Fetoprotein (MSAFP), but also, unconjugated Estriol (uE_3), as well as Human Chronic Gonadotropin (HCG). On the other hand, the new procedure has not been approved by the ACOG (American College of Obstetrics and Gynecologists), even though the FBR asserts that the new profile test is three times more accurate in predicting Down Syndrome.

My mission at Brooks Air Force Base, San Antonio, Texas, in the Epidemiologic Research Division, was to reinterpret the multivariate algorithm submitted by the Foundation for Blood Research (FBR) so that the new screening procedure could be implemented into Armstrong's existing laboratory data base.

ASSESSING THE IMPLEMENTATION OF A
MULTIVARIATE RISK SCREENING PROCEDURE
IN THE DIAGNOSING OF DOWN SYNDROME ASSOCIATED PREGNANCY (DSAP)

WILLIAM R. FLETCHER, JR.

INTRODUCTION:

Down Syndrome (DS) is a variant of nature. Defined by the Encyclopedia Britannica, "Down Syndrome was the first autosomal abnormality to be described in man. All persons afflicted with this disorder have all (or most) of the set of chromosomes conventionally labeled 21 represented triply rather than doubly, so that there are 47, rather than the normal 46, chromosomes in most cells." (See Fig. 1).

Fig. 1

Karyotype of a patient with Down's syndrome, or Mongoloid idiocy. The extra chromosome is usually assigned to position 21.

"As the most common autosomal aneuploidy (incidence at birth - 1.3 per 1,000) Down Syndrome has a significant survival rate beyond infancy. The disorder has been identified in all races (the term mongolism was used because the up-slanted eyes and epicanthal folds gave the appearance of an Oriental face) and usually occurs among the firstborn of women over 40 years of age. More than half of the affected children die during their first year, usually from heart defects or from infection. Life span, if the infant survives the first year, may be relatively normal, although sinus and lung infections are a recurrent hazard. Severe mental retardation requires special training, and individuals so affected do not ordinarily become self-supporting; they are, nevertheless, unusually sociable and affectionate." (1)

"Prenatal diagnosis is made by examination of fetal or placental chromosomes from sampling amniotic fluid (amniocentesis), from biopsy of the placenta [chorionic villi sampling (CVS)], and, less frequently, umbilical cord blood sampling. CVS is usually performed between 8 weeks and 12 weeks gestational age (G.A.); amniocentesis, $\geq$ 12 weeks G.A. Complications (e.g., fetal death) can occur from these invasive procedures; however, risk is relatively low in the U.S. (e.g., number of fetal deaths from amniocentesis - approximately 0.2% to 0.4%). Nevertheless, these procedures are performed when the risk of Down Syndrome or other abnormality is sufficiently high to justify them. Definition of sufficiently high risk is subjective and in 1983, the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Sciences recommended that genetic amniocentesis be made available to all pregnant women. In practice, due to cost (approximately $1,000), labor-intensity, and technical (e.g., shortage of cytotechnologists) constraints, most physicians recommend amniocentesis for those women and DS risk at term $\geq$ 1:365 [1:386; the risk at maternal age (M.A.) = 35 years], or 2nd trimester risk $\geq$ 1:270 (1:274). Thus, non-invasive, biochemical screening tests [e.g., maternal serum alpha-fetoprotein (MSAFP), human chorionic gonadotropin (hCG), unconjugated estriol (uE3), pregnancy-specific B1-glycoprotein (SP1)] that can provide better assessment of Down Syndrome risk than M.A. alone are desirable by the medical community." (2)

"The AFP Profile or alpha-fetoprotein-triple marker procedure (AFP-TMP) is a group of tests, performed on a single maternal serum sample, which will screen for open fetal defects, such as spina bifida, anencephaly, and abdominal wall defects, and also for fetal trisomy 21 (Down Syndrome). Along with maternal serum alpha-fetoprotein (MSAFP), two additional biochemical markers, unconjugated estriol (uE3) and human chorionic gonadotropin (hCG), are measured to calculate an increase or decrease in the patient's age-related risk for Down Syndrome. Compared to using age 35 alone as an indication of risk, or low MSAFP and age alone in the under 35 population, the AFP Profile can triple the number of cases of fetal Down Syndrome detected (from about 20% to 60%) following amniocentesis on about 5% of screened women.

The AFP-TMP is just as reliable as routine MSAFP screening for identifying open neural tube defects; the MSAFP level is still used by itself to determine risk of open fetal defects, and the protocols and recommendations associated with elevated MSAFP and open fetal defects are unchanged. The additional markers - uE3 and hCG - significantly enhance the sensitivity of Down Syndrome screening." (3)

As reported by Dr J. Canick, "with the discovery that pregnancies affected with Down Syndrome are associated with lower levels of MSAFP, we postulated that the levels of other biochemical markers of fetoplacental origin might also be abnormal in Down Syndrome pregnancy. Alpha-fetoprotein (AFP) is made primarily in the fetal liver during the second trimester of pregnancy and indirect data indicated that the livers of fetuses affected with Down Syndrome might be producing less than normal amounts of AFP. The fetal liver also participates in the synthesis of the fetoplacental steroid hormone, estriol.

Therefore, maternal serum unconjugated estriol (uE3) might also be lower than normal in pregnancies affected with fetal Down Syndrome and might also provide useful clinical information for prenatal screening. A research collaboration between Women & Infants Hospital, the Foundation for Blood Research, and the Medical College of St. Bartholomew's Hospital showed that maternal serum uE3 is, in fact, low in Down Syndrome pregnancy and that the measurement of uE3 along with AFP substantially enhances screening sensitivity." (4)

"Researchers at the University of California, San Diego, demonstrated at about the same time that maternal serum levels of the placental hormone, human chorionic gonadotropin (HCG), were also abnormal in pregnancies affected with fetal Down Syndrome (3). Interestingly, HCG levels were higher than normal in such cases. Immediately following this discovery, we were able to demonstrate that the measurement of maternal serum HCG in conjunction with uE3 and AFP could better define the patient-specific risk for fetal Down syndrome than any method previously described." (5)

DISCUSSION OF PROBLEM

The problem at Brooks Air Force Base's Epidemiologic Research Division (AOE), is also that of variance. Requests for the AFP triple-marker system has been increasing significantly, and AOE's Clinical Chemistry Laboratory has been considering a change to the proposed system for at least 6 months. On the other hand, because of a shortage of manpower and managerial transition (NCOIC, as well as OIC) very little time could be allotted to do the necessary research required. Moreover, AOE had no existing system that would actively involve the OB/GYN physicians from Wilford Hall USAF Medical Center in the implementation of the triple-marker AFP profile. Yet, having the reputation of being a state-of-the-art reference laboratory, AOE could ill afford to overlook the claims of various biomedical scientists, that the AFP triple-marker is 50-300 percent more accurate in predicting Down Syndrome than the current Maternal Age (MA)-MSAFP diagnostic system.

Cost is another consideration. AOE is on a "tight" budget. Thus, perhaps, the cost of such a system might be prohibitive. And, feasibility, is the instrumentation readily available and could they possibly write the

software for the triple-marker system, in house, utilizing a mathematician? Finally, what variances would be considered in the system redesign?

METHODOLOGY

The variances in this case are caused by AOE's socio-technical environment. (See Fig. 2). Indeed, the emergence of AOE's desire to meet the demand for the triple-marker profile is a typical case of change being brought about from the outside.

Fig. 2

THE SOCIO-TECHNICAL ENVIRONMENT
*(IS CONSTANTLY CHANGING)

STAKEHOLDER:

Military Physicians
(Army, Navy, Air Force)

STAKEHOLDER:

Military Personnel
and/or their dependents

RISK TAKER:

AOE
Brooks AFB

STAKEHOLDER:

Lackland OB/GYN
Medical Group
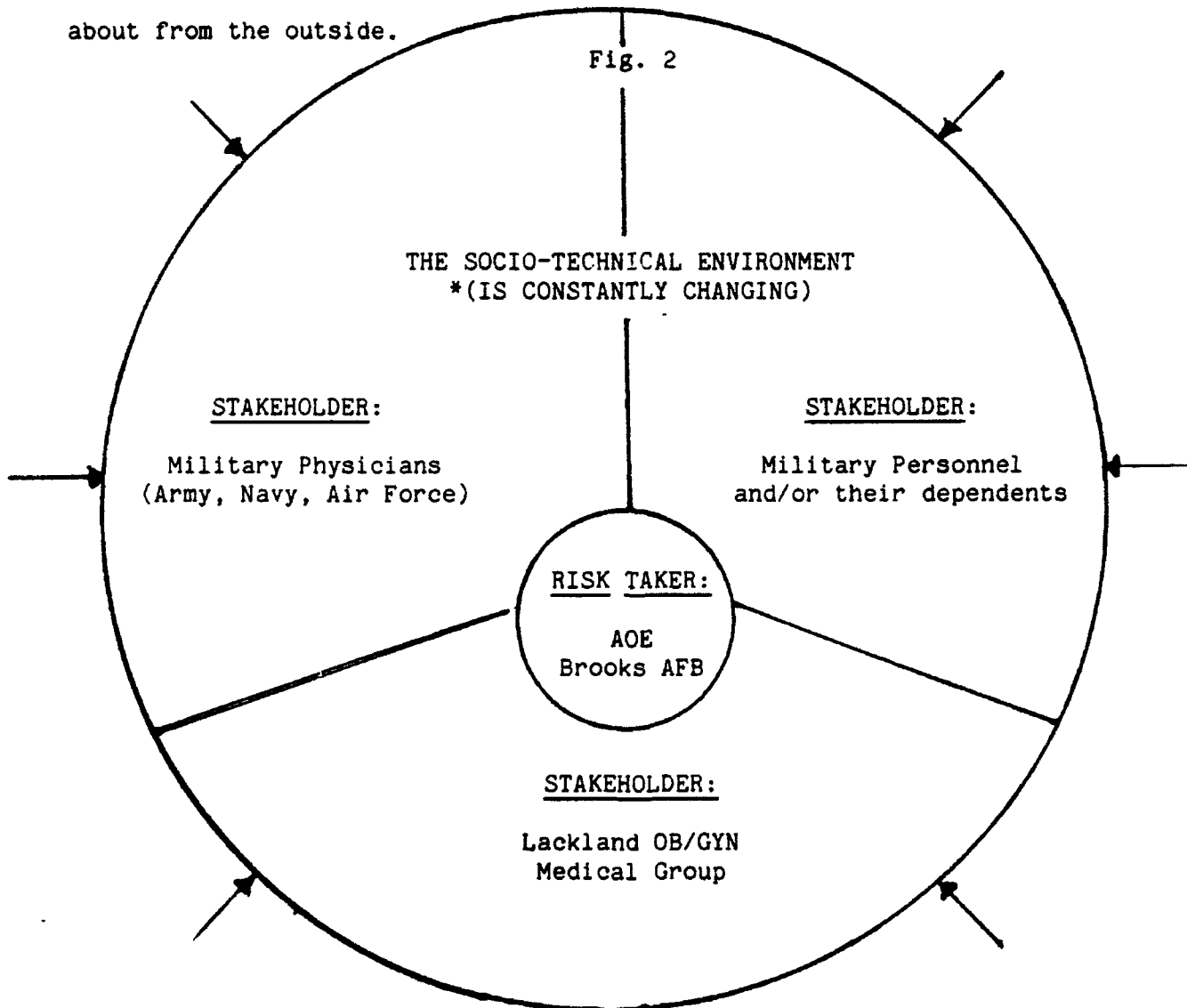
<u>Analysis of the sociological forces bringing about AOE's need to change its</u>

<u>current MSAFP system:</u>

* More military and/or professional women are having children at a later age.

* Stiffer abortion laws have increased the patients need-to-know (as to whether or not their baby will have DS, and, if so, when and where should they have an abortion) - if they make such a choice.

* As a result, in order to escape legal entanglement, more military doctors are requesting the triple-marker AFP profile for DS.

* Like any other top international, U.S.-based corporation, the USAF depends heavily on "leading edge technology." Consequently, in keeping with the USAF "image," AOE must continue to show that, like its civilian counterparts, it too can provide state-of-the-art medical information, quickly, efficiently, but at a lower economic cost.

* Indeed, the Japanese have shown that, in order to be a competitive "world-player," organizational group participation is a must. AOE seems to be courting this concept in its effort to involve the Wilford Hall OB/GYN medical group with the triple marker program. By involving the physicians, as well as, possibly, getting their business, AOE will further entrench their existence with the "new" Air Force.

<u>Analysis of the technical forces bring about AOE's need to change its current</u>

<u>MSAFP system:</u>

* Until biochemical screening became available, the most frequent method for selecting women at increased risk for DS was maternal age.

* In 1984, AFP was reported to be lower than average in maternal serum associated with DS pregnancy. The discovery was found to be independent of maternal age.

* In 1987, Bogart and Company at the University of California, San Diego, reported an association between abnormal human chorionic gonadotropin (hCG) levels and pregnancies with fetal aneuploidy.

* "Furthermore, in 1987, it was discovered that the fetal liver also participates in the synthesis of the fetoplacental steroid hormone, estriol (uE3). Therefore, maternal serum uE3 might also be lower than average in pregnancies affected with fetal Down Syndrome and might also provide useful clinical information for prenatal screening. In collaboration with the Foundation for Blood Research in Scarborough, Maine, and the Medical College of St. Bartholomew's Hospital in London, the WIH group showed that maternal serum uE3 is, in fact, low and that the measurement of uE3 along with AFP substantially enhances screening sensitivity." (6)

* "Again in 1987, several groups described methods of screening pregnant women for the risk of carrying a fetus with Down Syndrome; the screens utilized maternal age at expected date of delivery (EDD) and the measurement of maternal serum a-fetoprotein (mAFP). Cuckle et al. calculated that applying such a screening procedure, with all pregnancies having a risk greater than 1:250 receiving amniocentesis, would correctly detect 28% of affected pregnancies with 2.8% of non-affected pregnancies being detected as false positives. Although only a small proportion of Down's associated pregnancies could be detected with an acceptable false positive rate, this did represent an improvement over methods of screening not including maternal age within the risk calculation." (7)

* "In 1988 Wald published a further refinement of the method of screening using maternal age and three analytes: mAFP, human chorionic gonadotrophin (HCG) and serum unconjugated oestriol (uE_3). By selecting decision limits such that all women with a risk greater than 1:250 would be offered amniocentesis. It was demonstrated that a direction in excess of 60% of Down's associated pregnancies was possible whilst maintaining an overall amniocentesis rate of about 5%, a rate no higher than that currently achieved by centers screening for Down Syndrome solely on the basis of maternal age 36 years or over." (8)

RESULTS:

As a result, AOE had to become a "risk-taker." Sociological and technological advances have brought about the emergence of a AFP-Triple Marker Profile (AFP-TMP). With the notoriety of being an "advanced" reference laboratory, AOE could not afford to overlook a procedure that is, experimentally, 50-300% more accurate in predicting Down Syndrome Associated Pregnancies (DSAP) than the current Maternal Age (MA)-MSAFP diagnostic system. Thus, initial understanding of the mathematics behind the (AFP-TMP) was necessary.

DEFINITIONS

Risk ratio (or odds ratio): A ratio of 1:n implies that for n trials there is 1 chance of a specified event occurring.

Odds modifier (or likelihood ratio): This is a factor which is used to modify a risk ratio. For example if a risk ratio of 1:20 is modified by an odds modifier of 3 the modified risk ratio is 1:60.

Increase in risk: Increased risk means that there is a greater probability that an event will occur. For example a risk of 1:200 is greater than a risk of 1:300.

Decrease in risk: From the above a risk of 1:300 is less than a risk of 1:200.

Fig. 3

12-8

Matrix mathematics

What is a matrix?

A matrix is a group of numbers in a specific format which interacts with other such sets of numbers in a very specific way. Matrices are formed from a number of rows and columns. For a multivariate Gaussian distribution the core of the formula is the matrix $\underline{V}$ which has the same number of rows as columns.

Matrix notation:

$\underline{X}$: The line beneath the variable name signifies that it is in matrix format. $\underline{X}$ is the matrix containing all the patient's parameters. The matrix $\underline{\mu}$ contains the population means.

$\underline{V}^{-1}$: This is the inverse of the matrix $\underline{V}$. This concept will be explained later.

$|\underline{V}^{-1}|$: This is the determinant of a matrix. This will be explained later.

"The risk of Down Syndrome in relation to maternal age at EDD has been calculated by Cuckle. This risk is expressed as an odds ratio. Information concerning the progress of the pregnancy independent of maternal age is then used to modify this ratio. Providing the Gaussian distributions for a given parameter are known for both unaffected and Down Syndrome pregnancies, then for a given parameter value the chance of a Down Syndrome associated pregnancy is the ratio of the heights of the Gaussian distribution for Down Syndrome pregnancies and unaffected pregnancies, respectively: this has been termed the likelihood ratio by Wald. It may also be called the odds modifier; i.e., calculation of age related risk:

The formula for calculation of maternal age related risk is:

$$Risk = \frac{0.999373 \div exp[(-16.2395 + 0.286 \cdot Maternal\ Age)]}{0.000627 \div exp[(-16.2395 + 0.286 \cdot Maternal\ Age)]}$$

This formula gives a value for n such that Risk = 1:n.

Data required for screening test: parameters used by Wald et al. were Log_{10} of the multiple of the median for mAFP, HCG and the multiple of the median for uE3. Depending on the number of parameters, the appropriate Gaussian distribution function should be used. For one or two parameters the functions may be entered directly. For three parameters there is some precalculation required. The matrix $\underline{V}$, its inverse and the determinant of the inverse must be evaluated for both normal and Down's associated pregnancies. They need only be calculated once as they may then be stored in the computer as constants. The $\underline{V}$ matrix is composed of the population standard deviation values for each parameter and the correlation coefficients between each parameter." (9)

$$\underline{V} = \begin{pmatrix} \sigma_{x_1}^{2} & \rho_{x_1 x_2}\sigma_{x_1}\sigma_{x_2} & \rho_{x_1 x_3}\sigma_{x_1}\sigma_{x_3} \\ \rho_{x_1 x_2}\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^{2} & \rho_{x_2 x_3}\sigma_{x_2}\sigma_{x_3} \\ \rho_{x_1 x_3}\sigma_{x_1}\sigma_{x_3} & \rho_{x_2 x_3}\sigma_{x_2}\sigma_{x_3} & \sigma_{x_3}^{2} \end{pmatrix}$$

"For example the second element of the top row of the illustrated matrix is the product of the correlation coefficient relating parameters 1 and 2 with the standard deviations of parameter 1 and 2.

The patients' data are entered into the normal distribution function twice. Once using population data derived from Down's associated pregnancies, and once using the data from normal pregnancies. These two values f_{downs} and f_{normal} are used to calculate the odds modifier

$$= \frac{f_{normal}}{f_{downs}}$$

The final risk is the product of the age related risk and the odds modifier.
Thus in the case of screening for Down's associated pregnancy the risk would
be:

$$\text{Risk} = \text{Age-related risk} \times \frac{f_{\text{normal}}}{f_{\text{downs}}} \qquad (10)$$

"Whereas the estimation of racing odds is based on subjective qualitative
information and the computations are intuitive, based on experience, the
numerical data obtained from biochemical measurement, or other numeric
parameters of pregnancy influenced by Down Syndrome, can be processed
mathematically.

The Gaussian distribution of a single variable is visually presentable (in two
dimensions), and the related Gaussian distributions of two variables are also
presentable in three dimensions: the introduction of more variables results
in conceptual difficulty, which can be overcome by mathematical analysis using
matrix algebra and may be easily programmed in BASIC on any commercially
available personal computer.

The method of risk calculation is analogous to setting betting odds for horses
in a race. Established odds (equivalent to odds ratio) are based on past form
(equivalent to maternal age) but can be modified by further independent
information such as an observation of current form (equivalent to mAFP).
Consistent success in picking a winner requires a reliable source of
independent items of information (equivalent to HCG, uE_3). As with the horse
racing analogy many items of apparently independent information, but from the
same source add no more to the modification of the established odds than a
single item of information from that source. However, multip e sources of
truly independent information (especially if not available to the bookmaker)
improve success." (11)

On the other hand, to successfully accommodate the variations of its socio-
technical environment, an organizational redesign of AOE's DSAP procedure
appears necessary. AOE's basic mission is the providing of medical
information to the user. Yet, there are social issues attached to DSAP that
AOE must fully discuss, sooner or later, with its users- the military
physician group, especially with regard to the patient's need-to-know.
Indeed, the current MA-MSAFP system doesn't seem to account for the increasing
social concern surrounding DSAP and abortion laws. This increase in the
concern over DSAP by older professional civilian and/or military women and the

ability to have abortions, could translate into a need for more accurate DSAP testing. Already, more than 30 commercial laboratories are utilizing the AFP-TMP on a daily basis. On the other hand, though it is 50 percent more sensitive than the current MA-MSAFP procedure, the AFP-TMP still misses a significant number of DSAP's (approximately 40 percent). But then, AFP-TMP is an indirect biochemical procedure. Since DS is an autosomal aneuploidy, there will probably evolve an alternate laboratory procedure involving chromosomal genotyping and/or phenotyping of nucleated fetal cells with approximately 94-100 percent accuracy. Therefore, AOE, in order to successfully accommodate the variations of its socio-technical environment, must have an organizational redesign - with respect to DSAP - that allows for socio-technical change. By doing so, AOE will avoid the trauma and the inefficiency of crisis management, as well as maintain its reputation as a state-of-the-art facility.

One way to adjust to the socio-technical "change process" is the creation of a "Provider/User" steering committee (see Fig. 4). This committee should meet periodically to:

1. Discuss latest events in AFP testing.

2. Insure the integrity of the Air Force mission - the well being of Air Force personnel and their dependents by a dedicated commitment to quality with respect to patient care. This quality can be obtained by (a) setting "idealized goals," i.e., 100 percent forecasting for Down Syndrome and (b) perhaps inviting participation of enlisted staff by allowing them to attend and interact at "provider/user" steering committee meetings via offering suggestions.

3. Design a pilot program by which the AFP triple marker profile will be quantitatively tested and assessed.

4. Oversee the interfacing of the commercial software with the Laboratory Information Management System (LIMS), as well as how the final LIMS report will be reported.

5. Assess the data produced from the implementation process (as to how efficient).

6. Analyze new techniques and new innovations.

Fig. 4

DSAP PROCEDURE REDESIGN:

USER:
MILITARY PHYSICIANS

PROVIDER:
AL/AOELC

USER/PROVIDER
COMMITTEE

AL/AOELC
DSAP MISSION

TECHNICAL:
Maintain and/or
update computer
hardware, computer
software, and
AFP-TMP equipment
in order to
accomplish mission

SOCIAL:
Update and/or
establish-new
procedures for
personnel training
and utilization

MANAGERIAL:
Set idealized goals-
i.e., 100% accuracy;
no risk factor.
Evaluate new
procedures that will
enhance speed and
accuracy of any lab
process concerning
DSAP, i.e., AFP-TMP.
Maintain good
relations with the
primary users
(military physicians)

CONCLUSION:

We found three existing commercial software packages which perform the AFP-TMP and interface with AOE's data base. Consequently, the mathematical basis of multivariate risk screening was discussed in order to reaffirm, theoretically, that the AFP-TMP is more sensitive to DSAP than the current MA-AFP procedure (12). However, because of ever changing technical advances (i.e., flow cytometry, DNA probes, etc.) and, because of various social issues involved with the AFP-TMP (i.e., abortion, ACOG not yet approving the procedure [AFP-TMP] as a diagnostic standard, etc.), an Organizational Redesign, with respect to DSAP, was also submitted. This Organizational Redesign defines and procedurally establishes a pilot program with "idealized" managerial goals and expectations that coincide and/or contain the Air Force mission. Thereby, providing a format which not only can quantitatively test the claims of such research and development groups as FBR but also eases the transition from an experimental stage into a daily procedural process should any future diagnostic screening such as AFP-TMP for Down Syndrome be deemed acceptable and feasible by the "provider," (Armstrong Laboratory management and technical staff within the Chemistry Laboratory at Brooks AFB), as well as the "user," (military physicians from the Army, Navy, and Air Force bases worldwide, and/or the various medical groups affiliated directly or indirectly with Brooks AFB, [i.e., ACOG]). Thus, a greater commitment to quality of service is achieved when the testing and implementation processes are more defined and better understood by all involved. As a result, the following AOE proposal was written for the 18 June 1992 Wilford Hall USAF Medical Center MSAFP Committee. Please note the objective.

PROPOSAL FOR

WILFORD HALL USAF MEDICAL CENTER

(MSAFP COMMITTEE)

ARMSTRONG LABORATORY
EPIDEMIOLOGIC RESEARCH DIVISION
CLINICAL CHEMISTRY LABORATORY
BROOKS AFB TX 78235-5000

12-14

OBJECTIVE:

Review and consultation by the MSAFP Committee at Wilford Hall USAF Medical Center regarding the implementation of a multivariate risk screening procedure in the assessment of Down Syndrome risk.

INTRODUCTION:

The AFP triple-marker profile is a group of tests, performed on a single maternal serum sample, which will screen for open fetal defects, such as spina bifida, anencephaly, and abdominal wall defects, and also for fetal trisomy 21 (Down Syndrome). Along with maternal serum alpha-fetoprotein (MSAFP), two additional biochemical markers, unconjugated estriol (uE3) and human chorionic gonadotropin (hCG), are measured to calculate an increase or decrease in the patient's age-related risk for carrying a Down Syndrome-affected fetus. (Current ACOG guidelines require low MSAFP and age alone in the under 35 population) (1).

DESIRABILITY:

There is an increasing demand for use of the triple marker from the field.

The Epidemiologic Research Division (AOE) can't satisfy the demand for this testing. (Consequently, private commercial labs are being used, costing the Air Force extra monies).

The AFP Profile (triple marker profile) is just as reliable as routine MSAFP screening for identifying open neural tube defects: the MSAFP level is still used by itself to determine risk of open fetal defects, and the protocols and recommendations associated with elevated MSAFP and open fetal defects are unchanged. However, the additional markers - uE3 and hCG - significantly enhance the sensitivity of Down Syndrome screening. (See Table 1) (Ref 2).

Compared to using age 35 alone as an indication of risk, or low MSAFP and age alone in the under 35 population, the AFP Profile can triple the number of cases of fetal Down Syndrome detected (from about 20% to 60%) (Ref 1).

COST ANALYSIS:

	Cost/Pt.[1]	Cost/Month	Cost/Annum
Brooks[2]	$15.20	$15,200	$182,400
Commercial Lab (i.e., FBR)[3]	37.20	37,200	446,400

NOTES:

[1] Based on a projection of 1,000 patients/month

[2] AOE (maximum cost/patient - includes commercial software package costs amortized over a 1 year period): $15.20

[3] Commercial Lab, i.e., Foundation for Blood Research (FBR). July 1, 1991 List Price $62.00; Wholesale Price: $37.20

RESOURCE REQUIREMENTS:

Instrumentation - hCG: Stratus II; MSAFP: Hybritech QA; uE3: Amersham Amerlex - MSp System (Magnetic Isotopic I^{125}).

    Software: "AFP SMS" from Maciel Assoc.

    Hardware: 2 magnetic racks for the Amersham
                 1 PC that is hard-wired to LIMS. (Ref 3)

    Personnel Requirements (earned FTE's based on CAP Work Load Data and a manning factor of 7,730 CAP units/FTE/mo): 2.6

    CAP Points: 4.60 min/test MSAFP; 4.60 min/test hCG; 6.0 min/test uE3.

Budget: (Supply costs/month) @ 1,300 tests/mo (including repeats, calibrations, controls and dilutions)

MSAFP Kits	$2,795.00
hCG Kits	3,065.83
uE3 Kits	2,080.00
Cups	90.44
Tubes	77.82
Controls	225.00
	$8,334.09/mo. for 1,300 tests

SUMMARY/CONCLUSION:

With triple-marker screening, maternal age is no longer the prime determinant of risk but just one of four such variables. In fact, maternal age and AFF are the two weakest contributors to the Down Syndrome risk calculation. Thus, all pregnant patients, regardless of age, can be considered candidates for screening because it is just as valid to calculate the risk of an older woman downward as that of a younger woman upward.

Should the triple marker profile for Down Syndrome become standard procedure for prenatal testing, AOE will be ready to offer service quickly and efficiently.

REFERENCES (For the Wilford Hall MSAFP Committee Proposal)

1. AFP Office Update, Vol 4, No. 1, 1990, Pg 1

2. Contemporary OB/GYN, Vol 36, April 15, 1992

3. LIMS - Laboratory Informational Management System

TABLE 1
Combining markers

Case	AFP MoM	uE3 MoM	hCG MoM
1	0.32	0.82	1.15
2	0.60	0.61	1.44
3	0.91	·0.54	1.77
4	1.22	0.43	1.93

In the four cases shown, all women were 30 years of age and, despite disparate marker values, had an identical risk after screening of 1 in 190. The table illustrates the point that in triple-marker screening it is necessary to use all three assay values, regardless of concentration.

AFP—a-fetoprotein, hCG—human chorionic gonadotropin, MoM—multiple of the median, uE3—unconjugated estriol

(Results of the 18 June 1992 MSAFP Committee Meeting)

1. AOE will now do Wilford Hall's MSAFP's.

2. There will be regular periodic meetings with Wilford Hall's physician group and AOE (in particular, AOELC), also:

3. Follow-up communications - with regard to various patients, as well as, (for the first time):

4. Alternate meetings to be held at AOE.

5. Though the implementation of AFP-TMP was not supported by Wilford Hall's MSAFP Committee (at this time) the "need-to-know," with respect to the patient was discussed, as was:

    a. A future prospective DSAP procedure, introduced by Dr Barth, involving genotyping with DNA probes as being more favorable than AFP-TMP, (100% accuracy; no risk; this procedure to be developed within 3 to 6 years) (13).

    b. If, on the other hand, AFP-TMP is approved by ACOG, Wilford Hall's medical group (OB/GYN) has assured AOE continued informational ties.

6. On the next day (after the MSAFP Committee meeting) AOE conducted a nationwide Air Force telephone survey. Total number of Air Force clinical labs called: 62

<u>Questions Asked</u>:

    a. Have your OB/GYN clinicians asked about the availability of triple marker testing?

9 laboratories - yes:	53 laboratories - no
*Andrews AFB, MD	
Griffiss AFB, NY	
Hill AFB, UT	
Homestead AFB, FL	
*Travis AFB, CA	
Williams AFB, AZ	
Wright-Patterson AFB, OH	
*Keesler AFB, MS	*= Were going to establish
Wurtsmith AFB, MI	AFP-TMP in-house.

    b. Do you refer any requests for triple-marker screening for Down Syndrome Risk Assessment to a commercial or other reference lab?

6 laboratories - yes:	56 laboratories - no
Hill	
Homestead	
Keesler	
x Williams (closing)	x= OB/GYN clinicians were in
Griffiss	the process of being
x Wurtsmith (closing)	transferred

c. If so, approximately how many per month?

Hill	100: hCG, uE3
Homestead	3-4: AFP-TMP
Keesler	80: AFP
Williams	5: AFP-TMP
Griffiss	6-2: AFP-TMP
Wurtsmith	14-25: hCG; 1-5:uE3

d. Cost of Panel?

Hill	$42-uE3; $53-hCG
Homestead	$62-AFP-TMP
Keesler	$17-AFP
Williams (Closing)	?
Griffiss	$60-AFP-TMP
Wurtsmith (Closing)	$23.50-hCG

REFERENCES

1. Encyclopedia Britannica, Vol. 4, pg. 204.

2. American Association for Clinical Chemistry; 43rd National Meeting, July 28, 1991; Lt. Col. Wians, PhD; pg. 2.

3. Alpha Feto Protein (AFP) Office Update; Vol. 4, No. 1, 1990; pg. 1.

4. Screening for Down Syndrome Using Maternal Serum Alpha-Feto-Protein, Unconjugated Estriol, and hCG: Journal of Clinical Immunoassay; Volume 13, No. 1, Spring 1990; Jacob A. Canick, PhD.

5. ibid

6. Alpha Feto Protein (AFP) Office Update; Vol. 4, No. 1, 1990; pg. 2.

7. Ann Clin Biochem 1989; 27:452-457. T.M. Reynolds and M.D. Perry.

8. ibid; pg. 452

9. ibid; pg. 457

10. ibid; pg. 453

11. ibid; pg. 453

12. M.A.: Maternal Age; AFP: Alpha-Feto-Protein.

13. Trisomy 21: Association Between Reduced Recombination and Nondisjunction: Am. J. Hum. Genet., 1991, Sep; 49(3):608-20; Sherman, S.L.

THIS PAGE INTENTIONALLY LEFT BLANK

PROBABILITY AND CUE TYPE MANIPULATIONS IN A VISUAL ATTENTION TASK

Lawrence R. Gottlob
Department of Psychology

Arizona State University
Tempe, AZ 85287

Final Report for
Summer Research Program
Armstrong Laboratory

PROBABILITY AND CUE TYPE MANIPULATIONS
IN A VISUAL ATTENTION TASK

Lawrence R. Gottlob
Department of Psychology
Arizona State University

Abstract

In a previous study, it was found that observers could allocate attention to four locations in a display based on the probability that a peripheral (abrupt-onset) cue was valid. The first experiment presented in this report sought to investigate the effects of a central cue, and the second experiment compared central to peripheral cues.

Observers in both experiments were run in visual attention tasks. The target was a T-like character in different orientations, that was presented in one of four different locations (subtending a total of 12 degrees visual angle) on a computer screen. The target was preceded by a central (Experiments 1 and 2) or a peripheral (Experiment 2) cue, and followed by a mask. The probability that the cue was valid was varied over blocks of trials. The observers' task was to indicate the direction that the "T" was pointing. Proportion correct scores indicated that observers allocated more attention to more probable locations (across valid trials only) with a peripheral cue, but that there was no probability effect with a central cue.

PROBABILITY AND CUE TYPE MANIPULATIONS
IN A VISUAL ATTENTION TASK

Lawrence R. Gottlob

INTRODUCTION

In the area of visual attention, the nature of responses to both abrupt-onset peripheral cues and central cues have been studied. An abrupt-onset flash on a display has been found to initiate an automatic or transient response, while central cues initiate a purposive and sustained response (Jonides, 1981, Tepin & Dark, 1991, Nakayama & Mackeben, 1989, Posner & Cohen, 1984, Cheal & Lyon, 1991).

In a previous study, Gottlob (1992) used a peripheral cue, and found that observers were able to allocate different mixtures of a fixed quantity of attention over four locations of a display. The allocation was affected by cue validity expressed as the proportion of correctly cued trials. By allocating according to cue probability, observers were maximizing their chances for correct target discrimination. Even though the task used a peripheral cue, the study showed that observers did have control over a putatively automatic response.

This control was seen to have come from a top-down source. When cue validity is manipulated between blocks of trials, individual trials are indistinguishable between conditions; both valid and invalid trials are the same across

conditions. What varies is the mixture of trials across conditions. If a performance difference is seen across probability conditions, that difference must be necessarily due to top-down influences since the bottom-up information is identical across conditions.

The task used in Gottlob (1992) and in the present report is a location cuing task similar to that used by Cheal and Lyon (1989, 1991), and also Cheal, Lyon and Hubbard (1991). The location-cuing paradigm uses a cue to direct attention to one of four locations equidistant from a fixation point, followed by a target stimulus and a mask, whereupon the observer attempts to identify the target at the cued location. (See Figure 1 for a schematic order of events).

The dependent variable of accuracy is measured at different cue-target stimulus-onset asynchronies (SOA); for targets such as the "T" pictured in Figure 1, accuracy has been found to increase as a function of SOA until it begins to level off (approach asymptote) at 100 to 150 msec (Cheal and Lyon, 1989). The SOA by accuracy curve enables one to follow the timecourse of the build-up of attention at the cued location. Experiments have shown that, in order to discriminate certain types of targets, attention takes about 100 msec to "concentrate on" a cued location, (Cheal & Lyon, 1989).

The increase in accuracy as a function of SOA was obtained in trials where the cue was valid, and attention had time to be oriented to the cued location. In addition, Cheal et al., 1991, included an invalid cue condition in which the target appeared at a noncued location. Cheal et al. (1991) found lower accuracy for invalid trials than for valid trials, and the accuracy by SOA function for invalid trials was flat, unlike the rising SOA function for valid trials. Valid and invalid accuracy curves had different shapes as a function of SOA.

The performance difference between valid and invalid trials is due to the orienting effect of the cue. The addition of the factor of cue probability allows us to isolate top-down processes to determine whether top-down control can be exerted over attention in response to a peripheral cue. As mentioned above, with manipulation of cue validity probability, there is no difference in bottom-up information, so any changes in performance must be due to top-down influences.

The main finding of Gottlob (1992) is illustrated in Figure 2, which illustrates the growth in accuracy (proportion correct) as a function of cue-target SOA, across valid and invalid trials, and 75% and 50% cue probabilities. As was predicted, the 75% valid curve lies above the 50% valid curve, and the 75% invalid curve lies below the 50%

invalid curve. The results were in accordance with optimal search theory (Koopman, 1957), which prescribe that attention should be allocated with regard to a probability distribution of target location.

Experiment 1 in this report was designed to replicate the probability effect with a central cue instead of a peripheral cue. It was expected that the controlled nature of the central cue response would magnify the probability effect.

METHOD

Observers

Two male and one female observer between the ages of 19 and 22, with normal vision, were paid $7.00 per hour to participate. None were informed of the purpose of the study. Each observer practiced for a number of sessions before beginning the experimental trials. One observer was experienced in this type of experiment; two were not. Each was run in 24 one-hour sessions, for a total of 30,720 trials each.

Apparatus

Stimuli were displayed on an IBM-XT with an EGA color monitor. An adjustable head and chin rest fixed the eye-to-screen distance at approximately 37 cm. Eye movement was monitored with a video camera to ensure that the observers fixated their gaze at a single location, and that observers

maintained the proper head position. Responses were recorded
on the numeric keypad of a standard IBM keyboard.

Stimuli

Stimuli were presented as white pixels on a dark gray
background, with a total luminance of 80 cd/m^2. Targets
consisted of "T"s which could appear in one of four locations
six degrees in radius from a central fixation point, and in
one of four orientations (pointing right or left sideways, up
or down).

Procedure

Observers were seated in front of the screen, with their
head position fixed by the chin support. Screen refreshes
were made every 16.7 msec, so that the duration in msec of
any stimulus, target etc. was a multiple of 16.7.

The order of events in each condition was identical (see
Fig. 1). A fixation bar appeared for 668 msec, followed by
a 16.7 msec central arrow cue appearing at one of four
locations. The cue was followed by a target at one location,
and circles at the other three locations, after a variable
stimulus-onset asynchrony (SOA). SOAs used were (rounded to
the nearest msec) 0, 33, 67, 100, 133, 167, 200, 233, 267,
and 300 msec. Following the presentation of the target, a
mask was presented which was an outline of all possible
targets at all four locations. The observer's task was to
indicate, by pressing the corresponding arrow on the numeric

keypad, which direction the target was pointing. The observer was instructed that accuracy, and not speed, was required.

There were four blocked conditions which consisted of different percentages of cue validity; observers were informed of what condition was being run. In the 100% condition, the cue always indicated the correct position of the target. In the 75% condition, the cue indicated the correct target location on 75% of the trials, but on 25% of the trials the target would appear at any one of the noncued locations with equal (8.3%) probability. The 50% condition was similar to the 75% condition except for 50% correct (valid) cuing and 50% incorrect (invalid) cues, split 16.7% to each of the noncued locations. The 25% condition used a cue which was valid 25% of the time and invalid on 75% of the trials, split 25% to each of the noncued locations; this last type of cue gave essentially no information since the chance probability of a target appearing at any one location was 25%.

Each observer was first instructed in the task and run for a number of sessions at the 100% level for training purposes. Target durations were individually determined for each observer on the basis of performance in the training sessions: 33 msec (CT), 33 msec (CW), or 17 msec (MU). Each observer was exposed to 24 sessions of 1280 trials: six sessions each at the four probabilities, with an equal number

of trials at each SOA, and counterbalanced across cue and stimulus location, direction of target, and target duration. Total number of trials per observer per condition per SOA ranged from a maximum of 768 trials to a minimum of 187 trials per point.

Data were collected in the form of accuracy scores (by SOA) for valid and invalid conditions, for both individual observers and combined observers. An analysis of accuracy by SOA curves by condition was planned.

RESULTS AND DISCUSSION

Figure 3 illustrates accuracy by SOA for all four conditions. As can be seen in the figure, the curves for valid trials all lie above the invalid curves. indicating that performance is better at cued locations than at noncued locations with a central cue. This finding replicates Cheal et. al., 1991 and Gottlob, 1992. The manipulation of interest, probability, appears to have no effect, as opposed to the effect seen with the peripheral cue (figure 2). To directly contrast the present findings with those found with the peripheral cue, the 75% and 50% curves are presented alone in figure 4 which shows that the 75% valid curve does not lie above the 50% valid curve and that the 75% invalid curve does not lie below the 50% invalid curve.

Data were analyzed for all conditions, using the Hierarchical Log-linear analysis in BMDP4F, which is a multi-

dimensional nonparametric variant of the chi-square test, for dichotomous data. A summary table of partial chi-squares, along with significance values, is presented in Table 1. There were significant effects of SOA and validity. There was a significant effect across observers, which was to be expected since target durations and presumably, individual ability levels, differed across observers.

There was a significant validity x SOA interaction for combined and all individual observers, indicating that the invalid curves had different shapes than the valid curves. The valid curves rose as a function of SOA; the invalid curves were flat across SOAs. There was also a significant probability x validity interaction in two of the four tests; however, the interaction was not indicative of a probability effect upon visual inspection of the curves. There was no probability x validity x SOA interaction, indicating no significant non-parallelity of curves.

There was a significant validity x SOA x observer interaction, due to individual differences in shape and value of curves.

The results indicate that performance at cued locations improves as a function of SOA, and noncued performance does not improve with SOA. In contrast to the probability effect seen with a peripheral cue, there is no probability effect seen with the central cue, indicating that observers cannot

allocate attention in accordance with cue probability.

The inability to allocate can be due to several factors. The first is that processing of a central cue can impose a capacity limitation on the resources used for allocating attention. As discussed above, there is evidence that the response to a peripheral is at least partly automatic, while central cue response is voluntary. A central cue must be interpreted and attention must be moved, which may leave no resources for anything but movement to the cued location.

Another possible impediment to allocation may be a "mechanical" problem of moving attention in several directions simultaneously. In an "automatic" task like peripheral cuing, it may be possible to spread attention over an entire display at fixation (facilitating detection of the cue) and then allow attention to concentrate on the cued location while exerting some control over the amount remaining at the noncued locations. In contrast, with a central cue, if central fixation is required to interpret the cue, then attention may be concentrated at the center. An optimal allocation of attention for any given probability cue would require the simultaneous movement to four locations: the single cued and the three noncued. A purposive movement of attention to four separate locations might be too difficult to perform.

A third possibility is that the finding of no

probability effect is a type 2 error; that is, the experiment did not pick up the (real) effect. It is hard to compare results from peripheral and central cue designs since we didn't use a true within-subject design. Experiment 2 in this report was designed to remedy this problem by manipulating both probability and cue type (central vs. peripheral) in a within-subject design.

EXPERIMENT 2

Experiment 2 was designed to directly investigate the hypothesis that probability allocation is possible with a peripheral cue but not with a central cue. To that end, we used two probabilities (75% and 50%) and two types of cue (central and peripheral).

METHOD

Observers

Four observers, three male and one female, between the ages of 20 and 33, were used in this experiment. Two were experienced in this type of experiment; one observer (CT) had served in Experiment 1.

Apparatus and Stimuli

The apparatus and stimuli were the same as used in Experiment 1.

Procedure

The trials were similar to those in Experiment 1, with

the following changes: (1) There were 7 SOAs - 0, 50, 100, 150, 200, 250, and 300 msec. (2) Cue type was manipulated across session. A central cue consisted of a "V" presented at fixation that cued one of four target locations. A peripheral cue was the same "V" presented 1 degree to the outside of a single target location. (3) Only two probabilities were used for cue validity - 75% and 50%, since those did not represent any boundary conditions as 100% and 25% did. (4) Target durations ranged from 33 to 67 msec. Number of trials per observer, per point, ranged from 185 to 495.

RESULTS AND DISCUSSION

The data for one observer, MP, was not included in the analysis because he was unable to respond to a central cue. Therefore, the analysis will include data only from the other three observers.

The accuracy by SOA curves are presented separately for the central and peripheral conditions, in Figure 5. The peripheral cue data shows a possible probability effect for valid but not invalid trials, while the central cue data shows a possibly very small effect for valid trials but not for invalid trials.

Table 2 shows the results of the High-Log Linear analysis for all observers combined and for individuals.

Both central and peripheral show effects of validity, SOA, and observer. The interaction of most interest is the one of probability by validity - that, combined with an absence of probability by SOA interaction (indicating parallel curves) and visual inspection, can indicate further tests for valid and invalid curves separately. The P x V interaction was not significant for the central cue (indicating no effect of probability) but it was significant for the peripheral cue, which suggests further tests for the peripheral cue.

Table 3 shows the results for valid and invalid trials only, for the peripheral cue. There was no probability effect for invalid trials, which is not surprising given figure 5 which shows the two lines crossing. For valid trials, there was an effect of probability for all observers combined and for two of the three individuals.

Therefore, the findings for the second experiment are equivocal. It does replicate the lack of probability effect for the central cue, but it only partly replicates the effects found in last summer's experiment with the peripheral cue. The first peripheral cue experiment found probability effects for both valid and invalid trials, while the present experiment found the effect for valid but not invalid trials.

CONCLUSION

Both experiments 1 and 2 found that observers could not perform a probability allocation of attention in response to

a central cue. Experiment 2 partly replicated last summer's experiment in that it found a probability effect for valid trials (but not for invalid trials).

At this point it seems appropriate to investigate further the peripheral cue effect. The effect was significant at the 0.01 level for all observers combined; however the magnitude of the effect was small: the mean proportion correct for the 75% valid trials was 0.861, and the mean proportion correct for the 50% valid was 0.832. This yielded a small but consistent difference of 0.029 which only represents 3.4% of the average accuracy.

One way to boost the effect (if it is real) is to present two probabilities which are more discriminable by the observers. The difference between 75% and 50% is only one third of the higher value. In the next experiment, I will try to produce a larger effect by using the two probability levels of 66% and 33%, where one level is twice the other.

REFERENCES

Cheal, M. L., & Lyon, D. (1989). Attention effects on form discrimination at different eccentricities. Quarterly Journal of Experimental Psychology: A. Human Experimental Psychology, 41A, 739-746.

Cheal, M. L., & Lyon, D. (1991). Importance of precue location in directing attention. Acta Psychologica, 76, 201-211.

Cheal, M. L., Lyon, D., & Hubbard, D. (1991). Does attention have different effects on line orientation and line arrangement discrimination? *Quarterly Journal of Experimental Psychology: A. Human Experimental Psychology, 43A,* 825-857.

Gottlob, L.R. (1992). Spatial allocation of attentional resources varies with proportion of correctly cued trials. Unpublished Master's thesis.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long, & A. Baddely (Eds.), *Attention and performance* (pp. 187-203). Hillsdale, NJ: Erlbaum.

Koopman, B.O. (1957). The theory of search; Pt. III: The optimum distribution of searching effort. *Operations Research, 5,* 613-626.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 11,* 1631-1647.

Tepin, M. B., and Dark, V. J. (1991). Do abrupt-onset peripheral cues attract attention automatically? Unpublished manuscript.

Figure 2. Accuracy (proportion correct) by SOA for Gottlob, 1992. The two top lines represent performance for valid trials, and the bottom lines for invalid trials. Two probability conditions are represented: 75% and 50%.

Figure 1. The order of events in a location-cuing trial (adapted from Cheal et al., 1991). Observers first fixate on the central bar. The peripheral cue appears, followed by target presentation after a variable cue-target onset asynchrony (SOA). On valid trials, the target appears at the cued location. On invalid trials, the target appears at one of the three noncued locations.

Figure 1. Accuracy by SOA curves for Experiment 1. Probabilities are 75%, 50%, and 25%. There is an effect of validity but none for probability.

Figure 4. Experiment 1 accuracy curves for 75% and 50% only. There is no probability effect.

Table 1
Partial Chi-Squares for the Central Cue Condition - Exp 1

Effect	df	All Observers(O)	MU	CW	CT
Prob.	1	3.02	9.27*	2.31	2.53
Val.	1	1595.48**	154.13**	1232.85**	482.16**
SOA	7	484.08**	106.94**	154.49**	152.70**
Obs.	2	817.11**			
P x V	1	6.49*	5.49*	1.82	0.47
P x S	9	7.54	7.07	9.93	9.10
V x S	9	196.41**	15.93	114.71**	111.53**
P x O	2	10.85**			
V x O	2	273.21**			
S x O	18	28.93*			
PxVxS	9	6.43	14.01	15.09	7.82
VxSxO	18	45.99**			
PxVxO	2	1.29			
PxSxO	18	18.29			
PxVxSxO	18	17.66			

** p < .001, *p < .05
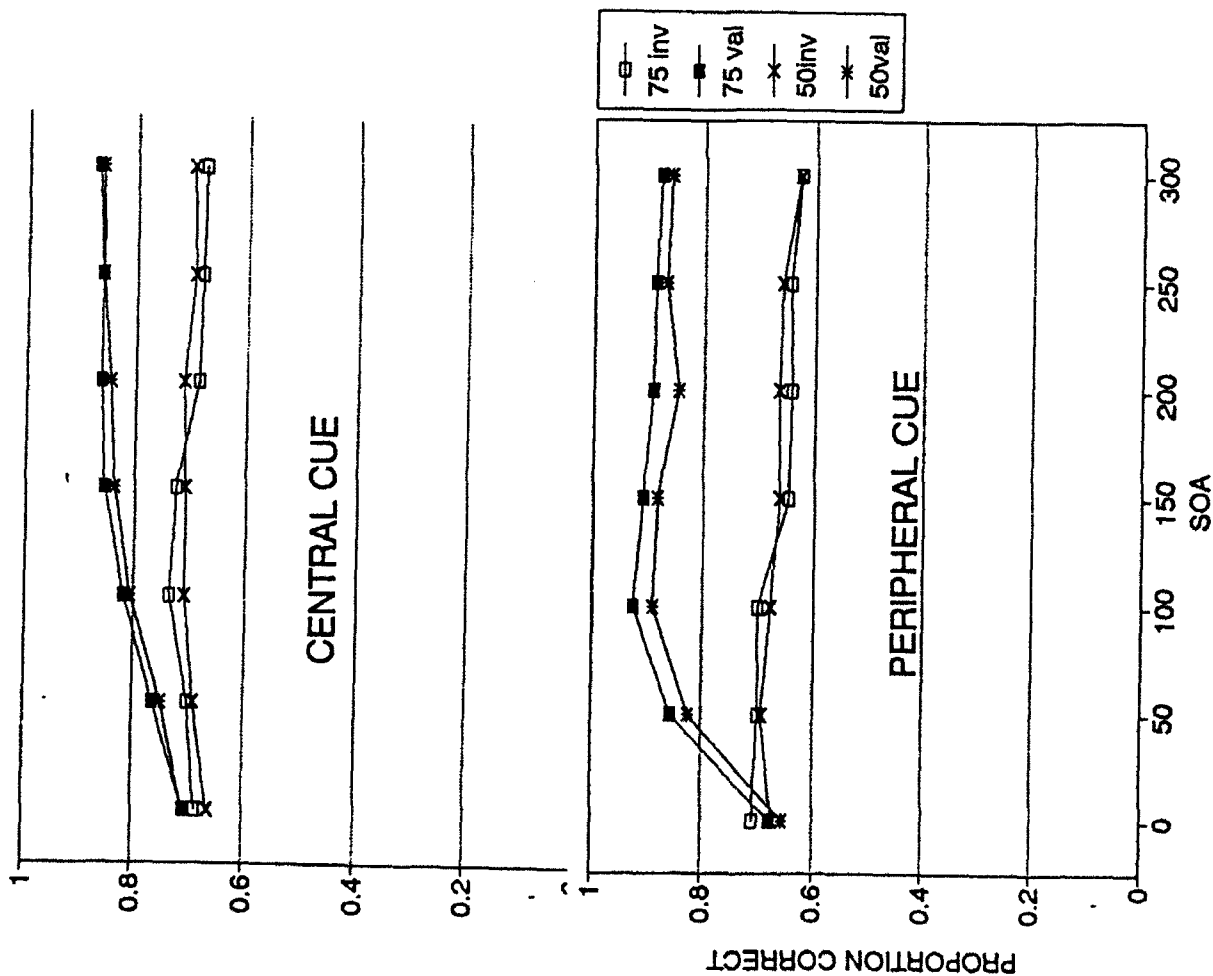
Figure 5. Accuracy by SOA curves for Experiment 2. The top graph is for central cues, and the bottom graph is for peripheral cues.

Table 2

Partial Chi-squares for Experiment 2 - Central and Peripheral

Effect	df	All Observers(O)	CT	TB	SB
Prob.	1	3.36 / 6.49	14.79** / 5.78#	28.50** / 3.53	0.80 / 0.03
Val.	1	755.57** / 2132.03**	120.66** / 499.79**	462.16** / 1011.32**	213.69** / 845.66**
SOA	6	460.27** / 539.76**	117.87** / 167.30**	195.70** / 219.54**	173.17** / 204.22**
Obs.	2	1272.88** / 603.42**			
P x V	1	3.75 / 15.25**	0.58 / 0.08	1.71 / 15.48**	1.49 / 6.90*
P x S	6	5.28 / 3.44	2.80 / 0.89	3.67 / 3.51	3.49 / 3.34
V x S	6	127.60** / 435.33**	55.30** / 190.70**	71.09** / 179.73**	25.09** / 284.58**
P x O	2	39.58** / 2.81			
V x O	2	30.39** / 15.21**			
S x O	12	22.35* / 33.90**			
PxVxS	6	5.08 / 15.50*	1.39 / 6.62	4.06 / 5.15	6.44 / 7.37
VxSxO	12	21.45* / 17.22			
PxVxO	2	0.06 / 7.57*			
PxSxO	12	9.21 / 3.58			
PxVxSxO	12	6.01 / 17.03			

Note: numbers in regular type are for central cues; bold type denotes peripheral cues.

**$p < .001$, *$p < .05$

Table 3

Partial Chi-Squares for Experiment 2 Peripheral Cue, Valid and Invalid Trials Only

		All Observers	
Effect	df	Valid	Invalid
Prob.	1	23.27**	0.81
SOA	6	1216.82**	30.45
Obs.	2	226.14**	401.97
S X O	12	28.29**	22.28*
P X S	6	2.70	2.75
P X O	2	1.77	8.74*
PxSxO	12	11.05	10.36

		Individuals					
		CT		TB		SB	
Effect	df	Val	Inv	Val	Inv	Val	Inv
Prob.	1	3.36	3.48	16.65**	2.65	4.74*	2.97
SOA	6	355.78**	15.65*	408.71**	15.60*	480.96**	22.18**
P X S	6	3.38	4.13	4.42	4.24	4.75	7.63

Note - See Table 1 for abbreviations.
**$p < .01$, *$p < .05$ .op

14-20

The Effects of Two Doses of Exogenous Melatonin on Temperature and Subjective Fatigue

Rod J Hughes
Graduate Research Assistant
Psychology Department

Bowling Green State University
Bowling Green, Ohio 43403

The Effects of Two Doses of Exogenous Melatonin on Temperature and Subjective Fatigue

Rod J Hughes
Graduate Research Assistant
Psychology Department
Bowling Green State University

Abstract

The effect of orally administered exogenous melatonin on oral temperature and subjective mood states was tested in a placebo-controlled, double blind design. Subjects were tested from 0800 - 1700. At 0915 subjects ingested either 0 mg, 10 mg or 100 mg of melatonin. Serum melatonin for both of the treatment conditions raised rapidly and remained well above physiological levels for the entire testing session. Exogenous melatonin administration yielded significant dose-dependent decreases in oral temperature. Exogenous melatonin treatment also significantly increased subjective ratings of fatigue.

The Effects of Two Doses of Exogenous Melatonin on Temperature and Subjective Fatigue.

Rod J Hughes

Introduction

The pineal hormone melatonin (MT) has been implicated as a possible "hypnotic" or sleep inducing agent. The present investigation used a placebo-controlled, double blind design to test some sedating qualities of two doses of orally administered exogenous melatonin. Compared to placebo, exogenous melatonin showed dose-dependent decreases in oral temperature. Exogenous melatonin also significantly increased subjective fatigue in a dose-dependent manner.

Endogenous melatonin is produced primarily in the pineal gland via metabolism of tryptophan (1). The vast majority of melatonin is produced and released into all parts of the brain and all parts of the body primarily at night (2). There is considerable evidence that melatonin is the "messenger of darkness", mediating the effects of photoperiod throughout the body (2 see also 3). In fact, endogenous melatonin is likely responsible for much if not all circadian rhythmicity in mammalian physiology.

Exogenous melatonin has been shown to have two types of effects on human physiology. First, exogenous melatonin has chronic circadian phase-setting effects. For instance, melatonin has been used to facilitate adjustment to jet-lag (e.g. 4,5). Second, exogenous melatonin has more acute sedating qualities (e.g. 6). The present investigation will focus on the immediate effects of exogenous melatonin administration.

Evidence supporting the contention that melatonin has sedating qualities comes from several sources. Lieberman et al showed that a large dose of melatonin (240 mg over two hours) produced significant tranquilizing effects.

15-3

After ingestion of MT their subjects reported increased subjective fatigue. They also showed behavioral decrements in the form of longer reaction times to psychomotor tasks (6). Others have reported similar subjective effects for medium (50 mg) doses of exogenous melatonin (7) and for smaller (2 mg) doses of exogenous melatonin (8).

The few investigations that have directly tested the effects of exogenous melatonin on sleep support the potential use of MT as a sleep inducing agent. For instance, Waldhauser et al reported that large doses of melatonin can significantly improve sleep (9). Studies using lower doses of MT have also reported promising findings (10,11,12).

The primary advantages of using MT as a sleep aid are that (a) MT is a natural substance that is metabolized by the body every night and relatedly that (b) MT does not appear to show the side effects associated with benzodiazepines (the most commonly used type of hypnotic drug). For instance, Temazepam has been recommended as a sedative for promoting crew rest (13,14). Although Temazepam has proven to be effective in promoting daytime sleep (13,14) it, like other benzodiazepines, has the potential to produce anterograde amnesia (e.g. 13). Further, because of its intermediate half-life (metabolic rate) accumulation and carry-over effects are other potential problems with this and other benzodiazepines (15).

It is hoped that since melatonin is a natural substance it will not show the negative side effects associated with typical hypnotic drugs. Therefore, for promoting crew rest exogenous melatonin may prove to be a safe and effective alternative to the benzodiazepines. Before implementation, however, more systematic research on the specifics of melatonin administration needs to be done. For instance, in order to choose an effective dose for sleep research some of the pharmacokinetic issues of melatonin administration need to be documented. These issues are rate of absorption or the speed at which exogenous melatonin is brought into the system and elimination half life or how quickly exogenous melatonin is metabolized. Therefore, aside from testing the effects of MT on temperature and

15-4

subjective fatigue the present investigation was designed to document some of the pharmocokinetics of two morning doses of exogenous melatonin.

Method

Subjects

Six male volunteers between the ages of 23 and 42 participated in three one day sessions spaced a week apart. Three of the subjects were paid for their participation. The remaining three subjects received time off from work as compensation. After the experiment it was discovered that one of the subjects suffered from a sleep disorder. Consequently, his data were not included in the final analyses.

Materials

Melatonin and Placebo

The pill conditions were 0 mg, 10 mg and 100 mg of synthetic melatonin (Sigma Chemical Co, Saint Louis, Missouri) mixed, in 500 mg gelatin capsules, with Konsyl (6.0 grams of Psyllium Fiber per teaspoon, Konsyl Pharmaceuticals Fort Worth, Texas).

Subjective Fatigue Scale

This is a seven point self-report fatigue scale. The scale ranges from 1 (fully alert; wide awake; extremely peppy) to 7 (completely exhausted; unable to function effectively; ready to drop). This scale and others like it (e.g. the Stanford Sleepiness Scale) have been shown to be sensitive to fatigue.

Profile of Mood States (POMS)

The POMS is a computer presented five point self-report scale in which subjects rate how well a given adjective describes their present feelings. 65 adjectives aggregate into 6 emotional factors; Anger, Confusion,

Depression, Fatigue, Tension and Vigor. The POMS has also shown to be a sensitive measure of fatigue as well as being a sensitive measure of emotional changes associated some drug administration (6).

Oral Temperature

Subjects took their oral temperature every hour using a Becton Dickinson (B-D) Digital Thermometer.

Melatonin Collection

Systemic levels of melatonin were measured from samples of saliva and serum. Every hour subjects extracted approximately 4 cc of saliva which was chilled in ice immediately. Within 10 minutes the saliva samples were placed in a Clay-Adams Sero-Fuge II centrifuge and centrifuged at 3491 rpms for 10-15 minutes. About 1 ml of sputum was then extracted and placed in microfuge tubes. These samples were then stored at -70 degrees Farenheit until they were shipped to an outside laboratory for assaying.

Blood (10 cc) was extracted every two hours via an indwelling IV catheter with a heparin lock. These samples too were chilled until being centrifuged. The serum was then separated into 1 ml aliquotes and placed in microfuge tubes. All other handling of the serum samples was the same as for the saliva samples

Melatonin Assay

Serum and saliva samples were taken to Dr. Reiter's laboratory at UTSA Health and Science Center for radioimmunoassy (16).

Procedure

Subjects arrived at the laboratory at 0700, at which time a 20 gauge indwelling IV catheter was inserted into one of their arms. Beginning at 0800 and ending at 1700 saliva, temperature and subjective fatigue were assessed hourly. Serum samples and the POMS were taken every two hours.

15-6

At 0915 subjects ingested a pill containing either 0 mg, 10 mg or 100 mg of melatonin. The order of pill administration was counter-balanced across each of the sessions. Subjects were required to remain in the laboratory for the duration of each testing session (0800 - 1700). While in the laboratory subjects could read, write, work on a computer or play computer/video games. Subjects had hourly access to juice, however, they refrained from eating for the duration of the testing session.

Analyses

Two-way (pill condition X time of day) repeated measures analyses of variance (ANOVA) were performed on the data. 3 (pill condition) X 10 (time of day) ANOVAs were run on oral temperature and subjective fatigue. 3 (pill condition) X 5 (time of day) ANOVAs were run on each of the factors that make up the POMS test (anger, confusion, depression, fatigue, tension and vigor).

Results

At the time of this publication only two thirds of the melatonin samples had been assayed, however, Figure 1 depicts clear dose-dependent effects for serum melatonin levels (saliva data show similar functions). As expected, melatonin levels in the placebo condition remained within physiological levels. The two treatment conditions, however, showed marked dose-dependent levels of melatonin. After ingestion of exogenous melatonin, serum melatonin levels quickly raised to pharmacological levels and remained high throughout the testing session.

Oral temperature revealed a significant main effect for dose $F_{(2,6)} = 13.22$, $p < .005$. There was also a significant main effect for time of day $F_{(8,24)} = 3.82$, $p < .005$. The dose by time of day interaction approached statistical significance $F_{(16,48)} = 1.75$, $p < .07$. Because of the high variability of the 0800 temperature data this time point was removed from this analysis. This did not significantly affect either the form of the data or

15-7

the statistical analysis. Figure 2 shows the temperature data represented as difference scores from the 0900 (baseline) time point.

Analysis of the subjective fatigue ratings (see Figure 3) also revealed significant main effects for dose $F(2,6) = 16.49$, $p < .005$ and for time of day $F(9,27) = 5.89$, $p < .005$. Overall, the 100 mg condition produced the highest mean fatigue ratings (0 mg = 1.55, 10 mg = 1.83, 100 mg = 2.27). The dose by time of day interaction was also significant $F(18,54) = 2.88$, $p < .05$.

The POMS fatigue ratings (see Figure 4) also revealed a significant main effect for dose $F(2,4) = 6.52$, $p .05$. Again the 100 mg condition produced the highest mean ratings of fatigue (0 mg = 35.07, 10 mg = 36.87, 100 mg = 44.26). The dose by time of day interaction was again significant $F(8,16) = 4.83$, $p < .05$.

POMS tension ratings yielded a significant main effect for time of day $F(4,8) = 7.22$, $p < .05$. For these data as well the dose by time of day interaction was significant $F(8,16) = 3.12$, $p < .05$. Figure 5 depicts the changes in POMS tension ratings across the day.

None of the remaining POMS factors yielded significant results (see Figures 6,7,8 & 9). However, although they were not statistically significant the data for subject ratings of POMS vigor (Figure 6) and POMS confusion (Figure 7) are consistent with the finding that subjects were more tired in the 100 mg condition.

Discussion

The results of this investigation add to the current body of evidence that exogenous melatonin has significant sedating qualities. Melatonin produced significant increases in subjective fatigue. This investigation also adds to the body of evidence that MT plays a significant role in thermoregulation (17). Czeisler has shown that going to sleep while body temperature is decreasing leads to shorter latency to sleep and longer sleep duration (18). Therefore, since sleep quality is known to be dependent on body temperature some of the

15-8

sleep promoting effects of exogenous MT administration may be secondary to its effects on body temperature.

Future research should extend the window of melatonin sampling in order to better assess some of the pharmocokinetics of exogenous melatonin administration. Ongoing collaborative investigations are focusing on this issue as well as studying some of the behavioral/performance decrements associated with daytime melatonin administration. In addition, ongoing research is directly assessing the efficacy of melatonin on polygraphically recorded sleep.

The present investigation was part of an ongoing evaluation of the hypnotic efficacy of exogenous melatonin. The goal of this research is to demonstrate that melatonin can be used as an effective and safe alternative to benzodiazepines for promoting crew rest. In addition, future research will focus on the combined strategy of daytime melatonin administration and nighttime bright light administration in order to (a) promote daytime sleep (b) increase nocturnal performance and (c) rapidly phase shift endogenous circadian rhythms.

References

1. Reiter, R.J. (1991). Pineal melatonin: Cell biology of its synthesis and of its physiolgical interactions. Endocrine Reviews, 12(3), 151-180.

2. Reiter, R.J. (1991). Melatonin: That ubiquitously acting pineal hormone. News in Physiological Sciences, 6, 223-227.

3. Wehr, T.A., (1991). The durations of human melatonin secretion and sleep respond to changes in daylength (photoperiod). Journal of Clinical Endocrinological Metabolism, 73(6), 1276-1280.

4. Arendt, J., Aldhous, M., English, J., Marks, V., Arendt, J.H., Marks, M. & Folkard, S. (1987). Some effects of jet-lag and their alleviation by melatonin. Ergonomics, 30(9), 1379-1393.

5. Samel, A., Wegmann, H-M., Vejvoda, M., MaaB, H., Gundel, A. & Schutz, M. (1991). Influence of melatonin treatment on human circadian rhythmicity before and after a simulated 9-Hr time shift. Journal of Biological Rhythms, 6(3), 235-248.

6. Lieberman, H.R., Waldhauser, F., Garfield, G., Lynch, H.J. & Wurtman, R.J. (1984). Effects of melatonin on human mood and performance. Brain Research, 323, 201-207.

7. Nickelson, T., Demisch, L., Demisch, K., Radermacher, B. & Schoffling, K. (1989). Influence of subchronic intake of melatonin at various times of the day on fatigue and hormonal levels: A placebo-controlled, double-blind trial. Journal of Pineal Research, 6, 325-334.

8. Arendt, J., Barbely, A.A., Franey, C. & Wright, J. (1984). The effect of chronic, small doses of melatonin given in the late afternoon on fatigue in man: A preliminary study. Neuroscience letters, 45, 317-325.

9. Waldhauser, F., Saletu, B. & Trinchard-Lugan, I. (1990). Sleep laboratory investigations on hypnotic properties of melatonin. Psychopharmacology (Berlin), 100(2), 222-226.

10. James, S.P., Mendelson, W.B., Sack, D.A., Rosenthal, N.E. & Wehr, T.A. (1990). The effect of melatonin on normal sleep. Neuropsychopharmacology, 1(1), 41-44.

11. James, S.P., Sack, D.A., Rosenthal, N.E. & Mendelson, W.B. (1990). Melatonin administration in insomnia. Neuropsychopharmacology, 3(1), 19-23.

12. Sack, R.L., Blood, M.L., Omerod, G.M., Rich, G.B. & Lewy, A.J. (1992). Oral melatonin reverses the alerting effects of nocturnal bright light exposure in humans. Abstract presented at the 6th annual meeting of the Association of Professional Sleep Societies. Phoenix, Arizona.

13. French, J., Boll, P. & Storm, W.A. (1990). Temazepam and performance following a sleep cycle shift. Annual Review of Chronopharmacology, 7, 41-44.

14. Storm, W.F. & Parke, R.C. (1986). FB-111A aircrew use of temazepam during surge operations. AGARD-CP-415, 12-1 - 12-11.

15. Kales, A. & Kales, J.D. Evaluation and Treatment of Insomnia. Oxford University Press, New York. 1984.

16. Fraser, S., Cowen, P., Fanklin, M. & Shaw, D.M. (1985). Direct radioimmunoassy of melatonin in saliva (letters). Clin. Chem., 31, 1412-1413.

17. Badia, P., Meyers, B. & Murphy, P. Melatonin and thermoregulation. In Reiter, R.J. & Yu, H-S. (Eds.) Melatonin: Biosynthesis, Physiological Effects, and Clinical Applications. (1992). CRC Press. Boca Raton, FL.

18. Czeisler, C.A., Weitzman, E.D., Moore-Ede, M.C., Zimmerman, J.C. & Knauer, R.S. (1980). Human sleep: Its duration and organization depend on its circadian phase. Science, 210, 1264-1267.
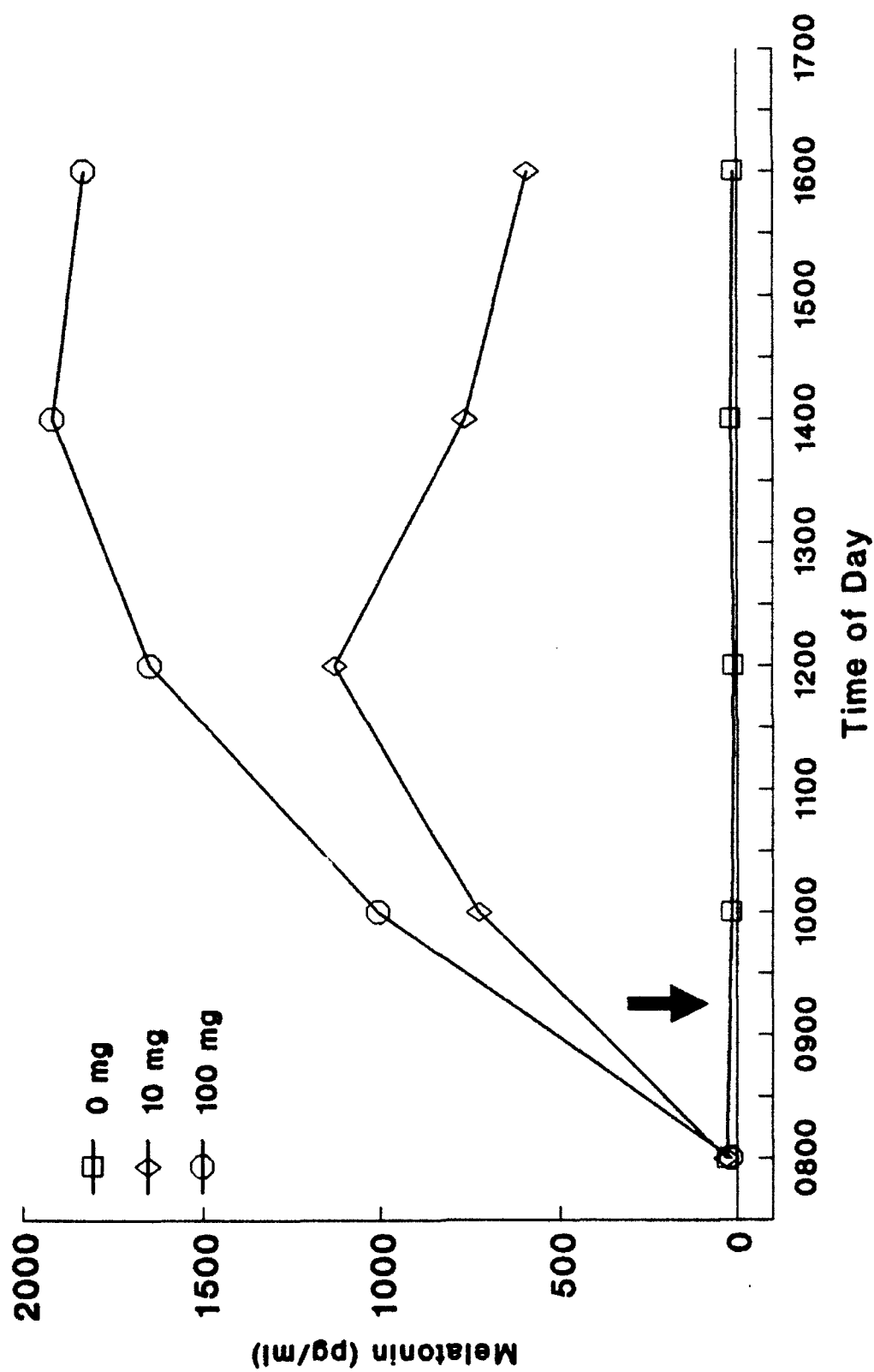
Figure 1 Dose-response changes in serum melatonin by time of day. Arrow indicates time of the dose.

15-12
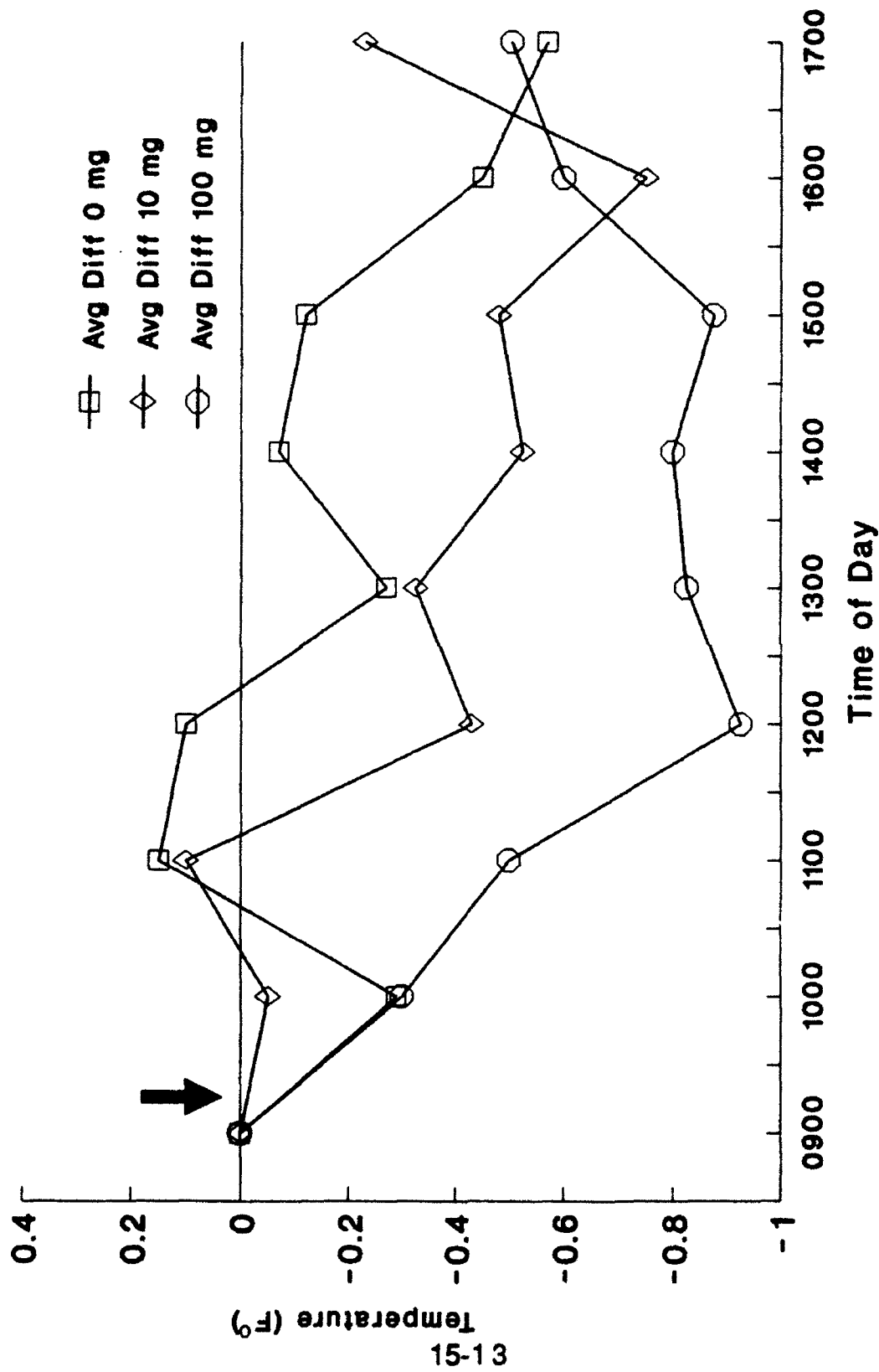
Figure 2 Dose-response changes in oral temperature by time of day.

15-51
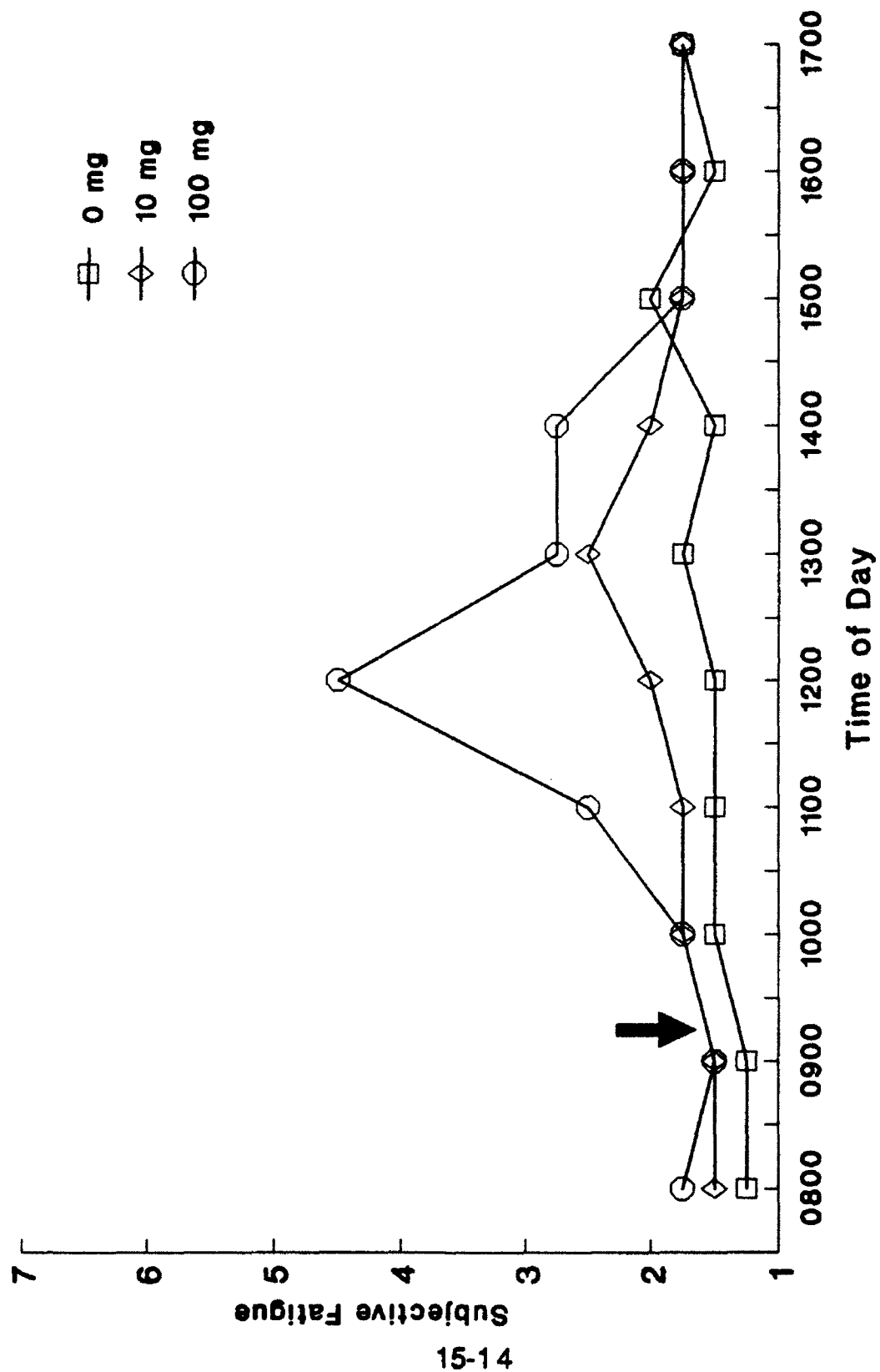
Figure 3 Dose-response changes of subjective fatigue by time of day.
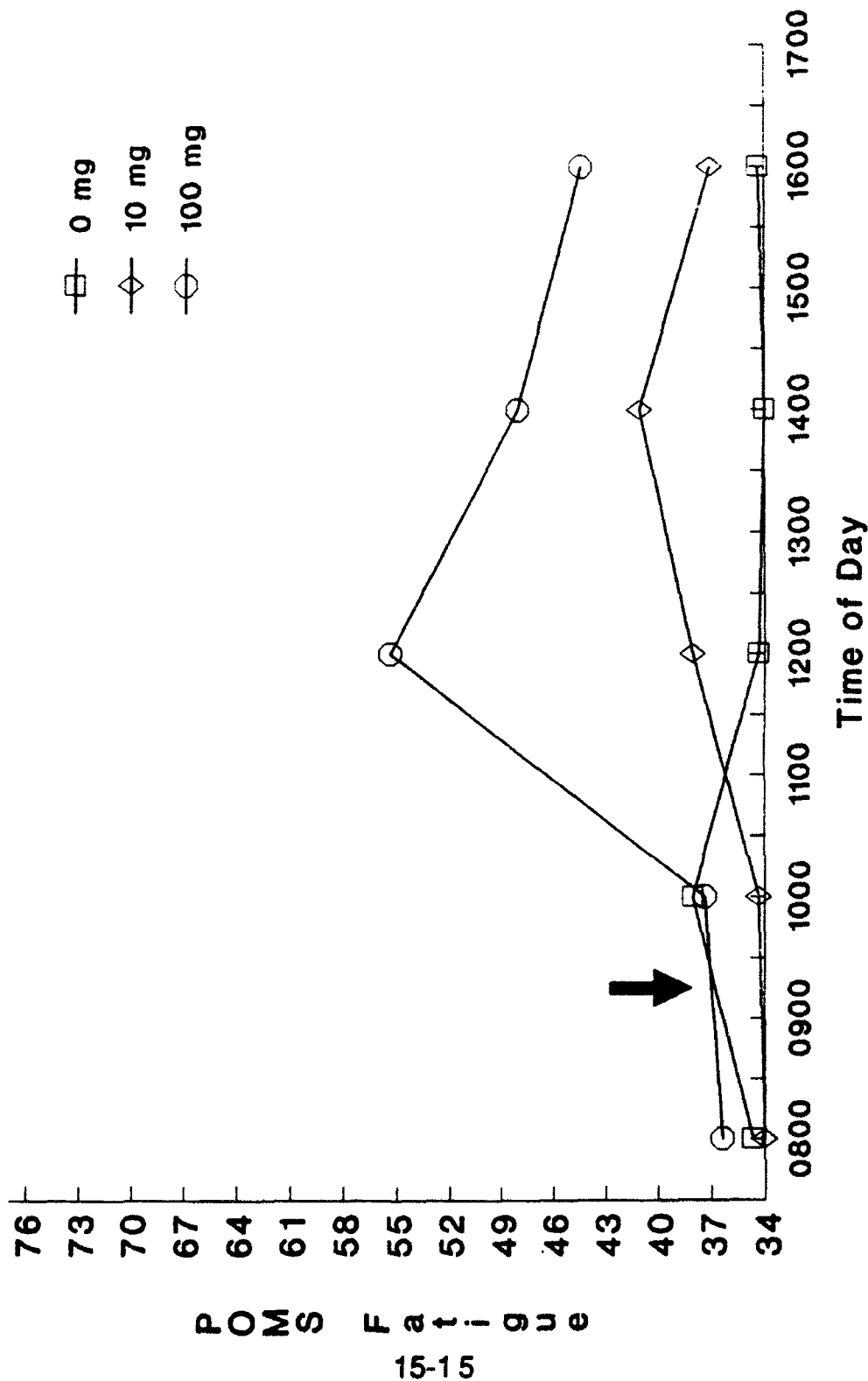
Figure 4 Dose-response changes in
POMS fatigue by time of day.

15-15
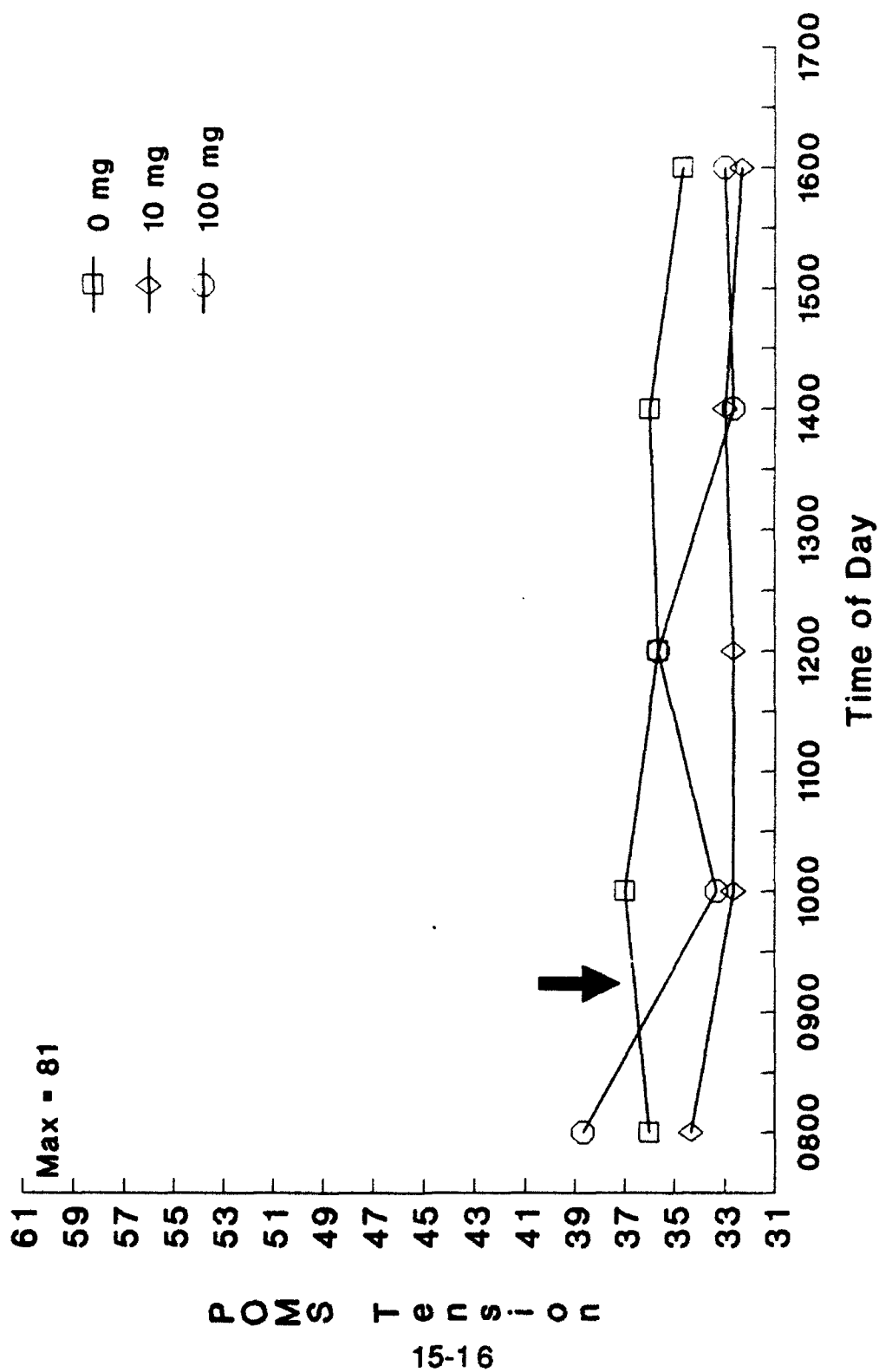
Figure 5 Dose-response changes in
POMS tension by time of day.

15-16
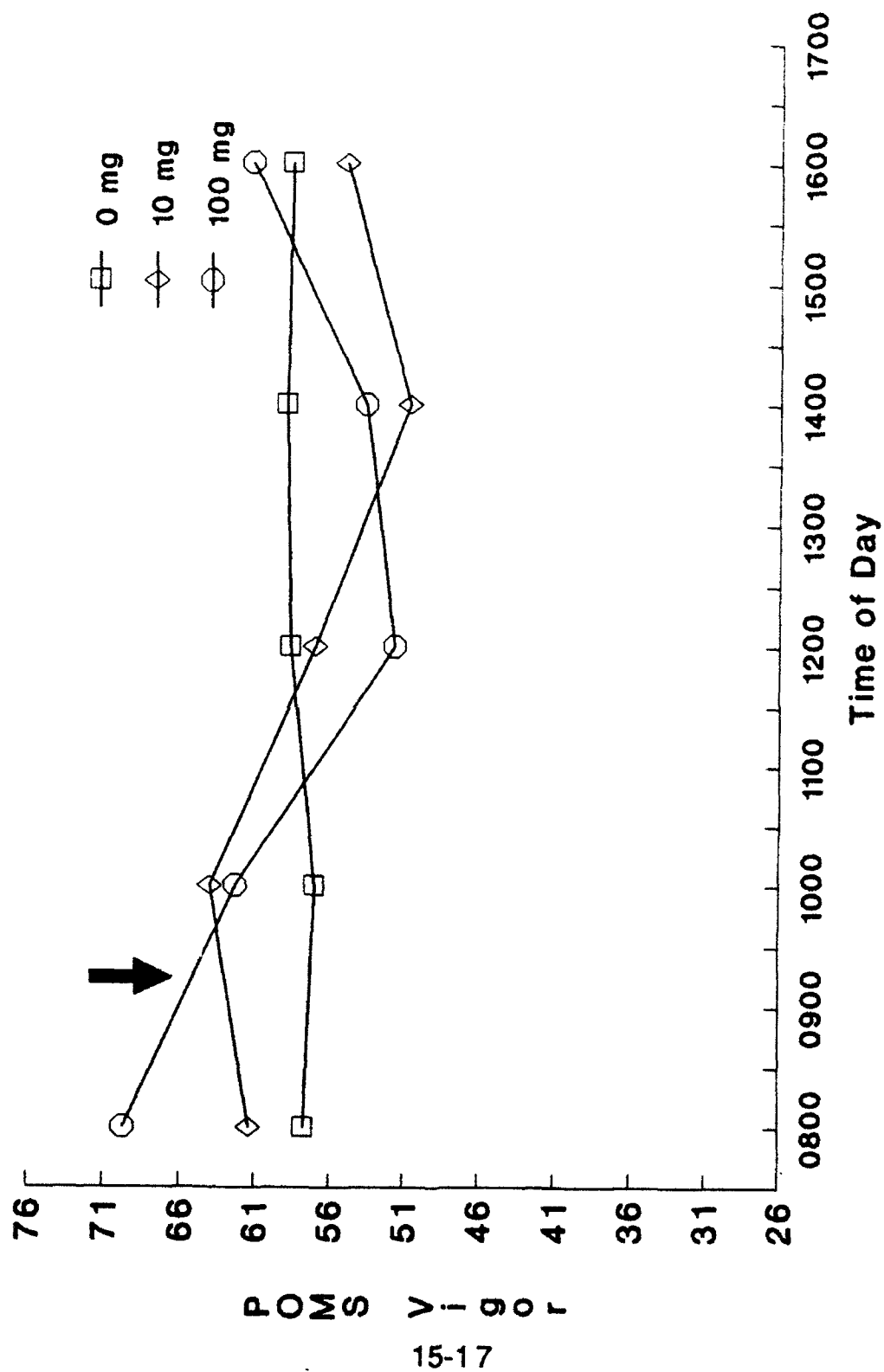
Figure 6 Dose-response changes in POMS vigor by time of day.

POMS Vigor

15-17
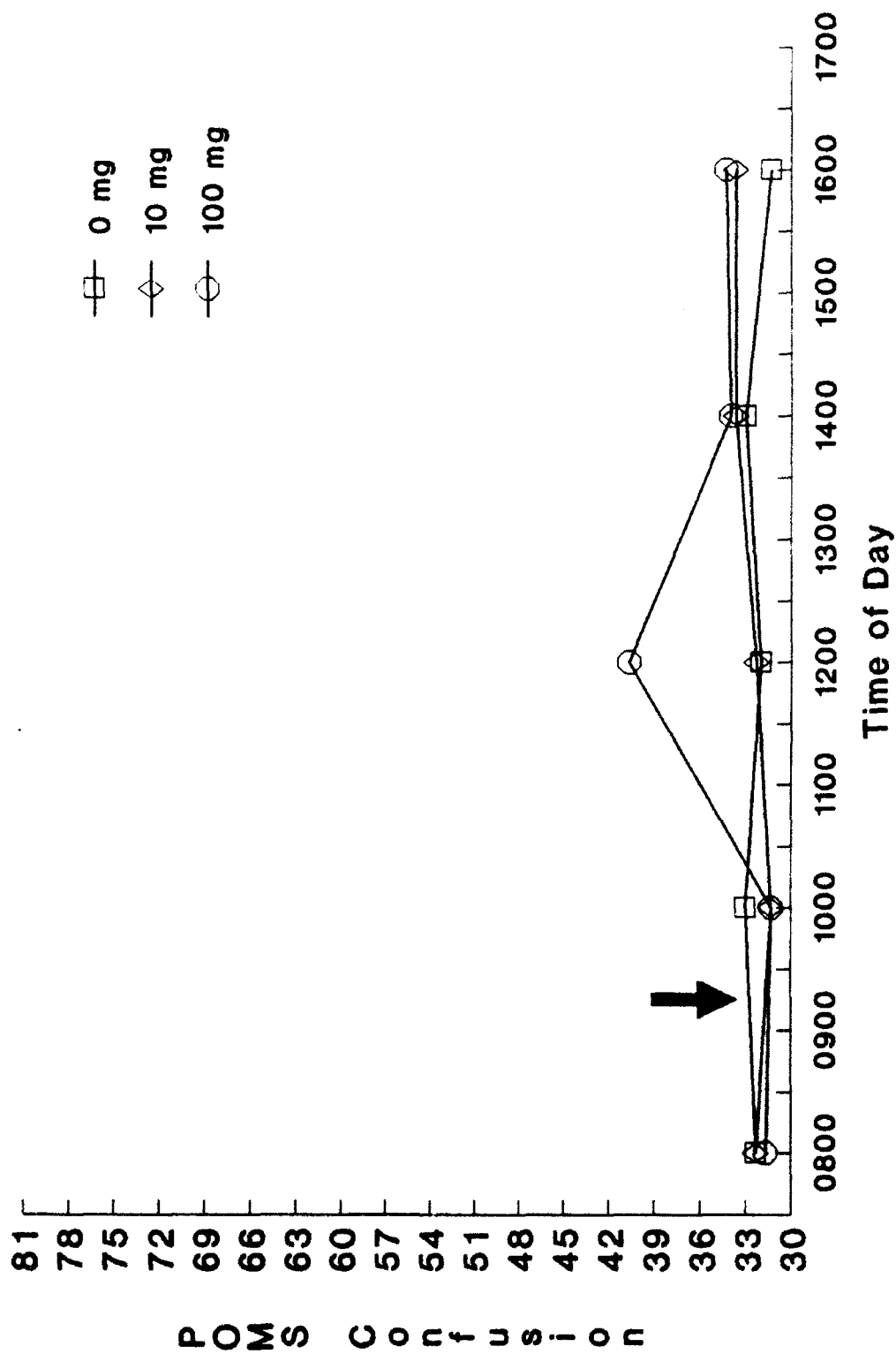
Figure 7 Dose-response changes in POMS confusion by time of day.

15-18
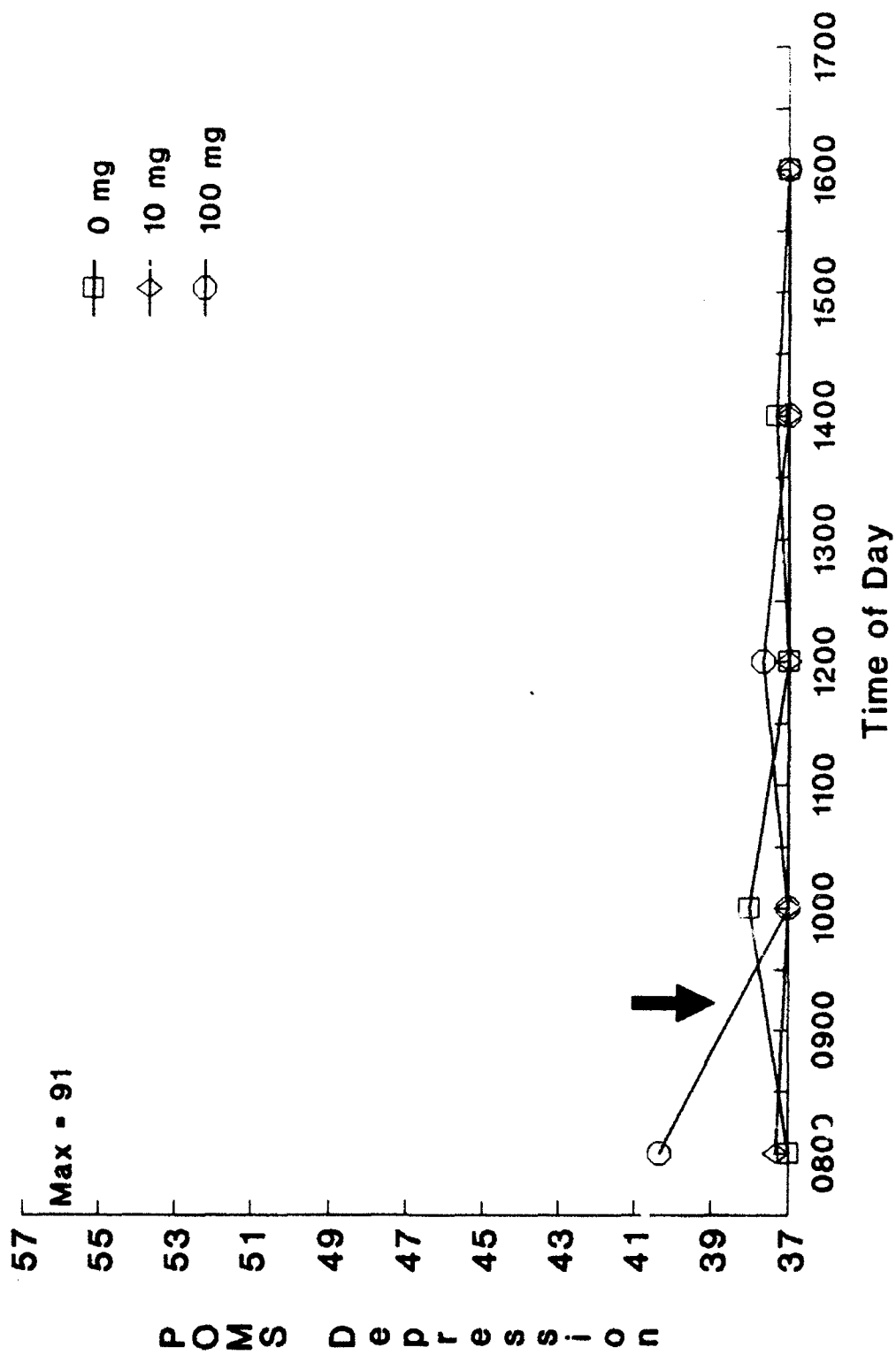
Figure 8 Dose-response changes in
POMS depression by time of day.

15-19
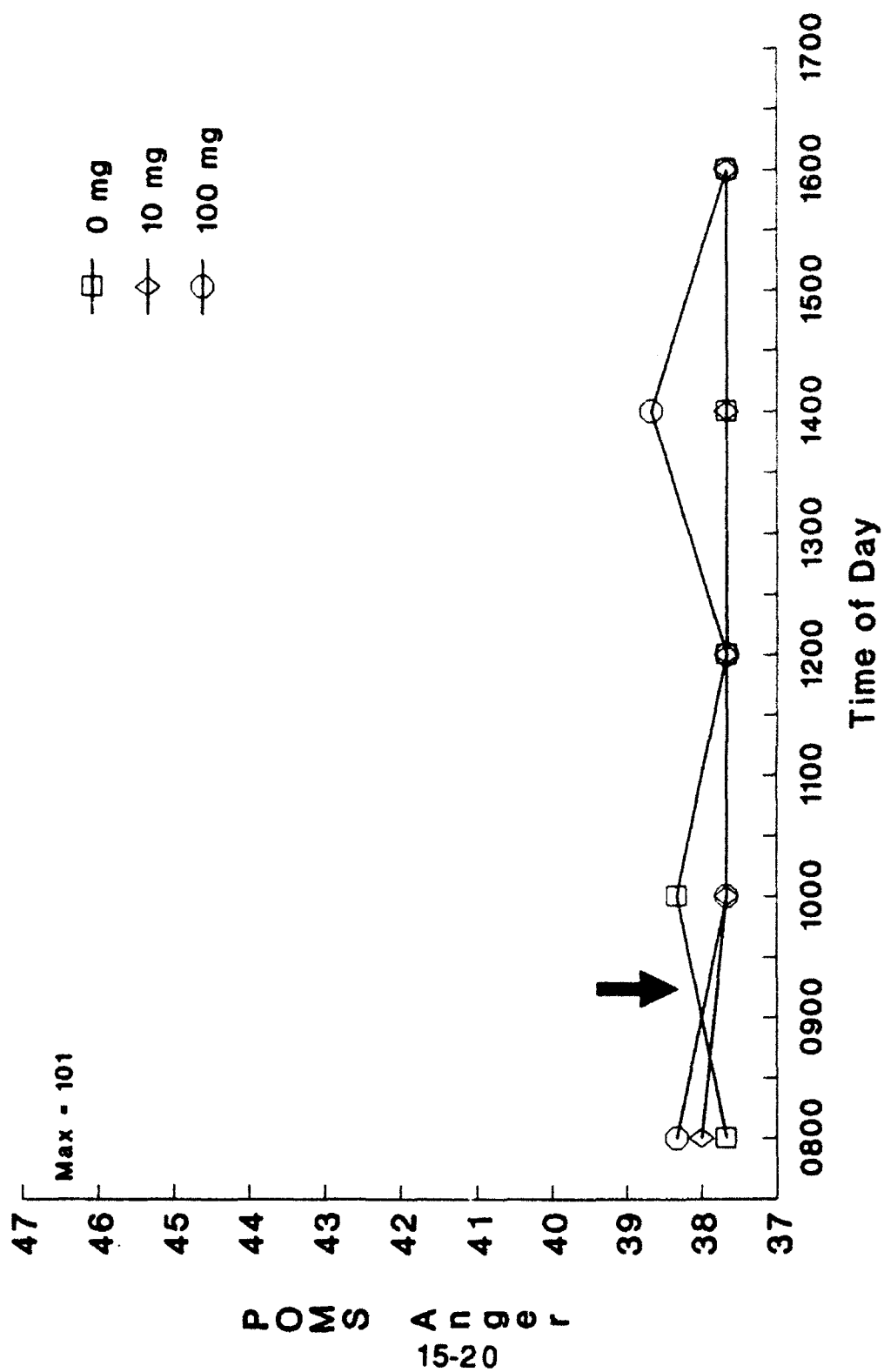
Figure 9 Dose-response changes in
POMS anger by time of day.

ASSISTING AIR FORCE INSTRUCTIONAL DESIGNERS

Tricia Jones

Graduate Student

Department of Educational Studies

University of Michigan

Ann Arbor, MI 48109-1259

ASSISTING AIR FORCE INSTRUCTIONAL DESIGNERS

Tricia Jones

Graduate Student

Educational Studies

University of Michigan

Abstract

The Instructional Design Branch (HRTC) of the Training Research Division at Brooks AFB is committed to providing Air Force instructional designers with tools that support a high level of design efficiency and quality. One of the primary thrusts is a project known as AIDA, the Advanced Instructional Design Advisor. Two short-term projects represent portions of the final project. XAIDA is the Experimental Instructional Design Advisor; it represents a limited prototype of the lesson content specification process. GAIDA stands for the Guided Approach to Instructional Design Advising. This is a case-based system which provides guidance to novice designers about lesson design and delivery, based on Gagné's Nine Events of Instruction (Gagné, Briggs, & Wager, 1992)

During the course of my summer research tenure, I was able to apply a broad range of my own expertise. I was able to combine knowledge of human factors and interface issues with instructional design expertise in order to evaluate XAIDA. Furthermore, I again used my instructional design expertise as well as my skills in software development to refine the design of GAIDA, implement a case for it, and write instructional design commentary for other cases.

XAIDA and GAIDA represent different approaches to advising instructional designers. XAIDA attempts to shield Air Force designers from the need to know principles of instructional design. GAIDA, on the other hand, presents the designers with a particular model and expects them to understand it and incorporate it into their own lessons. Furthermore, XAIDA represents a particular theory of learning and knowledge structure; GAIDA, on the other hand, represents a model of lesson delivery. The design of AIDA combines a variety of tools in order to provide support for novice instructional designers. The complementary strengths and weaknesses of XAIDA and GAIDA reinforce the need for such a structure.

Tricia Jones

1. Introduction

This summer, I was affiliated with the Instructional Design Branch (HRTC) of the Training Research Division at Brooks AFB. The Instructional Design Branch is committed to providing Air Force instructional designers with tools that support a high level of design efficiency and quality. One of the primary thrusts is a project known as AIDA, the Advanced Instructional Design Advisor. During my assignment at HRTC, I participated in the development of tools for Air Force personnel involved in instructional design. I contributed to the long-range goals of AIDA by evaluating XAIDA and participating in the development of GAIDA.

1.1 Overview of HRTC Work: AIDA

The Air Force faces a particular problem in regard to designed instruction. Instructors are chosen from the ranks of experts in the specialities on the basis of their knowledge of the domain. Many of these personnel do not have any particular expertise is designing instruction or delivering lessons. Although they attend an instructor's course before their tour of duty as an instructor, the course does not cover in depth the type of information that is necessary to become an expert instructional designer. Trainers are presented with a formal model of instructional system design (ISD) (Andrews & Goodson, 1980), but many find this too complex to use effectively.

To help overcome this problem, a long-term effort is underway within HRTC to develop tools that will assist novices in the instructional design process. AIDA, the Advanced Instructional Design Advisor, is scheduled to be completed in FY97. "AIDA would incorporate prescriptive advice about course authoring based on established theories of knowledge, learning, and instruction. It would contain a variety of tools. Some would automate established processes. Others would assist or advise authors about designing effective instructional materials. The tools might be used as stand-alone, special purpose tools, or as integrated tools using a shared database of course and content information. The purpose of the AIDA program is to provide reliable, effective, and efficient instructional design guidelines for USAF course designers" (Hickey, Spector, and Muraida, 1991, pp. 1-3). AIDA would incorporate artificial intelligence in the form of an expert system for planning courseware.

One segment of AIDA will contain a number of shells which pertain to different learning types or learning goals. The shells would incorporate elements that expert instructional designers know to consider. A subject-matter expert would interact with this shell to specify the content knowledge of the lesson. The shell would prompt this novice designer, perhaps through a structured series of questions, to

specify various parameters of the lesson, such as the type of learners, their motivation levels, prior knowledge, target performance, etc. The shell would then format and deliver the lesson using the most appropriate delivery methods.

AIDA represents the long-term product in the goal of assisting novice instructional designers. In the short term, two pieces have been under development. XAIDA, which stands for the Experimental Instructional Design Advisor, represents a limited prototype of the lesson content specification process. It incorporates instructional design theories and methods of David Merrill (Merrill, et al., 1990). Another interim product is GAIDA, the Guided Approach to Instructional Design Advising. This effort was initially directed by Dr. Robert Gagné and incorporates his nine events of instruction (Gagné, Briggs, & Wager, 1992). It is a case-based advisor focussing on lesson delivery.

1.2 Overview of My Tasks

My efforts this summer were varied. My initial task at the Laboratory was to become familiar with the various resources that are being used to assist instructional designers. I examined the ID Advisor (originally developed by Progressive Learning Systems under the Small-Business Innovative Research program). I learned the scripting language associated with ToolBook (Asymetrix, 1989), an authoring platform for IBM PCs running Windows (this is the development environment for GAIDA). Once familiar with the resources, I chose two projects to work on, projects which were do-able in the course of a summer yet which could contribute to the mission of the branch. I evaluated XAIDA in its current form and evaluating existing GAIDA modules. In addition, I participated in the redesign of GAIDA to reflect a case-based framework, created the content for one GAIDA case, and specified the instructional design guidance to accompany three other modules.

In this report, I will focus on the two primary tasks: the evaluation of XAIDA and the evolutionary design process of GAIDA. I will describe my efforts, and explain how they fit into the context of the ongoing efforts of the branch.

2. Evaluation of XAIDA

HRTC will be evaluating the use of XAIDA in various contexts. It was suggested to me that I evaluate this software from the perspective of an instructional designer, in order to provide formative feedback to Mei Technology, the contractor for the software. To this end, I decided to implement the same lesson content (end-to-center band marking color code for electrical resistors) in two formats: as a lesson generated by XAIDA, and as a stand-alone lesson for GAIDA.

Currently, two manuals and a fairly detailed example lesson are available for XAIDA. One manual is provided for students running the lesson, the other is a tutorial aimed at designers. To familiarize myself with the tool, I worked through both tutorials. In the process, I took extensive notes (22 pages) on the usability of the program and the support it provides for designers. The difficulties I

faced and the recommendations I made as a result have been transformed into a report for Mei Technology. In the section below, I discuss some of the theoretical perspectives which I applied to my evaluation.

2.1 Formative Evaluation

The types of questions that can be asked to examine support for computer use include: is the interface consistent? can computer novices negotiate the interface? is there sufficient support for "power users?" does the software warn about or provide opportunities to recover from drastic changes? and so on. Questions about support for domain activities include: is the user empowered to accomplish the task in ways that are both familiar yet efficient and effective? can the user make productive use of the tool to accomplish familiar tasks? can the user extend the use of the tool to new tasks? In my evaluation of XAIDA, I attempted to address these types of issues.

Shneiderman (1987) presents a model of user behavior known as the *syntactic/semantic model*. It proposes two kinds of knowledge: *syntactic*, which refers to precise and somewhat arbitrary bits of information about device-dependent details, and *semantic*, which has to do with general concepts independent of any particular system. The *semantic* knowledge is separated into task concepts and computer concepts; both sets of concepts consist of objects and actions. In order to effectively complete a task, there needs to be a mapping between semantic task and computer concepts. *Syntactic* knowledge is obtained through rote memorization, while semantic is acquired by meaningful learning.

Kieras and Polson, in their *cognitive complexity theory*, describe three categories of user knowledge: job situation, how-to-do-it, and how-it-works (Polson, 1987). These are similar to Shneiderman's categories, but the labels are somewhat more intuitive to understand. *Job situation* knowledge and *how-to-do-it* knowledge are closely related: the former refers to knowing how to do the task independent of system characteristics, while the latter is knowing how to use the system to accomplish the familiar task. *How-it-works* knowledge is a user's understanding of how the overall system functions.

These categories can be applied to XAIDA. *How-it-works* deals with learning to use the software itself. *How-to-do-it* focuses on how to use a tool to accomplish a class of specific tasks: in this case, designing a lesson. In the context of XAIDA, *job situation* knowledge could be one of two things: instructional design or knowledge decomposition. Most software designers can assume that users know the underlying task, that is, that they are familiar with the *job situation* knowledge. This assumption is not appropriate for XAIDA, since the purpose is to support novices in the domain.

Researchers at IBM, led by John Carroll, have studied people learning to use software and software manuals (Carroll, 1989; Carroll et al., 1985; Carroll & Mazur, 1984; Carroll et al., 1988). When using self-study manuals for software training, people often avoid reading because the manuals are too long, unnecessarily repetitive, and contain too many technical details. Instead, users set up goals for

themselves and explore the system in terms of these goals; that is, they exhibit problem-solving behavior. Most authors of manuals assume that the reader will follow the directions exactly and never make any mistakes; because of this assumption, there are very few error recovery procedures in the manuals. In summary, manuals traditionally are written with the implicit assumption that the primary goal of the user is to *learn the system*; the Carroll group found that most users just want to use the system to *accomplish a task* (Carroll and Mazur, 1984).

This points precisely to the frustration I experienced with XAIDA. I wanted to accomplish my own self-selected task, but nothing in the manuals supports such an effort. In addition, the interface and terminology are hard to decipher and often hard to use. The interface is hard to use, input mechanisms are inconsistent and buggy. The software uses unfamiliar terms, assuming too much domain knowledge. There is no support for error detection and recovery in the manual. In addition, task support (for design) is very limited, both in the software and in the manuals. I was not able to accomplish a task (and neither was I able to learn the system).

In short, the program was difficult to use. Although my content was specified beforehand, and although I had already created this lesson once, I was not successful in building the lesson with XAIDA. In addition, I was disappointed with the instruction it delivered. I recommend that the delivery mechanism be replaced entirely. XAIDA should serve as a knowledge specification tool. It should create well-specified and well-designed "instructional resource" objects that reside in a library. These resources could then be accessed by the delivery mechanism, according to parameters specifying the learning situation.

3.0 GAIDA

3.1 History of GAIDA

From 1991 to 1992, Dr. Robert Gagné served the Laboratory as a National Research Council Senior Fellow. Gagné is well-known for his Nine Events of Instruction (Gagné et al., 1992), a model of lesson delivery which has been used with great success and is widely taught at universities. Dr. Gagné envisioned a system where instructional design novices in the Air Force would examine cases representing good instructional design practices, cases built around the Nine Events. As a result, the novices would be more effective in their own instructional design. To fulfill this goal, Gagné defined the framework of GAIDA and directed its development at HRTC. The underlying philosophy can be found in the following quote:

> "The GAIDA approach assumes that novice designers are capable of understanding concepts of this sort in their precise meanings. The further assumption is made that directions of this sort can be 'followed', in the sense that concrete instances of such an abstraction as 'relevant knowledge' can be identified and selected from the domain of the equipment data base being dealt with. ... GAIDA represents an approach to delivery of instruction embodied in a computer-based lesson on how to design instruction" (Gagné, 1992, p. v).

Initial Development

As a result of the earliest efforts, GAIDA consisted of three products.

The first is a videotape consisting of 3 distinct segments. In the initial segment, Gagné introduces and explains the Nine Events. After this, a model lesson is presented, showing an instructor teaching the procedure used for handcuffing. Ultimately, Dr. Gagné provides voice-over commentary on the instructional design while replaying clips from the lesson video, describing where each of the Nine Events is used in the lesson.

In addition to the videotape, two software products have been created using ToolBook. One is a functional check of the electrical circuits of the gun in the F-16 aircraft, a procedure using a checklist. The second represents a procedure that must be memorized: conducting a certain pulmonary examination. In both programs, guidance to the instructional designer is embedded alongside the text of the lesson. The content of both lessons was specified by Dr. Gagné; programmers included Larry Whitehead and Steve Hancock. Each of these is a single "book", combining text of a lesson with advice to designers.

In the pulmonary lesson, the user (assumed to be a novice instructional designer) is called upon to make a duplicate of the lesson content, with the aid of some tools embedded into the lesson. The intention was to provide an opportunity for practice for these instructional designers.

Along with these finished products, a few additional scripts had been written for eventual integration into GAIDA. The types of learning represented are identification (naval ranks), verbal knowledge (the history of the Air Force), and rule learning (simplifying algebraic expressions).

Early Evaluations

The F-16 lesson has been examined by subject matter experts and instructional designers in a formative evaluation (Gagné, 1992). Results indicated that users found the level of guidance to be meaningful and helpful. Users had no difficulties in understanding the instructions, and were able to use them to design their own lessons. They considered the resulting lessons to be satisfactory for training new specialists.

At an more informal level, I assessed the programs both in terms of usability and content. In my opinion, while the ideas behind these products are sound, the initial implementation fell somewhat short of intentions. Readers become confused; it is hard to distinguish between the lesson and the commentary on the lesson, for they are closely intertwined. In addition, the practice contained in the pulmonary example is both awkward and unrealistic. It is not a real design situation; instead, the user merely copies and pastes from one book into another. Better support tools could be created in ToolBook; however, this authoring program is not used extensively by Air Force designers, so this type of practice may be superficial at best. Instead, it seemed more appropriate to redesign the case material presentation.

3.2 GAIDA: Redesign

After identifying some of the problems associated with the original computer-based implementations of GAIDA, and comparing these with the videotape component, I became convinced that a new model was needed. Fortunately, there was agreement among other members of the branch about this need. As a result, I proposed that a new model be used to direct the implementation of GAIDA. I was primarily responsible for designing this new model.

In contrast to the two software portions of GAIDA, I felt that the videotape represented a good model that could also be implemented in software. In addition, practice should be provided in the design and planning of a lesson, not in the use of any particular authoring environment or computer program. The latest model for the GAIDA architecture is shown in Figure 1.

In this model, a GAIDA case consists of two parts: a stand-alone lesson (which could be viewed independently of GAIDA), as well as instructional design guidance related to that lesson. These can be referred to as "the lesson" and "the design commentary". Users of GAIDA will first view information about the Nine Events of Instruction; after this, they would choose between a list of cases. While viewing the case, users would read commentary on how the lesson was designed and how it incorporated the Nine Events; in addition, they could switch to the actual lesson to experience the event in the context of the full lesson.

3.3 GAIDA: implementation

After developing a new design, I implemented the materials for one GAIDA case.

Stand-alone lesson on resistors

Based on the advice of Dr. Gagné, my first step in developing a case was to design the stand-alone lesson. Such a lesson represents the core of the case. Classification was chosen as the learning type in order to round out the suite of lessons. In a classification task, the learner first makes a discrimination or judgement, and then takes an action based on that decision. We determined to use a concept in basic electronics, that of converting the color code on a resistor into its numeric equivalent. One resulting action could be to set the range switch on an ohmmeter to ensure accurate readings. The content for this lesson came from two sources: the course materials for the Common Electronic Training Program (CETP), which is conducted at the USAF Technical Training School at Keesler Air Force Base by the 3410 Technical Training Group, and various electronics textbooks (Horn, 1990; Loper, Ahr, & Clendenning, 1979). The plan of instruction and course chart in the Air Force course materials provided a baseline for determining prior knowledge (Event 3 of the Nine Events) and experience of the target audience.

Although the content of my lesson was similar to that found in the course materials and the textbooks, it differed in at least two significant ways: through the explicit incorporation of the Nine Events of Instruction at the design stage, and by the varied range of practice opportunities afforded by

GAIDA Design Framework

Section 1:   The reader is presented with a general
             introduction to the Nine Events of Instruction,
             to the software, etc.

Section 2:   User is allowed to make a choice: which
             type of lesson do you want to see?

Section 3:   User reads the Instructional Design Guidance (case
             commentaries). This is approximately 11 pages long,
             and contains text specific to lesson. In addition, it
             allows user to "switch over" to the actual designed
             lesson.

Section 4:   Summmary, Conclusion, etc.
             Option to link to Design Template

Figure 1: GAIDA Design

the computer. A textbook contains a limited number of problems. Typically, the problem take the following form: "Suppose you have a resistor with the following marking: brown black yellow and gold; what is the value of the resistor?". Textbooks are usually black and white , so if an image appears with the problem statement, it is black and white with names of colors labelling the bands. With the computer based lesson, these and other limitations can be avoided.

To achieve Event 6, I provided two opportunities for practice in my lesson. In the first, the trainee is shown a resistor with colored bars. The trainee enters the numeric equivalent into a field. Feedback (Event 7) provides an evaluation of the correctness of the response along with the correct answer. Trainees can generate as many problems as they like by clicking on a button on the screen. When they are satisfied with their ability to solve the problems, they can go to the next page. The values for the resistors are generated randomly, so each use of the program will contain different problems.

The second practice opportunity requires the reverse application of the rule. The computer randomly generates a resistance value and the user is required to "color in" the bands on a resistor to match that number. Feedback takes the form of right or wrong, along with a hint as to which band is in error. Again, users are allowed to generate as many problems as they wish to view. In both settings, hints are available upon user request; these hints consist of the shortcut for the color code along with a small image which names the bands of the resistor.

For Event 8 (Assess Performance), there are two testing situations. In first one, the user drags a series of resistors into correct order. Feedback indicates which are in error; the student is not allowed to progress until all are in correct order. The second situation requires the user to set a range on an ohmmeter based on the value of a resistor.

3.4 Other GAIDA work

Instructional Design guidance for resistor lesson

After specifying the content for the resistor lesson, I created the GAIDA commentary. The commentary consists of a description of how each of the Nine Events of Instruction was incorporated into the lesson. Particular attention is paid to the demands of a classification lesson.

GAIDA driver design document

My implementation used the model developed at the beginning of the summer. Another case, the naval rank lesson, was simultaneously implemented by Mr. Larry Whitehead during the summer. Problems of integration and standardization became apparent. Thus, near the end of my tenure, the design was further refined, and I wrote a design document for this new model. (This is the design shown in Figure 1).

Instructional design guidance

Finally, I wrote (and extracted from other documents) the instructional design advice for the four modules which will be incorporated into GAIDA.

3.5 GAIDA: Future Work

Shortcomings

The stand-alone resistor lesson has not been evaluated, in terms of ease of use or content coverage. However, since its primary purpose is to serve as a design example and not as actual instruction, this task is not a priority. It is more important to evaluate the commentary on the lesson.

Similarly, the GAIDA interface has not received user testing. A help function has not been implemented. Testing may result in further refinements, such as a different mechanism for displaying the sample lesson, or may point out a need for additional features, such as a notebook.

Currently, GAIDA does not provide an opportunity for novice designers to practice. Appropriate practice would be to use the Nine Events (e.g., design or plan a lesson). We intend to incorporate an environment wherein the users can 'sketch out' their design, by describing the text and graphic content of their lesson, lesson flow, etc.

Future Work

In addition to usability evaluations, GAIDA should be evaluated as to its effectiveness. One approach would be to compare designers who use this case-based method to those who have access to the Gagné, Briggs, and Wager book; or to those who have received formal training on the ISD model. The groups would be required to design a lesson on some specified material. The designs would be evaluated in terms of quality by blind reviewers. We would hypothesize that those using GAIDA would develop better overall designs.

Variations on this study include: matched pairs in terms of subject-matter expertise (e.g. maintenance, electronics, and so on) across conditions, have subjects design a lesson for a learning type found in GAIDA and one not found in GAIDA. Another potential hypothesis is that GAIDA users might be more efficient designers.

The Armstrong Laboratory is planning to conduct these studies in collaboration with Dr. Gros of the University of Barcelona in 1993.

4. Summary

XAIDA and GAIDA represent different approaches to advising instructional designers. XAIDA attempts to shield Air Force designers from the need to know principles of instructional design. GAIDA, on the other hand, presents the designers with a particular model and expects them to understand it and incorporate it into their own lessons. Furthermore, XAIDA represents a particular theory of learning and knowledge structure; GAIDA, on the other hand, represents a model of lesson delivery.

While they are difficult to directly compare, it is possible to use them to shed light on one another's strengths and shortcomings. It is my opinion that the lessons delivered by XAIDA are lacking many of the features advised in GAIDA. At best, it only focuses on the middle events. In addition, the

delivery mechanism is severely limited. It is tied directly to the method used to input the knowledge, although some customization via parameters is possible. XAIDA does not advise at all about parameters. GAIDA, on the other hand, may be too optimistic – it provides advice but not really support, and support may be more essential for some personnel.

The intention of AIDA is to combine a variety of tools in order to provide support for novice instructional designers. The complementary strengths and weaknesses of XAIDA and GAIDA reinforce the need for such a structure. XAIDA focuses on a subset of the events of instruction: those of presenting the material and providing learning guidance. Instructional designers will need advice on the other aspects of delivering a lesson, thus GAIDA will play a role. As a minimum, what is still needed is a good delivery mechanism.

During the course of my summer research tenure, I was able to apply a broad range of my own expertise. I was able to combine knowledge of human factors and interface issues with instructional design expertise in order to evaluate XAIDA. Furthermore, I again used my instructional design expertise as well as my skills in software development to refine the design of GAIDA, implement a case for it, and write instructional design commentary for other cases.

References

Andrews, D.H. & Goodson, L.A. (1980). A comparative analysis of models of instructional design. *Journal of Instructional Development*, 3(4): 2–16.

Asymetrix. (1989-91). ToolBook Software Construction Set for Windows. Computer program.

Carroll, J. M. 1989. An overview of minimalist instruction. IBM Research Report #1035.

Carroll, J. M., Mack, R. L., Lewis, C. H., Grischkowsky, N. L., and Robertson, S. R. 1985. Exploring exploring a word processor. *Human-Computer Interaction*, 1, 283–307.

Carroll, J.M., and Mazur, S. A. 1984. Learning to use an office system with an on-line tutorial. IBM Technical Report #47740.

Carroll, J. M., Smith-Kerker, P. L., Ford, J. R., and Mazur-Rimetz, S. A. 1988. The minimal manual. *Human-Computer Interaction*, 3, 123–153.

Gagné, R.M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart and Winston.

Gagné, R.M. (April 1992). Tryout of an organizing strategy for lesson design: Maintenance procedure with checklist. Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate (AL-TP-1992-0016).

Gagné, R.M., Briggs, L.J., & Wager, W.W. (1992). *Principles of instructional design* (4th ed.). New York: Holt, Rinehart and Winston.

Hancock, S.W. (1991). Prototyping GAIDA: A learning and developmental process. Summer Research Report.

Hickey, A.E., Spector, J.M., & Muraida, D.J. (1991). Design specification for the Advanced Instructional Design Advisor (AIDA). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate (AL-TR-1991-0085).

Horn, D.T. (1991). *Basic Electronics Theory with Projects and Experiments* (3rd Ed.). Blue Ridge Summit, PA: TAB Books.

Loper, O.E., Ahr, A.F., & Clendenning, L.R. (1979). *Introduction to Electricity and Electronics*. Albany, NY: Delmar Publishers.

Merrill, M.D., Li, Z., & Jones, M.K. (1990). Second generation instructional design (ID-2). *Educational Technology*, 30(2).

THE ASSESSMENT OF
BIOLOGICAL VARIABILITY

Robert Craig Kundich, M.A.
M.S. /Ph.D. student
Department of Biomedical Engineering

University of Tennessee at Memphis
899 Madison Avenue, Suite 801
Memphis, TN 38163

THE ASSESSMENT OF
BIOLOGICAL VARIABILITY

Robert Craig Kundich, M.A.
M.S. /Ph.D. student
Department of Biomedical Engineering
University of Tennessee at Memphis

Abstract

The enhancement of the operational envelope of high performance aircraft has led to intensified physiological stresses being applied to the aviator. Much emphasis has been placed on combating and understanding the effects of sustained, elevated +Gz stress. While Combat EDGE has sought to extend the level of G tolerance, the G-LOC effort has attempted to understand the mechanisms of gravity induced loss of consciousness. A more complete analysis of the autonomic nervous system (ANS) under gravitational stress may aid in enhancing operator capability in hostile environments, and contribute to the G-LOC and Combat EDGE efforts. Computer algorithms were programmed using LabVIEW on a Macintosh IIfx computer to assess ANS regulation of biological variability (heart rate, blood pressure, and respiration). Implementation of these algorithms will include the assessment of biological variability in individuals during varying workloads, an orthostatic tolerance test, and potentially during human centrifuge studies.

THE ASSESSMENT OF BIOLOGICAL VARIABILITY

Robert Craig Kundich

INTRODUCTION

During a hostile engagement, the modern aviator may be required to operate under sustained, high +Gz acceleration. With a normalized curve of G tolerance among the aviator population (10), some aviators may not be able to compensate for the sustained gravitational stress. The next generation of aircraft are destined to enhance the magnitude and duration of +Gz acceleration. Certainly, engagements can ferret out those that have the capability to achieve excellence and withstand the ever increasing physiological stressors. However, such a Darwinian selection process would be costly to both man and machine.

In an effort to reduce mortality and maximize performance, the United States military has a long standing tradition of attempting to reduce the medical risk to aviators. Certain abnormalities are cause for rejection from flight training or removal from high performance aircraft. Yet for all of the effort expended, there still appears to be an unacceptable loss of aviators and aircraft.

To maximize aviator survivability and performance, some of the Armstrong Laboratory Crew Technology Division research efforts have focussed on Combat EDGE, G-LOC, and spatial disorientation. Of potential importance to the Combat EDGE and G-LOC efforts is how the Autonomic Nervous System regulates the peripheral circulation during gravitational stress. Although quite contested, it has been suggested that individuals who are highly conditioned may exhibit an inadequate response to +Gz and have altered HR variability under varying workloads (13).

Without countermeasures a relaxed individual will pass out at +5 Gz (10). However, the current generation of countermeasures can provide sustained protection up to +9 Gz (10). Regardless of the countermeasures involved, certain individuals are more susceptible to G-LOC that others, leading to a basic question of, "What factors imbue an individual with G tolerance?"

Heart rate variability is an appropriate place to initiate a comprehensive study of an individuals ANS, as it should theoretically be a sensor of all impingements on the circulatory system. Standards for the recording of an ECG signal have existed for over 50 years (2,12,19,25). However, there has been relatively

little standardization of the newest methodologies such as signal averaged ECG, and the assessment of heart rate variability. The purpose of this Summer research project was to develop an understanding of the current methods of HR variability analysis and develop the software to implement the findings at AL/CFTF for the assessment of HR, blood pressure, and respiratory variability.

BACKGROUND

Biological Variability

Although often described as beating in a normal rhythmic pattern, giving rise to such terms as normal sinus rhythm, the human heart exhibits a quantifiable beat to beat variability. There is also variability throughout the day, referred to as the circadian rhythm. By quantifying the standard deviation of the R-R intervals, some insight can be gained into the variable nature of the heart beat intervals. The use of modern mathematical tools has led to the ability to map the time domain signal into the frequency domain. It is the frequency domain that has provided the most insight into the beat to beat variability of the heart.

Data in the frequency domain is typically analyzed in terms of a power spectrum. Total power for R-R interval variability is measured as the integral from 0.01 Hz to 0.4 Hz. Power is also measured in three specific frequency ranges; 0.01-0.04 Hz, 0.04-0.15 Hz, and 0.15-0.4 Hz. These frequency ranges are influenced by neurohumoral, sympathetic plus parasympathetic influences, and parasympathetic influences, respectively.

Orthostatic tolerance is clinically assessed using a standing test or a tilt table test. This test is designed to evaluate an abnormal response to a shift in body fluids. Blood pressure and heart rate are assessed while the subject is in a supine position and again while in the upright position. A drop in systolic pressure of greater than 25 mmHg or diastolic drop of greater than 10 mm Hg is abnormal (1). These changes in pressure are primarily due to a drop in peripheral resistance (23). A variety of factors impinge on the regulation of blood pressure upon standing including; the baroreceptors, catecholamines, the renin-angiotensin-aldosterone system, and vasopressin release (1). All of these factors can have an influence upon the heart rate in addition to the BP. However, heart rate may not always be altered.

The measurement of HR and BP variability, should not be construed as the only method to assess the ANS. Many drugs influence ANS performance. Various neuropathies can also severely impinge on the adequate functioning of the ANS. Therefore, neurologists have an extensive protocol to establish the location of ANS disorders. Such a neurological profile may need to be considered to gain a more accurate representation of who may have reduced G- tolerance.

ECG Signal

The electrical signal recorded and referred to as the electrocardiogram is a function of the depolarization and repolarization of the heart. Two types of standard ECG's are typically conducted a 3 lead or a 12 lead. For heart rate variability analysis, the lead most typically used is the chest lead (Lead II). Before the electrical activity is recorded, it follows a treacherous path through many noise producing agents and signal strength reducers. The tissues, such as lung, bone, muscle, and fat act as resistors which serve to attenuate the signal and change its form. Intense muscular activity, power surges, and 60 Hz interference can all cause an increase in the noise picked up by the electrodes (8). If there is not a solid connection additional interference is encountered

BP Signal

A noncontinous BP measurement can be made using a sphygmomanometer cuff and listening for the sound of blood. Additional approaches exist for the continuous measurement of blood pressure. An indwelling arterial catheter can be used to get a continuous reading, although distortion from air bubbles and "catheter whip" are possible (24). Another approach is to use a noninvasive device to measure arterial pressure. This is the basis for the Finapress device, which uses a fingertip BP monitor.

Analog Prefilters

The purpose of prefiltering the analog signals in biological variability analysis is to avoid the introduction of errors due to aliasing. The ideal prefilter for variability studies should have minimal attenuation in the frequencies of interest (passband, 0 to 125 Hz) (25), and create a large signal attenuation at higher frequencies. The cutoff and stopband are two critical frequencies in the design of a prefilter. The cutoff frequency (Fc) is the point where signal attenuation is initiated (an attenuation of -3 dB). The

stopband frequency (Fs) is where attenuation is typically greater than -70 dB. The ratio of stopband frequency to cutoff frequency is often calculated to give a measurement of the rolloff from passband to stopband. In a lowpass filter, the passband region is before the cutoff frequency. In the passband region, minimization of ripples is important.

There are currently four popular types of analog filters; Butterworth, Chebyshev, Elliptic, and Bessel. The Butterworth offers a maximally flat passband, while the Chebyshev has the steepest transition from passband to stopband. The Bessel filter has a maximally flat time delay. The Butterworth filter trades off everything else for maximal flatness. However, the Butterworth has less than ideal phase characteristics. If 1 dB ripples in the passband are acceptable, then a Chebyshev filter can be used. It allows greater sharpness at the knee (3 dB), but has less than ideal phase characteristics. However, no matter what design is employed, flatness of the passband is never really possible. Design criteria calls for very tight tolerances of the resistors and capacitors (1%), but some ripple still occurs. The elliptic or Cauer filter is a variation of the Chebyshev filter and is easy to implement with CAD. The elliptic filter is indicated where the shape of the waveform is important.

Sampling Frequency

The sampling frequency is determined by multiplying the stopband of the prefilter by 2. This value is then multiplied by the number of channels to get the total sampling rate. Sampling at twice the stopband frequency allows for 1 bit resolution to be achieved if the stopband is at -72 dB. To simplify the calculation of storage space requirements, the sampling rate should also be a power of two.

Through the use of signal decimation techniques, it is possible to alter the sampling rate. For a signal sampled at 256 Hz and 1024 Hz, the errors for sampling a 100 BPM heart rate signal range from 2.34 to 0.59 msec, respectively. However, it is possible that both sampling rates may give the same information about the heart rate intervals. The only way to assess this is by doing a signal to noise analysis.

Pizzuti et al. did an analysis of sampling rate and ECG using a system with ± 9 bits accuracy (20). Sampling at 250 Hz rather than 500 Hz is acceptable for large interval measurements such as R-R interval (20). However, sampling at 125 Hz introduced significant errors in interval measurements (20). Merri et al.

studied signal to noise ratios of R-R interval measurements at various sampling frequencies (16). The highest quality signal was at 256 Hz, followed by sampling at 128 Hz and 64 Hz (16). Again, supporting the contention of sampling at greater than 250 Hz. It was suggested that the more variable the R-R intervals are, the greater the noise and therefore the higher sampling frequency.

To insure the integrity of the data analysis, it is appropriate to conduct a signal to noise test on each set of data. This will insure that no errors are introduced in the analysis through improper sampling rate selection.

Peak Detection Algorithms

Once an analog ECG signal has been pre-filtered and converted into digital form by sampling at an appropriate rate, the R peaks of each QRS complex must be detected. Over thirty seven different methodologies have been published concerning the detection of QRS peaks (8). One of the simplest approaches to use, if the ECG signal has a stable baseline, is a peak threshold detection of a negative derivative. When the derivative signal crosses the zero line (corresponding to a R peak) the peak detector records the time of occurrence. If the data is not free of noise or unstable this will not work accurately.

Friesen presents two approaches that appear to be very appropriate for the analysis of the ECG during centrifugation studies (8). The first QRS detector, referred to as AF1, uses an algorithm that includes the first derivative and amplitude in the analysis (8). This method is exceptional at the detection of a QRS complex even with high levels of EMG noise. The second filter is referred to as DF1 and uses digital filtering methods (8). This method is a good choice for peak detection, under a variety of noisy environments. The reader is referred to Friesen et al., for a more detailed discussion of QRS detection (8).

Instantaneous Heart Rate Determination

Although the heart rate power spectrum can be computed using just the R-R intervals, it tends to create errors in the spectrum. To obtain a more precise power spectrum representation, the heart rate needs to be sampled at regularly spaced intervals; in other words an instantaneous heart rate. Five methodological approaches exist for the determination of instantaneous heart rate (5, 6). Factors to be considered include correcting for the delay encountered in some of the methods and the errors induced by each methodological

approach. De Boar et al. present three methods; spectrum of intervals, spectrum of counts, and spectrum of inverse integrals (6). Berger et al. present two methods (5). The first method involves taking the reciprocal of the R-R interval duration and passing it through a low pass filter window with a width of 2 times the sampling frequency.

An equivalent approach to determine instantaneous rate is to let:
$R_i = f_r * n_i/2$ where R is the instantaneous rate, f_r is the sampling rate and n_i is the sum of all fractions of intervals covered by that window (5).

Instantaneous Rates of Other Biological Signals

Often data from other biological signals are gathered in conjunction with the ECG. These have included such factors as blood pressure and respiration. These signals present a unique processing challenge when compared to that of the heart rate signal. Typically, sampling would need to be reduced to 4 Hz for ECG, BP, and respiration. This would allow us to evaluate the interval from d.c. to 2 hz. However, Berger et al. suggest that the power spectrum is only accurate to the sampling frequency divided by 4 (5). This would produce a spectrum from 0 to 1 Hz. The major dilemma regardless of the final width of the power spectrum desired, is the manyfold reduction in sampling rate. In the case of the current protocol the signal was reduced from a sampling rate of 1024 Hz to 4 Hz, a factor of 256. Digital filtering techniques are necessary when the magnitude of biological signals is reduced.

Decimation is downsampling by lowpass filtering followed by compression (17). Berger et al, use a 31 point low pass filter followed by a compression of a factor of 4. This approach will be used to go from 1024 Hz-> 256 Hz -> 64 Hz-> 16 Hz -> 4 Hz. An ideal low pass filter is of the form (17):
$$h_{lp}(n) = sin(w_c n)/pn$$

Computing the Power Spectrum

The power spectrum can be computed after the biological signal is in digital form and preferably converted to an instantaneous sampling rate. Nine methods exist for the generation of power spectrums from time domain signals (11). One of these approaches has three variations, bringing the total to eleven possible methods with their own distinct advantages and drawbacks (11). The power spectrum is an

assessment of the relative contribution of each frequency to the signal. For instance, if a signal repeats at exactly a 2 Hz period, such as a sine wave a single sharp peak will appear. However, if a signal is composed of repeating components of various periods of duration, multiple peaks are possible.

The generation of a power spectrum is typically based on two approaches; the autoregressive approach and the fast Fourier transform (FFT) based approaches. Press et al provide C algorithms for both approaches (21). The AR approach has the advantage over the FFT in that it does not require a power of two for the number of samples (11). Additionally, less samples are required (11). On the other hand, more care must be employed in the use of the AR approach. The proper order of the AR approach can be determined using the Akiake Information Critewria (AIC) (15). The easiest method is the FFT based, periodogram approach. Minimal computational and software design strain is necessary to generate the power spectrum. A variation on this theme is the Blackman-Tukey approach, employed by Cohen's group (5). A practical application of the AR approach is in Bartoli et al (5) and an additional algorithm is in Marple (15).

Presentation of Power Spectrum Results

After the selection of a method for computing the power spectrum, the results must be presented in an undestandable format. On a two dimensional graph of the power spectrum, the frequency is along the xaxis. Conversely, the x-axis may also be in units of cycles/beat if the De Boar approach is followed. This may also then be converted to equivalent Hertz by multiplying the x-axis values bi the mean R-R interval. The values along the y-axis are unitless for normalized data and in units of $msec^2/Hz$ if not normalized. If the De Boar method is used the units are $msec^2/cycles/beat$. Normalization of the power spectrum occurs when the power spectrum values are divided by the mean of the interval squared. The next factor to consider is whether the amplitude or power spectrum is being plotted on the y axis. Amplitude is simply the square root of the power spectrum.

Once the various methodologies are sorted out, it is necessary to interpret the power spectrum density plots. Of considerable importance is if the subjects were under controlled breathing or free breathing. Controlled breathed at a regularly spaced interval will cause a large spectral peak at the

frequency of breathing. Whereas free breathing causes a more diffuse peak. This points out the stunning influence of vagal innervation on heart rate.

Resolution of the spectral plot is very important. The maximum resolution is a function of T^{-1}, with T equaling the period. For 5 minutes the resolution would be 1/300 Hz or .0033 Hz. A longer period is necessary to generate smaller interval values. Thus, resolution is enhanced significantly as the period is increased.

Before the methods section is presented, two of the published approached to the assessment of biological variability will be presented. Additonal approaches are also available (4,14,22).

De Boar Method of Power Spectrum Analysis (7)

1. For a period of 512 beats the R-R interval duration (I), Systolic Pressure (S), Diastolic Pressure (D),

   Mean Pressure (M), and Pulse Pressure (P) will be determined.

2. Determine the means for the above variables.

3. Subtract the d.c. component

4. do an FFT

5. Power Density Spectra will be FFT^2

6. Smooth with a 31 point triangle window. this is also known as a Bartlett window.

7. Convert to equivalent HZ by multiplying by the I\mean R-R interval. If this is not done the units are

   cycles/beat

8. Letting B = S,M,D,P then compute cross spectra for each blood pressure variable with heart rate.

   $C_{BI}(f)=X^*_B(f)^*X_I(f) = L(f)-iQ(f)$

   L(f) is called the cospectrum and Q(f) is the quadrature spectrum. Remember that these have been

   smoothed.

9. Compute the smoothed estimator of the phase spectrum for each variable:

   psi(f)=arc tan (-Q(f)/L(f))

   psi(f) over the range of ± 180 degrees

The Berger et al. method (5)

1. Determine R-R intervals

2. Plot as I/In for all n

3. Pass through a boxcar (rectangle window) with the following parameters:

$$W(f)=\{[\sin(2p f/f r)]/[2p f/f r]\}^2$$

where fr is the new sampling frequency and $2/fr$ is the width of the window in Hz. The value for fr is usually taken to be 4 Hz.

4. Compute mean and remove d.c. component

5. Compute $(FFT)^2$

6. apply a $1/W(f)$ correction to the $(FFT)^2$

7. Compute the square root of the values in number 6 to get amplitude

8. Plot

METHODS

Equipment

The data necessary to conduct the heart rate variability studies may be collected on a wide range of ECG machines. It is essential that each machine be tested using a device similar to that recommended by the AHA (12). The impedance of each electrode should also be evaluated to insure a solid connection, thus minimizing interference. Multiple types of continuous BP monitors are available. The Finapress (Ohmeda, Englewood, CO), is one of the more widely used machines. A perusal of the literature and manufacturer specifications must be done to ascertain any restrictions placed on the analysis by the internal signal processing of the chosen BP machine.

All data gathered from the BP and ECG machines should be passed through an analog prefilter prior to conversion to digital form. The prefilter stopband for each channel is based on the stopband of the channel requiring the fastest acquisition speed. These filters may be commercial or custom built. If custom built, a function generator and oscilloscope are necessary to evaluate their effectiveness.

The conversion of the analog signal to digital requires one of the many available A/D boards. Two stipulations exist. First, it must be compatible with the chosen computer system. Second, its throughput must

be fast enough to transfer data from all of the signals being processed. As a minimum, it should have 8 channels, 12 bit resolution, and a throughput of 10,000 Hz. Although, 16 channels, 12 bit resolution, and a 100,000 Hz would allow for the expansion of the number of signals simultaneously evaluated.

Many computer solutions exist for the implementation of the software analysis including; mainframe, DEC, IBM, and Macintosh. The AL/CFTF system was a Macintosh IIfx 40 Mhz computer running the LabVIEW instrument system. Software development can occur with packages such as LabVIEW, LabWINDOWS, or Hyperception. These environments make programming easier. However, experienced C programmers can use Think C, Borland C++, or Microsoft C/C++ version 7.0 to achieve the programming objectives.

As an intermediate step, a Honeywell or Raycal tape recorder can be used to store the analog signals prior to processing.

Recommended analysis protocol using the Fast Fourier Transform

1. Use a Macintosh computer, with at least 32 Mb of RAM, running LabVIEW

2. For a 4 channels system the A/D board should have a throughput of at least 10,000 Hz

3. A prefilter is required for each channel. These may be bought at considerable expense, or a lower quality filter can be built without too much difficulty. These filters should have a frequency cutoff of 100 or 125 Hz and a stopband that is approximately -70 dB.

4. **Reality check.** Compute the characteristics of the filter using a function generator and an oscilloscope. Compute the Nyquist frequency (2 times the cutoff frequency), and 2 times the stopband frequency. Two times the stopband frequency will be the sampling rate for each channel.

5. Send known analog signal (such as sine or cosine signals with high and low amplitudes and periodicities) through filters for A/D conversion.

6. **Reality Check.** Sample at different frequencies 64, 128, 256, 512, 1024 Hz and compute errors of magnitude and phase.

7. Send known ECG signal through filters and A/D to the peak detection algorithm (8).

8. **Reality Check**. Over plot peak detection with original signal. Check by visual inspection for the alignment of the peak detection with the R peak.

9. Send a known signal simultaneously through 4 channels or however many will be used in the acquisition.

10. **Reality Check**. Compute errors between channels due to sampling time selection. It has been referred to as non-significant. However, to obtain the most accurate results possible, this error needs to be quantified. Note, if the A/D board is designed specifically for simultaneous sample and hold this step is not necessary.

11. Digitize the data from the ECG, BP, Thermistor, and other channels.

12. Store data in binary format as this is the most compact.

13. Determine ECG peaks and R-R intervals, verifying by visual inspection.

14. Determine systolic, diastolic, pulse, and mean values along with their intervals, verifying by visual inspection.

15. **Reality check**. Send known signals through the blood pressure detection device and make sure measurements are valid.

16. Compute instantaneous rates using the Berger et al. method (5), the equivalent of a low pass window boxcar.

17. Make instantaneous plots of heart rate, SBP, DBP, PP, Mean Pressure.

18. **Reality check**. Also send a known signal through the boxcar filter and the steps that follow below.

19. Subtract D.C. component from data

20. Develop two sets of data and pass one through a Hanning window.

21. **Reality check**. Look at side lobe leakage from the application of the Hanning window.

22. Compute $(FFT)^2$ for the data with and without the Hanning filter and get total power for each case.

23. **Reality check**. Compare the known signal power spectrum results with those that have been published by Berger et al. (5).

24. Correct for Hanning window by multiplying by [total power of no Hanning/ total power Hanning].

25. Correct for boxcar by multiplying by 1/ W(f)

26. Plot PSD from 0 to 0.4 Hz. This is not normalized.

27. Take square root. Plot amplitudes for second set of graphs. This is not normalized.

28. Normalize 26 and 27 for a second set of graphs.

29. Letting B = S, M, D, P, where; S is the systolic pressure, M is the mean pressure, D is the diastolic pressure, and P is the pulse pressure, then compute cross spectra for each blood pressure variable with heart rate (7). I is the R-R interval measurement

$$C_{BI}(f) = X^*_B(f) * X_I(f) = L(f) - iQ(f)$$

L(f) is called the cospectrum and Q(f) is the quadrature spectrum.

30. Compute the squared coherence estimator (7).

$$k^2(f) = [(L(f))^2 + (Q(f))^2] / [C_{BB}(f)C_{II}(f)]$$

31. Compute the estimator of the phase spectrum for each variable (7):

$$f(f) = arc\ tan\ (-Q(f)/L(f))$$

f(f) over the range of ±180 degrees

32. Make cross spectra, phase, and coherence plots of SBP, DBP, PP, and Mean Pressure with heart rate..

SUGGESTIONS FOR IMPROVING THE ANALYSIS

Visual Instruments

The LabVIEW visual instruments (VI's) can undergo substantial improvement in the future. The algorithms for the autoregressive approach can be added to the series of VI's most probably through the linking of C code modules. Additionally, enhancing the memory allocation and the overall design of the VI's should make them more efficient in the data analysis. With the additional digital signal processing board, real time analysis may be a possible program extension. Algorithms for the peak detection under noisy conditions should be added. The R-R interval algorithms can be optimized to reduce the memory requirements of the programs and enhance efficiency.

Filters

If biological variability studies are to be conducted on a continuous basis, it may be appropriate to purchase commercially manufactured filters. Selecting a filter with less than 1% attenuation between 0 and

100 Hz, a cutoff of 125 Hz and a stopband of less than 256 Hz will allow each channel to be sampled at less than 512 Hz. This will reduce the storage requirements by 50% and reduce the computational burden by 50%.

Memory

When doing the variability analysis, lack of memory is crippling in the LabVIEW environment. The programs ran more comfortably (i.e. larger data files could be manipulated) with 20 Mb of RAM memory. However, the programs still had problems with very large data files. Therefore, 32 or 64 Mb of high speed RAM would be more appropriate, and compensate for inelegant programming.

LabVIEW

The LabVIEW programming environment is very efficient for rapidly developing applications. However, it has certain annoying problems that the manufacturer may consider rectifying. They are:

1. LabVIEW needs on-line help in addition to the wire paths

2. A 1-2 page form with pictures of the icons, their name and what page in the manual to find them, would be appropriate for novice system users.

3. Icon rotation. Due to the rigid wiring structure, the wires can become very messy.

4. A function to zoom in or out, to find cursor position in a large diagram (S. Stranges suggestion)

5. Easy development of a multiple windowing system like Microsoft Windows, the DEC Workstations, or even the Macintosh when its not running LabVIEW.

REFERENCES

1.   Aminoff MJ. Evaluation of autonomic function. *AAEM Didactic Program*, 1990; 23-27.

2.   Bailey JJ, Berson AS, Garson A, et al. Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. *Circulation*, 1990; 81(2): 730-739.

3.   Bartolli F, Baselli G, Cerruti S. AR identification and spectral estimate applied to the R-R interval measurements. *Int J. Bio-Medical Computing*, 1985; 16: 201-215.

4. Baselli G, Cerutti S, Civardi S, et al. Spectral and cross-spectral analysis of heart rate and arterial pressure variability signals. *Computers and Biomedical Research*, 1986; 19: 520-534.

5. Berger RD, Askelrod S, Gordon D, Cohen RJ. An efficient algorithm for spectral analysis of heart rate variability. *IEEE Trans BME*, 1986; BME-33(9): 900-904.

6. DeBoer RW, Karemaker JM, Strackee J. Comparing spectra of a series of point events particularly for heart rate variability data. *IEEE Trans BME*, 1984; BME-31(4): 384-387.

7. DeBoer RW, Karemaker JM, Strackee J. Relationship between short-term blood-pressure fluctuations and heart-rate variability in resting subjects I: a spectral analysis approach. *Med. & Biol. Eng. & Comput.*, 1985; 23:352-358.

8. Friesen GM, Jannett TC, Jadallah MA, et al. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans BME*, 1990; 37(1): 85-98.

9. Furlan R, Dell'Orto S, Crivellaro W, et al. Effects of tilt and treadmill exercise on short-term variability in systolic arterial pressure in hypertensive men. *J. of Hypertension*, 1987; 5(suppl5): S423-S425.

10. Harding RM, Bomar JB. Positive pressure breathing for acceleration protection and its role in prevention of in-flight G-induced loss of consciousness. *Aviat. Space Environ. Med.*, 1990; 61:845-9.

11. Kay SM, Marple SL. Spectrum analysis- a modern perspective. *Proc. IEEE*, 1981; 69(11). 1380-1419.

12. Kossmann CE, Brody DA, Burch GE, et al. Recommendations for standardization of leads and of specifications for instruments in electrocardiography and vectorcardiography. *Circulation*, 1967; 35: 583-602.

13. Krock LP. Enhanced aerobic capacity and increased acceleration stress: Is their interaction a problem for pilots? *Physiologist*, 1992; 35(1): S139-S142.

14. Lipsitz LA, Mietus J, Moody GB, Goldberger AL. Spectral characteristics of heart rate variability before and during postural tilt. *Circulation*, 1990; 81: 1803-1810.

15. Marple L. A new autoregressive spectrum analysis algorithm. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 1980; ASSP-28(4): 441-454.

16. Merri M, Farden DC, Mottley JG, Titlebaum EL. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Trans. BME*, 1990; BME-37(1):99-107.

17. Oppenheim AV, Schafer RW. *Discrete-time signal processing*. Prentice-Hall, 1989.

18. Perini R, Orizio C, Baselli G, et al. The influence of exercise intensity on the power spectrum of heart rate variability. *Eur. J. Appl. Physiol.*, 1990; 61:143-148.

19. Pipberger HV, Arzbaecher RC, Berson AS. Recommendations for standardization of leads and of specifications for instruments in electrocardiography and vectorcardiography. *Circulation*, 1975; 52:11-31.

20. Pizzuti GP, Cifaldi S, Nolfe G. Digital sampling rate and ECG analysis. 1985 *Journal unknown*.

21. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical recipies in C.* 1988. Cambridge.

22. Rottman JN, Steinman RC, Albrecht P. Efficient estimation of the heart period power spectrum suitable for physiologic or pharmacologic studies. *Am. J. of Cardiol.*, 1990;66:1522-1524.

23. Sprangers RLH, Wesseling KH, Imholz ALT, Wieling W. Initial blood pressure fall on stand up and exercise explained by changes in total peripheral resistance. *J. Appl. Physiol.*, 1991; 70(2):523-530.

24. Webster JG. *Medical Instrumentation Application and design*. Houghton- Mifflin. 1992.

25. Wolf HK, Graystone P, LeBlanc AR, et al. Suggested minimum performance characteristics of data acquisition instrumentation in computer-assisted ECG processing systems. *J. Electrocardiology*, 1976; 9(3):239-247.

MULTIPLE PHYSIOLOGICAL MEASURES OF PERFORMANCE ON A MULTIPLE TASK BATTERY

Scott H. Mills
Graduate Student
Department of Psychology

University of Oklahoma
455 W. Lindsey Street
Norman, OK 73026

MULTIPLE PHYSIOLOGICAL MEASURES OF PERFORMANCE
ON A MULTIPLE TASK BATTERY

Scott H. Mills
Graduate Student
Department of Psychology
University of Oklahoma

Abstract

Most studies of human performance and workload have been limited to single or dual task paradigms. Recent advances in desktop computers have made possible the development of multiple task batteries, which are more applicable to many real-world tasks. One such battery, the Multi-Attribute Task Battery (MATB), was investigated. Psychophysiological measures combined with performance data may provide useful information about workload, especially when used with multiple task batteries. The Psychophysiological Assessment Test System (PATS) can simultaneously present tasks and collect numerous channels of psychophysiological data. A demonstration study was designed to evaluate the feasibility of using the MATB and the PATS together to explore performance and workload. Two subjects performed different combinations of tasks from the MATB while EEG and heart rate data were collected by the PATS. Subjective ratings of workload and performance measures were also recorded. Both the MATB and the PATS were found to be powerful, user-friendly systems that should lead to new avenues of research in human performance and workload.

MULTIPLE PHYSIOLOGICAL MEASURES OF PERFORMANCE ON A MULTIPLE TASK BATTERY

Scott H. Mills

Human performance on standard laboratory tasks has long been used as a vehicle for understanding many aspects of workload. However, most studies have been limited to the use of single or dual task paradigms. While these research settings do allow for the investigation of many well-defined aspects of human performance, the findings can be difficult to apply to tasks performed routinely in the workplace. Most of these tasks involve several different processes, instead of just one or two. For example, driving a car requires simultaneously performing a large number of sub-tasks. Another difficulty of applying laboratory tasks exists because they are usually relatively short in duration (usually less than five minutes). This presents a problem with comparisons to real-world tasks that are often much longer in duration.

One solution to the problem of short, simple laboratory task paradigms is the use of multiple task batteries. Improvements in the capabilities of desktop computers over the last several years have made the development of these batteries more feasible. One such battery recently developed at NASA is the Multi-Attribute Task Battery (MATB; Comstock & Arnegard, 1992). This battery is capable of presenting multiple tasks on the same computer screen. With the use of programming scripts, the experimenter can vary the combination of tasks, their difficulty, and their duration. The MATB includes a system monitoring task, a tracking task, an audio communications task, and a resource management task (See Figure 1). The flexibility of this battery makes possible research for both single and multiple task performance within both short and long duration test periods. Short task periods allow greater control in the exploration of task effects (such as type of task and workload level) on psychophysiology and performance, while longer tasks allow greater ability to explore the effects of stressors and adaptive responses to workload.

Another feature of the MATB is its inclusion of a subjective workload rating instrument, the NASA Task Load Index (TLX; Hart & Staveland, 1988). The TLX is a multiple subscale measure that can be presented at any time during performance of the MATB. Thus, it can be easily used as a complement to performance data in the assessment of workload (See Figure 2).

Figure 1. The MATB screen.

Figure 2. The NASA-TLX workload rating screen.

Workload assessment usually takes the form of one of the following categories: task performance, subjective measures, or psychophysiological indices. Of these, task performance is most often cited in the literature. One disadvantage of looking only at task performance is that actual effort or workload may not always be reliably estimated. This is because subjects may be recruiting different amounts of resources in order to attain similar performance; that is, some subjects may have to work harder to achieve the same performance as others. An additional attempt to quantify differences in workload is the use of subjective rating scales. Subjective data, however, may be subject to bias and expectations. It also can be compromised by problems associated with the delay involved between actual task performance and responding to the ratings.

Psychophysiological methods of assessment have several advantages over performance and subjective measures. These measures can be made continuously, and therefore reveal changes in patterns of psychophysiological activity over time. Another advantage is that the collection of these measures does not intrude on the subject's primary task.

The management of such varied sources of task, subject, and operational environment data does pose as a significant problem. One important advance in addressing this problem has been the development of the Psychophysiological Assessment Test System (PATS), which has been developed by the psychophysiological laboratory of the Armstrong Aerospace Medical Research Laboratory at Wright-Patterson AFB. The PATS was designed to be a comprehensive test system for the presentation of tasks and for the measurement of psychophysiological data. It can record and analyze numerous channels of such data. It also has the capability of communicating with other computer systems such as those generating tasks, and other data output devices (such as digitized or analog signals for sensors or controllers of various sorts). These features make the PATS an ideal system for use with multiple task battery research.

The combination of the MATB and the PATS provides a flexible laboratory system for investigating human performance and workload under a wide variety of experimental designs. The MATB can present multiple combinations of different types of tasks for both short-interval and long-interval durations. The PATS can record up to sixteen channels of physiological data and provide several levels of analysis. Both these systems are designed to optimize usability by the researcher. The PATS system has a user interface based on HyperCard and requires no programming. The MATB includes an easy-to-use setup screen and requires only the writing of script files consisting of times and one or two-word event codes. This ease of use makes the combination of these two systems a very practical, useful research environment that may provide new avenues to investigate complex areas of human performance and workload.

The purpose of this study was to evaluate the feasibility of using the MATB and the PATS together for conducting research in human performance. An experiment was designed to present

subjects with different combinations of tasks and workload levels on the MATB and to record psychophysiological responses with the PATS for comparison to performance and subjective data.

Method

Subjects

Two males subjects, ages 25 and 21 participated in this study.

Apparatus

The Multi-Attribute Task Battery (MATB) provides the subject with a complex combination of tasks. The battery presents the tasks running in individual windows on a standard PC screen and comprises the four tasks described below.

System Monitoring task. This task is presented in the upper left portion of the screen, where two lights and four dials are presented. The subject's job is to monitor these objects for changes from their "normal" condition. When a change is detected, the subject must click with the mouse on the aberrant object. In the normal condition the left light is on and is green. A change is defined as when this light turns off. The right light is normally off, and a change in it is defined when it turns red. The pointers in each of the dials normally fluctuate around the center mark, staying within one mark below and one mark above. A change in a dial is defined as when the pointer begins fluctuating beyond one mark on either side of the center.

Tracking task. This task is presented in the upper center of the screen. The subject controls the target (the circle) with a joystick and must attempt to keep the target in the center of the window. This task has three difficulty levels. Tracking can also be switched to an automatic mode, in which the target stays near the center of the window without any input from the subject. During this mode the word "AUTO" appears at the bottom of the window.

Resource Management task. This task appears in the window at the bottom center of the screen. A diagram representing fuel tanks and pumps connecting them is presented. The objective of this task is to maintain the level of fuel in the upper two tanks at 2500 gallons. These tanks are drained of fuel at a constant rate; therefore the subject must turn on and off pumps to control the transfer of fuel from the lower tanks. Pumps that are turned on are green in color, and pumps that are off are clear. The subject can toggle pumps by using the mouse to click on them. The levels of fuel in four of the tanks are indicated by the number below each tank. The other two are supply tanks and have unlimited capacity. The flow rate for each pump is indicated in the "Pump Status" window in the lower right portion of the screen. When a pump if off, this rate is displayed as zero.

Any fuel pump can be made to fail by the program, at the discretion of the experimenter. When this occurs, the pump turns red and stops transferring fuel. The subject has no control over these failures and cannot control the pump during a failure. Pumps can also be repaired by the program. When this happens, the pump is returned to the off condition; therefore the subject must actively turn it back on if desired.

Communications task. The communication task is presented in the window in the lower left portion of the screen. This window displays the subject's call sign and the settings for four aircraft radios. The subject must listen through headphones for messages directed to his or her callsign and follow instructions to change one of the radio frequency settings to a specific new one. This task was *not* used in the present study.

Scheduling Window. This window at the top right of the screen allows subjects to anticipate when the performance of either the tracking or communication task will be required. It was *not* used in the current study.

Rating Scale Screen. This screen (See Figure 2) displays the NASA Task Load Index (TLX). It can be programmed to appear at any time during performance of the battery. While the TLX is presented, the battery is frozen and resumes when the subject exits the TLX screen or when the maximum time for completing the ratings (one minute) has been exceeded. The TLX measures subjective workload by requiring the subject to rate their experience of workload on six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The subject uses the mouse to select a rating value for each subscale. The analysis of these selections is explained in the Results section.

The special nature of this research requires a powerful and flexible system for the collection and management of psychophysiological data. The Psychophysiological Assessment Test System (PATS) uses a Macintosh IIx computer equipped with commercially available data acquisition boards and a high-capacity data storage device. The PATS has the ability to simultaneously present tasks and collect both performance and psychophysiological data for periods of up to several hours. It can digitize up to 16 channels of analog data at a 1000 Hz rate for each channel, and also provides several analysis tools such as digital filtering, spectral analysis, etc. The PATS can be set up to present tasks and collect data in a wide variety of combinations through the use of menu-driven screens. For the present study, only the psychophysiological collection and analysis components of the PATS were used.

Task Conditions

Eight task conditions were developed, using various combinations of the Monitoring, Tracking, and Resource Management tasks (See Figure 3). Three of these conditions consisted of one of the tasks being performed by itself. One condition was a baseline, in which no tasks were performed, but the subject viewed the MATB screen and sat quietly. Three conditions were designed to require low, medium, and high workload demands of the subject. Finally, one condition was designed to demand an extremely high level of workload on all three tasks for four minutes (overload), and then to require only a moderate level for 4 minutes (medium).

Task Condition Summary

| Condition | Task Status | | | Duration |
	System Monitoring	Tracking	Res. Management	
Baseline	automatic	automatic	automatic	4 minutes
Monitoring	2 stim/min	automatic	automatic	4 minutes
Tracking	automatic	Low	automatic	4 minutes
Res Management	automatic	automatic	On-No Failures	4 minutes
Low	2 stim/min	automatic	On-No Failures	4 minutes
Medium	2 stim/min	Low	On-No Failures	4 minutes
High	4 stim/min	Medium	1.25 Failures/min	4 minutes
Overload/	14 stim/min	High	8 Failures/min	8 minutes
Medium	2 stim/min	Low	On-No Failures	

Figure 3. Task Conditions.

Procedure

This study consisted of two phases, a training phase and a test phase. Each phase lasted approximately two-hours, and each phase occurred on a different day. Subjects were trained and tested individually. For the training session, the subject was provided with a demonstration of the Monitoring, Tracking, and Resource Management tasks, as well as the TLX rating screen. The subject was encouraged to ask questions and to practice each procedure. Then each subject practiced the MATB for a total of 72 minutes. This practice session comprised a combination of the (non-baseline) 4-minute conditions described above. The overload condition was not included. Each four-minute trial was followed by the TLX rating screen.

For the testing phase, each subject was instrumented to permit recording of several standard measures used in human performance research in workload assessment. These were heart rate, eyeblink, and four channels of EEG data. The subject was seated in an electrically shielded, acoustically insulated chamber, equipped with mouse and joystick controls and a window for screen viewing. Each of the four-minute task conditions of the MATB was presented in a random order, with the exception of the baseline condition, which was always presented as the fifth trial. After each four-minute condition, the subject was allowed three minutes to return to baseline. During this break period the door to the experimental chamber was opened, and the experimenter told the subject which of the tasks he would be performing in the next presentation. The subject was not told the difficulty of the upcoming tasks, only which ones would be required. Following the last four-minute condition, the eight-minute "overload-medium" condition was presented .

The NASA Task Load Index was presented immediately following completion of each task condition (during the break period--for no more than one minute) on the same screen. Following the eight-minute condition, the subject was asked to complete the TLX for the "first half" of the task, the overload portion. After completing the last TLX rating, the subject exited the chamber and performed the Aftermat program. Aftermat is a component of the TLX that requires the subject to make factor comparisons of each pair of subscale titles. Its results are used in the computation of the TLX weighted workload scores.

4-min	4 min	4-min	4-min	4-min baseline	4-min	4-min	4-min overload	4-min medium
TLX	TLX	TLX	TLX		TLX	TLX		TLX

Time →

Results

This study was designed as a pilot project to determine the feasibility of using the MATB and the PATS systems together. Because of the exploratory nature of the design and the low number of subjects, only descriptive statistics will be reported.

The following data were collected by the MATB: for Monitoring, response time for correct responses, number of false alarms, and number of missed stimuli; for Tracking, Root Mean Square error (RMS); and for Resource Management, mean deviation. This deviation was computed by the following method: Each time the status of a fuel pump changed, the absolute deviations from the criterion fuel level for both upper tanks were measured and added together. The mean of these numbers, over a four-minute trial, is reported as the mean deviation. These MATB performance data are presented in Figure 4. Several characteristics can be noted: First, performance for all single-task conditions was superior to that for any dual or mutli-task conditions. In contrast, task measures did not necessarily degrade linearly with the hypothesized increases in workload. Performance for the final medium condition was expected to return to approximately the level for the first medium condition. However, both Tracking and Resource Management appear to be anomalous. Tracking performance for both subjects was better than expected, and Resource Management performance appears to be worse. It should be noted, however, that the mean deviation reported for Resource Management probably reflects degraded performance from the Overload condition. This is because when a subject changes his or her performance, it is not reflected in the measure of fuel deviation until the tanks "catch up" with that change.

As previously mentioned, the psychophysiological measures were heart rate, eyeblink, and four channels of EEG data. The EEG data were recorded at locations O_z, P_z, C_z, and F_z (See 10-20 Electrode Placement System, Jasper, 1958). Spectral analysis was performed and mean power was calculated for each of the following bandwidths: Delta (up to 3 Hz), Theta (4-7 Hz), Alpha (8-13), and Beta (14-25). Ordinarily, an experimenter can use the PATS to subtract eyeblink signals from those of the EEG. Because of a configuration error, eyeblink data for this study was not usable. Therefore, it is not reported, and the reported EEG data includes some eyeblink artifact. For heart rate, the mean interbeat interval (IBI) for each task condition is reported. These data are presented in Figures 5 and 6. The eyeblink artifact makes interpretation of EEG data difficult. Heart rate, however, did appear to correspond to some differences in task condition. For both subjects, heart rate was relatively slow during the baseline condition and relatively fast during overload.

The NASA TLX program produces a weighted workload score for each rating performed. This score is computed from weighted subscale ratings. These scores are presented in Figure 7. These data are interesting because for certain conditions, they parallel some of the psychophysiological data. For example for the low, medium (first), and high conditions, Subject 1 rated medium as having the least amount of workload and had the lowest heart rate of those three conditions. Subject 2 rated these conditions in a more linear manner, and his heart rate followed the same trend (with heart rate increasing as reported workload increased).

Subject 1

Task Condition

Variable	Tracking	Res Mngmnt	Monitoring	Low	Medium	High	Overload	Medium*
Tracking[1]	13.5				35.15	37.32	61.79	29.07
Res Mngmnt[2]		193.2		405.22	439.94	610.78	529.39	674.74
Monitoring[3]			1.32	4.39	2.49	3.02	3.31	3.21
Mon. Errors[4]			2	7	7	7	7	0

Subject 2

Task Condition

Variable	Tracking	Res Mngmnt	Monitoring	Low	Medium	High	Overload	Medium*
Tracking[1]	8.29				19.88	25.78	39.68	16.62
Res Mngmnt[2]		72.95		95.41	159.03	225.95	699.19	555.02
Monitoring[3]			1.53	3.07	2.17	3.6	2.45	2.04
Mon. Errors[4]			0	2	1	1	3	1

[1]The Tracking dependent variable is Root Mean Square Error (deviation from center of tracking target in pixel units).
[2]The Resource Management variable is the mean deviation from criterion "fuel" level.
[3]The Monitoring dependent variable is response time, in seconds.
[4]Monitoring Errors include false alarms and missed stimuli.

*Four-minute medium condition that followed the Overload condition.

Figure 4. MATB Performance Means.

Subject 1

Task Condition

	Baseline	Tracking	Res Mngmnt	Monitoring	Low	Medium	High	Overload	Medium
Delta--EEG	1.78	1.77	1.39	1.63	2.12	1.54	1.52	1.65	1.54
Theta--EEG	.87	.80	.74	.76	.96	.78	.85	.77	.88
Alpha--EEG	.71	.66	.54	.58	.67	.51	.70	.51	.64
Beta--EEG	.24	.27	.25	.25	.46	.26	.50	.26	.32
IBI--heart	1116.9	1028.9	966.89	956.4	914.5	948.0	919.85	960.76	976.55
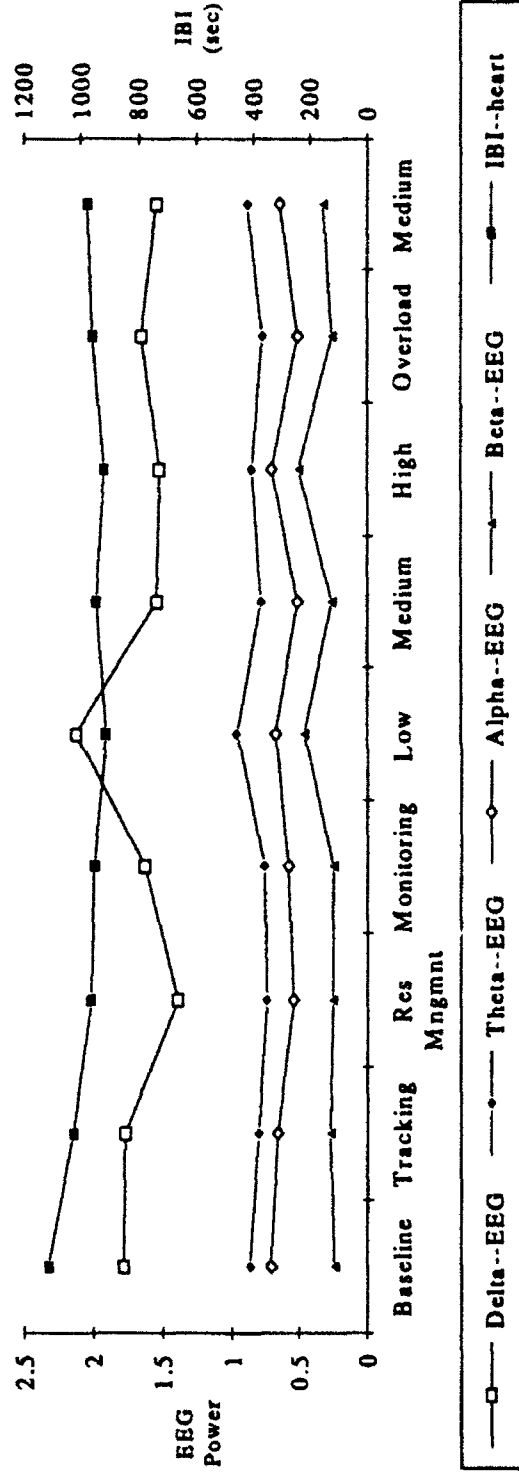
Figure 5. Psychophysiological Measures for Subject 1.

18--13

Subject 2

Task Condition

	Baseline	Tracking	Res Mngmnt	Monitoring	Low	Medium	High	Overload	Medium
Delta--EEG	1.32	1.32	1.31	1.32	1.39	1.46	1.45	1.57	1.57
Theta--EEG	1.33	1.18	1.34	1.20	1.53	1.29	1.53	1.35	1.35
Alpha--EEG	1.18	1.03	.87	.96	1.07	.90	1.06	1.01	1.00
Beta--EEG	.59	.42	.37	.45	.34	.40	.36	.46	.49
IBI--heart	933.6	871.8	904.3	951.8	931.8	854.7	832.2	862.1	847.7
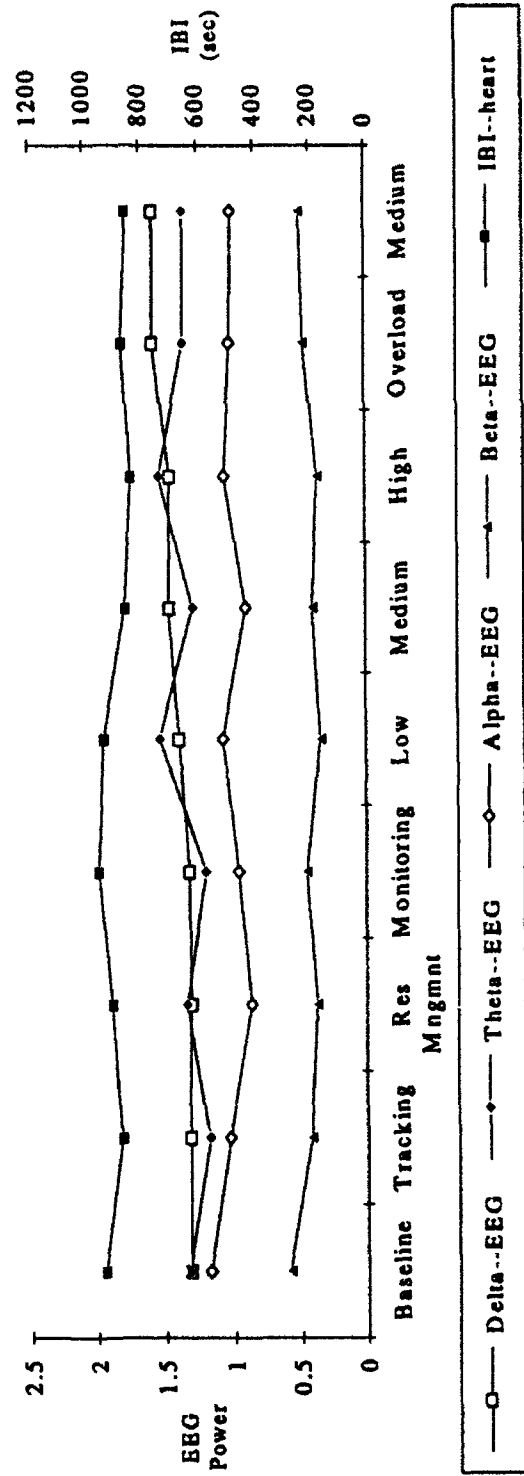
Figure 6. Psychophysiological Measures for Subject 2.

18--14

	Tracking	Res Mngmnt	Monitoring	Low	Medium	High	Overload
			Task Condition				
Subject 1	2	4.67	5.33	33.67	24	36.667	59
Subject 2	10.67	59.33	13	57	61.33	71.67	91.33
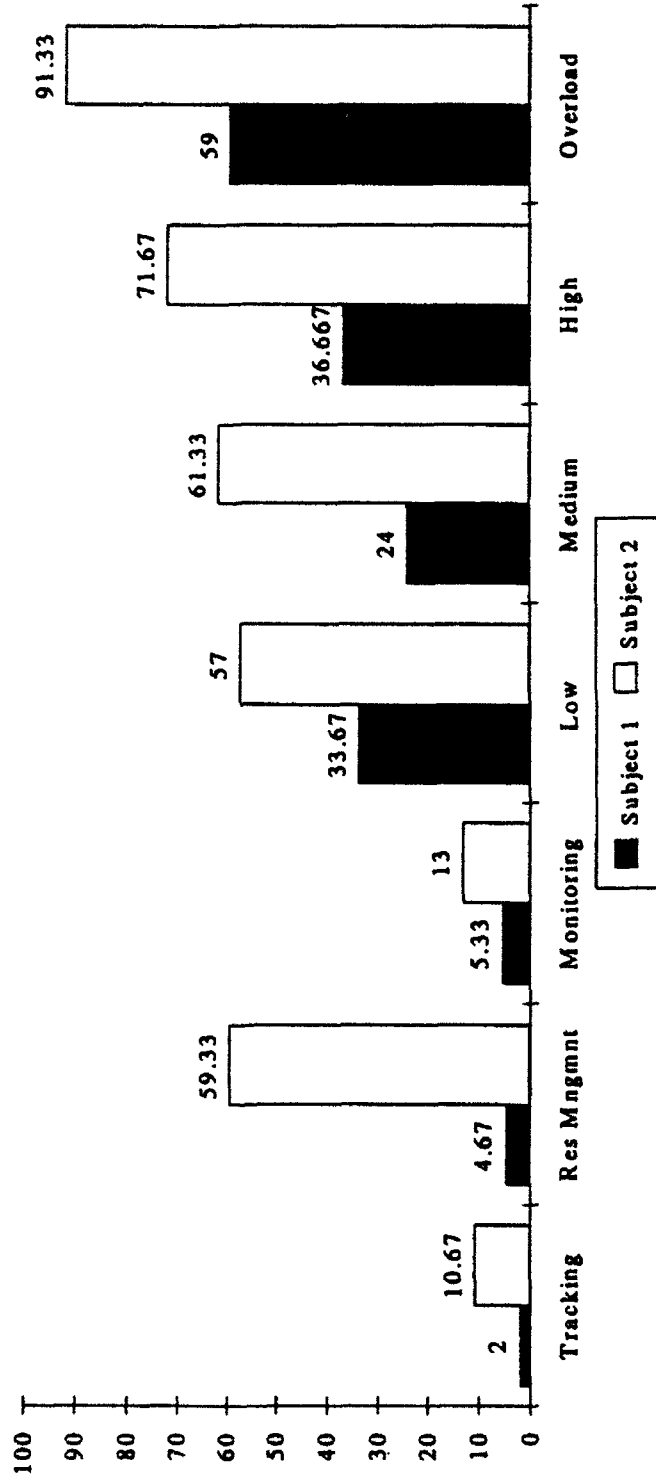
Figure 7. NASA TLX weighted workload scores.

18--15

Discussion

This study included only two subjects because of its purpose—to examine the feasibility of using the PATS and the MATB concurrently in the investigation of human performance and workload. This small sample size makes the formation of any conclusions about the observed data difficult. Regardless, some general trends in the data are encouraging. It is also possible that any meaningful patterns in the EEG data may have been obscured by interference from eyeblink signals.

Nonetheless, the primary purpose of the study was accomplished. Both the PATS and the MATB performed very successfully and their combination provided a robust, reliable research environment. The MATB was easily configured for a variety of task combinations. It presented tasks and recorded both performance and subjective workload rating data reliably. The PATS also proved to be flexible and easy to use, for both psychophysiological data collection and analysis. The user interface is based on Hypercard and uses a hierarchy of menus. This interface is very user-friendly and provides extensive error-checking of user input. The PATS also manages and analyzes large amounts of data with a maximum of ease for the user. These features make the PATS a powerful and versatile laboratory system with the potential of facilitating a wide variety of psychophysiological research.

Multiple task battery research combined with the analysis of multiple psychophysiological measures may lead to better understanding of many facets of human performance and workload. The PATS and MATB systems appear to be valuable tools when used both individually and in combination and should provide many new avenues of productive research.

References

Amegard, R. J. (1991). Operator strategies under varying conditions of workload. NASA Contractor Report 4385.

Comstock, J. R., & Amegard, R. J. (1991). The Multi-attribute Task Battery for human operator workload and strategic behavior research. NASA Technical Memorandum 104174.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371-375.

A NONRADIOACTIVE ASSAY FOR THE DETECTION
OF AP-1 COMPLEXES IN THE SCN OF HAMSTERS

Heather Panek
Graduate Teaching Assistant
Department of Biology

University of Scranton
Scranton, PA 18510

Final Report for:
Summer Research Program
Armstrong Laboratory

September 1992

A NONRADIOACTIVE ASSAY FOR THE DETECTION OF AP-1 COMPLEXES IN THE SCN OF HAMSTERS

Heather Panek
Graduate Teaching Assistant
Department of Biology
University of Scranton

Abstract

The primary pacemaker generating circadian rhythms is located in the suprachiasmatic nuclei (SCN) of the hypothalamus. Light and the circadian pacemaker interact to regulate a specific set of immediate early genes in the SCN that may participate in the entrainment of the circadian clock. Specifically, c-fos related proteins interact with a member of the Jun family to form the transcription factor complex AP-1 that can bind to a particular DNA sequence (TGACTCA). This DNA-protein complex interaction can be isolated in a gel mobility shift assay. To detect the AP-1 complex, a nonradioactive system, based on previously described radioactively labeled DNA probe protocols, has been researched. DNA probes labeled nonisotopically with fluorescein-11-dUTP can be detected with antibody-enzyme conjugates and enzyme substrates. This technique is desirable for several reasons including; short exposure times, long storage of labeled probes and the added safety of nonisotopic systems.

A NONRADIOACTIVE ASSAY FOR THE DETECTION
OF AP-1 COMPLEXES IN THE SCN OF HAMSTERS

Introduction

The SCN of the hypothalamus in mammals is the site of a circadian pacemaking system and is entrainable by light during subjective night. (1)   However, the specific mechanisms that set the rhythm keeping clock are unknown. It has been noted, that in response to light at appropriate circadian times, phase shifts and a corresponding increase in the mRNA of c-fos and jun-B occur.   This increase is only observed in response to light stimulation at times that also cause phase shifts. The resulting Fos proteins can interact with Jun proteins to form a transcription factor AP-1 complex.(2) This AP-1 complex contains a specific DNA binding site, TGACTCA. (3) The protein is able to bind to DNA outside of the helix without disturbing base pairing. These AP-1 complexes can be detected, after light stimulation during subjective night, by a gel mobility shift assay. (2) Use of this assay would be helpful to determine if there are AP-1 complex increases in response to stimuli other than light, such as certain drugs that alone can cause circadian phase shifting. If AP-1 is seen in response to other stimuli it will be further proof that c-fos and jun-B are part of the expression cascade in the generation of circadian rhythms However, present mobility shift assays have some disadvantages associated with them since they are based on radioactive detection systems. (2)(4)

There are nonisotopic labeling and detection systems available to replace radioactive methods, but they have not been applied specifically to mobility shift assays. Therefore an effort was made to adapt these nonisotopic kits to previously described mobility shift protocols and to determine the conditions necessary to detect AP-1 complex formation.

Gel mobility shift assays are based on the fact that the mobility of DNA complexed to a protein is greatly reduced in a polyacrylamide gel. (5) Because of this, non-bonded DNA and proteins will run, in a low percentage gel, well in front of any complexes, making the complex easily detectable.

To perform a mobility shift, DNA in short oligonucleotide form, containing the binding protein consensus site, is end labeled with terminal deoxynucleotidyl transferase, an enzyme that will repeatedly add nucleotides to the 3'OH termini of oligonucleotides. (6) Then the labeled oligonucleotide is incubated with the DNA binding protein and binding buffers. (4) The entire assay is run on a nondenaturing polyacrylamide gel and labeled DNA-protein complexes are then detected.

The first nonisotopic system tested was Amersham's ECL (enhanced chemiluminescence) 3' oligolabeling and detection system. It uses fluorescein-11-dUTP and terminal deoxynucleotidyl transferase to label the 3' end of standard oligonucleotides. A single labeling reaction labels from 25 to 100 picomoles of oligonucleotide in 60 to 90 minutes, adding

approximately 6-8 nucleotides. (7) Labeled probes can be stored in the dark at - 20 °C for up to five months. The labeled oligonucleotide is incubated with AP-1 containing SCN extract and separated on a 4% polyacrylamide gel. After electrophoresis the complexes must be transferred to a solid support. The complex is electroblotted onto a positively charged nylon membrane (Boehringer Mannheim) for one hour. The DNA of the complex is fixed to the nylon membrane by ultraviolet radiation that crosslinks thymine residues in the DNA to positively charged amino acids on the surface of the membrane. (9) After fixation, the membrane is blocked to prevent non-specific antibody-enzyme conjugate binding, a major source of high background. Then the membrane is incubated with anti-fluorescein horseradish peroxidase conjugate that will bind to the fluorescein labeled complexes. Detection occurs when the horseradish peroxidase catalyzes the oxidation of the detection substrate, luminol. Immediately following oxidation, luminol is in an excited state which can decay to the ground state by a light emitting pathway. Amersham's detection kit also contains an enhancer that magnifies the reaction 1000 fold and can be detected on blue light sensitive film, with a maximum emission of 428nm. The light emission reaction peaks after 1-5 minutes before decaying over 2-3 hours. (10) Sensitivity for these probes is 25 attomoles compared to 1-5 attomoles for ^{32}P.

The second system tested was Boehringer Mannheim's DIG

Oligonucleotide 3´-End Labeling Kit (Genius 5). The overall protocol for Genius 5 is very similar to the ECL system even though they utilize different labeling and detection reagents. In the Genius 5 system, digoxigenin (DIG), a steroid hapten, is linked via a spacer arm to a nucleotide. Oligonucleotides are labeled by the incorporation of a single DIG labeled nucleotide, DIG-ddUTP. (11) After the binding assay and electrophoresis, the DIG-labeled oligonucleotide probes are detected by an immunoassay utilizing antidigoxigenin-alkaline phosphatase conjugate. To visualize AP-1 and oligonucleotide complexed molecules. Lumigen, a component of Lumi- Phos 530, is dephosphorlyated by alkaline phosphatase forming an unstable intermediate. This intermediate decays and emits blue light in direct proportion to the amount of alkaline phosphatase present. Lumi-Phos also contains a fluorscein surfactant that emits bright yellow luminescence at 525 nm, greatly enhancing the detected signal. Lumi-Phos 530 allows multiple exposures of the blot since the light intensity increases for 7-8 hours, is maximal for 12 hours and slowly decreases over several days. (12) The Genius 5 system is reputed to have sensitivity equal to ^{32}P.

Methods and Materials

Paired SCN were removed from Syrian hamsters after two hour light stimulation, beginning at CT 19. The tissue from two hamsters was pooled into one sample, frozen on dry ice and

19-6

stored at – 80°C until sonication as described by Kornhauser et al. (2) Soluble proteins were obtained after 25 minutes of microcentrifugation at 10000g and at 4°C. AP-1 consensus site oligonucleotides were purchased from Promega. The oligonucleotides are double stranded, 21 nucleotides long and contain a single consensus site (5´-CGCTTGATGACTCAGCCAA-3´). Oligonucleotides were end labeled with fluorescein-11-dUTP and terminal transferase according to Amersham's 3´ oligolabeling and detection system. Labeling efficiency was checked by the manufacturer's rapid labeling assay. The labeled oligonucleotides do not have to be purified and are stable in a dark environment at -20°C for up to five months. Mobility shift binding assays were based on previously reported protocols. (2) (4) Ten microliters of tissue extract, 4 μl 5x binding buffer [20% glycerol, 5mM MgCl$_2$, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 50mM Tris-Cl (pH 7.5), 100 μg/ml denatured salmon sperm] and dH$_2$O to 20 μl (including labeled oligonucleotide) were incubated on ice for 15 minutes. Approximately 22ng of oligonucleotide was added and the reaction continued at room temperature for 45 minutes. Reactions were stopped with the addition of 10x loading buffer [40% glycerol, 250mM Tris-Cl (pH 7.5), 0.2% bromophenol blue]. The mobility shift was performed on a pre-run 4% polyacrylamide gel at 150 volts for approximately one hour in 25mM Tris, 192mM glycine buffer. The samples were then electroblotted onto a positively charged nylon membrane (Boehringer Mannheim) for one

hour at 90 volts in 0.5x TBE buffer. The DNA complexes were fixed to the membrane by ultra- violet crosslinking for three minutes. The membrane was blocked with Amersham's blocking agent for 30 minutes, incubated with anti-fluorescein horseradish peroxidase conjugate and washed as suggested in the ECL detection system protocol. The blot was exposed to ECL Hyperfilm (Amersham) for three hours or overnight and developed.

The Genius 5 system was used according to manufacturer's instructions which were similar to the ECL system's. A major difference with the Genius detection kit was the exposure times needed. The Lumi-Phos 530 gave off a much stronger signal for a much longer time. This cut the exposure time to minutes instead of hours and multiple exposures were possible.

Results and Discussion

There are several adjustments that must be made to the gel mobility shift assay before the switch to a nonisotopic system can occur. Both Amersham's and Boehringer Mannheim's systems require larger concentrations of oligonucleotide in the labeling reaction versus radioactive protocols. In some cases, purchased oligonucleotides did not come in great enough quantity to perform labeling reactions of optimal concentration, even though there is more than enough for radioactive methods. Another change for both kits was that the time needed for the binding assay was tripled compared to traditional protocol. (4)

The ECL system required much more fluorescein labeled probe for the binding assay than either radioactive methods or ECL literature predicted. This observation may stem from the fact that the Amersham labeled probe was not purified (per instructions). If the labeling reaction was not as efficient as expected, much unlabeled oligonucleotide could be included in the 22ng of probe used in the binding assay. Another possibility is that there is a loss of labeled oligonucleotide in the electroblot transfer and the antibody washing steps, which would result in decreased signal output. Detection times for Amersham were approximately two to three hours or overnight if desired, although no increase in detection time will be seen. Using plenty of probe resulted in a fairly strong signal and very little background. Unfortunately, the ECL detection system reaches a maximum very quickly (1-5 minutes) and decays rapidly so multiple or long exposures for weaker signals is impossible.

The Genius 5 system, which is technically almost identical to Amersham's, has a major advantage in that its detection signal has a much longer time course. This makes multiple exposures with widely varying times possible. The Genius system did not require as much labeled probe to be added to the binding assays. Approximately 8 ng were used. This is probably because the Genius instructions recommend purification of the labeled probe and include a method to estimate labeled probe yield. Being able to determine labeled probe concentration is also an advantage

over Amersham's system. High background has been a problem with the Genius systems, but it has been reduced with new blocking agents and increased suggested blocking times.

After finding the best way to apply these systems to mobility shifts there still remains the problem that no AP-1 was detected. Since this complex has been previously seen in response to light stimuli (2), there must be some technical reason AP-1 does not appear in the mobility assay. Either the protein is not being extracted from the SCN tissue or it is being degraded. Experimenting with nuclear extracts of the SCN tissue and more rigorous precautions against peptidases should produce Ap-1 protein. Also purified Ap-1 should be used in a binding assay to insure that the systems are capable of binding and detecting the complex.

If the problem of no AP-1 detection continues with the mobility shift assay, perhaps a slightly different approach is necessary. A southwestern hybridization (8) can be used. This technique is performed by running only the protein extract on a polyacrylamide gel and transferring the proteins to a solid support (western). A southern hybridization with labeled consensus oligonucleotide could then used to detect AP-1 protein.

In conclusion, with a little more work, the nonradioactive systems can be viable alternatives to the more dangerous and ultimately more expensive ^{32}P based mobility shift assays. The non- isotopic systems are faster and the five month probe storage is a real advantage. Once the nonisotopic systems have been

completely adapted for the gel mobility shift there will be no

reason to return to radioactive systems.

References

1. Meijer, J.H. and W.J. Reitveld. <u>Physiol. Rev.</u> 69, 671-707 (1989); B. Rusak and I. Zucker, <u>Physiol. Rev.</u> 59, 449-526 (1979).

2. Kornhauser, D.E. Nelson, K.E. Mayo and J.S. Takahashi, <u>Science</u> 255, 1581-1584 (1992).

3. Lee, W., P. Mitchell and R. Tjian, <u>Cell</u> 49, 741 (1987).

4. "Gel Shift Assay" Promega Technical Bulletin #106, Promega (1992).

5. Alberts, B. et al. <u>Molecular Biology of the Cell</u> second ed. Garland Publishing, Inc. NY (1989).

6. Chang, L.M.S. and F.J. Bollum, <u>Biochemistry</u> 10, 536-542 (1971).

7. "Nucleic Acid Labeling and Detection: a review of radioactive and nonradioactive technologies" Amersham International plc UK (1992).

8. Old, R.W. and S.B. Primrose, <u>Principles of Gene Manipulation</u> fourth ed. 131-133 Blackwell Scientific Publications, London (1989).

9. Church, G.M. and W. Gilbert. <u>Proc. Natl. Acad. Sci.</u> USA 81. 1991-1995 (1987).

10. Whitehead, T. P., et al. <u>Nature</u> 305 158-9 (1983); Thorpe. G.H.G., et al. <u>Clin. Chem.</u> 31. 1335-1341 (1985); Pollard-Knight, D., et al. <u>Anal. Biochem.</u> 184 (1990).

11. Schmitz, G.G., et al. <u>Anal. Biochem.</u> 192. 222 (1991).

12. Schaap, A.P., H. Akhavan and L.J. Romano. <u>Clin. Chem.</u> 35 1863-4 (1989); Martin, R., et al. <u>Biotechniques</u> 9, 762-768 (1990).

GROWTH OF MICROBUBBLES AT ALTITUDE

Joseph Pellettiere
Graduate Student
Department of Mechanical Engineering

Case Western Reserve University
Cleveland, OH 44106

GROWTH OF MICROBUBBLES AT ALTITUDE

Joseph Pellettiere
Graduate Student
Department Of Mechanical Engineering
Case Western Reserve University

ABSTRACT

The exact cause of decompression sickness (DCS) is not known. Several theories try to predict the risk of DCS through bubble growth models. These models describe the mechanics of bubble growth based on diffusion of gas from a tissue into a bubble due to a change in pressure. To assess the validity of these theoretical bubble growth equations, an in vitro experiment was performed to simulate a subject going to altitude. The experiment was set up to eliminate unknowns in the growth equations. A model altitude chamber containing water was subjected to various pressure differentials and the resulting microbubbles were observed through a microscope connected to a video camera. A computer based image analysis system was used to determine the growth rate from the video recordings. Three pressures 500 mmHg, 380 mmHg, and 250 mmHg, and 2 surface tensions 72 dynes/cm and 40 dynes/cm were used for growing bubbles. Although the solution with the lower surface tension produced more small bubbles, 5 μm in diameter, there was no significant difference in the bubble growth rate. When the data was compared to the theory, the theoretical dr/dt differed by a factor of 2 - 4. The bubbles that grew were preexisting in the fluid; only bubbles that could be seen with the microscope were measured for their growth rates. Some bubbles, 15 μm or less, did not grow in the time frame observed, while other bubbles that were slightly larger, 30 μm, did grow.

INTRODUCTION

Decompression sickness (DCS) afflicts many divers and aviators. It is attributed to an excess of nitrogen in the body that causes bubble formation because of a reduction in pressure. For a diver, this happens during his ascent from the bottom after his task is completed. While for an aviator it occurs on his ascent and duration at altitude, before the task is completed. For diving, DCS is prevented by avoiding certain supersaturations of nitrogen. This is done by the use of tables that prescribe the dive profile. Even if a diver deviates from the prescribed profile he can slow his ascent and even recompress if necessary. For the aviator, he can not avoid going to altitude so he must always accept some risk. If he shows signs of DCS, he may not be able to alter his profile and recompress by descending. To more safely plan flights, a model is needed that can predict the risk of getting DCS based upon the flight profile. This model could be used to determine how much prebreathe time would be necessary. If a model is developed and validated, it can then be used for monitoring the risk during the flight.

Because DCS is thought to be caused by nitrogen bubbles, the model should incorporate how these bubbles form and grow. Currently it is not clearly understood how bubbles form. They are thought to originate from bubble nucleation sites or preexisting bubbles. Two factors that are important for the formation of bubbles are the gas supersaturation and the surface tension (4). Tissues which become supersaturated due to the reduction in ambient pressure attempt to re - equilibrate by diffusion of gas out of the tissue and into the bubble, but this only happens if there is a bubble for the gas to diffuse into. The surface tension contributes the most to the pressure in the bubble. A high surface tension corresponds to high pressure; this hinders bubble formation. A lower surface tension will reduce this pressure and allow bubbles to form. The surface tension can be reduced from any number of surfactants in the body. The easiest places being the lungs, which are covered with surfactant and have gas exchange taking place, and in the joints, where the lubricating fluid could act as a surfactant.

After bubbles are formed, it is necessary to know how fast they will grow. This will give information as to how long it will take for the bubbles to reach a critical size where problems may occur. In research altitude chamber flights there is usually a latency period of 1/2 - 2 hours before a subject will develop bubbles and show symptoms of DCS. It is unclear whether it is the number of bubbles present, the size of the bubbles, the amount of gas which has evolved into the bubbles, or the location of the bubbles that causes DCS.

To gain a better understanding of DCS, the growth process of microbubbles will be observed. If this growth can be accurately described, maybe a relationship between bubbles and DCS could be developed to assess an aviators risk. This growth will occur at altitude because of a pressure difference.

THEORY

Diffusion of a gas is caused by a pressure gradient between a bubble in a fluid and the fluid. A gradient can be caused by the warming of the fluid; the fluid is going from a saturated state at one temperature to a supersaturated state at a higher temperature. The gradient can be caused by a decrease in ambient pressure. When the ambient pressure decreases, the partial pressure of the gases in the fluid become larger than the ambient pressure, this causes the gas to diffuse from the liquid into the bubble.

Several theories have been proposed to describe this process, but none of them have been validated. Both Van Liew's (2) and Gernhardt's (4) theory relate the growth of bubbles to DCS. Because these theories are dealing with physiological systems there are many unknowns: perfusion, solubilities, surface tensions, tissue half times, and even how to deal with the complexities of different growing medium.

Besides the uncertainties in describing the growth, there is a question of when does the growth start. The theories that describe growth do not describe how the bubbles form in the first place.

If a chamber of water is used for the growing medium of the bubbles, many uncertainties can be removed. There is no perfusion of the water; thus, there is no need for a tissue half time. The physical constants for water are known; therefore, there is no ambiguity in which constant should be used. If bubbles are present in this fluid, the chamber can be taken to an altitude where the bubbles will grow.

There will be two parts to the growth of the bubbles. The first is a Boyle's law free expansion due to the change in pressure. The next part is the growth due to diffusion of gas from the fluid into the bubble. This phase takes significantly longer. Diffusion will be affected by the

pressure, size, and the number of bubbles, while the free expansion is due solely to the change in pressure.

In order to describe how bubbles should grow, a theoretical derivation of the mechanics is included. This derivation will be used for comparing how bubbles grow in a simulated altitude chamber. In this derivation the bubbles will be assumed spherical. The volume of a sphere is:

$$V = 4\pi r^3/3 \tag{1}$$

When studying a bubble, it is necessary to know what the pressure in the bubble is. Three factors make up the pressure in a bubble, P_b; the hydrostatic pressure, or the barometric pressure, P_h, the pressure due to surface tension, $2\delta/r$, where δ is the surface tension and r is the radius of the bubble, and the pressure from the tissue elasticity, $H4\pi r^3/3$, where H is the Bulk modulus of tissue elasticity. For this derivation, the pressure from tissue elasticity will be ignored. Values of H for water were found, and when they were used in the equations they made no significance. It is assumed that H for a tissue would be less than H for water, so ignoring its affects does not alter the values given by the equations. Generally, for a DCS study, the pressure in the bubble is divided into two parts, the pressure due to nitrogen and the pressure due to the tissue gases (oxygen, carbon dioxide, and water vapor). For a model altitude chamber there are no metabolic processes to regulate the tissue gases; therefore, the pressure is:

$$P_b = P_h + 2\delta/r \tag{2}$$

The first part of the growth is a Boyle's law expansion. Boyle's law describes the behavior of an ideal gas moving from one state to another at a constant temperature.

$$P_1 V_1 = P_2 V_2 \tag{3}$$

When Boyle's law is applied to a spherical bubble and equations (1) and (2) are used for P and V an equation which describes the free expansion is derived.

$$f(r_2) = P_{h2} + 2\delta/r_2 - P_{b1}(r_1/r_2)^3 = 0 \tag{4}$$

- After the Boyle's law expansion gas will begin to diffuse into the bubble. Two equations have been derived to describe this process. One by Gernhardt and one by Van Liew. Gernhardt's equation is based on diffusion limited gas transport across the bubble / fluid interface based on a specified shell thickness.

$$\frac{dr}{dt} = \frac{\alpha D(P_t - P_b) P_s}{h(P_b - 2\delta/3r)} \tag{5}$$

α = Solubility of gas in the tissues

D = Diffusivity

P_t = Pressure of gas in the tissue

P_b = Pressure of gas in the bubble

P_s = Standard pressure

h = Effective thickness of the diffusion barrier between the bubble and the tissue

δ = Surface tension

r = Radius

While the equation derived by Van Liew describes the same process it has a slightly different form.

$$dr/dt = (\alpha D P_s)(1 - P_t/P_b)(1/r + \lambda) \qquad (6)$$

where $\qquad \lambda = \sqrt{(\alpha_b kQ)/(\alpha_t D)}$.

Because gas is diffusing into the bubble, the pressure of the gas in the tissue will be decreasing. This process can be described by doing a mole balance, (moles of gas in tissue)$_{state1}$ = (moles of gas in tissue) $_{state2}$ + (moles of gas diffused into bubbles).

Using the mole balance and the ideal gas law an equation which describes the change in pressure is derived.

$$P_{t2} = P_{t1} - NP_b 4\pi(r_2^3 - r_1^3)/3\alpha P_s \qquad (7)$$

Equations (4), (5), (6), and (7) can then be used in a computer program to simulate the growth of bubbles at altitude.

METHODS

Figure one shows a general set up of the equipment used, which includes: a model altitude chamber, a microscope, a microscope stand, a light, a camera, a T.V., and a VCR. The model altitude chamber was used to reduce the pressure. Inside the chamber was the fluid to be taken up to altitude. In this fluid there was a 100 um wire that was used for a reference for making measurements. The microscope allowed for the visualization of the bubbles as they grew.
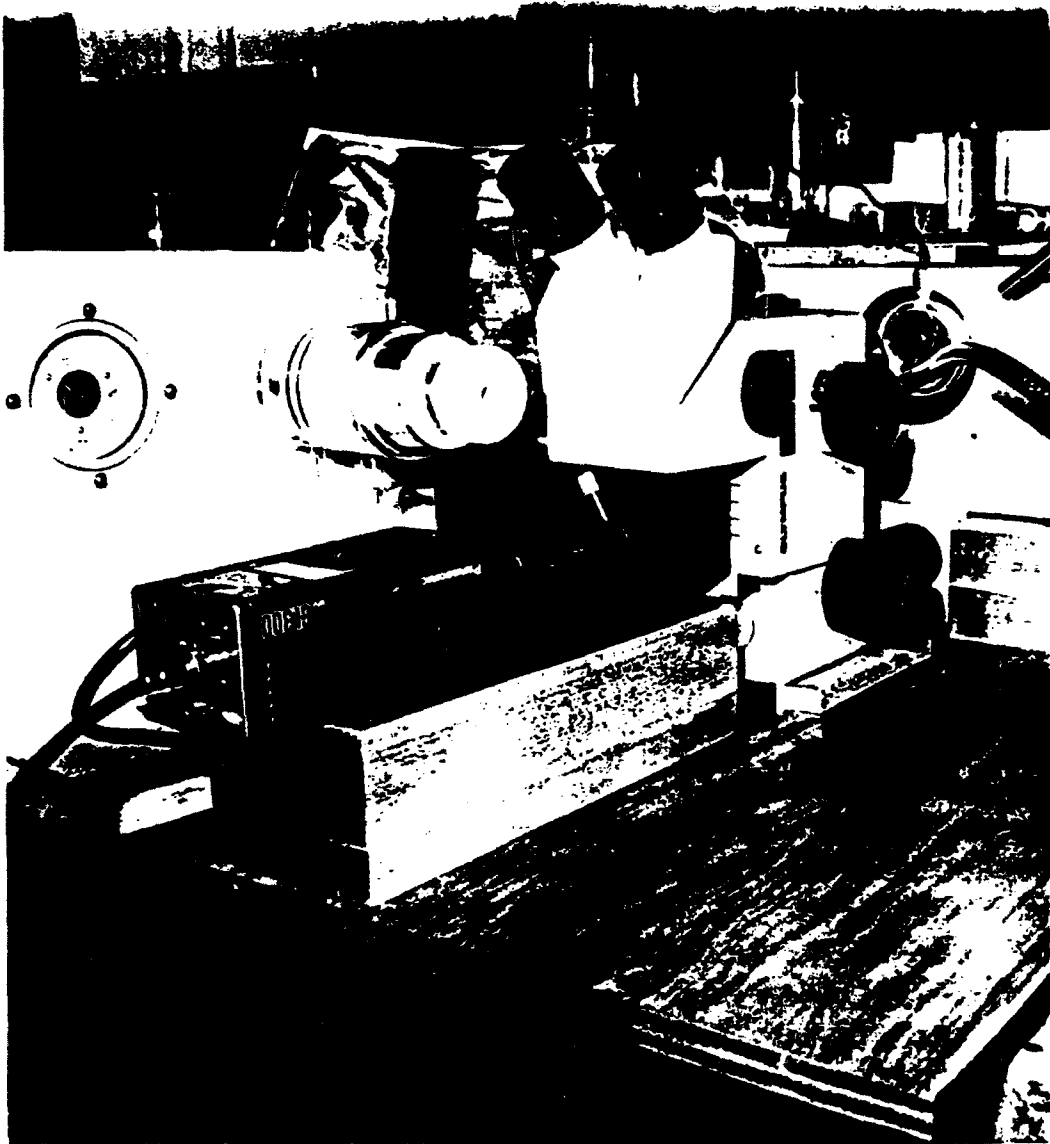
Fig. 1 Experimental apparatus used. Pressure device is on left, connected to the chamber of water. Microscope with attached camera is in center. The microscope was used to focus on a preexisting bubble, the pressure was dropped, while the camera recorded it.

Attached to the microscope was a camera, which was output to the T.V and VCR.

To record the growth rate of the bubbles the following procedure was used. The chamber was filled with the fluid to be taken to altitude. This fluid was either 100% distilled H_2O or a mixture of 80% distilled H_2O and 20% Acetone. The Acetone was added to reduce the surface tension of the water from 72 dynes/cm to 40 dynes /cm. The microscope was then focused on a bubble which would be followed as it was taken to altitude. It should be noted that all bubbles measured were preexisting bubbles stuck to the wire. The pressure in the chamber was then dropped to one of these three pressures, 250 mmHg, 380 mmHg, 500 mmHg. The pressure drop was a step change. This decrease in pressure set up the pressure gradient for air to diffuse from the fluid into the bubble, causing the bubble to grow. This bubble growth was recorded on the VCR. The bubbles were left to grow as long as possible; usually less than 30 minutes. As the bubbles became larger, they either moved out of the field of view or they pulled off the wire and floated away.

After the growth was recorded, the diameter as a function of time was measured. The tape was played back and digitzed into a Macintosh computer. The computer could grab images from the tape at a rate of 1/30th of a second. For the first two minutes of growth an image was taken every 1-2 seconds. After this time period, an image was taken every 6, 10, or 20 seconds.

A typical image is in figure two. To measure the diameter, a scale was first set up using the 100 μm wire. A line was drawn across the wire. Then a straight line was drawn across the bubble and measured. Care was taken to ensure that this line was a diameter. To make measuring easier, the image could be thresholded or have image processing software trace the edges. The resulting image was only black and white which allowed the edges to be easily detected (figure 3). The diameter was not the only possibility for measuring the size. An outline of the bubble could be traced and have its length measured; this would give the circumference which is related to the radius by $C = 2\pi r$. The area inside the outline could be measured; the area is related to the radius by $A = \pi r^2$. Another method to measure the radius could have been to calculate the equation of a circle by choosing three points on the edge of the bubble. The timer from the VCR was used as the time base (labeled time counter in figure 2).

The data for the diameter was then entered into the computer and a Nelder - Mead
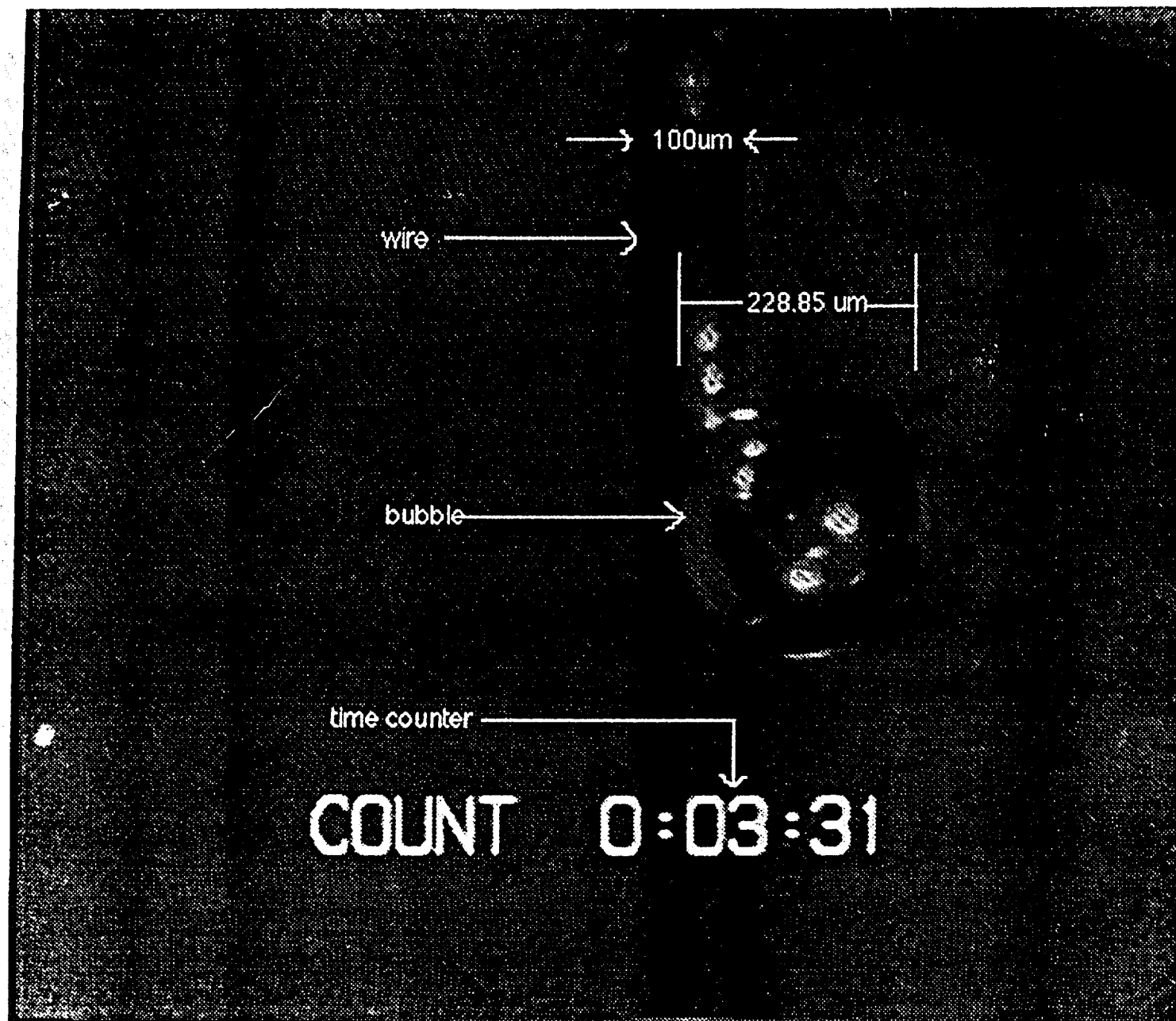
Fig. 2 This is a typical image after it is digitzed into the computer. The vertical band is the wire and the bubble is growing on the right side of the wire. The time counter is output from the VCR. No ennancements have been made to this image. Generally, the images appear much cleaner on the computer screen.

COUNT 0:03:00

Fig 3. This is an image of a growing bubble after it has been
enhanced. The image processing software has traced the
edges. It is easier to measure the diameter on this image.

simplex algorithm for a non-linear optimization routine was used to fit the data to an exponential equation of the form:

$$D_i(t) = C_1(1 - \exp(-C_2 t)) + D_o \qquad (8)$$

D_i = Diameter of the bubble

C_1 = Constant

C_2 = Constant

t = Time

D_o = Initial Diameter

To get an equation for the radius as a function of time, equation (8) should be divided by 2.

The data was then checked against Van Liew's and Gernhardt's equations. A computer program was used to simulate conditions in the lab. The program incorporates the ability for gas to wash out of the tissue based on a perfusion model. However, there was no perfusion in the lab; to remove its effects the tissue half time was set large so that perfusion could not take place during the time frame being observed. The values for inputs to the program are as follows:

α = .0125 1/atm

D = .00002 cm2/s (3)

δ = 72.0 dynes/cm (3)

%iner breathing gas = 100%

P_{N2} initial = 760 mmHg

Initial P_h = 760 mmHg

Half time = 100000 min

Initial radius = .5*D_o cm

Time step = 12 sec

By setting the initial Pn2 to 760 mmHg, air became the gas involved in the diffusion process and not just nitrogen. The data file, pressure.dat, contains some of the constants and the pressure profile. The program will read in the current pressure and check to see if there was a change in the pressure. If there was a change, a Boyle's law expansion was calculated. The difference in pressure set up the gradient for gas to diffuse into the bubble. A new radius was then calculated based on both rate equations. Because gas diffused into the bubble, the tissue pressure dropped. These calculations were performed for every 12 seconds and the output compared to the data. The profile used was the same as that in the lab. To compare the predicted

data with the measured data, the radius was multiplied by two and converted from cm to μm.

RESULTS

The lab data was fit to an exponential curve. The time constant of the exponential wa averaged for trials at each pressure (tables 1, 2, and 3). Without Acetone, surface tension is 72 dynes/cm and with the Acetone, surface tension is 40 dynes/cm. A t - test was performed on the data to see if there was any statistical significance with and without Acetone. As can be seen from the tables, there was not a difference in the exponential of the data when Acetone was added. The full set of coefficients is in appendix A. The coefficients represent the data after the Boyle's law expansion was completed. The first 5 seconds of the $D_i(t)$ was removed before the equations were fit to the data.

Table 1.
250 mmHg with and without Acetone.
No Acetone: C2 = .004890; SD = .0024837
With Acetone: C2 = .004857; SD = .0018978
Student t = .0253; No significant difference

Table 2.
380 mmHg with and without Acetone.
No Acetone: C2 = .0016863; SD = .0002998
With Acetone: C2 = .0022521; SD = .0007397
Student t = 1.825; No significant difference

Table 3.
500 mmHg with and without Acetone.
No Acetone: C2 = .0013222; SD = .0004793
With Acetone: C2 = .0012631; SD = .0004835
Student t = .021; No significant difference

From observation it was noted that the density of bubbles that grew was 5 - 10 bubbles per ml. When the surface tension was decreased, many small bubbles, 5 - 25 μm, were present. These small bubbles constantly floated in the fluid; however, they never stuck to the wire. Bubbles less than 15 μm diameter did not grow in the time frame observed. However, bubbles of 30 μm diameter did grow. With the surface tension reduced, bubbles of <5 μm persisted in the fluid. This was seen by taking a sample of the Acetone / water mixture and looking at it under a high power microscope. With plain water these 5μm bubbles were rarely seen.

Small bubbles formed on the surfaces of larger bubbles. A bubble of 300 μm diameter would have as many as 50 smaller bubbles on its surface. These bubbles did grow, but not at the same rate as the larger one to which they were attached.

When the computer program was run, neither equation matched the data. If the h, barrier length, in Gernhardt's equation was replaced with r, then both equations gave similar answers. The prediction for the Boyle's law expansion was similar to the expansion seen in the lab. The equation predicted an expansion that was smaller than what was observed. Because diffusion was taking place while the bubble was expanding, a larger expansion than expected was observed. The rate was off approximately by a factor of 2, 3, or 4 (figure 4). When dr/dt was multiplied by this factor, the equations more closely matched the data. However, the equations predicted the growth was too fast for the smaller bubbles.

DISCUSSION

Neither theory accurately predicts bubble growth for the lab. By replacing the effective diffusion barrier with the radius, both methods predict the same result. If the theoretical dr/dt is multiplied by a factor of 2 - 4, the equations match the data. Van Liew's equation does not leave room for adding such a factor for the lab. However, the h can be defined as a function of the radius. For example, at 250 mmHg, h could equal .25*r. This allows the theoretical equation to follow the data for the middle sized bubbles. Even with this correction the theory does not work for the smaller sized bubbles that grow. Since dr/dt is increased with smaller radii, a small bubble is predicted to grow very rapidly. This is not the case. A small bubble grows at approximately the same rate as a larger bubble.

The equations do not describe why some bubbles do not grow, or grow very slowly. If a bubble radius of 7 μm is put into the equation then it will grow rather large in a couple of minutes. A 7 μm bubble never grew while under observation in the lab. There appeared to be a threshold for bubbles to grow. In trials were there was not a preexisting bubble to focus on, the chamber was taken to altitude and left for as long as possible, 1-2 or more hours. In these trials, larger bubbles were not found while the smaller ones were present. A 30 μm diameter bubble grew while a 15 μm diameter bubble either did not grow or grew very slowly. Somewhere in between these two sizes could be a cut off, or a gray area where bubbles will

Growth of Bubbles at 500 mmHg

▲ Data from lab
▫ Van Liew dr/dt
+ Van Liew dr/dt*3

Diameter (μm)

Time (sec)
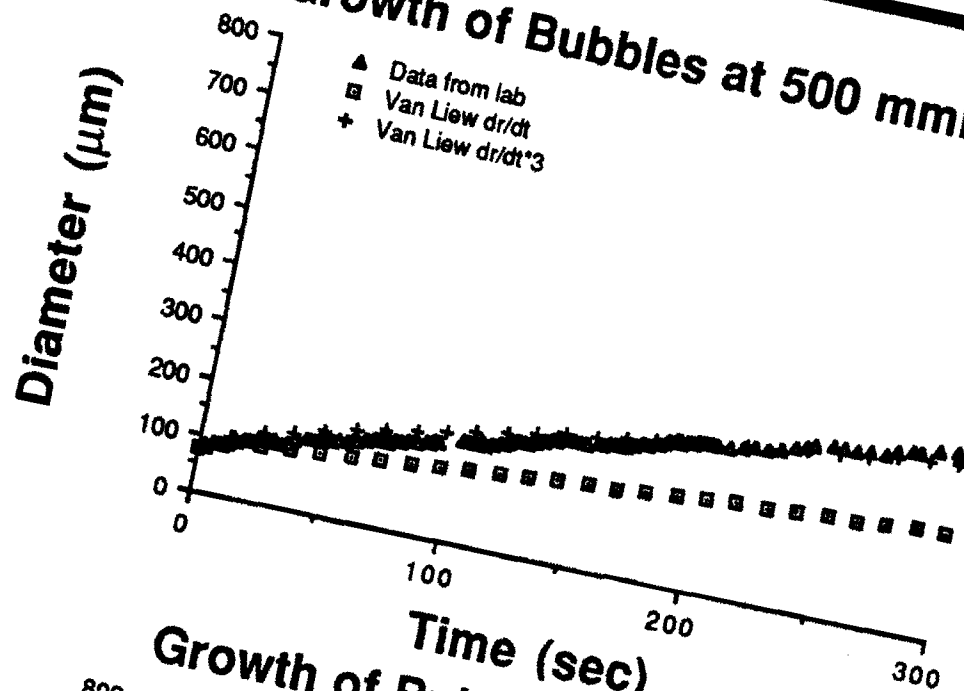
Fig 4.a dr/dt had to be multiplied by 3 to match the data

Growth of Bubbles at 380 mmHg

▲ Data from lab
■ Van Liew dr/dt
+ Van Liew dr/dt*2

Diameter (μm)

Time (sec)
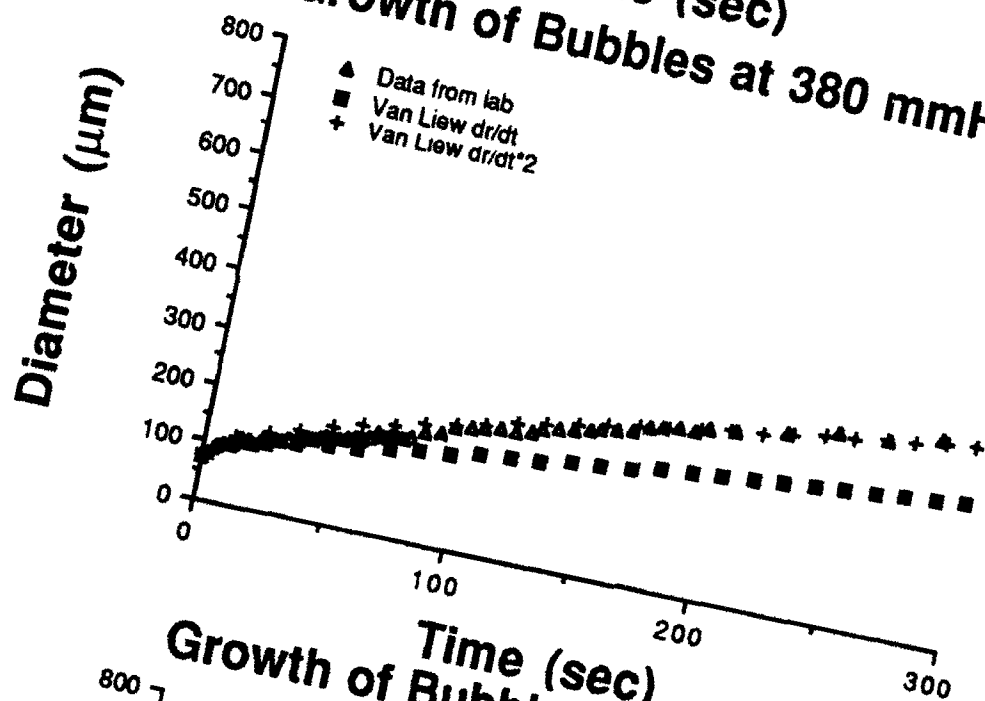
fig 4.b dr/dt had to be multiplied by 2 to match the data

Growth of Bubbles at 250 mmHg

▲ Data from lab
▫ Van Liew dr/dt
+ Van Liew dr/dt*4

Diameter (μm)

Time (sec)
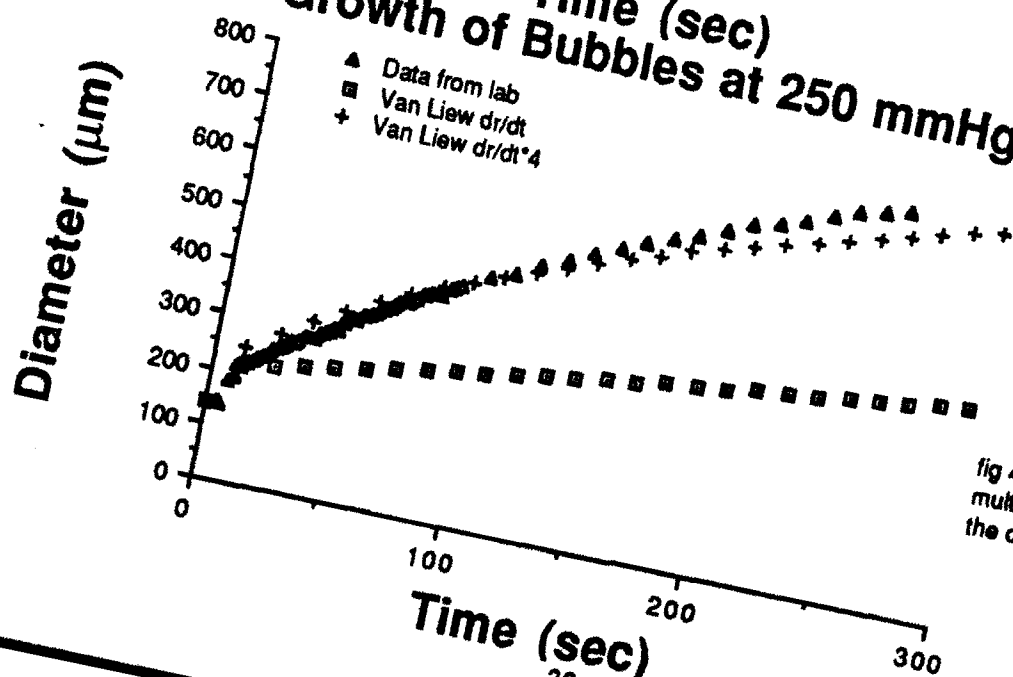
fig 4.c dr/dt had to be multiplied by 4 to match the data

20-14

either grow or not grow. If there is a cut off, it is probably around 20 μm in diameter. A possible explanation is that the pressure due to surface tension at this size is too large for a bubble to grow.

When the surface tension was reduced, bubbles of 5μm diameter existed, while they were rare with the higher surface tension. This supports the theory that the surface tension is the main inhibitor for bubble formation. Bubbles will tend to form in areas of reduced surface tension. Blood normally has a δ of 45-50 dynes/cm, physiological lipids have a δ of 20-30 dynes/cm, and in the lungs the surface tension can be as low as 5 dynes/cm. This allows the bubbles to form, but how they become large enough to grow is unknown. The addition of energy, heat or kinetic, to the system may be a possibility. If the ambient temperature rises, there will be heat flow into the system that could cause the bubble to grow. Kinetic energy was added to the system by striking the side of the chamber. The spontaneous generation of hundreds of bubbles was the result. Many bubbles were too small to grow; however,, others were of adequate size. In humans, this type of kinetic energy could arise from movement. Motion causes stress across the joints and the energy produced when joints impact can cause bubble formation. In test subjects, there was a higher incidence of DCS when the subject exercised.

Smaller bubbles always formed on the surface of the larger bubbles. While these bubbles were of the size that would normally not grow, <15 μm, they did grow. The difference might be that diffusion is now from a gas to a gas, one bubble to another. This could be a way that bubbles grow in the body. Small bubbles could be attached to red blood cells, and grow through diffusion of gas from the cells to the bubble. These bubbles were not measured for their growth rate because of their nonspherical shape and the difficulty in focusing on them. A possible reason for the existence of the these bubbles is that they were in the fluid, and as the larger bubble grew it began to occupy the space were these smaller bubbles existed. Another possibility is that the smaller bubbles floated by and were attracted to the larger bubble, this was observed on several occasions. A third possibility is that the smaller bubbles pinched off of the larger bubble in an effort to increase the surface area to volume ratio to continue growth.

Overall the bubbles exhibit three phases of growth. The first is the rapid Boyle's law expansion. The second is a phase of linear growth, and the third is a decaying rate of growth. The gas entering the bubble decreases the gas concentration in the fluid. This decreases the pressure gradient and the amount of gas available for diffusion, which causes the third phase. A

representation of the third phase is shown in equation (7). The ultimate size that a bubble can reach is dependent on this equation; thus it is dependent on the number of bubbles in the fluid. If there are many bubbles, they can not grow very large because of the competition between each other for the gas. The rate of linear growth is dependent on the altitude. The growth at 500 mmHg and 380 mmHg is significantly less than the growth at 250 mmHg. The relationship of growth rate to altitude does not have a linear relationship. This is seen by comparing the time constants, C2, from the curve fits. There is only a difference of .00036 between 500 mmHg and 380 mmHg, while the difference between 380 mmHg and 250 mmHg is .0032, almost a factor of ten difference for an almost identical change in pressure.

CONCLUSION

Growth of microbubbles is important to the study of DCS. Symptoms of DCS are attributed to the existence of these bubbles. If an understanding of DCS is desired, then an understanding of how bubbles form and grow must be known. The growth rate of bubbles was studied in the lab, and an attempt was made to correlate the data with the existing theory.

If bubbles form from bubble nuclei, how do they grow? Small bubbles did not grow. This suggests that there is some other catalyst for bubbles to reach a size where they will grow. This catalyst could be the link in describing the latency period between the time when a subject goes to altitude and the time when bubbles are seen in his body. In the laboratory, the time course for a bubble to grow to the size commonly found in the *inferior vena cava* (IVC), 100 μm, was only a couple of minutes. A subject can be taken to altitude and bubbles may not be seen for a couple of hours. Clearly there may be a process that prevents these bubbles from growing.

To further develop a model that predicts the risk of DCS, it is necessary to know how a bubble forms. If this information could be discovered, it would be known what types of flights would lead to bubble formation. A profile could then be developed that would manipulate the formation process to hinder bubble formation and decrease the risk of DCS. An important factor in bubble formation is the surface tension. Low surface tension allows bubbles to form easily and exist for a long period of time. Currently bubbles are seen in the IVC but their origin is unknown. They are too big, 50 - 100 μm, to have transported through small capillary beds. A probable scenario is that the bubbles form in an area of low surface tension, the lungs or

possibly the joints. These bubbles will be <15 μm; they can be transported through the capillaries and into larger vessels where some other process will allow them to grow.

REFERENCES

1.   Gernhardt, M. "Tissue gas bubble dynamics during hypobaric exposure."
        SAE Technical Paper Series 851337. Fifteenth Intersociety Conference
        on Environmental Systems, San Francisco, California, July 1985.

2.   Van Liew, H. D. and Hlastala, M. P. "Influence of bubble size and blood
        perfusion on absorption of gas bubbles in tissues." Respiration
        Physiology. 7:111-121, June 1969.

3.   Weast. CRC Handbook of Chemistry and Physics. 70th1989.

4.   Weathersby, P. K., Homer, L. D., and Flynn, E. T. "Homogeneous nucleation
        of gas bubbles in vivo." J. Appl. Physiol.: Respirat. Environ. Exercise
        Physiol. 53(4): 940-946, 1982.

APPENDIX A

This appendix contains all the coefficients for the data. The coefficients are for the equation:

$$D_i = C_1(1 - exp(-C_2 t)) + D_0$$

The coefficients are for each trial. The pressure was known, and D_0 was taken as the point after the Boyle's law expansion.

Table A1
Values for the results of the curve fits to the data when no Acetone was added

Pressure(mmHg)	C_1	C_2	D_0
250	526.52	.0064305	267.50
250	494.39	.0070759	65.00
250	567.27	.0076308	65.00
250	850.38	.0037429	213.22
250	1230.71	.0016308	289.09
250	1070.15	.0028302	238.18
380	627.63	.002198	90.70
380	735.60	.0014717	125.58
380	706.82	.0014808	102.13
380	534.13	.0016909	129.79
380	753.28	.00159	114.00
500	1160.1	.0007533	87.5
500	239.64	.0012018	23.08
500	536.7	.0012566	241.82
500	393.76	.0020615	63.64
500	768.14	.0013178	108.00

Table A2.
Values for the results of the curve fits when Acetone was added.

Pressure(mmHg)	C_1	C_2	D_0
250	1300.7	.0023253	318.18
250	500.94	.0075987	278.18
250	527.38	.0052597	339.62
250	968.30	.006105	170.91
250	1150.8	.0033635	174.55
250	991.05	.0044943	194.55
380	989.23	.0018978	200.00
380	676.80	.0026882	143.40
380	813.01	.0030355	249.09
380	1013.3	.0025543	207.27
380	773.00	.0029063	207.27
380	1011.90	.0010456	187.27
380	980.84	.001637	140.00
500	898.84	.0010972	135.85
500	607.19	.0015083	125.71
500	1076.1	.0010087	169.81
500	652.52	.0019992	178.18
500	992.55	.0007844	264.00
500	686.18	.0017099	136.21
500	1053.3	.00073418	212.23

THIS PAGE INTENTIONALLY LEFT BLANK

ANALYSIS AND SYNTHESIS OF WHISPERED SPEECH USING A FORMANT SYNTHESIZER

Edward L. Riegelsberger
Graduate Student
Department of Electrical Engineering
The Ohio State University
2015 Neil Ave.
Columbus, OH 43210

ANALYSIS AND SYNTHESIS OF WHISPERED SPEECH USING A FORMANT SYNTHESIZER

Edward L. Riegelsberger
Graduate Student
Department of Electrical Engineering
The Ohio State University

ABSTRACT

Whispering is a common mode of communication which has been overlooked in current speech synthesis research. Whispered speech is less affected by the complications inherent in voiced speech such as dynamic source-tract interaction, variability in phonation modes, and source-based speaker dependent differences. These properties simplify the analysis/synthesis task. Investigation of whispered speech should provide insights about vocal-tract dynamics without the complications of source dynamics and source-tract interactions. Much can be also be learned from the comparison of whispered speech and their phonated speech counterparts. This documents illustrates some of the potentials of whispered speech synthesis research and describes the initial stage of our investigation of whisper synthesis: the development of an analysis/resynthesis system for whispered speech. The basic system consists of a formant tracking algorithm interfaced to a formant synthesizer. Performance results are fair, although not as good as desired. Complications encountered during the developement process are discussed along with potential solutions and inherent limitations of the system.

ANALYSIS AND SYNTHESIS OF WHISPERED SPEECH USING A FORMANT SYNTHESIZER

Edward L. Riegelsberger

I. INTRODUCTION

Whispering is a common mode of communication which has been overlooked in current speech synthesis research. Whispered speech is less affected by the complications inherent in voiced speech such as dynamic source-tract interaction, variability in phonation modes, and source-based speaker dependent differences. Whispered speech typically is depicted as a white-noise source driving the vocal-tract filter. The nature of this model simplifies the analysis/synthesis task by allowing focus to be placed on vocal-tract dynamics without the complications of source dynamics and source-tract interactions.

The long-term goal of this research pursuit is the study of the synthesis of whispered speech. We expect that by studying the synthesis of whispered speech, knowledge will be gained that can be applied to the synthesis of speech in general.

This document primarily reports progress on the first stage of research: developing a frame-based speech analysis/resynthesis system designed for whispered speech. The first section provides background information about whispered speech and gives a review of whispered speech research. Some of our motivations are described here along with potential areas of exploration. The remainder of the report describes the design of a speech analysis/resynthesis system for whispered speech.

II. WHISPERED SPEECH

A. Background

Whispered sounds are produced by turbulence resulting from air rushing past the slightly open glottis. During a whisper, the glottis tenses forming a triangular opening through which turbulence is generated. Physiological studies report that the glottal opening is larger during whispered speech than during voiced speech and that the vocal folds during whispered speech are seen to move very slightly or not at all [1]. Acoustic studies have found whispered speech to have a flatter spectrum than voiced speech, with a level power spectrum between 200 and 2000Hz [2]. Whispering is an inefficient speaking mode, and is limited in the number of words than can be whispered in a single breath. A loud whisper's amplitude is typically 20 db less than in conversational speech and only slight variations in intensity are possible.

Previous research of whispered speech has primarily been concerned with perception. Kallail and Emanuel [3] [4] found the identification of whispered vowels to be 65% for males and 64% for females. This is about 15% less than for normally phonated vowels. Tartter [5] obtained a 64% overall identification accuracy for 18 consonants (both normally voiced and normally unvoiced). He also reported that whispered speech contains reliable cues to place and manner of articulation, and somewhat less prominent cues to voicing [5]. It also has been demonstrated that whispered speech is sufficient to perform sex discrimination and that listeners can discriminate between two speakers from their whispered speech [6].

Past studies have repeatedly observed that whispered phonemes have higher formants than their voiced counterparts [3]. Kallail and Emanuel found statistically significant differences in first formant frequencies between whispered and phonated vowels. Similar trends were observed in F2 and F3 for certain vowels. It is suspected that a glottis slightly more open during whispered than in phonated speech produces the constant rise in formant frequencies. This would help to explain why the effect

is most prominent for lower formants, which are more affected by tracheal coupling. A quantitative explanation of these effects remains to be given.

B. Motivation

Little has been reported on the synthesis of whispered speech although the simplifying assumptions that can be made about whispered speech make it an attractive subject for syrthesis. As a non-periodic waveform, there is no fundamental frequency, open-quotient or other voice source variables. There are no voice/unvoiced/mixed decisions or alternative modes of phonation.

The way in which the turbulence source varies spectrally for different modes and places of articulation is not precisely known. If we assume that the spectral variation is small, and that the turbulence spectrum is generally flat (white noise), then existing spectral estimation techniques can be applied to model the "true" vocal-tract transfer function. Even if the aspiration noise is not white, the assumption of a non time-varying source offers a new way of studying the dynamics of the vocal-tract. Knowledge of vocal-tract dynamics independent of source dynamics and source-tract interactions is valuable to voiced speech analysis and synthesis.

Additional insights applicable to synthesis should be obtainable from the comparison of whispered utterances to their voiced counterparts. Such comparisons should illustrate the effects of source-tract interaction and tracheal coupling on voiced utterances. It has been suggested that some speaker dependent characteristics may become apparent through mappings between phonated and whispered speech. As noted by Kallail and Emanuel [3], " an understanding of the formant differences between whispered and phonated vowels might aid in the development of optimally effective methods for identifying certain speaker characteristics (e.g. sex) from speech samples."

We plan on exploring many of the issues introduced above using synthesis techniques. First though, a synthesis method must be chosen and a reliable technique for analyzing whispered speech in terms of the synthesizer parameters must be developed. Automated speech analysis in this manner

is not a trivial task. The remainder of this document describes the first stage of our investigation of whisper synthesis: the development of an analysis/resynthesis system for whispered speech.

III.   THE FORMANT SYNTHESIZER

Our present objective is to create a system which analyzes speech and generates synthesizer parameters necessary to reproduce the original utterance. Such a system is correctly labelled a speech encoder/decoder. Substantial effort has been devoted to speech encoder/decoders but with the objective of compact and robust representations for communication purposes. For synthesis applications, the encoder/decoder system should encode speech in terms of acoustically and articulatorally significant features. Meaningful speech features are essential to the effective design of rule-based synthesizers. Additionally, these meaningful features can be used to better understand speech production and improve speech synthesis.

The formant synthesizer is a popular synthesis tool controlled by parameters that represent acoustic and articulator features. They are based on the source-filter theory of speech production and come in two different varieties: cascade and parallel. All are comprised of a number of resonators which correlate with formants observed in speech. In the cascade formant synthesizer, resonators are arranged in a serial manner to jointly define an all-pole filter. In a parallel formant synthesizer, resonators are arranged in parallel and a gain term is associated with each resonator. The parallel arrangement results in a pole-zero transfer function. Clearly, the parallel formant synthesizer is more versatile in that it can produce a greater variety of spectral representations. However, the cascade formant synthesizer requires fewer controlling parameters and its all-pole characteristic is sufficient to describe a majority of speech types.

The block diagram of Figure 1 depicts a simple cascade type formant synthesizer. The source block represents the glottis during voicing and aspiration. Source parameters include mode-of-
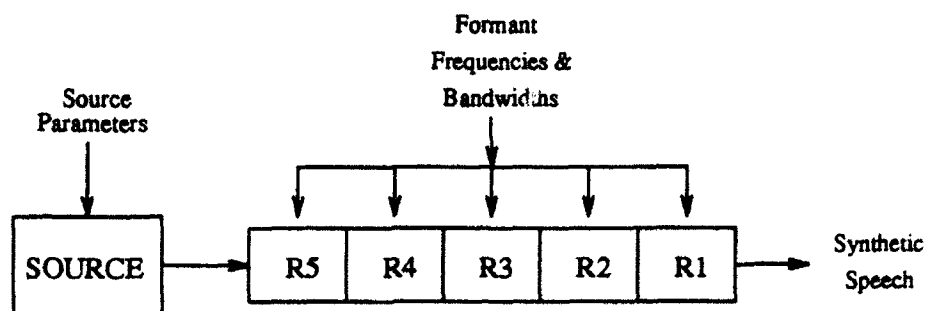
Figure 1: Block diagram of a simple cascade formant synthesizer.

voicing, amplitude of voicing, amplitude of aspiration, and fundamental frequency. The blocks labelled R1 through R5 are resonators, each representing a single formant, whose combination is the vocal-tract transfer function. A resonator is controlled by two parameters: formant frequency and formant bandwidth.

A. The Klatt Cascade-Parallel Formant Synthesizer

The Klatt formant synthesizer [7] is a hybrid formant synthesizer that incorporates both cascade and parallel synthesis methods. It is designed with parameters to control voice and noise source characteristics as well as filter parameters. It is capable of generating mixed modes of voicing and modeling source-tract interactions. It includes three different voice source models and provides control of the phenomena of diplophonia, F0 jitter, and tracheal coupling. In addition to the cascade synthesizer, a separate parallel branch exists for the generation of frication and burst noise. Extra pole zero pairs are provided to model nasal and tracheal resonances. It has been demonstrated to produce natural and intelligible synthesized speech and its sufficient to synthesize all of the sounds of English.

We are using a recent version of the Klatt synthesizer that uses the KLSYN88 cascade-parallel formant synthesizer first described in Klatt and Klatt [8]. Since we will be using the aspiration noise source exclusively for synthesis of whispers, some more detail of its generation is necessary. The

Klatt synthesizer produces aspiration noise using a random number generator. The simple generator provides a flat amplitude distribution between limits of -32768 and +32767. Note that this is a very basic random number generator whose output is not truly Gaussian, although Klatt claims that the spectral differences are barely perceptible. The random numbers are filtered through two zeros at $z=-0.75$ and $z=1$. These provide soft high-pass and low-pass filtering of the number stream. This stream is then amplitude adjusted to produce the final input to the resonators. The gain controlling synthesis parameter AH, "amplitude of aspiration", is in decibels.

IV. A FRAME-BASED ANALYSIS/RESYNTHESIS SYSTEM

In this section we describe the design of a speech analysis/resynthesis system for whispered speech. The routine analyzes whispered speech, and generates parameters to drive the Klatt formant synthesizer to reproduce the input utterance. The system is frame based (i.e. each frame is analyzed and synthesized independent of adjacent frames) and consists of three stages: a speech analysis (data collection) stage, a translation/encoding stage, and a synthesis stage.

The initial data collection stage employs a formant tracking algorithm to determine formant center frequencies and bandwidths for every frame. The tracking algorithm used is related to the one proposed by Secrest and Doddington [9]. The algorithm uses linear prediction analysis to generate formant frequency candidates and applies dynamic programming to smooth formant trajectory estimates by imposing frequency continuity constraints. For each frame, rms energy is also calculated.

The second stage generates parameters necessary to drive the Klatt synthesizer from the measurements and estimate of the first stage. We do not intend to modify the Klatt synthesizer (third stage of system) except in special cases so it is the second stage in which the majority of processing will occur. Since we will be analyzing whispered speech only, information about mode of voicing and voicing characteristics are not necessary. We are also assuming that source/tract interactions are

negligible. Therefore only formant frequencies F1-F5, formant bandwidths B1-B5, and amplitude of aspiration *AH* values must be generated by stage two. The update interval is chosen to be 5 msec, so synthesizer parameter values must be produced every 5 milliseconds. Window width, the data window over which estimates will be calculated, is independent of the update interval. Window width is generally larger than the update interval. Although the appropriate window width is debatable, a default width of 50msec is used.

A collection of VCV's and CVC's, both phonated and whispered, have been collected from a male and a female speaker. Selected whispered tokens from this collection have been used to evaluate the basic analysis resynthesis system. Surprisingly, the frame by frame method of encoding and synthesizing speech presents some difficulties in implementation. Analysis of the initial analysis/resynthesis routine on whispered steady-state vowels produces whispered speech containing click. burp, and gargle like transients. It appears that the inherent spectral variability of whispered speech demands very careful attention to factors such as smoothing, window width and mean filter power. Secondly, transients are introduced at frame boundaries by the formant synthesizer which required special corrective measures.

The performance of this initial version of an analysis/resynthesis system is unacceptable. The following section describes efforts to improve the performance of the system. Some of the inherent limitations of frame based analysis/resynthesis are also discussed.

V. SYSTEM EVALUATION AND IMPROVEMENT

Initial tests of the frame based analysis/resynthesis system demonstrate audible transient noise in steady-state whispered vowels. This is unacceptable whisper reproduction. Four potential sources of these transients are investigated: inappropriate window width, abrupt amplitude variations, synthesizer induced transients and inherent spectral variation. Each of these issues are addressed, and

some possible corrective measures are offered.

A. *Window width*

A steady-state whispered vowel /i/ was analyzed/synthesized using a range of window widths: 25msec, 50msec, 100msec, 250msec, and 500msec. As expected, transient activity decreases as window size increases. With the 25, 50, and 100 msec windows, transients can still be heard. The 250 and 500 msec windows produce rather steady sounding synthesized speech. Visually, the synthesized versions, even the large windowed ones don't appear as steady-state as the original.

In wide-band spectrograms of the original and 5 synthesized whispered vowels, transients manifest themselves as vertical lines most visible at high frequencies. This demonstrates the impulsive nature of these transients. Overall, there is significantly more spectral energy in the original whispered vowel above 4000Hz. The synthesized versions' lack of high frequency energy demonstrates some of the deficiencies of using the cascade synthesizer with a limited, fixed number of formants.

Certainly, having a larger window width reduces the variability of the pole estimates, but window widths larger than 100msec are clearly not practical for the analysis of continuous speech. As system behavior is studied for consonants, it may become necessary to incorporate a dynamic window width based on the rate of spectral change.

B. *Amplitude Variations*

It is possible that some of the transients result from abrupt changes in the amplitude parameter AH. To test this, a steady-state whispered vowel /i/ was resynthesized normally and with $AH = 60$ for all time. The constant amplitude version is just as inconsistent as the old. There is very little audible difference between the two. There is a very slight difference noticeable if the window width is set to 100msec or larger. The pops still exist, but the weaker ones are less noticeable.

There is another more subtle way in which amplitude variations could be introducing audible

transients. Filter parameters specified in terms of formant frequencies and bandwidths equivalently specify pole locations in an all-pole transfer function. No amplitude information is included. The mean power of the all-pole filter is a function of the location of poles. As a result. a constant amplitude source driving two different all-pole filters will produce outputs with different average powers. Amplitude control independent of filter function is desirable, therefore, the filters should be normalized with respect to average power.

Unfortunately, there is no simple expression for mean power in terms of linear prediction coefficients. A numerical evaluation technique is used instead. An FFT is used to calculate the power spectral density at 512 points. The sum of these points produces a rectangular rule approximation to the total filter power. The calculated average filter power is used to normalize the all-pole filter power to 1.0. As a result, the synthesizer amplitude control is independent of the filter function. [1]

The application of this normalizing function offers some improvement in synthesis quality due to amplitude smoothing. The time waveform is visually more consistent. In general, the adjustment made little difference but for the occasional odd filter estimate. compensation made a substantial improvement.

C. Synthesizer induced transients

The fact that transients can be generated at frame boundaries in formant synthesizers is well known. To reduce the effects of transients, the Klatt synthesizer has code to correct waveform irregularities caused by sudden changes to F and BW. A correction proposed by Fujisaki and Azami [10] is implemented in the synthesizer to alleviate the problem. Careful testing of this correction found

---

[1] Ultimately, the synthesizer gain is a function of the measured rms energy of the signal, the mean filter power, and an overall gain parameter. For a given signal with rms amplitude rms and its all-pole filter estimate with power spectrum $|H(\omega)|^2$ and total power g^2,

$$filter\_gain(db) = gain - 20\log g + 20\log rms \qquad (1)$$

little to no improvement in removing transients. When the error correction routine was removed, no audible difference could be heard. If the resonators were zero-stated at frame boundaries, the extreme magnitude transients were removed. Transients at frame boundaries still exist, but at least are not visible spikes in the waveform. One undesirable artifact of this method of reducing transients is that a periodic component is audible at the frame frequency. This effect is reduced by applying compensation only at locations where large transients are likely to occur.

D. Spectral Variation

The formant tracking algorithm is one unavoidable source of errors in the analysis stage. Many formant tracking algorithms exist, each performing best in different situations. Regardless of the algorithm, errors will be present. All that can be done is to detect and correct the obvious errors when they occur. The formant tracking algorithm presently in use is proposed by Secrest and Doddington [9]. Occasionally, this algorithm will incorrectly produce two near coincident formant tracts resulting in abnormally high amount of resonance at a given frequency. The best solution is for the formant tracking algorithm not to produce such estimates. A suboptimal solution is to watch for coincident formant tracts and substitute an adjacent frame estimate in place of the aberrant frame estimate.

Even good formant track estimates contain some variability in pole location estimates. Unfortunately, this variability is audible and very noticable in steady-state vowels. To qualitatively determine the extent of variation in pole estimates, random parameter slices for /i/ and /a/ analyzed with a 25ms window were used to synthesize non-varying whisper tokens. Although each token is a reasonable reproduction of the whispered vowel, there are substantial differences in whisper quality between the different slices. This is most likely due to variability in the bandwidth estimates. Vowel quality remains relatively consistent except for a few cases in which there is a large deviation in formant frequency values. To further investigate the effects of spectral variability, the above wave-
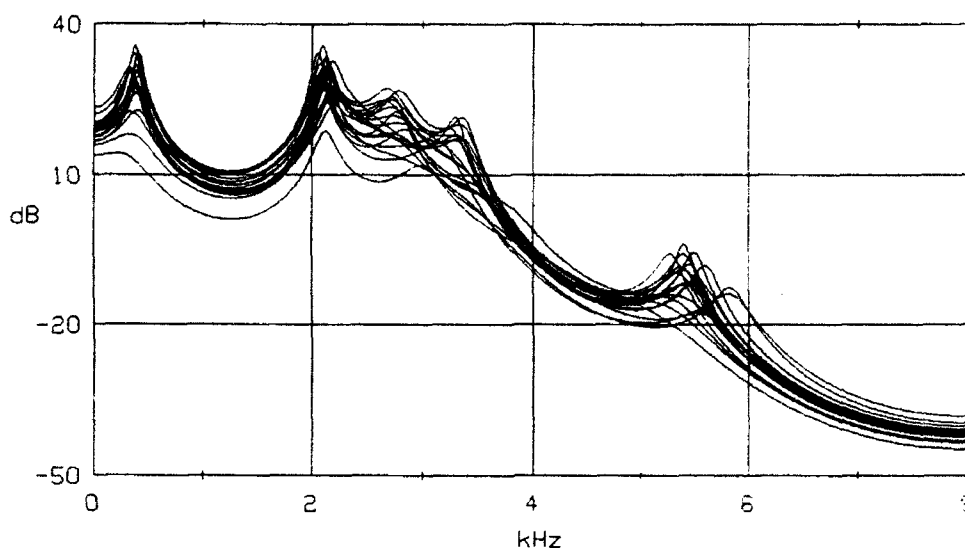
Figure 2: LPC spectrum for 50 msec frames of single steady-state vowel.

forms were concatenated together and played, simulating changes in synthesizer parameters across frame boundaries without potential filter transients. Admittedly the abrupt boundaries may have added some noise but hopefully not too much. The files sound much like the original synthesized tokens, full of burps and clicks. The conclusion is rather clear: significant variability in the spectral estimates (formant parameter selections) make a major contribution to the audible transients.

To illustrate spectral variability in formant estimates, the formant tracker derived lpc spectrum of 50msec frames of a steady state vowel are displayed in Figure 2. The general spectral shape is maintain in every trace but slight variations in formant center frequency and bandwidth alter the entire spectrum. As observed earlier, the overall quality and timbre of each spectrum is different and it is not surprising that all of these sounds jumbled together do sound poor.

VI. CONCLUSIONS

Whispered speech is an often overlooked mode of communication with properties that make it attractive for experiments in speech synthesis. The properties of whisper speech enable the study of vocal-tract dynamics independent of source-tract interactions. Also, by comparing the synthesis of whispered speech to that of phonated speech, some effects of source-tract interactions can be illustrated. A review of previous whisper research shows substantial study of the perception and acoustics of whispered speech but surprisingly little research of the synthesis of whispered speech. We believe that the study of whisper synthesis will provide insights into the synthesis of speech in general.

To commence our study, an analysis synthesis system for whispered speech is constructed. This document primarily describes the developement of the synthesis system. The analysis/resynthesis system encodes real whispered speech in terms of parameters that drive a speech synthesizer. Ideally, the speech synthesized from the encoded parameters should sound identical to the original utterance. In practice, this is not always possible.

Frame based analysis and synthesis has proven to be a more difficult task than expected, for both voiced and whispered speech. Initial attempts were disappointing. Synthesis of steady-state whispered vowels sound gargly and full of transients. The sources of much of the transient behavior have been identified: frame boundary transients, poor formant tracking, discrete amplitude variations, and the inherent variability of spectral estimates. Careful tuning of the algorithm that interfaces a formant tracking routine to the Klatt formant synthesizer results in a fair analysis/resynthesis system.

The ultimate difficulty lies in the spectral variability of the lpc estimates. Each estimate by itself produces good sounding, intelligible synthetic sounds, but putting a bunch of good but different estimates next to each other produces poor sounds. An improved formant tracking algorithm would

prevent many of these problems. In addition, better use of formant tracking results could also improve synthesis while relaxing performance requirements demanded from the tracking algorithm.

The analysis/resynthesis system will be used as a tool in the study of whispered speech. Since the system does not perform as well as expected, truly automatic analysis synthesis will not be attempted. Whisper synthesis will ultimately require some hand tuning of the analysis/resynthsis output.

VII. REFERENCES

[1] W. R. Zemlin, *Speech and Hearing Science: Anatomy and Physiology.* 1968.

[2] M. F. Schwartz, "Power spectral density measurements of oral and whispered speech." *J. Speech and Hearing Research*, vol. 13, pp. 445–446, June 1970.

[3] K. J. Kallail and F. W. Emanuel, "Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects," *J. Speech and Hearing Research*, vol. 27, pp. 245–251, June 1984.

[4] K. J. Kallail and F. W. Emanuel, "The identifiability of isolated whispered and phonated vowel samples," *J. Phonetics.* vol. 13, pp. 11–17, January 1985.

[5] V. C. Tartter, "What's in a whisper?," *Journal of the Acoustical Society of America.* vol. 86, pp. 1678–1683, November 1989.

[6] M. F. Schwartz and H. E. Rine, "Identification of speaker sex from isolated, whispered vowels," *Journal of the Acoustical Society of America,* vol. 44, pp. 1736–1737, 1968.

[7] D. H. Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America,* vol. 67, pp. 971–995, 1980.

[8] D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers." *Journal of the Acoustical Society of America,* vol. 87, pp. 820–857, February 1990.

[9] B. G. Secrest and G. R. Doddington, "An integrated pitch tracking algorithm for speech systems," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing,* (Boston, MA), pp. 1352–1355, 1983.

[10] H. Fujisaki and S. Azami, "Characteristics of digital pole circuits for simulating time-varying vocal tract transfer function," in *Annual Report Engineering Res. Inst. 30,* (Univ. Tokyo, Tokyo

Japan), pp. 89–94, 1971.

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

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.

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

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

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.

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

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.

PC Based Cardiovascular Model for Displaying Acceleration Stress

Frank C. Smeeks
Medical Student

Meharry Medical College
Nashville, Tennessee

September 1992

PC Based Cardiovascular Model for Displaying Acceleration Stress

Frank C. Smeeks
Medical Student
Meharry Medical College

Abstract

Acceleration on the human body is known to cause several physiological effects. A PC based interactive computer model of the cardiovascular system was developed for studying the system response to acceleration stress. The PC model can be used in educating pilots of high performance aircraft and visualizing acceleration stress tests without physically performing them. The model consists of simulations of the heart, arterial vasculature, venous system, and peripheral circulation. The calculations for each component of the model include the effects of gravitational acceleration and compensatory mechanisms(both physiological and externally applied). Using this model, pressures and flows at different points of the cardiovascular system can be calculated and displayed. By observing pressures and flows at different points of interest(e.g. the ophthalmic artery pressure), protective techniques can be investigated. The observations allow pilots and scientists to experiment on optimizing these protective procedures without performing unnecessary testing. Thus, the model saves undo stress on pilots and experimental costs. The model is useful in the identification of pilot dysfunction and in the development of new protective maneuvers in response to acceleration stress on the cardiovascular system. The availability and affordability of the PC make this model scientifically and educationally indispensable.

Title   PC Based Cardiovascular Model for Displaying Acceleration Stress

Author   Frank C. Smeeks, Meharry Medical College

Introduction

Under most conditions the acceleration due to gravity, g, is constant. In aviation, however this is not necessarily the case. During aviation maneuvers, the human body may be exposed to additional accelerations which may have dramatic physiological consequences. United States Air Force pilots of high performance aircraft are exposed to acceleration forces exceeding their physiological tolerance limit. At high acceleration g values, the pressure due to the column of blood between the heart and brain may exceed the pressure generated by the heart. The cerebral blood vessels collapse and the pilot suffers 'blackout' due to impaired retinal circulation.

The cardiovascular system is suited to modeling attempts because of its dependence on well understood hydraulic principles and because of the wealth of understanding about the various constituent organs involved in blood transport. A PC based compartmental model which models and displays the pressures in the human cardiovascular system was developed for multiple purposes. The model serves the function of educating pilots by visual display of physiologic phenomena associated with aerospace maneuvers, allows scientists to visualize the results of possible experiments without actually performing the test, and connects the world of cardiovascular modeling to the PC domain.

Description of the Model

The cardiovascular system is a distributed system composed of blood with varying non linear viscous and inertial properties interacting with vessels of complex morphology, non linear elastic walls, and multiple levels of homeostatic systems controlling many distributed properties. Material properties of components of vascular walls, blood flow in branching tubes, distributed impedance of arterial and pulmonary trees, force-length-velocity relationship in the heart, and hemodynamic control of the system are some of the components which must be considered. Guyton et al. developed refined computer models of the cardiovascular system over longer time scales, and has built an incredibly complex model which embodies, in over 350 equations, the behavior of interacting cardiovascular, renal, and respiratory systems (Guyton 1972). The use of linear approximations and

lumped parameters allows a much simplified description in terms of linear ordinary differential equations. Such a model was converted from a text based PC program to a user friendly interactive data display model (Moore and Jaron 1991).

The model consists of simulations of the heart, arterial vasculature, venous system, and peripheral circulation. A block diagram of the model is given in Figure 1. The model included compensatory mechanisms which pilots use to offset the affect of acceleration stress on the cardiovascular system (externally applied pressure, muscular straining, and positive pressure breathing). Physiological mechanisms including carotid baroreceptor control of the heart rate and venous tone were also included in the model. The ventricular model is based on the 1974 Suga and Sagawa variable elastance model. The arterial system and venous system model consisted of segments each with its own resistance, compliance, and inertance characteristics. The parameters for both were obtained using calculations based on the Navier-Stokes equations. The peripheral circulation connecting arterial and venous segments are modeled by lumped representations. The compensatory mechanisms were modeled by adding a pressure component to apply the appropriate external pressure to a vessel segment through its compliance. Physiologic mechanisms were simulated by an adaptation of a 1973 model by Green and Miller.

This model although straightforward for a scientist or computer programmer to use, was ineffective for an end user. The shortcomings of the program were the inability to visualize the data and the inability to chose data to report without program modification. The existing interface was a primitive menu driven interface with a relatively poor level of interaction. These shortcomings were overcome by creating a user friendly interactive interface between the mathematical component and the end user. The user interface consisted of a compartmental model which was displayed, a maneuver choice display, and a simulated chart recorder display. All selection input was controlled by point and click mouse function in a graphical environment and user input was accepted via keyboard in a graphic environment.

Results

In order for the model to be useful for teaching, it must be sophisticated enough to embody behavior being taught, paint a valid picture of the real world, and allow changes to system parameters. The revised PC based model provided a pseudo real time simulation with an intuitive interface with a "virtual patient" appearance. The user interface served the purposes of hiding implementation details, allowing students to manipulate the system to any conceivable hemodynamic pathology, structuring the model in an intuitive manner, reflecting the dynamic nature of the system, and maximizing
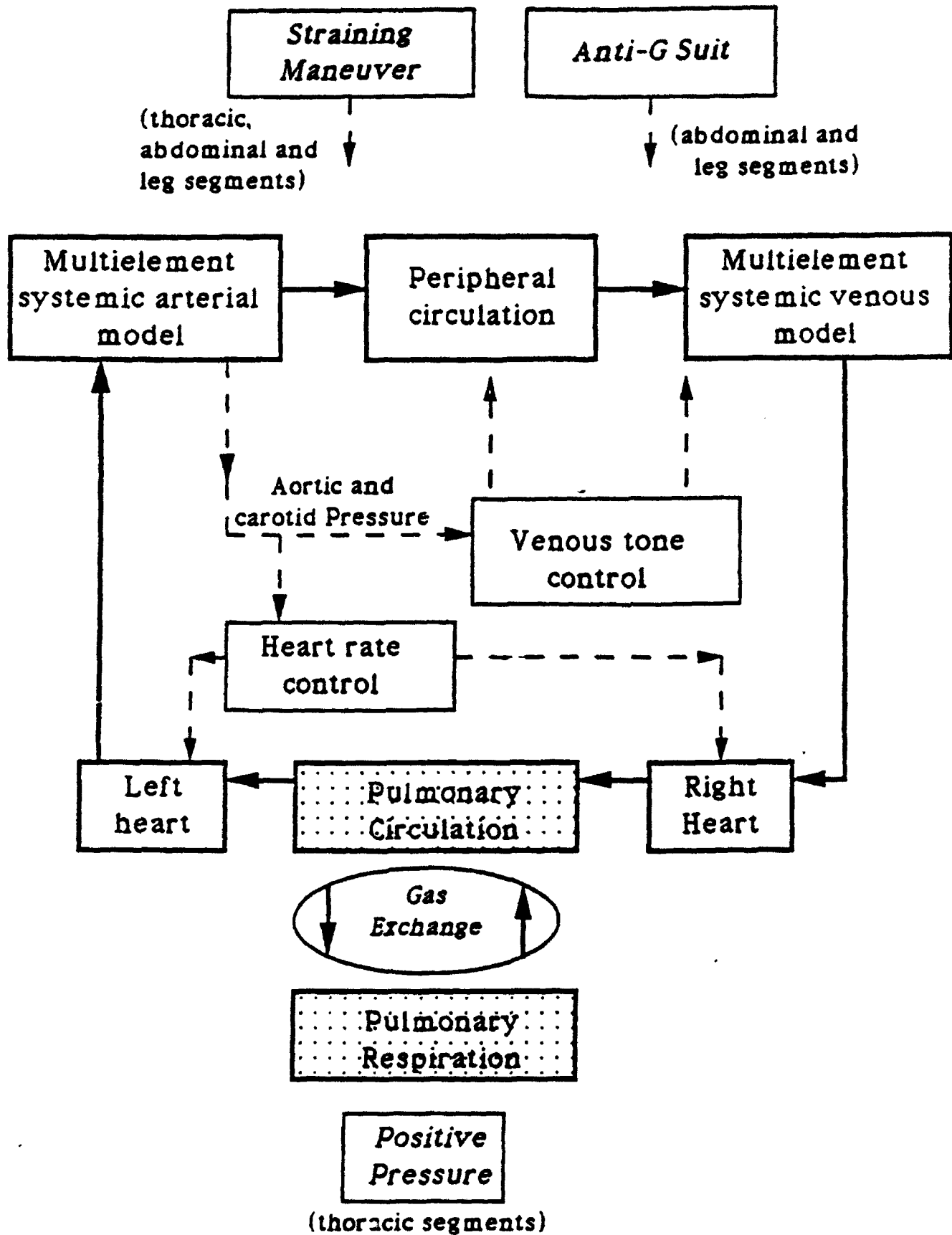
Figure 1. Block diagram of the cardiovascular model.

flexibility by allowing students to monitor any hemodynamic variable of interest. Several user interface objects were developed to satisfy these purr)ses. The window provides a view of some data object inside the computer (e.g. the compartmental model). An icon provides a reminder to the user(e.g. the F-16 aircraft). The dialog box receives input from the user(e.g. the main menu section).

High performance aircraft can withstand maneuvers that the human body is incapable of withstanding. Using the interactive PC based model the pilot can enter aircraft g accelerations and monitor the human physiology response. Pilots visualize on screen via the strip chart recorder what is occurring to the human body during the requested high performance aircraft maneuvers and compensatory mechanisms. This can help pilots in the decision of which compensatory mechanisms to employ or if no compensatory mechanisms would be of use. The dialog between the student(pilot) and the computer via the interface satisfies the educational objective.

With respect to research, studies of important hemodynamic events under conditions of acceleration stress are not easily accomplished. Invasive measurements in human subjects are often not possible and no suitable animal models exist. Acceleration studies require a controlled acceleration source and few human centrifuges exist. Most human experiments rely on indirect and subjective measurements yielding unreliable experimental results. The PC based mathematical model provides insight to the response of the cardiovascular system to acceleration stress. By observing the results, the model can provide insight to the causes of pilot impairment. This can lead to the postulation of solutions to pilot impairment. Protection techniques(effectual and ineffectual) can be modeled and applied to the subject without the requirement of human experimentation. Those techniques which prove mathematically effectual can then be tested using the human centrifuge. The use of a computer model to screen techniques saves undo human experimentation and cost. Cost savings may be realized by saving one high performance aircraft from loss due to undetected pilot impairment during a centrifuge experiment. In addition to these savings, is the possibility of design and optimization of new and existing techniques. The following paragraphs demonstrate some sample model results.

Figure 2 shows the model results(display) for a rapid onset run(ROR). This refers to a Haversian shaped increase in g from 1.0 to a plateau value of 4.5, a time at the plateau, and then a Haversian decline to 1.0 g. The first symptom of falling arterial pressure to the head is impairment of vision. Loss of peripheral vision occurs at 50 mmHg systole and complete loss of vision near 20 mmHg. The simulation indicates that the subject would experience 'blackout' and not recover until the acceleration stress is
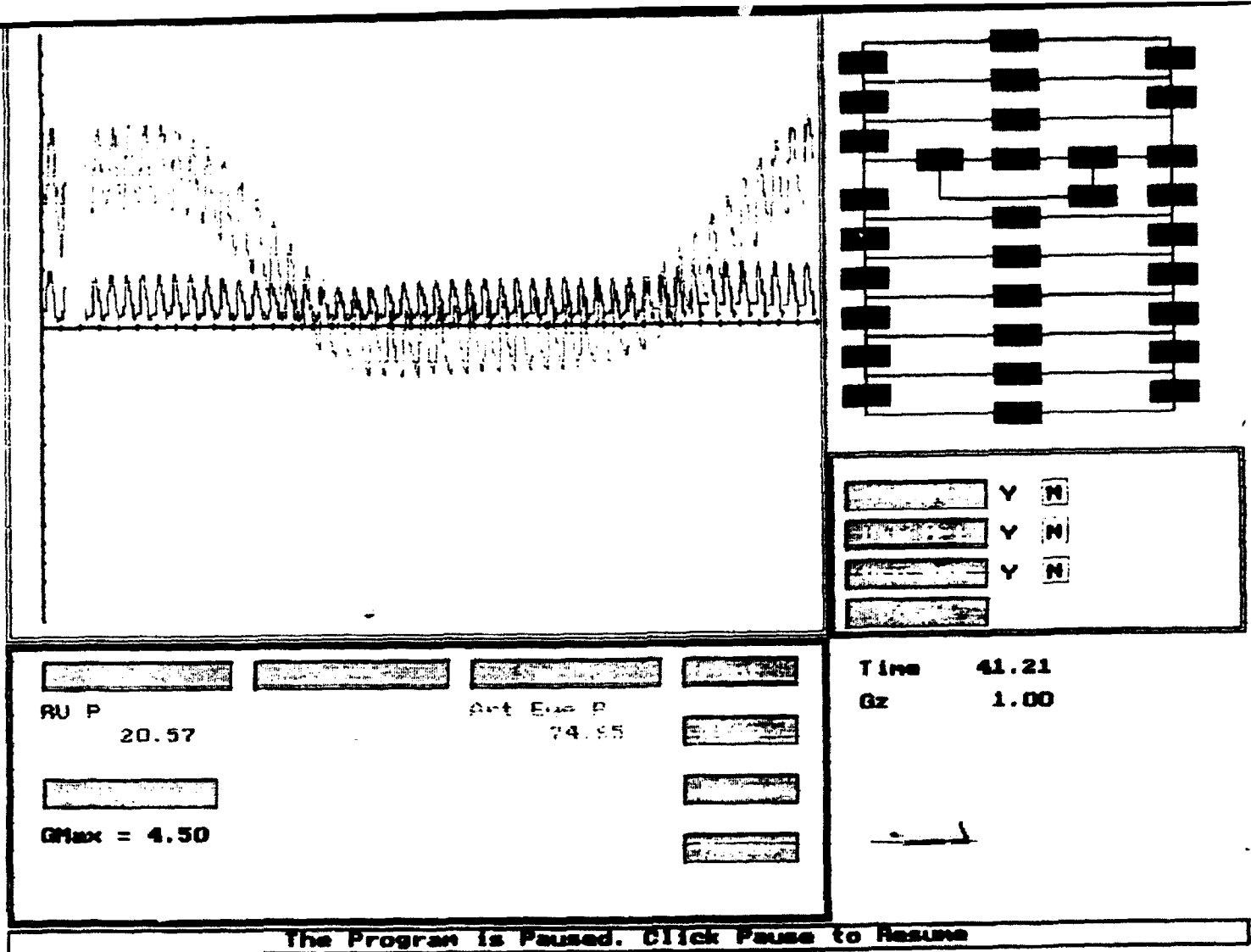
Figure 2. Model output under acceleration stress profile having a maximum of 4.5 g.
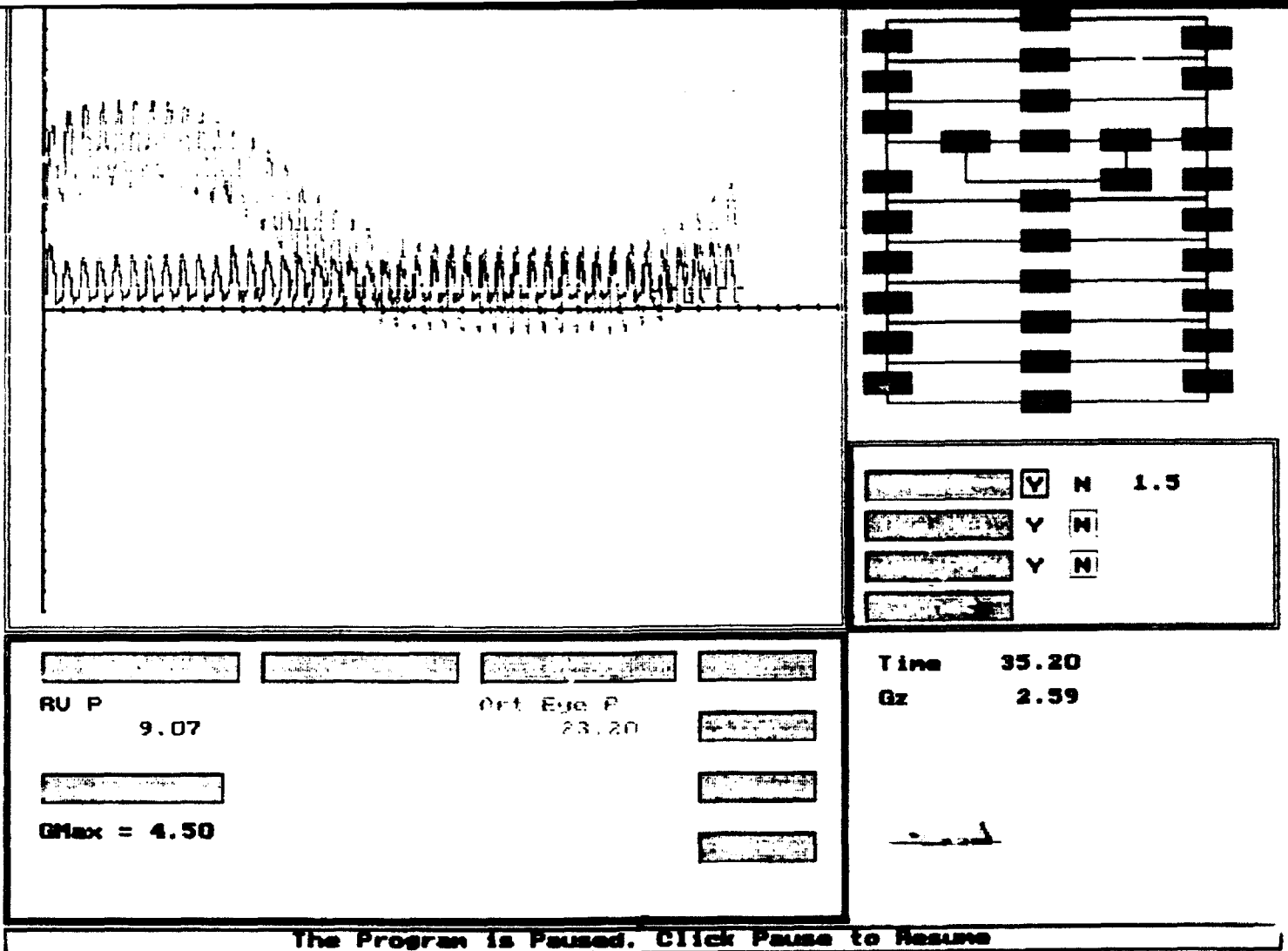
Figure 3. Model output under acceleration stress profile having a maximum of 4.5 g and standard G-suit applied at 1.5 g.

removed. Figure 3 shows the model display for the same 4.5 g Haversian profile with a standard G-suit as the employed compensatory maneuver. The prime mechanism of the standard G-suit is to apply 1.5 psi to the calves, thighs, and abdomen; therefore, maintaining a higher venous return. The model predicts that by wearing the standard G-suit only a loss of peripheral vision results as opposed to complete loss of vision without the suit. Further experimentation with a pulsating G-suit with the 4.5 g Haversian profile shows no impairment of vision. Experimental studies on the human centrifuge confirm all of the model findings.

Conclusions

Experimental results suggest that the PC based model is a useful tool for studying the cardiovascular system responses to acceleration stress. The complexity of the components mentioned earlier prevent an exact replica of the cardiovascular system; however, the model is valuable in the identification of sources of pilot dysfunction and as a guide to methods of improving tolerance to gravitational stress. The modular PC program can easily be modified to incorporate and test new methods. The availability and afford ability of the PC to physiologists in the lab and to pilots in the classroom make the PC based model indispensable.

Future Directions

Future directions for the program include the incorporation of alternate profiles(simulated air combat maneuvers), heart rate control, addition of a pulmonary system model, display of flow data concurrent with pressure data, and manipulation of parameters from the menu for particular systems compartments(resistances and capacitances).

Acknowledgment

References

Suga H, Sagawa K: Instantaneous pressure volume relationships and their ratio in the excised, supported canine left ventricle. *Circ. Res.*, 35:117-126, 1974.

Green JF, Miller NG: A model describing the response of the circulatory system to acceleration stress. *Ann. Biomed Eng.*, 1:455-467, 1973.

Beheken J: Some computer models in cardiovascular research. In Cardiovascular Fluid Dynamics I, 6:173-223, Academic Press, London, 1972.

Guyton A, Coleman T, Granger H: Circulation: Overall regulation, Annual Review of Physiology, 34(13), 1972.

Moore T, Jaron D: Cardiovascular Model for Studying Circulatory Impairment Under Acceleration. *IEEE Eng. in Med and Bio.*, 3:37-40, 1991.

Moore T, Jaron D, Chu C, Dinnar U, Hrebien L, White M, Hendler E, Dubin S: Synchronized External Pulsation for Improved Tolerance to Acceleration Stress: Model Studies and Preliminary Experiments. *IEEE Trans. Biomed Eng.*, 2:158-165, 1985.

Moore T, Jaron D: Positive Pressure Breathing During G Exposure. Report for U.S. Army Research Office, Research Triangle Park, NC.

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

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.

SITUATIONAL AWARENESS CORRELATES
(A PILOT STUDY)

Lorraine C. Williams
Graduate
School of Human Behavior

United States International University
10455 Pomerado Road
San Diego, CA

Final Report for:
Summer Research Program
Armstrong Laboratory

September 1992

SITUATIONAL AWARENESS CORRELATES
(A PILOT STUDY)

Lorraine C. Williams
Graduate
School of Human Behavior
United States International University

Abstract

The relationship between female and male situational awareness (SA) was studied. The test battery administered was comprised of the following: Armed Services Vocational Aptitude Battery (ASVAB), Space Fortress (SF), and Hartman Test (HT). Male Ss scored slightly higher than female Ss on the majority of testing however analysis measuring the significance of this was incomplete. Correlations that were significant were also quite low i.e. .1 to .4 suggesting the relationship was weak. While scores varied widely among both groups of Ss, further analyses is required to determine the potential for future pilot selection through such test tools.

SITUATIONAL AWARENESS CORRELATES
(A PILOT STUDY)

Lorraine C. Williams

Cognitive Psychologists have long been interested in the concepts and complex workings of attention and memory. Opposing theories regarding resources, mechanisms, processes, automaticity, workload and other issues remain unresolved, and as such require further study.

Aviation psychology also concerns itself with the above disputes. Particularly of interest is how these things relate to existing aviation problems of a cognitive nature such as accidents due to pilot error. The approach taken so far in the field is one that considers memory and attention as trainable, with the objective of making performance in a specific domain automatic. Automaticity is assumed to decrease workload thus allowing increased resources to be available, which is assumed to lead to an increased awareness state resulting in improved human performance. This increased awareness state is referred to as situational awareness (SA) and has several working definitions to date.

Situational Awareness (SA) in the aviation community can generally be defined as "the pilot's mental model of the world around him" (Endsley, 1990:41). It has also been referred to as "the knowledge that results when attention is allocated to a zone of interest at a level of abstraction" (Fracker, 1988:102).

For purposes of this study, the following will be used as a working definition: "Situational awareness is principally (though not exclusively) cognitive, enriched by experience" (Hartman & Secrist, 1991:1084). It "embodies a remarkable cognizance of the total combat situation, the capacity to anticipate rather than react to change, and the ability to make valid intuitive decisions under conditions of great time urgency and stress" (Secrist, et.al., 1991:1).

Situational awareness skills include the following: 1) keen sensitivity to critical cues, 2) early acquisition of cue patterns, 3) rapid cue integration, 4) high-speed pattern processing, 5) instantaneous situation assessment, 6) direct trend apprehension, 7) anticipatory judgment, 8)intuitive decisions, and 9) response automaticity. See Secrist, et.al., (1991) for details.

It has been proposed that these qualities may be already reasonably developed in some individuals such as those top pilots (approximately 5%) who have historically been responsible for 40% of total kills (Youngling, et.al., 1977). It is further suggested from previous research that these skills are trainable, and will contribute significantly to a highly competent level of performance (Hartman & Secrist, Endsley, 1987, 1988b, 1990).

THE PROBLEM

To increase the currently small number of top fighter aces, there is a need to make certain each aviator is performing at their full potential. To date limited studies with small samples have been done in developing the SA concepts and model of testing; therefore there is a need to conduct further studies of reliability and validity for existing protocols and compare them against each other. There are many approaches to this problem; I have chosen to examine the feasibility of selecting individuals into flying training programs who have situational awareness training potential.

METHODOLOGY

Two computer administered tests were used to assess for SA skills. These were comprised of the following:

Space Fortress (SF) - This tool was developed in the Cognitive Psychophysiology Laboratory at the University of Illinois at Urbana-Champaign as "an

28-4

experimental task for the study of complex skill and its acquisition" (Mane' & Dochin, 1989:17). This video game is comprised of multiple demanding tasks yielding a set of subscores as well as a total score. The goal is to keep one's ship (controlled by a standard joystick) in control at all times while maneuvering away from enemy fire and attempting to destroy the enemy's fortress (using a standard keyboard). See Mane' & Dochin, (1989) for details.

Hartman Test (HT) - This task, which measures visual identification in near-threshold processing, was developed by Bryce O. Hartman at the Armstrong Laboratory Aerospace Medicine Directorate, in collaboration with Grant Secrist. The task uses four visual targets (card suits) and a blank (no distractors) with a mask. Each subject will have two scores, 1) reaction time and 2) accuracy. See Secrist & Hartman (1992) for details.

Procedure - The tests, although self-administered, were monitored by experienced technicians. Each test comes with instructions and built-in practice sessions as part of the software.

The SF test were administered first over 45 minutes, followed by a three minute break, followed by 22 minutes of the HT test. There was one 15 minute break. The WT task was administered three blocks at a time with 20 trials per each block. Visual stimuli times were set at 66, 52, and 33 msec. There were two blocks given at each stimulus time for a total of six blocks.

Data Analysis - Descriptive statistics were calculated for ASVAB, SF, and SA scores. These data were analyzed using pearson product moment correlation coefficients to determine the relationship both within and among SF, SA and ASVAB subscores and total scores. Varied N's represent Ss whos testing was incomplete. A corrected matrix produced with the range-x program would have been appropriate to improve credibility of the estimates, however; this was not possible due to time constraints.

RESULTS

Demographics - Subjects were 191 male and 21 female basic AF recruits ranging in age from 18 to 26 with a mean age of 20. Subjects had previously been administered the Armed Services Vocational Aptitude Battery (ASVAB) test upon entry level, thus scores were obtained from data files (see Ree, et. al for details). Subjects were tested on the eleventh day of their six weeks of basic training.

Armed Services Vocational Aptitude Battery (ASVAB) - The sum of subtest scores or total score ranged from 446 to 644 for males with a mean of 550.2 and standard deviation of 36.3. Total scores for females ranged from 465 to 618 with a mean of 536.4 and standard deviation of 43.6. Males scored slightly higher on the subtests with the exception of three. Females scored slightly higher than males on numerical operational (NO) 57.5 vs. 55.3; coding speed (CO) 60.4 vs. 54.7; and mathematics knowledge (MK) 58.0 vs. 56.3 (See Table 1).

Male correlations were as follows. Significant correlations existed for general science (GS) and various subtests, however these were quite low ($r = .24$ to $r = .54$). The correlation between GS and word knowledge (WK) was $r = .57$, $p < .001$. Arithmetic reasoning (AR) was significantly correlated with various subtests ranging from $r = r = .19$ to $r = .39$. A significant correlation of $r = .6$, $p < .001$ existed for AR and MK. There were no significant correlations at $r = .6$ or greater for the remaining subtests (See Table 2).

Female correlations were as follows. General science (GS) was significantly correlated with AR at $r = .77$, $p < .001$; MK at $r = .65$, $p < .001$; MC at $r = .78$, $p < .001$; and EI at $r = .63$, $p < .001$. Other significant correlations were quite low as they were for AR and MC, WK and MC, WK and EI, PC with AS, MK, MC, and EI, NO with AS and MK, AS and EI, and Mk and MC. Other significant

correlations were for AR and MK, r = .80, p < .001; WK and PC, r = .70, p < .001; WK and AS, r = .71, p < .001; and AS and MC, r = .65, p < .001. No significant correlations existed between coding speed (CS) and any other subtest (See Table 2).

Table 1
Means and Standard Deviations for
Male's and Female's ASVAB
Scores

Source	Male	Female
General Science (GS)	54.7 (6.8)	51.8 (7.3)
Arithmetic Reasoning (AR)	55.0 (6.2)	55.5 (6.9)
Word Knowledge (WK)	54.2 (4.6)	54.3 (6.3)
Paragraph Comprehension (PC)	55.1 (5.2)	55.1 (5.3)
Numerical Operations (NO)	55.3 (6.4)	57.5 (4.9)
Coding Speed (CO)	54.7 (6.7)	60.4 (6.7)
Auto & Shop Info. (AS)	54.6 (7.5)	45.9 (7.0)
Mathematics Knowledge (MK)	56.3 (7.4)	58.0 (6.9)
Mechanical Comprehension (MC)	57.0 (7.1)	49.5 (8.6)
Electronics Info. (EI)	53.6 (7.5)	48.4 (6.2)
Sum of Subtests (SSS)	550.2 (36.3)	536.4 (43.6)

N = 184 males, 21 females
Note: SD in ( )

Table 2

Correlation Matrix for Male's and Female's
ASVAB Scores

	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	SSS
V14	1.0000 (0) P= .	.3532 (184) P= .000	.5723 (184) P= .000	.2375 (184) P= .001	-.0582 (184) P= .216	-.0504 (184) P= .248	.3332 (184) P= .000	.3115 (184) P= .000	.5431 (184) P= .000	.4973 (184) P= .000	.6768
V15	.3532 (184) P= .000	1.0000 (0) P= .	.3182 (184) P= .000	.2383 (184) P= .001	.3037 (184) P= .000	.2639 (184) P= .000	.0770 (184) P= .149	.6057 (184) P= .000	.3992 (184) P= .000	.1939 (184) P= .004	.6715
V16	.5723 (184) P= .000	.3182 (184) P= .000	1.0000 (0) P= .	.4719 (184) P= .000	.0272 (184) P= .357	.2639 (184) P= .000	.2321 (184) P= .001	.3943 (184) P= .000	.3220 (184) P= .000	.3687 (184) P= .000	.6399
V17	.2375 (184) P= .001	.2383 (184) P= .001	.4719 (184) P= .000	1.0000 (0) P= .	.0849 (184) P= .126	.0641 (184) P= .194	.0843 (184) P= .127	.2601 (184) P= .000	.2136 (184) P= .002	.1161 (184) P= .058	.4609
V18	-.0582 (184) P= .216	.3037 (184) P= .000	.0272 (184) P= .357	.0849 (184) P= .126	1.0000 (0) P= .	.5504 (184) P= .000	-.2219 (184) P= .001	.3531 (184) P= .000	-.0957 (184) P= .098	.1032 (184) P= .082	.3212
V19	-.0504 (184) P= .248	.2639 (184) P= .000	.2639 (184) P= .000	.1390 (184) P= .030	.5504 (184) P= .000	1.0000 (0) P= .	.0426 (184) P= .283	.3016 (184) P= .000	-.0043 (184) P= .477	.0368 (184) P= .310	.3902
V20	.3332 (184) P= .000	.0770 (184) P= .149	.2321 (184) P= .001	.0843 (184) P= .127	-.2219 (184) P= .001	.0426 (184) P= .283	1.0000 (0) P= .	-.1585 (184) P= .016	.5002 (184) P= .000	.5800 (184) P= .000	.4607
V21	.3115 (184) P= .000	.6057 (184) P= .000	.3943 (184) P= .000	.2601 (184) P= .000	.3531 (184) P= .000	.3016 (184) P= .000	-.1585 (184) P= .016	1.0000 (0) P= .	.2576 (184) P= .000	.0914 (184) P= .109	.6066
V22	.5431 (184) P= .000	.3992 (184) P= .000	.3220 (184) P= .000	.2136 (184) P= .002	-.0957 (184) P= .098	-.0043 (184) P= .477	.5002 (184) P= .000	.2576 (184) P= .000	1.0000 (0) P= .	.5529 (184) P= .000	.5913
V23	.4973 (184) P= .000	.1939 (184) P= .004	.3687 (184) P= .000	.1161 (184) P= .058	.1032 (184) P= .082	.0368 (184) P= .310	.5800 (184) P= .000	.0914 (184) P= .109	.5529 (184) P= .000	1.0000 (0) P= .	.6194
SSS	.6768 (184) P= .000	.6715 (184) P= .000	.6399 (184) P= .000	.4609 (184) P= .000	.3212 (184) P= .000	.3902 (184) P= .000	.4607 (184) P= .000	.6066 (184) P= .000	.5913 (184) P= .000	.6194 (184) P= .000	1.0000

Table 2 Cont.

	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	SSS
V14 (GS)	1.0000 (0) P= .	.7707 (211) P= .000	.4959 (211) P= .011	.4981 (211) P= .011	.3872 (211) P= .041	.1489 (211) P= .260	.5875 (211) P= .002	.5649 (211) P= .001	.7838 (211) P= .000	.6258 (211) P= .001	.9299 (211) P= .000
V15 (AR)	.7707 (211) P= .000	1.0000 (0) P= .	.1718 (211) P= .228	.1857 (211) P= .210	.3540 (211) P= .057	.1607 (211) P= .243	.2904 (211) P= .101	.8018 (211) P= .000	.6007 (211) P= .002	.1872 (211) P= .208	.7185
V16 (WK)	.4959 (211) P= .011	.1718 (211) P= .228	1.0000 (0) P= .	.6988 (211) P= .000	-.0538 (211) P= .408	-.1070 (211) P= .322	.7108 (211) P= .000	.2757 (211) P= .113	.5518 (211) P= .005	.4553 (211) P= .019	.6467
V17 (PC)	.4981 (211) P= .011	.1857 (211) P= .210	.6988 (211) P= .000	1.0000 (0) P= .	.1514 (211) P= .256	.0800 (211) P= .365	.4457 (211) P= .021	.4690 (211) P= .016	.4281 (211) P= .026	.3634 (211) P= .053	.6467
V18 (NO)	.3872 (211) P= .041	.3540 (211) P= .057	-.0538 (211) P= .408	.1514 (211) P= .256	1.0000 (0) P= .	.5759 (211) P= .003	-.1310 (211) P= .286	.4767 (211) P= .014	.0557 (211) P= .405	.2866 (211) P= .121	.4384
V19 (CS)	.1489 (211) P= .260	.1607 (211) P= .243	-.1070 (211) P= .322	.0800 (211) P= .365	.5759 (211) P= .003	1.0000 (0) P= .	-.1840 (211) P= .212	.2724 (211) P= .116	-.0710 (211) P= .380	-.0438 (211) P= .425	.2860
V20 (AS)	.5875 (211) P= .002	.2904 (211) P= .101	.7108 (211) P= .000	.4457 (211) P= .021	-.1310 (211) P= .286	-.1840 (211) P= .212	1.0000 (0) P= .	.1221 (211) P= .299	.6452 (211) P= .001	.6845 (211) P= .000	.6630
V21 (MK)	.5649 (211) P= .001	.8018 (211) P= .000	.2757 (211) P= .113	.4690 (211) P= .016	.4767 (211) P= .014	.2724 (211) P= .116	.1221 (211) P= .299	1.0000 (0) P= .	.4158 (211) P= .030	.0838 (211) P= .359	.7012
V22 (MC)	.7838 (211) P= .000	.6007 (211) P= .002	.5518 (211) P= .005	.4281 (211) P= .026	.0557 (211) P= .405	-.0710 (211) P= .380	.6452 (211) P= .001	.4158 (211) P= .030	1.0000 (0) P= .	.6137 (211) P= .002	.8076
V23 (EI)	.6258 (211) P= .001	.1872 (211) P= .208	.4553 (211) P= .019	.3634 (211) P= .053	.2866 (211) P= .121	-.0438 (211) P= .425	.6845 (211) P= .000	.0838 (211) P= .359	.6137 (211) P= .002	1.0000 (0) P= .	.6537
SSS	.9299 (211) P= .000	.7185 (211) P= .000	.6467 (211) P= .001	.6467 (211) P= .001	.4354 (211) P= .024	.2860 (211) P= .131	.6630 (211) P= .001	.7012 (211) P= .000	.8076 (211) P= .000	.6537 (211) P= .001	1.0000

Key:

V14 = (GS)
V15 = (AR)
V16 = (WK)

V17 = (PC)
V18 = (NO)
V19 = (CS)

V20 = (AS)
V21 = (MK)
V22 = (MC)

V23 = (EI)

28-9

Space Fortress (SF) - Block one total scores for males ranged from -41110.00 to 6936.00 with a mean of -15824.23 and a standard deviation of 9331.58. Block two totals ranged from -37582.00 to 21988.00 with a mean of -11719.75 and a standard deviation of 11832.86 (See Table 3).

Table 3

Means and Standard Deviations for
Male's Space Fortress Scores

Source	Block 1	Block 2
Points	-9023.68 (5922.78)	-7150.34 (8839.17)
Velocity	1556.60 (1191.76)	1969.22 (1799.35)
Control	-1919.56 (5969.21)	- 439.31 (6830.23)
Speed	-6437.59 (2347.94)	-6099.31 (3458.73)
Total	-15824.23 (9331.58)	-11719.75 (11832.86)

N = 191 Block 1, 189 Block 2
    Note: SD in ( )

Female total scores for block one ranged from -34212.00 to -16611.00 with a mean of -24168.67 and a standard deviation of 4755.76. Block two totals ranged from -34947.00 to -4497.00 with a mean of -20631.38 and a standard deviation of 7582.32 (See Table 4).

Females' scores improved from block one to block two with the exception of the points category which decreased. Males improved in all categories from block one to block two. Males scored higher overall for all categories with the exception of speed in block two (See Tables 3 and 4).

Table 4
Means and Standard Deviations for
Female's Space Fortress Scores

Source	Block 1	Block 2
Points	- 914.48	-9853.62
	(4378.35)	(1914.08)
Velocity	916.33	1206.67
	(533.67)	(1532.27)
Control	-8473.05	-6320.14
	(2859.21)	(4978.09)
Speed	-7470.48	-5664.29
	(1379.35)	(3870.34)
Total	-24168.67	-20631.38
	(4755.76)	(7582.32)

N = 21
Note: SD in ( )

For males, (See Table 5) block one points was negatively correlated with block one velocity at $r = -.26$, $p < .001$; and block two speed at $r = -.23$, $p < .001$. Block one and two points were positively correlated at $r = .64$, $p < .001$.

Block one velocity was negatively correlated with block two points at $r = -.15$, $p < .05$. Velocity one was positively correlated with control one and two at $r = .66$, $p < .0001$ and $r = .53$, $p < .0001$ respectively. Velocity one was also correlated with velocity two $r = .73$, and speed two $r = .37$, $p < .0001$.

Control one was correlated with speed one and two $r = .46$, $p < .0001$ and $r = .29$, $p < .0001$ respectively. Control one was also correlated with velocity two at $r = .49$, $p < .0001$. Control one and two were correlated at $r = .79$, $p < .0001$.

Speed one was negatively correlated with block two points at $r = -.27$, $p < .0001$. Speed one was positively correlated with velocity two $r = .39$; control two $r = .35$; and speed two $r = .55$, $p < .0001$ for all.

Block two points were negatively correlated with velocity two at $r = -.30$ and speed two at $-.52$, $p < .0001$. Velocity two was positively correlated with control two at $r = .61$ and speed two at $r = .62$, $p < .0001$. Control two was correlated with speed two at $.46$, $p < .0001$.

Table 5

Correlation Matrix for Male's Space Fortress
Scores

	Pnts 1	Vel 1	Cntrl 1	Spd 1	Pnts 2	Vel 2	Cntrl 2	Spd 2
Pnts 1	1.0	- .26*	.05	- .55	.64*	- .06	.10	- .23**
Vel 1	- .26*	1.0	.66*	.52*	- .15***	.73*	.53*	.37*
Cntrl 1	.05	.66*	1.0	.46*	.11	.49*	.79*	.29*
Spd 1	- .55*	.52*	.46*	1.0	- .27*	.39*	.35*	.55*
Pnts 2	.64*	- .15**	.11	- .27*	1.0	- .30*	- .05	- .52*
Vel 2	- .06	.73*	.49*	.39*	- .30*	1.0	.61*	.62*
Cntl 2	.10	.53*	.79*	.35*	- .05	.61*	1.0	.46*
Spd 2	- .23**	.37*	.29*	.55*	- .52*	.62*	.46*	1.0

* $p < .0001$
** $p < .001$
*** $p < .05$

Key: Pnts: Points
      Vel: Velocity
    Cntrl: Control
      Spd: Speed
        1:   Block 1
        2:   Block 2

N = 2 Missing for Block 2

For females, (See Table 6) block one points was correlated negatively with speed one $r = -.44$ and velocity one, $r = -.42$, $p < .05$. Points one and two were also correlated at $r = .59$, $p < .001$

Table 6

Correlation Matrix for Female's Space Fortress Scores

	Pnts 1	Vel 1	Cntrl 1	Spd 1	Pnts 2	Vel 2	Cntrl 2	Spd 2
Pnts 1	1.0	- .63	- .32	- .44****	.59**	- .42****	- .31	- .30
Vel 1	- .63**	1.0	.68*	.83*	- .28	.40****	.52***	.28
Cntrl 1	- .32	.68*	1.0	.75*	- .20	.31	.68*	.26
Spd 1	- .44****	.83*	.75*	1.0	- .21	.39****	.63**	.44****
Pnts 2	.59***	- .28	- .20	- .20	1.0	- .91*	- .58***	- .76*
Vel 2	- .42****	.40****	.31	.39****	- .91*	1.0	.68*	.88*
Cntrl 2	- .31	.52***	.68*	.63**	- .58***	.68*	1.0	.65**
Spd 2	- .30	.28	.26	.44****	- .76*	.88*	.65**	1.0

* p < .0001
** p < .001
*** p < .01
**** p < .05

Velocity one was correlated with control one at $r = .68$ and speed one at $r = .83$, $p < .0001$. Velocity one and two were correlated at $r = .40$, $p < .05$; and control two $r = .52$, $p < .01$.

Control one was correlated with speed one at $r = .75$, $p < .0001$. Control one and two were correlated at $r = .68$, $p < .0001$.

Speed one was correlated with speed two at $r = .44$, and velocity two $r = .39$, $p < .05$. Speed one and control two were correlated $r = .63$, $p < .01$.

Points two was negatively correlated with velocity two at $r = -.91$, and speed two at $r = -.76$, $p < .0001$. Points two and control were negatively correlated at $r = -.58$, $p < .01$.

Velocity two was correlated with control two at $r = .68$ and speed two at $r = .88$, $p < .0001$. Control two and speed two were correlated at $r = .65$, $p < .001$.

Hartman Task (HT) - Males scored a minimum of 0% and a maximum of 100%

correct at 66 msec for block one versus 1.6% to 100% for block two. Males

scored a minimum of 0% correct and a minimum of 100% for block one trials at 33

msec. They scored a minimum of 6.6% and a maximum of 96.7 correct for block two

trials at 33 msec.

Females scored a minimum of 5% correct at 66 msec. While their maximum

was 100% for block one. Block two yielded a 23% minimum and a 100% maximum.

Block one trials at 33 msec resulted in a 16% accuracy minimum and a 68.33%

maximum.

Females had slightly faster reaction times overall, however male Ss had

higher accuracy rates. Both male and female Ss improved their scores from block

one trials to block two except for males who scored slightly worse at 33 msec.

During block two trials (See Table 7).

Table 7
Means and Standard Deriations for Male
and Female Hartman Task Scores

	Male		Female	
Source	Block 1	Block 2	Block 1	Block 2
	%correct		%correct	
66 msec	63.18	71.91	57.58	69.75
	(21.86)	(23.36)	(20.92)	(24.32)
52 msec	65.09	67.58	60.67	66.83
	(23.52)	(26.66)	(20.94)	(21.87)
33 msec	45.02	45.33	38.25	40.33
	(22.82)	(21.54)	(13.28)	(17.56)
66 msec	1017.64	934.43	979.80	867.00
	(249.77)	(186.18)	(134.23)	(94.47)
52 msec	962.20	944.12	916.85	916.50
	(208.97)	(187.44)	(101.41)	(131.18)
33 msec	941.98	982.61	898.50	896.20
	(260.31)	(197.38)	(174.84)	(167.22)

N = 177 Males, 21 Females
NOTE: SD in ( )

Correlational Findings - Space fortress and ASVAB correlations for Males were as follows (See Table 8) here was a significant correlation between GS and control one, $r = .20$, $p < .001$. General science was correlated with velocity two, $r = .20$ and control two at $r = .20$, $p < .01$. Block one points and GS were correlated at .15, $p < .05$.

Table 8

Correlation Matrix for Males's ASVAB and
Space Fortress Scores

	Pnts1	Vel1	Cntrl1	Spd1	Pnts2	Vel2	Cntrl2	Spd2
GS	.15****	.10	.20**	-.02	.12	.20***	.20***	.06
AR	.09	.12****	.14****	.03	.10	.13****	.11	.09
WK	.15****	.13****	.21**	.02	.06	.21***	.20***	.18***
PC	.12****	.00	.10	-.04	.10	.05	.11	-.00
NO	.04	.08	.12****	.08	.06	.00	.04	.02
CS	.08	.01	.07	.06	.06	-.04	-.04	.06
AS	-.03	.10	.17	.13****	-.01	.08	.06	.07
MK	.25*	.15****	.25*	-.06	.15****	.20***	.24**	.10
MC	.07	.09	.30*	.10	.09	.16****	.32*	.14****
EI	.10	.12	.23**	.08	.04	.16****	.23**	.19**

*	$p < .0001$
**	$p < .001$
***	$p < .01$
****	$p < .05$

Block one and two scores were correlated with AR at $r = .12$ and $r = .13$ respectively; $p < .05$. Control one and AR were correlated at $r = .14$, $p < .05$.

Word knowledge was correlated with velocity one and two at $r = .13$, p $< .05$; and $r = .21$, p $< .01$. Control one and two were correlated with NK at $r = .21$, p $< .01$ and $r = .20$, p $< .001$. Block one points were correlated with NK at $r = .15$, p $< .05$ while WK and speed two were correlated at $r = .18$, p$< .001$.

Paragraph comprehevision was correlated with block one points at $r = .12$, p $< .05$. Numerical operations was correlated with control one at $r = .12$, p $< .05$. There were no significant correlations between coding speed and any SF subtest scores.

Block one speed and AS were correlated at $r = .13$, p $< .05$. Mathematical knowledge with points and control one were correlated both at $r = .25$, p $< .0001$. Mathematical knowledge was correlated with control two at $r = .24$, p $< .001$ and with velocity two at $r = .20$, p $< .01$. Block one velocity and points two were correlated with MK both at $r = .15$, p $< .05$.

Mechanical comprehension was correlated with control one and two at $r = .30$ and $r = .32$, p $< .0001$. Velocity two was correlated with MC at $.16$ and speed two and MC were correlated at $r = .14$, p $< .05$ for all.

Electronic information was correlated with control one and two, both at $r = .23$, p $< .001$. It was also correlated with speed two at $r = .19$, p $< .001$ and with velocity two at $r = .16$, p $< .05$.

Significant space fortress and ASVAB correlations for female Ss' were as follows (See Table 9). General science was correlated with points one at $r = -.39$ and control one at $r = .39$, p $< .05$ for both.

Numerical operations was negatively correlated with velocity one at $r = -.41$, p $< .05$. Coring speed was correlated with velocity two at $r = .39$, p $< .05$.

Mathematical knowledge was correlated with velocity and control one at $r = .39$ and $r = .38$, p $< .05$. There were no significant correlations between

Space Fortress subtest scores and arithmetic reasoning, word knowledge, para-
graph comprehension, auto and shop information, mechanical comprehension, or
electrical information.

Table 9

Correlation Matrix for Female's ASVAB
and Space Fortress Scores

	Pnts1	Vel1	Cntrl1	Spd1	Pnts2	Vel2	Cntrl2	Spd2
GS	-.39****	.28	.39****	.31	-.20	.15	.15	.25
AR	-.36	.33	.36	.23	-.30	.30	.27	.29
WK	-.17	-.05	.03	.01	-.19	.08	-.04	.24
PC	.06	.01	.17	.22	.11	-.05	-.11	.18
NO	-.33	-.41****	.36	.37	-.23	.26	.36	.23
CS	-.32	.37	.16	.33	-.31	.39****	.25	.17
AS	-.27	-.07	.14	-.01	-.11	-.07	-.19	.05
MK	-.21	.39****	.38****	.37	-.18	.27	.29	.29
MC	-.18	.18	.30	.34	-.10	.16	.13	.35
EI	-.21	.09	.29	.20	-.13	-.07	-.02	.19

*	p < .0001
**	p < .001
***	p < .01
****	p < .05

No significant correlations existed between SF subscores and HT response
times for male's block one testing. Significant correlations existed between
all SF subscores and percent correct at 52 msec, ranging from .14 to .39, p
< .05 - .0001. Significant correlations existed at 66 msec between % correct
and SF subscores except for points one ranging from .15 to .32 at p

< .05 - .0001. At 33 msec, correlations with SF subscores except for velocity ranged from .18 - .38, p < .01 - .0001.

For block two, male response times were negatively correlated with SF subscores i.e.: 66 msec with speed 2 at -.17, p < .05; 52 msec with points two at -.16, p < .05; and 33 msec with speed two at -.14, p < .05. For percent correct, significant correlations existed between 66 msec and SF subtests except for speed two; ranging from .26 - .29 at p < .0001. Correlations between 52 msec and SF subtests except for speed two ranged from .17 - .30 at p < .05 - .0001. For 33 msec significant correlations existed with SF subtests except for speed two, ranging from .12 - .31 at p < .05 - .0001.

For females, no significant correlations existed between HT and SF subscores for block one testing. For block two, there were no significant correlations between SF subscores and percent correct at 33 msec or response time at 52 msec percent correct for HT was correlated with control two at .47, p < .05 52 msec and r = .45, p < .05 for 66 msec.

For response time, HT and control two were correlated at r = -.44, p < .05 for 33 msec. At 66 msec, HT and SF subscores except control two were significantly correlated ranging from -.39 to .39, p < .05.

For male HT average scores and ASVAB scores, no significant correlations existed between WK and HT. Existing significant correlations were low (.13 to .24). Correlations between response times and ASVAB subscores tended to be negative as expected (See Addendum Table 10).

For females, no significant correlations existed between HT and the following ASVAB subscores: WK, PC, NO, CS, AS, and MK. Percent correct was correlated with GS at .56, p < .01 for 66 msec and at .49, p < .05 for 52 msec. Percent correct was correlated with AR at .46, p < .05 for 66 msec. Percent correct was correlated with MC at .59, p < .01 for 66 msec and at .53, p < .01

for 52 msec. Response time was correlated with EI at r = .39, p < .05 for 52

msec. There were fewer negative correlations between ASVAB subscores and HT

than for males (See Addendum Table 11)

DISCUSSION AND CONCLUSIONS

In general, males scored slightly higher than females on the majority of

tests and subtests. The significance of this however is not yet known due to

incomplete analysis. Large standard deriations and variances represent a wide

range of scores suggesting rather high as well as low scores for certain indi-

viduals. This also requires further testing for differences between means.

Significant correlations which existed between tests and subtests were

rather low for the majority i.e.: .1 to .4 with very few .5 or above. These

represent weak relationships although statistically significant.

As analyses remains in progress, it is difficult to come to a conclusion.

We do know however, that these tests produce a wide variety of scores which may

prove useful in future pilot selection through further test battery refinement.

REFERENCES

Endsley, M. R., Predictive Utility of an Objective Measure of Situation Aware-
ness. _Proceedings of the Human Factors Society 34th Annual Meeting_.
1990:41-57.

Endsley, M. R., Situation Awareness Global Assessment Technique (SAGAT), _Pro-
ceedings of the National Aerospace and Electronics Conference (NAECON)_.
May 1988b.

Endsley, M. R., SAGAT: A Methodology for the Measurement of Situation Aware-
ness. _Northrop Technical Report; NOR DOC 87-83_, August 1987.

Fracker, M. L., A Theory of Situation Assessment: Implications for Measuring
Situation Awareness. _Proceedings of the Human Factors Society 32nd Annual
Meeting_. October 1988: 102-106.

Hartman, B. O. and Secrist, G. E., Situational Awareness is More than Exception-
al Vision. _Aviation, Space, and Environmental Medicine_, November
1991:1084-1089.

Mane', A. M. and Dochin, E., The Space Fortress Game, _Acta Psychologica_ 71,
1989:17-22.

Mane', A. M., Acquisition of Perceptual Motor Skill: Adaptive and Part-Whole
Training. _Proceedings of the Human Factors Society 28th Annual Meeting_,
1984.

Ree, M. J., et.al., Equating and Implementation of Armed Services Vocational
Aptitude Battery. _AF-HR-LTP 85-21_. November 1985.

Secrist, G. E., et.al., Situational Awareness Training System (SATS): Descrip-
tion, Methodology and Applications. _USAF report no. AL-TP-1991-0054_.

Secrist, F. E., and Hartman, B. O., The Trainability of Certain Cognitive As-
pects of Situational Awareness. Report submitted to _The Department of
Aviation, Space and Environmental Medicine_.

Youngling, E. W., et.al., Feasibility study to Predict Combat Effectiveness for
Selected Military Roles: Fighter Pilot Effectiveness. St. Lois, MO:
McDonell Douglas Astronautics Company-East, April 1977, _MDC-E1634_.

ADDENDUM

Table 10

Correlation Matrix for Male's ASVAB
and Hartman Task (HT) Scores

	X̄=66	RT	X̄=52	RT	X̄=33	RT
GS	.22***	-.09	.21***	-.15****	.22***	-.01
AR	.07	-.03	.05	-.06	.13****	-.01
WK	.03	.05	.09	-.05	.10	-.00
PC	.09	.00	.14****	-.04	.09	-.02
NO	.10	-.09	.11	-.14****	.16****	-.11
CS	.11	-.13****	.13****	-.24**	.14****	-.16****
AS	.09	-.13****	.03	-.05	.01	-.08
MK	.16****	-.04	.17***	-.10	.19***	-.02
MC	.24***	-.09	.20***	-.02	.23**	.06
EI	.10	-.18***	.05	-.10	.08	-.05

*	$p < .0001$
**	$p < .001$
***	$p < .01$
****	$p < .05$

Table 11

Correlation Matrix for Female's ASVAB
and Hartman Task (HT) Scores

	X366	RT	X352	RT	X333	RT
GS	.56***	-.07	.49****	.03	.26	.02
AR	.46****	.24	.37	.08	.18	.04
WK	.34	-.22	.32	-.05	.08	-.04
PC	.29	-.29	.11	-.04	-.09	-.96
NO	.27	.24	.09	.21	-.17	.02
CS	.37	.08	.24	-.01	-.04	-.01
AS	.34	-.19	.31	.11	.17	.18
MK	.34	.02	.13	-.09	-.11	-.15
MC	.59***	-.07	.53***	.11	.21	.09
EI	.32	.04	.22	.39****	.03	.32

*	p < .0001
**	p < .001
***	p < .01
****	p < .05

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

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.