UNITED STATES AIR FORCE

1987 RESEARCH INITIATION PROGRAM

Conducted by

UNIVERSAL ENERGY SYSTEMS, INC.

under

USAF Contract Number F49620-85-C-0013

RESEARCH REPORTS

VOLUME IV OF IV

Submitted to

Air Force Office of Scientific Research

Bolling Air Force Base

Washington, DC

By

Universal Energy Systems, Inc.

April 1989

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 89-0829 32

6a. NAME OF PERFORMING ORGANIZATION: UNIVERSAL ENERGY SYSTEMS INC.

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research/XOT

6c. ADDRESS (City, State, and ZIP Code):
4401 Dayton Xenia Rd
Dayton OH 45432

7b. ADDRESS (City, State, and ZIP Code):
Building 410
Bolling AFB DC 20332

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL (If applicable): XOT

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-85-C-0013

8c. ADDRESS (City, State, and ZIP Code):
Building 410
Bolling AFB, DC 20332

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	3396	D5	

11. TITLE (Include Security Classification): USAF Research Initiation Program Volume 4 of 4

12. PERSONAL AUTHOR(S): Program Director Rodney C. Darrah

13a. TYPE OF REPORT: Interim

13b. TIME COVERED: FROM ________ TO ______

14. DATE OF REPORT (Year, Month, Day): April 1987

15. PAGE COUNT:

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number):

19. ABSTRACT (Continue on reverse if necessary and identify by block number):

(SEE REVERSE)

89 6 29 007

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Lt Col Claude Cavender

22b. TELEPHONE (Include Area Code): (202) 767-4971

22c. OFFICE SYMBOL: XOT

AFOSR-TR- 89-0829

32

INTRODUCTION

Research Initiation Program - 1987

AFOSR has provided funding for follow-on research efforts for the participants in the Summer Faculty Research Program. Initially this program was conducted by AFOSR and popularly known as the Mini-Grant Program. Since 1983 the program has been conducted by the Summer Faculty Research Program (SFRP) contractor and is now called the Research Initiation Program (RIP). Funding is provided to establish RIP awards to about half the number of participants in the SFRP.

Participants in the 1987 SFRP competed for funding under the 1987 RIP. Participants submitted cost and technical proposals to the contractor by 1 November 1987, following their participation in the 1987 SFRP.

Evaluation of these proposals was made by the contractor. Evaluation criteria consisted of:

1. Technical Excellence of the proposal
2. Continuation of the SFRP effort
3. Cost sharing by the University

The list of proposals selected for award was forwarded to AFOSR for approval of funding. Those approved by AFOSR were funded for research efforts to be completed by 31 December 1988.

The following summarizes the events for the evaluation of proposals and award of funding under the RIP.

A. Rip proposals were submitted to the contractor by 1 November 1987. The proposals were limited to $20,000 plus cost sharing by the universities. The universities were encouraged to cost share since this is an effort to establish a long term effort between the Air Force and the university.

B. Proposals were evaluated on the criteria listed above and the final award approval was given by AFOSR after consultation with the Air Force Laboratories.

C. Subcontracts were negotiated with the universities. The period of performance of the subcontract was between October 1987 and December 1988.

Copies of the Final Reports are presented in Volumes I through III of the 1987 Research Initiation Program Report. There were a total of 83 RIP awards made under the 1987 program.

UNITED STATES AIR FORCE

1987 RESEARCH INITIATION PROGRAM

Conducted by

UNIVERSAL ENERGY SYSTEMS, INC.

under

USAF Contract Number F49620-85-C-0013

RESEARCH REPORTS

VOLUME IV OF IV

Submitted to

Air Force Office of Scientific Research

Bolling Air Force Base

Washington, DC

By

Universal Energy Systems, Inc.

April 1989

TABLE OF CONTENTS

INTRODUCTION

Research Initiation Program - 1987

AFOSR has provided funding for follow-on research efforts for the participants in the Summer Faculty Research Program. Initially this program was conducted by AFOSR and popularly known as the Mini-Grant Program. Since 1983 the program has been conducted by the Summer Faculty Research Program (SFRP) contractor and is now called the Research Initiation Program (RIP). Funding is provided to establish RIP awards to about half the number of participants in the SFRP.

Participants in the 1987 SFRP competed for funding under the 1987 RIP. Participants submitted cost and technical proposals to the contractor by 1 November 1987, following their participation in the 1987 SFRP.

Evaluation of these proposals was made by the contractor. Evaluation criteria consisted of:

1. Technical Excellence of the proposal
2. Continuation of the SFRP effort
3. Cost sharing by the University

The list of proposals selected for award was forwarded to AFOSR for approval of funding. Those approved by AFOSR were funded for research efforts to be completed by 31 December 1988.

The following summarizes the events for the evaluation of proposals and award of funding under the RIP.

A. Rip proposals were submitted to the contractor by 1 November 1987. The proposals were limited to $20,000 plus cost sharing by the universities. The universities were encouraged to cost share since this is an effort to establish a long term effort between the Air Force and the university.

B. Proposals were evaluated on the criteria listed above and the final award approval was given by AFOSR after consultation with the Air Force Laboratories.

C. Subcontracts were negotiated with the universities. The period of performance of the subcontract was between October 1987 and December 1988.

Copies of the Final Reports are presented in Volumes I through III of the 1987 Research Initiation Program Report. There were a total of 83 RIP awards made under the 1987 program.

STATISTICS

Total SFRP Participants	159
Total RIP Proposals submitted by SFRP	117
Total RIP Proposals submitted by GSRP	7
Total RIP Proposals submitted	124

Total RIP's funded to SFRP	81
Total RIP's funded to GSRP	2
Total RIP's funded	83

Total RIP's Proposals submitted by HBCU's	11
Total RIP's Proposals funded to HBCU's	7

Laboratory	SFRP Participants	RIP's Submitted	RIP's Funded
AAMRL	13	12 (2 GSRP)	6
AFWAL/APL	8	6	4
ATL	9	8 (1 GSRP)	6
AEDC	6	4	3
AFWAL/AL	9	9 (1 GSRP)	5
LC	1	1	1
ESMC	1	0	0
ESD	1	1	0
ESC	8	8	6
AFWAL/FDL	9	8 (1 GSRP)	6 (1 GSRP)
FJSRL	9	5	5
AFGL	13	10 (1 GSRP)	7
HRL/OT	2	2	1
HRL/LR	3	3	2
HRL/MO	3	2	2
HRL/ID	0	0	0
LMC	3	1	0
AFWAL/ML	13	11 (1 GSRP)	6 (1 GSRP)
OEHL	5	3	3
AL	7	4	3
RADC	11	10	7
SAM	16	8	6
DEOMI	2	2	0
WL	7	6	4
Total	159	124	83

LIST OF UNIVERSITY THAT PARTICIPATED

Adelphi University	- 1	Meharry Medical College	- 1
Alabama, University of	- 1	Memphis State University	- 1
Alaska-Fairbanks, Univ. of	- 1	Metropolitan State College	- 1
Alfred University	- 1	Michigan State University	- 1
Arizona State University	- 1	Mississippi State University	- 4
Arkansas State University	- 1	Mississippi, University of	- 1
Arkansas, University of	- 1	Missouri-Kansas City, Univ.	- 1
Auburn University	- 1	Missouri-Rolla, Univ. of	- 3
Bishop College	- 1	Montana, University of	- 1
Capital University	- 1	Montclair State College	- 1
Catholic Univ. of America	- 1	Morehouse College	- 1
Cedarville College	- 1	Nazareth College	- 1
Central State University	- 1	Nebraska-Lincoln, Univ. of	- 2
Cincinnati, University of	- 5	New Mexico State University	- 1
Colorado, University of	- 2	New York State, Univ. of	- 3
Dayton, University of	- 7	N. Carolina A&T State Univ.	- 1
Dillard University	- 1	N. Carolina-Greensboro, Univ	- 1
Drury College	- 1	Northwestern University	- 1
Eastern Illinois University	- 1	Ohio State University	- 5
Eastern Kentucky University	- 1	Ohio University	- 2
Eastern New Mexico University	- 2	Oklahoma State University	- 1
Fairfield University	- 1	Oregon Institute of Tech.	- 1
Florida A&M University	- 1	Oregon State University	- 1
Florida, University of	- 2	Ouachita Baptist University	- 1
Fort Lewis College	- 1	Pace University	- 1
Gonzaga University	- 1	Pennsylvania State Univ.	- 1
Grambling State University	- 1	Point Loma College	- 1
Hampton University	- 1	Puerto Rico-Mayaguez, Univ.	- 1
Houston, University of	- 2	Purdue University	- 1
Howard University	- 1	Rochester Inst. of Tech.	- 1
Idaho, University of	- 1	Rose-Hulman Inst. of Tech.	- 2
Illinois-Chicago, Univ. of	- 2	Saint Paul's College	- 1
Indiana University	- 1	San Francisco State Univ.	- 1
Indiana Univ. of Pennsylvania	- 1	South Dakota State Univ.	- 1
Iowa, University of	- 1	South Florida, University of	- 2
Jackson State University	- 1	Southeastern Mass. Univ.	- 2
Jarvis Christian College	- 1	Southern Illinois University	- 2
Jesm Baromedical Res. Inst.	- 1	Southern Mississippi, Univ.	- 1
John Hopkins Evening College	- 1	Southern University	- 2
Kansas State University	- 1	St. Louis University	- 1
Kansas, University of	- 1	St. Mary's University	- 1
Kentucky, University of	- 1	Talladega College	- 1

Continued

LIST OF UNIVERSITY THAT PARTICIPATED
Continued

Lock Haven Univ. of Pennsyl.	- 1	Taylor University	- 1
Long Island University	- 1	Temple University	- 1
Louisiana State University	- 1	Tennessee Technical Univ.	- 1
Louisiana Tech. University	- 1	Tennessee, University of	- 1
Lowell, University of	- 4	Texas A&M University	- 2
Texas Southern University	- 3	Wichita State University	- 2
Texas Technical University	- 2	Wilberforce University	- 1
Texas-Austin, University of	- 1	Wisconsin-Eau Claire Univ.	- 2
Tuskegee University	- 1	Wisconsin-Madison, Univ. of	- 1
Utah State University	- 1	Wisconsin-Whitewater, Univ.	- 1
Walla Walla College	- 1	Wittenberg University	- 1
Washington State University	- 1	Worchester Polytech. Inst.	- 2
West Florida, University of	- 1	Wright State University	- 3
Western Michigan University	- 3	Xavier University	- 1

PARTICIPANTS LABORATORY ASSIGNMENT

PARTICIPANT LABORATORY ASSIGNMENT (Page 1)

AERO PROPULSION LABORATORY
(Wright-Patterson Air Force Base)

Dr. Suresh K. Aggarwal
Univ. of Illinois at Chicago
Specialty: Aerospace Engineering

Dr. Bryan R. Becker
Rose-Hulman Institute
Specialty: Engineering Science

Dr. Richard Tankin
Northwestern University
Specialty: Mechanical Engineering

Dr. Cheng-Hsiao Wu
Univ. of Missouri
Specialty: Solid State Physics

ARMAMENT LABORATORY
(Eglin Air Force Base)

Dr. Charles Bell
Arkansas State University
Specialty: Mechanical Engineering

Dr. Robert W. Courter
Louisiana State University
Specialty:Aerospace Engineering

Dr. Joseph J. Feeley
University of Idaho
Specialty:Electrical Engineering

Ms. Jennifer L. Davidson (1986), (GSRP)
University of Florida
Specialty: Mathematics

Dr. Mo Samimy (1986)
Ohio State University
Specialty: Mechanical Engineering

Dr. Elmer C. Hansen
University of Florida
Specialty:Mechanical Engineering

Dr. James Hoffmaster
Gonzaga University
Specialty:Physics

Dr. James Nail
Mississippi State Univ.
Specialty:Electrical Engineering

Dr. Meckinley Scott (1986)
University of Alabama
Specialty: Statistics

Mr. Jim S. Sirkis (1986),(GSRP)
University of Florida
Specialty: Engineering Mechanics

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
(Wright-Patterson Air Force Base)

Dr. Praphulla K. Bajpai
University of Dayton
Specialty: Immunology

Dr. Gwendolyn Howze
Texas Southern University
Specialty: Physics

Dr. Thomas Nygren
Ohio State University
Specialty: Psychology

Dr. Donald Robertson
Indiana University of PA
Specialty: Psychology

PARTICIPANT LABORATORY ASSIGNMENT (Page 2)

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
(Wright-Patterson Air Force Base)
(continued)

Dr. Noel Nussbaum
Wright State University
Specialty: Biology

Dr. John Westerkamp
University of Dayton
Specialty: Electrical Engineering

Dr. Jacqueline Paver (1986)
Duke University
Specialty: Biomechanical Engineering

ARNOLD ENGINEERING DEVELOPMENT CENTER
(Arnold Air Force Systems)

Dr. Suhrit K. Dey
Eastern Illinois University
Specialty: Aerospace Engineering

Dr. Surgounda Patil
Tennessee Technical University
Specialty: Math Statistics

Dr. William M. Grissom
Morehouse College
Specialty: Mechanical Engineering

ASTRONAUTICS LABORATORY
(Edwards Air Force Base)

Dr. Gurbux S. Alag
Western Michigan University
Specialty: Systems Engineering

Dr. Lawrence Schovanec
Texas Tech University
Specialty: Mathematics

Dr. John Kenney
Eastern New Mexico University
Specialty: Physical Chemistry

AVIONICS LABORATORY
(Wright-Patterson Air Force Base)

Dr. Vernon L. Bakke
University of Arkansas
Specialty: Mathematics

Dr. Narayan C. Halder
University of South Florida
Specialty: Physics

PARTICIPANT LABORATORY ASSIGNMENT (Page 3)

AVIONICS LABORATORY
(Wright-Patterson Air Force Base)
(continued)

Prof. William K. Curry
Rose-Hulman Inst. of Technology
Specialty: Computer Science

Dr. Verlynda S. Dobbs
Wright State University
Specialty: Computer Science

Dr. George W. Zobrist (1986)
University of Missouri-Rolla
Specialty: Electrical Engineering

Dr. Alastair McAulay
Wright State University
Specialty: Electrical Engineering

Dr. John Y. Cheung (1986)
University of Oklahoma
Specialty: Electrical Engineering

ENGINEERING AND SERVICES CENTER
(Tyndall Air Force Base)

Dr. William W. Bannister
University of Lowell
Specialty: Organic Chemistry

Dr. William M. Bass
The University of Tennessee
Specialty: Physical Anthropology

Dr. Peter Jeffers
S.U.N.Y.
Specialty: Chemistry

Dr. William T. Cooper (1986)
Florida State University
Specialty: Chemistry

Dr. William Schulz
Eastern Kentucky University
Specialty: Chemistry

Dr. Joseph Tedesco
Auburn University
Specialty: Civil Engineering

Dr. Dennis Truax
Mississippi State Univeristy
Specialty: Civil Engineering

Dr. Yong S. Kim (1986)
The Catholic Univ. of America
Specialty: Civil Engineering

FLIGHT DYNAMICS LABORATORY
(Wright-Patterson Air Force Base)

Mr. Thomas Enneking (GSRP)
University of Notre Dame
Specialty: Civil Engineering

Dr. Gary Slater
University of Cincinnati
Specialty: Aerospace Engineering

PARTICIPANT LABORATORY ASSIGNMENT (Page 4)

FLIGHT DYNAMICS LABORATORY
(Wright-Patterson Air Force Base)
(continued)

Dr. Oliver McGee
Ohio State University
Specialty: Engineering Mechanics

Dr. Forrest Thomas
University of Montana
Specialty: Chemistry

Dr. Shiva Singh
Univ. of Kentucky
Specialty: Mathematics

Dr. William Wolfe
Ohio State University
Specialty: Engineering

Dr. George R. Doyle (1986)
University of Dayton
Specialty: Mechanical Engineering

Dr. V. Dakshina Murty (1986)
University of Portland
Specialty: Engineering Mechanics

Dr. Tsun-wai G. Yip (1986)
Ohio State University
Specialty: Aeronautics-Astronautics Engineering

FRANK J. SEILER RESEARCH RESEARCH LABORATORY
(United State Air Force Academy)

Dr. Charles M. Bump
Hampton University
Specialty: Organic Chemistry

Dr. Howard Thompson
Purdue University
Specialty: Mechanical Engineering

Dr. Stephen J. Gold
South Dakota State University
Specialty: Electrical Engineering

Dr. Melvin Zandler
Wichita State Univ.
Specialty: Physical Chemistry

Dr. Henry Kurtz
Memphis State Univ.
Specialty: Chemistry

GEOPHYSICS LABORATORY
(Hanscom Air Force Base)

Dr. Lee A. Flippin
San Francisco State Univ.
Specialty: Organic Chemistry

Dr. Gandikota Rao
St. Louis University
Specialty: Meteorology

PARTICIPANT LABORATORY ASSIGNMENT (Page 5)

GEOPHYSICS LABORATORY
(Hanscom Air Force Base)
(continued)

Dr. Mayer Humi
WPI
Specialty: Applied Mathematics

Dr. Steven Leon
Southeastern Massachusettes
Specialty: Mathematics

Dr. Henry Nebel
Alfred University
Specialty: Physics

Dr. Timothy Su
Southeastern Massachusetts Univ.
Specialty: Physical Chemistry

Dr. Keith Walker
Point Loma College
Specialty: Physics

HUMAN RESOURCES LABORATORY
(Brooks, Williams and Wright-Patterson Air Force Base)

Dr. Patricia A. Carlson
Rose-Hulman Inst. of Technology
Specialty: Literature/Language

Dr. Ronna E. Dillon
Southern Illinois University
Specialty: Educational Psychology

Dr. Michael Matthews
Drury College
Specialty: Psychology

Dr. Stephen Loy (1986)
Iowa State University
Specialty: Management Information Sys.

Dr. Doris Walker-Dalhouse (1986)
Jackson State University
Specialty: Reading Education

Dr. John Uhlarik
Kansas State University
Specialty: Psychology

Dr. Charles Wells
University of Dayton
Specialty: Management Science

Dr. Charles Lance (1986)
University of Georgia
Specialty: Psychology

Dr. Jorge Mendoza
Texas A&M University
Specialty: Psychology

Dr. Billy Wooten (1986)
Brown University
Specialty: Philosophy, Psychology

LOGISTICS COMMAND
(Wright-Patterson Air Force Base)

Dr. Howard Weiss
Specialty: Industrial Engineering
Temple University

MATERIALS LABORATORY
(Wright-Patterson Air Force Base)

Dr. Bruce A. DeVantier
S. Illinois University
Specialty: Civil Engineering

Dr. Ravinder Diwan
Southern University
Specialty: Metallurgy

Dr. Bruce A. Craver
University of Dayton
Specialty: Physics

Dr. Robert Patsiga (1986)
Indiana Univ. of Pennsylvania
Specialty: Organic Polymer Chemistry

Dr. Gopal M. Mehrotra (1986)
Wright State University
Specialty: Metallurgy

Dr. John W. Gilmer
Penn State University
Specialty: Physical Chemistry

Dr. Gordon Johnson
Walla Walla College
Specialty: Electrical Engineering

Mr. John Usher (GSRP)
Louisiana State University
Specialty: Chemical Engineering

Dr. Nisar Shaikh (1986)
University of Nebraska-Lincoln
Specialty: Applied Mathematics

OCCUPATIONAL AND ENVIRONMENT HEALTH LABORATORY
(Brooks Air Force Base)

Dr. Richard H. Brown
Ouachita Baptist University
Specialty: Physiology

Dr. Elvis E. Deal
University of Houston
Specialty: Industrial Engineering

Dr. Kiah Edwards
Texas Southern University
Specialty: Molecular Biology

Dr. Ralph J. Rascati (1986)
Kennesaw College
Specialty: Biochemistry

ROME AIR DEVELOPMENT CENTER
(Griffis Air Force Base)

Prof. Beryl L. Barber
Oregon Institute of Technology
Specialty: Electrical Engineering

Dr. Kevin Bowyer
University of South Florida
Specialty: Computer Science

Dr. Ronald V. Canfield
Utah State University
Specialty: Statistics

Dr. Lionel R. Friedman
Worcester Polytechnic Inst.
Specialty: Physics

Dr. John M. Jobe (1986)
Miami University of Ohio
Specialty: Statistics

Dr. Louis Johnson
Oklahoma State Univ.
Specialty: Electrical Engineering

Dr. Panapkkam Ramamoorthy
University of Cincinnati
Specialty: Electrical Engineering

Dr. David Sumberg
Rochester Institute of Tech.
Specialty: Physics

Dr. Donald Hanson (1986)
University of Mississippi
Specialty: Electrical Engineering

Dr. Stephen T. Welstead (1986)
University of Alabama in Hunts.
Specialty: Applied Mathematics

SCHOOL OF AEROSPACE MEDICINE
(Brooks Air Force Base)

Prof. Phillip A. Bishop
University of Alabama
Specialty: Exercise Physiology

Dr. Mohammed Maleque
Meharry Medical College
Specialty: Pharmacology

Dr. Kurt Oughstun
University of Wisconsin
Specialty: Optical Sciences

Dr. Hoffman H. Chen (1986)
Grambling State University
Specialty: Mechanical Engineering

Dr. Ralph Peters
Wichita State University
Specialty: Zoology

Dr. Stephen Pruett
Mississippi State University
Specialty: Immunology

Dr. Wesley Tanaka
University of Wisconsin
Specialty: Biochemistry

Dr. Vito DelVecchio (1986)
University of Scranton
Specialty: Biochemistry, Genetics

PARTICIPANT LABORATORY ASSIGNMENT (Page 8)

WEAPONS LABORATORY
(Kirtland Air Force Base)

Dr. Jerome Knopp
University of Missouri
Specialty: Electrical Engineering

Dr. Barry McConnell
Florida A&M University
Specialty: Computer Science

Dr. Martin A. Shadday, Jr. (1986)
University of South Carolina
Specialty: Mechanical Engineering

Dr. Randall Peters
Texas Tech University
Specialty: Physics

Dr. William Wheless
New Mexico State University
Specialty: Electrical Engineering

RESEARCH REPORTS

MINI-GRANT RESEARCH REPORTS

1987 RESEARCH INITIATION PROGRAM

Technical Report Number	Title and Mini-Grant No.	Professor
Volume I		
Armament Laboratory		
1	Report Not Available at this Time 760-7MG-025	Dr. Charles Bell
2	Effects of Bending Flexibility on the Aerodynamic Characteristics of Slender Cylinders Determined from Free-Flight Ballistic Data 760-7MG-018	Dr. Robert W. Courter
3	Image Complexity Measures and Edge Detection 760-6MG-024	Ms. Jennifer L. Davidson (1986 GSRP)
4	Report Not Available at this Time 760-7MG-070	Dr. Joesph J. Feeley
5	Advanced Gun Gas Diversion 760-7MG-012	Dr. Elmer Hansen
6	A Physical and Numerical Study of Pressure Attenuation in Solids 760-7MG-002	Dr. James Hoffmaster
7	Pyroelectric Sensing for Potential Multi-Mode Use 760-7MG-026	Dr. James Nail
8	Gaseous Fuel Injection and Mixing in a Supersonic Combustor 760-6MG-059	Dr. Mo Samimy (1986)
9	Systems Effectiveness for Targets with Repair or Replacement Facilities of Damaged Components 760-6MG-025	Dr. Meckinley Scott (1986)
10	A Pattern Recognition Application in Elastic-Plastic Boundary Element, Hybrid Stress Analysis 760-6MG-142	Mr. Jim S. Sirkis (1986 GSRP)

Arnold Engineering Development Center		
11	Vectorized Perturbed Functional Iterative Scheme (VPFIS): A Large-Scale Nonlinear System Solver 760-7MG-037	Dr. Suhrit K. Dey
12	Liquid Film Cooling in Rocket Engines 760-7MG-022	Dr. William M. Grissom
13	Estimation of Autocorrelation and Power Spectral Density for Randomly Sampled Systems 760-7MG-085	Dr. Surgounda Patil
Astronautics Laboratory		
14	Report Not Available at this Time 760-7MG-042	Dr. Gurbux S. Alag
15	Report Not Available at this Time 760-7MG-019	Dr. John Kenney
16	Fracture in Soild Propellant: Damage Effects upon Crack Propagation 760-7MG-065	Dr. Lawrence Schovanec
17	Novel Conversion of Organometallics to Energetic Nitro Compunds 760-6MG-130	Dr. Nicholas E. Takach (1986)
Engineering and Services Center		
18	Correlations of Spontaneous Ignition Temperatures with Molecular Structures of Flammable Compounds 760-7MG-101	Dr. William W. Bannister
19	The Estimation of Stature from Fragments of the Femur: A Revision of the Steele Method 760-7MG-014	Dr. William M. Bass
20	Effects of Water Solubility and Functional Group Content on the Interactions of Organic Solutes with Soil Organic Matter 760-6MG-081	Dr. William T. Cooper (1986)
21	Report Not Available at this Time 760-7MG-038	Dr. Peter Jeffers

22	A Study of Semihardened Concrete Arch Structure Response Under Protective Layers 760-6MG-004	Dr. Yong S. Kim (1986)
23	Report Not Available at this Time 760-7MG-079	Dr. William Schulz
24	Stress Wave Propagation in Layered Media 760-7MG-034	Dr. Joseph Tedesco
25	Report Not Available at this Time 760-7MG-105	Dr. Dennis Truax

Volume II

Frank J. Seiler Research Laboratory

26	Report Not Available at this Time 760-7MG-076	Dr. Charles M. Bump
27	The Omnidirectional Torquer - Experimental Prototype Model I 760-7MG-123	Dr. Stephen J. Gold
28	Calculation of Nonlinear Optical Properties 760-7MG-030	Dr. Henry Kurtz
29	Report Not Available at this Time 760-7MG-071	Dr. Howard Thompson
30	Report Not Available at this Time 760-7MG-092	Dr. Melvin Zandler

Geophysics Laboratory

31	Report Not Available at this Time 760-7MG-056	Dr. Lee A. Flippin
32	Modelling and Prediction in a Nonlocal Turbulence Model 760-7MG-028	Dr. Mayer Humi
33	Report Not Available at this Time 760-7MG-036	Dr. Steven Leon
34	CO_2 (001) Vibrational Temperatures and Limb-View Infrared Radiances Under Terminator Conditions in the 60-100 Altitude Range 760-7MG-035	Dr. Henry Nebel

35	Comparison of SSM/I Rainrates and Surface Winds with the Corresponding Conventional Data in the North West Pacific Typhoons 760-7MG-072	Dr. Gandikota Rao
36	Report Not Available at this Time 760-7MG-040	Dr. Timothy Su
37	Development of a System for the Measurement of Electron Excitation Cross Sections of Atoms and Molecules in the Near Infrared 760-7MG-074	Dr. Keith Walker

Rome Air Development Center

38	Superconductor Testing 760-7MG-103	Prof. Beryl L. Barber
39	A Form and Function Knowledge Representation for Reasoning about Classes and Instances of Objects 760-7MG-003	Dr. Kevin Bowyer
40	Development and Evaluation of a Bayesian Test for System Testability 760-7MG-032	Dr. Ronald V. Canfield
41	Crystalline Silicon Electro-Optic Waveguides 760-7MG-040	Dr. Lionel R. Friedman
42	Measurements of a Slot Antenna Fed by Coplanar Waveguide and Solution of an Infinite Phased Array of Slots Fed by Coplanar Waveguide Over a Dielectric Half-Space 760-6MG-092	Dr. Donald Hanson (1986)
43	A New Measure of Maintainability/ Reliability and Its Estimation 760-6MG-019	Dr. John M. Jobe (1986)
44	Report Not Available at this Time 760-7MG-050	Dr. Louis Johnson
45	Signed-Digit Number System for Optical Adaptive Processing 760-7MG-015	Dr. Panapkkam Ramamoorthy

Contents.

Volume IV

Human Systems Division Laboratories

Harry G. Armstrong Aerospace Medical Research Laboratory

Human Resources Laboratory

94	Comparison of Supervisor's and Incumbent's Estimates of SDy 760-7MG-009	Dr. Michael Matthews
95	Report Not Available at this Time 760-6MG-136	Dr. Jorge Mendoza (1986)
96	The Role of Fourier Descriptions for Shape in Visual Form Perception 760-7MG-082	Dr. John Uhlarik
97	Comprehensibility of Technical Text 760-6MG-080	Dr. Doris Walker-Dalhouse (1986)
98	Report Not Available at this Time 760-7MG-046	Dr. Charles Wells
99	Mechanisms of Contrast and Lightness Constancy 760-6MG-051	Dr. Billy Wooten (1986)
Occupational and Environment Health Laboratory		
100	Phytotoxicity of Soil Residues of JP-4 Aviation Fuel 760-7MG-059	Dr. Richard H. Brown
101	An Impact Study for the Contracting Out of In-House Analytical Services at the USAF Occupational & Environmental Health Laboratory - Brooks AFB, San Antonio, Texas 760-7MG-096	Dr. Elvis E. Deal
102	Effects of Metal Mutagens on the Synthesis and Accumulation of Macromolecules 760-7MG-001	Dr. Kiah Edwards
103	Development of a Rapid and Sensitive Assay Procedure for the Detection of the Protozoan Parasite Giardia Lamblia in Drinking Water Supplies 760-6MG-062	Dr. Ralph J. Rascati (1986)
School of Aerospace Medicine		
104	Limitations to Heavy Work of Personnel Wearing at 21°C: U.S. Military Chemical Defense Ensemble. 760-7MG-067	Prof. Phillip A. Bishop (JHD)

105	Report Not Available at this Time 760-6MG-118	Dr. Hoffman Chen (1986)
106	Nucleic Acid Hybridzation - Dot Blot Test for the Presence of Ureaplasma Urealyticum and Mycoplasma Hominis 760-6MG-076	Dr. Vito DelVecchio (1986)
107	Report Not Available at this Time 760-7MG-078	Dr. Mohammed Maleque
108	The Asymptotic Description of Precursor Fields in a Causally Dispersive Medium 760-7MG-033	Dr. Kurt Oughstun
109	Report Not Publishable at this Time 760-7MG-091	Dr. Ralph Peters
110	Model Systems for Assessing the Effects of Microwave Radiation on the Immune System 760-7MG-060	Dr. Stephen Pruett
111	Report Not Available at this Time 760-7MG-043	Dr. Wesley Tanaka

Development of Implantable Devices for Sustained Delivery of Volatile Hydrocarbons in Rats

by

Praphulla K. Bajpai and Deborah E. Hollenbach

Abstract

A ceramic-glass reservoir delivery system (CGRS) was developed for studying the toxicity of 1,1,1-trichloroethane (TCE) in animals. Four factors, including the density of the ceramic, the blending of TCE with 5% mineral oil, the size of the opening at the neck of the glass reservoir, and the initial volume of solvent placed in the reservoir were studied *in vitro*. The density of the ceramic had no effect on the release rate of TCE, while blending the solvent in 5% mineral oil and varying the size of the neck opening of the glass tube altered the release rate of TCE. The initial volume of TCE placed in the CGRS also influenced the duration of the release of the solvent.

Fisher 344 rats (225-250g) were implanted subcutaneously behind the neck with CGRS devices. The empty implanted CGRS devices were injected with 0.25ml, 0.50ml, or 1.0ml TCE 4 days after implantation. CGRS devices charged with initial volumes of 0.25ml, 0.50ml or 1.0ml TCE maintained constant levels of TCE in the blood for 48 hours. A battery of behavioral experiments were conducted forty eight hours after injecting the solvent into the implanted CGRS devices. These consisted of two methods for detecting locomotor activity (Opto-Varimex and Figure Eight Maze) and monitoring of the auditory startle reflex. Only rats implanted with CGRS devices containing an initial volume of 1.0ml TCE showed a decrease in performance and activity.

Acknowledgements

Debbie Hollenbach (who completed her M.Sc. Thesis under this program) and I wish to thank the Air Force Toxic Hazards Division of the Armstrong Aerospace Medical Research Laboratory at Wright Patterson Air Force Base for the sponsorship of this research. We also wish to thank the Universal Energy Systems for the sponsorship of this program.

Specifically, we wish to thank SSgt. Greg Cason for teaching us how to use the various gas chromatographs and for technical assistance throughout these investigations. Thanks are due to Miss Brigitta Lusser for assisting us throughout the later stages of this project. We sincerely appreciate the time, technical support, and expertise provided by Mr. Michael Gargas. We are indebted to Dr. R. Drawbaugh for providing moral support, encouragement, and constructive suggestions. We want to thank Dr. J. J. Cooper for guiding us in the behavioral studies, and giving us the opportunity to work at the Veterinary Sciences Division of AAMRL. We are grateful to Dr. David R. Mattie for bringing to Dr. Andersen's attention the expertise available at The University of Dayton in Drug Delivery Systems, and providing help in all aspects of this project. Finally, we want to thank Dr. M. E. Andersen for giving us the opportunity to work at the Toxic Hazards Laboratory under the AFOSR, RIP Program, sponsored by UES, directing us in our endeavors, and providing the facilities and personnel to carry out the proposed studies.

1.0. INTRODUCTION:

A volatile chemical 1,1,1-Trichloroethane (TCE), is commonly used as a dry cleaning agent and as an organic solvent for natural and synthetic resins, oils, waxes, tar and alkaloids (1). It is also used as a cleansing solvent for electrical machinery and plastics, and is present as a contaminent in drinking water. Exposure to TCE and other volatile hydrocarbons is usually by inhalation or absorption through the skin (2-5). Since the potential for ground water contamination (especially in sandy soil areas) is high, dermal exposures to TCE can also occur through recreational use of contaminated waters.

The main effects of a single vapor exposure to TCE is a functional dose-related depression of the central nervous system (CNS) that may manifest itself in dizziness, loss of coordination, or respiratory distress (6). However, long term effects are virtually unknown due to the difficulty of maintaining constant levels of TCE, using inhalation experimental procedures. Recently, an alumino calcium phosphorous oxide (ALCAP) device was developed for in vitro delivery of chemicals and biologicals (7). The hollow ceramic cylinders have proven to be effective as sustained delivery devices for polypetides, steroids and other compounds (7). This investigation was conducted to develop a simple, implantable system which will deliver volatile solvents such as TCE in a sustained manner, using ALCAP ceramics. Both in vitro and in vivo procedures were used to develop the device. The release of TCE in air and blood was monitored by gas chromatography. The effectiveness of the delivery system was investigated by monitoring the behavioral performance of rats implanted with TCE containing CGRS devices. The behavioral performance was measured by two locomotor monitoring devices and the reflex responses of auditory startle.

2.0. OBJECTIVES OF THE RESEARCH EFFORT:

This project was a continuation of work initiated during the summer of 1987 as part of the SFR program. The primary objective of this project was to develop an implantable system for delivering volatile hydrocarbons, and to study the acute and chronic effects of TCE, delivered by such a device, in rats.

The objectives were accomplished by completing the following tasks:

a. Fabrication of ceramic(s) of appropriate density, porosity, and surface area for use in association with a non-porous reservoir.

b. Fabrication of non-porous tubes and refillable reservoirs with the capability of delivering TCE at various rates for both long and short term studies.

c. Continuation of the work initiated during the SFR program on the release rate of TCE *in vitro* and *in vivo* and studying its effect on the behavior and pathophysiology of animals.

3.0. MATERIALS AND METHODS:

The development of a ceramic-glass reservoir sustained delivery system (CGRS) for TCE was conducted in two phases to minimize the use of experimental animals. In phase I, the release of TCE from CGRS devices was examined in an *in vitro* free air flowing environment. In phase II, the delivery of TCE from CGRS devices and its effects on animals were investigated.

3.1. Chemical.

High purity 1,1,1-trichloroethane (99.9% pure) was obtained from the Aldrich Chemical Company (Milwaukee, WI). Some experiments in this study were conducted with TCE blended with 5% mineral oil.

3.2. Fabrication of ALCAP Ceramics.

Alumino calcium phosphorus oxide (ALCAP) ceramic capsules were fabricated by calcining a mixture (50:30:16 by weight) of aluminum oxide, calcium oxide, and phosphorus pentoxide powders (Fisher Scientific Co., Fairlawn, NJ) at 1315^0C, in a high temperature furnace (Thermolyne Inc. Dubuque, IO) for 12 hours. The calcined material was ground in a jar mill (U.S. Stoneware, Mahewah, NJ) and sized with a 400 mesh screen (No. 400, Fisher Scientific Co., Fairlawn, NJ) to obtain particles of less than 38um in diameter. Ceramic cylinders (green shape) were fabricated by compressing a mixture of one gram of ceramic particles (-38um) and 0.025g polyvinyl alcohol (PVA, 60um) in a 5/16" die at loads of 1365, 2275, and 3635Kg on a French Cell Press (American Instrument Co., Silver Spring, MD). The green shape cylinders were then sintered at 1400^0C for 36 hours to increase the density and mechanical strength of the ceramic. The dimensions and weight of each sintered ceramic cylinder were recorded and their densities were calculated using the following formula.

$$\text{Density } (g/cm^3) = \text{Weight/Actual Volume}$$

3.3. Glass Tubes and Reservoirs.

Eighteen 1cm x 4mm glass tubes, having an opening of 1.0mm were obtained from the glass blowing laboratory of the University of Dayton. The glass tubes were inserted into the ceramic, then each end of the ceramic cylinder was sealed with Silastic Medical Adhesive, Silicone Type A (Dow Corning, Midland MI). One and two milliliter Wheaton autosampler vials (Supelcoport Co., Bellefonte, PA) were modified into reservoirs by attaching 1cm x 4mm glass tubes with openings of 0.5mm, 1.0mm, 1.5mm, or 2.0mm respectively. Each reservoir was then closed with a crimp top, fitted with a rubber and teflon septa. 1,1,1-TCE was introduced into the reservoir by injecting through the septa into the sealed reservoir. Each reservoir was attached to a ceramic cylinder by inserting the 1cm x 4mm glass tube into the ceramic, and sealing the ceramic and the opening around the glass tube with Silastic Medical Adhesive (Figure 1).

3.4. Standardization of the Gas Chromatograph for Analysis of 1,1,1-TCE in Air.

A Hewlett Packard 5890 Gas Chromatograph (GC), containing a flame ionization detector (FID), and a 10', 10% SE-30, 80/100 Supelcoport^R column was standardized in the 10-100 parts per million (PPM) range, using 20 liter Mylar^R bags. The oven temperature was maintained at 100^{o}C, the injector temperature at 125^{o}C, the detector temperature at 300^{o}C, and the nitrogen carrier gas was maintained at a flow rate of 25ml/min. A Hewlett Packard 3393A Integrator and a Hewlett Packard 19405A Sampler/ Event Control Module were used in conjunction with the gas chromatograph for analysis of the 1.0ml samples injected into the GC at each sampling time. For conducting the *in vitro* experiments, each ceramic-glass tube or reservoir system was suspended inside a 108ml glass gas injection chamber on a silk string. The chamber was connected to a reciprocating electromagnetic piston air pump which was connected to the gas chromatograph. A constant air flow of 120ml/min was maintained through the glass chamber by the air pump. The HP Sampler/Event Control Module was utilized to provide automatic sampling of the glass chamber. In order to determine the concentration of TCE released by the delivery system in the gas chamber the control module was programmed to automatically sample the contents of the gas chamber, by the GC at five minute intervals.

3.5. Standardization of the Gas Chromatograph for Analysis of 1,1,1,-TCE in Blood.

A Hewlett Packard 5700A gas chromatograph was fitted with an electron capture detector (ECD), equipped with a glass column (8' x 2mm i.d), and packed with 10% SE-30, 80/100 Supelcoport (Bellefonte, PA). The oven temperature was maintained at 70^0C, the injector temperature was 125^0C, and the detector was set at 250^0C. The carrier flow rate was 25ml/min argon:ethane (95:5). Standards of TCE (0.05 to 10ppm) in hexane were prepared for constructing the standard curve. One milliliter of hexane and TCE standard was placed in a 2ml autosampler vial containing 100ul of unexposed rat blood. Following a one hour equilibration time period, 1ul of the hexane layer was analyzed by GC to detect the TCE. A Hewlett Packard 3392A Integrator was used in conjunction with the GC.

3.6. Behavior Experiments.

Three different experiments were conducted to assess the behavioral changes of rats in this investigation.

3.6.1. Opto-Varimex.

Four rats were introduced into individual Plexiglas boxes (17" x 17" x 8") placed on top of four Opto-Varimex (Columbus Instruments, Columbus, OH) platforms, with both vertical emitters and vertical detectors. The platforms were connected by an Opto-Varimex Autotract Interface to an IBM

Personal Computer for data collection and analysis. Opto-Varimex consists of infrared beams that divide the 17" square box into 16 smaller squares. Opto-Varimex records the time (in seconds) each rat takes to perform the following activities: distance traveled (DT), time resting (TR), time stereotypic (TS), time ambulatory (TA), small movements (SM). The data was obtained in three consecutive sessions.

3.6.2. Figure Eight Maze.

The Figure Eight Maze was constructed of aluminum and mounted on Plexiglas. Each Figure Eight Maze is 3'x 18" x 6" in size and contains eight photoelectric cells. The emitters and detectors in each maze are connected to 16 channel Multi-Varimex Interface master or expansion units (Columbus Instruments, Columbus, OH). The interface units are connected to an IBM Personal Computer which utilizes a Data Collection Center software package (Columbus Instruments, Columbus, OH) for recording and analyzing the activity. A total of eight mazes can be utilized simultaneously. Activity is measured by the number of times each individual beam is broken by the movement of the rat during three consecutive twenty minute sessions.

3.6.3. Auditory Startle Reflex.

Each startle Apparatus is enclosed in a large insulated box (24" x 20" x 18"). Each box has a Responder-X Force platform (Columbus Instruments, Columbus, OH) with a plexiglas box (11.5" x 6.5" x 6.5") placed over it to contain the rat. The Responder-X Force Platforms are connected to an Isolation Interface Unit and to an IBM Personal Computer. The computer uses a Micro-Responder Software (Columbus Instruments, Columbus, OH). An Audio generator (Columbus Instruments, Columbus, OH) is connected to a NAD Stereo amplifier #3020B. A Realistic Stereo Speaker (Tandy, Inc. Dallas, TX) is mounted on the inside wall of each large box. The audio generator emits a series of pre-programmed auditory stimuli of varying frequencies (3200-20000Hz), amplitudes, and durations (20-100msec). The Responder-X measures the responses of the animal to each stimulus. A pre-programmed series of learn, time, and amplitude (Amp) is used for each run. Learn consists of a series of auditory stimuli which acquaints the animal with a sound. The animal then responds to the sound. Time and Amp consists of a series of auditory stimuli which includes pre-pulses. The degree of the response, as well as the time it takes the rat to respond to the stimuli, are recorded.

TABLE 1

Experimental Design for the Effects of Density and Mineral Oil on the Release Rate of 1,1,1-TCE from Ceramic Cylinders and from Ceramic-Glass Tube Systems (CGTS).

Group (n=3)	Type of System	Compression Load (Kg)	Contents
1	Ceramic	1365	47mg 1,1,1-TCE
2	Ceramic	2275	47mg 1,1,1-TCE
3	Ceramic	3635	47mg 1,1,1-TCE
4	Ceramic	1365	44mg 1,1,1-TCE in 5% mineral oil
5	Ceramic	2275	44mg 1,1,1-TCE in 5% mineral oil
6	Ceramic	3536	44mg 1,1,1-TCE in 5% mineral oil
7	CGTS	1365	47mg 1,1,1-TCE
8	CGTS	2275	47mg 1,1,1-TCE
9	CGTS	3635	47mg 1,1,1-TCE
10	CGTS	1365	44mg 1,1,1-TCE in 5% mineral oil
11	CGTS	2275	44mg 1,1,1-TCE in 5% mineral oil
10	CGTS	3536	44mg 1,1,1-TCE in 5% mineral oil

n = number of replicates.

3.7. Experimental Design of In Vitro Experiment.

The in vitro release of TCE was studied in several ceramic, ceramic-glass tube, and ceramic-reservoir systems. The factors affecting the TCE release included the density of the ceramic, the storage of the solvent in a glass tube or reservoir, blending the solvent in mineral oil, the opening of the glass tube, and the volume of solvent placed into the system. Based on these investigations, one ceramic system was chosen for conducting experiments using animals.

3.7.1. Effect of Density and Mineral Oil on TCE Release from Ceramic Cylinders and Ceramic-Glass Tube Systems.

In order to study the effect of density on the in vitro release of TCE, 36 ceramic cylinders were compressed at 3 different loads (1365Kg, 2275Kg, and 3635Kg). Eighteen ceramics were converted to ceramic-glass tube systems (CGTS) and the remaining 18 ceramics were used as ceramic systems. Each of the ceramics or CGTS devices was loaded with either 47mg TCE or 44mg TCE in 5% mineral oil (Table 1). TCE release from each ceramic was individually analyzed by gas chromatography. The amount (concentration) of TCE released by individual ceramics was calculated by the Hewlett Packard integrator. The concentration of TCE released from the ceramics was converted to a rate of release using the following formula:

Conc. (ug/ml)	*	120 ml/min	=	ug/min
ug/min	*	1000 ug/mg	=	mg/min
mg/min	*	60 min/h	=	mg/h

3.7.2. Effect of Different Sized Reservoir Openings on TCE Release from CGRS.

Ceramics used in this study were compressed at a load of 3635Kg. The glass reservoirs had a capacity of 2ml and the 1cm x 4mm glass tube had openings of 0.5mm, 1.0mm, 1.5mm, and 2.0mm, respectively. Five hundred microliters of TCE (666mg) were injected into each ceramic-glass reservoir system (CGRS) through the rubber/teflon septa and the concentration of TCE released was analyzed for each system by gas chromatography. A rate of release was determined for each CGRS device. A total of 12 CGRS devices was distributed equally into 4 groups.

3.7.3. Effect of Different Volumes on the Release of TCE From CGRS.

The ceramics used in this study were compressed at a load of 3635kg. The glass reservoirs had a capacity of 1ml and the 1cm x 4mm glass tube had an opening of 0.5mm. Nine CGRS devices were distributed into 3 groups. Reservoirs in each group contained 0.25ml, 0.50ml, and 1.0ml TCE (333mg,666mg,and 1333mg), respectively. The concentration of TCE released from each CGRS device was analyzed by gas chromatography. The rate of release was calculated mathematically from the recorded concentrations.

3.8. Experimental Design of *in vivo* Experiments.

3.8.1. Animal Housing and Care.

Fisher 344 male albino rats, obtained from (Charles River, Wilmington, MA), were acclimatized for 2 weeks prior to surgery. Two rats from each shipment were assessed for standard quality control measures employed by the U.S. Air Force. Animals were housed 3 to a cage in facilities at the Veterinary Services Building, Wright Patterson Air Force Base, Dayton OH. Throughout this investigation, the rats were maintained in an environment of 68-70^{0}F and 40% humidity. A 12 hour light:dark cycle was maintained at all times. The rats were fed Rodent Lab Chow #5001 (Ralston Purina, St. Louis, MO) and given water ad libidum throughout this investigation.

3.8.2. Determination of the Concentration of TCE in the Blood.

Nine Fisher 344 male rats weighing 250-300g were anesthetized and each was implanted subcutaneously (S/C) behind the neck with a CGRS device. These rats were distributed equally into 3 groups and were given 120 hours to recover from surgery before 0.25ml, 0.50ml and 1.0ml (333mg, 666mg, and 1333mg) of TCE, respectively was injected into each device. At 24 hour intervals blood samples were taken from the tail vein following the procedures of Bober (8). A 100ul blood sample was placed into an auto sampler vial containing 1ml of hexane and allowed to equilibrate for 1 hour. One microliter of hexane layer was analyzed by gas chromatography to determine the concentration of TCE in the blood.

3.8.3. Behavior Experiment (Pilot Studies).

In the first pilot study, 8 Fisher 344 rats (Harlan-Sprague Dawley, Indianapolis, IN), weighing 500-525g were implanted with 2ml CGRS devices. Each rat was implanted subcutaneously behind the neck with one CGRS device which had been sterilized with ethylene oxide for four hours. The animals were allowed one week to recover from surgery and then were run through a series of behavioral experiments (described earlier) to record the values of baseline performances. Each animal was tested in the following order: Opto-Varimex, Figure Eight Maze, and Auditory Startle Tests. On the following 3 days, 0ml. 1.0ml, 1.5ml or 2.0ml of TCE was injected daily into the CGRS device. On the fifth day the animals were again run through the series of behavior experiments to record their activity level following exposure to TCE.

In the second pilot study, 8 Fisher 344 male rats (Harlan-Sprague Dawley, Indianapolis, IN) weighing 500-525g were implanted S/C behind the neck with 2ml CGRS devices. These CGRS devices had neck openings of 1.0ml, 1.5ml or 2.0ml respectively. Prior to surgery, the CGRS were sterilized with ethylene oxide for 4 hours. After allowing 4 days to recover from surgery, the baseline values were recorded for the animals during OPTO-Varimex, Figure-Eight Maze, and Auditory Startle Tests. Each animal then received a dose of 1ml/day for 3 days. Following 3 days of dosing the animals were run through the same series of behavioral experiments. One week following each of the pilot studies, the animals were sacrificed using an overdose of Halothane and examined for gross morphological changes.

3.8.4. Major Behavior Studies.

Fifty 25-300g male Fisher rats (Charles River, Wilmington, MA) were distributed equally into 5 groups (Table 2). Each sham control (empty vehicle control) animal and the animals to be treated with TCE were implanted with a 1ml CGRS device S/C behind the neck. The CGRS devices were steam autoclaved for 20 minutes prior to surgery. The animals were given 72 hours to recover from surgery and then run through a series of behavior experiments consisting of Opto-Varimex, Figure Eight Maze, and Auditory Startle to provide baseline performance data. Subsequently, the animals were randomly distributed into groups, consisting of 4 animals each, for the series of behavior experiments. On the day following the recording of baseline values, a total of 0.25ml, 0.50ml or 1.0ml TCE was injected into the CGRS device. After allowing the chemical to be released from the CGRS device for 48 hours, the animals were again run through the series of behavior experiments.

TABLE 2

Experimental Design for Determining the Concentration of TCE in the Blood of Fisher 344 Rats Exposed to Different Concentrations of TCE.

Group (n=3)	Volume of TCE
1	0.25ml
2	0.50ml
3	1.00ml

n = number of replicates.

3.9. Histological Evaluation of Tissues.

Following the behavior experiment the animals were euthanized by an overdose of Halothane. The tissues were examined for gross morphological changes and the lungs , heart, liver, spleen, and testes were harvested and fixed in formalin. The tissues were then processed for histological evaluation by the Toxicology Branch at the Toxic Hazards Division of Harry G. Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, Dayton, OH.

3.10. Data Analysis.

The data collected from all the experiments in this investigation were downloaded into Lotus 1,2,3 (Lotus Development Corp, Cambridge, MA) spreadsheets and then assessed by One Way Analysis of Variance (ANOVA) using Stagraphics (Statistical Graphics Inc., Rockville, MD) software on a Zenith 248 personal computer (Zenith Data Systems Corp., St. Joseph's MI).

4.0. RESULTS AND DISCUSSIONS:

4.1. *In Vitro* Studies.

4.1.1. Density of Ceramic Capsules.

The ceramic capsules used in the ceramic alone and CGTS devices had the following dimensions: inner diameter of 0.42cm, outer diameter of 0.80cm, height of 1.30±0.15cm, and a weight of 0.90±0.04g. Densities of the ceramics pressed at compression loads of 1365Kg, 2275Kg, and 3635Kg were 1.74±0.06g/cm^3, 2.32±0.02g/cm^3, and 2.36±0.04g/cm^3, respectively. The ceramics used in the CGRS devices had the following dimensions: inner diameter of 0.42cm, outer diameter of 0.80cm, height of 1.37±0.05cm, and a weight of 0.90±0.04g. Densities of these ceramics were 1.96±0.03g/cm^3. The densities of the ceramics compressed at a load of 1365Kg were significantly ($p<0.05$) lower than those compressed at loads of 2275Kg and 3635Kg. Differences between the densities of the ceramics compressed at loads of 2275Kg and 3635Kg were not significant.

4.1.2. Release of TCE from Ceramic Capsules.

The data obtained on the delivery of TCE and TCE in 5% mineral oil from ceramic capsules of three different densities is shown in Figure 2. Differences in the average release rates of TCE from ceramics of 1.74±0.06g/cm^3, 2.32±0.02g/cm^3, 2.36±0.04g/cm^3, respectively were not significant. However, the amount of TCE in 5% mineral oil delivered from ceramics of $2.36 \pm 0.04 g/cm^3$ densities was significantly lower than the amount of TCE delivered from ceramics of all 3 different densities. The amount of TCE in 5% mineral delivered from ceramics of $2.36 \pm 0.04 g/cm^3$ densities was also significantly lower than TCE in 5% mineral oil delivered from ceramics of $1.74 \pm 0.02 g/cm^3$ and $2.32 \pm 0.02 g/cm^3$ densities. Major differences in the release rate and amounts of TCE released from the ceramics occurred during the first hour of delivery. Differences in the release rates of TCE or TCE in 5% mineral oil from the ceramics of three different densities for the remaining two hours of this study were not significant.

4.1.3. Release of TCE from CGTS Devices.

The data obtained on the delivery of TCE and TCE in 5% mineral oil from the ceramics of three different densities in the CGTS devices is shown in Figure 3. Differences between the release rates of TCE from CGTS, consisting of ceramics of three different densities, were not significant. However, all the CGTS devices released significantly lower amounts of TCE when the solvent was incorporated in 5% mineral oil (Figure 3B).

Variation in the density of the ceramic had no effect on the rate of release of TCE (Figures 2 and 3). It is apparent that the volatile nature of TCE is responsible for the lack of differences in the release rates of the hydrocarbon from ceramics of 3 different densities. However, it should be noted that when the ceramic density was increased to $2.36 \pm 0.04 g/cm^3$ and TCE was incorporated in 5% mineral oil, the rate of release of TCE decreased significantly. Thus, the density and pore sizes of the ceramic cylinders may not have varied enough to affect the diffusion of TCE alone from the ceramics. Previous investigations conducted on release of dyes, steroids, and polypeptides have shown that these substances are released at a lower rate from higher density ceramics (9). In contrast to ceramics, porous polymer-glass reservoir systems released 0.25ml to 2.0 ml TCE was released within 30 seconds.

4.1.4. Effect of Different Sized CGRS Device Tube Openings on TCE Release.

Figure 4 shows the effect that the size of tube openings had on the rate of TCE release from the CGRS devices. Tube openings of 0.5mm, 1.0mm, and 1.5mm in diameter produced sustained release rates of 33.49$\pm$2.12mg/h, 30.20$\pm$0.86mg/h, and 32.45$\pm$0.51mg/h, of TCE respectively. Differences in the release rates of TCE from CGRS devices having the 3 different tube openings were not significant. At the end of 7.5 hours, the amount of TCE released from CGRS devices with tube openings of 0.5mm, 1.0mm and 1.5mm, was 308mg, 290mg, and 280mg respectively. Thus, only half the amount of TCE (666 mg) initially stored in the CGRS devices was delivered by 7.5 hours. However, a significantly higher rate of TCE was observed when the tube opening was increased to 2.0mm in diameter. The rate of release was more than twice (69.47$\pm$0.70mg/h) that of the CGRS devices having tube openings of 0.5mm to 1.5mm in diameter. Higher release rate of TCE delivered from TCE devices having tube openings of 2.0mm in diameter resulted in a total release of 496mg TCE by the end of 6 hours.

It is apparent that the variations in diameter (0.5mm and 1.5mm) of reservoir tube openings did not cause significant differences in the rate of delivery of TCE. However, increasing the tube opening diameter to 2.0mm increased (as expected) the rate of TCE delivery from the CGRS device and decreased the total time of delivery to 8 hours (calculated). Since the desired time of delivery in animals was more than 8 hours, this system was not selected for animal experiments. Lack of significant differences in TCE delivery between the CGRS device tube openings of 0.5 to 1.5mm in diameter suggested that any one of these systems would suffice for animal experiments. Thus, a CGRS device having a tube opening of 0.5mm in diameter was used for the remaining experiments.

4.1.5. Effect of Initial Volume of TCE in CGRS on Release of TCE.

Figure 5 shows the effect which the initial volume of solvent in the CGRS reservoir had on the release rate of TCE from the CGRS device. Sustained release rates of 13.74 ± 0.36 mg/h, 33.8 ± 2.28mg/h and 15.39 ± 0.43mg/h resulted from the storage of initial volumes of 0.25ml,0.50ml, and 1.0ml of TCE in the reservoirs of CGRS devices respectively. Thus, at the end of 7 hours, 103mg, 308mg and 180mg were released from the CGRS devices containing 333mg, 666mg and 1333mg of TCE, respectively. Differences in the release rates and total amounts of TCE released from CGRS devices containing 0.25ml and 1.0ml TCE were not significant, but the release rate from the CGRS devices containing 0.50ml TCE was significantly ($p<.05$) higher than that from the CGRS devices containing 0.25ml and 1.0ml TCE.

Large variations and lack of sustained release of TCE from CGRS devices containing 0.5ml TCE throughout the study suggested that either one of the ceramics was defective or that the caps of one of the reservoirs was not tightly sealed. Although the data obtained suggested that this study should be repeated, unavailability of the apparatus and gas chromatograph at WPAFB made this impossible. Yet the data obtained from CGRS devices containing 0.25ml or 1.0ml TCE suggest that the system is viable and can be used to deliver TCE in a sustained manner for long durations of time. We intend to repeat this study when a division of the U.S. Navy stationed at WPAFB can make the apparatus available to us.

4.2. Animal Studies.

4.2.1. Pilot Studies.

Two pilot studies were carried out to determine if TCE could be delivered in an animal model and if performance changes could be determined. Since this was a pilot study, the data obtained was not analyzed statistically. However, these studies provided the training needed for conducting various experiments. these studies also confirmed that the behavior of rats can be influenced by the TCE delivered from implanted CGRS devices.

4.2.1.1. Effects Due to Sterilization by Ethylene Oxide.

The CGRS devices in the pilot studies were sterilized for four hours by ethylene oxide and implanted intraperitoneally. Gas sterilization was used in this study since ceramic capsules used for delivering steroids and gossypol acetic acid were sterilized by gas sterilization (10, 11). Gross examination of CGRS devices revealed an unexpected soft tissue encapsulation of the device. It also revealed an acute inflammatory reaction at the implantation site characterized by yellow pus and fibrin strands parallel to the glass reservoir. Thus, in subsequent studies, CGRS devices were sterilized by steam autoclaving for 20 minutes.

4.2.1.2. Effect of Intraperitoneal Implantation of CGRS Devices.

Necropsy showed numerous gross morphologic changes. Adhesions and white sub-capsular foci on the liver, intraperitoneal fusions, as well as testicular hemorrhages and disruptions of the the testicular foot pad were present and attributed to the rapid delivery of concentrated solvent directly into the abdominal cavity. Thus, all subsequent studies were conducted by implanting the CGRS devices subcutaneously.

4.2.1.3. Pathological Examination.

Gross morphological examination of the animals in the pilot studies revealed slight testicular atrophy in the TCE dosed animals. Thus the total amount of chemical received by the animals were reduced by two thirds in the final animal experiments.

4.2.2. Major Animal Studies.

4.2.2.1. Gross Morphology and Histopathology.

An expected soft tissue encapsulation of approximately 2mm was observed around each subcutaneous CGRS implant. Histological examination of the heart, liver, lungs, kidney, spleen, and testes revealed no abnormalities or changes. Thus, short term exposure to 0.25ml, 0.50ml, and 1.0ml of TCE does not induce organic or histopathologic changes in Fisher 344 rats.

4.2.2.2. Concentration of TCE in Rat Blood.

Initially 9 animals were implanted with CGRS devices and 0.25ml, 0.50ml, or 1.0ml TCE was injected into each CGRS. Blood concentrations of TCE were monitored over 72 hours. In order to verify that the variability within the groups was a true phenomenon, a second group of 9 animals were implanted with CGRS devices. Five hundred microliters of TCE was injected into each CGRS device (second group) and the solvent concentration in blood was again monitored for 72 hours. Figure 6 shows the data collected from all 18 animals (3 with 0.25ml, 12 with 0.50ml, and 3 with 1.0ml of TCE in the reservoir of each implanted CGRS device).

The three animals implanted with CGRS devices containing 0.25ml TCE showed an average concentration of 1.65 ± 0.07ug/ml at 24 hours. The blood levels of

TCE decreased significantly to 1.23±0.09ug/ml within the next 24 hours (at 48 hours) and were barely detectable at 72 hours. The 12 animals implanted with CGRS devices containing 0.50ml TCE showed an average concentration of 1.89±0.42ug/ml, 2.04±0.52ug/ml, and 1.35±0.22ug/ml at 24, 48, and 72 hours, respectively. There were no significant differences between the blood solvent concentrations at the three different periods of observation.

The three animals implanted with CGRS devices containing 1.0ml TCE showed an average concentration of 2.20±0.43ug/ml at 24 hours. The amount of TCE increased to 4.33±1.55ug/ml at 48 hours, and then decreased to 1.12ug/ml by 72 hours in the surviving animal (two animals died during blood collection procedures at 48 hours). Schuman and Watanbe reported that a 6 hour inhalation exposure to 150ppm of TCE resulted in a rat blood solvent concentration of 2.6±0.9ug/g, which was similar to a value of 2ug/ml reported in man following inhalation exposure at 145 ppm for 4 hours (12). Rats in this study, implanted with CGRS devices containing 0.5ml TCE showed blood concentration values of 2.04±0.52ug/ml which are similar to those reported in the literature for inhalation exposure of TCE. Thus, rats implanted with CGRS devices containing 0.25ml solvent (blood concentration of 1.23±0.09ug/ml TCE) represent a low solvent exposure model, whereas rats implanted with CGRS devices containing 1.0ml TCE (blood concentration of 4.33±1.55ug/ml TCE) represent a high solvent exposure model. Since the differences between the blood levels of TCE in rats implanted with CGRS devices containing 0.25ml, 0.50ml and 1.0ml TCE were not significant at 24 and 48 hours at P<.05 level but significantly different at P<.06 at 48 hours, this experiment, on repetition with larger number of animals, may confirm the validity of the proposed low and high TCE exposure models.

4.2.2.3 Behavior Experiments.

The data reported on the effects of acute and chronic exposure to TCE on behavioral performance varies considerably (6, 13). It seems that these variations are due to the sensitivity of the techniques used and the type of activity chosen to test for behavioral performances. In a recent review published by MaCkay and coworkers it was noted that, in the studies reviewed, none of the behavioral changes correlated with blood or breath levels of solvents (14). In fact, in most of these studies, performance testing was monitored immediately after exposing the animals or humans to the solvent. Hence, in this investigation, attempts were made to correlate the blood solvent concentration levels with behavioral performance by closely coordinating the studies on blood solvent concentration and behavioral activity. Animals used for determining concentrations of solvent in the blood were implanted and injected with TCE on the same time schedule used for the behavioral study. Blood samples were taken at 9:30 A.M. because the behavioral studies were begun at 9:30 A.M. on day 2 (48 hours) of exposure. Therefore, the 48 hour blood solvent concentration values should reflect the blood solvent concentration responsible for changes in behavioral performance observed in this investigation.

4.2.2.3.1. Figure Eight Maze Study.

The number of movements of animal in different groups during the first session were compared to determine the effect of TCE on exploratory behavior. The total number of movements during the entire sixty minutes were also summed for total activity. Figure 7A shows the number of movements recorded for the 50 animals used in this investigation during the first 20 minutes. There were no significant differences between the number of movements made by animals designated as controls, shams, and experimentals implanted with CGRS devices containing 0.25ml, and 0.50ml TCE, respectively. This suggests that blood concentration of 1.23±0.09ug/ml and 2.04±0.52ug/ml achieved by 0.25ml and 0.50 ml TCE containing CGRS devices are not sufficient to influence the initial exploratory locomotor activity. The number of movements made by animals implanted with CGRS devices containing 1.0ml of TCE was significantly (P<.05) lower than those of all the other animals used in this study. Since these animals had a blood solvent concentration of 4.33±1.55ug/ml, the data suggests that higher doses of TCE are effective in reducing the initial number of movements made by rats implanted with CGRS devices containing 1.0ml TCE.

Figure 7B shows averages of total number of movements recorded for animals assigned to the different groups during the entire sixty minutes. Significant differences (p<.05) in movements over a total of 60 minutes occurred only between the animals implanted with CGRS devices containing 0.25ml and 1.0ml TCE respectively. The data obtained indicates that the number of movements made by the control, sham-operated rats and animals implanted with CGRS devices containing 0.25ml TCE did not vary significantly. The number of movements made by the animals implanted with CGRS devices containing 0.50ml TCE decreased somewhat, while the number of movements made by the animals implanted with CGRS devices containing 0.25ml TCE and controls did not differ appreciably. However, the differences between the number of movements made by the animals implanted with CGRS devices containing either 0.50ml or 1.0ml TCE were not significant (P<0.0582). Future studies with larger numbers of animals may show that increases in initial starting doses of TCE in the CGRS devices may have an inverse effect on the number of movements made by animals. Since other investigators have not used a Figure Eight Maze to determine the effects of chlorinated hydrocarbons or solvents, comparisons with other studies cannot be made. However, based on the data obtained in this investigation it appears that high doses of 1.0ml TCE in CGRS devices decreased both the the initial exploratory motor activity and the overall activity of the rat. A medium dose of 0.5ml TCE tended to decrease the overall locomotor activity of the rat, but not the initial exploratory locomotor activity.

4.2.2.3.2. Opto-Varimex Study.

The Opto-Varimex is an open field cage that is divided into 16 small squares by infrared beams. Opto-Varimex can measure a number of different activities and actions during three 5 minute sessions. The data obtained from the first 2 sessions were averaged together and analyzed.

The average distance traveled by rats in centimeters (cm) in each group is shown in Figure 8A. Differences between the distances traveled by the rats in different groups were not significant. Time resting (in seconds) should be an inverse of the distance traveled. Figure 8B shows that this is true. Significant differences in time resting occurred only between the animals implanted with CGRS devices containing 0.50ml and 1.0ml TCE, respectively. The animals implanted with CGRS devices containing 1.0ml TCE spent significantly more time resting ($p<0.05$) than the animals in other groups.

Figure 9A shows the stereotypic (in a fixed position) time spent by animals in different groups. Differences in the stereotypic time spent by animals in different groups were not significant. Figure 8B shows that the differences in time spent (ambulatory) by animals in different groups were not significant. Figure 9 shows the number of small movements made by animals distributed in the different groups. The differences in the number of small movements made by these animals were not significant. However, the overall data obtained on the 3 parameters (distance traveled,time at rest and time stereotypic) suggest (as expected) that TCE acts as a CNS depressant.

An expected observation was the increase in time spent ambulatory by animals implanted with CGRS devices containing 1.0ml TCE. It is not correlated with an increase in the distance traveled, since the distance traveled is least for the rats in this group. However, it does correspond to an increase in the number of small movements (Figure 10). The solvent, being lipophilic, probably acts on the lipids of the nerve cell membranes and increases their activity to a small degree. This would account for increases in small movements and for no increase in large movements (15). In general, TCE appears to have an effect similar to that of alcohol and/or prolonged work or boredom (14).

4.2.2.3.2. Auditory Startle Reflex.

Figure 11 shows the data recorded after giving pre-pulse signals to animals in the different groups. The time it took the animals to respond to the pre-pulse stimulus varied significantly ($p<0.05$) between the animals receiving 0.50ml and 1.0ml TCE, but not between the animals in other groups (Figure 11A). The peak amplitude response to the pre-pulse stimulus is presented in Figure 11B. 1,1,1-Trichlorethane had no effect on the peak amplitude response to pre-pulse signals.

Response to the high intensity stimulus is presented in Figure 12. In this instance, the time (sec) to peak response seemed to vary with the different amounts of solvent stored and delivered by CGRS to the animal (Figure 12A). However, the differences were not significant. Figure 12B shows the amplitude of the response (in grams of force) to the intensity stimulus. The animals receiving 1.0ml TCE showed a significantly ($p<0.05$) higher amplitude of response than animals in all other groups.

In this investigation, except for the difference in time to respond between the animals implanted with CGRS devices containing 0.50ml or 1.0ml TCE, there was essentially no change in the latency of pre-pulse signal to its response (Figure 11A). Exposing the animals to various levels of TCE had no significant effect on the amplitude of the startle response (Figure 11B). Upon the analysis of the data on the response latency to the stimulus response, an increase in the time to peak amplitude was observed, but the increase was not statistically ($p<0.074$) significant. However, the data obtained on corresponding startle response amplitude was quite variable (Figure 12B). The animals implanted with CGRS devices containing 1.0ml TCE showed a significantly ($p<0.05$) higher peak amplitude response in comparison to animals implanted with CGRS devices containing either 0.25ml or 0.50ml TCE.

An increase in the latency and a decrease in the startle response is known as pre-pulse inhibition (16). In comparison to the inhibition occurring in rats implanted with CGRS devices containing either 0.25ml, and 0.50ml TCE (blood levels of 1.23 ± 0.09ug/ml and 2.04 ± 0.53ug/ml), the inhibition was blocked in animals implanted with CGRS devices containing 1.0ml TCE (blood levels of 4.33 ± 1.55ug/ml).

It is known that the detection of a stimulus is inherently dependent on the precise temporal transmission of neural impulse in the auditory system (17).

It has been proposed that alcohol also increases the latency of the brain-stem evoked response. Therefore , a resulting decay in performance is the result of the disruption of temporal processing in the auditory system. According to Wecker and Ison, alcohol in 0.25g/Kg, 1.9g/Kg and 2.0g/Kg doses increases the latency of the response and decreases the startle amplitude in a dose-dependent fashion (17). But inhibition was reduced at the highest level of alcohol (2.0g/kg), which was thought to be due to a disruption in the temporal relationships along the auditory pathway (17). The TCE released from CGRS devices containing the highest amount of TCE (1.0ml) may be disrupting this temporal relationship and thus inducing the loss of the pre-pulse inhibitory pattern.

5.0. SUMMATION

The data obtained in this investigation indicates that CGRS devices are capable of delivering TCE in vitro as well as in vivo. The data also indicates that implantation of CGRS devices containing an initial dose of 0.50ml and 1.0ml TCE results in sustained concentrations of TCE in the blood for at least 48 hours. Mathematical modeling of both in vitro and in vivo release from the CGRS devices would be a valuable asset in designing and using these devices in the future.

The battery of behavior experiments conducted in this investigation leads to the conclusion that TCE does indeed decrease the locomotor activity in Fisher 344 rats and that high doses of solvent released from CGRS devices containing 1.0ml TCE increased the number of small movements and influenced the temporal relationship in the auditory neural circuit. The data obtained also confirmed the earlier findings that TCE is a CNS depressant and that it induces its effects in the same manner as alcohol (3, 6, 18).

6.0. CONCLUSIONS.

1. The incorporation of a non-porous container (a glass tube or reservoir) into an ALCAP ceramic resulted in a delivery device which was capable of releasing TCE in a sustained manner both in vitro and in vivo.

2. The density of the ceramic in the CGRS device had no effect on the rate of release of volatile hydrocarbons such as TCE.

3. Blending of TCE with 5% mineral oil reduces the rate of TCE release from CGTS.

4. The initial volume of solvent placed into the CGRS reservoir affects the rate of release of TCE from the CGRS device.

5. 1,1,1-Trichlorethane delivered by subcutaneous CGRS implants achieves blood solvent levels comparable to those produced by low level inhalation exposure to TCE in humans.

6. The highest dose level (1.0ml) of TCE in subcutaneous CGRS implants decreased the overall locomotor activity.

7. Both Figure Eight Maze and Opto-Varimex tests are sensitive in detecting locomotor behavior changes in laboratory rats exposed to the volatile hydrocarbon TCE.

7.0. REFERENCES:

1. De Ceaurriz, J., J.P. Desiles, P. Bonnet, B. Marignac, J. Muller and J.P. Guenier. Concentration-Dependent behavior changes in mice following short-term inhalation exposure to various solvents. Toxicol. Appl. Pharmacol. 1983. Vol. 67, pp. 383-389.

2. Moser, V C., J.A. Scimeca and R.L. Balster. Minimal tolerance to the effects if 1,1,1-Trichloroethane on fixed-ratio responding in mice. NeuroToxicology. 1985. Vol. 6 (1), pp. 35-42.

3. Moser, V.C. and R.L. Balster. Acute motor and lethal effects of inhaled toluene, 1,1,1-Trichloroethane, Halothane, and Ethanol in mice: Effects of exposure duration. Toxicol. Appl. Pharmacol. 1985. Vol 77, pp. 285-291.

4. Gargas, M.L., H.J. Clewell III, and M.E. Andersen. Metabolism of inhaled dihalomethanes in vivo: Differentiation of kinetic constants for two independent pathways. Toxicol. Appl. Pharmacol. 1986. Vol. 82, pp. 211-223.

5. Nilsson, K.B. Effects of 1,1,1-Trichloroethane on the cGMP metabolism in mouse brain. Acta. Pharmacol. &. Toxicol. 1986. Vol, 58, pp. 318-326.

6. Stewart, R.D. Methyl chloroform intoxication: Diagnosis and treatment. 1971. JAMA. Vol, 215 (11), pp. 1789-1792.

7. Bajpai, P. K. and H.A. Benghuzzi. Ceramic systems for long term delivery of chemicals and biologicals. J. Biomed. Mater. Res. 1988. Vol. 22, pp. 1245-1266.

8. Bober R. Technical Review: Drawing blood from the tail artery of a rat. Lab Animal 1988. Vol. 17, pp. 33-44.

9. Benghuzzi, H. and P. K. Bajpai. The effects of ALCAP mixed particle sizes on density and the delivery of proteins in vitro. 13th Northeast Bioengineering Conference, Philadelphia, PA.,March 12-13, 1987. IEEE., New York, NY., K. Foster (editor), 1987. Vol 2, 566-569.

10. Mcguire, M.A. and P.K. Bajpai. Effects of Long Term In vivo Release of Testosterone by ALCAP Ceramics. Proceedings of the 11th Northeast Bioengineering Conference. Worcester, Mass., March 14-15, IEEE., New York, NY., pp. 170-173, 1985.

11. Strasser, J. Jr. and P.K. Bajpai. Polymer Impregnated ALCAP Ceramics: Delivery of Gossypol Acetic Acid. Proceedings of the 11th Northeast Bioengineering Conference, Worcester, Mass., March 14-15, IEEE., New York, NY., pp. 166-169, 1985.

12. Schumann, A.M., T.R. Fox and P.G. Watanabe. Carbon-14 labeled methyl chloroform, 1,1,1-Trichloroethane pharmacokinetics in rats and mice following inhalation exposure. Toxicol Appl. Pharmacol. 1982. Vol, 62 (3), 390-401.

13. Tokelson, T.R., F. Oyen, D.D. McCollister and V.K. Rowe. Toxicity of 1,1,1iTrichlorethane as determined on laboratory animal and human subjects. Industrial Hygiene Journal. 1958. Vol, 19. pp. 353-362.

14. MacKay, C.J., L. Campbell, A.M. Samuel, K.J. Alderman, C. Idzikowski, H.K. Wilson, and D. Gompertz. Behavioral changes during exposure to 1,1,1-Trichloroethane: Time-course and relationship to blood solvent levels. American Journal of Industrial Medicine 1987. Vol, 11, pp. 223-239.

15. Kjellstrand, P., B. Holmquist, I. Jonsson, S. Romare and L. Mansson. Effects of organic solvents on motor activity in mice. Toxicology. 1985. Vol, 35 (1), pp. 35-46.

16. Use of Startle as a behavioral measure in research. Scientific and Professional Support Group, La Jolla, CA, pp. 1-10.1skl:ma5
17. Wecker, J.R. and R.1. Balster. Behavioral and lethal effects of combinations of oral ethanol and inhaled 1,1,1-Trichloroethane in mice. Toxicology and applied Pharmacology. 1984. Vol, 74, pp. 258-266.

18. NIOSH Current Intelligence Bulletin #48. U.S. Department of Health and Human Services. Organic Solvent Neurotoxicity. March 31, 1987, pp. 1-27.

Figure 1. Composite photograph of: (a) Ceramic cylinder, (b) Ceramic-Glass Tube System, and (c) Ceramic-Glass Reservoir Systems.

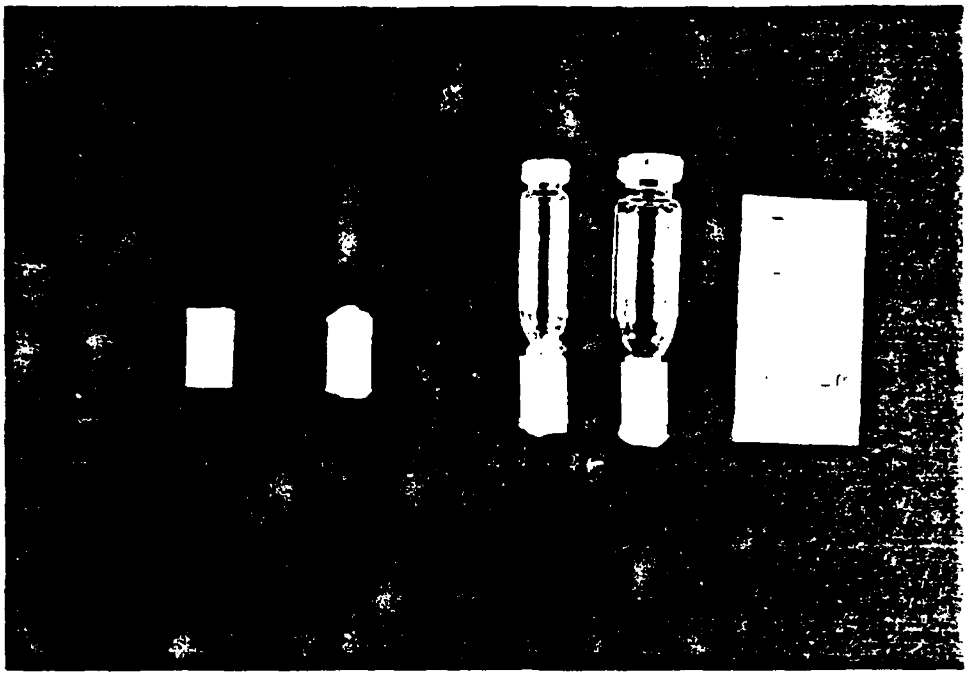
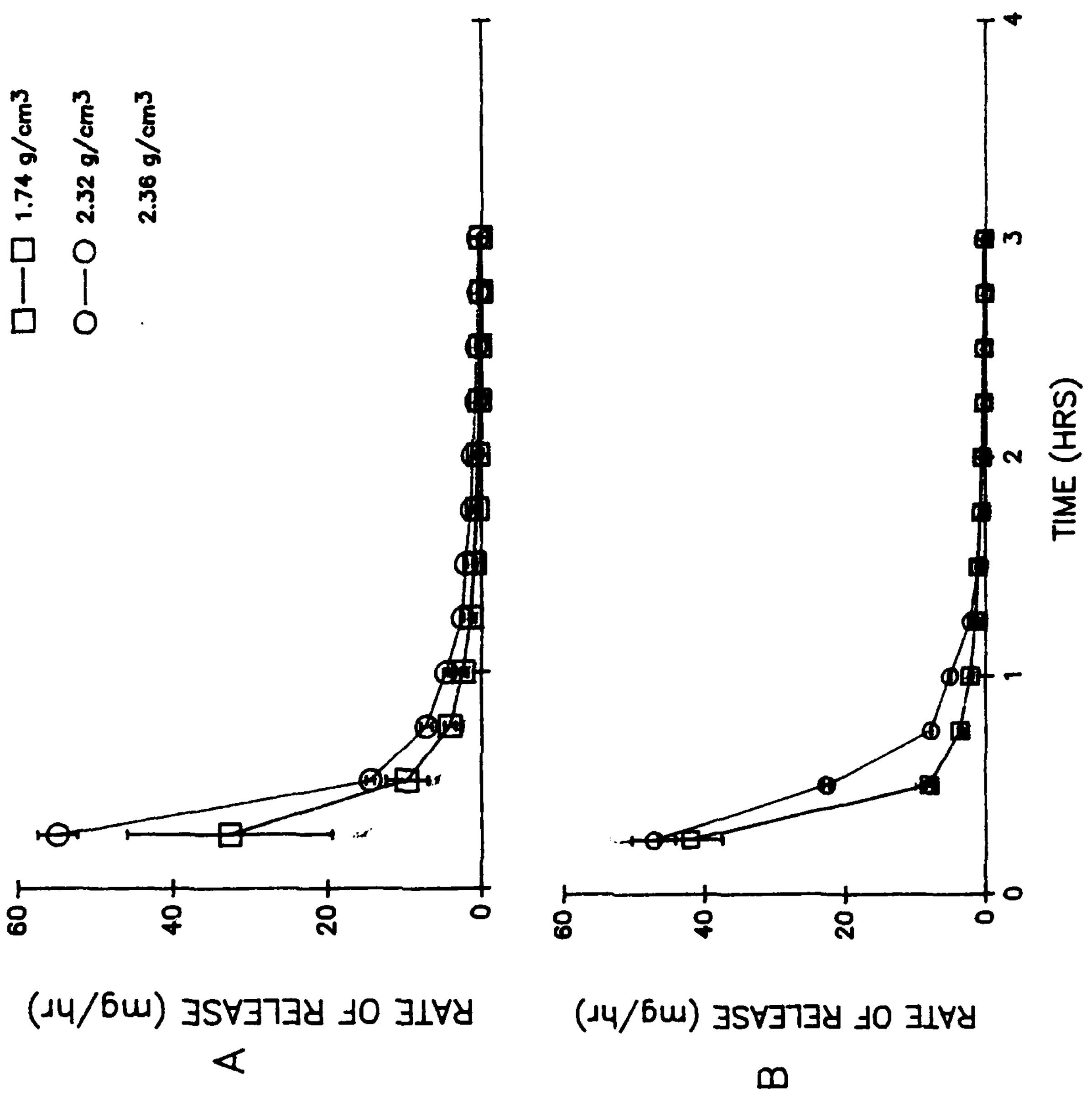


Figure 2. Release rates (mg/h ± SEM) of 1,1,1-TCE (A) and 1,1,1-TCE in 5% mineral oil (B) from ceramic capsules for a duration of 3 hours.



1.74 g/cm3
2.32 g/cm3
2.36 g/cm3
RATE OF RELEASE (mg/hr)
A
RATE OF RELEASE (mg/hr)
B
TIME (HRS)

Figure 3. Release rates (mg/h ± SEM) of 1,1,1-TCE (A) and 1,1,1-TCE in 5% mineral oil (B) from Ceramic-Glass Tube Systems (CGTS) for a duration of 5 hours.

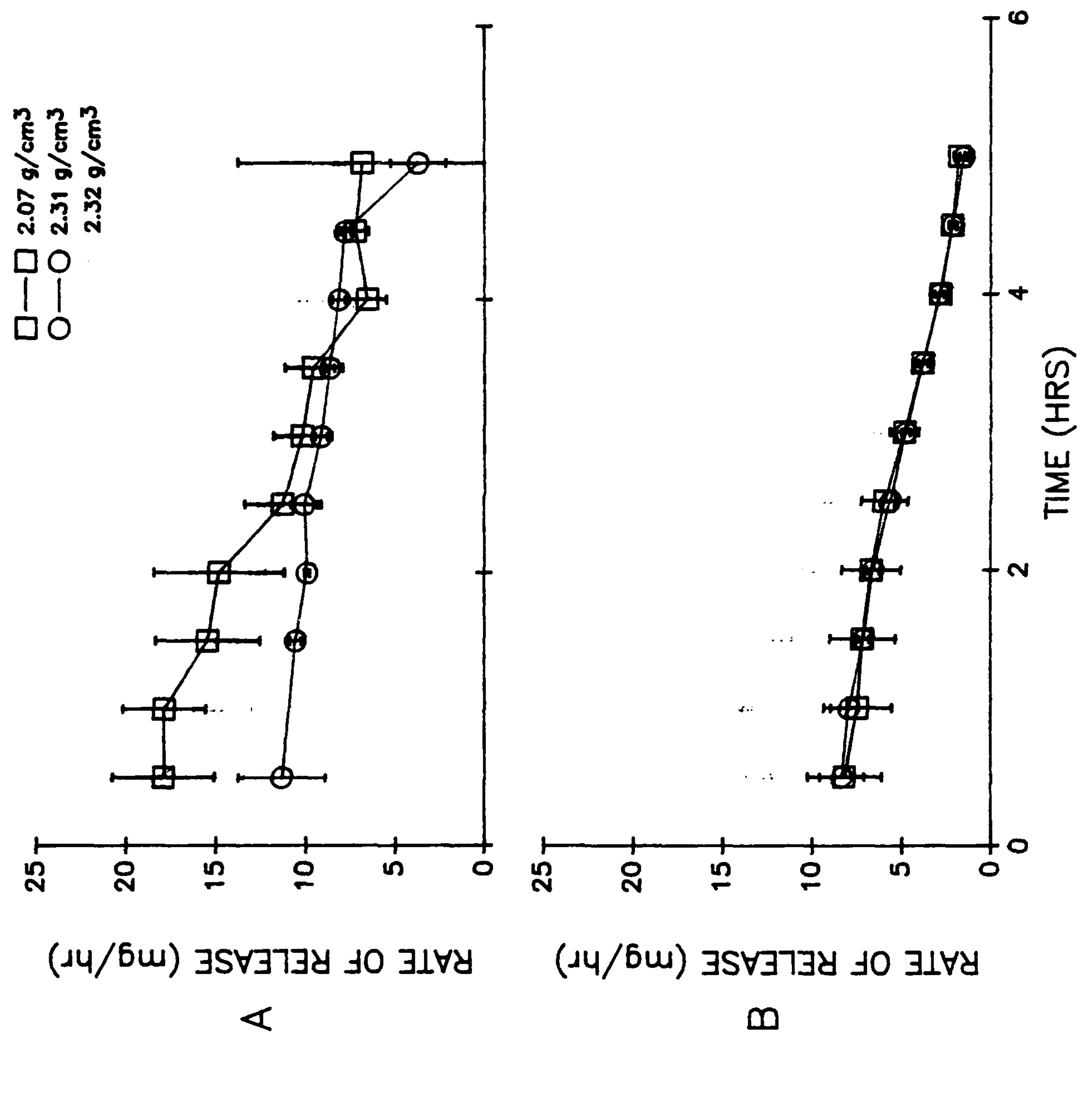
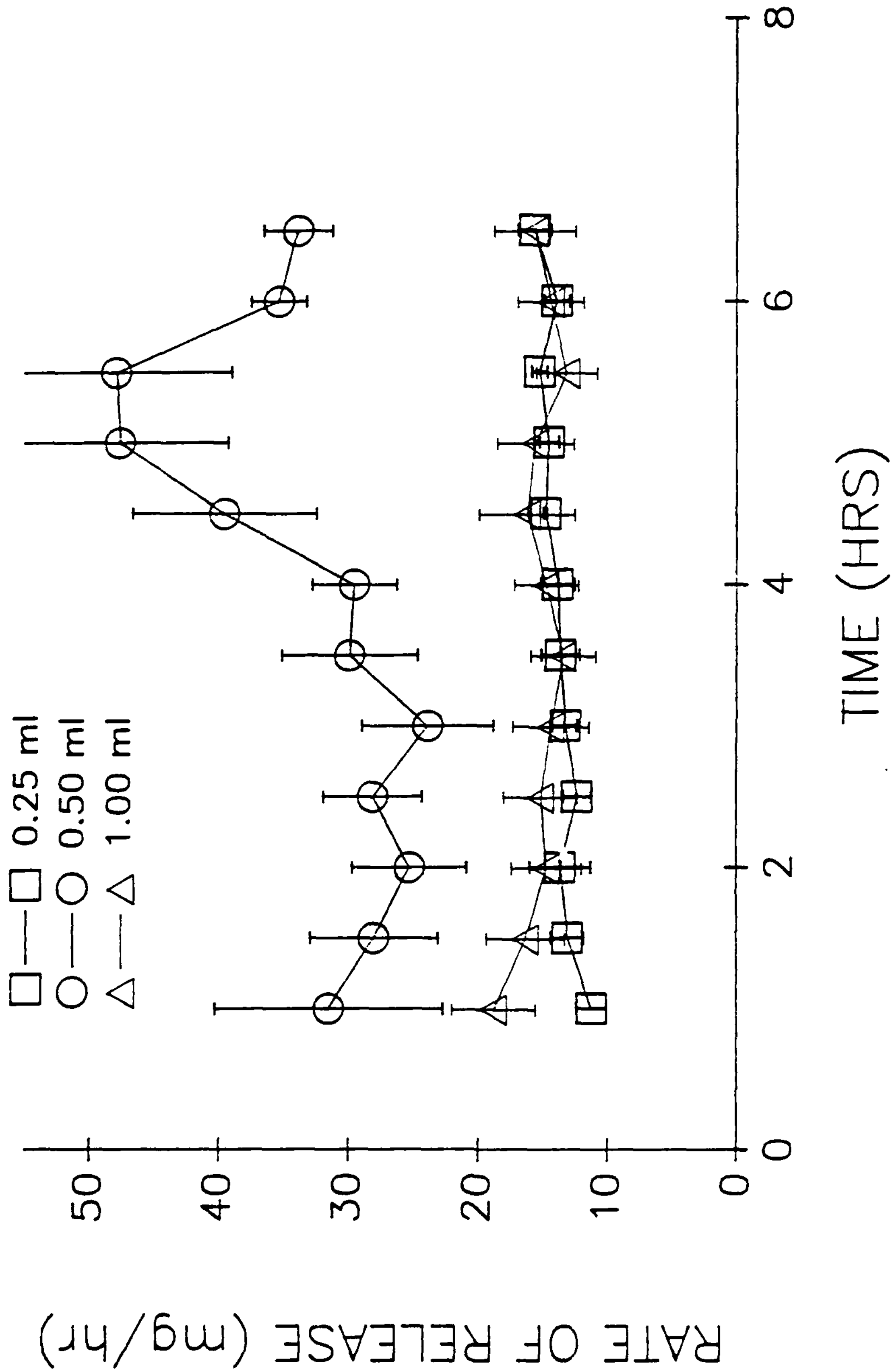
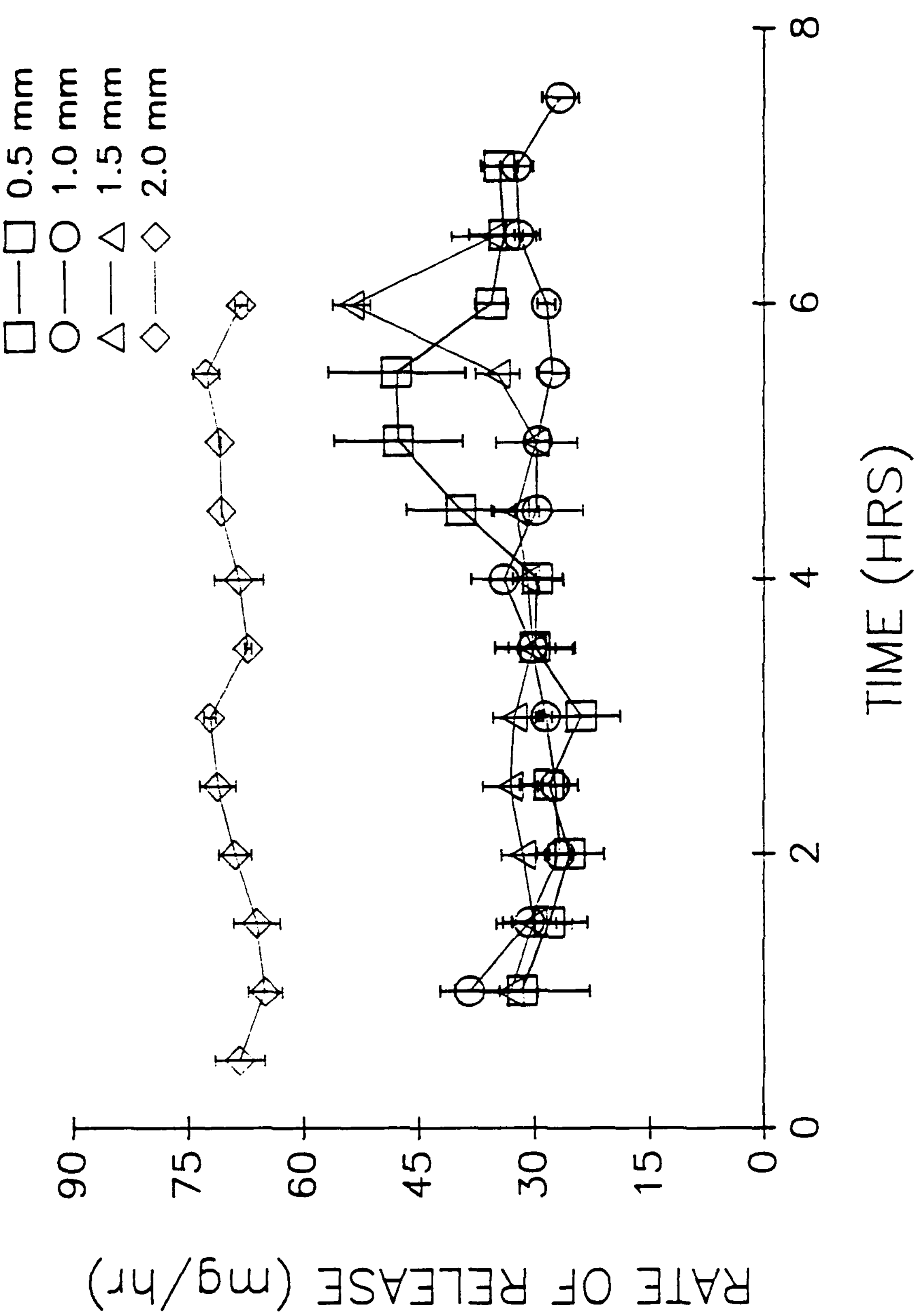

□—□ 2.07 g/cm3
○—○ 2.31 g/cm3
2.32 g/cm3
A
B
RATE OF RELEASE (mg/hr)
RATE OF RELEASE (mg/hr)
TIME (HRS)
0
5
10
15
20
25
0
2
4
6

Figure 4. Release rates (mg/h ± SEM) of 1,1,1-TCE from Ceramic-Glass Reservoir Systems (CGRS) with reservoir tube openings of 0.5mm, 1.0mm 1.5mm and 2.0mm in diameter.



□—□ 0.25 ml
○—○ 0.50 ml
△—△ 1.00 ml
RATE OF RELEASE (mg/hr)
0
10
20
30
40
50
TIME (HRS)
0
2
4
6
8

Figure 5. Release rates (mg/h ± SEM) of 1,1,1-TCE from Ceramic-Glass Reservoir Systems (CGRS) with initial volumes of 0.25ml (333mg), 0.50ml (666mg), and 1.0ml (1333mg) of the solvent for a duration of 7 hours.



□—□ 0.5 mm
○—○ 1.0 mm
△—△ 1.5 mm
◇—◇ 2.0 mm
RATE OF RELEASE (mg/hr)
TIME (HRS)
0
15
30
45
60
75
90
0
2
4
6
8

Figure 6. Blood levels (mg/h ± SEM) of 1,1,1-TCE in Fisher 344 rats implanted with CGRS devices containing 0.25ml (n=3), 0.50ml (n=12), and 1.0ml (n=3) of 1,1,1-TCE, respectively, over a duration of 72 hours.

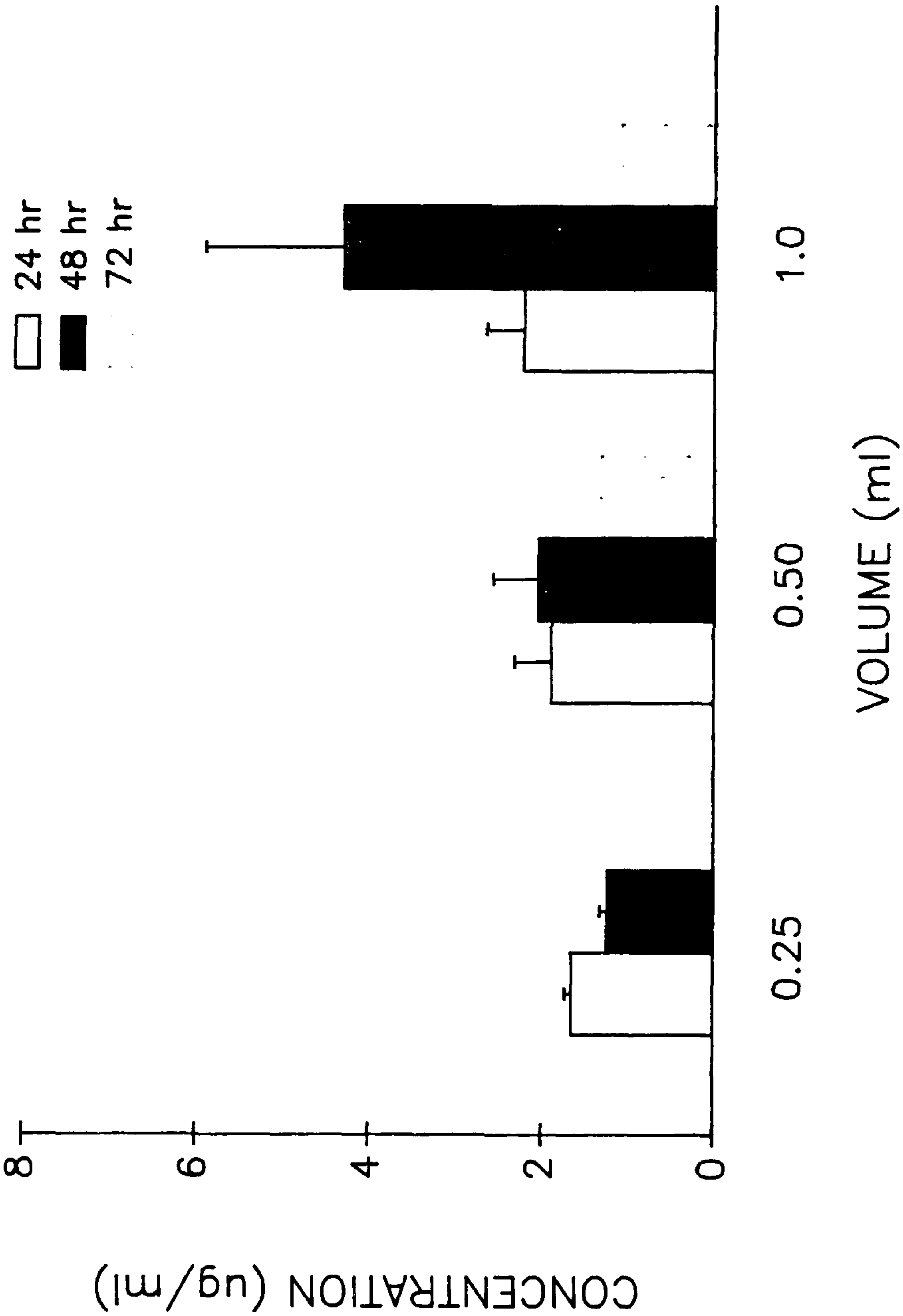
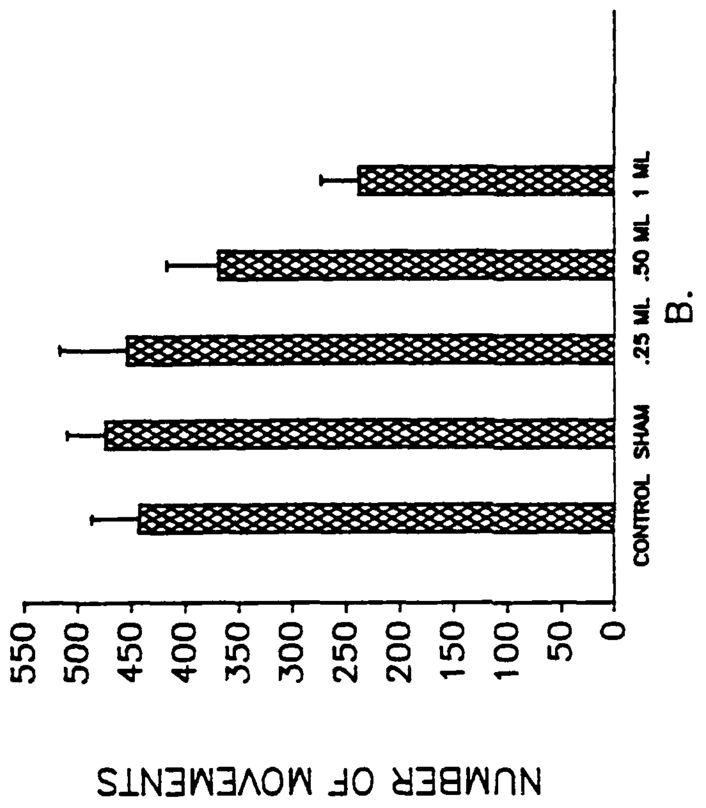

24 hr
48 hr
72 hr
CONCENTRATION (ug/ml)
0
2
4
6
8
0.25
0.50
1.0
VOLUME (ml)

Figure 7. The number of movements (number $\pm$ SEM) made by Fisher 344 rats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10). Movements detected in the first 20 minutes (a) and in 60 minutes (B) of a Figure Eight Maze behavior session. Control represents non-surgically altered animals (n=10).


NUMBER OF MOVEMENTS
550
500
450
400
350
300
250
200
150
100
50
0
CONTROL
SHAM
.25 ML
.50 ML
1 ML
B.

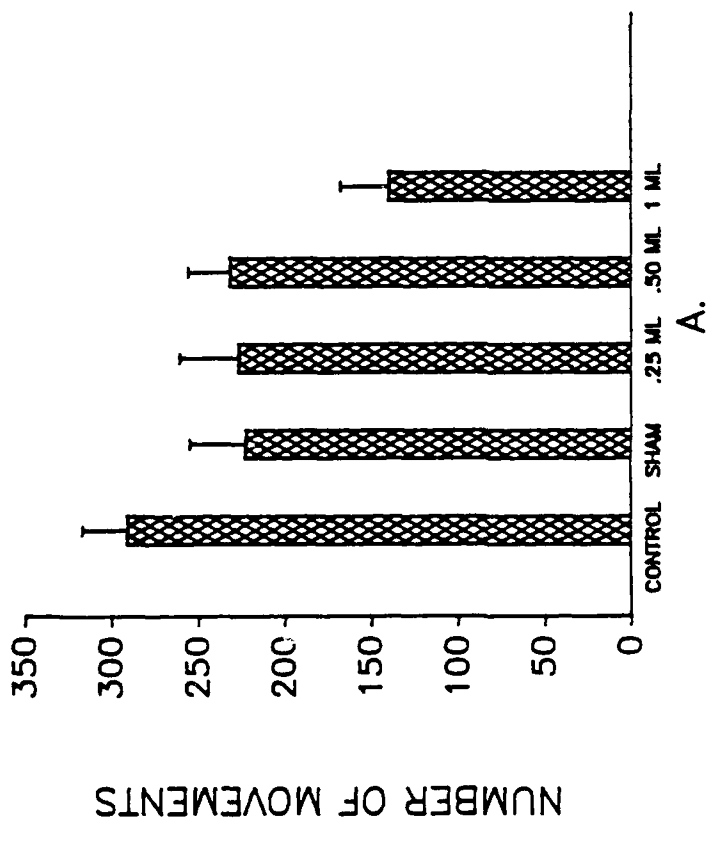

NUMBER OF MOVEMENTS
350
300
250
200
150
100
50
0
CONTROL
SHAM
.25 ML
.50 ML
1 ML
A.

Figure 8. Opto-Varimex data showing the distance (cm ± SEM) traveled (A) and the time spent (sec ± SEM) at rest (B) of Fisher 344 rats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10). Control represents non-surgically altered animals (n=10).

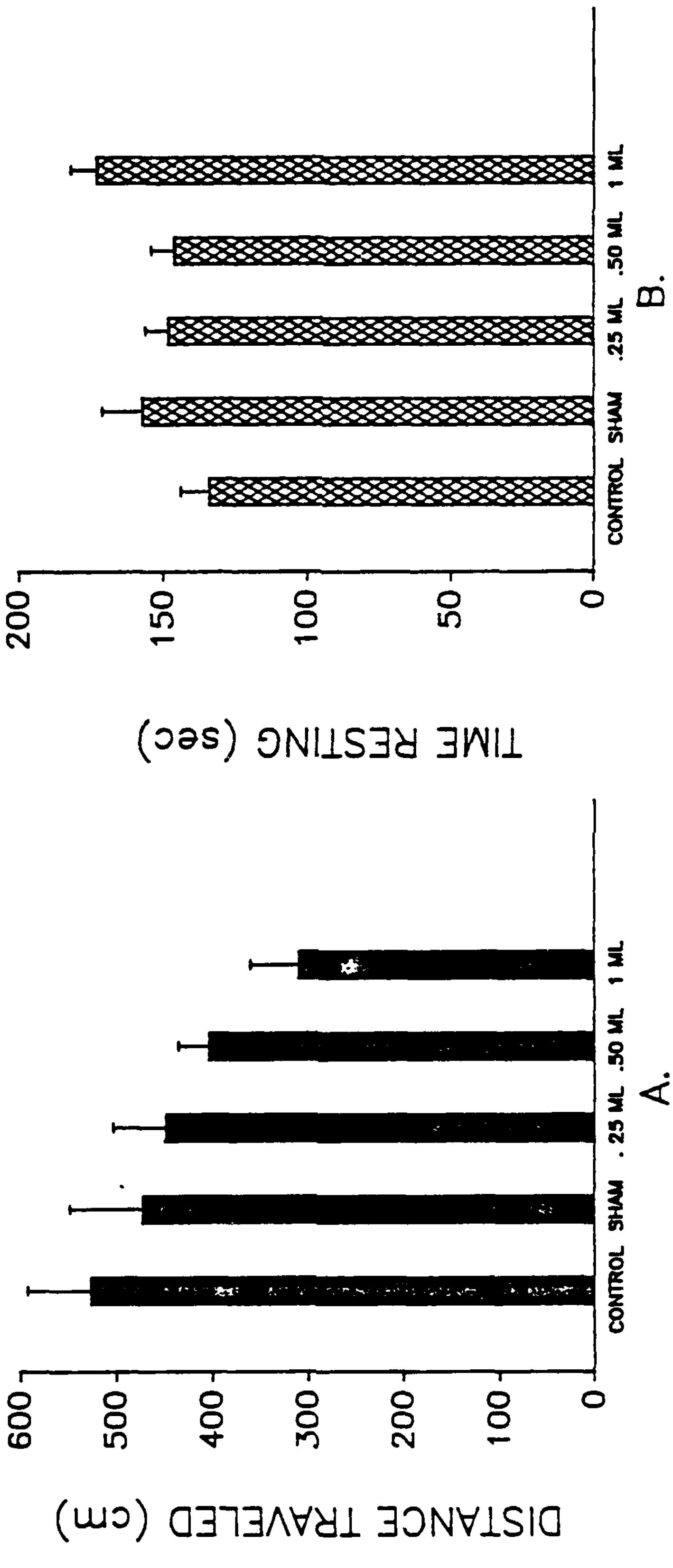

TIME RESTING (sec)
0
50
100
150
200
CONTROL
SHAM
.25 ML
.50 ML
1 ML
B.
DISTANCE TRAVELED (cm)
0
100
200
300
400
500
600
CONTROL
SHAM
.25 ML
.50 ML
1 ML
A.

Figure 9. Opto-Varimex data showing time (sec ± SEM) stereotypic (A) and the time (sec ± SEM) ambulatory (B) of Fisher 344 rats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10). Control represents non-surgically altered animals (n=10).

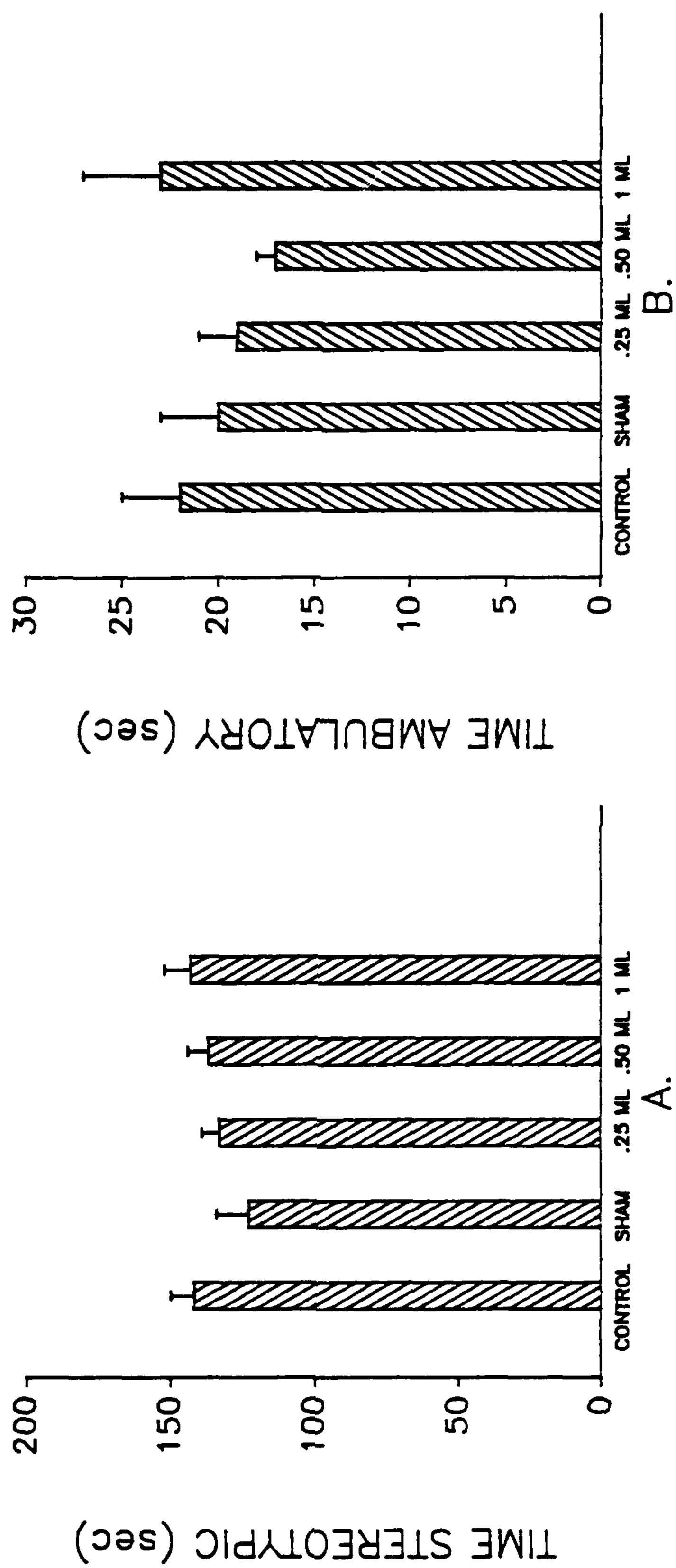

TIME AMBULATORY (sec)
0
5
10
15
20
25
30
CONTROL
SHAM
.25 ML
.50 ML
1 ML
B.
TIME STEREOTYPIC (sec)
0
50
100
150
200
CONTROL
SHAM
.25 ML
.50 ML
1 ML
A.

Figure 10. Opto-Varimex data showing the number (number $\pm$ SEM) of small movements (A) made by Fisher 344 rats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10). Control represents non-surgically altered animals (n=10).

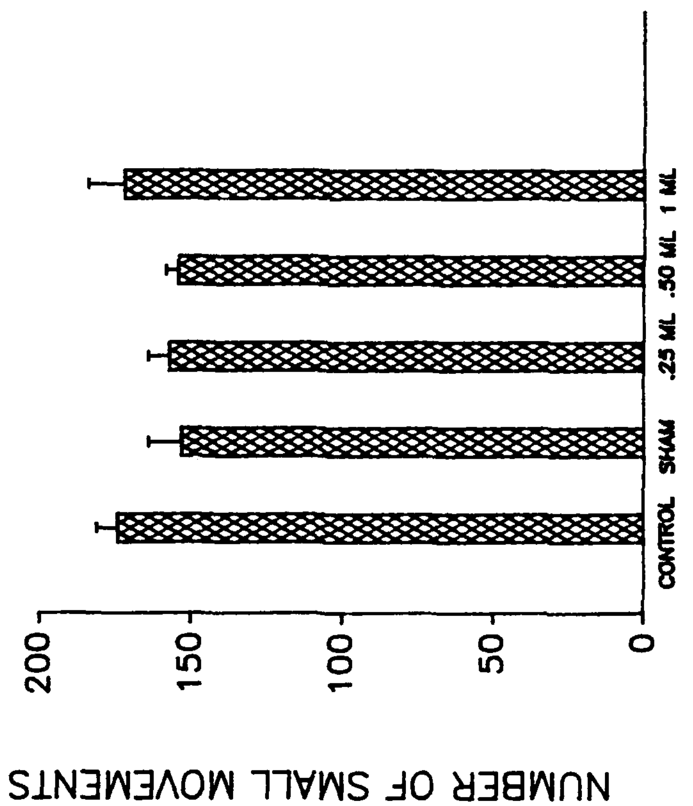
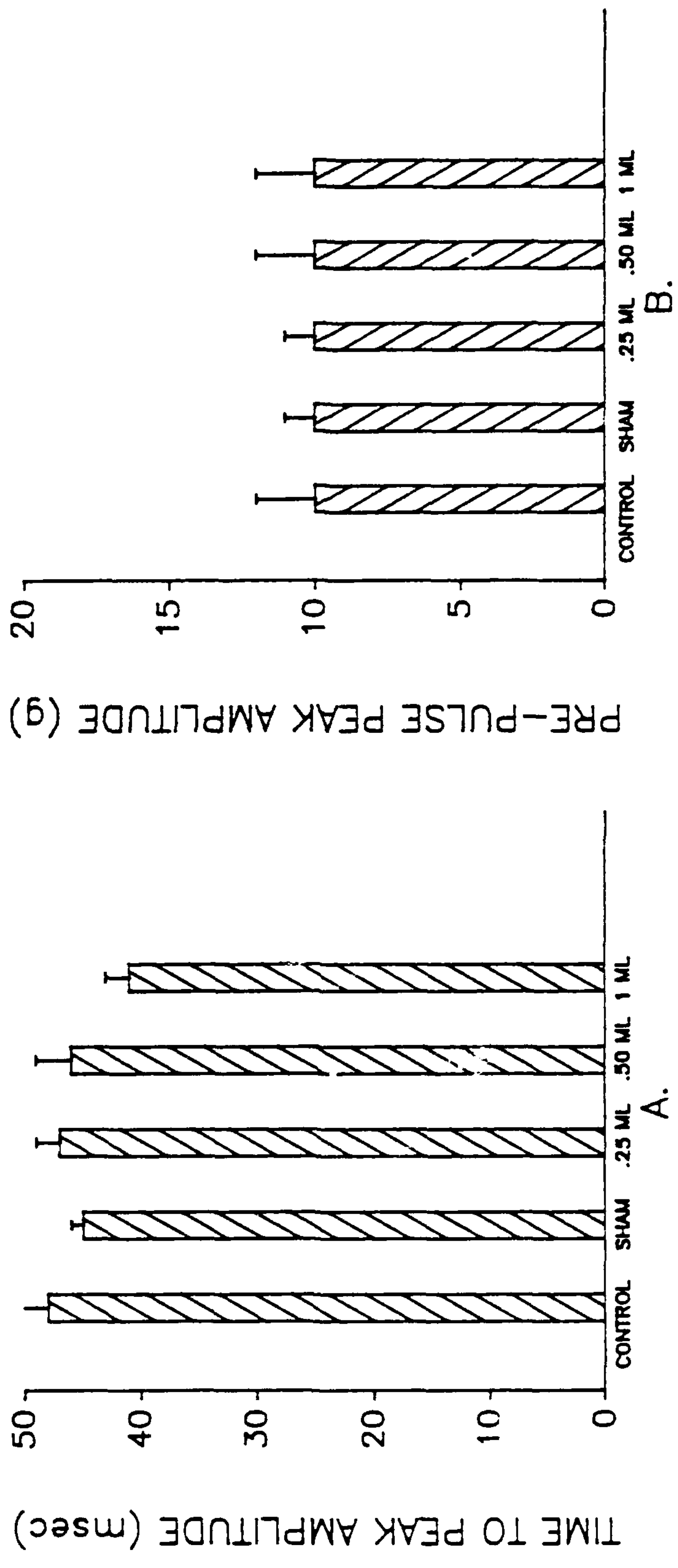
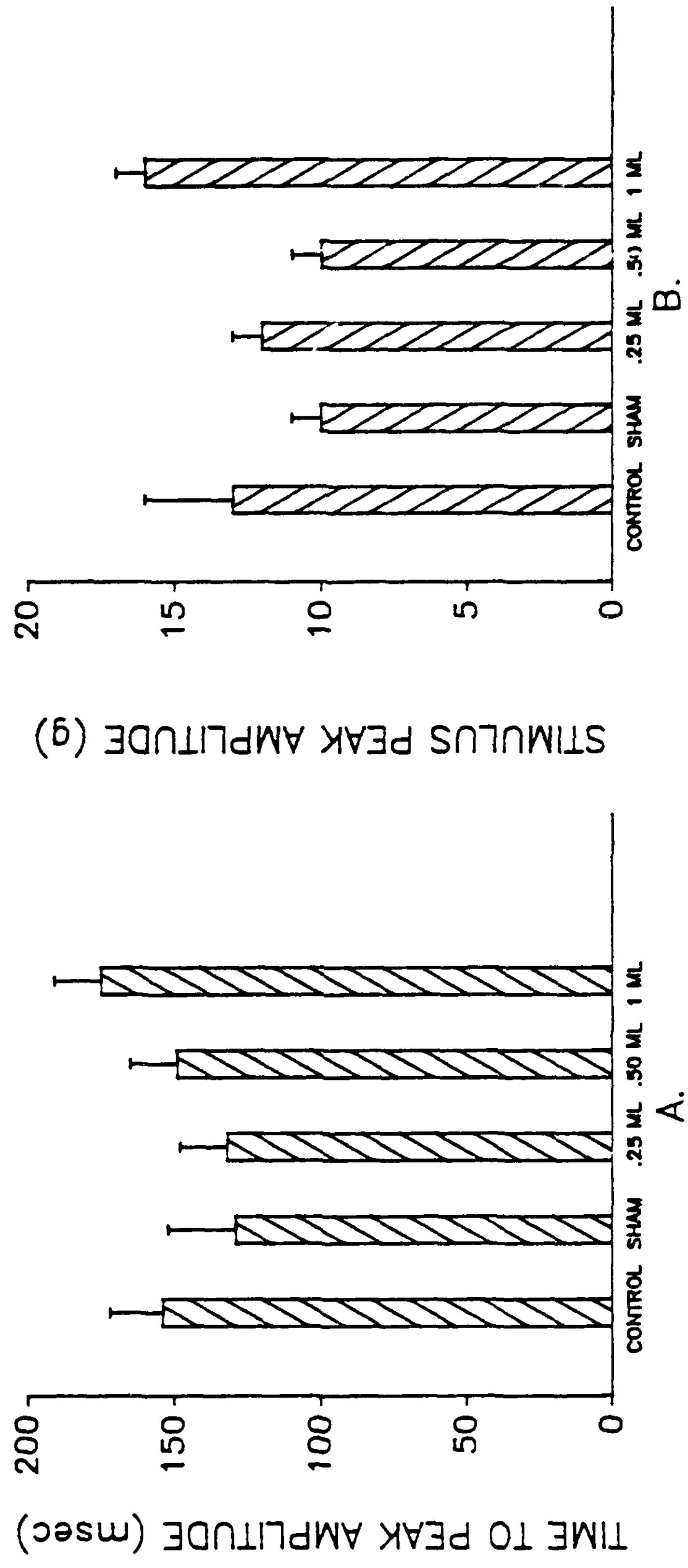

NUMBER OF SMALL MOVEMENTS
200
150
100
50
0
CONTROL
SHAM
.25 ML
.50 ML
1 ML

Figure 11. The auditory Startle data for the time (msec ± SEM) to response (A) and the amplitude (grams of force ± SEM) of response (B) of Fisher 344 rats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10) and subjected to a low intensity pre-pulse stimulus. Control represents non-surgically altered animals (n=10).



TIME TO PEAK AMPLITUDE (msec)
0
10
20
30
40
50
CONTROL
SHAM
.25 ML
.50 ML
1 ML
A.
PRE-PULSE PEAK AMPLITUDE (g)
0
5
10
15
20
CONTROL
SHAM
.25 ML
.50 ML
1 ML
B.

Figure 12. The auditory Startle data for the time (msec $\pm$ SEM) to response (A) and the amplitude (grams of force $\pm$ SEM) of response (B) of Fisher 344 r ats implanted with CGRS devices containing 0ml (shams), 0.25ml, 0.50ml, and 1.0ml of 1,1,1-TCE (n=10) and subjected to a high intensity stimulus. Control represents non-surgically altered animals (n=10).


STIMULUS PEAK AMPLITUDE (g)
0
5
10
15
20
CONTROL
SHAM
.25 ML
.50 ML
1 ML
B.
TIME TO PEAK AMPLITUDE (msec)
0
50
100
150
200
CONTROL
SHAM
.25 ML
.50 ML
1 ML
A.

1988 Research Initiation Program

Sponsored by the

Air Force Office of Scientific Research

Conducted by the

Universal Energy Systems, Inc.

REPORT - PROPOSAL NO. R-6563

Development of Implantable Devices for Sustained Delivery of Volatile Hydrocarbons in Rats

Prepared by:	P.K. Bajpai, Ph.D and D. E. Hollenbach, M.Sc.
Academic Rank:	Professor and Graduate Student
Department and	Biology Department
University:	University of Dayton
Research Location:	AAMRL/TH
	Wright Patterson AFB
	Dayton, OH 45433
USAF Researcher:	Dr. D.R. Mattie
Date:	December 30, 1988
P.O. No.:	S-760-7MG-098+

RESEARCH INITIATION PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

In situ Detection of Osteoprogenitor Cells in an Actively Growing Bone System

Prepared by:	Gwendolyn B. Howze, Ph.D.
Academic Rank:	Associate Professor
Department and University:	Department of Biology Texas Southern University
Research Location:	Texas Southern University 3201 Wheeler Ave. Houston, Texas 77004 (713) 527-7095
USAF Research	Dr. L.E. Kazarian, Wright-Patterson AFB, OHIO 45433
Date:	January 1, 1988 - December 16, 1988
Contract No.	

IN SITU DETECTION OF OSTEOPROGENITOR CELLS IN AN ACTIVELY GROWING BONE SYSTEM

by

Gwendolyn B. Howze

ABSTRACT

An attempt was made to detect osteoprogenitor cells without using radioactive tracers. The experimental approach utilized Hydroxyurea [HU], a DNA synthesis inhibitor and tetracycline type labeling. The fact that preosteoblast synthesize DNA on the pathway to becoming osteoblast, and the fact that tetracycline, Xylenol orange and calcein are deposited at the mineralization front during osteogenesis was exploited.

HU was used to inhibit "S" phase cells and to block cells at the GI/s transition. It was expected that the preosteoblast would be entrapped in an HU "net". According to this paradigm, HU, by inhibiting DNA synthesis, prevents stem cells from becoming osteoblast and thereby inhibit osteogenesis.

Hydroxyurea induced cellular changes in five bone compartments were detected; bone marrow, endosteum, periosteum, fibrocartilage, and hyaline cartilage of the epiphyseal plate. Several important expectations did not materialize: (1) a population of pyconotic cells did not appear and (2) high mitotic index was not detected. It is possible that the sampling regimen was not amicable to detecting these expected population.

The effect on bone marrow was very dramatic. The best correlation between cellular changes and osteogenesis linked labeling of bone matrix, however, was detected in periosteum. Two new large populations of cells were detected in periosteum. Multipolar cells with long processes and large structured nuclei were detected from the mid mass down to periosteal bone matrix interface. A second but less numerous population was seen only at the periosteal bone matrix interface. These cells were polygonal in contour and also contained structured nuclei. The few mitotic figures detected in periosteum were seen in the polygonal cells.

1.0 OBJECTIVES

When bone matrix is disrupted, a transient cell population called osteoblasts appears and carries out the job of restoration. The osteoblasts arise from pre-existing cells, the osteoprogenitors. The identity of the preosteoblast or osteoprogenitors is a major unsolved question. The proposed solutions are a matter of controversy since several cellular types have been suggested as being the precursor cells.

There are, therefore, two basic objectives: firstly, to identify the precursor cells which give rise to the osteoblasts which manufacture bone matrix in an active bone modeling system; secondly, to determine in which component of bone the osteoblast precursors (osteoprogenitors) reside.

This effort was begun during the 1987 USAF SUMMER Faculty Research Program (SFRP) and continued during the Mini-Grant period.

2.0 INTRODUCTION

The best known cells of bone are the osteocytes, osteoblasts and osteoclasts. The osteoblasts manufacture the bone matrix, change state and become osteocytes. The osteocytes are the mature bone cells which reside in the lacunas of the bone matrix. The osteoclasts which function in resorption are multinucleted giant cells derived from blood monocytes. Other cell types which have been described include: bone lining cells (13) and fibroblast (unpublished SEM studies).

The origins of the osteocytes and osteoclasts are agreed upon, the proposed origins of the osteoblast are, however, controversial. The proposed precursors to the osteoblasts, the so called osteoprogenitor cells include the following: undifferentiated mesenchymal cells (17), fibroblast (19), bone lining cells (13), and epithelium in certain experimental systems (7,8). If it is accepted that all of the cells in the list are indeed osteoprogenitors, it still is not clear whether they are all equally important in all systems. Nor is it clear whether there is a relationship between activating conditions and cell type.

In order to study this problem we utilized a continuously growing bone system, in which osteoblast are continuously manufacturing matrix. Presumably the osteoprogenitor population constantly generates osteoblasts according to some endogenous program. It was expected that the design would allow us to determine where in the bone the stem cells (osteoprogenitors) reside; and answer the question as to whether any or all of the cells referred to above act as stem cells in the system under study. According to this design the stem cells were both tagged and incapacitated. Thus if the incapacitated stem cells were not able to generate osteoblasts, bone matrix could not be produced, i.e., osteogenesis will be inhibited. By definition an osteoblast is a cell which lays down the osteoid which subsequently mineralizes to form bone matrix.

In many systems which are regenerated or restored from a stem cell, the stem or precursor cells follow a relatively set sequence of events after activation. One of the earliest events in the consecution is DNA synthesis, which is usually closely followed by mitosis. After their activation, at least one amplifying cell division (mitosis) is required of the precursors so as not to exhaust the stem cell line. In eukaryotic cells, mitosis is always preceded by DNA synthesis. There is evidence that osteoprogenitor cells follow the same pattern (3); for example, certain substances which induce bone formation (osteogenesis) in preosteoblast cultures also induce DNA synthesis in those cultures (4,3).

The basic assumption of this research proposal is that the following sequence of events will occur in the stem cell population: (1) activation, (2) DNA synthesis, (3) at least one amplifying cell division or mitosis, (4) the differentiation of at least one of the progeny cells into the osteoblast cell line, (5) osteogenesis by osteoblast. Events one through four are ascribed to the osteoprogenitor/stem cell line. The osteoblasts generated by those processes manufacture bone matrix, i.e., osteogenesis.

It was expected that the osteoprogenitors would be sensitive to DNA synthesis inhibitors. Additionally, it was expected that a careful choice of inhibitor would permit the *in situ* visualization of the osteoprogenitors engaged in the events which lead to the formation of osteoblasts. Hydroxyurea was used in these experiments. It is an inhibitor of DNA synthesis, which: (1) reversibly inhibits entry into the DNA synthetic period of the cell cycle, and (2) kills cells engaged in DNA synthesis of at the time the drug arrives at the site of action. These two features of hydroxyurea are expected to: (1) tag the activated osteoprogenitors by killing cells which are in the process of DNA synthesis, (2) prevent the osteoprogenitor from generating osteoblast (the DNA synthesis being a required event in this process), (3) inhibit new bone formation (osteoblast are required for new bone formation). Hydroxyurea (HU) is a simple compound with the formula: $H_2NCONHOH$ (16). It is relatively nontoxic at effective concentrations, and its effects are reversible by discontinuation of the dosing (16).

In order to monitor the effect of (HU) on osteogenesis (the manufacture of new bone matrix); tetracycline, calcium and xylenol orange labeling was used. Tetracycline and similar chemicals are deposited in vivo at the site of active bone formation (21b). Tetracycline chelates calcium and is deposited at the mineralization front in newly forming bone matrix (10, 18, 19, 21b). The incorporation of tetracycline, etc., at the mineralization front, permanently labels, and points out the bone that was manufactured during the tetracycline dosing period (10, 18, 21b). The label can be visualized by fluorescence microscopy of the undercalcified bone sections (2, 10, 18, 21b).

The experimental design did not permit the direct monitoring of the inhibition of DNA synthesis. Rather, one studied the effects of HU, a known inhibitor of DNA synthesis, on osteogenesis and bone histology. Detectable changes in osteogenesis, osteoblasts, bone lining cells [BLC] and osteocytes are presumed to be due to inhibition of DNA synthesis in osteoprogenitor cells.

In order to study bone histology, bone specimens were decalcified, sectioned and stained by hematoxylin and eosin. Changes in the following parameters were studied:

1. Number of cells with pyonotic nuclei
2. Number of degenerating cells
3. Number of bone lining cells in periosteum
4. Number of bone lining cells in endosteum
5. Number of osteoclast
6. Number of small blood vessels

3.0 OTHER RESEARCH PLAN

3.1 Specific Aims

In order to achieve the objectives (section 1.0) the experimental approach attempts to answer the following questions:

1. Did the treatment protocol inhibit osteogenesis? In order to answer this question, plastic embedded specimen was be studied by Histomorphometry.
2. What types of cells are modified by the treatment protocol? In order to answer this question paraffin embedded specimen were be processed by histological procedures and studied by light microscopy.
3. Does the cytological effect correlate with the inhibition of osteogenesis as indicated by Histomorphometry? In order to answer this question quantitave cytolgical data was be correlated with histomorphometric data.
4. Have the osteoprogenitors been detected by this method? How do the results obtained by this method compare with those reported in the literature?

3.2 Experimental Design

A. Distribution of the animals: forty animal were divided into two groups, a treated group which received hydroxyurea, a control group which received a saline placebo. Both groups received label compound at selected times to be described

(2) Day-6: All animals were treated with the label compound, Xylenol orange, 10 mg per Kg of body weight for one day (21b).

(3) Day-7 through Day-11: dosing of the treated group with (HU) and the controls with saline, five days of treatment.

(4) Day-12: All animals were treated with Calcein at 10 mg per Kg (ob body weight).

(5) Day-13 through Day-17: no treatment.

(6) Day-18: the animals were sacrificed. Carbon dioxide euthanasia was employed. Specimens were taken for histology, histomorphometry and electron microscopy.

Other experimental procedures

After euthanasia, the long bones were removed and prepared for utilization in the following procedures:

(1) histomorphometry to determine the amount of growth or inhibition of growth (18b), (2) histology and light microscopy to visualize the cells killed or injured by (HU), and to determine which cell populations participate in the recovery process; Hematoxylin and eosin stains were used. Differential cell counts were attempted.

3.3 Summary of Methods

Methods used in processing specimens for cytology/histology, scanning electron microscopy and histomorphometry are described below.

Photomicroscopy

Photomicroscopy of histological slides was done on a Zeiss Universal research microscope. The final magnification on the print is a product of the magnifications due to the following lenses objective x optivar x camera x enlarger. Thus the total magnification when the following objectives were used are listed below:

16x objective: 16 x 1.25 x 0.5 x 12.9 = 129.

40x objective: 40 x 1.25 x 0.5 x 12.9 = 322.5

100x objective: 100 x 1.25 x 0.5 x 12.9 = 806.25

4.0 RESULTS

The data will be presented in a format which reflects the categories outlined in the specific aims section. All of the data is from femur bone.

WHAT TYPES OF CELLS ARE MODIFIED BY THE TREATMENT?

Cellular charges have been detected in five compartments of femoral bone: bone marrow, endosteum, periosteum, epiphyseal plate cartilage and fibrocartilage.

The most striking visual changes were seen in bone marrow. In controls the bone marrow is characterized by the presence of the darkly staining nuclei of numerous leukocytes. Megakaryocytes are very obvious. Only a few sinusoids are visible, some with blood cells, some empty. Sinusoid congestion was not detected. Figure 1 shows bone marrow from 8d control. Figure 2 shows bone marrow from a 18d control. They are both typical of bone marrow from controls.

Bone marrow in treated cells is visibly different as early as 3 days after the start of the treatment, figure 3. The most visible early change is erythocyte congestion in sinusoids, figures 3 and 4. In time, the bone marrow becomes relatively depopulated of cells of any type. Figure 4 shows an early stage in the loss of blood cells. Figures 4, 5 and 6 show the increased number of empty sinusoids and the prominence of the reticulum which is normally hidden by the large leukocyte population. Figure six shows an enlargement of an area in figure five.

In addition, numerous cells which might have recently exited mitosis have been detected. Mitotic cells and putative mitotic cells are usually seen near the periosteum bone matrix interface.

Mitosis has been detected in both the hyaline cartilage of the epiphyseal plate as well as in the fibrocartilage of the control specimens. Although the mitotic index was not determined, increased numbers of mitosis were seen in both of those bone compartments. Figure 15 shows two telophases near the top of the field, and several cells which have probably recently exited telophase.

HU induced cellular changes are readily detectable by light microscopy. The periosteum of treated specimens is distinctly different from controls. In particular, the periosteum-matrix interface differs in the following ways: (1) it is more cellular; (2) cell outlines are more distinct; (3) most cells have processes; (4) the nuclei are structured; (5) more osteoclast are detectable. In summary, cellular contours and the cell distribution are indicative of an active tissue.

Figure 16 shows an osteoclast in which the ruffled border is detectable. The osteoclast is in a Howship lacuna. Although this specimen is from an 18d control, osteoclast are more plentiful in treated specimens.

It will be seen in the next section that the cellular changes induced in the periosteum are accompanied by changes in the labeling patter and consequently change in osteogenesis.

DOES THE CYTOLOGICAL EFFECT CORRELATE WITH THE LABELING PATTERN?

The labeling protocol, which has been successful in interest but unexplainable.

In general the treated specimens are distinguishable from the controls in the following ways: (1) the bands in the treated tend to be closer together; (2) the tetracycline band is distinctly different from that in controls in that it tends to be less coherent, or it may be incomplete; (3) the green band is usually more intense and thicker than in controls; (4) deposits of trabecular bone tends to contain much more green label than yellow label. Figure 22 shows a section through the endosteal bands in a treated specimen. There is very little unlabeled growth of bone between the yellow and green labels and the volume of the green label is greater than found in controls; as if the treated specimens are producing more bone during the labeling period in the absence of HU. Figure 23 also shows circumferential bonds of label. There is very little growth between the labels. This scene is from a different location in the same specimen.

Figure 24 shows a section through the epiphyseal plate and adjacent medullary cavity of a treated specimen. The yellow label, tetracycline, is suppressed. Most trabecular bone deposits in control specimens exhibit both labels, figure 25.

The best correlation between cytology and labeling pattern is seen in the periosteal compartment.

The magnification in figure 18 is x619. Otherwise, the total magnification for the fluorescence photomicroscopy is (10 x 1.25 x 3.2 x 12.9) = x516.

It was also possible to visualize changes in periosteum by light microscopy. Figures 7 and 8 show control periosteum. The view at low magnifications is similar to a compound epithelium. Most of the cells are spindle shaped in profile. The cell-to-cell interface is not sharply defined and the nuclei are ovoid, darkly staining and structureless. Figure 8 shows that there is a sharp change in cell morphology at the periosteum bone matrix interface. The cells are irregular in shape and the nuclei are more structured.

Even at low power, figure 9, the treated periosteum is visibly different. The cellular distribution is less regular. There is more extracellular matrix between the cells. A greater percentage of the cells exhibit processes. There are at least two types of cells with processes, some are obviously multipolar and some appear to be unipolar. Figure 10 is an enlargement of a region in figure 9, numerous cells with processes are seen. A multipolar cell is sharply defined in the center of the field.

Figure 11 and 12 illustrate two features of the treated periosteum; the cell outlines are more discrete and more structure is detectable in cell nuclei. An apparently monopolar cell is seen near the center of figure 12.

Figure 13 shows a cell with long branched processes as visualized by scanning microscopy. The cell is from untreated Rhesus periosteum.

Although mitoses are sometimes seen in the cartilage of controls, it is undetected in control periosteum. By contrast a few mitoses have been detected in treated periosteum, figure 14.

directing the deposition of three consecutive labels in Rhesus monkey bone, was not entirely successful. Even in control specimens, a maximum of two labels are detectable. Figure 17 shows two bands of label surrounding the marrow cavity and endosteum. The yellowish tetracycline label is nearest the cortical bone. Small Haversian systems are seen in the cortical bone. The bands are uninterrupted. Figure 18 shows labeling near an active thick periosteum. Small Haversian systems are seen between the two bands of label. Figure 19 shows labeling near a thin periosteal region. Both 18 and 19 are from control specimens. In controls, the bands surrounding the periosteum are also uninterrupted, i.e., two complete bands are detectable. The thickness of unlabeled bone between the labels is easily visualized and measurable. Figures 18 and 19 illustrate one of the complexities of the study. That is, periosteum is not equal in all areas of the femur. Periosteum nearer the epiphyseal plates tends to be thick and more cellular. The labeling results seems to indicate that it is also more active in matrix production. Thus, it is necessary to compare control thick periosteum with thick periosteum from treated femurs and similarly for the thin periosteums. Thus Figure 19 should be compared with figure 20 which shows thin periosteum from a treated femur. In order to quantify the difference volume measurements are necessary. In general the labels are more widely separated in control specimen than in treated. Figure 21 shows a treated thick periosteum and adjacent label. The green fluorescence associated with the skeletal muscle is of some

5.0 DISCUSSION AND SUMMARY

Figure 25 shows the basic organization of the femur. Hydroxyurea (HU), a known inhibitor of DNA synthesis induced cellular changes in five bone compartments: bone marrow which is contained in the medullary cavity; the endosteum which lines the medullary cavity; periosteum; epiphyseal plate hyaline cartilage and fibrocartilage. The epiphyseal plate is not shown in the diagram. In a growing bone, the epiphyseal plate [EP] is located at the junction between the shaft and epiphysis.

The effect on bone marrow is most dramatic. Bone marrow is a tissue with high cell turnover rate and a high rate of DNA synthesized. HU caused the cell population to decrease drastically. The decrease was presumably due to both cell maturation and inhibition of proliferation of stem cells. The HU effect was reversed by discontinuing the treatment, in which case the bone marrow regenerated. It is not known if the bone marrow effect is incidental or whether it is correlated with inhibition of osteogenesis.

The bone marrow reticulum cells and the endosteal BLC appear to form a continuum. In recovering bone marrow the normally single layered endosteum often contained from two to five layers of cells. Surprisingly, very few mitoses were found in the bone marrow. Since HU has a synchronizing effect in other systems it is possible that most of the mitoses occur as one or more large "bursts" within the first eight hours after ending the treatment. Since sampling was not done until 5 days after stopping the treatment. Most of the mitoses would have been missed.

An increased number of mitoses were observed in epiphyseal cartilage and also fibrocartilage. The mitotic figures were usually telophases. It is possible that earlier sampling would have revealed a much higher mitotic index in the cartilage compartments.

At this point in the study the best correlation between cytology and osteogenesis is seen in the periosteum. One detected an enrichment in the number of multipolar cells. The cells appear to be more from the midreogion of the periosteum toward the periosteum-matrix interface. The cells at the interface have an enlarged structured nucleus. Mitotic figures were detected at the interface. The mitotic figures were seen in a "plump" polygonal cell type. The interphase of this cell type exhibited large structured nuclei. The labeling pattern of the circumferential bands associated with the periosteum was modified by the HU treatment. The endosteal circumferential bands are separated by less unlabeled bone in the treated as compared with the controls.

In summary, HU, a DNA synthesis inhibitor, disrupts osteogenesis and points out a large cell with processes as a probable osteoprogenitor in the periosteal compartment.

Osteoclast were more active in the treated specimens than the controls.

Histomorphometry was initiated, and some measurements were made. The analysis could not be completed because the computer system was unable to carry out the volume determination functions at the time of my visit to Ohio.

A more definitive answer to the osteoprogenitor question, as well as a more effective tagging effect could probably be achieved by using monoclonal antibodies.

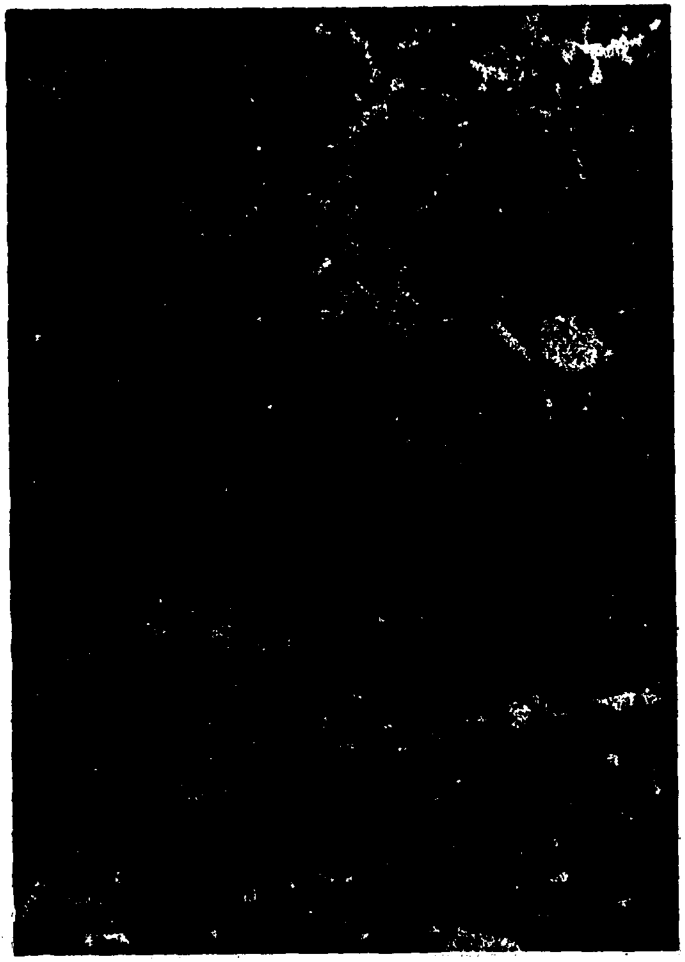


FIGURE 1: Bone Marrow 8d Control, 1 x 129

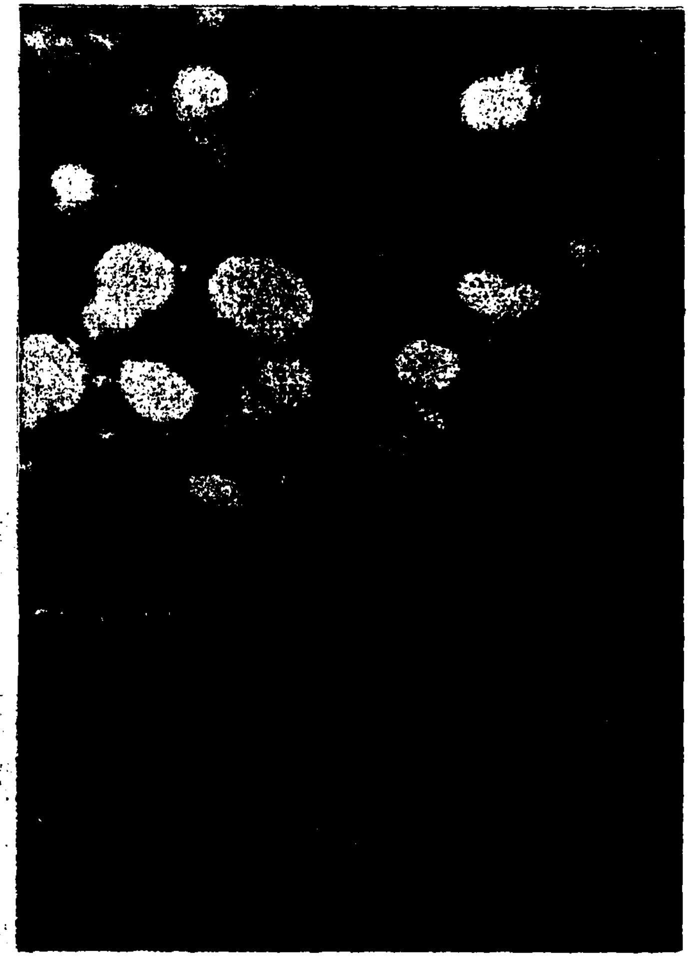


FIGURE 2: Bone Marrow 18d Control, x 129

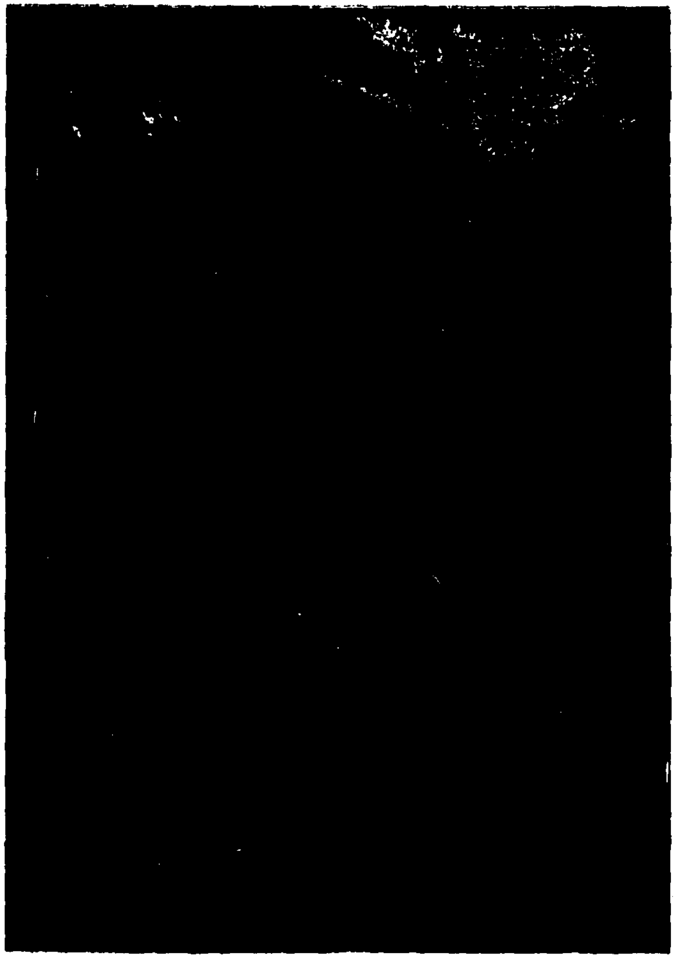


FIGURE 3: Bone Marrow 3d HU treatment x 129

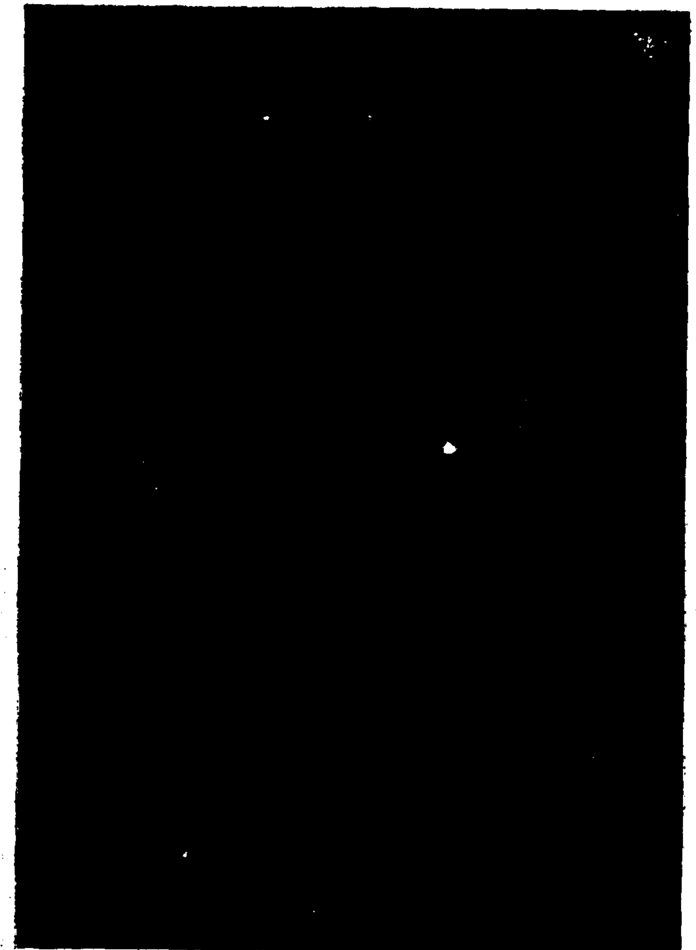


FIGURE 4: Bone Marrow 8d HU treatment x 129

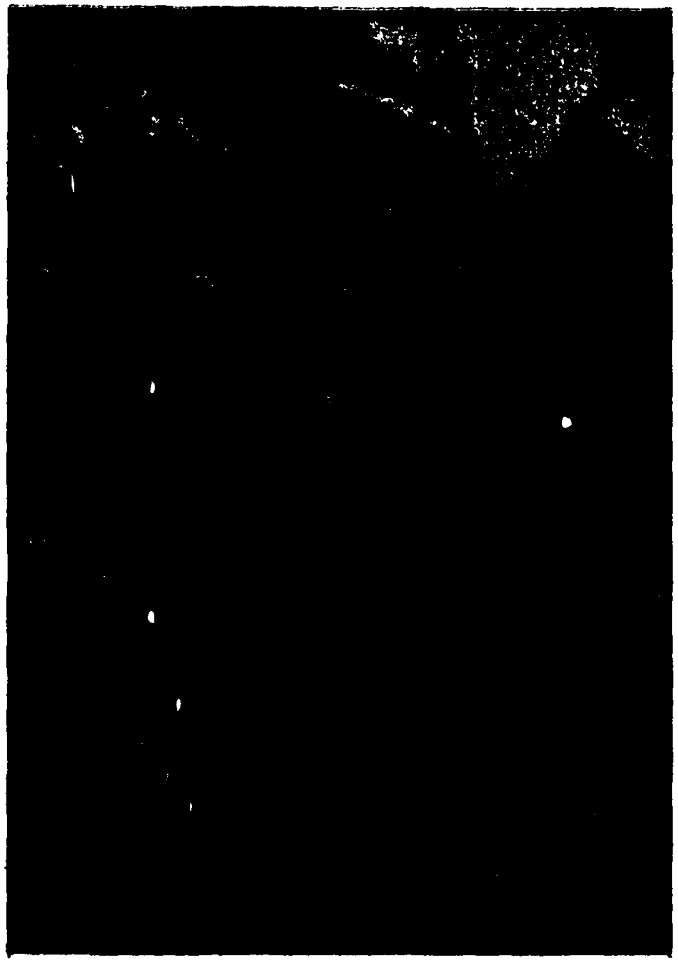


FIGURE 3: Bone Marrow 3d HU treatment
x 129

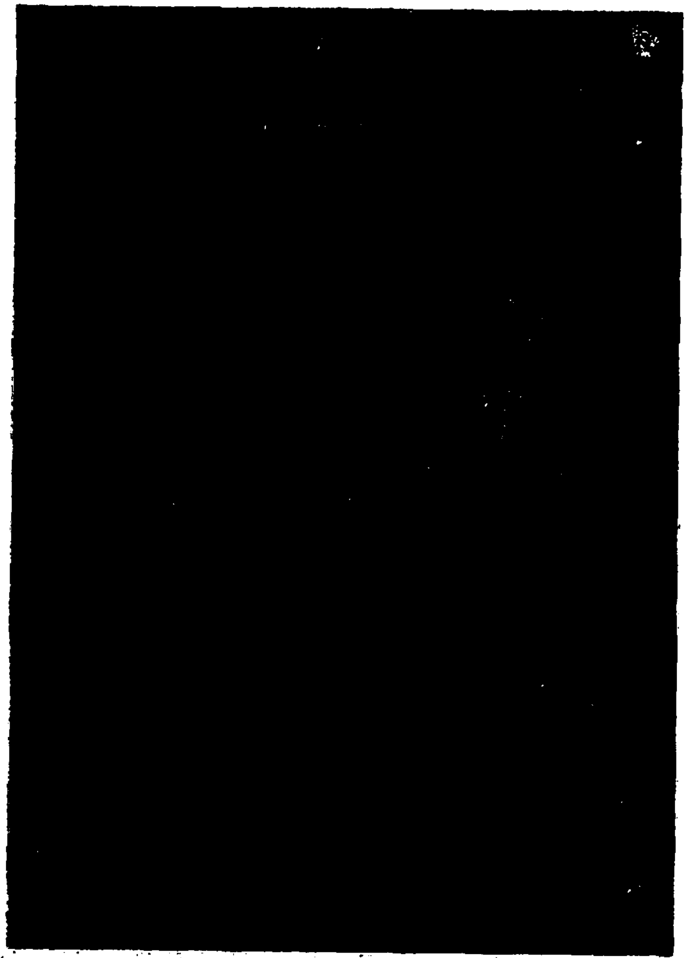


FIGURE 4: Bone Marrow 8d HU treatment
x 129

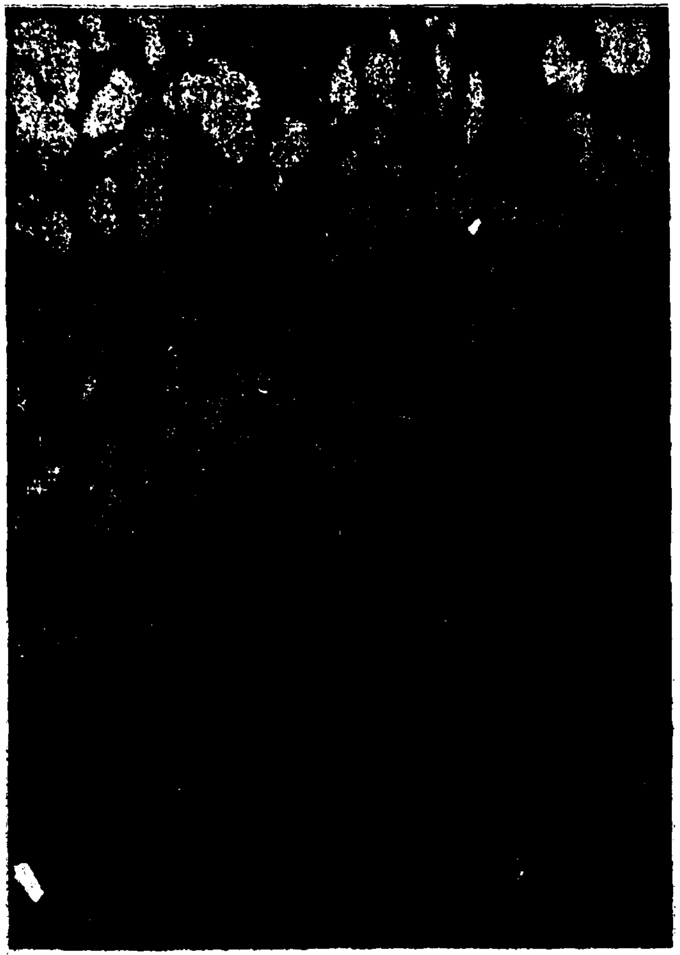


FIGURE 5: Bone Marrow 8d HU Treatment,
x 129

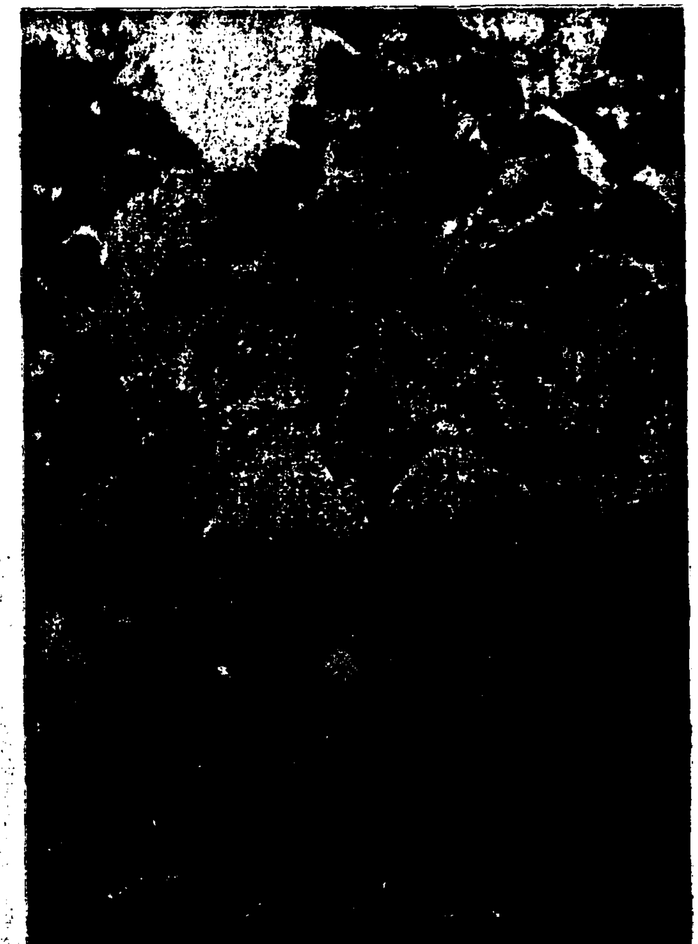


FIGURE 6: Bone Marrow 8d HU Treatment
x 322

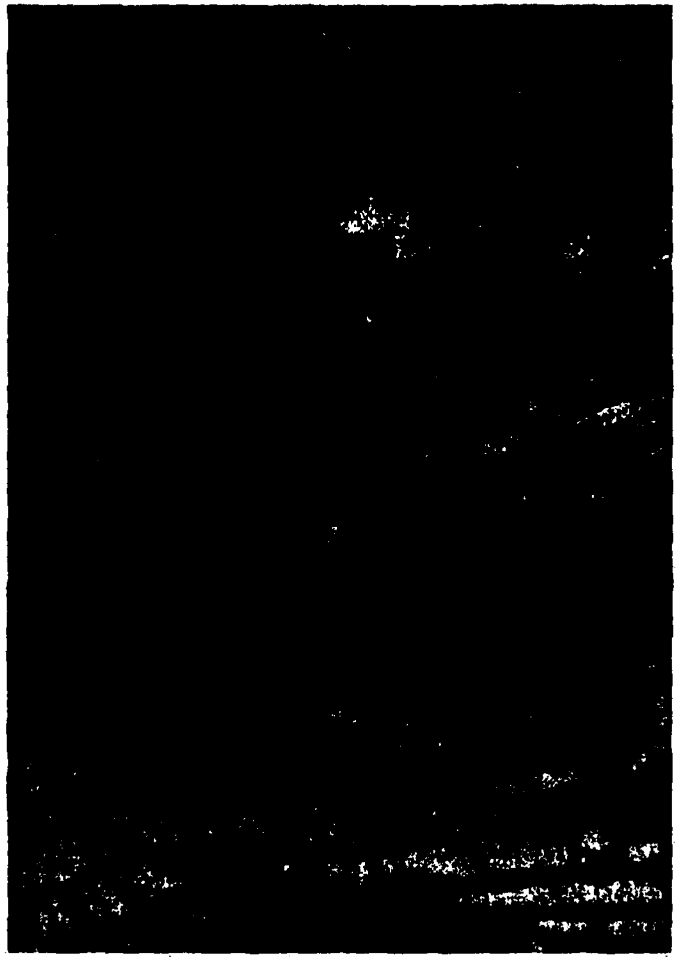


FIGURE 7: Periosteum 3d control, x 129

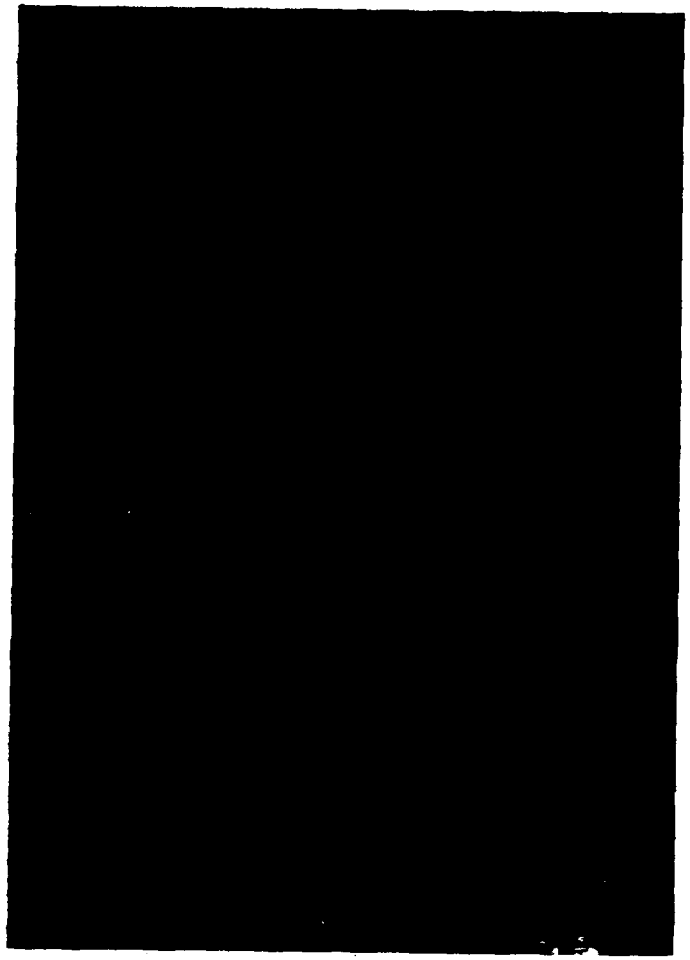


FIGURE 8: Periosteum 3d control, x 322

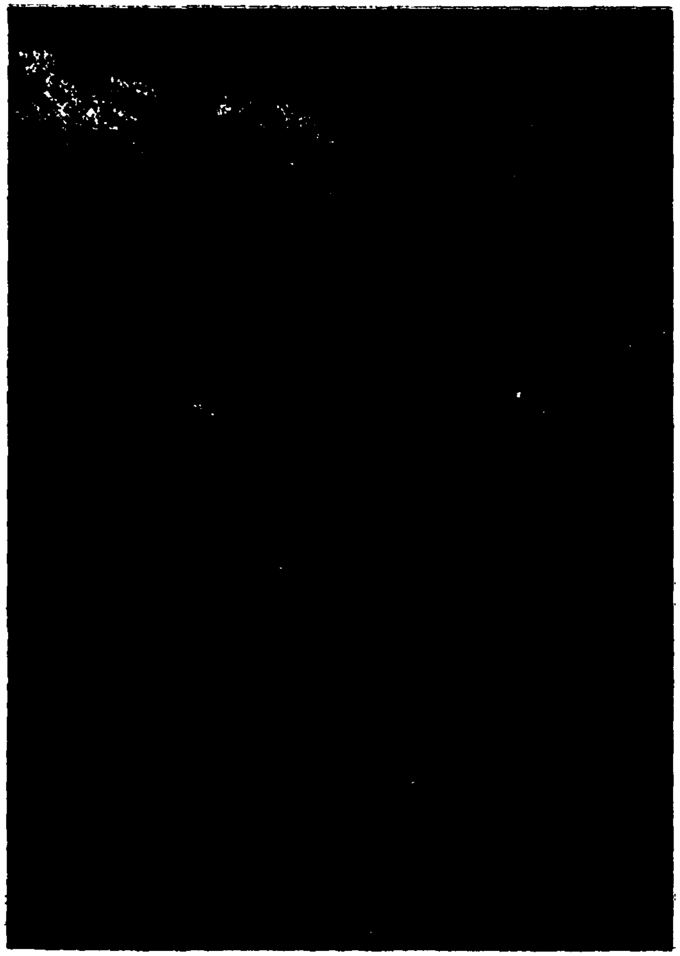


FIGURE 9: Periosteum 18d treated, x 129

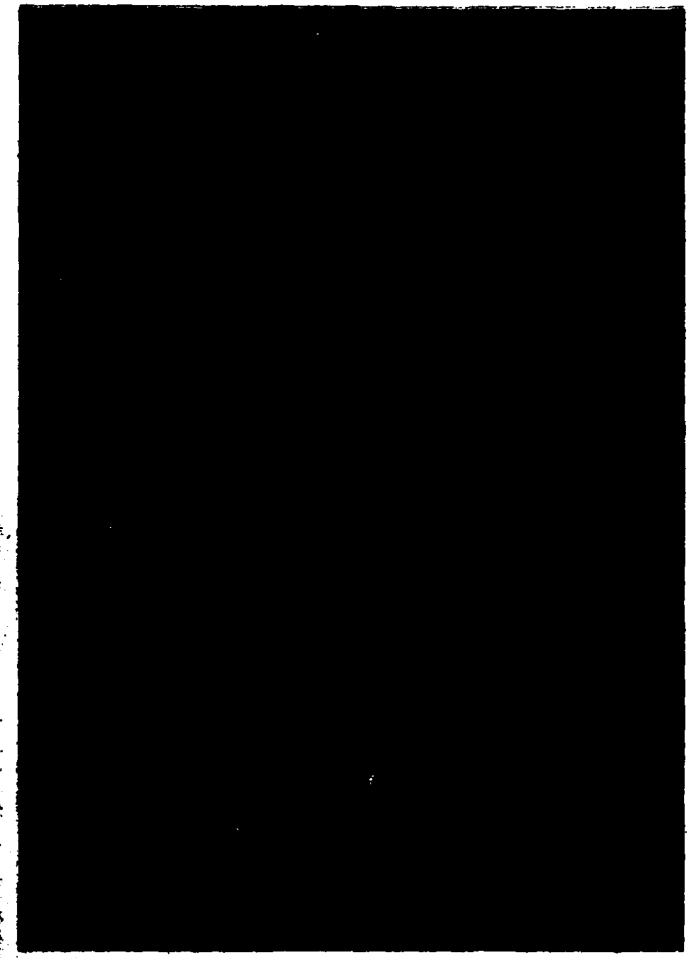


FIGURE 10: Periosteum 18d treated x 322

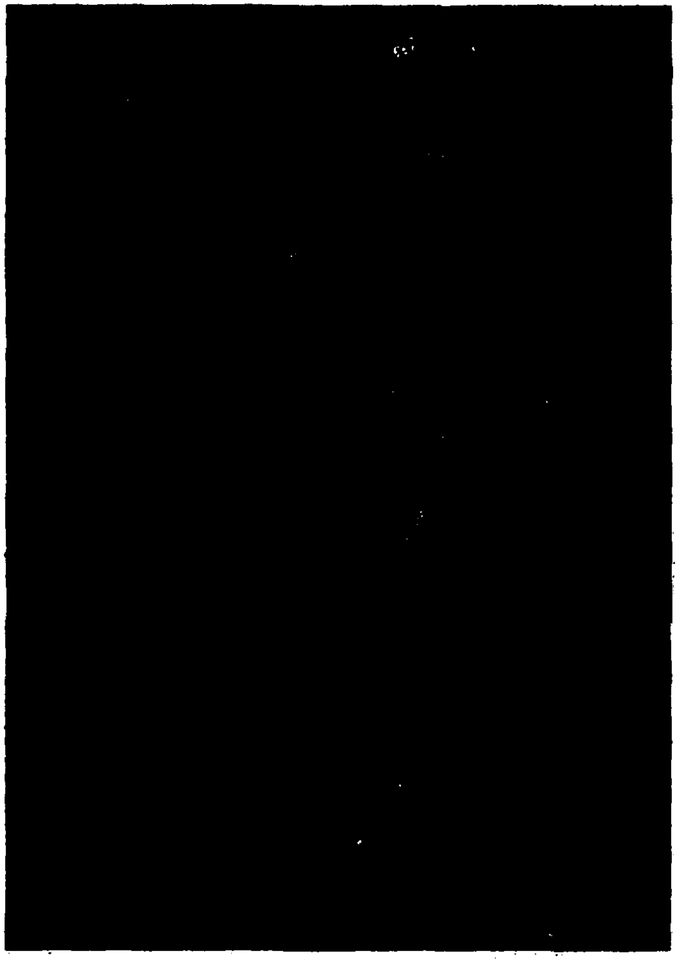


FIGURE 11: Periosteum 4d treated
x 322

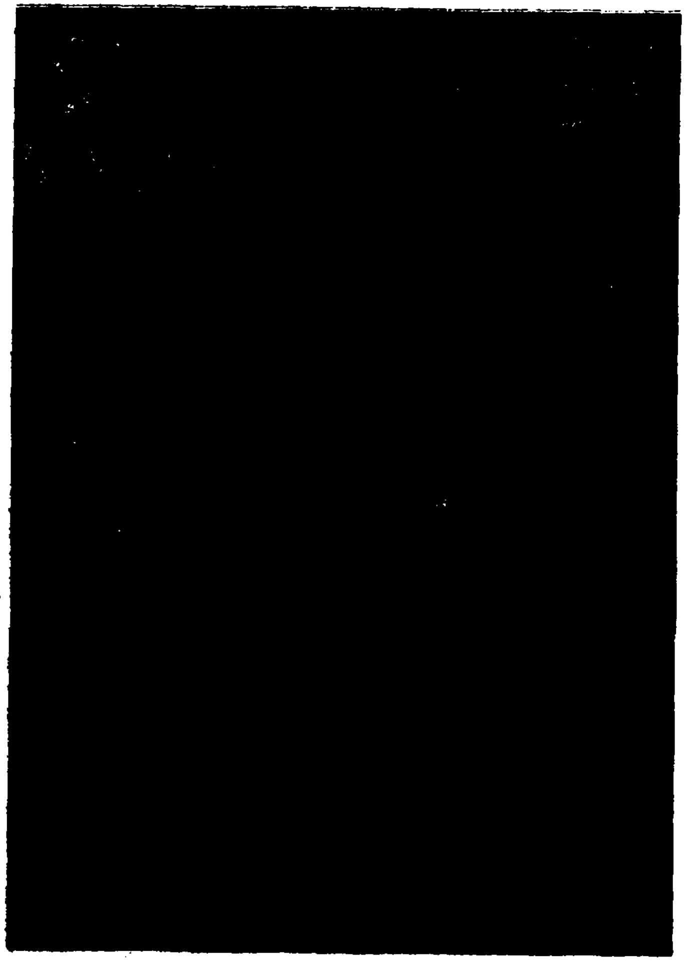


FIGURE 12: Periosteum 4d treated
x 322

FIGURE 13: Cell With Processes untreated, S.E.M.

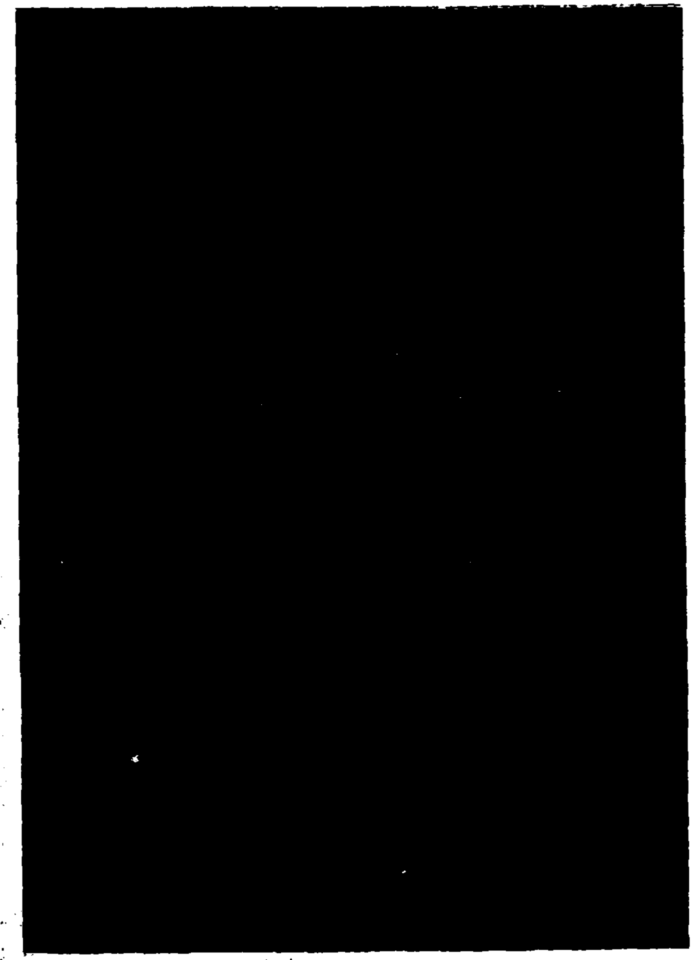


FIGURE 14: Prophase in Periosteum 18d treated, x 322

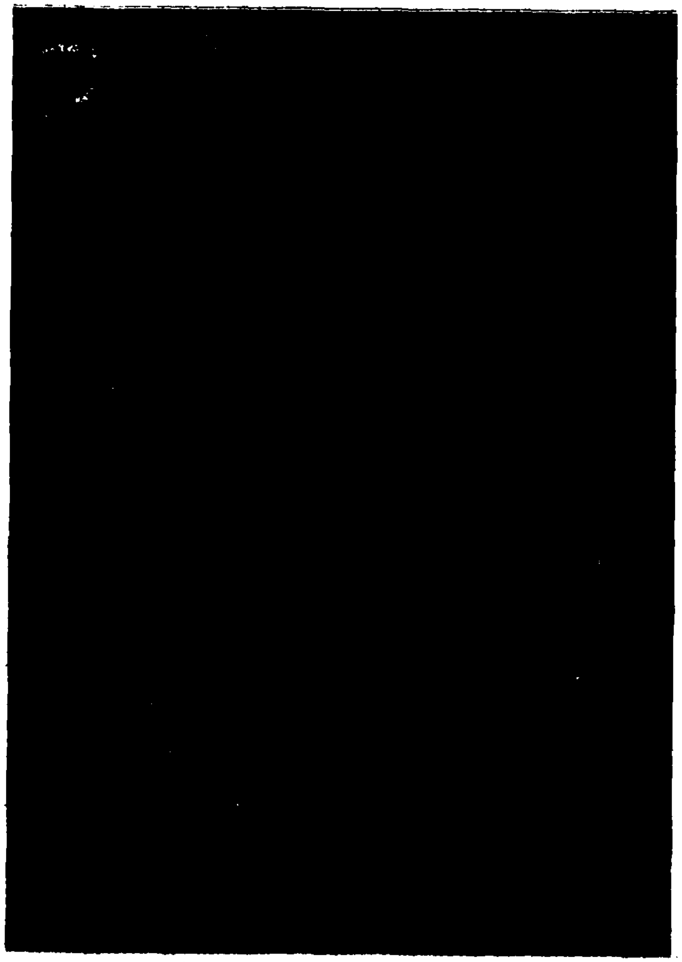


FIGURE 15: Telophase Hyaline Cartilage
18d treated, x 322

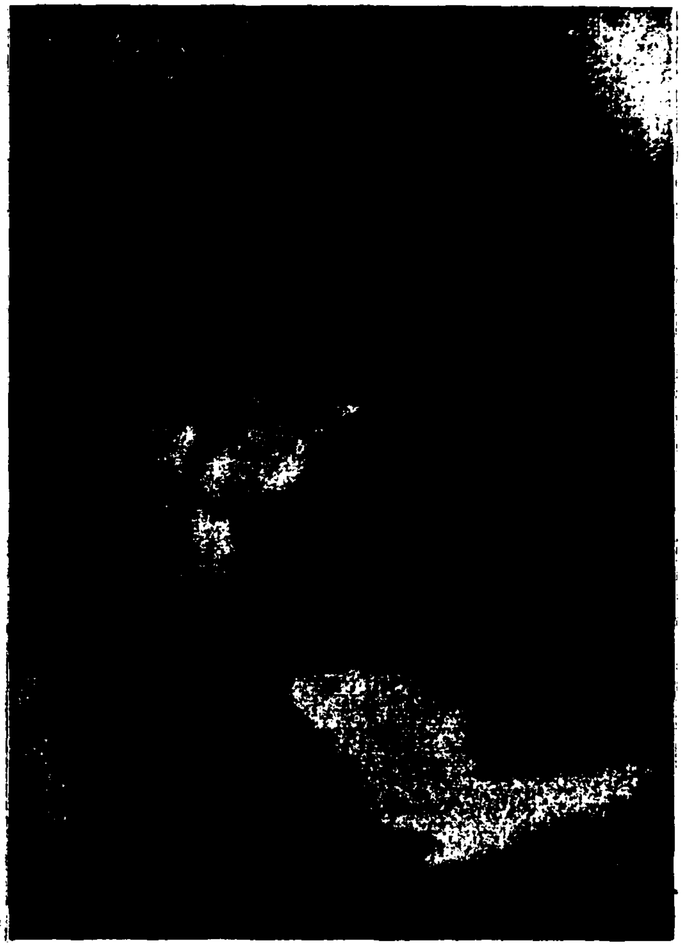


FIGURE 16: Osteoclast in Howship
Lacuna 18d control, x 806

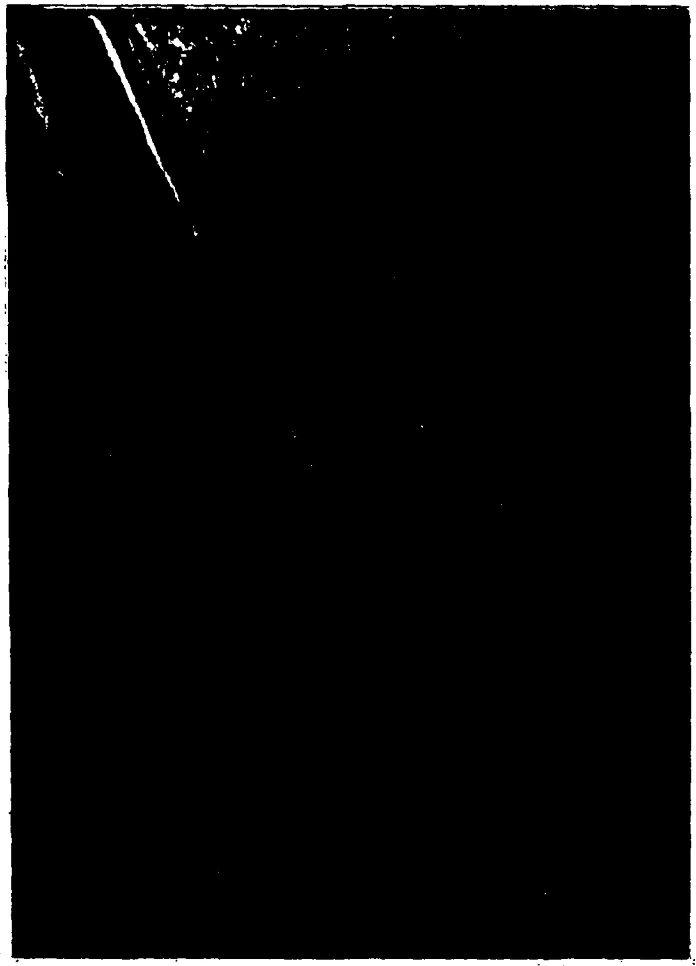


FIGURE 17: Endosteal Circumferential Labeling, 18d Control

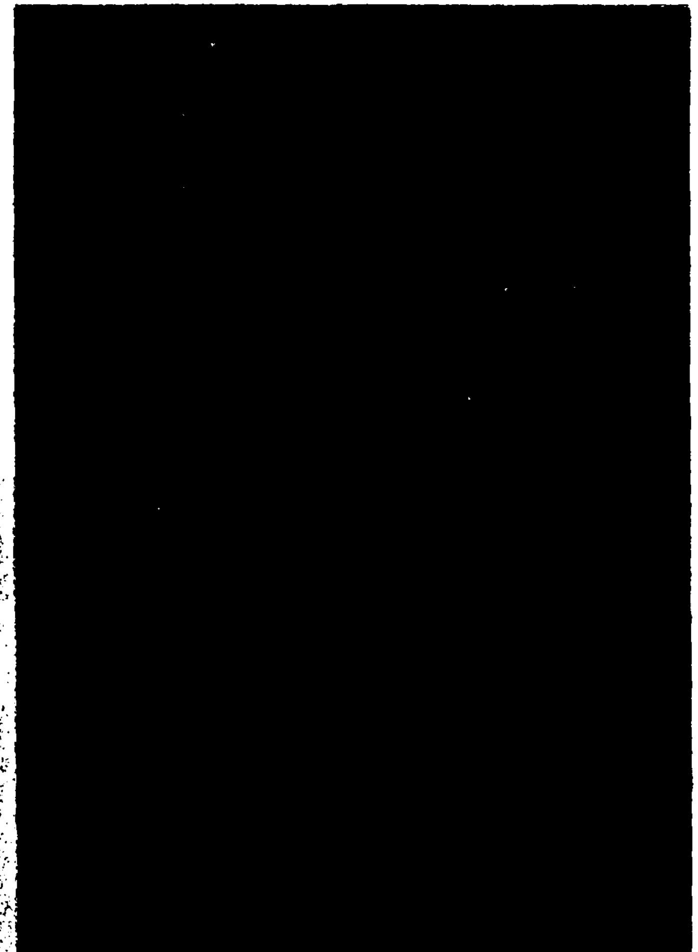


FIGURE 18: Periosteal Circumferential Labeling, 18d D Control

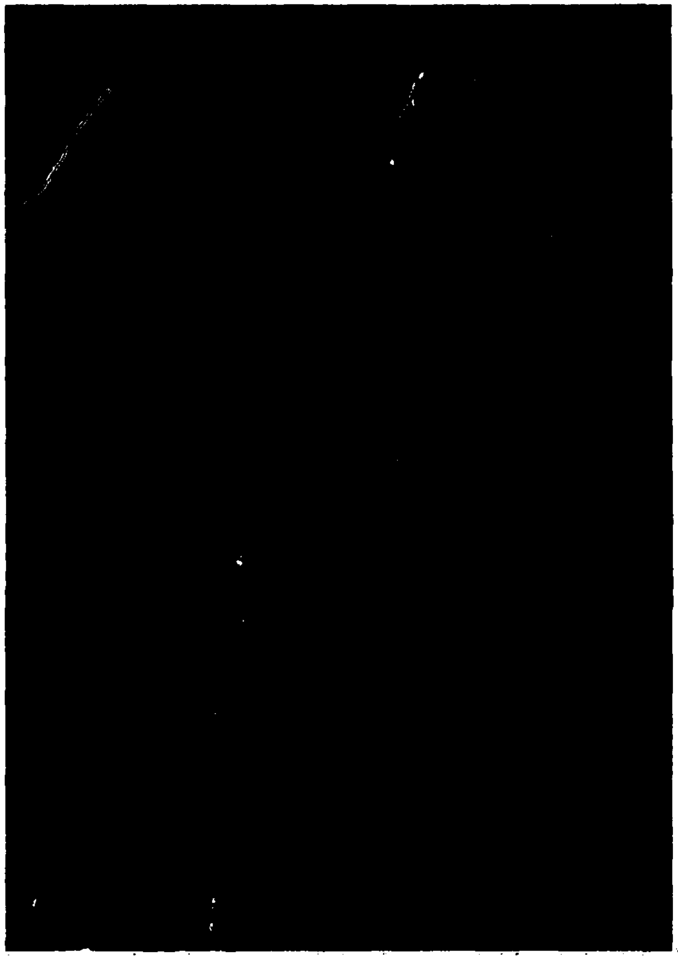


FIGURE 19: Periosteal Circumferential Labeling, 18d Control

FIGURE 20: Periosteal Circumferential Labeling, 18d Treated

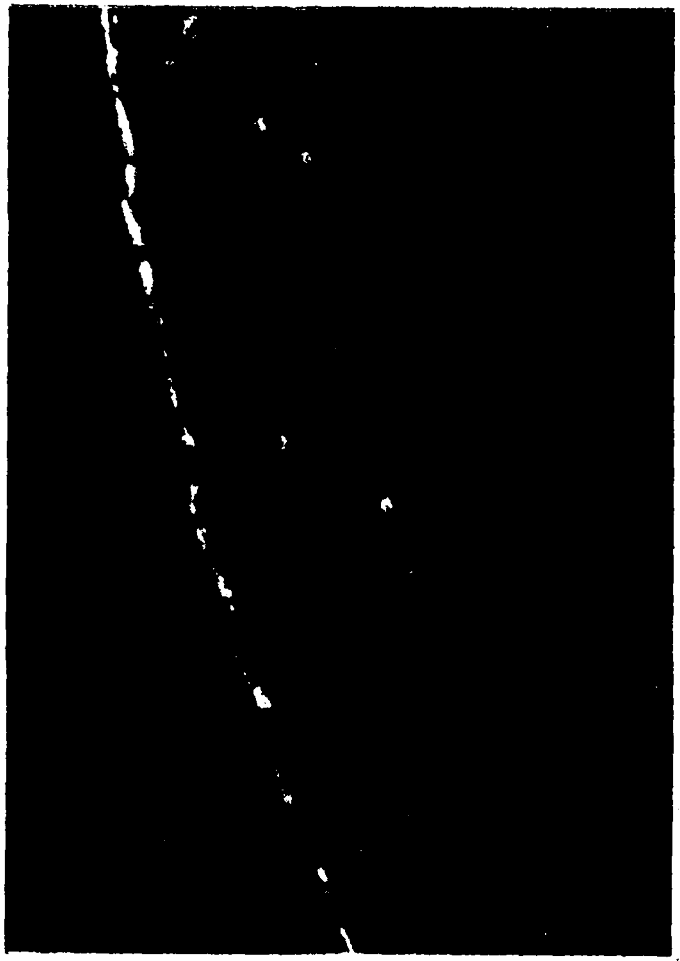


FIGURE:21: Periosteal Circumferential Labeling; 18d treated

FIGURE 22: Endosteal Circumferential Labeling, 18d treated

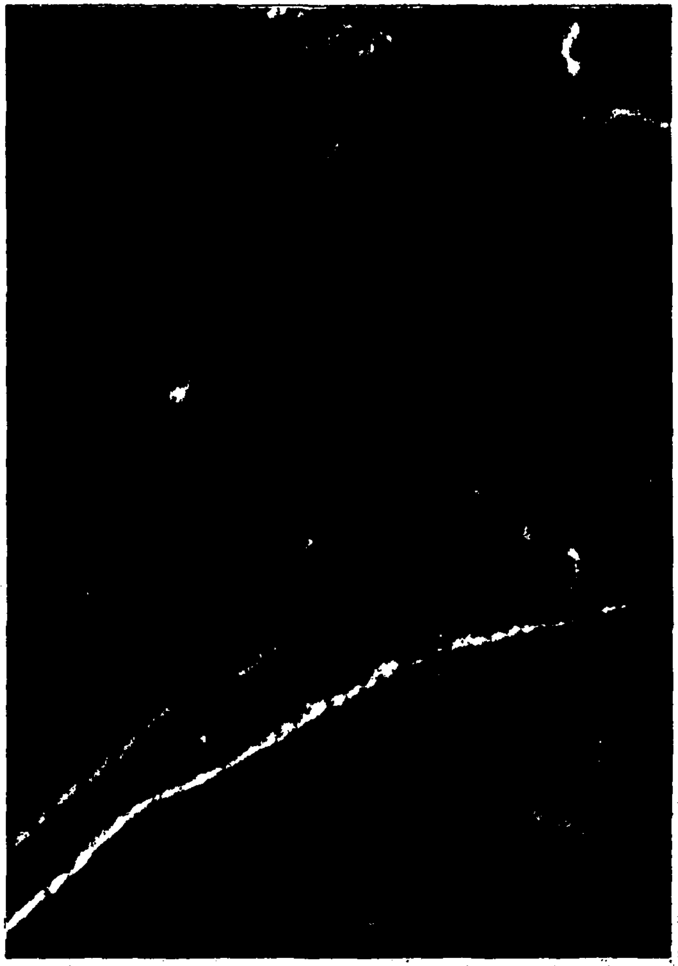


FIGURE 23: Endosteal Circumferential Labeling, 18d Treated

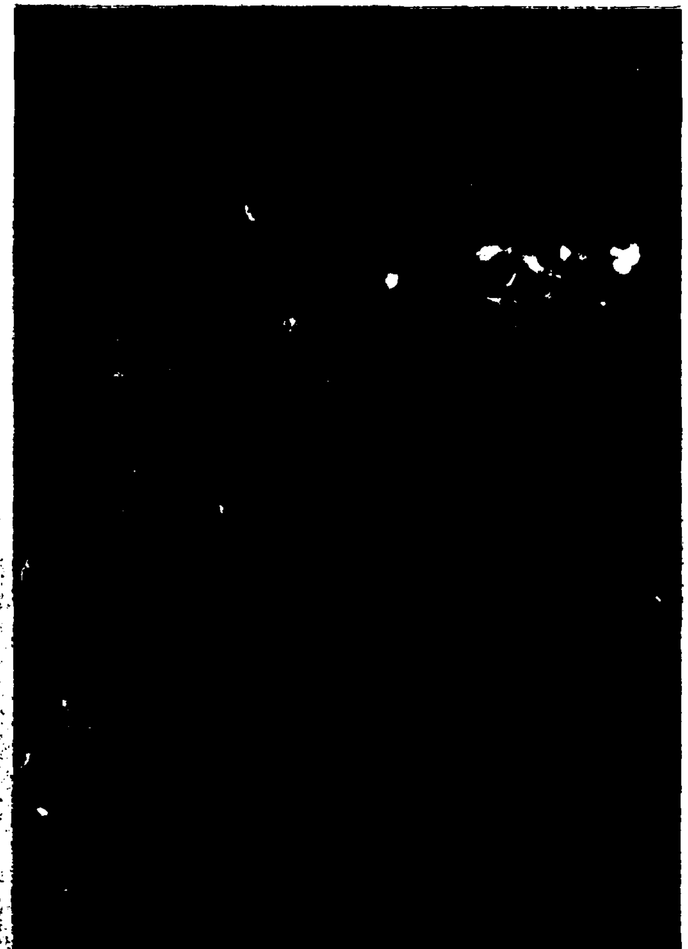


FIGURE 24: Epiphyseal Plate and Environs 18d Treated

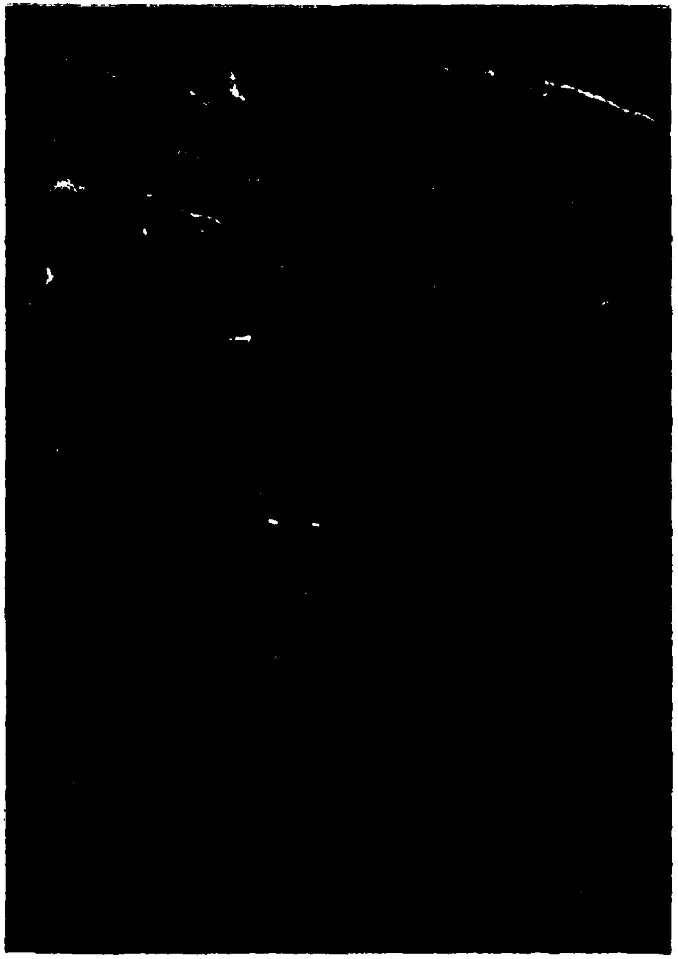
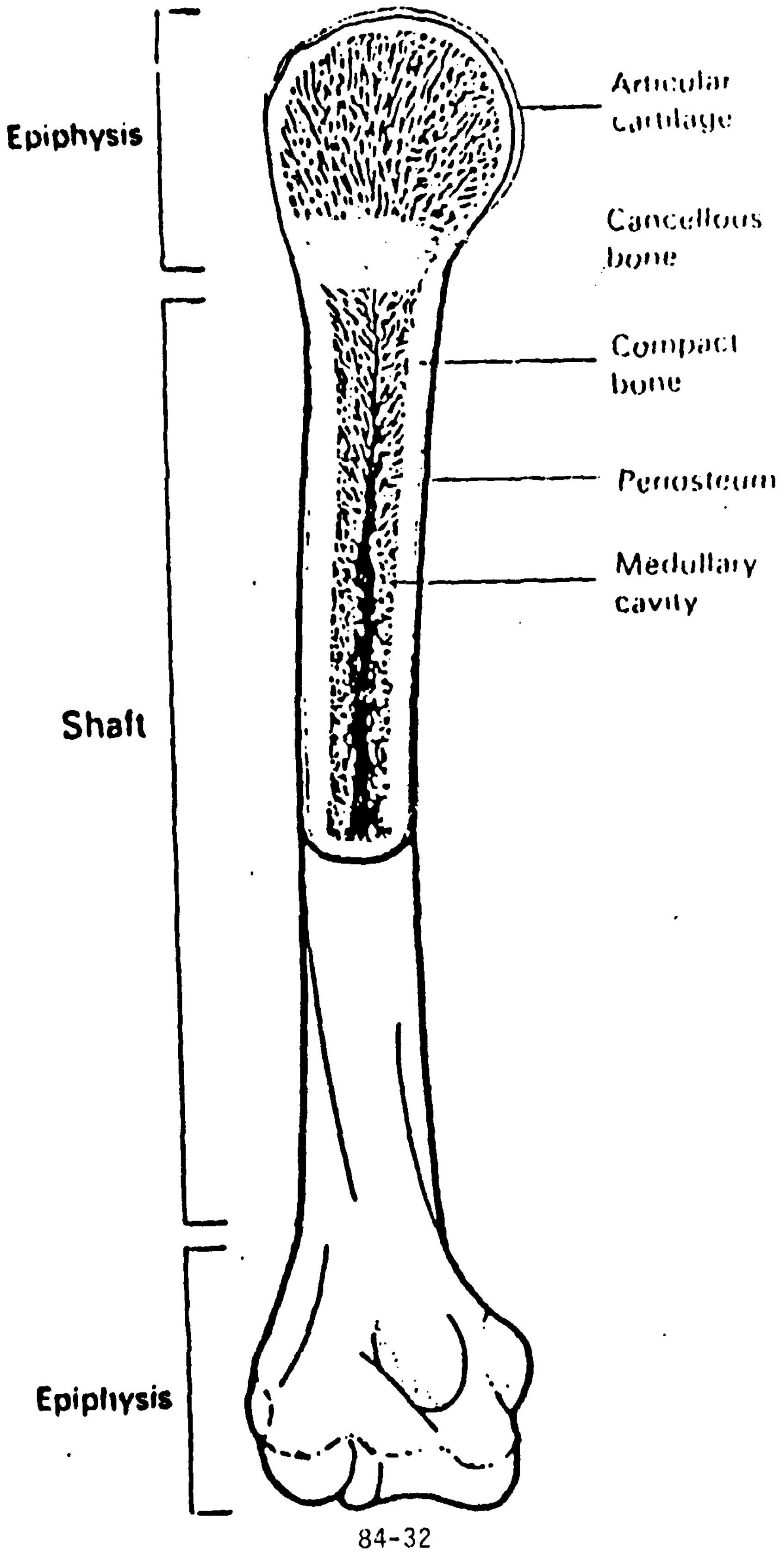


FIGURE 25: Trabecular Bone 18d Control

FIGURE 26


Epiphysis
Articular cartilage
Cancellous bone
Compact bone
Periosteum
Medullary cavity
Shaft
Epiphysis

6.0 REFERENCES

1. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, Molecular Biology of the Cell, pp 611-668, 918-947, Garland Publishing Company, Inc. New York, 1983.

2. Anderson, Colin, Manual for the Examination of Bone, pp 1-109, Boca Raton, FL, CRC Press, Inc., 1982.

3. Ashihira, E., et al., "3H-thymidine Autoradiographic Studies in Electrically Stimulated Osteogenesis", in C.T. Brighton, J. Black and S.R. Pollace (eds): Electrical Properties of Bone and Cartilage, New York, Grune and Stratton, pp 401-425, 1979.

4. Farley, J. R., T. Masuda, J.E. Wergedal and D. Baylink, Human Skeletal Growth Factor: Characterization of Mitogenic Effect on Bone Cells in vitro, Biochemistry, 21:3508-3513, 1982.

5. Frost, Harold, Tetracycline Based Histological Analysis of Bone Remodelling, Calc. Tiss. Res. 3: 211-237, 1969.

6. Goldstein, Auram, Biostatistics, pp 51-92, New York, The MacMillan Company, 1969.

7. Huggins, C.B., The Formation of Bone Under the Influence of Epithelium of the Urinary Tract, Archives of Surgery, 22: 377-407, 1931.

8. Huggins, C.B., Epithelium Osteogenesis-A Biological Chain Reaction, Proc. Amer. Philos. Soc. 113:458-463, 1969.

9. Huhmann, E.L., et al., Innervation of Periosteum and Bone by Sympathetic Vasoactive Intestinal Peptide-containing Nerve Fibers, Science, 232:868-871, 1986.

10. Jankovich, J.P., The Effects of Mechanicl Vibration on Bone Development in the Rat, J. Biomechanics,5:241-250, 1972.

11. Landis, W.J., "Special Techniques for Problem Tissues, Bone and Collagen, in B.R. Jones (ed.) **Electron Microscopy**, 41 Exercises by 17 Scientists, Monroe, New York, Library Research Associates, Inc. pp 328-339, 1985.

12. Linder, J.E., A Simple and Reliable Method for the Silver Impregnation of Nerves in Paraffin Sections of Soft and Mineralized Tissues, J. Anat., 127:543-551, 1978.

13. Menton, D.N., D.J. Simmons, S.L. Chng and B.Y. Orr, From Bone Lining Cell to Osteocyte--An S.E.M. Study, Anat. Rec., 209: 29-39, 1984.

14. Milch, R.A., D.P. Rall and J.E. Tobie, Basic Localization of the Tetracyclines, J. National Cancer Inst., 19: 87-93, 1957.

15. Smith, K.C., C.M. Oloff and L.E. Kazarian, AFAMRL-TR-81-122.

16. Timson, J., Hydroxyurea, Cancer Research, 32: 115-132, 1975.

17. Wlodarski, Krzysztof, Failure of Heterotrophic Osteogenesis by Epithelial Mesenchymal Cell Interactions in Xenogenic Transplants in the Kidney, Calcif. Tiss. Res. 25: 7-11, 1978.

18. Wronski, T.J. and E.R. Morey, Inhibition of Cortical and TRabecular Bone Formation in Long Bones of Immobilized Monkeys. NASA Technical Memorandum.

19. Urist, M.A., and K.H. Ibsen, The Chemical Reactivity of Mineralized Tissue with Oxytetracycline, Arch. Pathol. 76: 484-496, 1963.

20. Yagiela, J.A., and D. Woodbury, Osteoblast Isolation from Fetal Rat Calvaria, Anat. Rec., 188: 287-307, 1977.

21. Aerospace Medical Association, 58th Annual Scientific Meeting, May 10-14, Las Vegas, NV, 1987, abstracts:

21a. Girten, B., C. Oloff, L. Kazarian, S. Bloomfield, B. Smith, and R. Hamlin, Effects of Dobutamine on Suspension Hypokinesia/hypodynammia Deconditioning in Rats, #109, A19.

21b. Eveland, E.K. SMith, C. Oloff, K. Swenson, M. Souder, L. Kazarian, A Comparison of Bone Labelling Compounds Tetracycline, Calein, Xylenol Orange, and DCAF, #210, A34.

AFOSR RIP PROGRAM

1987-1988

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

"Trauma-Activated Periosteum Derived Osteogenic Cells: Response to Selected Growth Factors"

Prepared by: Noel S. Nussbaum, Ph.D.

Academic Rank: Associate Professor

Department and University: Physiology and Biophysics, Wright State University

Date: 25 September 88

Contract No: S-760-7MG-089

Trauma-Activated Periosteum Derived Osteogenic Cells: Response to Selected Growth Factors

by

Noel S. Nussbaum

ABSTRACT

Closed green-stick rib fractures were produced, by manual pressure, in anaesthetized white New Zealand male rabbits. After 5 days of healing, the enlarged periosteum of the fracture site was collected and subdivided, by sequential collagenase-trypsin digestion, into primary cell cultures. Semi-defined media (BGJ_b, Gibco) supplemented with hydrocortisone and insulin, was used to establish these non-transformed cells *in vitro*. Similar cultures were prepared from embryonic chick calvariae. Alkaline phosphatase secretion into the media was monitored and comparison of data derived from the two species supports the osteogenic identity of the cells isolated from the rabbit periosteum. By repetitive passage of high alkaline phosphatase producing cell lines, isolates were identified which could be maintained *in vitro*, in the presence of ILGF (Somatomedin C) and PDGF for up to three months. Subsequent exposure to transient low Mg^{++} Ca^{++} media resulted in structural changes consistent with osteocytic morphology *in vivo*. It is suggested that long-term, *in vitro* experiments with established cell populations, may accurately reflect osteogenic responsiveness *in vivo*.

ACKNOWLEDGEMENTS

The support of the Air Force Systems Command and the Air Force Office of Scientific Research is gratefully acknowledged. This work could not have been initiated without the interest and intellectual support of Leon Kazarian, Chief, Biodynamics Effects Branch, Armstrong Aerospace Medical Research Laboratory, whose continuing encouragement was always appreciated.

I also acknowledge the valuable technical assistance of Terri Caprio who performed her laboratory functions with intelligence, care, and speed.

Introduction

Periosteal osteogenesis has been investigated by: histologic analysis following heterotopic transplantation (Ham, 1930; Haldeman, 1932; Urist and McLean, 1952; Poussa and Ritsila, 1979), ultrastructural description of organ cultures (Tenenbaum et al, 1986) and by biochemical determinants of bone formation in vitro (Smith et al., 1973; Nijweide, 1975; Nijweide and vander Plas, 1979; Nijweide et al., 1981; Tenenbaum and Heersche, 1982, 1986). Studies by Fell (1932) and Gaillard (1934) demonstrated that the periosteum, or cells derived from it, could produce osteoid in vitro. This was later confirmed by Fitton-Jackson and Smith (1957), Goldhaber (1966), Binderman et al. (1964) and Marvaso and Bernard (1977).

While these studies demonstrated in vitro osteogenesis, the culture conditions and stimuli required for osteodifferentiation and osteogenesis were not investigated. Other workers have tried to examine the conditions necessary for in vitro osteogenesis. Raisz et al. (1976) implicated phosphates in enhanced osteogenesis in vitro and Brighton et al. (1976) demonstrated the significance of a low oxygen concentration in the oseodifferentiation of cultured periosteum.

Recently, the significance of "local factors" has been emphasized, in contrast to the general systemic influences of calcium regulatory hormones, in the osteoinductive process. These local regulators of skeletal growth have been classified into four major categories: 1) polypeptide growth factors; 2) cartilage and osteoinductive factors; 3) blood cell-derived factors; and 4) prostaglandins, (Centrella and Canalis, 1985). Less well identified tissue proteins have been studied by Urist and coworkers for several years (Syftestad et al., 1984) as chondrogenic and osteogenic active agents, while chemotactic agents that may be identical to Urist's "Bone Morphogenic Protein" have recently been partially characterized (Tenenbaum and Heersche, 1986).

A common source for most, if not all, of these agents are immune cells and formed blood elements associated with tissue injury and repair. It is not surprising therefore that recently studies have appeared that attempt to identify the "coupling factors" that integrate osteoclastic resorption and osteoblastic formation during bone remodeling, with these same local factors (Mundy, 1987). Such identity would greatly simplify our understanding of remodeling control and those disease states associated with its failure, e.g. osteopetrosis and osteoporosis.

If, as has been proposed, remodeling is in part a mechanism to repair microfractures produced by biomechanical forces (Parfitt, 1984; Baron et al., 1983) then the production of local factors in such situations is an extension of a normal response to tissue injury. Accordingly, a more extensive bone injury, i.e. fracture, should also be associated with the appearance and action of these same agents in greater amount (Simmons, 1980). Ham and Harris (1971) described an easily reproduced in vivo fracture repair model. By exploiting aspects of this model it is possible to produce in vivo, a high density osteoinductive cell population that can be harvested for the in vitro study of the response of such cell populations to polypeptide growth factors.

The present study presents the results of long-term exposure of "osteoblast-like" cells to Platelet Derived Growth Factor (PDGF) and Insulin-Like Growth Factor (ILGF) or Somatomedin C.

MATERIALS AND METHODS

Male White New Zealand rabbits (4-6 kg) were anesthetized with Ketamine (150 mg, I.M.) and Pentobarbital (32 mg, I.V.). Under anesthesia, rib fractures were produced by finger pressure. The animals were allowed to recover and observed for five days for signs of distress or pain. After five days, the animals were again anesthetized with Ketamine and Pentobarbital. The fracture site and contralateral control tissue was collected aseptically and placed in ice-cold, sterile Phosphate Buffered Saline (PBS). The animals were terminated with 0.5 ml of T-61. A total of 4 animals were used to establish fracture sites for tissue collection.

Calvaria were collected from 18-20 day chick embryos (Carey Farms, Inc., Marion, OH) and stored in ice-cold sterile PBS. Both rabbit and chick material were manually trimmed of non-osseus tissue and the appropriate material, (rabbit periosteum or chick bone) was minced and transferred to digestion chambers. Sequential enzymatic digestion with trypsin and collagenase (Cooper Biomedical) according to Cohn and Wong (1979), was employed to prepare separate cell populations at 20 minute intervals. Each separated population was seeded into a 75 cm^2 culture flask (1 x 10^6 cells) and 10 ml of media added. All cultures were fed with BGJ_b medium supplemented with 10% Fetal Bovine Serum (FBS) and other factors as indicated. All culture media was supplied by Gibco Laboratories, Grand Island, N.Y., and was supplemented with 200 U/ml Penicillin and 200 µg/ml Streptomycin. In addition, insulin, 0.5 units/ml (Squibb) and hydrocortisone acetate, 0.8 µg/ml (United States Biochemical Corp., Cleveland) were added to the BGJ_b medium as supplements for the first 24 hours *in vitro*. At the first media change after 24 hours *in vitro*, hydrocortisone was omitted from the medium.

Embryonic cultures were passaged to confluency then released into Trypsin-EDTA (Sigma) and subsequently split 1:3 for subculture. Each media fraction was collected every 2-3 days when media was replaced.

Biochemical assays were performed on media collected from each culture flask independently. Protein determinations were performed after Smith et al. (1985) with reagents supplied by Pierce (Rockford, IL). Alkaline Phosphatase was measured by the hydrolysis of p-Nitrophenylphosphate (Boehringer Mannheim Diagnostics, Houston, TX). Enzyme activity is expressed as units/hr/mg protein.

All values were normalized by subtraction of media controls. Assay quality control was established with assayed control serum (SeraChem, Fisher Scientific).

Assay data from each flask was evaluated separately at confluence plus 14 days. The rabbit-derived cell lines with the highest alkaline phosphatase production from each original explant were selected for further passage and maintained. Four high alkaline phosphatase periosteum derived cell lines were established. Assay data from these cell lines were statistically evaluated by a general linear model (SAS), found not to be different and randomly distributed for subsequent exposure to PDGF (Collaborative Research, Bedford, MA) or ILGF (Imcera, Terra Haute, IN).

The four cell lines were exposed to both growth factors. After preliminary testing, optimum concentrations of PDGF (8 units/ml) and ILGF (80 µg/ml) were added to either BGJ_b or MEM media. No differences due to media composition were observed and MEM (Gibco) was used for subsequent trials. Cultures were then maintained at confluence for 77 days (PDGF) and 92 days (ILGF). Before termination, flasks were treated with calcium and magnesium free saline (GKN, Merchant et al., 1960) to prepare the cells for Trypsin-EDTA dispersal.

Photographic records were maintained of all flasks throughout the culture period. Histological preparations were made from formic acid decalcified ribs, representing 5 and 10 day fracture healing.

Results

Histological examination of the rib fracture at five days confirms the features originally described by Ham and Harris (1971). Longitudinal sections show two enlarged tissue masses, cross sections of two collars of proliferating osteogenic cells derived from the periosteum, on either side of the fracture. These two collars become increasingly thicker, grow towards each other, meet and fuse, (Fig. 1).

Osteoblasts, identified as large, rounded cells containing eccentric positioned nuclei and basophilic cytoplasm, can be identified within the osteogenic tissue, (Fig. 2). The same features, altered by the two-dimensional environment of a culture flask, can be seen in the actively migrating cells seeded after enzymatic dispersion of the excised, osteogenic tissue (Fig. 3). If allowed to continue to develop *in situ*, this osteogenic mass would form a recognizable center of calcified tissue as osteoblast function is completed, (Fig. 4).

After 14 to 21 days in culture, confluent sheets of cells arranged in adjacent circular patterns fill the flask, (Fig. 5).

After long-term culture in growth factor supplemented media, followed by treatment with Ca^{++} and Mg^{++} free saline in preparation for final cell dispersion, cells with a new morphology appear. Identified by their refractile cell body at low magnification, these cells show a multi-polar array of cytoplasmic projections, and resemble osteocytic elements in cortical bone, (Fig. 6,7,8).

Alkaline phosphatase production by the growth factor-treated cultures is enhanced significantly over controls grown in the absence of such stimulation. After clonal selection of cell lines that secreted the highest levels of enzymic activity and preliminary dose response tests, it was found that confluent cultures could be maintained for up to 90 days when either ILGF or PDGF were in the media, (Fig. 9). The mean alkaline phosphatase activity over this time period is presented in Table I.

Table I. Effect of Growth Factors on Alkaline Phosphatase Secretion

	AlP/protein*	Days of Culture	% Difference
Control**	11.72 ± 0.05	20	-
PDGF (8 u/ml)	23.70 ± 1.92	80	50.05
ILGF (80 ng/ml)	32.62 ± 2.97	90	64.07

* u/Hr/mg x 10^{-3}

** Each value represents mean of 4 replicate cultures. Control values from Nussbaum (1988).

Confluent sheets of osteogenic cells were grown on membrane supports, (Millicell HA inserts, Millipore Corp., Bedford, MA). Microscopic examination of these cultures revealed that the extensive secreted matrix introduced large spaces between adjacent cells. This precludes the use of such preparations for ion transport studies.

Fig. 1 Site of fracture at 5 days healing. Hematoma in marrow cavity (H) bounded by lateral and pleural (bottom) cortical bone (B). Collar of enlarged periosteal derived cells (P) adjacent to external clot. 100X, H & E.

Fig. 2 Portion of periosteal collar adjacent to fracture site. Represents the cell population harvested for culture. 300X, H & E.

Fig. 3 Isolated osteoblast-like cells derived from periosteal collar after enzymatic dispersion and 48 hrs *in vitro*. Note ruffled margins of actively migrating cell and flat, thin cytoplasm. 800X, Interference Contrast.

Fig. 4 Fracture site at 10 days healing. Calcifying callus (C) forming within periosteal collar (P). 300X, H & E.

Fig 1

Fig 2

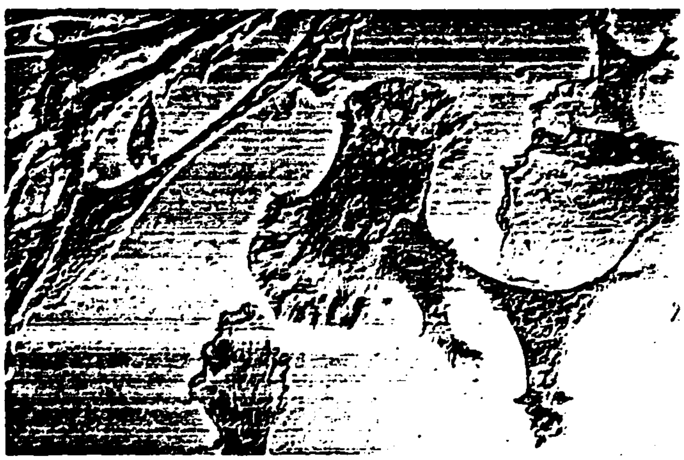


Fig 3

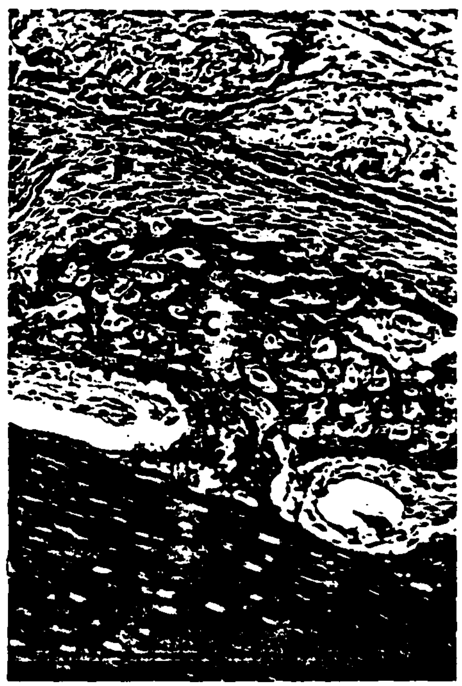


Fig 4

Fig. 5 Dense sheet of bipolar cells after 20 days in vitro. Typical field of confluent culture. Cells arranged in circular array about focal center. 80X, Phase Contrast.

Fig. 6-8 Examples of multipolar, refractile cells present after long-term culture. Compare with Figure 3. 800X, Phase Contrast.

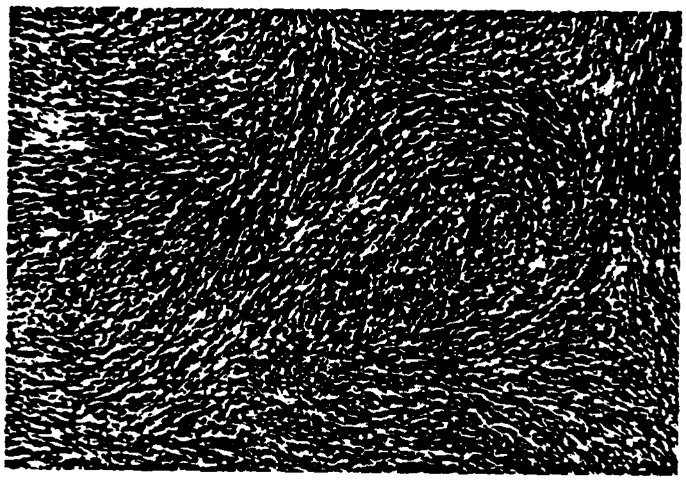


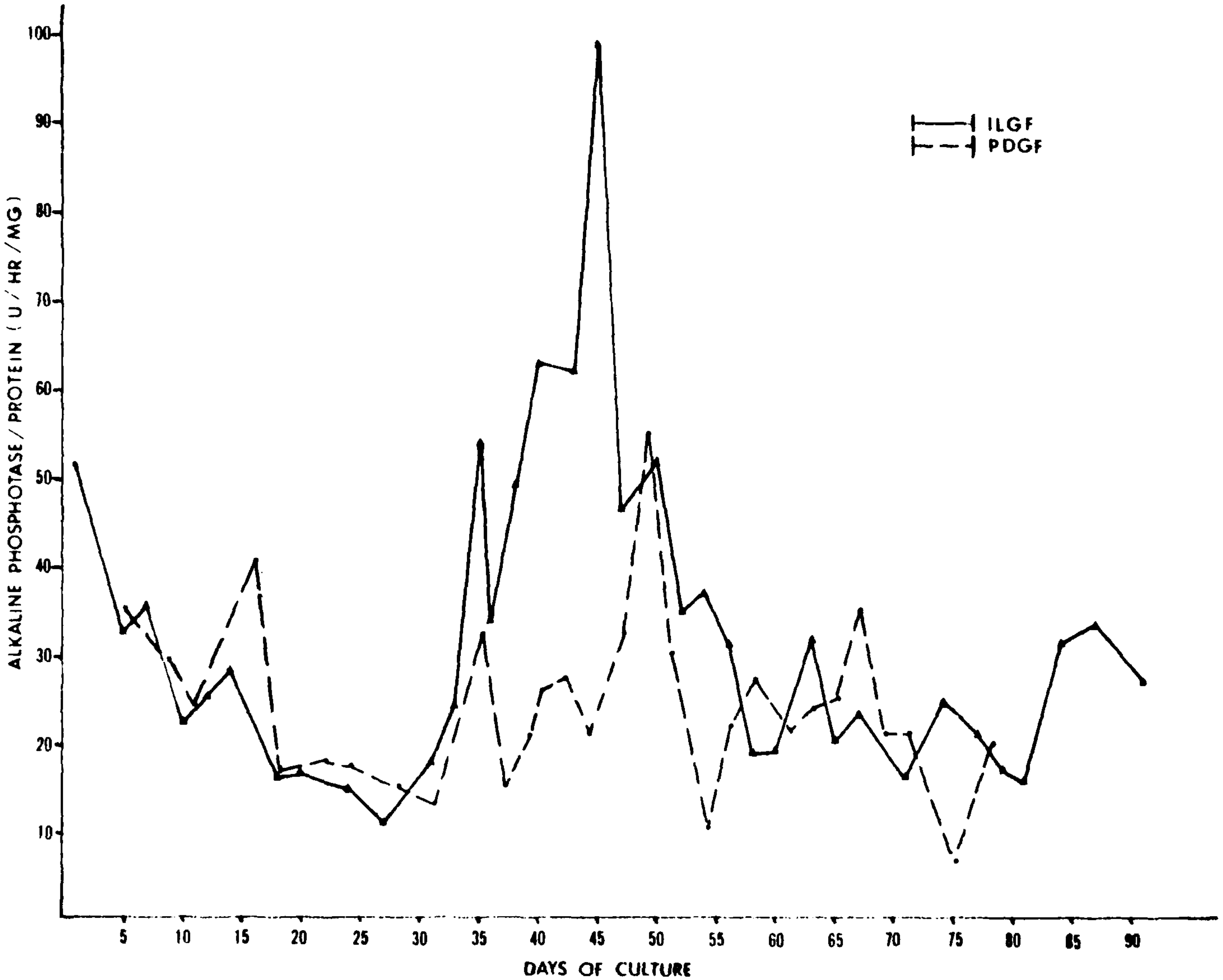
Fig 5

Fig 6

Fig 7

Fig 8

Fig. 9 ALKALINE PHOSPHOTASE PRODUCTION IN VITRO



Discussion

The histological description of rib fracture healing in the rabbit given above confirms in all details the process as presented by Ham and Harris (1971). In fact, Fig. 1 of this report corresponds to a great degree with Fig. 8 of their paper. Both represent the extent of the periosteal response after about 1 week of healing and demonstrates the enhanced growth of this tissue, contributing to formation of the soft callus. Characteristic osteoblastic structure and uniquely high alkaline phosphatase production identify the cultured cell populations as osteogenic (Farley and Baylink, 1986).

Two major innovations in the study of bone cell dynamics have recently been combined. First, the sequential enzymatic isolation of viable "osteoblastic-like" cells from fetal calvariae (Aubin et al., 1988; Cohn and Wong, 1979; Smith, 1973) and periostea (Bitz et al., 1981; Brighton et al., 1976; Fell, 1932; Ham, 1932; Nijweide, 1975; Nijweide and van der Plas, 1979; Poussa and Rilsila, 1979; Tenenbaum et al., 1986; and Uchida et al., 1988) has been used to establish transformed and non-transformed cells in vitro. These cell populations have been maintained for only about one week, while various morphological, biochemical and biophysical studies were performed.

Second, the availability of polypeptide growth factors, produced through recombinant DNA technology, has recently stimulated work to determine the influence of these "local factors" on bone cells. In particular, the work of Canalis and coworkers has been of significance (Canalis, 1980; 1985; Centrella and Canalis, 1985; 1987) in describing the mitogenic influence of Transforming Growth Factor beta on isolated osteoblast-like cells. Other workers have examined responses to PDGF (Hanks et al., 1986), ILGF (Ernst and Froesch, 1987), Epidermal growth factor (Hata et al., 1984), or Type-B transforming growth factor (Noda and Rodan, 1986). In most cases, these studies were also of brief duration or primarily concerned with determining the mitogenic potential of the various polypeptides. Only one recent study (Aubin *et al.*, 1988) has reported changes in hormone responsiveness after long term culture.

The heterogenous nature of the original isolates, whether from normal or malignant sources, has precluded repetitive examination of the same parameter under diverse environments. The procedures presented in this study demonstrate a methodology for the isolation of well-characterized, stable cell lines from adult, non-transformed tissues. Contrary to a mitogenic effect, these cells reveal a more appropriate physiological response to growth factor stimulation, e.g. sustained alkaline phosphatase activity. It is therefore proposed that two phenomena are involved in these results: 1) *in vivo* trauma and subsequent healing induces a local effect that initiates the mitogenic response, and 2) cell isolates of this activated osteogenic tissue will continue, when appropriately maintained *in vitro*, the differentiation process that began *in situ*.

These conclusions are supported by these observations: 1) periosteal-derived cells from non-fractured ribs fail to survive under culture conditions that support those from traumatized sites (Nussbaum, 1988). 2) In agreement with literature reports, there is a decrease in alkaline phosphatase production during the first 20 days of culture of osteogenic cells, but with continued maintenance, a restoration of activity occurs. 3) Survival in culture, for at least three months is, in part, dependent on the presence of ILFG or PDFG. 4) Under these conditions a portion of the osteogenic population will undergo norphological changes suggestive of osteocytic function.

REFERENCES

Aubin, J.E., I. Tertinegg, R. Ber, and Heersche, J.N.M. 1988. Consistent patterns of changing hormone responsiveness during continuous culture of cloned rat calvaria cells, J. Bone and Mineral Res. 3:333-339.

Baron, R., Vignery, A., Horowitz, M. 1983. Lymphocytes, Macrophages and the Regulation of Bone Remodeling. In: Bone and Mineral Research Vol. 2. Peck, W.A. (ed). Elsevier, pp. 175-242.

Binderman, I., O. Duksin, A. Harrell, E. Katzir, and Sachs, L. 1964. Formation of bone tissue in culture from isolated bone cells. J. Cell Biol. 61:427-439.

Bitz, D.M., J.E. Russell, D.J. Simmons, K. Loeffelman and Whiteside, L.A. 1981. Derivation, culture and osteogenic capacity of rabbit periosteal cells. Ortho. Trans. 5:327-328.

Brighton, C.T., J.L. Fox and Seltzer, D. 1976. In vitro growth of bone and cartilage from rat periosteum under various oxygen tensions (abstract), 22nd Annual ORS, January 28-30, p. 51.

Canalis, E. 1980. Effect of insulinlike growth factor I on DNA and protein synthesis in cultured rat calvaria. J. Clin. Invest. 66:709-719.

Canalis, E. 1985. Effect of growth factors on bone cell replication and differentiation. Clin. Orthop. 193:246-263.

Centrella, M. and Canalis, E. 1987. Isolation of EGF-dependent transforming growth factor (TGFμ-like) activity from culture medium conditioned by fetal rat calvariae. J. Bone and Mineral Res. 2:29-36.

Centrella, M. and Canalis, E. 1985. Local regulators of skeletal growth: a perspective. Endocrine Reviews 6:544-551.

Cohn, D.V. and Wong, G.L. 1979. Isolated bone cells. In: Skeletal Research. Vol. 1, Simmons, D.J. and Kunin, A.S. (eds). Academic Press, pp. 3-20.

Ernst, M. and Froesch, E.R. 1987. Osteoblastlike cells in a serum-free methylcellulose medium form colonies: Effects of insulin and insulin-like growth factor I. Calcif. Tissue Int. 40:27-34.

Farley, J.R. and Baylink, D.J. 1986. Skeletal alkaline phosphatase activity as a bone formation index *in vitro*. Metabolism 35:563-571.

Fell, H.B. 1932. The osteogenic capacity *in vitro* of periosteum and endosteum isolated from the limb skeleton of fowl embryos and young chicks. Anatomy LXVI (66):11-180.

Fitton-Jackson, S., and Smith, R.H. 1957. Studies on the biosynthesis of collagen. J. Biophys. Biochem. Cyto. 3:897-911.

Gaillard, P.J. 1934. Development changes in the composition of the body fluids in relation to growth and differentiation of tissue cultures. Protoplasma XXIII (23):145-174.

Goldhaber, P. 1966. Remodelling of bone in tissue culture. J. Dent. Res. 45:490-499.

Ham, A.W. 1932. The role of periosteum in the healing of fractures. Arch. Surg. 24:440-450.

Ham, A.W. and Harris, W.R. 1971. Repair and Transplantation of Bone. In: The Biochemistry and Physiology of Bone. Vol. 3. Bourne, G.H. (ed). Academic Press. pp. 337-399.

Hanks, C.T., J.S. Kim and Edwards, C.A. 1986. Growth control of cultured rat calvarium cells by platelet-derived growth factor. J. Oral Pathol. 15-476-483.

Hata, R., H. Hori, Y. Nagai, S. Tanaka, M. Kondo, M. Hiramatsu, N. Utsumi, and Kumegawa, M. 1984. Selective inhibition of type I collagen synthesis in osteoblastic cells by epidermal growth factor. Endocrinology 115:867-876.

Lucas, P.A., G.T. Syftestad and Caplan, A.I. 1986. Partial isolation and characterization of a chemotatic factor from adult bovine bone for mesenchymal cells. Bone 7:365-371.

Marvaso, V. and Bernard, G.W. 1977. Initial intramembraneous osteogenesis in vitro. Am. J. Anat. 149:453-468.

Merchant, D.J., R.H. Kahn, and Murphy, W.H. Jr. 1960. Handbook of Cell and Organ Culture. Burgess Publishing Co., Minneapolis, Minn.

Mundy, G.R. 1987. Bone resorption and turnover in health and disease. Bone 8:59-S16.

Nijweide, P.J. 1975. Embryonic chicken periosteum in tissue culture: osteoid formation and calcium uptake. Proc. Kon. Ned. Akad Wet. 78C:410-417.

Nijweide, P.J. and van der Plas, A. 1979. Regulation of calcium transport in isolated periosteal cells. Effects of hormones and metabolic inhibitors. Calcif. Tissue Int. 29:155-161.

Nijweide, P.J., A. van der Plas and Scherft, J.P. 1981. Biochemical and histological studies on various bone cell preparations. Calcif. Tissue Int. 33:529-540.

Noda, M. and Rodan, G.A. 1986. Type-B transforming growth factor inhibits proliferation and expression of alkaline phosphate in murine osteoblast-like cells. Biochem. Biophys. Res. Comm. 140:55-65.

Nussbaum, N.S. 1988. Isolation of osteoprogenitor cells from the trauma-activated periosteum. FASEB Journal 2:(5):7502.

Parfitt, A.M. 1984. The cellular basis of bone remodeling. The quantum concept re-examined in light of recent advances in the cell biology of bone. Calcif. Tissue Int. 36:537-545.

Parfitt, A.M. 1987. Bone and plasma calcium homeostasis. Bone 1:51-58.

Poussa, M. and Ritsila, V. 1979. The osteogenic capacity of free periosteal and osteoperiosteal grafts. A comparative study in growing rabbits. Acta. Orthop. Scand. 50:491-499.

Raisz, L.G., J.W. Dietrich and Canalis, E.M. 1976. Factors influencing bone formation in organ culture. Isr. J. Med. Sci. 12:108-114.

Rodan, S.B., G. Wesolowski, K. Thomas and Rodan, G.A. 1987. Growth stimulation of rat calvaria osteoblastic cells by acidic fibroblast growth factor. Endocrinology 121:1917-1923.

Rosen, D.M., S.A. Stempien, A.Y. Thompson, and Seyedin, S.M. 1988. Transforming growth factor-beta modulates the expression of osteoblast and chondroblast phenotypes in vitro. J. Cell. Physiol. 134:337-346.

Ross, R., Raines, E.W. and Bowen-Pope, D.F. 1986. The biology of platelet-derived growth factor. Cell 46:155-159.

Smith, D.M. 1973. Studies of the metabolism of separated bone cells I. Techniques of separation and identification. Calc. Tissue Res. 11:56-69.

Smith, P.K., R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, and Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. Analytical Biochemistry. 150:76-85.

Tenenbaum, H.C. and Heersche, J.N.M. 1982. Differentiation of osteoblasts and formation of mineralized bone in vitro. Calcif. Tissue Int. 34:76-79.

Tenenbaum, H.C. and Heersche, J.N.M. 1986. Differentiation of osteoid-producing cells _in vitro_: possible evidence for the requirement of a microenvironment. Calcif. Tissue Int. 38:262-267.

Tenenbaum, H.C., K.G. Palangio, D.P. Holmyard and Pritzker, K.P.H. 1986. An ultrastructural study of osteogenesis in chick periosteum _in vitro_. Bone 7:295-302.

Uchida, A., Kikuchi, T. and Shimomura, Y. 1988. Osteogenic capacity of cultured human periosteal cells. Acta. Orthop. Scand. 59:29-33.

Urist, M.R. and McLean, F.C. 1952. Osteogenic potency and new bone formation by induction in transplants to the anterior chamber of the eye. J. Bone Joint Surg. (Am.) 34:443-476.

Wong, G.L., R. Roberts and Miller, E. 1987. Production of and response to growth-stimulating activity in isolated bone cells. J. Bone and Mineral Res. 2:23-28.

Assessing the Attributes of Expert Judgment: Measuring Bias in Subjective Uncertainty Estimates

Dr. Thomas E. Nygren

The Ohio State University

Department of Psychology

Columbus, Ohio 43210

Universal Energy Systems, Inc. / Air Force Office of Scientific Research

Research Initiation Program Mini-Grant

Contract No: F49620-85-C-0013/SB5851-0360

Final Technical Report

December, 1988

Controlling Office: Air Force Office of Scientific Research

Bolling AFB, DC

ACKNOWLEDGEMENTS

I wish to thank the Air Force Office of Scientific Research and Universal Energy Systems, Inc. for sponsoring this research. In addition, I wish to thank the Harry G. Armstrong Aeromedical Research Laboratory and Gary B. Reid, in particular, for their support through the Summer Faculty Research Program and for this project, and for discussions that led to this productive research effort.

The views and conclusions in this document are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research of the United States Government.

ABSTRACT

It is argued that if our goal is to train individuals to become expert decision makers, regardless of the context of the judgment, it is of paramount importance to understand how individuals represent uncertainties as subjective probabilities. This paper reports the results of several studies that examined the role of the subjective probability function in models of human decision making. In addition, a computer-based methodology was introduced that has the potential of allowing for the measurement of subjective utility and probability functions at the individual decision maker level. Specifically, the studies and the methodology were used to begin building a model to describe how context, affect, and individual biases interact and influence judgments of subjective uncertainty. This report examines the studies in the framework of a dual probability function model where "good" and "bad" outcomes can differentially affect judged probabilities or probability weights.

1. INTRODUCTION: OBJECTIVES OF THE RESEARCH

The decisions being faced by trained Air Force pilots flying new and sophisticated aircraft are becoming more complex and demanding of many and varied cognitive processes. These individuals are being required to function as "experts" in their critical decision making judgments. However, the decision making literature is full of studies that show how people choose suboptimal alternatives in decision situations involving uncertainty. It is important, therefore, to understand the mechanism by which we may ultimately be able to efficiently train individuals to become expert decision makers and to exhibit the expertise that their work effort increasingly demands of them.

There are at least four major problems in research efforts on expert decision making. They are that (a) human experts are limited in their capacities to process all relevant information, (b) it is expensive to train individuals to obtain the level of expertise that is sometimes demanded, (c) experts sometimes disagree and give conflicting judgments, and (d) experts sometimes make mistakes because of inherent cognitive biases about the uncertainty or likelihood of outcomes associated with decision alternatives. Clearly, the study of experts and expert judgment must incorporate each of these issues if we are to establish what makes for a good expert.

The focus of this research has been associated with the last problem cited above -- cognitive biases. It is argued here that a critical dimension of expertise in any judgment situation is the ability to measure or evaluate uncertainties of events without bias or distortion. We make three assertions about these uncertainties. First, all uncertainties are inherently of the same kind. Second, probabilities are useful entities with which to numerically measure uncertainties of events. And, third, probabilities are personal beliefs about uncertain events.

The goal of the research to be described here was to begin to examine the questions of (a) "how do individuals relate uncertainties as measured by personal probabilities to observations of events?", (b) "how should we (or experts whom we must rely on) relate uncertainties to our observations of events so that biases are minimized?", and (c) "what biases exist in our judgments of uncertainties?" Project 1 consisted of three experiments that were designed to examine cognitive differences that exist in tasks of probability estimation and choice. These studies looked at subjective numerical responses to vague probability phrases in three different judgmental bias situations. The numerical interpretations for the probability phrases were expected to show differentially higher weighting when the outcomes were perceived as "bad" as opposed to "good".

Project 2 was designed to examine the influence of one's affective state on probability estimation and weighting. Specifically, the study was designed to show that the phenomenon of cautious optimism could be easily demonstrated and explained by positive affect induction.

Project 3 was designed to develop a modified methodology of the procedure used originally by Davidson, Suppes, and Siegel (1957) into a computer-based algorithm, DSS, (cf. Isen & Nygren, 1988) that could independently measure the utility and subjective probability functions (U and S) for an individual. Specifically, the procedure was designed to measure the subjective probability function separately as two functions, S^+, and S^-, in order to quantitatively assess the potential bias in judgments of the likelihoods of good (S^+) and bad (S^-) outcomes.

The results of each project point to the need for developing models that encompass both affect and context biases in judgments of uncertainties.

2. PROJECT 1: Estimating and Weighting Subjective Probabilities in Horse Race Gambling

Although subjective probabilities have long been integral components of virtually all mathematical models of decision making under uncertainty (Ramsey, 1931; Savage, 1954; Edwards, 1954, 1962; Handa, 1977; Einhorn & Hogarth, 1985; Hogarth & Einhorn, 1988), and have been studied and used extensively in a large variety of applied settings including such diverse areas as weather forecasting (Murphy & Winkler, 1974; Sanders, 1963) and medical decision making (Betaque & Gorry, 1971), the literature on cognitive aspects of subjective probability assessment has only recently expanded. Empirical studies including Kahneman and Tversky's (1972, 1973) studies on judgmental heuristics of representativeness and availability and related work on anchoring and adjusting strategies (Slovic, 1972; Tversky & Kahneman, 1982) have led to a greater emphasis on cognitive processes in decision making (cf. Arkes & Hammond, 1986; Kahneman, Slovic, & Tversky, 1982). The outgrowth of this research suggests that although subjective probability estimation *may* be a *necessary* initial component of some decision making under uncertainty, individual choice behavior suggests the utilization of a much more complicated process based on a monotonic function of these probabilities, one that Kahneman and Tversky (1979) have labeled a *decision weight* function.

Several models including extensions of the SEU models cited above reflect this distinction between subjective probability estimates and subjective decision weights in decision making under uncertainty

(Karmarkar, 1978; Kahneman & Tversky, 1979, 1981). In these models it is not so much the subjective probability of an event that affects the decision making process as it is the relative importance or weight that the decision maker attaches to this probability. Kahneman and Tversky (1979) explicitly state in their formalization of prospect theory that "decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree of belief (p.280)." Decision weights, then, assess the impact of the probabilities on the relevant outcomes, and they may reflect a number of interesting psychological influences including cognitive biases (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1973, 1974, 1981, 1983), ambiguity (Einhorn & Hogarth, 1985; Hogarth & Einhorn, 1988), dispositional tendencies and aspiration level (Lopes, 1987), or mood state (Isen, Nygren, & Ashby, 1988; Nygren & Isen, 1985).

An extension to a weighting model, however, begs the question of how an individual's decision weight function might be related to or derived from his or her initial subjective probability estimates over an event space. It is interesting that despite the increased attention that has been paid to the construct of decision weights and to Kahneman and Tversky's (1979) model in particular, no systematic effort has been made to empirically study in detail the relationship between these two subjective functions.

There are at least two reasons why this may be the case. First, it is clearly very difficult to measure either an individual's subjective probability or decision weight function. For the SEU model and its variants one needs first to have previously obtained or made simplifying assumptions about an individual's utility function (e.g., that it is linear over a restricted range) before the subjective probability or weight function can be derived. An extension of Davidson, Suppes, and Siegel's (1956) procedure based on the SEU model remains a major methodology for obtaining within an individual simultaneously inferred measures of subjective utility and subjective probability from preference judgments. Although the procedure has been greatly simplified via interactive computer algorithms, (e.g. Nygren & Morera, 1988) and can be used to obtain restricted (in the sense that $s(p) + s(1-p) = 1$) or unrestricted decision functions ($s(p) + s(1-p)$ does not necessarily add to one), it may not be possible to unambiguously separate the subjective probability and decision weight functions under this task. It seems reasonable to argue that in this Davidson et al. (1956) type choice paradigm, the probability scale that is derived from the individual's choices really reflects a subjective weighting of the probabilities that is influenced by any number of cognitive biases, rather than being a measure of the subjective probabilities themselves.

Secondly, much research relevant to the study of subjective probability is either concerned with individuals' *assessments* of probabilities for various events, or with inferences about the shape of their utility and subjective probability functions based on observed preference orderings. In the former case, the subjects are typically asked to give direct estimates of likelihoods of events, whereas in choice tasks using lotteries or applications of the SEU methodology, the probabilities are almost always explicitly stated numerically. In these latter choice tasks no real estimation on the part of the individual is required or assumed. As Kahneman and Tversky (1979) have pointed out,

> "It is important to distinguish overweighting, which refers to a property of decision weights, from the overestimation that is commonly found in the assessment of the probability of rare events. Note that the issue of overestimation does not arise in the present context, where the subject is assumed to adopt the stated value of *p*. In many real-life situations, overestimation and overweighting may both operate to increase the impact of rare events" (p. 281).

Hence, it is difficult to assess and no satisfactory attempt has yet been made to empirically show to what degree a decision maker's numerical assessment of the likelihood of an event (either obtained directly or inferred from her choices in an experimental setting) reflects a subjective probability estimate, a decision weight, or both.

Recently, however, Wallsten and Budescu and colleagues (Budescu, Weinberg, & Wallsten, 1988; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986) have examined in detail a promising alternative stimulus presentation mode to that of using explicitly stated numerical probability values. This stimulus mode entails using verbally expressed probabilities in displaying choice alternatives. They argue that these verbal expressions are meaningful, reliable, and valid representations of probabilities. I argue further that verbally stated probabilities have the potential to empirically differentiate between probability estimation and probability weighting. In several of their experiments Budescu et al. (1988) showed that despite large individual differences in the way people interpret phrases like "somewhat unlikely" or "very good chance", such judgments "yield reliable, internally consistent scales that demonstrated construct validity at the level of the individual subjects (p. 290)."

Given this demonstration of high reliability and consistency in subjects' assignments of numerical values to these verbally stated probabilities, it can be argued that verbally stated probabilities can be meaningfully used to independently examine probability estimation and decision weighting. Specifically, it is argued that when subjects are asked to provide a numerical interpretation to the various probability phrases in a context-free situation similar to what Budescu et al. (1988) have done, they are providing subjective *estimates* of the probabilities associated with the phrases. However, when individuals are presented with real choices under uncertainty that are based on these vague probability phrases and are asked to make decisions, they are no longer treating these probabilities as estimates in isolation, but are rather incorporating them into the decision making process as context-specific *weights.*

It is proposed here, as has been suggested elsewhere, that the initial subjective probability estimates are non-linearly, though monotonically, distorted to reflect the impact of the outcomes on the decision process. However, this weighting is more complex than has been suggested by Kahneman and Tversky (1979). It is context-specific in that "good" or positive outcomes are weighted differently from "bad" or negative outcomes in a manner similar to the good/bad distinction suggested by

Coombs and Lehner (1984) in the context of their model of risk. That is, it is suggested that a phrase like "very unlikely" will be given a different weight in a choice task when it is associated with a positive outcome as opposed to a negative outcome. Such a context-dependent, weighting mechanism would suggest an extension of the simple bilinear model like that used in prospect theory to a dual bilinear model with two rather than one subjective weighting functions. In fact, Luce and Narens (1986, 1988) have already proposed a formal mathematical model similar to this dual bilinear representation. It is important to note, however, that although mathematically similar, their weighting functions have a different psychological interpretation from what is being proposed here. Luce and Narens' (1986) model has two probability weighting functions, S^+ and S^-, that were proposed to reflect an ordering of the alternatives in the gambles rather than a good vs. bad or positive vs. negative context distinction.

The purpose of the following studies was to use verbal probability statements to study both probability estimation and probability weighting in the context of a dual bilinear model. The studies to be described here look at subjective numerical responses to vague probability phrases in three different situations. First, in Experiment 1 subjects were asked to simply make estimates of probabilities presented as phrases. In Experiment 2 the probability phrases were placed in a horse race gambling context, but were not associated with actual gambling. In Experiment 3 subjects actually played the horse race gambles in which the verbal phrases appeared. It was hypothesized that in the first two experiments the numerical judgments given by the subjects to the verbal phrases would be associated only with the estimation process and hence, would not exhibit any differentiation between winning and losing contexts. In the third experiment, however, where the phrases were an integral component of the subjects' actual horse race choices, decision weighting as well as estimation should be exhibited and the weighting function for phrases associated with losing should be significantly different from that for winning. Here, the numerical interpretations for the phrases were expected to show differentially higher weighting when the corresponding outcomes were perceived as "bad" as opposed to "good". Further, it was hypothesized that the degree of gambling success would affect this bias. The bias would be reduced for subjects who had experienced more wins or were "ahead" in their gambling and would be enhanced for those who had experienced more losses or were "behind."

EXPERIMENT 1

Method

Subjects. One hundred and seven undergraduates at The Ohio State University participated in this study as a partial fulfillment of the requirements for their introductory psychology course.

Procedure. Subjects were asked to evaluate the twenty different probability phrases presented in Table 2.1. They were asked to indicate what each phrase meant to them in terms of X chances out of 100 as in the following example.

It is ***somewhat unlikely*** that you will win $2.00.

How many chances out of 100 do you think this statement means in terms of the likelihood that you will win $2? _____ chances out of 100.

Table 2.1
Mean Single Stimulus Ratings of Probability Phrases
from Experiment 1

Probability Phrases	Mean: Win	Mean: Lose	Mean: All	Median: All	Stand Dev	Shanteau (1974)
No Chance	0.22	0.65	0.43	0.00	1.30	0
Extremely Unlikely	8.74	5.27	7.00	6.00	4.84	
Highly Unlikely	8.96	11.70	10.33	8.50	6.05	
Very Unlikely	16.08	13.67	14.88	14.50	7.10	
Unlikely	19.45	20.37	19.91	19.50	8.51	12
Fairly Unlikely	28.46	27.42	27.94	29.00	9.77	
Somewhat Unlikely	32.22	30.56	31.39	31.00	10.01	29
Worse Than Even Chance	39.01	38.57	38.79	40.00	7.96	38
Slightly Worse Than Even	40.90	42.72	41.81	43.50	8.92	
An Even Chance	49.97	49.98	49.98	50.00	0.28[a]	
Toss-Up	50.60	49.88	50.24	50.00	2.38	50
Somewhat Likely	54.80	56.83	55.82	59.50	13.79	
Slightly Better Than Even	56.19	57.78	56.98	55.00	5.02	
Better Than Even Chance	60.45	60.08	60.27	60.00	6.11	59
Fairly Likely	67.88	68.89	68.38	69.50	9.59	68
Likely	71.31	74.27	72.79	75.00	10.94	
Very Likely	84.05	84.44	84.24	85.00	7.74	
Highly Likely	86.33	85.28	85.80	87.50	7.26	
Extremely Likely	88.94	88.87	88.91	90.00	7.09	
Sure-Thing	98.56	97.93	98.24	99.00	3.14	100

$Mean_{Win} - Mean_{Lose} = -0.10$, $t(19) = -0.28$, $p > .20$

[a] The standard deviation for the phrase "An Even Chance" may be artificially small because it was used as a frame of reference in the explanation of the judgment task. Subjects were told as an example to think of it as representing an event like flipping a coin.

Each of the twenty phrases in Table 2.1 were presented in a random order twice, once as shown above and once where the phrase "that you will win $2" was replaced with the phrase "that you will lose $2".

Results

Table 2.1 also presents the summary data for the rated phrases in the study. The first two columns of data in Table 2.1 present the mean ratings given to each phrase when stated in this non-gambling context of simply winning $2.00 or losing $2.00. No significant differences were found between estimates obtained in these two contexts, indicating that merely presenting the phrases in terms of winning or losing does not appear to be sufficient to induce any bias in the subjects' estimates of the phrases (M_{diff} = -0.10, t_{19} = -0.28, $p > .25$). The ratings from both contexts were then combined to obtain an overall mean estimate for each phrase. Median values were also obtained and are presented in Table 2.1 in order to assess whether the degree of skewness in the distribution of estimates might be too severe to preclude the use of means. This does not appear to be the case for any of the phrases. Hence, the design of the last two studies was based on the mean estimates rather than medians and all analyses are reported in terms of means.

As a final comparison, mean judgments for common phrases used by Shanteau & Anderson (1971) are shown. These means are very similar for all shared phrases, supporting the argument that the judgments for these probability phrases are well-defined across subjects. Interestingly enough, even the standard deviations presented in Table 2.1 reflect, as Budescu et al. (1988) pointed out, that phrases near the typical anchor points (no chance, 0%; even chance, 50%; and sure-thing, 100%) are much less variable and thus reflect smaller individual differences than do phrases representing probabilities in the midrange between 0.0 and .5 and between .5 and 1.0.

The overall mean values in Table 2.1, then, represent stable aggregate *subjective estimates* of the probability phrases with no or minimal contextual bias. As such they serve as baseline values that can be compared with the probability *estimates* and *weights* obtained in the next two experiments.

EXPERIMENT 2

Method

Subjects. One hundred and eighty-three of 195 initial volunteer undergraduates at The Ohio State University completed participation in this study as a partial fulfillment of the requirements for their introductory psychology course. Subjects were randomly assigned to either an initial Winning condition (n = 60), a Losing condition (n = 60), or a Control condition (n = 63). They were separated and were individually seated in front of a terminal driven by a Data General Nova 3 computer. All stimulus materials and responses were presented and received on this terminal. Data from twelve subjects were stored incorrectly by the computer program driving the experiment and were inadvertently partially lost, resulting in the final count of 183 useable data sets.

Task 1: Betting. In this task subjects were instructed that they would be asked to bet on and make judgments about a set of hypothetical horse races. They were asked to imagine that they were at the racetrack, making $2 bets on various simplified races for which they could either win some amount of money, lose their $2 bet, or "break even", each with explicitly stated numerical probabilities. They would lose their $2 bet if their horse finished fourth through tenth, break even (i.e., get the $2.00 back) if the horse finished third, or win some stated amount of money if the horse finished second or first. All gambles, then, had a constant amount to lose of $2, but an amount to win that varied in such a way as to give the gambles an expected value of zero. A total of 20 of these horse races were presented in a different random order to each subject. The first three trials were the same for all subjects and for practice only; there was one additional filler item; and for the remaining 16 gambles the probabilities of winning and losing (P_+ and P_-) were factorially crossed with values of .1, .2, .3, and .4.

Before playing the gambles, subjects were allowed to actually manipulate the size of their bet. They were allowed to choose one of four options: (a) to not play the gamble, (b) to leave the dollar amounts as stated, (c) to multiply the dollar outcomes by 1.5 (i.e., make the bet a $3 bet), or (d) to double the outcomes of the bet (i.e., make it a $4 bet). If the subject chose not to bet on a particular gamble, the race was still played and its outcome was still shown to the subject.

At the start of each race following the betting, subjects entered a number from zero to nine inclusive to start the horse race. This response also served to determine their horse's final position. For example, if the probabilities for win, break even and lose in a race were 4/10, 2/10, and 4/10, respectively, then the computer algorithm would generate a random ordering of the numbers 0 to 9 and would divide them into three sets of four, two, and four values each. If the subject chose a number that was one of the four numbers in the first set generated by the computer, his or her horse was programmed to finish first or second and the subject would obtain the winning amount. Similarly, the horse was programmed to finish third if the subject's number matched one of the two break even values, or somewhere in the fourth through tenth positions if the subject's number matched one of the four losing numbers. As each race progressed, the subjects received continual feedback about the horse's position at the quarter, half, three-quarters, stretch, and finish marks of the race. This information, although programmed to be probabilistic in nature, was consistent with the ultimate outcome of the race. For example, no horse that was programmed to finish first or second was ever in less than fifth place at the stretch.

Group Manipulation. Participants were randomly assigned to one of three conditions. Subjects in the Control condition made estimates of vague probability phrases (Task 2 below) in a set of horse race gambles *before* doing this actual betting task, so as not to have this gambling experience affect their estimates of these phrases. Subjects in the Winning and Losing conditions, however, did the betting task first, followed by the probability estimation task. All subjects were told that the stated probabilities in the horse races were, in fact, the real probabilities associated with winning, losing, and breaking even in each race. It was explained to them that the horse race computer program used a "true and complex random number generator that was more precise, fair, and accurate than flipping coins or rolling dice so that when you see a stated probability of .2 to win, for example, that is the true probability."

For the Control subjects all of the gambles were set up to be fair with zero expected value (within rounding). For the *Winning* and *Losing* groups, however, the gambles in the betting task were manipulated in such a way as to insure probabilistically that these subjects would win or lose, respectively, more often than would be expected by chance. For subjects in the Winning condition the probability of winning was actually programmed to be .2 higher, and the probability of breaking even .2 lower than the numerical values shown to the subjects. For subjects in the Losing condition, the probability of losing was actually .2 higher, and the probability of breaking even .2 lower than the values seen by the subjects. All subjects were thoroughly debriefed following the experiment as to the nature of this manipulation. Subjects in the Losing condition especially were assured that their ill luck was in no way due to any suboptimal decision strategies on their part.

Task 2: Probability Estimation. In addition to betting, each subject was presented with 22 other zero-expected value three-outcome gambles similar to those described in the betting task except that the explicit probabilities were replaced with probability phrases. Subjects were told that these gambles represented a number of different horse races where the initial probabilities associated with the three outcomes were stated as verbal phrases rather than as precise numerical values. It was explained to subjects that at some racetracks initial estimates of a horse's odds of winning a race are stated in terms of these verbal phrases to make it easier for novice gambles to understand. Subjects were then told that we were interested in how they would interpret these horse race situations. They were told to do so by indicating what each of the phrases associated with winning, losing, and breaking even in the gamble would mean to him or her in terms of *X* chances out of 100. The subjects were told that since these were the only three possible outcomes, their numerical estimates were constrained to sum to 100. An example of this kind of horse race gamble is presented in Figure 2.1.

Sixteen gambles were constructed in this task so that each of four phrases (***extremely unlikely [EU], unlikely [UN], somewhat unlikely [SU],*** and ***worse than even chance [WTE]***) were presented with each other as the winning and losing probabilities in a 4x4 factorial design. These phrases were chosen because, based on the means obtained in Experiment 1, they approximately matched the probability values of .1 to .4 that were used in the betting task. Six additional gambles with other phrases were included as filler items to eliminate possible subject awareness of the factorial nature of the presentation of phrases in these 16 gambles.

The zero-outcome phrase for each gamble was also chosen on the basis of the results from Experiment 1 so that the sum of the means for the three phrases in each gamble was as close to 100 as possible, making each gamble a plausible combination for most subjects. The 16 constructed gambles and these sums are presented in Table 2.2. It is important to note that although the constructed gambles could not be produced to have sums that were exactly 100, control was maintained in that the twelve gambles with

Amount	Probability
Lose $ -2.00	Extremely Unlikely
Get $ 0.00	Better than Even Chance
Win $ +1.00	Somewhat Unlikely

What does each of these phrases mean to you?

"Extremely unlikely" that you will lose $-2.00
means _______ chances out of 100.

"Better than even chance" that you will get $0.00
means _______ chances out of 100.

"Somewhat unlikely" that you will win $1.00
means _______ chances out of 100.

Use a number between 0 and 100 in each case.

Your total must add to 100.

Fig. 2.1. An Example of a Three-Outcome Gamble in Experiments 2 and 3.

Table 2.2

Three-Outcome Gambles Used in Experiments 2 and 3 with Mean Summed Estimates of Total Gamble Probabilities

Win	Break Even	Lose	Est. Sum
* Extr Unlikely	Very Likely	Extr Unlikely	98.24
* Extr Unlikely	Likely	Unlikely	99.70
* Extr Unlikely	Better Than Even	Swhat Unlikely	98.66
* Extr Unlikely	Sl Better Than Even	Worse Than Even	102.77
Extr Unlikely	Worse Than Even	About Even	95.77
Mean		99.03	
* Unlikely	Likely	Extr Unlikely	99.70
* Unlikely	Better Than Even	Unlikely	100.09
* Unlikely	Sl Better Than Even	Swhat Unlikely	108.28
* Unlikely	Sl Worse Than Even	Worse Than Even	100.51
Unlikely	Swhat Unlikely	About Even	101.28
Mean		101.97	
* Swhat Unlikely	Better Than Even	Extr Unlikely	98.66
* Swhat Unlikely	Sl Better Than Even	Unlikely	108.28
* Swhat Unlikely	Worse Than Even	Swhat Unlikely	101.57
* Swhat Unlikely	Swhat Unlikely	Worse Than Even	101.57
Swhat Unlikely	Unlikely	About Even	101.28
Mean		102.27	
* Worse Than Even	Sl Better Than Even	Extr Unlikely	102.77
* Worse Than Even	Sl Worse Than Even	Unlikely	100.51
* Worse Than Even	Swhat Unlikely	Swhat Unlikely	101.57
* Worse Than Even	Unlikely	Worse Than Even	97.19
Worse Than Even	Extr Unlikely	About Even	95.77
Mean		99.62	
About Even	Worse Than Even	Extr Unlikely	95.77
About Even	Swhat Unlikely	Unlikely	101.28
About Even	Unlikely	Swhat Unlikely	101.28
About Even	Extr Unlikely	Worse Than Even	95.77
About Even	No Chance	About Even	100.39
Mean		98.90	

Note: Gambles marked with *'s were the 16 used in Study 2.

different P_{W} and P_{L} phrases formed six pairs of opposites where the winning and losing phrases were reversed.

Results and Discussion

Manipulation Check. Because the primary manipulation in this study concerned the effect of actually winning and losing on the likelihood judgments for the phrases, subjects in the losing and winning conditions were regrouped as necessary according to their actual performance. The probabilistic mechanism built into the computer algorithm to induce winning and losing was successful. Of the 60 subjects programmed to lose money, 49 actually did and 11 won money. Similarly, 52 of the 60 subjects who were programmed to win, did so and eight lost. Two new groups of actual winners and losers were then formed for the remaining analyses; these groups consisted of 63 winners (52 + 11) and 57 losers (49 + 8). The control group of 63 subjects, of course, remained intact. For the three groups the mean earnings after the last horse race in the betting task were +\$17.06 for winners, -\$12.85 for losers, and +\$10.22 for controls with standard deviations of \$16.00, \$7.94, and \$18.62, respectively. All of these means were significantly different from one another $p < .05$).

Gambling Data. These manipulated differences in performance produced only marginal differences in actual gambling behavior. The mean amounts of money bet per horse race were \$2.31, \$2.34, and \$2.55 for winners, losers, and controls, none of which were significantly different from one another ($p > .10$). The mean number of times bet (out of 17 trials) was also not significantly different for the three groups ($p > .10$; $M_{Winners} = 13.19$, $M_{Losers} = 13.53$, and $M_{Controls} = 13.65$, respectively). In addition, there were no differences in the variances of the amount bet or number of times bet among the three groups (all *p's* > .10). Although these differences did not reach significance, there is some evidence to suggest that had there been more gambling trials, significant group differences may have emerged. Correlations between the mean amount bet by each subject and his or her final earnings were -.22 for Losers, +.23 for Winners, and +.16 for Controls. A test of differences between the correlation coefficients for Losers and Winners was significant ($Z = 2.42$, $p < .02$); for Losers the more they were losing, the more they tended to increase there bet; while for Winners, the more they were winning, the more they tended to increase their bet. Finally, although the task was a hypothetical betting on horse races, the fact that all subject groups bet on the average significantly less than the largest possible amount each time ($4.00) and the maximum number of times (17), indicated that they were, in fact, making realistic discriminations.

Phrase Estimation Data. Because the gambles in Table 2.2 had sums that deviated slightly from 100, the means of the judgments across the four replications that were given by each subject to each of the four phrases (*EU, UN, SU,* and *WTE*) when they were associated with the winning and then with the losing outcomes were obtained separately. These eight mean values for each subject (4 phrases x 2 contexts) were then used in a 3x(4x2), Groups X (Phrases X Win/Lose Context) mixed ANOVA. The results indicated no overall significant differences among the three groups ($F_{2,180} < 1$). Averaged across the winning and losing contexts, the mean ratings given to the phrases by the Winning, Losing, and Control groups were 24.74, 24.97, and 24.66. Comparably, this means that the mean ratings given to the zero-outcome phrases (50.52, 50.06, and 50.68) were not significantly different for the three groups. The bias that exists in probability estimation appears to be observed only when the possible outcomes have a positive/negative or good/bad effect, a point that will be discussed in more detail later.[1]

Table 2.3

Mean Ratings Across Probability Phrases in Experiment 2 When Phrase Is Associated with Winning and with Losing

	Winning Group		Losing Group		Control Group		All Subjects	
Phrase	ProbW	ProbL	ProbW	ProbL	ProbW	ProbL	ProbW	ProbL
EU	11.07	11.31	11.04	13.68	9.64	11.21	10.57	12.02
UN	21.17	21.75	21.34	22.69	20.65	21.11	21.04	21.82
SU	29.82	29.93	29.85	30.40	30.86	30.96	30.19	30.43
WTE	35.80	37.10	35.00	35.78	35.77	37.07	35.54	36.68
Mean	24.46	25.02	24.31	25.39	24.23	25.09	24.33	25.24
Mean Diff	-0.56		-1.08		-0.86		-0.91	

EU = extremely unlikely, UN = unlikely, SU = somewhat unlikely, WTE = worse than even chance.

As shown in Table 2.3, the winning vs. losing context, however, did affect the judgments that subjects made ($F_{1,180} = 4.52$, $p < .035$, MSE = 67.78) with higher estimates given to the phrases when they were associated losing as compared with winning. Closer examination by gambles and across subjects indicated that this held true for all of the 16 matched pairs of phrases (e.g., EU in gambles [EU, L, UN] and [UN, L, EU]); there was a consistent bias to associate a higher judged likelihood to each phrase when it is associated with losing as opposed to winning, despite the fact that in four of the gambles the same phrase was used for P_W and P_L and in the remaining six pairs of matched gambles P_W and P_L were simply reversed. The mean difference score across the 16 matched gamble-phrases when averaged over subjects were -0.56 ($F_{1,15}$ = 6.15, $p < .05$), -1.33, ($F_{1,15}$ = 18.32, $p < .001$), and -0.86 ($F_{1,15}$ = 13.10, $p < .01$) respectively, for the Winners, Losers, and Controls. It was expected that this bias would be influenced by the previous gambling success with Winners showing less of a bias. This hypothesis was not supported in that the Context x Groups interaction was not significant ($F_{2,180} < 1$) indicating similar differences for all three groups. The Phrase X Context interaction was marginally significant ($F_{3,540}$ = 2.50, $p < .06$, MSE = 9.98) suggesting that the contextual effect of winning vs. losing was greatest across subjects for the more extreme probabilities (M_{EU} = -1.45, M_{UN} = -0.78, M_{SU} = -.24, and M_{WTB} = -1.14).

The results of Experiment 2 indicate that the judgments of the phrases made by the subjects were essentially probability estimates rather than decision weights. Despite the fact that some subjects won or lost a lot of money in the betting task, they did not exhibit any large differences in their judgments of the phrases. This suggests that the ratings task, although placed in the context of betting on horses, was actually perceived as being independent of the betting task, and hence reflected subjective judgments that were more like estimates than weights. It was somewhat surprising though that some consistent bias toward overestimating "bad" outcomes occurred in all three groups; the bias was even present for the winning subjects, even though they had just completed a task where they had successfully made a net average gain of $17.06. Experiment 3 was designed to make the judgments of the probability phrases a more integral part of the gambling process. As such, these judgments about the phrases were expected to be much more greatly influenced by subjects' winning and losing histories.

EXPERIMENT 3

It was hypothesized that had the estimation task been an integral part of the betting task in the horse races, judgments of the phrases would have been different in that they should have reflected individual *decision weight biases*. In this situation, actual losses and wins should have a much greater impact on the subjects' judgments. To test this hypothesis, a third experiment was done where the two tasks were completely interrelated. Subjects had to make gambling decisions on horse races where the probabilities were initially presented as vague phrases. In this situation both probability estimation and weighting should be necessary in such a gambling task. It was expected that this modification to the gambling task would produce two effects: (1) a different, though monotonic, *weighting* function of the probabilities than that found in Experiment 2, and (2) a large overweighting of the P_W and P_L phrases and consequently a large underweighting of the P[illegible] phrases relative to the baseline estimates found in Experiment 1 and to those given by subjects in Experiment 2. In addition, the differential bias between P_W and P_L judgments that was found in Experiment 2 with $P_W < P_L$ was expected to be enhanced for Losers and reduced or even reversed for winners.

Method

Subjects. Eighty-two undergraduates at The Ohio State University participated in this study as a partial fulfillment of the requirements for their introductory psychology course; complete data was obtained from 79 of them. They were randomly assigned to either a Winning (n = 40) or Losing condition (n = 39).

Procedure. The procedure in this study was similar to that of Experiment 2 except for the following modifications. A total of 35 races were randomly presented to each subject. This time, however, for both the gambling and phrase judgment tasks only phrases were used to indicate the likelihoods of the horse race outcomes. In the first ten trials only the betting task was done. These ten gambles were again used to set up the two conditions of winning and losing. In a manner comparable to that for Experiment 2, for each of the 35 horse races, the subjects were allowed to either skip the race, leave the bet as is ($2), double the bet ($4), or even triple the bet ($6). As in Experiment 2, the actual probabilities of winning and losing were adjusted upward or downward, respectively, by a value of .2 for the two sets of subjects in order to produce about an equal number of subjects actually winning or losing substantial amounts.

In the last 25 trials, subjects bet on the horse race and then subsequently gave their ratings to the three phrases, subject to the constraint that their numbers summed to 100. The final 25 gambles in this task were constructed factorially so that each of five phrases (***extremely unlikely, unlikely, somewhat unlikely, worse than even chance,*** and ***about an even chance***) were presented with each other as the winning and losing probabilities in a 5x5 design. This complete set of gambles is also presented in Table 2.2. No additional filler items were included, since the debriefing of subjects in Experiment 2 indicated that the random presentation of the stimuli and the use of a total of thirteen different phrases was sufficient to give the appearance of an unsystematic nature to the races.

Results and Discussion

Manipulation Check. Initial analysis of the gambling data indicated that again the manipulation of the probabilities had been successful, with 41 subjects being actual winners and 38 being losers (M_{Win} = \$26.07, s_{Win} = \$23.41; M_{Lose} = -\$23.17, s_{Lose} = \$15.48; M_{Diff} = \$49.24, $F_{1,77}$ = 119.46, $p <$ .001). In addition, the average "holdings" (i.e., mean winnings or losses) on any given trial during the final 25 trials when the probability phrase judgments were being given was \$16.74 for winning subjects and was -\$15.61 for losing subjects.

Betting Task. The subjects' betting behavior was examined two ways, first for each subject averaged over trials and second for each trial averaged over subjects. Data were found to replicate that of Experiment 2. There were no differences between the Winning and Losing groups in the mean amount bet (\$2.99 vs. \$3.07, $p >$ .10) or the mean number of times bet across the 25 trials (20.41 vs. 19.03, $p >$ .10). Both groups did, however, show a clear preference for *low-risk* over *high-risk* gambles. Of the 25 gambles, ten had win/lose phrase combinations such that the ratio of P_W / P_L was greater than one (labeled low-risk) and ten had this ratio being less than one (labeled high-risk). However, since the gambles were set up to have approximately zero expected value, the ratios of the payoffs were, of course, in the opposite direction.

Both groups of subjects exhibited a willingness to bet more money and with greater frequency for low-risk gambles. For winning subjects the mean amount bet on low and high risk gambles were \$2.55 and \$3.49 (M_{Diff} = \$0.94, $F_{1,40} = 18.92$, $p < .001$); for losers these values were \$2.65 and \$3.64 (M_{Diff} = \$0.99, $F_{1,37} = 14.98$, $p < .001$). Thus, although no strong differences were observed in actual betting behavior, subjects were clearly indicating preferences among the horse races on the basis of probabilities rather than amounts.

Estimation Task. The mean of the judgments across the five replications that were given by subjects to each of the five phrases (*EU*, *UN*, *SU*, *WTE*, and *AEC*) when they were associated with the winning and losing outcomes was obtained as in Experiment 2 yielding ten probability scores for each subject. These values were then used in a 2x(5x2), Groups X (Phrases X Win/Lose Context) mixed ANOVA. The ANOVA indicated no overall significant difference between the two groups ($F_{1,77} = 2.89$, $p > .05$, $MSE = 31.82$). Across the 25 gambles, the mean ratings given to the phrases when collapsed across context were 31.86, and 31.13 for the Winning and Losing groups, respectively. Note that again this is equivalent to there being no significant difference between the mean judgments for the zero-outcome phrases for the two groups (36.28 and 37.74).

As in Experiment 2 the win/lose context effect was significant ($F_{1,77}$ = 4.55, $p < .037$, MSE = 66.41). A summary of the ten means across phrases for both groups is presented in Table 2.4. Again, across all phrases and for both groups there is the tendency to give a higher value to a phrase when it is associated with losing than for winning.

Table 2.4

Mean Ratings Across Probability Phrases in Experiment 3 When Phrases Are Associated with Winning and with Losing

	Winning Group		Losing Group		All Subjects	
Phrase	ProbW	ProbL	ProbW	ProbL	ProbW	ProbL
EU	19.94	20.26	16.19	17.93	18.14	19.14
UN	24.03	25.41	23.94	25.00	23.99	25.21
SU	29.72	30.49	28.96	32.07	29.35	31.25
WTE	36.06	37.43	35.13	36.99	35.61	37.22
AE	47.25	47.57	47.34	47.79	47.29	47.67
Mean	31.40	32.23	30.31	31.96	30.88	32.10
Mean Diff	-0.83		-1.65		-1.22	

EU = extremely unlikely, UN = unlikely, SU = somewhat unlikely, WTE = worse than even chance, AE = about an even chance.

The Context x Group interaction was not significant ($F_{1,77} < 1$); However, overall difference scores when summarized in a different way over the 25 gambles and within subjects indicated more stable results. For each subject a mean difference score was computed across all 25 gambles between the rating given to a phrase when it was associated with winning and the matched rating when the phrase was associated with losing in the reverse gamble. The means of these values across subjects are, of course, the same -0.83 and -1.65 values shown in Table 2.4 for the Winning and Losing groups. When examined this way, the mean of -0.83 for the Winning group is not significantly different from zero ($F_{1,40} < 1$). The difference of -1.65 is significant for the Losing group, however ($F_{1,37} = 8.04$, $p < .01$).

Comparison of Estimates from Experiments 2 and 3.

It was somewhat surprising that the win/lose probability bias, although again clearly significant in Experiment 3, did not appear to be any stronger than the effect found in Experiment 2. Instead, the more striking finding was the difference in the magnitudes of the estimates of the phrases between the two experiments. Figures 2.2, 2.3, and 2.4 illustrate these differences. Mean estimates in Experiment 3 are much larger than those in Experiment 2 for *both* P_+ and P_- phrases. But this overweighting only occurred for the more extreme probabilities associated with *extremely unlikely* and *unlikely*. Because no overall group differences were found in Experiment 2, Figure 2.2 shows the plot of the mean ratings given to the winning and losing phrases averaged over all three groups when compared with the baseline values of Experiment 1. Figures 2.3 and 2.4 show comparable plots for the Winning and Losing groups separately in Experiment 3. Consistent with what has often been suggested in the decision making literature, the curves in all three figures reflect higher values relative to the "context-free" or baseline values for small probabilities and lower values for more moderate probabilities. However, the curves in Figures 2.2-2.4 differ somewhat from the typical probability functions associated with SEU models or the function proposed by Kahneman and Tversky (1979) in their prospect theory. Figure 2.2 has a probability crossing point, PCP (PCP is the point on the curve where there is neither over- nor underestimation), of about .26 for winning phrases and .27 for losing phrases. These values are much lower than what has been typically suggested in SEU models -- a PCP of about .50. The decision weight functions in Figures 2.3 and 2.4 clearly match the shape of that proposed by Kahneman and Tversky (1979, Fig. 4, p. 283), but Figures 2.3 and 2.4 have much higher PCP values, about .30 instead of .10 as suggested by the Kahneman and Tversky figure. In other words, the Kahneman and Tversky (1979) figure suggests that only very small probabilities get overweighted; our results suggest that much larger probabilities, even levels up to values around .30 are overweighted.

GENERAL DISCUSSION

The studies presented here were designed to examine differences between the processes of subjective probability estimation and decision weighting and how they are related to the functions of probabilities that are integral components of bilinear generalizations of SEU models like those proposed by Kahneman and Tversky (1979) and by Luce and Narens (1986). The argument has been made here that the ratings given to the phrases in Experiments 1 and 2 were basically subjective probability estimates and that those given in Experiment 3 were decision weights formed in the context of gambling decisions. It was expected that the judgments

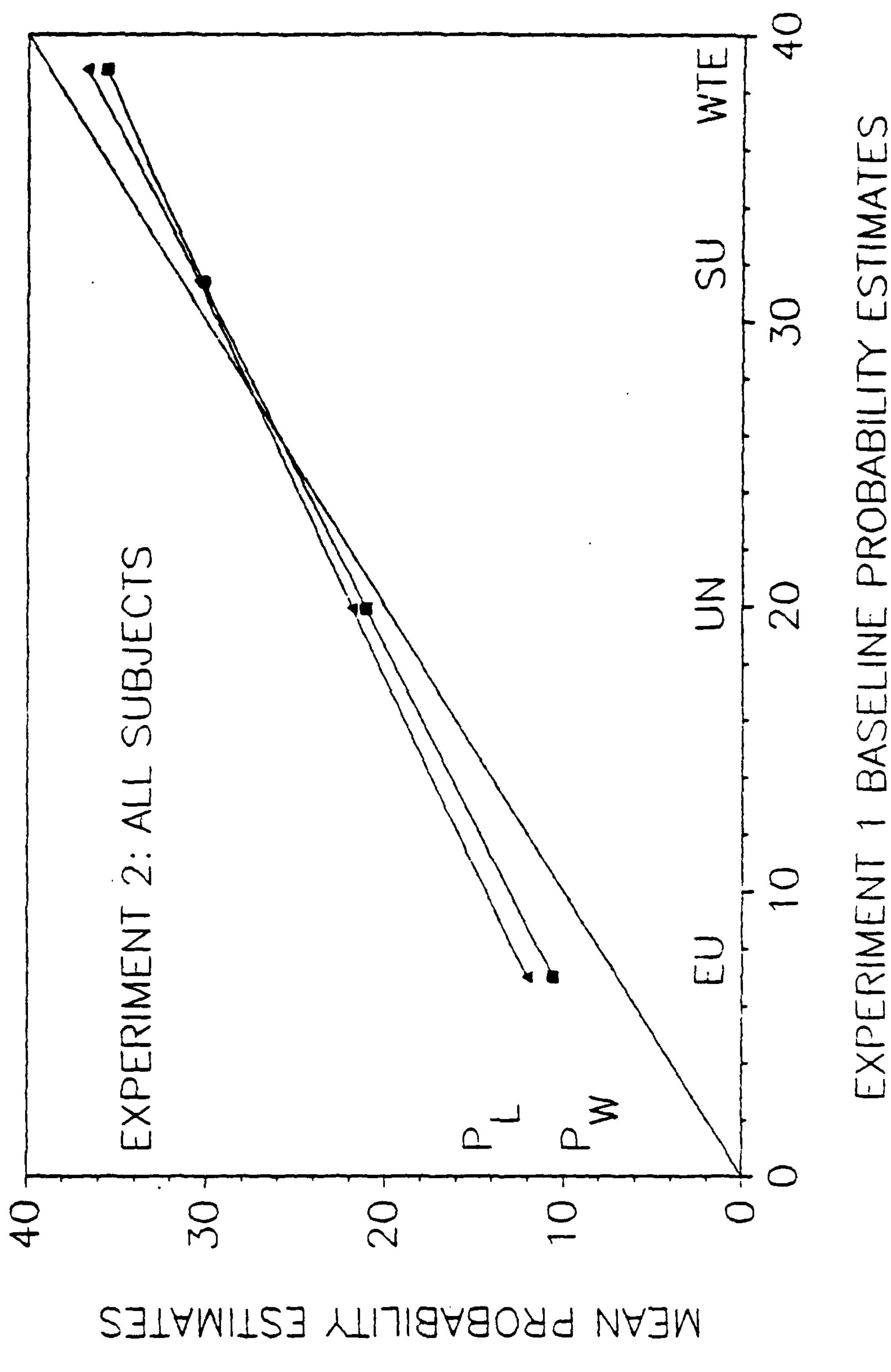


Fig. 2.2. Subjective probability functions for winning and losing vague probability phrases averaged across all subjects in Experiment 2.

Fig. 2. 3. Subjective weighting functions for winning and losing vague probability phrases for Winning Group subjects in Experiment 3.

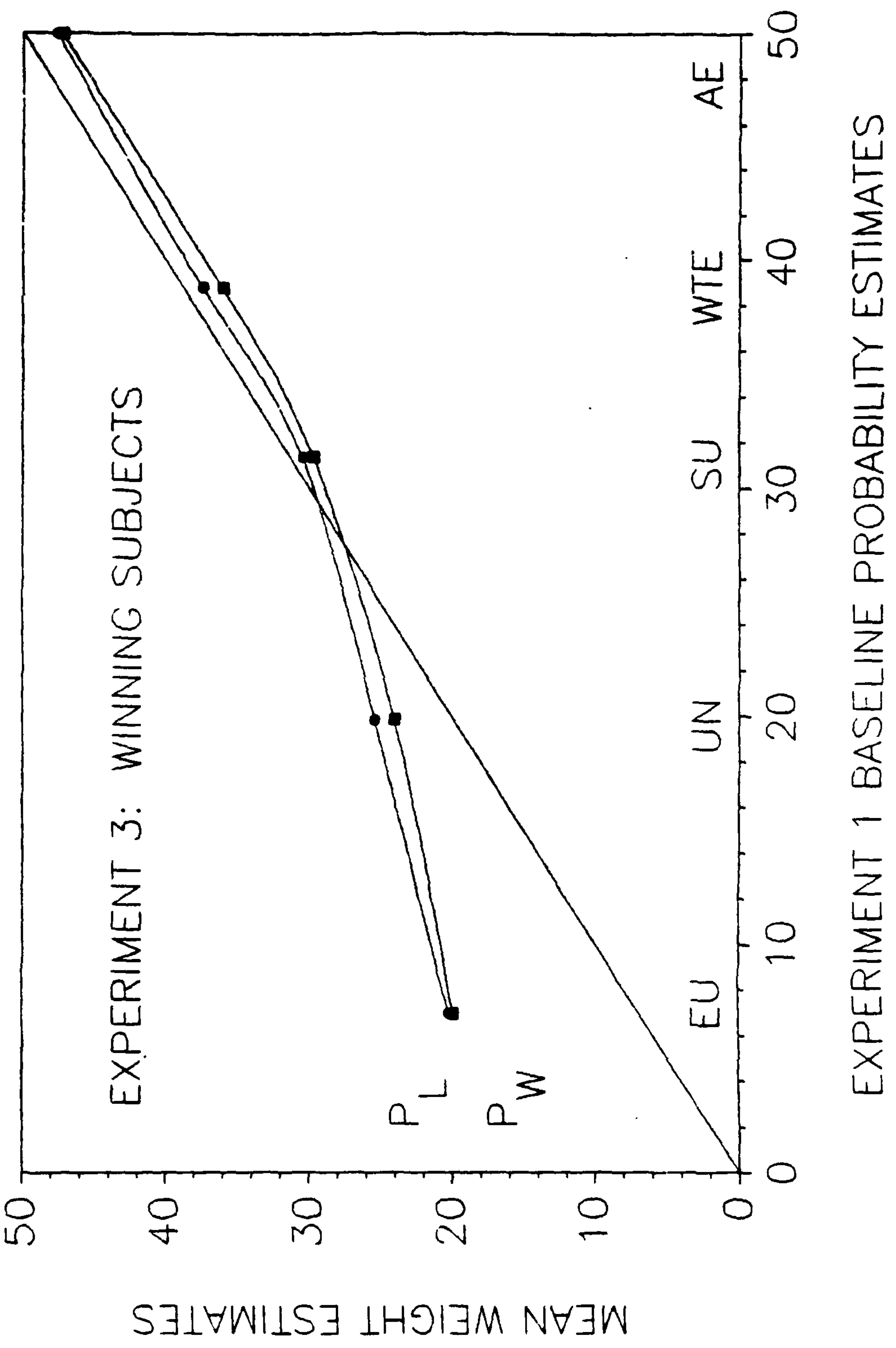
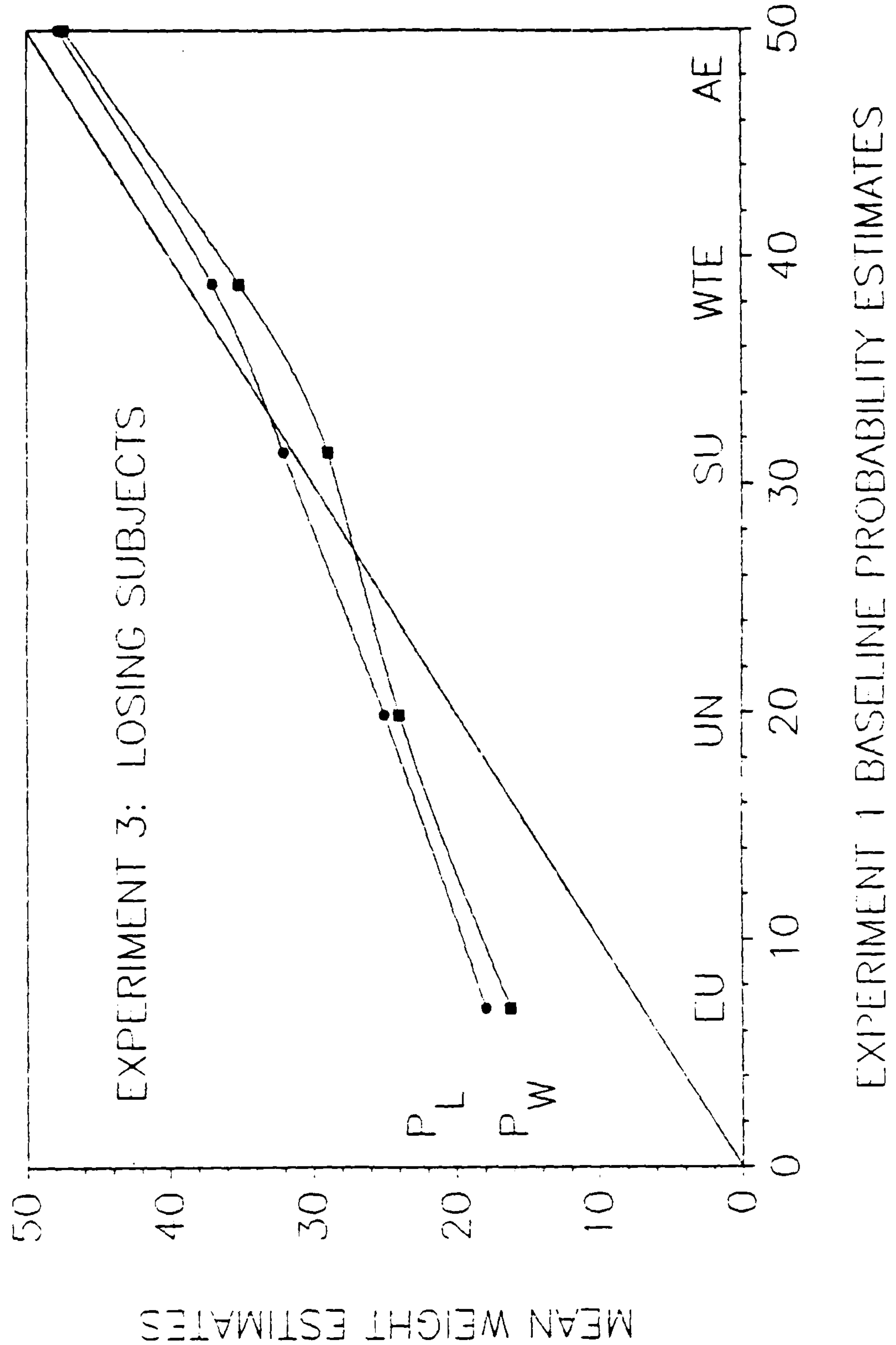


Fig. 2.4. Subjective weighting functions for winning and losing vague probability phrases for Losing Group subjects in Experiment 3.



across subjects in Experiment 2 would match the baseline values obtained in Experiment 1 with no win/lose bias effect or equivalently that $P_W = P_L$. This expectation was based on the assumption that since the phrase judgment task was completely independent of the actual gambling task in Experiment 2, it should have exhibited no affective component. In Experiment 3 where the evaluation of the phrases was a necessary first step in the gambling task, the $P_W < P_L$ ordering was expected, since decision weighting is also necessary in the evaluation process. Instead, subjects in *both* Experiments 2 and 3 demonstrated this differential biasing effect.

The $P_W < P_L$ ordering found in Experiment 2 suggests that the affective nature of the good and bad outcome comparison appears to be sufficient to produce a differential biasing in the *estimation* of probabilities when, due to the ill-defined nature or the vagueness of the events being evaluated, estimation is first required. In Experiment 1 the phrases were presented separately with no direct good/bad or positive/negative comparison; and hence no biasing was found. In Experiment 2, even though the phrases were presented in a context where the gamble was not being played, a good/bad comparison was still explicit. It is perhaps this comparison of good and bad outcomes that is necessary for the biasing effect to occur. Further, this biasing effect may be explained by an anchoring and adjustment process that is largely due to differences in the initial anchors rather than to differences in the adjustment process. Figure 2.2 provides some support for this idea; the curves for P_W and P_L remain fairly parallel over the range of phrases used. This would suggest an initial difference in anchor points for the losing and winning phrases, but a similar adjustment process for both contexts.

The results from Experiment 3 provide further support for this argument. Despite the fact that in this case both estimation and weighting were necessary, the magnitude of the differential effect between the P_w and P_l curves was not greatly enhanced. Although the size of the effect, as predicted, was clearly smaller for winning subjects than for losing subjects, it is surprising that the winning subjects still showed a slight negative bias with $P_w < P_l$. Neither having to actually play the gambles nor experiencing a major shift in the reference point on any given gamble (average status quo values of $16.74 and -$15.61 for winners and losers respectively) appeared to greatly affect the *magnitude* of the P_w vs. P_l bias. Hence, it is suggested that this difference is largely a function of an initial anchoring mechanism in the probability estimation process.

Probability weighting as depicted in Figures 2.3 and 2.4 appears to have its greatest subjective effect for small probabilities. The functions for both groups fit very nicely with Kahneman and Tversky's (1979) assertion in prospect theory that small probabilities tend to be overweighted and moderate and larger probabilities tend to be underweighted, but the underweighting is less so. It is important to note that in both Experiments 2 and 3 the sum of the estimates that subjects gave to each gamble was constrained to be 100. This in no way constrains a relationship between the P_w and P_l values in the three-outcome gambles, but does allow for an upper bound to be placed on the function. What is clear then from Figures 2.3 and 2.4 is that the probabilities associated with *either* outcome that could cause a change in the subjects' current assets were, in the framework of a bilinear model, given much more weight for small values at the expense of the zero outcome that could not change the status quo. It is also important to note, however, and would be the source of future studies, that because the gambles used in Experiment 3 were designed to have essentially zero expected value, these same small probabilities were tied directly to the largest outcomes, both good and bad. This suggests that overweighting may interact with outcome magnitude. One might argue that this is to be expected in most people when they are considering losses because they tend to be largely loss-averse, but the relatively equal effect for gains seems somewhat more puzzling.

The size of the $P_+ < P_-$ differential effect found in Experiments 2 and 3, however, does not appear to be affected by the magnitudes of the associated outcomes. The effect is fairly constant across phrases and hence across sizes of the good and bad outcomes, again because the gambles were zero expected value. It was expected that the differential effect would be stronger at the smaller probability levels. Since it was not, an important question that arises is whether the bias representing a difference of about .02 across the phrases used here, although reliably statistically significant, is practically significant. I argue that the answer is yes for several reasons. First, for a phrase like *extremely unlikely* which had a context-free mean rating of about .07, a difference up or down of .02 because of an anchoring and adjustment bias in some probability estimation application represents almost a 30% distortion, a change that could easily force preference reversals in later decision tasks. Secondly, and related to this issue is that phrases for very extreme probabilities where differential biasing could have its most dramatic effect on preferences were not used here. It would have been valuable to use phrases for events that had probabilities like 1/10,000 or 1/1000 or 1/500, since events with these probabilities are in many ways much more interesting because they are associated with the kinds of important choices that individuals often have to make in medical decisions, investments, insurance options, lotteries, etc. Further, overestimation or overweighting of these probabilities by an absolute magnitude even as little as .01 (e.g., from 1/10,000 to 1/100) may really be equivalent to a change by a factor of 100 in the choice process. However, there does not appear to be any empirical evidence that unique verbal descriptions and orderings on them for these extreme events are well-defined for most individuals. How then may these psychologically interesting very rare events be meaningfully ordered and represented by an individual? It is suggested that a phrase like "extremely unlikely" or "almost impossible" may, if placed in extreme situations like million dollar lotteries, nuclear disasters, or airplane accidents exhibit a relatively more dramatic good/bad differential effect than was obtained in the context of the less extreme horse race studies used here.

Nevertheless, the differential biasing of individuals' P_W and P_L subjective weighting functions found here and that we have demonstrated in other decision making frameworks as well (Isen et al., 1988; Nygren & Isen, 1985; Nygren & Morera, 1988) raises problems for simple bilinear models like SEU and prospect theory. For example, Nygren and Isen (1985) found that a positive affective state may differentially influence winning and losing probability estimates for the same event in decision making under uncertainty. In that study it was found that induced positive affect produced "cautious optimism"; that is, it increased the likelihood of subjects being more risk-averse for moderate to high risk gambles (caution), although at the same time they reported subjective judgments of the likelihood of winning and losing that were actually increased and decreased, respectively (optimism) relative to control subjects.

Since the simple bilinear models are bisymmetric representations, they do not allow for any kind of differential weighting of an event's probability for good and bad outcomes. That is, they do not allow for the possibility that a decision maker might evaluate or weight a probability differently by using two different functions, depending on the perceived goodness or badness of the outcome with which it is associated. Luce and Narens (1985) have presented a formal mathematical alternative representation for utility that would allow for such differential weighting. Their model encompasses the model underlying prospect theory, although it accounts for problematic results found here by allowing for two subjective probability or weight functions rather than imposing constraints on the shape of the utility function. They refer to their model as the dual bilinear utility representation.

In the Luce and Narens (1985, 1988) model there is a weighting function S^+ that would map, for example, the phrase *extremely unlikely* into one real value in the interval (0,1) if its corresponding outcome x is greater than that for outcome y and a separate function S^- that would map the phrase into a different value if the phrase were associated with an outcome x where $x < y$. This is mathematically very similar to, though psychologically different from, what is being proposed here. Namely, it is suggested that S^+ and S^- are associated with outcomes, not on the basis of their relative values (i.e., $x > y$), but rather on whether they are perceived to be good or bad. That is, here the psychological probability weighting function for an event A, S(A), takes on the form

$$S(A) = S^+(A) \text{ if } x > g \text{ and}$$

$$S(A) = S^-(A) \text{ if } x < g, \tag{2.1}$$

where g is some cutoff value, usually zero.

Thus, in Luce and Naren's (1985) dual bilinear model, both functions could be invoked in *any* nontrivial gamble. In the proposed dual bilinear model, the two separate probability functions would be used only when the gamble or prospect involves what the decision maker perceives to be both potentially good and bad outcomes. Hence, any nonregular prospect or gamble for which $x > y > 0$ or $x < y < 0$ should exhibit characteristics that are explained by a simple bilinear model, such as proposed in prospect theory. The present studies were not designed to compare the dual bilinear representations of Luce and Narens (1986) and that presented in Eq. 2.1. Rather, the point to be made is that several rather straightforward extensions of the simple bilinear model may account for the influence of affect variables in decision making.

In sum, evidence that the affective nature of the outcomes in a decision task may significantly bias the way in which probabilities are initially estimated or ultimately weighted was found. Further, subjective probability functions and decision weight functions were found to be different in shape and affected differently by context. The affective nature of outcomes appeared to be more influential in the initial probability estimation process, while the relative magnitude of the estimates were more influential in the over/underweighting process. These results support the argument for the necessity of encompassing other psychological factors into our decision making models. As Lopes (1987) has convincingly pointed out, "psychologists who study risky choice *don't* talk about a surprisingly large number of factors that are psychologically relevant to choosing among risks... fear, hope, safety, danger, fun, plan, conflict, time, duty, custom" (p. 286). The studies presented here and continuing work are attempts to move in this direction.

Footnote

Because three-outcome gambles have two degrees of freedom in the probability assessment, looking at the probability of winning and probability of losing judgments in the same analysis is entirely permissible. However, an overall nonsignificant main effect for groups in the ANOVA is equivalent to no group differences in the means for the probability associated with the zero-outcome because $P_{[illegible]}$ was constrained in Experiments 2 and 3 to be equal to 1 - ($P_{[illegible]}$ +$P_{[illegible]}$).

3. PROJECT 2. The Influence of Positive Affect on Probability Estimates and Gambling Behavior: A Case for Cautious Optimism

The phrase *cautious optimism* is commonplace in our vocabulary. It is used to describe a state of expectation of increased likelihood that a "good" or positive outcome will occur (optimism), yet the belief is tempered by an increased sensitivity to or awareness of placing oneself at risk of loss in choosing among courses of action (caution). Such a definition suggests two important points that are relevant to current models in judgment and decision research. First, from the perspective of a rational decision making model like subjectively expected utility theory (Savage, 1954), the term cautious optimism seems somewhat contradictory. If a person has reason to exhibit optimism about an increase in the likelihood of positive outcomes relative to negative ones, then we should expect to see a propensity for her to be more likely to take risks (from an expected value perspective) rather than to be cautious in her choices.

Secondly, it can be argued that optimism is associated with being in a positive affective state. Results from several recent studies indicate that persons in whom positive affect has been induced tend to be conservative or self-protective in situations where there is a focus on meaningful loss or where loss is highly likely (Isen & Patrick, 1983; Arkes, Herren, & Isen, 1988). These studies suggest that one possible mediator of the observed differences between positive-affect and control subjects is that a change in the perceived negative value or utility of losses takes place, such that the anticipated impact associated with any given loss is greater for a person in a positive-affect condition than for someone in a more neutral state. This idea is compatible with work in the social psychology literature suggesting that people who are feeling happy become motivated to maintain their positive states and thus may have more to lose than controls in the same situation (e.g., Isen & Simmonds, 1978; Mischel, 1973; Mischel, Ebbesen, & Zeiss, 1976).

Isen, Nygren, and Ashby (1988) also found that in a risky decision situation where real loss was possible (subjects were gambling with their credit hour for participating in the experiment), positive affect, in comparison with a control state, was associated with an increased tendency to avoid losses. In that study a procedure based on a modification of Davidson, Suppes, and Siegel (1956), was used which allowed Isen et al. to specifically isolate potential effects of positive affect on the subjective utilities associated with winning and losing outcomes. Both control and affect subjects (affect was induced by presenting subjects with a small gift) were asked to make choices between pairs of simple 50-50 gambles in such a way that an indifference point could be found between one fixed two-outcome and one variable two-outcome gamble in each of eight pairs. In this way individual utility functions were constructed for subjects in both groups. The average utility curves were computed for the two groups and were found to diverge significantly in the losses end where the affect subjects showed a steeper curve indicating a greater degree of loss-aversion. However, the curves were not significantly different in the gains end. These divergent utility functions indicated that a positive affective state accentuates an aversion to choosing riskier options as possible loss increases, by altering the perceived disutility associated with losses or negative outcomes, but appears to have little effect on positive utilities.

However, it is important to note that only utilities were measured in this study and probabilities were not allowed to change. In fact, the probabilities could *not* be changed in order that the utility estimation procedure could be done. It is quite possible, then, that affect may influence probabilities associated with the positive and negative outcomes as well as the utilities for losses.

This hypothesis has been indirectly supported in a recent study by Arkes et al. (1988). In one experiment where a meaningful loss was non-existent, affect subjects exhibited relatively more risk-prone behavior when compared to controls, even when risk was high, and this difference was accentuated as the probability of winning increased. In a second experiment dealing with insurance buying behavior where subjects were forced to focus on potential loss, positive affect subjects exhibited a greater risk-aversion tendency than did controls, and this difference was also magnified as the probability of the occurrence of the loss increased.

How might both risk-seeking and risk-averse behavior within the same individual, and cautious optimism in general, be described in the framework of a single cognitive processing model of affect and decision making? Kahneman and Tversky (1979, 1981) have proposed constructs in a model underlying their prospect theory that may partially account for some of these effects. In the model, Kahneman and Tversky (1979, 1981) suggest that alternatives or prospects are actually evaluated in a manner similar to that suggested by SEU theory where the alternative with the highest value (utility) is chosen. The critical difference is that whereas in SEU theory the utility of each outcome is multiplied by a *subjective estimate of its probability* and these expected utilities are summed across outcomes, in prospect theory the values are multiplied by *decision weights* which are not probabilities, but are monotonic with them. Unlike probabilities, the sum of the decision weights for complementary events with probabilities *p* and *1-p* are not bounded and will not necessarily add to one but will generally be less than one.

It is this decision weight function that is important to explaining the influence of positive affect on decision making. We suggest that positive affect results in cautious optimism because probability estimation and outcome weighting are independent processes that are uniquely influenced by affect. We know that probability estimation is subject to a number of cognitive biases. Probability estimates, even for simple commonly used phrases, have been found to be highly variable both between and within subjects (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986), they are generally overestimated for very low values and underestimated for high values (cf. Kahneman, Slovic, & Tversky, 1982), and they are generally overestimated for negative outcomes relative to positive ones (Nygren & Isen, 1985). We propose that positive affect also alters these estimates to produce optimism in such a way that probabilities associated with "good" or positive outcomes are increasingly overestimated and probabilities associated ith "bad" or negative outcomes are increasingly underestimated. Hence, when asked to make judgments that are framed strictly in terms of probability estimates and not choices, we would expect an individual in a positive affective state to exhibit optimism about the likelihood of the relevant events.

On the other hand, we propose, as suggested by Kahneman and Tversky (1979), that in real choice situations where loss is possible, the probability estimates are additionally weighted (not necessarily linearly) to reflect the individual's predisposition toward risk-seeking or risk-aversion behavior. Positive affect modifies the weighting functions in the following way. First, as argued above, there is strong evidence to suggest that positive affect causes decision makers to be particularly sensitive to potential loss. Consequently, not only is there a modification of the losses end of the utility function itself (Isen et al., 1988), but also the weights associated with losing outcomes are magnified relative to those associated with positive outcomes. Hence, even though affect may raise one's expectations that a good outcome will occur and lower one's expectations that a bad outcome will occur (optimism), his or her choice behavior may still reflect risk-aversion or loss-aversion because the *overweighting* of the negative outcomes in the choice situation due to the affective state change will have more than compensated for the *overestimation* of good and *underestimation* of bad outcomes in forming an overall evaluation of the choice alternative.

The present study was designed to test this hypothesis by independently examining the influence of positive affect on probability estimation and decision weight formation within the same individuals. We designed the study to show that the cautious optimism phenomenon could be easily demonstrated and explained by positive affect induction. Specifically, we designed one betting task where real loss was possible and probabilities were explicitly stated, and hence, the estimation process was eliminated. We proposed that in this betting task positive affect would induce cautious behavior relative to that shown by control subjects by differentially altering the weights of the probabilities for good and bad outcomes to reflect greater risk and/or loss-aversion. We then designed a second probability estimation task, where gambling was not involved, to test our prediction that the same positive-affect subjects would exhibit the hypothesized greater optimism relative to control subjects in their probability estimates.

Method

Subjects.

Subjects were 109 undergraduates from The Ohio State University who were fulfilling a requirement for an introductory psychology course. These subjects were randomly assigned to one of two conditions, Positive Affect (n = 51) or No Affect (n = 58).

Procedure.

All subjects were seated in individual cubicles in front of a CRT controlled by a Data General Nova 3 computer. All stimulus presentations and responses were made on the CRT. Subjects were informed that the experiment consisted of two parts -- an estimation task and a gambling task. Subjects were required to make probability estimates for twelve different three-outcome gambles with commonly used probability phrases and to make actual gambling decisions about a comparable set of eleven three-outcome gambles where probabilities were explicitly stated as decimal values.

Affect Manipulation. Affect was induced in this study by presenting one group of subjects with a small bag of candy at the time that each was seated in the booth with the computer terminal. Subjects were told that the gift was "just a small token of appreciation for their willingness to volunteer for this study." The receipt of such a gift has been used in a number of affect studies and has been shown through different manipulation checks to be a successful positive affect inducer (c.f., Isen & Patrick, 1983; Arkes et al., 1988).

All subjects were then asked to do the two different short tasks, a betting task and a probability estimation task. Because affect induction of this type is known to be temporary (usually lasting about one-half to one hour), the two tasks were counterbalanced, so that approximately one-half of the subjects in each group did either the gambling or the estimation task first. This allowed us to examine whether the task presented second, regardless of which type, was less influenced by the affect manipulation than when it was presented first.

Betting Task. In the betting task, each subject was presented with a different random ordering of the eleven three-outcome gambles presented in Table 3.1. These gambles were all zero expected value gambles in which the ratio of probability of winning to losing (and consequently the ratio of amount to lose to amount to win) were either Low P_W/P_L (.20, .30, .40, .50), Medium P_W/P_L (.80, 1.00, 1.25), or High P_W/P_L (2.00, 2.50, 3.33, 5.00).

In order to make the gambling situations realistic, subjects were told upon entering the experiment that they were going to do some gambling, but not for money. They were informed that although they could not gamble for money, the experimenter had special permission to allow them to gamble with something else that was valuable to them -- their credit hour for participating. In actuality, the experiment was programmed so as to insure that no subject would lose his or her credit.

Subjects were told that they were going to play eleven three-outcome gambles in which they could either win, lose, or break even on points. Forty points represented their participation in the experiment. If they finished the gambling task with more than their 40 points, they would "win" two credit hours for participating rather than the customary one. If they finished with between one and 39 points (i.e., if they still had any points

Table 3.1

Mean Bets for Gambles Used in the Betting Task

PROB RATIO	WIN PTS	WIN (PR)	BREAK EVEN PTS	BREAK EVEN (PR)	LOSE PTS	LOSE (PR)	VAR	P_W/P_L	AFFECT MEAN	AFFECT STD	CONTROL MEAN	CONTROL STD
Low P_W/P_L	20	(.10)	0	(.40)	-4	(.50)	48	.20	.72	(.74)	1.00	(.89)
	13	(.15)	0	(.35)	-4	(.50)	33	.30	.51	(.62)	.72	(.86)
	10	(.20)	0	(.30)	-4	(.50)	28	.40	.66	(.73)	.96	(.89)
	8	(.25)	0	(.25)	-4	(.50)	24	.50	.68	(.66)	.80	(.74)
Med P_W/P_L	5	(.40)	0	(.10)	-4	(.50)	18	.80	.70	(.78)	.89	(.79)
	4	(.45)	0	(.10)	-4	(.45)	14	1.00	.77	(.70)	1.11	(.84)
	4	(.50)	0	(.10)	-5	(.40)	18	1.25	.87	(.77)	1.04	(.82)
High P_W/P_L	4	(.50)	0	(.25)	-8	(.25)	24	2.00	1.00	(.98)	1.37	(1.20)
	4	(.50)	0	(.30)	-10	(.20)	28	2.50	1.15	(.78)	1.52	(1.02)
	4	(.50)	0	(.35)	-13	(.15)	33	3.33	1.06	(.87)	1.20	(.96)
	4	(.50)	0	(.40)	-20	(.10)	48	5.00	1.23	(.73)	1.37	(.78)
Mean									.85		1.09	

Note 1: All but Gambles 2 and 10 have zero expected value; these two gambles have EV = -.05 and +.05 respectively, because the number of points were rounded from -12.5 and +12.5 to -13 and +13. The gambles have been grouped into "High", "Medium", and "Low" on the basis of the ratio of Probability of Winning to Probability of Losing.

Note 2: Higher numbers indicate a greater willingness to bet for bigger outcomes. A mean of 1.00 would indicate that, on average, the outcomes of the gambles were left alone. Standard deviations for each gamble are in parentheses.

left) they would "break even" and receive their normal credit hour. If they lost all of their points, however, they would lose their normal credit hour and would have to participate in an additional experiment for their course requirement. At this point all subjects were given the opportunity to withdraw from the study with one credit hour and without any penalty. A total of four subjects chose to do so and were thoroughly debriefed as to the nature of the study. All other subjects were debriefed as to the necessity of this manipulation and mild deception at the completion of the experiment.

The subjects' task each time was to decide for each gamble whether to play it as it was presented, to double the outcomes, to triple the outcomes, or to not play the gamble at all. These judgments were recorded as 1, 2, 3, or 0, respectively. Thus, it was still possible for a subject to participate yet insure that she would keep the 40 points and get her one credit hour by simply responding "0" (no play) each time. However, no subject ever chose this strategy.

We followed the procedure suggested by Davidson et al. (1956), and told subjects that because of time constraints, all eleven of the gambles would actually be played after all gambling decisions had been made. Thus, subjects were not given continuing feedback as they made their decisions. This was done to insure that it was not possible for gambling outcomes to effect a possible shift in reference point during the gambling task and hence, confound the affect manipulation. We did not want the gambling results to produce possible positive or negative affect in and of themselves.

Probability Estimation Task. The second task consisted of presenting subjects with a random ordering of twelve three-outcome gambles of the same form described earlier, but for which the probabilities were actually stated in terms of commonly used phrases rather than proportions. These twelve gambles were presented individually but were actually formed from six pairs of gambles in which the probability of winning phrase and the probability of losing phrase were reversed. The same four phrases (*extremely unlikely (EU), unlikely (UN), somewhat unlikely (SU),* and *worse than even chance (WTE)*) were each associated with the winning and losing outcomes three times. Each winning/losing phrase combination was then paired with a different probability phrase for the "breaking even" or zero outcome. This third phrase was chosen on the basis of results from a pilot study such that the sum of the mean ratings given to the three phrases in each gamble was as close to 100 as possible. These mean values were found to be similar to those recently obtained by Wallsten et al. (1986).

The point values to win and lose in each gamble were again chosen so that the gambles would have approximately zero expected value. An example of one of the six reverse gamble pairs that were presented to subjects are:

GAMBLE 1

Points	Probability
+20	Extremely Unlikely
0	Worse than Even Chance
-4	Even Chance

GAMBLE 2

Points	Probability
+4	Even Chance
0	Worse than Even Chance
-20	Extremely Unlikely

Subjects were asked to evaluate each gamble by indicating what the phrases meant to them in terms of X chances out of 100. To simplify the task they were asked to think of a roulette wheel with 100 slots and to decide how many of the slots would be used to represent each phrase. The three responses on each trial were forced to add to 100; if they did not, the computer program so informed the subjects and asked them to revise their estimates appropriately.

Results

Betting Task. The betting data for the eleven gambles was first examined via a 2 x 2 x (11) two-between (Group: Affect vs. Control and Ordering of Tasks: Betting vs. Estimating first), one-within subjects (Gambles) ANOVA. As expected, all effects involving the ordering variable were not significant ($p > .05$); for subsequent analyses, then, the data were collapsed across this ordering factor. This non-significant ordering effect supported our assertion that (1) the two tasks were of short enough duration so that the affect manipulation did not "wear off" after the first task, and (2) that performance of the first task was sufficiently independent of the second task so as not to confound our analyses.

The ANOVA of the betting data indicated a strong group difference with affect subjects being less willing to bet across all gambles than were controls (M_{AFFECT} = .85, $M_{CONTROL}$ = 1.09, $F(1,99) = 7.27$, $p < .01$; MSE = 2.16).[3] There was also a significant main effect for Gambles ($F(10,990)$ = 10.10, $p < .001$; MSE = .56), but no Group X Gamble interaction ($F(10,990) < 1$). Figure 3.1 plots the mean bet values for each of the eleven gambles against the ratio of P_W/P_L. These results clearly show that the probability ratios strongly influence the betting behavior, more so than the amount ratios. Both groups were much less likely to gamble on low P_W/P_L gambles, even though the ratio of Points to Win to Points to Lose was inversely high. Mean bet values were *positively* correlated across gambles with the P_W/P_L ratio (r = .88 for Affect, df = 9, $p < .01$; and r = .70 for Control, df = 9, $p < .02$) but not with the Points to Win/Points to Lose ratio or variance, although the functions in Figure 3.1 are clearly not simple monotonic ones for either group.

There was, however, no indication of an interaction where affect subjects were expected to be more likely to gamble relative to controls on high P_W/P_L gambles. As seen in Figure 3.1, for every one of the eleven gambles, the mean bet value was smaller for affect subjects. This suggests that probability weights like utility values are influenced by affect in the betting task. Whereas the likelihood of

Table 3.2

Mean Bet Values for Low, Medium, and High

P_W/P_L Three-Outcome Gambles

Probability Ratio	Affect (n=47) Mean	Affect (n=47) StErr	No Affect (n=54) Mean	No Affect (n=54) StErr	Diff	p-value
Low P_W/P_L	0.64[1]	(.07)	0.87	(.09)	-0.23	$p < .05$[2]
Med P_W/P_L	0.78	(.09)	1.01	(.08)	-0.23	$p < .05$
High P_W/P_L	1.11	(.09)	1.37	(.10)	-0.26	$p < .05$

[1]Higher numbers indicate a greater willingness to bet for bigger outcomes. A mean of 1.00 would indicate that, on average, the outcomes of the gambles were left not changed. Standard errors for each mean are in parentheses.

[2]Comparison between Low and High levels within each group were also found to be sigmificant, $p < .01$ in each case.

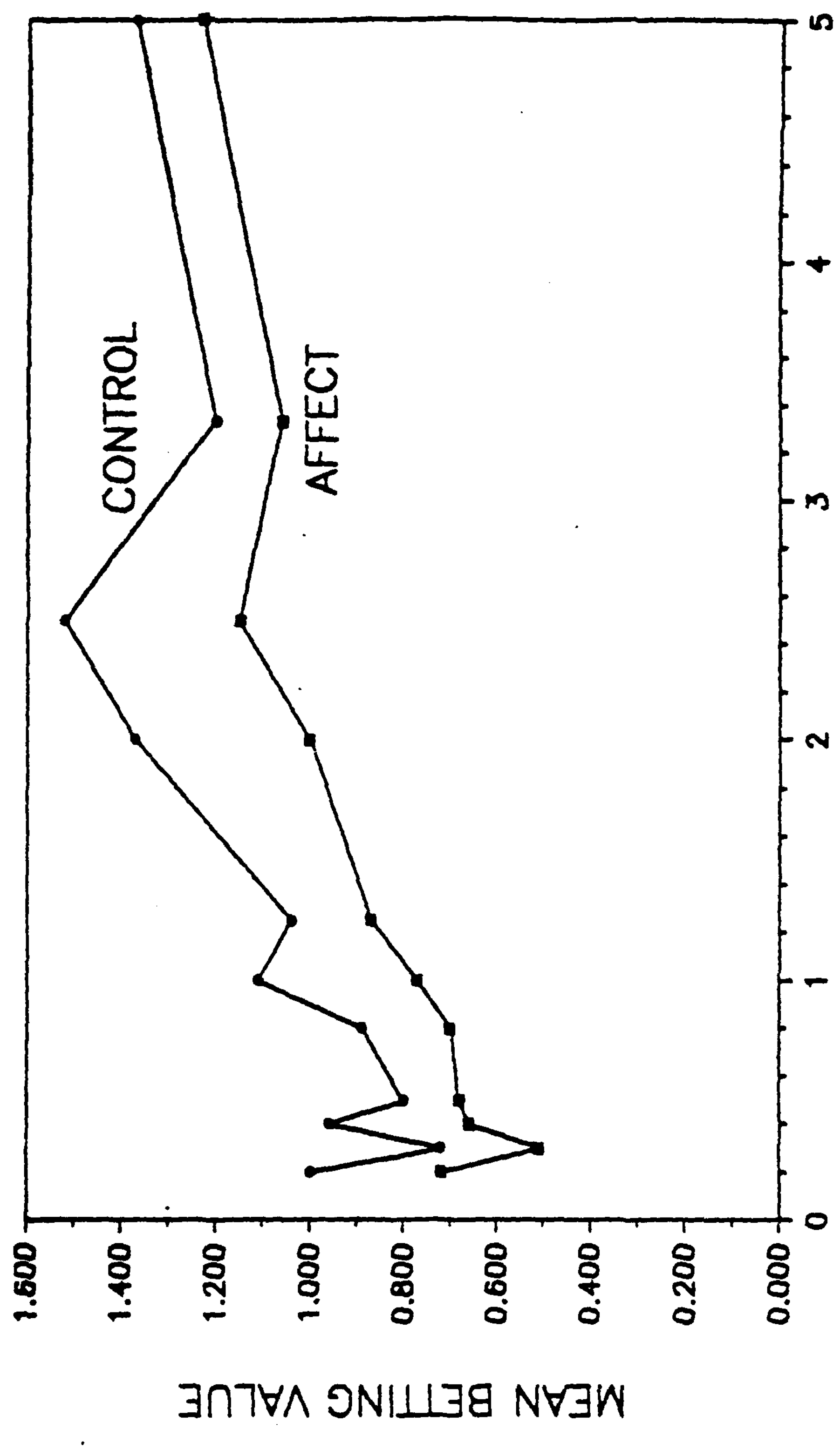


Figure 3.1. Plot of the ratio of the probability of winning to losing vs. the mean bet value (on a scale of 0 to 3) for affect and control subjects.

betting increases as the P_W/P_L ratio increases for both groups, the loss of a large number of points (e.g., -20 in Gamble 11) outweighs the large 5:1 P_W/P_L ratio enough to keep the affect subjects from betting as much as controls. This result is summarized more clearly in Table 3.2 where the gambles were grouped into Low, Medium, and High P_W/P_L categories. The differences between the two groups were significant for all three of these categories ($p < .05$). **Probability Estimation Task.** Data from the twelve gambles in the estimation task were first summarized for each subject. Average values for the three independent estimates given by each subject to the phrases "extremely unlikely", "unlikely", "somewhat unlikely", and "worse than even chance" were obtained separately for these phrases when they were associated with winning and losing. These data were then submitted to an initial 2 x 2 x (4 x 2) two-between subjects (Group: Affect vs. Control, and Ordering of Tasks: Betting vs. Estimating first) two-within subjects (Phrases, and Context: Win vs. Lose) ANOVA. As expected, an initial analysis of all effects associated with the Order variable were not significant ($p > .10$) and these data were then collapsed in subsequent analyses.

We hypothesized that the resulting 2 x (4 x 2) ANOVA would yield no overall Group main effect. It did not ($M_{AFFECT} = 24.53$, $M_{CONTROL} = 24.31$, $F(1,95) < 1$), but a significant overall Context main effect ($M_{WIN} = 25.49$, $M_{LOSE} = 23.36$, $F(1,95) = 15.66$, $p < .001$; MSE = 56.82) and a Group by Context interaction ($F(1,95) = 5.31$, $p < .05$; MSE = 56.82) were found. The overall Group main effect was not significant because, as expected, the overestimation of probability phrases associated with the positive outcomes by affect subjects were offset by their underestimation of the probability phrases for the negative outcomes, resulting in no difference between the overall means for affect and control subjects. It is important to recognize, however, that because the gambles were three-outcome gambles, this offsetting tendency was not constrained to be so. The estimates for the "zero-outcome" phrases could have changed as well. They did not. Across all five phrases that were associated with the zero-outcome in the gambles (***likely (L), better than even chance (BTE), even chance (EC), worse than even chance (WTE),*** and ***somewhat unlikely (SU)***) none of the mean differences were significant when compared between the two groups ($p > .10$ in each case). This result implies as Kahneman and Tversky (1979) and others have suggested that zero-outcomes (i.e., status quo) are essentially ignored. Hence, it is not surprising that their corresponding probabilities should not be influenced by positive affect.

Table 3.3

Mean Ratings of Probability Phrases When Phrases Are Associated With Winning and With Losing

	POSITIVE AFFECT (n=48)			NO AFFECT (n=49)		
Probability Phrase	Winning Ratings	Losing Ratings		Winning Ratings	Losing Ratings	
Extremely Unlikely	11.97	9.32	p <.013[1]	10.52	10.95	ns
Unlikely	24.64	21.03	p <.013	22.04	20.66	ns
Somewhat Unlikely	32.26	28.27	p <.013	30.01	29.09	ns
Worse Than Even	36.05	32.74	p <.013	36.46	34.76	ns

[1] The eight comparisons (four Win vs. Lose for each group) are based on a Bonferroni test with the family-wise error rate set at .10. This actually sets each per-comparison test at an alpha of .0125.

The means in Table 3.3 show the critical Group by Context interaction in detail. For each of the four phrases, affect subjects estimated the probability values associated with each phrase as being higher when they were associated with the winning context as opposed to the losing context. For the control subjects these differences were not significant. Figure 3.2 shows these means graphically and illustrates their differences when compared with mean estimates obtained for the phrases in a simple non-gambling or win/lose context. That is, the means in Table 3.3 were compared with those obtained in from a pilot study of 107 undergraduates who were presented with the same phrases in a non-gambling context and simply asked to interpret them in terms of X chances out of 100. Figures 3.2a and 3.2b show that for both of our groups there appears to be the typical overestimation of low and underestimation of moderate probabilities. However, the over- and underestimation crossover points for the two groups are quite different. In Figure 3.2a the curves indicate that for affect subjects when the phrase is placed in a winning context, the crossover point is about .33; for the losing context it is much lower, about .23. For control subjects, the two crossover points in Figure 3.2b are much closer, about .27 for winning and .23 for losing. It is interesting to note that it is the winning crossover point that differs for the two groups, not the losing crossover point. Hence, it appears that although in a real gambling situation, positive affect more strongly biases the *overweighting* of the likelihood of losing than winning, in an estimation task it more strongly influences the *overestimation* of winning as opposed to losing.

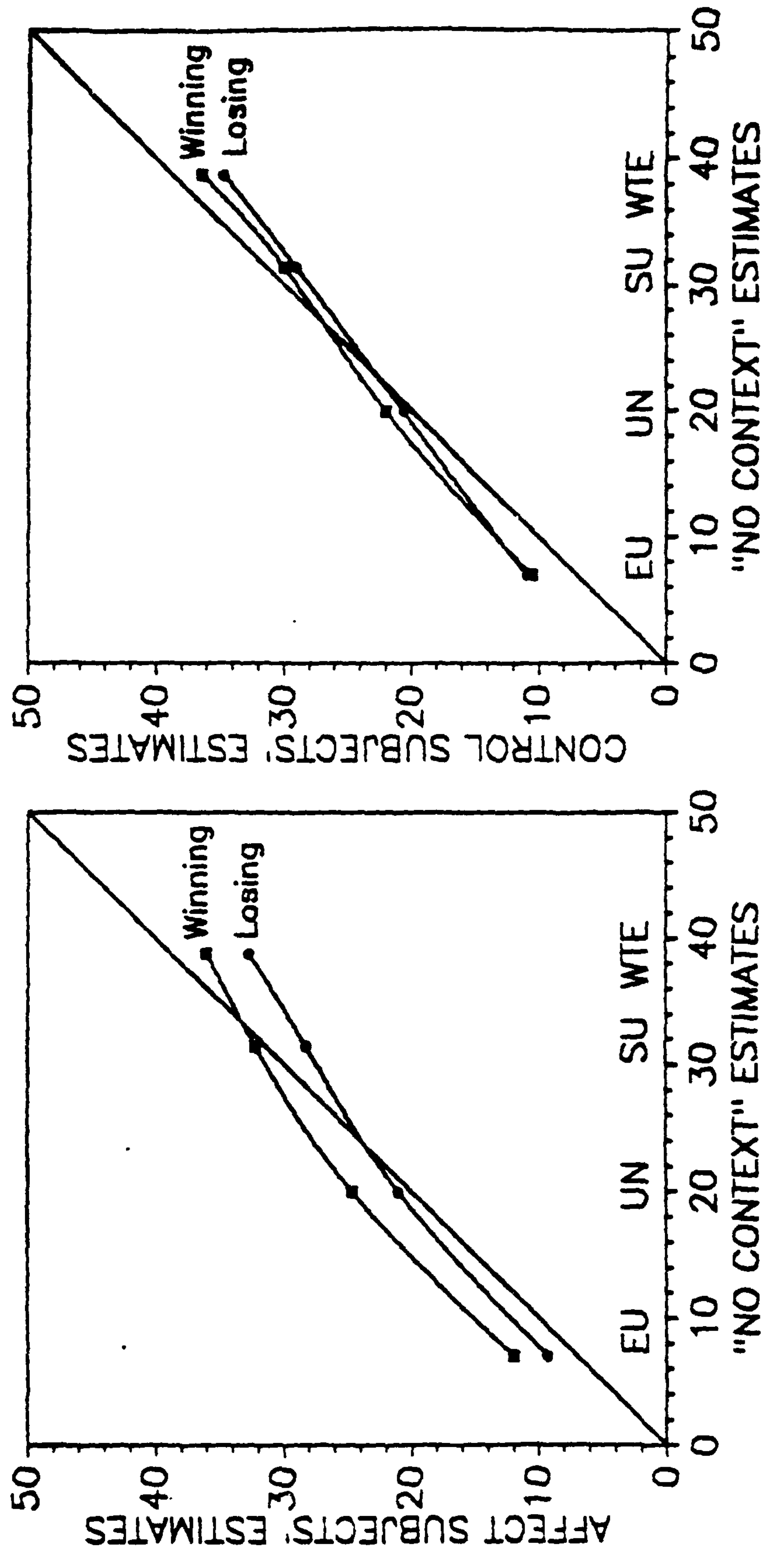


Figure 3.2. Subjective probability functions for winning and losing contexts for affect and control subjects.

Discussion

Positive affect has been found in a number of studies to produce both risk-seeking and risk-aversion behavior. Although these results seem at first look to be somewhat contradictory, the focus of the present study allows us to explain the results of these studies in a single processing model. We argue that a critical point of concern is the distinction between the estimation process for event probabilities associated with "good" and "bad" outcomes and the actual choice process where weightings of the "good" and "bad" outcomes are obtained.

The results from our study pose some interesting problems for simple bilinear models like SEU or prospect theory. The gambling data taken with the probability estimation data illustrate cautious optimism. The results of the study indicate that the effect of positive affect on subjects' betting is to decrease their willingness to bet. At the same time their subjective *estimates* of the likelihoods of winning and losing appear to be increased and decreased, respectively by the affective state. Any SEU-type model would have difficulty explaining how subjects can become more risk-averse as the ratio of the probabilities of winning to losing increase. In an SEU-type model, not only would the utilities of the outcomes have to be changed to compensate for changes in the probabilities in the betting data, but the probabilities of winning and losing would have to be differentially affected.

A more plausible explanation is to suggest that a clear distinction between the estimation process and the weighting process in gambling situations must be made. In our probability estimation task, subjects were not gambling, they were merely indicating interpretive estimates of probabilities for vague phrases. These estimates tended to exhibit the familiar cognitive bias of being somewhat inflated for low probabilities and somewhat depressed for higher probabilities when the probabilities were associated with a risky gambling context as opposed to a riskless one. Positive affect not only tended to increase this cognitive bias, but also enhanced the differentiation in the use of two separate probability functions, one for positive or good outcomes or contexts and one for negative or bad outcomes or contexts. Hence, we found a feeling of optimism in our affect subjects that was exhibited by their enhancement of subjective estimates of the probability of winning relative to losing.

In our gambling task the subjects were gambling for something very important to them -- their credit hour for participating in the experiment. That this gambling task was taken seriously and regarded as being "real" to the subjects was clear from observations that: (1) many of the subjects deliberated over whether or not to participate and risk the loss of their credit hour, with some actually deciding not to participate, and (2) very few subjects actually doubled or tripled the outcomes in even the "low risk" gambles, a observation that one might expect to find if subjects felt that they really "had nothing to lose." Thus we argue that the perception of experiencing a real loss was apparent to them. As Isen et al. (1988) have suggested, this is especially true of subjects in a positive affective state. These subjects tend to be more risk-averse in their betting behavior, when compared to control subjects -- a given loss seems more aversive to someone who is feeling happy or optimistic than to a control subject.

This suggests that the gambling task is very different from the probability estimation task. In the estimation task, subjects were providing estimates about probabilities. Affect subjects were relatively more optimistic about these judgments. In the betting task, where a real loss was possible, affect subjects appeared to be more sensitive to this loss. In our betting task, then, what were important, as suggested in Kahneman and Tversky's (1979) prospect model, were not the subjective probabilities per se, but rather the *weights* or importance values that individuals associated with these probabilities. If we interpret what is going on in the choice process, as Kahneman and Tversky (1979) do, in terms of decision weights rather than probabilities, then the risk-averse behavior of our subjects fits nicely with results of previous affect studies. Affect subjects, being more sensitive to potential loss, weight these losses more heavily relative to potential gains. Hence, the observation in the betting task of risk-averse or cautious gambling behavior is obtained. Although affect subjects are relatively more optimistic in their assessments of probabilities, they are also relatively *more cautious* in their gambling. They tend to *overestimate* the same probability statement when it is associated with winning as compared to losing, but they tend to *underweight* it when it is associated with winning as compared to losing.

Thus although our results provide support for Kahneman and Tversky's (1979) prospect theory concept of decision weights in the simple bilinear model, they also seem to point to the necessity of Luce and Narens (1986) proposal of a dual-bilinear model. A dual-bilinear model is one that allows for two separate and independent probability weighting functions, S^+ and S^-, one for good outcomes and one for bad. Our subjects gave different estimates to the same phrases in the same three-outcome gamble combinations depending on whether these phrases were associated with winning or losing. At the same time, those subjects who indicated a greater optimism in these judgments were more risk-averse in their betting. This suggests the differential weighting of the probabilities when they are associated with winning and losing that is permitted in a dual-bilinear model.

We suggest, as Coombs and Lehner (1984) suggested in their analysis of bilinear models, that there are two separate probability weighting functions relevant to decision making under risk, one for good and one for bad outcomes. We argue that the difference between these two functions can be enhanced by one affective state. Further, they are generally enhanced for probability weighting in a manner opposite of that associated with direct probability estimation -- positive affect positively biases estimation of the likelihood of good events and negatively biases the estimation of bad events, but amplifies the weighting of bad outcomes relative to good ones in the decision process.

Thus, we explain the cautious optimist as one who, because of some temporarily induced positive affective state, believes in a greater likelihood ratio of good to bad probabilities than might otherwise be expected, but who is also distinctly aware of his or her potential loss and consequently differentially weights potential losses relatively higher than equally likely potential gains, so much so that he or she exhibits a greater risk-aversion than would normally be observed.

Further, under this model we would expect that in a situation of no danger or potential loss, positive-affect subjects would think more about positive outcomes, stimuli, and options and exhibit more risk-seeking behavior than would normally be expected, since losses are not relevant and cannot be overweighted relative to gains. We did not test this prediction in our study, but other work by Isen and her colleagues cited here (cf. Arkes et al., 1988) strongly support it. For example, in the Isen and Patrick (1983) study where gambling preference was assessed in terms of amount bet in a game of roulette, and the items being wagered were points representing the subject's credit for participation in the experiment, it was found that persons in whom positive affect had been induced by receipt of a small gift, bet more on a gamble with a high probability of winning (83% chance of winning), but significantly less on a gamble with a high probability of losing (17% chance of winning), than did a control group (Isen & Patrick, 1983).

In summary, the results presented here, together with those reported elsewhere (e.g., Isen et al., 1988; Nygren & Morera, 1988), are resulting in a more complete picture of the complex roles that positive affect, probability estimation, and decision weighting play in decision-making under risk or uncertainty. The results of the present study suggest a mechanism underlying that complex state of affairs. A dual-bilinear model seems necessary to explain how persons who are feeling happy are more likely to overestimate the probability of good outcomes relative to bad ones, but are less likely to take risks if their perceived losses are high.

Footnotes

[1] We are using the construct *risk-seeking* as it is defined in the statistical decision theory literature; it is defined as a preference for a gamble over a sure-thing of equal or even slightly higher expected value. The construct *risk-averse* refers to having a preference for a a sure-thing over a gamble of equal or even slightly higher expected value. Loss-aversion refers to the phenomenon in which individuals demonstrate a greater disutility for a given loss than a positive utility for a comparable gain value.

[2] The six pairs of gambles and their summed ratings from the pilot study were [EU, L, UN; 99.7], [EU, BTE, SU; 98.7], [EU, EC, WTE; 95.8], [UN, EC, SU; 101.3], [UN, WTE, WTE; 97.5], [SU, SU, WTE; 101.6]. The first and third phrases were reversed in each pair to create the twelve gambles. It is critical to note that it is not necessary in this study for each gamble to have phrases that add exactly to 100; it is sufficient that each gamble is paired with another that reverses the winning and losing phrases, and that the sums be as close to 100 as possible to make the phrase combinations plausible.

[3] Due to a programming problem with the multi-tasking feature of the Data General Nova computer, the data from the second task, though collected, were not stored on disk for several subjects. This is the reason for there being a reduction to 101 subjects (47 affect and 54 controls) for the betting data and 97 subjects (47 affect and 48 controls) for the ratings data.

4. PROJECT 3. A Method for Modeling Bias in a Person's Estimates of Likelihoods of Events

It is of paramount importance in decision situations involving risk to train individuals to transform uncertainties into subjective probability estimates that are both accurate and unbiased. We have found that in decision situations involving risk, people often introduce subjective bias in their estimation of the likelihoods of events depending on whether the possible outcomes are perceived as being "good" or "bad". Until now, however, the successful measurement of individual differences in the magnitude of such biases has not been attempted. In this paper we illustrate a modification of a procedure originally outlined by Davidson, Suppes, and Siegel (1957) to allow for a quantitatively-based methodology for simultaneously estimating an individual's subjective utility and subjective probability functions. The procedure is now an interactive computer-based algorithm, DSS, that allows for the measurement of biases in probability estimation by obtaining independent measures of two subjective probability functions (S^+ and S^-) for "winning" (i.e., good outcomes) and for "losing" (i.e., bad outcomes) respectively for each individual, and for different experimental conditions within individuals. The algorithm and some recent empirical data are described.

It is argued that, if in decision situations involving substantial risk or potential loss, our goal is to train individuals to become expert decision makers, it is important to understand how people subjectively evaluate and represent uncertainties or probabilities. Decision theorists have argued for some time that any decision analysis under risk must involve the assessment of uncertainties, and that uncertainties can best be measured by subjective probabilities that represent the decision maker's degree of belief about the relevant uncertain events. The decision maker must somehow transform uncertainties into subjective probability estimates that are both accurate and unbiased. There is, however, convincing evidence that in many complex decision situations involving risk or uncertainty, people use heuristics that often introduce bias in the subjective estimation of the likelihoods of events relevant to the outcomes of a decision (Tversky and Kahnemann, 1974; Hogarth, 1981). They may judge the probability of an event by its representativeness of a set or class of events, by its availability in memory as a relevant example, or on the basis of an adjustment from a numerical anchor point.

Other recently published edited volumes and papers by Arkes and Hammond (1986), Coombs and Lehner (1984), Einhorn and Hogarth (1985), Kahneman and Tversky (1972; 1979a; 1979b; 1982), Kahneman, Slovic, and Tversky (1982), Luce and Narens (1985), and Tversky and Kahneman (1973; 1974; 1980; 1981) reflect this new direction in the field of judgment and decision making research. Studies have shown that people generally do not make good probability estimates (Kahneman, Slovic, & Tversky, 1982; Fischoff, 1977; Slovic, Fischoff, & Lichtenstein, 1980, 1982; Slovic & Lichtenstein, 1983). They overestimate low and underestimate high probabilities and they ignore base-rate information (Kahneman & Tversky, 1973; Bar-Hill, 1980); they revise opinions too conservatively (Edwards, 1968); they indicate excessive confidence in their judgment (Fischoff & Slovic, 1980); and they are influenced by their affective mood state (Nygren & Isen, 1985; Isen & Nygren, 1987).

However, only recently have systematic research efforts investigating the cognitive mechanisms by which biases are generated been reported. For example, Nygren and Isen (1985) have shown that a positive mood state can lead decision makers to exhibit "cautious optimism" in risky choice situations.

They become optimistic in the sense that they tend to overestimate the likelihood of "good" events and underestimate the likelihood of "bad" events; but at the same time they exhibit a cautious shift toward risk-aversion in their actual choices. Clearly, such findings on mood state are relevant to the understanding of cognitive biases in decision making.

These findings imply the need for models that interrelate cognitive processes and judgmental biases. Wickens (1987) has argued that without an understanding of these biases in such a framework, it is difficult to predict how specific decisions are being made by individuals. But, how can such biases be quantitatively measured? Most models of decision making under risk assume that there are four basic questions that remain the focal issues in decision analysis. They are: (1) what are the possible courses of action? (2) what are the outcomes associated with these courses of action? (3) what is the utility associated with each outcome?, and (4) what is the probability associated with each outcome? Much empirical research continues to focus on Questions 1, 2, and 3, and, in particular, the measurement of utility (Einhorn & Hogarth, 1981; Kahneman & Tversky, 1982). This paper describes a method to take a closer look at Question 4, the role of the subjective probability function.

MATHEMATICAL MODELS

There are two leading models of risky decision making upon which this research is based, subjective expected utility (SEU) theory and Kahneman and Tversky's (1979) prospect theory. In SEU theory the overall utility of a course of action or "gamble" is found by taking, for each possible outcome in the gamble, the product of the utility of the outcome multiplied by the subjective probability associated with that outcome's occurrence, and summing these terms across all outcomes. The decision maker is assumed to choose the gamble/option with the highest overall utility. Kahneman and Tversky's (1979b) prospect theory proposes that the decision process is, in fact, completed in two phases, with the potential courses of action first being "framed" for the choice process. This framing often may constitute a preliminary look at the outcomes, and this look sometimes results in a simplified representation of the choice alternatives, particularly if the alternatives are complex. Following this initial phase, the alternatives are actually evaluated in a manner similar to that suggested by SEU theory, where the alternative with the highest value (utility) is chosen.

Both models take the same general form, then, in that overall preference for a course of action or gamble (G) is assumed to be a function of (1) the values or utilities of the possible outcomes and (2) the subjective probabilities (in SEU theory) or decision weights (in prospect theory) associated with these outcomes. Expressed mathematically, for a simple gamble of the form $G = (x, p; y, 1-p)$ where one obtains outcome x with probability p or outcome y with probability 1-p, the subjective value of the gamble (G) is assumed in these models to be

$$V(G) = V(x) * S(p) + V(y) * S(1-p), \qquad (4.1)$$

where V(x) is the utility of outcome x and s(p) is a subjective probability that is associated with outcome, x. In prospect theory the overall value is based on multiplying the utility of each outcome by a decision weight rather than a probability estimate, per se. The decision weights are assumed to increase monotonically with objective probabilities of events, but are larger than the objective probabilities for extremely unlikely outcomes and smaller than the objective probabilities for more likely outcomes. In prospect theory the decision weights for complementary events with probabilities p and 1-p need not necessarily add to one, but will generally be less than one, a property that Kahneman and Tversky (1979b) label as subcertainty. However, since both models are simple bisymmetric representation, they do not allow for a differential weighting of an event's probability in winning versus losing or "good" versus "bad" contexts (Isen & Nygren, 1987; Nygren & Isen, 1985). That is, the models do not allow for the possibility that a decision maker might weight or even evaluate a probability like .2 or .8 differently, depending on the outcome with which it is associated. To account for such findings one needs a modification of SEU with a dual probability function. Such a model has been formally proposed by Luce and Narens (1985). A dual bilinear model would allow for the measurement of probability bias, where "good" and "bad" outcomes can differentially affect subjective judgments of the same explicitly stated probabilities.

THE QUANTITATIVE METHOD

A modification of the procedure used originally by Davidson, Suppes, and Siegel (1957) is now a computer-based algorithm, DSS, (cf. Isen & Nygren, 1988) that independently measures the utility and subjective probability functions (U and S) in Eq. 4.1 above. SpeCifically, the S function is measured separately as two functions, S^+, and S^-, in order to assess potential bias in judgments of the likelihoods of good (S^+) and bad (S^-) outcomes. To the extent that the same function is obtained for S^+, and S^-, no dual bilinear bias is present in a decision maker's probability estimates. To the extent that the functions obtained for S^+, and S^- differ, a cognitive bias exists in the individual's probability estimation process.

The DSS procedure involves determining the indifference point in sequences of pairs of gambles so that Eq. 4.1 can be found for each gamble;

these equations are then equated and solved in order to first estimate the subjective utility associated with various outcomes. This utility function is then used in a second phase to estimate the subjective probability functions. On each trial, an individual is presented with two two-outcome gambles and is asked to indicate which of the two he/she prefers. The two-outcome gambles are set up as follows: Individuals are told that in each gamble, one outcome would be obtained if the event E occurs and the other outcome would be obtained if the event E does not occur. The event E is never specified, but individuals are informed that it has a true computer-generated probability of one-half. (Data from several studies have indicated that such instructions produce no bias between these two alternatives of E and not E; individuals indeed weight the two events equally.) On each trial one gamble, Gamble 1, has both outcomes fixed at specified values (e.g.. some amount of money); the other gamble, Gamble 2, has one fixed outcome and one that is varied. For each of eight series of trials, the individual is asked to compare the fixed-outcome gamble (No. 1) with Gamble 2, which can be modified by changing the value in the variable outcome. The decision maker's task on each trial is simply to indicate which gamble s/he prefers. If Gamble 1 is preferred, the variable outcome in Gamble 2 is adjusted upward by DSS to make this gamble more attractive; if Gamble 2 is preferred, the variable outcome in Gamble 2 is lowered to make this gamble slightly less attractive. The amount of adjustment made by DSS depends on whether the individual indicates that one gamble is either slightly or strongly preferred to the other. Since events E and not E have probabilities fixed at .5 and these events are weighted as equivalent in probability, DSS determines subjective utilities by noting in the variable-outcome gamble the value/amount necessary for a subject to change his/her preference ordering between the fixed-outcome and variable-outcome gambles (indicating an indifference or equivalence point for that pair of gambles). That is, DSS notes the amount that the individual assigns to the variable outcome in Gamble 2 such that she no longer has a clear preference for either Gamble 1 or 2.

DSS is currently set up to handle gambles with money amounts although other attributes (e.g., points to earn prizes) have been used successfully. One sequence of pairs of gambles we have used with DSS is presented in Table 4.1; we will use these values throughout the remainder of this paper as an illustration. Each of the eight situations presented in Table 4.1 actually consists, then, of a series of adjustments to Gamble 2 that lead to the estimation of a subjective utility scale. For example, for the sequence presented in Table 4.1, in the first situation, the individual is faced with one gamble for which she would lose $10 with p = .5 (i.e., if E occurs) and would lose $10 with p = .5 (if not E occurs). This, then, is a sure-loss gamble. The alternative gamble in the pair is described as resulting in a loss of A dollars with p = money amount associated with A is initially randomly set to a large negative value or to a large positive value making one gamble initially more attractive. The individual adjusts the variable value up or down as necessary to reach indifference between the gambles (-10, -10) and (-A, +10). The final dollar amount associated with A is recorded for the individual by DSS so that this information can be used to determine other utility values in subsequent trials.

To facilitate the estimation process, we first assign a utility of +1 to +10 dollars and a utility of -1 to -10 dollars. (Because in SEU theory the subjective utility scale is unique up to an affine transformation, that is, it is interval, we can without any loss of generality assign the utilities of +1 and -1.) Then, after -A is found, we can determine the utility value

Table 4.1
Construction Sequence for Trials 1 - 8
Used to Obtain Individual Utility Functions

Trial	Gamble 1 Get	Gamble 1 Get	Gamble 2 Get	Gamble 2 Get	Subjective Utility
1	-10	-10	- A	+10	V(-A) = - 3
2	+10	+10	-10	+ B	V(+B) = + 3
3	-10	+ B	- A	+ C	V(+C) = + 5
4	- A	+10	- D	+ B	V(-D) = - 5
5	- D	+ B	- E	+ C	V(-E) = - 7
6	- A	+ C	- D	+ F	V(+F) = + 7
7	- D	+ F	- E	+ G	V(+G) = + 9
8	- E	+ C	- H	+ F	V(-H) = - 9

associated with -A (which now corresponds to an empirically determined amount of money), by substituting in the formula in Eq. 4.1

$$S^*[.5] * (-1) + S^*[.5] * (-1) =$$

$$S^*[.5] * (+1) + S^*[.5] * V(-A) \qquad (4.2)$$

making V(-A) = -3. In a manner comparable to that for Trial 1 in Table 4.1, a value for +B can be found next by comparing the gambles (+10, +10) and (-10, +B), yielding the amount that is associated with a utility of +3. These utilities of -A and +B are then used in situations 3 through 8 to determine other points on the subjective utility scale. Currently, DSS finds an estimate for +C, -D, -E, +F, +G, and -H, which have utility value of -5, -5, -7, +7, +9, and -9, respectively.

PROBABILITY ESTIMATION

Once these eight points on the utility function have been obtained, a nonlinear regression analysis is done to find the best fitting utility curve for the estimated point values. A typical observed curve is shown in Figure 4.1. Given this best fitting curve, other utility values can be estimated for the individual. A new series of gambles, with different probability values,

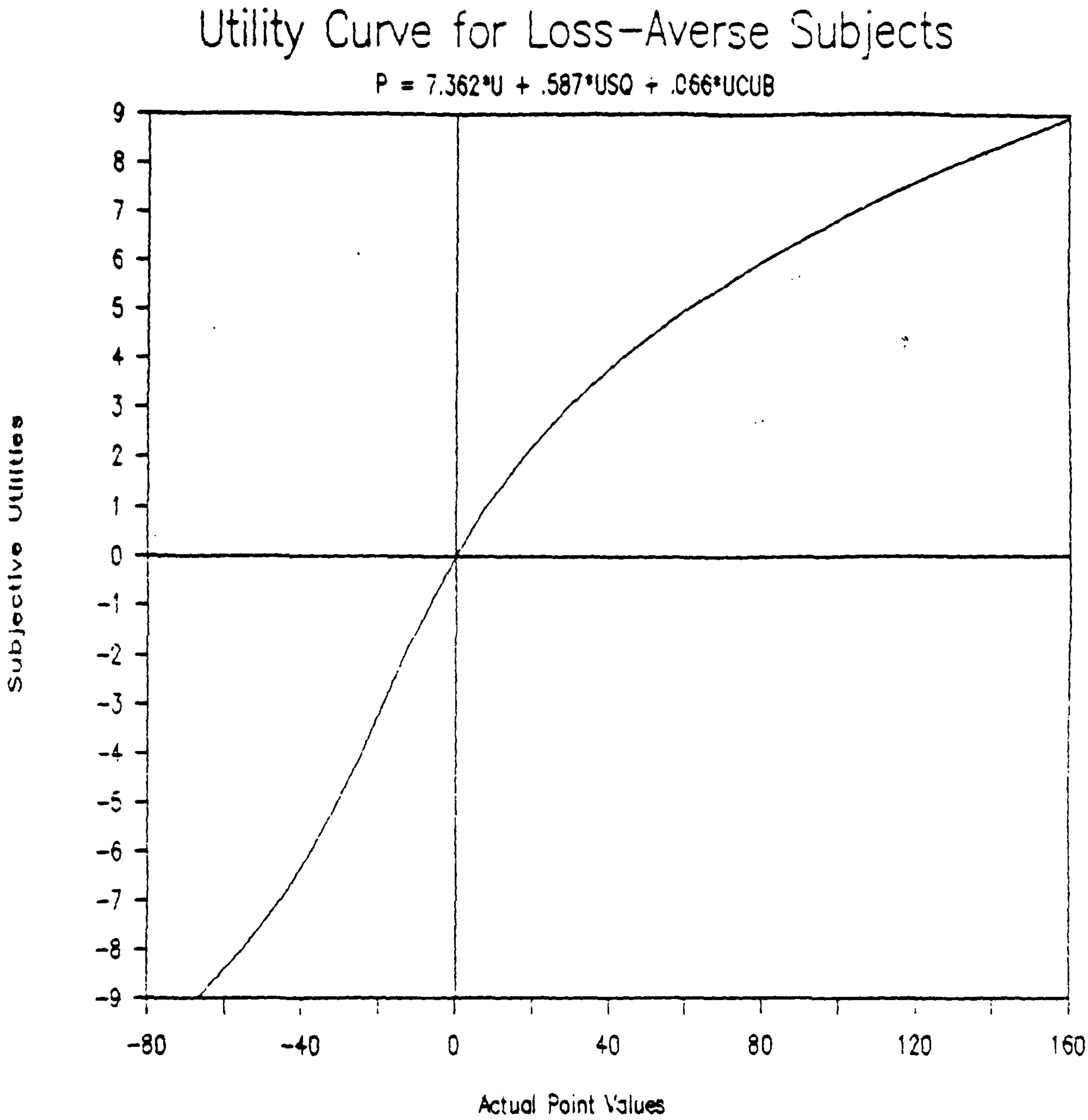


Fig. 4.1. A typical observed utility function.

are used to obtain the subjective probability functions for winning (i.e., "good" outcomes) and losing (i.e., "bad" outcomes). Again, in the variable gamble, Gamble 2, one outcome is obtained if the event E (s(p) = .5) occurs and the other outcome is obtained if the event E does not occur (also s(p) = .5). The fixed gamble is similar to those presented in the utility estimation phase, except that now the probabilities of winning and losing are either .2/.8, .4/.6, .6/.4, or .8/.2. The variable outcome in Gamble 2 is again modified in a series of steps until the individual indicates that the two gambles are equally attractive. An example of such a gamble is shown in Figure 4.2.

Table 4.2 shows an illustrative series of eight trials used to estimate the probability functions. Note that on two trials a subjective probability of winning value (S^+) and probability of losing value (S-) is estimated for each of the values 2., with loss of generality, S(.5) = .5. Equations 4.3 and 4.4 illustrate two examples of trials where $S^-(.8)$ and $S^-(.2)$ are determined as follows:

$$S^+[.5] * (-1) + S^-[.5] * (-1) =$$
$$S^+[.8] * (+1) + S^-[.2] * V(-A) \qquad (4.3)$$
$$S^+[.5] * (-1) + S^-[.5] * (-1) =$$
$$S^+[.8] * (+1) + S^-[.2] * V(-A). \qquad (4.4)$$

If we assume that these two subjective probabilities act like objective probabilities and add to one, then by substituting in utility values obtained from the utility curve estimated in the first sequence of trials, then two independent estimates of $S^+(.8)$ and $S^-(.2)$ can be obtained. If we do not wish to assume that these two subjective probabilities add to one, we can still estimate each since we have two equations and two unknowns.

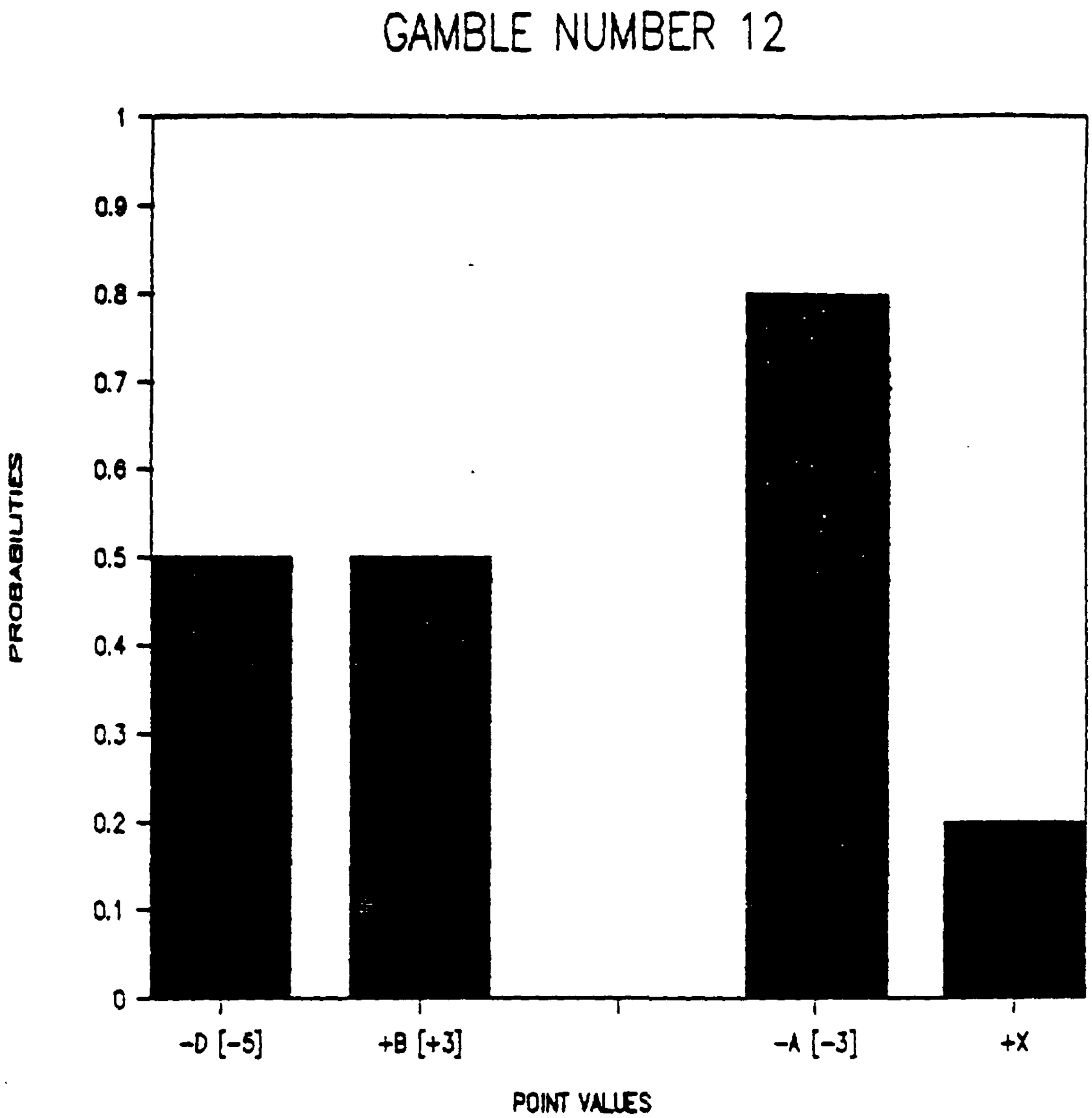


Fig. 4.2. An example of a variable gamble.

Table 4.2

Construction Sequence for Trials 9 - 16 Used to Obtain Individual Probability Functions

	Gamble 1		Gamble 2	
Trl	Get	Get	Get	Get
9	- D (.5)	+ F (.5)	- A (.6)	+ X (.4)
10	- D (.5)	+ B (.5)	- H (.2)	+ X (.8)
11	- A (.5)	+ C (.5)	- D (.4)	+ X (.6)
12	- D (.5)	+ B (.5)	- A (.8)	+ X (.2)
13	- E (.5)	+ C (.5)	- D (.6)	+ Y (.4)
14	- A (.5)	+ C (.5)	- D (.2)	+ Y (.8)
15	- D (.5)	+ B (.5)	- E (.4)	+ Y (.6)
16	- A (.5)	+ C (.5)	-10 (.8)	+ Y (.2)

In a comparable manner we can find restricted or unrestricted estimates of $S^-(.8)$ and $S^+(.2)$ from two new equations, and for $S^+(.6)$ and $S^-(.4)$ and $S^-(.6)$ and $S^+(.4)$. Other probability values could also easily be estimated. Regardless of whether a restricted or unrestricted model is fit, and regardless of the shape of each individual's subjective utility function, the S^+ and S^- functions should be the same if no probability bias due to a good vs. bad context exists in an individual's data. That is, if no bias exists, we should expect $S^-(.2) = S^+(.2)$, $S^-(.4) = S^+(.4)$, $S^-(.6) = S^+(.6)$, and $S^-(.8) = S^+(.8)$.

AN EMPIRICAL EXAMPLE

Subjects. Fifty-three male and female undergraduate students at The Ohio State University volunteered to participate in this study in partial fulfillment of the requirements of their introductory psychology course. On each trial, subjects were presented with two two-outcome gambles and were asked to indicate which of the two they preferred. In this way both subjective utilities and subjective probabilities for .2 to .8 were independently estimated.

Procedure. Subjects were run in groups of two or three and were individually seated in front of a computer terminal. They were initially instructed that they would be asked to make choices between pairs of gambles, and after they had indicated all of their preferences, ten of the choice situations would be randomly sampled and played. The gamble that they would play in each pair would be randomly picked between the two gambles indicated to be equally attractive on each trial. Each subject was given 100 points, representing his or her credit for participating and was told that s/he would be gambling with this credit (not money) in the randomly selected gambles. Subjects were told that as a result of the gambling, they might either lose their credit hour, retain it, or gain an additional credit. At this point, all subjects were given the opportunity to withdraw from the study without penalty. None did so.

Table 4.3

Median Estimates of Winning anosing Probabilities: Restricted Case

Actual	Prob L	Prob W	Diff
.200	.347	.208	.139
.400	.462	.401	.061
.600	.599	.539	.060
.800	.793	.653	.040

Median Estimates of Winning and Losing Probabilities: Unrestricted Case

Actual	Prob L	Prob W	Diff
.200	.312	.238	.074
.400	.469	.400	.069
.600	.567	.508	.059
.800	.815	.514	.301

Results. Table 4.3 presents the estimated subjective probabilities that were obtained for .2, .4, .6, and .8 for both the winning and losing contexts. The data labeled "restricted" represents as discussed above the average of two estimates that are based on the restriction that the estimates for complementary events sum to one. The data labeled "unrestricted" are based on the estimates found when the sums are not restricted to add to one. Regardless of which of these models is assumed, the estimates in the table indicate that indeed across individuals a strong biasing effect exists. The same objectively stated probability values (.2, .4, .6, and .8) when presented to individuals, elicit consistently different subjective values or weights in their deci_ion making process. At all four estimated levels of probability, the estimates that were associated with winning outcomes were consistently weighted lower than the corresponding values for losing. This represents a very strong affective bias in the estimation and/or weighting processes for subjective likelihoods.

SUMMARY

When an event has an established probability associated with it, it should be irrelevant whether that event is associated with a "good" or "bad" outcome; the affective nature of the outcomes should not influence the decision process. Two probability functions were estimated with different values assigned to winning and losing events. This differential weighting of the same event suggests that there is an affective influence on the judgmental process. Individuals make choices between alternatives by assigning different subjective probabilities or weights to the same explicit event depending on whether it has a positive affective component (winning) or negative affective component (losing).

The proposed quantitative methodology is important because it has the potential of not only quantitatively measuring this probability bias but also of explaining how some biases in probability estimation may cause suboptimal decisions, and how such bias can be reduced or eliminated in training decision makers to become more "expert" judges. It is designed to lead to programmatic research that has the ultimate application of developing training procedures that can: (a) standardize probability estimation methods in decision making under risk, (b) eliminate estimation biases such as over- and underestimation, (c) reduce individual differences in probability estimation, and (d) develop a scale for assessing a decision maker's accuracy and unbiasedness in subjective probability estimation. Research in this area is necessary if we are to go beyond merely describing suboptimal decision making behavior. The present research is an attempt to begin programmatic research that will allow us to predict suboptimal behavior due to biases in uncertainty estimates and to train individuals to reduce bias in their judgments. In particular, several issues seems to be initially relevant to continued research. First, how strong and how generalizable is the differential weighting effect for probabilities? Second, what factors influence the strength of this effect? And, third, can a method like the DSS procedure coupled with the dual bilinear model begin to allow us to predict and quantitatively measure suboptimal decision making strategies?

The research reported here, together with that reported elsewhere (e.g., Isen & Nygren, 1987; Nygren & Isen, 1985), are resulting in a more complete picture of the complex role that estimation bias and affect plays in decision-making under risk or uncertainty. Our findings suggest that the dual bilinear model is a model worth pursuing. It has the potential to explain a number of difficult findings in the decision making literature including the framing effect (Kahneman & Tversky, 1979b), the differential weighting effect

(Nygren, 1987), and the "cautious optimists" effect (Nygren & Isen, 1985). Finally, the DSS procedure offers a quantitative measurement procedure for actually measuring rather that simply describing biases in the judgment and decision making process.

5. RECOMMENDATIONS FOR FUTURE RESEARCH

There were two kinds of bias observed in the reported studies and these are psychologically very distinct. One, the overestimation vs. underestimation phenomenon, suggests a general misrepresentation and distortion of events that are either quite likely or quite unlikely. It is clearly a property of decision making that is undesirable since it can lead to suboptimal decisions. The second biasing effect that was found, the differential weighting of the same event, is a much more serious undesirable effect and is more interesting from a cognitive perspective. It suggests that there is an affective influence on the judgmental process. Individuals make choices between alternatives by assigning <u>different</u> subjective probabilities or weights to the <u>same</u> explicit event depending on whether it has a positive affective component (winning) or negative affective component (losing). The third biasing effect is that of base rate -- individuals tend to associated a higher likelihood with a probability statement when it is associated with an event that has a high base rate and to associate a lower likelihood to a probability statement when it is associated with a low base rate event.

I have argued that all decision making under uncertainty necessitates that a decision maker must somehow associate each relevant uncertainty with a probability value. The studies presented here indicates that humans may introduce strong affective, context, and base rate biases in this process. Clearly, such biases may lead to suboptimal decisions being made. Since the quality of expert judgment or the design of an expert system hinges on the quality of the experts' assessments of uncertainties, further research in this area seems critical. In particular, several issues seems to be initially relevant to continued research:

a. How strong and how generalizable is the differential weighting effect for probabilities that was found?

b. What factors influence the strength of this effect, and can decision makers be trained to reduce or eliminate it?

c. In decisions that involve uncertainties, what is the best way to assign probabilities to uncertain events so as to reduce bias and increase concordance among expert judges (e.g., pilots)?

d. What expert systems development tools might prove most useful in training individuals to become expert judges in situations involving uncertainty?

Clearly, the demand for individuals, and in particular pilots, to become "expert" decision makers in their work environments will only continue to increase. One aspect of what makes a person an expert decision maker or what is necessary in an expert system is a function of how well probabilities can be assigned to uncertainties that are critical components of the decision situation. It is argued here that the results of this research suggest that the four questions cited above constitute a necessary initial step in developing a better understanding expert judgment.

6. References

Arkes, H., Herren, L., & Isen, A. M. The role of potential loss in the influence of affect on risk-taking behavior. *Organizational Behavior and Human Decision Processes, 42*, 181-193.

Arkes, H. R., and Hammond, K. R., (Eds.) (1986). *Judgment and decision making: An interdisciplinary reader.* New York: Cambridge University Press.

Bar-Hill, M. (1980). The base rate fallacy in probability judgments. *Acta Psychologica, 44*, 211-233.

Becker, G. M., & McClintock, C. G. (1967). Value: Behavioral decision theory. *Annual Review of Psychology, 18*, 239- 286.

Betaque, N. E., & Gorry, A. (1971). Automating judgmental decision-making for a serious medical problem. *Management Science, B17*, 421-434.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes, 36*, 391-485.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 281-294.

Coombs, C. H., & Lehner, P. E. (1984). Conjoint design and analysis of the bilinear model: An application to judgments of risk. *Journal of Mathematical Psychology, 28*, 1-42.

Davidson, D., Suppes, P., & Siegel, S. (1956). *Decision making: An experimental approach.* Stanford: Stanford University Press.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51,* 380-417.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69,* 109-135.

Einhorn, H. (1986). Expert judgment: Some necessary conditions and an example. In H. R. Arkes and K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader.* New York: Cambridge University Press.

Einhorn, H., & Hogarth, R. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92,* 433-461.

Handa, J. (1977). Risk, probabilities, and a new theory of cardinal utility. *Journal of Political Economy, 85,* 97-122.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association, 70,* 271- 289.

Hogarth, R. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* 197-217.

Hogarth, R. M., & Einhorn, H. (1988). Venture theory: A model of decision weights. Center for Decision Research, University of Chicago.

Isen, A. M., & Daubman, K. A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology ,47,* 1206-1217.

Isen, A. M., & Geva, N. (1987). The influence of positive affect on acceptable level of risk and thoughts about losing: The person with a large canoe has a large worry. *Organizational Behavior and Human Decision Processes, 39,* 145-154.

Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). The influence of positive affect on the subjective utility of gains and losses: It's not worth the risk. *Journal of Personality and Social Psychology, 55,* 710-717.

Isen, A. M., & Patrick, R. (1983). The effect of positive feelings on risk-taking: When the chips are down. *Organizational Behavior and Human Performance, 31,* 194-202.

Isen, A. M., & Simmonds, S. F. (1973). The effect of feeling good on a helping task that is incompatible with good mood. *Social Psychology* (now *Social Psychology Quarterly*), *41,* 345-349.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrika, 47,* 263-291.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz, (Ed.). *Advances in Experimental Social Psychology, Vol. 20,* 255-295.

Luce, R. D. (1988). Rank-dependent subjective expected utility representations. *Journal of Risk and Uncertainty, 1,* 1988, 305-332.

Luce, R. D., & Narens, L. (1986). Classification of concatenation measurement: Structures according to scale type. *Journal of Mathematical Psychology, 29,* 1985, 1-72.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252-283.

Mischel, W., Ebbesen, E., & Zeiss, A. (1976). Selective attention to the self: Situational and dispositional determinants. *Journal of Personality and Social Psychology*, *27*, 129-142.

Murphy, A. H., & Winkler, R. L. (1974). Credible interval temperature forecasting: Some experimental results. *Monthly Weather Review*, *102*, 784-794.

Nygren, T. E. (1981). Examination of a context effect on subjective estimates of probabilities. Paper presented at the Judgment and Decision Group meetings, Philadelphia, PA, November, 1981.

Nygren, T. E., & Isen, A. M. (1985). Effect of positive affect on judgments of likelihoods of events and on gambling behavior. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL, May, 1985.

Nygren, T. E., & Isen, A. M. (1985). Examining probability estimation: Evidence for dual subjective probability functions. Paper presented at the Psychonomic Society Meetings, Boston, MA, November, 1988.

Nygren, T. E., & Morera, O. (1988). Davidson, Suppes, and Siegel revisited: Evidence for a dual bilinear model. Paper presented at the Society for Mathematical Psychology Meeting, Northwestern University, July 1988.

Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays.* Harcourt, Brace: New York.

Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology*, *2*, 191-201.

Savage, L. J. (1954). *The foundations of statistics.* Wiley: New York.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 29,* 1-39.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions. *Science, 211,* 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Wallsten, T. S., & Budescu, D. V. (1986). Encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29,* 151-173.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R. and Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348-365.

Wickins, C. D., Stokes, A., Barnett, B., & Davis, T. (1987) Componential analysis of pilot decision making. USAF Technical Report, June 1987.

1986 USAF-UES RESEARCH INITIATION PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

Prepared by: Jacqueline G. Paver

Academic Rank: Research Assistant Professor

Department: Biomedical Engineering Department

University: Duke University

Research Location: Duke University and AAMRL/BBM, Wright-Patterson Air Force Base

USAF Researcher: Dr. Ints Kaleps

Date: March 23, 1989

Contract No.: F49620-85-C-0013/SB5851-0360

1. INTRODUCTION

In recent years, there has been an increasing awareness of the serious consequences that can occur with head and neck injuries and the effectiveness of biomechanical studies to reduce the likelihood of these injuries. Each year, more than 140,000 Americans die from trauma. Approximately 80,000 other Americans suffer unnecessary, but permanently disabling, injuries of the brain and spinal cord. Highway-related injuries have stimulated a considerable number of studies leading to the development of crashworthy structures, restraint systems, and helmets. The incidence of injuries during ejection from high-speed aircraft has also generated interest in the study of trauma biomechanics for military personnel.

Information about head and neck injury and prevention is generated from several sources: (1) studies of actual accidents involving humans; (2) experiments with human vounteers; (3) experiments with human surrogates; and (4) accident simulations using mathematical models. Each of these sources has its advantages and disadvantages. Since trauma victims are rendered unconscious or dazed and since eyewitness reports are rarely complete and often contradictory, information from accident investigations and field studies is usually limited. Volunteers must be subjected to loads considerably below tolerance levels and may vitiate normal responses by anticipatory action. Cadavers can be subjected to loads beyond the onset of serious trauma but their properties may differ from that of the living human, particularly when preservatives have been used. Animals do not exhibit the same anatomical features as humans and the validity of scaling trauma data for the head-neck region has not been firmly established. Poor repeatability of results from subject to subject is not uncommmon. The use of inanimate devices reduces the repeatability problems associated with cadavers and live animals but raises questions as to biofidelity. Attempts to improve human surrogate biofidelity are well documented. When a new manikin design is incorporated into a crash test program, data bases are generated which describe the geometric and inertial properties and mechanical behavior of these manikins. Compliance tests are developed and performance evaluations are conducted to ensure uniformity of results among different laboratories. Mathematical modeling is also an accepted technique of scientific research. It is a process in which an actual physical system, consisting of elements with properties that interact according to certain relationships, is simplified by making "reasonable" assumptions and idealizations to obtain a representative model. Once validated by comparison with experimental results, mathematical models are useful, economical, and versatile engineering tools. They can, in lieu of direct experimentation with actual physical systems, evaluate the effects of varying parameters on the responses of systems to a wide variety of input conditions. In particular, with validated human and manikin data sets, laboratory tests and actual accident environments, where the head and neck are in jeopardy, can be simulated.

One goal of the research program at the Armstrong Aeromedical Research Laboratory (AAMRL) is to measure, analyze, and simulate the kinematic and dynamic head-neck responses of existing anthropomorphic dummies in crash environments in order to develop an improved biofidelic manikin head-neck system. To accomplish this goal, theoretical and experimental studies were performed by Paver and Doherty[1-3] with AAMRL.

The mathematical models utilized for this research were the Head-Spine Model (HSM) and the Articulated Total Body (ATB) Model. The Hybrid II and Hybrid III manikins were selected as the test specimens. HSM and ATB data sets, which accurately predict the mechanical behavior of the Hybrid II manikin head-neck structure, were developed and validated. The Part 572 Head-Neck Compliance Test[4], of the Code of Federal Regulations, was simulated to validate these data sets. Parametric studies were conducted to assess the effects of variations in neck characteristics upon the predicted head-neck responses. Modifications of the HSM program source code were made in order to model the asymmetric bending properties of the Hybrid III manikin neck in flexion and in extension. HSM data sets were also developed for the Hybrid III head-neck structure. The effect of the occipital condyle nodding block joint on head-neck kinematics and dynamics was investigated. The CFR Amended Part 572 Head-Neck Compliance Test[4] was simulated to validate these data sets.

As part of the SFRP and RIP, Paver and Doherty conducted experimental studies of the Hybrid II, modified Hybrid II, and Hybrid III dummy head-neck assemblies and the Part 572 and Amended Part 572 head-neck compliance tests. A test plan was written. Initially, the standard head-neck compliance tests were to be performed to verify the operation of the experimental setup. Tests at different impact velocities were planned to validate data sets for mathematical models. Measurements were made of the geometric and inertial properties of the test specimens and apparatus. Mounting plates, test fixtures, and apparatus modifications were made. All transducers were wired; some were calibrated. The test specimens and apparatus were assembled. The velocimeter was modified and calibrated. Attempts were made to determine the drop heights required for the specified impact velocities. Preliminary measurements of pendulum strike plate deceleration pulses were made; the effects of honeycomb precrush were investigated. Several recommendations for modifications of the experimental system were suggested.

Hopefully, the information generated by this study will help both in the understanding of the head-neck biomechanics of existing manikins and in future manikin development efforts. In particular, the proposed data sets will be useful to researchers modeling experiments with these dummies. Ideally, these experiments and computer simulations will help us understand the responses of humans in similar crash environments. It is also hoped that the results of this research will improve the head and neck injury prediction potential of the HSM and ATB models. The data set development and parametric studies will serve as preliminary efforts toward the development and validation of human data sets with accurate neck descriptions. The human data sets will be based upon the results of viscoelastic studies of the biomechanical properties of human cadaver spines[5].

2. BACKGROUND

2.1 Mathematical Modeling

A variety of mathematical models have been developed to predict human biodynamic responses and injury potential in impact environments. The mathematical models utilized for this research were the Head-Spine Model (HSM) and the Articulated Total Body (ATB) Model. The HSM and ATB models were developed by AAMRL at Wright-Patterson Air Force Base.

2.1.1 The Head-Spine Model

The HSM is a three-dimensional matrix structural analysis computer program[6]. It is an internal body structure type model, which deals with the detailed structure of various body subsystems and predicts stresses, strains, localized deformations, and injury potential. The human body is represented by a collection of rigid bodies connected by deformable elements. The deformable elements can be beam elements, spring elements, hydrodynamic elements, or elastic surfaces. The rigid bodies generally represent bones (e.g., the head, vertebrae, pelvis, and ribs); the deformable elements represent soft tissues (e.g., intervertebral discs, ligaments, and viscera). The program integrates the equations of motion in time, either implicitly or explicitly. The analysis accommodates large displacements and small strains of the rigid bodies, nonlinear stiffnesses, and viscous (percent critical) and structural (stiffness proportional) damping forces.

The data set defines the structure to be modeled. In general, an HSM data set consists of six types of inputs: (1) the coordinate system definitions; (2) the geometry of the nodes and connectivity data; (3) the boundary conditions and the constraints on the motion; (4) the inertial properties of the rigid bodies and deformable elements; and (5) the material properties of the deformable elements. The excitation and initial conditions are defined in the ICIF subroutine, which is linked with the code prior to execution.

2.1.2 The Articulated Total Body Model

The ATB Model is a modified version of the Calspan Crash Victim Simulator (CVS) Model[7–10] which accomodates several Air Force applications[11] such as encumberance effects on crewman performance, vibration loading, and ejection from disabled aircraft (e.g., retraction, head-canopy impacts, windblast, and parachute-opening shock). The model is a whole-body gross-motion simulator model. The equations of motion and constraint are formulated from Euler's rigid body equations of motion with Lagrange-type constraints. These equations are solved using a Vector Exponential Integrator and quaternions. This model differs from most other three-dimensional occupant models, which are formulated from Lagrange's equations of motion. In most applications, the crash victim, shown in Figure 2.1, is represented by fifteen segments and fourteen joints; the neck joints are typically ball-and-socket types with three degrees of freedom.

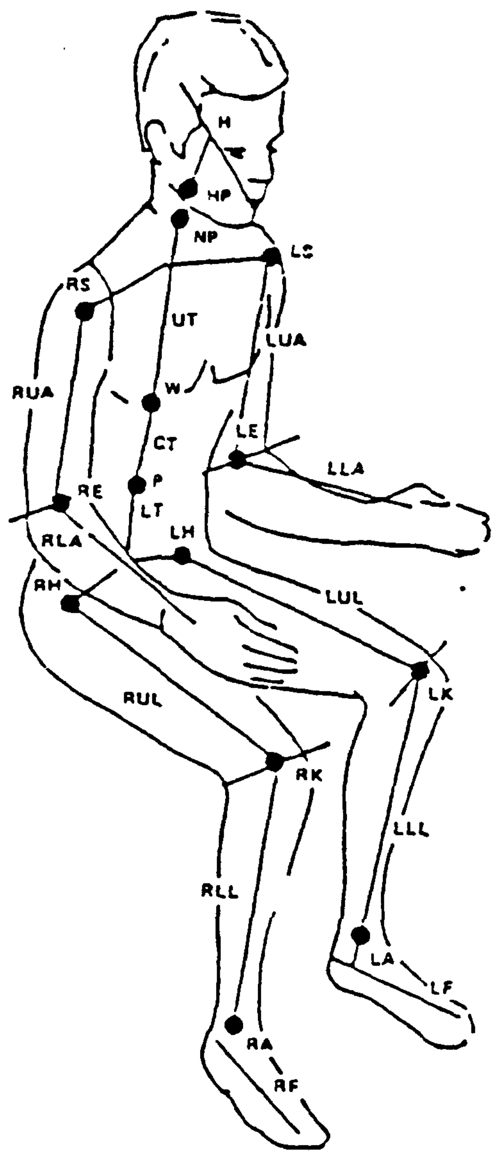


Figure 2.1

Three-Dimensional Head/Neck Injury Model Crash Victim (Abstracted from Bartz, 1972)

2.2 Experimental Aspects

2.2.1 Test Specimens

The Hybrid II and Hybrid III manikin head-neck assemblies were selected as the test specimens for this study.

The Hybrid II manikin was the first GM design to have acceptable repeatability and good durability and serviceability[12,13]. The head is a hollow two-piece aluminum casting. It is instrumented with a triaxial array of uniaxial accelerometers mounted at its center-of-gravity (cg). Both the skull and rear cap are covered with a rubber skin. The neck is a right circular cylinder constructed of butyl rubber elastomer. It is solid, except for a small hole through the middle. Aluminum plates are molded into each end to facilitate head-neck and neck-thorax attachment.

The Hybrid III manikin is a state-of-the-art biofidelic anthropomorphic dummy[14,15]. The head is a hollow two-piece casting of 356-T6 aluminum with a vinyl skin cover. The neck, which consists of aluminum and asymmetric butyl rubber layers with a central steel cable, exhibits different mechanical responses in flexion and extension. The cable provides axial strength. A load cell and occipital condyle joint, with butyl rubber nodding blocks, connect the head and neck. An adjustable steel bracket connects the neck and pendulum.

2.2.2 The Part 572 Compliance Tests

All Hybrid II and III compliance test procedures and performance standards are described in the Code of Federal Regulations[4], Title 49, Part 572 and Part 572 Amended, respectively. The environmental conditions, test apparatus (geometric and inertial properties), instrumentation requirements, and test procedures are specified.

The head-neck test consists of a pendulum drop (see Figure 2.2). At the bottom of the pendulum's swing, the arm impacts a block of honeycomb; this produces a near square wave pendulum deceleration pulse. The head-neck system, which is mounted to the end of the pendulum, undergoes no impact.

Only a neck flexion test is required by Part 572. The impact velocity and acceleration pulse at the pendulum strike plate are specified (i.e., a range for the average deceleration level above 20g's and minimums for the slopes of the deceleration curve between 5 and 20g's). There are three ways that the head-neck response is specified: (1) head rotation vs. time; (2) chordal displacement of the head cg vs. time; and (3) peak head acceleration and average head acceleration above 20 g's. The chordal displacement at time t is defined in Figure 2.3; it is the straight line distance between (1) the position of the head cg relative to the pendulum arm at time zero (the dashed lines); and (2) the position of the head cg relative to the pendulum arm at time t (the solid lines).

Both neck flexion and neck extension tests are required by Part 572 Amended. For these tests, the standard specifies the maximum pendulum strike plate deceleration level and deceleration level ranges at 10, 20, and 30 msec. The specifications for the head-neck mounting position and the pendulum strike plate deceleration pulse and impact velocity differ for flexion and extension. The pendulum strike plate deceleration level maximum and impact velocity are lower for the extension test than for the flexion test. There are two ways that the head-neck response is specified: (1) head rotation vs. time; and (2) moment about the occipital condyles vs. time. Ranges of amplitudes and times are specified for the peak values and ranges of times are specified for the zero crossings.

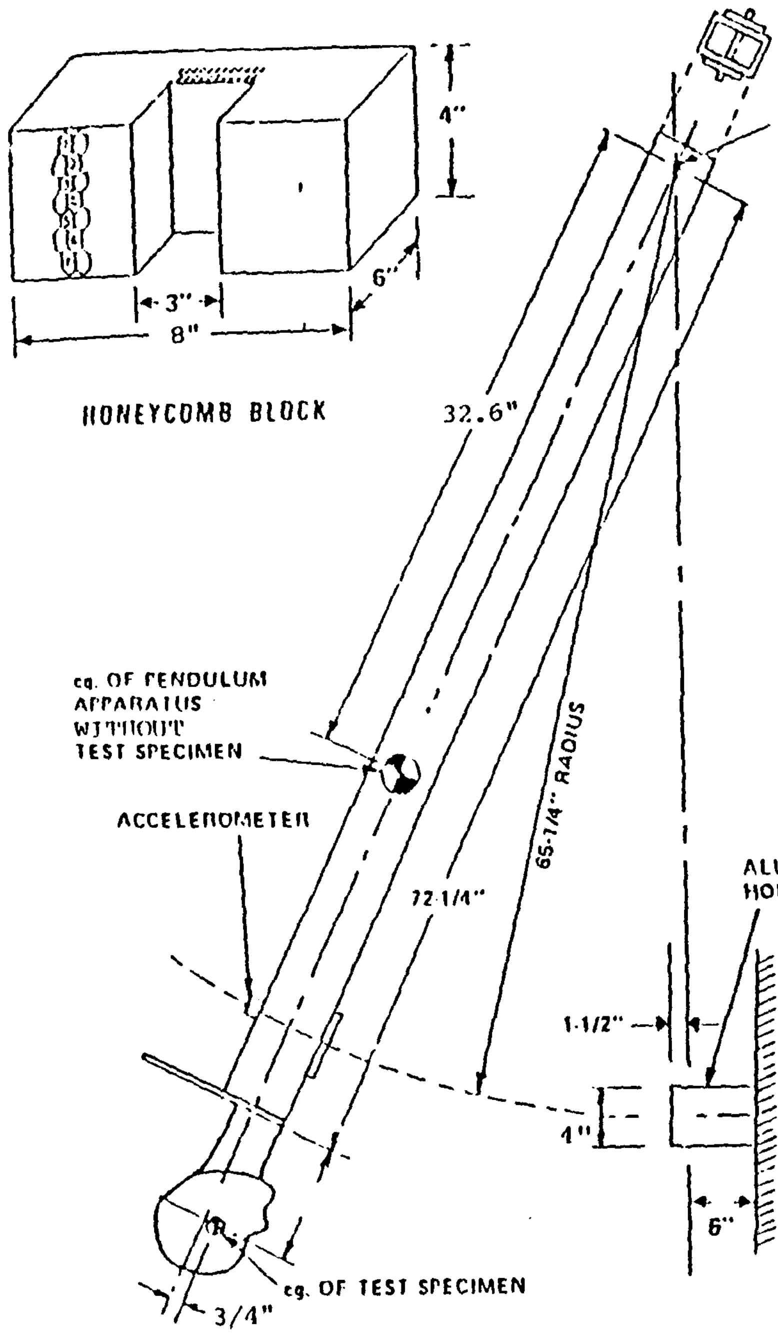
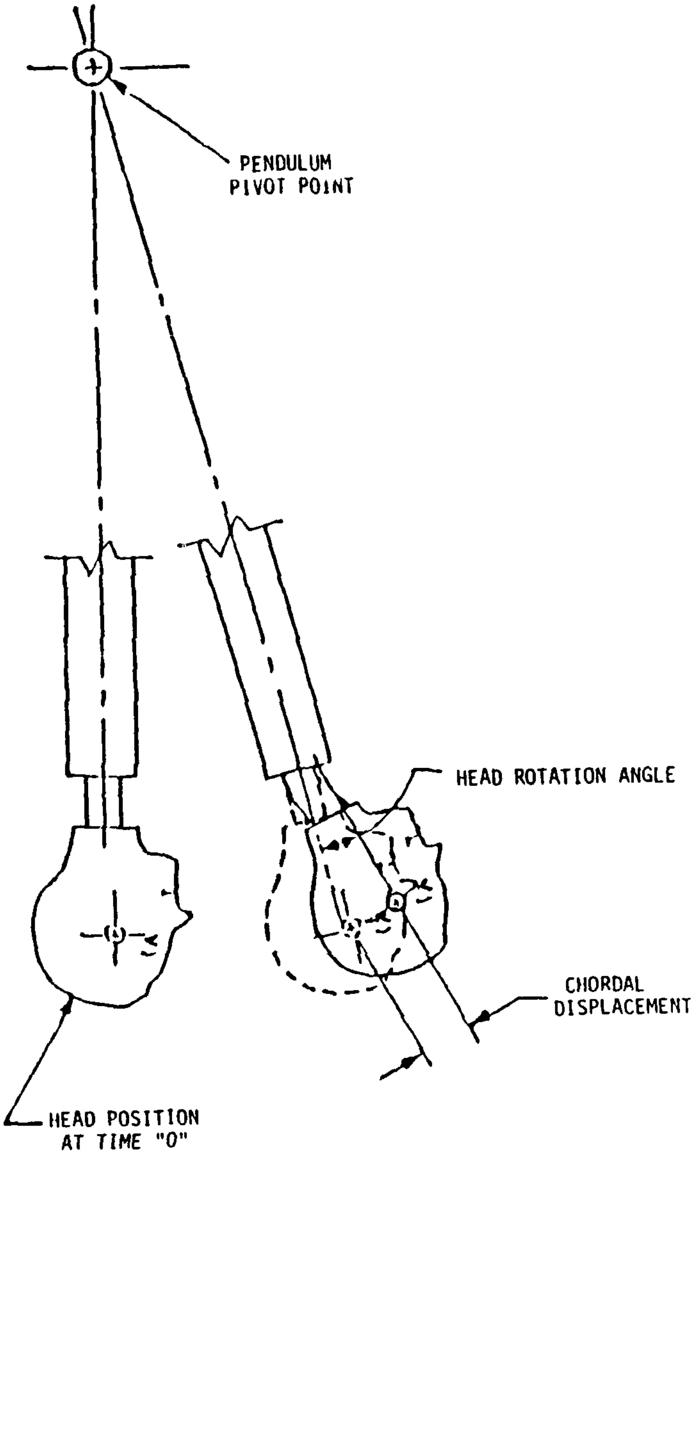


Figure 2.2: Pendulum Test Apparatus

Figure 2.3: Response Parameters for Part 572



3. THEORETICAL STUDIES

Theoretical studies were conducted of the Hybrid II and III dummy head-neck assemblies and the Part 572 and Amended Part 572 head-neck compliance tests. Detailed anthropometric and inertial test results and Part 572 and Amended Part 572 test specifications and performance standards were analyzed to determine inputs for the HSM and ATB Models.

HSM and ATB data sets, which accurately predict the mechanical behavior of the Hybrid II manikin head-neck structure, were developed and validated. The Part 572 Head-Neck Compliance Test[4], of the Code of Federal Regulations, was simulated to validate these data sets. Parametric studies were conducted to assess the effects of variations in neck joint characteristics upon the predicted head-neck responses.

Modifications of the HSM program source code were made in order to model the asymmetric bending properties of the Hybrid III neck in flexion and in extension. Two HSM data sets were developed for the Hybrid III manikin head-neck structure. One data set utilized a 1-element neck; the second data set utilized a 4-element neck. The effect of the occipital condyle nodding block joint on head-neck kinematics and dynamics was also investigated and modeled. Static and dynamic tests were conducted to measure the stiffness of this joint in flexion and extension. A data set with a 1-element neck and nodding joint was developed. The CFR Amended Part 572 Head-Neck Compliance Test[4] was simulated to validate these data sets. Efforts to develop, tune, and validate a Hybrid III data set for the ATB model are currently in progress.

3.1 Head-Spine Model

3.1.1 Hybrid II Data Set Development

The coordinate system was defined so that the positive x axis was the A-P direction, the positive y axis was the L-R direction, and the positive z axis was the S-I direction. The pendulum arm was free to rotate about the y axis only. Since the response is measured only from the time of impact to the time the head returns to the pre-impact position, the end of the pendulum (i.e., the base point) does not rotate during the test. The motion of this point was modeled as a pure translation in the x direction; no translations in the y and z directions or rotations about the x or z axes were allowed for this point. All other points were allowed to translate and rotate in the x-z plane.

The HSM data set was made up of two elements (Figure 3.1). A rigid body represented the head; a spinal disc beam deformable element represented the neck. The beam element was a Kelvin solid with linear stiffness and viscous damping. The pendulum arm was not explicitly defined; the boundary conditions of the node at the base of the neck were made to reflect the presence of the pendulum.

The geometric and inertial properties of the rigid body were calculated by lumping the upper half of the neck with the head. Since the Part 572 specification requires that the head and neck properties be defined by the SAE Recommended Practice J963, in which the head and neck are lumped together differently, this data was inappropriate for this simulation. Instead, data from Hubbard and McLeod[13], where the geometric and inertial properties of the head are described separately from those of the neck, were utilized.

A near square-wave strike plate acceleration pulse (Table 3.1) was the excitation for the model. This pulse complied with Part 572. The initial conditions were the velocities of the primary nodes, which were calculated from the pendulum strike plate impact velocity specified in Part 572. These two quantities were not included in the data set itself; they were inputs to the subroutine ICIF. This subroutine is called by the main program after the data set is read.

The response of the model did not comply with Part 572 using these initial values; it was necessary to increase the amplitude and decrease the period of the response. The data set was tuned to make the HSM head-neck responses comply with the Part 572 performance standards. Since the geometric and inertial properties of the Hybrid II head and neck are well documented, these constants remained fixed; they were not used to tune the data set. The material properties of the neck, however, are not well documented. The values used initially for the bending, axial, and torsional stiffness and damping and the bending shear parameter were estimated from a material specification abstracted from a Sierra blueprint[16] and from static test data[12,17]. The material properties of the neck were varied in a systematic manner to optimize the model response with Part 572 specifications.

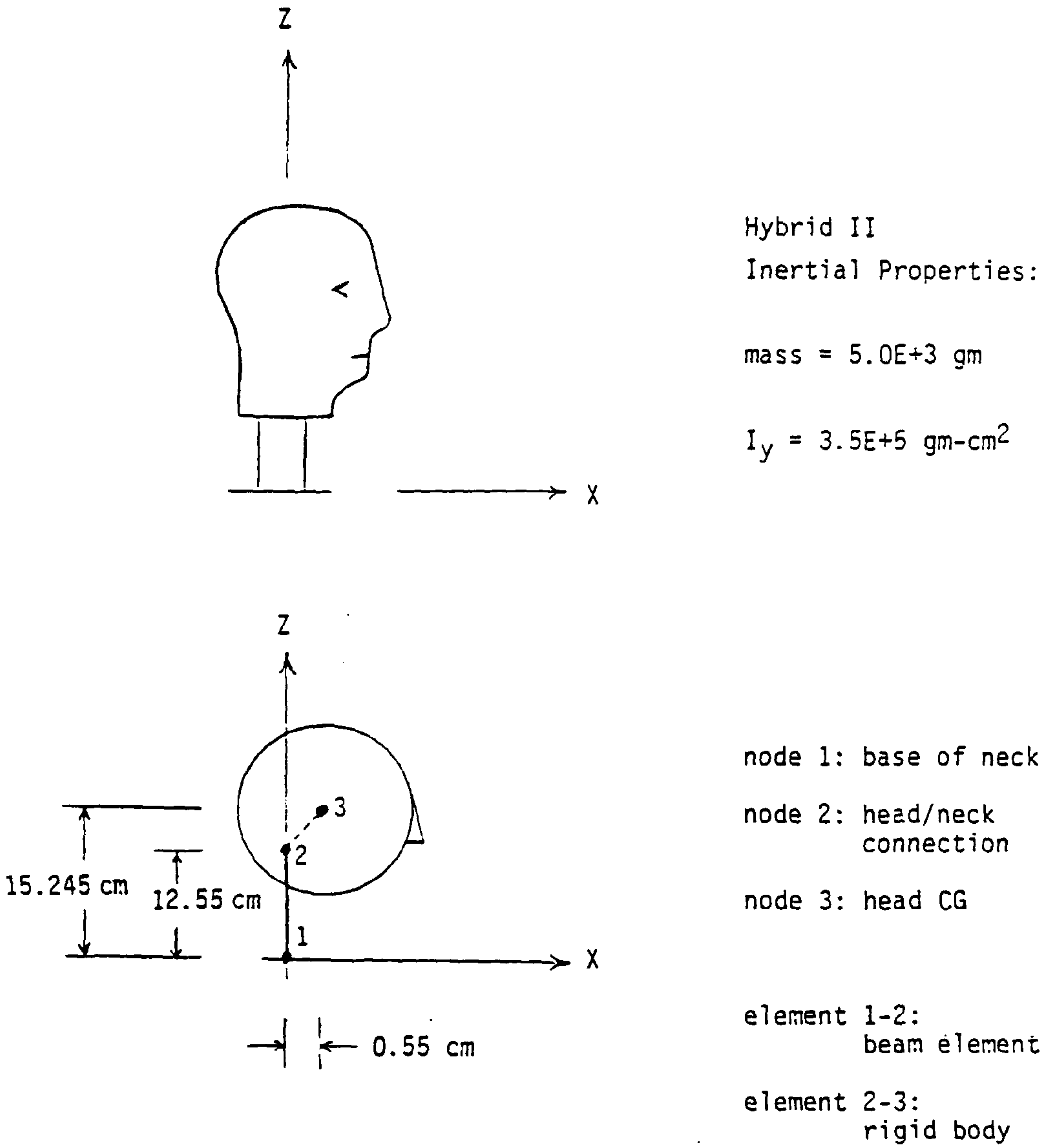


Figure 3.1: HSM Hybrid II Data Set Development

3.1.2 Hybrid II Simulation Results

Figures 3.2-7 show the results of the parameter tuning process. Plots labeled 572 illustrate the final data set results. Plots labeled A illustrate the predicted responses when a higher bending stiffness value is used; plots labeled B illustrate the predicted responses when a lower bending stiffness value is used. Plots labeled C illustrate the predicted responses when a higher bending damping value is used. Plots labeled D illustrate the predicted responses when a higher shear parameter value is used. Table 3.2 shows the values used for the selected parameters.

Listings of the final(572) input data set and relevant outputs are included in Appendix A. The final data set predicted head rotation and peak head acceleration data which complied with Part 572 specifications. The predicted peak chordal displacement was slightly lower than that specified by Part 572.

It was discovered that an increase in the bending stiffness decreased both the amplitude and the period of the head rotation (Figure 3.2) and chordal displacement (Figure 3.3). Increasing the bending stiffness also decreased the peak head acceleration (Table 3.2). An increase in the bending damping from 1×10^{-5} to 1×10^{-2}% critical had no effect on the predicted performance of the Hybrid II head-neck assembly. An increase to 1% critical, however, lowered the amplitude and increased the period of the head rotation (Figure 3.4) and chordal displacement (Figure 3.5). Increasing the bending damping also decreased the peak head acceleration (Table 3.2). An increase in the bending shear parameter increased the amplitude and delayed the head rotation (Figure 3.6), increased both the amplitude and period of the chordal displacement (Figure 3.7), and increased the peak head acceleration (Table 3.2). Changes in the axial and torsional stiffness and damping over several orders of magnitude had *no effect on the predicted performance of the head-neck assembly.*

Table 3.1: Acceleration Pulse Data

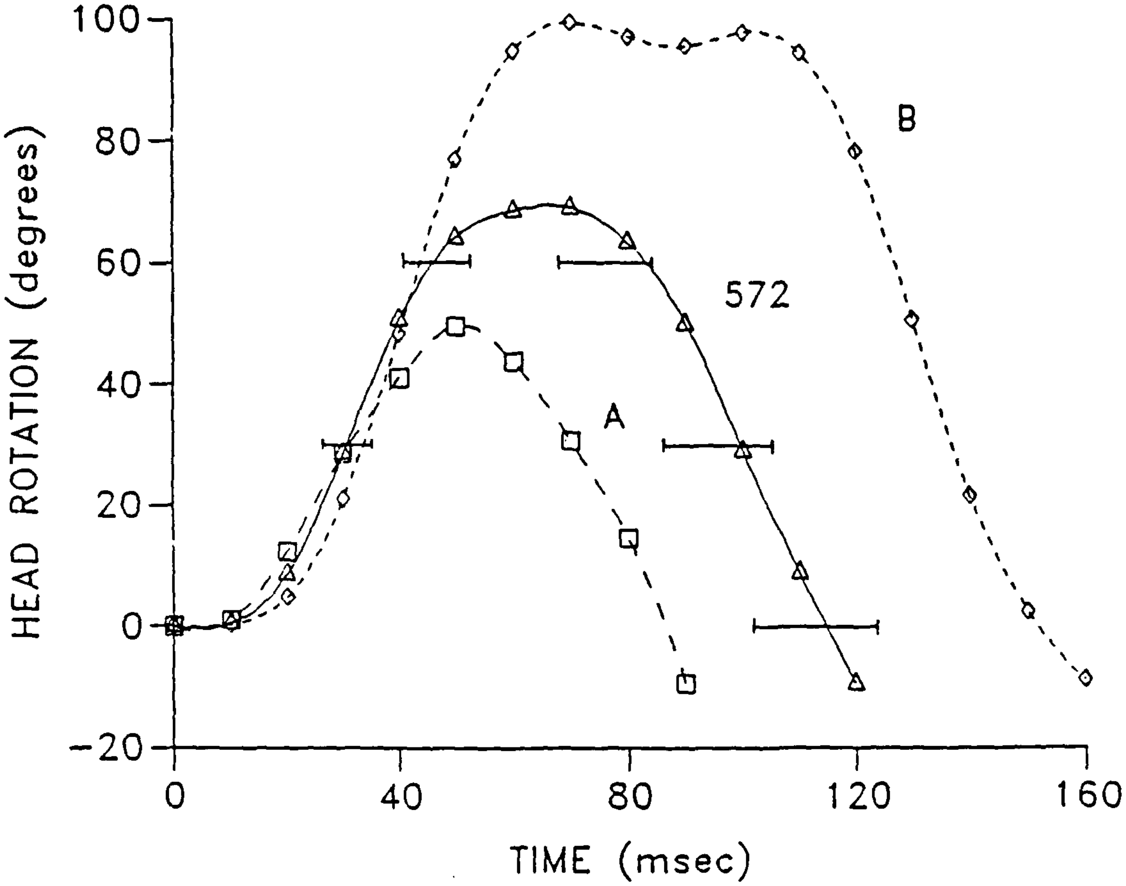
TIME (sec)	ACCELERATION (g's)	ACCELERATION-TIME SLOPE (g's/sec)
0.0	0	0.
0.0025	-5	-8.
0.0175	-26	-4.
0.005	-26	0.
0.012	-22	0.
0.020	-22	0.
0.02933	-22	2.
0.0335	-5	4.
0.03475	0	2.667
0.036	0	0.
0.063	0	0.
0.068	0	0.
0.073	0	0.
0.08	0	0.
0.085	0	0.
0.09	0	0.
0.1	0	0.

Table 3.2: Variations in Bending Stiffness, Damping, and Shear Parameter and Peak Head Acceleration Response

LABEL	STIFFNESS (dyne/cm)	DAMPING (% critical)	SHEAR PARAMETER	PEAK HEAD ACCELERATION (g's)
572	7.7 E+8	1E-5 to 1E-2	0	23.53
A	15.4 E+8	1E-5	0	21.34
B	3.85 E+8	1E-5	0	26.50
C	7.7 E+8	1E+0	0	20.63
D	7.7 E+8	1E-5	5	27.21

Figure 3.2: Effect of Bending Stiffness Variations on Head Rotation

Figure 3.3: Effect of Bending Stiffness Variations on Chordal Displacement


HEAD ROTATION (degrees)
TIME (msec)
B
572
A
100
80
60
40
20
0
-20
0
40
80
120
160

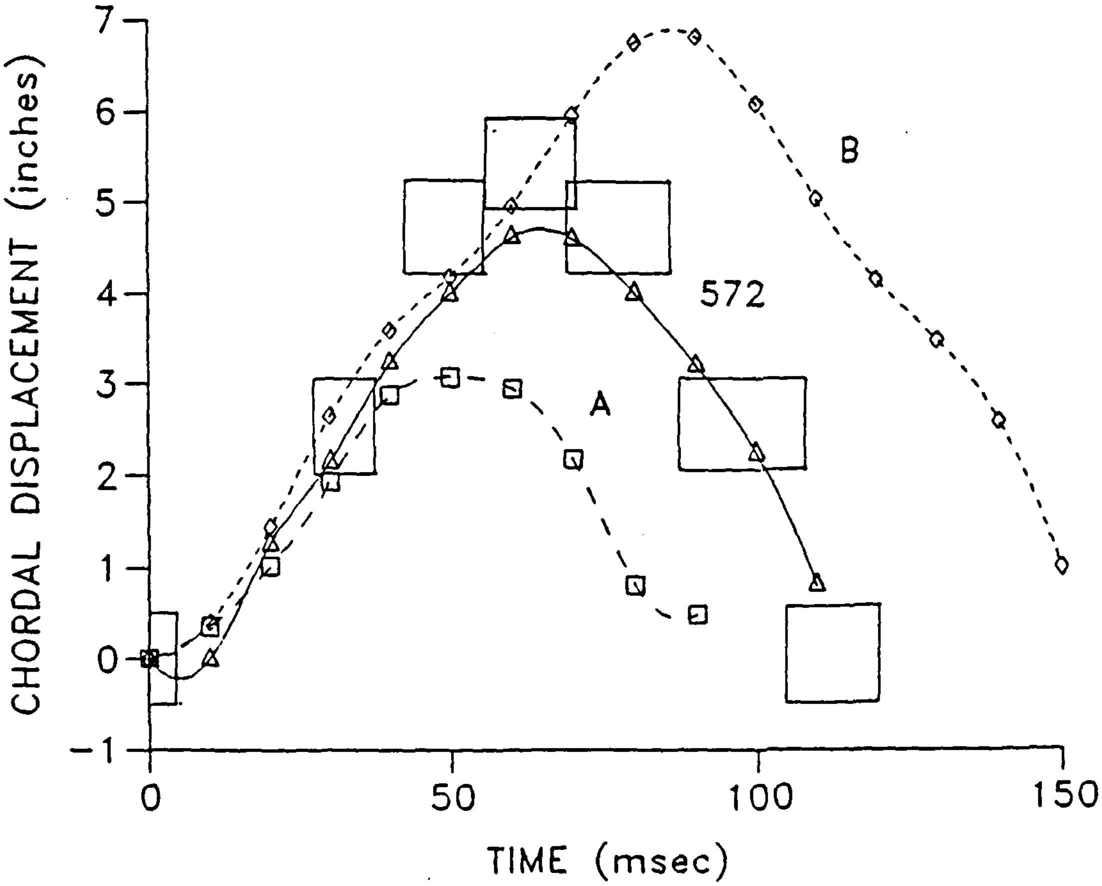
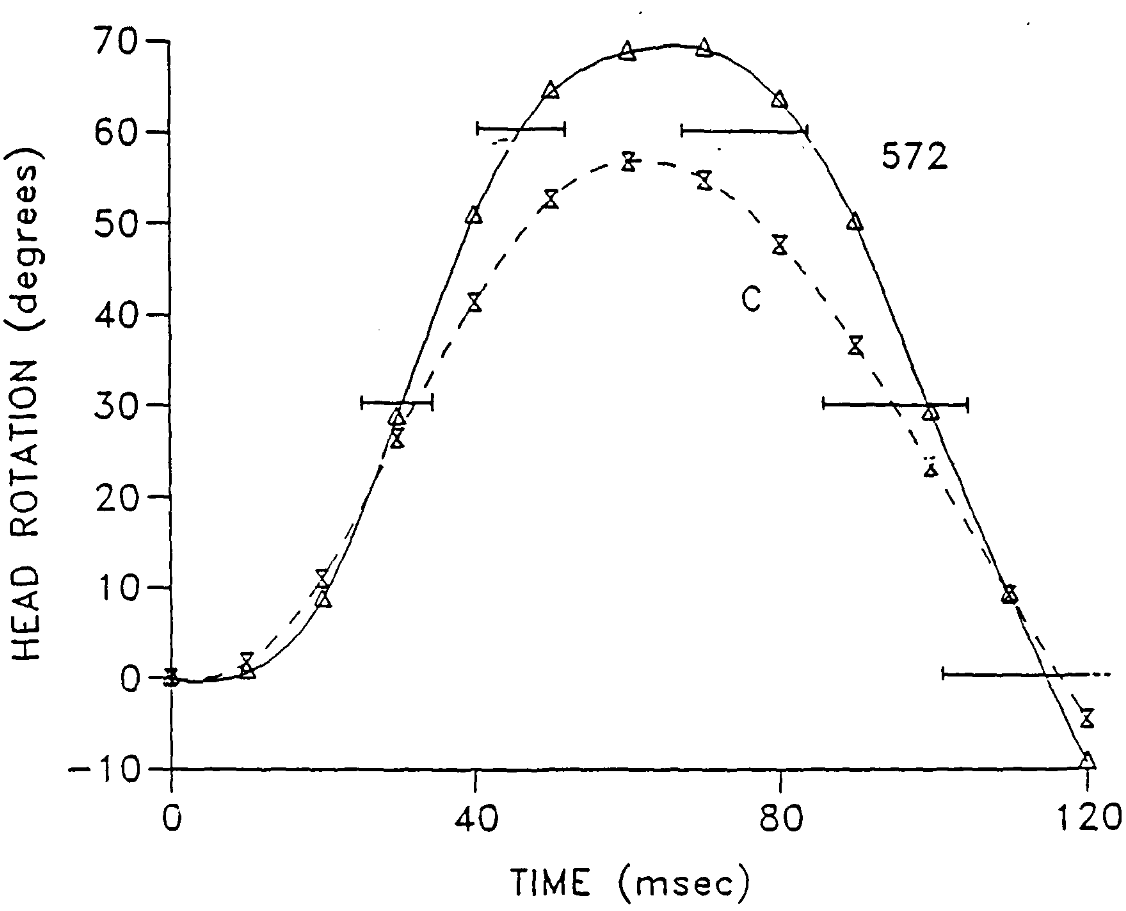
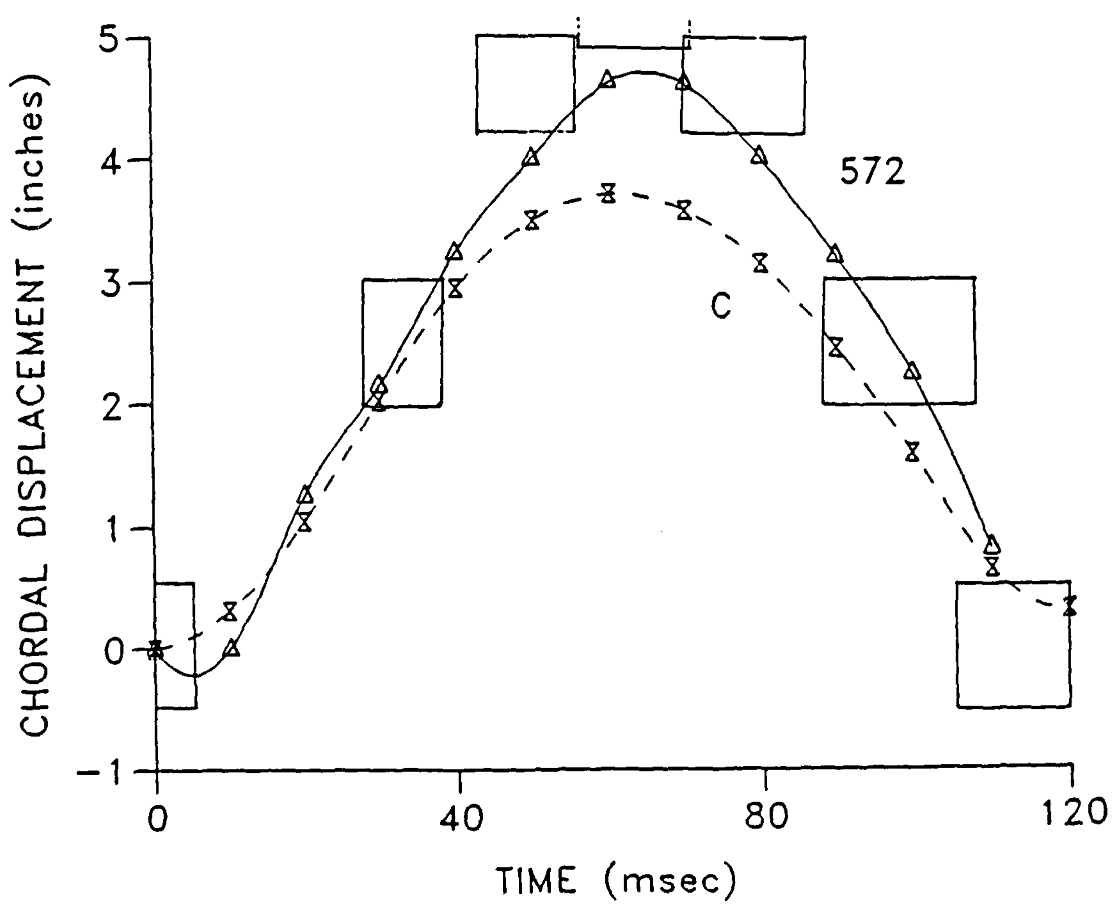

CHORDAL DISPLACEMENT (inches)
TIME (msec)
B
572
A
7
6
5
4
3
2
1
0
-1
0
50
100
150

Figure 3.4: Effect of Bending Damping Variations on Head Rotation

Figure 3.5: Effect of Bending Damping Variations on Chordal Displacement



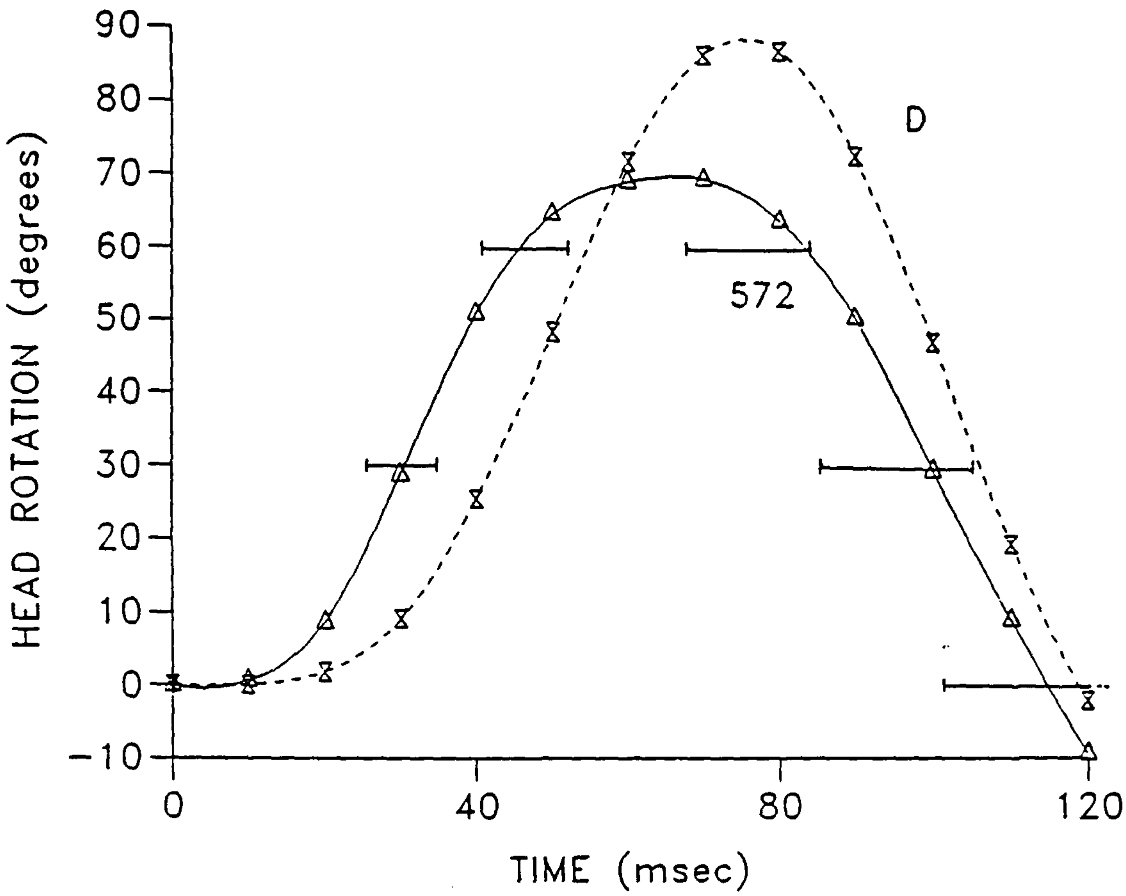
HEAD ROTATION (degrees)
70
60
50
40
30
20
10
0
−10
0
40
80
120
TIME (msec)
572
C



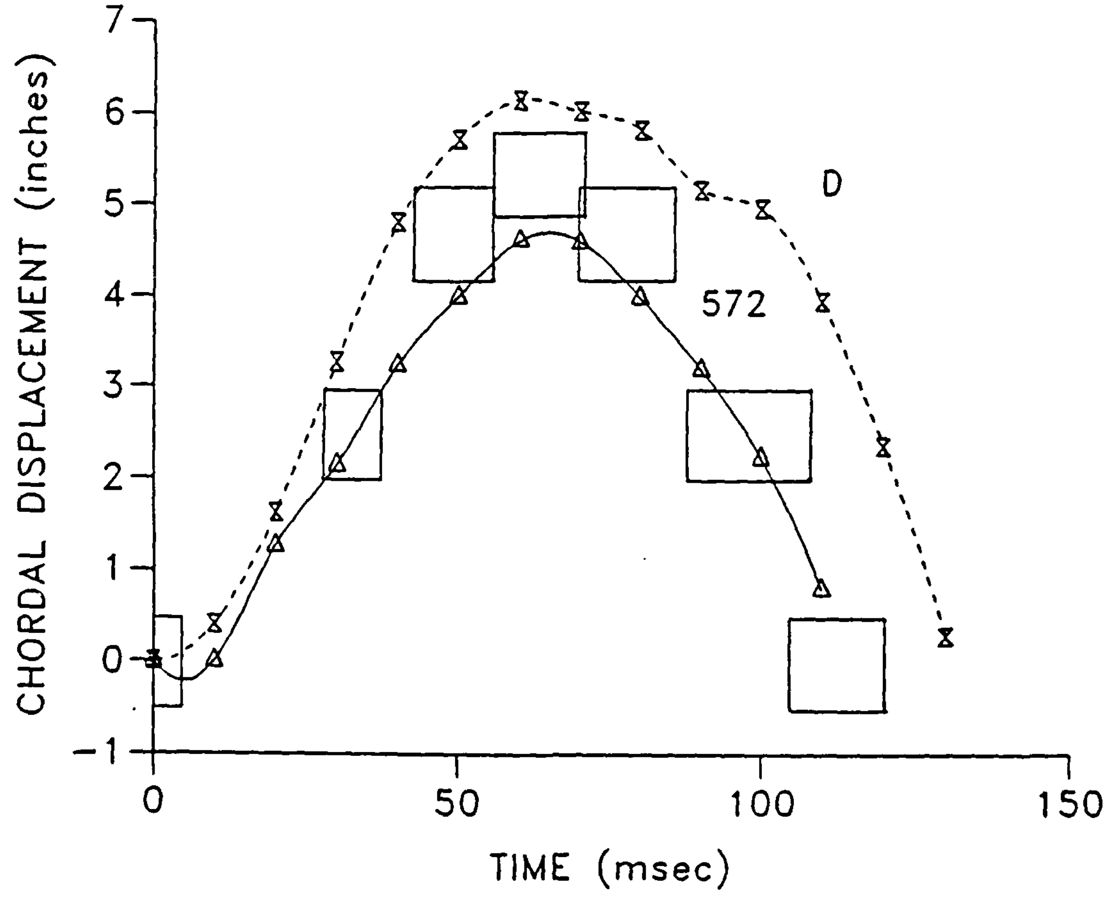
CHORDAL DISPLACEMENT (inches)
5
4
3
2
1
0
−1
0
40
80
120
TIME (msec)
572
C

Figure 3.6: Effect of Bending Shear Parameter Variations on Head Rotation

Figure 3.7: Effect of Bending Shear Parameter Variations on Chordal Displacement



D
572
HEAD ROTATION (degrees)
TIME (msec)



D
572
CHORDAL DISPLACEMENT (inches)
TIME (msec)

3.1.3 Program Modifications Required for Hybrid III Simulations

Modifications were made in the program source code in order to model the asymmetric bending properties of the Hybrid III manikin neck in the A-P plane. Different stiffnesses can be input for positive and negative bending and positive and negative torsion of the beam elements. In a similar way, different stiffnesses can be input for positive and negative torsion of the beam elements.

3.1.4 Hybrid III Data Set Development

Figure 3.8 shows the discretization of the Hybrid III head-neck structure. The coordinate system was defined so that the positive x, y, and z axes were the A-P, L-R, and S-I directions, respectively. During an actual test, the pendulum arm only rotates about the y axis. Since the response is specified from the time of impact to the time the head returns to the pre-impact position, it was assumed that the end of the pendulum (i.e., node 1) does not rotate. The motion of this point was modeled as a pure translation in the x direction; no translations in the y and z directions or rotations about the x or z axes were allowed for this point. All other points were allowed to translate and rotate in the x-z plane.

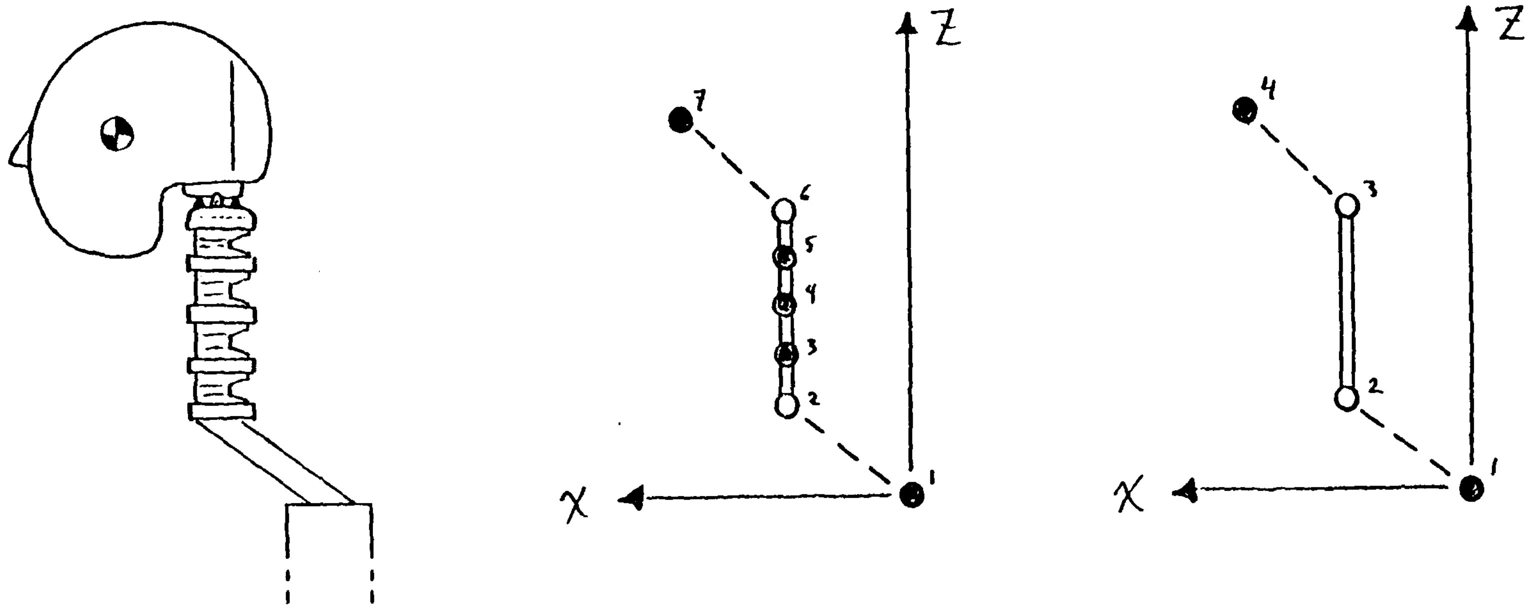
As a preliminary effort, the details of the construction of the Hybrid III neck were not explicitly modeled. The head was modeled as a rigid body; the neck was modeled as a single spinal disc beam deformable element (i.e., a Kelvin solid with linear stiffness and viscous damping). The dark circles show the locations of the lumped masses of the rigid bodies (i.e., the primary nodes). The open circles show the locations of the secondary nodes. The dotted lines represent rigid links between the primary and secondary nodes. The shaded bar represents the deformable element. The pendulum was not explicitly defined. The boundary conditions at node 1 were made to reflect the presence of the pendulum and the constraints of the experimental design.

The geometric and inertial properties of the head and neck were derived from GM design specifications[14]. The values used initially for the bending stiffness and damping of the neck were derived from results of a study conducted by AAMRL and Systems Research Laboratories for the Department of Transportation[18]. Ranges for the stiffness were calculated from the measured static and dynamic stiffnesses of two Hybrid III necks. The shear parameter and the axial and torsional stiffnesses and damping of the neck were derived from the bending parameters and appropriate geometrical considerations.

Next, the layered aluminum/butyl rubber construction of the neck was incorporated into the Hybrid III data set. The 1-element neck was divided into four rigid bodies connected by beam elements. Each aluminum disc was represented by a rigid body. Each rubber disc was represented by a beam element. Masses and moments of inertia of the rubber discs were lumped at the head cg or the cg's of the aluminum discs. The bending stiffnesses and shear parameters of the 4-element neck, which are length sensitive, were scaled from values used for the 1-element neck.

Figure 3.8

DISCRETIZATION OF HYBRID III HEAD-NECK STRUCTURE



Hybrid III

4-element neck

1-element neck

The excitation for the model, applied at node 1, was the specified pendulum strike plate deceleration profile. The initial conditions, which were the velocities of the primary nodes, were calculated from the specified pendulum strike plate impact velocity.

The original data sets were tuned to make the model responses comply with the performance standards. Since the geometric and inertial properties of the Hybrid III head and neck are well documented, these constants remained fixed; they were not used to tune the data set. The material properties of the neck, however, are not well documented. The bending properties of the neck were varied within the ranges of the experimental static and dynamic test data. By varying these properties in a systematic manner, optimization of the model response with the specifications was possible (Table 3.3).

Table 3.3 Hybrid III Material Properties

1-Element Neck Stiffnesses

axial:	9.000×10^9 $dynes/cm$
torsional:	5.594×10^9 $dyne - cm$
bending:	8.000×10^8 $dyne - cm$ (flex)
	3.550×10^8 $dyne - cm$ (ext)

Other 1-Element Neck Parameters

shear parameter:	5
axial damping:	0.2%
bending damping:	0.2%

4-Element Neck Stiffnesses

axial:	875.6×10^6 $dynes/cm$
torsional:	5.594×10^9 $dyne - cm$
bending:	4.500×10^9 $dyne - cm$ (flex)
	2.000×10^9 $dyne - cm$ (ext)

Other 4-Element Neck Parameters

shear parameter:	125
axial damping:	0.2%
bending damping:	0.2%

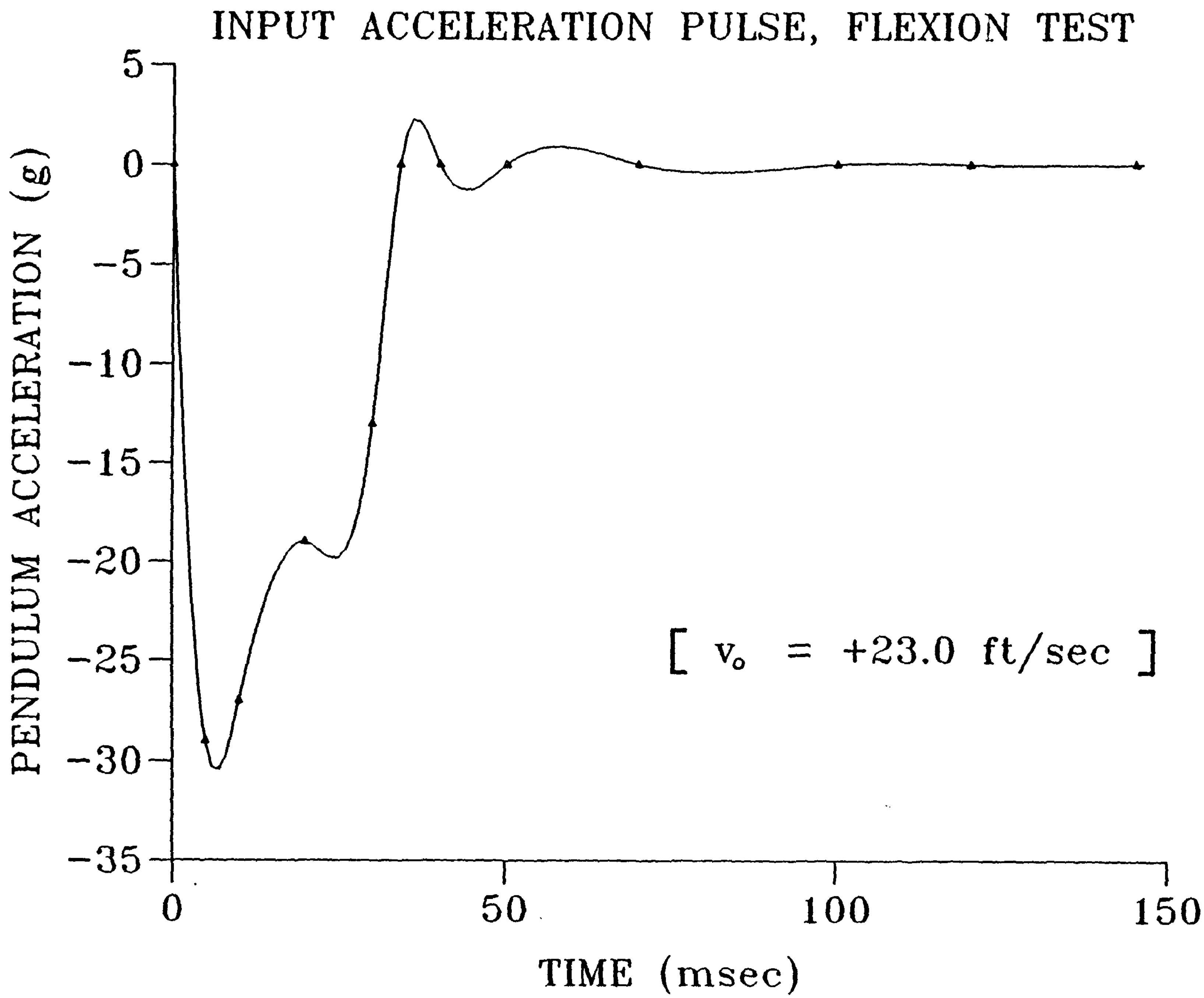
3.1.5 Hybrid III Simulation Results

Figures 3.9-3.14 show the results of the parameter tuning process. Both the flexion test and extension tests were simulated using the 1-element and 4-element neck data sets. Listings of the relevant input data sets and output files are included in Appendix B.

Figure 3.9 shows the model excitation and the specified pendulum strike plate velocity for the flexion test simulation. Figure 3.10 is a plot of the head rotation vs. time. Figure 3.11 is a plot of the moment at the occipital condyles vs. time. The dashed line is the response of the 1-element neck. The solid line is the response of the 4-element neck. Differences were observed between the responses of the 1-element and 4-element neck data sets: (1) the 4-element neck exhibited a slightly larger peak rotation than the 1-element neck; and (2) the 1-element neck exhibited a larger peak moment. The periods of rotation and moment were similar for both data sets.

Figure 3.12 shows the model excitation and the specified pendulum strike plate velocity for the extension test simulation. Figure 3.13 is a plot of the head rotation vs. time. Figure 3.14 is a plot of the moment at the occipital condyles vs. time. The dashed line is the response of the 1-element neck. The solid line is the response of the 4-element neck. Differences were observed between the responses of the 1-element and 4-element neck data sets: (1) the 4-element neck exhibited a larger peak rotation than the 1-element neck; and (2) the 4-element neck exhibited a slightly larger peak moment. The periods of rotation and moment were slightly longer for the 4-element neck.

Figure 3.9


INPUT ACCELERATION PULSE, FLEXION TEST
5
0
-5
-10
-15
-20
-25
-30
-35
PENDULUM ACCELERATION (g)
[v₀ = +23.0 ft/sec]
0
50
100
150
TIME (msec)

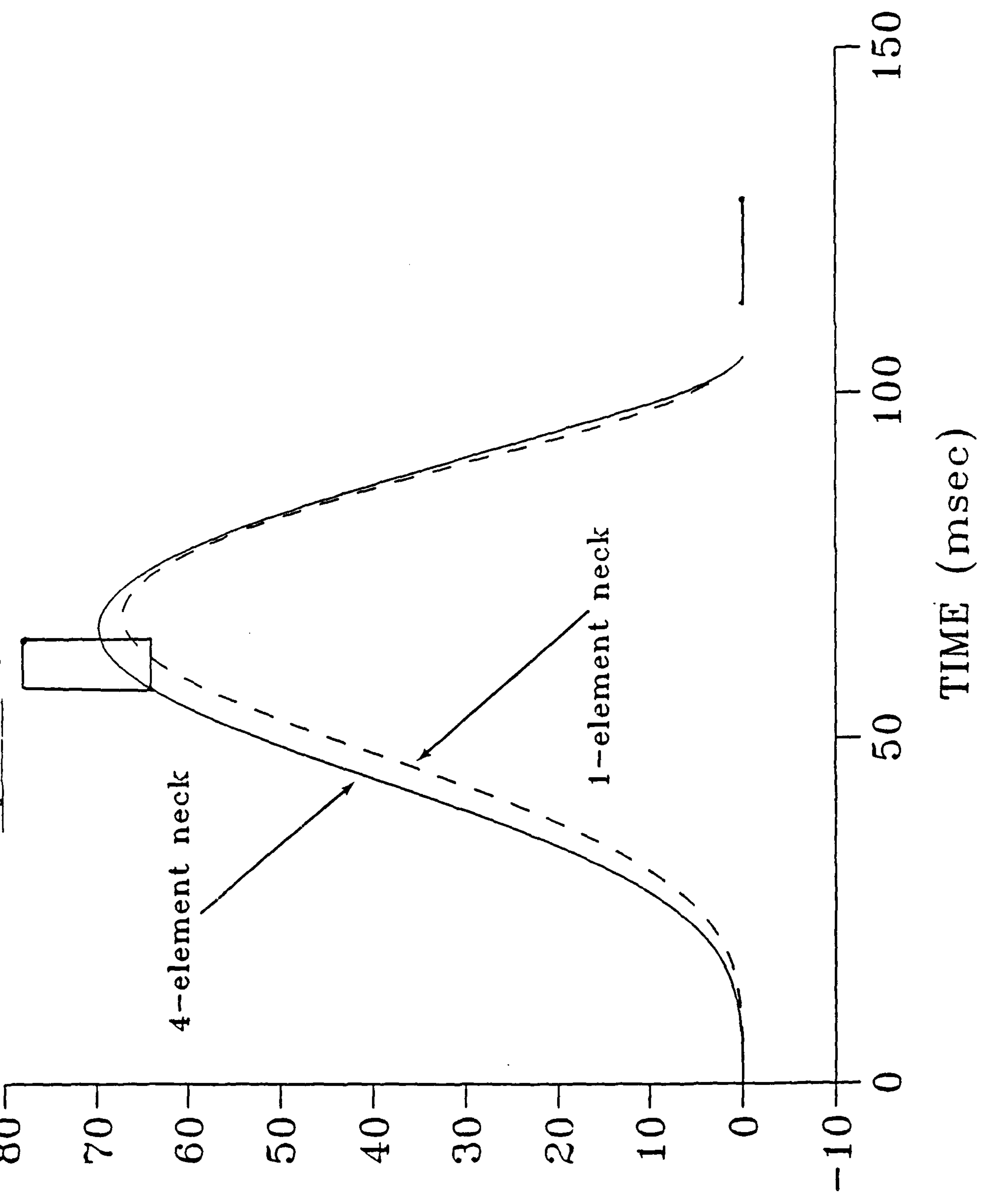


Figure 3.10: Hybrid III Flexion Test

Figure 3.11: Hybrid III Flexion Test

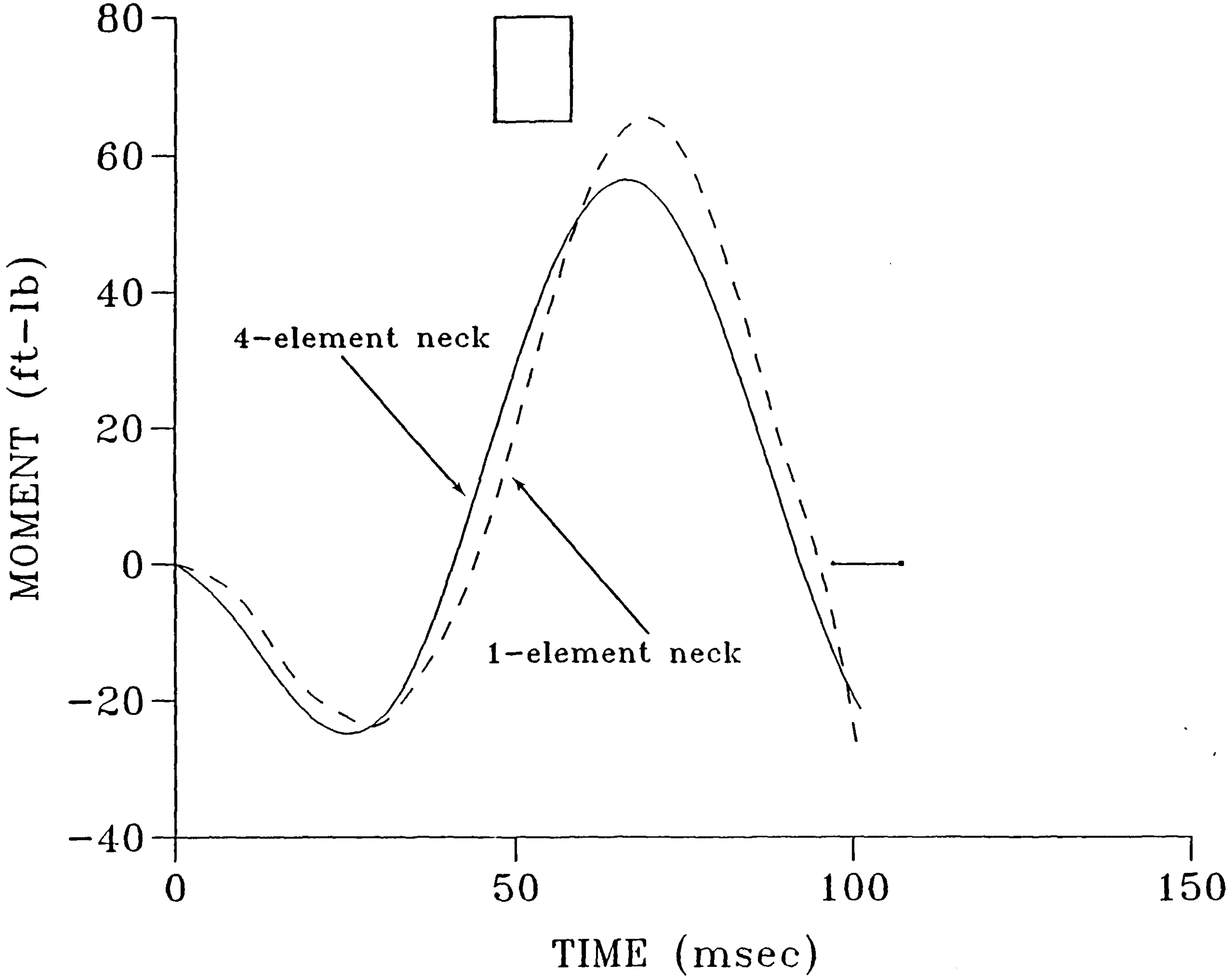


Figure 3.12

INPUT ACCELERATION PULSE, EXTENSION TEST

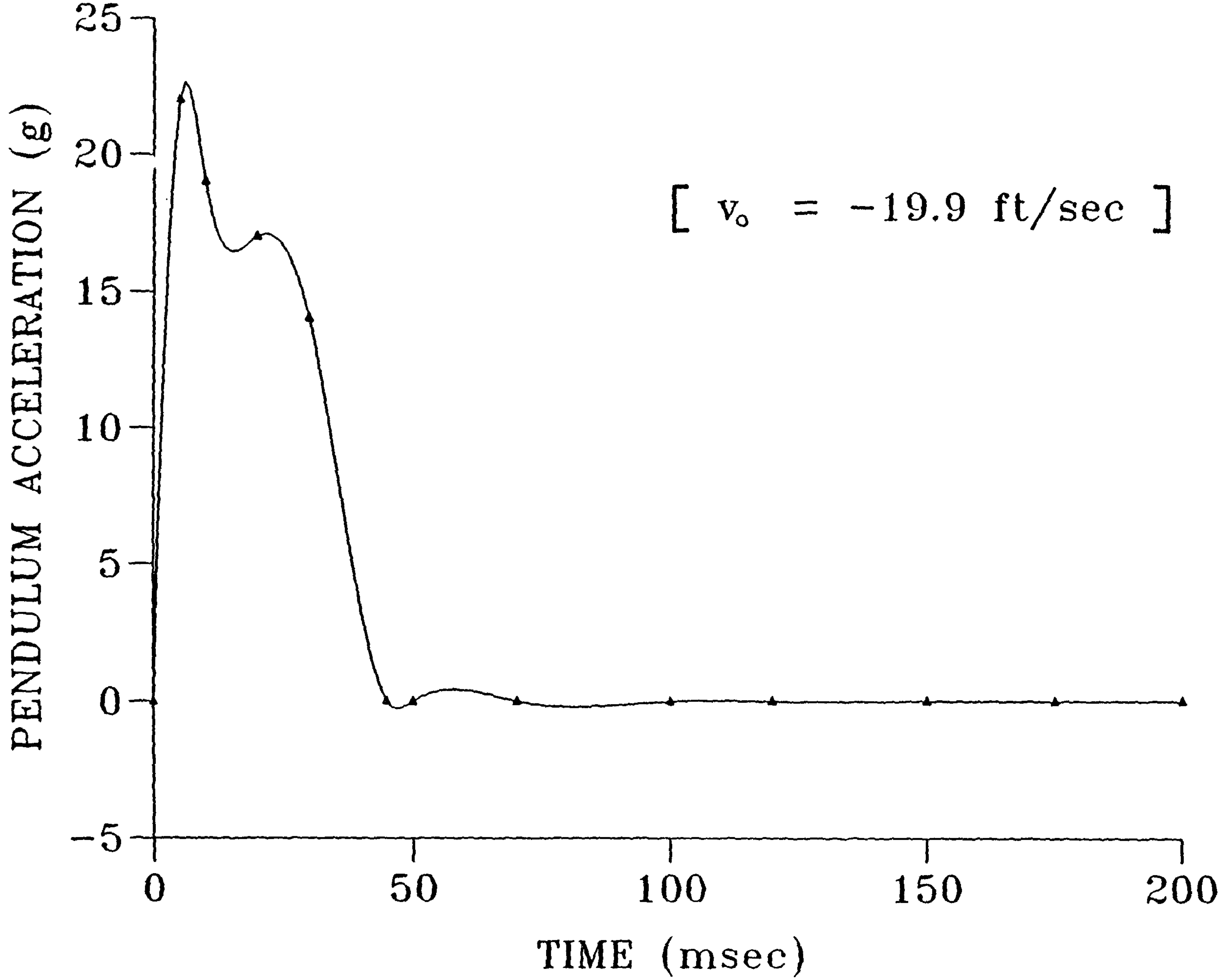


Figure 3.13: Hybrid III Extension Test

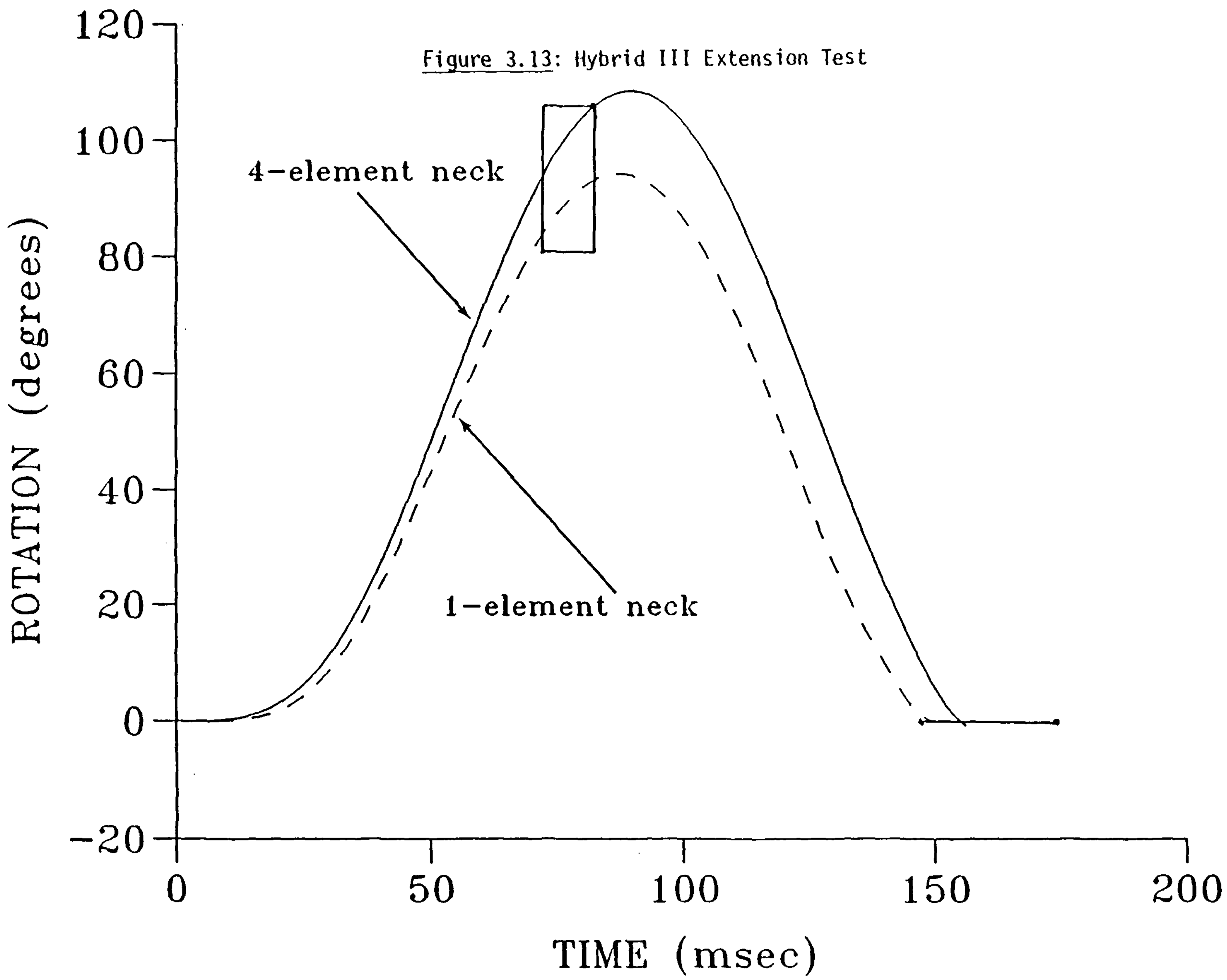
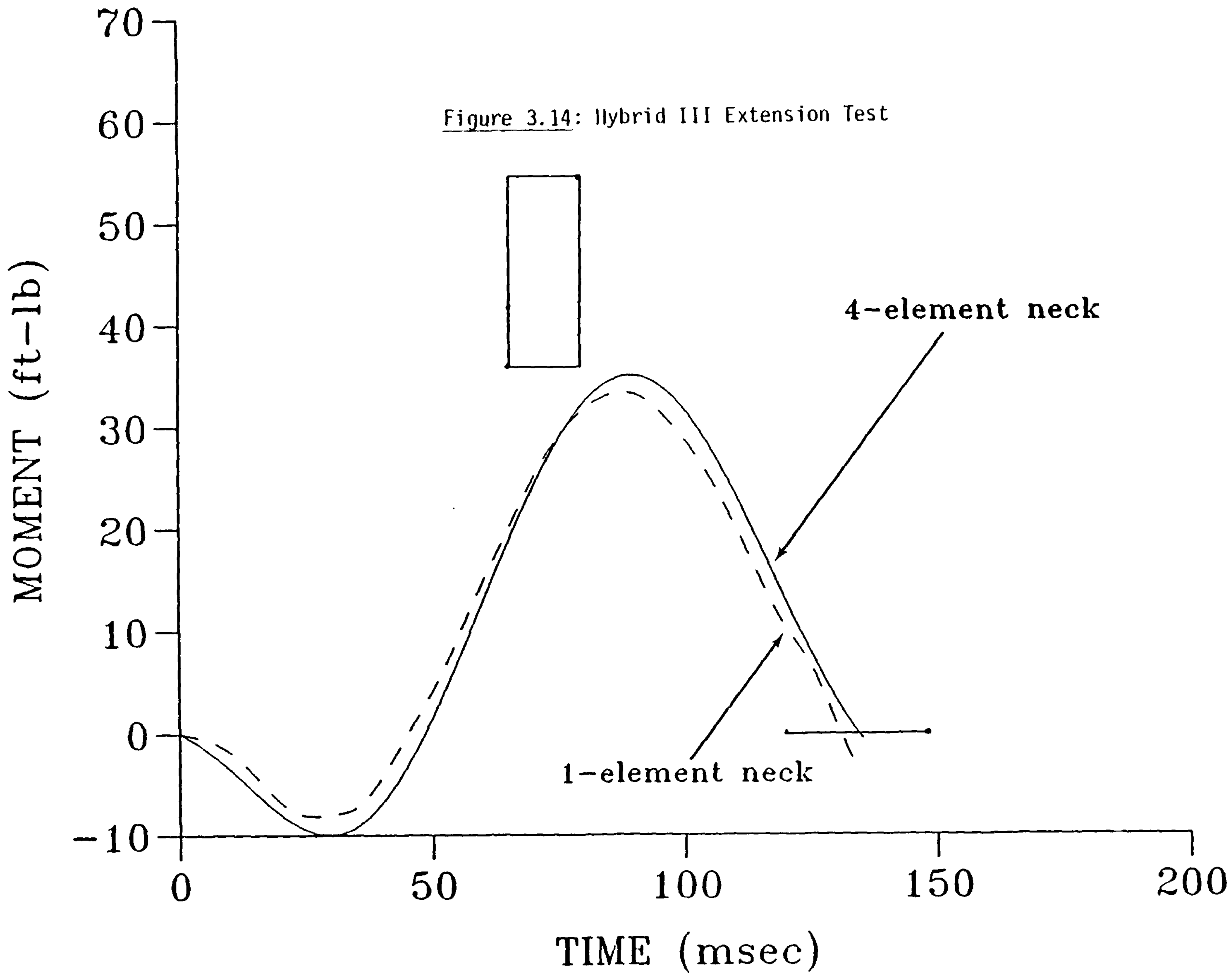


Figure 3.14: Hybrid III Extension Test



3.1.6 Nodding Joint Stiffness Tests

Static and dynamic nodding joint stiffness tests were performed. The joint was mounted horizontally for the static tests. The neck endplate was fixed and a moment arm was attached to the top of the load cell. Moments were applied by hanging weights at the end of the moment arm. The moment and rotation at the end of the moment arm were measured. Small differences were observed between the flexion and extension stiffnesses. For the dynamic tests, the joint and moment arm were mounted vertically. A weight was rigidly attached to the end of the moment arm. Either a force was applied to the weight or the weight was displaced and released. The moment-time histories were recorded. The system was analyzed as a torsional pendulum (i.e., the joint stiffness was calculated from the frequency of the vibration). Differences were observed between the static and dynamic stiffnesses. Further studies are being conducted.

3.1.7 Nodding Joint Data Set Description

The 1-element neck data set was modified to incorporate the occipital condyle nodding block joint into the neck description. Basically, this joint acts like a torsional spring; the spring stiffness is due to the rubber nodding blocks within the joint. Since there is no torsional spring element in the HSM code, the joint was represented by a suitable modification of the beam element.

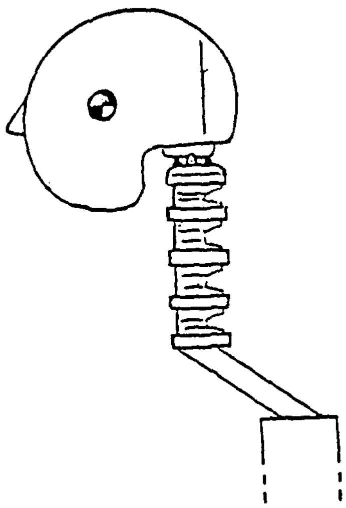
Figure 3.15 shows the geometry of the nodding joint. The physical descriptions of the system are shown on the left; the model descriptions are shown on the right. The two shaded horizontal links represent the joint, essentially two beams on their sides. These beams are compliant only in torsion. As a preliminary effort, the measured static stiffnesses were utilized for these simulations. The values were approximately one order of magnitude larger than the bending stiffness of the 1-element neck.

3.1.8 Nodding Joint Simulation Results

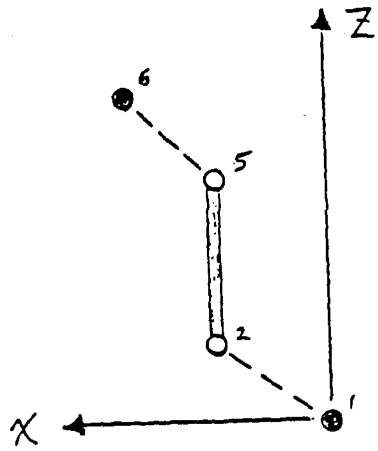
Simulations were performed, using the 1-element neck with the nodding joint, and compared to previous results. Figures 3.16-3.17 show the flexion response. Similar results were obtained for the simulation of the extension test. The dashed line illustrates the response of the 1-element neck without the nodding joint. The solid line illustrates the response of the 1-element neck with the nodding joint. Figure 3.16 is a plot of the head rotation vs. time. The differences, at the beginning of the simulation and at the peak rotation, were slight. Figure 3.17 is a plot of the moment at the occipital condyles vs. time. The occipital condyle nodding block joint had little effect on the response.

Figure 3.15

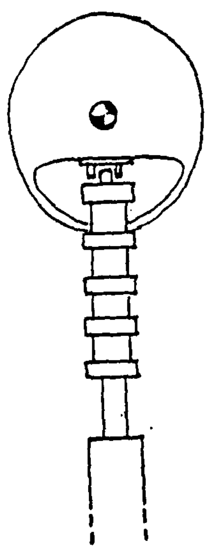
NODDING JOINT GEOMETRY



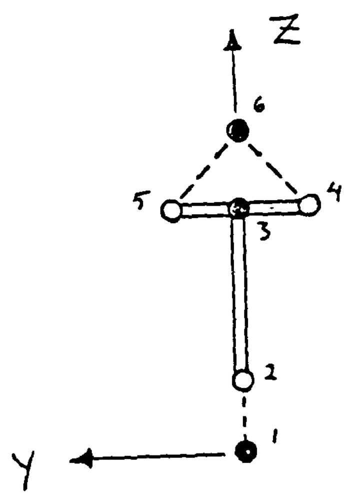
HYBRID III
side view



1-element neck
with nodding joint



HYBRID III
rear view



1-element neck
with nodding joint

Figure 3.16

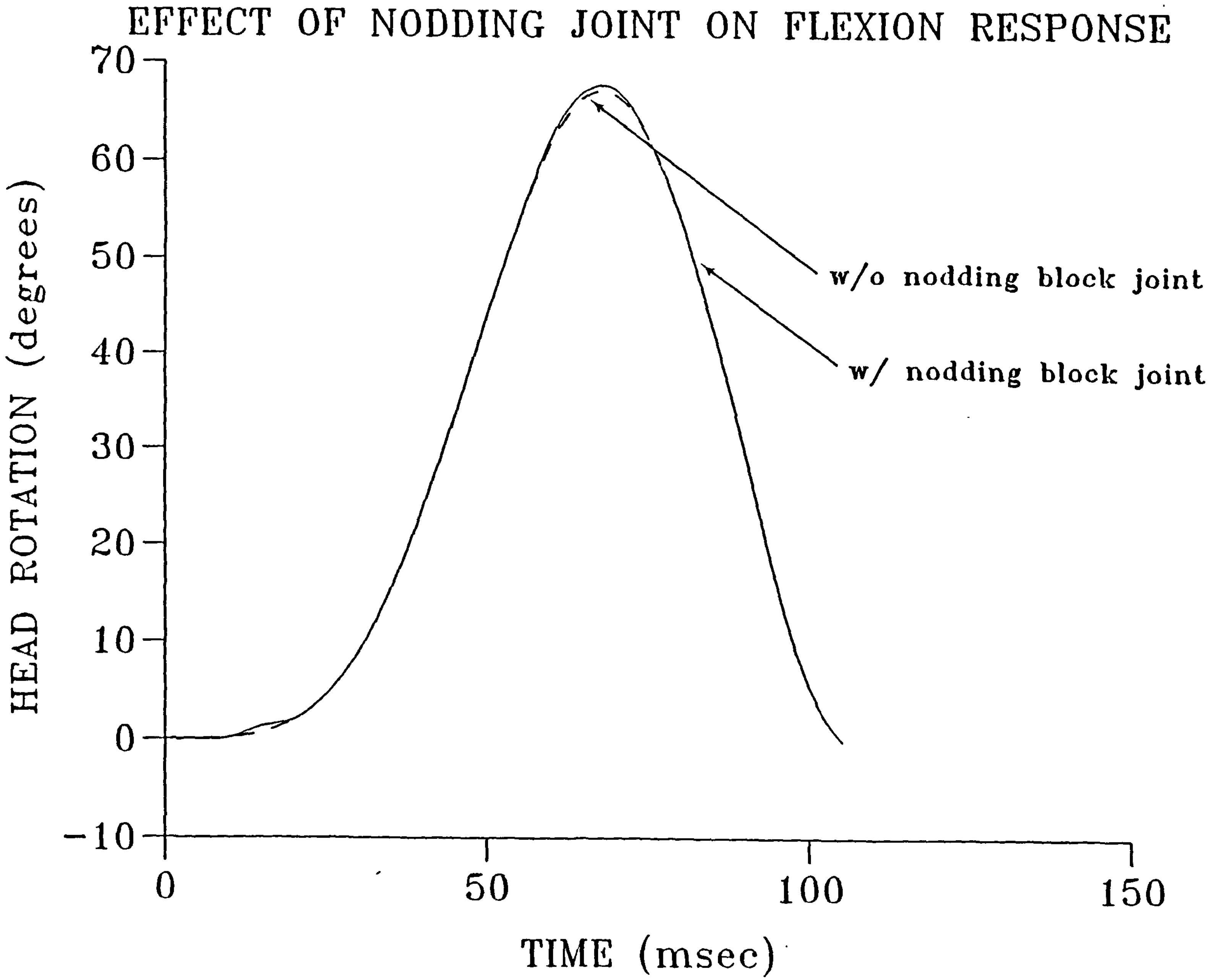
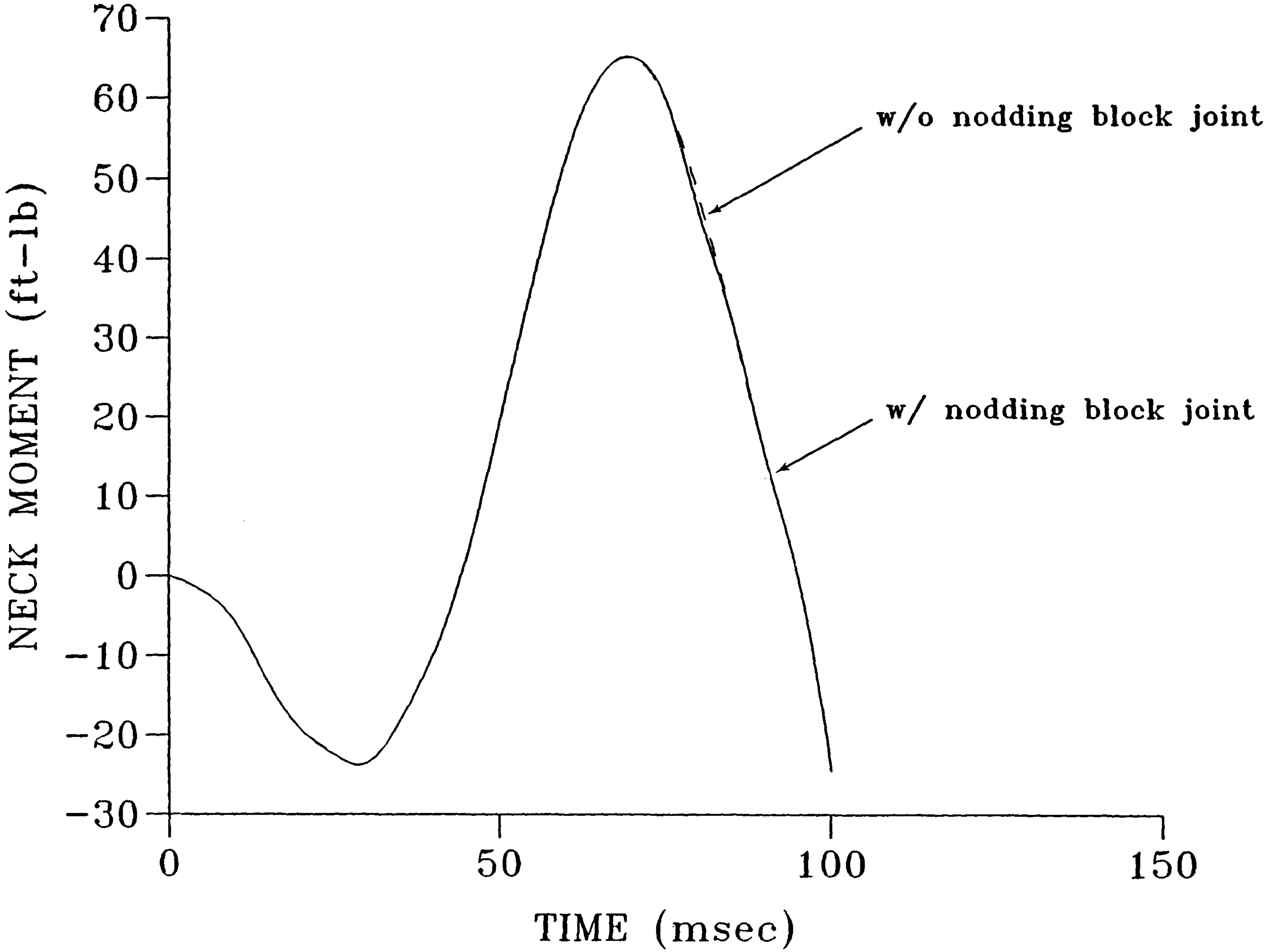

EFFECT OF NODDING JOINT ON FLEXION RESPONSE
HEAD ROTATION (degrees)
70
60
50
40
30
20
10
0
-10
0
50
100
150
TIME (msec)
w/o nodding block joint
w/ nodding block joint

Figure 3.17

EFFECT OF NODDING JOINT ON FLEXION RESPONSE



3.1.9 Execution Times

Execution times were considered in the evaluation of the different data sets. Table 3.4 shows the approximate CPU times required for typical simulation runs on a Microvax II.

Table 3.4 Execution Times for Typical Simulations

SIMULATION	CPU TIME
1-element neck w/o joint	appx. 2 min
4-element neck w/o joint	appx. 2 hr
1-element neck w/ joint	min. 20 hrs

3.2 ATB Model

3.2.1 Hybrid II Data Set Development

The Hybrid II data set consisted of three segments (the pendulum, head, and neck segments) and three joints (the head-neck or HN, neck-pendulum or NP, and pendulum-vehicle or PV joints). Ball-and-socket joints were defined between the head and neck and between the neck and pendulum. A pin joint was defined between the pendulum and the vehicle. The geometric and inertial properties of the segments and pendulum-vehicle joint characteristics were abstracted from operational data sets at AAMRL or estimated from the literature[9,10,12,13]. The neck joint characteristics used initally were estimated from CVS manuals[9,10], operational data sets at AAMRL, Part 572[4], and static test data[12,17]. The honeycomb was defined as a contact plane.

Two types of ATB simulations were performed. The first type was a full-drop simulation, where time zero represented the initial release of the pendulum. The second type were simulations of only the impact phase, where time zero was the time just prior to the impact of the pendulum against the aluminum honeycomb. Figure 3.18 shows the VIEW program output from the full-drop simulation. The final 150 msec of the full-drop simulations corresponded to the impact phase. The initial conditions for the impact phase simulations were the positions and velocities of the segments, which were calculated from the Part 572 pendulum strike plate impact velocity. The specified pendulum strike plate acceleration-time history was achieved by defining a plane-ellipsoid contact between the pendulum and honeycomb and force-deflection characteristics for that contact. Figure 3.19 shows a typical pendulum strike plate acceleration pulse for an impact velocity of 23.5 ft/sec.

The impact phase simulations revealed two problems with the initial definition of the pendulum-honeycomb contact. The pendulum was defined as a segment with an attached ellipsoid and the honeycomb was defined as a plane; plane-ellipsoid contact resulted. Since contact forces are defined in terms of the penetration between contact surfaces and since the strike plate was located at a distance almost equal to the major axis length of the pendulum ellipsoid, contact was not observed at the right point in space if the dimensions of the pendulum and honeycomb were defined correctly. Both the amount of penetration and the contact time were low. This problem was eliminated by shifting the contact ellipsoid attached to the pendulum so that its center coincided with the center of the strike plate. The second problem related to the contact algorithm, which assumes that no contact occurs and hence no forces are generated after full penetration of the pendulum into the honeycomb. Since this algorithm is not valid for this simulation, the pendulum thickness was increased with no change in its inertial properties. For realistic graphics, the pendulum ellipsoid was removed from the plots and replaced by a rectangular segment in the VIEW program input file. Figure 3.20 shows a typical VIEW program output.

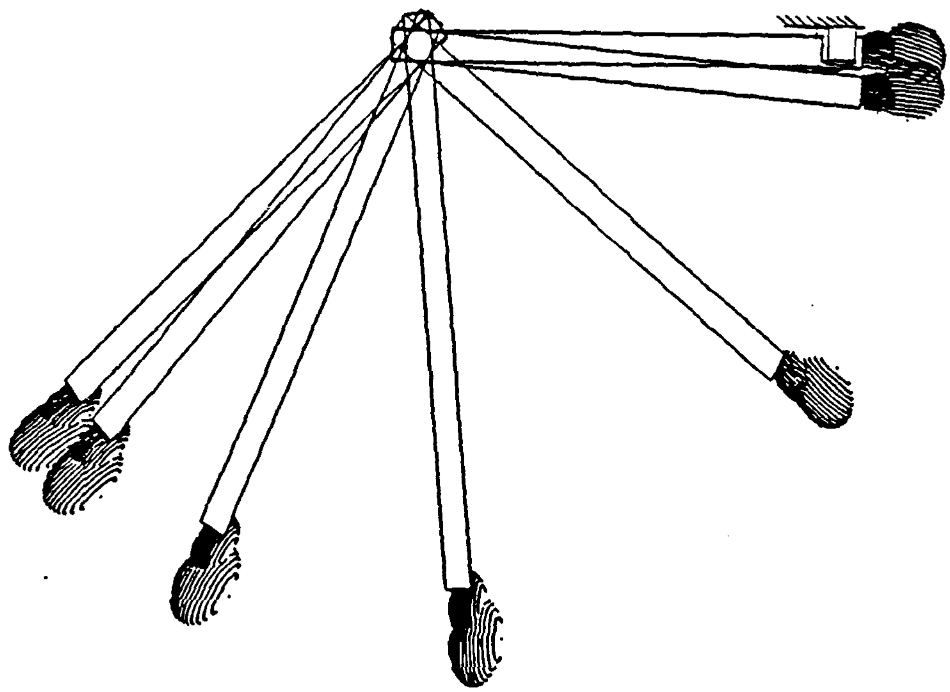
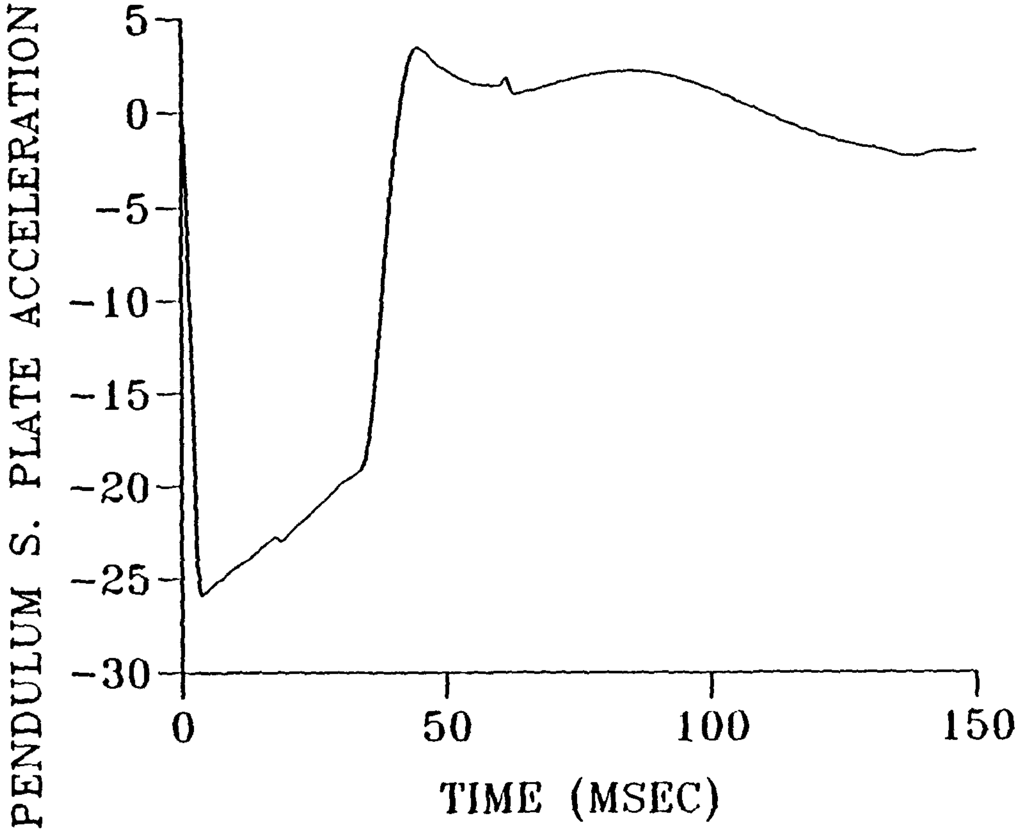


Figure 3.18: Typical VIEW Program Output

Figure 3.19

TYPICAL PEND. S. PLATE ACC. PROFILE



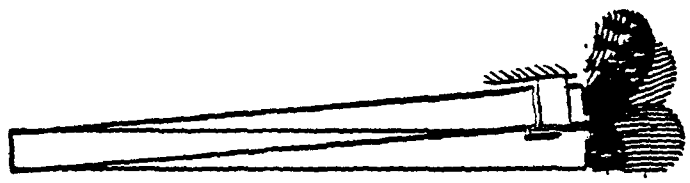
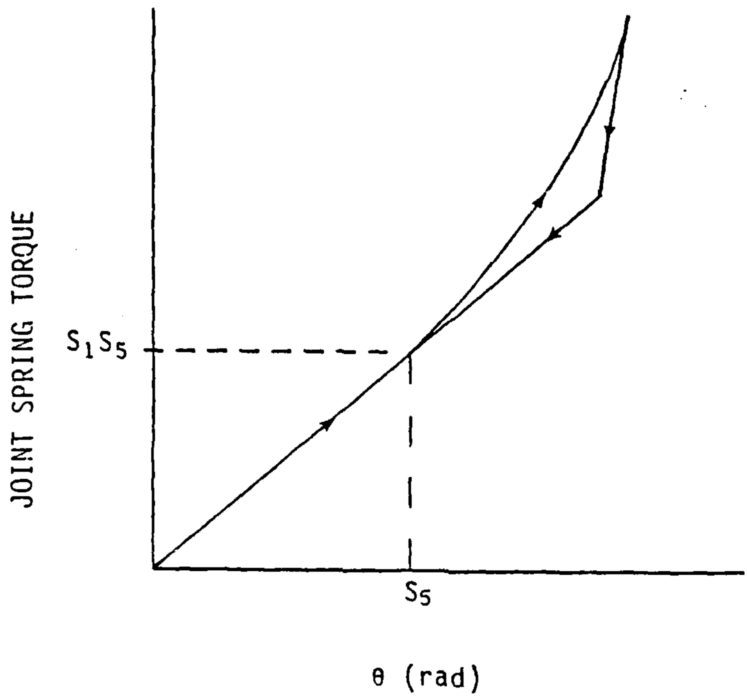


Figure 3.20: Typical VIEW Program Output

The responses of the model did not comply with the Part 572 performance requirements using the initial data set. The data set was tuned to make the ATB head-neck responses comply with Part 572. Since the geometric and inertial properties of the Hybrid II head and neck are well documented, these constants remained fixed; they were not used to tune the data set. The values used for the joint charactistics of the neck, however, are not well documented. Parametric studies were conducted to assess the effects of variations in neck joint characteristics upon the predicted head-neck responses. Since the neck joint was modeled as a ball-and-socket type, the relevant joint characteristics are: (1) flexural linear, quadratic, and cubic spring coefficients, energy dissipation coefficient, and joint stop; (2) torsional linear, quadratic, and cubic spring coefficients, energy dissipation coefficient, and joint stop; (3) viscous coefficient; (4) coulomb friction coefficient; (5) full friction angular velocity; (6) maximum torque for a locked joint; (7) minimum torque for an unlocked joint; (8) minimum angular velocity for a unlocked joint; and (9) impulse restitution coefficient. By varying these neck properties in a systematic manner, optimization of the model response with the specifications was possible.

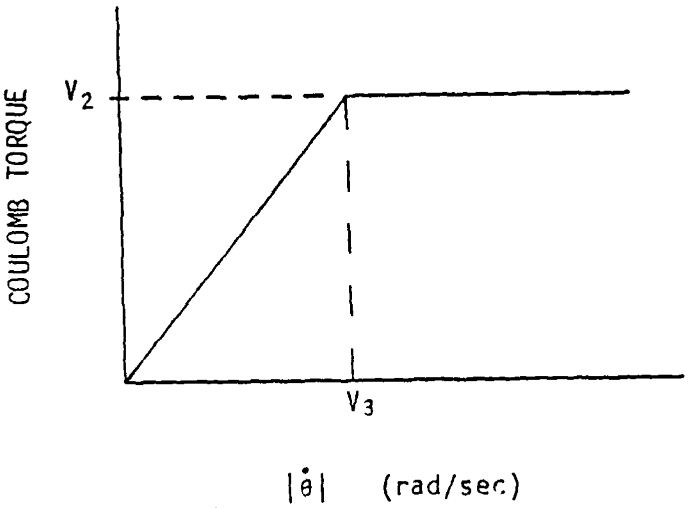
Joint torques consist of spring, coulomb friction, and viscous components. For the ball joint, both flexural and torsional torques are computed. When a joint angle exceeds the user-specified joint stop, the joint locks. The ball joint is either completely free or completely locked. If the locking torque exceeds a prescribed level, the joint unlocks. The user specifies the minimum torque and minimum relative angular velocity at which the joint remains unlocked. If the locking torque or the velocity fall below these specified levels, the joint will relock. If the joint goes from a free to a locked state or a locked to a free state, an impulse subroutine may be called to set the relative angular velocity of adjoining segments to zero.

Figure 3.21 shows the algorithm used to calculate the spring and joint stop restoring torques from the spring coefficients S_1 thru S_4 and joint stop angle S_5. A linear torque vs. angle is prescribed until the specified joint stop angle is reached. For angles greater than the joint stop, a quadratic and cubic restoring torque is added. If the angular motion of the joint exceeds the specified stop, a progressively increasing restoring torque will be applied. The algorithm used to calculate the coulomb friction torque from the magnitude of the relative angular velocity $|\frac{d\theta}{dt}|$, the constant coulomb friction torque V_2, and the limit of the linear coulomb friction torque V_3 is shown in Figure 3.22. The algorithm used to calculate the viscous torque from the relative angular velocity $\frac{d\theta}{dt}$ and the linear coefficient V_1 is shown in Figure 3.23.



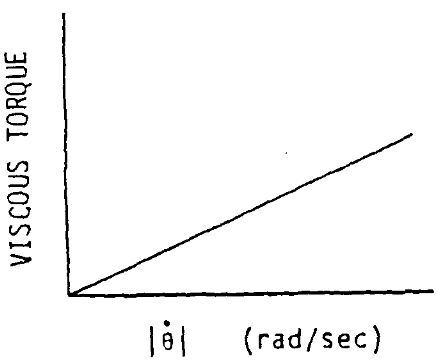
$$T_S = \begin{cases} S_1\theta \; ; \text{ for } \theta < S_5 \\ S_1\theta + S_2(\theta - S_5)^2 + S_3(\theta - S_5)^3 \; ; \text{ for } \theta > S_5 \text{ and } \dot{\theta} > 0 \\ S_1\theta + S_4[S_2(\theta - S_2)^2 + S_3(\theta - S_5)^3] \; ; \text{ for } \theta > S_5 \text{ and } \dot{\theta} < 0 \end{cases}$$

Figure 3.21: JOINT SPRING TORQUE (Abstracted from Fleck, et al., 1975)



$$T_C = \begin{cases} \dfrac{V_2 \dot{\theta}}{V_3} & ; \text{ for } |\dot{\theta}| < V_3 \\ V_2 & ; \text{ for } |\dot{\theta}| > V_3 \end{cases}$$

Figure 3.22: JOINT COULOMB FRICTION TORQUE



$$T_v = V_1 \dot{\theta}$$

Figure 3.23: JOINT VISCOUS TORQUE

3.2.2 Hybrid II Simulation Results

Figures 3.24-3.35 show the results of the parameter tuning process. Plots labeled 572 illustrate the final data set results. Plots labeled H illustrate the predicted responses when a higher value is used for the selected parameter. Plots labeled L illustrate the predicted reponses when a lower value is used for the selected parameter. Table 3.5 shows the values used for the selected parameters.

Listings of the final (572) input data set and relevant outputs are included in Appendix C. The final data set predicted head rotation and chordal displacement data which complied with Part 572 specifications. The predicted peak head accelerations were slightly higher than that specified by Part 572.

It was found that an increase in the HN linear flexural spring coefficient decreased the amplitude and period of the head rotation but increased the amplitude and decreased the period of the chordal displacement (Figures 3.24-3.25). The effects of changes in the HN viscous and HN coulomb friction coefficients were similar (Figures 3.26-3.29). An increase in the parameter above the 572 value decreased the amplitude of the head rotation with little change in the chordal displacement. A decrease in the parameter below the 572 value produced little change in the predicted responses. Increasing the NP linear flexural spring coefficient decreased the amplitude and period of both the head rotation and chordal displacement (Figures 3.30-3.31). The effects of changes in the NP viscous and NP coulomb friction coefficients were similar (Figures 3.32-3.35). An increase in the parameter decreased the amplitude but increased the period of both the head rotation and chordal displacement. Decreasing the HN joint stop had little effect on the predicted responses. Decreasing the NP joint stop, however, decreased the period of the head rotation and decreased both the amplitude and period of the chordal displacement. Increases in the energy dissipation coefficients with the smaller joint stop angles had little effect on the predicted responses. The effects of changes in the full friction angular velocity and changes in the torsional coefficients over several orders of magnitude were negligible. Finally, increasing the NP linear flexural spring, viscous, or coulomb friction coefficients reduced the predicted head accelerations by approximately 10%. Changes in all other parameters over several orders of magnitude had little effect on the predicted head accelerations.

Table 3.5: Neck Joint Characteristics

Parameter	Joint	L	572	H
Linear Flex. Spring Coeff. (in-lb/deg)	HN	7.	35.	3500.
	NP	10.	20.	40.
Joint Stop (deg)	HN	30.	90.	
	NP	45.	90.	
Energy Diss. Coefficient	HN		0.63	1.0
	NP		0.66	1.0
Viscous Coefficient (in-lb-sec/deg)	HN	0.0015	0.15	1.5
	NP	0.0020	0.20	0.4
Coulomb Friction Coeff. (in-lb)	HN	0.5	50.	500.
	NP	0.125	125.	250.
Full Friction Ang. Velocity (deg/sec)	HN	3.	100.	300.
	NP	3.	100.	300.

Figure 3.24

HN LINEAR FLEXURAL SPRING COEFFICIENT VARIATIONS

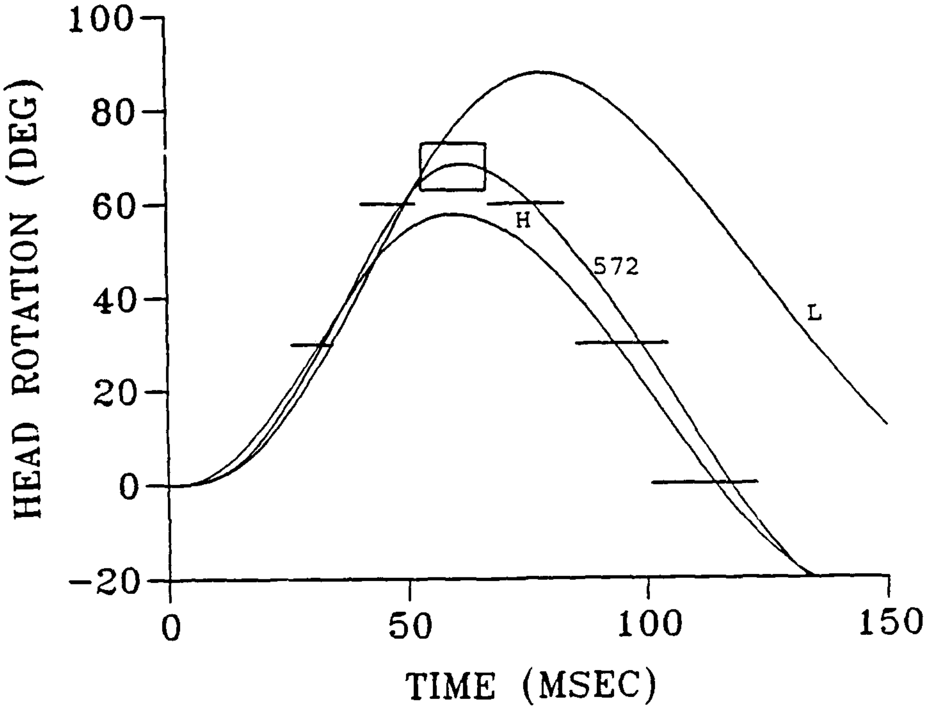


Figure 3.25

HN LINEAR FLEXURAL SPRING COEFFICIENT VARIATIONS

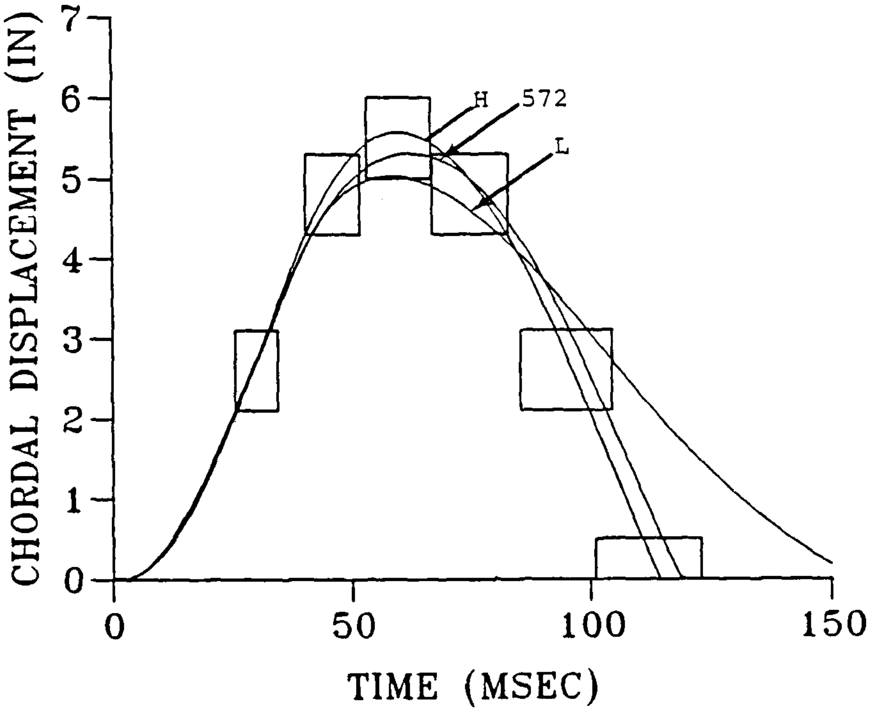


Figure 3.26

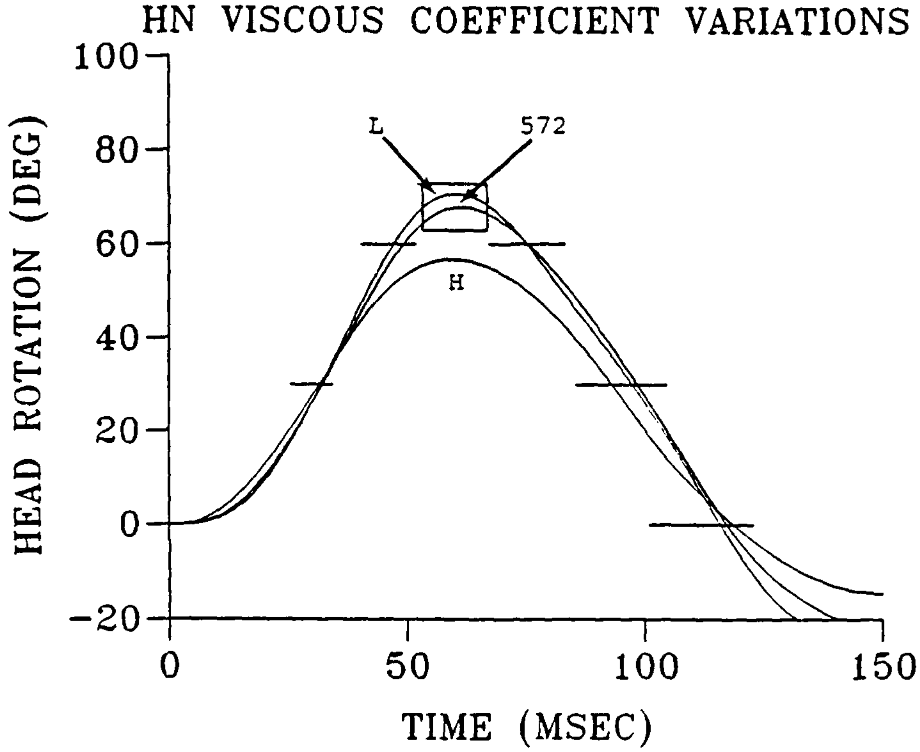


Figure 3.27

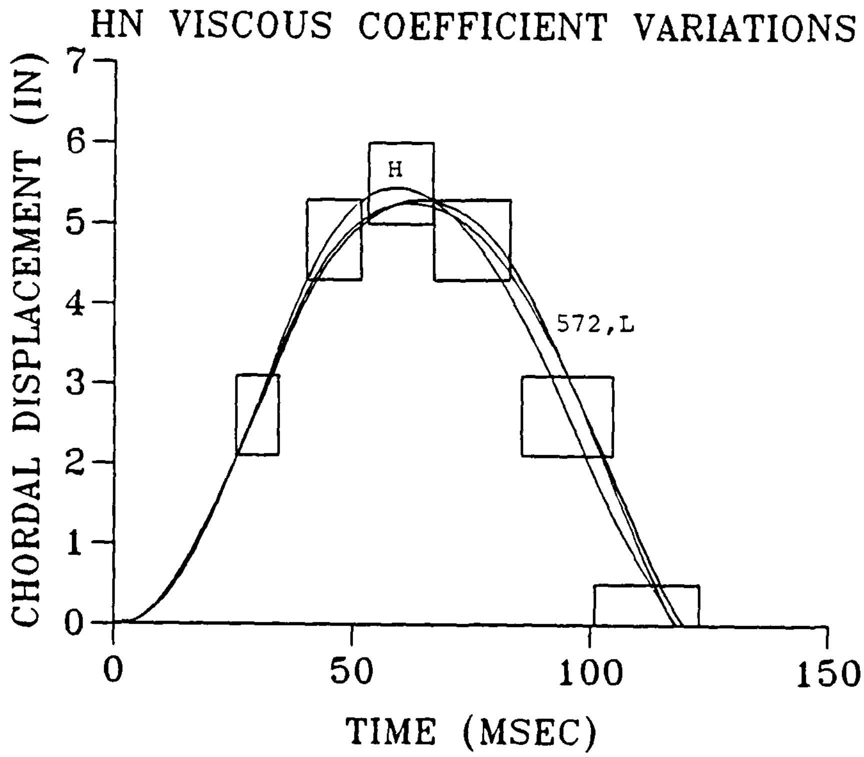


Figure 3.28

HN COULOMB FRICTION COEFFICIENT VARIATIONS

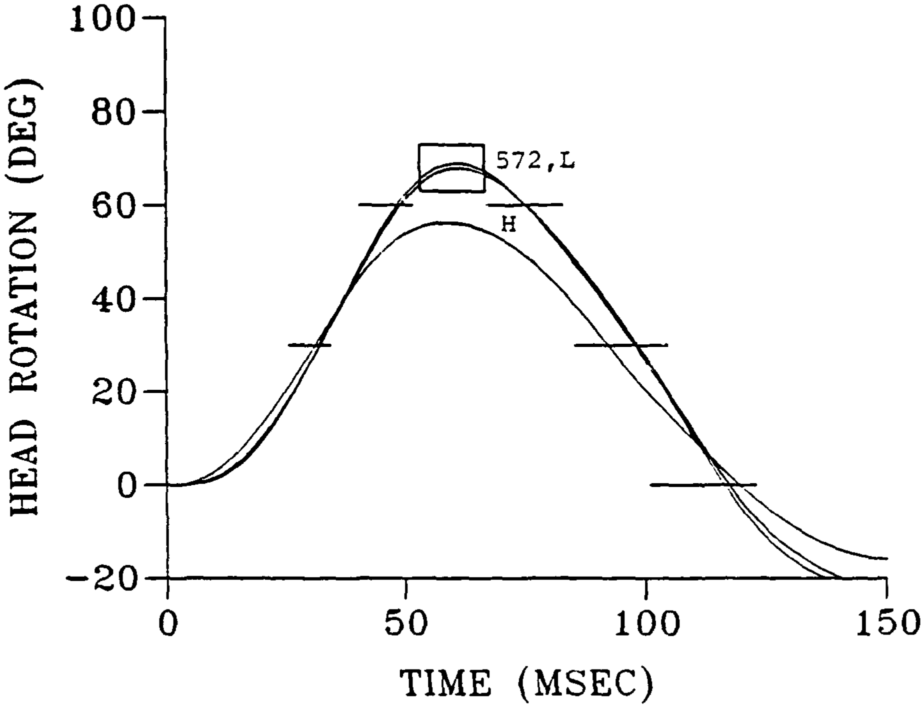


Figure 3.29

HN COULOMB FRICTION COEFFICIENT VARIATIONS

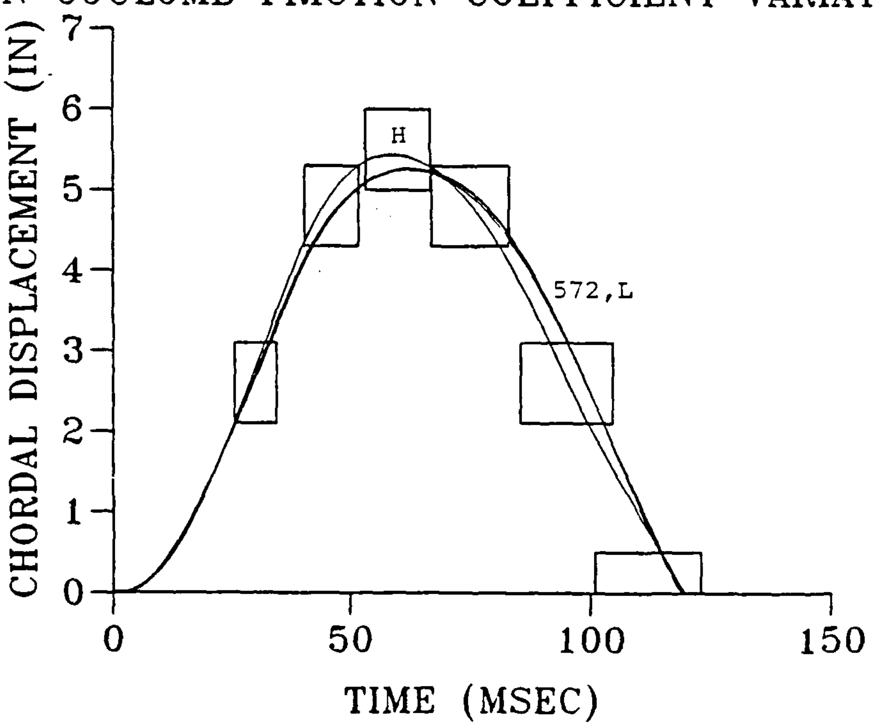


Figure 3.30

NP LINEAR FLEXURAL SPRING COEFFICIENT VARIATIONS

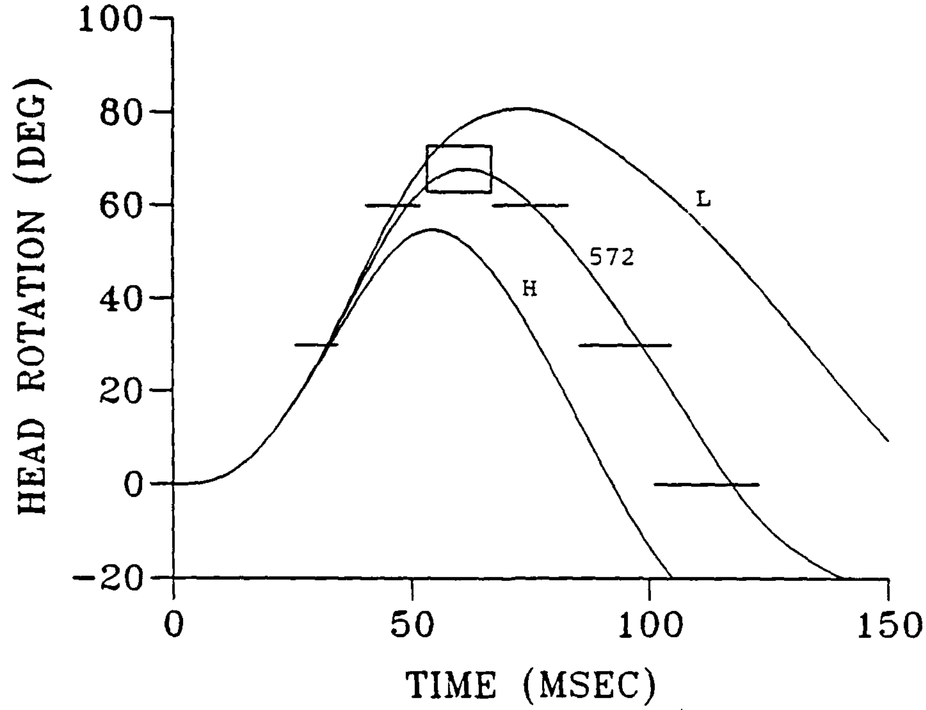


Figure 3.31

NP LINEAR FLEXURAL SPRING COEFFICIENT VARIATIONS

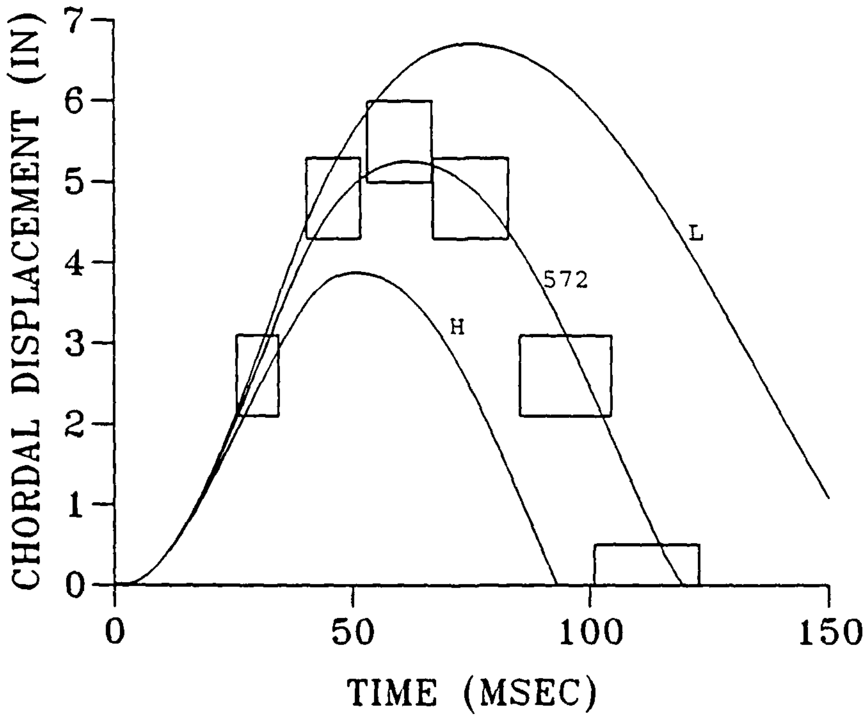


Figure 3.32

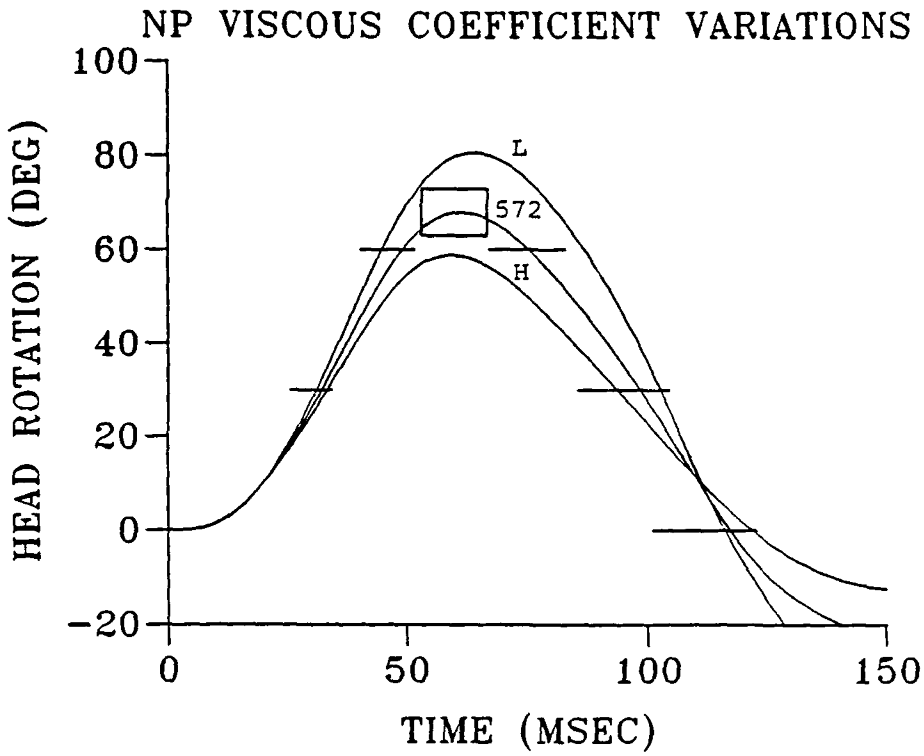


Figure 3.33

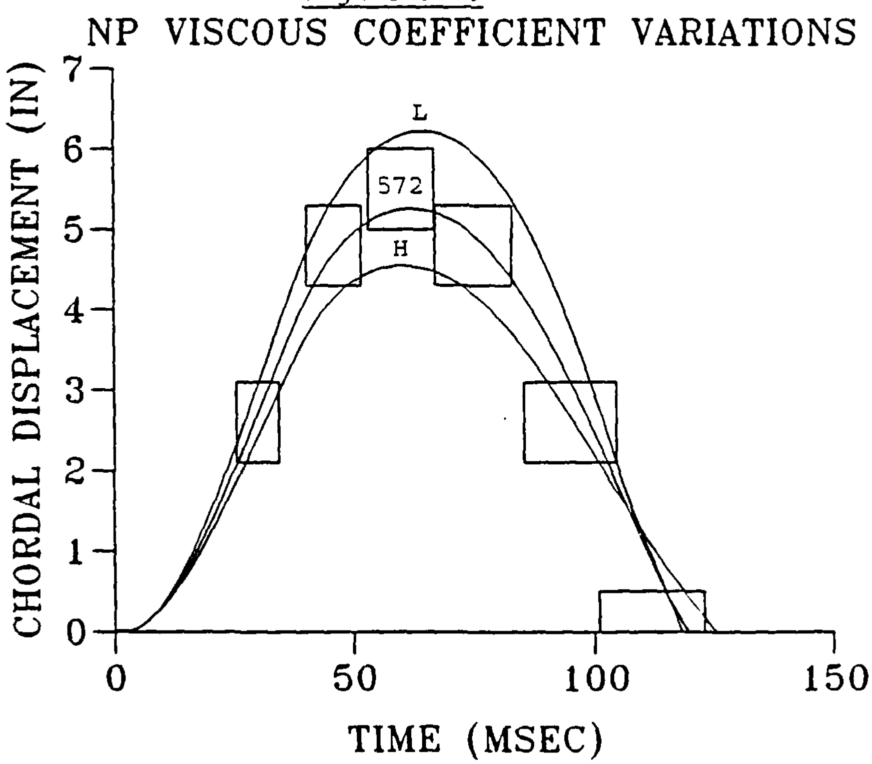


Figure 3.34

NP COULOMB FRICTION COEFFICIENT VARIATIONS

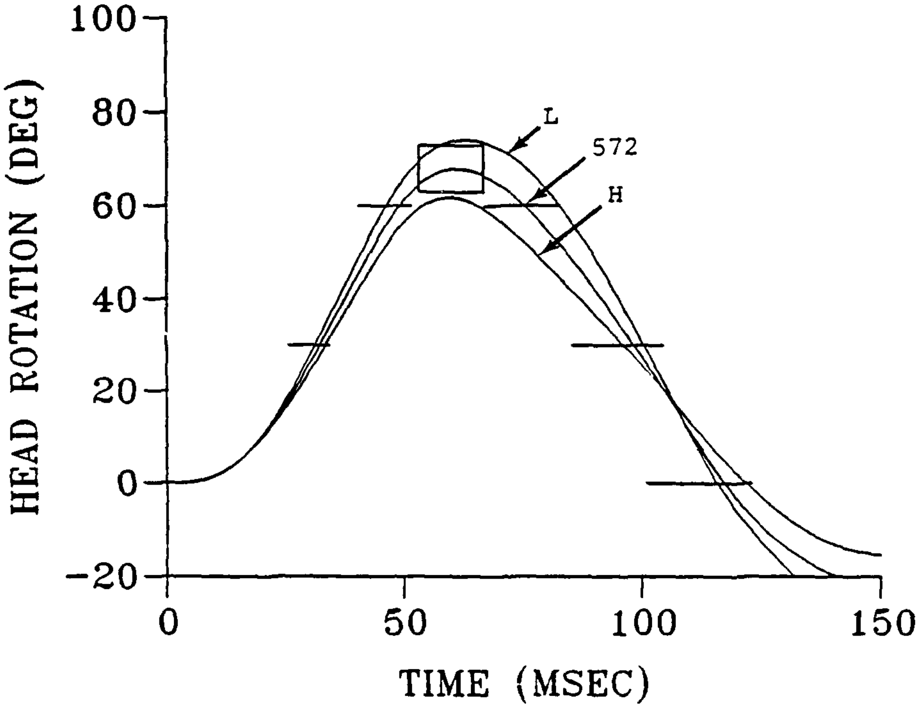
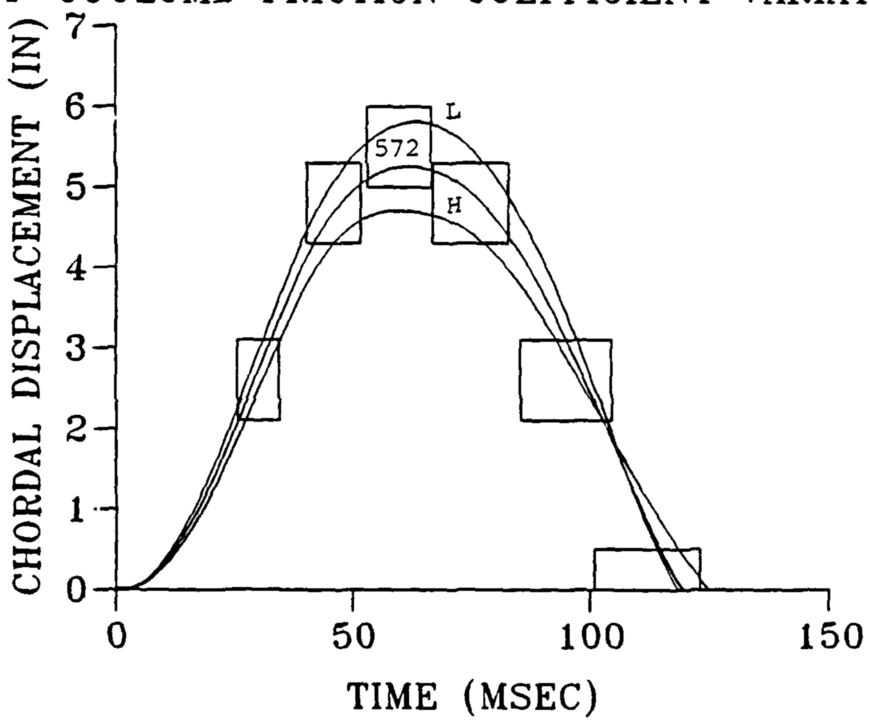


Figure 3.35

NP COULOMB FRICTION COEFFICIENT VARIATIONS



3.2.3 Hybrid III Data Set Development

A reasonable step in extending this work was to develop a data set which represents the Hybrid III. As a preliminary effort, the data set did not contain explicit representations of every part of the neck. Instead, the effects of the nodding blocks, central cable, and the layered construction of the Hybrid III neck were lumped together. The Hybrid III data set consisted of three segments (the pendulum, head, and neck segments) and three joints (the head-neck or HN, neck-pendulum or NP, and pendulum-vehicle or PV joints). Ball-and-socket joints were defined between the head and neck and between the neck and pendulum. A pin joint was defined between the pendulum and the vehicle. The geometric and inertial properties of the segments and pendulum-vehicle joint characteristics were abstracted from operational data sets at AAMRL or estimated from the literature[14,15]. The neck joint characteristics used initally were estimated from operational data sets at AAMRL, Amended Part 572[4], and static test data from a study conducted by AAMRL for the Department of Transportation. The honeycomb was defined as a contact plane.

Both flexion and extension tests were simulated. Simulations of the impact phase were performed, where time zero was the time just prior to the impact of the pendulum against the aluminum honeycomb. The initial conditions were the positions and velocities of the segments, which were calculated from the Amended Part 572 pendulum strike plate impact velocities. The specified acceleration-time histories were achieved by defining a plane-ellipsoid contact between the pendulum and honeycomb and force-deflection characteristics for that contact.

The responses of the model did not comply with the Amended Part 572 performance requirements using the initial data set. Efforts are currently being made to tune the data set in order to make the ATB head-neck responses comply with Part 572.

4. EXPERIMENTAL STUDIES

As part of the SFRP and RIP, Paver and Doherty conducted experimental studies of the Hybrid II, modified Hybrid II, and Hybrid III dummy head-neck assemblies and the Part 572 and Amended Part 572 head-neck compliance tests. A test plan was written. Initially, the standard head-neck compliance tests were to be performed to verify the operation of the experimental setup. Tests at different impact velocities were planned to validate data sets for mathematical models. Measurements were made of the geometric and inertial properties of the test specimens and apparatus. Mounting plates, test fixtures, and apparatus modifications were made. All transducers were wired; some were calibrated. The test specimens and apparatus were assembled. The velocimeter was modified and calibrated. Attempts were made to determine the drop heights required for the specified impact velocities. Preliminary measurements of pendulum strike plate deceleration pulses were made; the effects of honeycomb precrush were investigated. Several recommendations for modifications of the experimental system were suggested.

4.1 Test Specimens

The test specimens were the head and neck assemblies of Hybrid II, modified Hybrid II, and Hybrid III manikins. Noteworthy modifications of the Hybrid II include:

(1) a shorter molded neck than that specified by DOT for Part 572 dummies

(2) a Hybrid III type head-neck bracket which permits use of the six-axis load cell, nodding blocks, and occipital condyle pin

(3) head accelerometers and an aluminum mounting plate unlike the devices supplied by Humanoid Systems.

These modifications are significant and will probably alter the performance of the head-neck system. Modifications of the Hybrid III include:

(1) head accelerometers and a steel mounting plate unlike the devices supplied by Humanoid Systems

(2) an occipital condyle pin replacement which permits attachment to the three-potentiometer unit.

These modifications should not significantly alter the performance of the head-neck system. Tests performed with the Hybrid II exclude the neck-thorax bracket; tests performed with the Hybrid III include the neck-thorax bracket.

4.2 Pendulum Test Apparatus

All tests were conducted with the Humanoid Systems Model #TE-200-HIII pendulum test apparatus at AAMRL. A release trigger mechanism, mounted to the pendulum structure, allows repeatable release heights and release mechanics. The air brake mechanism, which was designed to eliminate rebound of the pendulum after impact, was not utilized; other laboratories have found that this mechanism can adversely affect the pendulum deceleration. The test specimens were rigidly fixed to the pendulum with a steel pendulum-neck mounting plate. Aluminum commercial grade honeycomb, Type ACG-3/4-1.8-0.003, produced by Hexcell, was used as the impact material.

4.3 Geometric and Inertial Properties Measurements

The inertial properties of the dummy head and neck segments were determined experimentally using balsa wood holding boxes, a C & S electronic scale, a moment table, a Space Electronics Model #KGR1000-3 Mass Properties Torsion Pendulum, a Hewlett-Packard Model 85 microcomputer, and a Micro Control Systems perceptor three-dimensional digitizer linked to the AAMRL Perkin Elmer computer. The scale was used to determine the segment-box weight to the nearest 0.01 lbs. The moment table, which had two parallel knife edges, was used to measure the segment-box cg location. One knife edge was placed on the scale and the other on a fixed stand. With the scale tared to zero, the first moment about the fixed blade edge was read directly from the scale. Then, using the known distance between the knife edges, the exact location of the cg of the segment-box combination was calculated. The torsion pendulum was used to determine the moment of inertia of the segment-box combination about its cg. By performing this task with the segment-box in six different attitudes, the inertia tensor elements at the cg were determined. Next, specified points on the segment were located in box coordinates using the perceptor. The principal moments of inertia and the direction cosines of the respective principal axes were determined from the tensor elements and the perceptor data. The effects of the box were removed and the inertial properties of the segment were calculated relative to the specified points on the segment.

The geometric and inertial properties of the pendulum were also measured. The weight was measured on a high-capacity scale. A knife edge was used to determine the cg location. The torsion pendulum and HP85 microcomputer were used to determine the moment of inertia. Although the pendulum was designed using the Part 572 specifications as a model, some differences exist. The effects of these differences were estimated to be negligible with respect to the dynamics of the head-neck system. This matter was discussed with the members of the SAE Dummy Users Group; it appears that the pendulum strike plate impact velocity and deceleration pulse are the important requirements of the head-neck pendulum compliance test.

4.4 Pendulum Test Instrumentation

The instrumentation used in this study consists of piezoresistive accelerometers, a six-axis load cell, a three-potentiometer device, and a velocimeter with a photocell sensor.

An Entran Devices Model #EGA-F-5OD triaxial accelerometer measures the pendulum arm deceleration. This transducer is mounted 65.25 inches from the pivot point of the pendulum along the pendulum arm centerline.

Three single-axis accelerometers (Endevco Model #7264-200) measure the Hybrid II head accelerations. These devices are mounted orthogonally in the head assembly on an Endevco Model #7954 mounting block, which is attached to the accelerometer mounting plate.

Each accelerometer will be full comparison calibrated every 9 months. Before each test, the accelerometers will be shunt calibrated by passing a standard signal for each channel through the instrumentation system. When any device shows a deviation of more than 5% from the standard shunt output, the device will be full comparison calibrated.

A three-potentiometer unit, manufactured by Humanoid Systems, measures head rotation and displacement data in order to obtain a complete description of head motion in the mid-sagittal plane. This unit attaches to the pendulum-neck mounting plate. The potentiometers will be calibrated every 9 months.

A six-axis load cell (Denton Model #1716) measures forces and moments at the modified occipital condyle pin of the Hybrid III manikin. Prior to delivery, this transducer was comparison calibrated by the manufacturer. Maintenance calibrations will be performed every 9 months.

A photocell device, manufactured by Humanoid Systems, measures the pendulum velocity just prior to impact. A velocimeter, built by SRL, counts the time between the rising edges of two sequential photocell pulses which result from a light beam passing through a uniformly perforated wand attached to the pendulum.

A high-speed 16mm camera will record photogrammetric data in order to document head motion. This camera will be mounted to the facility floor such that an unobstructed perpendicular side view of the head-neck assembly is obtained upon impact. Standard photometric targets will be mounted on the head, neck, and pendulum. The zero reference switch, strobe flash, and the light-emitting-diode TBAR display timing system in the camera will be used to synchronize the photographic data with the electronic data. Each test will be positively identifiable by appropriate numeric characters visible within the initial camera view. A 35mm camera will be utilized to document the pre-test configurations.

4.5 Velocimeter Modifications and Calibration

The SRL velocimeter and an MTS velocimeter were comparison calibrated using drop angles of 50° ± 3° (manual release) and 94.5° (first hole in release mechanism). Since the wands were positioned at different locations along the pendulum, efforts were made to insure that the photocell beams were blocked at the same instant in time.

The photocell device of the SRL velocimeter produces a pulse when the light beam passes through a slot in the wand. The original SRL velocimeter measured the time between the rising and trailing edges of each photocell pulse in order to calculate velocity. This system was modified to measure the time between the rising edges of two sequential photocell pulses. Since the distance between sequential slots on the SRL wand is 0.375 inches, the SRL velocity is calculated as follows:

$$v_{SRL} = \frac{0.375in}{count} \times \frac{1ft}{12in} \times \frac{10^6 counts}{sec}$$

$$v_{SRL} = \frac{31250}{count} \frac{ft}{sec}$$

The photocell device on the MTS velocimeter produces a pulse when the light beam is blocked by the wand and the velocimeter measures the time between the rising and trailing edges of the photocell pulse in order to calculate velocity. For comparison, a 0.375 inch wide aluminum wand was constructed for the MTS velocimeter. Since the pendulum angular velocity ω was constant *all along the beam at any instant in time*, the velocities were scaled as shown:

$$\omega = \frac{v}{r}$$

$$\omega = \frac{v_{SRL}}{r_{SRL}} = \frac{v_{MTS}}{r_{MTS}}$$

where $r_{SRL} = 59.039$ inches and $r_{MTS} = 56.039$ inches. Thus,

$$v_{SRL} = \frac{r_{SRL} \times v_{MTS}}{r_{MTS}}$$

$$v_{SRL} = 1.0535 \times v_{MTS}$$

Ten drops were made at each angle. The results of this comparison study are summarized in Appendix D.

Prior to these tests, the MTS velocimeter was calibrated according to manufacturer's instructions by measuring the internal clock frequency with a scope and frequency counter and comparing the measured results with specifications. The MTS photocell/velocimeter system was then calibrated on a turntable at 33 rpm, 45 rpm, and 78 rpm. The lower speeds were verified using the internal strobe; the highest speed was verified with an external strobe. For these measurements, a 0.753 inch wide wand was utilized and the photocell beam was positioned at a 16.75 inch radius from the center of the turntable. The results of these measurements are summarized in Appendix D.

The impact velocity v_{sp} at the pendulum strike plate, which is mounted $r_{sp} = 65.25$ inches from the pivot pin, is calculated as described above:

$$\omega = \frac{v_{wand}}{r_{wand}}$$

$$\omega = \frac{v_{sp}}{r_{sp}}$$

so that

$$v_{sp} = \frac{v_{wand} \times r_{sp}}{r_{wand}}$$

4.6 Drop Height Determinations

Attempts were made to measure the drop heights required for the specified impact velocities. The Hybrid II and III head-neck systems were assembled and mounted on the pendulum, released from the 1st (lowest), 4th, 8th, and 12th hole of the release mechanism, measured from the bottom. Velocities were calculated from the two velocimeter counters. An increase in drop height by four holes increased the velocity by approximately 0.5 ft/sec. For the Hybrid II, an angle of approximately 112 degrees relative to the vertical was required to achieve the specified impact velocity. For the Hybrid III, an angle of approximately 117 degrees relative to the vertical was required to achieve the specified impact velocity for the flexion test. An angle of approximately 95 degrees relative to the vertical was required to achieve the specified impact velocity for the Hybrid III extension test.

Basically, the release mechanism proved unstable and awkward. The need to manually lift the pendulum to the required height was also awkward and potentially dangerous. It was decided that a new hoist and release mechanism should be designed and built to eliminate these problems.

4.7 Honeycomb Precrush/Pendulum Deceleration Profile Studies

Preliminary measurements of pendulum strike plate deceleration pulses were made. Since the accelerometer outputs were connected directly to an oscilloscope, there was a considerable amount of ringing in the signals with poor resolution to assess compliance with the standard. The actual tests require signal filtering according to SAE recommended practice J211. So, it was suggested that the required signal conditioners and filters be designed and built by SRL. A more sophisticated digital data acquisition system was also recommended. Attempts made to determine the effects of honeycomb precrush proved to be of neglible value. Other possible variations to be considered in an attempt to get an acceptable acceleration profile are stretching and compressing the hexcell to vary the number of cells contacted by the honeycomb and changing the orientation of the cells.

4.8 Pendulum Test Data Acquisition, Storage, Processing, and Analysis

Up to sixteen channels of analog data shall be collected during each test. All data channels shall conform to the SAE recommended practice J2ll. The pendulum deceleration data shall be processed using an SAE Class 60 filter, the head acceleration data shall be processed using an SAE Class 1000 filter, and the neck response data channels shall be processed using an SAE Class 180 filter. The analog data will be stored on magnetic tape, digitized, and processed using a data package on the PDP 11/34. A completed data sheet, time histories of the primary response variables, and other relevant plots shall be generated.

Photogrammetric data will be recorded using a high-speed 16mm camera. After the high-speed films have been processed, the photogrammetric data will be digitized by the Automatic Film Reading System. Displacement-time histories of the designated photometric targets will be generated and stored on magnetic tape or on a PDP 11/34 disk. The pre-test configurations will be documented by still photographs.

A log will be maintained which identifies the transducer, filters, and associated electronics, and the sensitivities, gains, pre-and post-test calibration values, and processing parameters for each data channel. Records will also be maintained describing the model, lenses, positions, and speed of the photogrammetric cameras, the positions of the photometric targets mounted on the test objects, and the analysis techniques. Records of the magnetic tapes and disks will indicate the test type, test number, and parameter identification.

4.9 Pendulum Test Procedures

Initial test procedures shall be established using, as models, existing DOT specifications for Part 572 and Hybrid III dummy compliance testing. Hybrid II and III test procedures and performance standards are described in the Title 49, Part 572 and Part 572 Amended, respectively. The comprehensive evaluation of manikin head-neck responses may involve testing in other impact directions at various impact velocities. Some tests will be designed to assess the effects of three-potentiometer devices, in two different mounting configurations, on head-neck kinematics and dynamics.

5. SUMMARY

Data sets of the Hybrid II manikin head-neck systems were developed for the ATB and HSM Models. The Part 572 Head-Neck Pendulum Compliance Test, of the Code of Federal Regulations, was simulated to validate these data sets. Parameterizations were performed to assess the effects of changes in neck joint characteristics upon the model responses.

Modifications were made in the HSM program source code in order to model the asymmetric bending properties of the Hybrid III neck in flexion and extension. Data sets of the Hybrid III manikin head-neck system were developed for the HSM Model. The Amended Part 572 Head-Neck Pendulum Test was simulated to validate these data sets. Small differences were observed between the responses of the 1-element and 4-element necks. The model, which uses experimental data, proved to be a valid predictor of the manikin head-neck kinematics and dynamics. Tests were conducted to measure the stiffness of the occipital condyle nodding block joint in flexion and extension. Data sets were modified to incorporate this joint. The Pendulum Test simulations indicated that its effect on Hybrid III manikin head-neck kinematics and dynamics is negligible.

As part of the SFRP and RIP, Paver and Doherty conducted experimental studies of the Hybrid II, modified Hybrid II, and Hybrid III dummy head-neck assemblies and the Part 572 and Amended Part 572 head-neck compliance tests. A test plan was written. Initially, the standard head-neck compliance tests were to be performed to verify the operation of the experimental setup. Tests at different impact velocities were planned to validate data sets for mathematical models. Measurements were made of the geometric and inertial properties of the test specimens and apparatus. Mounting plates, test fixtures, and apparatus modifications were made. All transducers were wired; some were calibrated. The test specimens and apparatus were assembled. The velocimeter was modified and calibrated. Attempts were made to determine the drop heights required for the specified impact velocities. Preliminary measurements of pendulum strike plate deceleration pulses were made; the effects of honeycomb precrush were investigated. Several recommendations for modifications of the experimental system were suggested.

6. RECOMMENDATIONS FOR FUTURE STUDIES

The following are recommendations for future work:

(1) conduct tests to measure the stiffness and damping of a single Hybrid III neck element in order to verify the bending parameters used for the 4-element neck of the HSM data set.

(2) modify the HSM program source code to incorporate a true torsional element.

(3) continue tuning the proposed Hybrid II and III data sets by additional validation studies. The Part 572 and Amended Part 572 pendulum tests validated the bending parameters of the manikin head-neck structures. Next, the axial and lateral properties of the Hybrid III and torsional properties of the Hybrid II and III system should be studied. Vertical acceleration or sled tests could be used to determine the neck axial stiffness and damping. Other test modes could be utilized to derive values for the torsional stiffness and damping of the neck.

(4) develop and validate data sets of human and other mechanical head-neck systems

(5) compare dummy responses with human volunteer and cadaver data to assess the biofidelity of these systems.

ACKNOWLEDGMENTS

This research was sponsored by the Air Force Office of Scientific Research/AFSC, United States Air Force, under Contract F49620-85-C-0013. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

7. REFERENCES

1. Doherty, B.J.; Paver, J.G.: A Computer Simulation of the Hybrid II Manikin Head-Neck System. Proceedings of the 24th Annual Symposium of the Safe Association, 1986.

2. Doherty, B.J.; Paver, J.G.: Mathematical Modeling of the Hybrid III Manikin Head-Neck Structure. Proceedings of the International Conference on Mathematical Modeling, 1987.

3. Paver, J.G.; Fishburne, B.: The Prediction of Hybrid II Manikin Head-Neck Kinematics and Dynamics. Proceedings of the 25th Annual Symposium of the Safe Association, 1987.

4. Code of Federal Regulations, Title 49, Part 572 and Amended Part 572.

5. Paver, J.G.: The Biomechanics of Head and Neck Injury and Prevention. Ph.D. Dissertation, Duke University, Durham, N.C., 1985.

6. Belytschko, T.; Schwer, L.; Schultz, A.: A Model for Analytical Investigation of Three-Dimensional Head-Spine Dynamics. NTIS Report #AD-A025-911, April 1976.

7. McHenry, R.R.; Naab, K.N.: Computer Simulation of the Crash Victim – A Validation Study. Proceedings of the 10th Stapp Car Crash Conference, SAE PAPER #660792, 1967.

8. Bartz, J.A.: Development and Validation of a Computer Simulation of a Crash Victim in Three Dimensions. Proceedings of the 16th Stapp Car Crash Conference, SAE PAPER #720961, 1972.

9. Fleck, J.T.; Butler, F.E.; Vogel, S.L.: An Improved Three-Dimensional Computer Simulation of Vehicle Crash Victims. Calspan Corporation, NTIS #PB-241 692-695, April 1975.

10. Butler, F.E.; Fleck, J.T.: Validation of the Crash Victim Simulator. Calspan Corporation, DOT Report #DOT-HS-806 279-282, March 1982.

11. Butler, F.E.; Fleck, J.T.: Advanced Restraint System Modeling. Report #AFAMRL-TR-80-14, May 1980.

12. Miller, J.S.: Performance Evaluation of the General Motors Hybrid II Anthropomorphic Test Dummy. NTIS Report #PB-224-005, Department of Transportation Report #DOT-HS-800-919, September 1973.

13. Hubbard, R.P.; McLeod, D.G.: Geometric, Inertial, and Joint Characteristics of Two Part 572 Dummies for Occupant Modeling. Proceedings of the 21st Stapp Car Crash Conference, SAE PAPER #770937, 1977.

14. General Motors Corporation: Hybrid III Quality and Performance Qualification Manual. Safety Research and Development Laboratory, GM Proving Grounds, Milford, MI.

15. Foster, J.K.; Kortge, J.O.; Wolanin, M.J.: Hybrid III - A Biomechanically Based Crash Test Device. Proceedings of the 21st Stapp Car Crash Conference, SAE PAPER #770938, 1977.

16. Sierra Drawing #73051-3, 9 April 1973.

17. Piziali, R.A.: An Evaluation of the Performance Characteristics of Anthropomorphic Test Devices - Volume I. NTIS Report #PB-222-691, Department of Transportation Report #HS-800-869, June 1973.

18. Kaleps, I.; White Jr., R.P.; Beecher, R.M.; Whitestone, J.; Obergefell, L.A.: Measurement of Hybrid III Dummy Propert'es and Analytical Simulation Data Base Development, in preparation.

Appendices can be obtained from
Universal Energy Systems, Inc.

FINAL REPORT NUMBER 88
REPORT NOT AVAILABLE AT THIS TIME
Dr. Donald Robertson
760-7MG-094
760-7MG-075

LEARNING BEHAVIOR OF ADAPTIVE FILTERS FOR EVOKED BRAIN POTENTIALS

John J. Westerkamp
Janet L. Slifka

PREPARED FOR:

Universal Energy Systems, Inc.
Dayton, Ohio

on

Subcontract No. F49620-85-C-0013
Air Force Office of Scientific Research
Bollings Air Force Base, D.C. 20332

Final Report for Period 1 January 1988 to 31 December 1988

December 1988

Signal Processing Laboratory
Department of Electrical Engineering
The University of Dayton
Dayton, Ohio

ACKNOWLEDGMENTS

The authors would like to thank Dr. Thomas J. Moore for providing the opportunity to investigate the adaptive filter for processing human electrical brain activity. The authors would also like to thank Tim Anderson for his assistance during this research.

Finally, the authors would like to thank the Air Force Office of Scientific Research for funding this research and Universal Energy Systems, Inc., for administrating the contract.

ABSTRACT

An in-depth analysis of the learning behavior of the least-mean-square (LMS) adaptive algorithm was performed. Simulation studies show that the learned filter coefficients decay at a rate proportional to the convergence constant following the disappearance of the underlying signal. This conflicts with the need for rapid convergence when the signal changes as is the case for the evoked potential. Studies involving mean-square error measurements show that significant improvements to weight retention can be obtained by using a "gating" algorithm which updates the weights only when the signal is present. A symmetric noncausal format for the weights was shown to extract the initial peak of the signal with the most consistency. After investigating several filter configurations, averaging was found to function approximately as well as each filter when attempting to estimate the single-response evoked potential. Yet, for the cases when detection of loss of the signal is most crucial, the adaptive filter offers more advantages than averaging.

1. INTRODUCTION

Adaptive filters, specifically those using the least-mean-square (LMS) algorithm, have been studied extensively for well-behaved signals. Very few in-depth studies have been made involving nonstationary or transient signals. Primarily this is due to their intractable forms for analysis.

The evoked potential (EP), an electrical signal produced by the brain in response to some effective sensory stimulus, is a transient nonstationary signal. When recorded, these brain responses occur for a short duration, die away, and then occur again at the next stimulus presentation. This constantly changing signal environment makes the adaptive filter an attractive solution to EP estimation because it avoids a long and complicated design process by designing itself. Yet at the same time the EP presents several problems to adaptation. A nonstationary signal does not have a single optimum set of weights (Widrow 1976). As the signal statistics change, so does the optimum set of weights. This means that only under special circumstances can the weights track close to the current optimum and achieve near Wiener filter results (Widrow 1976,Gardner 1987). A second problem, the transitory nature of the signal, is almost a greater detriment to convergence. When the signal is present only in short duration bursts, the weights have a limited time to converge before the signal is gone. In essence, the adaptive filter is faced with a new system when the signal stops and only noise is present at the input. As will be shown in succeeding sections, the weights decay away from their previous nearly-converged values toward values corresponding to the new system's correlation statistics. This means that at least the first portion of the signal will be lost to the converging process (Westerkamp 1987).

Studies aimed at correcting this problem by speeding up the convergence process have been conducted by Westerkamp (Westerkamp 1987). Westerkamp's research raised questions about the learning behavior of the LMS adaptive filter. The purpose of this follow-up research effort was to obtain that knowledge and then use it to develop a modified algorithm that would optimally learn and retain the weights.

Section II presents a brief review of related research as found in existing literature. Following that in section III is an intensive discussion of the signal. The development of the modified algorithm is closely tied to the signal characteristics and understanding those is essential to finding an optimal solution. Section IV contains a summary of the adaptive linear combiner and the LMS algorithm. Section V examines the theoretical analysis of the weight deterioration behavior and Section VI presents the derivation of the modified algorithm. Results from applying the modified algorithm to simulated evoked potential buried in prestimulus electroencephalogram (EEG) and actual human visual evoked potential are contained in section VII. The final section, Section VIII, is a discussion of results.

2. LITERATURE REVIEW

Existing literature has provided a broad range of background material. The following is a brief summary of these resources with pertinent observations regarding the present research.

The visual evoked potential is buried within the on-going EEG. With a signal-to-noise ratio (SNR) reaching -6 dB, the EP is difficult to view without processing. Based on the random nature of the background EEG, response averaging has been traditionally used for EP estimation. Averaging assumes that the scalp-recorded response is time-locked to the stimulus and identical in both latency and contour from one stimulus to the next. When consecutive responses are averaged, the spontaneous EEG will cancel as the EP is enhanced. The problems with this approach are that several hundred responses are required to obtain one estimate and averaging can obscure meaningful variations existing in the single-response and between different responses.

Attempts to estimate single-response evoked potentials have centered on the use of *a priori* information. One approach has been to examine pre-stimulus EEG and create a filter that attempts to cancel any components in the post-stimulus recording that are predictable from pre-stimulus statistics. Results from this focus on reducing the number of responses needed for averaging to obtain a usable estimate (Krieger 1986).

Another method involved designing an optimum linear time-varying filter based on *a posteriori* estimates of the correlation matrices assuming that the EEG is zero-mean and uncorrelated with the EP. This research, conducted by Westerkamp and Aunon (Westerkamp 1987), provides the best to-date estimates of the single-response EP and will be used for comparison with the results of this research. In this approach, a multielectrode structure was used. Scalp-recorded data from each electrode were passed through a time-varying filter. A time-varying filter possesses a different impulse response at each time instant. The filtered results from each electrode were summed to produce the final EP estimate. This work makes the assumption that a series of recorded responses are available for analysis and design, and that it is these responses which will be filtered using the final design. If responses outside this set were passed throughout the filter, performance could deteriorate. Also, the intense post-processing involved makes this unsuitable for real-time applications.

The challenge is to find a way to estimate the single-response EP with a minimum amount of *a priori* information so that real-time processing could take place. This naturally leads to the use of adaptive filters. This incredibly powerful tool designs itself, eliminating the need for extensive *a priori* statistical knowledge and time-consuming design processes, and can adapt automatically as the signal changes. Madhavan proposed using adaptive noise cancellation techniques for estimating the EP (Madhavan 1984). In this structure, the noise in the desired channel is assumed to be correlated with the noise in the input channel. Promising simple simulation results were followed up with more complicated simulations involving crosstalk (Madhavan 1986). For scalp-recorded data it is not possible to record EEG at the time of the EP that doesn't contain some signal that is correlated with the EP. This produces crosstalk within the data recorded at the electrodes. Madhavan used a three-stage Weighted Least Squares Lattice (WLSL) structure to cancel the effects of crosstalk and effectively estimate the EP. In this set-up, the weights were updated when no signal was present and held constant when the signal was present. Significant improvements for SNR in simulated data were obtained. This implementation, however, has a high degree of complexity which may make it unsuitable for real-time applications and has not been tested on human data.

The LMS algorithm seems most attractive for its simplicity. In a recent paper by Thakor (Thakor 1987), the LMS algorithm was applied to the evoked potential signal. In this system, the final output was the result of averaging previous outputs. As the number of averaged outputs increases, the system is subject to the same constraints that conventional averaging carries in terms of distortion of single-response characteristics. In a correspondence on this article, Madhavan (Madhavan 1988) points out that any averaging done before the signal is input to the channels introduces unremovable distortions in the output. He also analyzes the application of the multichannel adaptive signal enhancer (MASE) structure proposed by Ferrara and Widrow (Ferrara 1981). He states the requirement that correlation must exist between channels but that identical signals within the channels is an uneffective filtering method. This same structure was further investigated by Westerkamp (Westerkamp 1987). In this case, the LMS was applied to the EP using the standard transversal filter in the MASE format. The major problem found with the LMS was that the initial peaks of the EP were lost to the converging process. Attempts were made to speed up the convergence rate using lattice and Fast Fourier Transform (FFT) methods that orthogonalize the data to remove the eigenvalue spread. Results continued to show loss of signal components in the early peaks to the converging process. This established the need to understand the learning and extinguishing behavior of the LMS algorithm.

A follow-up research effort was proposed with two major goals:

(1) characterize the learning behavior of the LMS algorithm;

(2) create a modified algorithm capable of filtering nonstationary and transient signals such as the EP.

As stated in the proposal for this research effort, understanding the learning behavior encompasses various factors such as the effects of filter length and number of reference channels in the multichannel structure, and generally quantifying the extinguishing behavior of the weights (Westerkamp 1987). These objectives required further pursuit of the literature. Although many papers have been written concerning the convergence of the LMS algorithm, it is best to concentrate on those relating to the nonstationary signal since the EP is nonstationary. The standard transversal LMS filter in the nonstationary environment functions well for slowly-varying nonstationarities (Widrow 1976). In fact, the LMS performs as well as the exact least-squares for certain nonstationary problems (Widrow 1984). Gardner provided in-depth analyses of the nonstationary problem by examining the system identification problem (Gardner 1987). Unfortunately the EP is highly nonstationary and buried in noise. It is not known exactly what the EP looks like since the only reference is estimated. Obtaining statistics is not only difficult for nonstationary signals but also memory and processor intensive. Even collecting accurate and clean data in large quantities is difficult due to the extreme measures that must be taken to avoid noncerebral and unwanted artifacts from appearing.

Basic analyses of the weights as presented by Widrow and Stearns (Widrow 1985) provided a start for analyzing the learning behavior of the LMS and extracting key ideas from available literature proved useful in designing a modified algorithm. This modified algorithm must retain the learning of the weights so that reconvergence of the algorithm is not required with each new response. Since one of the variable factors in the response is latency, unwanted phase distortion must be minimized by the filter. Noise must be effectively cancelled when the EP is not present so that the response may be clearly viewed. The system must also operate at near real-time speed for efficient use of the responses. Among the key ideas available from literature were some variations on the LMS. These included a variable-step LMS (Harris 1986), a lattice structure (Griffiths 1979,Savoji 1987), a complex LMS algorithm (Fisher 1983, Sherwood 1986), a frequency domain algorithm (Lee 1987), and a time-sequenced algorithm (Ferrara 1981). Each modification added its own degree of complexity to the solution. It is necessary to keep the complexity to a minimum so that processing can present a near real-time output. After analysis and testing, the final modified algorithm borrowed ideas from several areas. From the WLSL came the idea of turning learning on or off, which also borrowed from neural networks the concept of training records to establish the weights (Rumelhart 1986). The necessity of minimizing phase distortion lead to a noncausal form for the filter weights. Finally previous promising work with the multichannel system suggested that it be maintained.

3. SIGNAL ATTRIBUTES AND SIMULATION

When designing, modifying, or simply applying an algorithm to a set of data, it is essential to understand the signal and its properties. If this understanding is present, optimal choice of algorithm can be made, distortion of the output can be avoided, and results can be properly evaluated. For this study, a complete analysis of both the EEG and the EP is necessary. Several areas were researched: the auto- and cross-correlations of the EEG, the autocorrelation function and the power spectral density of the actual EP and a simulated EP, and the averages over many records of the EP and EEG.

When collecting visual evoked potential data, the subject views a patterned stimulus and electrodes attached to the scalp record the electrical response of the brain. This recorded signal has two parts -- the evoked potential which measures the dynamic response of the brain to the stimulus, and the EEG which represents the on-going brain activity. The peak amplitudes and latencies in the individual EP are highly sensitive to many factors, especially to the pattern of stimulus presentation and the subject's visual acuity. If the stimuli are presented too quickly, the evoked response deteriorates as successive responses overlap. Because of this problem, there must be a suitable pause between stimulus presentations. The recorded data then consist of short duration "bursts" of signal separated by longer periods of noise only -- the EEG. The EEG itself tends to obscure the EP such that it is usually impossible for the eye to discern. The evoked potential can be a complicated series of peaks, however, a rather simple response is obtained for visual stimuli. An appropriate checkerboard pattern stimulus produces a large amplitude response (Eason 1970). It is that response which will be examined here. This lower checkerboard stimulus produces a response with a positive peak approximately 85 msec after the stimulus, a negative valley approximately 100 msec after the stimulus and a positive peak with a 185 msec latency. The total duration of the signal is approximately 300 msec. The peak amplitudes and latencies of these components vary, however, from one response to the next (Brazier 1964). Shown in Figure 1 is the average of 100 responses with all peaks labeled. Each data record contains 375 points which were obtained at a sampling frequency of 250 Hz. The stimulus was presented at sampling point 125.

Ep data were collected from five human subjects during a checkerboard flash experiment. The data from two of those subjects (one female and one male) were analyzed in this study. The subjects were seated comfortably in a soundproofed and electromagnetically shielded booth. The booth was dimly lit with direct current ls to eliminate 60 cycle interference within the booth. The subjects viewed a CONRAC video monitor from a distance of 1.5 m. The video display subtended 11.5° (horizontal) by 8.5° (vertical) of visual angle. The experiment timing was controlled by a microprocessor system which triggered the analog-to-digital (A/D) conversion and the stimulus presentation. A Grass pattern generator produced checkerboard patterns with check size 18 min (256 checks in a full display) which appeared on the CONRAC monitor. The stimulus parameters were chosen in an attempt to elicit maximal amplitudes in the average EP's (Eason 1970). Prior to the experiment, the intensity of the visual display was measured and adjusted to 6 ft · lamberts for both the stimulus on and blank screens. This was to ensure that no flash response would contaminate the pattern response due to changes in the intensity. Beckman Ag/AgCl EEG electrodes were applied with conductive paste to scalp locations *Pz* and *Cz* according to the International Federation "10-20" system (Jasper 1958), of which the *Pz* data will be examined here. Monopolar recordings were made at each electrode with reference to electrically linked mastoids. The forehead was used to ground the subject. An eye channel (EOG) was included to monitor eye blink artifact. Electrode impedances were measured before the experiment and were found to be below 10 kΩ.

It has been previously assumed that the EEG is uncorrelated from one response to the next. This assumption is necessary for the proper functioning of the LMS algorithm. If some correlation did exist then the algorithm would find it and the evoked potential estimate would be distorted. Time domain techniques were used to estimate the autocorrelation function and cross-correlation functions of the EEG. As seen in Figure 2(a), the data are uncorrelated after about 15 samples and the EEG has a variance or noise power of about 40. Figure 3(a) shows the Fast Fourier Transform (FFT) of the windowed autocorrelation function estimate. This is the power spectral density of the EEG.

It is only possible to obtain quantitative measures of the performance of a filter if the desired output is exactly known. If that signal is defined, the filter output can be compared to the ideal output and statistics can be collected. For this research the evoked potential was simulated with a series of raised cosines with random peaks and latencies. This method of simulation is not physiologically accurate but does provide a relatively simple and clean signal for analysis. Generation was done utilizing a program developed by Aunon and McGillem (McGillem 1985) for their investigations into the visual evoked potential. The peaks and latencies have a zero-mean unit-variance Gaussian distribution centered around entered values for mean and variance. The means and variances used to simulate the evoked potential were estimated from latency-corrected average (McGillem 1985). These values are shown in Table 1 and the values computed from the generated signals are shown in Table 2. A few typical simulated responses are shown in Figure 4.

While visually this appears to approximate the form of the average signal shown in Figure 1, it is also important that they share similar autocorrelation functions as well as power spectral densities since for testing, this signal will be buried within the prestimulus EEG. The removal of noise that overlaps in frequencies with the desired signal is one of the difficulties the filter must attempt to overcome. To assure this time-domain techniques were used to estimate the autocorrelation function of both the simulated EP and the actual averaged EP. The results are shown in Figure 2(b). The FFT was used to achieve the power spectral densities of these two signals and can be seen in Figure 3(b). The actual EP can be seen to have a slightly different magnitude in its power spectrum than that in the simulated EP. The simulated EP has a power spectrum that overlaps completely the frequency range of the actual EP spectrum although differences exist in amplitude.

For testing purposes, the simulated EP was added to the 125 points of prestimulus EEG that is at the beginning of each data record. Having actual EEG as the noise for the simulation was considered to present a more accurate representation of the actual situation rather than using a generated noise such as white noise. The drawback to this decision is that only 125 points of noise were available. This means that the simulated EP must have a shorter duration than the actual signal in order to have proportionate durations of noise before and after the signal is present. These noise portions are essential since the filter will be noncausal. When generating the first signal output point, one-half of the weights will be on prestimulus EEG and one-half will be on the signal. Because of this, at least one-half the maximum number of filter weights of EEG data points must be present before the simulated EP begins. For the data generated here, the earliest simulated EP begins at point 31. The maximum filter length tested was 65 due to these same data constrictions.

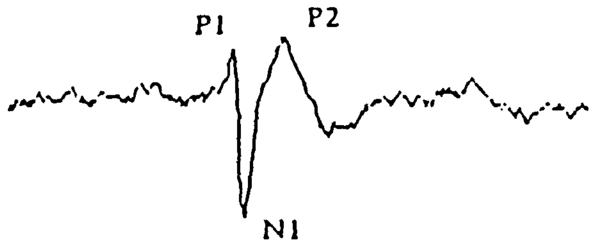


Figure 1. Average of 100 Human Visual Evoked Potential Responses Collected at Electrode *Pz* (Lower Checkerboard Stimulus).

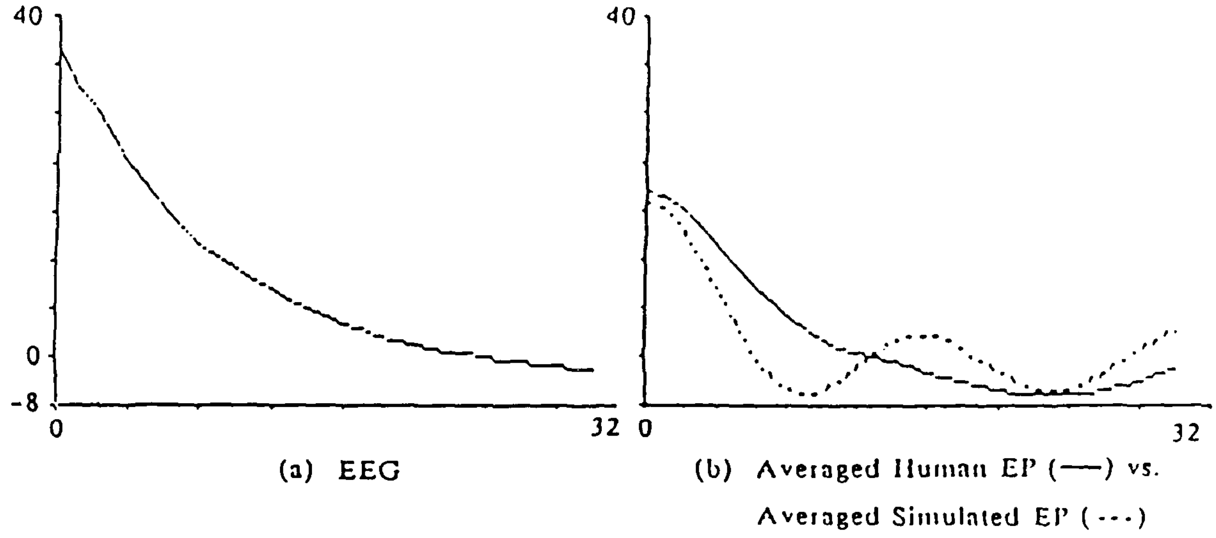


Figure 2. (a) Estimate of the Autocorrelation Function of the EEG from 125 Points of Prestimulus Data, (b) Estimate of the Autocorrelation Function of the Averaged Human EP and the Averaged Simulated EP.

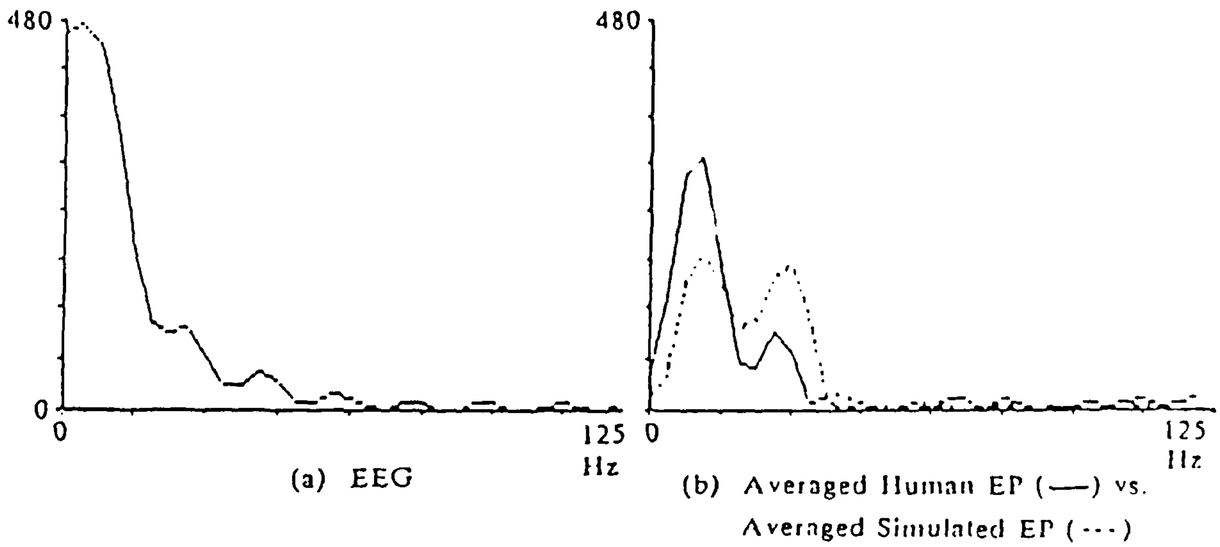


Figure 3. (a) Estimate of the Power Spectral Density of the EEG from the Autocorrelation Function of Figure 2a, (b) Estimate of the Power Spectral Density of the Averaged Human EP and the Average Simulated EP from the Autocorrelation Function of Figure 2b.

Peak	Latency(ms)	Latency St.Dev.(ms)	Amplitude(μV)
1	74.14	7.41	6.28
2	111.60	6.65	-13.07
3	146.87	8.99	3.11
4	170.00	7.81	-2.22
5	202.95	11.32	8.42

TABLE 1. Latency Corrected Averaging Values Used for Simulating the EP.

Peak	Latency(ms)	Latency St.Dev.(ms)	Amplitude(μV)
1	75.06	6.94	6.45
2	111.52	6.28	-13.16
3	145.50	9.07	3.20
4	169.55	6.78	-2.34
5	202.33	10.37	8.59

TABLE 2. Computed Mean and Variance Values from Simulated EP.

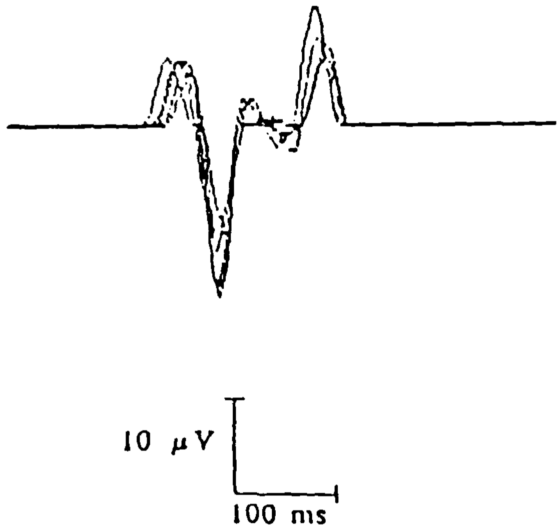


Figure 4. Examples of the First Five Simulated EP Responses.

4. THE LMS ALGORITHM AND FILTER STRUCTURE

The adaptive linear combiner as seen in Figure 5 is the basic form for the implementation of the LMS algorithm. This simple structure, known as a tapped delay line, derives the output signal from a weighted sum of the input signals. For this application, the input signals are the last L+1 samples of the same signal, where L+1 is the number of filter weights. The inputs at the kth sampling instant can be denoted by the vector

$$\mathbf{X}_k = (x_k, x_{k-1}, \ldots, x_{k-L})^T \tag{1}$$

where T denotes the vector transpose and boldface denotes a vector quantity. The filter weights can be represented in a similar vector,

$$\mathbf{W}_k = (w_{0k}, w_{1k}, \ldots, w_{Lk})^T . \tag{2}$$

The output sample y_k is computed as

$$y_k = \mathbf{W}_k^T \mathbf{X}_k = \mathbf{X}_k^T \mathbf{W}_k . \tag{3}$$

The primary channel contains the signal d_k known as the desired response. Within this channel is an underlying signal correlated in some way to the input in the secondary channel, $\mathbf{X}_k$. The adaptive filter will use this correlation to estimate that underlying signal. The error signal is the difference between the desired response and the filtered output,

$$\varepsilon_k = d_k - y_k . \tag{4}$$

The LMS algorithm performs as the name implies -- it attempts to minimize the mean square error (MSE). As shown this error is represented by equation (4) and can be written using (3) as

$$\varepsilon_k = d_k - \mathbf{W}_k^T \mathbf{X}_k = d_k - \mathbf{X}_k^T \mathbf{W}_k . \tag{5}$$

Assuming that the weights are fixed at some value and computing an expression for this mean square error,

$$\varepsilon_k^2 = d_k^2 - W^T X_k X_k^T W - 2d_k X_k^T W . \qquad (6)$$

The expected value of this expression, assuming ε_k, d_k, and X_k are all statistically stationary, is

$$E[\varepsilon_k^2] = E[d_k^2] - W^T E[X_k X_k^T] W - 2E[d_k X_k^T] W . \qquad (7)$$

The matrices $E[XX_k^T]$ and $E[d_k X_k^T]$ are commonly defined as **R** and **P** . The matrix **R** is the autocorrelation matrix of the input and **P** is the cross- correlation matrix between the input and the desired. Now,

$$E[\varepsilon_k^2] = E[d_k^2] - W^T \mathbf{R} W - 2\mathbf{P}W . \qquad (8)$$

The mean square error is a quadratic function of the weights. For a two-weight case, plotting the MSE as a function of the weights would produce a three-dimensional parabola. The minimum MSE would be the bottom of the "bowl" and steepest descent gradient search techniques, from which the LMS is derived, attempt to find this bottom. As the weights converge, the gradient estimate would move down the bowl toward a zero value for the gradient. At that point the weights reach their optimum value and mean square error is minimized. For the highly nonstationary case, the optimum weight values change, causing the shape of the bowl to change as well as move around in the error space. It is desirable to have the algorithm track close to the bottom of the bowl as it moves. In this way, the LMS could stay close to the optimum weights in a way the fixed Wiener filter cannot. For nonstationary inputs, the Wiener filter still attempts to provide an optimum solution. Estimates of **R** and **P** can be made for nonstationary data. Computing

$$\mathbf{W}^* = \mathbf{R}^{-1}\mathbf{P} \qquad (9)$$

where the * indicates the optimal value, then yields the optimum filter impulse response. For the EP+EEG, this solution is a bandpass filter and an estimate of the Wiener filter impulse response using **R** and **P** estimated from the averaged EP of Figure 1 and 100 records of prestimulus EEG is shown in Figure 6(a). The filter sees the signal as moving around within a band of frequencies and decides that the best it can do is to pass that band as seen in Figure 6(b). Unfortunately, as already shown, the frequencies of the EP overlap those of the EEG and a great deal of noise would still make it through the filter.

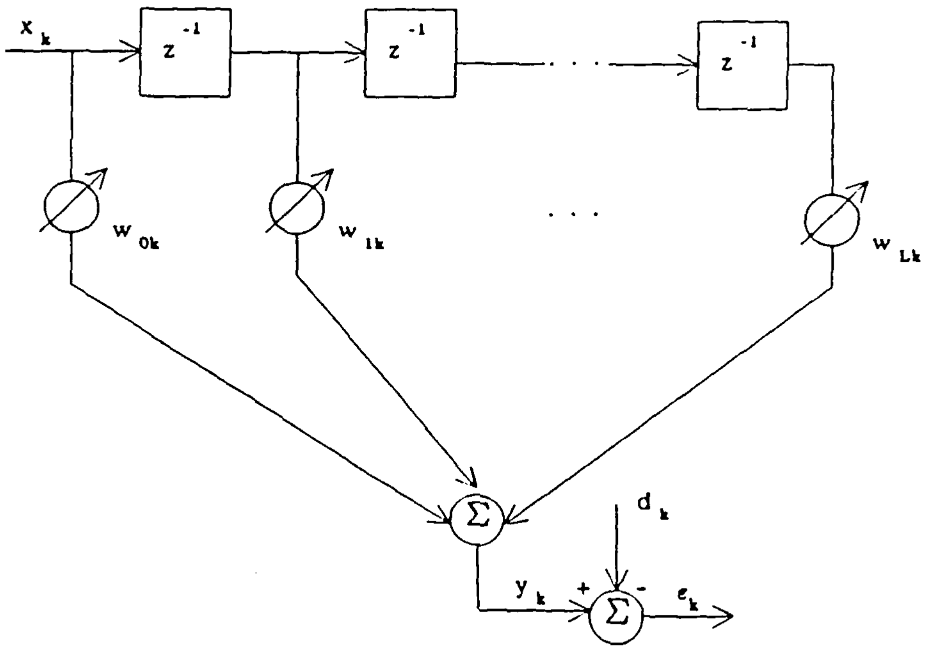


Figure 5. Adaptive Linear Combiner Structure.

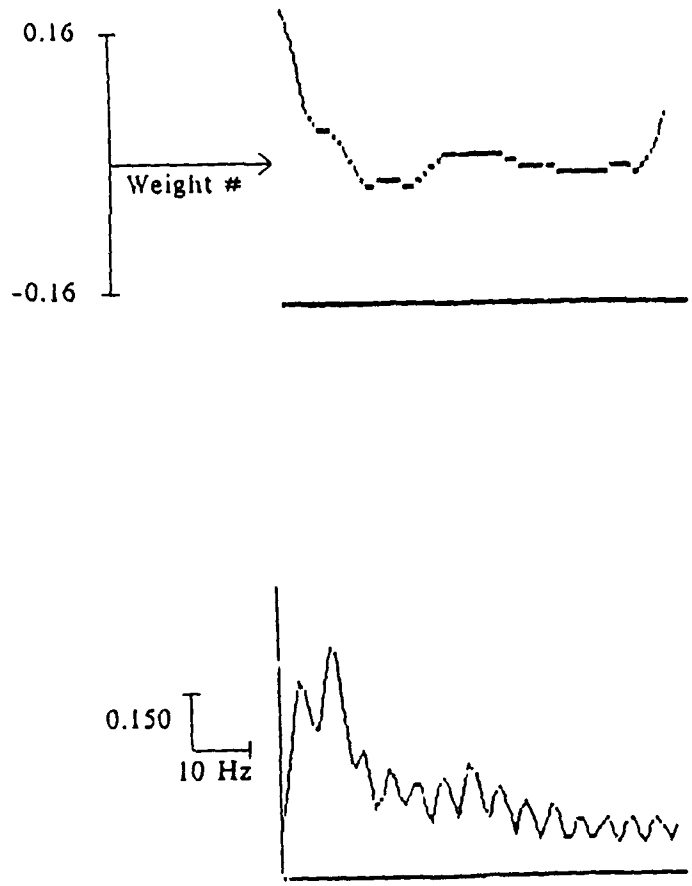


Figure 6. (a) Optimal Wiener Filter Impulse Response Computed from Estimates of R and P, (b) Frequency Domain Representation of the Wiener Filter Estimate.

5. WEIGHT DETERIORATION

The adaptive filter iteratively reaches an optimum set of weights by estimating the gradient and following it down the error surface toward a solution. Steepest descent methods of gradient estimation define the error surface as

$$\varsigma = E[\epsilon_k^2] . \tag{10}$$

The LMS algorithm estimates this value as

$$\varsigma = \epsilon_k^2 . \tag{11}$$

Recalling the definition of the gradient (Widrow 1985),

$$\hat{\nabla}_k = \begin{bmatrix} (\partial \epsilon_k^2 / \partial w_0) \\ \vdots \\ (\partial \epsilon_k^2 / \partial w_L) \end{bmatrix} = 2\,\epsilon_k \begin{bmatrix} (\partial \epsilon_k / \partial w_0) \\ \vdots \\ (\partial \epsilon_k / \partial w_L) \end{bmatrix} = -2\,\epsilon_k X_k . \tag{12}$$

Substitution yields a noisy gradient estimate of the form

$$\hat{\nabla} = -2\,\epsilon_k X_k , \tag{13}$$

and a corresponding weight update equation of

$$W_{k+1} = W_k + 2\mu\,\epsilon_k X_k \tag{14}$$

where μ is the convergence constant.

A simple case would be an input that is signal plus noise and a desired that is another instance of the signal such that the signals in the two channels are correlated and noise exists only in the input channel. For simplicity, assume that the noise is white noise with a power denoted by ϕ .

If the system is running as desired, after some time the weights will converge to an optimum value called W^* . At this point let the two signals be turned off. The weight behavior can be traced as follows:

$$W_{k+1} = W_k + 2\mu \epsilon_k X_k . \tag{15}$$

Examine the expected value of these weights:

$$E[W_{k+1}] = E[W_k] + 2\mu \, E[\epsilon_k X_k] , \tag{16}$$

but using (4)

$$E[W_{k+1}] = E[W_k] + 2\mu \, E[(d_k - y_k) X_k] \tag{17a}$$

$$= E[W_k] + 2\mu \, E[d_k X_k - y_k X_k] . \tag{17b}$$

As stated, the desired channel signal is now uncorrelated to the input channel signal and allows the above equation to be written as

$$E[W_{k+1}] = E[W_k] - 2\mu \, E[y_k X_k] \tag{18a}$$

$$= E[W_k] - 2\mu \, E[X_k X_k^T W_k] \tag{18b}$$

Assume that the weights and the input are uncorrelated.

$$E[W_{k+1}] = E[W_k] - 2\mu \, E[X_k X_k^T] \, E[W_k] \tag{19}$$

But $E[X_k X_k^T]$ is the autocorrelation matrix $\mathbf{R}$. Substituting and regrouping yields:

$$E[W_{k+1}] = (\mathbf{I} - 2\mu \mathbf{R}) E[W_k] , \tag{20}$$

and extending this another point gives

$$E[\,W_{k+2}\,] = (\,I - 2\mu R\,)\,E[\,W_{k+1}\,] = (\,I - 2\mu R\,)^2\,E[\,W_k\,]\,. \tag{21}$$

In general,

$$E[\,W_{k+n}\,] = (\,I - 2\mu R\,)^n\,E[\,W_k\,]\,. \tag{22}$$

Assume the time instant k is the last appearance of the signal before it is turned off. If the system has converged, then at k:

$$W_k = W^*\,, \tag{23}$$

and for samples beyond that occurring at k,

$$E[\,W_{k+n}\,] = W_p^*(\,I - 2\mu R\,)^n\,, \tag{24}$$

where the p subscript on the weight vector indicates the optimal set of weights as pertaining to the previous system when the signal was present. Suppose the system had just one weight. Recalling that **R** is the autocorrelation matrix of the input and the input is white noise then

$$E[\,w_{k+n}\,] = w_p^*(1 - 2\mu\,\phi)^n\,. \tag{25}$$

This shows an exponential decay of the weight toward zero. The rate of that decay is dependent on the power in the noise and the convergence constant. If both numbers are << 1, the weights will deteriorate very slowly. As either the noise power or μ increases, the decay rate increases. An example of the weight trace behavior for a two weight case is shown in Figure 7. In this example, the white noise has a power of 0.01 and μ is 0.05. The signal to be detected is turned off at point sample 600 and turned back on at sample point 1100. When the signal disappears, the weights head toward a value of zero and must reconverge when the signal reappears. Both weights were initially zero for this test.

Does this apply to the evoked potential? When the EP is not present, the desired channel does contain a signal, the EEG. Now assuming that the EEG in the desired is uncorrelated with the EEG in the input produces the following. Rewriting equation (11)

$$W_{k+1} = W_k + 2\mu X_k(dk - X_k^T W_k) \quad (26a)$$

$$= W_k + 2\mu d_k X_k - 2\mu X_k X_k^T W_k. \quad (26b)$$

But taking the expected value causes the $2\mu d_k X_k$ term to become zero and the results are the same. This says that learning is lost for those periods between stimulus response. If those periods were considered by the algorithm to not exist then the filter would have better retention of the learning.

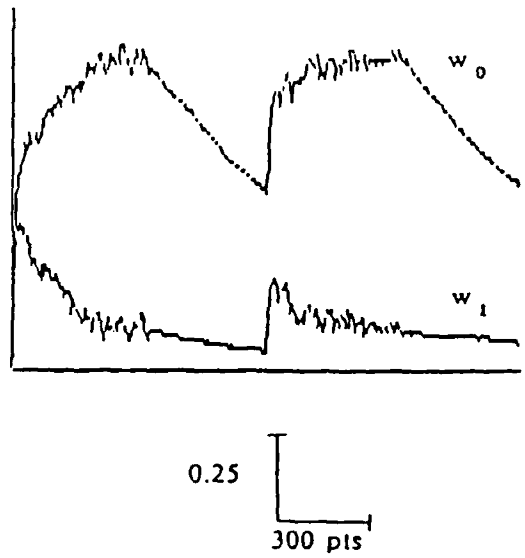


Figure 7. Weight Trace Behavior for Two-Weight Case. Sinusoid in White Noise for 600 Points and Disappearance of Sinusoid for 500 Points. Repeat Cycle.

6. THE MODIFIED ALGORITHM

The following subsections discuss the development of the modified algorithm and the choise of critical parameters.

6.1 BASIC ALTERATIONS

The basic requirements for the modified algorithm were as follows:

(1) Retain learning of the weights
(2) Make a 'good' estimate with as few records as possible (near real time).
(3) Minimize distortion of the EP
(4) Baseline cancellation of the EEG when the EP is not present.

Assuming that the design begins with the LMS algorithm, four levels of modification were added. The basic structure of the filter was chosen to be the adaptive linear combiner in the multichannel format as seen in Figure 8. Additional channels provided several references for the filter to detect correlation and produce an output. Summing these various outputs yields additional noise cancellation as well as reinforcement of EP characteristics. The second level of modification is to define the filter weights to be noncausal giving a linear-phase response for the filter. To retain the learning of the weights, adaptation was gated with a wrap around function of the weights for successive records. Finally, a bias weight acting as a high pass filter was applied for those portions of the input for which it was determined that no EP would be present. The width of the gate was based on previous knowledge of the probable duration of the EP and data outside that window was assumed to be noise to be cancelled with the bias weight.

The multichannel structure was recently investigated in relation to the EP by Westerkamp (Westerkamp 1987). This research recommended use of the Multichannel Adaptive Signal Enhancer developed by Widrow and Ferrara in 1981. As previously stated, this same structure was recently evaluated in terms of the EP by Madhavan (Madhavan 1988) and stressed that proper usage requires the signals in the channels to be correlated but not identical. The structure consists of several separate finite impulse response filters. The output of each of these filters is then summed to form the final output of the system. Each filter finds the correlation between its input signal and the signal in the primary channel, where the noise in each channel is assumed to be uncorrelated with the noise in every other channel. To achieve this, the current record is placed in the primary channel and the secondary channels contain previous records. Before a new response is read into the desired channel, the current data is all shifted down in the structure. This brings the previous desired response into the first of the secondary channels. The EP is highly correlated from response to response even though the actual shape may differ in amplitude and latency. When these various channel outputs are summed, any random noise from the EEG that may have made it through the filter is further reduced in the averaging process.

Reducing distortion is essential to preserve variations in latency of the response peaks. With phase distorted at the output, amplitude can also be affected if output components are delayed to other frequencies. A noncausal finite-impulse response filter will have linear-phase response. For this reason, the weights of the LMS were forced to this response. An odd number of weights was always used to allow one weight to be centered upon the point for which the output was being computed. One-half of the remaining weights applied to data preceding the center point and one-half applied to data succeeding the center point. In the same vein, the weights in this configuration were also forced to be symmetric producing a zero-phase response. This format would have greater applicability if the variations in peak latency were small. The symmetric weights are updated according to the following equation:

$$w_{-l,k+1} = w_{l,k+1} = w_{lk} + \mu \epsilon_k (x_{k+l} + x_{k-l}) \qquad 1 \leq l \leq L. \tag{27}$$

The center weight, labeled w_0 in Figure 9, is updated according to the standard LMS weight update equation.

Gating the learning borrows ideas from neural networks. Neural networks have to be trained before they produce a usable output. Applying several records to the LMS and wrapping around the weight values from record to record allows the weights to converge close to an optimum value. The weights in succeeding updates can still change and adapt but have a reasonable base from which to work. The gating works by only adapting when the EP is present and taking the weight values reached at the end of the window and using them for the previous weight value in the update of the first point of the next response. In this way the filter appears to see a continuous stream of evoked responses. This should prevent the extinguishing behavior of the weights and leave the filter to tackle the still difficult problem of tracking a highly nonstationary signal.

6.2 EEG BASELINE CANCELLATION

Because the EP signal occurs only in response to a stimulus, it is expected that the signal will occur within a certain time window following the stimulus. If the filter could then cancel the EEG outside this window as well as estimate a clear EP within the window, the estimate would be easily identified. However, the LMS weights are like any other filter -- inputing a strong spike will produce an impulse response. Also, the frequencies of the EP overlap those found in the EEG. Combined, these allow some signal to appear at the output expected. This leads to the application of a bias weight.

The configuration for the bias weight is shown in Figure 10. This weight attempts to cancel the current DC value and in the optimum sense would produce a flat line response outside of the EP window. To better understand the bias weight performance, it is useful to derive a transfer function.

The desired function is that of a high pass filter. For the configuration shown, the weight is updated for each noise point with the following equation:

$$w_{bk+1} = w_{bk} - y_{final\ k}, \tag{28}$$

where $y_{final\ k}$ is the final sum of each of the outputs of the secondary channels with the bias weight already added. Transferring this equation to the z-domain, yields a transfer function of the following form, with the sum of the output channels before addition of the bias weight denoted as $Y_{summed}(z)$,

$$H_{wb}(z) = \frac{Y_{final}(z)}{Y_{summed}(z)} = \frac{z-1}{z}. \qquad (29)$$

This function has a zero on the real axis at one and a pole at the origin. A plot of the transfer function and phase response of this fixed filter is shown in Figure 11. As can be seen, this is simply a high pass filter and should effectively block the predominantly low frequencies found in the recorded biological signals due to artifacts such as eye blinks and baseline drift. This filter is used on all simulated and actual response filtering presented in the results section of this paper and provides excellent baseline cancellation of the EEG.

6.3 THE CONVERGENCE CONSTANT

Although the LMS filter adaptively designs itself, certain parameters must still be determined. Among these are the filter length, the number of channels, and μ, the adaptive feedback constant. The last of these, μ, controls the convergence rate of the filter weights. This rate of convergence is inversely proportional to μ; a large value would produce the desirable trait of fast convergence. However, the convergence constant also determines the amount of excess mean square error (MSE) of the final solution as compared to the optimal Wiener solution. This same large μ would now produce a larger error at the converged solution, known as misadjustment.

As given with the derivation of the algorithm, in order to assure convergence of the algorithm:

$$1/\lambda_{max} > \mu > 0\ , \qquad (30)$$

where the parameter λ_{max} is the largest eigenvalue of the autocorrelation matrix, **R** . However, as previously mentioned, the goal here is to avoid the requirement of *a priori* statistical information. To this end, an alternate expression of equation (30) can be used as follows:

$$(1/tr[\mathbf{R}]) > \mu > 0\ . \qquad (31)$$

The tr[R] is the sum of the eigenvalues of R which is also the total input power. The parameter μ could then be determined from an estimate of the input power.

There are still two other parameters to determine, filter length and number of channels, both of which will cause the value for the total input power to change. When comparing different combinations of these values, it is necessary to keep all conditions the same except for the one being varied. This cannot be done if changing the filter length necessitates a change in μ. However, it has been shown that keeping the misadjustment, M, constant is sufficient for comparing cases (Widrow 1985). The misadjustment for the LMS algorithm is defined as

$$M = \frac{\text{excess MSE}}{\zeta_{\min}} . \tag{32}$$

This can also be written as

$$M = \mu \operatorname{tr}[\mathbf{R}] , \tag{33}$$

which can be used to rewrite the expression for a stable μ :

$$\mu = M/\operatorname{tr}[\mathbf{R}] . \tag{34}$$

Keeping the misadjustment constant for all tests allows the results for different filter structures to be equally compared. Determining the tr[R], however, is not straightforward.

The signal is nonstationary which means the average power within the signal can vary over time. A conservative estimate for the tr[R] would be to multiply the maximum expected power by the number of filter taps. This large value would yield a relatively small μ. The value for μ cannot be too small or else the weights could not track rapid variations in amplitude and would require a large number of training records to bring the weights up to a usable base value.

Another desired trait is to keep the algorithm simple for the user. It might not be feasible for the user to obtain an accurate value for the maximum possible input power. For this reason, it was decided to keep a running estimate of the signal power which is revised with each new data point to be adaptively filtered. The convergence constant was then updated correspondingly for each point. The power update equation is as follows:

$$pwr = 0.99pwr + 0.01d_k^2, \tag{35}$$

where d_k is the current desired sample and pwr is the estimate for the input power to each filter weight. This method of power estimation might at first seem contradictory to the definition that requires an estimate of the power in the input channel rather than the desired channel. Yet, this format offers simplicity. Instability is avoided because the desired channel from which the update is made will be an input channel in the next test and variations in the actual power estimate are small. Yet the system is still able to track long term trends in the power. In this way the power value really is a running average of the *input* channel power. The convergence constant is then updated as

$$\mu = M \,/[\, nfl{*}pwr{*}nch \,] \;, \tag{36}$$

with nfl equal to the number of filter taps and nch equal to the number of secondary channels. The misadjustment has been denoted by M. Although it is common practice to set this value at 0.1, the performance of the filter is limited by this value. Since the filter will only be presented noise+EP, the weights should be able to approach a smaller misadjustment given enough training records.

Because the noise varies throughout such a wide range, one of the inputs required of the user is an initial guess at the input power. If an estimate were made from say the first point in the desired channel, a large peak or valley at this point could seriously distort the estimate for a relatively long period of time. This entered level is the beginning point from which future estimates will be made.

A spike such as that described above could also cause problems for the convergence constant. This spike will artificially raise/lower the power estimate and with the update equation having a coefficient of 0.99, its effect will take time to die away. Conversely, if the update equation were looser to avoid this problem, as in

$$pwr = 0.8pwr + 0.2d_k^2, \tag{37}$$

the system can become unstable for the same reason of signal variation. For this case, it is better to have a temporary variation in convergence efficiency rather than a potentially unstable system.

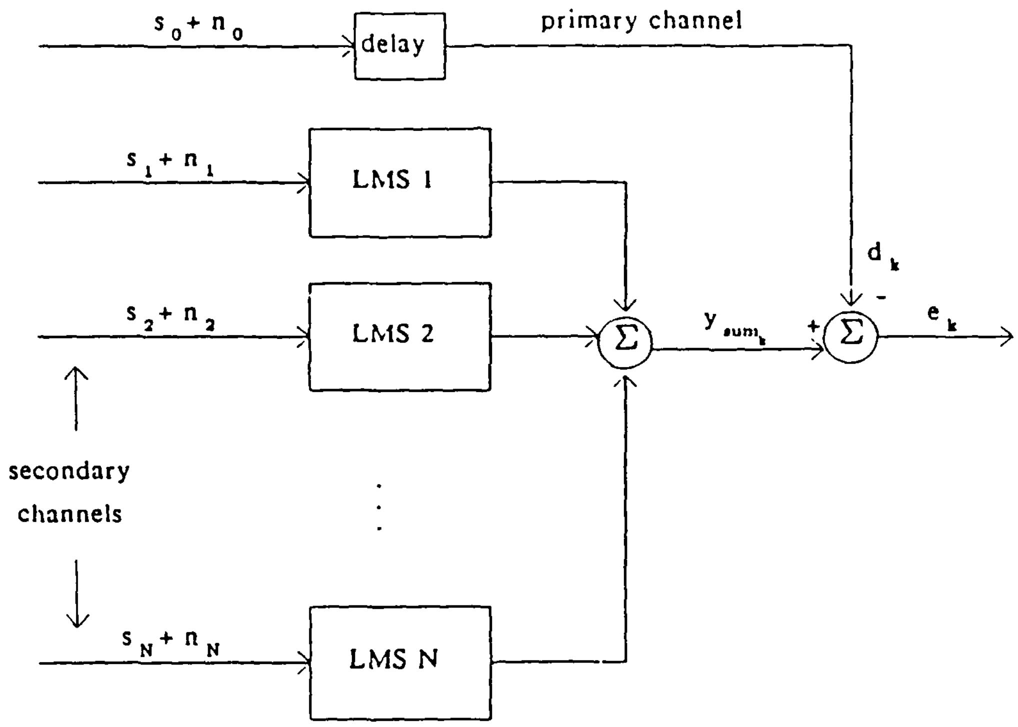


Figure 8. Multichannel Adaptive Signal Enhancer Format.

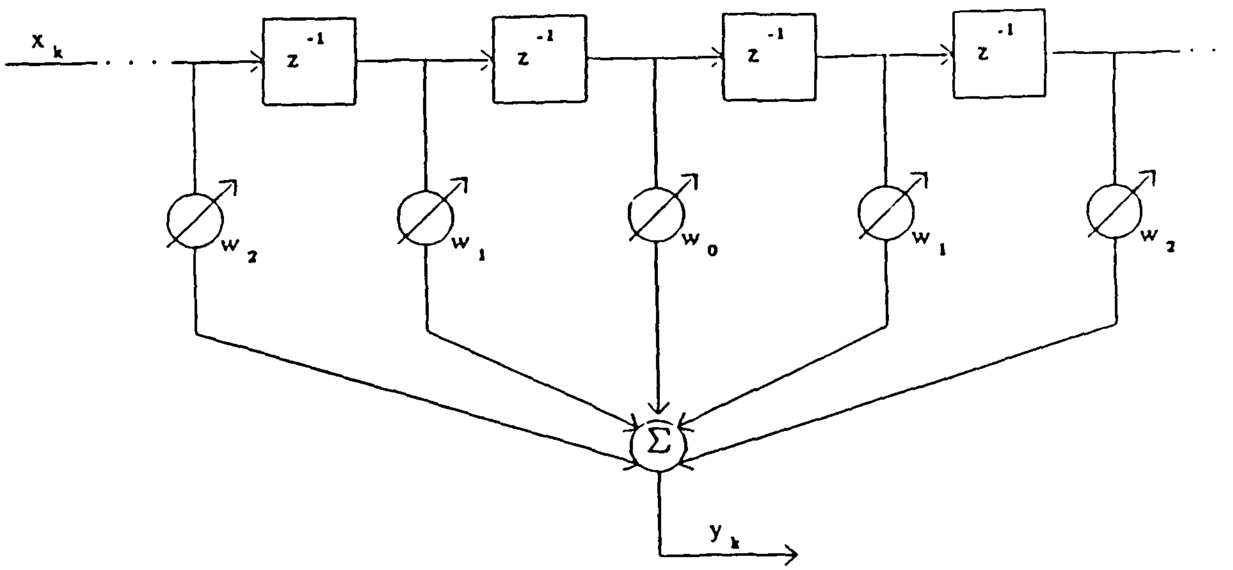


Figure 9. Symmetric Noncausal Weight Structure.

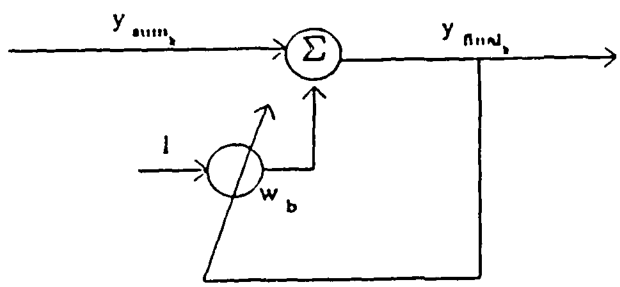
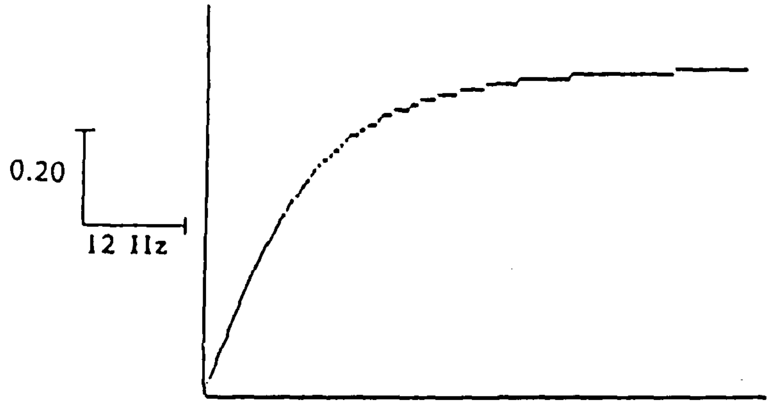
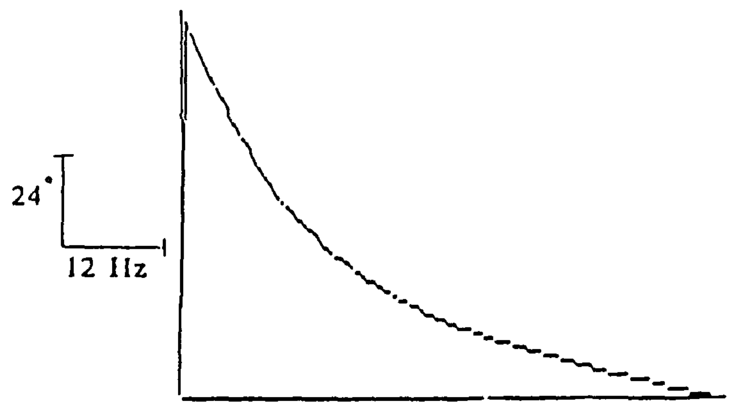


Figure 10. Bias Weight Configuration.



(a) Magnitude Response



(b) Phase Response

Figure 11. High Pass Filter for EEG Cancellation Magnitude and Phase Response.

7. RESULTS

Simulation results using an EP model and human EP results are discussed for various implementations of the modified algorithm.

7.1 SIMULATION RESULTS

In order to determine which implementation -- noncausal, noncausal gated, or symmetric noncausal gated -- and which configuration -- number of channels, number of filter weights, and convergence constant -- provides the best filtering results, performance statistics were computed using the known underlying signal for reference. These indicators of performance were Mean-Square-Error (MSE), noise reduction (NR), and filter bias (FB). Average mean square error was defined as the squared and averaged sum of the difference between the filtered output and the actual signal divided by the squared average of the actual underlying signal.

$$MSE = \frac{\sum_{k=nsa}^{k=nea} (1/N) \sum_{i=1}^{N} (y_k - ref_k)^2}{\sum_{k=nsa}^{k=nea} (1/N) \sum_{i=1}^{N} ref_k^2} \tag{38}$$

In the above equation, 'nsa' denotes the number of the sample to start adaptation and 'nea' denotes the number of the sample to end adaptation. The noiseless signal contained within the desired channel is represented by ref_k and N is the number of data records processed. This provides a measure of how close the output is to the ideal result. Noise reduction of the filter is measured by computing the MSE for the case when the input and the desired signals are both prestimulus EEG and the weights are fixed to those values obtained in the final run of the system utilizing all records when both signal and noise are present. These fixed weights were stored at point nea, the end of the adaptation window. This number is then converted to dB for evaluation purposes. The final statistic is the bias of the filter and is computed by obtaining the MSE when both the input and the desired are noiseless signal, again using the fixed weights from the final run of the system. In this case, the filter should distort the signal as little as possible.

All computations for statistics were done only within the expected EP duration window. For the simulated data, this gate covered points 33-82. As previously mentioned, the parameters of filter length and number of channels are two variables that must be chosen for the filter. To get a quantitative understanding of the effects of these factors, the mentioned statistics were collected for a wide range of these values. The number of channels was varied from one to ten while the number of filter weights was varied throughout the odd lengths from 35 to 65. The maximum number of weights was determined by the limitation imposed from using prestimulus EEG as the noise for the simulated EP. (Section III) All available records were used in the process, meaning that in order to use ten secondary channels, 90 tests were run. Each test produces an output estimate. When 90 tests are executed, the final output is an estimate of the signal in the 100th record.

Using MSE as the major benchmark, the minimum value was found to occur for the symmetric filter at 10 channels, 49 weights, a misadjustment of 0.01, and a power update constant (PUC) of 0.99. Appendix A, tables A.1 through A.9, contain the collected results. As a measure of the effect of the running power estimate, misadjustment was kept at 0.01 and the MSE statistics were collected for a fixed power value. This power value was estimated from an autocorrelation estimate of the unaveraged EP region and had a value of approximately 60. This produced very similar statistics (table A.10), and leads to the belief that the power update procedure is valid and has some merit since the MSE is slightly improved when the update is used. It must be kept in mind, though, that all of the statistics presented are estimates from a relatively small data set and as such will have their own variance.

Considering other research, specifiaclly the best to-date results as provided by Westerkamp and Aunon (Westerkamp 1987) gives perspective on the collected statitics. A MSE of 0.156 for the single channel time-varying filter was achieved with a corresponding bias of 0.068 and noise reduction of -18.312 dB. Obviously none of the tested methods come close to these values. Yet is is still useful to explore their properties for what they yield in terms of the learning properties of the algorithm.

One of the goals of this research effort was to explore the learning behavior of the LMS algorithm and each implementation provides its own perspective on how the filter learns. For this reason, it is useful to look at each in a more in-depth manner.

7.1.1 MASE

In the MASE implementation the weights adapted across the entire record and were wrapped around to succeeding records. The filter was presented a prestimulus section of data in which no correlation exists between the channels and the desired data. As previously shown the weights will tend toward zero. Following this prestimulus section, is the EP for which the weights must quickly try to adapt to a filter capable of extracting the signal. The final data section is again uncorrelated noise, pulling the weights back toward zero.

This problem is obvious in the results. Using a standard case of 10 channels, 49 weights, a misadjustment of 0.01, and a PUC of 0.99, the statistics in table A.1 show a large error in estimating the signal as compared to an excellent job of noise reduction. This seems to indicate that the weights have adapted to values closer to those dictated when just the EEG is in the system rather than when the EP is present. The weights are now apparently better suited to cancelling the noise than to estimating the signal.

Figure 12 shows the filtered output when the 100th record is in the desired channel. As expected, the signal is essentially cancelled. Partially, this is due to the very small convergence constant produced by having a misadjustment of 0.01. This hinders the rapid convergence required when the signal suddenly becomes present. Letting the convergence constant have a larger value by increasing the misadjustment to 0.10 allows the weights to converge faster to the EP and actually manages to reduce the MSE. This is due mainly to the fact that the weights are passing a signal and not merely hovering around zero. An improved estimate of the underlying signal can now be obtained as in Figure 14.

The weights for the case when the misadjustment was 0.01 are essentially zero. However, letting the misadjustment increase to 0.10 yields the weights shown in Figure 13, which, as can be seen from the frequency domain representation are similar in shape to the power spectral density. The weights shown were taken at the end of the EP-window. This means that the weights have had the entire length of the EP to adapt to this configuration and this is about as good as the filter will get.

7.1.2 NONCAUSAL GATED MASE

By gating the period of adaptation, the weights are not allowed to decay as in the previous configuration. The noncausal gated MASE produced the much-reduced MSE shown in table A.4 with a minimum of 0.47488 at 10 channels and 53 weights. This is still a high value but greatly improved from the standard MASE. The noise reduction has decreased because the weights have now converged to a form that is not all essentially zero. The frequency bands that the filter is now able to pass coincide with frequencies also found within the EEG. As the EEG is filtered with the fixed weights, noise passes through to the output. The NR for the 10-49 case is -10.65 dB as compared to -24.48 dB for the ungated MASE. Several waveforms are shown in Figure 15 comparing the filtered output with the actual underlying signal contained in the desired channel for four separate records, #70, 80, 90, and 100. Part of the goal of this research was to improve estimation of the intial peak. As can be seen, the first peak is consistently detected, yet the waveform varies from the shape of the desired. However, information from the desired signal shown in the plots will not affect the weights until the next record. Referring back to the weight update equation (14) this is obvious. Now it might be tempting to compare the current output to the previous desired signal but in this configuration the previous desired signal is now the input to the first of the secondary channels.

Another useful comparison is the average of the filtered responses as compared to the average of the untouched records. For the noncausal case, this is shown in Figure 16. Assuming the EEG noise is uncorrelated with the EP and zero-mean, averaging should produce reliable results. Averaging the filtered signal would then show any distortions introduced in the processing. The filter implemented here introduces no extra distrotion in the initial peak but does attenuate the depths of the other peaks.

7.1.3 SYMMETRIC NONCAUSAL GATED MASE

This final configuration, symmetric noncausal gated MASE, produced a slightly lower MSE than the unsymmetric form. The value of 0.44247 at 10-49 is still very high if the goal of the application is to examine the slight variations in amplitude and latency from record to record. Figure 17 shows a display of waveforms in the same format as for the noncausal gated case. The waveforms appear slightly smoother but offer no significant increase in peak detection. The comparison of the filtered average with the unprocessed average gives the result in Figure 16. Again the initial peak is, on the averavge, undistorted.

Since this case offered the lowest MSE, the weights were examined for each of the ten channels and found to have the forms given in Figure 18. The variations between each channel are more easily visible when viewed in the frequency domain as in Figure 19. The filter attempts to pass the bands in which the frequencies of the EP exist. The basic shape of the filter is similar to that shown in Figure 3(b) for the estimate of the power spectral density of the simulated EP.

7.1.4 AVERAGING

The question still remains as to whether the LMS method has advantages over what is currently in use. A widely used method is averaging. For comparison, the tne channels structure was used to average ten records at a time. No weights were applied excpet for the high pass filter in the regions outside the EP window. The data in all the channels were averaged to procduce the current estimate. This method produced statistics shown in Appendix A, tables A.11 through A.13. Overall, the performance statistics for the symmetric noncausal gated form are better than those for averaging. However, when viewing a selection of the filtered waveforms, the eye seems to see these as better estimates of the signal than previously presented results (Figure 20). It is a difficult case to argue in favor of the additional processing the adaptive filter requires.

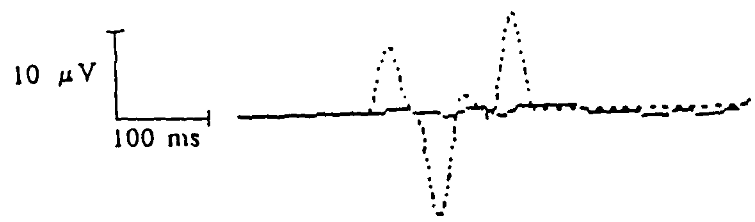


Figure 12. MASE, M = 0.01, 10 channels, 49 weights, desired record #100.

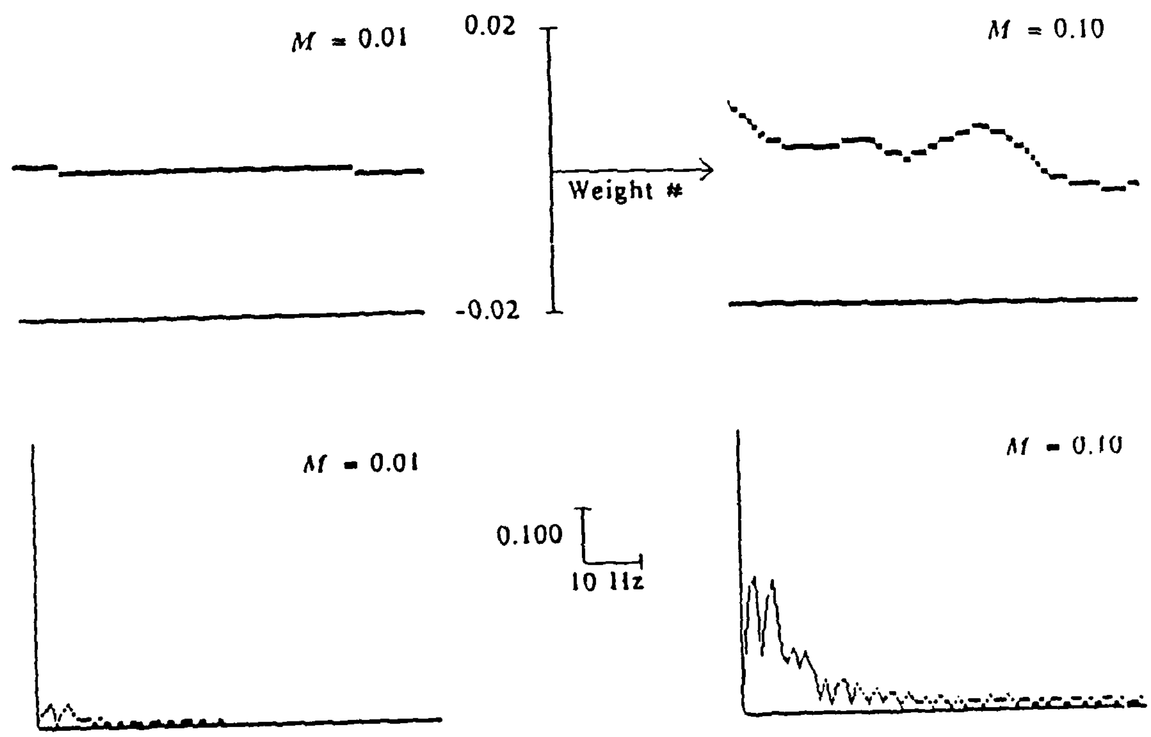


Figure 13. Weight Trace Comparison, (a) M = 0.01, Time Domain, (b) M = 0.01, Frequency Domain, (c) M = 0.10, Time Domain, (d) M = 0.10, Frequency Domain.

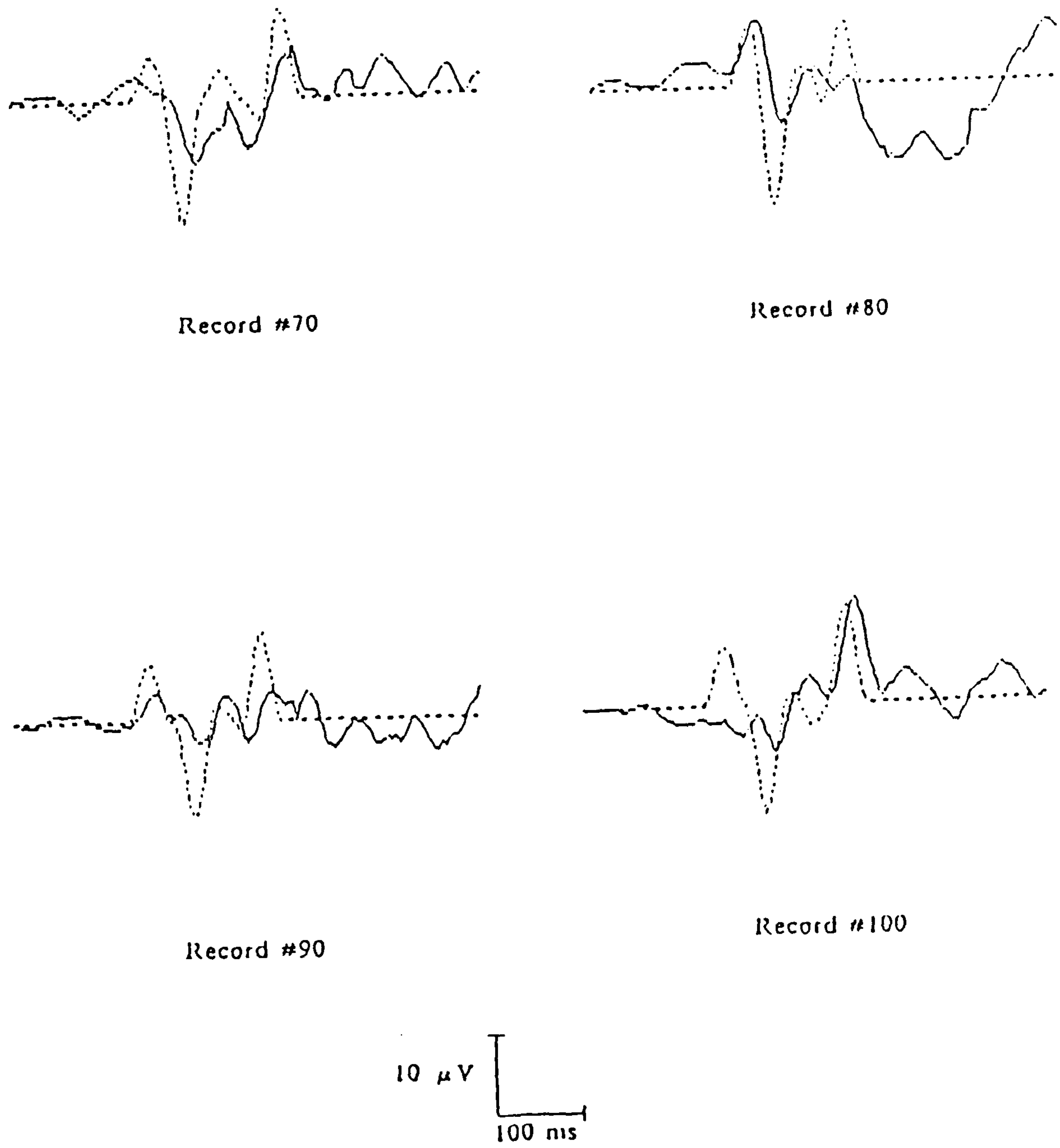


Figure 14. Output Waveforms, MASE, M = 0.10, 10 channels, 49 weights.

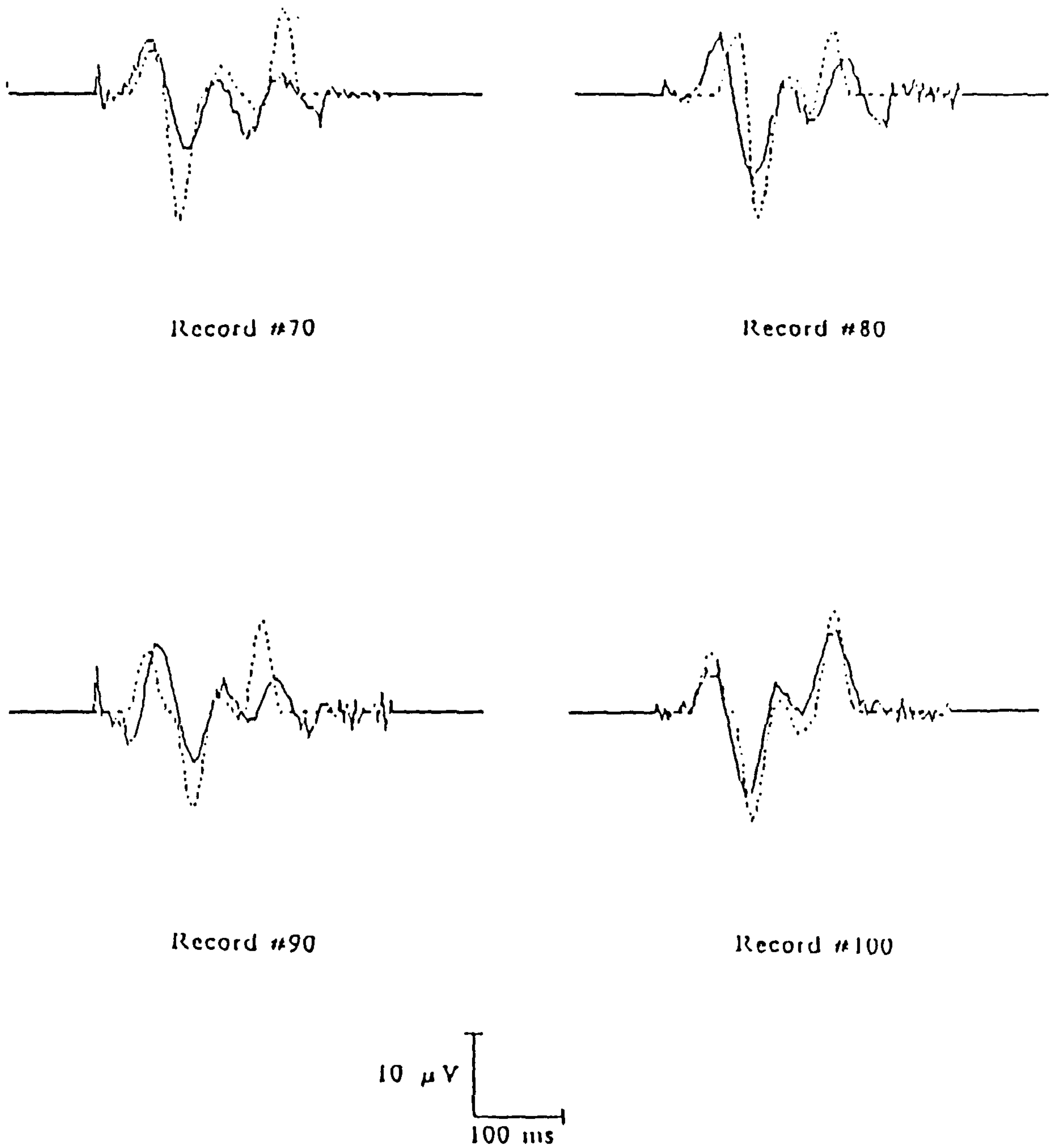
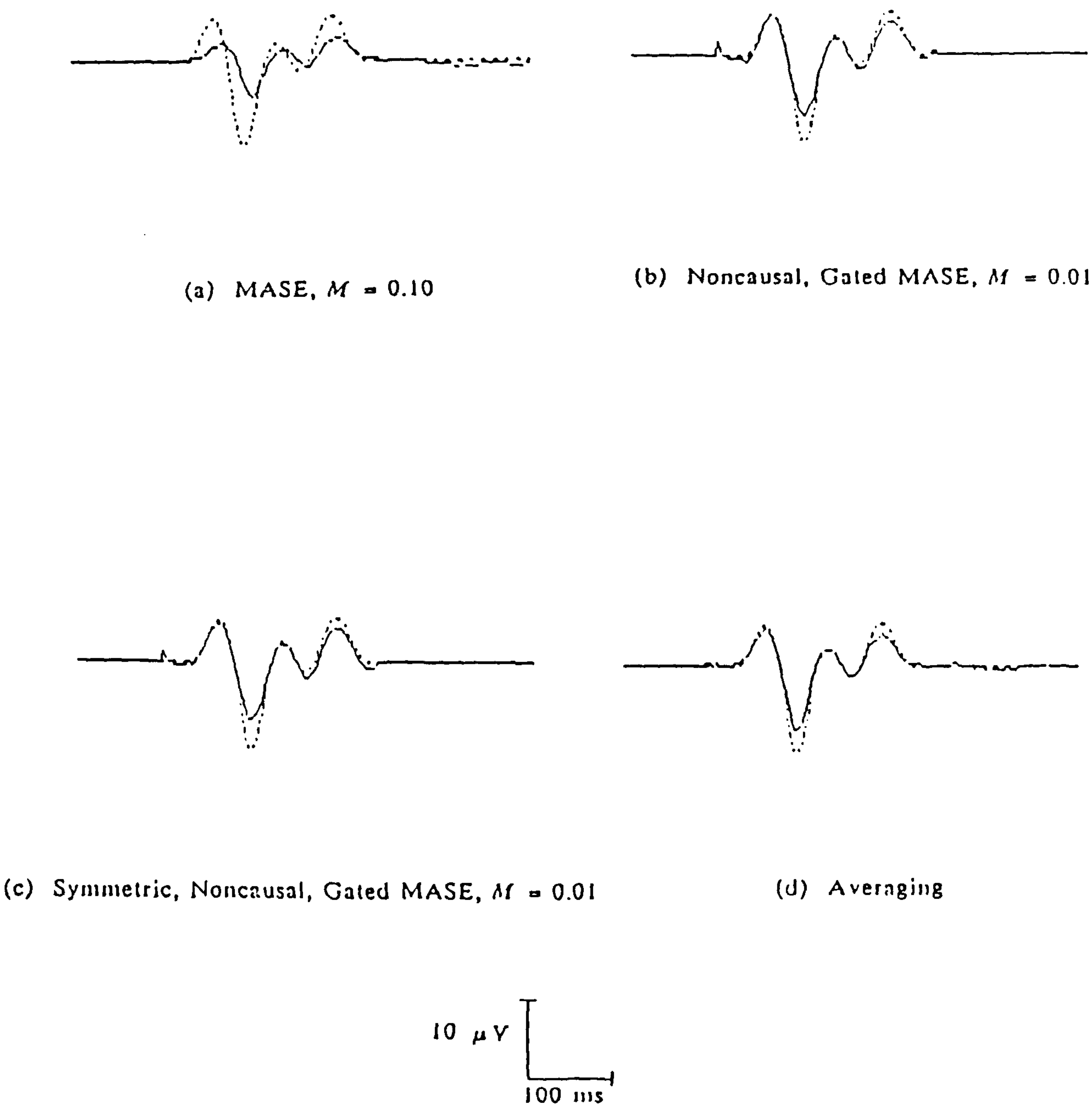


Figure 15. Output Waveforms, Noncausal Gated MASE, M = 0.01, 10 Channels, 49 Weights.



(a) MASE, M = 0.10

(b) Noncausal, Gated MASE, M = 0.01

(c) Symmetric, Noncausal, Gated MASE, M = 0.01

(d) Averaging

Figure 16. Comparison of Averages, Solid Line - Average of Filter Outputs, Dashed Line - Average of Unprocessed Simulated EP.

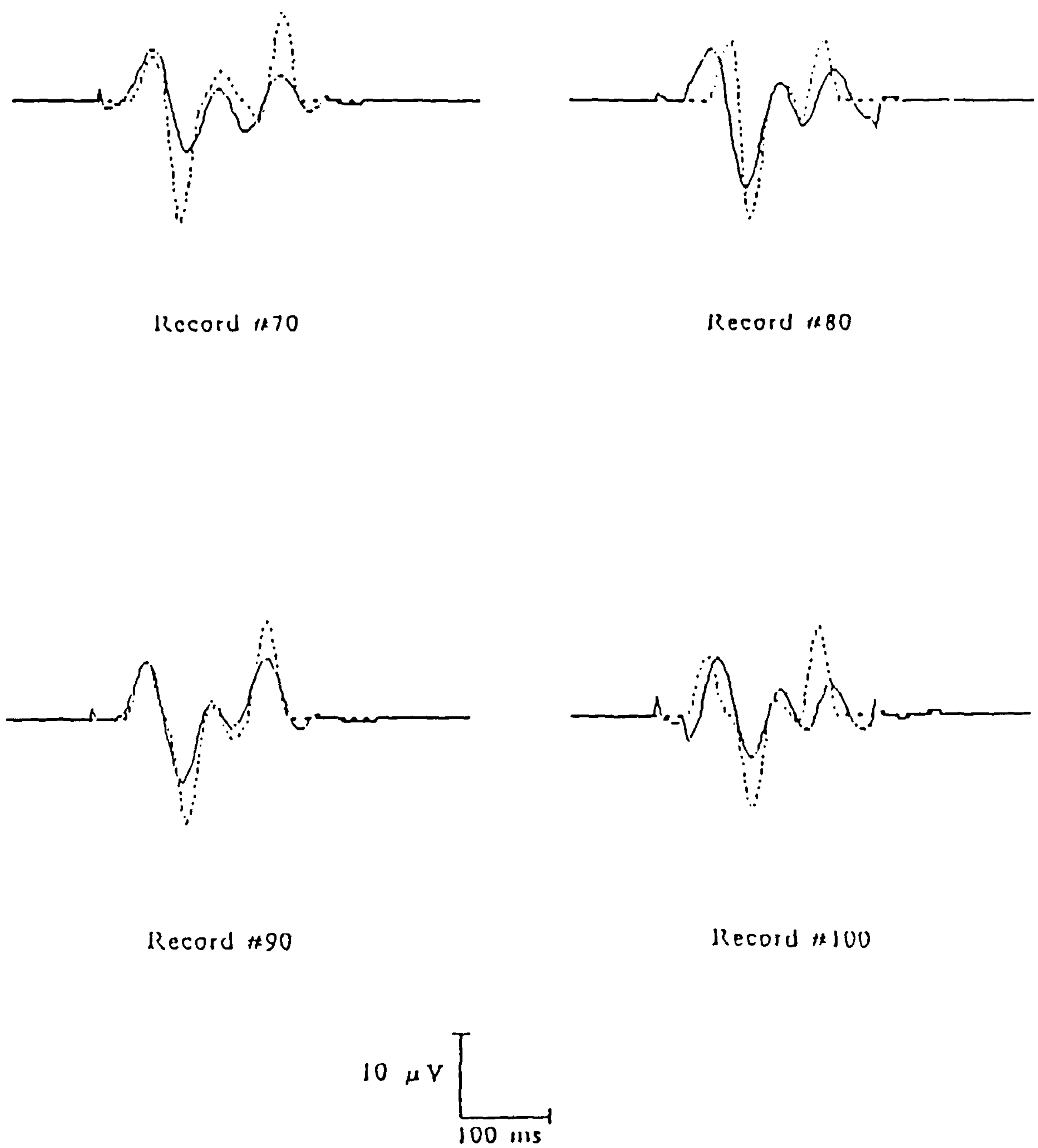


Figure 17. Output Waveforms, Symmetric Noncausal Gated MASE, μ = 0.01, 10 Channels, 49 Weights.

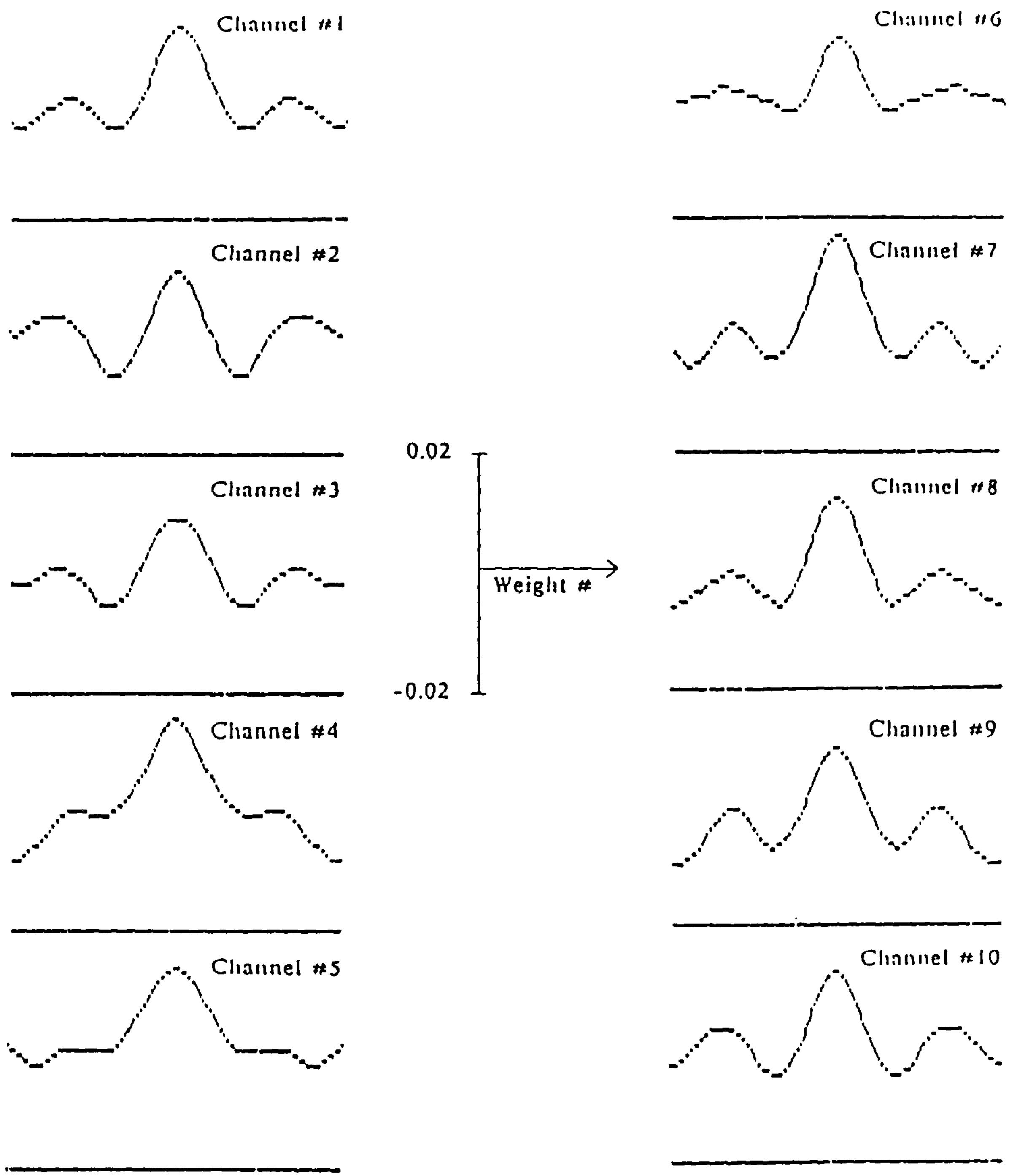


Figure 18. Time Domain Weights for Each of the Ten Channels, Symmetric Noncausal Gated Weights, $M = 0.01$.

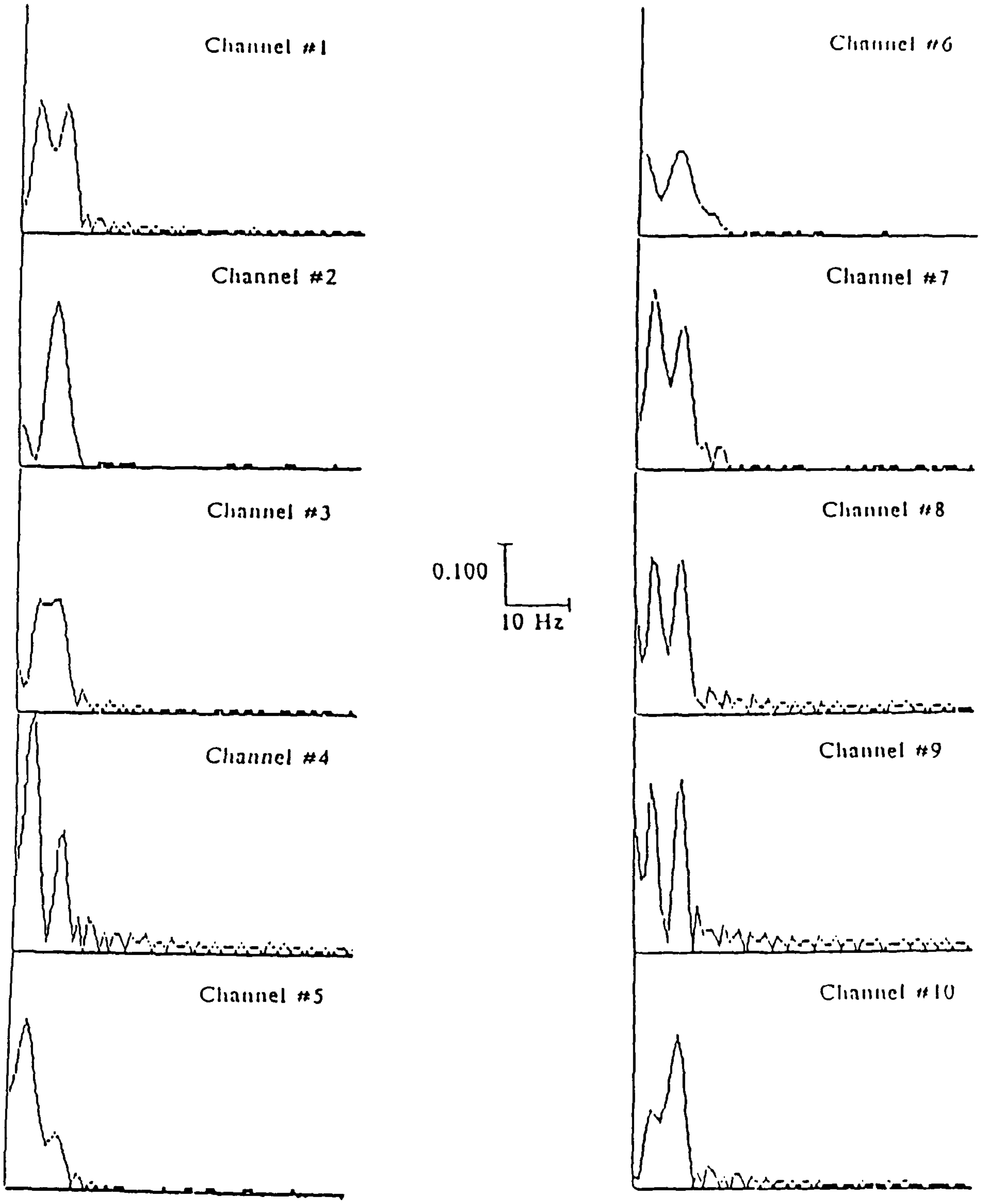


Figure 19. Frequency Domain Weights for Each of Ten Channels, Symmetric Noncausal Gated Weights, Δf = 0.01.

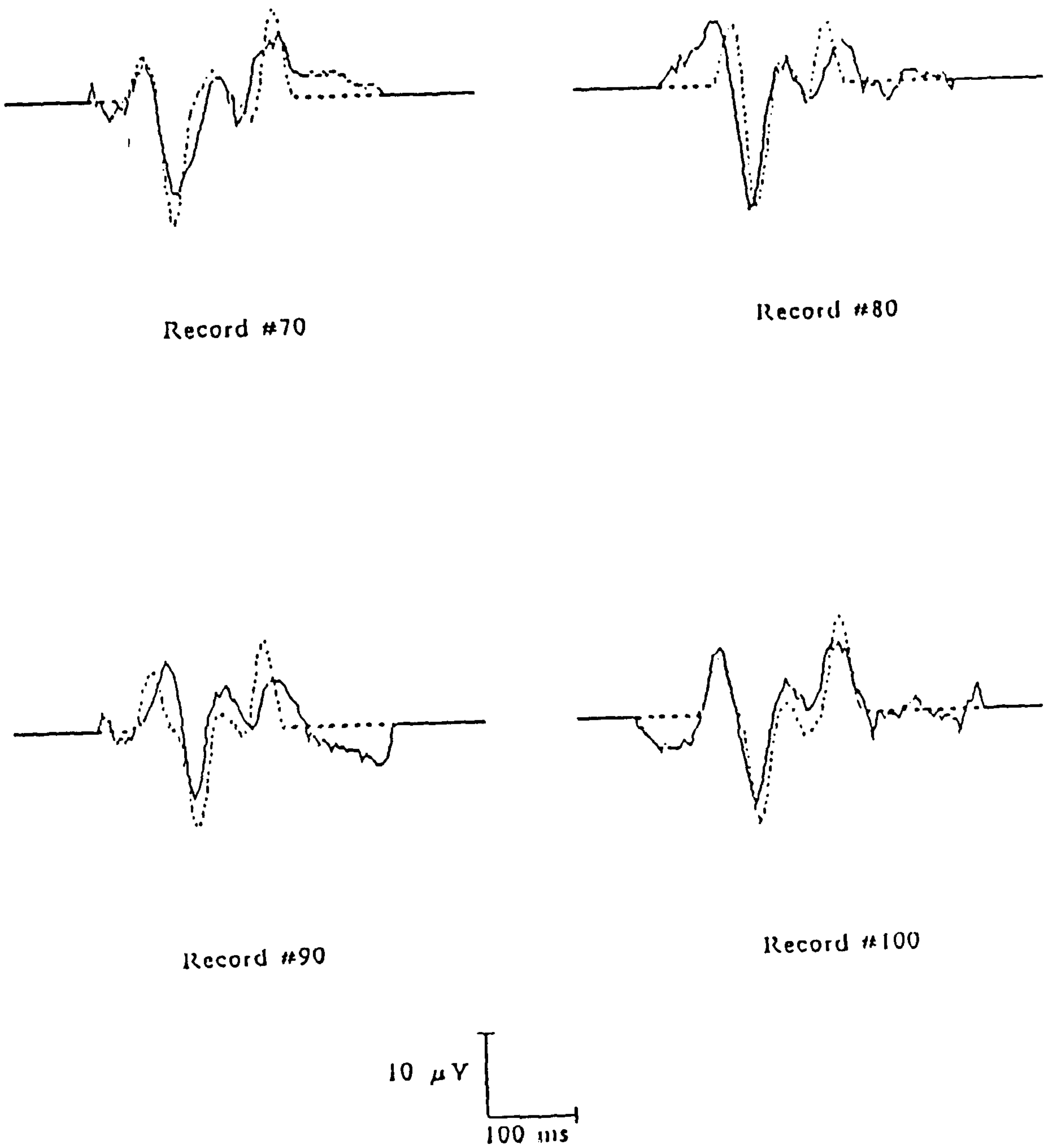


Figure 20. Output Waveforms, Averaging, 10 Channels.

7.2 HUMAN DATA

7.2.1 SYMMETRIC GATED NONCAUSAL MASE

Similar tests were performed using actual human visual evoked potential data. Looking at the optimum case of symmetric noncausal gated filter 10-49, misadjustment 0.01, and PUC 0.99, produced the filtered waveforms given in Figure 21 for records #70, 80, 90, 100. The waveforms don't appear to be greatly distorted in shape, yet comparison of the averages in Figure 22 shows a significant decrease in peak amplitudes. For this reason, the same configuration was used across a range of convergence constants to determine its effect. Figures 23 and 24 show the output and averages as misadjustment varies from 0.01 to 0.10 for 90 tests. As misadjustment increases, the distortion apparent in the average comparison decreases and noise in the estimate increases. The best tradeoff between noise and distortion appears to occur for a misadjustment of 0.05. For comparison, records number 70, 80, and 90 now have the estimates shown in Figure 25. The noise has visibly increased in each case as have the peak amplitudes.

Again it is useful to examine the weights. For the range of misadjustment given, 0.01 to 0.10, the weights of the first secondary channel along with their frequency domain representation are shown in Figure 26. Essentially this illustrates the effect of increasing the convergence constant. As this value increases, the filter tends toward the shape of the EP power spectral density (Figure 27), yet more noise from the EEG also passes through the filter. If the misadjustment of 0.01 case had more records with which to adapt, would that small μ eventually produce the best filter? It is hard to determine without additional data available.

7.2.2 NONCAUSAL GATED MASE

The idea behind using a noncausal filter was to improve alignment of the peaks before the output signals were summed. Looking at a sample of the records, Figure 28, it appears that the filter is attempting to accomplish this objective. Yet with the noncausal filter, unwanted spikes may get into the estimate. If the portion of the filter overlapping into just EEG happens to encounter a region of large sudden amplitude, regardless of the weight values, some of that voltage will appear in the output for the current point. This is a danger not present for averaging since only those points within the window will effect the estimate.

As misadjustment is increased, the noncausal filter appears to lose the signal characteristics to the point where it is simply passing the signal as is and letting the channel summation cancel the noise (Figures 29 and 30). This can be seen in Figure # where the small misadjustment values pull up the initial peak and larger values let the waveform pass with little change in shape. In contrast, the symmetric filter tended to get unstable when the misadjustment reached the higher values. This comparison seems to imply that choice of misadjustment and therefore convergence constant is very dependent on implementation and configuration.

Figure 31 contains the time-domain weights for the noncausal filter as misadjustment increased from 0.01 to 0.10 for channel #1. For comparison the Fast Fourier Transforms of those weights are shown in Figure 32. As the misadjustment increases the passed band of frequencies expands and the power content at those frequencies smooths out to a more uniform value -- losing the shape of the power spectral density of the EP.

7.2.3 MASE

From Figure 33, the MASE format appears to perform essentially a low pass filter operation. This accounts for the apparent minimal distortion of the peaks in Figure # but does not provide a clear estimate of the individual response.

7.2.4 AVERAGING

Averaging, as in the simulated EP case, appears to be more accurate in detecting the heights of the peaks. At the same time, the signals do not appear as clear as for the filtered case. The peaks are there to be seen if the observer is familiar with the data (Figure 34). Any oberservations beyond the subjective are difficult to make.

It is useful to compare the averages of all the filtered outputs with the average of the untouched data. In Figure 22, it can be seen that the noncausal and symmetric filters, with a misadjustment of 0.01 introduce distortion above that attached to simple averaging. In referring back to Figures # and #, when the misadjustment has a value of 0.05, the distortions on the average were minimized. Again this points out the application-specific nature of the filtering methods. If the desire is to reduce distortion on the average, a higher misadjustment is required. Minimizing the mean square error produces a distorted signal in the average. Overall, inconsistencies between the results for the simulated data and the human data may lead to re-evaluation of the methods for simulation.

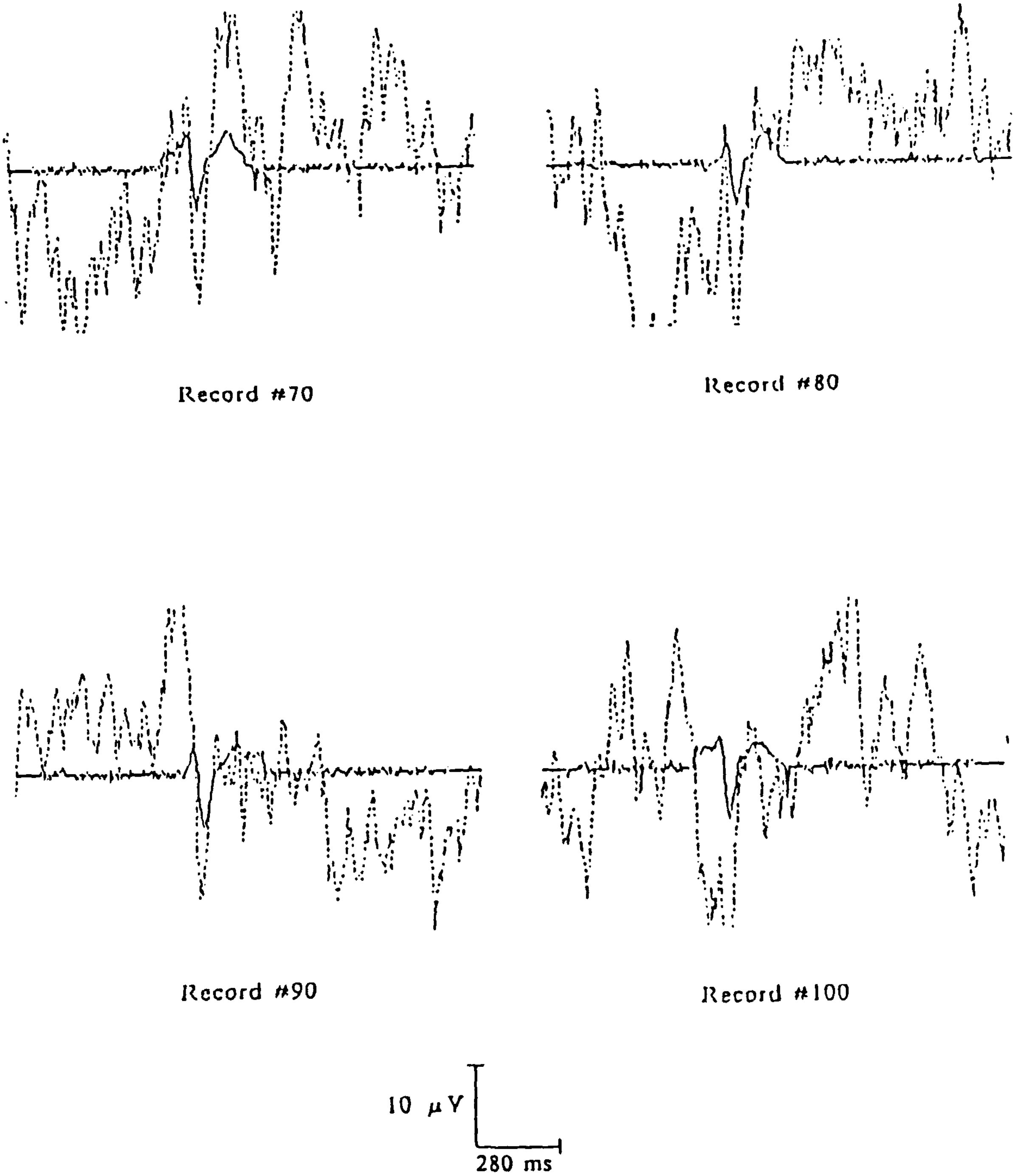
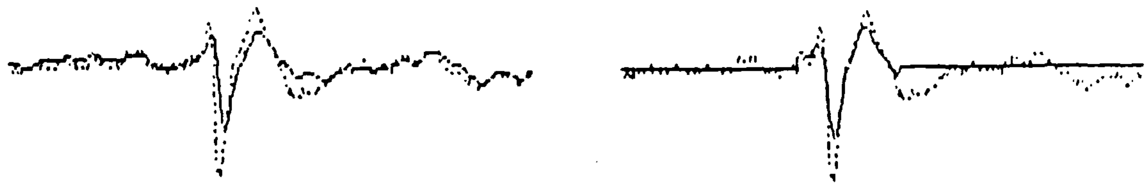
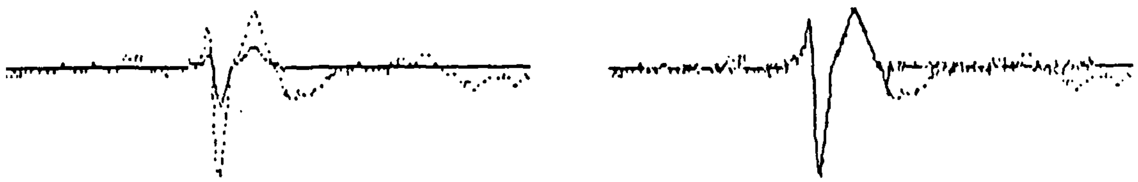


Figure 21. Output Waveforms, Symmetric Noncausal Gated MASE, *M* 0.01, 10 Channels, 49 Weights.



(a) MASE, μ = 0.10

(b) Noncausal, Gated MASE, μ = 0.01



(c) Symmetric, Noncausal, Gated MASE, μ = 0.01

(d) Averaging

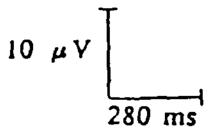


Figure 22. Comparison of Averages, Solid Line - Average of Filter Outputs, Dashed Line - Average of Unprocessed Human EP.

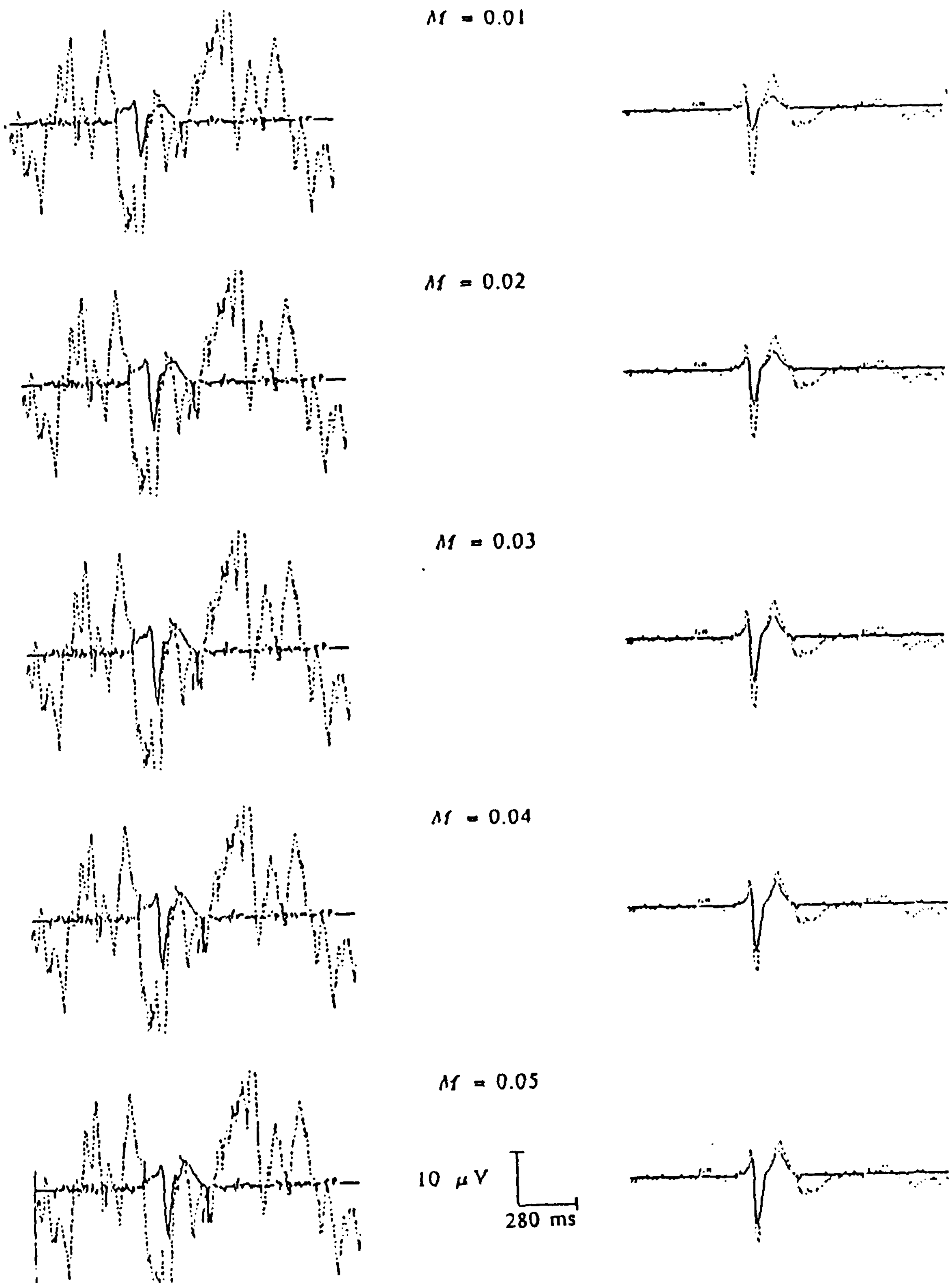


Figure 23. Increasing Misadjustment, Symmetric Noncausal Gated MASE, Desired Record #100 and Comparison of Averages, $M = 0.01$ to $M = 0.05$.

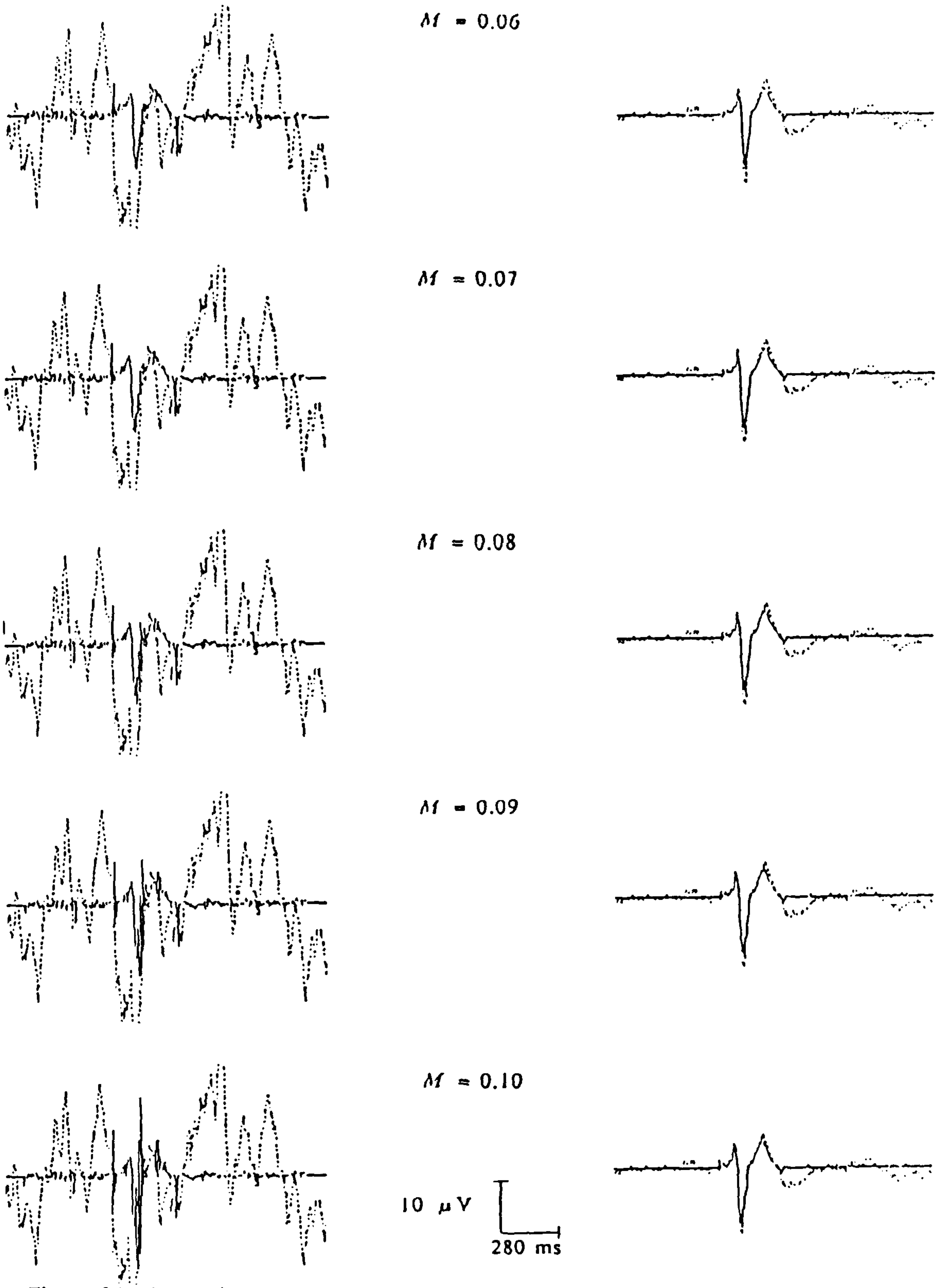


Figure 24. Increasing Misadjustment, Symmetric Noncausal Gated MASE, Desired Record #100 and Comparison of Averages, $M = 0.06$ to $M = 0.10$.

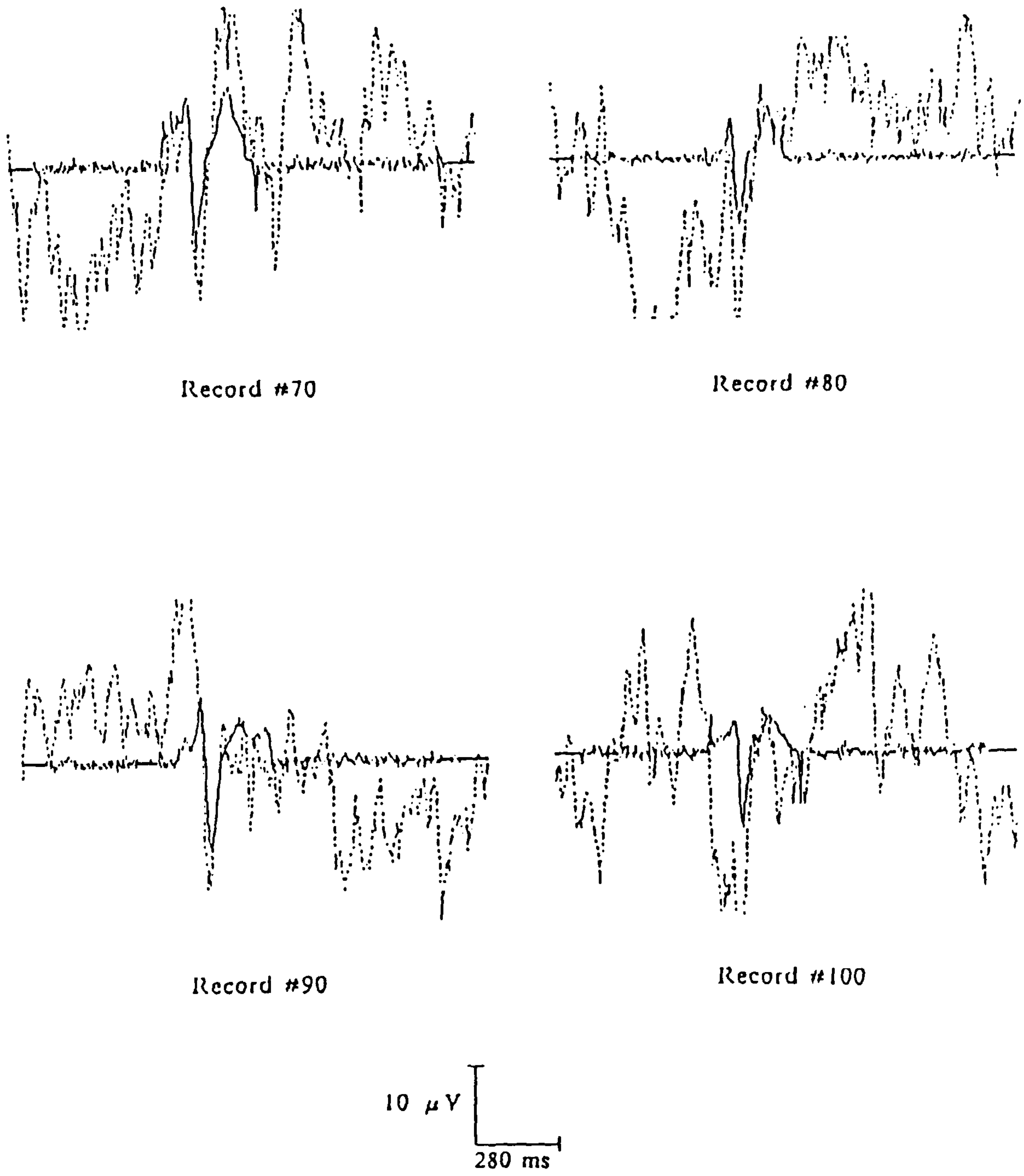


Figure 25. Output Waveforms, Symmetric Noncausal Gated MASE, M = 0.05, 10 Channels, 49 Weights.

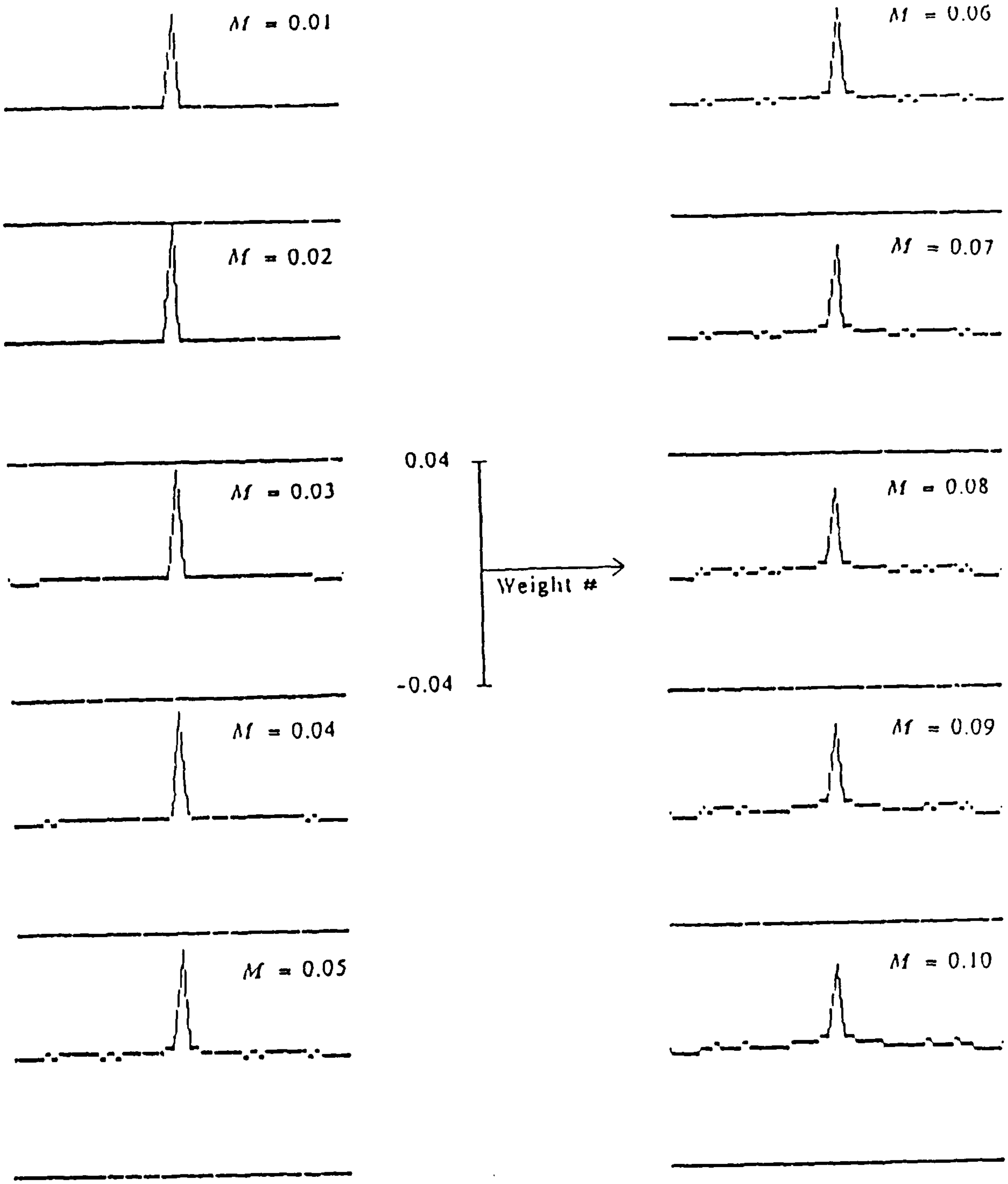


Figure 26. Time Domain Weights for Channel #1, Symmetric Noncausal Gated MASE, Increasing Misadjustment from 0.01 to 0.10, 10 Channels, 49 Weights.

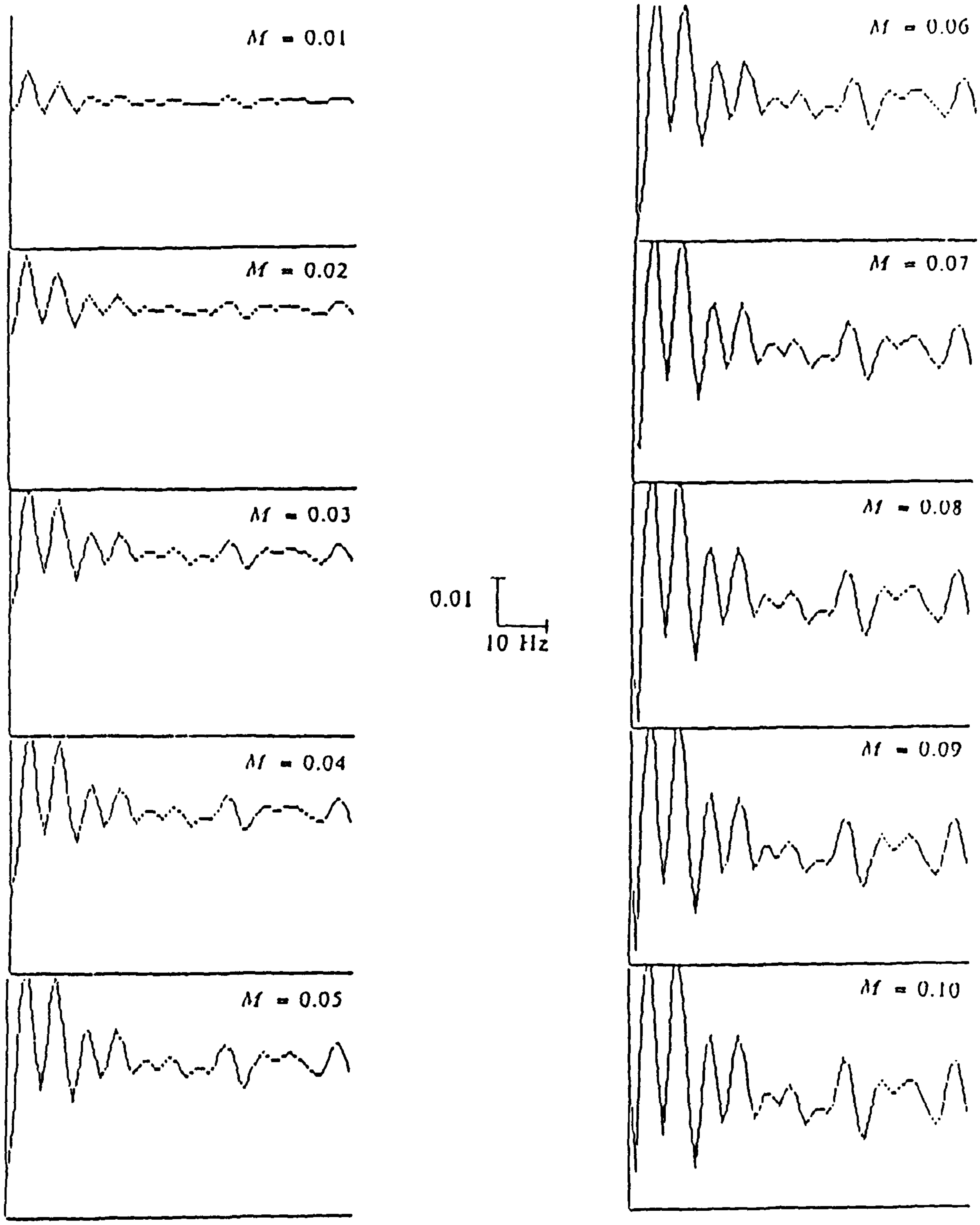


Figure 27. Frequency Domain Weights for Channel #1, Symmetric Noncausal Gated MASE, Increasing Δf from 0.01 to 0.10.

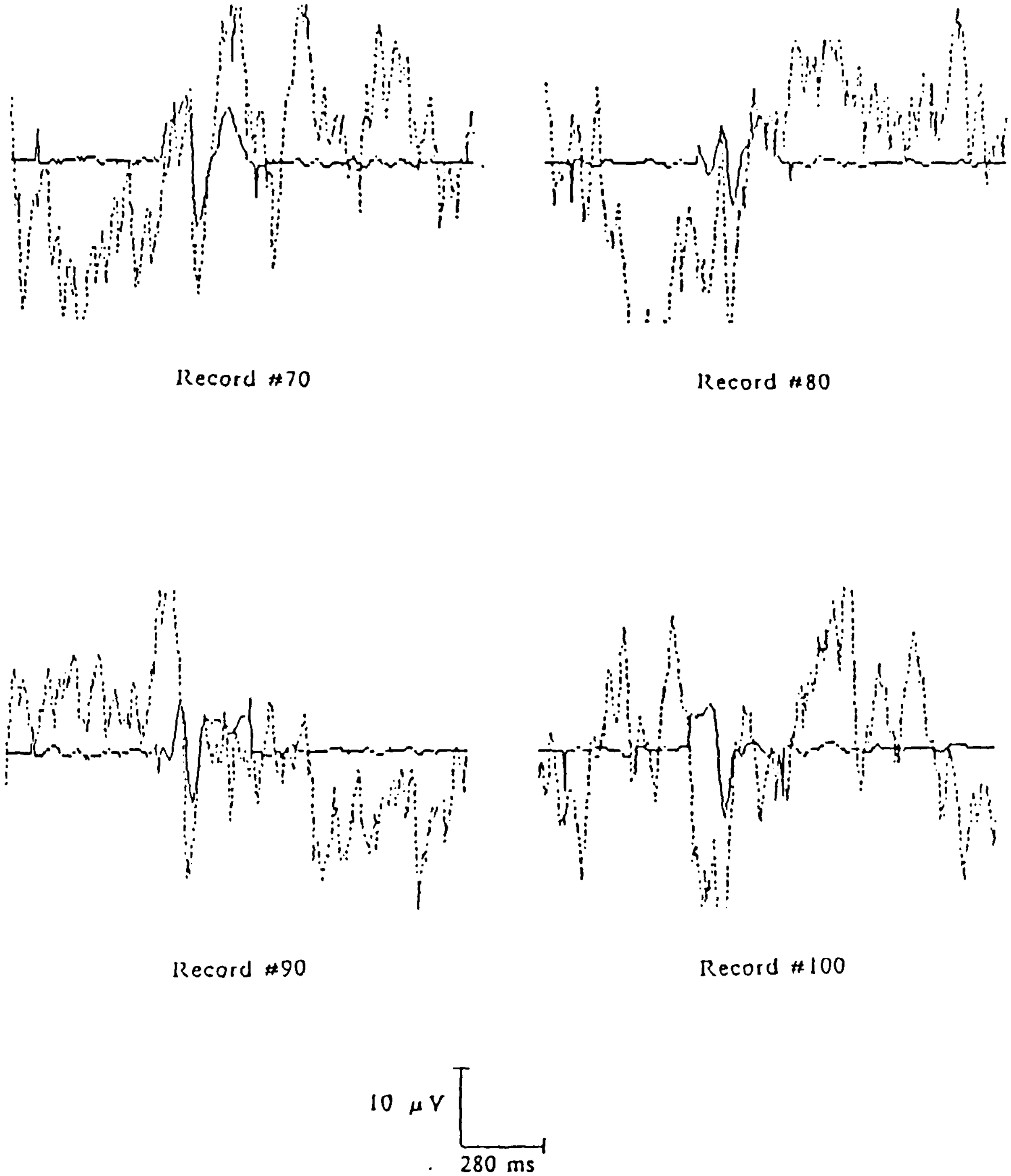


Figure 28. Output Waveforms, Noncausal MASE, μ = 0.01, 10 Channels, 49 Weights.

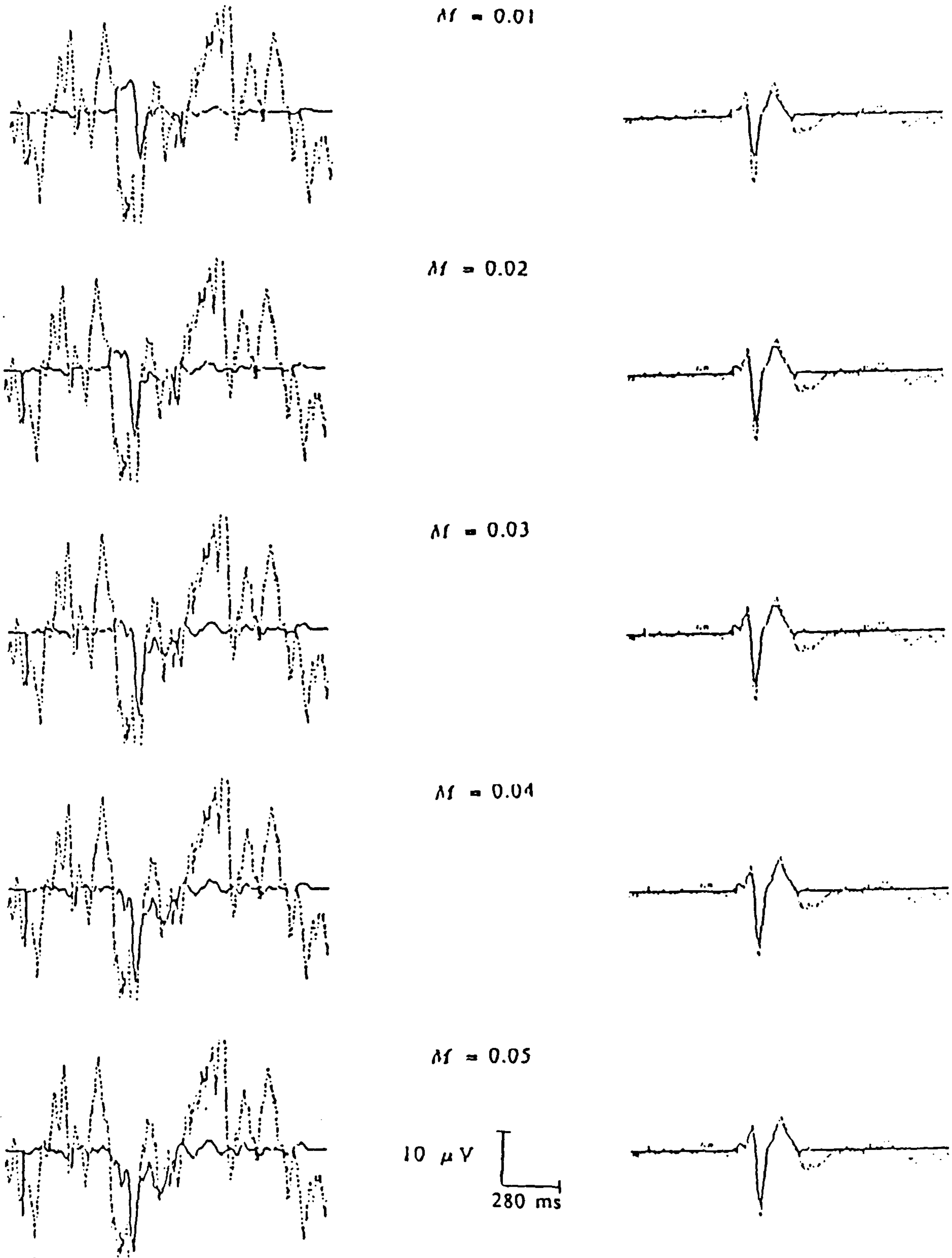


Figure 29. Increasing Misadjustment, Noncausal Gated MASE, Desired Record #100 and Comparison of Averages, $M = 0.01$ to $M = 0.05$.

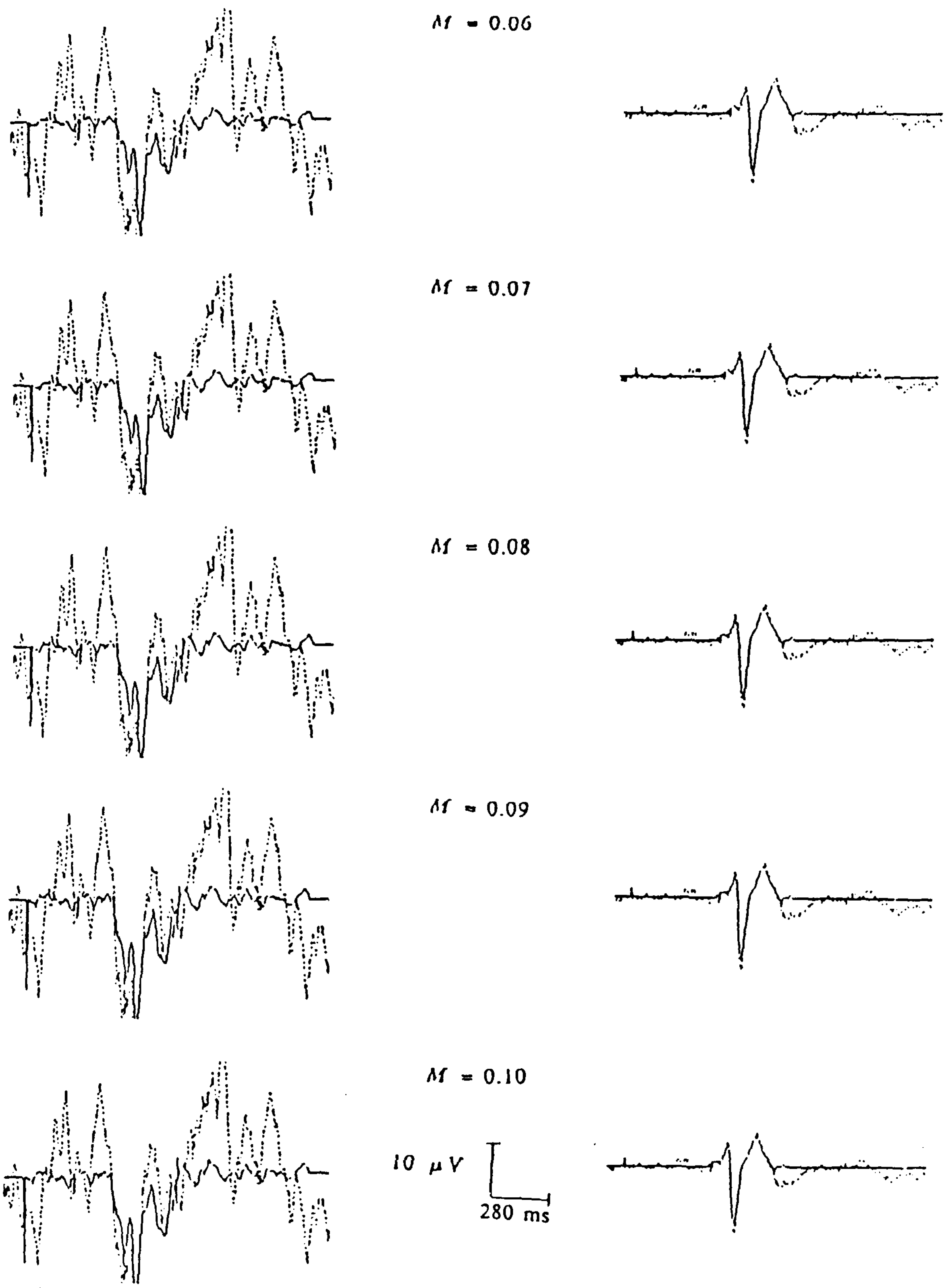


Figure 30. Increasing Misadjustment, Noncausal Gated MASE, Desired Record #100 and Comparison of Averages, $M = 0.06$ to $M = 0.10$.

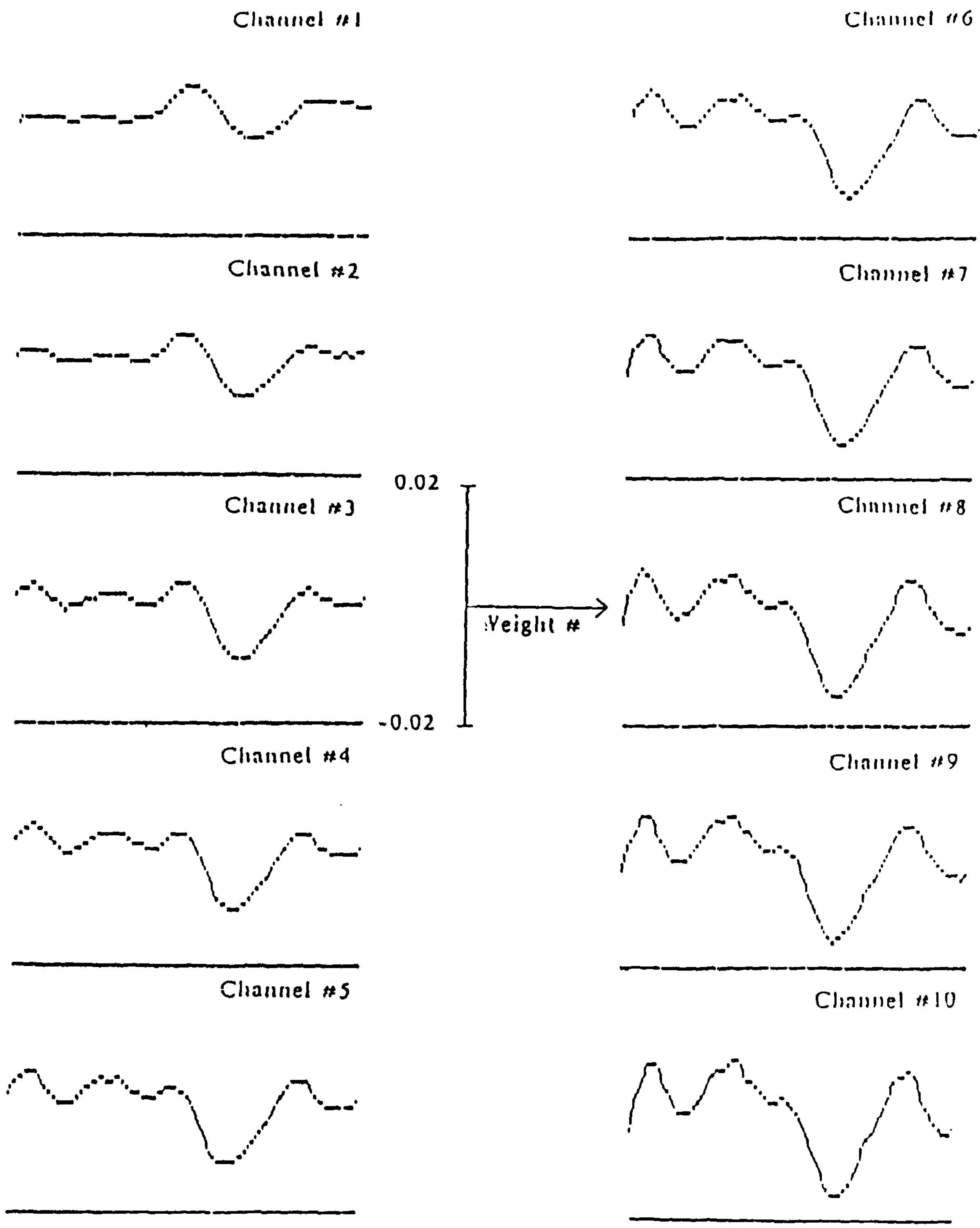


Figure 31. Time Domain Weights for Channel #1, Noncausal Gated MASE, Increasing Misadjustment from 0.01 to 0.10, 10 Channels, 49 Weights.

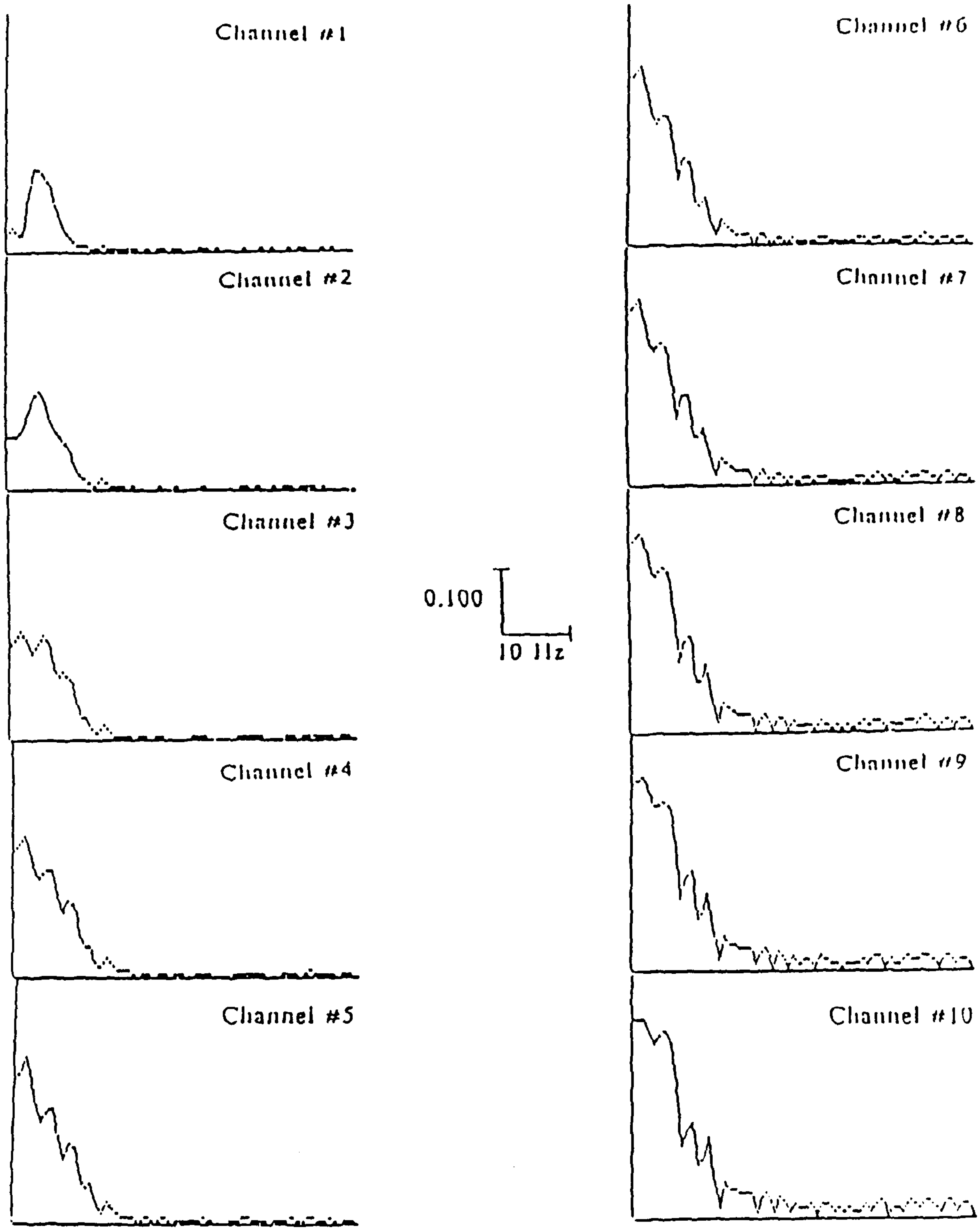


Figure 32. Frequency Domain Weight Magnitude for Channel #1, Noncausal Gated MASE, Increasing M from 0.01 to 0.10.

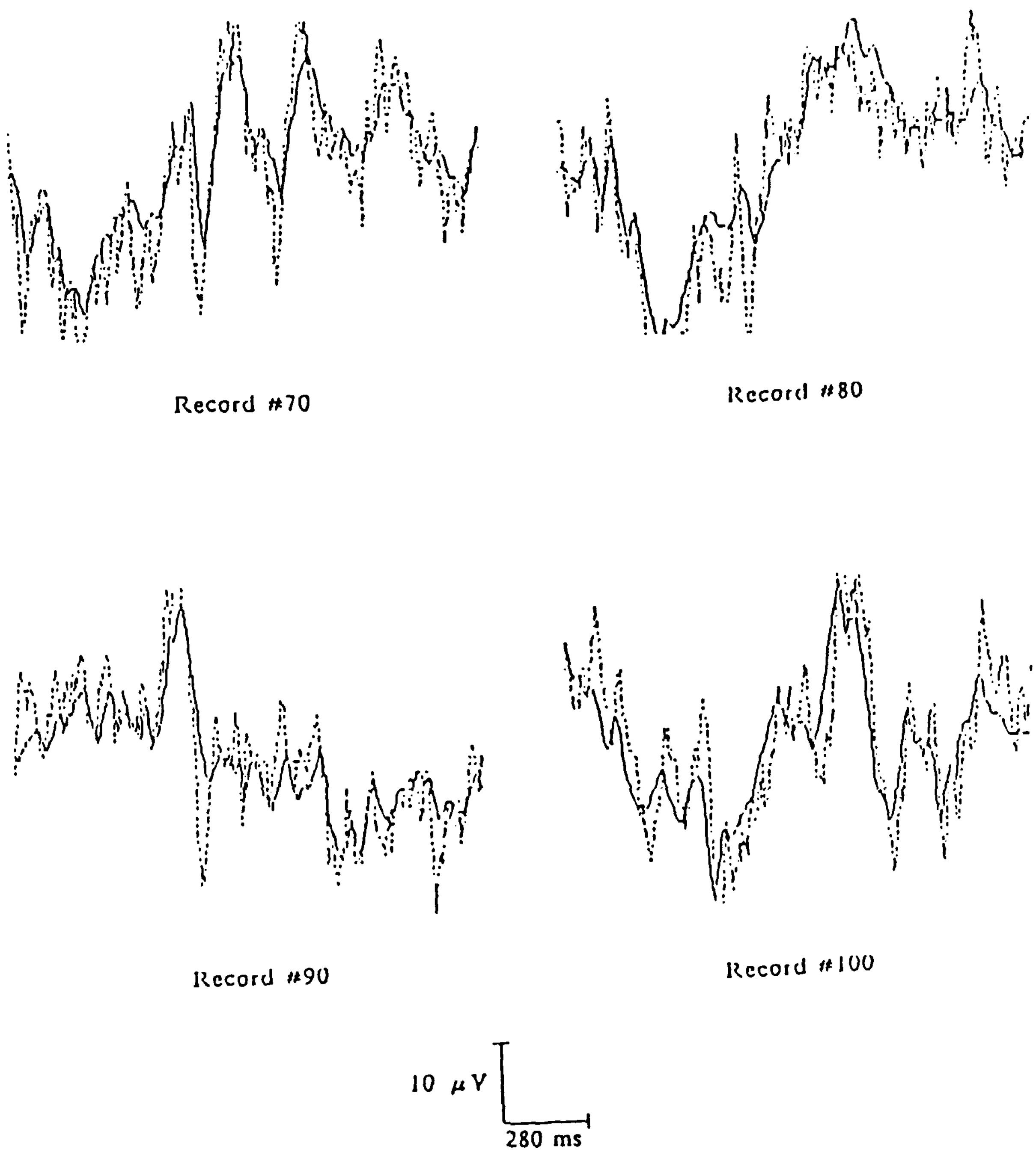


Figure 33. Output Waveforms, MASE, M = 0.10, 10 Channels, 49 Weights.

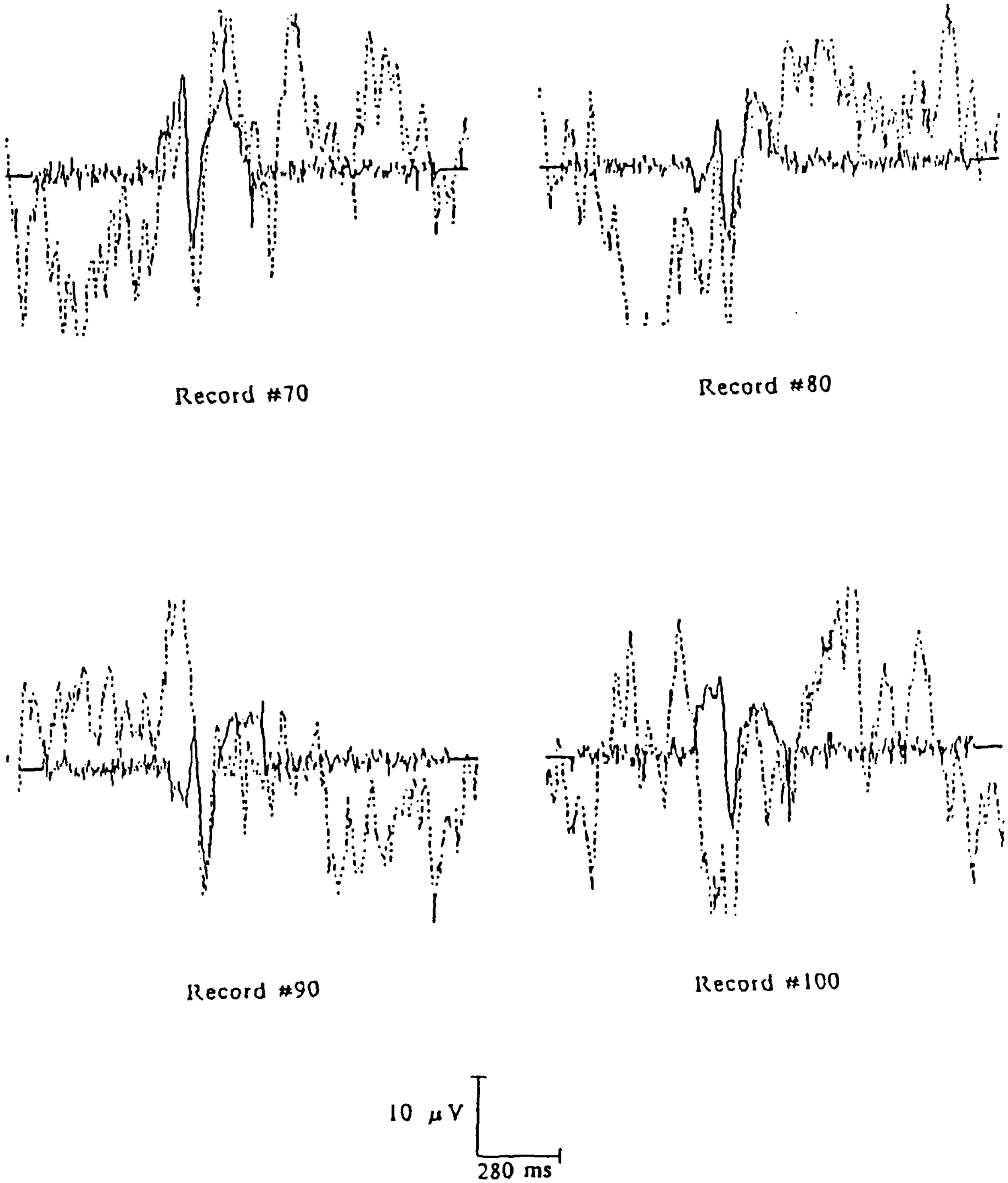


Figure 34. Output Waveforms, Averaging, 10 Channels.

7.3 SPECIAL CASE

In a recent correspondence by Madhavan (Madhavan,1987), it was stated that if the signals are the same in all the channels of the multichannel structure, the result at the output is a distortion of the signal above that produced by simple averaging. The filter configurations presented here have different but correlated signals in all of the channels, yet they perform essentially on the level of averaging. In part his could be due to the fact that the signal's nonstationarities vary so rapidly that the filter cannot detect the various differences within the channels and attempts to average them. However, examinations of the weights has shown that this is not quite the case.

Within the subject of how the filter learns is what happens if the signal suddenly disappears. How soon could that be detected? To examine this case, simulated EP plus EEG was presented to the filter for 50 records and the EEG was presented for the remaining 50 records. For averaging the results are given in Figure 35. Corresponding tests for the symmetric noncausal gated form are shown in Figure 36. It is again hard to determine the better case. The human eye tends to see a signal in the smoother output even when none is present for the sole reason that it *is* smooth. The choppy noise in the averaging may have a greater amplitude but be construed sooner as loss of signal due to the randomness. For interest, the by-record squared error was computed when the signal was present and results are contained in Table 3 for the averaging and Table 4 for the symmetric noncausal gated MASE. When just EEG was present in the desired channel, noise reduction was computed. Statistically the filter appears to give better results.

Neither situation appears to present a quick and accurate result. No sigificant change was expected in the output when the first just-noise record entered the desired channel because all of the secondary channels contained EP and the affected weights would not be applied until the next record. Again this is not promising. If the application did not require an exact estimation of the signal but rather desired a general feel for the integrity of the signal, as in monitoring the deterioration of the EP due to some outside factor, then a larger misadjustment could be tolerated. Increasing the misadjustment will increase the convergence constant and allow for quicker convergence to a changed desired signal. For this reason, the same tests as above were run for the symmetric noncausal gated filter with a misadjustment of 0.10. This produced the series of plots in Figure 37 and the by-point MSE/NR in Table 5. Two things are immediately obvious, the signal estimation is worse and the apparent loss of signal is more easily detected. This is finally a case to argue for use of the LMS.

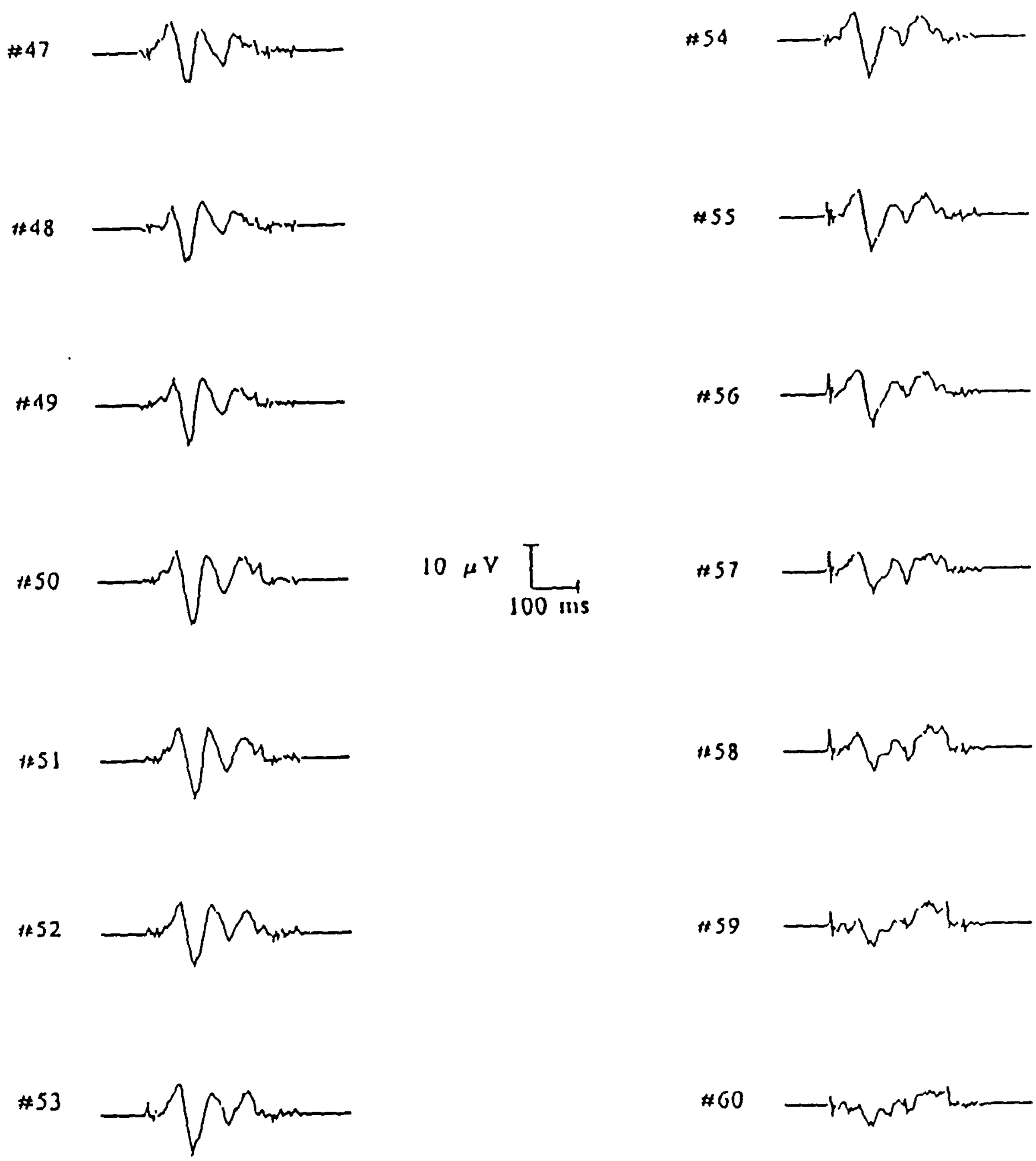


Figure 35. Averaging, Signal in Desired Channels Disappears at Record #51.

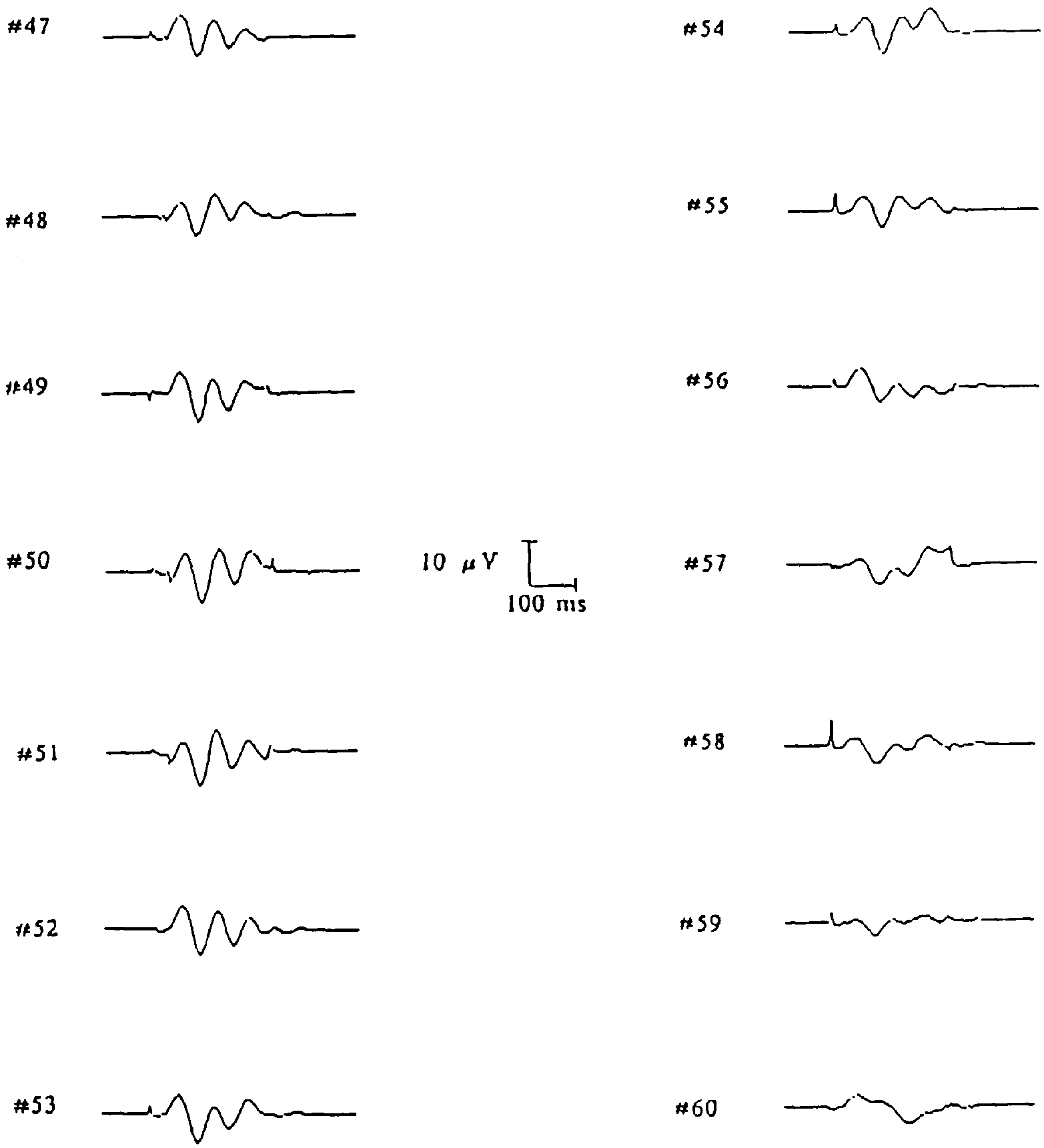


Figure 36. Symmetric Noncausal Gated Mase, Δf = 0.01, Signal in Desired Channels Disappears at Record #51, 10 Channels, 40 Weights.

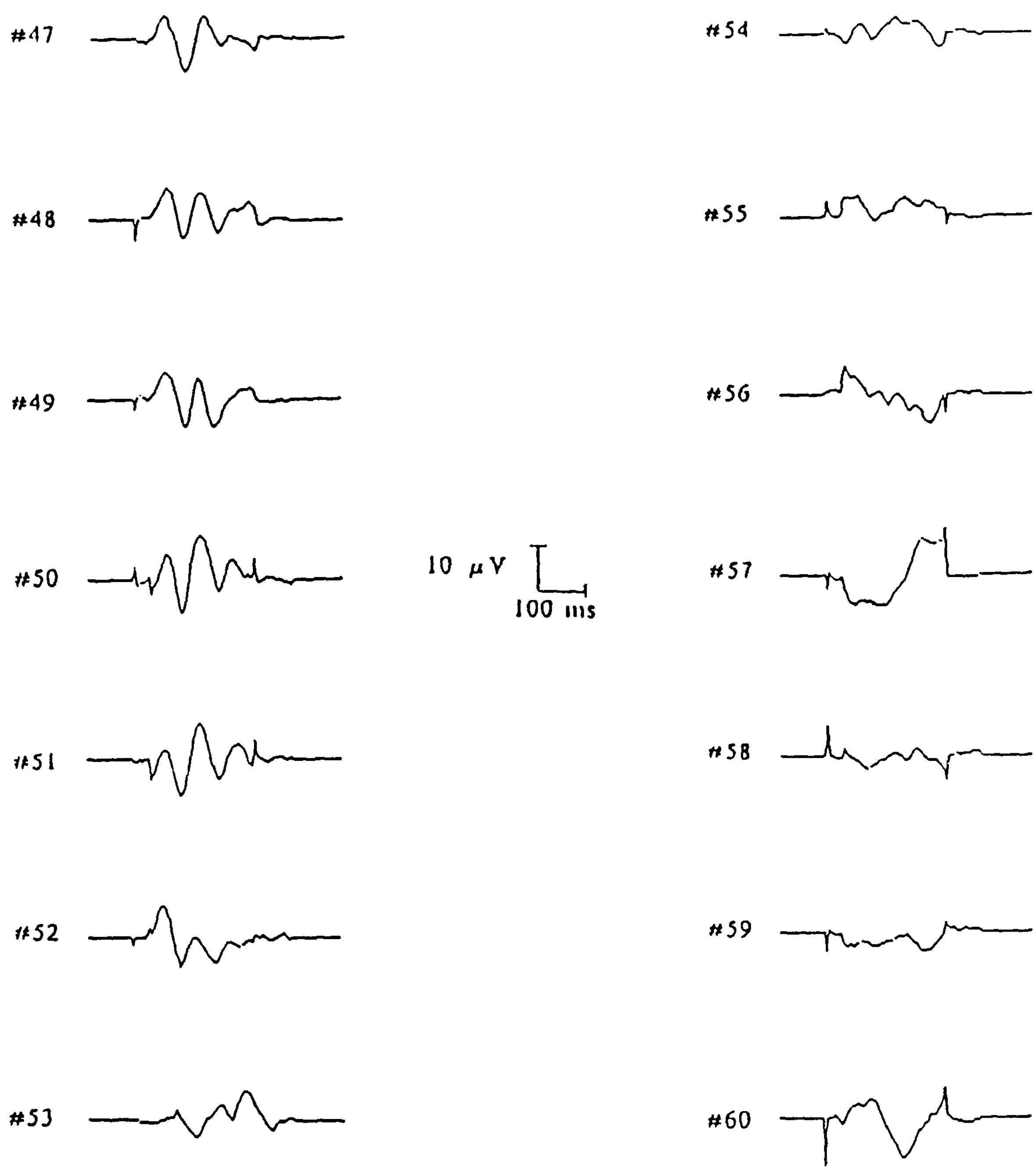


Figure 37. Symmetric Noncausal Gated Mase, M = 0.10, Signal in Desired Channels Disappears at Record #51, 10 Channels, 40 Weights.

AVERAGING					
Run #	MSE/NR	Run #	MSE/NR	Run #	MSE/NR
21	0.52682	44	0.45093	67	-10.06864
22	0.71769	45	0.60700	68	-6.59960
23	0.58854	46	0.63068	69	-3.91606
24	1.04170	47	0.54053	70	-13.32461
25	0.77030	48	0.59380	71	-8.44514
26	0.60413	49	0.59104	72	-12.04266
27	0.73491	50	1.03080	73	-8.36099
28	1.21102	51	1.83705	74	-3.49368
29	1.07186	52	-6.04286	75	-8.99705
30	0.73051	53	-8.05640	76	-7.94157
31	0.75709	54	0.19395	77	-10.85638
32	1.01893	55	-2.52206	78	-12.57829
33	1.71072	56	-8.53705	79	-5.04555
34	1.56205	57	-10.53634	80	-2.71478
35	1.05381	58	-7.62157	81	-7.14302
36	1.16949	59	-6.56419	82	-4.82081
37	0.65790	60	-10.31386	83	-6.78604
38	1.17555	61	-9.79256	84	-7.66760
39	0.90897	62	-11.82333	85	-3.74411
40	0.47410	63	-7.36623	86	-6.14336
41	0.63096	64	-7.18891	87	-6.72137
42	0.68391	65	-8.05051	88	-9.58566
43	0.37815	66	-10.05289	89	-6.16756
				90	-13.03666

TABLE 3. By-Point MSE/NR for Signal Disappearance shown in Figure 35.

SYMMETRIC FORMAT $M = 0.01$					
Run #	MSE/NR	Run #	MSE/NR	Run #	MSE/NR
21	0.49542	44	0.23765	67	-12.46205
22	0.19640	45	0.47488	68	-10.46952
23	0.22110	46	0.39202	69	-4.40554
24	0.44540	47	0.23519	70	-18.08762
25	0.29544	48	0.30844	71	-18.15525
26	0.22390	49	0.32752	72	-20.57792
27	0.28457	50	0.51936	73	-14.70378
28	0.62526	51	-0.67428	74	-9.71301
29	0.48534	52	-9.42541	75	-17.86863
30	0.26035	53	-10.83104	76	-11.68523
31	0.39180	54	-2.35974	77	-16.28502
32	0.47158	55	-6.97167	78	-21.69371
33	0.81439	56	-12.84527	79	-10.11691
34	0.66488	57	-10.97223	80	-6.11197
35	0.50787	58	-11.65540	81	-12.44602
36	0.66424	59	-14.31330	82	-10.11445
37	0.31237	60	-11.19173	83	-12.45706
38	0.50925	61	-14.91491	84	-14.20138
39	0.47246	62	-10.97873	85	-7.63046
40	0.21267	63	-15.72426	86	-16.00113
41	0.34385	64	-9.84952	87	-17.87591
42	0.33044	65	-12.78507	88	-17.42786
43	0.18267	66	-15.31941	89	-10.64988
				90	-21.53238

TABLE 4. By-Point MSE/NR for Signal Disappearance shown in Figure 36.

SYMMETRIC FORMAT $M = 0.10$					
Run #	MSE/NR	Run #	MSE/NR	Run #	MSE/NR
21	0.46814	44	0.68768	67	-6.65334
22	0.51775	45	0.95580	68	-1.30609
23	0.69693	46	0.82880	69	0.58324
24	1.59086	47	0.41513	70	-7.46701
25	0.60456	48	0.73361	71	-4.89845
26	0.53123	49	0.55809	72	-9.86911
27	0.95805	50	1.06656	73	-8.01972
28	1.54112	51	1.67637	74	-4.76887
29	0.76251	52	-7.34047	75	-9.79245
30	0.52245	53	-10.65088	76	-2.42404
31	0.96291	54	-5.10979	77	-8.47380
32	1.23263	55	-4.72802	78	-8.20130
33	1.41015	56	-8.27822	79	-2.47149
34	0.85198	57	-3.38952	80	2.50257
35	0.62148	58	-13.71996	81	-4.54602
36	1.20008	59	-6.96977	82	-3.15440
37	0.22843	60	-4.60890	83	-4.98670
38	0.62473	61	-12.09960	84	-9.78133
39	0.72252	62	-4.25799	85	-4.59840
40	0.25769	63	-6.57282	86	-6.97511
41	0.80354	64	-2.94769	87	-8.39828
42	0.70133	65	-10.04713	88	-11.50206
43	0.92012	66	-6.49417	89	1.13765
				90	-13.53428

TABLE 5. By-Point MSE/NR for Signal Disappearance shown in Figure 37.

8. SUMMARY

Investigations into the learning behavior of the LMS algorithm were presented along with attempts to modify this behavior in order to improve weight retention and signal estimation. The weights were shown to decay toward zero when only uncorrelated noise is present in the filter channels. The rate of decay is dependent on the noise power within the channels. Gating the LMS learning and wrapping around the weight values to successive records was shown to be an effective means of counteracting the loss of learning and improving estimation of the initial EP peak.

Of the various filter configurations tested, the symmetric noncausal gated MASE was shown to offer the best performance, which did not visually exceed the performance of simple averaging. This is due to the limitations in the LMS for highly nonstationary situations. With the time for adaptation available for the EP, the weights do not have sufficient time for close convergence. If a relatively large number of records were to be available for training of the filter, then a time-squenced application would be reccomended as a method that might exceed the performance of averaging.

In the case where the filter operates to detect the presence of an EP, the LMS holds prompise. In the simple tests presented here, the filter gives clear visual indication of the peak loss sooner than averaging. Further research needs to be undertaken to determine if relaible statistical indication could be obtained.

9. PROJECT PERSONNEL

John J. Westerkamp, Ph.D., Assistant Professor of Electrical Engineering, Department of Electrical Engineering, University of Dayton

Janet L. Slifka, Graduate Research Assistant, Department of Electrical Engineering, University of Dayton

10. DEGREES GRANTED

Janet L. Slifka, Master's Thesis: "Learning Behavior of the Multichannel Adaptive Signal Enhancer in Relation to the Evoked Brain Potential," April 1988.

11. PROPOSED PUBLICATIONS

IEEE Transactions on Biomedical Engineering

NAECON

LIST OF REFERENCES

Brazier1964. M. A. B. Brazier, "Evoked Responses Recorded from the Depths of the Human Brain," Annals of the New York Academy of Science, Vol. 112, pp. 35-59, (1964).

Eason1970. R. G. Eason, C. T. White, and N. Bartlett, "Effects of Checkerboard Pattern Stimulation on Evoked Cortical Potentials in Relation to Check Size and Visual Field," Psychon. Sci., pp. 113-115, (1970).

Ferrara1981. Earl R. Ferrara and Bernard Widrow, "The Time-Sequenced Adaptive Filter," ASSP Vol. 29(3) pp. 679-683 (June 1981).

Ferrara1981. Earl R. Ferrara and Bernard Widrow, "Multichannel Adaptive Filtering for Signal Enhancement," ASSP Vol. 29(3) pp. 766-770 (June 1981).

Fisher1983. Ben Fisher and Neil J. Bershad, "The Complex LMS Adaptive Algorithm - Transient Weight Mean and Covariance with Applications to the ALE," IEEE Trans. on ASSP Vol. 31(1) pp. 34-44 (Feb. 1983).

Gardner1987. William A. Gardner, "Nonstationary Learning Characteristics of the LMS Algorithm," IEEE Trans. on Circuits & Systems Vol. 34(10) pp. 1199-1207 (October 1987).

Griffiths1979. Lloyd J. Griffiths and Raymond S. Medaugh, "Convergence Properties of an Adaptive Noise Cancelling Lattice Structure," Proc. IEEE, pp. 1357-1361 (January 1979).

Harris1986. Richard W. Harris, Douglas M. Chabries, and F. Avery Bishop, "A Variable Step (VS) Adaptive Filter Algorithm," IEEE Trans. ASSP Vol. 34(2) pp. 309-316 (April 1986).

Jasper1958. Herbert H. Jasper, "The Ten Twenty Electrode System of the International Federation," Electroen. Clin. Neuroph., Vol. 10, pp. 371-375, (1958).

Krieger1986. D. Krieger and W. Larimore, "Automatic Enhancement of Single Evoked Potentials," EEG and Clinical Neuro. Phys. Vol. 64(6) pp. 569-572 (Dec 1986).

Lee1987. Jae C. Lee and Sanjit K. Mitra, "On Frequency-Domain Least Squares Adaptive Algorithms," ICASSP '87 Vol. 1 pp. 411-414 (April 1987).

Madhavan1984. G. Madhavan, H. deBruin, and A. R. M. Upton, "Evoked Potential Processing and Pattern Recognition," IEEE Frontiers of Engineering and Computing in Health Care, pp. 699-702 (1984).

Madhavan1986. G. Madhavan, H. DeBruin, and A. R. M. Upton, "Improvements to Adaptive Noise Cancellation," IEEE 8th Annual Conf. in Med. & Bio., pp. 482-486 (September 1986).

Madhavan1988. G. P. Madhavan, "Comments on Adaptive Filtering of Evoked Potentials," IEEE Transactions on Biomedical Engineering Vol. 35(4) pp. 273-275 (April 1988).

McGillem1985. C. D. McGillem and J. I. Aunon, "Detection, Estimation, and Multidimensional Processing of Single-Response Evoked Potentials," Final Report for AFOSR Contract #F49620-83-1C-0031, (August 1985).

Rumelhart1986. David E. Rumelhart and James L. McClelland, "," pp. The MIT Press in Parallel Distributed Processing Volume 1: Foundations, , Cambridge (1986).

Savoji1987. M. H. Savoji, "Performance of the Variable-Length Stochastic Gradient Lattice Filter in Adaptive Noise Cancellation of Nonstationary Signals - A Simulation Case Study," Signal Processing Vol. 13(4) pp. 361-375 (Dec. 1987).

Sherwood1986. Douglas T. Sherwood and Neil J. Bershad, "The Complex LMS Adaptive Algorithm, A New Result for Transient Mean Square Error," 19th Asimolar Conf. On Circ., Sys., & Computers, pp. 330-333 (1986).

Thakor1987. Nitish V. Thakor, "Adaptive Filtering of Evoked Potentials," IEEE Transactions on Biomedical Engineering Vol. 34(1) pp. 6-12 (January 1987).

Westerkamp1987. J. J. Westerkamp and J. I. Aunon, "Optimum Multi-Electrode A Posteriori Estimates of Single-Response Evoked Potentials," IEEE Transactions on Biomedical Engineering Vol. 34(1)(January 1987).

Westerkamp1987. J. J. Westerkamp, "Adaptive Filtering of Evoked Brain Potentials," Final Report for AFOSR Contract #F, (July 1987).

Westerkamp1987. J. J. Westerkamp, "Learning Behavior of Adaptive Filters for Evoked Brain Potentials," Proposal to Universal Energy Systems, Inc., (September 1987).

Widrow1976. Bernard Widrow, John M. McCool, Michael G. Larimore, and C. Richard Johnson, Jr. , "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," Proc. IEEE Vol. 64(8) pp. 1151-1162 (August 1976).

Widrow1984. Bernard Widrow and Eugene Walach, "On the Statistical Efficiency of the LMS Algorithm with Nonstationary Inputs," Proc. IEEE Vol. 30(2) pp. 211-221 (March 1984).

Widrow1985. Bernard Widrow and Samuel D. Stearns, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, (1985).

Appendices can be obtained from
Universal Energy Systems, Inc.

The Rhetoric of Hypertext:

An Examination of Document Database Concepts and the Integrated Maintenance Information System (IMIS)

Final Report

Air Force Office of Scientific Research

RESEARCH INITIATION PROGRAM

Contract #F49620-85-C-0013/SB5851-0360

Purchase Order S-760-7MG-021

Submitted by

Patricia Ann Carlson
Principal Investigator
Rose-Hulman Institute of Technology
Terre Haute, Indiana 47803

November 30, 1988

Abstract

This study adopts as a fundamental orientation the view that hypertext may eventually bring about a paradigm shift in text delivery and in human information processing. However, paradigm shifts do not occur overnight; they are evolutionary rather than revolutionary. Because of the considerable commitment of Western knowledge and culture to the written word and linear text, it seems likely that successful hypertext systems will -- at least in the beginning -- electronically emulate many of the strategies a sophisticate reader uses in dealing with hardcopy. The demonstration software for this project modeled the functionality an intelligent reader would bring to the paper text.

Theoreticians of hypertext have promoted the concept as an enhancement for reading through and learning a complex body of knowledge and for capturing ideas spontaneously created by the mind during the reading process. A rich body of literature is growing up around the theory of hypertext, and trade publications are beginning to extol the potential of hypertext. Additionally, hypertext has recently been used in read-only systems intended for more utilitarian purposes, such as the Symbolics Document Examiner (which allows users of a LISP machine to navigate easily through online system manuals) and TIES, an automated "book-on-a-stick" which allows museum goers to browse through explanations while viewing exhibitions.

However, many questions about application are yet to be answered. For example:

- **User Interface**: After centuries of dealing with paper text, humans have developed a number of tactics for quickly extracting meaning from linear text (e.g., glancing, skimming, browsing, writing in the margins, "dog ears") as well as retaining a sense of place (through visuo-spatial cues). Any electronic information delivery systems must be able to duplicate this functionality.

- **Text design**: Document designers have carefully examined paper delivery for better ways of presenting substance, syntax, structure, and style of text written for a specific function. Very little, however, is known about the "rhetoric" of non-linear text. Establishing contextuality, measuring readability, and sustaining meaning must be approach in new ways once text is measured in terms of "chunks" rather than pages and paragraphs.

This report describes a stackware version of a portion of an aircraft maintenance manual and considers issues of storage models, user interface, and information retrieval methods for online text.

1. Introduction

Online documentation has matured from the early days where terse and often cryptic error messages constituted the only form of online assistance. Today, there is a growing interest in electronic documentation as a means of enhancing performance, increasing user satisfaction, and improving cost-effectiveness.

One specific research effort is the Air Force Human Resources Laboratory (AFHRL), Logistics and Human Factors Division's work on a prototype online system for aircraft maintenance. The Integrated Maintenance Information System (IMIS) concept is to provide the technician with all logistical, operational, technical, training, and diagnostic information for aircraft repair. The automation of Technical Orders (maintenance manuals, troubleshooting guides, operational theory, procedures, and the like) for weapons systems is central to IMIS research.

Ideas on how to transform paper documentation libraries to electronic support systems vary. A relatively simple approach is to use the word processor files from the paper documentation and run them through a reformatting program to facilitate screen presentation. More complex approaches make use of sophisticated database management techniques and deconstruct the paper text into a collection of fragments embedded in a knowledge base which can then be accessed by specially designed retrieval software.

"*Hypertext*" -- a *model* based *on the* assumption that human idea processing occurs through association -- has received much attention as a framework for effective and efficient communication of knowledge. A hypertext system uses electronic capabilities to overcome the limitations of the linear nature of printed text. Paper text (or flat text) provides only two-dimensions of information processing: linear and hierarchical.

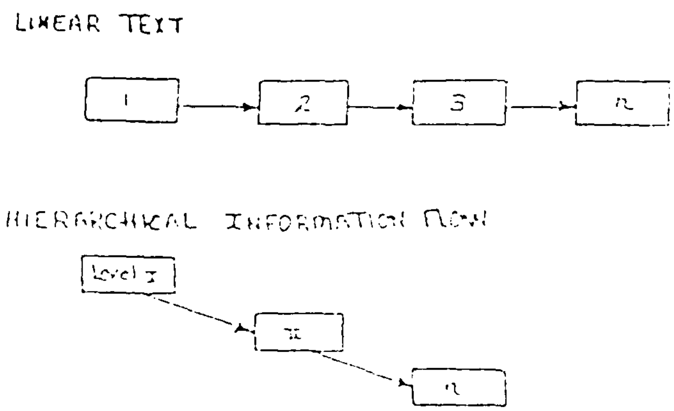


Figure 1: 2-D Information Processing

A hypertext system more closely models the deep structure of human idea processing by creating a network of nodes (modules) and links (webs), allowing for three-dimensional navigation through a body of data. The two major elements of a hypertext system are

- o **Modules:** pools of information collected in one anthology, labelled or typed, and electronically stored as nodes in a database.
- o **Webs:** the pattern or links among the nodes. The links can be predefined by the hypertext system designer or the user(s) can establish the links as part of walking through the information domain.

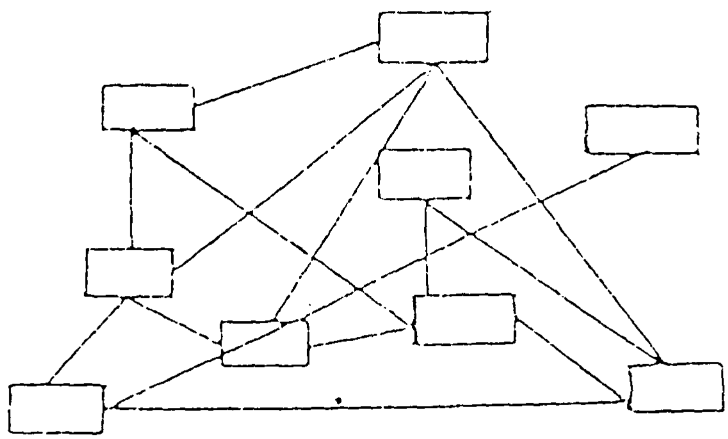


Figure 2: 3-D Information Processing

This much, almost all theoreticians and practitioners of hypertext would agree to. However, when one considers the dimensions of application and implementation, the company starts to part ways. The truth of the matter is that *hypertext* is a collective term covering different instantiations -- analogous to the term *book* -- whose manifestatic.› are numerous and varied.

For a parallel picture of what has happened to the notion of hypertext over its brief forty-three year history, consider the history of the book as technology. In its earliest form, a "book" might have been a collection of clay tablets containing cuneiform, or it might have been a bundle of notched sticks. Later versions of "books" included papyrus rolled into scrolls and tree bark bound with leather thongs. (The English word book is derived from the Old German word for birch.) Parchment, the fervor of monks in medieval monasteries, and advancements in calligraphy and illumination turned books into an art form. Gutenberg's serendipitous connection between the mechanism of a wine press and the function of a signet ring produced moveable type, and the world was on the road to the Information Age. Certainly similar variations driven by human ingenuity, improved technical capabilities, and user demands will determine the future of hypertext.

Even though we are fairly close to the root of hypertext, it is already possible to see variations developing. One method of classifying emergent hypertext applications is to consider the cognitive activities they support.

- o *Reading*: goal-oriented (information seeking) navigation through a large, perhaps unstructured library of information or casual browsing through a pool of text and graphics.

- o *Annotating*: recording ideas dynamically generated while reading text (including critiquing); explicating difficult passages; sorting user-produced mnemonic aids; communicating with the library manager and/or other users.

- o *Collaborating*: electronic conferencing and/or multiple authoring of complex documents.

- o *Learning*: accommodating varying learning styles, varying speeds of ingesting materials, and personalizing structures of bodies of information.

Jeff Conklin, whose comprehensive survey of the field is a starting point for any serious inquiry into hypertext, categorizes current hypertext systems based on their function.

- o *Macro Literary Systems*: a large, online library in which interdocument links are machine supported; all reading, writing, collaboration, and criticism take place via the hypergraph.

- o *Problem Exploration Tool*: supports early unstructured thinking on a problem, in which many disconnected ideas come to mind, such as early authoring and outlining ("idea processors'), problem solving, and design work.

- o *Information Collection and Management Tool*: an extensible environment designed to help knowledge workers gather, sort, and display the structure of information in a particular domain; a tool for discerning patterns of meaning in a body of data.

Yet a third categorization scheme is based on the nature of the software platform and the design philosophy which inform the system. Software packages which look like hypertext to the end user may have radically different storage models, data structures, software design, and information retrieval mechanisms. Although not mutually exclusive, four general categories suggest themselves:

- o *Static Text Based on a Word Processor Platform*: Some applications -- notably the "SuperBook" project at Bell Labs -- take "as input ordinary text in a standard text formatting language, and automatically [convert] it into a multi-windowed browser with rich search, navigation, annotation and display enhancements" (Remde et al., p. 175). Such systems make heavy use of the paper document's table of contents for hierarchical searching and rely on the software's intrinsic markup language for locating and formatting the new representation.

- *Notecard Metaphors or Fragmented Text Paradigms*: First fully implemented in Xerox PARC's NoteCards, this paradigm has become synonymous with hypertext because of the commercial success of Apple Computer's HyperCard^{tm} software. Text, graphics, code, or data are stored on separate "cards" (files), which are collected in aggregates called stacks. When this fragmentation model is implemented with powerful software, aggregates may be hierarchical structures as well as semantic networks whose relationship is determined by the user.

- *Database Management Methodology*: A third category intended to overcome the limitations of linear text displayed in an electronic format stores the text as a document database and implements traditional information retrieval formalisms, such as string searches, keyword recognition, and Boolean refinements. Higher-level models of databases management (including the entity-relationship concept developed by Chen, the idea of aggregation and generalization propounded by Smith and Smith, and Hammer and McLeod's semantic data model), do not depend upon static records stored in a machine-oriented form. These models see a database as a highly flexible, dynamic entity supporting multiple views, allowing for object- rather than record-orientation, and encouraging inferential patterns of processing through semantic networks. The end result of a modern-day document database is a system whose attributes closely resemble the functionality and flexibility of a hypergraph.

- *Artificial Intelligence Research*: Conjoining certain AI models with information retrieval research has produced intriguing hybrid systems. Of particular interest are expert systems which constitute intelligent interfaces to a document database or to a collection of fragmented text. Several efforts attempt to capture the expertise of search intermediaries in online systems. More ambitious efforts integrate browsing and automated retrieval, even to the point that the system has rudimentary models of categories of users and can select search strategies based upon a specific situation (Fox, p. 163).

2. Project Objectives

At the time the proposal was made for this project (September 1987), interest in hypertext was just beginning to gain momentum. Additionally, HyperCard -- the software platform used in the project -- had been released for only two months. The original proposal focused on general issues of task requisites and the nature of linear text as the prime elements in a hypertext definition. As work on the project began and as articles began to appear in the literature, it became clear that the rhetoric of hypertext starts with the design of the software platform and ends with the screen display for the end user.

2.1 Background

As a Summer Faculty Research Fellow for AFOSR working with AFHRL/LRC at Wright-Patterson AFB, Ohio, during the Summer of 1987, I investigated hypertext as a potential design philosophy for the IMIS project. As a background for my project, I read numerous accounts of the hypertext philosophy and of hypertext-based systems. The results of my study were reported in "Hypertext and the Integrated Maintenance Information System (IMIS)," (Universal Energy Systems, Contract No. F49620-85-C0013).

My summer study suggests many ways in which the hypertext concept and the IMIS concept are mutually complementary. However, working with abstracts has its dangers: the compatibility of hypertext and IMIS may result from the fact that, as forward-looking visions, both share the same wish-list for future information processing systems. Only a demonstration project can uncover the elementary problems and potentials of translating the text of aircraft maintenance manuals to hypertext.

Funded for 267 hours of research during the period of December 1, 1987 - March 4, 1988, by the Office of Scientific Research, I designed and built a hypertext version of a segment of a Technical Order (TO). My project had two general objectives:

- o To translate the section of the Technical Order (TO) used in the CMAS II (Computerized Maintenance Assistance System) demonstration into a "hypertext" information delivery system and thereby provide a demonstration of the concept of nonlinear text applied to a familiar domain. [The specific TO used is "Technical Manual: Field Maintenance for Radio Receiver Transmitters RT-727/APX-64(V); RT-728/APX-64(V); RT-731/APX-64(V); RT-727A/APX-64(V); RT-728A/APX-64(V); or RT-731A/APX-64(V) part of Transponder Set AN/APX-64(V).]

- o To examine how much and what type(s) of text redesign are necessary to transform a specific portion of a TO into a hypertext environment.

2.2 Questions of Interest

My original proposal suggested four potential areas of investigation:

- o *Minimal requisites*: The basic elements of hypertext are nodes and links (chunks of text threaded together, either on an **ad hoc** basis by the user(s) or in predefined patterns by information designers). However, is the presence of a database and a retrieval mechanism enough to constitute a hypertext system? How much functionality must the system incorporate in order to reach critical mass? Additionally, if the chunks have been rigorously threaded in predefined webs, has the hypergraph become just another form of sequential processing?

- *Effects of preprocessed text*: Proponents of hypertext point out that the hypergraph (or web) captures the deep structure of a body of information. This is especially useful in situations where the information is complex or where the relationship of its parts either is not clear or cannot be agreed upon. However, is hypertext a useful approach when the web (or knowledge structure) of the text has already been defined?

  As an illustration, a Fully Proceduralized Job Aid (FPJA) has been carefully designed to represent its knowledge structure (by mapping chunks of text to a task orientation). While this threading of text may not be adequate for all the user's information needs (for example, those who try to use the FPJA for reference purposes will find its tutorial approach inhibiting), its web is appropriate for the majority of users. Few technicians, for instance, would suggest that the procedure steps be performed backwards -- although the degree of detail in each step and the deletion of some steps (or conflating them with others) might be open to individual preference.

- *Variations in task environment*: Hypertext helps in dealing with open-ended problems where creativity, spontaneity, and large bodies of information are involved, but how useful is hypertext to a task domain where documentation is intended for cueing/prompting purposes (such as it is in aircraft maintenance manuals)?

- *Optimal time for integration*: As the above questions suggest, not all treatments of documentation as a database need result in a hypertext. The advantages of hypertext are flexibility (especially the ability to create individual views of a body of information) and functionality (for example, rapid traversal through pools of data, marking text for different levels of users, and increasing potential for knowledge synthesis and improved representation). However, not all applications require all these features. At what point in IMIS development does the additional functionality become a useful dimension? At what point does the increased power offset the cost of development and the overhead of implementation?

These questions all have ramifications for human factors and user acceptance. As such, they are treated in segment 3.3 -- User Interface.

However, in constructing the demonstration software package, I found that the project lent itself to an examination of a related -- but different -- set of questions arising from the nature of the HyperCard platform.

- To what degree can the mapping of existing text onto canonical notecard format be automated. In other words, can pre-existing texts be digitized and deconstructed into meaningful units based on machine-supported algorithms?

- o Once the choice of fragmented text has been made, to what extent can the system be enhanced to incorporate abstraction methodologies used in modern database management. For example, can aggregations be made?

- o To what extent can information be filtered -- in the sense that representations can be tailored for the individual user? Can the system support multiple views and information retrieval through a query language as in database management systems?

- o Does the design complement current procedures in AI by mirroring patterns of knowledge representation?

Dealing with these inquiries was more appropriate to the scope and length of the 10-week research effort. Additionally, increased knowledge about basic design issues provides a foundation for more informed commentary on the four, more global, questions of the original proposal. These questions -- which in essence deal with the data storage model used in a hypertext system and its ability to provide flexibility and extensibility -- are considered in section 4. -- System Enhancements.

3. Project Description

Computer systems can be described from three different vantage points:

- o Software: the view from inside the skin, including data structures, storage models, language capabilities, program design, and processing mechanisms.

- o System Design: the complex process of integrating features and functions into a logical pattern which models the task or behavior being supported.

- o User Interface: the front-end which facilitates access to system services and which is predicated on a model rooted in the user's prior experience or *a prior* conceptualization of the task.

3.1 Software

One of the deliverables of the project is a test hyperweb, constructed with HyperCard software. This commercial package has been available for a short period, thus it is much too early to assess its impact. Indeed, most computer journalists admit that it is difficult to pinpoint just exactly what HyperCard is --

> Has anybody figured out what *HyperCard* is yet? An application? System software? A universal data standard that lets you store text, data, graphics, sound and even logical structures in a common format that can be shared by diverse applications? An intuitive programming tool that does to programming languages what calculators did to log tables?
> -- Johnson and Juarez, "Flash Cards," *MacUser*, February 1988, p. 276.

Despite the early confusion, developers and advocates predict major impacts on all domains of information processing -- from software engineering to educational technology.

In brief, HyprCard consists of an authoring and a presentation or browsing system, as well as a high-level, object-oriented programming language. The system uses the controlling metaphor of a deck of cards. The five building blocks of HyperCard are

- *Stacks*: The central concept for information development. A stack is a collection of cards, which may be interlinked. The term "stackware" also conveys the idea of dynamic and modularized information processing implicit in hypertext.

- *Backgrounds*: The placement of information is guided by the background. A stack may consist of any number of cards, which may share a single background or which may have a background specifically designed to suit the type of information or to facilitate retrieval and processing.

- *Cards*: The fundamental unit of HyperCard. Cards are identified in three ways: a unique numerical ID; the sequence number (position of the card in the stack); a unique alphanumeric ID given to the card by the designer.

- *Fields*: Slots, or predefined areas on cards which hold information. Fields may be mouse-sensitive, meaning they may initiate a command when clicked on.

- *Buttons*: mouse-sensitive areas of the screen which -- when clicked on -- initiate some action. This may be movement to another card or stack, retrieval of information stored elsewhere, or some sort of data processing -- to mention only a few possibilities.

HyperTalk^{tm}, the English-like, high-level language, provides a programming environment suitable for novices. The language is declarative rather than procedural. Furthermore, it has an Object-Oriented Programming (OOP) feel because programs are built around *objects*, *messages*, *message handlers*, *classes* (backgrounds), and a rudimentary for of *inheritance*.

However, the current version of HyperCard (1.2) has limitations. Notably, it does not support windowing and the screen size remains constant across monitors. Because HyperTalk is not a compiled language, response time for many applications is unacceptable. Additionally, *scripting* is not adequately treated in the manual issued with the software. Danny Goodman's 720 page The Complete HyperCard Handbook -- which can be purchased as a supplement -- is comprehensive but tedious. Crucial to this project is the limitation that HyperCard is a database management system only in the loosest sense of the term.

3.2 System Design

System designs are influenced by many constraints -- things such as availability of resources, state of knowledge, and inventiveness of the researcher. This segment discusses two conditions which influenced the demonstration software's specifications:

- o the implicit and explicit logic patterns of the paper text being reworked into a hypertext representation, and
- o the information representation mandated by stackware as well as the system functionality made possible by object-oriented programming.

From this combination of text requisites and software capabilities comes the basic architecture of the frames making up the system.

3.2.1 Data Preparation: Knowledge Structures of TOs

Translating text into hypertext is a crucial area of concern. The written word -- as it exists today -- has been molded to suit a paper delivery system. While in some applications it may be feasible to construct text especially for a hypertext representation, most implementations will only be cost-effective if documents authored for paper can be modified for a hypertext environment (Raymond and Tampa, p. 143). This translation process raises questions of rhetoric and knowledge structure.

These are not trivial problems. Extracting a workable hyperweb from an existing document can be difficult. In most cases, the author(s) are not available for consultation, so answers to questions of meaning and form may have to be inferred from the paper representation. In many cases, form does not reflect logic because the requisites of paper and of the press exert powerful influences on document design and text formatting. For example, the original paper *Oxford English Dictionary* was restricted to a fixed number of pages by the publishing contract. Therefore, every effort was made to conserve space -- hence the dense typesetting, extensive use of abbreviations and symbols, and limited use of meaningful white space (Raymond and Tompa, p. 146).

Additionally, not all texts are suitable for hypertext representation. Hypertext makes the implicit knowledge structure of a text explicit. Although there is no rigorous model for hypertext implementation, experience suggests that if the document is closely interwoven through rhetorical devices, then decomposition into chunks and links will be difficult, with lose of information and confusion of meaning a result. For some documents, this conversion is either impossible or not desirable because it destroys the subtle interconnection of themes, argument, metaphor, and word choice.

Paper versions of reference documents and instruction sets already have knowledge structures that facilitate access and interpretation. Thus, they are more amenable to deconstruction. The particular TO modeled in this project consists of different types of information, each with its own domain-specific pattern of presentation.

- *Checkout and Analysis*: A highly modularized set of procedural instructions, presented in a step-wise fashion, interspersed with Warnings, Cautions, and Notes, and partially enriched by graphics.

- *Illustrated Parts Breakdown (I.P.B.)*: Hierarchical graphics and alphanumeric tables which are keyed to the graphics by callouts. The tables contain the system's model number, a vendor-supplied part number, the federal supply category number, a brief description, the number of parts per assembly, and the applicable tail numbers.

- *Theory*: A narrative/descriptive presentation of the engineering concepts behind the equipment's specifications and operation. This material is supplemental background for the procedural instructions in the Checkout and Analysis section.

- *Troubleshooting Guide*: Usually a tabular arrangement mapping observable symptoms to components and fault rectification procedures.

- *Schematics*: Visual presentation of system components -- could include such things as wiring diagrams and logic circuits.

Although no fully specified model of how a technician uses a TO exists, conventional wisdom suggests that Section VI, "Checkout and Analysis" is the starting point. For the Transmitter/Receiver TO used in the study, "Checkout and Analysis" is made up of numerically labeled headings and paragraphs -- sequenced 6-1 through 6-57 -- containing instructions for a battery of tests to determine minimum performance of the unit. Subparagraphs (usually a single instruction) are indicated with an *a* through *z* designator. Although there is some integration of graphics, in general the representation is text-intensive.

After a few introductory comments, the logic structure bifurcates, with two separate streams based on which of two standard test sets is being used (AN/UPM-137A or AN/UPM-98A). Although there may be a dozen or more steps in each instruction for a particular test, only three basic activities are being described: *connecting cables*, *setting dials and switches*, and *interpreting readings*. After each test, procedures for returning the equipment to its original configuration are given; these are essentially a reversal of the instruction set. Figures 3 and 4 give a graphical representation of the hierarchical structure and recursive patterns implicit in the text.

MINIMUM PERFORMANCE TEST

AN/UPM-137A	AN/UPM-98A
Prelim. Op. Check	Interrogation PRF
Receiver Sensitivity	Interrogation Frequency
Bandwidth & Center Freq.	Interrogation Pulse Width
Trans. Frequency	Prelim. Op Check
Trans. Power	Receiver Sensitivity Check
Auto Overload Control	Bandwidth & Center Freq.
SLS Check	Trans. Frequency Check
Standby Enable Check	Transmitter Power
Mode Decoding Check	Automatic Overload Control
Mode Enable Check	SLS Check
Test Mode Enable	Standby Enable Check
Clock Pulse & Frequency	Mode Decoding Check
Reply Encoding Check	Mode Enable Check
I/P Enable	Test Mode Enable Check
Auxiliary Trigger Check	Clock Pulse & Frequency
Emergency Enable Check	Reply Encoding Check
Suppression Out Check	I/P Enable Check
Suppression In Check	Emergency Enable Check
Mode 4 Decoding	Suppression Out Check
Mode 4 Check	Suppression In Check

Figure 3: Knowledge Structure of Candidate TO

MODE ENCODING CHECK. To check mode decoding first perform preliminary procedure contained in paragraph 6-10, then perform the following steps in the order given.	**Introduction -- Common to all tests**
Steps a - f all have action verb beginning (a) = set; (b) = adjust; (c) = adjust; (d) = set: (e) = set; (f) = set	**Instructions for equiptment set-up and interpreting readings**
NOTE If continuing with performance test of receiver-transmitter, return set-up to initial condidition by performing steps g and h.	**Choice point -- Common to all tests**
Steps g and h begin with action verb set; return AN/UPM-137A to preliminary configuration	**Instructions for return to preliminary configuration**

Figure 4: Knowledge Structure of Individual Test

3.2.2 Basic Architecture

HyperCard's basic metaphor for information management is the stack. Data is stored on electronic cards, which can then be threaded and sorted in different ways.

Demonstration Stackware

The data arrangement in the demonstration software mirrors the hierarchical model of the paper TO. The user enters the system at a global map (see Figure 5), which serves as an embedded menu. For convenience, only four tests from each of the two branches are included in the demonstration software.

Each test description has two tracks:

> *More* -- an information-rich, graphic-intensive version intended for the novice
>
> *Less* -- a sparse rendition, perhaps a tabular arrangement of connections, settings, and readings for the expert who needs only to refresh her memory about specific details.

Figure 6 depicts the basic architecture of the system.

While in the modularized card stacks of "Checkout and Analysis," the user can also access appropriate modules from the other four bands of information (Troubleshooting, Theory, I.P.B, and Schematic), either by following the pre-defined links or by requesting a specific frame. Figure 7 conceptualizes the entire system and its potential web.

Nomenclature

HyperCard identifies cards in three ways:

- o A unique ID randomly generated for each card by the system software -- a four digit number permanently assigned to the card
- o The number representing the position of the card in the stack, which changes with additions and deletions over time
- o A designer-specified alphanumeric, whose form should convey meaning about the stack's content or structure, or both

Because of the nature of HyperTalk, using a pattern in the designer-assigned ID tags increases the system's functionality and expansibility. The demonstration project uses a nomenclature modeled on the Dewey Decimal System -- a library indexing system now almost universally supplanted by the Library of Congress System. Using this

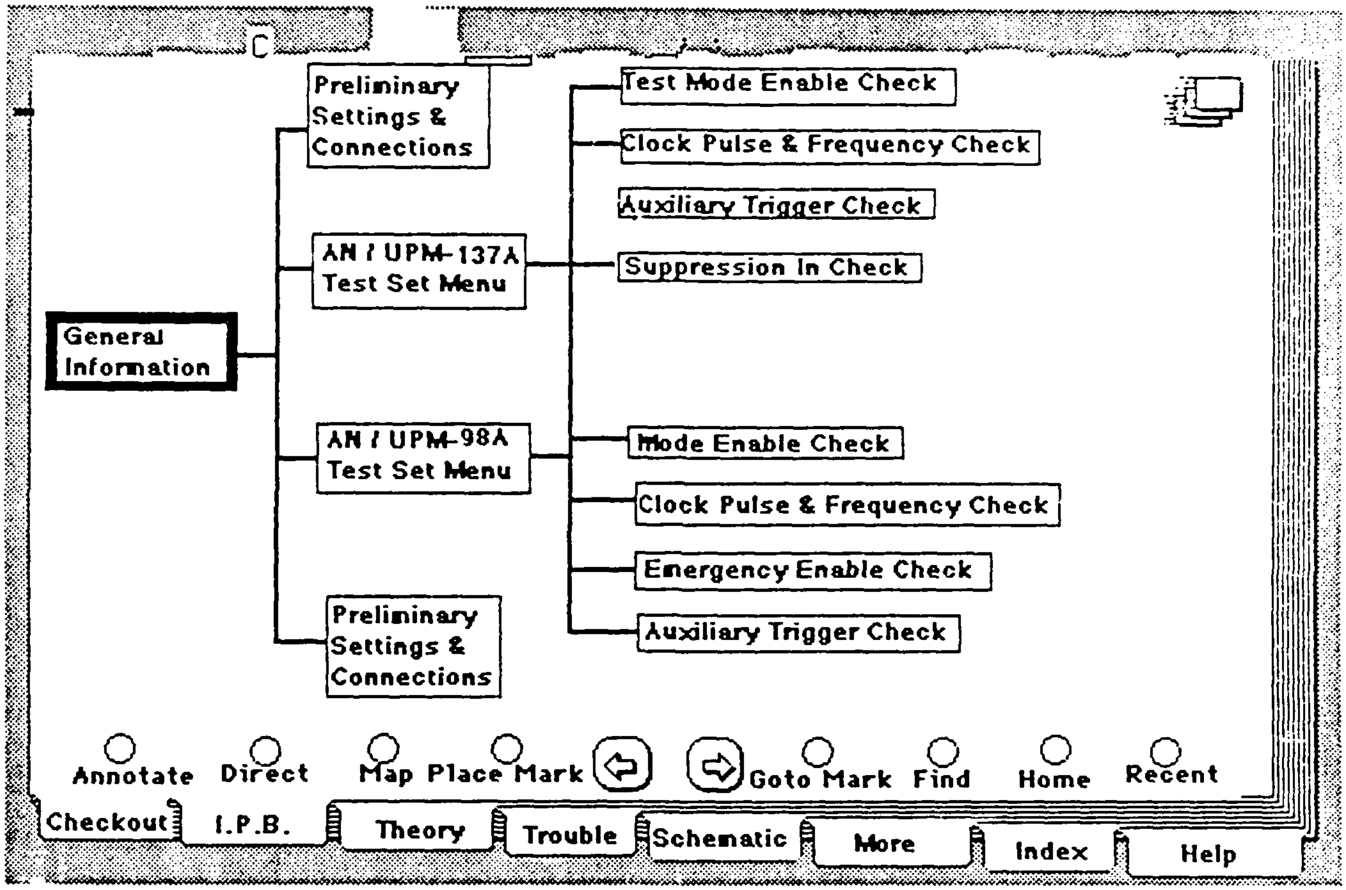


Figure 5: Global Map

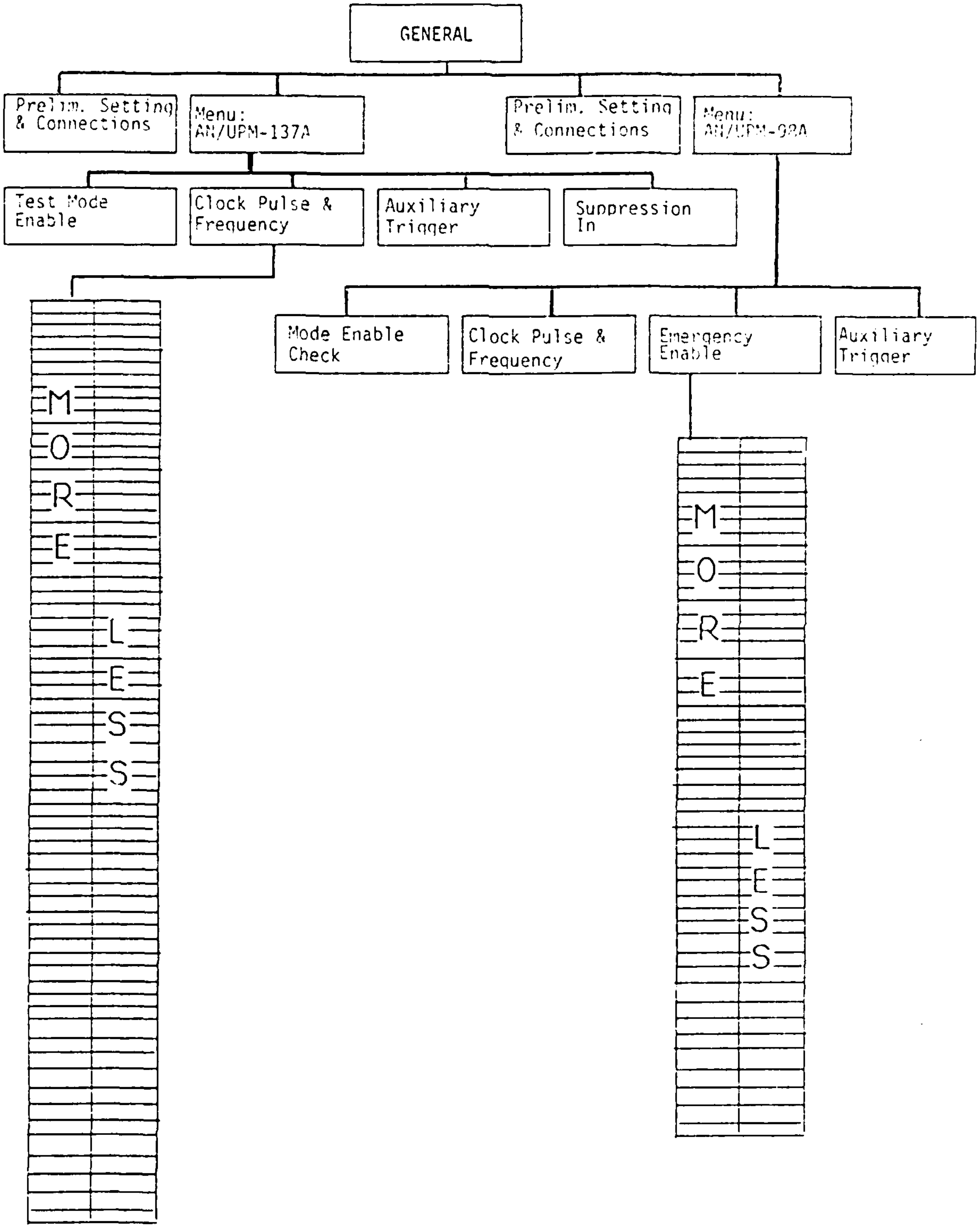


Figure 6: Backbone Architecture

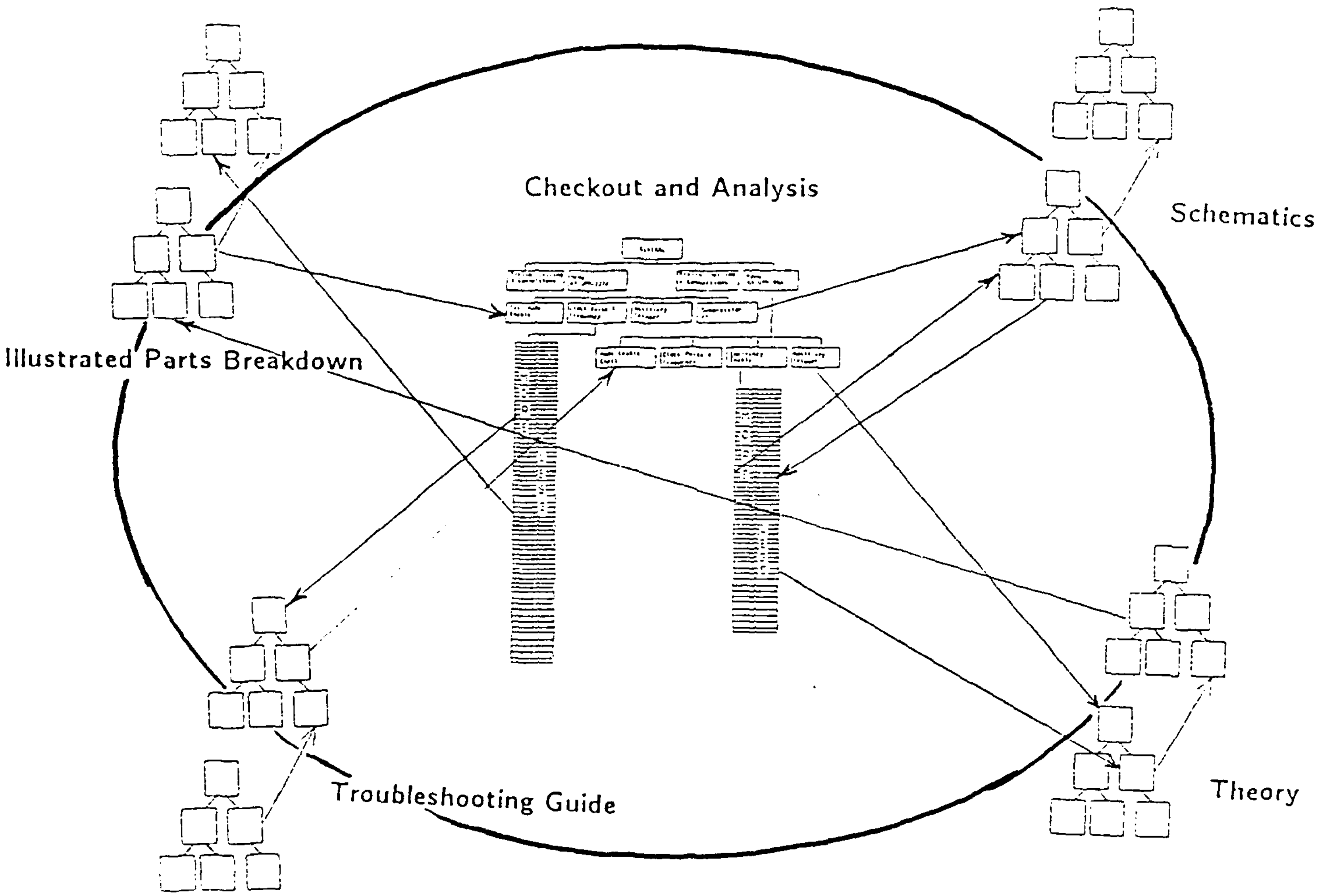


Figure 7: System Architecture

patterned signature is valuable because it echoes the knowledge structure of the system, is open-ended and thus expansible, and has mnemonic value.

Each stack has a specific identifier of the generic form:

C	137A	- SIC	- M
1	2	3	4

where

position 1 is the first letter of the content area (Checkout and Analysis, Theory, TRoubleshooting, Schematic, Illustrated Parts Breakdown),

position 2 is the test set designation,

position 3 is an abbreviation of the specific test,

position 4 is a designator for the More or Less track.

Individual cards in each stack are given the same letter as their step indicator in the procedure sequence (e.g., step a, step b, . . . step z). Because some steps were broken down and placed on separate cards, these indicators no longer match the paper TO nomenclature.

The user can call a stack directly, by using its alphanumeric, and can call a specific card by using the stack name plus an alphabetical designator (entered separately in two dedicated dialogue boxes). Once the nomenclature system becomes familiar, the user can intuit structure and place in the knowledge web. Even a relative novice can guess at position in the web.

The utility of having an open-ended, expansible nomenclature was not fully exploited in the demonstration project. However, some suggested functionality which can be generated with very low overhead include *filters* which help with rapid transversal through the web and automation of various *sorting operations* in individual stacks.

3.3 User Interface

Designing a coherent, consistent, and cogent user interface may be the single most important aspect of hypertext development. Three danger areas must be addressed:

o *Disorientation*: freedom and flexibility create confusion unless there are many navigational aids in the user interface.

- o *Modularization*: determining what constitutes a node (establishing sets and subsets); addressing issues of style, structure, substance, syntax for individual frames (establishing contextuality, readability, consistent meaning).

- o *Cognitive Overhead*: in a free-form system, the user must constantly make meta-level judgments about the flow of ideas; even in relatively structured systems, the user runs the risk of needing to monitor too many levels or modes of presentation.

Figure 8 shows a typical information screen used in the demonstration software.

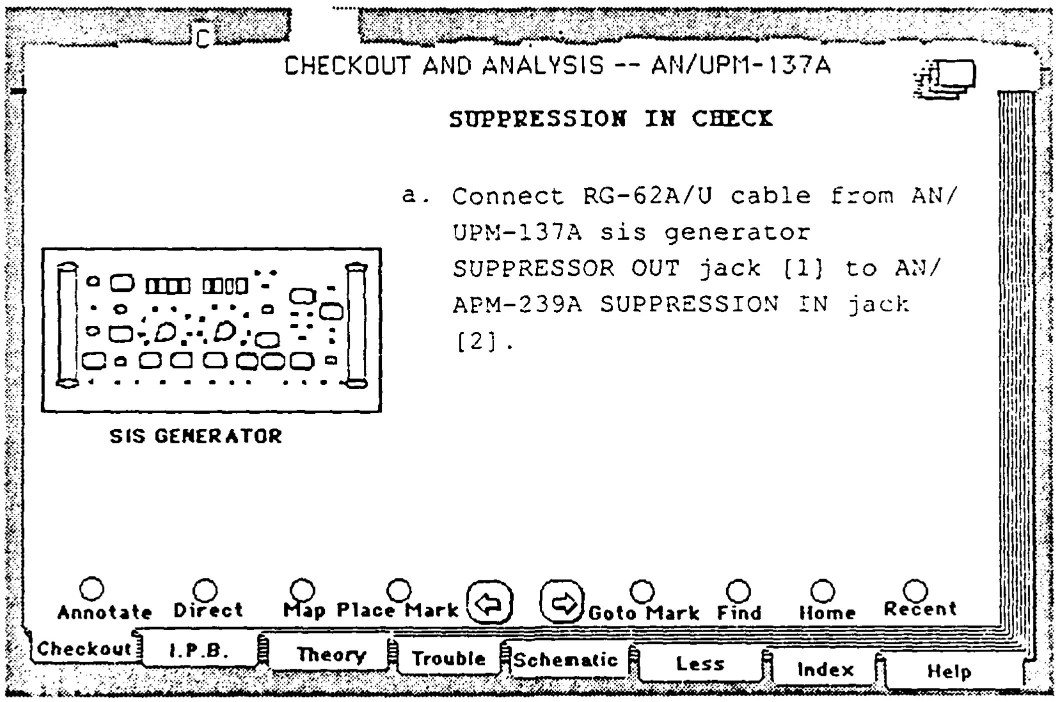


Figure 8: Canonical Information Screen

Of the items on the tabs at the bottom of the card, only *Index* and *Help* need further explanation. *Index* gives the user an alphabetical list of the abbreviations for specific tests. *Help* contains assistance on the operation of the demonstration software. *Checkout*, *I.P.B.*, *Theory*, *Trouble*, and *Schematic* take the user to the indicated band of information. *More/Less* is a toggle switch for selecting one of the two tracks of presentation (lean or rich).

Items represented by the "radio buttons" provide functionality and are modeled on the information management strategies used in paper delivery systems. Many proponents of hypertext overstate the case against "flat" text. They concentrate on the inflexibility of the printed word, the sequential nature of the presentation, and the sheer bulk of text mandated by conventional composition patterns, such as the paragraph.

However, few documents are read from cover-to-cover in the lock-step, serial fashion depicted by hypertext advocates. To the contrary, good documents are designed to facilitate rapid perusal. Furthermore, the discipline of document design has flourished over the past decade and has produced new techniques and new awareness for meeting reader's needs in paper text.

Sophisticated readers know how to navigate through a text without savoring each individual word. Skilled readers have a number of strategies and tactics for extracting meaning quickly from paper documentation. Some examples of functionality gained through training or experience include:

- o Various methods of perusal -- e.g. thumbing, glancing, skimming, and browsing.

- o Individual markup techniques -- e.g. writing in the margin or on the end papers corrections, summaries, mnemonic devices, commentary, conversion of prose to tables and graphs, lists, cross-references, key terms, diagrams, indexing aids.

- o Placing bookmarks or "dog ears," marking trails in various manners.

- o Intuiting place and content through visuo-spatial cues.

- o The psychological benefit of tactile contact with paper documentation (a form of hand-eye-brain relationship not yet fully researched).

In addition, modern methods of document design greatly increase ease-of-use, timeliness of information, and accessibility in a paper delivery system. Design techniques include:

- o Alternatives to the paragraph as the basic unit of prose (e.g. structured writing and information mapping, other devices for information compression).

- o Improved printing and graphics technology, including the emergent capabilities of desktop publishing.

- o Improved techniques for indexing, referencing, and branching.

- o Integrated graphics which signal knowledge structure (e.g. visual table of contents and logos as signposts).

- o Mapping presentation to user needs (e.g. task orientation versus reference function).

Any electronic information delivery system must be able to -- at a very minimum -- duplicate the capabilities provided by the combination of an experienced reader and a well-designed text. Anything less degrades the system: leaving at best. an electronic page-turner; at worst, even less than the paper version. Furthermore, in order to justify abandoning the conventional method and medium, an electronic system should offer improvements, such as increased flexibility, reduction in storage, and convenient document development/maintenance.

The remainder of the section describes and assesses the 10 features in the prototype intended to allow the end user to more productively interact with the fragmented textbase.

- o *Annotate*: One of the most powerful strategies users of paper documentation have is the ability to annotate text. These activities include such things as writing in the margin or on the end papers; corrections; summaries; mnemonic devices; comments; conversion of prose to graphics; lists; cross-references; underlining key terms; marking important passages; making or altering diagrams; and indexing. These capabilities constitute an individual markup language which aids in the processing of information.

  The prototype system supports annotation by allowing the user to create a card at any time and enter commentary. These cards are kept in a separate stack, called "Annotate" and draw their nomenclature (card ID) from the card visible on the screen when the user presses the Annotate button.

- o *Link*: Linking is the essence of hypertext. Although complete freedom in constructing a web may be neither necessary nor desirable for automated TOs, there is utility in allowing readers to make links:

  - -- Capturing associations made while reading/browsing the default hypergraph.

  - -- Electronic support for checking/scanning information.

  - -- Individualized pathways through a familiar body of information.

  A linking icon does not appear on the screen because this is a constant function of HyperCard and not of the prototype software. The user selects the New Button option on the Objects pop-down menu and then enters the appropriate information in the Button Info dialogue box.

- o *Direct*: Allows a user to go directly to a card in given stack. Since the nomenclature has mnemonic value, an experienced technician should be able to enter an alphanumeric and move directly to the card she wants.

Having a tutor mode available at this point, makes the feature even more powerful. If unsure of the card name, the user can request a listing for the stack. This listing could be interlinked with other listings, so that a user could thread a path through the entire system using the nomenclature dimension.

- *Scan*: Skilled readers vary their rate and have many methods of perusing text to extract meaning -- e.g., thumbing, glancing, skimming, and browsing. The *show all cards* command causes a HyperCard stack to flip through the deck, emulating a rapid browsing mode. A "speeding cards" icon is available on all cards in the prototype. Clicking on this button activates the script; a second click stops the scan. The exposure time of each card on the screen can be adjusted in the script. Additionally, the programmer can specify how many cards should be scanned.

- *Map*: User overload is a major human factors concern in any information integration system. Empirical evidence supports the contention that visuo-spatial (e.g., use of icons) representations and retrieval are easier than symbolic (words and numbers), and the success of workplace metaphors in commercial systems warrants considering spatial management devices.

  The "map" feature gives a rudimentary iconic representation of the knowledge structure. As such, it acts as an aid in navigation and as a visuo-spatial system summary. The map, as delineated here, also acts as an embedded menu.

- *Back* and *Forward*: These navigational scripts are constants in HyperCard and provide the basic mobility of moving sequentially through the card file.

- *Find*: This button calls a HyperCard constant -- a string search capability. Clicking on the find button brings up a dialog box. The user enters the specific text between the quotation marks.

- *Home*: This feature is a system constant. Clicking the home button takes the user to the highest card in the system software.

- *Place Mark* and *GoTo Mark*: Readers of paper systems are able to make rapid comparisons of one page with another page. (This is usually accomplished by holding the various places in the volume with one or more fingers.) Easy comparison of data, text, or graphics in one location with information in another node is crucial for some maintenance operations.

  With this feature, the user can mark a card (by clicking on the *Place Mark* button), progress to any other card in the stack, and then return to the marked card by clicking on the *GoTo Mark* button. Successively clicking on the *GoTo Mark* button causes the system to "flip" back and forth between the two cards.

- *Recent*: In a given information retrieval task, a reader of paper text visits several different segments of the document. Keeping track of the trail requires a high cognitive overhead, and is not normally something a reader consciously stores without mechanical aids -- such as "dog-earing" or placing bookmarks. The prototype software includes two variations on mechanical trail-markers to help the user conceptualize a path through the fragmentations.

  Recent is a menu command in HyperCard which helps in revisiting cards during an interactive session. As the user moves from card to card and from stack to stack, the system keeps a "picture" log of each frame. Selecting "Recent" on the Go pop-down menu shows up to 42 cards in miniature.

  Only one representation of a card appears, no matter how many times the user has accessed a particular frame. Thus, the record is not a true session log, nor are the cards necessarily represented in the sequence in which they were opened. A second limitation of menu *Recent* for this application accrues from the canonical screen design which causes all the miniature representations to look very much alike. This can be overcome by rewriting the primitive script for *Recent* so that the names of the cards appear, rather than the image.

  The demonstration software's *Recent* button activates a script which scans through -- in full size -- the last ten cards visited.

4. System Enhancements

Designing and building the prototype suggested areas of improvement, some of which could be added to the current demonstration and others of which may be beyond the scope of HyperCard. This segment considers three generic categories where additional features and functionality might be woven into the web. They are (a) screen display, (b) electronic storage paradigms, and (c) smart retrieval mechanisms.

4.1 Screen Display of Text -- Graphical Interface

Even in early, static representation of text for screen display, designers recognized that the rhetorical strategies for paper copy would not translate directly to text for online use. Deconstructing linear prose poses even more difficulties for consistency, contextuality, access, and mobility. However, storing text in small, information-rich chunks (perhaps not more than a sentence to each module) may not have as many disadvantages as has been suggested in the literature, provided the user interface is enriched by visuo-spatial enhancements.

Location is one of the most difficult problems in hypertext design. Electronic presentation of text removes most of the location cues available in paper. Adding to the confusion, hypertext encourages the proliferation of small nodes. The result can create chaos for the end user. Adopting some form of visualization vastly improves system usability.

Various frames of Figures 9 show how information from a TO can be accessed graphically. Each element on the diagram is a mouse-sensitive button. The user clicks on the appropriate item to call up information which appears in the scroll box.

An alternative approach to TO automation is to store text as independent nodes in a database and to design different types of user interfaces (iconic, fisheye or distorted view, synoptic or information map, graphic flyover, and representational graphics), which can then be bolted on as a front end to the system based on the user's preference.

4.2 Electronic Storage -- Object-Oriented Language Environment

Early online documentation was stored as a sequential file. Access was provided by selecting from a menu (whose structure mirrored the document's table of contents), by keyword search coupled with a fuzzy match algorithm, and by string search. Additional refinements included Boolean operators and successive iterations in the search.

While this approach is attractive for some applications and while there are some hypertext-like systems which have adopted the method (see Remde et al.), there are significant limitations to adaptability and functionality. On the other hand, by storing text as "chunks" and by using an object-oriented programming environment, the basic structure lends itself to incorporation of software advancements in semantic networks and logic programming.

In procedural programming, code is written as procedures which operate on certain types of data. This data is stored separately from the procedures or functions which operate on it. In object-oriented programming (OOP), both data and source code are stored together in packages called *objects*. Four other concepts are central to OOP:

- o *Method*: the code or instructions stored within an object.
- o *Message*: analogous to a procedure or function call -- initiated by the user or by an event in the software environment.
- o *Class*: groups of objects with common characteristics.
- o *Inheritance*: objects are arranged in hierarchical order and inherit behavior potentials (notably methods) from their ancestors.

Clearly, object-oriented programming offers enormous potential to hypertext designers. Not only are chunks of text linked together in static webs -- thereby providing the dimension of 3-D structure to the textbase -- but *a fourth dimension* (a means for dynamic reconfiguration) is now available. In other words, the mobility and knowledge structure of the web are augmented by chunks of text (nodes) which

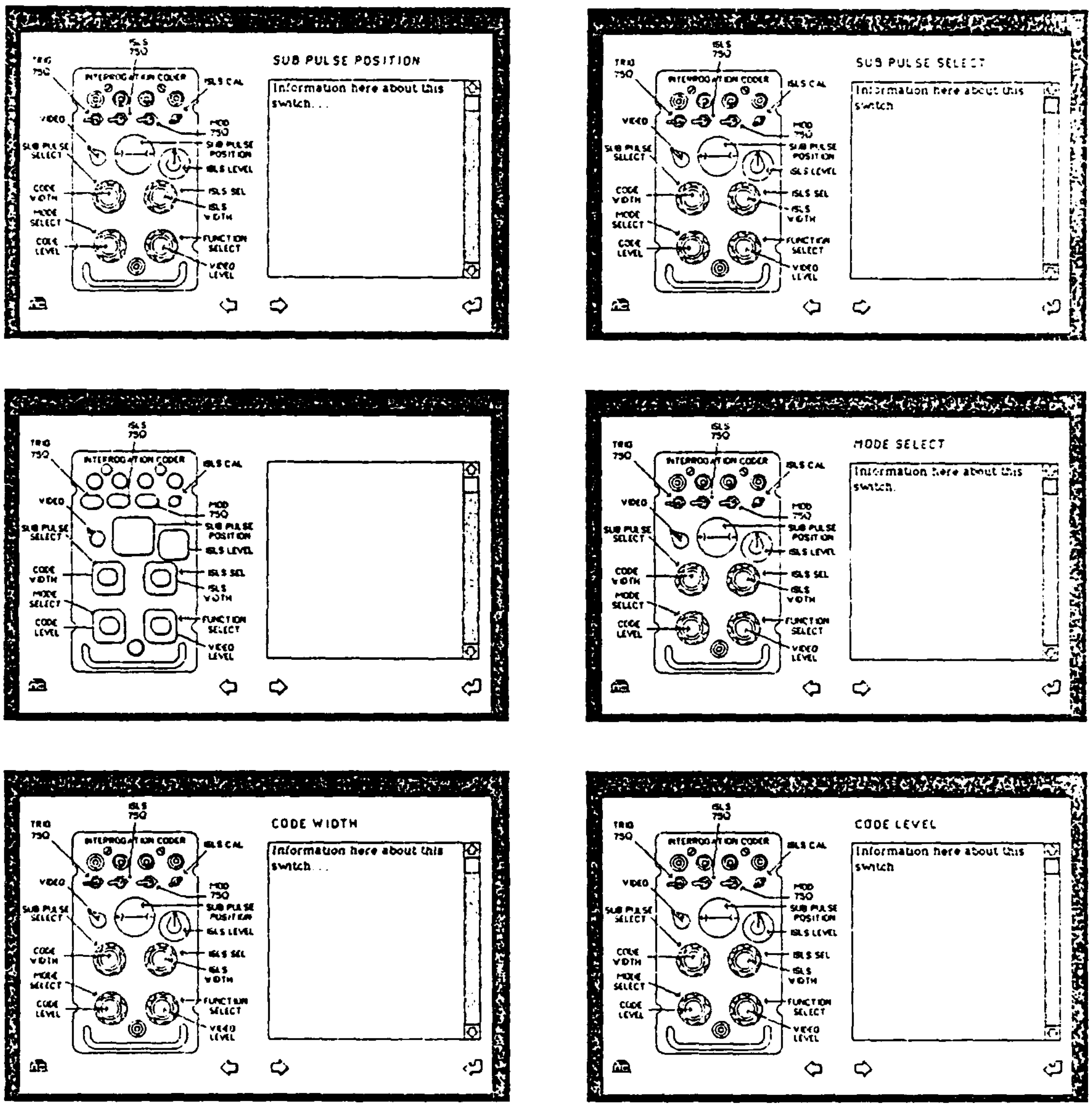
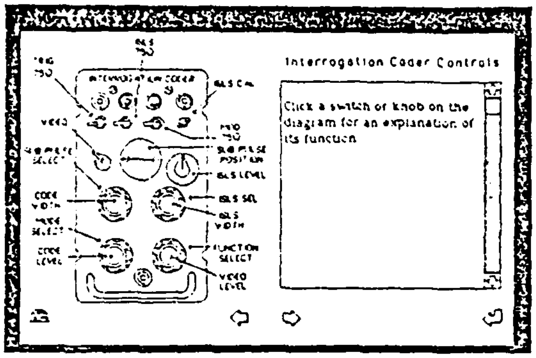
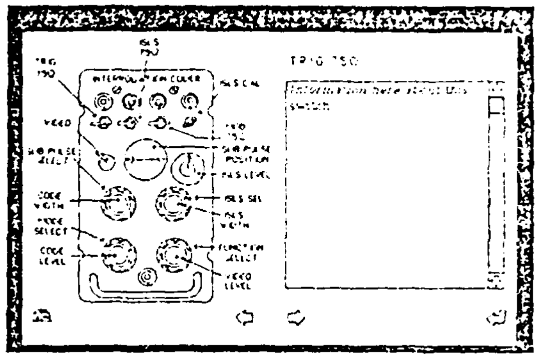
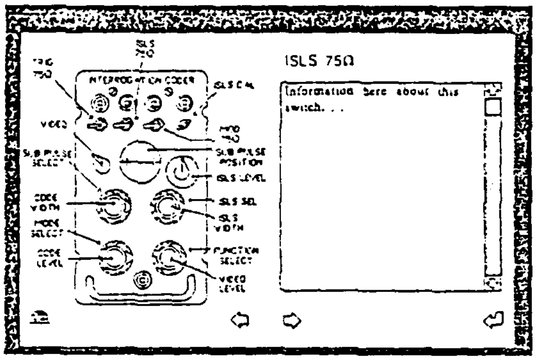
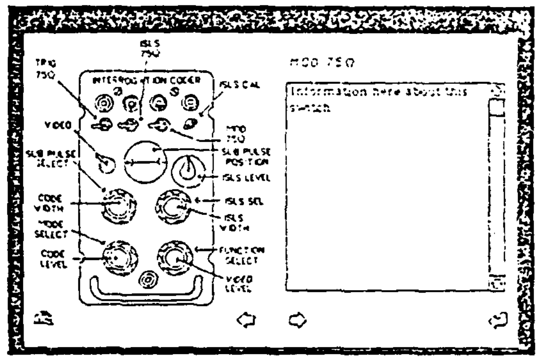


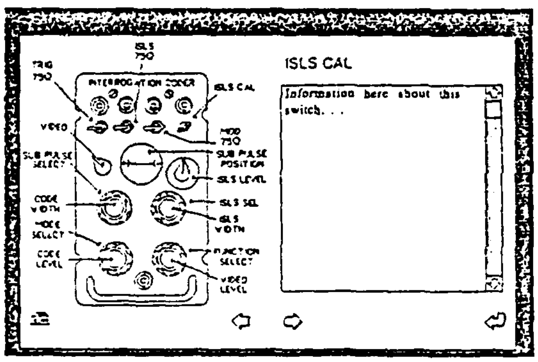
Figure 9: Graphical Interface

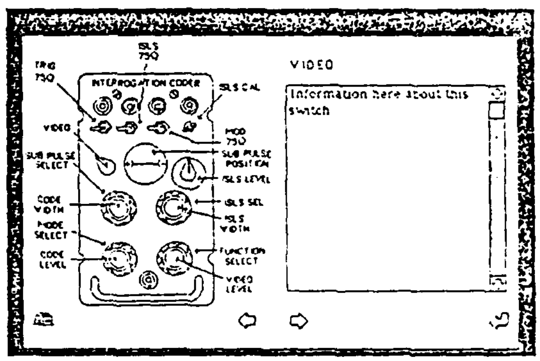

Interrogation Coder Controls
Click a switch or knob on the diagram for an explanation of its function

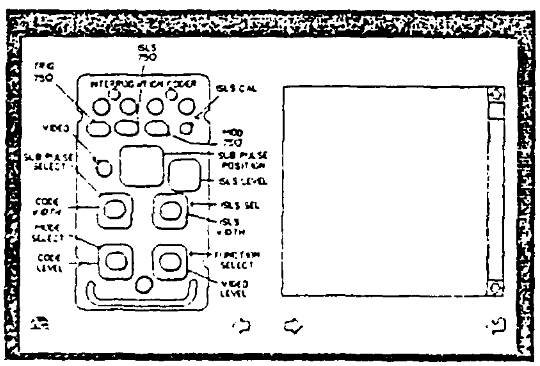


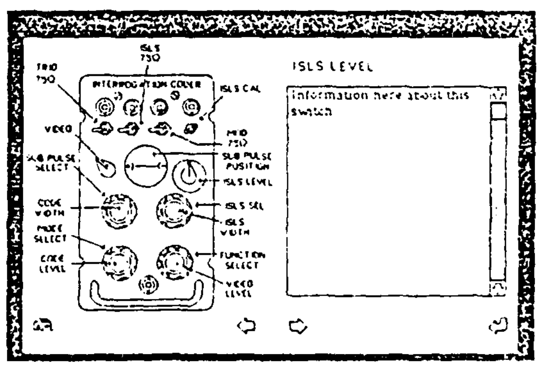

ISLS 750
Information here about this switch. . .


Information here about this switch


ISLS CAL
Information here about this switch. . .


VIDEO
Information here about this switch




ISLS LEVEL
Information here about this switch

can talk to one another or trigger events in the software environment. For example, in the demonstration software, each card is an active node capable of accepting, processing, and emitting messages traveling through the network. Therefore, carefully designed scripts could emulate expert systems by passing the user from node to node in an intelligent fashion.

4.3 Retrieval Mechanisms for Text -- Smart Interfaces

Because brute force or blind search through all nodes in a database produce a combinatorial explosion, almost all research into automated information extraction focuses on ways to more fruitfully direct the search algorithm. Attempts to produce a "superbook" have merged the forces of information retrieval, database management systems, and artificial intelligence to focus on the single crucial issue of implementation: how to provide search procedures which are accurate, complete, and comfortable for the end user.

In general, a hyperweb allows for two very different types of retrieval activities, depending on whether one concentrates on the nodes and their content or whether one focuses on the arrangement of the nodes and the structure of their connecting links.

If one elects to use the node content, then techniques of modern database management are of value. Areas of intensive research today include natural language query, vector and probabilistic models, and the use of a "seed" (statistical profiling of a sample which is then used as a query template).

On the other hand, if one elects to concentrate on the web, then a more graphic approach exploiting knowledge representation techniques from AI is appropriate. Because of its structure and connectivity, a hypergraph incorporates intelligence, perhaps more than is possible with most current, static database storage representations. Retrieval facilities that will squeeze as much functionality out of the web as possible can be grafted onto the system, making a hypertext document database a symbiotic relationship between the intelligence built into the web and the capabilities of a smart interface. These techniques may draw from the current research on semantic networks.

5. Hyper-Rhetoric

This segment briefly examines -- through the methodology of discourse analysis -- how an experienced reader determines patterns of meaning, form, and coherence in a piece of prose. Based upon these observations, ways in which hypertext systems can support the more subtle aspects of human text process are indicated.

Although rhetoric is a 2,000 year old discipline, only in the past two decades has scholarly research in linguistics, psycholinguistics, and cognitive psychology given us a rudimentary understanding of the concept Freshman Composition teachers refer to as *coherence* -- or the sense of structure in a prose passage. Clearly, we do not read sentences as separate entities nor do we simply extract meaning, on a one-for-one basis, from the printed word. We construct meaning. At its most encompassing level, this concept of form includes all the social, psychological, and environmental factors impinging on the reading episode. However, this analysis will consider only aspects of the text itself in deriving a model of what determines *form* in conventional continuous prose.

Rhetoricians suggest that text structure is an attribute of several different signalling devices and that a sense of overarching design is based on the total experience of these multi-determiners. Because of this redundancy, our writing can be less than perfect and still be understood as having a connected meaning. As elements are either mishandled or missing , the sense of coherence erodes, until enough of the system has broken down that the reader loses the sense of the message.

Gregory G. Colomb and Joseph M. Williams suggest a four-layered model, "with multiple strata contributing to the experience of coherent text structure" (p. 90). In adapting their analysis, the present study first describes the signaling cues as they exists in conventional prose representation and then speculates on ways in which the fragmented text of a hypergraph can retain or simulate enough of these cues that the logic of the prose does not disappear in the decomposition.

5.1 Lexical and Syntactical Cueing

Starting with the elements of word choice and sentence structure, this model progresses from microanalysis to marcroanalysis.

- *Functional Sentence Perspective (FSP)*: Sometimes called the "given-new" contract, in its simplest definition this means that the writer front-loads the sentence with information already referred to or information for which the reader already has a sense of meaning. The writer then introduces information that is newer or less familiar than the previous discourse.

- *Topic Strings*: The old/new pattern is important for individual sentences, as well as for giving a sense of connectivity to aggregates of sentences. Another attribute which weaves ideas together is the notion of *topic*. This is not the grammatical idea of sentence subject, but the item which the reader tacitly takes as the psychological subject of the sentence. The echoing of this psychological subject across syntactical structures creates the sense of "aboutness" which molds sentences into paragraphs.

- *Lexical Strings*: These are strings of related words which cue the reader about the discourse domain and thus bring to mind a frame of reference which minimizes the potential for ambiguity in the interpretation.

In the hypertext world, this reciprocity between "given" and "new" can take forms other than verbal entities. In canonical types of writing (such as the formulaic patterns used in fully proceduralized job aids), the reader can pick up contextuality by the structure of the prose. Additionally, integrated graphics are powerful devices for providing referential coherence. For users who have experience with the knowledge domain, these visual cues may suffice. For novices, other more specific textual cues may be necessary. Already, most text-oriented displays repeat the heading and the subheading of a unit on individual frames. More fine-grained cueing might include a small, unobtrusive "pane" which displays a keyword or phrase extracted from a previous frame or two.

5.2 Discourse Unit

A unit of discourse (or d-unit) is "any stretch of continuous text -- a whole text, a section, a paragraph, even a small group of related sentences -- that functions as a unit and whose parts are more related to each other than to those outside the d-unit" (Colomb and Williams, p. 102). In traditional paper text, discourse units are represented by formal devices such as indentation of paragraphs and topic or subtopic headings. Other conventions of rhetoric contribute to the definition of a discourse unit. Typically, a well-formed passage has an opening segment which specifically announces what is "at issue" in the d-unit. The remainder of the unit unpacks the issue by explaining, describing, illustrating, drawing conclusions from, or otherwise developing the issue (Colomb and Williams, p. 103).

Hypertext fragmentation threatens to nullify many of the d-unit definitional characteristics, which are obviously highly dependent upon contiguity. To retain some sense of cohesion present in the paper form, hypertext presentations need to preserve contextuality and to provide a means by which the user can *see* the fragment's place in the coherent whole.

One easily implemented feature is rapid scanning, both forward and backward. Another approach is to implement a windowing configuration so that the entire d-unit is displayed as a scrolling background and the content of the specific node being accessed appears in an overlay window, graphically represented as exploding out from its original context. Yet a third technique for sustaining the logic of the d-unit while fragmenting its elements is to present two windows on the display screen. One window contains the text of the node. The other widow contains a dynamically updated graphical representation of the logic structure of the unit, with the area the user is currently viewing highlighted in some manner.

6. Project Management

The bulk of the project was accomplished from December 1, 1987 to March 4, 1988. In addition to the Principle Investigator, six students worked on the demonstration software. An accounting of student wage-hours and a comparison of the cost estimate with the final budget are given in Table 1 and Table 2.

Table 1: Summary of Student Assistances' Time and Wages

Name	Total Time	Per/Hr	Total Wages
Mike Allard	15	$3.35	$ 50.25
Dale Blocher	16	3.35	53.60
Bill Jurasz	2	3.35	6.70
Christ Kuntz	38	3.35	127.30
Nelson Nissley	2	3.35	6.70
Mark Taylor	89	3.35	298.15
	162		$ 542.70

Table 2:
Comparison of Estimated Budget and Actual Expenditures

	AFOSR		ROSE-HULMAN	
	Proposed	Spent	Proposed	Spent
a. Salaries & Wages				
Principle Investigator Dr. P. A. Carlson (2/9 x $35,900)	7978.00	7978.00		
Research Assistants (two undergraduates: 160 hrs. x $3 35)			536.00	542.70
b. Permanent Equipment				
Macintosh SE Memory Upgrade Image Writer Expanded Keyboard	3406.06	3505.02		
c. Expendable Supplies			300.00	300.00
c. Travel				
Hypertext Conference Brooks AFB, Feb., 1988	630.00	504.42		
e. Publication and Report Costs				
Page costs (professional journal) and/or printing (campus shop)			300.00	--------
f. Computer Costs				
Additional to Macintosh			200.00	200.00
g. Employee Benefits				
28.5% of $7978.00 (Principal Investigator's salary)			2273.73	2273.73
h Indirect Costs				
49.6% x $7978.00	3957.09	3957.09		
	15,971.69	15,945.43	3,609.73	3,316.43

Project Total	Estimated	Spent
	$19,581.42	$19,261.86

7. Selected Bibliography

Armstrong, Anne A. "Navigating with Search Software," Federal Computer Week, July 13, 1987, Vol. 1, no. 16, pp. 26-27, ff.

Brandt, Craig M., et al. "Training Technology Scenarios for an Integrated Automated Job Aiding/Training System," Draft Version, 1986. Systems Exploration, Inc. Contract No. F33615-85-C-0010.

Bush, Vannever. "As We May Think," The Atlantic Monthly. July 1945, Vol. 176, no. 1, pp. 101-108.

Colomb, Gregory G. and Joseph M Williams. "Perceiving Structure in Professional Prose: A Multiply Determined Experience." in Lee Odell and Dixie Goswami, eds. Writing in Nonacademic Settings. New York: The Guilford Press, 1985, pp. 87 - 128.

Colvin, Gregory. "The Current State of Text Retrieval," CD ROM: The New Papyrus. Microsoft Press, 1986, pp. 131 - 136.

Conklin, Jeff. "A Survey of Hypertext," MCC Technical Report No. STP-356-86, Rev. 1, February 9, 1987.

Duffy, Thomas. "Preparing Technical Manuals: Specifications and Guidelines," in The Technology of Text: Principles for Structuring, Designing, and Displaying Text. Vol. 2 David H. Jonassen, ed. Englewood Cliffs, New Jersey, Educational Technology Publications, 1985, pp. 370-392.

Engelbart, D. C. "A Conceptual Framework for the Augmentation of Man's Intellect," in Vistas in Information Handling, P. Howerton, ed. Washington, D.C., 1963, pp. 1-29.

Fox, Edward A. "Information Retrieval: Research into New Capabilities," CD ROM: The New Papyrus. Microsoft Press, 1986, pp. 143 - 174.

Frenzel, Louis E., Jr. Understanding Expert Systems. Indianapolis: Howard W. Sams, 1987.

Frisse, Mark Edwin. "Searching for Information in a Hypertext Medical Handbook," Proceedings of Hypertext '87, University of North Carolina, Chapel Hill, pp. 57 - 66.

Furnas, George W. "Generalized Fisheye Views," Proceedings CHI'86, Boston, Massachusetts, April 13 - 17, pp. 16 - 23.

Gill, Jock and Toby Woll. "Full Text Management," CD ROM: The New Papyrus. Microsoft Press, 1986, pp. 137 - 141.

Halasz, Frank G. "Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems," Proceedings of Hypertext '87, University of North Carolina, Chapel Hill, pp. 345 - 365.

Herot, Christopher. "Graphical User Interface," in Human Factors and Interactive Computer Systems: Proceedings of the NYU Symposium on User Interfaces. Y. Vassiliou, ed. Norwood, N.J.: Ablex Publishing Corporation, 1984, pp. 83 -103.

IMIS: Integrated Maintenance Information System -- A Maintenance Information Delivery Concept, unpublished paper, AFHRL/LRC, Wright-Patterson AFB, August 1986.

James, Geoffrey. "Artificial Intelligence and Automated Publishing Systems," in Text, ConText, and HyperText. Cambridge: MIT Press, 1988, pp. 15 - 24.

Johnson, Laura and Victoria Juarez. "Flash Cards." MacUser, February, 1988, pp. 276 - 295.

Koved, Larry and Ben Shneiderman. "Embedded Menus: Selecting Items in Context," Communications of the ACM, April 1986, Vol. 29 (4), pp. 312 - 318.

Marchionini, Gary and Ben Shneiderman. "Finding Facts vs. Browsing Knowledge in Hypertext Systems," IEEE Computer, January, pp. 70 - 80.

Monarch, Ira and Jaime Carbonell. "CoalSORT: A Knowledge-Based Interface," IEEE Expert, 1987, Vol. 2 (1), pp. 39 -53.

Nelson, T.H., Computer Lib/Dream Machines, Chicago, Illinois, Hugo's Book Service, 1974.

Nelson, T.H., "A Conceptual Framework for Man-Machine Everything," AFIPS National Computer Conference and Exposition Proceedings. 1973, Vol. 42, Montvale, New Jersey, AFIPS Press, pp. M21-M26.

Oren, Tim. "The Architecture of Static Hypertexts." Proceedings of Hypertext '87, University of North Carolina, Chapel Hill, pp. 291 - 301.

Raskin, Jef. "The Hyper in Hypertext." Proceedings of Hypertext '87, University of North Carolina, Chapel Hill, pp. 325 - 330.

Raymond, Darrell R. and Frank Wm. Tompa. "Hypertext and the 'New Oxford English Dictionary.'" Proceedings of Hypertext '87. University of North Carolina, Chapel Hill, pp. 143 - 153.

Remde, Joel R., Louis M. Gomez, Thomas K. Landauer. "'SuperBook': An Automatic Tool for Information Exploration -- Hypertext?" Proceedings of Hypertext '87. University of North Carolina, Chapel Hill, pp. 175 - 188.

Shafer, Dan. "Hypermedia and Expert Systems: A Marriage Made in Hyper Heaven," HyperAge: The Journal of HyperThinking, May-June 1988, pp. 26 - 33.

Shneiderman, Ben. "User Interface Design for the HyperTies Electronic Encyclopedia," Proceedings of Hypertext '87, University of North Carolina, Chapel Hill, pp. 189 - 194.

Shneiderman, Ben. "The Future of Interactive Systems and the Emergence of Direct Manipulation," Human Factors and Interactive Computer Systems: Proceedings of the NYU Symposium on User Interfaces. Y. Vassiliou, ed. Norwood, N.J.: Ablex Publishing Corporation, 1984, pp. 1 - 27.

Wells, H.G., World Brain, Garden City, New York, Doubleday, Doran & Co., 1938.

FINAL REPORT NUMBER 91
REPORT NOT AVAILABLE AT THIS TIME
Dr. Ronna E. Dillon
760-7MG-100

STRUCTURAL REPRESENTATIONS OF MULTIDIMENSIONAL CRITERION CONSTRUCT SPACE

Charles E. Lance Ph.D.
Department of Psychology
University of Georgia

7 April 1989

Final report prepared under Universal Energy Systems Project 760 (No. S-760-6MG-031), Air Force Office of Scientific Research Contract No. F49620-85-0013/SB5851-0360.

SUMMARY

Literature relating to models of proficiency true score and measurement method effects on proficiency measures was reviewed. Eight structural models were identified and the appropriateness of each to the U. S. Air Force Job Performance Measurement System (JPMS) were evaluated along several criteria. The empirical fits of three models were tested against Jet Engine Mechanic Job Performance Measurement (JPM) data and none were found to be wholly adequate. Two respecified models (a first-order factor model and a hierarchical model) were subsequently tested and a hierarchical model appeared to provide a more defensible structure to the data. Prototype statistical methods were also demonstrated for estimating proficiency true score components and controlling for measurement method effects. However, these procedures were not successful. An attempt to assess the generalizability of results from the Jet Engine Mechanic Specialty to the Air Traffic Control Operator Specialty failed due to imprecise specification of performance dimensions and/or allocation of tasks to performance categories. Subsequent analyses in a third Specialty, (Aerospace Ground Equipment Mechanic) substantially confirmed findings from the Jet Engine Mechanic Specialty. Implications of findings for modeling the proficiency construct, the measurement of proficiency at different levels of analysis, and the theoretical structure underlying proficiency measures are discussed.

PREFACE

This research was conducted under Universal Energy Systems (UES) Project 760 (No. S-760-6MG-031), Air Force Office of Scientific Research Contract No. F49620-85-C-0013/SB5851-0360. The author thanks Universal Energy Systems, the Air Force Office of Scientific Research, Dr. Jerry W. Hedge, Mr. Terrance M. Donnelly, Capt. Marty Pellum, and Mr. Mark Teachout for their support in this research.

I. INTRODUCTION

Two primary goals of a Joint-Service Job Performance Measurement/Enlistment Standards (JPM) Project begun in the early 1980's by the U. S. Military were to (a) develop comprehensive job performance measurement technologies, and (b) to assess the predictive efficiency of the Armed Services Vocational Aptitude Battery (ASVAB) as it relates to on-the-job proficiency of first-term enlisted personnel. The U. S. Air Force (USAF) was the lead Service in the JPM project for developing and demonstrating a Walk-Through Performance Testing (WTPT) work sample methodology for assessing job proficiency (Hedge & Teachout, 1986) and for determining the suitability of alternative surrogate proficiency measures (e.g., self-, supervisor-, and peer-ratings of proficiency). The USAF Job Performance Measurement System (JPMS) was designed to include proficiency measures taken at various levels of specificity (task, dimensional and global levels) and from various measurement sources (Walk Through Performance Testing, Self, Peer and Supervisory proficiency ratings). Figure 1 illustrates this cross-classification.

The JPMS includes self, peer and supervisory ratings each of job proficiency at the task, dimensional and global levels (Hedge & Teachout, 1986). Also included are high fidelity "hands-on" and interview WTPT measures of proficiency (see Hedge, 1984; Hedge & Lipscomb, 1987; Hedge & Teachout, 1986; Lipscomb, 1984). Walk Through measures were designed to be benchmark measures of proficiency against which alternative measurement methods are validated (Wigdor & Green, 1986).

Level of Specificity	Self	Peer	Supervisor	WTPT
Global	Global & USAF-wide Ratings	Global & USAF-wide Ratings	Global & USAF-wide Ratings	- - -
Dimension	Dimensional Ratings	Dimensional Ratings	Dimensional Ratings	- - -
Task	Task Ratings	Task Ratings	Task Ratings	WTPT Scores

Measurement Source

Figure 1: JPMS Source x Specificity Cross-Classification

The 1986 National Academy of Sciences (NAS) Committee on the Performance of Military Personnel report on the JPM Project (Wigdor & Green, 1986) identified several research and development needs, including: (a) methodologies for the assessment of overall job competency, (b) development of strong theoretical rationale for the measurement of overall job proficiency, (c) methodologies for deriving overall job proficiency measures, and (d) comprehensive specification of underlying constructs of job proficiency. Kavanagh, Borman, Hedge and Gould (1986) also called for formal specifications of the performance construct space that identify (a) relevant dimensions of performance, (b) sources of measurement method effects in job proficiency measures, and (c) measurement methods that best assess various aspects of the total criterion space.

The purposes of the present work were to address several of the research needs identified by Kavanagh et al. (1986) and Wigdor and Green (1986) by (a) reviewing literature relating to proficiency true score and measurement method effects on measures of job proficiency, (b) deriving structural models of proficiency true score and measurement method effects on proficiency measures appropriate to the USAF JPM research effort, (c) determining the empirical fit between structural models and sample data from one Air Force Specialty (AFS), Jet Engine Mechanic (AFSC 426x2), (d) developing prototype means for statistical control of measurement method effects and statistical estimates of proficiency true score components, and (e) determining the generalizability of results to a second AFS.

II. LITERATURE REVIEW

Scope of the Review

Campbell (1982) and Naylor (1982) were optimistic over the prospects for "modeling" performance. The present research concerns structural models of proficiency true score and measurement method effects on proficiency measures (Kavanagh et al., 1986). In this section, literature relating to performance modeling that was excluded from the present review is identified. Subsequently, relevant literature is reviewed.

There are many important considerations for modeling performance, but several of these were beyond the scope of the present research. These included: (a) predictor models (Ghiselli, Campbell & Zedeck, 1981) or models of causes of performance (e.g., Schmidt, Hunter & Outerbridge, 1986), (b) human factors models of information processing, man-machine interface, (e.g., Alluisi & Morgan, 1976) and ability requirements (e.g., Fleishman, 1978), (c) models of rater cognitive processes (e.g., Borman, 1978; Cooksey & Freebody, 1985; DeCotiis & Petit, 1978; DeNisi, Cafferty & Meglino, 1984; Feldman, 1981; Lance & Woehr, 1986; Nathan & Lord, 1983; Zedeck & Kafry, 1977), (d) models of response sets/styles (e.g., Bentler, Jackson & Messick, 1971; Cronbach, 1946; Givon & Shapira, 1984; Rorer, 1965), (e) performance rating as one component of more general organizational behavior models (e.g., Campbell, Dunnette, Lawler & Weick, 1971; Katz & Kahn, 1978; Landy & Farr, 1980), (f) psychophysical scaling models (e.g., Givon & Shapira, 1984; Guilford, 1954), (g) multidimensional scaling, lens, and policy capturing models (e.g., Cooksey & Freebody, 1985; Einhorn & Hogarth, 1981; Hobson & Gibson, 1983; Schmitt, Noe & Gottschalk, 1986; Slovic, Fischhoff & Lichtenstein, 1977; Zedeck & Kafry, 1977), (h) item response models of rating (e.g., Hulin, Drasgow & Parsons, 1982), (i) models of rater training and performance feedback (e.g., Ilgen, Fisher & Taylor, 1979; Lee, 1985; Spool, 1978), and (j) models of leadership processes concerned with the performance appraisal process (e.g., Dienesch & Liden, 1986; Green & Mitchell, 1979). These topics were excluded from the present review because they (a) do not relate explicitly to modeling of underlying performance constructs, (b) concern performance appraisal processes, rather than the actual content of performance measures, and/or (c) tend to include only a single method of performance measurement.

The present review was restricted to literature relating to statistical/structural models of performance true score and measurement method effects on job performance measures. The review drew largely on literature relating to assessment of construct validity of performance measures, statistical theories of performance rating, and statistical estimation and control of rater bias. Relevant literature was identified from (a) literature citations in previous reviews of the performance measurement literature (e.g., Landy & Farr, 1980, 1983), (b) indexing appropriate subject keywords in Psychological Abstracts, and (c) by cross-referencing relevant literature in the Social Science Citation Index.

Structural Models of Performance True Score and Measurement Method Effects

In this section, several structural models of performance identified in the literature review are described, major theoretical propositions of each are discussed, and important differences in theoretical structures of performance true score and measurement source effects are identified. Models are arbitrarily identified by Roman numerals and each is related to a common notational system for comparative ease. For algebraic simplicity, models' components are expressed in standard score form except where noted.

Model I: Multitrait-Multimethod (MTMM) Matrices: Analysis of Variance (ANOVA) Approach

Limitations to Campbell and Fiske's (1959) "qualitative" approach to analysis of MTMM data are now well known (Alwin, 1974; Boruch & Wolins, 1970; Browne, 1984; Kalleberg & Kluegel, 1975; Schmitt, 1978; Schmitt & Stults, 1986). An ANOVA approach to the analysis of MTMM data, developed by Boruch and Wolins (1970), Boruch, Larkin, Wolins and MacKinney (1970) and Kavanagh, MacKinney and Wolins (1971) defines each observed variable, in raw score form, to be a function of four components:

$$\underline{X}_{ijm} = \mu_m + \underline{W}_T \underline{T}_i + \underline{W}_M \underline{M}_j + \underline{E}_{ijm} \qquad (1)$$

where $\underline{X}_{ijm}$ (in raw score form) represents the $\underline{m}$th ($\underline{m} \rightarrow \underline{M}$) person's score on the $\underline{i}$th ($\underline{i} \rightarrow \underline{N}$) trait ($\underline{T}_i$) measured by the $\underline{j}$th ($\underline{j} \rightarrow \underline{J}$) method ($\underline{M}_j$), μ_m (the grand mean) represents a general factor that contributes to each person's score on all $\underline{N}$

traits measured by all J methods, Ws are weights relating trait and method factors to observed scores, and E_{ijm} are nonsystematic measurement errors. Measurement errors and trait and method factors are defined to be independent of the general factor and each other. ANOVA of MTMM data yields estimates of mean trait and measurement method effects. The MTMM ANOVA model may be written in standard score form as:

$$Z_{ijm} = W_T T_i + W_M M_j + E_{ijm} \qquad (2)$$

where Z_{ijm} is the mth person's standardized score on the ith trait measured by the mth method and T_i, M_j, E_{ijm} and the Ws are defined as in Equation 1. Figure 2 illustrates an ANOVA modeling application to the USAF Job Performance Measurement System (JPMS). Here, Ts represent trait (dimensional proficiency) factors, Ms represent method (measurement source) effects, and Ds and TRs represent, respectively, dimensional and task-level proficiency measures. Task and dimension-level proficiency measures are both assumed to reflect dimensional proficiency but at different levels of generality. For clarity of presentation, nonsystematic measurement error effects (E_{ijm}s) are not shown. The ANOVA modeling approach defines only orthogonal dimensional true score and general measurement method effects. Schmitt and Stults (1986) provide an extended review and discussion of limitations to the ANOVA MTMM approach.

Model II: MTMM Matrices: Confirmatory Factor Analysis (CFA) Approach

CFA has become the preferred approach to the analysis of

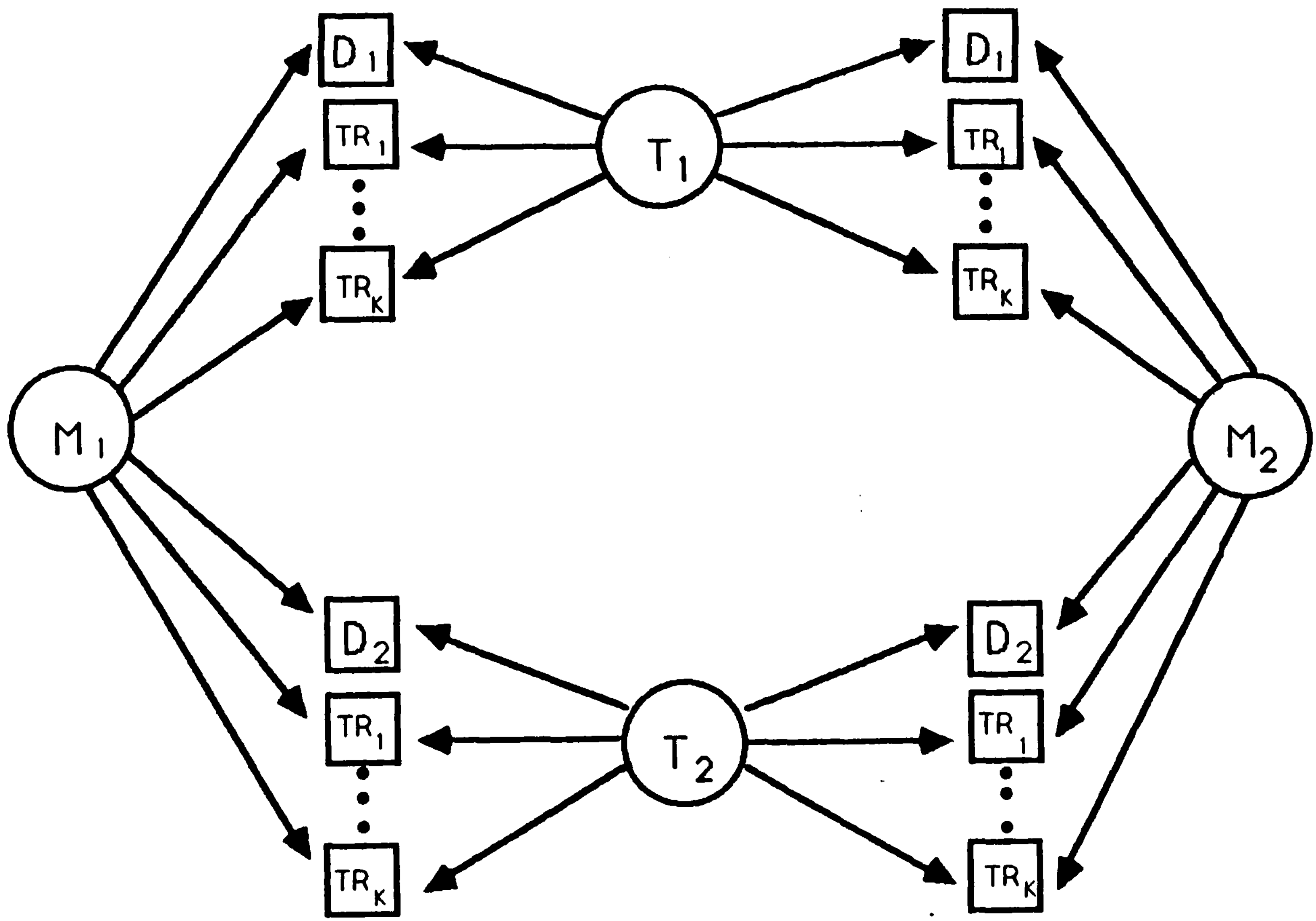


Figure 2: Model I: Multitrait – Multimethod: Analysis of Variance

MTMM data (e.g., Arora, 1982; Browne, 1984; Burnkrant & Page, 1982; Marsh & Hocevar, 1988; Schmitt & Stults, 1986; Widaman, 1985). CFA has path analysis of MTMM matrices as a special case (e.g., Avison, 1978; Kalleberg & Kluegel, 1975; Schmitt, 1978). CFA and ANOVA approaches to MTMM data specify identical trait and measurement method effects. The CFA approach, however, allows for the estimation of correlations among trait and method factors:

$$\underset{\sim}{R}ZZ = \left|\underset{\sim}{\Lambda}T : \underset{\sim}{\Lambda}M\right| \left|\begin{array}{c|c} \underset{\sim}{\Phi}_{TT'} & \underset{\sim}{\Phi}_{TM} \\ \hline \underset{\sim}{\Phi}_{MT} & \underset{\sim}{\Phi}_{MM'} \end{array}\right| \left|\begin{array}{c} \underset{\sim}{\Lambda}'_{T} \\ \hline \underset{\sim}{\Lambda}'_{M} \end{array}\right| + \underset{\sim}{\Theta}_{\delta} \qquad (3)$$

Here, $\underset{\sim}{R}_{ZZ}$ ($\underline{N}$*$\underline{J}$ x $\underline{N}$*$\underline{J}$) is the MTMM correlation matrix, $|\underset{\sim}{\Lambda}_T : \underset{\sim}{\Lambda}_M|$ (($\underline{N}$*$\underline{J}$ x ($\underline{N}$+$\underline{J}$)) is the partitioned matrix of observed variables' factor loadings on trait ($\underset{\sim}{\Lambda}_T$) and method ($\underset{\sim}{\Lambda}_M$) factors, $\underset{\sim}{\Phi}$ contains correlations among trait ($\underset{\sim}{\Phi}_{TT'}$) and/or method factors ($\underset{\sim}{\Phi}_{MM'}$ and $\underset{\sim}{\Phi}_{MT}$), and $\underset{\sim}{\Theta}_{\delta}$ ($\underline{N}$*$\underline{J}$ x $\underline{N}$*$\underline{J}$, diagonal) contains uniquenesses. Widaman (1985) presented a taxonomy of CFA parameterizations of MTMM models that allow (a) no, (b) one, (c) multiple orthogonal, or (d) multiple correlated trait and method factors. An example of the most general of these parameterizations extended to the JPMS is shown in Figure 3 where curved, double-headed arrows indicate correlations among model components. Model components are defined as in Figure 2, and again, measurement errors are not depicted for clarity of presentation.

Both ANOVA and CFA general approaches for the analysis of MTMM matrices but CFA is the more general analytic strategy (see Schmitt & Stults, 1986; Widaman, 1985), and seeks to explain

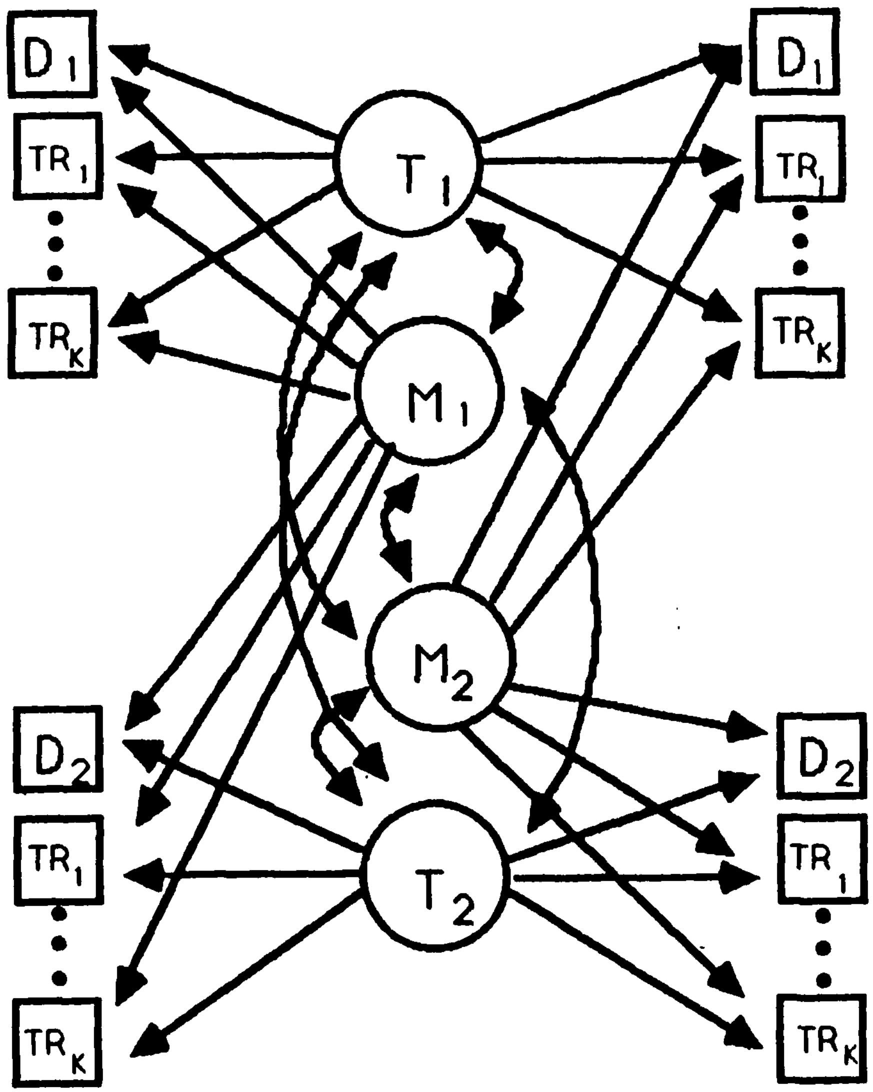


Figure 3: Model II: Multitrait - Multimethod Confirmatory Factor Analysis

observed correlations among dimensional and task level proficiency measures as functions of correlated trait (dimensional) true score factors and global measurement method factors.

Model III: Kenny and Berman's (1980) Model for Ratings

Kenny and Berman (1980) proposed the rating model:

$$X_{ijkm} = \mu_{ijk} + \beta_{ijk} \gamma_{im} + \epsilon_{ijkm} \qquad (4)$$

where model components are expressed in raw score form and X_{ijkm} refers to rater j's ($j \rightarrow J$) rating of the mth ($m \rightarrow M$) ratee on the kth ($k \rightarrow K$) item belonging to the ith ($i \rightarrow N$) performance dimension, μ_{ijk} is the mean of rater j's ratings on all the ith performance dimension's items across all M ratees, β_{ijk} is a weight relating the influence of ratee true performance (γ_{im}) on the ith performance dimension to X_{ijkm}, and ϵ_{ijkm} is presumed to contain measurement error. This model may be represented in standard score form as:

$$Z_{ijkm} = W_T T_{im} + E_{ijm} \qquad (5)$$

Kenny and Berman (1980) assumed that performance true scores were correlated, that is, $r(T_{im}, T_{i'm}) \neq 0$. Rater bias (measurement source effects) was assumed to be indicated by correlated within-source measurement errors. Specifically, E_{ijkm}s were assumed to correlated if $j = j'$ but zero if $j \neq j'$. Thus Kenny and Berman (1980) modeled rater bias as a within-rater source of covariation among observed ratings above and beyond covariation due to effects of correlated ratee true scores. This model is

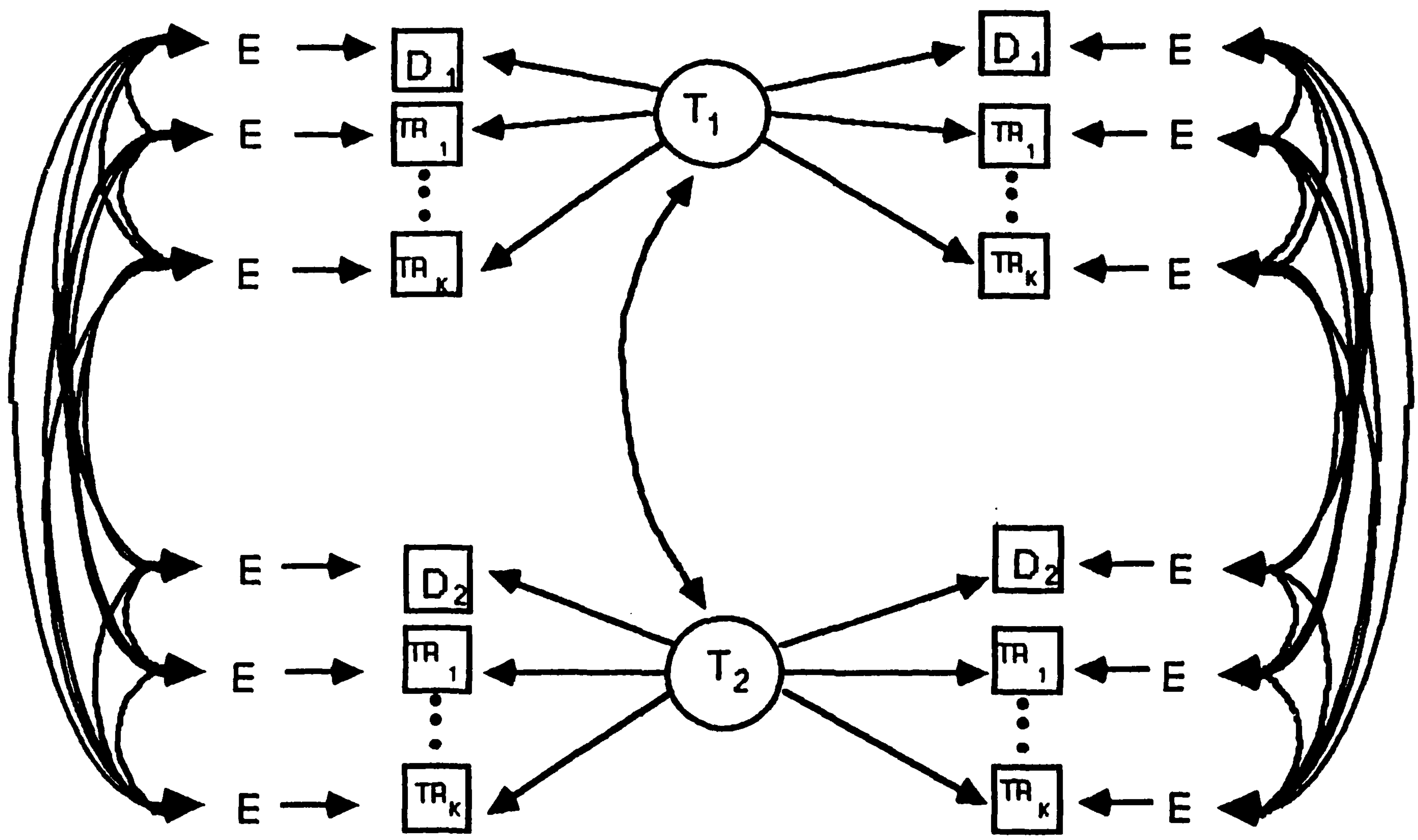


Figure 4: Model III: Kenny & Berman's (1980) Model for Rating

shown in Figure 4.

According to this model, correlations among JPMS task and/or dimensional proficiency ratings within a rating source are due effects of (a) correlated dimensional true scores ($\underline{T}$s) and (b) correlated within-source measurement errors (rater bias, shown as curved, double-headed arrows connecting $\underline{E}$s). Cross-source correlations among rating items are presumed to result from effects of correlated dimensional true scores.

<u>Model IV: General Source Effects</u>

A more parsimonious model of correlated within-source measurement errors in Model III includes a specification of general measurement source effects:

$$\underline{Z}_{ijkm} = \underline{W}_T \underline{T}_{im} + \underline{W}_M \underline{M}_{jm} + \underline{E}_{ijkm} \quad (5)$$

where $\underline{W}$s are weights relating correlated dimensional true score ($\underline{T}_{im}$) and measurement source effects ($\underline{M}_{jm}$) to standardized proficiency measures $\underline{Z}_{ijkm}$. Measurement errors ($\underline{E}_{ijkm}$s) in this model are assumed to be uncorrelated. Model IV is shown in Figure 5. This more parsimonious model seeks to explain correlated within-source measurement errors in Model III in terms of general measurement source effects, often referred to as halo effects (e.g., Prien & Liske, 1962). This model is a special case of the general CFA MTMM Model II (Widaman's, 1985 model 3B') with $\Phi_{TM} = \Phi_{MT} = 0$, $\Phi_{MM'} = I$ in Equation 3.

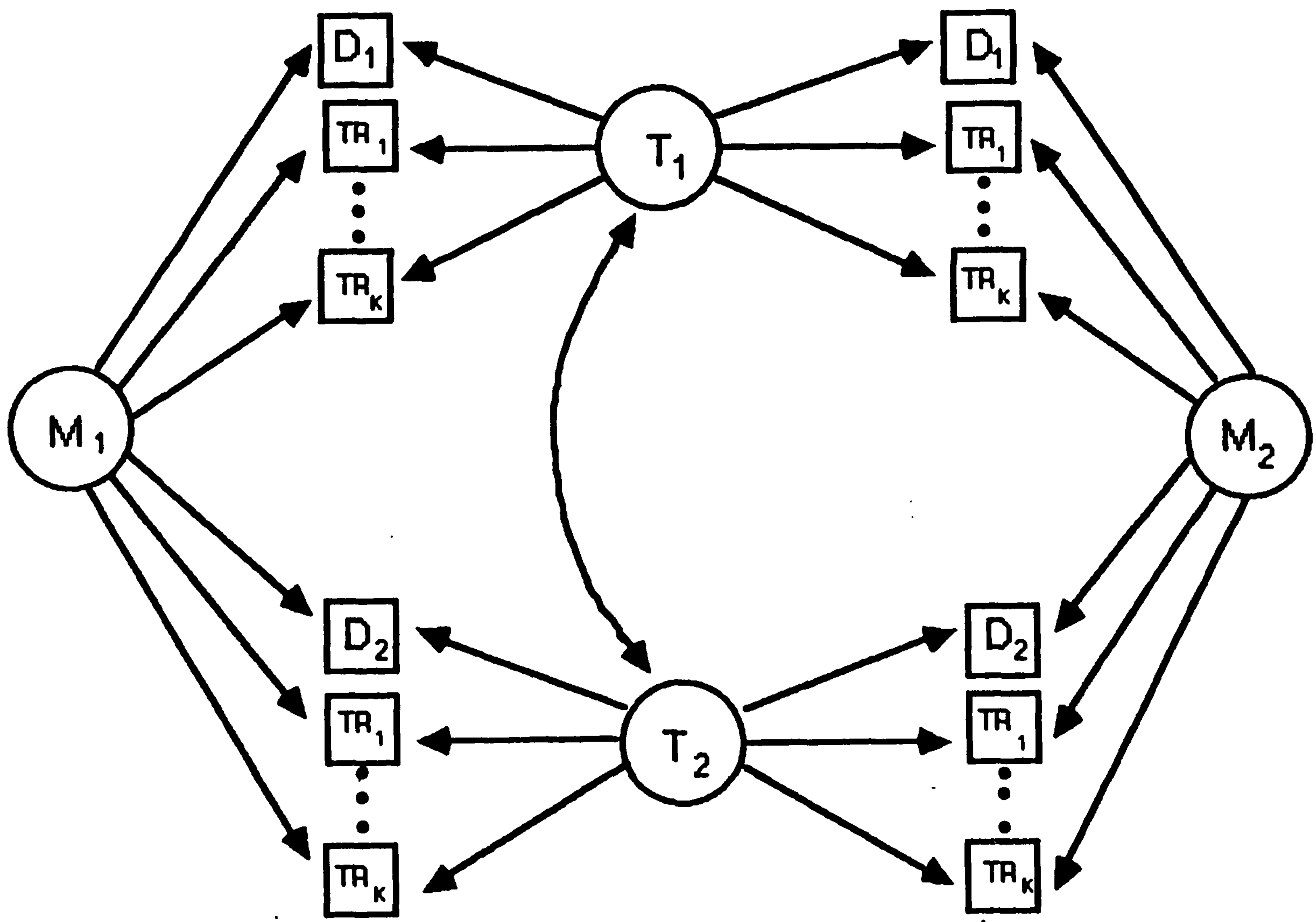


Figure 5: Model IV: General Source Effects

Model V: Guilford's (1954) and King, Hunter and Schmidt's (1980) Psychometric Rating Model

King, Hunter and Schmidt (1980) proposed a modification of Guilford's (1954) psychometric rating model that specified measurement source, as well as overall and dimensional proficiency true score effects:

$$Z_{ijkm} = W_g g_m + W_T T_{im} + W_M M_{jm} + E_{ijkm} \quad (6)$$

Here Z_{ijkm} refers to rater (source) *j*'s rating of ratee *m* on the *k*th item belonging to the *i*th performance dimension, *W*s are weights relating ratings (proficiency measures) to a general proficiency factor (g_m), dimensional true scores (T_{im}) and rater bias (source) effects (M_{jm}). Measurement errors (E_{ijkm}) and all other model components were assumed to be mutually orthogonal. This model is shown in Figure 6.

Figure 6 shows general (*g*) and dimensional proficiency true score (*T*s), and general measurement method effects (*M*s) on global- (*GR*s) dimension- (*D*s) and task-level (*TR*s) proficiency measures. Correlations among within-source measures of proficiency on the same performance dimension are assumed to be due to common effects of (a) dimensional true scores (*T*s) (b) the general proficiency factor (*g*) and, (c) measurement source effects (*M*s). Correlations among within-source measures of proficiency on <u>different</u> performance dimensions are assumed to be due to common effects of (a) the general proficiency factor and (b) measurement source effects. Correlations between cross-source proficiency measures are seen to result from common

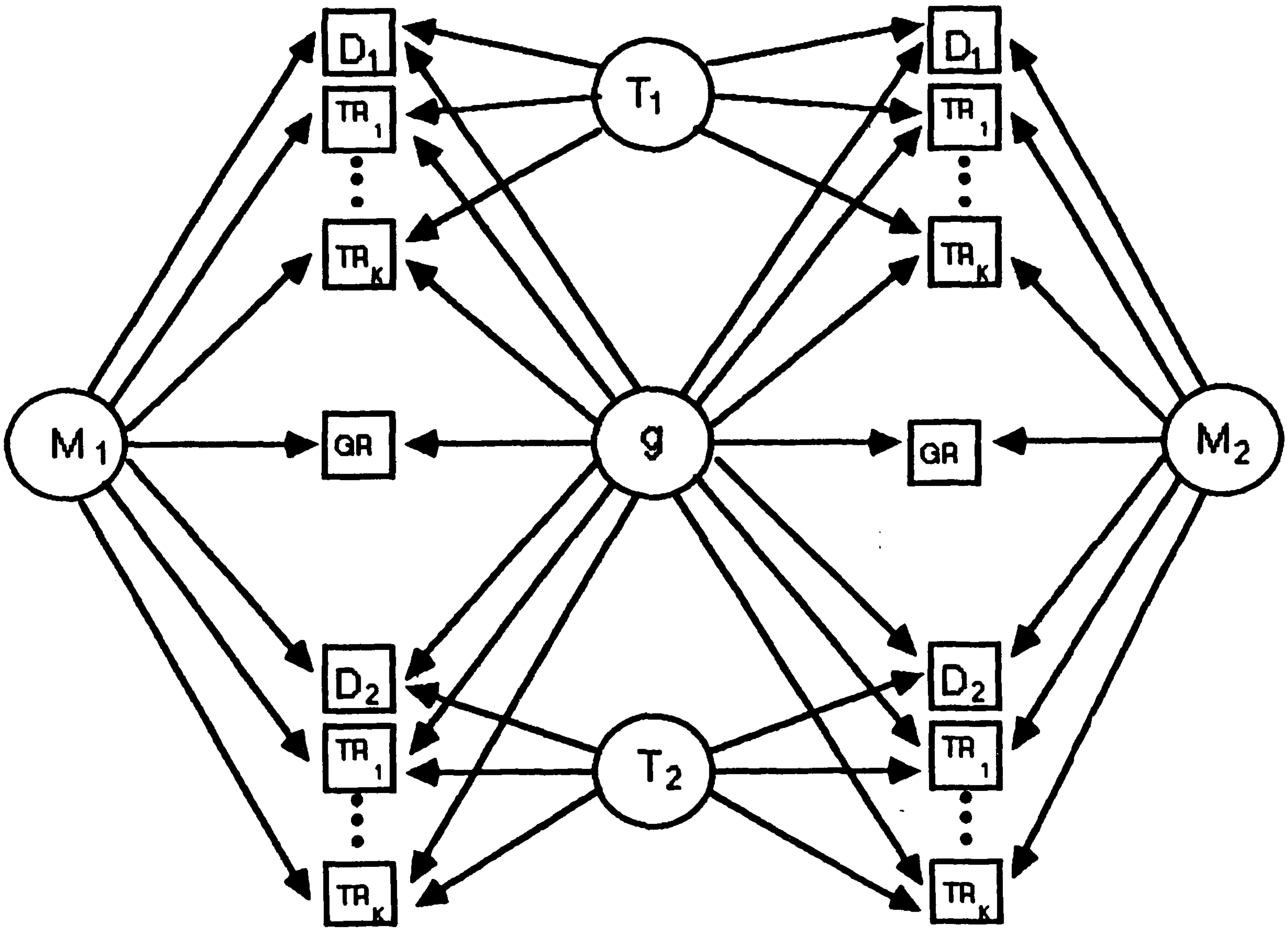


Figure 6: Model V: Guilford's (1954) and King, Hunter, & Schmidt's (1980) Psychometric Model

effects of (a) dimensional true proficiency and/or (b) the general proficiency factor, depending on whether the measures tap the same or different performance dimensions. Global ratings are seen to be functions of general proficiency and measurement source effects, and are expected to be correlated to the extent that global ratings from different rating sources reflect the general proficiency factor.

Model VI: An Extension of the King et al (1980) Model

King et al. (1980) alluded to the possibility of dimension-specific measurement source effects that influence ratings above and beyond general source effects. Dimension-specific measurement source effects have been interpreted as (a) areal rater biases (Wherry & Bartlett, 1982) and (b) aspects of the total criterion space assessed uniquely by particular measurement methods (Kavanagh et al., 1986). Equation 6 may be augmented to include these effects as:

$$\underline{Z}_{ijkm} = \underline{W}_g \underline{g}_m + \underline{W}_T \underline{T}_{im} + \underline{W}_M \underline{M}_{jm} + \underline{W}_{TM} \underline{TM}_{ijm} + \underline{E}_{ijkm} \qquad (7)$$

Equation 7 expresses standardized proficiency measures $\underline{Z}_{ijkm}$ as a function of a general ($\underline{g}_m$) and dimensional true proficiency ($\underline{T}_{im}$) factors; and general ($\underline{M}_{jm}$) and areal ($\underline{TM}_{ijm}$) measurement source effects. As in Model V, model components are assumed to be uncorrelated. This model is shown in Figure 7. Again, measurement errors ($\underline{E}_{ijkm}$) are not shown for the sake of clarity.

Model VII: Wherry's (1952) Theory of Rating

Wherry's (1952) theory of rating was summarized by Wherry and Bartlett's (1982) equation 8e:

$$\underline{Z}_{XR} = \underline{W}_T \underline{Z}_T + \underline{W}_{BRA} \underline{Z}_{BRA} + \underline{W}_{BRO} \underline{Z}_{BRO} + \underline{W}_I \underline{Z}_I + \underline{W}_{EA} \underline{Z}_{EA} + \underline{W}_{EP} \underline{Z}_{EP} + \underline{W}_{ER} \underline{Z}_{ER} \quad (8)$$

Here, $\underline{Z}_{XR}$ is the rating response, $\underline{W}$s are weights relating the rating response to (dimensional) ratee true performance scores ($\underline{Z}_T$), areal (dimensional) rater biases ($\underline{Z}_{BRA}$), overall rater biases ($\underline{Z}_{BRO}$), opportunity bias ($\underline{Z}_I$) and nonsystematic errors in: (a) ratee performance ($\underline{Z}_{EA}$), (b) rater perceptions of ratee behavior ($\underline{Z}_{EP}$), and (c) rater recall of ratee behavior ($\underline{Z}_{ER}$). Since all error components were assumed to be nonsystematic (Wherry & Bartlett, 1982, p. 559), they may be represented as a single aggregate source of nonsystematic error. Also, since the present research was not concerned with opportunity bias effects, Equation 8 may be rewritten:

$$\underline{Z}_{ijkm} = \underline{W}_T \underline{T}_{im} + \underline{W}_M \underline{M}_{jm} + \underline{W}_{TM} \underline{TM}_{ijm} + \underline{E}_{ijkm} \quad (9)$$

Here, $\underline{Z}_{ijkm}$, $\underline{T}_{im}$, $\underline{M}_{jm}$, and $\underline{E}_{ijkm}$ are defined as in equation 8, and $\underline{TM}_{ijm}$s are defined as areal rater biases. Measurement errors $\underline{E}_{ijkm}$ are assumed to be mutually uncorrelated, and uncorrelated with other model components. Dimensional true scores ($\underline{T}_{im}$s) are assumed to be mutually correlated, but uncorrelated with $\underline{M}_{jm}$s and $\underline{TM}_{ijm}$s. Overall biases ($\underline{M}_{jm}$s) are assumed to be correlated (to allow for the possibility of "shared halo" effects) but orthogonal to the $\underline{T}_{im}$s and the $\underline{TM}_{ijm}$s. The $\underline{TM}_{ijm}$s are assumed to be correlated when $\underline{i} = \underline{i}'$ (i.e. shared intersource biases in rating the same trait) but orthogonal when $\underline{i} \neq \underline{i}'$, and orthogonal to the $\underline{T}_{im}$s and $\underline{M}_{jm}$s.

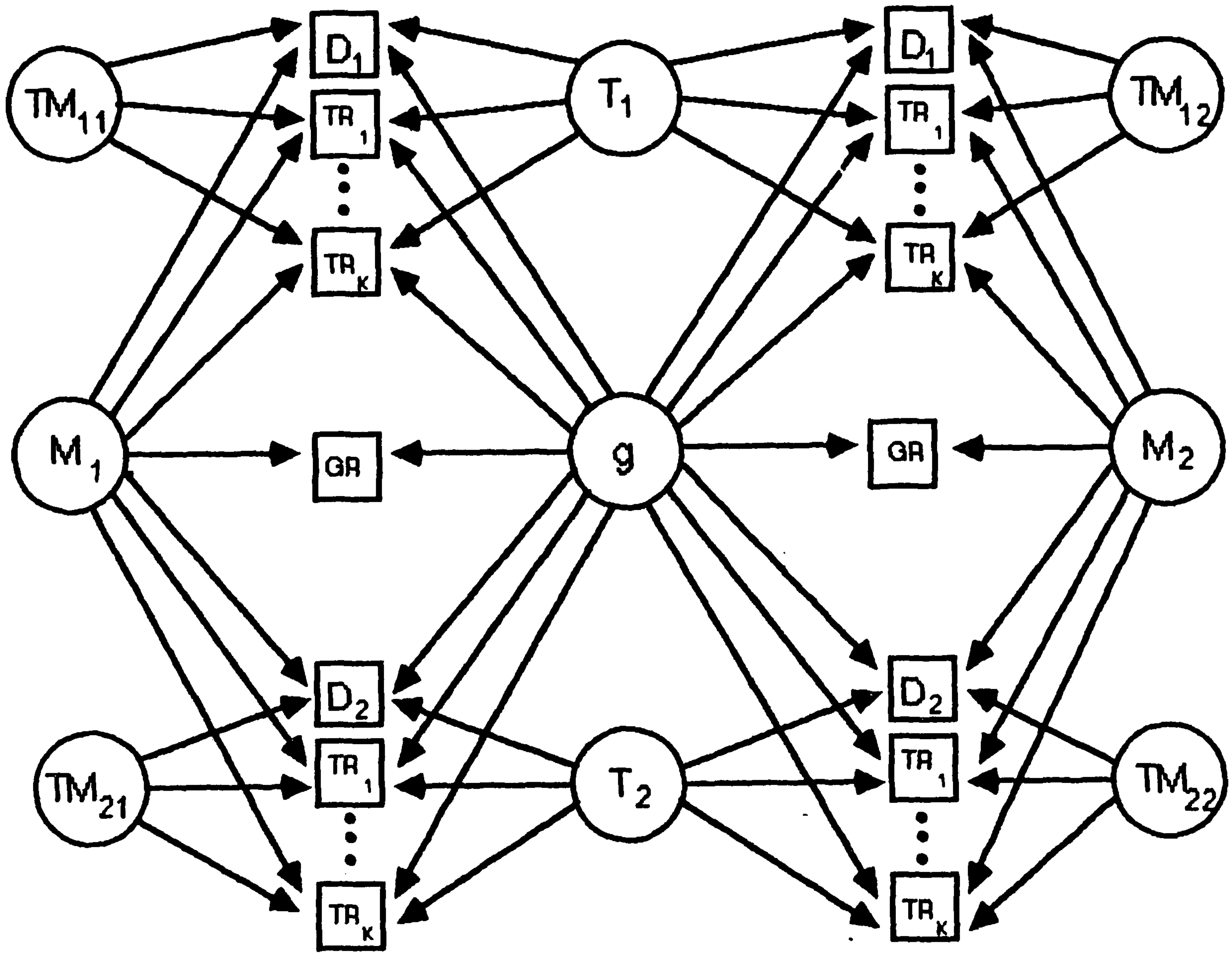


Figure 7: Model VI: An Extension of King et al.'s (1980) Model for Rating

Figure 8 shows several distinctive features of this performance rating model. First, there is no explicit consideration of general proficiency effects. Instead, the weaker assumption of correlated dimensional true score effects is opted for. Second, as in model VI, multiple measurement source effects are considered orthogonal within measurement sources, and measurement source effects are considered to be orthogonal to dimensional true score effects. However, unlike the models above, overall rater biases (measurement source effects) are assumed to be correlated as are same-dimension areal biases. These assumptions lead to the specification of correlated method variance effects (Widaman, 1985), which may be interpreted as cross-source shared implicit theories of the structure of performance (Lance & Lautenschlager, 1987), or common frames of reference among measurement sources (Kavanagh et al., 1986).

Model VIII: A Hierarchical Model

Hulin (1982) suggested that performance, as well as human abilities (Thorndike, 1985) may be organized hierarchically. According to this notion, global proficiency and measurement source effects influence observed measures indirectly through more specific components. That is, a hierarchical model seeks to explain correlations among dimensional true score and areal measurement source effects in terms of their common dependency on more global proficiency and measurement source effects:

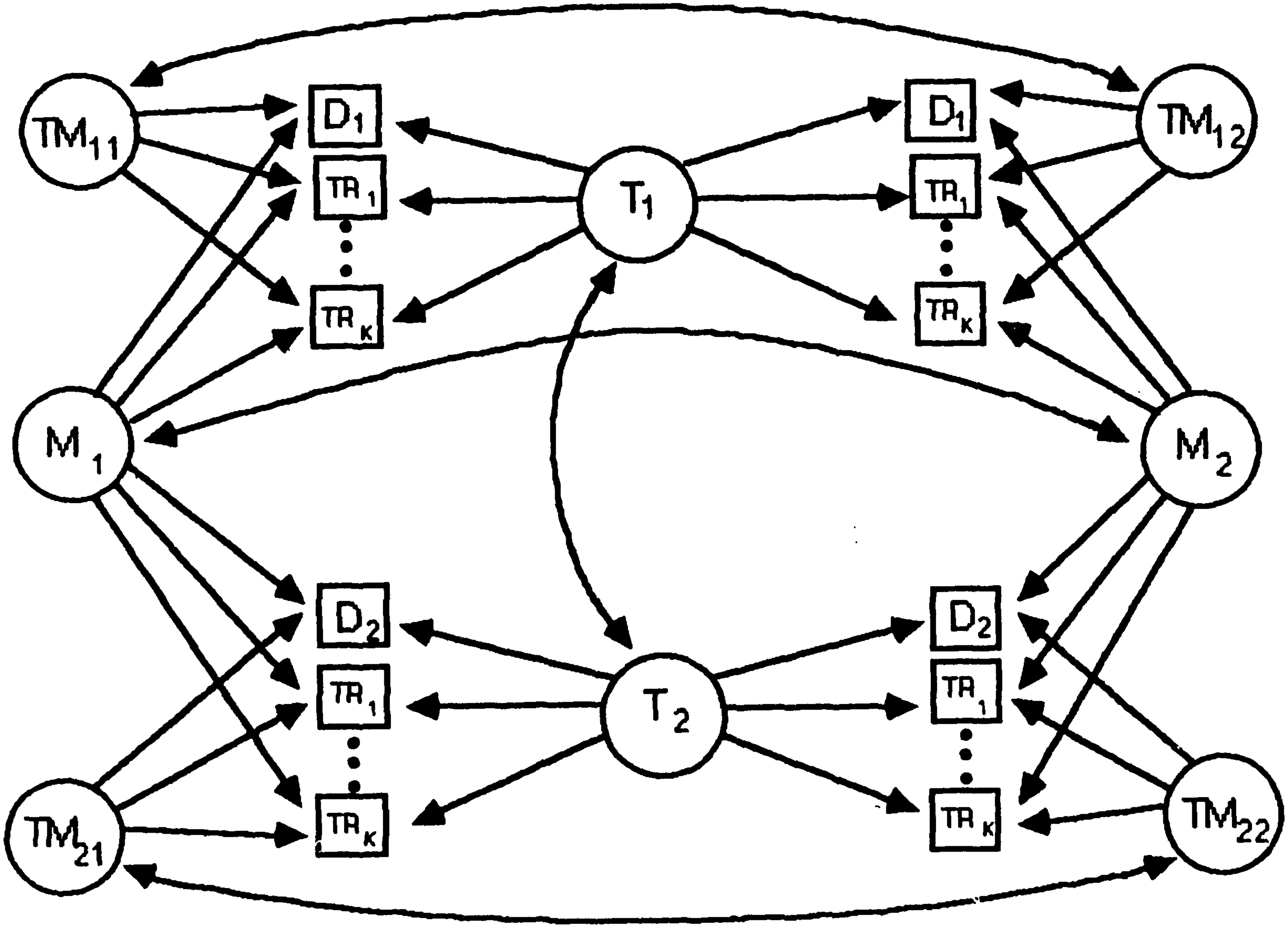


Figure 8: Model VII: Wherry & Bartlett's (1982) Theory of Rating

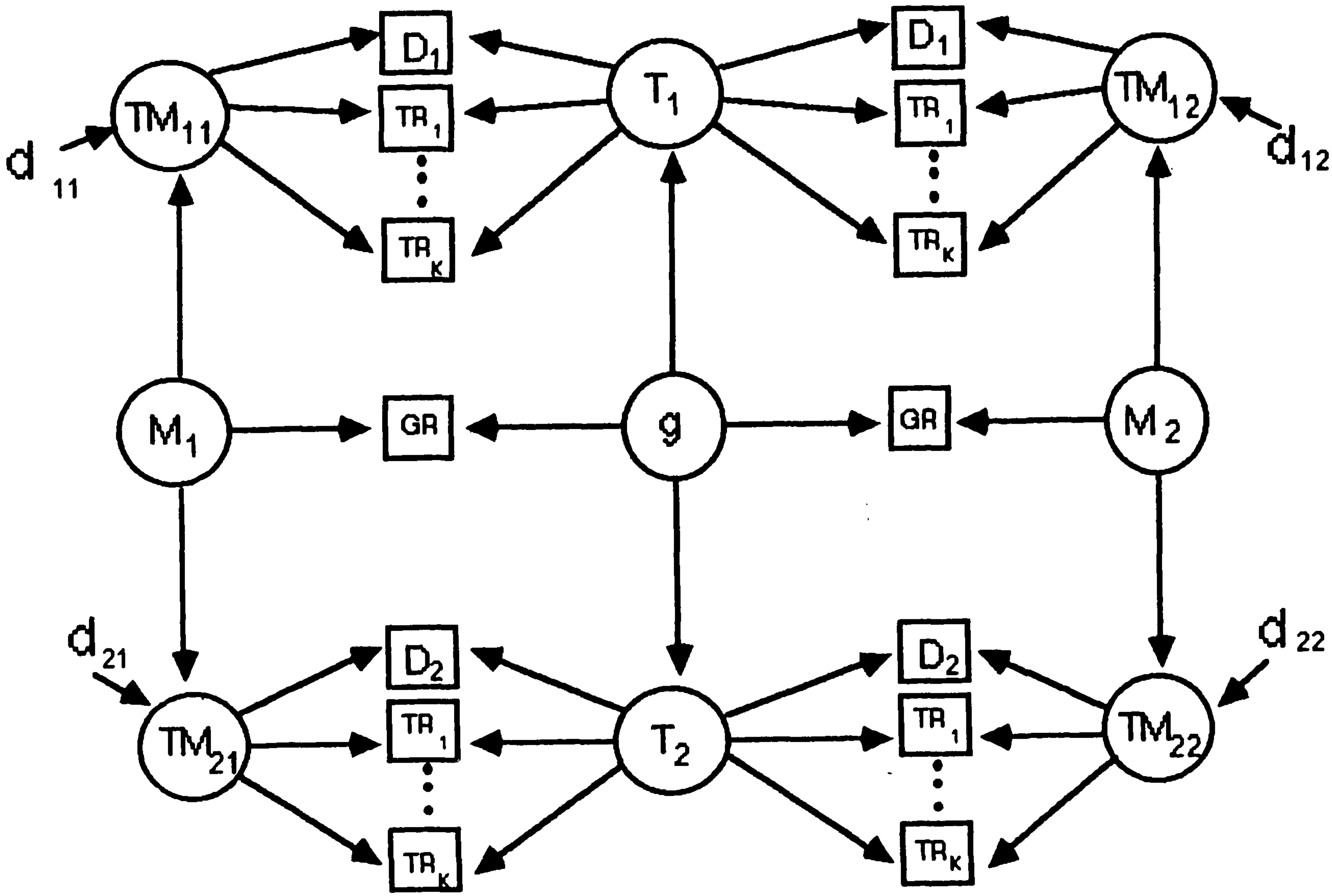


Figure 9: Model VIII: A Hierarchical Model of Performance and Measurement Source Effects

$$\underline{Z}_{ijkm} = \underline{W}_T \underline{T}_{im} + \underline{TM}_{ijm} + \underline{E}_{ijkm} \quad (10a)$$

$$\underline{T}_{im} = \underline{W}_g \underline{g}_m + \underline{d}_T \quad (10b)$$

$$\underline{TM}_{ijm} = \underline{W}_M \underline{M}_{jm} + \underline{d}_{TM} \quad (10c)$$

Here, dimensional true scores ($\underline{T}_{im}$) and areal measurement source effects ($\underline{TM}_{ijm}$) are presumed to affect dimensionally specific proficiency measures ($\underline{Z}_{ijkm}$) directly. The $\underline{T}_{im}$s are assumed to be correlated due to their (imperfect) common dependency on a second-order general proficiency factor ($\underline{g}_m$). Similarly the $\underline{TM}_{ijm}$s are assumed to be correlated within a single measurement source due to their common dependency on more global measurement source effects ($\underline{M}_{jm}$s). The $\underline{d}_T$s and $\underline{d}_{TM}$s are residual variables and contain variance unaccounted for by $\underline{g}_m$ and the $\underline{M}_{jm}$s, respectively. To the extent that measurement sources share common frames of reference, the $\underline{M}_{jm}$s may also be correlated across sources.

Figure 9 illustrates Model VIII, where Ds, TRs and GRs represent, respectively, dimensional, task and global proficiency measures. Dimension-specific proficiency measures are viewed as direct functions of dimensional true scores and areal measurement source effects, and as being affected indirectly by general proficiency and source factors. Overall proficiency is specified explicitly as a general proficiency factor ($\underline{g}$) and, along with general measurement source effects (Ms) are assumed to have direct effects on global proficiency ratings. Correlations among dimensional true scores and areal measurement source effects are assumed to be accounted for by common effects of the general proficiency factor and general measurement source effects, respectively.

Appropriateness of Structural Models for the USAF JPMS

In this section, several criteria are suggested for determining the appropriateness of structural Models I - VIII for describing relations among job proficiency measures in the USAF JPMS. Figure 10 summarizes these criteria and the extent to which Models I - VIII satisfy each.

Criterion 1: Does the model specify dimensional true score effects? Job performance criteria have long been considered to be multidimensional (Brogden & Taylor, 1950; DeNisi et al., 1984; Ghiselli, 1956; James, 1973; Schmidt & Kaplan, 1971). The JPMS reflects this concern with the collection of proficiency measures at various levels of measurement specificity (Hedge & Teachout, 1986). Job performance dimensions are considered to represent conceptually distinct clusters of job tasks or duties. Thus, dimensional proficiency measures should reflect general proficiency on each job performance dimension as well on constituent tasks. All models partition at least some of the total true proficiency space into dimensional true score effects. This desirable feature of each model supports the specification of separate dimensional true score and measurement method effects.

Criterion 2: Does the model specify performance dimension-specific measurement source effects? Wherry (1952) referred to dimension-specific measurement source effects as areal rater biases, or dimensional judgments based on performance non-

Modeling Consideration	MODEL							
	I	II	III	IV	V	VI	VII	VIII
1. Does model specify dimensionaltrue score effects	yes	yes	yes	yes	yes	yes	yes	yes
2. Does model specify dimension-specific measurement sources?	no	no	no	no	no	yes	yes	yes
3. Does model explicitly consider the overall true performance construct?	no	no	no	no	yes	yes	no	yes
4. Does model explicitly consider overall measurement source effects?	yes	yes	no	yes	yes	yes	yes	yes
5. Does model consider a possible hierarchical structure of performance?	no	no	no	no	no	yes	no	yes
6. To what extent is the model derived from theory?	small	small	small	small	mod.	mod.	large	mod.

Figure 10: Considerations for Modeling Performance

relevant factors (Wherry & Bartlett, 1982).

On the other hand, Kavanagh et al (1986) suggested that these effects might also be interpreted as aspects of the total criterion space that are assessed uniquely by particular measurement sources. This implies that certain measurement sources assess certain aspects of performance more effectively than do other sources. Others (e.g., Holzbach, 1978; Kane & Lawler, 1978; Klimoski & London, 1974; Mabe & West, 1982; Steel & Ovalle, 1984; Thornton, 1980; Zammuto, London & Rowland, 1982) have also suggested that raters occupying different organizational positions (e.g., incumbents, peers, supervisors) view ratee performance from different frames of reference. Specification of dimension-specific measurement source effects allows a test of the hypothesis that some measurement methods assess certain aspects of the total criterion space more effectively than do other measurement methods. Thus this criterion would seem to be an important one in determining the appropriateness of structural models for the JPMS.

Only Models VI through VIII specify dimension-specific measurement source effects. MTMM approaches (Models I and II) and Model IV restrict specification of measurement source effects to the global level (however, see Marsh & Hocevar's [1988] recent development), and Kenny and Berman's (1980) model (Model III) addresses measurement method effects only from the standpoint of correlated measurement errors.

Criterion 3: Does the model explicitly consider the overall true performance construct? This is an important criterion for at least two reasons. First, from a pragmatic standpoint, overall performance measures may be necessary for purposes such as test validation, and promotion and retention decisions. Direct measurement of overall performance would obviate concerns over the appropriateness of aggregating specific criterion elements (e.g., Schmidt & Kaplan, 1971). Second, from a theoretical standpoint, the overall true performance construct (Cooper, 1981; Hulin, 1982; Lance & Woehr, 1986) is achieving the status of general ability (Thorndike, 1985) as a broad explanatory construct in the study of work behavior. Thus, a model appropriate to the JPMS should specify general proficiency effects. Models V, VI and VIII explicitly consider a general proficiency factor. Other models (except MTMM-ANOVA, Model I) consider general proficiency only implicitly in terms of correlated dimensional true scores.

Criterion 4: Does the model explicitly consider overall measurement source effects? Global measurement source effects have been interpreted to indicate halo in ratings (Holzbach, 1978; Landy, Vance, Barnes-Farrell & Steele, 1980; Prien & Liske 1962; Wherry, 1952; Wherry & Bartlett, 1982). Others (e.g., Bingham, 1939; Cooper, 1981; Feldman, 1981; Kozlowski, Hirsch & Chao, 1986; Lance & Woehr, 1986), on the other hand, have provided compelling arguments for interpreting at least some portion of general within-source effects as representing a performance-based general impression, or "valid halo." Cross-source differences in ratings may reflect differences in overall frames of reference (Holzbach, 1978; Kane & Lawler, 1978;

Klimoski & London, 1974; Mabe & West, 1982). Thus, it is important that general within-source measurement effects be separable from cross-source and dimensionally specific measurement source effects. All models, except Model III, specify general measurement source effects.

Criterion 5: Does the model consider the possible hierarchical structure of performance? This question concerns (a) whether the model incorporates overall, as well as, dimensional true proficiency and measurement source effects, and (b) whether dimensionally specific true score and measurement method effects are viewed explicitly to be dependent on more global true proficiency and measurement method factors. The USAF JMPS reflects a hierarchical conception of performance. Consequently, a useful structural model of performance would allow an assessment of the extent to which hierarchical relations are actually represented among proficiency measures.

Model VIII meets both requirements for a hierarchical structure of performance, while Model VI meets only the first. Model VI specifies overall proficiency and measurement method effects on proficiency measures that are direct, and separate from dimension-specific method and true score effects. Model VIII, on the other hand, proposes that overall proficiency and measurement method effects on proficiency measures are indirect, and are mediated by dimension-specific effects. In this case, dimension-specific effects are presumed to be subsumed by overall effects.

Criterion 6: To what extent is the model derived from theory? The concern here is not over statistical theory, but the extent to which Models I - VII were derived from substantive performance theory. Data-driven models usually provide closer initial approximations to sample data than do theoretically derived models. Data-driven models, however, may be sample-bound (James, Mulaik & Brett, 1982; Judd, Jessor & Donovan, 1986; Mosier, 1951), while theoretically-based models may lead to more generalizable findings and broader explanantory power. To some extent all of the models described above are derived from theories of performance. The MTMM models (Models I and II) represent well developed, general data analytic strategies in the assessment of convergent and discriminant validity of performance measures (Schmitt & Stults, 1986; Widaman, 1985). However, Models I and II have little to say about the theoretical structure of the performance construct. Models III and IV are based on the classical test theory notion of true and error scores (Lord & Novick, 1968) and are intended to account for sources of systematic errors in performance measures, but the sources of the constant errors remain undefined. Models V and VI are rooted in Guilford's (1954) psychometric model for rating data and Model VIII derives from theoretical notions of ability hierarchies (Hulin, 1982; Thorndike, 1985). Model VII is most firmly rooted in theory, namely Wherry's (1952) theory of the rating process. Model VII was derived from Wherry's (1952) explicit mathematical representation of ratee true performance and rater bias effects in performance ratings.

Conclusions. A review of literature relating to structural models of proficiency true score and measurement method effects on proficiency measures identified several plausible models. However, the focus of the empirical research described below was limited to Models VI, VII, and VIII. Empirical analyses were limited to these models because, among the plausible models considered (a) they were most firmly rooted in performance theory, (b) they made competing specifications of the underlying structure of the total criterion space, (c) they most closely parallel the design of the JPMS, and (d) they contained simpler, perhaps more parsimonious models as special cases.

III. INITIAL EMPIRICAL TESTS OF MODELS VI - VIII: ANALYSES OF JET ENGINE MECHANIC (AFSC 426x2) JPM DATA

Method

Rationale and procedures for collection of the Jet Engine Mechanic JPMS sample data are described in detail elsewhere (Hedge, 1984; Hedge & Teachout, 1986). Measures adapted from the Jet Engine Mechanic (JEM) data base for the present research were self, supervisory and peer ratings each of (a) task-level technical proficiency, (b) dimension-level technical proficiency, (c) global-level task and interpersonal proficiency, and (d) several Air Force-wide factors relating primarily to interpersonal proficiency. WTPT measures of task-level technical proficiency were also used in the present research.

Simplified representations of models derived from King et al. (1980) (Model VI), Wherry and Bartlett (1982) (Model VII), and a third, hierarchical model (Model VIII) of proficiency true score and measurement method effects, are shown in Figures 11 through 13. These representations are simplified in the sense that (a) only two of the four measurement sources studied (self, supervisor, peer and WTPT) are shown (arbitrarily, self and supervisory ratings), (b) only two of the three technical proficiency dimensions studied (Completion of Forms (COF), Removal/Replacement of Engine Components, and Inspect Engine) are shown, and (c) only an illustrative subset of the measures actually used in this study are shown. In Figures 11 through 13 (a) encircled elements represent latent proficiency and/or measurement method factors, (b) elements in boxes are observed

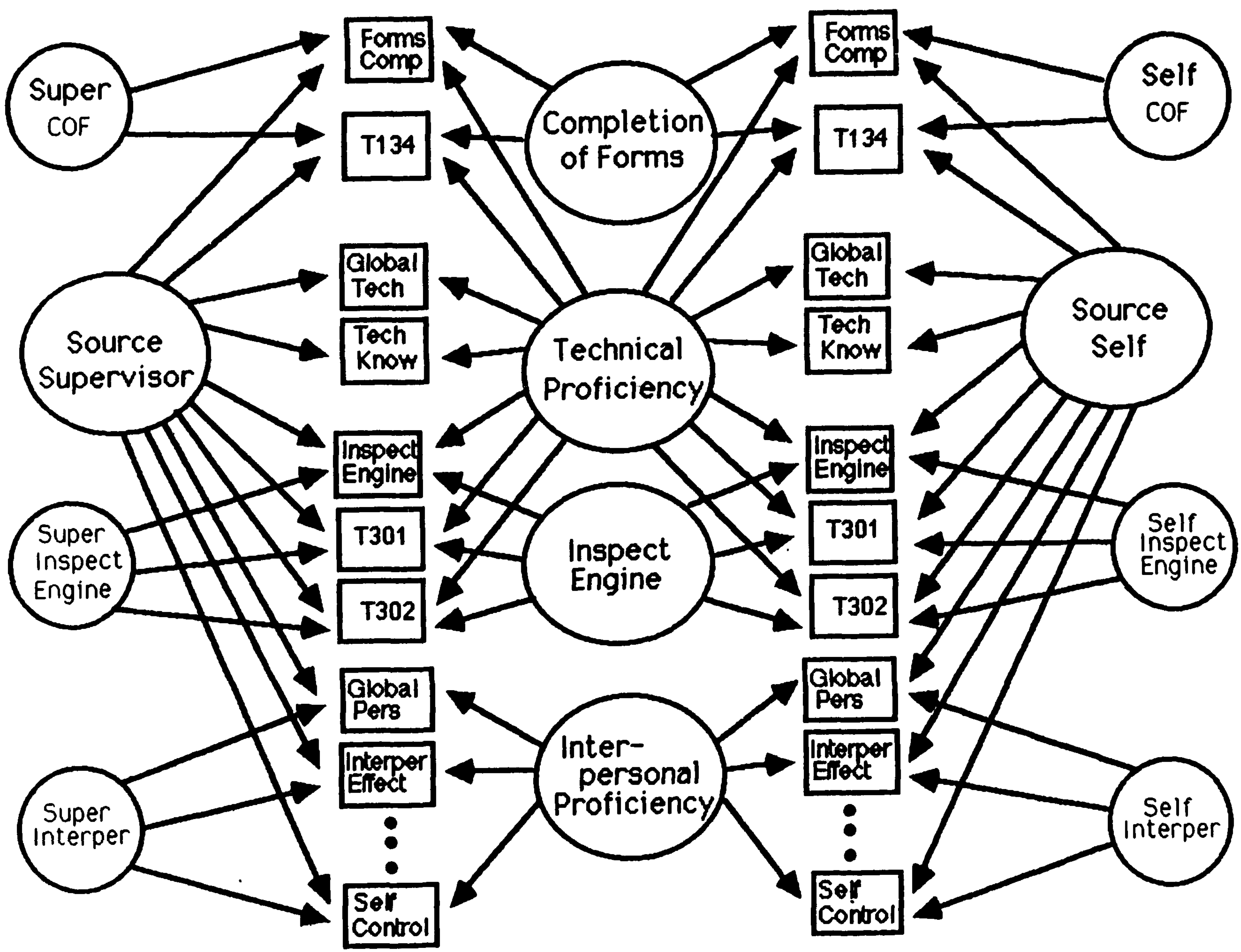


Figure 11: Model Derived From King et al. (1980) for AFSC 426x2

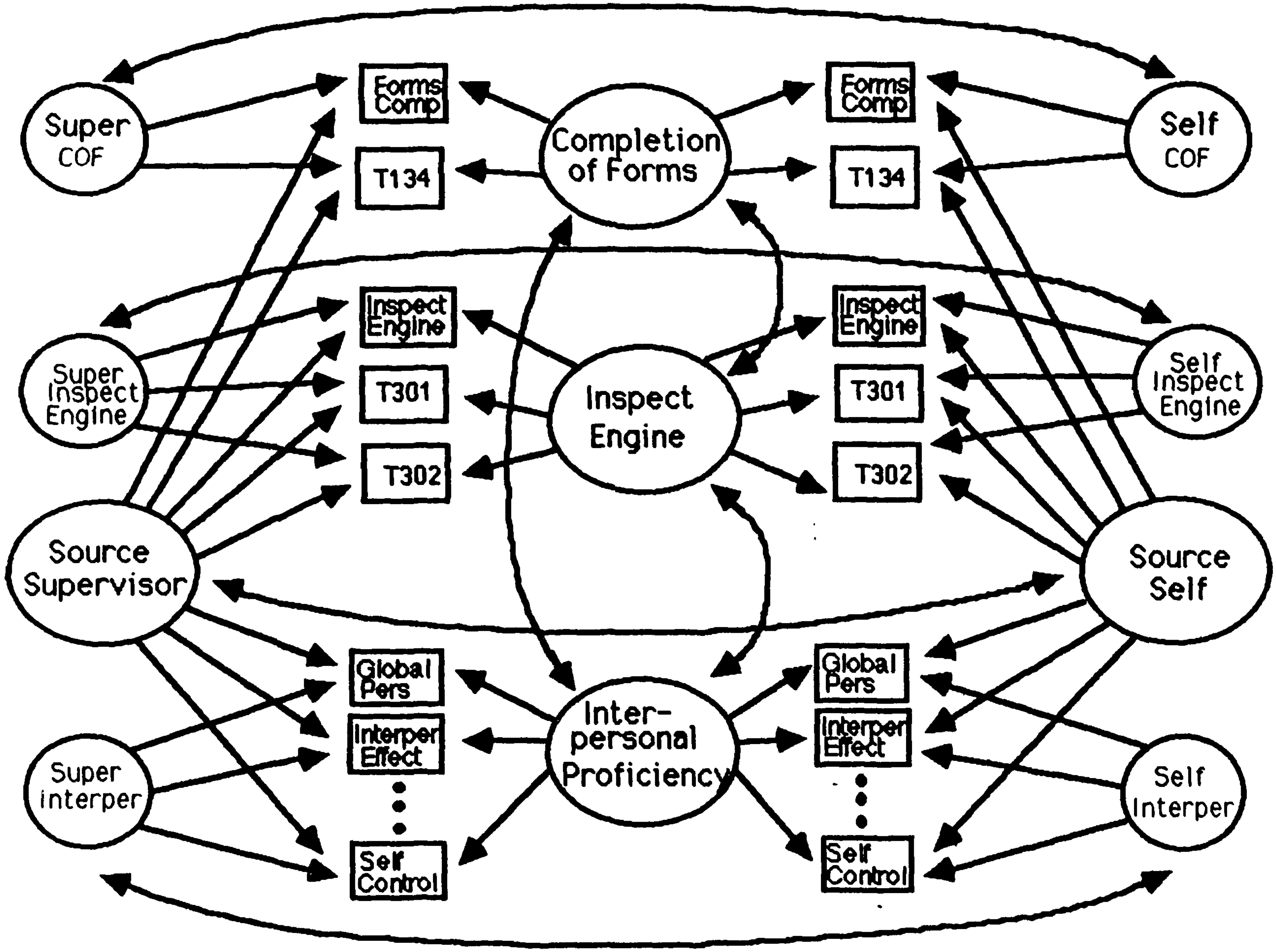


Figure 12: Model Derived From Wherry and Bartlett (1982) for AFSC 426x2

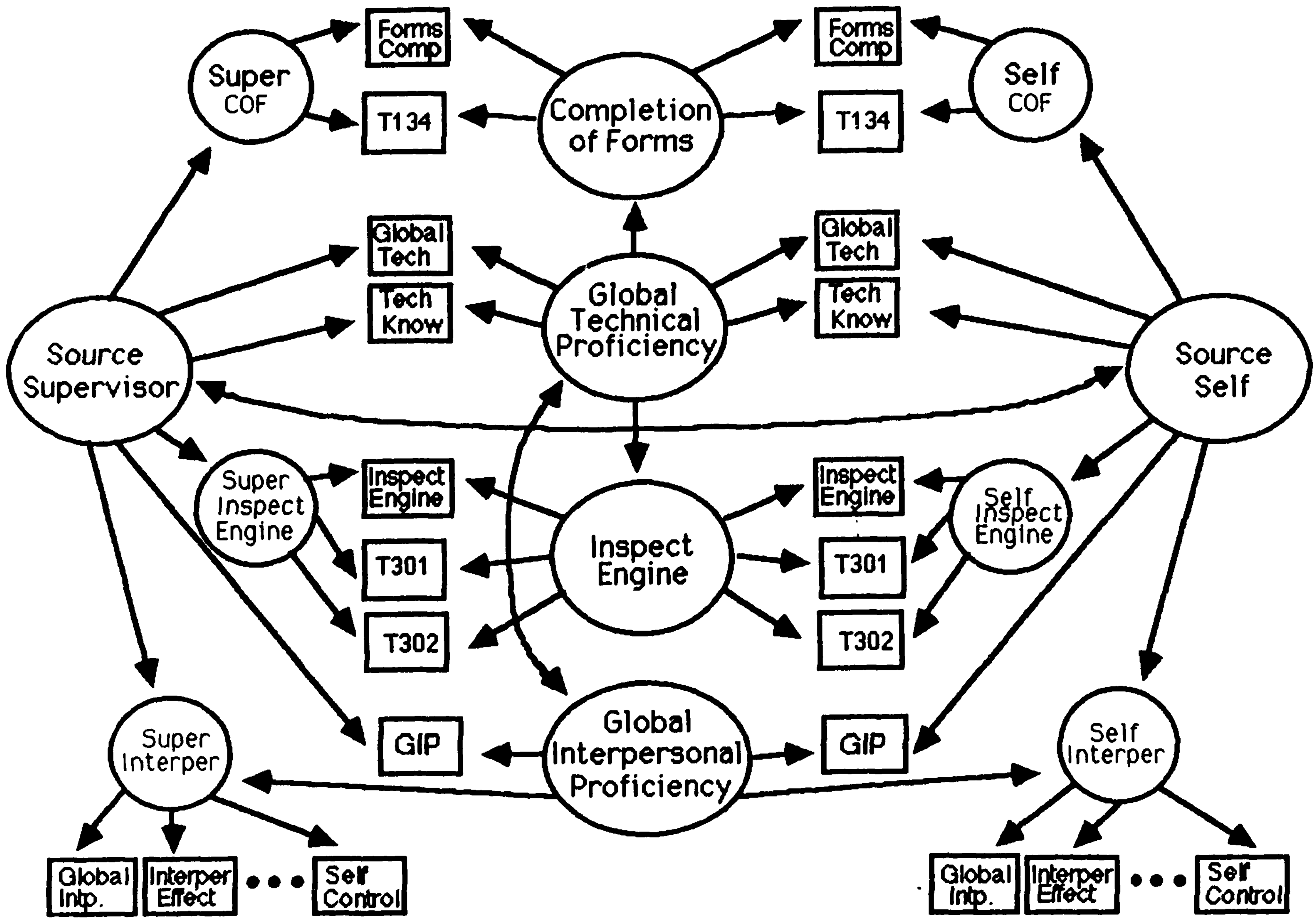


Figure 13: Hierarchical Model for AFSC 426x2

proficiency measures, (c) single-headed arrows indicate hypothesized effects from factors to observed variables and/or other factors, and (d) curved, double-headed arrows indicate correlations.

Measures

Dimensional Proficiency

Technical proficiency. Three JEM technical proficiency dimensions were chosen for study (Completion of Forms [COF], Removal and Replacement of Engine Components [RREC], and Inspect Engine [IE]). These three dimensions were chosen (a) to limit the scope of data analytic requirements and (b) because the JEM JPMS contained common WTPT and rating measures of proficiency on these dimensions. Self, supervisor and peer rating measures included for each dimension were: (a) Completion of Forms - dimensional-level COF rating, and proficiency rating on JEM task 134 (T134), (b) Removal and Replacement of Engine Components - dimensional-level RREC rating and proficiency ratings on JEM tasks T347, T351, T353, T360 and T363, (c) Inspect Engine - dimensional-level IE rating and proficiency ratings on JEM tasks T301 and T302. WTPT measures were: (a) COF - the Hands-on measure of task T134 proficiency, (b) RREC - Hands-on measures of proficiency on tasks T347, T353, T360 and T363, and an Interview measure of task 351 proficiency, and (c) IE - Hands-on measures of proficiency on tasks T301 and T302. Task descriptions are shown in Table 1.

Interpersonal proficiency. Measures of interpersonal proficiency dimensions were self, supervisor and peer Air Force-wide

Table 1

Descriptions of Tasks Included in Jet Engine Mechanic (AFSC 426x2) Analyses

Dimension	Task #	Task Description
Completion of Forms	134	Completes AFTO Forms 349 (Maintenance Data Collection Record), AFTO Forms 350 (Reparable Item Processing Tag), and AFTO Forms 781A (Maintenance Discrepancy and Work Document).
Remove/Replace Engine Components:	347	Installs starters.
	351	Installs engine constant speed drives (CSD).
	353	Installs engine forward top anti-icing ducts.
	360	Installs engine exhaust gas temperature thermocouple harness.
	363	Installs engine ignition exciter boxes.
Inspect Engine:	301	Inspects engine plumbing.
	302	Inspects 3000 series trailers for serviceability.

ratings of (a) Initiative/Effort (INTEF), (b) Knowledge of/Adherance to Regulations/Orders (REG ORD), (c) Integrity (INTEG), (d) Leadership (LSHIP), (e) Military Appearance (MIL AP), (f) Self Development (S DEV), and (g) Self Control (S CNT).

Global Proficiency

Global technical proficiency measures were self, supervisor and peer ratings of (a) Global Technical Proficiency (GL TECH), and (b) Air Force-wide rating of Technical Knowledge/Skill (TECH KN). Global Interpersonal Proficiency (GL INTP) self, supervisor and peer ratings were used to indicate general interpersonal proficiency.

Summary. A total of 71 JEM proficiency measures were included in this research - self, supervisor and peer ratings each on two global proficiency scales, three dimensional proficiency scales, eight task proficiency scales and eight Air Force-wide rating scales. WTPT scores on eight JEM tasks were also included. Figure 14 summarizes the measures adapted from the JEM JPMS data base.

Procedures

A 71 x 71 matrix of correlations among the variables listed above and in Figure 14 was constructed using the SAS CORR procedure (SAS Institute, 1985). Missing data were deleted listwise yielding a matrix based on a working sample of n=175.

The 71 x 71 correlation matrix was then input to the LISREL-VI computer program (Jöreskog & Sörbom, 1984) for Full Information Maximum Likelihood (FIML) confirmatory first- and second-order factor analysis (Long, 1983a, 1983b) estimation of

Measurement Source	Dimensional Level: Technical Proficiency	Dimensional Level: Interpersonal Proficiency	Global Level: Technical Proficiency	Global Level: Interpersonal Proficiency
Self	3 Dimension Level and 8 Task Level Ratings	7 Air Force Wide Ratings	Global Technical Proficiency and Tech Knowledge /Skill Ratings	Global Interpersonal Proficiency Rating
Supervisor	3 Dimension Level and 8 Task Level Ratings	7 Air Force Wide Ratings	Global Technical Proficiency and Tech Knowledge /Skill Ratings	Global Interpersonal Proficiency Rating
Peer	3 Dimension Level and 8 Task Level Ratings	7 Air Force Wide Ratings	Global Technical Proficiency and Tech Knowledge /Skill Ratings	Global Interpersonal Proficiency Rating
WTPT	8 Task Level Proficiency Measures	----	----	----

Figure 14: Technical and Interpersonal Proficiency Measures for AFSC 426x2

parameters in models illustrated in Figures 11, 12 and 13. Results from factor analyses (e.g., nonsignificant factor loadings, goodness-of-fit indices, estimates of correlations among factors) were used to derive respecified (trimmed) models. Parameters in respecified models were then estimated, again using first- and second-order factor analysis application of the LISREL program.

Next, regression estimated factor scores (Mulaik, 1972) were computed for respecified factors using the SAS MATRIX procedure (SAS Institute, 1985b). These were treated as estimates of scores on the underlying factors. Residualized observed variables, with estimates of measurement method effects theoretically removed (Lance & Lautenschlager, 1987) were also computed.

Finally, composites of observed scores, regression estimated factor scores, and observed scores corrected for measurement method effects were correlated with measures of several factors theoretically related to proficiency. These were measues of job experience (Months on Engine [MOE]), aptitude (ASVAB Mechanical Aptitude Index [MECH-AI]), success in training (Technical School grade) and situational constraints on performance (a composite of self-ratings of the availability of tools and equipment, clarity and availability of technical manuals and other written materials). It was hypothesized that (a) composites of true proficiency factor score estimates and (b) composites of observed variables corrected for estimated measurement method effects, would exhibit stronger correlations with variables theoretically related to proficiency, than would composites of observed proficiency scores.

Results

Confirmatory First- and Second-Order Factor Analyses: Initial Specifications

Models derived from King et al. (1980) and Wherry and Bartlett (1982) (Models VI and VII in Figures 11 and 12, respectively) were tested using first-order confirmatory factor analysis (Jöreskog & Sörbom, 1984). The hierarchical model (Model VIII, shown in Figure 13) was tested using confirmatory second-order factor analysis. Many of the same first-order factor loadings were specified by all three models. Table 2 shows the pattern of fixed and free elements in the factor pattern matrix (LISREL's lambda-X matrix) used to test the model derived from King et al. (1980), so as to provide an overall perspective on the elements that were estimated. Factors were hypothesized to be mutually orthogonal in this model. The model derived from Wherry and Bartlett (1982) proposed nonzero correlations (a) among general source factors, (b) among dimensional technical proficiency and interpersonal proficiency factors, (c) between within-source technical proficiency factors and same-source interpersonal proficiency factors, and (d) between same-dimension, within-source technical proficiency factors, but corresponding to different sources. Table 3 shows this pattern.

Due to the many similarities between first-order factor pattern loadings proposed by the hierarchical model and the patterns proposed by the models derived from King et al. (1980)

Table 2

Factor Pattern Matrix: Model Derived From King et al. (1980) for AFSC 426x2

Factor:	Global Tech. Prof.	Global Intp. Prof.	COP	MRBC	IS	Source Self	Source Super.	Source Peer	Source WTPT	Self COP	Self MRBC	Self IS	Self Intp	Supervisor COP	Supervisor MRBC	Supervisor IS	Supervisor Intp	Peer COP	Peer MRBC	Peer IS	Peer Intp	WTPT COP	WTPT MRBC	WTPT IS
Self Ratings:																								
GL TECH	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GL INTP	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COP	XXX	0	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MRBC	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
IS	XXX	0	0	0	XXX	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0
T134	XXX	0	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T301	XXX	0	0	0	XXX	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0
T347	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
T302	XXX	0	0	0	XXX	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0
T351	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
T353	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
T360	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
T363	XXX	0	0	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0
TECH EX	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INTRP	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
RNG ORD	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
INTRG	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
LSHIP	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
MIL AP	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
S DEV	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
S CNT	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0
Super. Ratings:																								
GL TECH	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GL INTP	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COP	XXX	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0
MRBC	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
IS	XXX	0	0	0	XXX	0	XXX	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0
T134	XXX	0	XXX	0	0	0	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0
T301	XXX	0	0	0	XXX	0	XXX	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0
T347	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
T302	XXX	0	0	0	XXX	0	XXX	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0
T351	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
T353	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
T360	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
T363	XXX	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0
TECH EX	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INTRP	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
RNG ORD	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
INTRG	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
LSHIP	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
MIL AP	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
S DEV	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
S CNT	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0
Peer Ratings:																								
GL TECH	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GL INTP	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COP	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0
MRBC	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
IS	XXX	0	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0
T134	XXX	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0	0
T301	XXX	0	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0
T347	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
T302	XXX	0	0	0	XXX	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0
T351	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
T353	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
T360	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
T363	XXX	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0	0	0
TECH EX	XXX	0	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INTRP	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
RNG ORD	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
INTRG	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
LSHIP	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
MIL AP	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
S DEV	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
S CNT	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	0
WTPT:																								
T134	XXX	0	XXX	0	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0
T301	XXX	0	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX
T347	XXX	0	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0
T302	XXX	0	0	0	XXX	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX
T351	XXX	0	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0
T353	XXX	0	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0
T360	XXX	0	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0
T363	XXX	0	0	XXX	0	0	0	0	XXX	0	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0

Note: "XXX" indicates a parameter whose value is free to be estimated, "0" indicates a parameter's value fixed equal to zero.

Table 3

Hypothesized Pattern of Correlations Among Factors: Model Derived From Wherry and Bartlett (1982) for AFSC 426x2

1.	Source-Self	1.0																					
2.	Source-Super.	XXX	1.0																				
3.	Source-Peer	XXX	XXX	1.0																			
4.	Source-WTPT	XXX	XXX	XXX	1.0																		
5.	COF	0	0	0	0	1.0																	
6.	RREC	0	0	0	0	XXX	1.0																
7.	IE	0	0	0	0	XXX	XXX	1.0															
8.	Intp.-Self	0	0	0	0	XXX	XXX	XXX	1.0														
9.	Intp. Super.	0	0	0	0	XXX	XXX	XXX	XXX	1.0													
10.	Intp.-Peer	0	0	0	0	XXX	XXX	XXX	XXX	XXX	1.0												
11.	Self-COF	0	0	0	0	0	0	0	XXX	0	0	1.0											
12.	Self-RREC	0	0	0	0	0	0	0	XXX	0	0	0	1.0										
13.	Self-IE	0	0	0	0	0	0	0	XXX	0	0	0	0	1.0									
14.	Super.-COF	0	0	0	0	0	0	0	0	XXX	0	XXX	0	0	1.0								
15.	Super.-RREC	0	0	0	0	0	0	0	0	XXX	0	0	XXX	0	0	1.0							
16.	Super.-IE	0	0	0	0	0	0	0	0	XXX	0	0	0	XXX	0	0	1.0						
17.	Peer-COF	0	0	0	0	0	0	0	0	0	XXX	XXX	0	0	XXX	0	0	1.0					
18.	Peer-RREC	0	0	0	0	0	0	0	0	0	XXX	0	XXX	0	0	XXX	0	0	1.0				
19.	Peer-IE	0	0	0	0	0	0	0	0	0	XXX	0	0	XXX	0	0	XXX	0	0	1.0			
20.	WTPT-COF	0	0	0	0	0	0	0	0	0	0	XXX	0	0	XXX	0	0	XXX	0	0	1.0		
21.	WTPT-RREC	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	XXX	0	0	XXX	0	0	1.0	
22.	WTPT-IE	0	0	0	0	0	0	0	0	0	0	0	0	XXX	0	0	XXX	0	0	XXX	0	0	1.0

Note: "XXX" indicates an element whose value is free to be estimated; "1.0" and "0" indicate respectively, elements fixed to these pre-specified values.

and Wherry and Bartlett (1982), it is not shown. However, (a) patterns of manifest variables (indicants') loadings on second-order factors, and (b) patterns of first-order factors' loadings on the second-order factors are shown in Tables 4 and 5, respectively. Nonzero second-order factor correlations were proposed (a) between Global Technical and Interpersonal Proficiency factors, and (b) among global source factors.

LISREL's automatic starting values option (which produces two-stage least squares [2SLS] initial parameter estimates) was chosen, since there were no prior estimates that might have been thought to be closer to full information maximum likelihood (FIML) estimates than 2SLS estimates. A maximum of 60 minutes CPU time was allocated for each analysis.

The FIML solution for the hierarchical model converged within four iterations, but contained many estimates that lay outside admissible parameter space (e.g., negative error variances, standardized parameter estimates larger in absolute value than 1.00). The initial 2SLS solution also contained many "improper" estimates.

"Reasonable" 2SLS estimates were obtained both for models derived from King et al. (1980) and Wherry and Bartlett (1982). However, in neither case did the FIML solution converge within 67 iterations or within the 60 CPU minutes allocated for the problems. Initial 2SLS estimates, however, pointed to several sources of model confirmation, and to some sources of potential model misspecification.

Variables' loadings on the Global Technical Proficiency and

Table 4

Second-Order Factor Pattern Matrix: Hierarchical Model for AFSC 426x2

Factors:	Global Tech. Prof.	Global Intp. Prof.	Global Source Seelf	Global Source Super.	Global Source Peer	Global Source WTPT
<u>Indicants</u>:						
Self-GL TECH	XXX	0	XXX	0	0	0
Self-GL INTP	0	XXX	XXX	0	0	0
Self-TECH KN	XXX	0	XXX	0	0	0
Super.-GL TECH	XXX	0	0	XXX	0	0
Super.-GL INTP	0	XXX	0	XXX	0	0
Super.-GL-TECH KN	XXX	0	0	XXX	0	0
Peer-GL TECH	XXX	0	0	0	XXX	0
Peer-GL INTP	0	XXX	0	0	XXX	0
Peer-TECH KN	XXX	0	0	0	XXX	0

Note: "XXX" indicates an element whose value is free to be estimated;
"0" indicates an element whose value is fixed equal to zero.

Table 5

Pattern of First Order Factors' Loadings on Second-Order Factors: Hierarchical Model for AFSC 426x2

Second-Order Factors:	Global Tech. Prof.	Global Intp. Prof.	Global Source Self	Global Source Super.	Global Source Peer	Global Source WTPT
First-Order Factors:						
Intp.-Self	0	XXX	XXX	0	0	0
Intp.-Super	0	XXX	0	XXX	0	0
Intp.-Peer	0	XXX	0	0	XXX	0
Self-COF	XXX	0	XXX	0	0	0
Self-RREC	XXX	0	XXX	0	0	0
Self-IE	XXX	0	XXX	0	0	0
Super.-COF	XXX	0	0	XXX	0	0
Super.-RREC	XXX	0	0	XXX	0	0
Super.-IE	XXX	0	0	XXX	0	0
Peer-COF	XXX	0	0	0	XXX	0
Peer-RREC	XXX	0	0	0	XXX	0
Peer-IE	XXX	0	0	0	XXX	0
WTPT-COF	XXX	0	0	0	0	1.0
WTPT-RREC	XXX	0	0	0	0	XXX
WTPT-IE	XXX	0	0	0	0	XXX

Note: "XXX" indicates an element whose value is free to be estimated; "1.0" and "0" indicate respectively, elements fixed to these pre-specified values.

Table 6

Global Technical and Interpersonal Proficiency 2SLS Factor Loadings: Model Derived From King et al. (1980) for AFSC 426x2

Factor:	Global Tech. Prof.				Global Intp. Prof.			
Source:	Self	Super	Peer	WTPT	Self	Super	Peer	WTPT
Variable:								
GL TECH	.188	.268	.369	--	0	0	0	--
GL INTP	0	0	0	--	.051	.109	.079	--
COF	-.025	.083	-.041	--	0	0	0	--
RREC	.428	.579	.454	--	0	0	0	--
IE	.030	.258	-.020	--	0	0	0	--
T134	-.094	-.084	.016	.000	0	0	0	0
T301	-.047	.126	-.009	.037	0	0	0	0
T347	-.316	-.156	-.094	.134	0	0	0	0
T302	-.203	-.241	.145	-.131	0	0	0	0
T353	-.088	-.180	-.111	-.406	0	0	0	0
T351	.110	-.182	-.052	-.056	0	0	0	0
T360	-.003	-.083	-.227	.321	0	0	0	0
T363	-.036	.005	.161	-.401	0	0	0	0
TECH KN	.291	.265	.291	--	0	0	0	--
INT EF	0	0	0	--	.102	.031	-.035	--
REG ORD	0	0	0	--	.149	-.000	.031	--
INTEG	0	0	0	--	-.096	-.056	-.085	--
LSHIP	0	0	0	--	-.124	-.049	.081	--
MIL AP	0	0	0	--	-.028	.174	.040	--
S DIS	0	0	0	--	.073	.170	.061	--
S CNT	0	0	0	--	.121	-.088	.017	--

Global Interpersonal Proficiency factors hypothesized by the King et al. (1980) model tended to be small and inconsistent. These are shown in Table 6. Columns for the two proficiency factors in Table 6 are "collapsed" for interpretive ease. That is, these loadings actually formed only one column for each factor in the original factor pattern matrix. These results lent little support for hypotheses of direct global technical and interpersonal proficiency effects on proficiency measures. However, they are consistent with prior research that has found only marginal convergence between measurement sources in evaluating performance (e.g., Holzbach, 1978; Prien & Liske, 1962).

All other hypothesized patterns of factor loadings were the same for the King et al. (1980) and Wherry and Bartlett (1982) models, and 2SLS factor pattern estimates were similar in the two solutions. Table 7 shows factor loadings (from the solution for Wherry and Bartlett [1982] model) for the inter-source dimensional technical proficiency factors (e.g., "Completion of Forms", and "Inspect Engine" in Figure 12). Columns from the original factor pattern matrix are again "collapsed" for interpretive ease. Most factor loadings were small and nonsignificant. These results provide little support for the inter-source dimensional technical proficiency true score factors hypothesized by the hierarchical model, and the models derived from King et al. (1980) and Wherry and Bartlett (1982).

There was, however, reasonably strong support for global measurement source effects. 2SLS estimates of variables' loadings on these factors (again from the solution for the model

Table 7

Inter-Source Dimensional Technical Proficiency 2SLS Factor Loadings: Model Derived From Wherry and Bartlett (1982) for AFSC 426X2

Factor:	Completion of Forms				Remove/Replace Eng. Comp.				Inspect Engine			
Source:	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT
Variable:												
GL TECH	0	0	0	---	0	0	0	---	0	0	0	---
GL INTP	0	0	0	---	0	0	0	---	0	0	0	---
COF	.067	.130	.165	---	0	0	0	---	0	0	0	---
RREC	0	0	0	---	.164	.239	.280	---	0	0	0	---
IE	0	0	0	---	0	0	0	---	.289	.196	.356	---
T134	.156	.029	.119	.023	0	0	0	0	0	0	0	0
T301	0	0	0	0	0	0	0	0	.105	.056	.152	.068
T347	0	0	0	0	-.008	.079	.170	.055	0	0	0	0
T302	0	0	0	0	0	0	0	0	-.132	.076	-.250	.185
T353	0	0	0	0	.000	.052	-.008	.333	0	0	0	0
T351	0	0	0	0	.230	.079	.030	-.157	0	0	0	0
T360	0	0	0	0	.073	-.045	.009	.207	0	0	0	0
T363	0	0	0	0	-.101	-.068	-.127	.046	0	0	0	0
TECH KN	0	0	0	---	0	0	0	---	0	0	0	---
INTEF	0	0	0	---	0	0	0	---	0	0	0	---
REG ORD	0	0	0	---	0	0	0	---	0	0	0	---
INTEG	0	0	0	---	0	0	0	---	0	0	0	---
LSHIP	0	0	0	---	0	0	0	---	0	0	0	---
MILAP	0	0	0	---	0	0	0	---	0	0	0	---
S DIS	0	0	0	---	0	0	0	---	0	0	0	---
S CNT	0	0	0	---	0	0	0	---	0	0	0	---

Table 8

Global Source 2SLS Factor Loadings: Model Derived From Wherry and Bartlett (1982) for AFSC 426x2

Source Factor:	Self	Supervisor	Peer	WTPT
Variable:				
GL TECH	.845	.876	.913	--
GL INTP	.346	.585	.422	--
COF	.473	.617	.568	--
RREC	.698	.797	.711	--
IE	.544	.715	.703	--
T134	.464	.487	.613	.161
T301	.566	.688	.640	.091
T347	.426	.698	.566	.200
T302	.283	.518	.550	-.119
T353	.478	.705	.523	.181
T351	.545	.627	.475	.437
T360	.513	.691	.529	-.196
T363	.426	.599	.527	.553
TECH KN	.986	.913	.919	--
INT EF	.305	.616	.237	--
REG ORD	.168	.550	.133	--
INTEG	.074	.551	.257	--
LSHIP	.615	.763	.449	--
MIL AP	.148	.405	.110	--
S DIS	.354	.533	.128	--
S CNT	.293	.566	.344	--

derived from Wherry & Bartlett [1982]) are shown in Table 8. Here, columns correspond exactly to columns in the original factor pattern matrix but rows of fixed-zero elements are truncated to facilitate interpretation. Variables' loadings on global measurement source factors tended to be quite large for self, supervisor, and peer rating sources, but not for the WTPT measurement source. These results strongly support global self, supervisor and peer measurement source effects on proficiency measures, but support for a global WTPT source factor was weak.

There was also strong support for within-source dimensional proficiency effects, for example, Self-COF, Self-Inspect Engine and Self-Interper. factors in Figure 12. Table 9 shows estimates for variables' loadings on these factors, again from the solution for the model derived from Wherry and Bartlett (1982). Here again, columns correspond exactly to columns in the original factor pattern matrix but rows of zero-restricted elements have been omitted for interpretive ease. Factor loadings were generally high and in the expected direction.

Finally, there was mixed support for the pattern of factor correlations hypothesized by the model derived from Wherry and Bartlett (1982). 2SLS estimates of factor correlations (elements in LISREL's phi matrix) are shown in Table 10. Rating source factors were positively intercorrelated but essentially orthogonal to the WTPT source factor. Inter-source technical proficiency factors (COF, RREC and IE) were essentially orthogonal, and uncorrelated with interpersonal proficiency factors. Interpersonal proficiency factors were positively correlated and were

Table 9

Within-Source Dimensional Proficiency 2SLS Factor Loadings: Model Derived From Wherry and Bartlett (1982) for AFSC 426x2

Factor:	Completion of Forms				Remove/Replace Eng. Comp.				Inspect Engine				Interpers. Prof.		
Source:	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super	Pee
Variable:															
GL TECH	0	0	0	---	0	0	0	---	0	0	0	---	0	0	0
GL INTP	0	0	0	---	0	0	0	---	0	0	0	---	.565	.428	.53
COF	.809	.700	.812	---	0	0	0	---	0	0	0	---	0	0	0
RREC	0	0	0	---	.484	.269	.357	---	0	0	0	---	0	0	0
IE	0	0	0	---	0	0	0	---	0	0	0	---	0	0	0
T134	.792	.805	.752	1.00*	0	0	0	0	.562	.499	.512	---	0	0	0
T301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T347	0	0	0	0	.677	.461	.605	.683	.641	.573	.643	.715	0	0	0
T302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T353	0	0	0	0	.764	.509	.754	.528	.647	.600	.517	.587	0	0	0
T351	0	0	0	0	.567	.412	.651	.083	0	0	0	0	0	0	0
T360	0	0	0	0	.691	.489	.676	.663	0	0	0	0	0	0	0
T363	0	0	0	0	.670	.475	.616	.194	0	0	0	0	0	0	0
TECH KN	0	0	0	---	0	0	0	0	0	0	0	---	0	0	0
INTEF	0	0	0	---	0	0	0	0	0	0	0	---	.651	.439	.71
REG ORD	0	0	0	---	0	0	0	0	0	0	0	---	.768	.472	.76
INTEG	0	0	0	---	0	0	0	0	0	0	0	---	.765	.379	.67
LSHIP	0	0	0	---	0	0	0	0	0	0	0	---	.390	.292	.57
MILAP	0	0	0	---	0	0	0	0	0	0	0	---	.623	.499	.76
S DIS	0	0	0	---	0	0	0	0	0	0	0	---	.594	.493	.66
S CNT	0	0	0	---	0	0	0	0	0	0	0	---	.578	.409	.55

Table 10

2SLS Estimates of a priori-Specified Factor Correlations: Model Derived From Wherry and Bartlett (1982) for AFSC 426x2

1. Source-Self	1.00																					
2. Source-Super.	.40	1.00																				
3. Source-Peer	.37	.55	1.00																			
4. Source-WTPT	.12	.03	-.01	1.00																		
5. COF	0	0	0	0	1.00																	
6. RREC	0	0	0	0	.03	1.00																
7. IE	0	0	0	0	.04	.09	1.00															
8. Intp.-Self	0	0	0	0	.04	.02	.06	1.00														
9. Intp.-Super.	0	0	0	0	.07	.10	.10	.36	1.00													
10. Intp.-Peer	0	0	-0	0	.01	.05	.04	.26	-.43	1.00												
11. Self-COF	0	0	0	0	0	0	0	.23	0	0	1.00											
12. Self-RREC	0	0	0	0	0	0	0	.25	0	0	0	1.00										
13. Self-IE	0	0	0	0	0	0	0	.35	0	0	0	0	1.00									
14. Super.-COF	0	0	0	0	0	0	0	0	.31	0	.21	0	0	1.00								
15. Super.-RREC	0	0	0	0	0	0	0	0	.13	0	0	.30	0	0	1.00							
16. Super.-IE	0	0	0	0	0	0	0	0	.44	0	0	0	.16	0	0	1.00						
17. Peer-COF	0	0	0	0	0	0	0	0	0	.30	.22	0	0	.23	0	0	1.00					
18. Peer-RREC	0	0	0	0	0	0	0	0	0	.31	0	.07	0	0	.15	0	0	1.00				
19. Peer-IE	0	0	0	0	0	0	0	0	0	.32	0	0	.05	0	0	.08	0	0	1.00			
20. WTPT-COF	0	0	0	0	0	0	0	0	0	0	.17	0	0	.02	0	0	.05	0	0	1.00		
21. WTPT-RREC	0	0	0	0	0	0	0	0	0	0	0	.26	0	0	.31	0	0	.18	0	0	1.00	
22. WTPT-IE	0	0	0	0	0	0	0	0	0	0	0	0	.08	0	0	.09	0	0	.12	0	0	1.00

also positively correlated with same-source technical proficiency factors. Finally, correlations between same-dimension source-specific technical proficiency factors (analogous to correlations among trait-method interaction effects in MTMM analysis) tended to be small to negligble. These findings, along with findings that indicated lack of support for inter-source dimensional technical proficiency factors, suggested little convergence among measurement sources in assessing common aspects of the technical proficiency criterion space.

Model Respecifications

Results from initial model tests suggested several misspecifications in the models derived from King et al. (1980) and Wherry and Bartlett (1982), and indirectly, the hierarchical model. Among these were: (a) hypothesized direct global technical and interpersonal proficiency effects on proficiency measures were disconfirmed (see Table 6), (b) hypothesized inter-source technical proficiency effects on proficiency measures were disconfirmed (see Table 7), (c) factors were neither orthogonal, as hypothesized by the model derived from King et al. (1980), nor did factor correlations conform precisely to the pattern hypothesized by the model derived from Wherry and Bartlett (1982).

These findings guided respecification of two models shown in Figures 15 and 16. Figure 15 represents many aspects of the models derived from King et al. (1980) and Wherry and Bartlett (1982), but with (a) global technical and interpersonal proficiency factors absent, (b) inter-source technical proficiency factors

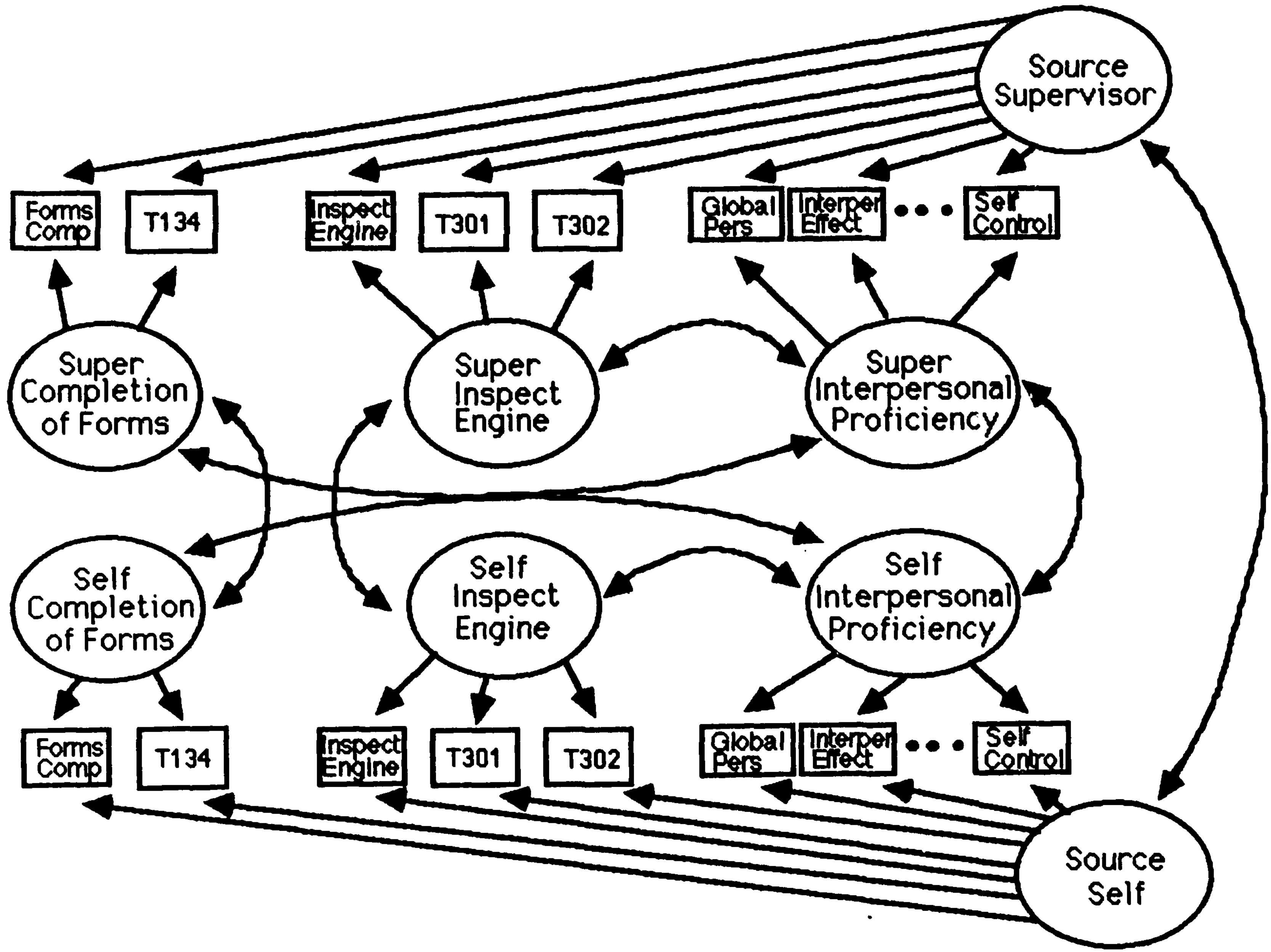


Figure 15: Respecified First-Order Model for AFSC 426x2

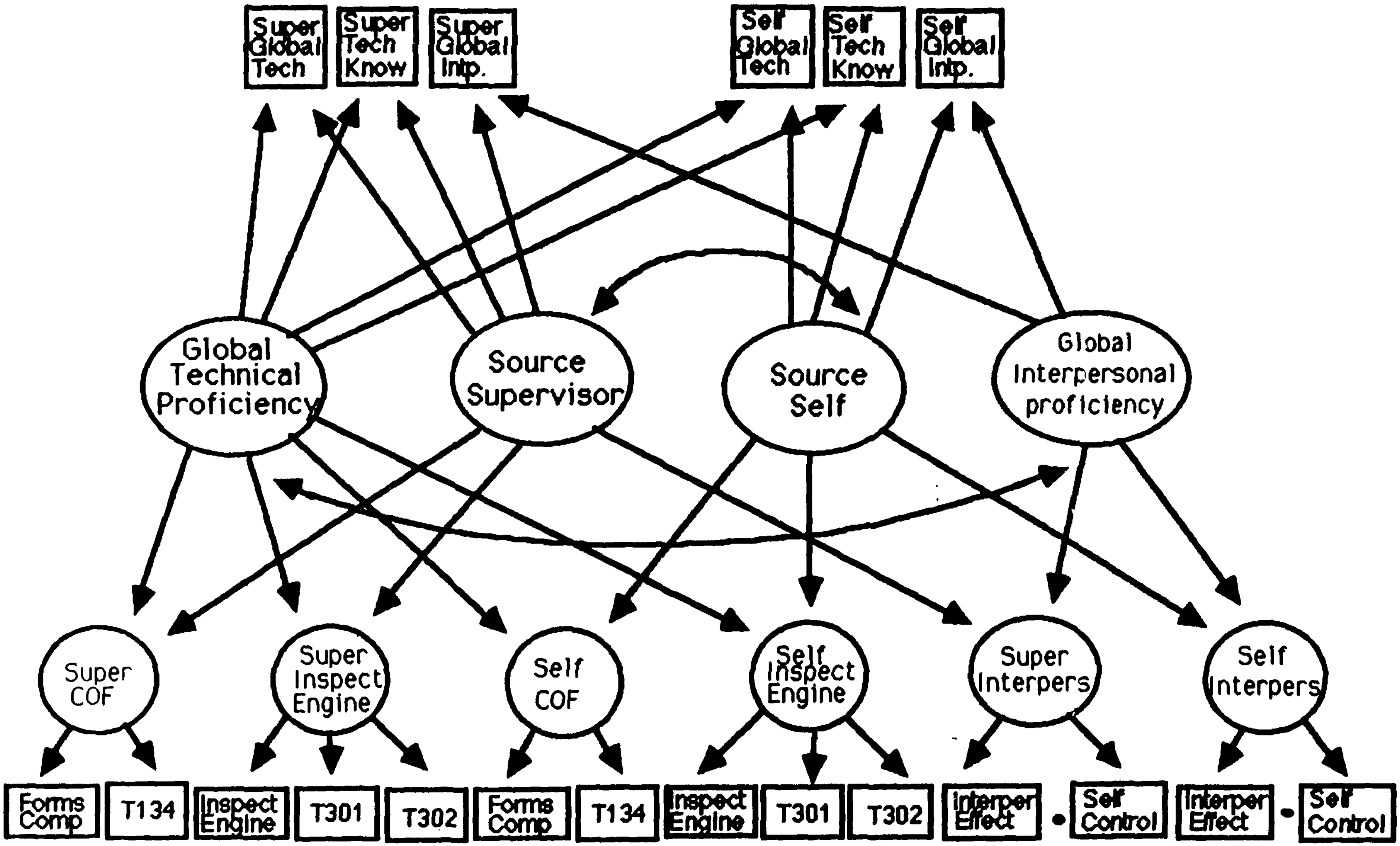


Figure 16: Respecified Hierarchical Model for AFSC 426x2

absent, and (c) several correlations among specific factors hypothesized to be non-zero. These factor correlations were estimated to assess potential convergence (a) among interpersonal proficiency factors, (b) between same-dimension within-source technical proficiency factors, and (c) between measurement source factors.

Figure 16 shows a respecified hierarchical model. Results above did not support the specification of inter-source dimensional technical proficiency factors. Second-order global technical and interpersonal proficiency factors were retained on a theoretical basis to account for correlations among specific proficiency factors. Figure 16 shows specifications of (a) task- and dimension-level technical proficiency measures' loadings on source-specific technical proficiency first-order factors, (b) Air Force-wide ratings' loadings on source-specific interpersonal proficiency first-order factors, (c) global technical proficiency and interpersonal proficiency ratings' loadings on general technical and interpersonal proficiency second-order factors, and (d) first-order technical and interpersonal proficiency factors' loadings on the second-order technical and interpersonal proficiency factors.

Tests of Respecified Models

Parameters in the respecified first-order factor model (Figure 15) and the respecified hierarchical model (Figure 16), were estimated using the LISREL-VI program, and allowing a maximum of 90 minutes CPU time. FIML estimates for the respecified hierarchical model converged in 82 iterations and within 81 minutes CPU computing time. All but one of the standardized FIML parameter estimates were within admissible parameter space. The respecified first-order factor model converged in 72 iterations and in just over 68 minutes CPU computing time. However, several FIML parameter estimates for the respecified first-order model were "improper." 2SLS estimates for parameters in the first-order model were reasonable, however, and were used for further analyses.

Only the hierarchical model converged in "proper" FIML estimates, so comparable goodness-of-fit statistics were not directly calculable between it and the first-order factor model. However, approximations to Jöreskog and Sörbom's (1984) goodness-of-fit index (GFI) were calculated for both models using 2SLS estimates. The SAS MATRIX procedure (SAS Institute, 1985b) was used to calculate the reproduced correlation matrix from 2SLS estimates in the first-order model:

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}_X \underset{\sim}{\Phi} \underset{\sim}{\Lambda}'_X + \underset{\sim}{\Theta}_\delta \tag{11}$$

and the hierarchical model:

$$\underset{\sim}{\Sigma} = \left|\begin{array}{c|c} \underset{\sim}{\Lambda}_Y(\underset{\sim}{\Gamma}\underset{\sim}{\Phi}\underset{\sim}{\Gamma}' + \underset{\sim}{\Psi})\underset{\sim}{\Lambda}'_Y + \underset{\sim}{\Theta}_\epsilon & \underset{\sim}{\Lambda}_Y\underset{\sim}{\Gamma}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'_X \\ \hline \underset{\sim}{\Lambda}_X\underset{\sim}{\Phi}\underset{\sim}{\Gamma}'\underset{\sim}{\Lambda}'_Y & \underset{\sim}{\Lambda}_X\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'_X + \underset{\sim}{\Theta}_\delta \end{array}\right| \tag{12}$$

Approximate GFIs were computed as:

$$GFI = 1 - \frac{tr(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S} - \underset{\sim}{I})^2}{tr(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S})^2} \tag{13}$$

where "tr" indicates the trace of a matrix, $\underset{\sim}{S}$ is the sample matrix of correlations among observed variables, and $\underset{\sim}{I}$ is an identity matrix dimensioned as $\underset{\sim}{S}$ and $\underset{\sim}{\Sigma}$. GFIs for the respecified first-order and hierarchical models were .435 and .818, respectively. These are low by Bentler and Bonett's (1980) standards. However, when evaluated against a chi-square to degrees-of-freedom ratio standard (Hoelter, 1983), the hierarchical model (3890.69/2098 = 1.85) appeared to fit the data well. (This ratio was not calculable for the first-order factor model since the solution converged in improper estimates). This discrepancy reflects the fact that a very parsimonious (highly overidentified) hierarchical model was very effective, but imperfect, in accounting for interrelationships among observed variables. The first-order model, while also highly overidentified, was less effective in accounting for observed relations. Since (a) only the respecified hierarchical model resulted in a convergent FIML solution containing "proper" estimates, and (b) 2SLS hierarchical model parameter estimates reproduced relations among observed variables more effectively than first-order factor model parameter estimates, the hierarchical model was accepted as a more defensible representation of the latent structure underlying relations among observed proficiency measures. Standardized FIML estimates of first- and second-order factor loadings, and correlations among second-order factors are shown in Tables 11 - 14.

Table 11 shows variables' standardized loadings on the first-order factors. Columns in Table 11 correspond exactly to columns in the actual factor pattern matrix, but many rows of

Table 11

Standardized FIML First-Order Factor Loadings: Respecified Hierarchical Model for AFSC 426x2

Factor:	Completion of Forms				Remove/Replace Eng. Comp.				Inspect Engine				Interp. Prof.		
Source:	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer
Variables:															
COF	.848	.890	.878	---	0	0	0	---	0	0	0	---	0	0	0
RREC	0	0	0	---	.770	.684	.999	---	0	0	0	---	0	0	0
IE	0	0	0	---	0	0	0	---	.628	.785	.801	---	0	0	0
T134	.774	.764	.876	.767	0	0	0	0	0	0	0	0	0	0	0
T301	0	0	0	0	0	0	0	0	.817	.807	.727	.439	0	0	0
T347	0	0	0	0	.709	.859	.763	.353	9	0	0	0	0	0	0
T302	0	0	0	0	0	0	0	0	.475	.593	.572	.507	0	0	0
T353	0	0	0	0	.835	.888	.906	.828	0	0	0	0	0	0	0
T360	0	0	0	0	.754	.858	.827	.365	0	0	0	0	0	0	0
T363	0	0	0	0	.747	.793	.776	.451	0	0	0	0	0	0	0
INTEF	0	0	0	---	0	0	0	---	0	0	0	---	.608	.857	.781
REG ORD	0	0	0	---	0	0	0	---	0	0	0	---	.716	.866	.795
INTEG	0	0	0	---	0	0	0	---	0	0	0	---	.681	.798	.762
LSHIP	0	0	0	---	0	0	0	---	0	0	0	---	.637	.842	.712
MILAP	0	0	0	---	0	0	0	---	0	0	0	---	.612	.715	.771
S DIS	0	0	0	---	0	0	0	---	0	0	0	---	.658	.831	.671
S CNT	0	0	0	---	0	0	0	---	0	0	0	---	.585	.820	.614

zero-restricted elements have been deleted toward clarity of presentation. Nearly all factor loadings are high, corroborating evidence above that supported strong effects of within-source dimensional proficiency factors.

Table 12 shows the actual factor pattern matrix of variables' (indicants') loadings on the second-order factors. Strong loadings on each of the Measurement Method factors indicate strong measurement source effects on global ratings. Self global ratings of Technical Proficiency and Technical Knowledge were larger than Supervisory or Peer ratings. This indicates that self ratings may be the most appropriate source from which obtain to global assessments of incumbent technical proficiency. On the other hand, Supervisory and Peer ratings of global Interpersonal Proficiency had larger loadings on the Global Interpersonal Proficiency factor than did the Self rating, indicating that Supervisors and Peers might best assess Interpersonal Effectiveness.

First-order factors' loadings on the second-order factors, shown in Table 13, show strong Global Method factors' effects on respective within-source proficiency factors. Somewhat weaker Global Interpersonal Proficiency factor effects on within-source Interproficiency factors suggest only a moderate degree of convergence among sources. The magnitudes of within-source Technical Proficiency factors' loadings on the General Technical Proficiency factor were varied. However, relatively higher loadings on the WTPT and Self within-source proficiency factors suggested some convergence between sources in assessing overall

Table 12

FIML Estimates of Indicants' Loading on Second-Order Factors: Hierarchical Model for AFSC 426x2

Factors:	Global Technical Proficiency	Global Interpers. Proficiency	Global Method		
			Self	Super.	Peer
<u>Variables</u>:					
Self-GL TECH	.524	0	.511	0	0
Self-GL INTP	0	.152	.607	0	0
Self-TECH KN	.508	0	.608	0	0
Super.-GL TECH	.250	0	0	.768	0
Super.-GL INTP	0	.420	0	.768	0
Super.-TECH KN	.233	0	0	.814	0
Peer-GL TECH	.268	0	0	0	.743
Peer-GL INTP	0	.394	0	0	.650
Peer-TECH KN	.232	0	0	0	.803

Table 13

FIML Estimates of First-Order Factors' Loadings on Second-Order Factors: Hierarchical Model for AFSC 426x2

Second-Order Factors:	Global Technical Proficiency	Global Interpers. Proficiency	Global Method		
			Self	Super.	Peer
First-Order Factors:					
Interpers.-Self	0	.284	.925	0	0
Interpers.-Super.	0	.395	0	.873	0
Interpers.-Peer	0	.534	0	0	.630
Self-COF	.227	0	.522	0	0
Self-RREC	.683	0	.561	0	0
Self-IE	.468	0	.663	0	0
Super.-COF	.140	0	0	.732	0
Super.-RREC	.280	0	0	.806	0
Super.-IE	.368	0	0	.866	0
Peer-COF	.132	0	0	0	.693
Peer-RREC	.041	0	0	0	.788
Peer-IE	.191	0	0	0	.926
WTPT-COF	.548	0	0	0	0
WTPT-RREC	.445	0	0	0	0
WTPT-IE	.226	0	0	0	0

Table 14

FIML Estimates of Correlations Among Second-Order Factors: Hierarchical Model for AFSC 426x2

Factor:	1	2	3	4	5
1. Global Tech Prof.	1.000				
2. Global Intp. Prof.	.365	1.000			
3. Global Method-Self	0	0	1.000		
4. Global Method-Super.	0	0	.320	1.000	
5. Global Method-Peer	0	0	.257	.524	1.000

technical proficiency.

Finally, Table 14 shows FIML estimates of correlations among the second-order factors. Source and global proficiency factors were restricted to be orthogonal. Rating source factors were moderately intercorrelated, suggesting some shared perspective on evaluation of proficiency. The moderate correlation between global proficiency factors suggests the presence of distinct, but related general abilities.

Summary. Tests of the respecified (trimmed) hierarchical model, and the models derived from King et al. (1980) and Wherry and Bartlett (1982) lent more support to a hierarchical structure of proficiency. Compared to the respecified first-order factor model, the hierarchical model (a) converged in more reasonable structural model parameter estimates, and (b) more effectively reproduced observed correlations among proficiency measures. These results support Hulin's (1982) contention that performance factors, like ability factors (Thorndike, 1985), may be hierarchically arranged. They also support an assumption underlying the design of the Air Force Job Performance Measurement System, namely that proficiency may be operationalized at multiple levels of generality/specificity.

Statistical Estimation of Proficiency True Scores and Control of Measurement Method Effects

Brogden and Taylor (1950) described several sources of criterion bias. Using a classical test theory analogy, observed proficiency score variance may be seen as consisting of (a) variance due to differences in individuals' true proficiencies,

(b) variance due to systematic factors unrelated to true proficiency (bias), and (c) nonsystematic error variance. Theoretically, the ultimate criterion construct is better represented to the extent that proficiency measures reflect variance source (a), and not sources (b) and (c). It may also be expected that proficiency measures containing substantial true score variance would exhibit higher correlations with measures of factors theoretically related to proficiency, compared to more fallible measures. This hypothesis was addressed by comparing correlations of (a) regression estimates of scores on proficiency factors, (b) proficiency scores corrected for measurement source effects, and (c) observed proficiency scores, with measures of (a) job experience, (b) aptitude, (c) success in training, and (d) situational constraints on performance.

Using estimates of parameters in the respecified hierarchical model, regression estimated factor scores (Kim & Mueller, 1978; Mulaik, 1972) on the second-order factors (global proficiency and measurement source factors) were computed, using the SAS MATRIX procedure (SAS Institute, 1985b) as:

$$\hat{\xi} = \underset{\sim}{X}\,\underset{\sim}{R}_{XX}\,\underset{\sim}{\Lambda}_{X}\,\underset{\sim}{\Phi} \qquad (14)$$

where $\hat{\xi}$ (175 x 5) contains estimated scores on the five second-order factors each for 175 airmen, X (175 x 9) contains 175 observations each on the nine manifest indicators of the factors, $\underset{\sim}{R}_{XX}$ (9 x 9) contains correlations among the $\underline{X}$s, $\underset{\sim}{\Lambda}_{X}$ is the second-order factor pattern matrix (Table 12), and $\underset{\sim}{\Phi}$ contains estimates of correlations among the ξs (Table 14). Scores for the first-order factors were computed as:

$$\bar{\eta} = Y R_{YY} \Lambda_{Y} (\Gamma \Phi \Gamma' + \Psi) \tag{15}$$

where $\bar{\eta}$ (175 x 15) contains estimates of scores on the 15 first-order factors, Y (175 x 58) contains observations on the manifest indicators of the ηs, R_{YY} (58 x 58) contains correlations among Ys, Γ (15 x 5) contains estimates of first-order factors' loadings on second-order factors (Table 13), and Ψ (15 x 15) contains residual variances from the second-order factor solution.

Estimates of observed variables' scores, corrected for measurement source effects were computed as:

$$\bar{X} = X - X R_{XX} \Lambda_{X(m)} \Phi_{m} \Lambda'_{X(m)} \tag{16}$$

for manifest indicators of second-order factors, and:

$$\bar{Y} = Y - Y R_{YY} \Lambda_{Y} (\Gamma_{m} \Phi_{m} \Gamma'_{m} + \Psi) \Lambda'_{Y} \tag{17}$$

for manifest indicators of first-order factors where $\bar{X}$ (175 x 9) and $\bar{Y}$ (175 x 58) contain estimates of scores on manifest indicators with measurement source effects removed, $\Lambda_{X(m)}$ (9 x 3) and Γ_{m} (15 x 3) are portions of Λ_{X} and Γ defined above with columns corresponding to measurement sources and Φ_{m} (3 x 3) contains correlations among measurement source factors.

<u>Derived Global Technical Proficiency Measures</u>. General technical proficiency true scores were approximated by regression estimated scores on the Global Technical Proficiency factor. An observed score composite was also computed by summing over scores

on manifest indicators of the Global Technical Proficiency factor (i.e., Self-, Supervisor- and Peer ratings of Global Technical Proficiency and Technical Knowledge). A composite of observed scores corrected for measurement source effects, was also computed by summing over residualized scores on manifest indicators created as shown in Equation 16.

Derived Dimensional Technical Proficiency Composites. Composites of dimensional proficiency true score estimates were computed, for each measurement source, by summing over dimensional ("Completion of Forms", "Remove/Replace Engine Components", and "Inspect Engine") factor score estimates. Observed score composites were computed by summing scores on manifest indicators of dimensional proficiency factors. Composites of observed scores corrected for measurement source effects (created from Equation 17) were similarly computed.

Tables 15 - 18 compare these criterion measures' correlations with measures of (a) experience - Months on Engine, (b) aptitude - ASVAB Mechanical Aptitude Index, (c) success in training - Technical School Grade, and (d) situational constraints - a composite of self-ratings on three items indicating the extent to which: technical manuals and other written materials are (i) clear and (ii) available, and (iii) tools and equipment are available.

As expected, correlations between experience and (a) global proficiency true score estimates, and (b) scores corrected for measurement effects, were larger than the correlation between experience and a global proficiency composite of observed scores

Table 15

Correlations Between Months-on-Engine and Technical Proficiency Measures

	True Score Estimates	Observed Score Composite Corrected For Msmt. Method	Observed Score Composite
Global Proficiency	.393**	.428**	.348**
Dimensional Proficiency Source:			
Self	.045	.407**	.401**
Supervisor	.079	.270**	.337**
Peer	.077	.030	.211**
WTPT	.173*	.243**	.237**

*$p < .05$, **$p < .01$, $n = 175$

Table 16

Correlations Between ASVAB Mechanical Aptitude Index and Technical Proficiency Measures

	True Score Estimates	Observed Score Composite Corrected For Msmt. Method	Observed Score Composite
Global Proficiency	.037	.070	.150
Dimensional Proficiency Source:			
Self	.058	.058	.090
Supervisor	.158*	-.010	.175*
Peer	.081	-.070	.016
WTPT	.090	.047	.098

Table 17

Correlations Between Technical School Grade and Technical Proficiency Measure

	True Score Estimates	Observed Score Composite Corrected For Msmt. Method	Observed Score Composite
Global Proficiency	.037	.079	.153
Dimensional Proficiency Source:			
Self	-.038	.136	.054
Supervisor	-.007	.159*	.137
Peer	.066	.006	.125
WTPT	.080	.176	.188*

Table 18

Correlations Between Situational Constraint Composite and Technical Proficiency Measures

	True Score Estimates	Observed Score Composite Corrected For Msmt. Method	Observed Score Composite
Global Proficiency	-.082	.090	-.081
<u>Dimensional Proficiency Source</u>:			
Self	.105	-.096	.033
Supervisor	.048	-.045	-.002
Peer	.095	-.076	.050
WTPT	-.003	-.042	.032

However, correcting for measurement method in dimensional proficiency composites effects had little effect, or an attenuating effect, on correlations with MOE (see Table 15). Dimensional proficiency true score estimates' correlations with MOE were uniformly smaller than observed score counterparts, and most were nonsignificant. These results were, of course, counter to hypotheses. Two possible interpretations of these findings are that dimensional proficiency true score estimates reflect (a) specific proficiencies that are not related to a global predictor such as MOE (b) or areal biases (Wherry & Bartlett, 1982) that in fact do not relate to true proficiency.

Tables 16 through 18 show further that neither (a) correcting observed scores for measurement source effects, nor (b) using proficiency true score estimates in lieu of observed score counterparts, had any consistent effect on correlations between proficiency measures and measures of other factors theoretically related to performance. In fact, derived scores' correlations with factors theoretically related to performance were often lower in absolute value than correlations with uncorrected proficiency scores. These results were counter to hypotheses and suggest that neither statistical estimation of latent proficiency true scores, nor statistical correction for measurement method effects had the expected beneficial effects of disattenuating correlations with other factors theoretically related to job proficiency.

Discussion

None of the models tested received unequivocal empirical support as they were initially specified. Many of the models' major propositions were supported, but one key source of disconfirmation concerned inter-source proficiency factors. The specification of inter-source proficiency factors reflects the hypothesis that different measurement sources converge in representing incumbent proficiency on multiple performance dimensions and this hypothesis received little support.

Evidence favored a hierarchical conceptualization of the underlying structure of performance, over alternative models that considered only first-order performance factors. Support was obtained for the hypothesis that within-source proficiency effects are mutually dependent on higher-order global technical and interpersonal proficiency factors. These dependencies were, however, imperfect. This is consistent with the notion that proficiency measures obtained from different measurement sources are, in part, capturing common aspects of the total criterion space, but at the same time reflect different proficiency effects. These findings also support assumptions underlying the design of the JPMS: (a) that proficiency may be operationalized at different levels of specificity/generality, and (b) different measurement sources may provide distinct and complementary assessments of airman performance.

Strong measurement source effects were identified. This supported hypotheses proposed earlier (e.g., King et al., 1980; Wherry & Bartlet, 1982) that observed performance measures reflect substantial influences of the measurement methods.

Prototype procedures for (a) statistical control of measurement method effects, and (b) estimation of proficiency true scores, appeared to be ineffective. These procedures were derived under the rationale that correlations between job proficiency and other factors theoretically related to proficiency might be attenuated because of either (a) measurement unreliability, and/or (b) contamination by extraneous measure method effects. Thus, it was expected that correlations between proficiency true score estimates and measures of other factors theoretically related to performance would be disattenuated, relative to correlations based on observed measures of proficiency. Results were largely inconsistent with this expectation.

Summary. Analyses of JEM (426x2) JPMS data supported a hierarchical structure of job proficiency in which dimensional proficiency is dependent, in part, on more global technical proficiency, interpersonal proficiency, and measurement source factors. The next phase of this research examined the generality of these findings and the efficacy of statistical correction procedures in a second AFS.

IV. GENERALIZABILITY OF FINDINGS: EXTENSION TO AIR TRAFFIC CONTROL OPERATOR (AFS 272x0) JPM DATA

The primary objectives of this research phase were to determine the generalizability of research findings from the Jet Engine Mechanic AFS (AFSC 426x2) in a second AFS. The Air Traffic Control Operator (ATC) AFS was chosen because (a) of the timing of JPMS data collection for the ATC AFS, (b) of the operational importance of effective performance in this AFS, (c) the ATC AFS represented a different ASVAB selector area (General), and (d) of the relatively large JPMS sample size (191 airmen).

Specific research plans were to assess the goodness-of-fits of the respecified first-order factor model (derived from King, et al., 1980 and Wherry & Bartlett, 1982) and the respecified hierarchical model investigated in the Jet Engine Mechanic (JEM) sample. While the first-order factor model had not received strong support in the JEM sample, its tenability was addressed in the ATC sample to determine whether its lack of fit might have been due to characteristics peculiar to the JEM sample. The tenability of the hierarchical model was investigated in the ATC sample to determine its transportability to a second AFS. The final goal of this research phase was to replicate the statistical correction procedures effected in the JEM sample and to determine the generalizability of these results to a second AFS.

Measures

Dimensional Proficiency

Technical Proficiency. Tasks in the ATC JPMS were examined to identify those tasks (a) which were commonly performed by incumbents both in the control tower and radar, (b) clearly represented performance dimensions defined for the ATC JPMS. Tasks and performance dimensions which were subsequently chosen for the present analyses are listed in Table 19. Self,

Table 19

Descriptions of Tasks Included in AIR Traffic Control Operator (272x0) Analyses

Dimension	Task #	Task Description
Coordination:	274	Issue or transmit enroute clearances using FAA procedures.
	320	Relay Meanconing, Intrusion, Jamming, & Interference (MIJI) information.
	293	Perform interfacility communications.
Controlling Aircraft:	253	Direct aircraft to alternate airports.
	260	Initiate emergency assistance procedures.
	366	Provide traffic advisories to VFR aircraft.
	278	Issue weather advisories.
Administration:	172	Annotate flight progress strip. (FAA Form 7230-8)
	318	Relay information from flight information publications (FLIP).
	319	Relay information from runway visual range (RVR) readings.

supervisor, and peer rating measures included for each of three performance dimensions were (a) Coordination (COORD) - dimensional-level COORD rating, and proficiency ratings on tasks 274, 320, and 293, (b) Controlling Aircraft (CNTRL) - dimensional-level CNTRL rating, and proficiency ratings on tasks 253, 260, 366, and 278, and (c) Administration (ADMIN) - dimensional-level ADMIN rating, and ratings on tasks 172, 318, and 319. WTPT measures included were Hands-on measures of proficiency on tasks 274, 293, 366, 278, 172, 318, and 319, and Interview measures of proficiency on tasks 320, 253, and 260.

Interpersonal proficiency. As for the JEM analyses, interpersonal proficiency measures included self, supervisory, and peer ratings of (a) Initiative/Effort (INTEF), (b) Knowledge of/Adherance to Regulations/Orders (RGORD), (c) Integrity (INTEG), (d) Leadership (LSHIP), (e) Military Appearance (MIL AP), (f) Self Development (S DEV), and (g) Self Control (S CNT).

Global Proficiency.

Global technical proficiency measures included self, supervisory, and peer ratings of (a) Global Technical Proficiency (GL TECH), and (b) Air Force-wide ratings of Technical Knowledge/Skill (TECH KN). Global Interpersonal Proficiency (GL INT) self, supervisor, and peer ratings were used to indicate general interpersonal proficiency.

Summary. A total of 79 ATC proficiency measures were included in this research phase - self, supervisor, and peer ratings each on two global proficiency scales, three dimensional proficiency scales, ten task proficiency scales, and eight Air

Force-wide scales, along with WTPT proficiency scores on ten tasks.

Procedures

A 79 x 79 matrix of correlations among proficiency measures was obtained as optional output from the SPSS-X FACTOR procedure. Missing data were deleted listwise leaving a correlation matrix based on 191 observations.

The correlation matrix was input to the LISREL-VI computer program for confirmatory first- and second-order FIML factor analysis. Two models initially fit to the data were a first-order factor model similar to the respecified model for the JEM sample shown in Figure 15, and a second-order, hierarchical model similar to the model shown in Figure 16. TSLS initial estimates were sought for FIML final estimates and a maximum of 60 minutes CPU time was allocated for each analysis.

Results

Neither of the models initially specified for the ATC JPMS converged in a proper solution within the allocated CPU time. Only TSLS initial estimates were obtained for the hierarchical model and many of the parameter estimates were "improper" (i.e., negative error variances, correlations larger in absolute value than 1.00). FIML estimates were obtained for the first-order factor model but many of its estimates too, were "improper." Several alternative specifications were attempted (for example, estimating separate factor models for each measurement source) but these too resulted in nonconvergent solutions and improper parameter estimates. Together, these results indicated that neither the factor structure proposed by the revised first-order model nor the second-order hierarchical model was consistent with the actual ATC JPMS data structure. That is, these results suggested the presence of significant model misspecifications.

A final estimation strategy was attempted which some literature indicates is less sensitive to model misspefications. This strategy suggested by Burt (1973, 1976) and Anderson and Gerbing (1982) is a limited information approach in which (a) separate single-factor models are estimated first (e.g., "Coordination" assessed by self ratings, or "Controlling Aircraft" assessed by the WTPT) to estimate relations between manifest variables and the underlying, latent variables, (b) in a second step, factor loadings are fixed to values estimated from the single-factor analyses and correlations among the factors (latent variables) are estimated, and (c) in a third step, hypothesized causal relations among the latent variables (factors, if any), are estimated from factor correlations. However, even this estimation strategy produced many improper estimates, suggesting (a) severe model misspecifications and (b) that neither of the revised models were consistent with the ATC JMPS data.

A final exploratory principal components (PCA) analysis was conducted to determine potential sources of model misspecifications. Correlations among task proficiency scores were analyzed separately for each measurement method (WTPT scores and self, supervisor, and peer ratings). Since three performance dimensions were purportedly represented by the tasks selected for this research phase, three components were retained in each analysis. Retained components were rotated to a direct OBLIMIN criterion (Jennrich & Sampson, 1966) to (a) preserve possible correlations among performance dimensions, and (b) achieve as simple a factor structure as possible (Kim & Mueller, 1978).

Direct OBLIMIN-rotated factor pattern loadings and component correlations are shown in Table 20. Shown next to each task number are acronymns for the performance dimension to which each task purportedly belongs. Tasks purportedly representing the same performance dimension did not cluster with one another on the same principal component. These results pointed to one likely source of specification error which prevented confirmatory factor analyses (CFAs) from converging in proper solutions. One critical assumption for CFA is that manifest indicators are correctly assigned to the underlying factor or latent variable that they represent (Long, 1983b). Violation of this assumption can result in misattribution of relations among latent variables and the types of nonconvergent solutions encountered in the present research phase. Thus the exploratory principal components analysis suggested the following potential problems in the ATC JPMS: (a) inappropriate allocation of ATC tasks to performance dimensions, (b) insufficiently concrete and/or distinct definition of the performance dimensions, (c) potentially overlapping performance dimensions, and (d) consequently, potential allocation of tasks to multiple performance categories.

Table 20

Air Traffic Control Operator Direct OBLIMIN Exploratory Principal Components Analysis Results

Source	Task(Dimension)	Principal Component I	II	III	Component Correlations I	II	III
Self Rating	320(COORD)	.850	.089	-.159	1.00		
	319(ADMIN)	.811	.018	.053	-.66	1.00	
	274(COORD)	.739	-.190	.446	-.19	.10	1.00
	318(ADMIN)	.734	-.132	.010			
	253(CNTRL)	.693	-.128	-.205			
	293(COORD)	-.088	-.963	-.082			
	172(ADMIN)	-.044	-.886	.166			
	366(CNTRL)	.186	-.764	-.118			
	278(CNTRL)	.235	-.676	-.117			
	260(CNTRL)	.362	-.328	-.587			
Supervisor Rating	293(COORD)	1.000	-.087	.023	1.00		
	278(CNTRL)	.780	-.046	.219	.75	1.00	
	366(CNTRL)	.674	.127	.212	.80	.63	1.00
	260(CNTRL)	.672	.325	-.019			
	318(ADMIN)	.594	.460	-.086			
	319(ADMIN)	-.033	.922	.084			
	320(COORD)	.172	.637	.200			
	253(CNTRL)	.291	.484	.257			
	274(COORD)	.034	.035	.915			
	172(ADMIN)	.229	.247	.533			
Peer Rating	(OBLIMIN rotation failed to converge)						
WTPT	320(COORD)	.750	-.232	.106	1.00		
	366(CNTRL)	.616	.102	-.107	.29	1.00	
	318(ADMIN)	.585	.171	-.246	.16	.13	1.00
	260(CNTRL)	.510	.183	.148			
	293(COORD)	-.154	.778	-.041			
	253(CNTRL)	.021	.689	.010			
	274(COORD)	.233	.577	.074			
	172(ADMIN)	.249	.461	.287			
	319(ADMIN)	-.192	.091	.855			
	278(CNTRL)	.482	-.017	.573			

In summary, an unambiguous mapping of ATC tasks into performance dimensions precluded confirmatory analysis using the ATC JPMS data. Optionally, confirmatory analyses could have proceeded using reformulated performance dimensions inferred from the exploratory principal components analysis. However, this approach would have been inconsistent with (a) the general programatic approach to confirmatory analysis (James et al., 1982), and (b) the design of the ATC JPMS itself. Also, there was no assurance that the likelihood of achieving substantively meaningful convergent solutions would have increased by using. performance dimensions redefined on the basis of exploratory analysis. Rather, JPM data from a third AFS were obtained for assessment of the generalizability of results from analyses based on the JEM JPM data.

V. GENERALIZABILITY OF FINDINGS: EXTENSION TO AEROSPACE GROUND EQUIPMENT MECHANIC (AFSC 423x5) JPM DATA

The primary objectives of this research phase were the same as those intended for the Air Traffic Control Operator JPM data, namely to determine the generalizability of research findings from the Jet Engine Mechanic AFS (AFSC 426x2) in a second AFS. The Aerospace Ground Equipment Mechanic (AGE) Specialty was chosen primarily because (a) of the timing of JPMS data collection for the AGE AFS, and (b) performance dimensions for the AGE AFS appeared to more distinctly defined and clearly distinguishable than for other AFSs in the overall JPM data base.

As intended for the ATC Specialty, specific research plans for the AGE Specialty were to assess the goodness-of-fits of the respecified first-order factor model (derived from King, et al., 1980 and Wherry & Bartlett, 1982) and the respecified hierarchical factor model investigated in the Jet Engine Mechanic (JEM) sample, and to replicate the statistical correction procedures effected in the JEM sample.

Measures

Dimensional Proficiency

Technical Proficiency. Tasks in the AGE JPMS were examined to identify those tasks (a) which were assessed by hands-on measures in the WTPT, and (b) clearly classifiable into appropriate performance dimensions. A pilot study was conducted by an AFHRL/ID scientist (Mr. T. M. Donnelly) at Chanute AFB in which subject matter experts (SMEs) classified tasks into performance dimensions defined for the AGE JPMS. Using a criterion of 80% agreement, all of the tasks assessed by hands-on measures in the AGE WTPT were unambiguously classified into respective performance dimensions with the exception of task #421 - "Remove or install hydraulic lines or fittings." Tasks and performance dimensions which were subsequently chosen for the present analyses are listed in Table 21. Self, supervisor, and peer rating measures included for each of three performance dimensions were (a) General AGE Maintenance (GENM) - dimensional-level GENM rating, and proficiency ratings on tasks 300 and 264, (b) AGE Electronic System Maintenance (ELEC) - dimensional-level ELEC rating, and proficiency ratings on tasks 209, 215, and 238, and (c) AGE Pick-up, Delivery and Service Functions (PDSF) - dimensional-level PDSF rating, and ratings on tasks 549, 155, and

Table 21

Descriptions of Tasks Included in Areospace Ground Equipment Mechanic (423x5) Analyses

Dimension	Task #	Task Description
General AGE Maintenance (GENM):	300	Remove or install fuel lines or fittings.
	264	Isolate engine, motor, or generator mechanical malfunctions.
AGE Electronic System Maintenance (ELEC):	209	Measure resistance of AGE electrical circuits.
	215	Perform AGE electrical system operational checks.
	238	Splice electrical system wiring.
AGE Pick-up, Delivery and Service Functions (PDSF):	549	Inspect vehicles for safety of operation.
	155	Perform load bank service inspections.
	154	Perform generator service inspections.

154. WTPT measures included were Hands-on measures of proficiency on tasks 300, 264, 209, 215, 238, 549, 155, and 154.

Interpersonal proficiency. As for the JEM and ATC analyses, interpersonal proficiency measures included self, supervisory, and peer ratings of (a) Initiative/Effort (INTEF), (b) Knowledge of/Adherance to Regulations/Orders (RGORD), (c) Integrity (INTEG), (d) Leadership (LSHIP), (e) Military Appearance (MIL AP), (f) Self Development (S DEV), and (g) Self Control (S CNT).

Global Proficiency.

Global technical proficiency measures included self, supervisory, and peer ratings of (a) Global Technical Proficiency (GL TECH), and (b) Air Force-wide ratings of Technical Knowledge/Skill (TECH KN). Global Interpersonal Proficiency (GL INT) self, supervisor, and peer ratings were used to indicate general interpersonal proficiency.

Summary. A total of 71 AGE proficiency measures were included in this research phase - self, supervisor, and peer ratings each on two global proficiency scales, three dimensional proficiency scales, eight task proficiency scales, and eight Air Force-wide scales, along with WTPT proficiency scores on eight tasks.

Procedures

A 71 x 71 matrix of correlations among proficiency measures was obtained from the SPSS-X FACTOR procedure. Missing data were deleted listwise leaving a correlation matrix based on 211 observations. This correlation matrix was input to the LISREL-VI computer program for confirmatory first- and second-order FIML

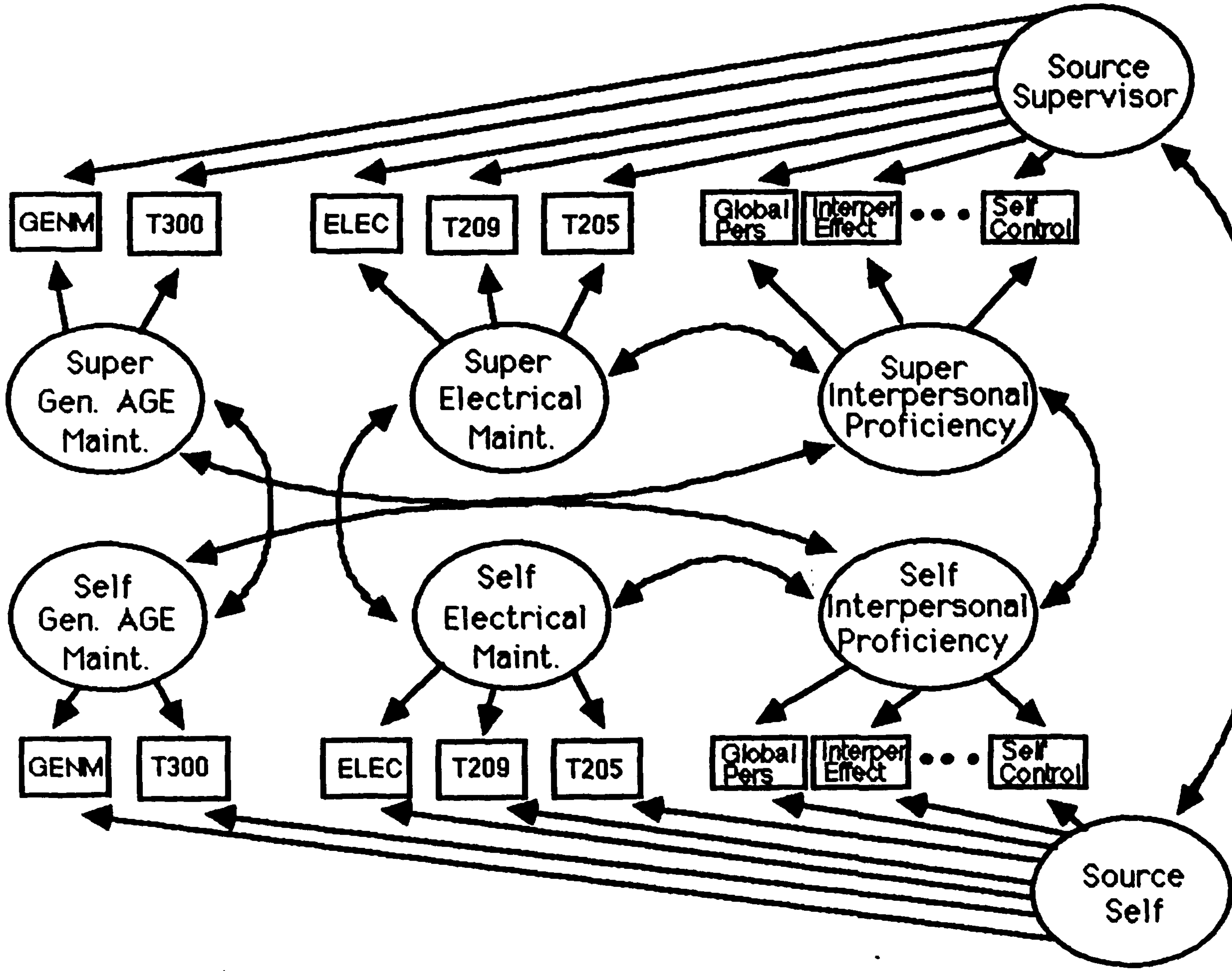


Figure 17: Respecified First-Order Model for AFSC 423x5

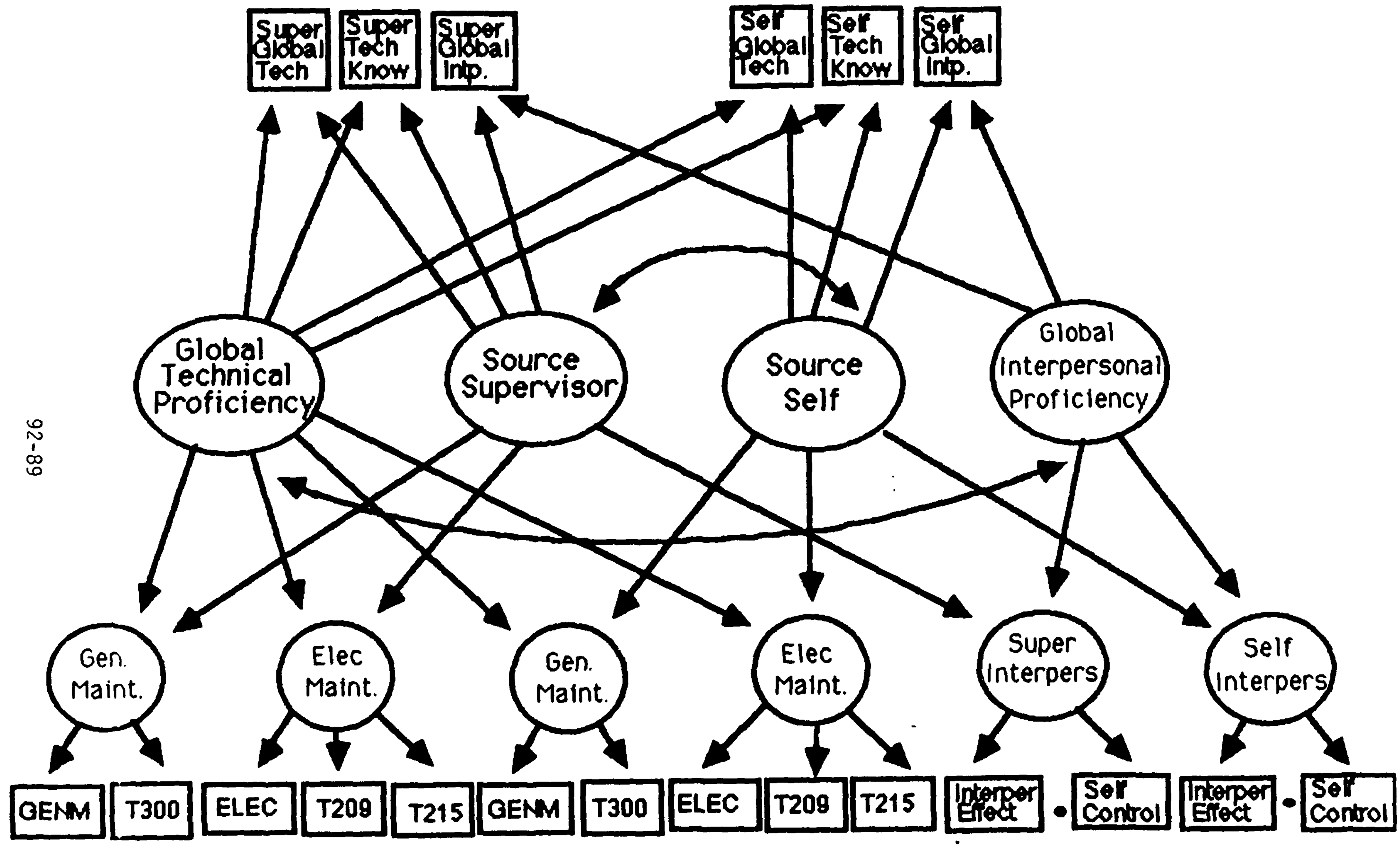


Figure 18: Respecified Hierarchical Model for AFSC 423x5

factor analysis. As in the ATC analyses, a first-order factor model similar to the respecified model derived from analysis of the JEM JPM data and a second-order, hierarchical model similar to the respecified model for the JEM sample were fit to the data. Simplified versions of these models are shown in Figures 17 and 18. These models are simplified in the senses that (a) only two measurement sources are shown (arbitrarily, self and supervisor rating sources), (b) only two of the three performance dimensions actually studied are shown, and (c) only a subset of the rating measures actually studied is shown. Starting values of 0.4 were specified in the LISREL-VI program to obtain TSLS initial estimates and FIML final estimates for model parameters. A maximum of 30 minutes CPU time was allocated for each analysis.

Results

The respecified first-order model failed to converge a in proper solution after 76 iterations for TSLS and FIML estimates, and within the maximum allowed CPU time. A further respecified (less restrictive) first-order model in which correlations among all factors (latent variables) were freely estimated parameters also failed to converge in a proper solution after a total of 72 iterations and within the allowed CPU time. These results parallel those from analysis of the JEM JPMS data in suggesting that respecified first-order factor model based on King et al.'s (1980) and Wherry and Bartlett's (1982) models were not consistent with the data. Consequently, additional analyses were directed toward assessing the generalizability of earlier findings relating to a second-order, hierarchical model of proficiency.

The respecified hierarchical model (a simplified version of which is shown in Figure 18) converged in FIML estimates in 52 iterations. All parameter estimates were "proper" with the exception of one residual variance estimate which was negative but not significatly different from zero. The model's overall Chi-square statistic (Chi-square [2374] = 3921.84, $p < .0001$) and the overall Goodness-of-Fit statistic provided by the LISREL-VI program indicated that the model could be rejected both on statistical and practical grounds. However, an Chi-square/degrees-of-freedom ratio of 1.65 indicated that the fit of the model per parameter estimated was strong. Three other fit indices discussed by Marsh, Balla, and McDonald (1988): (a) the root mean square of the elements in the matrix of residuals from fitting the model to the sample correlation matrix (RMSR = .061), (b) the Tucker-Lewis (1973) reliability index (TLI = .890), and (c) Marsh et al.'s (1988) χ^2/I2 index (χ^2/I2 = .893), also indicated that the model was successful in accounting for a substantial proportion of covariation among the observed measures. Thus, the respecified model was regarded as a useful explanation for relations among proficiency measures.

Proficiency measures' loadings on the first-order proficiency factors are shown in Table 22. With very few exceptions, proficiency measures had strong and significant loadings on respective first-order factors representing technical and interpersonal proficiency assessed by alternative measurement sources. These results corroborate earlier findings in the Jet

Table 22

Standardized FIML First-Order Factor Loadings: Revised Hierarchical Model for AFS 423x5

Factor:	General AGE Maintenance				Elec. Systems Maintenance				Pick-up, Service, Delivery				Interpers. Proficiency		
Source:	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer	WTPT	Self	Super.	Peer
Variables:															
GENM	.780	.829	.867	---	0	0	0	---	0	0	0	---	0	0	0
ELEC	0	0	0	---	.789	.907	.854	---	0	0	0	---	0	0	0
PDSF	0	0	0	---	0	0	0	---	.682	.835	.700	---	0	0	0
T300	.717	.602	.768	.554	0	0	0	0	0	0	0	0	0	0	0
T264	.821	.789	.821	.606	0	0	0	0	0	0	0	0	0	0	0
T209	0	0	0	0	.769	.820	.882	.629	0	0	0	0	0	0	0
T215	0	0	0	0	.855	.873	.892	.569	0	0	0	0	0	0	0
T238	0	0	0	0	.727	.806	.826	.292	0	0	0	0	0	0	0
T549	0	0	0	0	0	0	0	0	.749	.800	.756	.681	0	0	0
T155	0	0	0	0	0	0	0	0	.709	.814	.808	.324	0	0	0
T154	0	0	0	0	0	0	0	0	.791	.857	.849	.435	0	0	0
INTEF	0	0	0	---	0	0	0	---	0	0	0	---	.761	.849	.833
REG ORD	0	0	0	---	0	0	0	---	0	0	0	---	.697	.841	.813
INTEG	0	0	0	---	0	0	0	---	0	0	0	---	.702	.813	.812
LSHIP	0	0	0	---	0	0	0	---	0	0	0	---	.797	.829	.853
MILAP	0	0	0	---	0	0	0	---	0	0	0	---	.708	.752	.778
S DIS	0	0	0	---	0	0	0	---	0	0	0	---	.766	.849	.838
S CNT	0	0	0	---	0	0	0	---	0	0	0	---	.574	.702	.651

Table 23

FIML Estimates of Indicants' Loading on Second-Order Factors: Hierarchical Model for AFSC 423x5

Factors:	Global Technical Proficiency	Global Interpers. Proficiency	Global Method		
			Self	Super.	Peer
Variables:					
Self-GL TECH	.411	0	.723	0	0
Self-GL INTP	0	.333	.660	0	0
Self-TECH KN	.443	0	.718	0	0
Super.-GL TECH	.542	0	0	.707	0
Super.-GL INTP	0	.455	0	.737	0
Super.-TECH KN	.555	0	0	.741	0
Peer-GL TECH	.557	0	0	0	.675
Peer-GL INTP	0	.537	0	0	.659
Peer-TECH KN	.514	0	0	0	.730

Engine Mechanic (JEM) JPMS data.

Table 23 shows global technical and interpersonal proficiency measures' loadings on the second-order technical proficiency, interpersonal proficiency, and measurement method factors. Consistent with findings from the JEM Specialty, strong measurement method effects were found for self, supervisor, and peer rating measures for the AGE Specialty. Also consistent with earlier findings were results indicating that supervisor and peer ratings were stronger indicators of the global interpersonal proficiency factor (second column of loadings in Table 23). The patterns of technical proficiency ratings' loadings on the global technical proficiency factor differed somewhat between the JEM results and the AGE results: self technical profieciency ratings loaded most highly in the JEM JPMS data, whereas loadings were more homogeneous in the AGE results (see Table 12 vs. Table 23). Generally, however, patterns of global rating items' loadings were consistent between the JEM and AGE JPMS data sets.

First-order factors' loadings on the second-order factors are shown in Table 24. These results too, were generally consistent with earlier findings from the JEM JPMS data: (a) strong global method effects contributed to covariances among the first-order factors, (b) first-order interpersonal proficiency factors' loadings on the global, second-order factor were somewhat larger for the peer and supervisor first-order interpersonal proficiency factors, and (c) loadings of the first-order technical proficiency factors on the global, second-order technical proficiency factor were varied. As might be expected,

Table 24

FIML Estimates of First-Order Factors' Loading on Second-Order Factors: Hierarchical Model for AFSC 423x5

Second-Order Factors:	Global Technical Proficiency	Global Interpers. Proficiency	Global Method		
			Self	Super.	Peer
Variables:					
Interpers.-Self	0	.470	.818	0	0
Interpers.-Super.	0	.524	0	.798	0
Interpers.-Peer	0	.576	0	0	.777
Self-GENM	.434	0	.887	0	0
Self-ELEC	.349	0	.816	0	0
Self-PDSF	.138	0	.843	0	0
Super.-GENM	.476	0	0	.878	0
Super.-ELEC	.504	0	0	.808	0
Super.-PDSF	.268	0	0	.888	0
Peer-GENM	.432	0	0	0	.913
Peer-ELEC	.246	0	0	0	.913
Peer-PDSF	.117	0	0	0	.927
WTPT-GEMN	.255	0	0	0	0
WTPT-ELEC	.251	0	0	0	0
WTPT-PDSF	.237	0	0	0	0

loadings on the global technical proficiency factor tended to be higher for the more technically-oriented performance dimensions (General AGE Maintenance [GENM], and Electronic System Maintenance [ELEC]) regardless of the measurement source. This pattern of loadings supports the inference that the first global factor does reflect general technical proficiency.

Finally, Table 25 shows correlations among the global, second-order factors. These results diverge from earlier findings on the JEM Specialty in that correlations were substantially higher between (a) the global technical and interpersonal proficiency factors, and (b) the global self rating method factor and the supervisor and peer rating method factors. Thus, general technical proficiency and general interpersonal proficiency appeared to be more closely related in the AGE Specialty than in the Jet Engine Mechanic Specialty, and Self-Method effects were more highly related to Supervisor and Peer Method effects.

Statistical Estimation of Proficiency True Scores and Control of Measurement Method Effects.

The purpose of this research step was to replicate results obtained from the JEM JPMS data in which measures of factors theoretically related to proficiency (experience, aptitude, success in training, and situational constraints on performance) were correlated with (a) observed proficiency score composites, (b) composites of regression estimated scores on proficiency true score factors, and (c) composites of proficiency scores corrected for measurement source effects. The general hypothesis tested

Table 25

FIML Estimates of Correlations Among Second-Order Factors: Hierarchical Model for AFSC 423x5

Factor:	1	2	3	4	5
1. Global Tech Prof.	1.000				
2. Global Intp. Prof.	.702	1.000			
3. Global Method-Self	0	0	1.000		
4. Global Method-Super.	0	0	.612	1.000	
5. Global Method-Peer	0	0	.501	.499	1.000

was that measures of factors theoretically related to job proficiency would correlate more strongly with either (a) estimates of scores on underlying true score factors, or (b) observed proficiency score composites corrected statistically for measurement method effects, than with composites of observed proficiency measures. Results for the Jet Engine Mechanic Specialty had disconfirmed this hypothesis.

The SAS MATRIX procedure was used to obtain regression estimated scores on the first- and second-order factors in the AGE hierarchical model using Equations 14 and 15. However, the SAS MATRIX procedure indicated that the matrix of correlations among the manifest indicators of the latent endogenous variables (R_{YY} in Equations 15 and 17) was singular. Consequently, neither estimated first-order factor scores, nor task or dimensional scores corrected for measurement effects could be computed. However, factor score estimates were obtained for the second-order factors.

General technical proficiency true scores were estimated from factor scores on the global second-order technical proficiency factor (from Equation 14). Observed score composites representing general technical proficiency were also formed by summing self, supervisor, and peer global technical proficiency ratings, and Air Force-wide ratings of Technical Skill/Knowledge. Finally, composites of these variables were also computed but correcting statistically for measurement method effects (from Equation 16).

Table 26 shows the correlations between the three derived general technical proficiency measures and measures of (a) job experience (Months in AGE Shop), (b) Aptitude (ASVAB Mechanical Aptitude Area Index), (c) Final Rating in Technical Training, and (d) a Situational Constraint composite (a sum of self ratings on three items indicating the extent to which tools and equipment are available, and technical manuals and other written materials are clear and available). Results paralleled those from the Jet Engine Mechanic Specialty: statistical estimation of true score components or correction for measurement method effects actually had attenuating effects on the estimated relation between general technical proficiency and other factors theoretically related to proficiency. These results were again counter to original hypotheses.

Discussion

Results from analysis of AGE JPMS data corroborated earlier findings from the Jet Engine Mechanic Specialty: (a) a hierarchical structure of proficiency and measurement source effects, rather than a first-order model, appeared to be the more defensible representation of the actual structure of the JPMS, (b) hypothesized within-source dimensional proficiency effects received strong support, (c) second-order general technical proficiency, interpersonal proficiency, and measurement method effects were also supported, (d) general technical proficiency and interpersonal proficiency factors were positively correlated, (e) measurement source effects were also positively correlated, and (f) statistical procedures for estimating true proficiency scores and scores corrected for measurement source effects met

Table 26

Correlations Between Alternative Global Technical Proficiency Measures and Factors Theoretically Related to Performance

	True Score Estimate Composite	Observed Score Composite Corrected For Msmt. Method	Observed Score Composite
Job Experience (Months in AGE Shop)	.084	.159*	.249**
ASVAB MECH-AI	.065	.106	.211**
Final Tech School Grade	.370**	.686**	.905**
Situational Constraint Composite	.065	.022	.003

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

with little of their intended success.

VI. GENERAL DISCUSSION AND CONCLUSIONS

Three of the primary goals of this research were to (a) identify theoretical models of job proficiency and measurement source effects on proficiency measures, (b) assess the empirical goodness-of-fit of one or more of these models in one USAF JPMS data set, and (c) determine the generalizability of results to a second AFS. Eight plausible models were identified from a review of relevant literature and among those tested in the present research, a hierarchical model provided the most defensible representation of job proficiency and measurement source relations in the Jet Engine Mechanic (JEM) and Aerospace Ground Equipment Mechanic (AGE) JPM data sets. These results should be viewed as corroborative of the general design of the USAF JPMS. The USAF JPMS reflects a hierarchical structure of performance in the sense that data are collected from several measurement sources and at several levels of measurement specificity, including the task, dimensional, and global level. Yet to be resolved, however, are questions relating to (a) the relationship between proficiency at various levels of analysis, and (b) determinants of proficiency at different levels of analysis.

The question of the relationship between proficiency at different levels of analysis relates to more general literature on cross-level and mulit-level research (Lance, Hedge, & Alley, in press; Mossholder & Bedeian, 1983; Rousseau, 1985). The present research implicitly modeled relations between proficiency at different levels of analysis in the form of a "top-down" relationship (Diener, 1984) in which proficiency at more global levels of analysis determine proficiency at more micro levels. This is consistent with literature relating to hierarchical models of intelligence (Thorndike, 1985), and performance (e.g., Hulin, 1982), and dispositional models of attitudes (e.g., Lance, Lautenschlager, Sloan, & Varca, in press; Staw & Ross, 1985; Watson & Clark, 1984). The general idea here is that proficiency at more specific job activities are dependent on levels of more general job proficiency.

However, Lance & Woehr (1986) also considered other plausible relations between proficiency at micro and macro levels of analysis. For example, they also considered a "bottom-up" conception in which proficiency at more micro levels of analysis determines, rather than results from, proficiency at more macro levels of analysis. The idea here is that overall job proficiency is some function of proficiency at specific job elements.

While interesting in its own right, relationships between proficiency at different levels of analysis are also important from an operational perspective, and for determining appropriate means for assessing overall job proficiency (Wigdor & Green, 1986). Operationally, proficiency measures at different levels of analysis are required for different purposes. For example, task proficiency measures are appropriate for assessing training outcomes, dimensional proficiency measures are appropriate for assessing training needs, and global proficiency measures are appropriate for allocating merit raises and determining the predictive efficiency of cognitive ability measures. One key operational problem concerns forming appropriate linkages among proficiency measures at different levels of analysis. For example, if a link is established between cognitive ability and general job proficiency, and similarly, a link is established between a number of training interventions and proficiency on several job tasks, how should aptitude requirement/training time tradeoffs be appropriately weighed in terms of their overall effect on proficiency payoff to the Air Force? Future conceptual and empirical work is needed to address questions such as these.

From the standpoint of the JPM project, the question of how to appropriately assess overall job proficiency remains an important issue, and one whose resolution may be contingent on the theoretical structure of proficiency that is accepted. For example, if a "bottom-up" structure of proficiency is adopted, then overall job proficiency is only appropriately assessed by generating internally hetergeneous composite measures of proficiency on specific job elements (e.g., tasks). Assessments of overall job proficiency using global measures (e.g., global technical proficiency ratings) would overlook the contributions of proficiency on specific job elements to overall effectiveness. If, on the other hand, a "top-down" structure is adopted, then overall job proficiency would only be appropriately assessed using direct, global measures of overall proficiency. Assessments of overall proficiency using aggregates of specific proficiency measures would confound the measurement of overall proficiency with the more micro outcomes (i.e., proficiency on specific tasks) of general job proficiency. While the present research supported a hierarchical model based implicitly on a "top-down" structure of proficiency, additional research should also determine the empirical plausibility of alternative corss-level relations between proficiency at different levels of theory and analysis.

Distinguishing between proficiency at different levels of analysis is also important for identifying determinants of proficiency because proficiency at different levels of analysis may have different determinants. Performance determinants themselves may also be operationalized at different levels of analysis. For example, Lance, Hedge, and Alley (in press) found that specific task experience was a more potent predictor of task proficiency than was global job experience. Task proficiency may be most strongly affected by specific training interventions and experience in performing the task. Proficiency at the level of the performance dimension may more strongly reflect interests in, or preferences for, performing certain clusters of job duties over others, and general job proficiency may more strongly reflect the contributions of general aptitudes or prior educational attainment. Research on these questions could also have operational implications for, for example, prioritizing training interventions for those (clusters of) tasks which have the most severe consequences for ineffective performance. In summary, support in the present study for a hierarchical structure of proficiency and measurement source effects has implications for (a) the measurement of performance at multiple levels of analysis, (b) linking performance across multiple levels of analysis, and (c) identifying different sets of determinants of proficiency considered at different levels of analysis.

A second general finding from the present research concerns the extent of convergence among different measurement sources in representing job proficiency. Much of the theoretical work on performance measurement has based on the fundamental assumptions that (a) a (set of) performance true score(s) exists, (b) measures of performance obtained from different measurement sources reflect identical performance true scores, but different measurement source effects and nonsystematic factors, and (c) inter-source performance measures converge to the extent that they reflect performance true scores (e.g., Guilford, 1954, Kenny & Berman, 1980; King et al., 1980; Schmitt, Noe & Gottschalk, 1986; Wherry & Bartlett, 1982). Consequently, this literature suggests that measures of performance obtained from different sources will corrlelate highly to the extent that the measures reflect performance true score variance and not systematic biases. Under this logic, the present research demonstrated significant lack of convergence among measurement sources in finding little support for (a) first-order general technical proficiency or interpersonal proficiency factors, or (b) first-order cross-source dimensional technical proficiency factors.

Kavanagh et al. (1986) suggested an alternative view. They argued that different measurement sources may tap different aspects of the total criterion space. In effect, performance measures obtained from different measurement sources may reflect the influences of different true scores. This conception is consistent with recent literature relating to ecological perspectives on social perception (e.g., Funder, 1987; McArthur & Baron, 1983). From this perspective, measures obtained from different sources may be equally valid, but at the same time inter-source agreement may be low. Results from the present study are consistent with this idea. Reliable within-source effects were found for assessments of technical and interpersonal proficiency, but inter-source agreement was moderate.

One of the key implications of this interpretation is that multiple measurement sources are necessary for a more complete representation of the total criterion construct space. Results for the hierarchical model tested here in the JEM and AGE samples support the idea that proficiency is multidimensional by finding reliable, and well defined first-order source-specific dimensional proficiency factors. Many of these first-order factors also had significant loadings on more general second-order technical and interpersonal proficiency factors. This indicates that some aspects of proficiency that are assessed by each of several measurement sources relate to proficiency at a more global level, but that also some aspects of proficiency are assessed more or less uniquely by specific measurement sources. In particular, some measurement sources may be more appropriate for assessing general technical and/or interpersonal proficiency than others. For example, peer first-order technical proficiency factors did not tend to have high loadings on a second-order general technical proficiency factor in either the JEM or AGE samples, while WTPT, supervisor, and self first-order technical proficiency factors' loadings were higher. On the other hand, supervisor and peer first-order interpersonal proficiency factors tended to have higher loadings on a second-order general interpersonal proficiency factos in both samples (see Tables 13 and 24). Thus WTPT measures, and supervisor and self ratings may be more effective in representing general technical proficiency, while peer and supervisor ratings may more appropriately assess general interpersonal effectiveness.

A third general finding from the present research was of strong measurement method effects for self, supervisor, and peer rating sources. Large general, within-source factors are commonly interpreted as reflecting halo (Landy, Vance, Barnes-Farrell, & Steele, 1980). Other interpretations emphasize "true halo" effects or differing frames of reference for evaluating performance (Bingham, 1939; Lance, Woehr, & Fisicaro, 1989; McArthur & Baron, 1983). It was not possible to distinguish among between interpretations from the present results. The fact that cross-source general and dimensional proficiency factors (indicating inter-source agreement) were not supported makes distinguishing among these interpretations particularly difficult. In this case, general within-source factors may represent source-specific (a) halo, or (b) assessments of different aspects of general true proficiency.

Recent literature relating to ecological perspectives on social perception is consistent with the latter interpretation.

From this perspective, social perceivers construct different frames of reference concerning a social object as a function of their goals for interacting with the social object. From an organizational standpoint, this means that different true scores may exist depending on the organizational perspective taken by the rater. Self ratings, for example, may represent incumbents' judgments of their own abilities to perform their job successfully and, to a lesser extent, the actual outcomes of their job behaviors. Supervisor judgments, on the other hand, may be based more on the results of incumbent behaviors and perceptions of motivation and loyalty to the supervisor. Peer ratings may reflect incumbent collegiality, willingness to assist others, or incumbent status in the workgroup as determined by tenure, expert power, or the power to administer social rewards. Viewed this way, correlations among measurement method factors represent an alternative assessment of convergence among measurement sources. Correlations in Tables 14 and 25 indicate that convergence among measurement sources assessed in this manner is moderate to substantial, especially between peers and supervisors.

Finally, prototype procedures for estimating scores on underlying proficiency true score factors and for correcting observed proficiency measures for measurement method effects (bias) were unsuccessful. There may be several reasons why.

First, it may be the case that job experience, aptitude, success in training and situational constraints truly do not relate to incumbent job proficiency. This is counterintuitive, but proficiency may be determined primarily by factors such as motivation, work group norms, or the quality and amount of on-the-job training.

A second, more plausible, explanation concerns the ambiguity with which proficiency true scores and measurement method bias effects were identified. The ambiguity of interpreting general within-source factors as indicating halo or method bias was indicated above. Similarly, within-source dimensional proficiency factors may represent aspects of the criterion true score space or they may indicate "areal biases" (Wherry & Bartlett, 1982). In the absence of "pure" true score and bias measures, either interpretation remains plausible. The statistical correction procedures proposed here are based on the assumption that proficiency true score effects and method bias effects have been correctly identified. To the extent that this assumption has not been met, the statistical correction procedures attempted here are inappropriate.

A third possibility relates to the homogeneity-heterogeneity of the general proficiency measures derived in the statistical estimation process. Estimates for general proficiency true scores reported in Tables 15 - 18 and 26 are, theoretically, estimates of values on underlying unidimensional constructs. Similarly, general technical proficiency composites corrected for measurement method effects should have a lower underlying dimensionality as a result of having "purged" the composite of variance attributable to the three rating source effect factors. If proficiency is multiply determined, then it might be expected that factors that are theoretically related to proficiency would correlate more highly with heterogeneous, multifaceted proficiency measures, than with measures reflecting a lower dimensionality. This hypothesis should receive further research attention.

In summary, support for a hierarchical conception of proficiency and measurement method effects on proficiency measures supports the design of the USAF JPMS to assess proficiency from multiple perspectives (measurement sources), including more objective perspectives such as through the WTPT methodology. Counter to some earlier performance measurement literature, substantial inter-source agreement on the measurement of proficiency was not supported. Instead, results indicated that either (a) proficiency measures obtained from different measurement sources contain substantial amounts of source-specific bias, or that (b) different measurement sources assess somewhat different aspects of the total criterion construct space. Recent performance measurement and social cognition literature supports the latter interpretation. This conclusion argues for the importance of continued collection of performance-related information from multiple measurement sources for both research and operational uses.

Future research should also attempt to specify theoretically and demonstrate empirically linkages between proficiency at various levels of analysis. Work in this area may help to clarify some of the questions concerning the potential substitutability of alternative criterion measures for the high fidelity WTPT measures. For example, supervisor global technical proficiency judgments might be readily substitutable, and even preferable, for certain purposes such as test validation, while suitable substitutes may simply not be available for other purposes (e.g., assessment of proficiency on tasks that are critical for effective job performance).

The present research suggests that there is still much to be learned about the latent structure underlying criterion measures. For the present, however, a hierarchical structure in which alternative measurement sources converge moderately in assessing performance seems useful.

VII. REFERENCES

Alluisi, E. & Morgan, B. (1976). Engineering psychology and human performance. Annual Review of Psychology, 27, 305-330.

Alwin, D. (1974). An analytic comparison of four approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.) Sociological Methodology 1973-1974. San Francisco: Jossey-Bass.

Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measruement models to obtain unidimensional construct measurement. Journal of Marketing Research, 19, 453-460.

Arora, R. (1982). Validation of an S-O-R model for situation, enduring, and response components of involvement. Journal of Marketing Research, 19, 505-516.

Avison, W. R. (1978). Auxiliary theory and multitrait-multimethod validation: A review of two approaches. Applied Psychological Measurement, 2, 431-447.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P. M., Jackson, D. N. & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. Psychological Bulletin, 76, 186-204.

Bingham, W. V. (1939). Halo, invalid and valid. Journal of Applied Psychology, 23, 221-228.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Boruch, R. F., Larkin, J. D., Wolins, L. & MacKinney, A. C. (1970). Alternative methods of analysis: Multitrait-multimethod data. Educational and Psychological Measurement, 30, 833-853.

Boruch, R. F. & Wolins, L. (1970). A procedure for estimation of trait, method, and error variance attributable to a measure. Educational and Psychological Measurement, 30, 547-574.

Brogden, H. E. & Taylor, E. K. (1950). The theory and classification of criterion bias. Educational and Psychological Measurement, 10, 159-186.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology, 37, 1-21.

Burnkrant, R. E. & Page, T. J. Jr. (1982). An examination of the convergent, discriminant, and predictive validity of Fishbein's behavioral intention model. Journal of Marketing Research, 19, 550-561.

Burt, R. S. (1973). Confirmatory factor analytic structures and the theory construction process. Sociological Methods & Research, 2, 131-190.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. Sociological Methaods & Research, 5, 3-52.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, J. P. (1982). Some possible implications of "modeling" for the conceptualization of measurement. In F. J. Landy, S. Zedeck & J. Cleveland (Eds.), Performance measurement and theory. Hillsdale, NJ: Erlbaum.

Campbell, J. P., Dunnette, M. D., Lawler, E. E. & Weick, K. E. (1970). Managerial behavior, performance and effectiveness. New York: McGraw Hill.

Cooksey, R. W. & Freebody, P. (1985). Generalized multivariate lens model analysis for complex human inference tasks. Organizational Behavior and Human Decision Processes, 35, 46-72.

Cooper, W. H. (1981). Ubiquitous halo. Psychological Bulletin, 93, 218-244.

Cronbach, L. J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.

DeCotiis, T. & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 3, 635-646.

DeNisi, A. S., Cafferty, T. P. & Meglino, B. M. (1984). A cognitive view of the performance appraisal perocess: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.

Diener, E. (1984). Subjective well-being. Psychological Bulletin, 95, 542-575.

Dienesch, R. M. & Liden, R. C. (1986). Leader-member exchange model of leadership: A critique and further development. _Academy of Management Review_, _11_, 618-634.

Einhorn, H. J. & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. _Annual Review od Psychology_, _32_, 53-88.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. _Journal of Applied Psychology_, _66_, 127-148.

Fleishman, E. L. (1978). Relating individual differences to the dimensions of human tasks. _Ergonomics_, _21_, 1007-1019.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. _Psychological Bulletin_, _101_, 75-90.

Ghiselli, E. E. (1956). Dimensional problems of criteria. _Journal of Applied Psychology_, _40_, 1-4.

Ghiselli, E. E., Campbell, J. P. & Zedeck, S. (1981). _Measurement theory for the behavioral sciences_. San Francisco, CA: Freeman.

Givon, M. M. & Shapira, Z. (1984). Response to rating scales: A theoretical model and its application to the number of categories problem. _Journal of Marketing Research_, _21_, 410-419.

Green, S. G. & Mitchell, T. R. (1979). Attributional processes in leader-member interactions. _Organizational Behavior and Human Performance_, _23_, 429-458.

Guilford, J. P. (1954). _Psychometric methods_. New York: McGraw-Hill.

Hedge, J. W. (1984, August). The methodology of walk-through performance testing. Paper presented at the meeting of the American Psychological Association, Toronto, Ontario, Canada.

Hedge, J. W., & Lipscomb, M. S. (1987, September). Walk-through performance testing: An innovation to work sample testing. (AFHRL-TP-87-8). Air Force Human Resources Laboratory, Training Systems Division, Brooks AFB, TX.

Hedge, J. W. & Teachout, M. S. (1986). Job performance measurement: A systematic program of research and development. (AFHRL-TP-86-37) Air Force Human Resources Laboratory, Training Syctems Division, Brooks AFB, San Antonio, TX.

Hobson, C. J. & Gibson, F. W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. Academy of Management Review, 8, 640-649.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. Sociological Methods & Research, 11, 325-344.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. Journal of Applied Psychology, 63, 579-588.

Hulin, C. L. (1982). Some reflections on general performance dimensions and halo rating error. Journal of Applied Psychology, 67, 165-170.

Hulin, C. L., Drasgow, F. & Parsons, C. K. (1982). Item response theory: Applications to psychological measurement. Homewood, IL: Irwin.

Ilgen, D. R., Fisher,, C. D. & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.

James, L. R. (1973). Criterion models and construct validity for criteria. Psychological Bulletin, 80, 75-83.

James, L. R., Mulaik, S. A. & Brett, J. (1982). Causal analysis: Assumptions, models and data. Beverly Hills, CA: Sage.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. Psychometrika, 31, 313-323.

Jöreskog, K. G. & Sörbom, D. (1984). LISREL-VI: Analysis of linear srtuctural relations by maximum likelihood and least squares methods. Mooresville, IN: Scientific Software.

Judd, C. M., Jessor, R. & Donovan, J. E. (1986). Structural equation models and personality research. Journal of Personality, 54, 149-198.

Kalleberg, A. L. & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. Journal of Applied Psychology, 60, 1-9.

Kane, J. S. & Lawler, E. E. III (1978). Methods of peer assessment. Psychological Bulletin, 85, 555-586.

Katz, D. & Kahn, R. L. (1978). The social psychology of organizations. New York: Wiley.

Kavanagh, M. J., Borman, W. C., Hedge, J. W. & Gould, R. B. (1986). A classification scheme, literature review and directions for research. (AFHRL-TP-85-51) Air Force Human

Resources Laboratory, Manpower and Personnel Division, Brooks AFB, San Antonio, TX.

Kavanagh, M. J., MacKinney, A. C. & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 75, 34-49.

Kenny, D. A. & Berman, J. S. (1980). Statistical approaches to the correstion of correlational bias. Psychological Bulletin, 88, 288-295.

Kim, J-O, & Mueller, C. W. (1978). Factor analysis: Statistical methods and practical issues. Beverly Hills, CA: Sage.

King, L. M., Hunter, J. E. & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. Journal of Applied Psychology, 65, 507-516.

Klimonski, R. J. & London, M. (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59, 445-451.

Kozlowski, S. W. J., Kirsch, M. P. & Chao, G. T. (1986). Job knowledge, ratee familiarity, conceptual similarity and halo error: An exploration. Juornal of Applied Psychology, 71, 45-49.

Lance, C. E., Hedge, J. W., & Alley, W. E. (in press). Joint relationships of task proficiency with aptitude, experience, and task difficulty: A cross-level, interactional study. Human Performance.

Lance, C. E. & Lautenschlager, G. (1987). Applications of confirmatory factor analysis to performance rating data. Manuscript submitted for publication.

Lance, C. E., Lautenschlager, G. J., Sloan, C. E., & Varca, P. E. (in press). A comparison between bottom-up, top-down, and bidirectional models of relationships between global and life facet satisfaction. Journal of Personality.

Lance, C. E. & Woehr, D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. Journal of Applied Psychology, 71, 679-685.

Lance, C. E., Woehr, D. J., & Fisicaro, S. A. (1989). Cognitive categorization processes in performance evaluation: Confirmatory tests of two models. Manuscript submitted for publication.

Landy, F. J. & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Landy, F. J., & Farr, J. L. (1983). The measurement of work performance. New York: Academic.

Landy, F. J., Vance, R. J., Barnes-Farrell, J. L. & Steele, J. W. (1980). Statistical control of halo error in performance ratings. Journal of Applied Psychology, 65, 501-506.

Lee, C. (1985). Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. Academy of Management Review, 10, 322-331.

Lipscomb, M. S. (1984, August). A task-level domain sampling strategy: A content valid approach. Paper presented at the meeting of the American Psychological Association, Toronto, Canada.

Long, J. S. (1983a). Confirmatory factor analysis. Beverly Hills, CA: Sage.

Long, J. S. (1983b). Covariance structure models: An introduction to LISREL. Beverly Hills, CA: Sage.

Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mabe, P. A. III & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67, 280-296.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. Journal of Applied Psychology, 73, 107-117.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social eprception. Psychological Review, 90, 215-238.

Mosier, C. I. (1951). Problems and designs of cross-validation. Educational and Psychological Measurement, 11, 5-12.

Mossholder, K. W., & Bedeian, A. G. (1983). Cross-level inference and organizational research: Perspectives on interpretation and application. Academy of Management Review, 8, 547-558.

Mulaik, S. A. (1972). The foundations of factor analysis. New York: McGraw-Hill.

Nathan, B. R. & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102-114.

Naylor, J. C. (1982). Modeling performance. In F. J. Landy, S. Zedeck & J. Cleveland (Eds.) Performance measurement and theory. Hillsdale, NJ: Erlbaum.

Prien, E. P. & Liske, R. E. (1962). Assessments of higher-level personnel: III. Rating criteria: A comparative analysis of supervisor ratings and incumbent self-ratings of job performance. Personnel Psychology, 15, 187-194.

Rorer, L. G. (1965). The great response-style myth. Psychological Bulletin, 63, 129-156.

Rousseau, D. M. (1985). Issues of level in organizational research: Milti-level and cross-level perspectives. Research in Organizational Behavior, 7, 1-37.

SAS Institute Inc. (1985a). SAS version 5. Cary, NC: Author.

SAS Institute Inc. (1985b). Technical report: P-135, the matrix procedure. Cary, NC: Author.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 432-439.

Schmidt, F. L. & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. Personnel Psychology, 24, 419-434.

Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 2, 157-178.

Schmitt, N., Noe, R. A. & Gottschalk, R. (1986). Using the lens model to magnify raters' consistency, matching, and shared bias. Academy of Management Journal, 29, 130-139.

Schmitt, N. & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.

Slovic, P., Fischhoff, B. & Lichtenstein, S. 91977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

Spool, M. D. (1978). Training programs for observers of behavior: A review. Personnel Psychology, 31, 853-888.

Staw, B. M., & Ross, J. (1985). Stability in the midst of change: A dispositional approach to job attitudes. Journal of Applied Psychology, 70, 469-480.

Steel, R. P. & Ovalle, N. K. 2d (1984). Self-appraisal based on supervisory feedback. Personnel Psychology, 37, 667-685.

Thorndike, R. L. (1985). The central role of general ability in prediction. Multivariate Behavioral Research, 20, 241-254.

Thornton, G. C. III. (1980). Psychometric properties of self-appraisals of job performance. Personnel Psychology, 33, 263-271.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. Psychological Bulletin, 96, 465-490.

Wherry, R. J. (1952). The control of bias in ratings: VIII. A *theory of rating*. PRB Report No. 922, Contract No. DA-49-083 OSA69, Department of the Army.

Wherry, R. J. & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, *35*, 521-551.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*, 1-26.

Wigdor, A. K. & Green, B. F. Jr. (Eds.) (1986). *Assessing the performance of enlisted personnel: Evaluation of a joint-service research project*. Washington, DC: National Academy Press.

Zammuto, R. F., London, M. & Rowland, K. M. (1982). Organization and rater differences in performance appraisals. *Personnel Psychology*, *35*, 643-658.

Zedeck, S. & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance*, *18*, 269-294.

FINAL REPORT NUMBER 93
REPORT NOT PUBLISHABLE AT THIS TIME
Dr. Stephen Loy
760-7MG-134

Comparison of Supervisor's and Incumbent's Estimates of SD_y

Michael D. Matthews

Drury College

Final Report: Contract #F49620-85-C-0013/SB5851-0360

Comparison of Supervisor's and Incumbent's Estimates of SD_y

A major problem in manpower and personnel research within the military community involves transferring basic research findings to operational settings. Hunter and Schmidt (1983) state that one reason that such a difficulty exists is that scientists communicate in a somewhat "arcane" language that managers find difficult to thoroughly understand. Hunter and Schmidt suggest that in order for this communication process to be improved, scientists should translate their terms into those with economic implications, which in turn would be more meaningful to managers.

Matthews and Sewell (1987) recently reviewed attempts to quantify the results of manpower and personnel research in economic terms and to assess how relevant these attempts are to military personnel research programs. This literature review revealed that the developments in tnis domain most germane to the military have occured in the field of "utility analysis."

Since 1979, there has been a substantial amount of research in utility analysis. While basic utility models have existed for some time (e.g., Cronbach & Gleser, 1965), practical methods for estimating key parameter values were developed only within the past decade (Schmidt, Hunter, McKenzie, & Muldrow, 1979). The basic formulas have been extended to include vectors for economic variables (Boudreau, 1983a), employee flow (Boudreau, 1983b), recruitng practices (Boudreau & Rynes, 1985), effects of rejected offers (Murphy, 1986), and extentions for evaluating the utility of selection tests to post-selection organizational interventions

(Landy, Farr, & Jacobs, 1982; Schmidt, Hunter, & Pearlman, 1982).

The most important developments in utility analysis pertain to methods for estimating the standard deviation of job performance in dollars and cents terms. This variable is a key one in utility models, and is symbolized by SD_y. Schmidt et al. (1979) proposed a way of estimating this variable based on obtaining expert judgments of the dollars and cents worth of workers at the 50th, 15th, and 85th percentiles of performance. If job performance is normally distributed, then by subtracting the estimates of the 50th percentile from the 85th, and the 15th from the 50th, then an estimate of SD_y is obtained. A number of subsequent studies examined the validity of this approach (e.g., Burke & Frederick, 1986; Weekley, Frank, O'Connor, & Peters, 1985), and have found it to be farily accurate in the populations studied, although some psychometric questions do persist. Additionally, Hunter and Schmidt (1983) report that estimates of SD_y tend to fall near 40 percent of the mean annual salary for the job incumbents, a technique referred to as the "salary percentage technique," (Eaton, Wing, & Mitchell, 1985).

The Schmidt et al. (1979) method for estimating SD_y relies on obtaining estimates from supervisors concerning the value to the organization of the incumbents. In order to obtain an accurate estimate, a relatively large sample of supervisors is required. For example, Weekley et al. (1985) examined a sample of 196 supervisors employed by a national convenience store. Because of the large number of supervisors required to obtain these estimates, applying utility analysis methods in small organizations could be difficult due to there being an insufficient number of supervisors to provide performance estimates. This could limit the applicability of this technology to relatively large organizations.

A solution to this limitation exists if one could obtain the required estimates from the incumbents themselves. Conceptually, one could question whether or not incumbents have a sufficiently broad view or understanding of their contributions to the organization as a whole to provide the basis for accurate judgments. On the other hand, as Tuttle, Wilkinson, and Matthews (1985) have argued, no one knows their job as well as the workers themselves. Perhaps they are indeed in a good position - perhaps even the best position - to make such estimates of worth. This issue may be tested empirically, by obtaining estimates of value of workers from both supervisors and incumbents from the same organization, and comparing these estimates.

The purpose of the present study was to make just such a comparison. Estimates of the dollar value to the organization of workers at the 50th, 15th, and 85th percentiles of performance were obtained from a sample of supervisors and job incumbents. No basis exists for an a priori prediction of the outcome of the study, but the implications of the outcome are clear. If supervisors and incumbents were found to make similar estimates, then future utility analysis studies could be conducted in smaller organization, which lack the number of supervisors needed to make accurate estimates, but which have sufficient numbers of incumbents. Indeed, data from supervisors and incumbents could be pooled to allow for a sufficently large sample to accurately estimate utility parameters.

Method

Subjects. The subjects used in the current study were task scientists and their supervisors from a government human resources research laboratory. Eighteen incumbents were included in the sample. Their average age was 39.6, 16 were male, and two were female. Their average job tenure was 9.8 years. Their GS ratings ranged from GS-11 to GS 13. Thirteen supervisors were studied. Their average age was 42.2, 10 were males, and three were females. Their average tenure in the organization was 14 years, and they had been supervisors for an average of 6.5 years. Their grade ranged from GS/GM 13 to GS/GM 14.

Materials. Each respondent was asked three questions. These were modeled after the questions used by Schmidt et al. (1979), and required the supervisors and incumbents to provide estimates of the dollars and cents worth of a "task scientist" who performed at the 50th, 15th, or 85th percentile. A copy of the questionnaire is included in Appendix A. In addition, the subjects were asked their age, gender, time of service in the organization and (for supervisors) how long they had been in a supervisory position.

Procedure. A master list of the personnel assigned to the target organization, the manpower and personnel research division of a military research organization, was obtained and task scientists and their supervisors were identified. All such personnel assigned to the division were surveyed. The researcher made personal contact with each subject, explained that he was conducting a project designed to develop ways of quantifying the performance of workers in economic terms, reviewed the questionnaire with the subject, and left it with her/him to complete. The researcher would only explain the questions as posed, and provided no guidance to the respondents about how to make their judgments. That is, each respondent, after examining the questionnaire, independently generated a response using her/his own personal metric.

The day following the distribution of the questionnaires, the completed forms were gathered and collated. Three respondents were away from the organization at the time of the initial survery. Copies of the questionnaire were left with their supervisors for them to complete when they returned. All three completed the questionnaire and mailed them to the researcher.

Results

The main findings from the study are shown in Figure 1, which shows the mean dollar estimates given by supervisors and incumbents for workers at the 15th, 50th, and 85th performance percentiles. Supervisors ($\bar{X}$ = \$26,962) and incumbents ($\bar{X}$ = \$24,670) gave similar estimates for the dollar value of the 15th percentile workers (t=1.02; df= 29; $p > .05$). The estimates of the value of 50th percentile workers were less similar, with supervisors ($\bar{X}$ = \$75,923) providing a larger estimate than incumbents ($\bar{X}$ = \$65,511). The difference between estimates given by incumbents and supervisors for 50th percentile workers was signficant (t=2.88; df=29; $p < .05$). Finally, the greatest disparity between value estimates of incumbents and supervisors was for workers at the 85th percentile. Supervisors provided a mean estimate of $278,150 for these superior workers, while the mean estimates of the incumbents was $160,878. This difference was also significant (t=18.58; df=29; p <.05).

Estimates of SD_y can be obtained by subtracting the mean estimates given for the 50th percentile from those of the 85th percentile, and those of the 15th from the 50th. Such estimates for both supervisors and incumbents are given in Table 1. These data suggest that estimates of SD_y based on subtracting the 15th percentile from the 50th are similar for both supervisors and incumbents. However, estimates derived from subtracting the 50th from the 85th percentile are quite disparate, with supervisors providing a much larger estimates than incumbents. These estimates suggest that performance, as estimated by the dollar value of workers, is not distributed symmetrically around the mean (i.e., the 50th percentile).

Because of the apparent asymmetrical nature of these estimates, it could be argued that the mean does not provide the most accurate estimate of central tendency. Accordingly, median values for estimates by supervisors and incumbents of workers at the three performance percentiles were obtained. These medians are shown in Figure 2. The median values given are substantially less than corresponding mean values. For example, the median estimate of workers at the 85th percentile given by supervisors was $95,000, compared to a mean estimate of $278,150. Also, while there appears to be some difference between median estimates of supervisors and incumbents, especially for workers at the 15th and 85th percentiles, analysis by Mann-Whitney U test indicates that none of median estimates given differed significantly. (See Table 2 for summary of Mann-Whitney U test analysis.)

Discussion

The interpretation of the results of the current study depend on whether or not a parametric analysis is viewed as appropriate for these data. If it is felt that the assumptions required by the *t* test are met, then it may be concluded that supervisors and incumbents do not render similar judgments of the worth of workers (except at the 15th percentile), with supervisors providing significantly and substantially higher estimates of the dollar value of average and superior workers. If this is true, then it would it would likely be argued to accept the estimates of supervisors as being more valid, since they should have a more comprehensive picture of the role of workers in the organization as a whole and should be in a better position to compare the output of workers of different performance levels.

However, if it is felt that the assumptions required of parametric analysis are not met, then it would be concluded that supervisors and incumbents provide similar estimates of the value of workers at the performance levels tested. This would allow researchers to view estimates of incumbents with greater credibility, and make utility analysis more amenable to application to smaller organizations which lack the number of supervisors needed to provide accurate SD_y estimates.

If the parametric analysis is accepted, what can be made of the SD_y estimates obtained by the two groups? Hunter and Schmidt (1983) reported that most estimates of SD_y obtained using their procedure fall within a range of 40 to 70 percent of the mean salary of the job incumbents. This approach is known as the "Salary Percentage Technique," and provides a basis for evaluating the estimates obtained in the current study. Informatin on the actual salary of the incumbent sample in the present study was not obtained. However, their average GS rating was approximately a mid-level GS-11. The pay for this grade is $32,700 per year. Using the salary percentage technique, then, yields an expected value of SD_y of $13,080 to $22,890. Both of these values are substantially less than than SD_y estimates obtained by either the supervisors or incumbents in the current study. This disparity may be a reflection of the skewed nature of the distribution of performance estimates. Alternatively, the larger estimates obtained in the current study may be related to the nature of the job performed by these incumbents. Previous published studies of SD_y, with one exception (Eaton et al., 1985), have all examined workers from private enterprise and most have been blue-collar workers. Additionally, other researchers (e.g., Weekley et al, 1985) report that the salary percentage technique yields estimates of SD_y that are substantially less than those obtained by the original Schmidt et al. (1979) procedure.

Further research comparing estimates of supervisors and incumbents is needed. The population of workers sampled in the current study were highly educated (most with Masters degrees, many with doctorates) specialists in behavioral sciences. The current sample of incumbents had considerable experience at their jobs. Moreover, the job of task scientist is multifaceted and often involves regular interaction with supervisors on the planning and execution of research projects and related tasks. It may be that somewhat different results would be obtained if the sample used consisted of workers with less diverse jobs who were also less educated.

It may also be the case that the methods used in this study are not the most accurate way of generating the desired estimates. Tuttle et al. (1985) described a procedure based on group dynamics principles used to develop productivity indexes for organizations. This method could also be used to generate SD_y estimates. In the standard Schmidt et al. (1979) procedure, respondents produce their estimates independently, and performance estimates are obtained by calculating group averages. A more structured approach using groups of supervisors *and* incumbents, utilizing methods described by Tuttle et al. might provide even more accurate estimates.

References

Boudreau, J. W. (1983a). Economic considerations in estimating the utility of human resources productivity improvement programs. *Personnel Psychology, 36*, 551-576.

Boudreau, J. W. (1983b). Effects of employee flow on utility analysis of human resources productivity improvement programs. *Journal of Applied Fsychology, 68,* 396-406.

Boudreau, J. W., & Rynes, S. L. (1985). Role of recruitment in staffing utility analysis. *Journal of Applied Psychology, 70,*, 354-366.

Burke, M. J., & Frederick, J. T. (1986). A comparison of economic utility estimates for alterntive SD_y estimation procedures. *Journal of Applied Psychology, 71*, 334-339.

Cronbach, L. J., & Glaser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Champaign, Il: University of Illinois Press.

Eaton, N. K., Wing, H., & Mitchell, K. J. (1985). Alternative methods of estimating dollar value of performance. *Personnel Psychology, 38*, 27-40.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity. *American Psychologist*, *38*, 473-479.

Landy, F. J., Farr, J. L., & Jacobs, R. R. (1982). Utility concepts in performance measurement. *Organizational Behavior and Human Performance, 30,* 15-40.

Matthews, M. D., & Sewell, M. T. (1987). Application of utility models to the evaluation of military manpower and personnel research programs: A critical review and analysis. Brooks Air Force Base, TX: Air Force Human Resourses Laboratory, in press.

Murphey, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. Psychological Bulletin, 99, 133-138.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on work-force productivity. Personnel Psychology, 35, 333-347.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Tuttle, T. C., Wilkinson, R. E., & Matthews, M. D. (1985). Field test of a methodology for generating efficiency and effectiveness measures. AFHRL-TP-84-54. Brooks AFB, TX: Air Force Human Resources Laboratory.

Weekley, J. A., Frank, B., O'Connor, E. J., & Peters, L. H. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. Journal of Applied Psychology, 70, 122-126.

Table 1

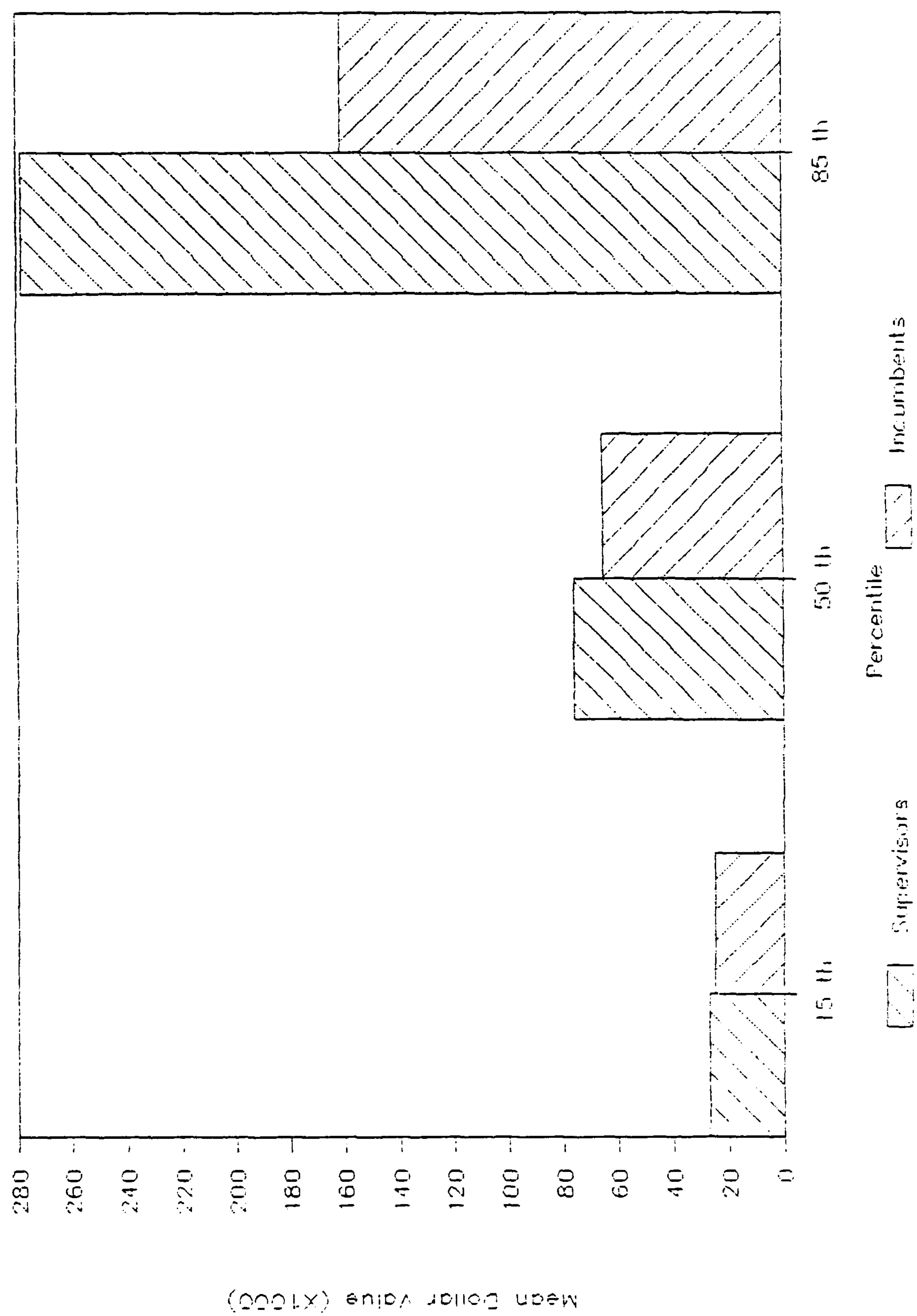
Estimates of SD_y Based on Supervisor's and Incumbent's Data

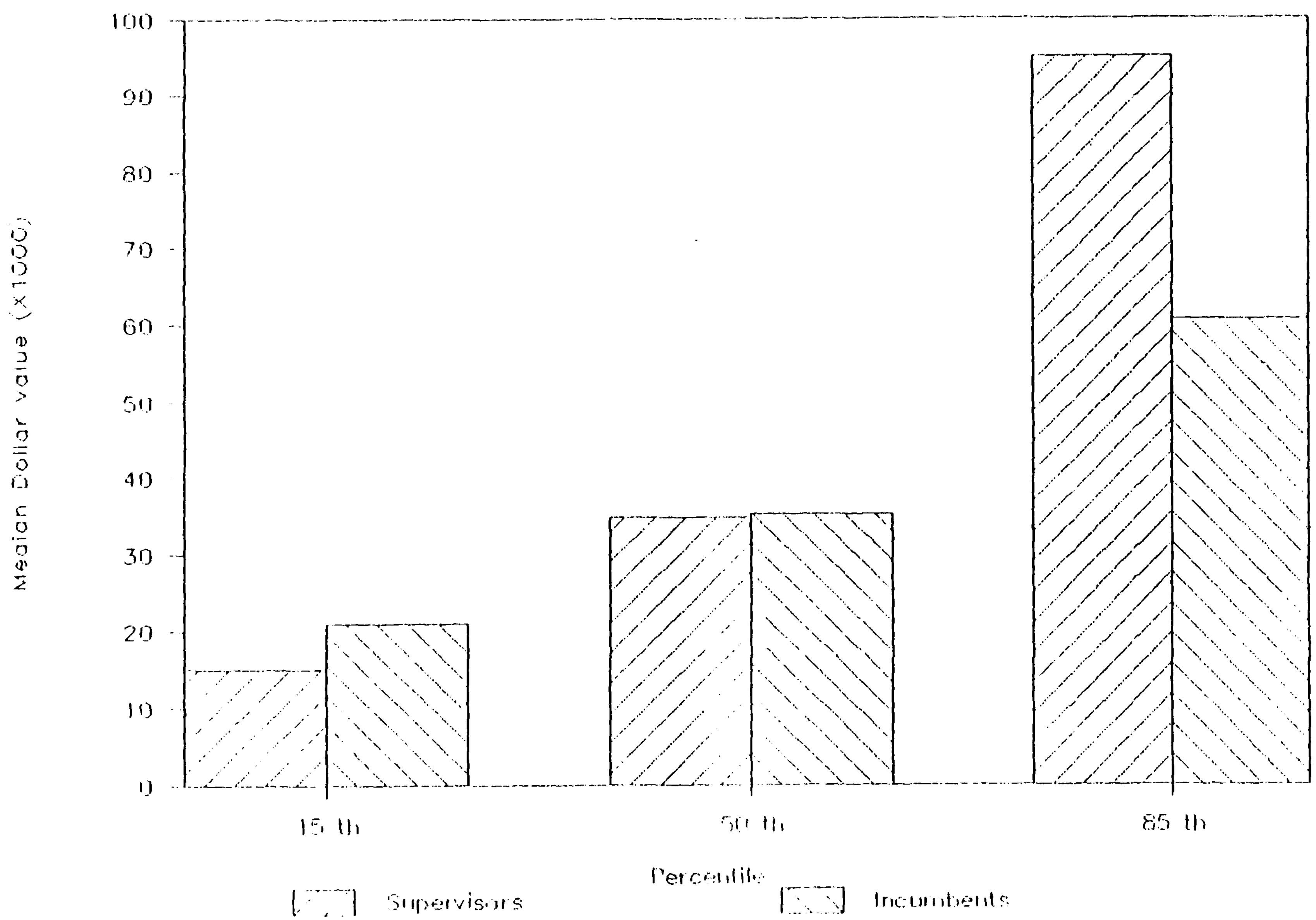
	85th - 50th	50th - 15th
Supervisors	$202,227	$48,961
Incumbents	$95,367	$40,841

Table 2

Results of Mann-Whitney U test analysis

	U	n_1	n_2	p
15th Percentile	114	13	18	.05
50th Percentile	120.5	13	18	.05
85th Percentile	87.5	13	18	.o5


Mean Dollar Value (x1000)
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
15 th
50 th
85 th
Percentile
Supervisors
Incumbents



Median Dollar Value (x1000)
100
90
80
70
60
50
40
30
20
10
0
15 th
50 th
85 th
Percentile
Supervisors
Incumbents

FINAL REPORT NUMBER 95
REPORT NOT AVAILABLE AT THIS TIME
Dr. Jorge Mendoza (1986)
760-6MG-136

1988 USAF-UES RESEARCH INITIATION PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

UNIVERSAL ENERGY SYSTEMS, INC.

Project Title: THE ROLE OF FOURIER DESCRIPTORS FOR SHAPE IN VISUAL FORM PERCEPTION

Principal Investigator: JOHN UHLARIK, PH.D.
DEPARTMENT OF PSYCHOLOGY
KANSAS STATE UNIVERSITY
MANHATTAN, KS 66506

Project Period: 1 JANUARY 1988 thru 31 DECEMBER 1988

Project Amount: $19,996

Contract Number: F49620-85-C-0013/SB5851-0360

ABSTRACT

Fourier Descriptors (FDs) have been used by investigators in computer pattern to discriminate and/or identify planar shapes. These descriptors provide for both the analysis and synthesis of closed curves by obtaining Fourier coefficients of chain-encoded contours. Discrimination thresholds were obtained for FD stimuli for shape in a manner similar to contrast sensitivity functions for grating stimuli. Adaptation to a specific FD harmonic frequency produced a selective decrease in sensitivity (i.e., increase in threshold) for the adapting frequency and closely related harmonic frequencies. These results suggest that FDs are an important and promising tool for research on human form perception.

THE ROLE OF FOURIER DESCRIPTORS FOR SHAPE IN VISUAL FORM PERCEPTION

INTRODUCTION

During the Summer of 1987 I participated in the Summer Faculty Research program at the Human Resources Laboratory at Williams AFB. Exploratory research was initiated that suggested that Fourier Descriptors (FDs) for shape were a promising avenue for investigating human form perception (Uhlarik, 1989). Using the classical psychophysical method of constant stimuli visual discrimination thresholds were obtained for FD stimuli in a manner analogous to contrast sensitivity functions (CSFs) that have been obtained for one-dimensional sine wave grating stimuli (cf., Ginsburg, 1984; Cornsweet, 1970). Furthermore, the initial research suggested that adaptation to a specific FD harmonic frequency (viz., the sixth) produced a selective threshold elevation for the adapting frequency and closely related even harmonic frequencies.

The RIP research extended the Summer Faculty research and attempted to determine the extent to which FDs for planar shape provide a relevant dimension by which the human visual system processes form. FDs represent transforms of boundary curvature that are popular in comuter pattern recognition because they are relatively easy to program and are backed by the well-developed Fourier mathematical theory (Pavlidis, 1980). These descriptors provide for both the analysis and synthesis of closed curves by obtaining Fourier coefficients of chain-encoded contours. Using this approach a curve is represented parametrically as a function of arc length by the accummulated change in the direction of the curve from the starting point. This function is expanded in a Fourier series and the coefficients are arranged in the amplitude/phase angle form (cf. Kuhl and Giardina, 1982; Zahn and Roskies, 1972). Thus any two-dimensional shape is fully described by its set of FDs; the low frequency terms provide the global properties of the shape and the addition of higher frequency terms fill in more and more local detail.

There is physiological evidence suggesting that neurons in striate cortex of infrahumans are selective for stimuli of specific boundary curvature (Schwartz, Desimone, Albright, and Gross, 1983). If features like FDs are involved in the coding of shape in the human visual system they might exhibit selective adaptation. The purpose of the present study was first to determine amplitude discrimination threshold functions for FD stimuli in a manner similar to those obtained for contrast sensitivity in the spatial frequency domain (Davidson, 1968), and second then determine if these thresholds for harmonic frequency can be selectively elevated using the adaptation paradigm in a manner similar to threshold elevation for spatial frequency (cf., Blakemore and Campbell, 1969).

The follow-on research was designed in part to remedy some procedural and methodological limitations of the preliminary research. For example there were some technical difficulties with "aliasing" when the FD images were presented on the computer display. This was alleviated by doubling both the image size and viewing distance. The net effect of this modification was that the displayed images had greater resolution and the visual angle subtended by the images remained the same. The second area of improvement has to do with the psychophysical method. The method of constant stimuli is very reliable and not prone to influence from bias factors. Thus, it was an ideal choice for the initial research. Once the basic parameters of the phenomenon were established the follow-on research utilized a more efficient modification of this method in order to facillitate data collection.

In addition to these methodological improvements, the purpose of the follow-on research was to extend the preliminary findings. Specifically, it was designed:

1.) to replicate the basic threshold function;

2.) to determine if the selective adaptation effects occurred for other harmonic frequencies;

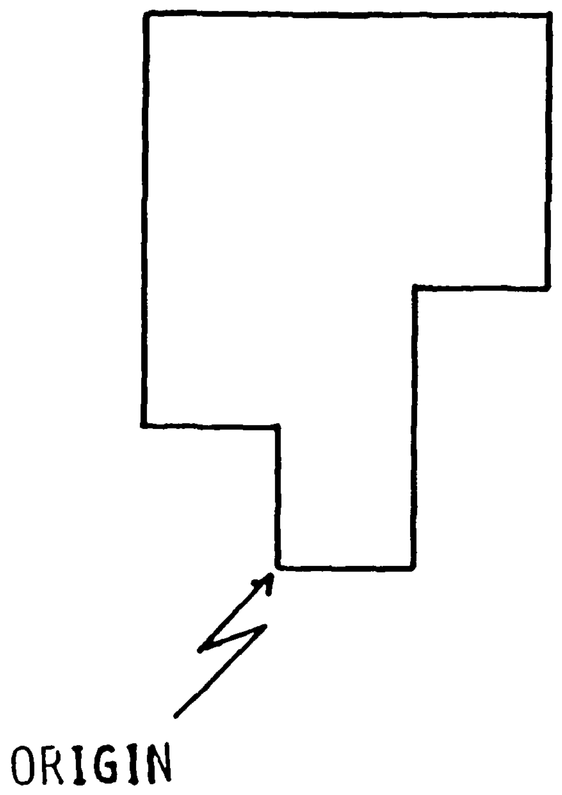
3.) to determine if the initial differences between the odd and even harmonic maintained (i.e., does adaptation to even harmonics influence only even harmonics and adaptation to odd harmonics influence only odd harmonics).

METHOD

Stimuli and Design

Stimulus patterns were generated using an algorithm developed by Kuhl and Giardina (1982). Using an arbitrarily chosen closed contour as a source (shown in Figure 1A), this shape was chain-encoded in order to obtain Fourier coefficients. Figure 1B shows the chain code link labels which approximates the continuous contour by a sequence of piecewise linear fits (cf., Freeman, 1974). The Fourier series representation is appropriate for the chain code because the code repeats on successive traversals of the contour, and each term in the Fourier expansion is a descriptor associated with a particular frequency (defined in units of cycles/perimeter), amplitude (defined as a square root of the sum of the squared coefficients), and relative phase angle which can be manipulated orthogonally. The resulting FDs are invariant with rotation, dialation, and translation of the source contour, but lose no information about the shape of the contour (Kuhl and Giardina, 1982).

Using a modified version of the psychophysical method of constant stimuli, observers were required to judge which of two FD shapes was the fundmental frequency. The fundamental was consider the <u>standard</u> in the method of constant



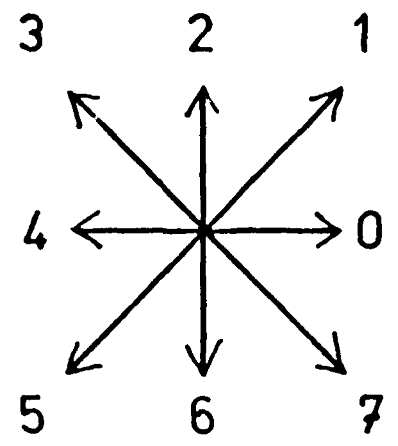


Figure 1. Figure 1A (on left) shows the source figure used to derive the FD stimuli. Figure 1B (on right) shows the eight standardized line segments used to chain-encode the source figure. Thus from the point of origin the chain code for this particular source figure would be: 2, 2, 2, 2, 2, 4, 4, 4, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 4, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 4, 4, 4, 4, 4.

stimuli and was presented on every trial. This first-order harmonic was always an ellipse and was always assoicated with a fixed amplitude which determined its size. This ellipse shared only the most global shape and size of the source shape shown in Figure 1A. Comparison figures consisted of the sum (Fourier synthesis) of the fundamental and another higher-order harmonic frequency that could be varied in FD amplitude. As the amplitude of the higher-order component increased it became more discriminable in that features emerged that made it less like the elliptical shape of the standard.

A trial consisted of presenting a pair of high amplitude adapting stimuli for a period of 2 s, followed by a noise screen for 1 s, and then the target stimuli. The target stimuli consisted of simultaneous presentation of the standard and comparison for 150 ms which were followed by the noise screen. On any given trial the standard (ellipse) could be either the leftward or rightward side of the pair, and the observer's task was to choose the side of the standard. The discrimination threshold for a given frequency was defined as that amplitude associated with the observer making the correct determination 50% of the time.

Eleven different higher-order, harmonic frequencies (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, & 12) were used for the comparison figures, and each comparison figure could be associated with each of eight different harmonic amplitudes (.05, .10, .15, .20, .25, .30, .35, &.40). Figure 2 illustrates examples of some of the harmonic frequencies used for comparison stimuli; this figure shows each frequency at three different levels of harmonic amplitude.

Four different frequencies (1, 3, 5, 8,) were used for the adapting stimuli. For a given adaptation frequency, the adapting stimuli were always an identical pair of shapes. When the adapting frequency was the first harmonic the pair were both ellipses identical to the standard target figure. It was presumed that using the fundamental frequency for adapting stimuli would not

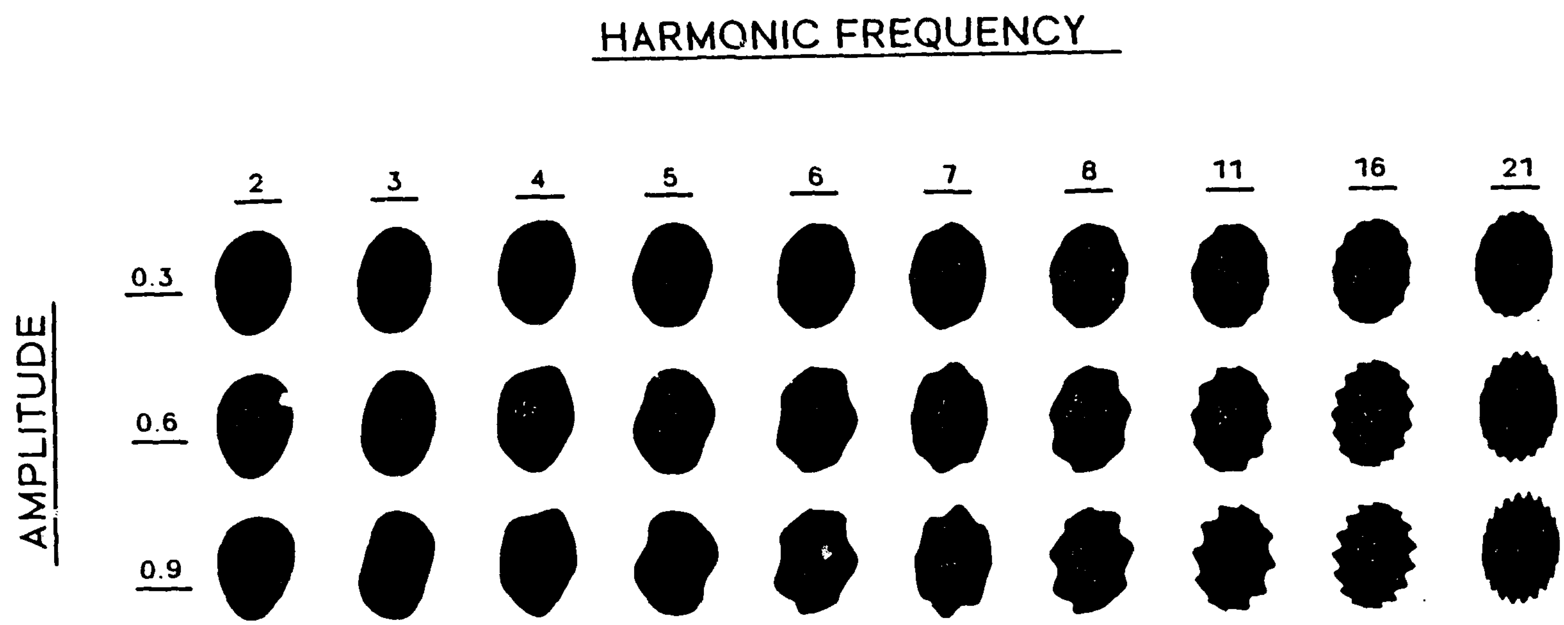


Figure 2. Examples of the "comparison" figures used as the independent variables for harmonic frequency and amplitude. Individual figures consisted of the fundamental harmonic frequency summed with a higher-order harmonic frequency. Three levels of FD amplitude are illustrated for each frequency.

exert any systematic effects of discrimination of the higher-order harmonic frequencies. To the extent that this fundamental adaptation condition replicates the basic threshold functions from previous research using no adaptation stimuli (Uhlarik, 1989), it can provide a baseline against which the effects of selective adaptation to higher order harmonic frequencies can be compared. When the third, fifth or eighth harmonics was used for adaptation, the figures were well above the discrimination threshold in that the adapting stimuli were composed of the fourier synthesis of the fundamental and a higher-order harmonic frequency with an amplitude of 1.5 which was almost four times that of the highest amplitude comparison stimulus.

The harmonic frequencies of the adapting and comparison stimuli were constant for each block of 80 trials. Within each block, the amplitude of higher-order harmonic component of the comparison figure varied so that each of the eight levels of amplitude (.05, .10, .15, .20, .25, .30, .35, .40) was presented ten times. Within each block the presentation order of these amplitudes was random as was whether the standard was the left or right side of the target pair.

The stimuli were generated and displayed on an IBM-AT style computer using an Orchid Turbo PGA graphic system with horizontal resolution of 640 pixels and a vertical resolution of 480 pixels. The stimuli were displayed on a NEC Multisync monitor (model #JC-1501VMA). A small fixation cross was located in the center of the display. The standard and comparison figures were filled closed contours that were centered about a point .75 deg to the left and right of fixation; each figure subtended .75 deg visual angle. On half of the trials the standard figure was on the left side of each pair, and on the other half it was on the right.

Procedure

The observer viewed the display binocularly in a darkened room from a distance of 3.0 m after dard adapting for 10 min. The observers were instructed to fixate a small cross in the center of the screen and not to make gross eye movements until a trial was completed. A tone signaled the start of a trial. After the presentation of the stimuli the observer indicated the side of the standard by pressing one of two button switches. Auditory feedback indicated whether the choice was correct or incorrect.

The noise screen consisting of a random dot pattern (50% of the pixels illuminated) was always present on the display, except during presentation of target and adapting stimuli. The purpose of the noise screen was to eliminate persistence of target and adapting stimuli and to reduce contour masking. The space average luminance of the noise screen was 0.2 log ft lamberts; the luminance of the adapting and target figures was 0.62 log ft lamberts.

Four male observers participated in the experiment. One had emtropic vision and the others had vision corrected to at least 20/30 for the viewing conditions involved in the experiment.

RESULTS

In any given block of 80 trials the higher-order component of the comparison could be one of eight different FD amplitudes (viz., .05, .10, .15, .20, .25, .30, .35, or .40) and each amplitude was presented ten times. The amplitude discrimination threshold was defined as that FD amplitude that allowed the standard to be correctly choosed 50% of the time. Figure 3 shows these amplitude discrimination thresholds for harmonic frequencies 2 thru 12 averaged over observers. Note that the vertical axes are reversed so that the resulting functions represent relative sensitivity rather than thresholds *per se*. The dashed curves in each of the three panels represent the baseline conditions in

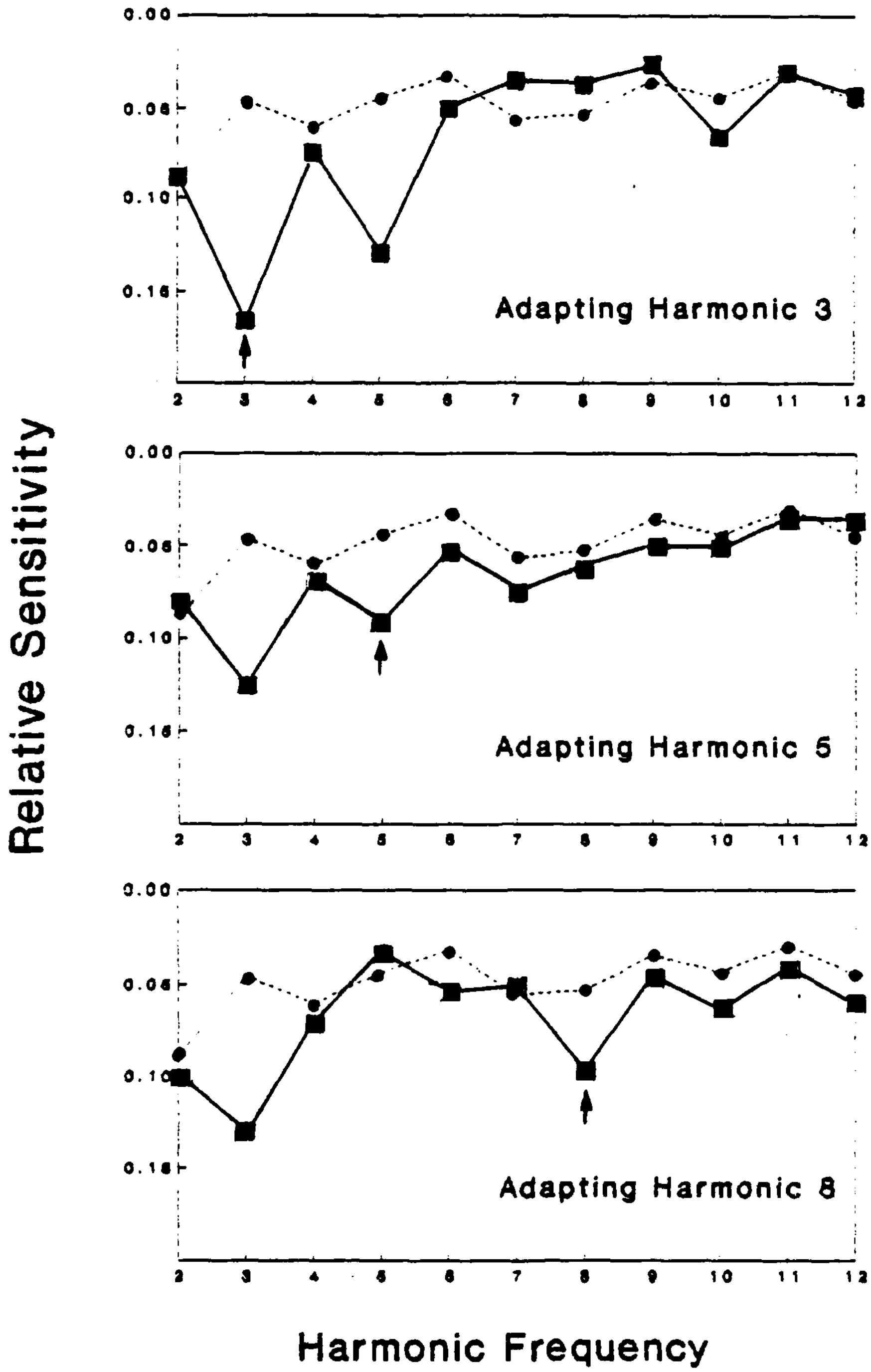


Figure 3. FD amplitude discrimination thresholds for the adapting harmonics 3, 5, and 8 (solid lines) relative to when the first (funamental) harmonic was used for adaptation (dashed line). Note the vertical axes are reversed so these functions represent relative sensitivities.

which the pair of adapting stimuli consisted of fundamental (or first order) harmonics and was based on data from all five observers. The relatively smooth and flat shape of these functions indicated that the fundamental did not selectively reduce sensitivity (i.e., elevate threshold) at any of the higher order harmonics, these functions are similar to the basic threshold functions involving no adapting stimuli found in the preliminary research (Uhlarik, 1989). On the other hand, adaptation to the third harmonic (solid curve in upper panel of Figure 3) produced a decrease in sensitivity for discrimination of the third harmonic and to a lesser extent the fifth harmonic. Similarly, adaptation to the fifth (solid curve in middle panel) and eight harmonics (solid curve in lower panel) produced analogous selective depressions in discrimination sensitivity. In both of these two latter cases there was also a marked depression in sensitivity for discrimination of the third harmonic. The "adaptating harmonic 3" curve was based on data from two observers and the other two adapting conditions were based on data from three observers. Thus, the dashed curves in Figure 3 are conceptually similar to contrast sensitivity functions (CSFs) that have been obtained for sine wave grating stimuli in the spatial domain, and the selective depression in sensitivity due to adaptation to a specific higher order harmonic is similar to CSF adaptation studies where spatial frequency rather than harmonic frequency is the independent variable (cf., Blakemore and Campbell, 1969). In both cases it is presumed that selective adaptation produces a temporary loss of sensitivity of those visual neurons responsible for processing particular parameters of the adapting stimulus.

Figure 4 replots the data in Figure 3 in terms of mean threshold increase. For each harmonic frequency this was obtained by subtracting the baseline condition in which the adapting stimuli consisted of fundamental harmonics (dashed curved in Figure 3) from the adaptation conditions involving higher order

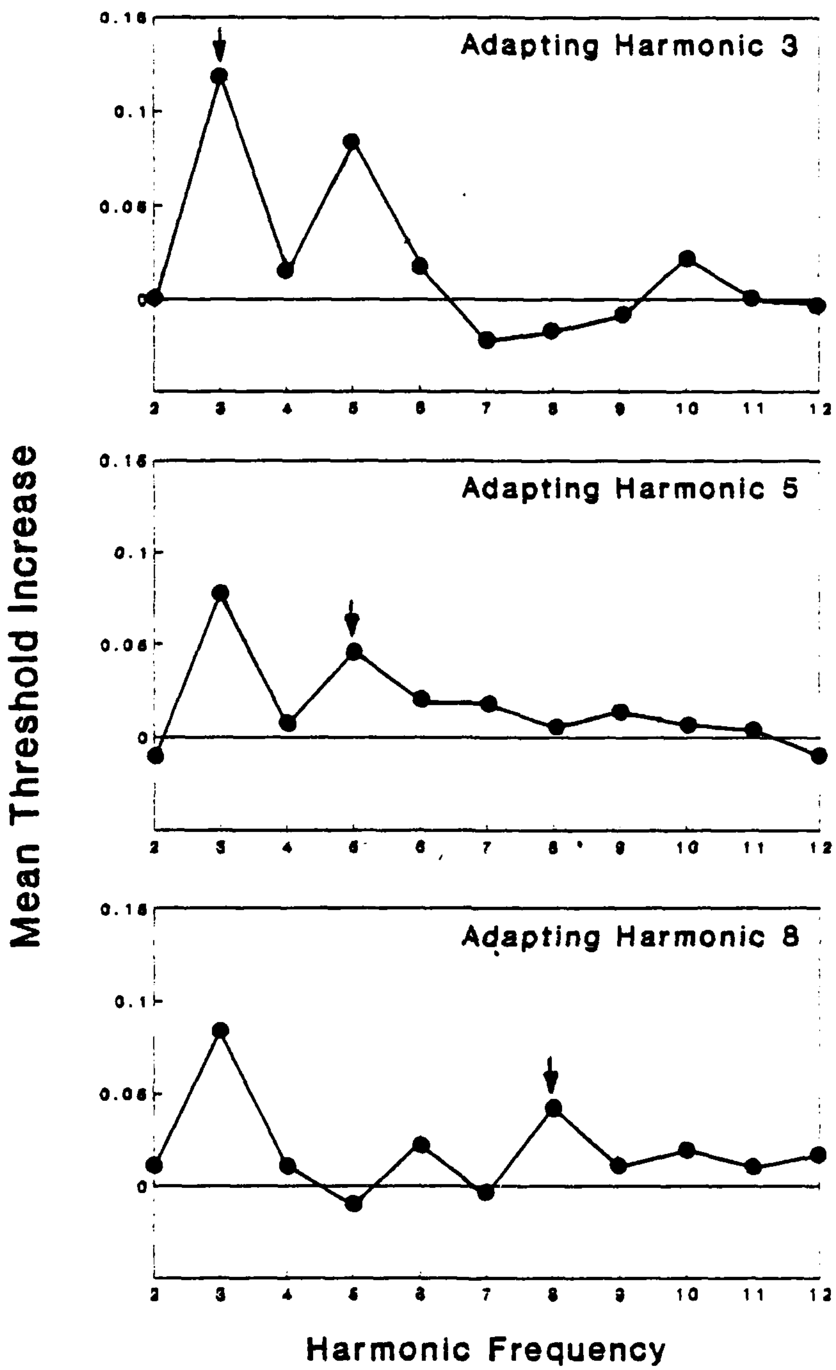


Figure 4. Increase in amplitude discrimination thresholds as a function of harmonic frequency subsequent to adaptation to the 3rd, 5th, and 8th harmonic frequencies.

harmonics (solid curves in Figure 3). Figure 4 shows that for all three adapting frequencies there was marked increases in discrimination thresholds for the harmonic frequency of the adapting stimuli. In the case of "adapting harmonic 3" this effect spread to the adjacent odd harmonic (i.e., the fifth), and for "adapting harmonic 5" the effect was manifested for the third harmonic. In the case of "adapting harmonic 8" there were small peaks for the sixth and tenth harmonics. Inexplicitedly, however, this latter condition also manifested a large effect at the third harmonic. In fact, all three panels in Figure 4 show large systematic elevations in thresholds for discriminations involving the third harmonic regardless of the adapting harmonic. At the present time it is not clear why this should occur.

DISCUSSION

The results of the present research suggests that FDs are an important tool for research on human form perception. A persistent problem in the study of form perception and pattern recognition has been the lack of a complete and reliable measure of the physical stimulus. Voluminous research has addressed this issue but reviews indicate that this fundamental issue has never been resolved (e.g., Triesman, 1987; Brown and Owen, 1965; Michels and Zusne, 1965). Fourier descriptors have the potential to provide such a metric because they are powerful and efficient algorithm for representing and classifying shapes. These results also suggest that to the extent that the adaptation paradigm provides an indication of neural coding, the finding that prolonged exposure to a specific harmonic frequency elevates threshold in the region of that frequency suggests the processing of form in the human visual system may involve the decomposition of shape into FDs related to boundary curvature.

In addition to having important implications regarding the afore mentioned basic research issues in human visual perception, examining the role of FD stimuli in form perception has direct relevance to the USAF in terms of potential for evaluating display systems. A major problem with creating performance related metrics for visual displays has been the lack of "analytic throughput" (Ginsburg, 1983). In this context analytic throughput refers to the ability to specify target information in the same language used to specify system capability, visual processes, and performance metrics. As in the case for contrast sensitivity, FDs for shape have promise because of their potential to create unified performance-based metrics for evaluating display systems.

Further research is needed to determine the extent to which the present findings generalize to other viewing conditions (target duration, retinal location, etc.) and when other more complex source figures are used to generate the FD stimuli. It would also be important to explore the basis of the systematic and pervasive effects of adaptation on perception of the third harmonic frequency.

REFERENCES

Blakemore, C., and Campbell, F.W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. Journal of Physiology, 203, 237-260.

Brown, D.R., and Owen, D.H. (1967). The metrics of visual form: Methodological dyspepsia. Psychological Bulletin, 68, 243-259.

Cornsweet, T.N. (1970). Visual Perception. New York: Academic Press.

Davidson, M.L. (1968). Pertubation approach to spatial brightness interaction in human vision. Journal of the Optical Society of America, 58, 1300-1309.

Freeman, H. (1974). Computer processing of line drawing images. Computing Survey, 6, 57-97.

Ginsburg, A.P., and Cannon, M.W. (1983). Comparisons of three methods for rapid determination of threshold contrast sensitivity. Investigations in Ophthalmological and Visual Sciences, 24, 798-802.

Kuhl, F.P., and Giardina, C.R. (1982). Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing, 18, 230-258.

Michels, K.M., and Zusne, L. (1965). Metrics of visual form. Psychological Bulletin, 63, 74-86.

Pavilidis, T. (1980). Algorithms for shape analysis of contours and waveforms. Transactions on Pattern Analysis and Machine Intelligence, 2, 301-312.

Schwartz, E.L., Desimore, R., Albright, T.D., and Gross, C.G. (1983). Shape recognition and inferior temporal neurons. Proceedings of National Academy of Sciences, 80, 5776-5778.

Treisman, A. (1986). Properties, parts and objects. In K.R. Boff, L. Kaufman and J.P. Thomas Handbook of Visual Perception, Vol. II, New York: John Wiley & Sons.

Uhlarik, J. (1989). Fourier descriptors for shape: Effects on adaptation on discrimination thresholds. Bulletin of the Psychonomic Society, (in press).

Zahn, C.T., and Roskies, R.Z. (1972). Fourier descriptors for plane closed curves. IEEE Transactions on Computers, C-21, 281-296.

Department/Project No.: 0210507-414
Project Description: Comprehensibility of Technical Text
Project Director: Dr. Doris J. Walker-Dalhouse

INTRODUCTION

The following document is being submitted as a final report. The delay in the submission of the final report for AFOSR Contract No. F49620-85-C-0013/SB5851-0360, Subcontract number S-760-6MG-080 is due to difficulties resulting from my decision to leave Jackson State University (JSU) before the end of the December 31, 1987 granting period.

After deciding to leave JSU in July 1987 for employment at Moorhead State University (MSU) a request was made to Universal Energy Systems (UES) by Dr. James A. Hefner, at my request, transfer the balance of the grant to MSU. I received a copy of Dr. Hefner's letter, but never received any official notification from UES relative to any action being taken toward the transfer request.

After numerous calls to both JSU and UES to inquire about the status of Dr. Hefner's request, I was told that a no cost time extension would be given while the request for transfer was being considered. I was, however, unable to pay for data analysis, consultant services, and prepare the data for dissemination due to the lack of access to grant funds. Although Jackson State was notified by UES that the transfer of the grant had been approved (which I learned about through telephone coorespondence with UES personnel) in a letter dated June 17, 1988, they did not respond to verbal and a written requests from me concerning the procedures for expending funds. Through the assistance of UES personnel, I did receive a purchase order from JSU on November 18, 1988 to handle billing until December 31, 1988. In spite of the fact that the Air Force extension given verbally amounted to a year, in actuality, I was only given funds to complete data analysis and consultation for approximately forty-three days. The brevity of this report reflects the extent of data analysis that was permitted during the time frame allowed.

PROCEDURES

Subjects

Sixty (60) students from two introductory computer science classes at Jackson State University were used as subjects. Data gathering was done in two sessions a week apart. Only thirty-six (36) students showed for the second data gathering session, thus complete data was collected from 36 of the 60 students. Thirty-four (34) of the original sample were males and twenty-six (26) were females. Of the subjects for which complete data was available nineteen (19) were female and seventeen (17) were males. A total of eightteen (18) subjects were contained in the group that received the original passage (control group) and eighteen in the group that received the rewritten passage (experimental group).

Methods

The principal investigator met with both classes of students enrolled in the introductory computer science classes and explained the purpose of the study. The classes were taught by different instructors, one male and one female. During the first data gathering session subjects were given two cloze passages selected from a US Air Force technical training manual for information systems programming training.

During the second data gathering session, one half of the subjects in each class were given envelopes containing a passage whose content dealt with general purpose registers. This passage was designated as the original text passage because it was taken verbatim from a United States Air Force information systems training manual.

The other half of the subjects were given a rewritten version of the original passage in which an attempt was made to simplify the vocabulary and sentence structure. (See Appendix A). The intent was to provide the reader with background information about the topic thereby reducing the amount of information to be retrieved from long-term memory. The subjects were instructed to read the passage and upon completion to replace the passage in the envelope. After the passage was replaced in the envelope they were asked to respond to a comprehension test by identifying statements as representing "old"(verbatim statements from the original passage or paraphrases thereof) or "new"information

(information completely false or slightly altered in meaning.)

Phase 2 of the study which was to involve the students in recalling, in writing, the main points contained in the passages was not completed due to problems encountered in scheduling another data collection session prior to final exams.

Instumentation

Two cloze tests were constructed from the text of a USAF training manual for information systems personnel. Every fifth word was deleted from passages of approximately two hundred and fifty words in length. The first and last sentences were left intact. Students were given the passages and asked to fill in the missing words. Only exact replacements were accepted as correct in determining reading levels. The following criteria were applied in ascertaining reading levels and designating readers as "good", "average", or "poor" readers (Vacca, Vacca & Gove, 1987):

Accuracy in word replacement	Reading Level
Above 60 %	Independent
40 - 60 %	Instructional
Below 40%	Frustration

A third passage pertaining to central hardware registers was selected from a different USAF information systems training manual for propositional analysis. Dr. Belita Gordon's services were used in analyzing the number of propositions found in the passage on general purpose registers.

A sentence verification test (Royer, Hastings & Hook, 1979) consisting of fourteen (14) statements was constructed as a measure of comprehension of the original and re-written passages. Test sentences were designated as containing "old" information original statements or paraphrases of original statements contained in the text, or "new" information, distracting statements or ones in which a meaning change occurred. See Appendix B for a copy of this test.

Results

The results of the two cloze tests are in Table 1.

Insert Table 1 about here

It reveals that the passage used in cloze test 1 was on the instructional reading level for subjects receiving the verbatim passage (41.27) and those receiving the rewritten passage (47.6) The passage used in the second cloze test represented the frustration level for both groups, 30.33 and 29 respectively.

Based upon the following cut-off scores specified by Vacca, Vacca & Gove, 1987; above 60% comprehension - independent (above average readers who can read material without assistance); 40% - 60% comprehension - instructional (average readers who will find the material challenging if given some form of guidance in reading), and below 40% comprehension - frustration (poor readers who need considerable guidance to understand the material or easier material) the students were classified as "g" (good readers), "a" (average readers)and "p" (poor readers). The scores on the first cloze test show four subjects as good readers, eightteen (18) as average readers, and fourteen (14) as poor readers. Cloze test 2 shows the majority of subjects as poor readers (28) with eight as average and none as good readers. The average of both cloze test scores, shows no good readers, eighteen (18) average, and eighteen (18) poor readers. Because of the greater apparent difficulty of passage two, passage one was used as the basis for determining the subjects' reading level. This permitted the placement of subjects in all three reading categories.

On the basis of cloze test one data, four of the subjects in the original passage group were classified as good readers and one subject in the rewritten passage group. Seven subjects in the original passage group and twelve subjects in the rewritten passage group were classified as average readers, and eight subjects in the original passage group and five in the rewritten passage group were classified as poor readers.

Table 2 shows the mean comprehension scores for good, average, and poor readers based on the sentence verification test which was administered after exposure to their respective passage.

Insert Table 2 about here

No meaningful comparisons can be made between the two groups of subjects classified as good readers because the rewritten passage group had only one subject in that category. Since there was only one subject in the good readers category of the experimental group, and three in the good readers category of the control group, the subjectd in the good and average readers groups were combined and an analysis of variance was performed on the data. No significant main or interaction effects were found. However, average readers did slightly better with the rewritten passage than with the original passage, 74.3% and 73.3% respectively. Poor readers also did better with the rewritten passage than with the original passage, 75.6% and 71.4% respectively. (See Table 3)

Insert Table 3 about here

Although great caution is advised in drawing conclusions from nonsignificant data, it would seem that the poorer the subjects' reading ability, the better they did with the rewritten passage. These findings seem to be consistent with general findings that poor readers are deficient in metacognitive skills which allow the reader to employ strategies during reading that facilitate understanding of the text. Consequently, assistance or instruction is needed in using or understanding textual aids, text organization, writing style, and/or vocabulary.

Discussion

The present study attempted to answer the following questions:

1. Will the reading comprehension of good readers differ significantly from that of average or poor readers after reading an original version of a technical passage taken from a US Air Force Information Systems training manual in one of four select training areas?

2. Will the reading comprehension of good readers differ significantly from that of average or poor readers after reading a rewritten technical passage. The rewritten passage will be simplified based on information obtained from the Kinstch and van dijK model of comprehension.

The proposed research was initiated to provide data to achieve the following objectives:

1. To investigate the comprehensibility and theoretical text cohesion of Air Force technical training manuals.

2. To rewrite technical text by attempting to reduce the number of text propositions, and determining the subsequent comprehension levels of good, average, and poor readers on a test of comprehension using the sentence verification technique (SVT).

3. To suggest instructional strategies for improving the comprehensibility of technical prose.

It is possible that the sample size and inaccessibility of the subjects for further testing may have been responsible for the inability to find significant differences between the groups before any definitive conclusions can be drawn. Additional passages should be analyzed, rewritten and administered to a larger sample of subjects. Although no significant differences were found, the differences noted were in favor of the group receiving the rewritten passage.

It is also possible that if other techniques such as word frequency count (American Heritage word frequency list), discourse pointers, ensuring passage cohesiveness by employing strongly connected propositions, are employed to enhance text comprehensibility of the rewritten passage, significant differences may be observed between the groups. (Zakaluk & Samuels, 1988). Other techniques that have been reported to foster understanding and comprehension of technical passages that could be used in a replication of this study are the use of graphic aids such as tables, charts, and diagrams to elucidate main points and teaching readers to use semantic webs (Johnson, 1987), a graphic aid which consists of the following aspects:

1. a core - question which serves as the focal part of the web.

2. web strands - subordinate or secondary ideas associated or related to each web strand.

3. strand supports - the details, inferences, and generalizations associated or related to each web strand.

4. strand ties - connections made by students between and among strands.

Structured overviews, another graphic arrangement to show connections between terms, have been found to be effective in guiding interactions with text. Repeated readings of narrative/ story type text has been found to affect comprehension.(Taylor et. al, 1985; O'Shea et al., 1985) Guided repeated readings of text for unspecified number of times in which readers are encouraged to read for specific purposes or to mentally answer questions about information expected to be answered in the content is an additional strategy worthy of consideration for improving text comprehensibility.

Other suggested strategies for promoting the comprehension of technical text include a wriiten preview, or advance organizer, of ideas covered in a segment of technical text and in which clearly stated links are made between prior knowledge or experiences and the content of the text.

Finally, it is also possible that no significant differences were observed between the groups because the material in the USAF technical manual used was so well written that it was difficult to make the rewritten passage easier to understand.

References

Johnson, K.L. Improving reading comprehension through pre-reading and post-reading excercises. Reading Improvement, vol. 24, no.2, 1987, 81-83.

O'Shea, Lawrence J., Paul T. Sindelar and Dorothy J. O'Shea. "The Effects of repeated readings and attentional cues on reading fluency and comprehension." Journal of Reading Behavior, vol. 17, no.2, 1985, 129-42.

Royer, J.M. , C.N. Hastings, and C. Hook. A sentence verification technique for measuring reading comprehension. Journal of Reading Behavior, vol. 11, 1979, 355-363.

Taylor, Nancy, Margaret R. Wade and Frank R. Yekovich. "The effects of text manipulation and multiple reading strategies on the reading performance of good and poor readers." Reading Research Quarterly, vol. 20, Fall 1985, 566-74.

Vacca, Jo Anne, Richard T. Vacca and Mary Gove. Reading and Learning to Read. Boston: Little, Brown and Co., 1987.

Zakaluk, Beverley and S. Jay Samuels. Readability: Its Past, Present, & Future. Delaware: International Reading Association,1988.

Table 1 Cloze Tests Results & Reading Status

	Technical Passage						Rewritten Passage					
Ss#	C1	RS	C2	RS	Average Cloze	RS	C1	RS	C2	RS	Average Cloze	RS
01	48	A	40	A	44	A	30	P	24	P	27	P
02	28	P	34	P	31	P	48	A	28	P	38	P
03	28	P	60	A	44	A	60	A	20	P	40	A
04	29	P	24	P	26.5	P	58	A	24	P	41	A
05	56	A	30	P	43	A	42	A	30	P	36	P
06	32	P	18	P	25	P	60	A	42	A	51	P
07	48	A	34	P	41	A	52	A	34	P	43	A
08	52	A	36	P	44	A	34	P	24	P	29	P
09	52	A	32	P	42	A	58	A	40	A	49	A
010	62	G	44	A	53	A	52	A	28	P	40	A
011	24	P	16	P	20	P	34	P	20	P	27	P
012	22	P	12	P	17	P	62	G	40	A	51	A
013	64	G	44	A	54	A	36	P	34	P	35	P
014	16	P	08	P	12	P	54	A	34	P	44	A
015	42	A	14	P	28	P	48	A	20	P	34	P
016	48	A	46	A	47	A	54	A	26	P	40	A
017	30	P	14	P	22	P	42	A	24	P	33	P
018	62	G	40	P	51	A	34	P	30	P	32	P
Mean	41.27		30.33		35.80		47.6		29.0		38.33	

C1: Cloze Test One
C2: Cloze Test Two

RS: Reader Status
G: Good Reader
A: Average Reader
P: Poor Reader

Table 2 Results of Sentence Verification Test
Reading Status & Treatment

Reading Ability	AF Original Passage*	Rewritten Passage+
Good	010 64%	012 79%
	013 86%	
	018 93%	
Mean	81%	
Average	01 79%	02 57%
	05 78%	03 93%
	07 71%	04 79%
	08 86%	06 86%
	09 78%	07 79%
	015 78%	09 64%
	016 43%	05 64%
		010 93%
		014 64%
Mean	73.3%	015 71%
		016 78%
		017 64%
		74.3%
Poor	02 79%	01 79%
	03 57%	08 93%
	04 64%	011 71%
	06 57%	013 57%
	011 93%	018 78%
	012 71%	
	014 64%	75.6%
	017 86%	
Mean	71.4%	

* Passage entitled "General Purpose Registers" taken from Information Systems Programming Specialist Manual, AFSC 49152, Vol. 2, Extension Course Institute, Air University.
\+ Rewritten version of the original passage

Table 3 Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	45.204	2	22.602	.132	.877
Passg	4.367	1	4.367	.026	.874
Rabil	35.176	1	35.176	.206	.653
2-Way Interactions	58.802	1	58.802	.344	.562
Passg Rabil	58.802	1	58.802	.344	.562
Explained	104.006	3	34.669	.203	.894
Residual	5476.744	32	171.148		
Total	5580.749	35	159.450		

Appendix A

GENERAL PURPOSE REGISTERS

There are 16 hardware registers located in the central processing unit of the IBM System/360. These are known as the general purpose registers (GPRs), as opposed to the floating-point registers. The GPRs are used in fixed-point arithmetic operations, addressing, data movement, and as intermediate storage areas. Because these registers are hardware devices, all operations between these registers or between registers and the arithmetic unit are approximately four times faster than similar operations in main memory. The registers are numbered 0 through 15, or 0 through F in HEX notation. For normal use in addressing, arithmetic operations, etc., you will use registers 1 through 12. Registers 0, 13, 14, and 15 are used for special applications and by the operating system. Figure 1-3 shows the layout of information in the GPRs. Each register holds 32 bits of information. Note that the bits are numbered 0 through 31, where the 0 bit is also the sign bit. The 0 bit is called the high-order position of the register, and bit 31 is the low-order position of the register. Of course, 32 bits can also be represented as a 8 hexadecimal digits. The largest positive value that can be represented in a register is $2^{31} - 1$, and the smallest negative value is -2^{31}. At this point, there is one rule that you should keep in mind when programming in assembler language on a machine with hardware registers: use the registers whenever possible. This is the most efficient method of programming because you gain speed and save core storage space.

GENERAL PURPOSE REGISTERS

The IBM 360 is one type of computer hardward system. It contains 16 hardware registers known as general purpose registers (GPRs) in its central processing unit (CPU). The GPRs are used for many operations. Because they are used in fixed-point arithmetic operations, addressing and data movement, they are different from floating point registers.* GPRs are hardware devices. This means all operations between these registers and between registers and the arithmetic unit located in the central processing unit are fast. They are about four times faster than operators like them in main memory.

Registers are numbered 0 through 15 or 0 through F in hexadecimal (HEX) notation. A hexidecimal digit is made up of four bits and can range from zero (0) to 15 (F).

Addressing or performing arithmetic operations are normal operations for the computer. Register 1 through 12 are used for normal operations. Registers 0, 13, 14 and 15 are used for special uses and by the operating system. A diagram would help you to understand the information found in the GPRs. Thirty-two (32) bits (smallest pieces) of information are held in the registers of all computers. Positive and negative values can be shown in a register. The largest positive value is $2^{31}-1$ and the smallest negative value is -2^{31}.

Practice helps when using assembler language to program the IBM System 360 or any computer with hardware

*(where the computer must keep track of the point in each numerical value

registers Assembler language is the most productive way of programming for two reasons. It lets you gain speed and save space in core storage, or internal unit made of magnetic cores.

Appendix B

COMPREHENSION CHECK 1

Old New 1. Every computer has a input output unit containing 16 hardware devices.

Old New 2. General purpose registers (GPRs) are hardware internal computer components that can store data and accept data rapidly in the IBM 360.

Old New 3. GPR's differ from floating-point registers in the arithmetic operations which they perform.

Old New 4. The Central Processing Unit of the IBM 360 has 16 hardware register.

Old New 5. GPR's peforms operations between them or between registers and arithmetic unit slower than similar operations in main memory because they are hardware devices.

Old New 6. General Purpose Registers are used in fixed-point arithmetic operations, addressing, data movement, and as intermediate storage areas.

Old New 7. A hexidecimal digit is made up of four bits of information.

Old New 8. For addressing, arithmetic operations, and data movement, registers 0, 13, 14, 15 should be used.

Old New 9. Thirty-two bits are equivalent to 4 hexadecimal digits.

Old New 10. Thirty-one (31) is the high-order position of the register while 0 is the low-order position of the register.

Old New 11. $2^{31}-1$ and -2^{31} are the largest positive and smallest negative values found in a register.

Old New 12. The 0 register and the three highest registers are used randomly by the programmer when compared to the other registers.

Old New 13. Registers are numbered 0 through 15, or 0 through F in HEX notation.

Old New 14. Gaining speed and saving core storage space are advantages which result from using assembler language for programming on a machine with hardware registers.

Name: ______________________ SS#: ____________

Test No: ______________

FINAL REPORT NUMBER 98
REPORT NOT AVAILABLE AT THIS TIME
Dr. Charles Wells
760-7MG-046

1986 USAF-UES SUMMER RESEARCH PROGRAM

MINI-GRANT

(CONTRACT # F49620-85-C-0013/SB 5851-0360)

SPONSORED BY THE
AIR FORCE OFFICE OF SCIENTIFIC SERVICES
BOLLING AFB, D.C.

CONDUCTED BY THE
UNIVERSAL ENERGY SYSTEMS

FINAL REPORT

MECHANISMS OF CONTRAST AND LIGHTNESS CONSTANCY

PREPARED BY: BILL R. WOOTEN
ACADEMIC RANK: PROFESSOR
DEPARTMENT: PSYCHOLOGY
UNIVERSITY: BROWN UNIVERSITY
LOCATION: 89 WATERMAN STREET
PROVIDENCE, RI 02912

Preliminary Note: This research required an achromatizing lens that corrects for the chromatic aberration of the eye. It was delayed because the company making the lens (Applied Physics Laboratory) encountered technical problems. They were finally solved and the lens was delivered. The work reported here is the summary to date that was made possible by the mini-grant that allowed us to buy the lens and conduct the research.

The research is concerned with the problem of lightness constancy. In keeping with Beck's (1972) definition, I define lightness constancy as the invariance of color perception which occurs with changes in the intensity of the illumination. Historically, there are two theoretical approaches to the problem of lightness constancy. The first approach suggests that a constant perception of color is subserved by sensory mechanisms which encode luminance ratios for adjacent regions across the visual field. This theoretical approach has been most clearly articulated by Wallach's (1948) ratio principle. The second theoretical position follows from Helmholtz's notion of "discounting the illumination". From a mechanistic point of view, this approach is perhaps best represented by Helson's adaptation level theory (Helson, 1943) and by current computational approaches in which the goal of the system is to disentangle illumination from reflectance information (e.g., Horn, 1974). More cognitive approaches to this kind of explanation are suggested by Beck's attempt to isolate cues used to identify the illuminant (Beck, 1961).

Early psychophysical studies took a predominantly sensory approach to the problem of constancy. Wallach (1948) showed that when the luminance ratio between a test region and it's surrounding field was held constant, matches to a comparison stimulus were invariant. Although several other studies also suggested a contrast explanation of constancy (Horeman, 1963; Heinemann, 1955; Leibowitz, et al., 1955; Fry and Alpern, 1953), several studies showed an independence between contrast effects and the tendency toward constancy (Jameson and Hurvich, 1961; Heinemann, 1955; Stevens and Stevens, 1960; Oyama, 1962; Kozaki, 1963). Jameson and Hurvich (1961), using a relatively complex stimulus, found diverging functions under ratio conditions; darker test regions became blacker as the luminance increased, lighter regions appeared whiter, and grey test regions on grey backgrounds tended to remain perceptually constant with increases in luminance. Although well-referenced, this study has been criticized for its small luminance range (Jacobsen and Gilchrist, preprint) and for its lack of reproducibility (Flock and Noguchi, 1970). However, due to variations in experimental conditions, these studies have been difficult to evaluate in relation to one another. The experimental conditions include variations in matching techniques, stimulus configurations, background luminance conditions, contrast and luminance ranges and instructions to the observers. Several of these studies have used binocular matching paradigms, thus introducing changes in adaptation level caused by shifting from standard to comparison stimuli. Heinemann's (1955) data suggest the importance of stimulus luminance and contrast range. His results support both Wallach's ratio principle and Jameson and Hurvich's diverging functions, depending upon the luminance level and degree of contrast of the stimuli. In relation to instructions, few of these experiments distinguish between lightness and brightness. Consequently, subjects may have matched according to either a dim to bright or a black to white dimension. Two recent studies (Arend and Reeves, 1986; Jacobsen and Gilchrist, 1988) emphasize this potential confound. Arend and Reeves (1986) report constancy for lightness in both center/surround and mondrian conditions, for brightness in mondrian conditions, and a failure of brightness constancy given a center/surround stimulus configuration. Jacobsen and Gilchrist found constancy for a black to white dimension but not for brightness.

Although these two studies suggest that achromatic stimuli vary along two interacting dimensions, they are inconclusive in relation to constancy. Arend and Reeves (1986) minimally describe the stimulus conditions for their lightness and brightness measures, do not determine the contrast relations in their mondrian conditions, and yet conclude that contrast relations have no effect on (color) constancy. Rather, they seem to suggest that constancy is a result of a cognitive or adaptive process serving to discount the illumination. In an effort to replicate Jameson and Hurvich (1961), Jacobsen and Gilchrist produced constant judgments of achromatic colors across a six log unit illumination range. However, it is not clear that they approached the contrast range used by Jameson and Hurvich, precluding the conditions within which Jameson and Hurvich showed a decrease in brightness with an increase in illumination.

The experiments I have complete thus far have identified simplified conditions under which constancy occurs. The second set of experiments will be a series of parametric studies to further identify the conditions underlying invariant achromatic color perception. Stimulus duration, adaptation level and higher contrast conditions will be independent variables. The conditions of the final experiment follow from a model of retinal encoding of luminance differences; it will measure constancy as a function of stimulus asynchrony.

<u>The First Series</u>:

The initial experiments were concerned with distinguishing between adaptation level and contrast explanations of constancy and included a reassessment of the conditions under which constancy occurs. In each experiment, stimuli consisted of a center/surround region presented in Maxwellian view. Eleven different contrast ratios were presented at six different luminance levels, producing 66 stimuli. The luminance varied across slightly more than 2.5 log units and the contrast ratios between center and surrounds varied between .08 (high contrast) and .90 (low contrast). Two reverse contrast conditions, in which the luminance of the center was higher than that of the surround, were also used. This is illustrated in Figure 1.

Experiment 1:

In the first experiment, each trial was presented for 300 msecs and subjects were asked to judge the percent blackness (along a dimension varying from white to black) of the center, test region. The average results from three subjects is shown in Figure 2. The results show constant judgements of blackness in the extreme low and high contrast conditions. However, in the mid-contrast range, subjects produced U-shaped functions in which judged blackness was high at low and high luminance levels and relatively low in the mid-luminance range. This is especially apparent for the 0.37 contrast ratio.

The observed constancy under the high and low contrast conditions is in keeping with a simple contrast explanation. In the highest contrast conditions, the induction from the surround to the center is so great as to produce a ceiling effect of nearly 100% blackness. In the low and reverse contrast conditions, the lack of contrast from the inducing field precludes the perception of blackness. Our results under these conditions are consistent with Wallach's ratio principle and are in contradiction to Jameson and Hurvich (1961).

On the other hand, the U-shaped functions apparent in the mid-contrast conditions are in disagreement with both Wallach's results and those of Jameson and Hurvich. Both of these studies used prolonged viewing conditions and we wondered if our results reflected the nature of transient responses. Secondly, the 300 msec and random presentation of stimuli precluded adaptation to the overall luminance level of the stimuli. The following two experiments were designed to address these possibilities.

Experiment 2:

The second experiment was an early attempt to test for interactions between adaptation level and perceived blackness. An adapting field was on at all times except immediately before (400 msecs), during (300 msecs) and immediately after (approx. 1500 msecs) presentation of the stimulus. Six different adapting fields were used, varying in 0.6 log unit steps across 3 log units. A contrast ratio of .37 was intermixed with distractors. This ratio was chosen because it showed the greatest departure from constancy in Experiment 1. At each adaptation level, our results showed an increase in perceived blackness as a function of increasing luminance. This suggests that adaptation level cannot account for the U-shaped function observed in Experiment 1.

Experiment 3:

The identical stimuli were presented in this experiment as in Experiment 1, however, we increased the stimulus duration and presented the stimuli in ascending order according to luminance level. The stimulus duration was 7 seconds, and the ratios were presented in random order within a particular luminance condition. This allowed the subjects to adapt, to some extent, to the luminance level of the particular stimulus. The results from this experiment are shown in Figure 3. These data represent the average from 3 subjects. The functions represent the best-fitting straight lines for each contrast ratio, and demonstrate a high degree of constancy in all conditions. Thus, for simplified stimuli, the conditions which produce a constant judgment are constant luminance ratios between adjacent regions and sustained viewing conditions. This suggests that adaptation to the overall luminance is required for constant judgments.

In this case, the results are consistent with Wallach's ratio principle, and in contradiction to Jameson and Hurvich's diverging functions. This may reflect a problem stated earlier, namely, that it is unclear as to whether subjects were judging lightness or brightness in the earlier studies. The following experiment was intended to differentiate between these two hypothetical dimensions of achromatic stimuli.

Experiment 4:

The stimuli were identical to those used in the previous experiment and were presented for 7 secs. However, in this experiment subjects were asked to judge the brightness of the test region, relative to a standard stimulus shown at the beginning of each block of trials. We anticipated increases in brightness with increases in luminance, despite constant ratios. The results are shown in Figure 4.

As expected, the conditions in which the test region was higher in luminance than the surround show the highest degree of constancy. Most notably, the test regions for these ratios increase in brightness with increases in overall luminance. These are the conditions in which induction onto the test region is negligible, or non-existent, and in which our lightness experiment produced flat functions.

At the other extreme, the stimuli with the highest contrast ratios show approximate constancy, indicating that the induction, as before, is large enough to override the perception of overall changes in luminance. Another way to look at this is to suggest that induction favors a mode of perceiving surfaces or reflectance. In this case, perception in the surface mode implies black, grey or white. However, without contrast induction, and at the higher end of the luminance range, a region may appear luminous and brightness may be the most salient dimension. If brightness is the perceptual analogue of illumination, we would expect to see a breakdown of constancy under these conditions. Presumably, this explains Jameson and Hurvich's increasing function for the two higher luminance test patches used in their study.

The Second Series:

This series will attempt to more thoroughly define the parameters and conditions that produce a constant blackness judgment across changes in the overall luminance. In three experiments, the independent variables will be stimulus duration, adaptation level and increased contrast ratio. In the first case, a block of trials will consist of stimuli with a contrast ratio of .37 and distractors across three luminance levels, spanning a 2.5 log unit range. In the second case, stimuli will be preceded by an adapting field, as before, however, the stimulus duration will be that of our sustained condition above (about 7 secs). The final experiment in this set will be an extension of Experiment 3. Stimuli with greater contrast ratios will be defined in an attempt to replicate Jameson and Hurvich's finding that dim test regions decrease in brightness as overall luminance increases.

Testing a Model:

The final experiment will be designed to test a model currently being discussed. The model is an attempt to account for the differing results found in Experiments 1 and 3. In the first case, transient stimuli produced U-shaped functions whereas in the sustained condition (Experiment 3), the identical contrast ratios produced invariant judgments of blackness. The model emphasizes lateral interactions and is based on two fundamental assumptions. First, it assumes that increases in stimulus intensity decrease the response latency of both receptor potentials and lateral inhibitory interactions. Secondly, the model assumes that at higher stimulus intensities, response duration is increased. Both of these assumptions are well-supported by physiological studies showing response characteristics of receptors (Fain and Dowling, 1973) and horizontal cells (Dowling and Ripps, 1971). A third assumption is that the latency (as a function of intensity) for signals prior to lateral interactions is different from the latency of cells mediating lateral activity.

The fundamental principle of our model is based on the response asynchrony produced by adjacent regions of differing intensity. At low luminance levels, the combined response latencies of receptors encoding the surround region and lateral interactions is equivalent to the response latency of the receptor potentials corresponding to the dim center region. In this case, the signals are effectively synchronous and maximal cancellation, or inhibition of the center signal, may occur. On the other hand, under mid-luminance conditions, the combined latencies of the surround and lateral signals are not equivalent to the signal latency from the center region. If this is the case, a portion of the center signal is not inhibited by lateral activity, producing a stronger white signal. In the high luminance conditions, the latencies of the surround signals and lateral activity "catch up" to the center signal. Again, maximal cancellation occurs, and the stimulus appears blacker.

The assumption of different time constants for the latency curves describing direct and lateral signals is the most questionable part of the model. In order to test the possibility that signal asynchronies mediate the variation in perceived blackness as a function of luminance changes and short stimulus durations, the last experiment will stagger the onset time of the center and surround regions. We predict that at low and high luminance levels, a stimulus asynchrony will produce a decrease in perceived blackness. Theoretically, this would be a consequence of less synchrony between signals, and thus, less cancellation from the surround, or inhibitory field. Alternatively, stimulus asynchronies (within bounds) in the mid-luminance range will show an increase in perceived blackness, indicating greater cancellation from the surround. The details of this experiment are yet to be determined.

REFERENCES

Arend, L. & Reeves, A. (1986) Simultaneous color constancy. Journal of the Optical Society of America, 3(10), 1743-1751.

Beck, J. (1961) Judgment of surface illumination and lightness. Journal of Experimental Psychology, 58(4), 267-274.

Flock, H.R. & Noguchi (1970) An experimental test of Jameson and Hurvich's theory of brightness contrast. Perception and Psychophysics, 8(3), 129-136.

Fry, G.A. & Alpern, M. (1953) The effect of peripheral glare source on the apparent brightness of an object. Journal of the Optical Society of America, 43, 189-195.

Heinemann, E.G. (1955) Simultaneous brightness induction as a function of inducing and test field luminances. Journal of Experimental Psychology, 50(2), 89-96.

Helson, H. (1943) Some factors and implications of color constancy. Journal of the Optical Society of America, 33(10), 555-567.

Horeman, H.W. Inductive brightness depression as influenced by configurational conditions. Vision Research, 3, 121-130.

Horn, B.K. (1974) Determining lightness from an image. Computer Graphics and Image Processing, 3, 277-299.

Jacobsen, A. & Gilchrist, A. (1988) The ratio principle holds over a million-to-one range of illumination. Perception and Psychophysics, 43, 1-6.

Jameson, D. & Hurvich, L. (1961) Complexities of perceived brightness. Science, 133, 174-178.

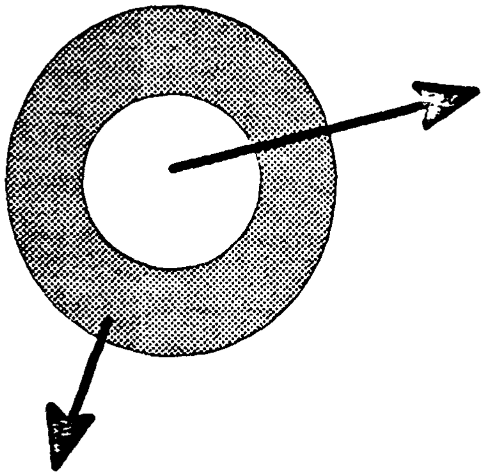
Kozaki, A. (1963) A further study in the relationship between brightness constancy and contrast. Japanese Psychological Research, 5, 129-136.

Leibowitz, H., Myers, N.A. & Chinetti, P. (1955) The role of simultaneous contrast in brightness constancy. Journal of Experimental Psychology, 30, 15-18.

Oyama, T. (1968) Stimulus determinants of brightness constancy and the perception of illumination. Japanese Psychological Research, 10(3), 146-155.

Wallach, H. (1948) Brightness constancy and the nature of achromatic colors. Journal of Experimental Psychology, 38, 310-324.

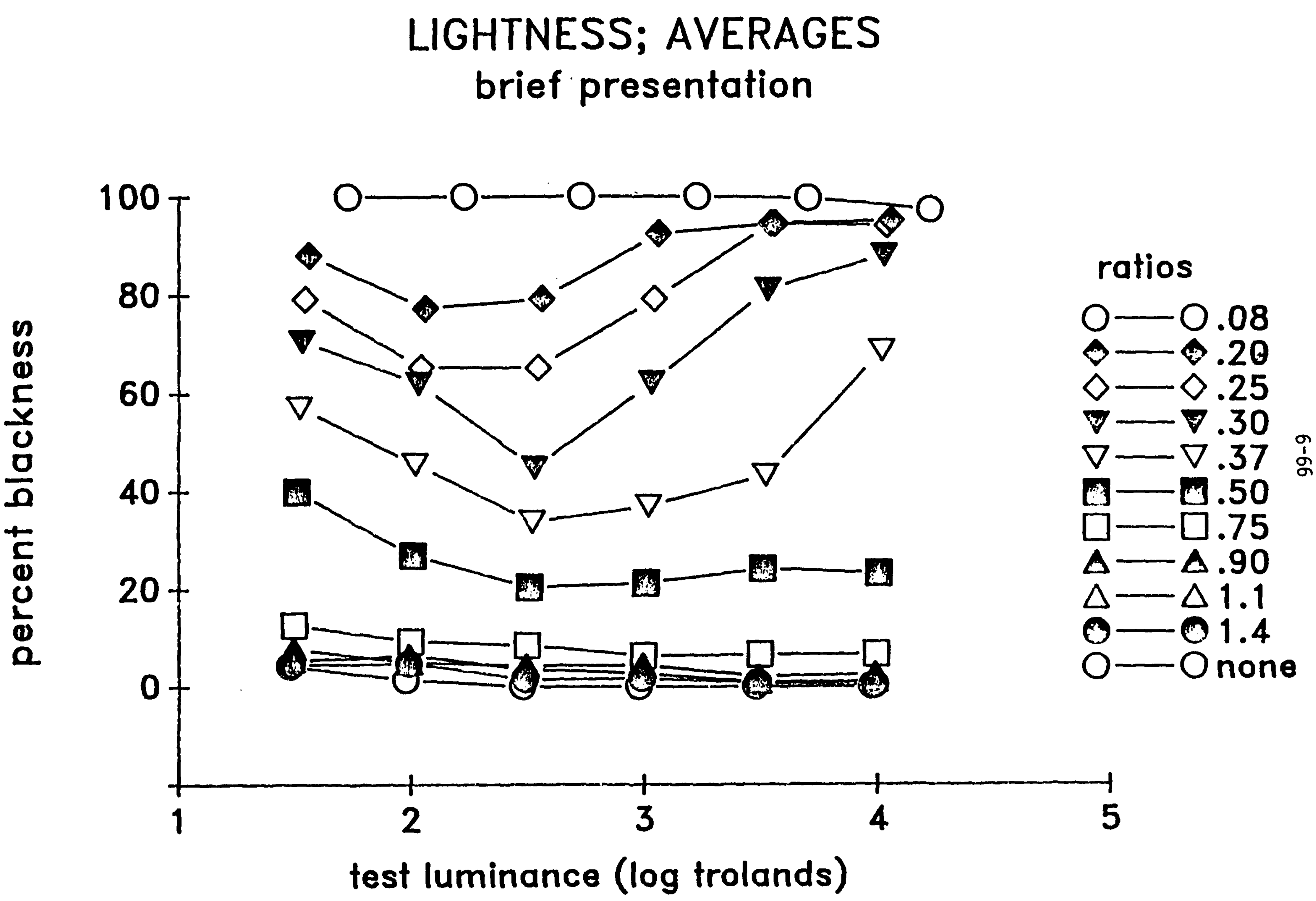
STIMULUS CONDITIONS
MAXWELLIAN VIEW



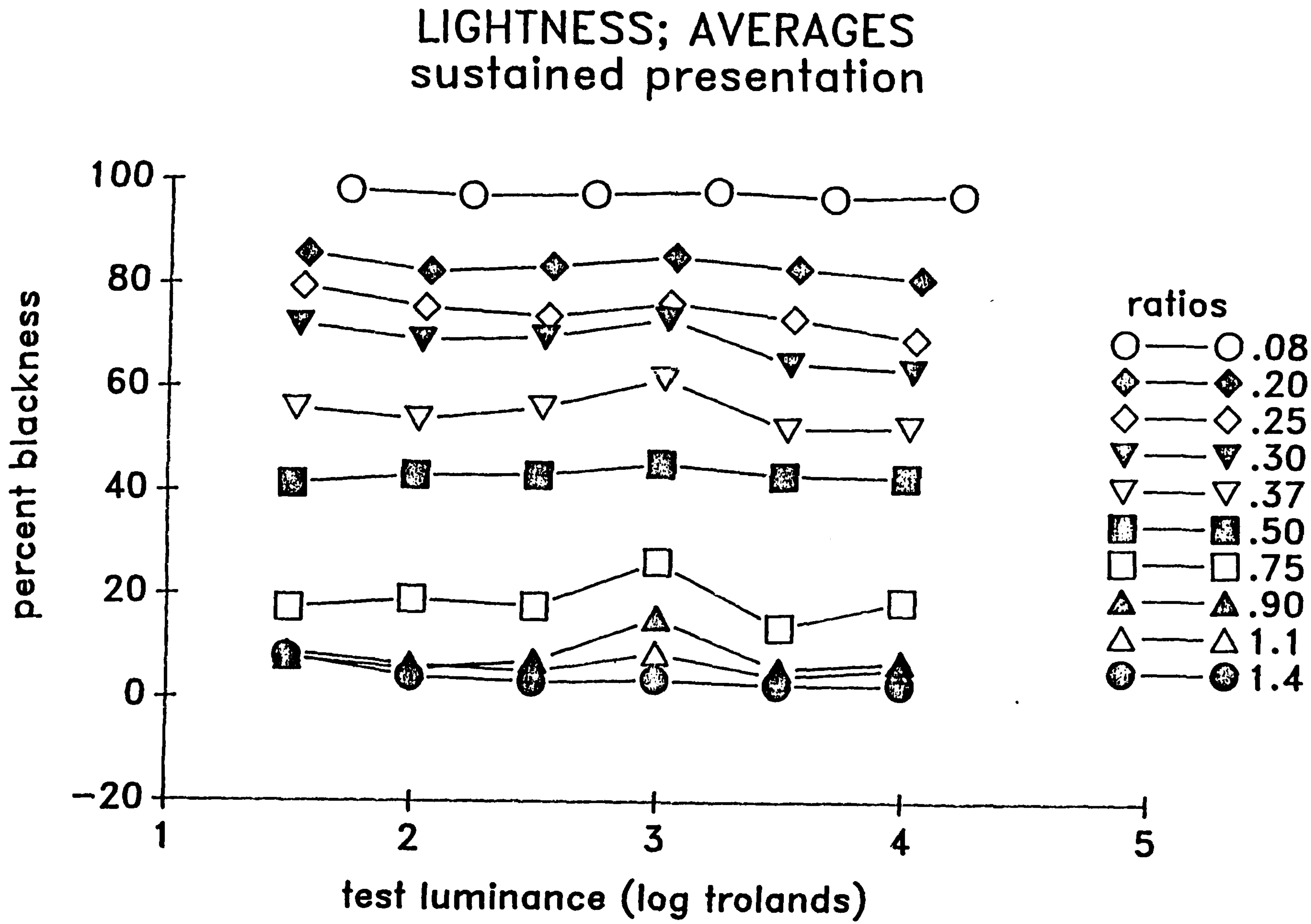
TEST REGION:
varied between
1.5 & 4.2 log trolands
in six steps

SURROUND REGION:
varied over
10 luminance levels
within each
TEST luminance level
producing

CONTRAST RATIOS:
.08
.20
.25
.30
.37
.50
.75
.90
1.10
1.40

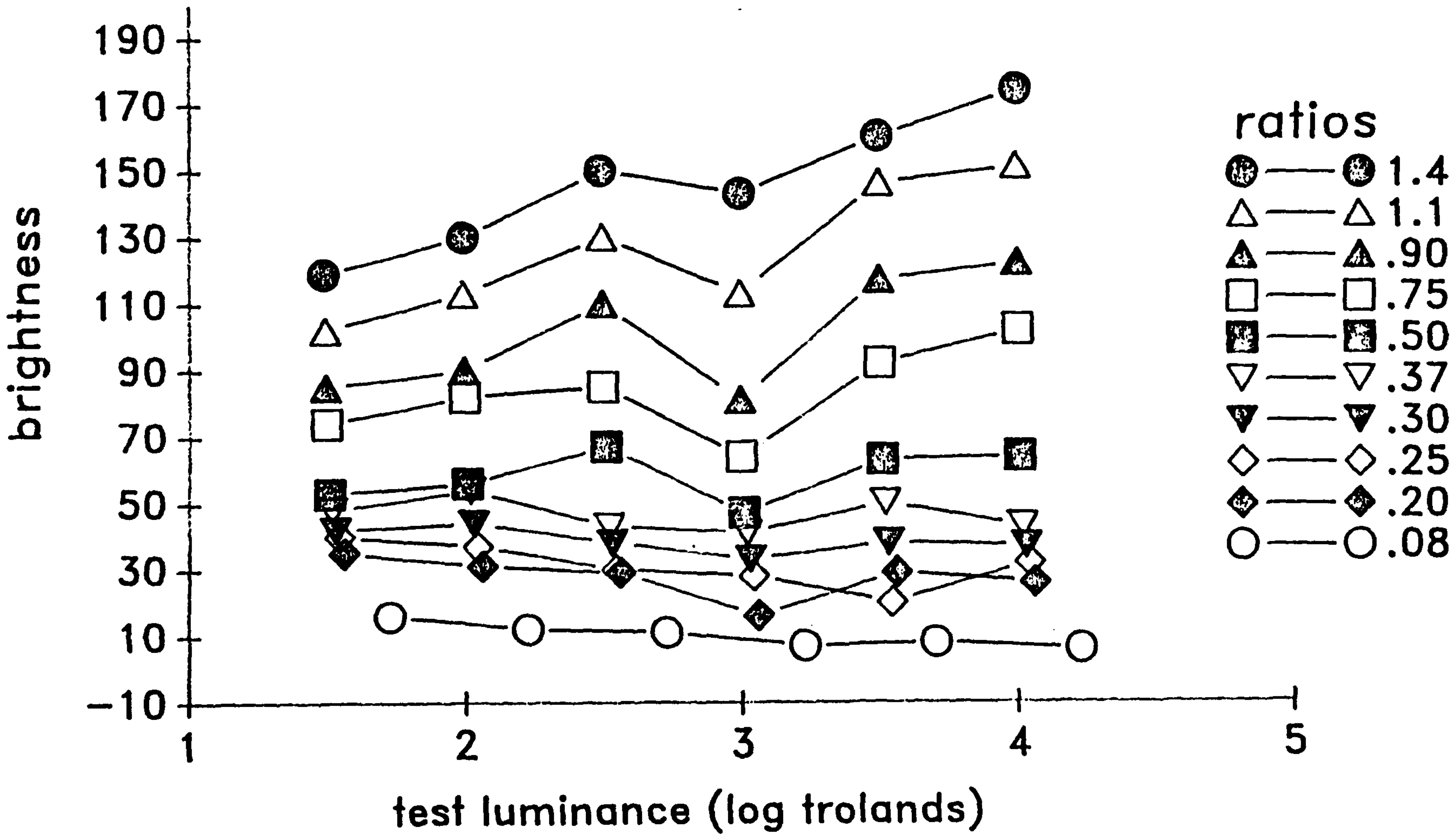


LIGHTNESS; AVERAGES
brief presentation
percent blackness
100
80
60
40
20
0
1
2
3
4
5
test luminance (log trolands)
ratios
.08
.20
.25
.30
.37
.50
.75
.90
1.1
1.4
none



LIGHTNESS; AVERAGES
sustained presentation
percent blackness
100
80
60
40
20
0
-20
1
2
3
4
5
test luminance (log trolands)
ratios
.08
.20
.25
.30
.37
.50
.75
.90
1.1
1.4

BRIGHTNESS; AVERAGES
sustained presentation



1988 USAF-UES RESEARCH INITIATION PROGRAM

Sponsored by the

AIRFORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

PHYTOTOXICITY OF SOIL RESIDUES OF JP-4 AVIATION FUEL

Prepared by: Richard H. Brown

Academic Rank: Associate Professor

Department and University: Biology Department, Ouachita Baptist University

Research Location: Ouachita Baptist University
Arkadelphia, Arkansas 71923

Primary Researcher: Richard H. Brown

Date: December 15, 1988

Contract No.: F49620-85-C-0013/SB5851-0360

Purchase Order: S-760-7MG-059

ACKNOWLEDGEMENTS

I would like to express my appreciation to the Air Force Systems Command and the Air Force Office of Scientific Research for sponsoring this research and to Universal Energy Systems for administering the program through which this study was conducted.

A number of persons have been instrumental in the operation and completion of this project. Special appreciation is extended to Ms. Missy Tomlin and Ms. Lorinda Dickey of Universal Energy Systems for the assistance they have so frequently and graciously provided. The word processor instruction of Mr. Jon Kelly and the technical assistance of Dr. Alton Crawley was invaluable and the assistance of Mrs. Wilma Maddox and Mr. Joe franz in the acquisition of supplies and materials is deeply appreciated. A special love is extended to my wife, Alta, for allowing me to go a little insane during this study.

ABSTRACT

Fifteen bioassays were conducted under two different environmental conditions and in three soil types to assess the toxicity of JP-4 aviation fuel soil residues to sorghum (Sorghum vulgare), bean (Phaseolus vulgaris), oat (Avena sativa), cucumber (Cucumis sativus), wheat (Triticum aestivum), corn (Zea mays), and soybeans (Glycine maximus). Wet shoot weight and shoot length was stimulated in most plants grown in soils contaminated with less than 2 ppmw and was inhibited in soils receiving between 2,000 and 300,000 ppmw of JP-4. Growth patterns were similar for most plants in all soils and conditions studied. An ED-50 was experimentally determined to be between 20,000 and 200,000 ppmw for most plant groups regardless of soil type or conditions employed. Four of these bioassays indicated that by 90 days after contamination, the concentration of JP-4 residue in commercial potting soil had diminished to a level equivalent to less than 20 ppmw.

I. INTRODUCTION:

JP-4 aviation fuel is the primary fuel used in aircraft of the United States Air Force and is composed of a complex mixture of about 300 aliphatic and aromatic hydrocarbon compounds varying in length up to C_{16}. Jettisoning of fuels, aircraft crashes, and fueling operations may effect a release of aviation fuels into the environment thus effecting some degree of ecological perturbation. It has been reported by Clewell (1980) and Bishop (1983) that upon release, those fuel components on the lower boiling light end, C_8 and below, tend to evaporate leaving the heavier C_9 - C_{16} compounds as residues. The phytotoxicity of these residues in soil is the focus of this study.

Previous studies have reported the toxicity and mutagenicity of jet fuels to dogs, monkeys, mice, rats, and humans (Kinkead, 1974 and Kapp, 1979 and Knave, 1979). Still others have addressed the impact of these fuels on aquatic ecosystems (Jenkins, 1977 and Doane, 1980 and Bombick, 1982). In contrast, there has been limited research conducted on the effects of aviation fuels on terrestrial ecosystems although these effects are certainly just as important. One study has indicated that JP-4 soil residues are substantially less problematic than previously thought and at low concentrations (.002-2 ppmw) even stimulate growth in some plants (Brown, 1987).

This study was sponsored by the United States Air Force Office of Scientific Research through Universal Energy Systems, Inc. in order to continue the work begun during the summer of 1987 comparing new results with those already obtained and to study the effects of JP-4 soil residues on additional plant species and soil types.

II. OBJECTIVES OF THE RESEARCH EFFORT:

1. A continued literature search for information pertinent to the project.
2. To continue the study begun during the summer of 1987 comparing new results with those already obtained.
3. To study the effects of JP-4 soil residues on additional plant species.
4. To study the longevity of JP-4 residues in soil.
5. To study the effect of other soil types on JP-4 residue action.

III. MATERIALS AND METHODS:

In bioassays 1-3 (Figures 1-6), commercial topsoil (Southland) was air dried and sieved through a number 7 U. S. Standard Sieve then adjusted to a pH between 6.5 and 7.5. One thousand gram samples of this soil were placed in large plastic bags and contaminated with 0, 0.002, 0.02, 0.2, 2, 20, 200, 2000, 20000, and 200000 ppmw of JP-4 aviation fuel dissolved in a methyl alcohol carrier by slowly dripping the contaminant solution into a depression in the surface of each sample. These samples were ventilated for 24 hours to allow evaporation of the carrier and the volatile components of the fuel then thoroughly mixed by closing and inverting each bag 25 times (Lavy, 1986). One hundred and twenty-five gram aliquots of each sample were transferred to a series of labeled, plastic pots and watered to field capacity.

Seeds of corn (Zea mays, var. Funk's 4765), wheat (Triticum daestivum, var. Florida 302), and bean (Phaseolus vulgaris, var. Improved) were pregerminated for 24 hours to provide seeds with uniform radicle lengths. Five similar sized, pregerminated seeds were planted in each of four cups per concentration per species so that the surface of each seed was level with the soil surface. These pots were then placed in a greenhouse for 14 days during which the temperature varied from 22 degrees C. at night to 28 degrees C. for the 14 hour days. Each cup was watered daily to field capacity and the relative humidity was allowed to vary between 60 and 90 percent.

In bioassays 4-7, the same procedure was employed except the plants were grown in Biotronette Mark III environmental chambers in which the temperature varied from 18 degrees C. at night to 25 degrees C. during the 14 hour day at full illumination. Also the contaminated soils were ventilated for 90 days before being transferred to pots for planting. In bioassay 4 (Figures 7 & 8), bean (*Phaseolus vulgaris*, var. Improved) was the test species and data obtained in this series was compared to that of a replicate series in bioassay 5 (Figures 9 & 10). Bioassay 6 (Figures 11 & 12), used corn (*Zea mays*, var. Funk's 4765) as the test species and data obtained was compared with that from a replicate series in bioassay 7 (Figures 13 & 14).

In bioassays 8 and 9 (Figures 15 - 18), clay-loam soil was contaminated 24 hours prior to planting with pregerminated seeds of soybean (Glycine maximus var. Davis) and sorghum (Sorghum vulgare var. Fox 525) respectively. These plants were grown in Biotronette Mark III chambers under the conditions described above. Bioassays 10 and 11 (Figures 19 - 22) employed the same procedures except the pregerminated seeds were planted in sand-loam soil.

In bioassays 12 - 15 (Figures 23 - 30), commercial potting soil was contaminated with 0, 50000, 100000, 150000, 200000, 250000, and 300000 ppmw of JP-4, ventilated for 24 hours, and planted with pregerminated seeds of soybean (Glycine maximus var. Davis), wheat (Triticum aestivum var. Florida 302), oat (Avena sativa var. Bob), and cucumber (Cucumis sativus var. Poinsett 76) respectively. These plants were grown in the Biotronette Mark III chambers as outlined above.

In each series, the above ground plant shoots were harvested on the 14th day of growth. Fresh shoot weights (wet weights) and shoot lengths (soil surface to the apex of the leaf providing the longest measure) were determined and recorded. These data are summarized in Tables 1 - 5 and Figures 1 - 30.

Data from all assays were analyzed for means, standard deviations, standard errors, and confidence limits using Analysis of Variance (ANOVA, $p<0.01$).

IV. Results and Discussion:

In bioassays 1 - 3 in which corn, bean and wheat were planted in soils contaminated with 0 - 200,000 ppmw of JP-4, plant growth (shoot length and wet shoot weight) was generally stimulated by concentrations up to 2 ppmw and inhibited by concentrations that exceeded 20 ppmw. An ED-50 was experimentally indicated between 20,000 and 200,000 ppmw. These data are summarized in Table 1 and in Figures 1 - 6 and confirm the results of the work begun during the summer of 1987 (Brown, 1987) at Brooks Air Force Base, Texas.

In bioassays 4 - 7, soil was contaminated with 0 - 200,000 ppmw of JP-4 then ventilated for 90 days before being planted. In #4 in which bean was the test organism, results indicate that growth is unaffected in soil originally contaminated with up to 2,000 ppmw and only slightly inhibited in samples contaminated with up to 200,000 ppmw. These results were confirmed in bioassay 5. In bioassay 6, corn growth was stimulated in soils originally treated with up to 20,000 ppmw and no inhibition was observed even at the highest levels of JP-4 contamination. Bioassay 7 confirmed these findings. These results indicate that by 90 days after treatment with up to 200,000 ppmw of JP-4, only an amount equivalent to less than 20 ppmw remains. Table 2 and Figures 7 - 14 summarize these data.

In bioassays 8 and 9, clay-loam soil was combined with 0 - 200,000 ppmw of JP-4, ventilated for 24 hours, then planted with pregerminated soybean and sorghum seeds. At no concentration was stimulation of growth observed but growth was generally diminished in those samples grown in soil contaminated with 2,000 - 200,000 ppmw. In bioassays 10 and 11, similar results were obtained when soybean and sorghum were planted in sand-loam soil treated in the same fashion. In each series, an ED-50 is indicated between 20,000 and 200,000 ppmw. These results are presented in Tables 3 and 4 and in Figures 15 - 22.

In a final series of assays (12 - 15), an attempt was made to more precisely determine an ED-50 for JP-4 residues in soil. For some reason, the results of these were inconclusive as shown in Table 5 and in Figures 23 - 30.

Fresh shoot weight and shoot length dose-response curves were remarkably complimentary in all tests involving similarly treated soils. Mortality was not considered a factor due to nearly 95% survivorship even at the higher concentrations.

Throughout the study, only fresh shoot weight and shoot length were measured and recorded; but in each series, root growth was observed. It was evident that as the concentration of JP-4 increased above 200 ppmw, root growth diminished until at 20,000 - 200,000 ppmw root growth was practically nonexistent. It is suggested that any shoot growth occurring at these higher concentrations is due to cotyledon nourishment and when that nourishment is depleted the plants will die. It is suggested that more accurate assessment of plant growth would result from 30 - 60 day bioassays.

Although inclusion of field studies of the effect of JP-4 soil residues were originally planned, interference by departmental personal resulted in their omission. Such studies should be conducted.

V. Recommendations:

As is typical of scientific investigations, this study has raised more questions than it has answered. Although these 14-day bioassays indicate a relatively low toxicity for JP-4 soil residues, longer assay times could change this assessment. The following suggestions are presented for consideration:

a. A new series of assays based on 30 or 60 day growth periods or even to maturity should be conducted. These studies would test growth after cotyledon depletion and could also test crop productivity.

b. In order to study the stimulatory effects of very low levels of JP-4 contamination, bioassays that include growth periods to maturity should be initiated.

c. Field studies should be devised to determine the effects of JP-4 under natural conditions in various types and using numerous test species.

d. Long term studies tc assess JP-4 persistence in natural soils should be conducted.

REFERENCES

1. Bailey, D. R., "Comparison of Oats and Lemna oligorhiza for Diuron Bioassay," Queenland Journal of Agricultural and Animal Sciences, Vol. 27, 1970. pp. 395-400.

2. Behrins, R., "Quantitative Determination of Triazine Herbicides in Soils by Bioassay," Residue Reviews, Vol. 32, 1970, pp. 355-369.

3. Bishop, E. C., M. Macnaughton, R. deTreville, and R. Drawbaugh, Rationale for a Threshold Limit Value (TLR)R for JP-4 Jet B Wide Cut Aviation Turbine Fuel, USAFOEHL Report No. 83-128EH111DGA, Brooks Air Force Base, TX, 1983.

4. Bombick, D. W. and L. G. Arlian, Comparative Toxicity of Selected Aviation Fuels as Measured by Insect Bioassay, USAFAMRL Report No. AFAMRL-TR-82-31, Wright-Patterson Air Force Base, OH, 1982.

5. Brown, Richard H., Phytotoxicity of Soil Residues of JP-4 Aviation Fuels, USAF/UES Report Contract No. F49620-85-C-0013 Brooks Air Force Base, TX, 1987.

6. Clewell III, H. J., The Effect of Fuel Composition Groundfall From Aircraft Fuel Jettisoning, USAFAFESC Report No. ESL-TR-81-13, Tyndall Air Force Base, FL, 1981.

7. Doane, T., USAFOEHL Guide to Aquatic Bioassay Techniques, Terminology, and Capabilities, USAFOEHL Report No. 80-18, Brooks Air Force Base, TX, 1980.

8. Dowler, C. C., "A Cucumber Bioassay Test for the Soil Residues of Certain Herbicides," Weed Science, Vol. 17, Weed Science Society of America, 1969, pp. 309-310.

9. Gerber, H. R., "Biotests for Herbicide Development," Crop Protection Agents - Their Biological Evaluation, (N. R. McFarlane, ed.), Academic Press, N.Y., 1977, pp.307-321.

10. Hurle, K., "Biotests for the Detection of Herbicides in the Soil," Crop Protection Agents - Their Biological Evaluation, (N. R. McFarlane, ed.), Academic Press, N.Y., 1977, pp.285-306.

11. Jacques, G. L., and R. G. Harvey, "Persistence of Dinitroaniline Herbicides in Soil," Weed Science Vol. 27, Weed Science Society of America, 1979, pp. 660-665.

12. Jayakumar, R., and A. Mohamed Ali, "Response of Cucumber, Finger Millet, Foxtail Millet, and Sunflower to soil Residues of Atrazine and 2,4 D," Indian J. Weed Science, Vol. 16, 1984, pp. 133-135.

13. Jenkins, D., S. A. Klein, and R. C. Cooper, "Fish Toxicity of Jet Fuels--I. The Toxicity of the Synthetic Fuel JP-9 and its Components," Water Research, Vol.11, 1977, pp. 1059-1067.

14. Kapp, R. W. and C. E. Piper, In Vitro and In Vivo Mutagenicity Studies Jet Fuels, A Final Report, Hazelton Laboratories America, Vienna, VA, 1979.

15. Kimball, K. D. and S. A. Levin, "Limitations of Laboratory Bioassays: The Need for Ecosystem Level Testing," Bioscience, Vol. 35, 1985, pp. 99-129.

16. Kinkead, E. R., L. C. DiPasquale, E. H. Vernot and J. D. MacEwen, Chronic Toxicity of JP-4 Jet Fuel, Report No. 11, AFAMRL-TR-74-125, Wright-Patterson Air Force Base OH, 1974.

17. Knave, B., P. Mindus, and G. Struwe, "Neurasthenic Symptoms in Workers Occupationally Exposed to Jet Fuel," Acta Psychiatrica Scandinavica, Vol. 60, 1979, pp. 39-49.

18. Kohn, G. K., "Bioassay as a Monitoring Tool," Residue Reviews, Vol. 76, Springer-Verlac, N.Y., 1980, pp. 99-129.

19. Kratky, B. A. and G. F. Warren, "The Use of Three Simple Rapid Bioassays on Forty-two Herbicides," Weeds Research, Vol. 11, 1971,pp. 257-262.

20. Lavy, T. L. and P. W. Santelmann, "Herbicide Bioassay as a Research Tool," Research Methods in Weed Science, 3rd Ed. (N. D. Camper, ed.) Southern Weed Science Society, Champaign, IL, 1986, pp. 201-217.

21. Leasure, J. K., "Bioassay Methods for 4-Amino-3,5,6-trichloropicolinic Acid," Weeds, Vol. 12, Weed Society of America, 1964, pp. 232-233.

22. Santelmann, P. E., J. B. Weber, and A. F. Wiese, "A Study of Soil Bioassay Technique Using Prometryne," Weed Science, Vol. 19, 1971, pp. 170-174.

Table 1. Summary of results in bioassays 1 - 3 in which commercial potting soil was contaminated with 0, .002, .02, .2, 2, 20, 200, 2000, 20000, and 200000 ppmw of JP-4 and then after 24 hours of ventilation was planted.

Bioassay	Plant studied	Parameter of study	Range of stimulation*	Range of inhibition*	Range of ED-50*
1	Corn	ww**	.002-2	200-200,000	20,000-200,000
		l**	.002-2	200-200,000	Undetermined
2	Bean	ww	2	20-200,000	20,000-200,000
		l	.002-2	20-200,000	20,000-200,000
3	Wheat	ww	.002	200-200,000	Near 200,000
		l	None	20-200,000	Undetermined

Table 2. Summary of results in bioassays 4 - 7 in which commercial potting soil was contaminated with 0, .002, .02, .2, 2, 20, 200, 2000, 20000, and 200000 ppmw of JP4 and then after 90 days of ventilation was planted.

Bioassay	Plant studied	Parameter of study	Range of stimulation*	Range of inhibition*	Range of ED-50*
4	Bean	ww**	None	2,000-200,000	Undetermined
		l**	None	200,000	Undetermined
5	Bean	ww	2	20,000-200,000	Undetermined
		l	None	2,000-200,000	Undetermined
6	Corn	ww	2-20,000	None	Undetermined
		l	20-20,000	None	Undetermined
7	Corn	ww	2-20,000	None	Undetermined
		l	2-2,000	None	Undetermined

* - Parts per million by weight (ppmw).
** - ww = wet weight of plant shoot and l = shoot length.

Table 3. Summary of results in bioassays 8 - 9 in which clay loam soil was contaminated with 0, 2, 20, 200, 2000, 20000, and 200000 ppmw of JP-4 and then after 24 hours of ventilation was planted.

Bioassay	Plant studied	Parameter of study	Range of stimulation*	Range of inhibition*	Range of ED-50*
8	Soybean	ww**	None	2,000-200,000	20,000-200,000
		1**	None	20,000-200,000	20,000-200,000
9	Sorghum	ww	None	2,000-200,000	20,000-200,000
		1	None	2,000-200,000	20,000-200,000

Table 4. Summary of results in bioassays 10 -11 in which sand loam soil was contaminated with 0, 2, 20, 200, 2000, 20000, and 200000 ppmw of JP-4 then after 24 hours of ventilation was planted.

Bioassay	Plant studied	Parameter of study	Range of stimulation*	Range of inhibition*	Range of ED-50*
10	Soybean	ww**	None	20,000-200,000	20,000-200,000
		1**	2	2,000-200,000	20,000-200,000
11	Sorghum	ww	None	2,000-200,000	20,000-200,000
		1	None	2,000-200,000	20,000-200,000

* - Parts per million by weight (ppmw).
** - ww = wet weight of plant shoot and 1 = shoot length.

Table 5. Summary of results in bioassays 12 - 15 in which commercial potting soil was contaminated with 0, 50000, 100000, 150000, 200000, 250,000 and 300,000 ppmw of JP-4 then after 24 hours of ventilation was planted.

Bioassay	Plant studied	Parameter of study	Range of stimulation*	Range of inhibition*	Range of ED-50*
12	Soybean	ww**	None	50,000-300,000	Undetermined
		l**	None	50,000-300,000	Undetermined
13	Wheat	ww	None	50,000-300,000	Undetermined
		l	None	50,000-300,000	Undetermined
14	Oat	ww	None	50,000-300,000	Near 200,000
		l	None	50,000-300,000	Near 300,000
15	Cucumber	ww	None	50,000-300,000	Near 200,000
		l	None	50,000-300,000	Near 250,000

* - Parts per million by weight (ppmw).
** - ww = wet weight of plant shoot and l = shoot length.

Fig. 1. The effects of JP-4 on the wet weight of corn shoots grown in commercial potting soil.

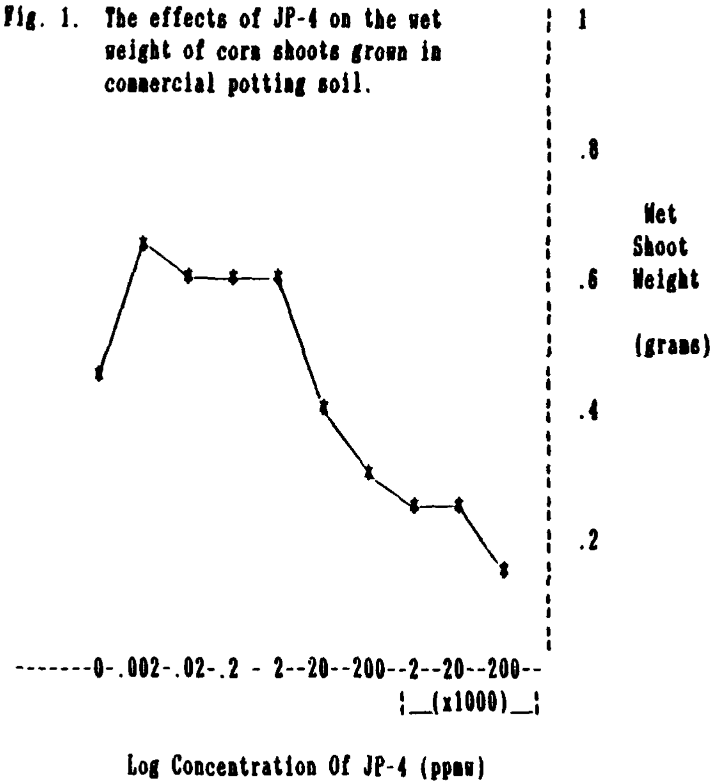
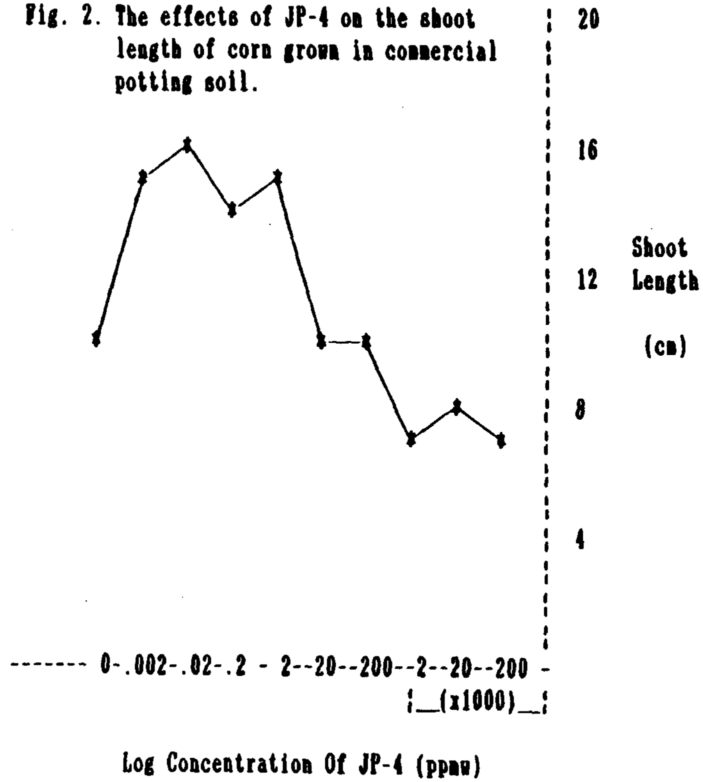


Fig. 2. The effects of JP-4 on the shoot length of corn grown in commercial potting soil.



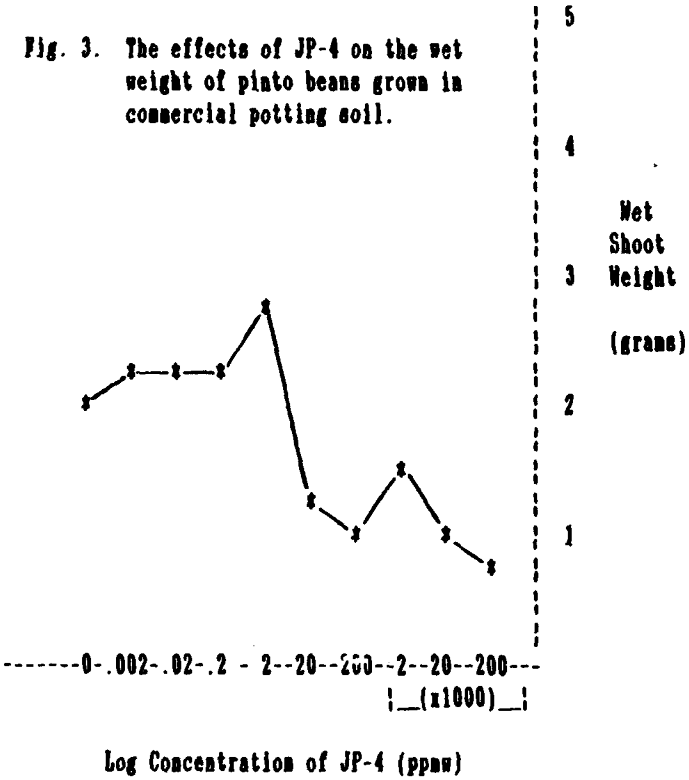


Fig. 3. The effects of JP-4 on the wet weight of pinto beans grown in commercial potting soil.

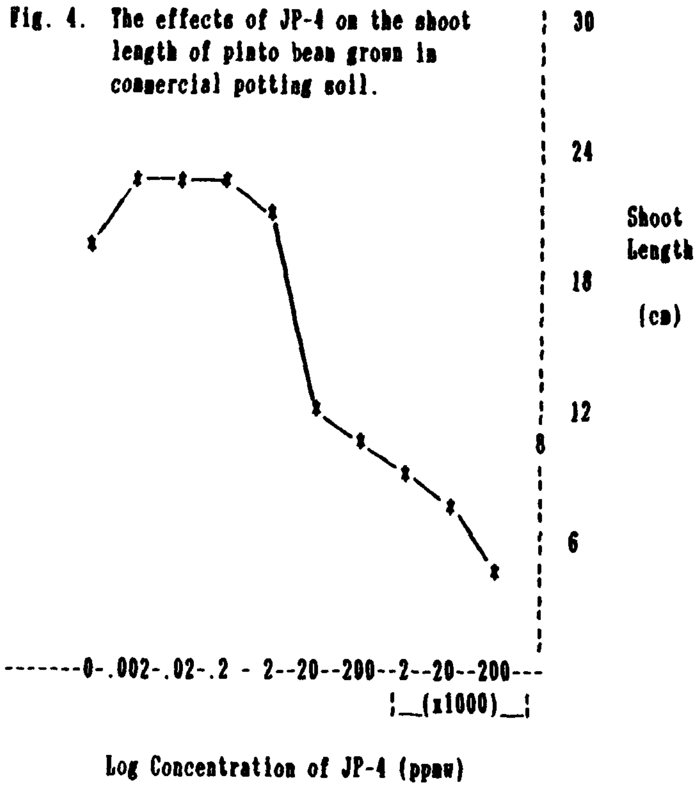


Fig. 4. The effects of JP-4 on the shoot length of pinto bean grown in commercial potting soil.

Fig. 5. The effects of JP-4 on the wet shoot weight of wheat grown in commercial potting soil.

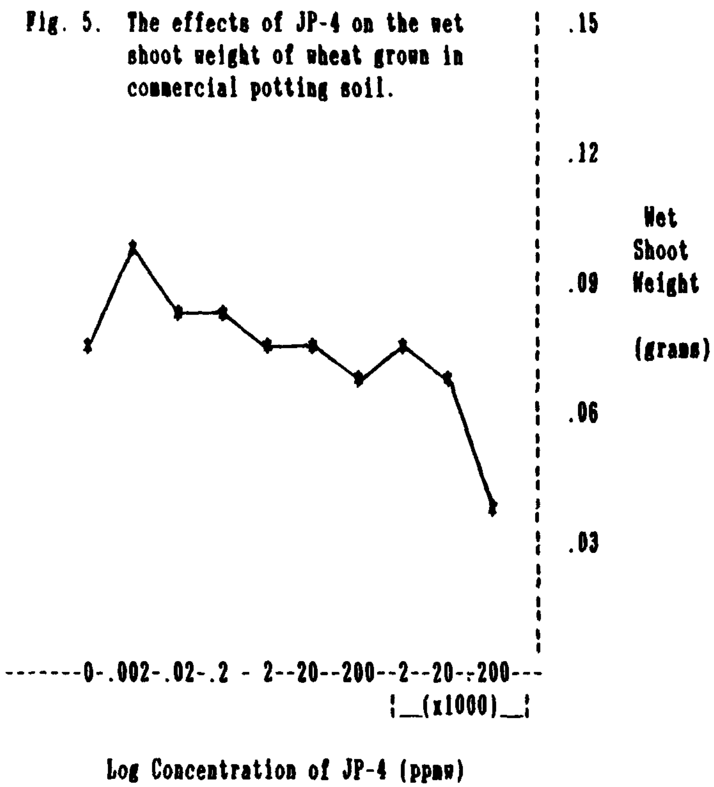


Fig. 6. The effects of JP-4 on the shoot length of wheat grown in commercial potting soil.

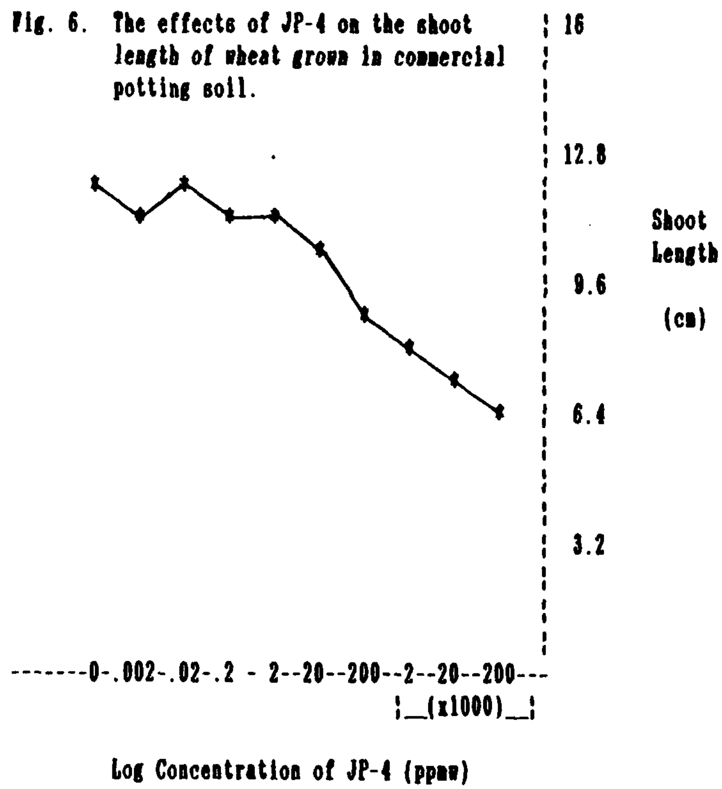


Fig. 7. The effects of JP-4 on the wet weight of pinto bean shoots grown in commercial potting soil contaminated 90 days prior to planting.

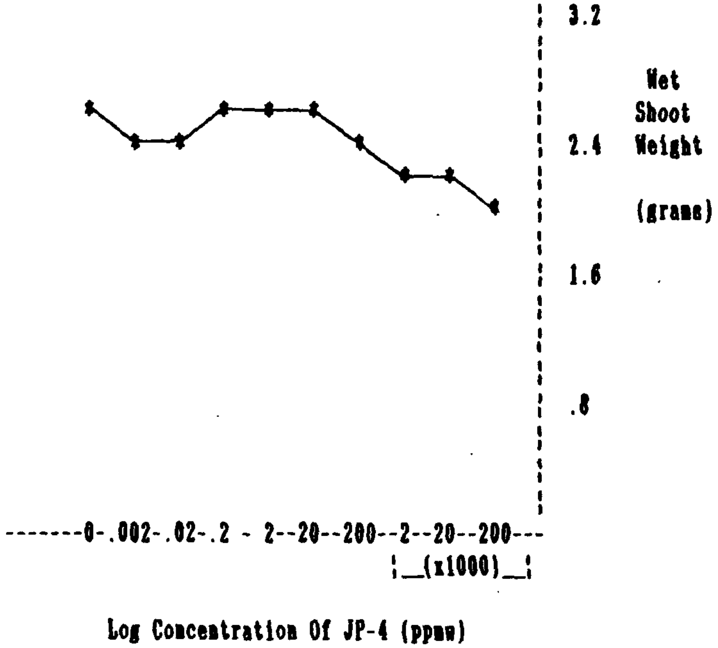


Fig. 8. The effects of JP-4 on the shoot length of pinto beans grown in commercial potting soil contaminated 90 days prior to planting.

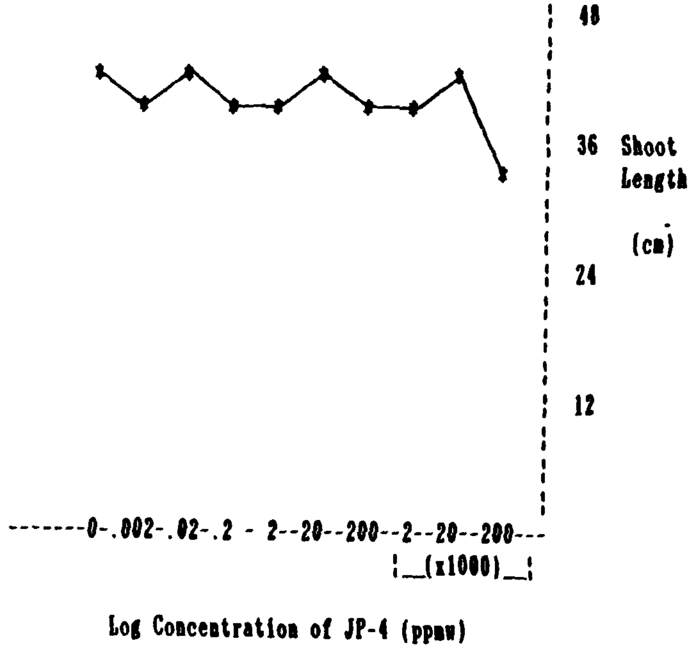


Fig. 9. The effects of JP-4 on the wet weight of pinto bean shoots grown in commercial potting soil contaminated 90 days prior to planting.

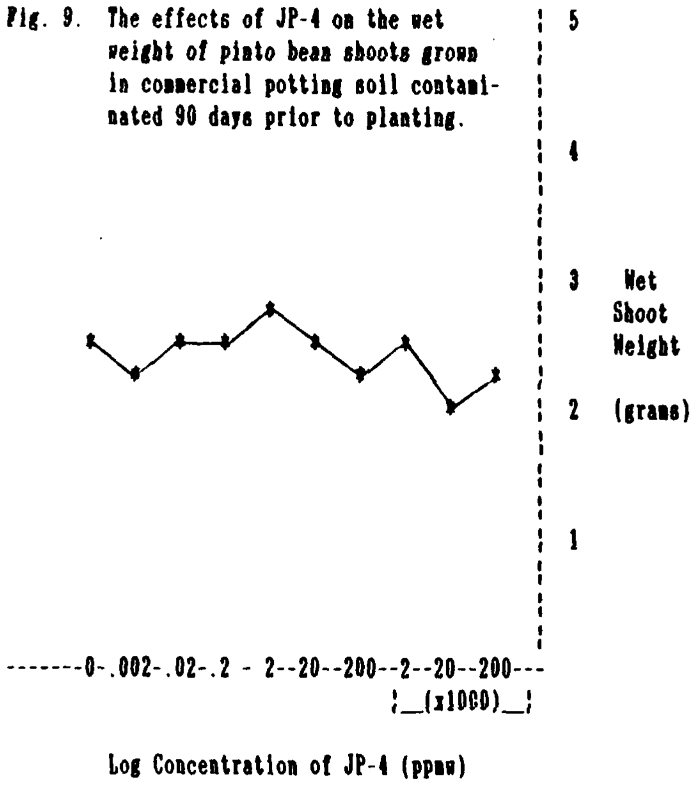


Fig. 10. The effects of JP-4 on the shoot length of pinto beans grown in commercial potting soil contaminated 90 days prior to planting.

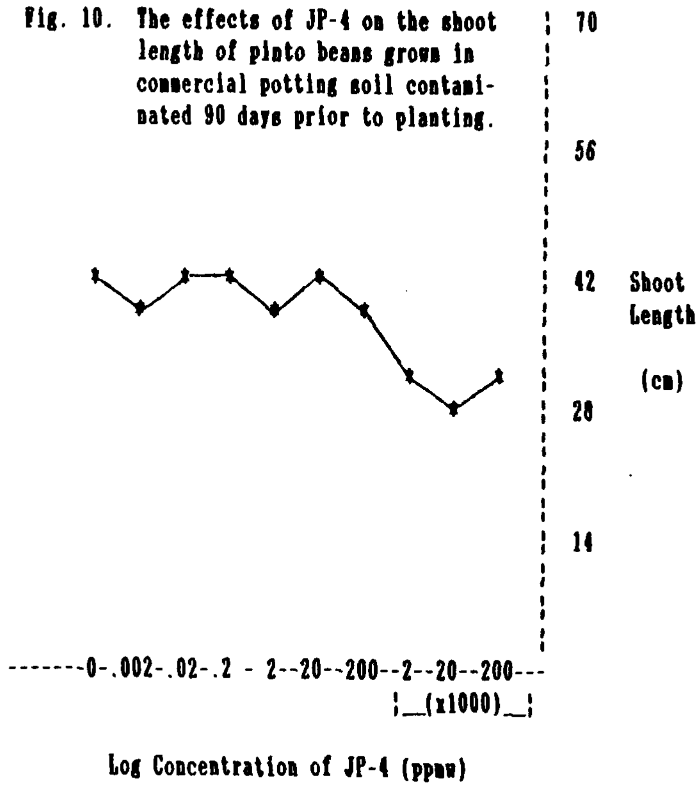


Fig. 11. The effects of JP-4 on the wet weight of corn shoots grown in commercial potting soil contaminated 90 days prior to planting.

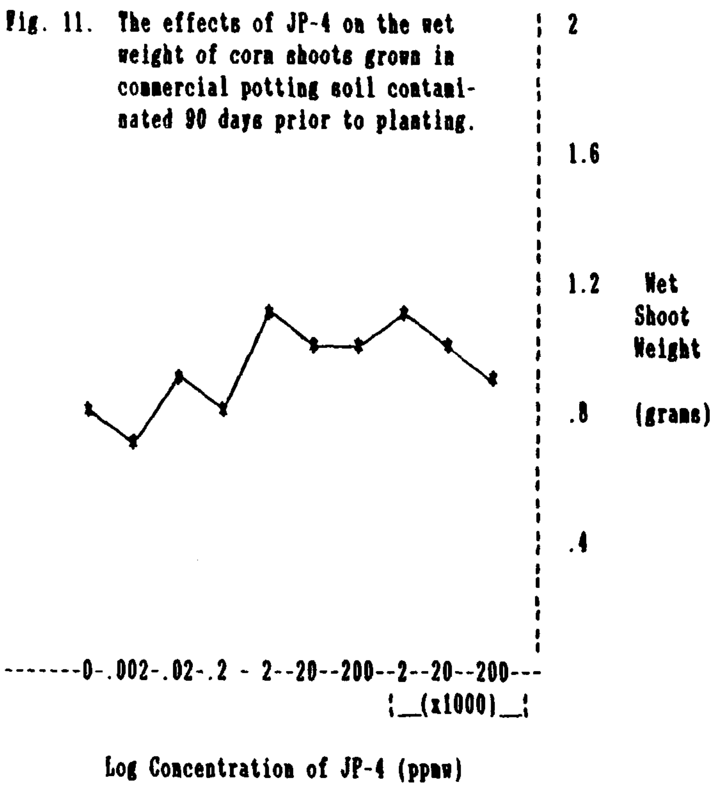


Fig. 12. The effects of JP-4 on the shoot length of corn grown in commercial potting soil contaminated 90 days prior to planting.

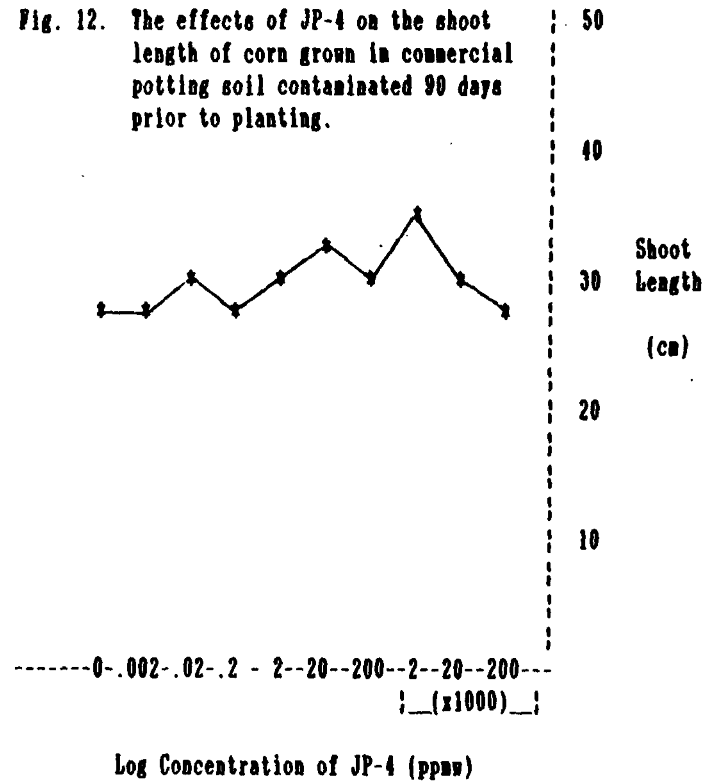


Fig. 13. The effects of JP-4 on the wet weight of corn shoots grown in commercial potting soil contaminated 90 days prior to planting.

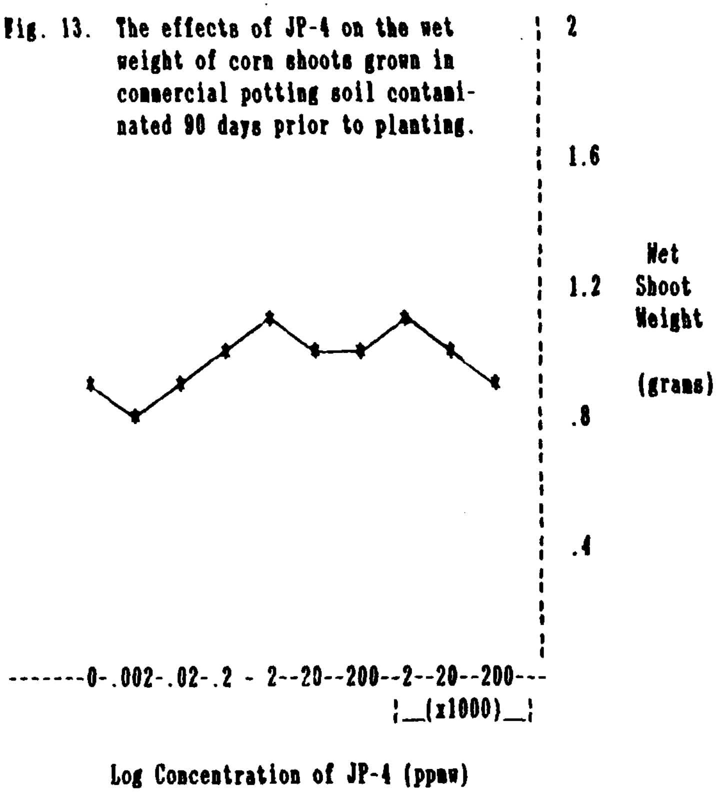


Fig. 14. The effects of JP-4 on the shoot length of corn grown in commercial potting soil contaminated 90 days prior to planting.

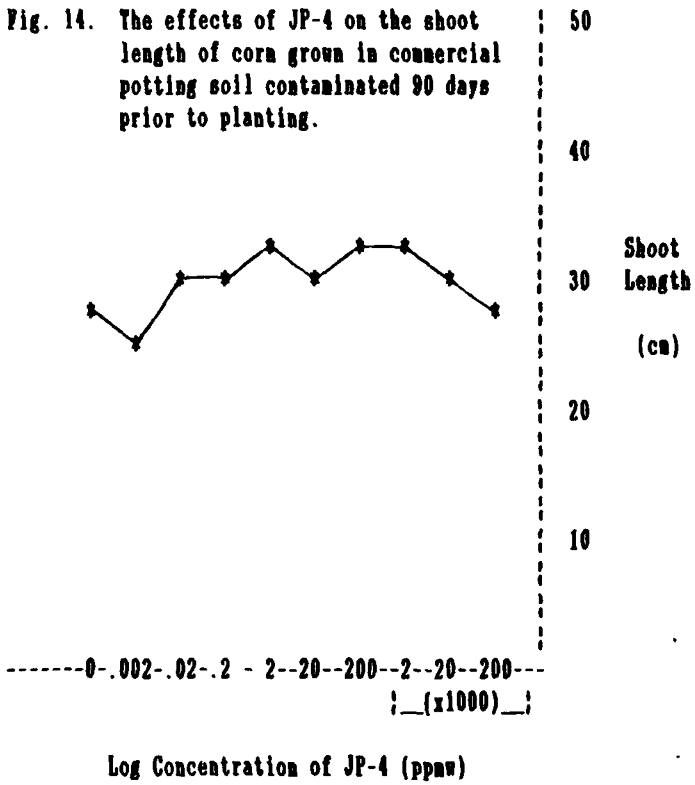


Fig. 15. The effects of JP-4 on the wet weight of soybean shoots grown in clay loam soil.

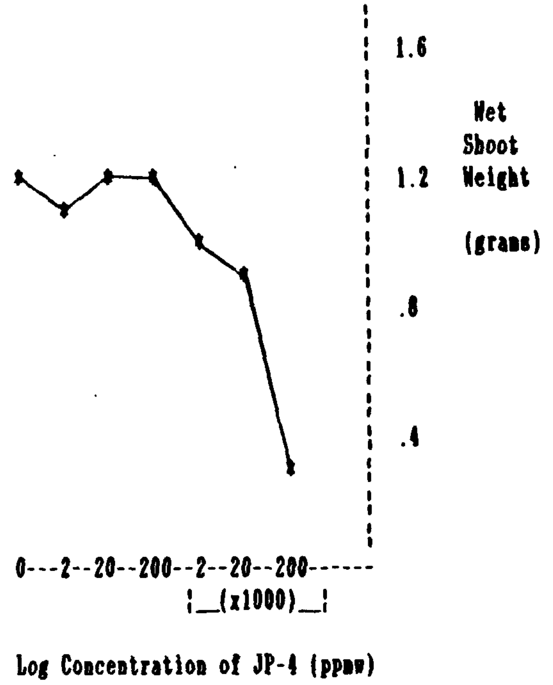
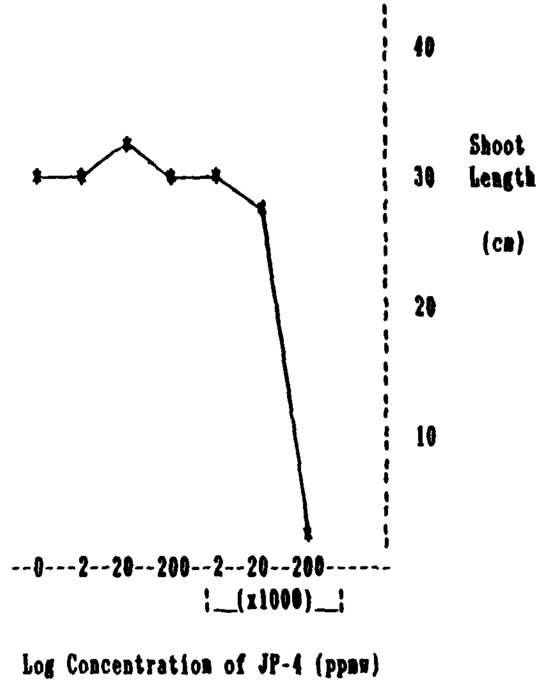


Fig. 16. The effects of JP-4 on the shoot length of soybeans grown in clay loam soil.



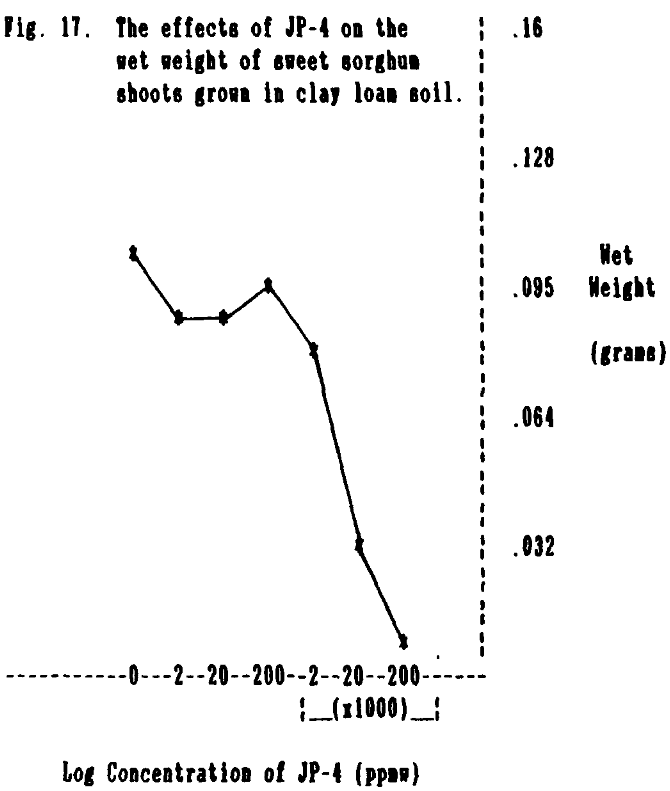


Fig. 17. The effects of JP-4 on the wet weight of sweet sorghum shoots grown in clay loam soil.

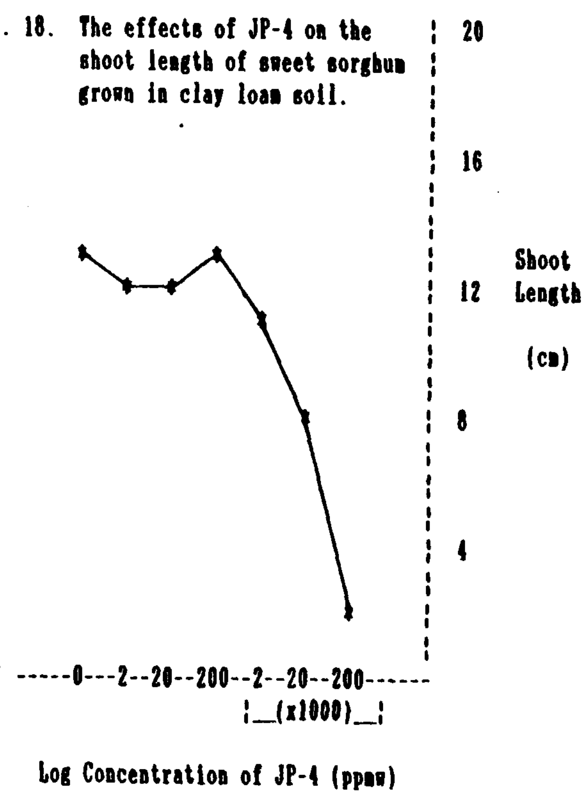


Fig. 18. The effects of JP-4 on the shoot length of sweet sorghum grown in clay loam soil.

Fig. 19. The effects of JP-4 on the wet weight of soybean shoots grown in sandy loam soil.

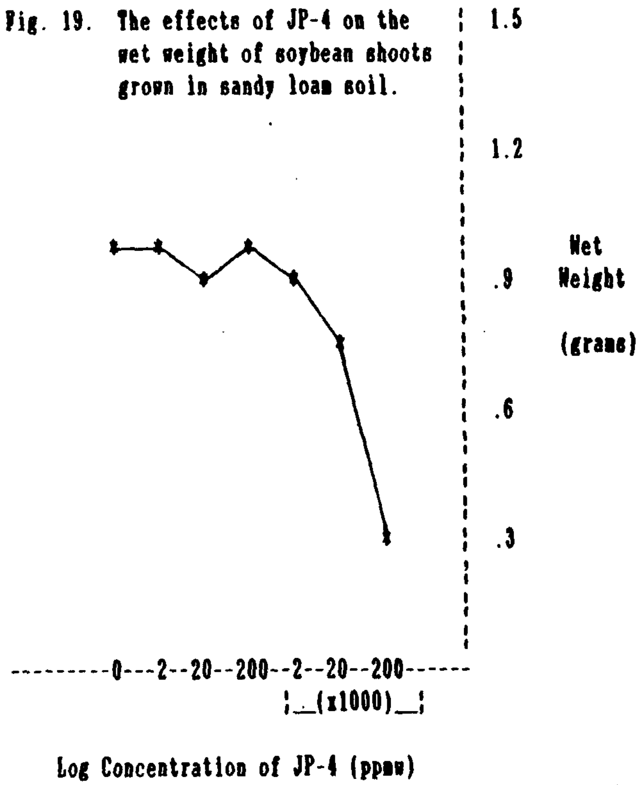


Fig. 20. The effects of JP-4 on the shoot length of soybeans grown in sandy loam soil.

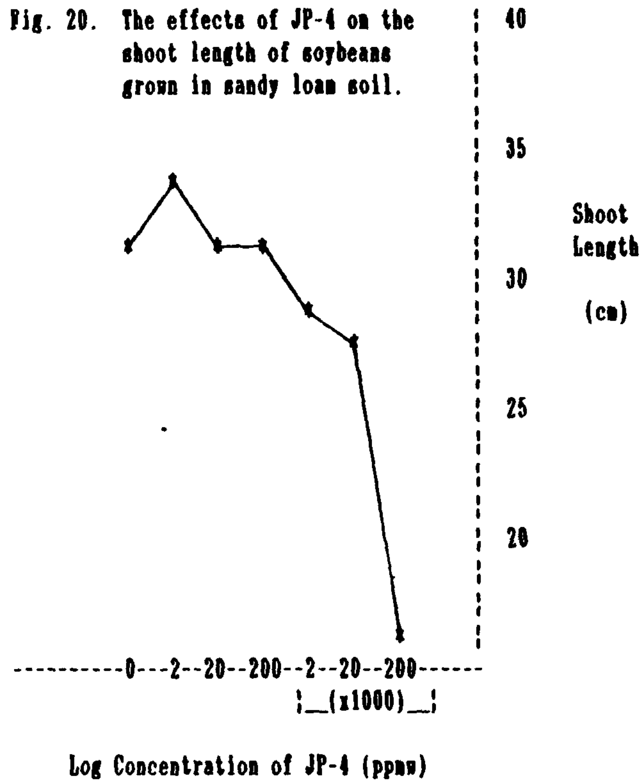


Fig. 21. The effects of JP-4 on the wet shoot weight of sweet sorghum grown in sandy loam soil.

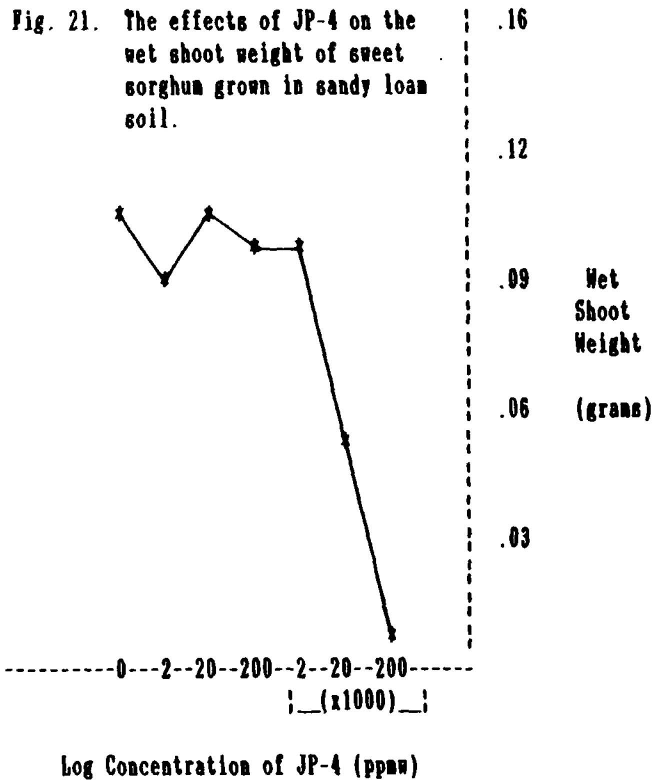


Fig. 22. The effects of JP-4 on the shoot length of sweet sorghum grown in sandy loam soil.

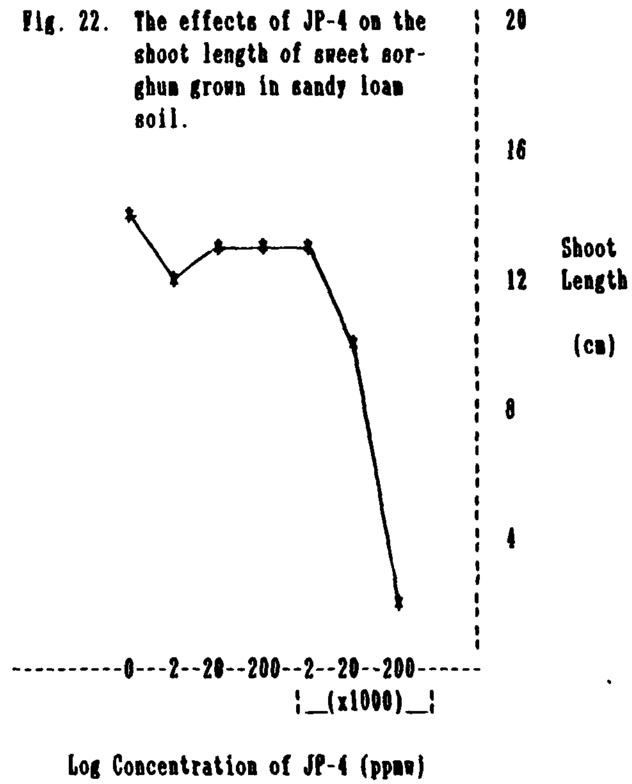


Fig. 23. The effects of JP-4 on the wet shoot weight of soybeans grown in commercial potting soil.

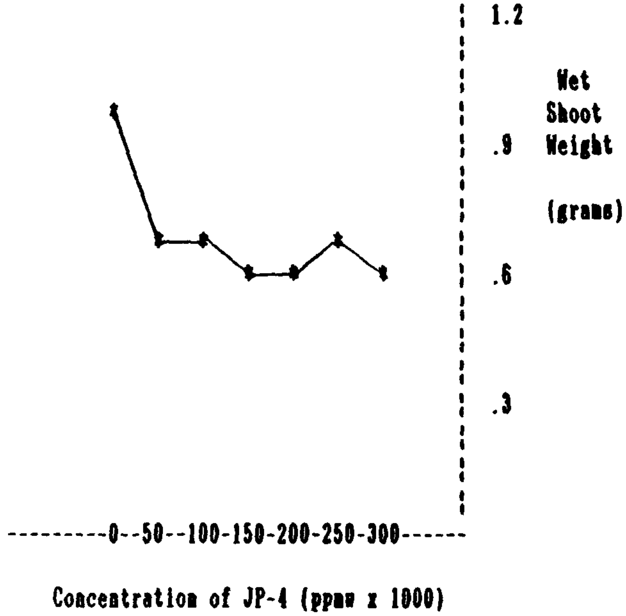


Fig. 24. The effects of JP-4 on the shoot length of soybeans grown in commercial potting soil.

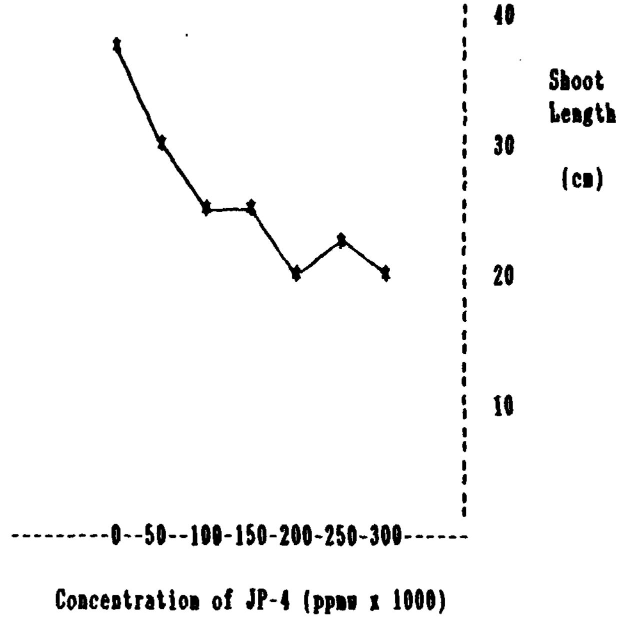


Fig. 25. The effects of JP-4 on the wet shoot weight of wheat grown in commercial potting soil.

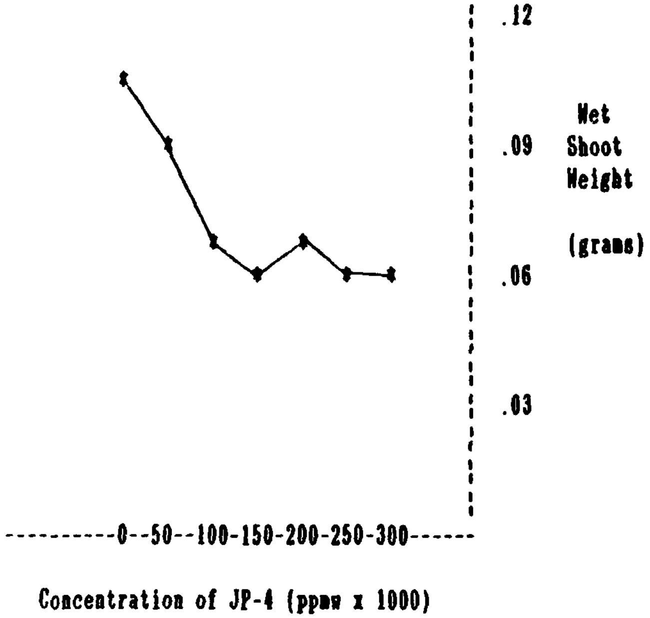


Fig. 26. The effects of JP-4 on the shoot length of wheat grown in commercial potting soil.

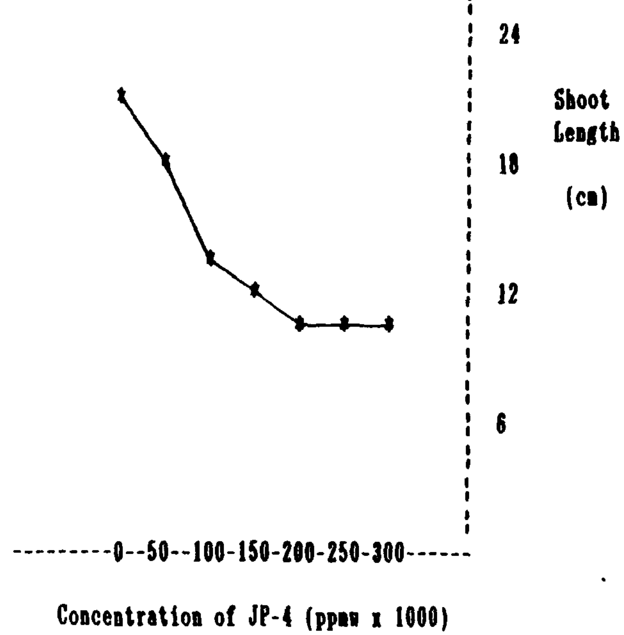


Fig. 27. The effects of JP-4 on the wet shoot weight of oats grown in commercial potting soil.

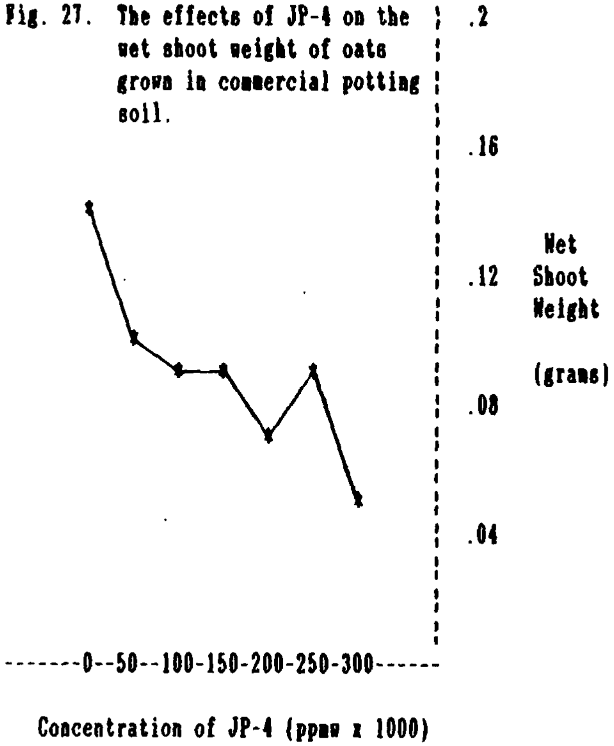


Fig. 28. The effects of JP-4 on the shoot length of oats grown in commercial potting soil.

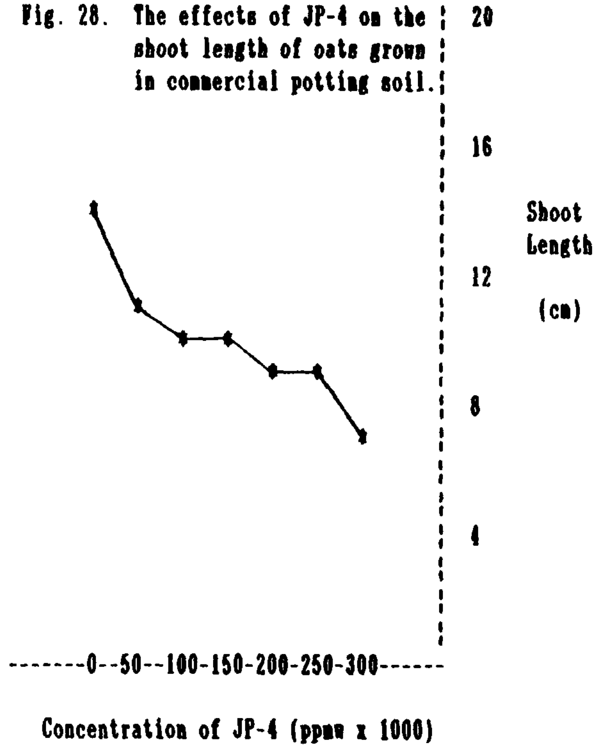


Fig. 29. The effects of JP-4 on the wet weight of cucumber shoots grown in commercial potting soil.

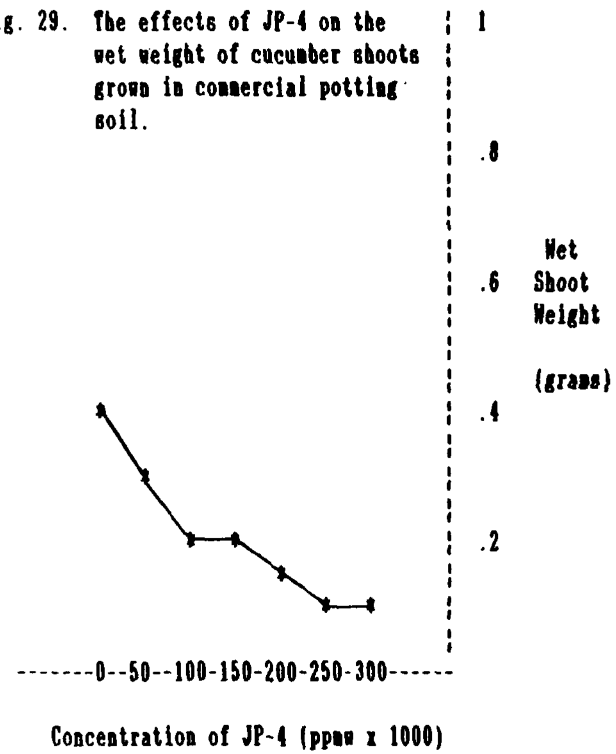
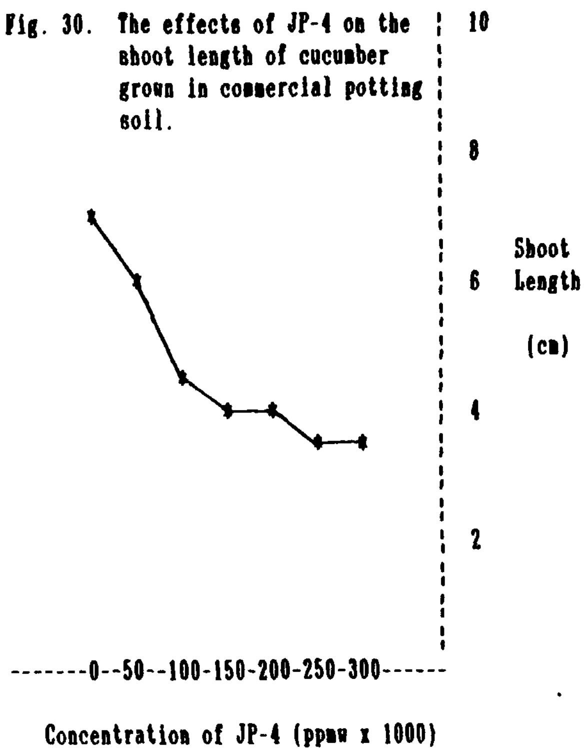


Fig. 30. The effects of JP-4 on the shoot length of cucumber grown in commercial potting soil.



1988 USAF-UES RESEARCH INITIATION PROGRAM

Sponsored by the
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Conducted by the
UNIVERSAL ENERGY SYSTEMS, INC.

FINAL REPORT

AN IMPACT STUDY FOR THE CONTRACTING OUT OF IN-HOUSE ANALYTICAL SERVICES AT THE USAF OCCUPATIONAL & ENVIRONMENTAL HEALTH LABORATORY - BROOKS AFB, SAN ANTONIO, TEXAS

Prepared by:	Don E. Deal, Ph.D.
Academic Rank:	Assistant Professor
Department and University:	Industrial Engineering, University of Houston
Date:	27 September 1988
Contract No.:	S-760-7MG-096

UNIVERSITY OF HOUSTON
DEPARTMENT OF INDUSTRIAL ENGINEERING

Don E. Deal
Assistant Professor
Industrial Engineering

James P. Ignizio
Professor & Chair
Industrial Engineering

AN IMPACT STUDY FOR THE CONTRACTING OUT OF IN-HOUSE ANALYTICAL SERVICES AT THE USAF OCCUPATIONAL & ENVIRONMENTAL HEALTH LABORATORY- BROOKS AFB, TEXAS

by

Don E. Deal

ABSTRACT

A study was performed, the key elements of which are outlined in this report, to assess the feasibility of acquiring a small group of contract personnel on-site at the USAF-OEHL to assist in the processing of various sample analyses requests, and to estimate the cost-savings (if any) of such an action and its effects on sample turn-around times. Projections suggest that costs can be reduced significantly and turn-around times improved with the addition of a small number of contract chemists and lab technicians. Specifically, through the approach described herein, a projected cost savings of approximately $652,000 in the single year of focus in this analysis, is expected.

I. INTRODUCTION:

The USAF/OEHL at Brooks AFB, Texas is that central body in the Air Force Systems Command whose responsibilties are concerned with environmental health and occupational safety. The laboratory's activities represent a wide range of duties, including evaluation of procedures and hazards, review of standards and legislation and comprehensive analysis of air, water, soil, and biological samples and industrial materials. It is this last activity which represents the focus of the laboratory's operations and which employs the efforts of a number of civil servant and military chemists, technicians and supervisors in the Analytical Services Division.

Over the past several years the laboratory has witnessed a substantial growth in the number of sample analysis requests received per year. This growth is comprised of a steady increase in the demand for traditional analytical services plus a widening in scope of services requested. During most of this period it has been difficult to acquire new personnel to handle the increased workload owing to hiring freezes and staffing limitations (in both civil and military quarters). While the OEHL has added workspace, purchased new equipment, and initiated a modernization program --- the result of which has been an observable increase in productivity --- the manpower shortage has had a major impact. Transshipment of samples to contract labs for analysis reached record levels last year; and sample turn-around times, though showing evidence of improvement in some areas, still exceed nominal guidelines by a significant margin.

Given the staffing restrictions noted above, addition of on-site personnel via contracted labor appears to be a viable option for improving operating performance. A significant advantage to be considered in evaluating this possibility is the fact that contract personnel numbers may be varied according to need; this flexibility does not now exist with government employees to any major degree. This study, then, has been undertaken to ascertain what areas of the laboratory most require additional support and what number of persons might be appropriate at present; these determinations are made with regard to their impact on cost and receipt to check-out times.

II. BACKGROUND: PRESENT STATE OF OPERATIONS

To assess trends in and the present state of analysis operations, a variety of sample processing statistics (for both in-house samples and transshipments to contract labs) were collected and reviewed for a 46-month period from May 1984 through February 1988. Table 1 below shows the annual figures for total sample receipts over this 4 years of history; during this period the number of sample analysis requests increased by 15.4% for an average per annum growth rate of 4.9%.

TABLE 1: SAMPLE RECEIPTS (MARCH 84 - FEBRUARY 88)

12 Month Period	Total Samples Received	Growth	Percent Growth
Mar 87 - Feb 88	84,259	1362	1.6%
Mar 86 - Feb 87	82,897	3957	5.0
Mar 85 - Feb 86	78,940	5940 (est.)	8.1 (est.)
Mar 84 - Feb 85	73,000 (est.)		

The plot on the following page depicts month-by-month sample receipt totals and numbers retained in-house and transshipped for the period. Trend lines superimposed suggest that of the growth in samples received, about 45% are assigned to contract laboratories and 55% analyzed on-site.

Service time guidelines call for a nominal 10 days turn-around for sample analyses. Growth in the workload while work force levels have remained essentially constant has made this target unachievable under current conditions; a more realistic goal of about 15-17 days seems more appropriate with present volumes. A review of the previous 46 month's data for in-house processing does show that a 15-day average turn-around time for all samples has been realized about 1/2 the time (Figure 2); adding transshipped samples, whose check-in to check-out durations are more lengthy, we see that a 17-day or better turn-around has been achieved approximately 40% of the time (Figure 3).

Returning to the in-house processing times in Figure 2, we might take special note of trends during selected periods in the 46 months. From months 1 to 25 (May 84 through May 86) an upward trend in average sample turn-around times is observed; probable reasons for this positive slope are the rather large growth in samples received over this period plus disruptive effects felt from lab expansion efforts (e.g., relocation of various stations, change-out of equipment, etc.). From June 86 through May 87 (months 26 to 35) the trend is fairly flat. In the final period, starting from April 87, turn-around times are somewhat erratic; however, we should note that several values are among the shortest receipt to check-out times recorded in the 46 months.

From our cursory review of the above data we have some suggestion that there may have occurred an improvement in service times during the last year of history. Additional evidence for gains in productivity is indicated when we analyze information reflecting analyses requested per sample and analyses performed per day. Figure 4 displays a dual plot of these parameter values for the same May 84 to February 88 time frame. Up through month 30 the average number of analyses requested per sample was aproximately 3.25 (bottom curve); since that point analyses per sample has been climbing and is now [illegible] about 4.0-4.25. Turning to analyses performed per day in-house (top

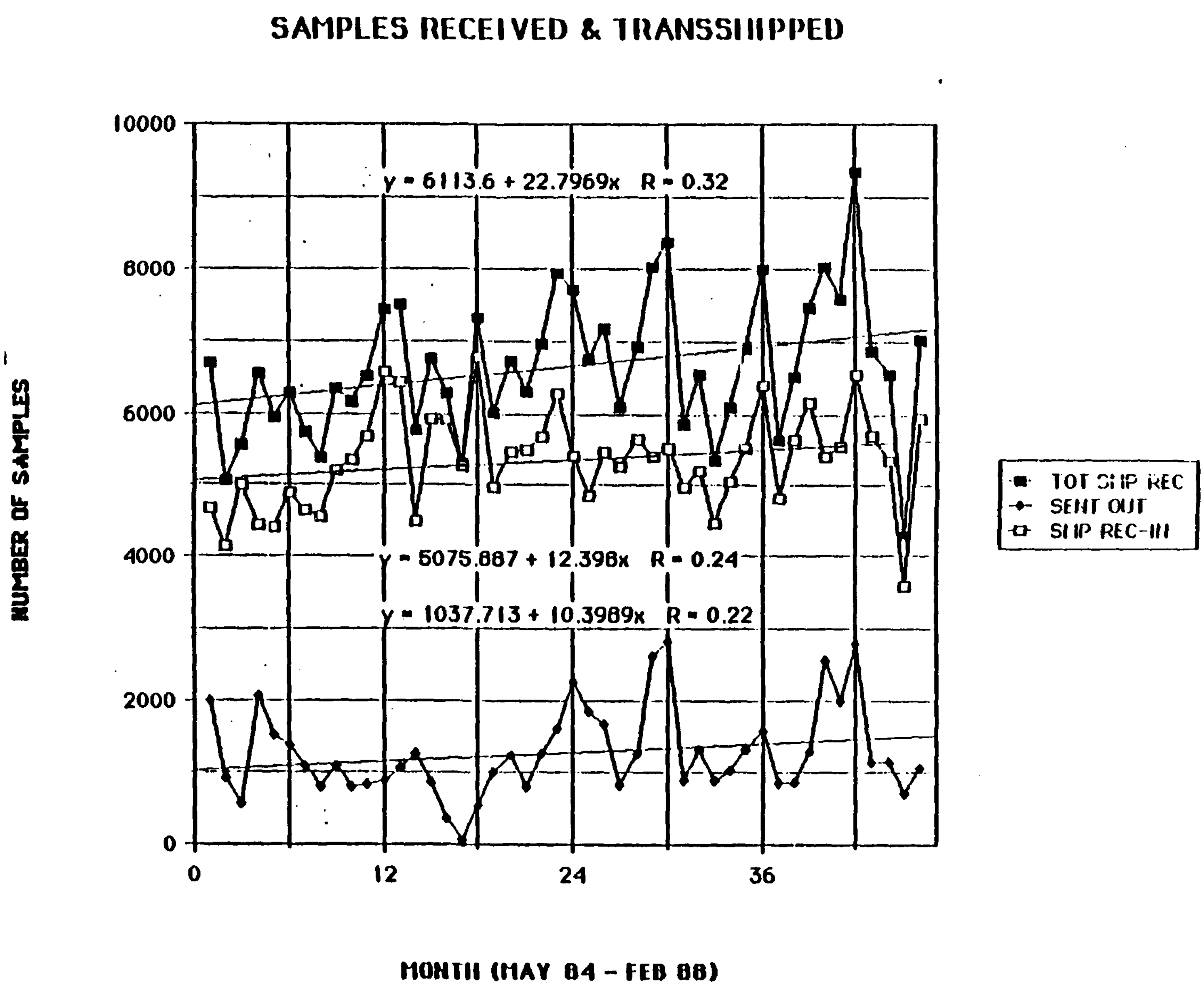


Figure 1. Monthly Sample Receipts: In-House, Transshipped, and Total

(May 1984 - February 1988)

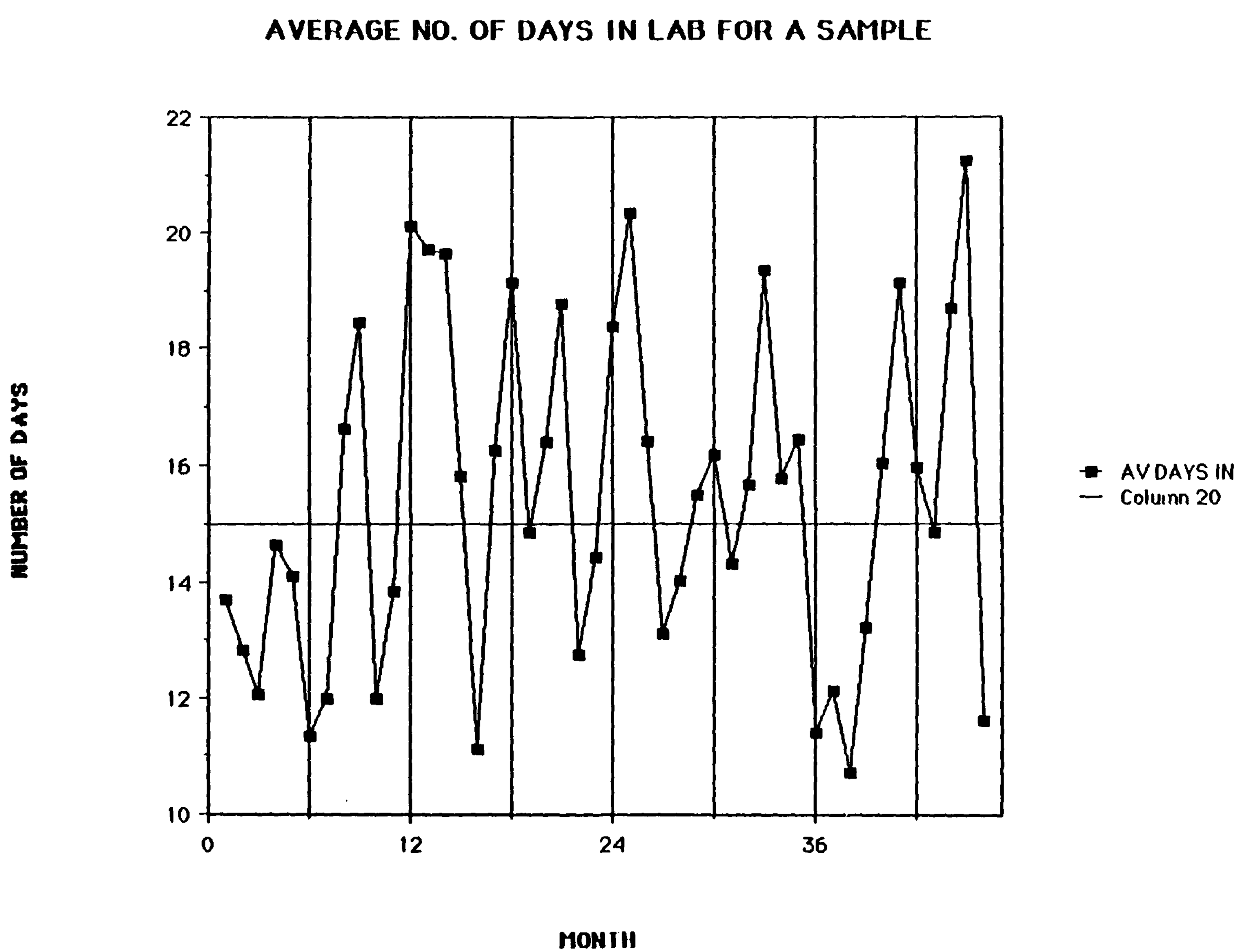


Figure.2. Monthly Average Turn-Around Times For In-House Samples

(May 1984 - February 1988)

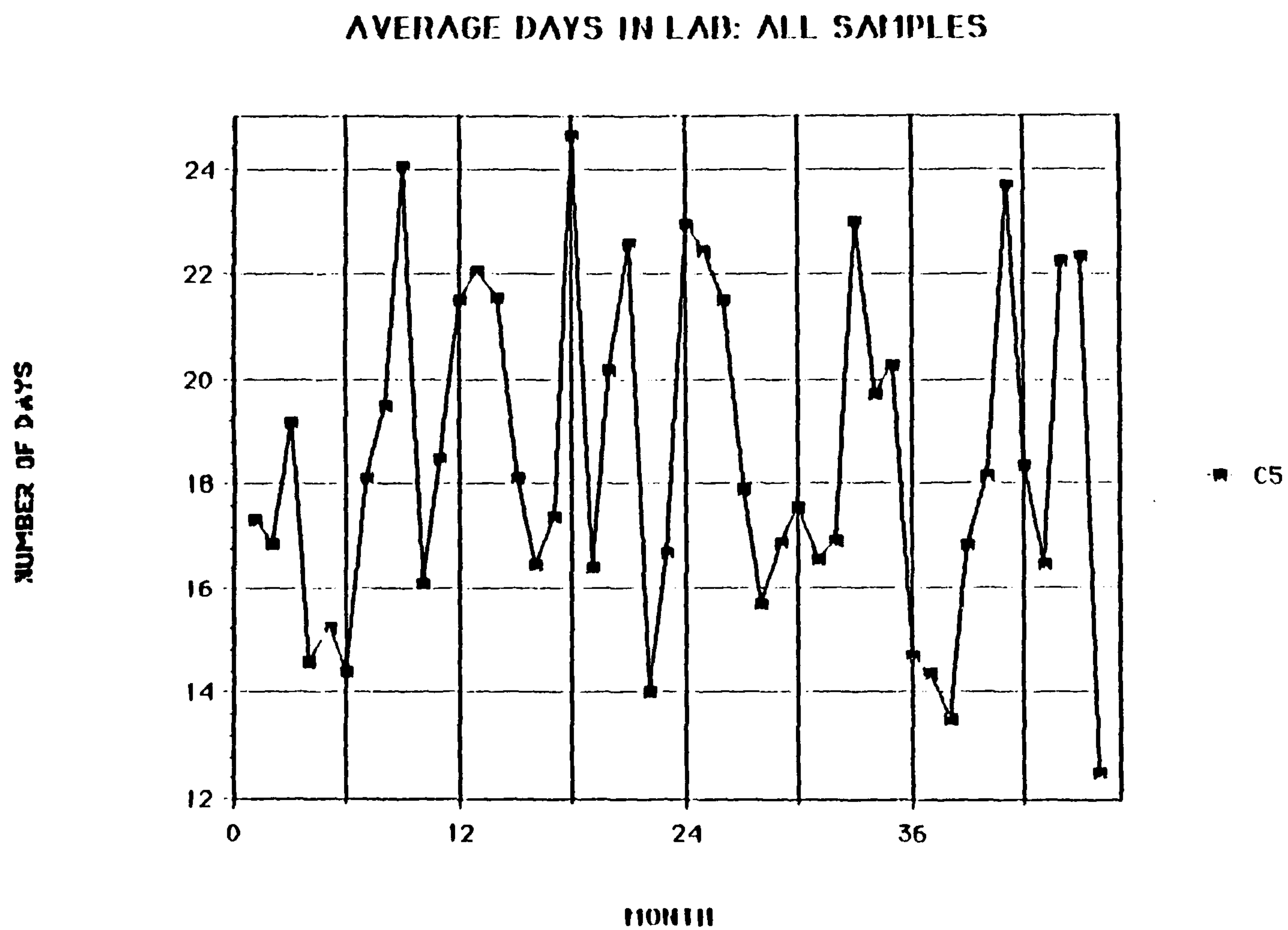


Figure 3. Monthly Average Turn - Around Times - All Samples

(May 1984 - February 1988)

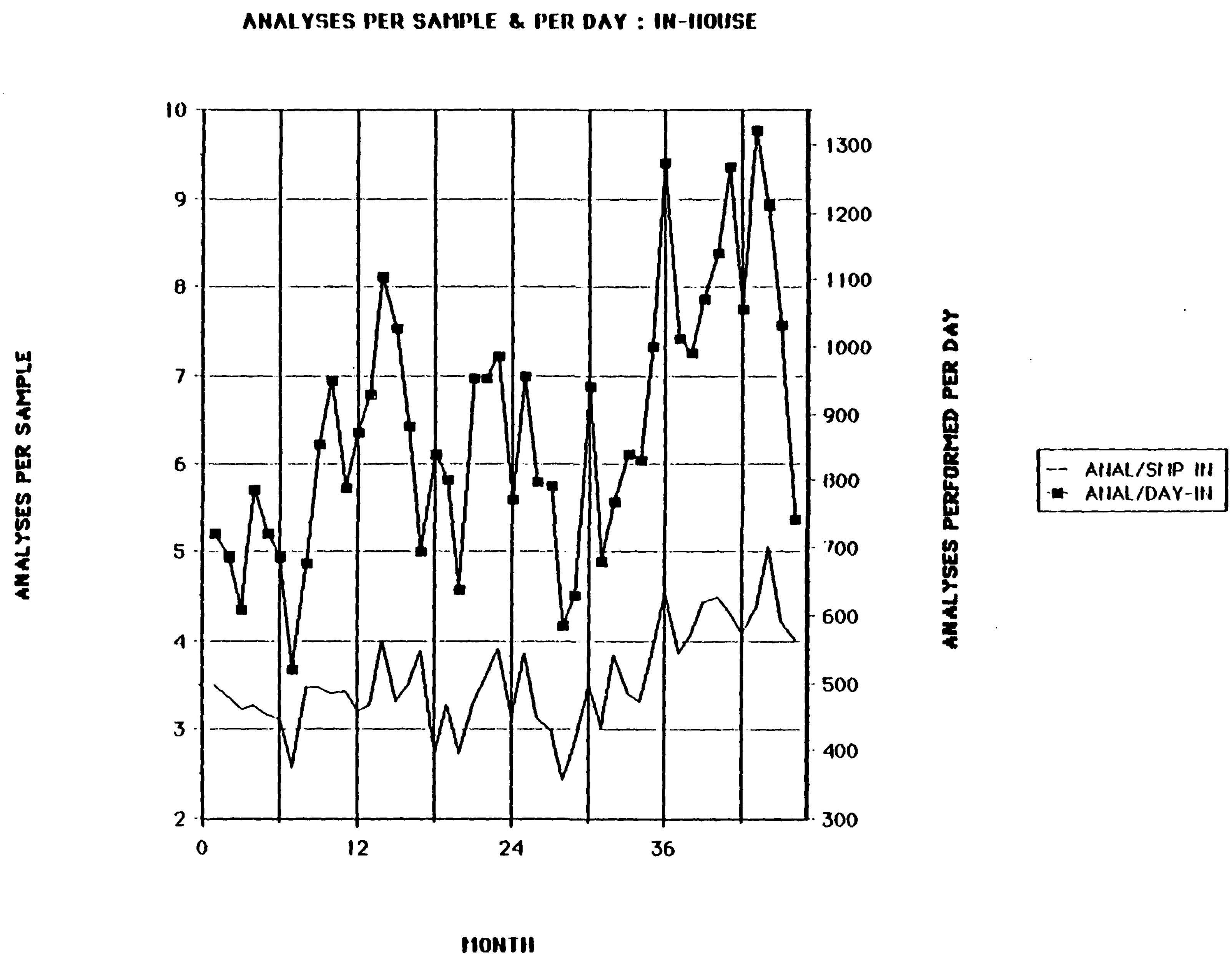


Figure 4. Analyses Requested per Sample and Analyses Performed per Day

(May 1984 - February 1988)

curve), we note that through month 34 the average number performed per day was around 825. However, since then a marked increase in the analyses completed per day has occurred and has averaged about 1025 over the final 12 months, indicating a 24% increase in capacity (the low value posted for February 88 is attributable to the small number of samples received in that final month).

Despite the observed growth in the number of samples received and number of analyses requested per sample, during about the last year of the historical period there is some evidence that average turn-around times may be improving, though still exceeding desirable service time upper bounds. The reason for this improvement is the recent substantial increase in the number of analyses that can be peformed in a work day; this increase, in turn, seems to coincide with the completion of lab expansion efforts and various equipment upgrades.

III. AREAS TARGETED FOR ADDITION OF CONTRACT PERSONNEL

In determining what sections in the Analytical Services Division might be candidates for personnel additions, we focused on two objectives: (1) reducing turn-around times in those lab areas where times appeared excessive, and (2) reduction of proportion of samples transshipped, thereby cutting sample processing costs.

Data was collected and compiled for the various sample codes and indices computed which could be used to compare need for increased work force among the several lab sections. Chief among these indicators were the following:

(1) Average Turn-Around Time - a weighted average (weighted by number of samples) of in-house and contract lab turn-around times

(2) Backlog in Months of Work - in-house + contract lab backlogs as of February 88 divided by the average number of samples completed per month over the latest 1-year period

(3) Ratio of Average Turn-Around Times to Holding Times - for each sample type there is a holding time; analysis of the sample must be completed within this period to avoid sample deterioration.

(4) Ratio of Samples Completed on Contract to Samples Completed In-House

(5) Volume of Samples Completed on Contract

Appendix A contains a series of graphs which list values by sample code for the indicators utilized in reviewing and assessing section need. With this information a prioritization matrix was produced which displayed for each code a ranking in every index category. From this, a preliminary group of sections was chosen for consideration in receiving additional contract workers. At this point a

TABLE 3: SAMPLE CODE GROWTH ESTIMATES

Sample Group	Percent Growth Over 19 Months	Annual Growth Rate
Group 1	11.0%	6.8%
Group 2	6.1	3.8
Group 3	1.8	1.1
HF	0.0	0.0
HK	0.1	0.1
MK	9.0	5.5
All Above	5.8	3.6

In order to attain any real cost savings while bringing in additional work staff, it will be necesary to reduce the proportion of samples transshipped to contract laboratories below current levels in these areas. We have assumed, however, that some fraction of samples in each code group will still need to be contracted out for analysis due perhaps to special testing requirements or temporary backlog problems. Table 4 lists most recent data for proportions transshipped along with target values for the future period of focus. These goals were discussed and deemed reasonable in talks with ASD section supervisors.

TABLE 4: HISTORICAL AND TARGET PERCENTAGES FOR PROPORTION OF SAMPLES TRANSSHIPPED

Sample Code Group	Past Percent Transshipped	Future Percent Transshipped
Group 1	4.5%	1.0%
Group 2	14.0	5.0
Group 3	14.9	5.0
HF	34.0	20.0
HK	23.7	15.0
MK	45.4	30.0

Workload capacities for use in base year projections were determined from per-person sample analysis capabilities in the final year of the historical period. For example, if during the time frame March 87 - February 88 two persons averaged completion of 100 samples per month, then three persons could average 150 samples per month. Reductions in average sample turn-around times were computed analagously --- i.e., on a proportion basis --- with the assumption that present service times are a result of backlogs rather than lengthy analytical procedures. Thus, in the example above, if samples received increased by 50% at the same time that one worker is added, no change in the average turn-around time would be expected.

A final piece of information required in the computation of net change in costs was current charges for contract analyses. As one might imagine, charges vary from sample type to sample type, so that it was necessary to estimate costs for each individual sample code. These estimates are given in the table below. Shipping charges in the amount of $3.50 per sample should be added to obtain total average estimated cost.

TABLE 5: AVERAGE ANALYSIS COSTS FOR SAMPLES TRANSSHIPPED

Sample Code	Average Cost Per Sample
NA,PA	$ 70
NB,ND,NE,PB,PD,PE	40
NC,PC	90
NG,PG	160
ST,PT,NT	250
NH,PH	175
HF	100
HK,MK	125

V. STUDY RESULTS

Detailed calculations as outlined in the previous section were carried out, yielding the results summarized in Table 6. Projected staffing in the sections listed were set at levels as determined by sample growth projections and need to reduce average service times. Turn-around times are averages for all samples (weighted average in-house and transshipped). Estimated cost savings are transshipment savings only --- that is, samples processed in-house that would otherwise have been contracted out times average contract analysis costs; increased employee salaries and other operational costs are balanced against these savings in computations which follow the tabular summary.

TABLE 6: RESULTS SUMMARY - PROJECTED STAFFING, SERVICE TIME REDUCTIONS & TRANSSHIPMENT COST SAVINGS

CODE	EXISTING STAFF	PROJECTED STAFF	TURN-AROUND TIME (DAYS)		COST SAVINGS ($1,000'S)
			CURRENT	PROJECTED	
Group 1	6	8	11.3-14.7	9.6-13.0	146
Group 2	4	6	6.0-28.4	6.4-19.9	338
Group 3	3	6	26.2-29.7	16.9-18.9	74
HF	1.5	2.5	24.0	17.5	69
HK	2	3	18.3	14.7	273
MK	1	2	19.3	15.1	299

Reductions in sample service times through addition of contract workers averages approximately 20% for the groups displayed in Table 6, and total transshipment savings for all groups sums to $1,199,000. Increased operational costs as a result of adding personnel must now be estimated to arrive at projected net cost savings. Turning our attention first to increased salary expenditures, we note that bench personnel levels have been stepped up to 27.5 workers, up from 17.5 persons previously. Division managerial personnel judged that the composition of this increase should be two chemists and eight lab technicians, whose salaries we have estimated at $35,000 and $25,000 respectively. Heading up this contingent of contract staff will be a contract supervisor at a salary of $42,500, bringing annual contract salary totals to $312,500. Employee benefits (e.g., insurance, retirement, etc.) were projected to be 40% of salaries or $125,000; and to cover increased overhead and contractor profits, a figure of 35% of annual salaries ($109,375) was used. We therefore arrive at a total of about $547,000 in increased operational costs as a result of employing on-site contract personnel, suggesting a net savings of $1,199,000 - $547,000 = **$652,000** for this one year period alone.

VI. CONCLUSIONS

The analysis performed herein is based on "normal" case assumptions. Computations show that for samples previously projected for transshipment which could, instead, be processed in-house with on-site contract labor support, analysis costs run, on average, about $62 per sample versus about $135 per sample when transshipped. A "worst" case scenario analysis was also performed based on significant workload increases in the PT and PH sample code groups resulting from ramifications of the Safe Drinking Water Act, and data for this case indicate even larger savings (see Appendix B).

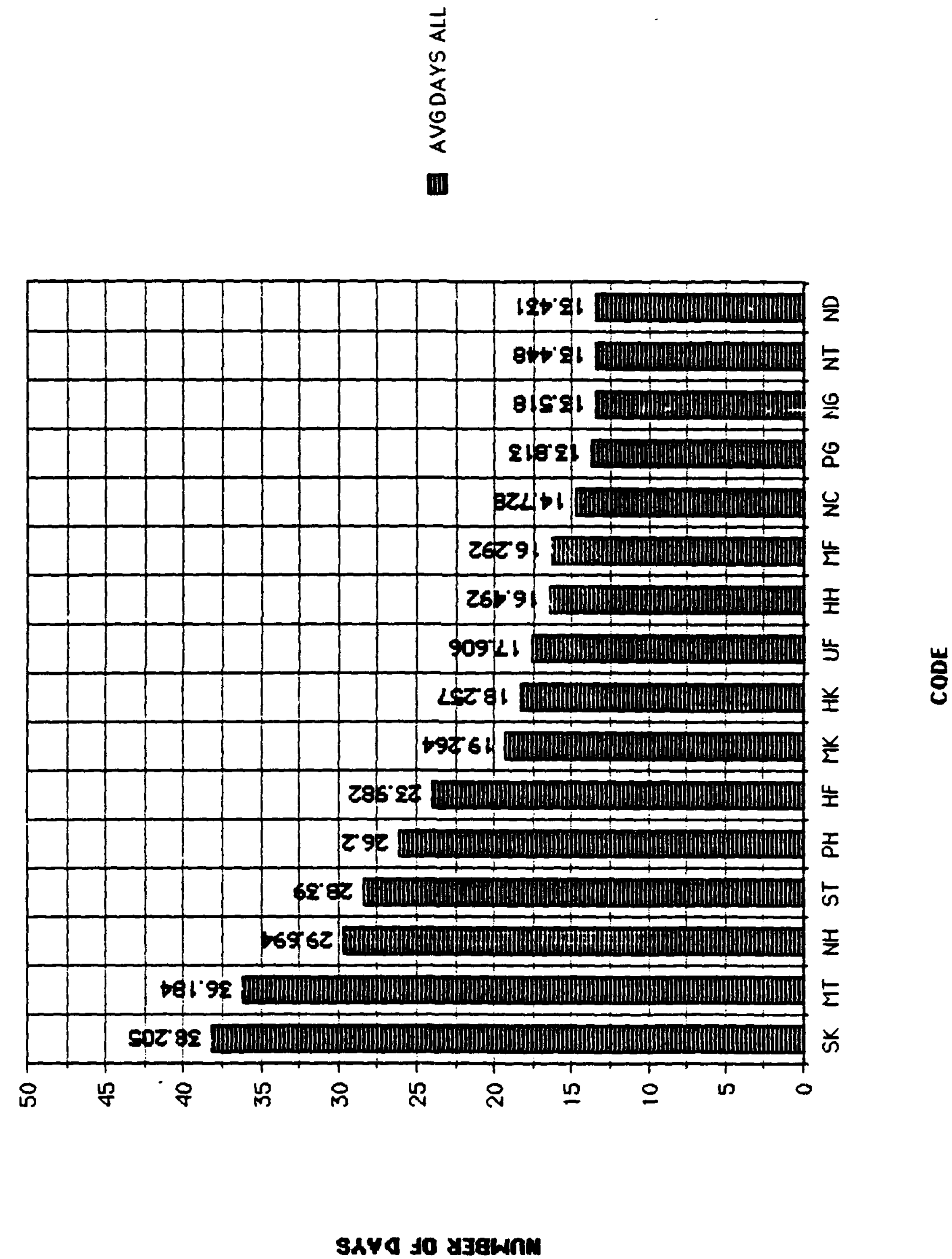
The positive results arrived at in this analysis surely lend support to testing of this strategy. We are convinced that need is real and are optimistic that performance will further improve with their use. However, if the growth projected in the various sample classes is not realized, the flexibility retained by OEHL regarding utilization of contract services permits their release. Thus, there appears to be little risk in employing a contract contingent, if only for a "trial" period.

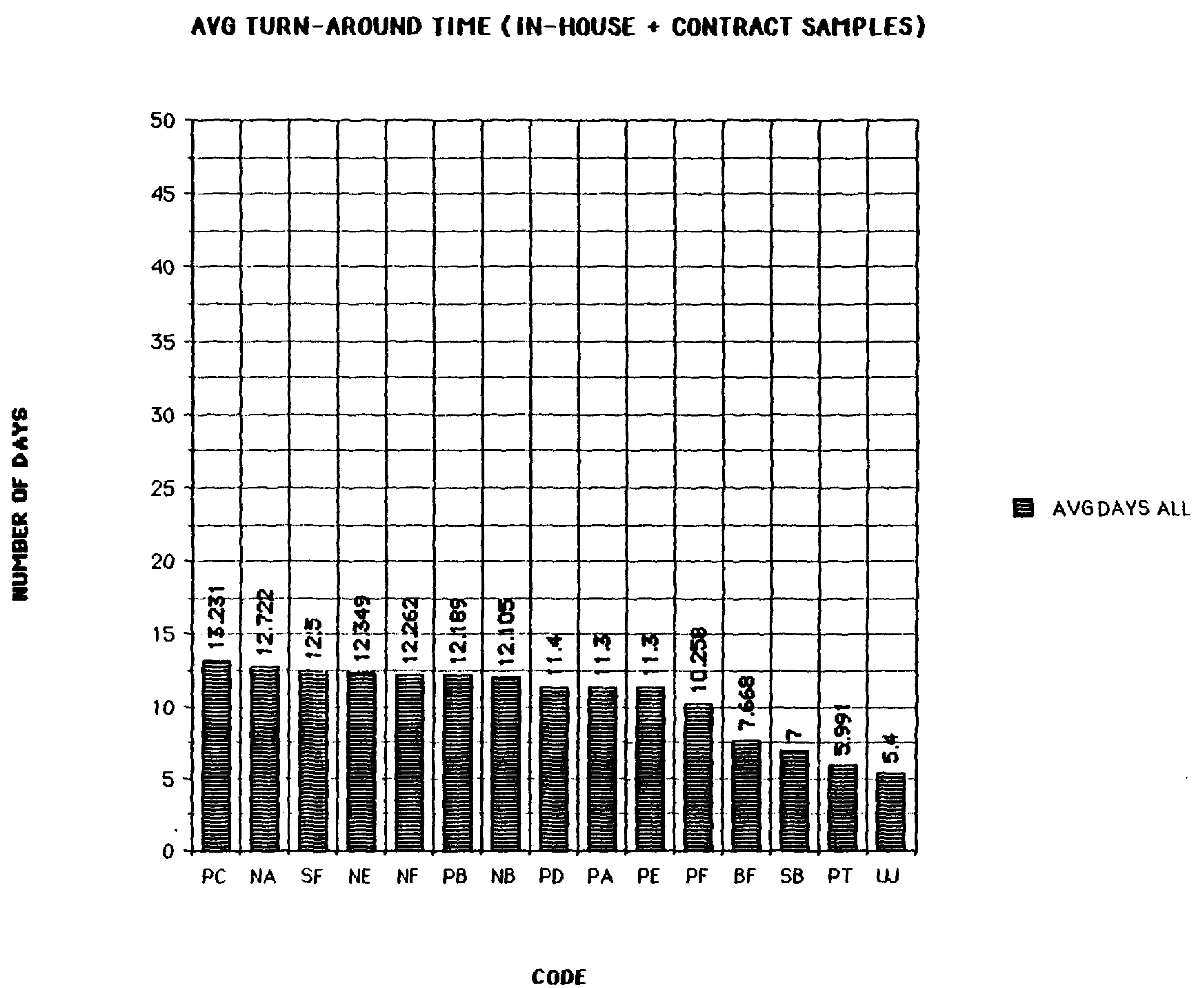
Given recent trends in the budgeting for services derived from contract laboratories, we note in a final word that it remains to study the problem of transshipment assignment control. Whereas in the past budgeting levels permitted freer use of analytical services through transshipment, present uncertainties underscore the need for tools and procedures through which assignments are made. We shall address specific problems and suggest solution directions in a forthcoming white-paper.

APPENDIX A: INDICATOR GRAPHS

As explained in Section **IV** of this report, data was collected and compiled from which various indices were computed for each sample code. These indices reviewed together aided the determination of which codes most required added personnel support in handling present and projected workloads. The graphs which follow display indicator values computed for the final 1-year period of historical data (March 87 - February 88).

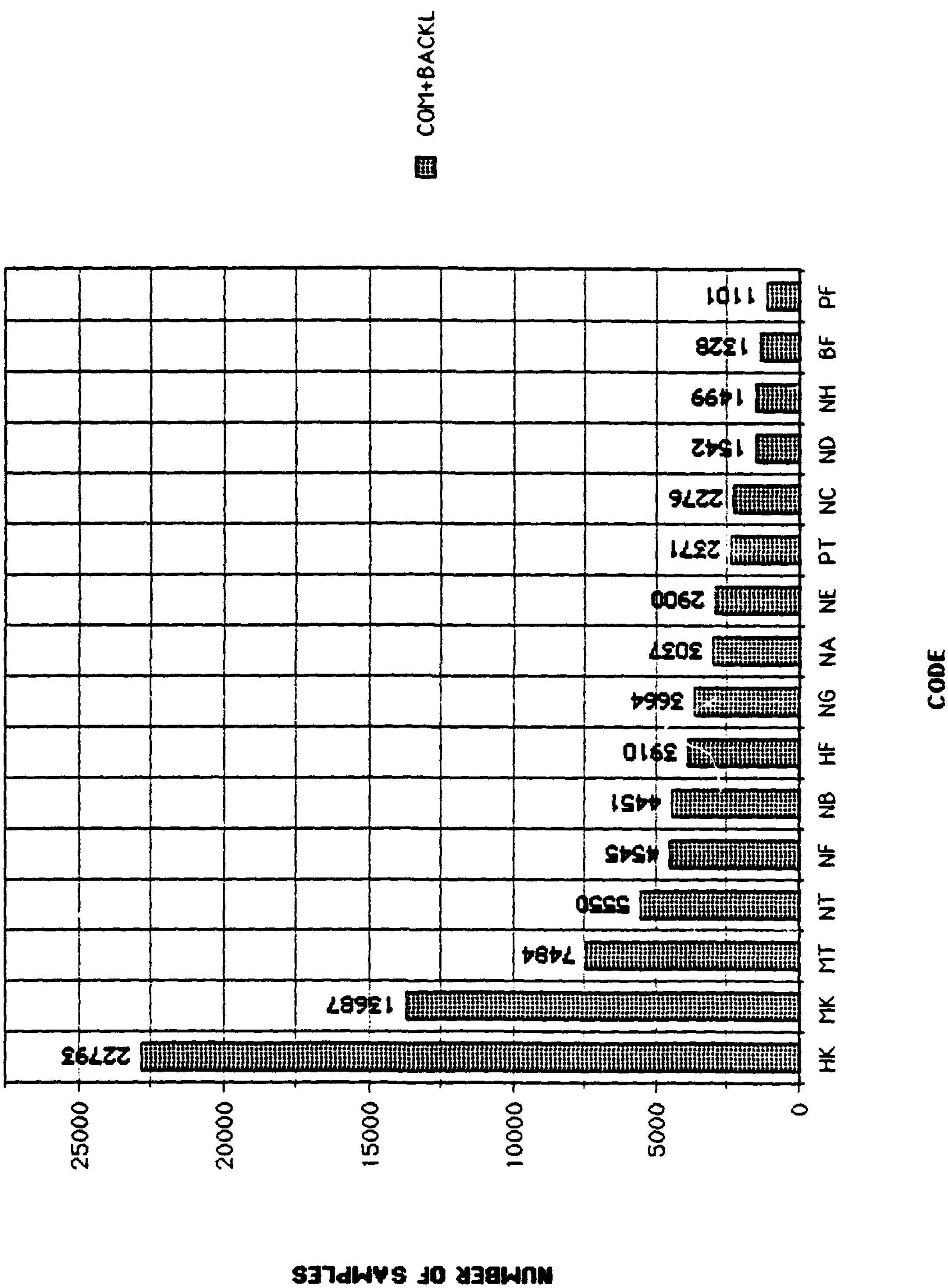
AVG TURN-AROUND TIME (IN-HOUSE + CONTRACT SAMPLES)



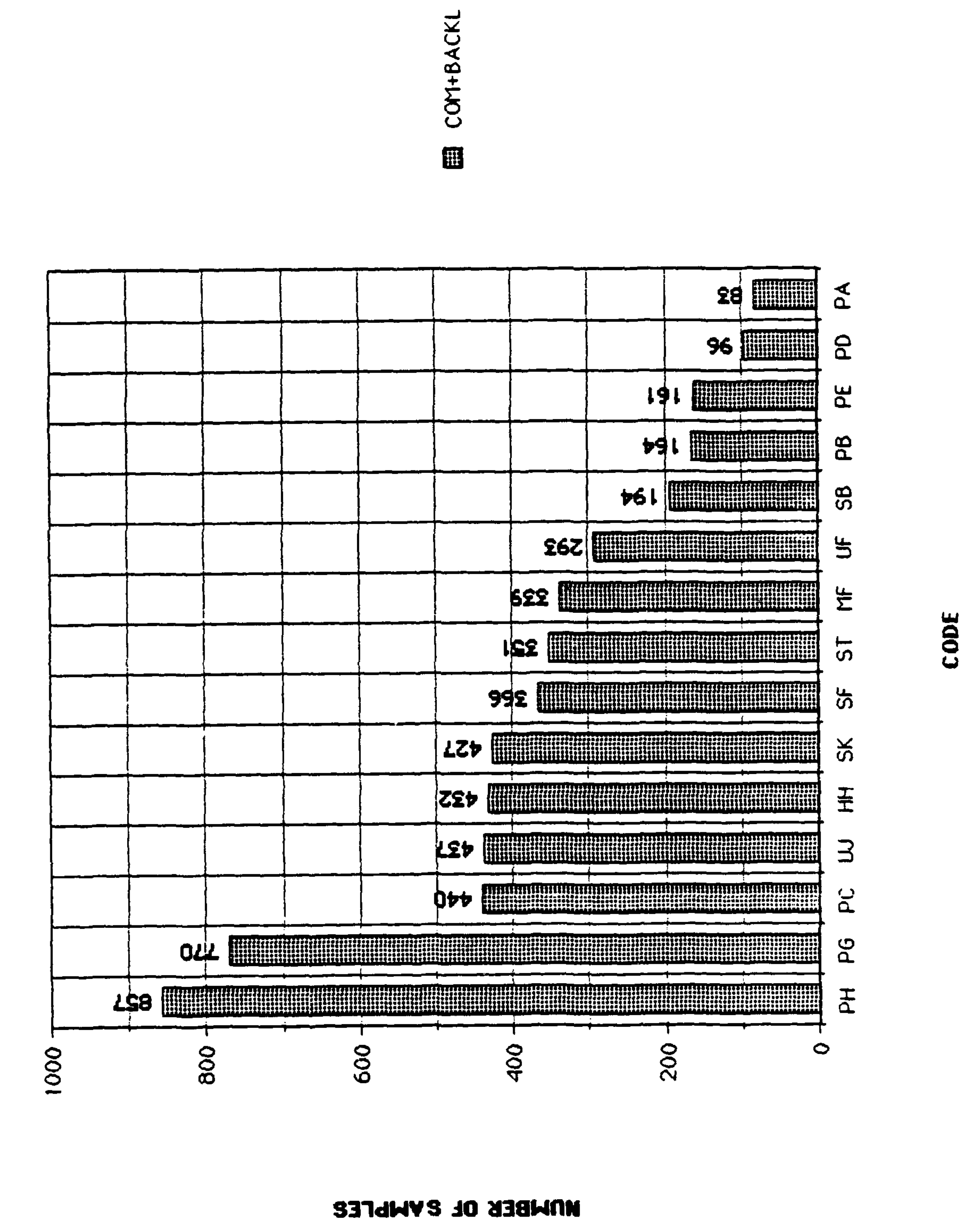


AVG TURN-AROUND TIME (IN-HOUSE + CONTRACT SAMPLES)
NUMBER OF DAYS
50
45
40
35
30
25
20
15
10
5
0
13.231
12.722
12.5
12.349
12.262
12.189
12.105
11.4
11.3
11.3
10.258
7.668
7
5.991
5.4
PC
NA
SF
NE
NF
PB
NB
PD
PA
PE
PF
BF
SB
PT
UJ
CODE
AVGDAYS ALL

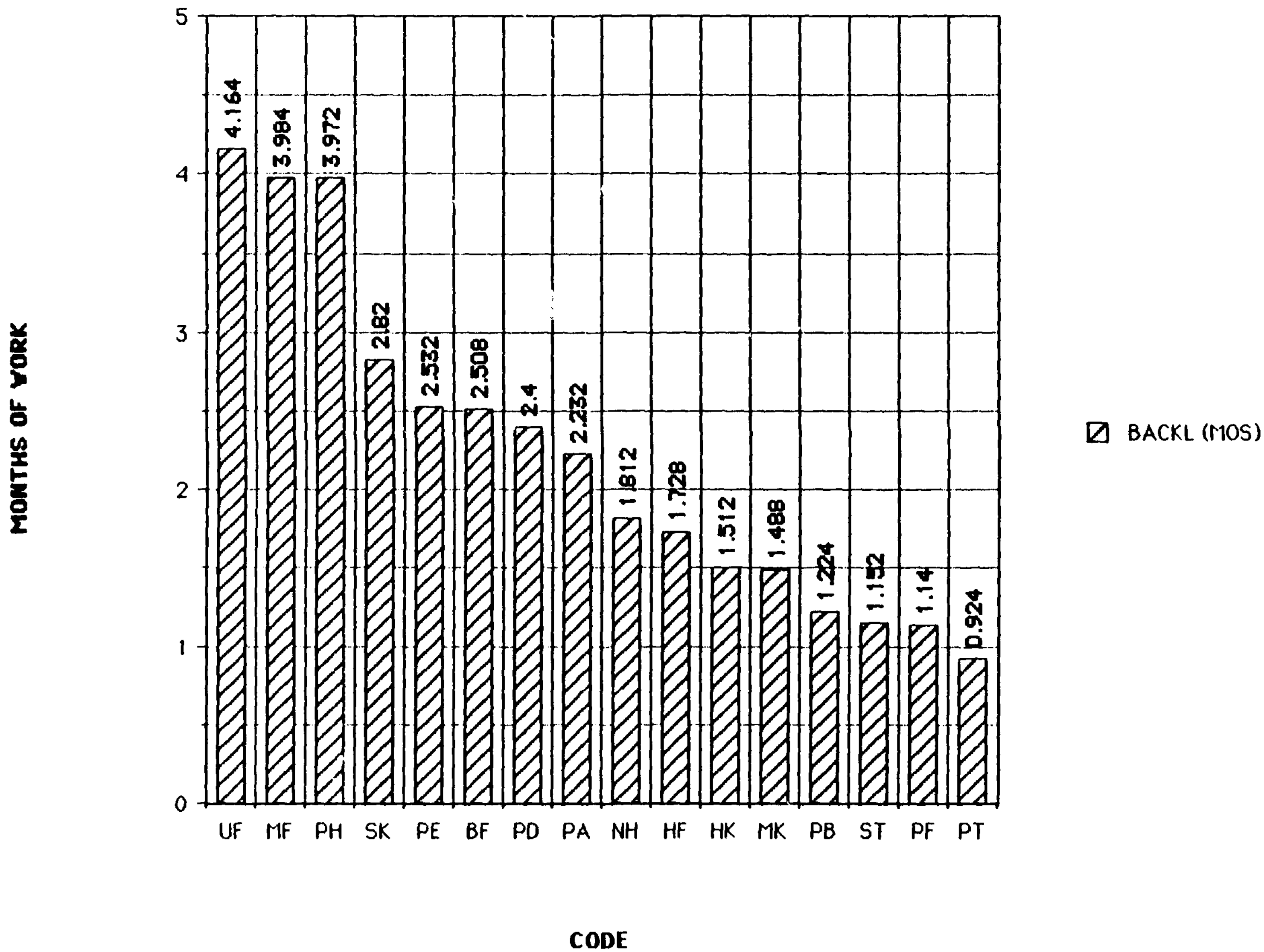
COMPLETED + BACKLOG: MAR 87 - FEB 88



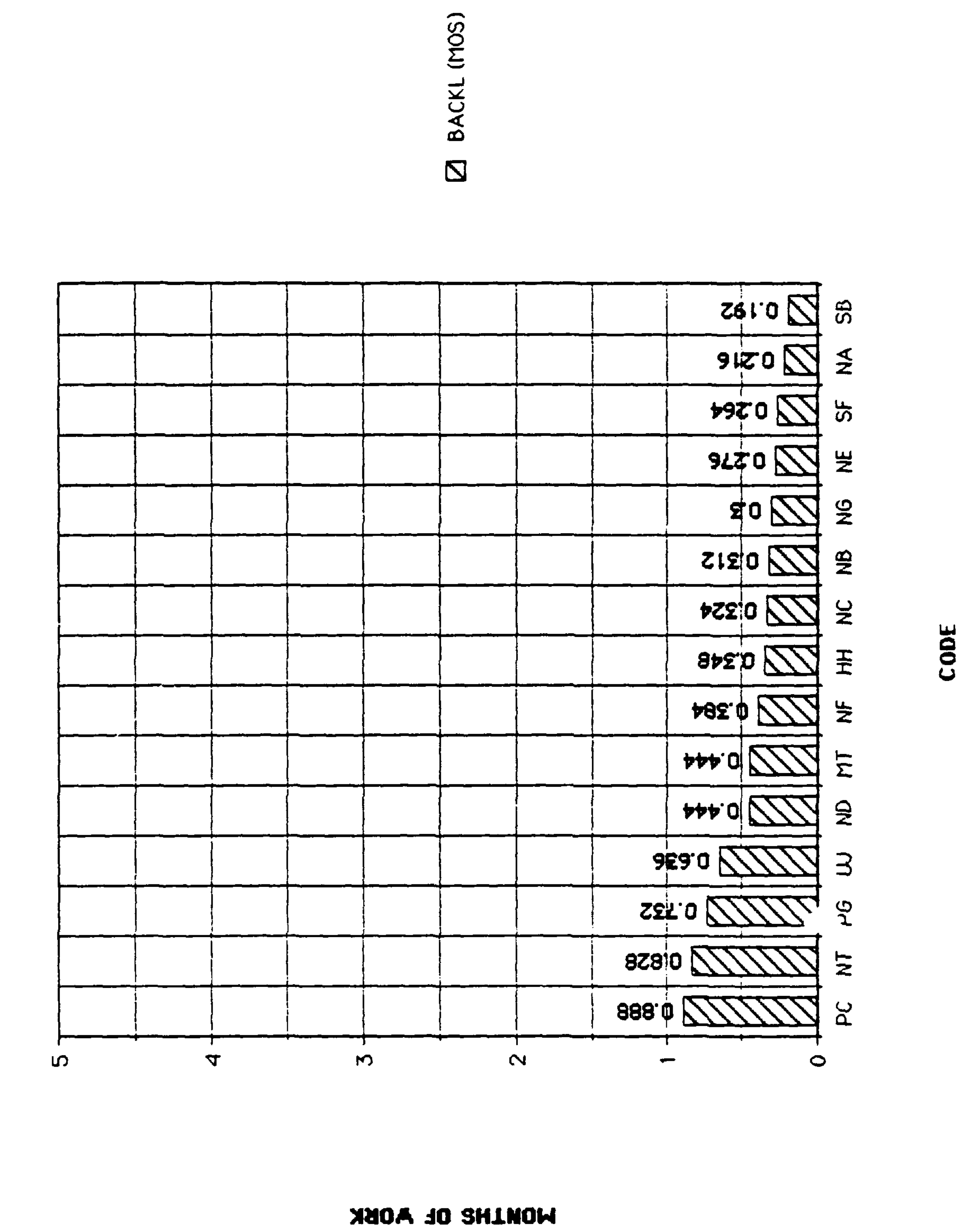
COMPLETED + BACKLOG: MAR 87 - FEB 88



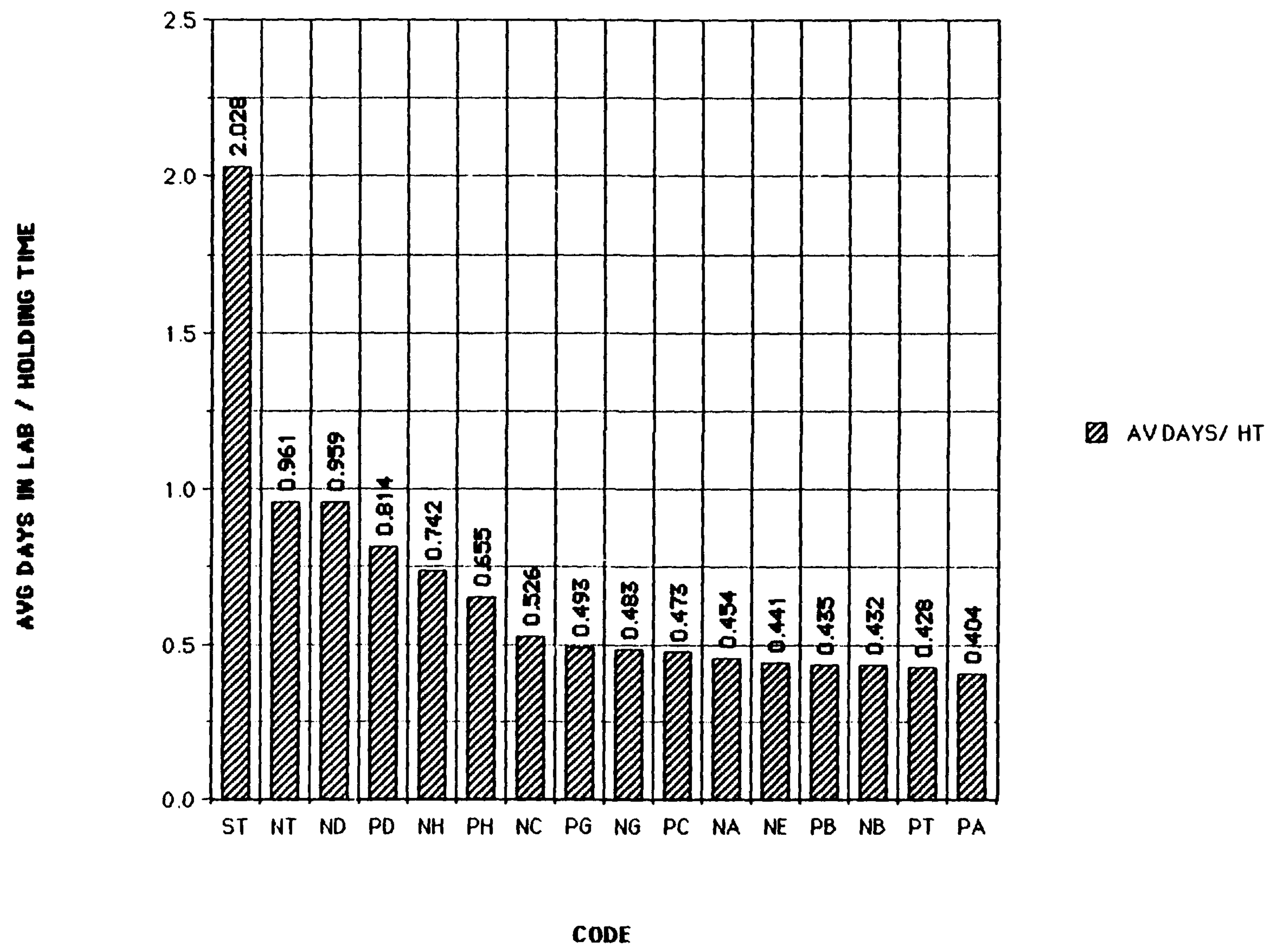
BACKLOG IN MONTHS OF WORK: IN-HOUSE + TRANSSHIPPED

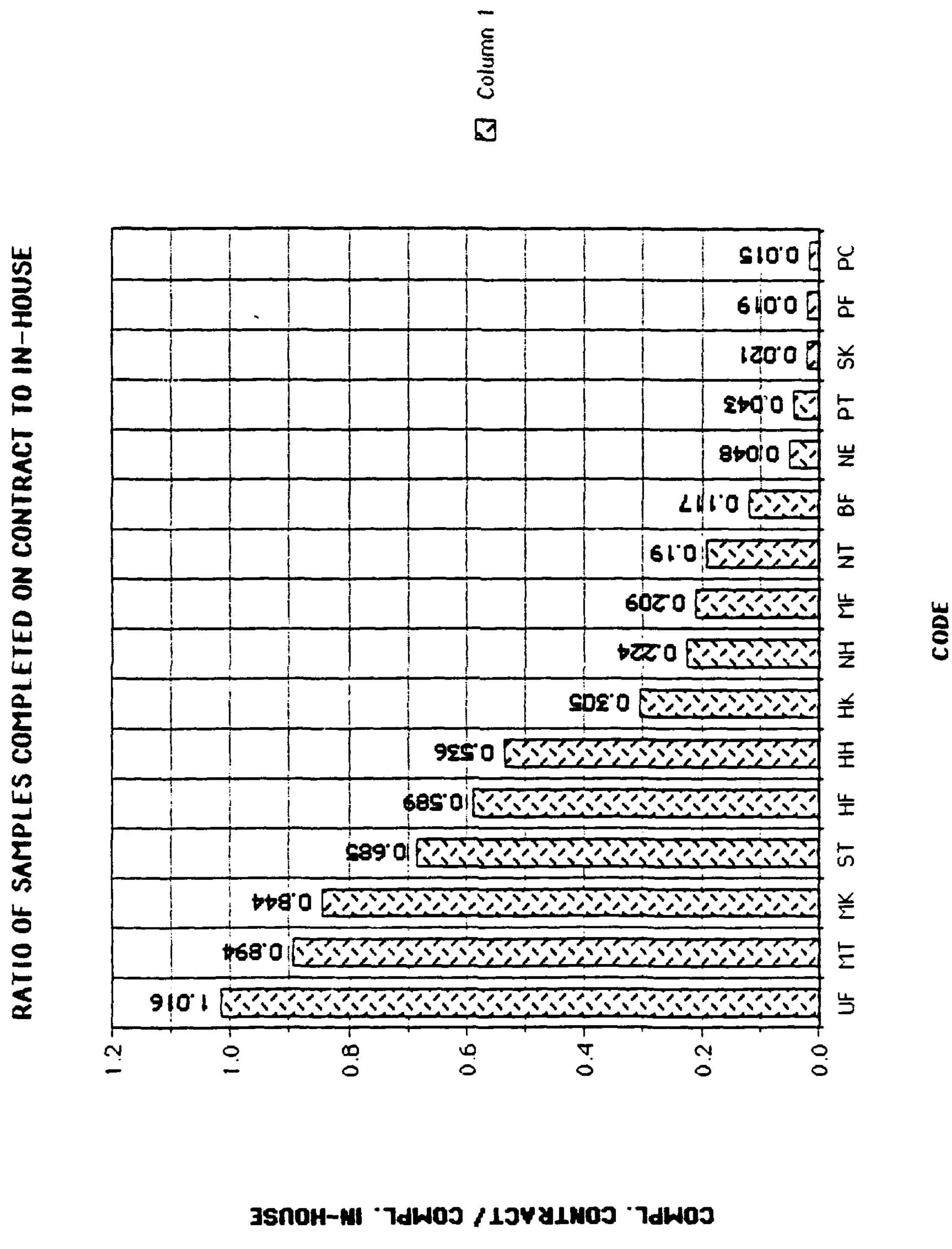


BACKLOG IN MONTHS OF WORK: IN-HOUSE + TRANSSHIPPED

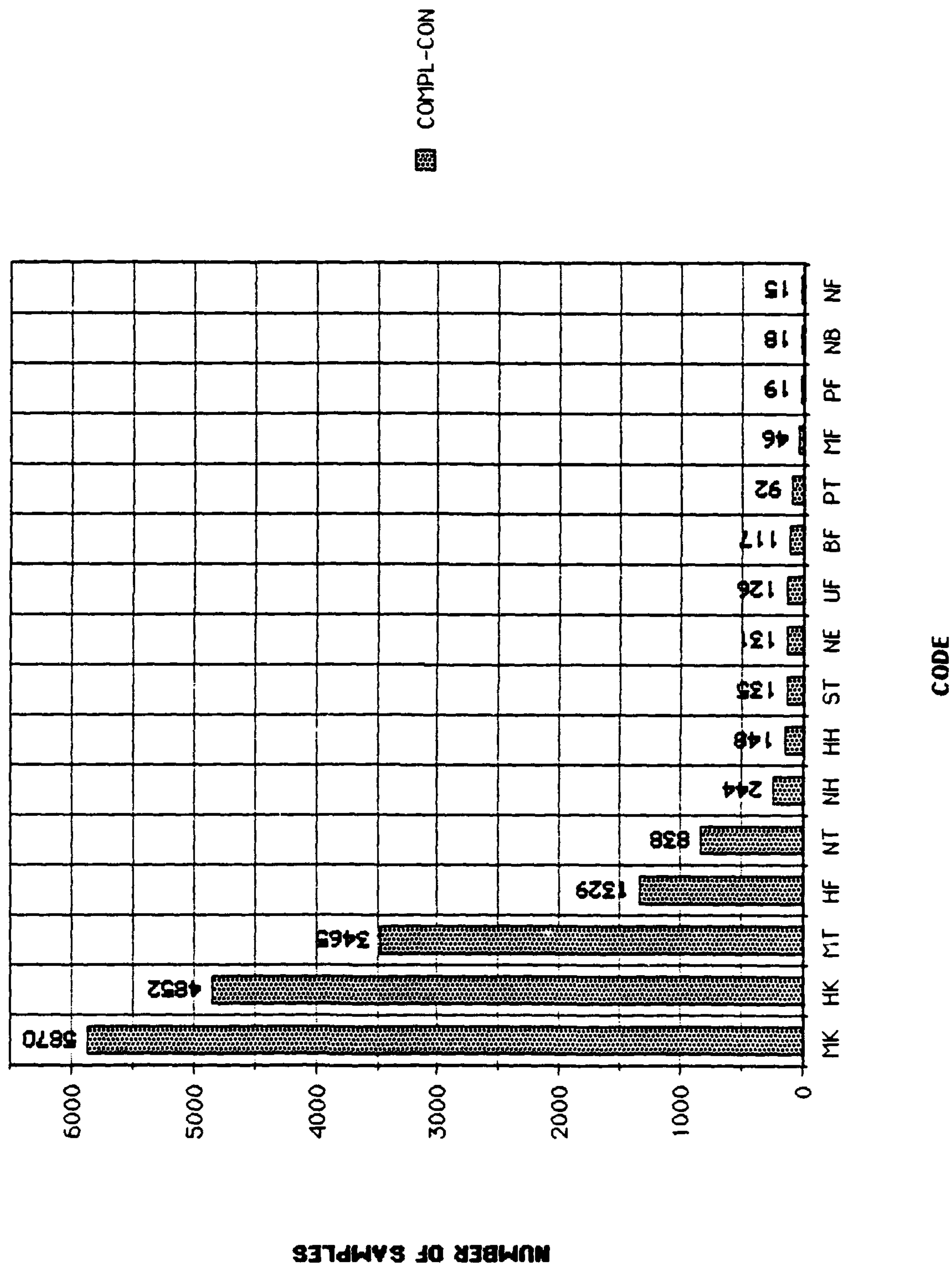


RATIO OF TURN-AROUND TIMES TO HOLDING TIMES




RATIO OF SAMPLES COMPLETED ON CONTRACT TO IN-HOUSE
Column 1
COMPL. CONTRACT/ COMPL. IN-HOUSE
CODE
1.2
1.0
0.8
0.6
0.4
0.2
0.0
UF 1.016
MT 0.894
MK 0.844
ST 0.685
HF 0.589
HH 0.536
Hk 0.305
NH 0.224
MF 0.209
NT 0.19
BF 0.117
NE 0.048
PT 0.043
SK 0.021
PF 0.019
PC 0.015

SAMPLES COMPLETED BY CONTRACT LABS



APPENDIX B: SAFE DRINKING WATER ACT ANALYSIS

Analysis results presented earlier in the body of this report were posited on the assumption of normal growth in samples received over the period of time considered in this study. As an addendum we offer here a brief estimate of the impact of growth resulting from increased requests under the Safe Drinking Water Act.

This legislation could translate into tremendous increases (on the order of 350%) in the number of samples received under particular potable water codes as shown below in data representing projected estimates of new receipts in this category:

1989 - 8400 (PT)   1990 - 14,300 (PT & PH)   1991 - 10,400 (PH)

Since we have focused on the 12 months ending September 90 for purposes of this study, the numbers above were broken down and re-figured for this period and apportioned between the two codes listed above (PT & PH). Nominal case increases were then replaced by these new values. These changes affected earlier computations for sample code Groups 2 & 3, so that new calculations were carried out with the following results:

CODE GROUP	EXISTING STAFF	PROJECTED STAFF	TURN-AROUND TIME (DAYS)		COST SAVINGS ($1,000'S)
			CURRENT	PROJECTED	
Group 2	4	11	6.0-28.4	7.7-14.5	1,888
Group 3	3	13	26.2-29.7	18.9-20.5	409

Adding these adjusted savings figures to previous savings for groups unaffected by the act (Table 6), we compute a total transshipment savings of $3,105,000. This figure certainly seems incredible, but note that approximately 60% of this value derives from savings in Group 2 wherein some 6300 samples which would have otherwise been contracted out at around $275 per sample are now earmarked for processing in-house. Note, also, that in our analysis before, only 10 aditional bench personnel were required; in the present case 22 are needed as follows: 6 chemists, 16 lab technicians. We should register here our concern regarding space limitations and the difficulties in accomodating this number of additonal workers in the event that this scenario proves accurate. Increased annual salary figures now total $652,500, and benefits, increased overhead, and contractor profits come to $489,375. Net savings accrue to approximately $1,963,000.

1987 USAF - UES FACULTY MINI-GRANT RESEARCH PROGRAM

Sponsored by the Air Force Office of Scientific Research

Conducted by the Universal Energy Systems, Incorporated

FINAL REPORT

Prepared by:	Kiah Edwards
Academic Rank:	Professor of Biology
Department and University:	Biology - Texas Southern University
Research Location:	Texas Southern University 3100 Cleburne Ave. Houston, Texas 77004
USAF Researcher:	Dr. Kiah Edwards, III

Period of Performance:	December 15, 1987 - August 31, 1988
Contract No.:	F-49620-85-0013/SB5851-0361

ABSTRACT

EFFECTS OF METAL MUTAGENS ON THE SYNTHESIS AND ACCUMULATION OF MACROMOLECULES

KIAH EDWARDS

Cellular and Molecular Biology, Department of Biology,
Texas Southern University
Houston, Texas 77004

In an effort to further elucidate the profound effects of environmental pollutants on biological systems, this study was undertaken to determine the effects of metal ions on the rate of synthesis and accumulation of total RNA using rabbit uterine DNA as a template. Eight of the metals (cadmium, cesium, cobalt, copper, lead, mercury, selenium and silver) accentuated the initiation of RNA synthesis at concentrations that inhibited overall RNA synthesis. These data indicate that metal mutagens and carcinogens not only decrease the fidelity of DNA replication but also exert their effects at the level of RNA initiation and total RNA synthesis. Additional findings indicate the inhibitory effects of cadmium and beryllium on the synthesis of specific protein in rabbit uterine.

INTRODUCTION

Some metals have been shown to react with electron donor sites on nucleic acids and alter the synthesis of gene products, cause base substitution and point mutations and promote carcinoma and/or tumorous growth. Others such as lead, cadmium and copper have been implicated in the stimulation of RNA synthesis (1), inhibition of DNA synthesis (2), misincorporation of bases during DNA replication (3) and superinducing the synthesis of specific proteins while suppressing others (4). Earlier data (5) have indicated that potassium is essential for the transcription of T4 phage DNA and is believed to enhance the activity of the RNA polymerase by promoting the activity of sigma factor and the premature release of elongated RNA chains. Despite this very dramatic effect of metals on DNA synthesis and RNA synthesis few studies have been conducted on the *in vivo* effects of metals on uterine tissues. In an attempt to study the effects of lead, cadmium and other metals on the synthesis of uterine proteins, the activities of several species of macromolecules critical to the synthesis of proteins were studied, both *in vivo* and *in vitro*. Blastokinin, a protein that is found in the uterine fluid of mammals and serves as biochemical marker for the progesterone-mediated stimulation of uterine endometrium during early pregnancy in rabbits, was a specific protein studied (6). Blastokinin has been indicated to have a profound effect on the growth of rabbit blastocysts *in vitro* and *in vivo*. Because of the abundance of blastokinin in rabbits, its hormone dependent synthesis and the fidelity of its regulation, both replication and transcription, the synthesis of this protein could serve as an outstanding indicator of mutagenic and carcinogenic environmental contaminants, such as metals. An attempt was made to correlate the *in vitro* data with the *in vivo* findings. The results presented in this paper indicated that the heavy metals inhibited the overall synthesis of RNA, while stimulating the initiation of RNA synthesis and suppressed the accumulation of blastokinin and total proteins in the uterine secretions. The data indicated, also, that metals required for nucleic acid synthesis, suppress the activity when used in varying concentrations.

MATERIALS AND METHODS

Endometrial Flushings

Young, sexually-mature New Zealand White rabbits were obtained from a commercial supplier and maintained under laboratory conditions at least seven days prior to use. Estrous does/or those at progressive stages of pregnancy, were killed by cervical dislocation. The uteri were rapidly excised, washed with chilled saline and cut open along the mesometrial surface. The exposed endometrium was scraped from the surface with a scalpel blade, and the tissue used for extraction. Total time of recovery was usually less than two minutes at $4^{o}C$.

Extraction and Fractionation of DNA

DNA was extracted from tissue (uterine endometrium before and during pregnancy), following the phenol method of Kirby (7). In some preparations the chloroform-isoamyl alcohol method was used to determine if specific DNA;s are selectively lost during deproteinization with phenol (8). Homogenized tissues were resuspended in 0.15M NaCl, 0.015M sodium citrate, pH 7.0 (SSC) and lysed by adding sodium lauryl sulfate (SLS) (Final concentration 0.4%) to the mixture and stirring for 15-20 minutes. The lysate was extracted twice at room temperature with an equal volume of phenol preequilibrated with SSC and centrifuged at 20,000 xg for one hour. The aqueous supernatant was dialyzed, overnight, against SSC at $4^{o}C$ on a magnetic stirrer. Thereafter, it was subjected to sequential digestion first with ribonuclease T1 (Grade III: Sigma Chemical Co.) at a final concentration of 300 g/ml for 2 hr at $37^{o}C$ with 50 g/ml pancreatic RNase (Calbiochem) for 1 hr at $37^{o}C$ and finally with 200 g/ml pronase (Calbiochem) for 2 hr at $40^{o}C$ in 2 x SSC. After two more SLS-phenol extractions, 2 volumes of 05% ethanol were added, the DNA collected by winding, redissolved in 0.1 x SSC, dialyzed and stored at $4^{o}C$.

These procedures yielded DNA of about 300 million daltons, as determined by analytical band sedimentation (9) from previous extractions. The T4 DNA was a gift from Dr. Errol Archibald, Morehouse College, Atlanta, Georgia.

In vitro Synthesis of RNA

The experiments to determine the effects of metal mutagens on the synthesis of RNA were carried out using rabbit uterine DNA as a template. Each metal (cadmium, lead, silver, cesium, etc.) was tested at a minimum of six different concentrations, 0.1-2mM, for effects on overall RNA synthesis. The effects of the metals was monitored by the incorporation of ^{3}H-AMP, 32P-ATP and ^{32}p-GMP into RNA. The reaction mixture (0.2ml) for measuring the overall rate of RNA synthesis consisted of 10mM Tris (pH 8.1); 20mM-mercapto-ethanol; 10mM $mgCl_2$, ^{3}H-ATP, GTP, CTP, and UTP at 0.25 mM each; 10 g rabbit DNA and 10 g RNA polymerase. Background counts were determined with reaction mixtures devoid of enzymes or DNA. The overall determination of incorporation was assayed on filters with the Beckman LS 800 system and plotted, the percent incorporation of ^{3}H-AMP versus the concentration of metal. The enzymes and RNA precursors were obtained from New England Nuclear, Boston.

In vivo Sysnthesis of RNA

In order to study in vivo protein we injected a saline (0.85%) suspension of the metal into the lumen of one uterine horn and for the control, the saline carrier alone into the other horn. The uterine fluids were harvested on day five of pregnancy and the protein content analyzed by cellulose chromatography and the Lowry method. The endometrium from each uterus was scraped and the total RNA, protein and DNA determined. These data compliment the levels of blastokinin found in the uterine cavity.

Column Chromatography

Blastokinin was separated from other proteins in uterine flushing using column chromotography. Approximately 2 ml of sample were loaded onto a 1 x 20cm column of Sephadex G-200 and 0.1 ml fractions collected at 10^{o}C. The samples were monitored at 750 nm. The amount of blastokinin per milligram of total protein was determined using the Lowry method (10) and pooled fractions were further characterized using Sephadex 6B.

Determination of Proteins and Nucleic Acids

To determine the amount of protein in uterine flushings, 0.2 ml of sample was mixed with the Lowry reagent and the absorbance read in the Gilford 250 Spectrophotometer at a wavelength of 250nm, using Bovine Serum Albumen (BSA) as a standard. RNA and DNA were determined using the modified Resorcinol and Diphenylamine Assays (11) respectively.

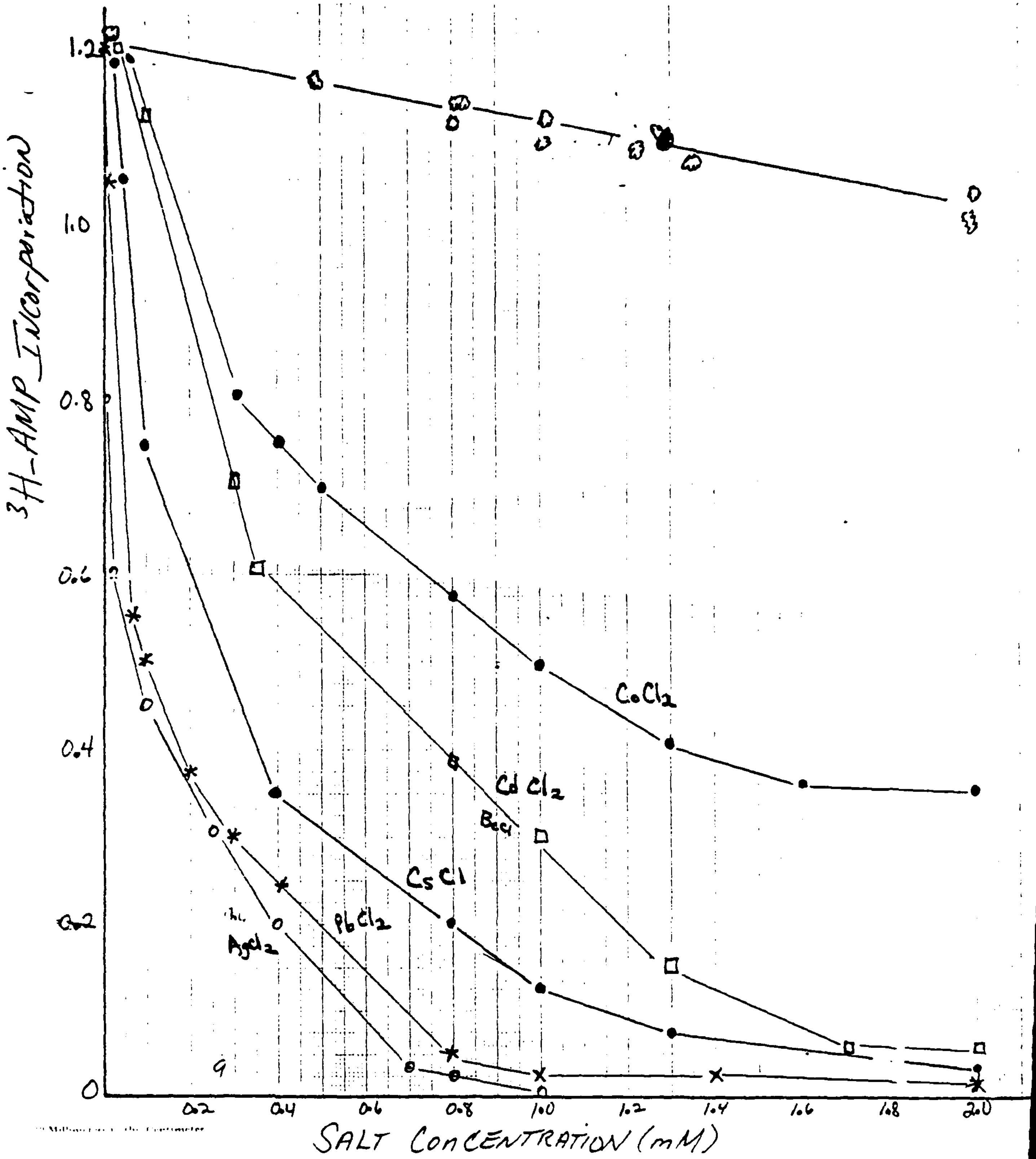
RESULTS AND DISCUSSION

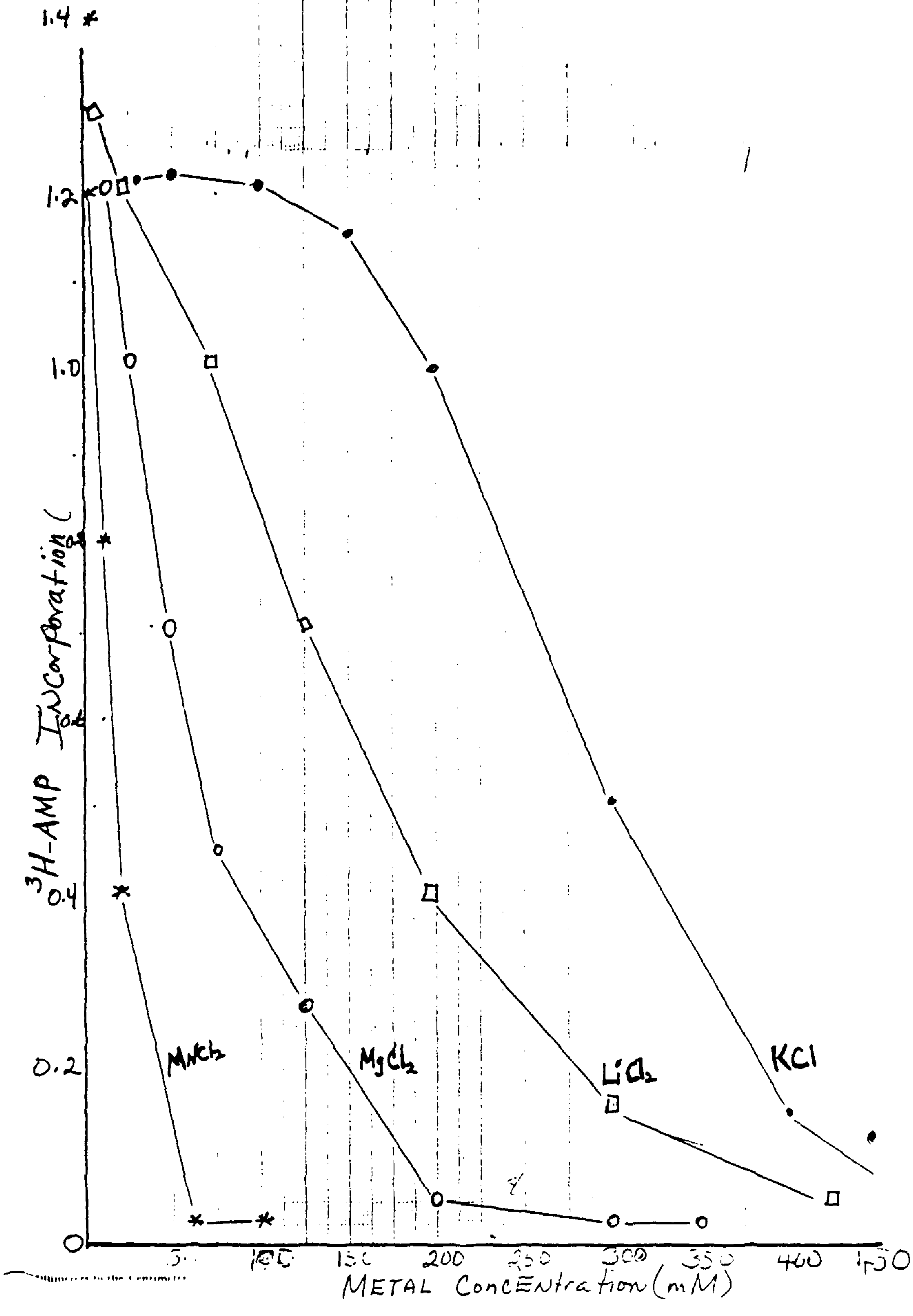
In vitro Procedures

The experiments to determine the effects of metals on the synthesis of RNA, using rabbit uterine DNA as a template, were monitored by the incorporation of ^3H-adenosine monophosphate into acid precipitable molecules. Each heavy metal (cadmium, lead, mercury, silver, cobalt and cesium) was tested at a minimum of six different concentrations, 0.1-2.0mM for effects on total RNA synthesis and the rate of initiation. The overall synthesis of total RNA was inhibited by all six of the metals and the order of greater inhibition for each of the metals was: $Ag^{+}>Hg^{++}>Pb^{++}>Cs^{+}>Cd^{++}>$ and Co^{++} and occurred in the same order with either template (Fig. 1 and Table 1). The inhibitory effects were apparent with increasing concentrations of metals and in some instances precipitates were observed in the reaction vessels, especially silver and mercury. Table 1 and Figure 1 clearly indicates the effects of increasing concentrations of metals on RNA synthesis to a maximum of total inhibition. These concentrations exceed the trace amounts found in the food chain and water of most humans.

The effects of metals in the rate of RNA initiation was measured _in vitro_ by the incorporation of ^{32}P-Adenosine and ^{32}P-Guanosine Triphosphate because both of these nucleotides are incorporated at the 5' ends of newly synthesized RNA polymers and guanosine residues specifically in the cap of heterogeneous RNA. Some of the radiolabel in the RNA polymer represents the poly (A^{+}) tails found at the 3' end of messenger RNA molecules and hnRNA. At the concentrations of metals that inhibited total RNA synthesis, stimulation of RNA transcription was evident by the incorporation of ^{32}P-ATP and $^{32}PGTP$. The rate of initiation was greater with ^{32}P-ATP than with ^{32}P-GTP a concentrations of metals which produced 50-60% overall RNA inhibition. These effects were evident with both DNA templates (Tables 1 and 2). The effects of several metals that are not considered to be mutagenic or carcinogenic were examined under similar conditions using both DNA templates (Fig. 2). The concentration of each metals (Mn^{++}, Mg^{++}, Li^{++}, K^{+}) was greater than the heavy metals and at concentrations that inhibited total RNA synthesis by 50-60%, there was no evidence of stimulation of initiation (Table 1 and Fig 1). The incorporation of ^{32}P-ATP and ^{32}P-GTP into RNA decreased and a more pronounced effect was observed on the incorporation of ^{32}P-ATP (Tables 1 and 2). The results also indicate that concentrations of metals that inhibited total RNA synthesis by 5-10% also caused decreased chain initiation. These results clearly indicate that the effects on the incorporation of ^{32}P-ATP were greater than on ^{32}PGTP incorporation during the initiation of RNA transcription and that metal mutagens promote RNA initiation at new sites on the DNA template while light metals used in this study do not. The order of inhibition of these metals on RNA synthesis may be explained by the affinity of the metal for the pyrimidine and purine bases of the template. Density gradient data have shown that heavy metals, such as lead, silver, cesium and mercury bind to the nitrogenous bases and accentuate changes in the buoyant density. This study further suggests that alterations in gene expression may be caused by heavy metals binding to the DNA template and affording new sites for the initiation of RNA transcription and producing RNA chains out of sequence and/or activating foreign genes that have been incorporated into the genome. These data lend additional support to the role of metals in cominc cancer.

To determine if heavy metals would affect the transcription of proteins _in vivo_, we examined the effects of cadmium chloride, a carcinogen and teratogen, on the synthesis of a specific uterine protein, blastokinin.


3H-AMP Incorporation
SALT CONCENTRATION (mM)
CoCl2
CdCl2
CsCl
PbCl2
AgCl2
1.2
1.0
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0



1.4
1.2
1.0
0.4
0.2
0
^{3}H-AMP Incorporation
MnCl2
MgCl2
LiCl2
KCl
200
300
400
METAL ConcEntration (mM)

TABLE 1

EFFECTS OF METAL CHLORIDES ON RNA SYNTHESIS, USING RABBIT DNA AS TEMPLATE

		PERCENT INCORPORATION		
Metal Chloride	Concentration (mM)	^{3}H-AMP	-^{32}P-ATP	^{32}P-GTP
Control	Saline	100.0	100.0	100.0
$HgCl_2$	0.10	44.6	150.0	138.6
$AgCl_2$	0.10	43.8	148.0	133.4
$CdCl_2$	0.33	50.0	186.4	130.1
$PbCl_2$	0.10	45.0	160.0	140.0
$CoCl_2$	0.75	44.0	319.0	160.0
CsCl	0.15	40.0	142.0	161.0
$MgCl_2$	38.0	54.0	68.0	70.0
$MnCl_2$	5.0	50.0	——	——
BeCl	0.05	40.2	47.5	83.2

The effects of heavy metal salts on total RNA synthesis and the initiation of RNA transcription. Total RNA synthesis was measured by the incorporation of ^{3}H-ATP while initiation was monitored by the amount of radioactive ^{32}P-ATP and ^{32}P-GTP.

TABLE 2

EFFECTS OF METAL CHLORIDES ON RNA SYNTHESIS, USING T4 DNA AS A TEMPLATE

Metal Chloride	Concentration (mM)	PERCENT INCORPORATION ^{3}H-AMP	$-^{32}$P-ATP	^{32}P-GTP
Control	Saline	100.00	100.00	100.0
$HgCl_2$	0.10	44.00	148.21	133.20
$AgCl_2$	0.10	43.80	145.00	121.10
$CdCl_2$	0.33	48.40	160.20	127.11
$PbCl_2$	0.10	62.0	134.00	118.20
$CoCl_2$	0.75	46.21	230.04	141.00
CsCl	0.15	——	——	——
$MgCl_2$	38.0	40.10	43.10	60.00
$MnCl_2$	5.0	43.01	49.50	72.30
BeCl	0.05	40.20	47.0	85.2

The effects of heavy metal salts on total RNA synthesis and the initiation of RNA transcription. Total RNA synthesis was measured by the incorporation of ^{3}H-ATP while initiation was monitored by the amount of radioactive ^{32}P-ATP and ^{32}P-GTP.

Sexually mature New Zealand White rabbits were mated and sacrificed at progressive stages of pregnancy. The uteri were flushed with chilled phosphate buffer, excised and cut along the mesometrial surface. The uterine flushings were chromatographed and analyzed. The data indicate that four major classes of proteins are present in the uterine flushings of 5-day pregnant rabbits. The four major peaks represent high molecular weight proteins, hemoglobin, albumin and blastokinin. In the 5-day control, blastokinin represents 54% of the total protein in uterine flushings. However, when animals are injected with cadmium chloride, at levels that inhibit total RNA synthesis in vitro by 50%, there is a decrease (7%) in the amount of total protein in 5-day flushings as well as a decrease (38%) in the amount of blastokinin (Table 3).

When beryllium chloride was injected into the uterine horns of pregnant does a remarkable increase in total protein accumulation was observed. This level of protein accumulated in the uterine cavity remained high from day-5 through day eight. However, a significant difference in observation was observed in the accumulation of blastokinin. The level of blastokinin in 5 day pregnant does was only 38% of the total protein as compared to control animals that represented 54.26% of the total protein (Table 3). These observations with beryllium and cadmium clearly indicate a significant accumulation of blastokinin and may represent the suppression of specific genes by these metals, via premature termination or other mechanisms.

TABLE 3

EFFECTS OF METALS ON THE SYNTHESIS OF UTERINE PROTEINS

Stage of Pregnancy (Days)	Total Protein (mg)			Blastokinin* (% Ttl. Prot.)		
	Ctrl.	$CdCl_2$	BeCl	Ctrl.	$CdCl_2$	BeCl
Estrus	2.26	2.26	2.30	--	--	--
3	2.35	2.33	2.36	19.60	--	--
4	2.40	2.46	2.40	38.45	--	39.2
5	2.41	2.26	2.40	54.26	33.6	35.1
6	2.42	2.30	2.40	37.00	33.0	31.0
7	2.40	2.41	2.40	--	30.9	31.2
8	2.37	2.40	2.40	10.21	--	--
9	2.12	2.10		7.93	--	--
10	2.10	2.10		--	--	--

*Expressed as percent of Total uterine protein.

The effects of cadmium and beryllium on the synthesis of uterine proteins. The RNA and DNA determinations are expressed as micrograms per milligramof protein. Cadmium was used at 50 ppm, 100 ppm and 0.33 mM; however, the table indicates the effects at 0.33 mM; Beryllium was used at 0.05 mM.

FIGURE 1

The effect of increased concentrations of heavy metal salts on the synthesis of RNA 0———0 $AgCl_2$, x———x $PbCl_2$, .———. CsCl, []——— $CdCl_2$, {}———{} $CoCl_2$. The incorporation of ^{3}H-ATP is used as the measure of synthesis.

FIGURE 2

The effect of non-mutagenic or carcinogenic metal salts on total RNA synthesis using rabbit DNA as template.

TABLE 1 and 2

The effects of heavy metal salts on total RNA synthesis and the initiation of RNA transcription. Total RNA synthesis was measured by the incorporation of ^{3}H-ATP while initiation was monitored by the amount of radioactive ^{32}P-ATP or ^{32}P-GTP.

TABLE 3

The effects of cadmium and berylluim on the synthesis of uterine proteins. The RNA and DNA determinations are expressed as micrograms per milligram of protein. Cadmium was used at 50 ppm, 100 ppm and 0.33 mM; however, the table indicates the effects at 0.33 mM.

These findings may reflect a stimulation of RNA transcription and an increase in the premature termination of the RNA chains, even though the concentration is higher than the amounts normally found in human (14). From these findings, it is apparent that further studies are required to establish the effects of metals on the in vivo synthesis of RNA and proteins in eukaryotic systems. However, the effects of lead and cadmium have been reviewed thoroughly in the laboratory mouse (12, 13, 14) and humans (15) and it is known to be transferred across the placenta in pregnancy in humans (15), in the rat (16) guinea pig (17) and the mouse. Previous studies have clearly demonstrated the deposition of lead in the bone and indicate that it may affect the synthesis of macromolecules (proteins, etc.) and affect the developing embryos in pregnant mothers (18, 19). Our data confirm earlier findings in that the expression of specific proteins is suppressed.

REFERENCES

1. Hoffman, D. and S. Niyogi. 1977. Science 198:513.

2. Witschi, H. 1970. Biochem. J. 120:623.

3. Bassler, R. 1965. C. R. Searic. Soc. Biol. 159:1620.

4. Luke, M., L. Hamilton and T. Hollocher. 1975. Biochem. Biophys. Res. Commun. 62:497.

5. Richardson, J. 1970. J. Mol. Biol. 49:235.

6. Krishnan, R. and J. C. Daniel. 1967. Science 158:490.

7. Kirby, K. S. 1957. Biochem. J. 66:495.

8. Skinner, D. and L. Triplett. 1967. Biochem. Biophys. Res. Comm. 28:892.

9. Studier, F. 1965. J. Mol. Biol. 63:409.

10. Lowry

11. Amveller, G. 1985. Histochemistry 83:413.

12. Donald, J., M. Cutler and M. Moore. 1986. Environ. Res. 41:420.

13. Angell, N. and J. Lavery. 1982. Am. J. Obstet. Gynecol. 142:40.

14. Daniellson, B., L. Dencker and A. Lindgreen. 1983. Arch. Toxicol. 54:97.

15. Harris, P. and M. Holley. 1986. Pediatrics 49:606.

16. Green, M. and N. Gruener. 1974. Res. Commun. Chem. Pathol. Pharm. 8:735.

17. Kelman, B. and B. Walter. 1980. Proc. Soc. Exp. Biol. Med. 163:278.

18. Clarren, S., D. Bowden and S. Astley. 1987. Teratology 35:345.

19. Neubert, D., I. Chahoud, T. Platzek and R. Meister. 1987. Arch. Toxicol 60:238.

20. Baranski, B. 1987. Environ. Res. 42:54.

21. Edwards, K. 1988. MARC-NIH Symposium, Washington, D.C.

ACKNOWLEDGEMENTS

The author wishes to express his appreciations to the Air Force Office of Scientific Research and Universal Energy Systems, Dayton, Ohio, for sponsoring this research project and for exposing me to opportunities available in federal environmental research laboratories. Special thanks are extended to Dr. Bruce Poitrast and Colonel J. Rock of the Occupational and Environmental Health Laboratory at Brooks Air Force Base, San Antonio, Texas.

ASOFR-UES RESEARCH INITIATION PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

Development of a Rapid and Sensitive Assay Procedure for the Detection of the Protozoan Parasite *Giardia lamblia* in Drinking Water Supplies

Prepared by: Ralph J. Rascati

Academic Rank: Associate Professor

Department and University: Department of Biology, Kennesaw College

Date: April 26, 1988

Subcontract: 760-6MG062

Development of a Rapid and Sensitive Assay Procedure for the Detection of the Protozoan Parasite *Giardia lamblia* in Drinking Water Supplies

by

Ralph J. Rascati

ABSTRACT

Procedures for concentration and detection of the protozoan parasite *Giardia lamblia* were developed. Membrane filtration was used for concentration of the cysts from water supplies and an immunoassay was developed for the detection of cysts. Testing of the membrane filtration procedure for concentration of cysts was undertaken using water samples deliberately seeded with known amounts of *Giardia* cysts. It was found that filters with a pore size of 3 um were more efficient (80-90% recovery) than filters with a pore size of 5 um (≤ 45% recovery). Although the larger pore filters could process slightly larger samples before filter clogging occurred the difference in sample size did not fully compensate for the decreased recovery efficiency. A sensitivity coefficient was defined which allows direct comparisons of procedures with different sample processing capacities and different recovery efficiemcies. A comparison of the membrane filtration method developed under the auspices of this grant indicates that the sensitivity of the procedure is equal to that of the standard EPA recommended procedure for finished water supplies. When testing raw water supplies the membrane filtration procedure is somewhat less sensitive but is within the range of sensitivities of the EPA procedure. Processing with the membrane filtration procedure is more convenient. The immunofluorescence assay is faster and less subjective than phase contrast microscopy and is less stressful to the technician when examining samples containing many objects of the same approximate size and shape as *Giardia lamblia* cysts.

INTRODUCTION

Giardia lamblia is a flagellated protozoan responsible for the gastrointestinal disorder known as giardiasis (see ref 1 for review). The organism has a two-stage life cycle. In the cyst stage, a dormant stage, the organism is encased in a relatively tough covering which provides resistance to environmental degradation and to stomach acid. When ingested the organism survives in the stomach and is excysted in the small intestine where the conditions are no longer acidic. The organism then enters into the reproductive vegetative stage (trophozoite stage). Humans infected with Giardia can either be asymptomatic or may develop acute or chronic symptoms, including diarrhea, abdominal distention, nausea, flatulence, belching, anorexia, vomiting, fatigue, and/or abdominal cramps. As the organisms pass through the small intestine into the large intestine they are again encysted and the cysts are shed in the feces.

Waterborne transmission of giardiasis was suggested in 1946 when an outbreak in Tokyo was traced to sewage contamination of a water supply. Documentation of waterborne transmission was obtained in Aspen, Colorado in 1965 when cysts were detected in the water supply during an outbreak. Since the time of the Aspen outbreak over 80 outbreaks affecting more than 25,000 persons have been reported in the U.S. alone. Most outbreaks are relatively small, affecting <200 persons, but several of the outbreaks have been quite large. Cysts were recovered from the water supplies in only some of the outbreaks. In others the causative agent of the disease was determined by examining fecal specimens from patients showing symptoms of gastroenteritis while the waterborne etiology of the disease was suggested from epidemiological data only. In two of the largest outbreaks (Rome, NY, >4800 cases, >10% of the population; Berlin, NH, >7000 cases, >40% of the population) multiple cysts were detected in raw and treated water. In

Berlin, NH contamination resulted from joint leakages and from a poor design which allowed filtered and unfiltered water to mix, thus bypassing the filter system. In fact, most of the outbreaks have been associated with the use of surface water without filtration or with inadequate filtration. Many have occurred in New England, the Rocky Mountain states and the Pacific states where the untreated water quality is often considered good enough that many water supplies are not filtered but are only chlorinated or are filtered without coagulation pretreatment. This may not be adequate for the removal of Giardia, even though it is adequate for many other pathogens, because Giardia cysts are relatively resistant to chlorination and because filtration without coagulation has been shown to be ineffective for removal of the cysts (1). In the plains states there have been fewer outbreaks; possibly because the water quality is so poor that both filtration (with coagulation) and chlorination must be used and therefore Giardia is effectively removed (2-4).

Detection of Giardia in water supplies suspected of being contaminated by the organism has been hampered by the lack of a concentration/detection procedure with sufficient sensitivity to detect low levels of contamination. For any detection method to be useful it must be able to process large samples of water and concentrate Giardia cysts so that they may be identified. It has been estimated that raw sewage should contain between 9,600 and 240,000 cysts per liter assuming average per capita water consumption and an incidence of infection of 1-25%. Calculation of cyst levels in drinking water supplies will depend, therefore on the degree of raw sewage contamination of that supply. They should however, be considerably lower than the concentration in the raw sewage (5). Where contamination of the water supply is from some source other than raw sewage leakage, the level of cysts in the water cannot be predicted but might be expected to be even lower. A number of methods have been used to concentrate cysts in order to detect possible contamination by Giardia lamblia.

The most commonly used method for concentration of Giardia cysts is the method developed by the Environmental Protection Agency (EPA) and which is widely accepted as the standard method (6). Briefly, large samples of water can be filtered through a cartridge filter of wound orlon yarn. A water meter is connected downstream from the filter and connections are made with standard garden hose fittings. The requisite volume of water (100 gal) is passed through the filter and the filter is then shipped on ice to the laboratory. The filter fibers are cut and are then homogenized in a blender to extract the cysts. The extract is then subjected to flocculation and/or screening to further concentrate the cysts. The advantages of this method are portability, ease of use in the field, and the ability to process large water samples. The disadvantages include a complex processing procedure for extraction of the cysts from the orlon yarn fibers and a relatively low efficiency of cyst recovery in the absence of flocculation (Mean = 6.3%; Range = 3-15%) with an additional 1 log loss of recovery occurring if flocculation is used.

An alternative procedure is to use thin, flat microporous membranes to trap the particulates from the water. These particulates can then be easily eluted from the filter by a simple washing and agitation procedure. Recently Wallis and Buchanan-Mappin (7) reported the use of 5 um Nucleopore membranes (110mm diameter) to concentrate and detect cysts in untreated stream water which had been seeded with Giardia cysts in known amounts. After filtration the membrane was rinsed with small amounts of water (total = 80 ml). The cysts were then concentrated further by centrifugation and detected by microscopic examination. Using this method (MF) samples of 100-200 liters

(25-50 gal) were processed. Cyst recovery in different trials indicated that the mean recovery was 71% with a range of 58-84% at optimum pressure. However, the recovery efficiency for this method has not yet been determined for raw water sources of poor quality (2-10 NTU) with high concentrations of algae and protozoa present.

Once water samples have been concentrated they must be accurately assessed for the presence of Giardia cysts. Several methods are available for detecting the presence of the cysts. Some methods are time-consuming and require highly skilled technicians while others are rapid and rely on an objective endpoint rather than the judgment of the technician. The most common method for the detection of Giardia cysts in processed fecal or water samples is examination of stained preparations using a microscope equipped with phase contrast optics (5). After appropriate concentration procedures, samples are transferred to a microscope slide or hemacytometer and stained with dilute Lugol's Iodine. The preparation is then scanned under the microscope for the presence of objects with the correct size and shape. Since many different objects which may be present in the sample can have the correct size and shape, objects meeting these criteria must be closely examined for the presence of characteristic identifying internal structures (two to four nuclei, axonemes, and/or median bodies). This method is time consuming and fatiguing both because of the low numbers of cysts found and because of the poor visual contrast between the Giardia cysts and other contaminants of similar size and shape. This problem is even more acute in water samples with high algal and protozoan growth. The fewer the number of cysts present the longer it will take to scan enough of the sample to obtain accurate results. Negative samples could take as long as 20-134 hours for examination of the entire sample. Furthermore, this method requires a highly skilled and experienced technician to distinguish cysts from other similar objects since the internal structures are not always readily discernible.

Alternatively immunofluorescence methods have been applied to the detection of various aquatic microorganisms deposited on membrane filters (8). The filter is reacted with specific antibodies against the organism. These antibodies are coupled to a fluorescent dye. When incident ultraviolet irradiation (UV) is used those microorganisms bound by the conjugated antibodies can be located by their bright green fluorescence in contrast to a dark background. Judith Sauch (9) reported the use of this method for the detection of Giardia cysts on membrane filters. Confirmation of the identity of fluorescing objects was obtained by switching to phase contrast optics and discerning the existence of the appropriate internal structures. In a comparison of this method with the simple phase contrast microscopy described earlier it was found that for comparable sample sizes the combination method was 28 times faster (9). The method can be complicated however, by the presence of non-specific fluorescence. Furthermore, although faster than simple phase contrast microscopy it still takes some time to scan an entire sample. Although this would not be necessary with water samples contaminated with a high concentration of cysts, it would be necessary for low level contamination or for negative samples (10,11).

SPECIFIC AIMS

The goal of this investigator was to develop a procedure for the concentration, detection and identification of Giardia cysts that is convenient to process, reliable, sensitive, and which can readily distinguish Giardia cysts from among the many organisms which might be present in some raw water supplies. Ultimately, this procedure will be used to assess the occurence of this organism in water supplies for U.S. Air Force Bases and to assess the ability of various water treatment procedures to remove this organism from raw water in order to produce uncontaminated finished water for distribution and consumption. Membrane filtration was selected as the method for concentration of the cysts because of the ease of processing and the high recovery efficiencies. The problem of sample size for turbid waters will be discussed later. Immunofluorescence was chosen as the detection method because it gives an easily discernible endpoint which can be observed at low magnification (100x - 200x) while phase contrast identification requires more interpretive skill on the part of the microscopist and also requires higher magnification (400x - 1,000x) thus necessitating a longer period of time to examine the specimen.

RESULTS AND CONCLUSIONS

1. Preparation of Antiserum

Cysts of *Giardia lamblia* obtained from Dr. Frank W. Schaeffer, III of the the Health Effects Research Laboratory at the Environmental Protection Agency in Cincinnatti, Ohio were used to stimulate production of specific antibodies for use in the detection assay. Three New Zealand White (NZW) rabbits were injected in the thigh muscle of the right hind leg with an emulsion containing 2.5 million cysts suspended in 0.5 ml distilled water mixed with an equal volume of Freund's complete adjuvant. One month later the rabbits were hyperimmunized with the same dose of cysts in incomplete Freund's adjuvant. The rabbits were then bled 2, 4, and 6 weeks after the second injection. The sera collected were pooled and stored at -70 C. An indirect immunofluorescence is used for dectection of *Giardia* cysts. Concentrated samples are reacted with the rabbit-anti-Giardia antibodies (RaG). This mixture is then washed free of unbound antibody and reacted with goat-anti-rabbit-immunoglobulin antibodies which have been conjugated with the fluorescent dye, fluorescein-isothiocyanate (GaR-FITC) obtained from Sigma Chemical Company (St. Louis). If no cysts are present in the sample all of the RaG will be washed from the sample and therefore the GaR-FITC will not be able to bind and will also be washed away. Thus no fluorescent objects will be seen when the sample is observed. However if cysts are present the RaG will be bound to them and not washed away. Subsequently, the GaR-FITC will bind to the cyst-RaG complex and will fluoresce a bright green when illuminated with UV light.

2. Concentration of Cysts by Membrane Filtration

Initial investigation focused on measuring the recovery efficiency and the maximum sample size which could be processed with various membrane filter porosities using seeded water samples of different turbidity levels. These results could then be compared to published results using the EPA-developed cartridge filter procedure in order to determine the relative sensitivities of the two methods. The initial experiments were done before the RaG had been prepared and recovery efficiencies were determined by phase-contrast microscopy (PCM) on iodine stained samples. Later experiments were done using the immunofluorescence-phase contrast combination (IFPC). In experiments where both procedures were used the results were similar. Typically, cysts were seeded into water samples (2.5 - 25 gal) and filtered through the designated filter (Millipore Corp; 142 mm diameter). The filters were washed and agitated with approximately 80 ml distilled water. The cysts (and other particulates) were them sedimented by centrifugation and resuspended in a small volume (<= 1.0 ml). Aliquots (0.01 - 0.1 ml) were then examined by either PCM or IFPC microscopy and the number of cysts recovered was determined. Table 1 shows the results of an experiment in which 5 gal of finished water (0.18 NTU) were seeded with 1,000,000 Giardia cysts. The sample was divided in half and one half was processed through a filter with a 5 um pore size. The other half was processed through a filter with a 3 um pore size. Cyst recovery was 43.0% with the 5 um filter and 90.6% when the 3 um filter was used. When the filtrate from the 5 um filter was subsequently processed through a 3 um filter no significant numbers of cysts were recovered from the latter filter. This indicated that the unrecovered cysts from the 5 um filter did not pass through the filter but rather were wedged into the pores of the filter and were not eluted by the agitation/wash procedure. On the other hand when the 3 um filter was used the cysts remained at the surface of the filter (could not readily become trapped within the pores) and therefore were more easily eluted.

It was necessary however, to determine if the increased recovery efficiency obtained with the 3 um filters was negatively counterbalanced by a decrease in the maximum sample volume which can be processed through the filter. Table 2 gives the results obtained with several different water sources which were processed through both 5 um and 3 um filters. For finished water neither filter had reached capacity even when as much as 25 gallons were processed. For all of the other water sources tested the maximum volume which could be processed through the 3 um filter was 16-22% less than the maximum volume which could be processed through the 5 um filter. Therefore, an increase in the sensitivity of the procedure is obtained by using 3 um filters since the capacity is only reduced by <= 22% while the recovery efficiency is increased 2-fold. All subsequent experiments are done using 3 um filters.

3. Comparison of Immunofluorescence and Phase Contrast Microscopy

In order to test the utility of the membrane filtration - IFPC procedure it was necessary to measure the recovery efficiency from seeded raw water samples. Table 3 presents the results of one such experiment. *Giardia* cysts were analyzed by both PCM and IFPC to compare the two detection/identification methods. The numbers of cysts detected in replicate samples was consistently lower when the IFPC procedure was used. However the percentage recovery measured by either method was approximately the same (PCM - 81.8%; IFPC - 80.3%). Furthermore, since these raw water samples contained millions of algae and protozoa (some similar to Giardia in size and shape) the IFPC procedure (at 200x magnification) was much faster and less strain on the eyes since the fluorescent Giardia stood out in bright contrast to the dull background of the other organisms. Spotchecking of selected fluorescing objects by switching to phase-contrast optics and higher magnification (400x) indicated that all of the fluorescing objects of the appropriate size and shape were Giardia cysts. In contrast, for the PCM procedure the sample had to be scanned entirely at 400x magnification and each object of the appropriate size and shape had to be closely examined for the characteristic internal structures. This proved to be more time-consuming and tedious and led to periodic eyestrain.

4. Comparison of Membrane Filtration and EPA Procedures

If this procedure is to be useful for monitoring Giardia occurrence in various water supplies and in monitoring the efficiency of water treatment facilities at removing these parasites the procedure needs to be sensitive enough to detect cysts which may be present at low concentration. The current standard for comparison is the recommended EPA procedure. Various characteristics of the EPA and MF procedures are compared in Table 4. The values for the EPA procedure are taken from the published literature (1). Both procedures are convenient for sampling. The filtration apparatus for each of the methods is compact and can be easily shipped to the appropriate site for sampling. A pump or a pressurized water source can be used to push the requisite volume through the filter. Once the filter has been returned to the laboratory the elution of the trapped cysts is a much more time-consuming and tedious process for the EPA method than for the MF method. For the former the fibers of the filter cartridge must be cut, homogenized, and agitated in relatively large volumes of distilled water and the cysts must then be concentrated by coagulation and/or screening onto flat nylon membranes, once more eluted by agitation and finally concentrated to a very small volume by centrifugation. In contrast the cysts can be eluted from the membrane filters by agitation with a relatively small volume of distilled water (80 ml) and concentrated to a very small volume by a single centrifugation step. Recovery efficiencies are extremely variable with the EPA procedure but are seldom higher than 15%. In contrast, the recovery efficiencies for the MF procedure are routinely 80-90%. However, the sensitivity of any procedure is not dependent solely upon the recovery efficiency but is also greatly influenced by the volume of the water sample which is processed. I have defined a <u>sensitivity coefficient</u> (SC) which can be used to compensate for the variability in both recovery efficiency and sample volume. This sensitivity coefficient is defined as the multiplicative product: % recovery x volume processed. This coefficient more accurately describes the sensitivity of a given procedure and therefore the probability of detecting contaminating <u>Giardia</u> cysts in a given sample. For example, if by one procedure the % recovery were 80% and the maximum volume which could be processed were 10 gallons then SC = 800. If by a different procedure the % recovery were only 8% but the maximum volume which could be processed was 100 gallons then the SC would still be 800. Therefore, the probability of detecting contaminated cysts would be the same with either method. As shown in Table 4, for the EPA procedure the SC has a wide range (30-1,500) depending upon the recovery efficiency obtained. The sample size is held constant at 100 gallons since this is the recommended sample size. There is little, if any, difference between raw and finished water with this procedure. The MF procedure is at least as sensitive as the EPA procedure for finished water samples. A 25 gallon sample gives a sensitivity coefficient (2,000-2,265) which is somewhat higher than the upper level for the EPA procedure. Furthermore, since the maximum sample volume which may be processed could be substantially higher than 25 gallons the sensitivity coefficient could be even greater. However, the major drawback for the MF procedure is in the analysis of raw water sources. In this case the quality of the particular sample (turbidity; COD) can greatly reduce the maximum volume which can be processed. Even with the increased recovery efficiency of this method this reduced capacity produces sensitivity coefficients in the range from 200-775. While greater than the lower limits of the EPA procedure it is lower than the upper limits. Therefore, for raw water samples the probability of detecting contaminating Giardia cysts cannot yet be said to be as high with the MF procedure as with the EPA procedure. The use of multiple parallel filters or the use of a serial prefilter to remove large particulates (pore size >= 40 um) should improve the maximum volume which can be processed by the MF procedure. It remains to be seen however, what effect this will have on the recovery efficiency and, therefore, on the sensitivity coefficient.

SUMMARY

In summary, it has been suggested that the failure to identify Giardia lamblia in drinking water supplies is due, in large part, to the lack of a sensitive assay system for detecting the presence of the organism. In order to accurately assess the presence of this organism in water supplies where outbreaks have occurred it is necessary to employ a concentration and detection procedure which will conveniently and rapidly process (with high sensitivity) water samples such that the organism can be detected even when

present at low concentrations. Furthermore, with the incidence of giardiasis increasing it is desirable to have an assay which is sensitive and convenient enough to be used on a routine basis to monitor the efficiency with which water treatment facilities remove the organism during the processing of raw water sources to produce a finished drinking water supply. Only in this way can standards for the removal of the organism be rationally developed and can compliance with those standards be monitored.

In this report I have described the development and preliminary testing of a method for detecting *Giardia* cysts in water samples which is convenient to use, convenient to process, and which is at least as sensitive for detecting cysts in finished water samples as the more awkward but currently accepted standard procedure. With some modifications the procedure may also become as sensitive for raw water samples. Even without these modifications the procedure as described is within the range of sensitivities available with the standard EPA procedure. This procedure can be used for monitoring water supplies for the occurrence of *Giardia lamblia* in water supplies at U.S. Air Force bases.

LITERATURE CITED

1. Erlandsen, L. & E.A. Meyer (Eds.) *Giardia* and Giardiasis, Plenum Press, New York, 1984.

2. Craun, G.F. (1979) Waterborne Giardiasis in the United States: A Review, Amer. J. Public Health, 69:817-819.

3. __________ (1985) Giardiasis: The New Waterborne Disease, A.W.W.A., 73:33.

4. __________ (1985) Waterborne *Giardia*: It's Enough to Make You Sick, A.W.W.A. 77:14

5. Jakubowski, W. & T.E. Ericksen (1979) Methods for detection of *Giardia* cysts in water supplies, In: "Waterborne Transmission of Giardiasis" (W. Jakubowski & J.C. Hoff, eds.) EPA 600/9-79-001, pp.193-210.

6. Standard Methods for the Analysis of Water and Wastewater 16th Edition APHA, AWWA, WPCF.

7. Wallis, P.M. & J.M. Buchanan-Mappin (1985). Detection of _Giardia_ cysts at low concentrations in water using nucleopore membranes. Water Res. _19_:331-334.

8. Schieman, D.A. (1981) Advances in membrane filter applications for microbiology. In: "Membrane Filtration" (F.C. Dutka, ed.) Marcel Dekker, Inc., New York, pp.527-566.

9. Sauch, J.F. (1985) Use of immunofluorescence and phase-contrast microscopy for detection and identification of _Giardia_ cysts in water samples. Appl. Environ. Microbiol. _50_:1434-143[illegible].

10. Jakubowski, W. (1984) Detection of _Giardia_ cysts in drinking water. In: "_Giardia_ and Giardiasis" (S.L. Erlandsen & E.A. Meyer, eds.) Plenum Press, New York, pp.26[illegible]-28[illegible].

11. Schaeffer, F.W. & E.W. Rice (1982) _Giardia_ methodology for water supply analysis. In: "Proceedings of the American Water Works Association Water Technology Conference. American Water Works Association, Seattle, pp.143-147.

TABLE 1

Recovery of *Giardia* cysts from Membrane Filters[a]

	5 um Filter	3 um Filter
# Cysts Applied	500,000	500,000
# Cysts Recovered[b]	215,000	453,000
% Recovery of Cysts	43.0	90.6

a) 1,000,000 *Giardia* cysts were suspended in 5 gallons of finished water (0.18 NTU). 2.5 gallons were filtered through a Millipore filter (142 mm diameter) with a nominal pore size of 5 um. The remaining 2.5 gallons were filtered through a similar sized filter with a nominal pore size of 3 um. Particulate material was washed from the filters into 80 ml of distilled water and concentrated by centrifugation (1,000 x *g*; 15 min). The final resuspended volume was 1.0 ml.

b) A 100 ul aliquot of each sample was mixed with 10 ul of Lugol's Iodine and cysts were counted in a hemacytometer using phase-contrast optics to identify *Giardia* cysts from other particulates in the concentrated sample.

TABLE 2

Maximum Volumes of Various Water Samples Which Can be Filtered Through Microporous Membranes

Sample	Turbidity NTU	COD[a] mg/l	Maximum Volume (gal) 5 um	3 um	Reduction[b] %
Finished Water	0.10	<1.0	>25.0	25.0	-[c]
Diluted Raw Sewage	18.0	30.0	[illegible]	[illegible]	10.9
" " "	43.4	67.0	2.3	1.9	17.4
Reservoir[d]	1.3	35.0	10.5	8.5	19.1
Raw Water[e]	3.0	9.5	4.2	3.5	16.7
" "[e]	3.9	11.3	3.2	2.5	21.9

a) COD - Chemical Oxygen Demand -- a measurement of the amount of organic material in a water sample

b) The % Reduction of the maximum sample volume which could be processed through a 3 um filter compared to a 5 um filter

c) Maximum capacity was not reached for either filter

d) Reservoir located at Makah AFS, Makah, Washington

e) Raw Water taken from the Marietta-Cobb County Water Authority Treatment Facility

TABLE 3

Recovery of *Giardia* Cysts from Seeded Raw Water (3.0 NTU) Samples[a]

	PCM[b]	IFPC[c]
# Cysts Applied	780,000	685,000
# Cysts Recovered	638,000	550,000
% Recovery of Cysts	81.8	80.3

a) *Giardia* cysts were suspended in 3.5 gallons of raw water from the Marietta-Cobb County Water Authority Treatment Facility and filtered through a 3 um filter. Particulate matter was eluted from the filter by agitation and concentrated to 1.0 ml by centrifugation.

b) PCM - Phase Contrast Microscopy -- For determination of cysts a 10 ul sample of either the original seed stock of *Giardia* cysts or the concentrated material was mixed with 10 ul of dilute (1:10) Lugol's Iodine on a microscope slide. A coverslip was applied and the sample was examined by phase-contrast microscopy at 400x magnification. Counts were made in 20 representative fields and these numbers were used to calculate the # cysts in each sample.

c) IFPC - Immunofluorescence Phase Contrast -- For determination of cysts a 10 ul sample of either the original seed stock of *Giardia* cysts or the concentrated material was pipetted onto a microscope slide. The sample was air-dried and briefly heat fixed. RaG (1:40 dilution) was applied and the slides incubated in a humidified chamber for 30 min @ 37 C. Excess antibody was removed by washing with phosphate buffered saline (PBS) and GaR-FITC (1:20 dilution) was added. The sample was again incubated in a humidified chamber for 30 min @ 37 C. Excess antibody was again removed by washing with PBS. A coverslip was applied and the sample was examined by epifluorescence optics at 200x magnification. Counts were made in 10 representative fields and these numbers were used to calculate the # cysts in each sample.

TABLE 4

A Comparison of the EPA and MF methods of *Giardia* Detection

Characteristic	EPA (published) Raw or Finished Water	MF (measured) Finished Water	Raw Water
Max Sample Vol (gal)	100 [a]	≥ 25 [b]	2.5 - 8.5 [c]
Convenient Sampling	yes	yes	yes
Convenient Processing	no	yes	yes
% Recovery (Range)	0.3 - 15.0	80.0 - 90.6	80.0 - 91.2
Sensitivity Coefficient [d]	30 - 1,500	2,000 - 2,265	200 - 775

a) 100 gallons is the sample size recommended by the EPA procedure

b) In several experiments with finished water the process was terminated after 25 gallons had been filtered. However, in none of the experiments had the filter capacity been exceeded at that sample volume.

c) The volume which may be filtered is dependent upon the quality of the particular water sample being tested.

d) Sensitivity Coefficient = % Recovery x Sample Volume -- This represents a way of comparing the sensitivity of detection between two different procedures with different recovery efficiencies and different processed sample volumes.

Limitations to Heavy Work of Personnel Wearing at 21° C U.S. Military Chemical Defense Ensemble

Phillip A. Bishop, M.S.Ed., Ed.D.
R. E. Pieroni, B.S., M.D.
J. F. Smith, M.S., Ed.D.

Human Performance Laboratory
P.O. Box 870312
University of Alabama
Tuscaloosa, AL 35487-0312

Running Head: Limitations to Work in CDE

Address correspondence to: Phillip A. Bishop
P.O. Box 870312
Tuscaloosa, AL 35487-0312

tele. 205-348-8370

The purpose of this study was to determine the limitations to work of personnel performing heavy work at 21°C while wearing the U.S. military chemical defense ensemble (CDE). Male volunteers (N=17) wearing the CDE performed an arm and leg work task yielding a time-weighted energy consumption rate of 450W, under environmental conditions of 21°C and 65-70% relative humidity. Each work bout was continued until one of the following limits: volitional fatigue, rectal temperature of 39°C, or heat storage (S) of 140Wh. (Each termination of work marked completion of a cycle.) Subjects then rested to an S of zero or less, then repeated work until refusal to restart work after rest. Over a total of 37 work cycles, subjects terminated only two work cycles for S, with the remainder almost equally divided between fatigue and high rectal temperature. Intersubject variability was high for work and rest times, S, heat loss, and sweat production. USAF Reg. 355-8, which provides guidance for CDE use, appears to be overly conservative regarding estimated work tolerance under study conditions.

Phillip A. Bishop is assistant professor and director of the Human Performance Laboratory, University of Alabama

Key Words: Protective Clothing, Response to Heat, Heat storage, Rate of Heat Storage, Variability of Response

The presence of physical, biological, or chemical hazards necessitates the use of protective clothing. Current designs of some protective clothing, such as the U.S. military chemical defense ensemble (CDE) result in high insulation and low permeability characteristics. The combination of high insulation and low permeability creates a situation in which the body's heat dissipative capacity is severely restricted, resulting in thermal storage consequent to work at even mild environmental temperatures. Thermal storage results in elevated core temperatures which also pose a health hazard which is additional to that of any environmental hazard. To facilitate the optimum management of personnel for maximal work productivity, attempts have been made to predict maximal safe work times. One such prediction has been attempted by the U.S. Air Force in Regulation 355-8.

Predictions of human performance are very difficult because a number of variables must be incorporated into a prediction model, and some of these variables interact. Some of the key variables include body size, work rate, environmental temperature, radiant load, humidity, and wind speed. The management and prediction problems are further complicated because of possible individual differences in size, work rate, work capacity, initial core temperature, body specific heat, and hour to hour variations in environmental conditions and work requirements.

Bell & Walters (1969) for example studied the physiological responses of 31 subjects to various hot, humid conditions. Times to "imminent collapse" were quite variable with the maximum tolerance time being anywhere from 1.8 to 2.7 times as great as the minimum tolerance time. The mildest environment tested was 37.5/33.9°C (dry/wet bulb) with a mean tolerance time of 87 $\pm$ 20 min. The authors were unable to derive satisfactory predictions of imminent collapse.

The purpose of this study was to determine the limitations to work of personnel performing tasks requiring an energy output rate of 450 W in mild (21°C) conditions while wearing the CDE.

MATERIALS AND METHODS

Subjects: Subjects for this study were 17 male volunteers. All subjects provided written informed consent. Subjects were selected for participation based upon age, physical fitness, and health criteria. They were unacclimatized and selected to provide a range of cardiorespiratory fitness levels. Subjects were paid for participation with a bonus awarded based upon the duration of their work performance. Subject physical characteristics are displayed in Table 1.

Procedures: Subjects were first advised of the nature of the experiment and then given a physical examination and EKG. They then performed a maximal graded incremented treadmill run to volitional fatigue. Maximal oxygen uptake (VO_2) was measured using a Sensormedics metabolic measurement cart. Criteria for attainment of maximal VO_2 were heart rate within 10 beats of age-predicted maximum and respiratory quotient greater then 1.0. Subjects also donned a U.S. military chemical defense ensemble (CDE) including rubber gloves and protective over-boots. Subjects practiced the work protocol which consisted of walking at 3 mph at a grade that elicited an energy output rate of approximately 490W (VO_2 of about 1.4 liters per minute). They also lifted a 14.5 kg weight bar from a table using a flexing-extending motion (biceps curl) at a rate which elicited approximately 315W (VO_2 0.9 liters per minute). Environmental conditions were maintained within one degree of $21^{o}C$, and relative humidity was maintained between 65-70%. All data was collected in April through June when mean environmental temperatures were close to experimental temperatures.

Measurements: Rectal temperature (Tre) was measured from a thermister inserted approximately 8 cm beyond the rectal sphincter. Chest, forearm, thigh and calf skin temperatures were measured with surface thermisters. Temperatures were automatically measured each minute with a computerized system (Columbus Instruments, Isothermex). Water intake was measured and pre and post nude and totally clothed body weights were measured with a calibrated beam balance scale. Heart rate was measured via direct connection to a Quinton EKG or Narcobiosystem physiograph.

To help motivate subjects, they were paid on a graduated scale beginning at $5.00 per hour, with a one dollar pay raise for each hour of the test completed.

Work Test: Walks of 15 minutes were followed by 5 minutes of arm work in an effort to simulate a generic work task. Subjects wore the complete CDE including MK 17 chemical protective masks for the work test. They were allowed water ad libitum and continued work until one of the following criteria was reached: heart rate greater than 185 bpm, volitional fatigue, Tre greater than 39^{o}C, or heat storage (S) of 140 Wh (120 kcal). For the purposes of this study, reaching a termination criterion marked the end of a work "cycle." After reaching a criterion limit subjects were seated and allowed to cool to an S of zero or less relative to the initial S before any work was begun. Initiation of another work cycle marked the end of the previous rest "cycle." After recovery, subjects repeated the routine until one of the aforementioned limits was reached again and then allowed to recover in the same manner. Subjects continued the work-rest sequence until they were unwilling to restart work at the end of a rest cycle or until total experiment time, including subject preparation, reached eight hours.

Data Analysis: Mean skin temperatures (Msk) were calculated using the formula of Burton (1935). Heat storage was calculated using the formula .98 (0.2 X Msk(^{o}C)+0.8 Tre(^{o}C)) x body weight (Kg) (Shapiro et al., 1982). Test-retest reliability was evaluated with intraclass correlations and dependent t-tests. An alpha level of 0.05 was used for all statistical tests.

RESULTS

Mean (± sd) work VO_2 shown in Table 1 represents 35 (±5.1) percent of maximal work capacity and is equivalent to an energy production rate of 450 W. Mean work time for the first work cycle was 108 (±44.1) min with 8 subjects stopping work due to elevated Tre (39^{o}C), 7 stopping because of fatigue, and only 2 stopped due to cumulative S of 140 Wh (120 Kcal). No subjects were stopped on any test due to high heart rate. Mean S for the first cycle was 105 (±25.7) Wh, yielding a mean rate of S of 73 (±41.2) W. Mean rest time to a mean net S of -9 (±17.5) Wh [i.e. total heat loss of 114 (± 28.6) Wh] was 61 (±23.7) min. Mean rate of heat loss for the first rest cycle was 125 (± 31.8) W.

Fifteen subjects did at least one additional work-rest cycle (cycle 2). Mean work time for the second cycle was 78 (±31.8) min which produced a mean S of 87 (±42.0) Wh. On the second work cycle, seven subjects had to stop due to high Tre, and eight stopped because of fatigue and none were limited by S. Mean recovery time to -7 (±19.6) Whr was 51 (±26.9) min. Mean rates of S (N=15) and heat loss (N=12) for the second work and rest cycles were 71 (±39.1) W and 134 (±43.6) W, respectively.

Only four subjects did three full work-rest cycles. Mean work time for the third work cycle was 59 (± 33.1) min with a resulting net S of 59 (± 44.9) Wh (relative to start). One subject terminated work because of high Tre, one because of fatigue and two subjects were able to continue work to the eight-hour experiment time limit. The two subjects who terminated work because of Tre or fatigue took 52 (±2.1) min to recover to -2 Wh (relative to start).

Mean total work time for all subjects was 190 ± (62.5) min and mean total rest time was 111 (±50.2) min. One of the two subjects who's work was terminated due to test time limits was able to accomplish 336 min of work with only 57 min of rest and the other accomplished 314 min of work on 125 min of rest. In contrast one subject accomplished only 67 min of work. After 72 min of rest, he was unwilling to restart another work cycle. One particularly noteworthy subject reached a Tre of 39°C in only 35 min of the first work cycle and took 59 min to recover. He actually completed four work cycles and averaged 34 min per work cycle and averaged 51 min of recovery per rest cycle.

In total, 37 work cycles were completed. Of those only 2 were limited by S of 140 Wh, both cases occuring on the first work cycle at 95 and 67 min. Calculatable data for obtaining intermediate S levels were obtained for twelve subjects who stored 93 Wh of heat in a mean of 78 ($\pm$35) min in the first cycle. Five observations of S of 93 Wh were obtained in the second cycle, with mean time to 93 Wh of 48 ($\pm$13) min. Only one subject provided usable data at the third work cycle, requiring 45 min to store 93 Wh of the 37 work cycles, 18 provided usable data for 93 Wh calculations; on 8 cycles, subjects never reached 93 Wh; and insufficient usable data occurred in 11 cycles. Data loss was chiefly a result of electrode displacement or failure.

Mean sweat production rate for 15 subjects was 10.3 (± 3.2) ml/min. Mean sweat loss for these subjects was 5.2 (± 1.5) ml/min. Because some unquantifiable small drippage occurred, all of this loss could not be attributable to evaporation.

Correlations of key variables are shown in Table 2. Complete correlational analysis for all variables results in an unwieldly matrix. Table 2 contains practically all the significant correlations.

Results of the test-retest reliability determination are included in Appendix A. In brief, measures of work and rest times and S showed generally high degrees of repeatability, however, heat loss repeatability was lower.

DISCUSSION

The purpose of this study was to determine the limitations to work of personnel performing heavy work (requiring energy production rate of 450 W) in mild (21^{o}C) environmental conditions in the CDE. It was anticipated that the chief limitation to work would be S of 140 Wh. This assumption was based upon USAF Regulation (USAFR) 355-8 (1987) and the theoretical research upon which that regulation is based (Hinch, 1982). In USAFR 355-8, for the protective clothing conditions of this study, (MOPP4), for heavy work rates

(1.3 l/min oxygen uptake of 450W), the projected time to S of 93 Wh (safe level) at 70% humidity is 0.7 hours and 1.1 hours to S of 140 Wh ("maximum") (Hinch, 1982). The safe level is suggested to be "quite uncomfortable" but safe, whereas the maximum S level, it is suggested, will "produce some casualties", but will not render most workers unable to function. Surprisingly, our subjects, albeit somewhat more fit than the USAF population, were on the first work cycle able to work for 1.8 hours (163% of projected time to S of 140 Wh) with an S of only 105 Wh (75% of 140 Wh). Similarly, 93 Wh (safe level) was reached in 1.3 hours (186% F time projected by Reg. 355). Although the influence of fitness on tolerance is unclear, the correlations between first cycle work time and absolute VO_2, VO_2 relative to body weight, work VO_2, and work VO_2 as a percentage of max VO_2, were all insignificant with the highest r being .38 ($p = .15$). Reasons for quitting the first work cycle were essentially equally divided between high Tre and fatigue, but only two subjects were able to reach 140 Wh before termination for other reasons.

It could be argued that Reg. 355 is overly conservative to protect workers as much as possible. If this is intentional, it is suspect in terms of both safety and efficiency. With regards to safety, one subject reached a Tre of 39^{o}C in only 0.6 hours or 83% of the time projected for the "safe" level. A field study by Joy and Goldman (1968) observed that under study conditions, casualties rose precipitously when Tre rose above 38.5^{o}C. For our longer work durations (relative to Reg. 355 projections), no casualties occurred, although in some cases, subjects were "quite uncomfortable". Under these particular environmental conditions, fatigue was as important a variable as Tre or S.

On the second work bout, work time was reduced to 1.3 hour (0.5 hour reduction from first cycle). Again, Tre and fatigue each accounted for about one-half of the terminations, but no subject was able to reach S of 140 Wh. Two subjects were unable to go beyond the first cycle whereas two other subjects were still able to continue when their experiments were stopped due to eight total hours of experiment time. The loss of subjects after the second cycle, prohibits further meaningful evaluation.

In many cases during rest cycles, subjects' mean skin temperatures fell below their initial pre-exercise skin temperatures. This was attributed to the time lag between donning the protective suit and the onset of sweating. During a rest cycle evaporative heat losses plus other heat losses must have exceeded the rate of heat delivery to the skin, causing skin temperatures to fall. This observation bears further study in that optimizing the blood flow - sweat evaporation matchup is essential in optimizing body cooling. Proppe (1981) demonstrated in baboons that skin temperature exerted a strong influence on mean iliac blood flow and iliac vascular conductance.

The sweat production rate observed (10.3 ± 3.2 ml/min) was lower than that of unclothed subjects in warmer temperatures (38°C) and similar humidity who were working at 100W (Alber-Wallersrom & Holmer, 1985). Mean sweat rate for that study was 19.2 (± 7.7) g/min. Of note is the large variability of their six subjects and also the subjects of the present study.

One of the key findings of this study is the large inter-subject variability in all physiological responses which is consistent with the observations of others. Mar'yanovich, Balandin, Bekuzarov and Lapikov (1984) observed that individual responses to heat and work were independent of physical work capacity or sweat production and were quite variable. Likewise, Wyndham et al. (1970) citing three previous studies, observed that for a given oxygen consumption rate, there were large interindividual differences in the ultimate steady-state body temperature. They also pointed out that the rank order of subjects did not change in differing environmental conditions. They observed in their own data a noteworthy spread of rectal temperature in response to exercise at each of two different absolute oxygen consumption rates, despite the homogeneity of their subjects in body weight. These authors also suggest that VO_2 relative to surface area might be the best predictors of work tolerance in the heat. In the present study, VO_2 relative to surface area did not correlate significantly with any of the work response variables.

Large intersubject variabilities in response to work in heat has also been reported in other studies (Bell, Crowder and Walters, 1971; Craig, Garren, Frankel and Blevins, 1954; Vogt et al., 1983; Krajewski, Kamon, Avellini, 1979). These observations of large intersubject variability in response to work in heat storage conditions suggests that a priori group predictions would be difficult.

The prediction of recovery times (heat loss rates) is also difficult. The estimated rest time to recover from a maximum body S (140 Wh) according to Reg 355-8 is 1.2 hours. Although the subjects in this study only reached, on the average, 75% of that S, mean recovery time was still 1.0 hr. (83% of estimated). Total heat loss was 115 Wh (81% of 140). Although this is good correspondence, the large variability ($\pm$24 min) suggests that close adherence to the regulation might not be practical. It should also be kept in mind that the heat loss rate is not linear. During the second rest cycle, recovery time fell to 0.9 ($\pm$0.45) hr (71% of 1.2 hr.). Heat loss during the second rest cycle was lower also, reflecting the fact that less heat was stored the second work cycle. This is verified in that the rate of heat loss in the first and second rest cycles were similar, 125 ($\pm$32) Wh and 134 ($\pm$ 43) Wh, respectively.

As an alternative to a priori group prediction models of response to work in protective clothing a simpler and more accurate approach might be to attempt to model individual performance based upon responses to a single work cycle. This alternative would require that some simple measurable observations be made empirically. These responses would then be employed in a simple model which would predict individual total work tolerance. Individual work tolerance for a range of similar, reasonable environment conditions might be predictable. The observations of Wyndam et al. (1970) discussed earlier, would support this approach. Additionally, the high test-retest correlations suggest the usefulness of such predictions. The utility of this procedure awaits further testing and would undeniably have some disadvantages. Since the military and some industries already keep some performance records, the record-keeping requirements might not be impractical. Should prediction methods prove inadequate, individual monitoring of Tre or other variables might provide a viable alternative.

Because of speculation that maximal oxygen uptake or aerobic fitness might be a key variable in determining response to these conditions, further analyses were conducted. Subjects were divided into one group consisting of the six subjects with lowest maximal VO_2's and a second group of six subjects with the highest VO_2's. Mean maximal VO_2 was 3.2 ($\pm$3.1) l/min and 4.2 ($\pm$.41) l/min for the low and high fit groups, respectively. These respective means were 1.4 standard deviations (of the total group) apart. Although mean work time for the low fit group was 32 min lower than that of the high fit group [means of 98 ($\pm$44) min vs. 130 ($\pm$43) min], the variability was sufficient that the ANOVA F was only 1.63 (df = 1,10) and p = .23, indicating that a group difference was not detected. Tests of total work time and total rest time yielded similar results. Given the correlations shown in Table 2, these findings are not surprising. Apparently, aerobic fitness only influences physical endurance and not overall tolerance of these conditions. This was substantiated in that 5 of the 6 low-fit subjects terminated their first cycle because of fatigue, whereas, 5 of 6 high-fit terminated because of Tre or S.

The correlation matrix (Table 2) suggests that the rate of first cycle S is a good predictor of total work time, as is first cycle work time. Sweat production rate and first cycle rest time also correlate moderately with total work time. Likewise, total rest time is highly correlated with first and second cycle rest times and also with first cycle work time. It appears that first cycle responses might be predictive of overall response to these conditions, but that is not clear, because, in the present study, the first cycle work and rest times dominate total work and rest times. Further research is need to determine the true relationships among these variables.

Austin and Lansing (1986) found a weak but positive relationship between heat tolerance and surface area to weight ratio, in a computer simulation. This ratio did not correlate significantly with any performance variable except one, time to 93 Wh on the second work cycle ($r=.99$, $N=5$).

It could be argued that the present study is a conservative estimate of performance, that is, it underestimates work times because subjects were not acclimatized. Acclimatization training was not employed in the protocol for two reasons. Primarily, one of the key populations to which results are intended to be generalized is the U.S. Air Force. Depending upon the state of chemical defense training, the time of year, and the local climate, these populations may not be acclimatized. Secondly, the variability in response to a set acclimatization program may have introduced even more intersubject variability. The results of this study should then be viewed as conservative relative to comparisons with acclimatized workers who are habituated to protective clothing and their particular work tasks. The lower fitness of AF personnel would suggest that fatigue might be a greater limitation than S or Tre.

In summary, USAF Reg. 355-8 was found to be a conservative predictor of mean performance time. Work cycle times were limited by Tre and fatigue rather than S of 140 Wh. Intersubject variability was so high as to suggest that the use of a single work cycle time for all workers is inappropriate. Predictions of individual performance for personnel wearing protective clothing based upon representative individual responses should be explored as an alternative to group prediction. The importance of fatigue as a limitation to performance rather than S should also be examined.

Table 1. MEAN ($\pm$ SD) SUBJECTS PHYSICAL CHARACTERISTICS (N=17).

Age (yrs)	HT (cm)	WT (kg)	VO_2 MAX* (l/min)	VO_2 MAX* (ml/min.kg)	Work VO_2 (l/min)
24.7 (4.1)	180.5 (6.1)	78.8 (8.3)	3.7 (0.5)	47.6 (8.1)	1.30 (.02)

*(N=16) one subject failed to meet criteria for maximal oxygen uptake

Table 2. Pearson Correlation Coefficients/Prob > |R| Under C.J. HO:RHO=0 / Number of Observations for Key Variables

	VOM	PCT	HTSTO1	REST1	TM801	WKTM1	WKTM2	HEATS2	REST2	TM802	TWT	TRT	SWTPRO	RHEATS1	RHEATS2
VOM	1.00000	-0.95915	0.24561	-0.23775	0.67476	0.38129	0.59983	0.12121	-0.11133	-0.51149	0.56043	-0.18979	0.16425	-0.19962	-0.09854
	0.0000	0.0001	0.3592	0.3753	0.0277	0.1451	0.0234	0.6798	0.7305	0.3784	0.0371	0.5547	0.5586	0.4586	0.7375
	16	16	16	16	11	16	14	14	12	5	14	12	15	16	14
PCT		1.00000	-0.31640	0.16724	-0.61914	-0.27289	-0.56731	-0.18992	0.12375	0.62386	-0.47525	0.18851	-0.12164	0.09808	0.02264
		0.0000	0.2325	0.5359	0.0422	0.3065	0.0344	0.5155	0.7016	0.2607	0.0859	0.5574	0.6659	0.7178	0.9388
		16	16	16	11	16	14	14	12	5	14	12	15	16	14
HTSTO1			1.00000	0.53819	-0.34316	-0.46048	0.52521	0.76281	0.70401	0.01596	-0.8010	0.68408	-0.35114	0.55676	0.43335
			0.0000	0.0258	0.2748	0.0629	0.0444	0.0009	0.0072	0.9797	0.7766	0.0099	0.1994	0.0203	0.1066
			17	17	12	17	15	15	15	5	15	13	15	17	15
REST1				1.00000	-0.48766	-0.70860	0.17641	0.75908	0.88582	0.72776	-0.42522	0.9867	-0.24668	0.56650	0.51647
				0.0000	0.1078	0.0015	0.5294	0.0010	0.0001	0.1634	0.1141	0.0001	0.3754	0.0177	0.0487
				17	12	17	15	15	13	5	15	13	15	17	15
TM801					1.00000	0.90778	0.73571	-0.17105	-0.29060	0.67400	0.88071	-0.37877	0.54054	-0.82740	-0.62679
					0.0000	0.0001	0.0099	0.6150	0.3860	0.2122	0.0003	0.2507	0.0860	0.0009	0.0390
					12	12	11	11	11	5	11	11	11	12	11
WKTM1						1.00000	0.29074	-0.51096	-0.61337	0.47172	0.87343	-0.67996	0.58893	-0.90434	-0.79641
						0.0000	0.2931	0.0516	0.0258	0.4225	0.0001	0.0106	0.0209	0.0001	0.0004
						17	15	15	13	5	15	13	15	17	15
WKTM2							1.00000	0.34956	0.07447	0.81506	0.71986	0.09081	0.50004	-0.18822	-0.32481
							0.0000	0.2016	0.8090	0.0928	0.0025	0.7680	0.0818	0.5017	0.2375
							15	15	13	5	15	13	13	13	15
HEATST2								1.00000	0.72972	-0.14821	-0.19278	0.77857	-0.52817	0.53890	0.64962
								0.0000	0.0046	0.8120	0.4912	0.0017	0.0635	0.0382	0.0088
								15	13	5	15	13	13	15	15
REST2									1.00000	0.41155	-0.41909	0.97331	-0.40089	0.57219	0.49287
									0.0000	0.4912	0.1540	0.0001	0.2217	0.0410	0.0870
									13	5	13	13	11	13	13

VOM = VO_2 max; PCT = percent of VO_2 max used in work; HTSTO1 = first cycle heat storage; REST1 = first cycle rest time; TM801 = first cycle time of 93Wh of S; WKTM1 = first cycle work time; WKTM2 =second cycle work time; HEATST2 = second cycle absolute S; REST2 = second cycle rest time

ACKNOWLEDGEMENT

This study was supported by Air Force Office of Scientific Research, Bolling AFB, DC # F 49 620-85-C-0013/ SB5851-0360.

REFERENCES

1. Austin, D. M. and M. W. Lansing. 1986. Body size and heat tolerance: a computer simulation. Hum. Biol. 58:153-169.
2. Alber-Wallerstrom and I. Holmer. 1985. Efficiency of sweat evaporation in unacclimatized man working in a hot humid environment. Eur. J. Appl. Physiol. 54:480-487.
3. Bell C. R., and J. D. Walters. 1969. Reactions of men working in hot and human conditions. J. Appl. Physiol. 27:684-686.
4. Burton, A.C. 1935. Human Calorimetry II. The average temperature at the tissues of the body. J. Nutr. 9:261-280.
5. Craig, F. N., H. W. Garren, H. Frankel and W. V. Blevins. 1954. Heat load and voluntary tolerance time. J. Physiol. 6:634-644.
6. Hinch, J. H. 1982. Performance of Airbase Personnel in chemical protective clothing. AF Studies and Analyses, Nov.
7. Joy, J. T., and R. F. Goldman. 1968. A method of relating physiology and military performance--A study of some effects of vapor barrier clothing. Mil. Med. :458-470.

8. Krajewski, J. T., E. Kamon, B. Avellini. 1979. Scheduling rest for consecutive light and heavy work loads under hot ambient conditions. Ergo. 22:975-987.

9. Mar'yanovich, A. T., V. S. Balandin, A. K. Bekuzalov, and G. M. Lapikov. 1984. Individual features of responses to a combination of heat and physical exertion. Hum. Physiol. 10:49-55.

10. Proppe, D. 1981. Influence of skin temperature on central thermoregulatory control of leg blood flow. J. Appl. Physiol. 50:974-978.

11. Shapiro, Y., K. B. Pandolf, M. N. Sanka, M. M. Toner, F. R., Winsman, and R. E. Goldman. 1982. Auxillary coding: comparison of air-cooled vs. water-cooled vests in hot-dry and hot-wet environments. Aviat. Space Environ Med. 53:785-789.

12. USAF REGULATION 335-8. 1978. Department of the Air Force, Washington D.C.

13. Vogt, J. J., J. P. Libert, V. Candas, F. Duall and Ph. Mairiaux. 1983. Heart rate and spontaneous work-rest cycles during exposure to heat. Ergo. 26:1173-1185.

14. Wyndham, C. H., N. B. Strydan, A. J. van Rensburg, A.J.S. Barrade, and A. J. Heyns. 1970. Relation between VO_2 max and body temperature in hot humid air conditions. J. Appl. Physiol. 29:45-50.

APPENDIX A

Test-Retest Results

TEST-RETEST MEASUREMENT RELIABILITY OF HUMAN PHYSIOLOGICAL RESPONSE TO HEAVY WORK IN PROTECTIVE CLOTHING

Phillip A. Bishop
Robert E. Pieroni
Joe F. Smith
Thomas W. Merritt

The University of Alabama
Box 870312
Tuscaloosa, AL 35487-0312

Although test-retest measurement reliability is essential to generalizability of research results, there have been no published reports of reliability determination of physiological response to work in protective clothing. Seven male volunteers, on two separate occasions, wore protective clothing while performing two work bouts with each work bout followed by a rest period. Measurements of work and rest durations and heat storage and loss were made for each work and rest bout. For analysis, data order was counter balanced. Intraclass correlations were generally high (R_{xx}=.71 to .98) for all measures except rate of heat loss. The p values for dependent t-tests were >.05 for all test-retest results except total work time. The reliability of these measures was acceptable.

INTRODUCTION

Fundamental to all measurement is the concept that the variable of intent is stable in the absence of external influences. If a measurement yields essentially the same numbers with repeated measurements, we logically assume that subsequent generalizations about that measurement will be accurate. Conversely, the absence of measurement repeatability (reliability) makes generalizations about that measurement invalid. For some common ergonomic measurements the test-retest measurement reliability has been well established. However, the reliability of physiological response to work in protective clothing is not well established.

The purpose of this study was to determine the measurement reliability of certain physiological variables associated with prolonged work while wearing the U.S. military chemical defense ensemble (CDE).

METHODS

Subjects

Subjects for this study were seven male volunteers from whom informed consent was obtained. Subjects were paid on a graduated scale for the initial test, beginning at $5.00 per hour. A one dollar per hour pay raise was provided for each additional hour worked. On the retest, subjects were paid at the rate of $7.00 per hour. The mean (±sd) subject physical characteristics are given in Table 1 below.

Table 1. Physical characteristics of Subjects and Energy Costs. Means (±sd); N=7.

AGE (YRS)	HT (CM)	WT (KG)	VO_2 MAX (l/min)	VO_2 MAX (nl/kg)	WORK VO_2 (l/min)	PCT OF MAX (%)
25.0	181.2	81.0	3.9	48.4	1.31	33.6
(3.8)	(5.6)	(8.7)	(0.5)	(7.1)	(0.0)	(4.1)

Work Task

The work task for all subjects was a 15 minute treadmill walk followed by 5 minutes of elbow flexion and extension (biceps curls) performed in a standing position with 14.5 Kg of weight. Treadmill speed was 1.34 m/s (3.0 mph). Treadmill grade and biceps curling rate was determined through experimentation to provide a time-weighted energy production rate of approximately 450 W. Subjects also performed a maximal oxygen uptake test from which energy production percentages were calculated. Environmental temperature was maintained within 1°C of 21°C for work and test and relative humidity was 65-70°C. Water was supplied ad libitum.

Subjects worked continuously until one of three criteria was met: rectal temperature (Tre) at 39°C or volitional fatigue, or heat storage of 140 W.h. Termination of work by any of these criteria marked the end of the first work "cycle". Subjects were then seated and allowed to recover to a heat storage of zero or less relative to their heat storage level at the initiation of the first work cycle. After recovery, subjects then restarted the work under the same conditions as before. Restart of work (or refusal to restart work in one case) marked the end of that rest "cycle". Two work-rest cycles were completed on the initial test.

After at least one week, subjects were retested under the same conditions. All subjects completed two complete work-rest cycles except one subject who refused to complete more than one work-rest cycle on the retest.

Clothing

The U.S. Armed Forces chemical defense ensemble (CDE) was worn for all tests except the VO_2 maximum determination. The suit consisted of a MK17 chemical protective mask, charcoal impregnated jacket and trousers and rubber gloves and over boots. A military cotton fatigue shirt and trousers and cotton gloves were worn underneath the suit. Footwear of the subject's choice was worn underneath the boots. The total outfit insulation value was clo=2.55 and im/clo=0.28.

Measurements

Rectal temperature was measured with a thermister positioned approximately eight cm beyond the rectal sphincter. Chest, forearm, thigh, and calf skin temperatures were measured with surface thermisters. All temperatures were recorded each minute with a computerized system (Isothermex, Columbus Instrument). Mean skin temperature was calculated using Burton's equation (1934).

Pre and post nude and clothed body weights were measured on calibrated beam balance scales. Heat Storage (S) and loss values were calculated using the method of Shapiro, et al., (1982). Water intake was measured. Heart rate was measured with a direct wire system. Maximal oxygen uptake and the oxygen costs of the work task were measured with a metabolic measurement cart (Sensormedics). Because the first test session was part of a longer experiment, the work-rest cycling was continued. Consequently sweat production and sweat loss data could not be compared between test and retest.

Data Analysis

Intraclass correlation coefficients were used to assess reliability. This correlation coefficient is appropriate in a univariate, repeated measures analysis (Baumgartner & Jackson, 1982). It was computed from mean squares generated by the General Linear Models procedure of the Statistical Analysis System. To reduce ordering influence, one half of the data was input in reversed order. That is, for four subjects, selected randomly, the actual retest was entered as if it were the first test. The focus of the research question as on test-retest reliability independent of the influence of habituation.

As a further indirect test of reliability, a dependent t-test was conducted on the differences between the means for variables on the test and retest. Also, rates of heat storage were calculated and tested for reliability.

RESULTS

In Table 2 below, mean (sd), work and rest times, heat exchanges and rate of heat exchange are shown for both conditions, along with intraclass correlation coefficients (R_{xx}) and p values for paired t-tests.

Table 2. Mean (sd), work (wk) and rest (rst) times, (tm) heat storage (S) and loss storage ($\dot{s}$) and loss, intraclass correlation coefficients (R_{xx}) and p values for paired t-tests. (N=7).

	Test	Retest	R_{xx}	p
1st WK cycle tm (min)	125(33)	139(41)	.90	.18
1st cycle S (W.h)	107(26)	108(36)	.79	.92
1st Rst cycle tm (min)	55(32)	62(28)	.98	.12
1st Rst ht loss (W.h)*	127(25)	110(43)	.85	.62
2nd WK cycle tm (min)*	83(27)	98(37)	.88	.14
2nd cycle S (W.h)*	117(32)	86(52)	.96	.07
2nd Rst cycle tm (min)*	52(30)	49(28)	.97	.46
2nd Rst ht loss (W.h)***	126(40)	101(58)	-	-
Total WK tm (min)	201(46)	222(49)	.96	.03
Total Rst tm (min)	100(64)	124(55)	.99	.52
1st cycle $\dot{s}$ (W)	56.4(26)	54(31)	.93	.72
1st cycle ht loss rate (W)**	132(26)	109(20)	.19	.06
2nd cycle $\dot{s}$ (W)*	76(20)	50(20)	.71	.07
2nd cycle ht loss rate (W)***	136(48)	139(20)	-	-

*(N=6) **(N=5) ***(N=3) R_{xx} and p not computed.

CONCLUSIONS

The primary focus of this research was to determine the measurement reliability of certain physiological variables associated with prolonged work in protective clothing. As can be seen in Table 2, correlation coefficients were, for the most part, high. Although this sample was homogeneous on size, age, and fitness relative to the population of potential users of protective equipment (men and women of all ages in the U.S. military, for example), the intersubject variability for all measures was large. Large intersubject variability has also been reported by others (Mar'yanovich et al., 1984). In that the heterogeneity of the sample positively influences the magnitude of correlation coefficients, the sample used in this study may underestimate the true population test-retest correlations.

The low correlation obtained for rate of heat loss during the 2nd rest cycle is difficult to explain. The Mean Square for the model was 13.2 (df=1) compared with the Mean Square for error of 10.7 (df=4). It was very difficult to maintain complete data sets in this calculation and others due to subject refusal, loss of skin electrodes, malfunction of computer interface, with each of these problems exacerbated by the relatively long experiment durations.

The dependent t-tests show that most test-retest means were not significantly different. Again, this is partly attributable to relatively large intersubject variability. But, again, the sample was homogeneous relative to the population of interest. Total work time showed the greatest probability of test-retest difference. Of the 26 work cycles completed, 13 were terminated by the subject because of fatigue, 12 were terminated by the investigators due to high subject Tre, and 1 was terminated by the investigators due to subject S. Examination of chronologically ordered data indicates on the true first test day, seven subjects terminated due to Tre versus five terminations for fatigue. On the second test day, five terminated due to Tre versus eight due to fatigue. Since Tre is an objective measure, and fatigue is a subjective measure, perhaps subjects were more willing to terminate earlier than necessary on their second trial. This "early" termination may have artificially amplified test-retest differences in work times and possibly in some of the heat exchange rates, since heat exchange rates are non-linear.

In summary, given the large number of interrelated factors inherent in human physiological responses, the test-retest reliability was quite good. It appears that valid generalizations of these measurements can be made to situations with the same conditions.

REFERENCES

Baumgartner, T. A. and Jackson, A. S., 1982, Measurement for Evaluation in Physical Education, 2nd ed. (Dubuque, IA, Wm. C. Brown), pp. 98-102.

Burton, A. C., 1934, A new technique for the measurement of average skin temperature over surfaces of the body and the changes of skin temperature during exercise. The Journal of Nutrim, 7 (5), 481-495.

DuBois, D. and DuBois, E. F., 1916, A formula to estimate the approximate surface area if height and weight be known. The Archives of Internal Medicine, 17, 863-871.

Mar'yanovich, A. T., Baladin, V. S., Bekuzavov, A. K., and Lapikor, C. M., 1984, Individual features of responses to a combination of heat and physical exertion. Human Physiology, 10 (1), 49-55.

Shapiro, T., Pandolf, K. B., Sanka, M. N., Toner, M. M., Winsman, F. R., and Goldman, R. F., 1982, Auxiliary cooling: comparison of air-cooled vs. water-cooled vests in hot-dry and hot-wet environments. Aviation, Space and Environmental Medicine, 53 (8), 785-789.

FINAL REPORT NUMBER 105
REPORT NOT AVAILABLE AT THIS TIME
Dr. Hoffman Chen (1986)
760-6MG-118

FINAL REPORT

Nucleic acid hybridzation -
Dot blot test for the presence of
Ureaplasma urealyticum and _Mycoplasma hominis_.

Vito G. DelVecchio, Ph.D.

INTRODUCTION

Mycoplasma hominis and *Ureaplasma urealyticum* are two mycoplasmas which belong to the Class Mollicutes and as such lack a cell wall and represent the smallest free living prokaryotes. Mycoplasmas are characterized by a rather small genome of only some 5 x 10^8 Da and a low guanine and cytosine content.[1] Various mycoplasmas have been continually implicated as the causative agents of a diverse array of diseases in humans.[2] *M. hominis* and *U. urealyticum*, specifically, have been identified as urogenital tract infectants where they have been implicated in urethritis, epididymitis, Reuter's syndrome, Bartholin's abscess, vaginitis, cervicitis, salpingitis, acute pyelonephitis, nongonococcal urethritis, and pelvic inflammatory disease.[3-5] In addition, infertility, spontaneous abortion, premature births, and other reproductive failure may be due to infection by these organisms.[6-11] The need, then, for a rapid, sensitive, and specific clinical test for these organisms is clearly indicated.

However, the current clinical test for *M. hominis* and *U. urealyticum* is rather time-consuming, expensive, and often of low specificity. the test involves the isolation and indirect identification of the two specimens on two different culture media which are easily contaminated by other organisms present in clinical samples and which require long incubation periods for mycoplasmal growth. The test has also been shown to cause a high number of false positives and to have a low sensitivity.[12-13]

Fortunately, the use of DNA probes for detection of mycoplasmas has been shown to be effective. Recently, species-specific probes for *M. pneumoniae* and *M. genitalium* have been developed by Hyman et al[14], and Taylor reported a species-specific probe for M. hyorhinis.[15] In light of the fact that previous studies have suggested that DNA homologies among mycoplasmas are limited[16], the possibility of developing species-specific probes for other mycoplasmas is favorable. The aim of this research is the development of species-specific DNA probes for M. hominis and U. urealyticum.

METHODS AND MATERIALS

Cultures

Mycoplasma hominis (ATCC# 23114) and Ureaplasma urealyticum (ATCC# 27619) were purchased form Hana Biologics, Inc. M. hominis was grown in the MBM medium described by Ghosh[17], and U. urealyticum was cultured in Shepard's[18] urea-containing medium. U. urealyticum was cultured under non-strict anaerobic conditions in a jar depleted of oxygen by a burning candle. All cultures were incubated at 37°C.

DNA Extraction

Genomic DNA was extracted from M. hominis and U. urealyticum by a modification of the method of Marmur.[19]

Cloning of M. hominis and U. urealyticum Genomic DNA

The isolated DNA was cloned into E. coli using three different cloning strategies involving the restriction endonucleases PstI, EcoRI, and DraI (all from Bethesda Research Laboratories). The recognition sites for these three enzymes are as follows:

PstI	GTGCA G
EcoRI	G AATTC
DraI	TTT AAA

PstI has the greatest number of G + C in its recognition sequence and EcoRI has less, but both give rise to sticky ends. DraI, since it has no G or C in its sequence gives the most complete digestion of mycoplasmal DNA , but it generates blunt ends. The three cloning strategies take these factors into account.

A. PstI/pBR322 system

PstI digestion of *M. hominis* and *U. urealyticum* genomic DNA were ligated into dephosphorylated PstI cut pBR322 (New England Biolabs). Pst I digestion was carried out for 1 hour at 37°C, followed by phenol extraction. The restricted DNA was then diluted to 100 ng/uL and added in a 1:3 ratio to dephosphorylated PstI cut pBR322.[20] Ligation with T4 DNA ligase (BRL Laboratories) was carried out for 4 hours followed by transformation into DH-5 competent *E. coli* cells (BRL Laboratories) by the method of Hanahan.[21] The transformed cells were grown on Luria agar plates containing tetracycline (12.5 ug/mL). This assured the growth of pBR322 containing cells since pBR322 contains a tetracycline resistance (Tet) gene. The presence of mycoplasmal inserts was determined by insertion inactivation of the ampicillin resistance (Amp) gene using Ampscreen (BRL Laboratories). pBR322 has a unique PstI site in the Amp gene. Ampscreen discs produce a yellow pigment in the presence of B-lactamase which is responsible for Amp resistance. Therefore, Amp^- clones which contain mycoplasmal inserts are easily selected since they do not produce this yellow pigment.

B. EcoRI - pBR325 system

Using a similar protocol, a parallel cloning strategy involved the insertion of fragments from EcoRI digestion of genomic DNA from each of the mycoplasma species into pBR325.

pBr325 has a unique EcoRI restriction site in the chloramphenicol resistance (Cm) gene. Thus, transformed cells grown on Luria + tetracycline plates were screened for mycoplasmal inserts by replica plating onto Luria + chloramphenicol (10ug/mL) plates.[22] Colonies which did not grow on the Luria + chloramphenicol plates were presumed to contain mycoplasmal inserts.

C. Poly C/G Tailing System

The third cloning strategy involved the use of DraI generated mycoplasmal fragments. As noted, these fragments have blunt ends. Moreover, DraI cuts pBR322 and pBR325 at several sites. To overcome these two problems, the fragments were tailed with poly-C tails using DNA terminal transferase (BRL Laboratories). DraI cut DNA dissolved in dH_2O at 100 ng/uL was combined with 50uL of tailing buffer (14mM sodium cacodylate, 3mM Tris-HCl, 0.1mM calcium chloride, 0.01mM dithiothreitol, 0.002mM dCTP) and 45 units terminal transferase and incubated at 37^oC for one hour.[23] Tailed DNA was recovered by phenol extraction followed by ethanol precipitation. Ligation with poly-G tailed PstI cut pBR322 (BRL Laboratories) was carried out using T4 DNA ligase as above followed by transformation by the method of Hanahan.[21] Clones containing mycoplasmal inserts were again selected by insertion inactivation of the Amp gene via Ampscreen following growth on Luria + tetracycline plates.

Screening of Clones

Clones generated by each of the systems were screened for presence and size of insert. The method of Birnboim[23] was used to isolate plasmid DNA from E. coli clones. The plasmids were then digested with the restriction endonuclease used to form them and were electrophoresed in agarose gels as detailed by Maniatis.[22] Sizing of inserts was accomplished by comparing with a 1kb ladder (BRL Laboratories). Over 250 clones were obtained with fragments ranging in size from 200 to 1600 base pairs.

Isolation and Biotinylation of Probes

Clones containing recombinant plasmid molecules were grown in preparative cultures on "terrific broth"[24], and the plasmids were isolated using the pZ523 method (3'-5' Incorporated).[25] It was decided to use a non-radioactive labeling system to visualize probe/target interaction. Biotin-11-dUTP was incorporated into the DNA of the recombinant molecules via nick-translation according to the method of Langer.[26]

Dot-Blot Studies

These biotinylated molecules were then tested for sensitivity and specificity as probes for *M. hominis* and *U. urealyticum* DNA. The target DNA's were immobilized on nitrocellulose paper by the SSC method of Schleiden and Schuell.[27] Visualization was through a streptavidin-alkaline phosphatase system as described by Leary.[28] In the presence of its homologous sequence in target DNA, a biotinylated probe hybridizes with that sequence as detailed by Meinkoth et al.[29] Using the protocol accompanying the Blu Gene Nonradioactive Nucleic Acid Detection System (BRL Laboratories), streptavidin with its conjugated alkaline phosphatase then attaches to the probe because of its tremendous affinity for biotin. Alkaline phosphatase, in turn, acts on the substrates BCIP (5-Br-4-Cl-3-indolyl phosphate) and NDP (nitroblue tetrazolium) to produce a blue colorimetric reaction which is directly visualized on the nitrocellulose membranes.

Three probes characterized by these dot-blot studies were chosen for application to the in situ system. Probes pMD37 and pMD22 were generated via the poly C/G tailing strategy, and pUP18 was generated via the PstI - pBR322 strategy. pMD37 and pUP18 were determined to be specific for M. hominis and U. urealyticum respectively, and pMD22 was determined to be a sensitive, but general mycoplasma probe. These data were also corroborated by Southern [30] blotting techniques.

RESULTS

In the dot-blot system, pUP18 detected U. urealyticum genomic DNA down to 247 pg, and detected M. hominis DNA only at 11.25 ng. pUP18 was also positive for U. urealyticum cells and negative for M. hominis cells. These results were corroborated by Southern blotting which showed that pUP18 reacted positively against fragment of PstI digested U. urealyticum DNA and negatively against PstI digested M. hominis DNA electrophoresed on agarose gels and blotted onto nitrocellulose paper. pMD22 was found to be a general, yet sensitive probe for both M. hominis and U. urealyticum, and is of value as a general Mollocute probe. pMD22 detected 82pg of U. ureaplasma DNA and 595 pg of M. hominis DNA in the dot-blot system. In a Southern blot, biotinylated M. hominis DNA in the dot-blot system. In a Southern blot, biotinylated M. hominis genomic DNA also reacted positively against pMD22. pMD37 detected M. hominis DNA down to 763pg, and U. urealyticum only at 60ng. It showed positive reactivity for M. hominis cells, and negative reactivity for U. urealyticum cells.

DISCUSSION AND CONCLUSION

DNA probes which differentiate between *M. hominis* and *U. urealyticum* DNA at high sensitivity have been developed.

A system of plasmid screening and dot-blot hybridization has been successfully developed through which sensitivity and specificity of other probes can easily be assessed.

REFERENCES

1. Razin, S. (1978). The Mycoplasmas. Microbiol. Rev. 42:414-470.

2. Freundt, E.A. (1974). Present status of the medical importance of mycoplasma. Pathol. Microbiol. 40:155-187.

3. Mardh, P.A. (1983). Mycoplasma PID: A review of natural and experimental infections. Yale J. Biol. Med. 56: 529-539.

4. Taylor-Robinson, D., W.W. McCormack. (1979). Mycoplasma in human genito-urinary infections. 308-366. In J. G. Tully and R.B. Whitcomb, eds., The Mycoplasmas, vol. 2. Academic Press, New York.

5. Taylor-Robinson, D. (1983). The role of mycoplasma in non-gonoccocal urethritis. Yale J. Biol. Med. 56: 537-543.

6. Cassell, G.H., J.B. Younger, and M.B. Brown. (1983). Microbiologic study of infertile women at the time of diagnostic laparoscopy: Association of Ureaplasma urealyticum with a defined sub-population. New Eng. J. Med. 308: 502-505.

7. Swenson, C.E., A. Toth, and W.M. O'Leary. (1979). Ureaplasma urealyticum and human infertility: The effect of antibiotic therapy on semen quality. Fertil. Steril. 31: 660-665.

8. Kundsin, R.B. (1981). Ureaplasma urealyticum incriminated in pre-natal morbidity and mortality. Science. 213: 474-476.

9. Styler, M. and S.S. Shaprio. (1985). Mollicutes in infertility. Fertil. and Steril. 44: 1-12.

10. Lamont, R.F., D. Taylor-Robinson, M. Newman, J. Wigglesworth, and M.G. Elder. (1986). Spontaneous early pre-term labor associated with abnormal genital bacterial colonization. Br. J. Obstet. Gynecol. 93: 804-810.

11. Friberg, J. (1983). Mycoplasma and Ureaplasma in reproductive failure. Contemporary OB/GYN.

12. Clyde, W.A. (1983). Laboratory diagnosis of chlamydial and mycoplasma infection. 511-528. In Balows, A., W.J. Hausler, eds., Diagnostic Procedures for Bacterial, Mycotic, and Parasitic Infections. American Public Health Associations, Washington, D.C.

13. Tully, J.G. (1983). Evaluation of culture media for the recovery of Mycoplasma hominis from the human urogenital tract. Sex. Trans. Dis. 10: 256-260.

14. Hyman, H.C., D. Yogev and S. Razin. (1987). DNA probes for detection and identification of Mycoplasma pneumoniae and Mycoplasma genitalium. J. Clin. Microbiol. 25: 726-728.

15. Taylor, M.A., K.S. Wise and M.A. McIntosh. (1984). Species-specific detection of Mycoplasma hyorhinis using DNA probes. Israel J. Med. Sci. 20: 778-780.

16. Stanbridge, E.J. and M.E. Reff. (1979). The molecular biology of mycoplasma. 157-187. In Barile, M.F. and S. Razin, eds., The Mycoplasmas, vol.1. Academic Press, New York.

17. Ghosh, A., J. Das and J. Mainiloff. (1977). Lack of repair of UV light damage in Mycoplasma gallisepticum. J. Mol. Biol. 337-344.

18. Shepard, M.C. Culture medium for ureaplasma. 377-378. In Razin, S. and J.G. Tully, eds., Methods in Mycoplasmology, vol.1. Academic Press, New York.

19. Marmur, J. (1961). A procedure for the isolation of DNA from microorganisms. J. Mol. Biol. 3: 208-218.

20. Goodman, H.M. and R.J. MacDonald. (1979). Cloning of hormone genes from a mixture of cDNA molecules. 75-90. In R. Wu, ed., Methods in Enzymology, vol. 68. Academic Press, New York.

21. Hanahan, D. (1983). J. Mol. Biol. 155: 557-580.

22. Maniatis, T. E.F. Fritsch and J. Sambrook. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press.

23. Birnboim, H.C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. 243-255. In Wu, R., L. Grossman and K. Moldave, eds., Methods in Enzymology, vol. 100. Academic Press, New York.

24. Tartoff, K.D. and C.A. Hobbs. Improved media for growing plasmid and cosmid clones. Focus. 9: 112.

25. Zervas, P.H., L.M. Morris, and R.J. Hellwig. (1988). A novel method for rapid isolation of plasmid DNA. Biotechniques. 6: 238-242.

26. Langer, P.R., A.A. Waldrop and D.C. Ward. (1981). Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. Proc. Nat. Acad. Sci. 78: 6633-6637.

27. Transfer and Immobilization of Nucleic Acids to S & S Solid Supports. Schleiden and Schuell, Inc., pp. 24-25.

28. Leary, J.J., D.J. Brigati and D.C. Ward. (1983). Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA and RNA immobilized on nitrocellulose: Bio-blots. Proc. Nat. Acad. Sci. 80: 4045-4049.

29. Meinkoth, J. and G. Wahl. (1984). Hybridization of nucleic acids immobilized of solid supports. Annal. Biochem. 138: 267-283.

30. Southern, E. (1975). Detection of specific sequences using DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503

FINAL REPORT NUMBER 107
REPORT NOT AVAILABLE AT THIS TIME
Dr. Mohammed Maleque
760-7MG-078

1987 USAF-UES RESEARCH INITIATION PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by the

Universal Energy Systems, Inc.

FINAL REPORT

The Asymptotic Description of Precursor Fields in a Causally Dispersive Medium

Prepared by: Kurt Edmund Oughstun, Ph.D.
Academic Rank: Associate Professor
Department: Computer Science and Electrical Engineering
University: University of Vermont
Date: December 31, 1988
Contract No. F49620-85-C-0013/SB5851-0360

THE ASYMPTOTIC DESCRIPTION OF PRECURSOR FIELDS IN A CAUSALLY DISPERSIVE MEDIUM

by

Kurt Edmund Oughstun

Department of Computer Science and Electrical Engineering

University of Vermont

ABSTRACT

The asymptotic theory of the dynamical structure of the precursor fields in a homogeneous, isotropic, locally linear, temporally dispersive medium that is causal (i.e. a medium whose dielectric permittivity satisfies the Kramers-Kronig relations) is considered in two parts. First, the accuracy of the asymptotic description of the Sommerfeld and Brillouin precursor fields in a single resonance Lorentz medium is considered by comparison with exacting numerical calculations of this dispersive field structure for both an input delta function pulse and an input unit step-function modulated signal of fixed carrier frequency. In the second part, the asymptotic description of the precursor field evolution in a double resonance Lorentz medium is considered. The appearance of a new precursor field is obtained and the conditions for its evolution are described. It is then seen that for a general multiple resonance Lorentz medium there will be the possibility of an additional precursor field structure for each additional resonance frequency present in the medium dispersion. The condition for the appearance of each such dispersive field structure may readily be obtained from the frequency dependence of the energy transport velocity for a monochromatic field in the multiple resonance medium.

I. Introduction

A complete understanding of the dynamics of electromagnetic pulse propagation in a causally dispersive medium is a fundamental problem in electromagnetic theory. Although the basis physical processes involved are well known for the time-harmonic (or continuous wave) case, their combined effects for a pulsed wave field are not in general known due to the associated mathematical complexity of the problem. Prior to this research, the dynamical evolution of an electromagnetic pulse in a homogeneous, isotropic, temporally dispersive medium has only been fully described for one specific model of the material dispersion, namely that given by the classical Lorentz theory, and then only for the simplest case of a single resonance frequency. The foundational analysis of this problem [1-3], which is now classical, clearly showed the existence of two distinctive dispersive field structures, called precursors, that are excited by an instantaneous rise-time signal, whose attenuation with propagation distance in the dispersive medium is much less than the attenuation of the signal itself. Hence, these precursor fields will persist at greater propagation distances in the medium even after the main body of the signal has been almost completely attenuated.

The most recent analysis [4-6] of this fundamental problem has yielded a complete, continuous uniform asymptotic description of both the individual precursor fields and the total field evolution in a single resonance Lorentz medium. At a more fundamental level, this analysis has led to a new physical description [5] of dispersive pulse propagation that is based on the causal energy transport velocity of monochromatic fields and replaces the old, inaccurate group velocity description. Although derived from the results obtained for a single resonance Lorentz model medium, it has been postulated that this new model has direct applicability to the description of dispersive pulse dynamics in general dispersive media and systems. A rigorous proof of the validity of this postulate would clearly be a significant contribution to mathematical physics.

We have begun this proof through an investigation of the dynamical structure of the precursor fields in a general homogeneous, isotropic, linear, temporally dispersive medium that is causal. This analysis, initiated under the Summer Faculty Research Program of the Air Force Office of Scientific Research, yields the evolutionary structure of the Sommerfeld and Brillouin precursor fields that is model independent; i.e. that is independent of the model used to describe the material dispersion. These results also provide a convenient means by which a given dispersive medium may be classified according to the types of precursor fields it supports. The results of this analysis have immediate application to the problem of microwave radiation exposure to biological tissue structures as well as to the analysis and design of low observable surfaces.

II. OBJECTIVES AND APPROACH

The overall objective of the proposed research is to develop a detailed asymptotic description of the dynamical evolution of the precursor fields in causally dispersive media that have loss. Emphasis has been placed on both the accuracy of the asymptotic theory and on the dynamical structure of the precursor fields. Particular attention has been given to the differences in the dynamical evolution of the precursor fields in Lorentz-type and Debye-type media. The effects introduced by any static or frequency-dependent conductivity that may be present in the medium will be considered in future research.

The analysis is based upon the exact Fourier-Laplace integral representation of the propagated field in a linear dispersive medium due to an input plane wave pulse, viz.

$$A(z,t) = \frac{1}{2\pi} \int_{ia-\infty}^{ia+\infty} \tilde{f}(\omega) e^{i[\tilde{k}(\omega)z-\omega t]} d\omega, \quad (1)$$

for $z \geq 0$. Here $a > \gamma_a$, where γ_a is the abscissa of absolute convergence for the integral transform. The function $\tilde{f}(\omega)$ is the temporal frequency spectrum of the initial field at the input plane at $z=0$, where

$$A(o,t) = f(t), \quad (2)$$

and where the function $f(t)$ identically vanishes for all $t<0$, so that

$$\tilde{f}(\omega) = \int_o^\infty f(t) e^{i\omega t} dt. \quad (3)$$

The medium effects are completely contained in the dispersion relation

$$\tilde{k}(\omega) = \frac{\omega}{c} n(\omega), \quad (4)$$

where $\tilde{k}(\omega)$ is the complex wavenumber at the angular frequency ω and $n(\omega)$ is the complex index of refraction of the homogeneous, isotropic, linear medium.

The representation (1) of the solution is most appropriate for numerical simulations that rely, for example, on the fast Fourier transform algorithm. With the dispersion relation (4) this representation may be written in the following form that is more suitable for asymptotic methods of approximation:

$$A(z,t) = \frac{1}{2\pi} \int_{ia-\infty}^{ia+\infty} \tilde{f}(\omega) e^{\frac{z}{c}\phi(\omega,\theta)} d\omega, \tag{5}$$

where the complex phase function $\phi(\omega,\theta)$ is defined by

$$\phi(\omega,\theta) = i\omega[n(\omega) - \theta]. \tag{6}$$

Here

$$\theta = \frac{ct}{z} \tag{7}$$

is a dimensionless parameter that characterizes a given space-time point in the field, where c is the vacuum speed of light.

From the constitutive relations for electromagnetic fields in linear, homogeneous, isotropic, temporally dispersive causal media, admissable models that describe the behavior of the complex dielectric permittivity $\epsilon(\omega) = \epsilon_r(\omega) + i\epsilon_i(\omega)$ in the complex ω-plane must obey the symmetry relation [7]

$$\epsilon^*(\omega) = \epsilon(-\omega^*), \tag{8}$$

where the complex index of refraction is given by the principal branch of the square root expression

$$n(\omega) = [\epsilon(\omega)]^{1/2}. \tag{9}$$

Furthermore, due to the analyticity properties of $\epsilon(\omega)$ as expressed by Titchmarsh's theorem [7], the real and imaginary parts of the complex dielectric permittivity are required to satisfy the dispersion relations

$$\epsilon_r(\omega) = 1 + \frac{1}{\pi} ⨍_{-\infty}^{\infty} \frac{\epsilon_i(\zeta)}{\zeta - \omega} d\zeta, \tag{10a}$$

$$\epsilon_i(\omega) = -\frac{1}{\pi} ⨍_{-\infty}^{\infty} \frac{\epsilon_r(\zeta) - 1}{\zeta - \omega} d\zeta. \tag{10b}$$

The corresponding equations for the complex index of refraction are referred to as the Kramers-Kronig relations. Similar relations hold for the frequency-dependent conductivity. An unambiguous definition of the precursor fields that are associated with dispersive pulse propagation may be directly obtained through the asymptotic description of the field given by Eq. (5). In that description each contribution to the propagated field is precisely defined according to its source term in the asymptotic expansion [4,6]. The precursor fields are then seen to arise from the saddle points of the complex phase function $\phi(\omega,\theta)$ which depends explicitly upon the medium properties through the complex dielectric permittivity and is independent of the input field structure*. The dynamical evolution of the precursor fields is then completely determined by the dynamics of these saddle points in the complex frequency plane, and these satisfy the equation of motion

$$n(\omega) + \omega n'(\omega) - \theta = 0 \quad , \tag{11}$$

where the prime denotes differentiation with respect to ω.

The dynamics of both the near and distant saddle points that are characteristic of any causally dispersive medium satisfying the dispersion relations (10) and such that the frequency dependence of $\epsilon_i(\omega')$ along the positive real ω'-axis is such that the dielectric loss is significant only within a finite frequency domain $[\omega_o,\omega_m]$, where

$$0 << \omega_o < \omega_m << \infty \quad , \tag{12}$$

have been derived in the initial phase of this research [8] for both perfect dielectrics and dielectric media with a nonvanishing static conductivity

*This does not mean that the form of the input field has nothing to do with the precursor fields, for their appearance is clearly dependent upon both the envelope spectrum and the signal frequency of the input field.

(imperfect dielectrics). Notice that the left-hand inequality in (12) does not imply that there is no loss at positive frequencies below ω_o, but only that the dielectric loss in this frequency interval is small in comparision to that for frequencies in the internal $[\omega_o, \omega_m]$. Furthermore, it is important to note that this inequality only applies to the dielectric loss and has no bearing on the loss due to any conductivity that may be present in the medium. The distant saddle points, which evolve in the region of the complex ω-plane above ω_m, give rise to the so-called Sommerfeld precursor field, while the near saddle points, which evolve in the region about the origin of the complex ω-plane below ω_o, give rise to the so-called Brillouin precursor field. The initial part of this research has shown that under proper initial field excitation in a perfect dielectric medium:

(i.) the classical Sommerfeld precursor field will evolve in any perfect dielectric medium provided that the real part of the complex dielectric permittivity approaches its infinite frequency value from below as $\omega' \to \infty$;

(ii.) the classical Brillouin precursor field will evolve in any perfect dielectric medium that is dispersive;

while in an imperfect dielectric with a nonvanishing static conductivity σ:

(iii.) the classical Sommerfeld precursor will evolve if σ is sufficiently small such that the real part of the complex dielectric permittivity approaches its infinite frequency value from below as $\omega' \to \infty$; and

(iv.) a modified Brillouin precursor field will evolve in a fashion that is essentially the reverse of that in a perfect dielectric.

The analysis performed under this research initiation program has considered both the accuracy of the asymptotic theory of precursor fields as well as the generalization of the asymptotic description to a general, multiple resonance Lorentz medium. These results are presented in detail in the following two preprints of papers that have been accepted for publication in the Journal of the Optical Society of America. The first paper will be published in the 1989 Feature Issue of the Journal of the Optical Society of American A on Mathematics and Modeling in Modern Optics, while the second paper will be published in the May, 1989 issue of the Journal of the Optical Society of America B.

IV. REFERENCES

[1.] Arnold Sommerfeld, "Uber die Fortpflanzung des Lichtes in Disperdierenden Medien," Ann. Physik, 44, 177-202 (1914).

[2.] Leon Brillouin, "Uber die Fortpflanzung des Licht in Disperdierenden Medien," Ann.Physik, 44, 203-240 (1914).

[3.] Leon Brillouin, Wave Propagation and Group Velocity, (Academic Press, N.Y., 1960).

[4] Kurt Edmund Oughstun, Propagation of Optical Pulses in Dispersive Media, Ph.D. Thesis, University of Rochester (University Microfilms International, Ann Arbor, Michigan, 1978).

[5.] George C. Sherman and Kurt Edmund Oughstun, "Description of Pulse Dynamics in Lorentz Media in Terms of the Energy Velocity and Attenuation of Time-Harmonic Waves," Phys.Rev.Lett. 47, 1451-1454 (1981).

[6.] Kurt Edmund Oughstun and George C. Sherman, "Propagation of Electromagnetic Pulses in a Linear Dispersive Medium with Absorption (the Lorentz Medium)," J.Opt.Soc.Am.B (to be published).

[7.] H.M. Nussenzveig, Causality and Dispersion Relations, (Academic Press, New York, 1972).

[8.] Kurt Edmund Oughstun, "On the General Existence of Precursor Fields in a Causally Dispersive Medium," Final Report, 1987 USAF-UES SFRP (F49620-85-C-0013).

[9.] C.H. Durney, T.G. Stockham, and K. Moten, "Electromagnetic Pulse Propagation, "Final Report (Subcontract A-3529-240, Georgia Institute of Technology, 1987).

Appendices can be obtained from
Universal Energy Systems, Inc.

FINAL REPORT NUMBER 109
REPORT NOT PUBLISHABLE AT THIS TIME
Dr. Ralph Peters
760-7MG-091

MODEL SYSTEMS FOR ASSESSING THE EFFECTS OF MICROWAVE RADIATION ON THE IMMUNE SYSTEM

(Contract #F49620-85-C-0013/SB5851-0360)

A FINAL REPORT

submitted to

THE RESEARCH INITIATION PROGRAM
UNIVERSAL ENERGY SYSTEMS, INC.

by

Stephen B. Pruett, Ph.D.
Department of Biological Sciences
Mississippi State University

Submitted December 13, 1988

INTRODUCTION

The experimental plans described in the original proposal ("Model Systems for Assessing the Effects of Microwave Radiation on the Immune System", Appendix) were generally followed. The proposal contained four specific goals, and results will be reported here as they relate to these goals. An additional line of investigation was also pursued, and results from that investigation are reported. At the end of each Results section is a brief discussion and recommendation for future studies. One such recommendation (p.20) has been expanded and presented as a Grant Proposal to the Air Force Office of Scientific Research.

RESULTS AND RECOMMENDATIONS FOR FUTURE STUDIES

Specific Goal 1 (Continue studies of the relationship between toxic forms of oxygen and thiols in macrophage-like cell lines)

The initial experiments with colloidal glucose oxidase-horseradish peroxidase (GOHRP) particles were conducted essentially as described in the proposal. Briefly, particles were pre-incubated with RAW 264.7 (macrophage-like) cells. The cells were washed and placed in culture for various periods of time, and the thiol content of the culture supernatant fluid was determined using our modicfication of the DTNB reduction assay (1). The data in Table 1 suggest that exposure to colloidal GOHRP particles increases thiol production by RAW 264.7 cells. Since tests of the particles with the Limulus Amebocyte Lysate Assay for bacterial endotoxin indicated that the particles contained a maximum of 0.0006μg/ml of endotoxin (LPS) when used at a 1/50 dilution, RAW 264.7 cells were pre-incubated with LPS at the same time as other samples were pre-incubating with GOHRP particles. The results indicate that pre-incubation with 0.04μg/ml LPS could yeild increases in thiol production as great as noted with GOHRP treatment (Exp. 2), but pre-incubation with a

TABLE 1. Effects of Glucose Oxidase Horseradish Peroxidase Colloidal Particles (GOHRP) and Bacterial Endotoxin (LPS) on Thiol Production by RAW 264.7 Cells

Exp.	Cell Type	Initial Cell Density	Treatment	Thiols[c] (µM)
1	RAW 264.7	$1.5x10^6$/ml	-	31.7±1.1
			LPS (.04-pre)[a]	31.3±1.1
			LPS (.04)	41.9±2.1
			GOHRP (1/50-pre)[b]	38.6±1.1
			GOHRP (1/100-pre)	40.0±0.4
			GOHRP (1/200-pre)	39.2±1.0
2	RAW 264.7	$1.3x10^6$/ml	-	10.6±0.6
			LPS (.04-pre)	21.4±1.0
			LPS (.04)	45.7±0.1
			GOHRP (1/50-pre)	22.3±2.1
			GOHRP (1/100-pre)	14.7±1.2
			GOHRP (1/200-pre)	17.2±6.4
3	RAW 264.7	$1.3x10^6$/ml	-	28.4±2.9
			LPS (.0016-pre)	37.4±2.1
			LPS (.0016)	49.7±1.0
			GOHRP (1/50-pre)	37.6±1.0
			GOHRP (1/100-pre)	42.6±2.7
4	RAW 264.7	$1.3x10^6$/ml	-	41.7±0.8
			LPS (.00016-pre)	44.2±0.8
			LPS (.00016)	53.4±5.7
			LPS (.005-pre)	53.0±0.9
			LPS (.005)	76.5±2.9
			GOHRP (1/200-pre)	46.7±1.3
			GOHRP (1/800-pre)	47.1±1.9

a) Values in parentheses are concentrations (µg/ml); -pre indicates a 3 hour pre-incubation with the compound followd by a wash and subsequent 20 hour incubation. Samples not designated -pre were incubated with the compound for the entire 23 hour culture period.

b) GOHRP was at an initial concentration of 1mg/ml and was used at the indicated final dilutions.

c) All values reported are means±standard deviation of triplicate samples.

concentration of LPS more similar to that noted in the GOHRP particles did not have as great an effect on thiol production as did the GOHRP particles (Exp. 4). Thus the enhancement of thiol production by GOHRP particles cannot be attributed solely to contamination of the particles with LPS, which is an excellent stimulator of thiol production. These results suggest that the toxic forms of oxygen produced by the GOHRP particles may be involved in the stimulation of macrophage-like cells to produce thiols. However, it remains possible that pinocytosis of virus-sized particles in general is a weak stimulator of thiol production. This will be investigated by determining the effects of control particles composed of human serum albumin on RAW 264.7 cells. These particles have been produced in Dr. Johnathan Kiel's lab (Division of Radiation Physics, U.S. Air Force School of Aerospace Medicine) and are on hand for use in these experiments.

The relationship between stimulation for generation of toxic forms of oxygen and stimulation for thiol production was also investigated with an agent which blocks intracellular mobilization of calcium ions and generation of toxic forms of oxygen (2). The effect of this inhibitor (8-(diethylamino)octyl 3,4,5-trimethoxybenzoate hydrochloride, TMB-8) on the increased thiol production stimulated by LPS was examined in several experiments (Table 2). The data indicate that TMB-8 at 75 or 37.5 μM decreases the response of RAW 264.7 cells to LPS. At higher concentrations TMB-8 kills most of the cells (results not shown), and at 75μM it prevents growth. Since LPS also inhibits growth, cultures treated with TMB-8 at 75μM, with LPS, or with both compounds have comparable numbers of cells at 24 hours. Thus, the lower levels of thiols in cultures with LPS and TMB-8 than in cultures with LPS only are not due to decreased cell

Table 2. Effects of LPS and TMB-8 (an Inhibitor of Intracellular Calcium Ion Mobilization) on Thiol Production by RAW 264.7 Cells

Exp.	Treatment	Final Cell Density ($x10^{-5}$)	Thiols[a] (μM)
6	-	19.0/ml	<4
	LPS (.2μg/ml)	11.0/ml	79.6±3.3
	LPS (.02μg/ml)	12.0/ml	58.3±6.7
	TMB-8 (75μM)	13.0/ml	11.6±1.6
	LPS (.2)+ TMB-8	11.0/ml	24.2±1.6
	LPS (.02) + TMB-8	9.6/ml	20.2±2.8
9	-	8.9/ml	7.2±1.3
	TMB-8 (75μM)	6.9/ml	7.2±0.6
	LPS (0.02μg/ml)	5.4/ml	41.7±0.8
	LPS + TMB-8	6.4/ml	5.7±3.4
10	-	14.0/ml	<1
	LPS (.02μg/ml)	7.0/ml	44.2±2.1
	TMB-8 (75μM)	7.1/ml	6.8±0.4
	TMB-8 (37.5μM)	8.1/ml	8.0±2.1
	LPS + TMB-8 (75)	5.6/ml	23.0±1.5
	LPS + TMB-8 (37.5)	7.1/ml	31.9±0.4
11	-	10.0/ml	4.5±0.2
	LPS (1μg/ml)	4.8/ml	23.1±1.1
	LPS (.1μg/ml)	4.7/ml	24.9±1.3
	LPS (.01μg/ml)	7.7/ml	19.4±0.7
	TMB-8 (75μM)	8.8/ml	<1
	LPS (1) + TMB-8	4.3/ml	20.9±1.3
	LPS (.1) + TMB-8	5.4/ml	16.1±2.9
	LPS (.01) + TMB-8	8.8/ml	14.1±3.1

a) Cells were incubated with agents at the concentrations indicated for 24 hours before analysis for thiols in culture supernates. Values shown are means±standard deviation for triplicate samples numbers caused by TMB-8. These results demonstrate that decreased mobilization of intracellular calcium ions (or possibly some other unknown function of TMB-8) decreases the response of the cells to LPS. This could be due to direct interference with signal transduction for thiol production, or it could be the result of interference with signal transduction for another function (e.g., generation of toxic forms of oxygen) which subsequently causes the increase in thiol production.

If thiol production were a secondary response to a particular function of activated cells, one would expect that only stimuli of that function would lead to increased thiol production. Conversely, if thiol production is directly stimulated by any agent which leads to intracellular calcium mobilization, many types of stimuli might lead to increased thiol production. It has been reported that lymphokine-rich supernates from Concanavalin A activated mouse splenocytes cannot activate RAW 264.7 cells to become tumoricidal, whereas LPS at concentrations >0.01μg/ml can accomplish this (3). One indication of macrophage activation is increased production of lactic acid accompanied by decreased pH in the culture medium (4). As indicated in Table 3, Con A supernates and LPS decrease the pH of the medium in RAW 264.7 cultures. Therefore, this constitutes a system in which the cells show indications of activation but are presumably different in terms of tumoricidal capacity (3). The results in Table 3 indicate that thiol production is increased in both Con A supernate- and LPS-treated cells. Interestingly, the relative increase in thiol production caused by the two treatments paralleled the relative decrease in pH. This suggests the possibility that thiol production (and decreased pH) may be a generalized indicator of macrophage activation and not secondary to any particular function of activated macrophages. The data in

Table 3. Effects of Lymphokine-Rich Splenocyte Supernates on RAW 264.7 Cells.

Exp.	Initial Cell Density	Treatment	pH	Thiols (μM)
20[a]	10^6/ml [b]	-	7.11±.04	8.5±1.1
		20% LK[c]	6.78±.05	17.8±2.2
		10% LK	6.84±.03	17.4±1.1
		5% LK	6.86±.02	16.1±1.1
		20% CS[d]	7.03±.02	10.7±0.6
		10% CS	7.08±.05	12.2±0
		5% CS	7.12±.01	12.8±0.7
		LPS (.02μg/ml)	6.42±.02	62.9±5.9

a) All samples except LPS contained polymyxin B (5μg/ml).

b) Final cell density was not significantly different from control, except in LPS-treated cultures in which cell density was about one-half that of controls.

c) LK=lymphokine-rich supernate from Concanavalin A (2μg/ml). stimulated mouse splenocytes ($2x10^6$/ml, 24hr.).

d) CS=control supernate made by incubating mouse spleen cells as above except without Concanavalin A. After supernate was harvested, Concanavalin A was added to make a 2μg/ml solution.

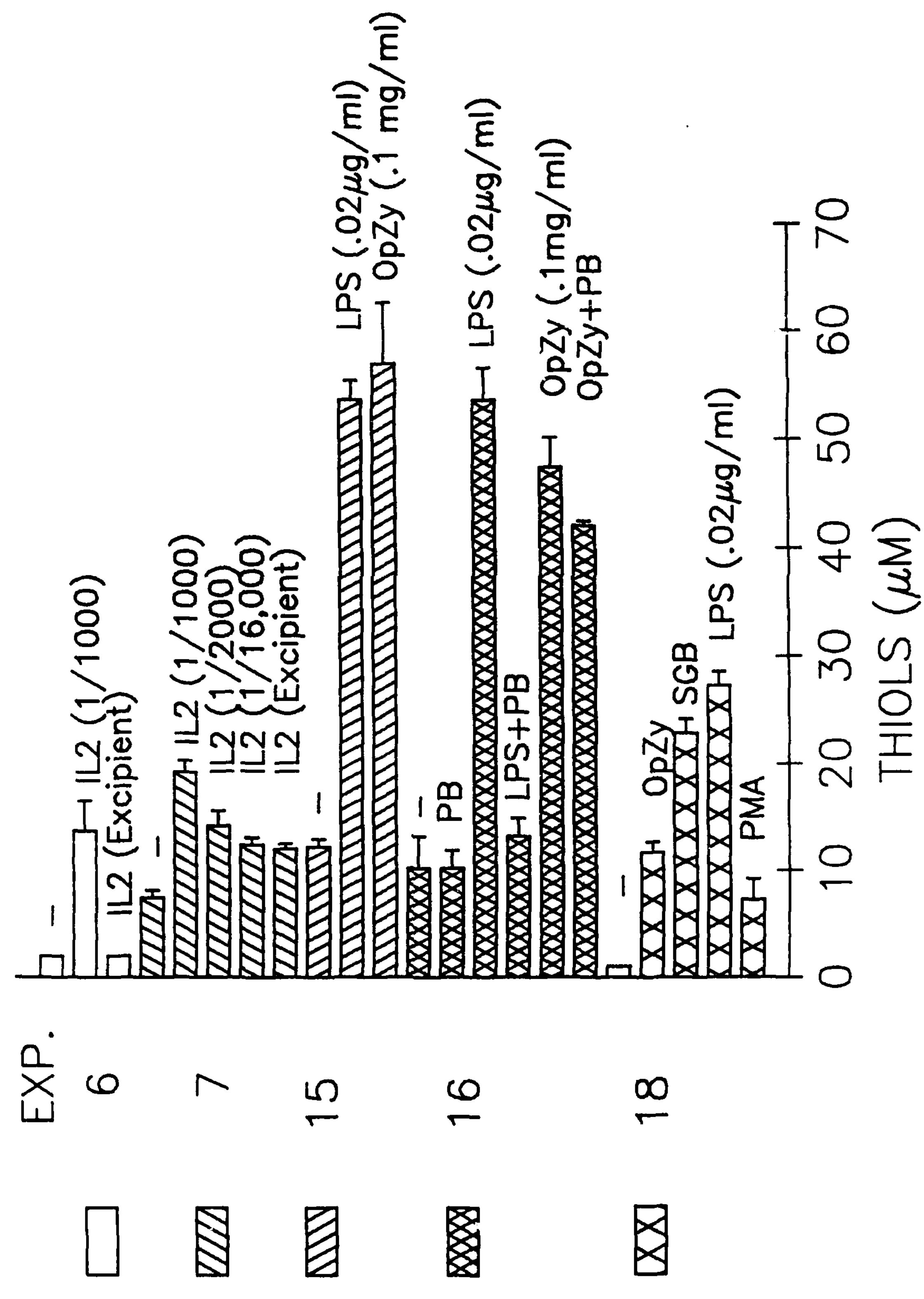
Figure 1 which demonstrate increased thiol production with all stimuli tested are consistent with this possibility. However, many of these stimuli also lead to increased production of toxic forms of oxygen by macrophages (3,5,6) and it is possible that all of them do. Therefore, it is not possible to exclude the option that toxic forms of oxygen stimulate production of thiols (perhaps as a protective measure against oxidative damage).

Future studies in this area should directly examine the effect of various stimuli on production of toxic forms of oxygen and other functions of activated macrophages and the association or dissociation of these functions and thiol production. Consistent association of thiol production and generation of toxic forms of oxygen (or another function) could indicate a causal relationship or a common signal pathway, whereas any case of dissociation would refute such a relationship. The results obtained might support a model in which thiol production is a a sensitive indicator of macrophage priming which increases quantitatively as full activation is induced. If this were the case thiol production might serve as a direct indicator of the activation status of the cell and would suggest a system in which various activation functions are triggered in a manner which is dependent on the intensity of the stimulus and subsequent second messenger signals (7).

Specific Goal 2 (Develop a practical method of assessing the effects of microwave radiation on the ability of macrophages to support lymphoid cells, and assess the involvement of thiols in any effects noted)

Since the support of lymphoid cells by macrophages is dependent on production of thiols by the macrophages (1), a method was sought by which to examine the effects of microwave radiation on thiol production by RAW 264.7 cells as an indicator

Figure 1. Effects of various stimulators on thiol production by RAW 264.7 cells and mouse peritoneal macrophages. The concentration of thiols (μM) in supernates at the end of 24 hour cultures is reported. Values represent means for triplicate samples, and the bars indicate the standard errors. Abbreviations for stimulators (which were present during the entire culture period) are as follows: IL2= Cetus recombinant human Interleukin 2 at an initial concentration of approximately 400,000 U/ml; IL2 excipient= Cetus preparation containing all components of IL2 preparation except IL2. Used as a control to insure effects noted are due to IL2 and not possible LPS contamination from bacteria in which IL2 gene is expressed; LPS= lipopolysaccharide, from E. coli 0111 phenolic extract;OpZy= opsonized (normal mouse serum 1/2, 30 min.) zymosan; PB= polymyxin B at 5μg/ml to inhibit activity of any LPS present in the other stimulators. As noted in Exp. 16, PB inhibits LPS at concentrations up to 0.02μg/ml; SGB= Streptococcus Group B (heat killed); PMA= phorbol myristate acetate. All of these compounds were selected on the basis of reports that they are able to activate at least one type of macrophage function. RAW 264.7 cells were used in Exp. 6, 7, 15, and 16, whereas thioglycollate-elicited mouse peritoneal macrophages were used in Exp. 18.


EXP.
6
7
15
16
18
–
IL2 (1/1000)
IL2 (Excipient)
–
IL2 (1/1000)
IL2 (1/2000)
IL2 (1/16,000)
IL2 (Excipient)
–
LPS (.02µg/ml)
OpZy (.1 mg/ml)
–
PB
LPS (.02µg/ml)
LPS+PB
OpZy (.1mg/ml)
OpZy+PB
–
OpZy
SGB
LPS (.02µg/ml)
PMA
0
10
20
30
40
50
60
70
THIOLS (µM)

of possible effects which would be observed in macrophage lymphocyte co-cultures. Since the location of the facilities for microwave exposure of cells (Brooks AFB) and the location where culture supernates were to be analyzed for thiols (Mississippi State University) are a considerable distance apart, a method for short term microwave exposure followed by removal and freezing of supernate samples for shipping and analysis was needed. Initial experiments indicated that thiols in culture supernates of RAW 264.7 cells were stable for over 3 months at -70 C (results not shown). Several experiments, including two shown in Table 4 (Exp. 8 & 9), involved a 4 hour incubation in the presence of LPS followed by analysis of thiols in supernate samples. The results of these experiments were inconclusive, in that differences in thiol levels between negative controls and LPS-treated cultures were noted in some, but not all experiments. This approach would also be impractical, since the microwave transmitter to be used in these experiments is not suitable for the continuous 4 hour exposures which would be required (J.L. Kiel, personal communication). Subsequent experiments involved shorter incubation times. In exp. 14, 16 and 17, the cells were pre-incubated at high density for 35 min. or 1.5 hr. in the presence or absence of LPS, then washed and incubated at low density for 20 hours for assessment of thiol production. Microwave exposure during the pre-incubation period would be expected to alter subsequent thiol production if microwave radiation affects early activation events. The large difference in thiol production between negative control and LPS-treated cultures indicates that this protocol should be suitable for detection of any such effects. The effects of microwave radiation on thiol production by cells which have already been activated cannot be determined by this approach, since the microwave exposure must occur during

Table 4. Effects of Various Incubation and Pre-Incubation Times on the Stimulation of Thiol Production by LPS

Exp.	Pre-Incubation			Incubation		Thiols[b] (μM)
	Cell Density	Time	Treatment[a]	Cell Density	Time	
16	10^7/ml	35 min.		$5x10^5$/ml	20 hr.	10.3±0.6
		35 min.	LPS (.01)		20 hr.	35.9±1.2
		1.5 hr.	-		20 hr.	12.4±2.7
		1.5 hr.	LPS (.01)		20 hr.	54.6±0.4
17	10^7/ml	35 min.	-	$5x10^5$/ml	20 hr.	15.6±1.6
		35 min.	LPS (.01)		20 hr.	39.0±9.9
		1.5 hr.	-		20 hr.	15.2±4.0
		1.5 hr.	LPS (.01)		20 hr.	45.9±9.9
14	10^7/ml	2.0 hr.	-	$5x10^5$/ml	22 hr.	11.3±2.0
		2.0 hr.	LPS (1)	$5x10^5$/ml	22 hr.	18.0±1.1
		2.0 hr.	LPS (.1)	$5x10^5$/ml	22 hr.	13.8±1.7
		2.0 hr.	LPS (.01)	$5x10^5$/ml	22 hr.	17.5±1.3
12	10^7/ml	1.5 hr.	-	-	-	21.3±1.6
		1.5 hr.	LPS (1)	-	-	24.1±3.0
		3.0 hr.	-	-	-	25.2
		3.0 hr.	LPS (1)	-	-	27.2
13	10^7/ml	3.0 hr.	-	-	-	25.3±2.5
		3.0 hr.	LPS (.02)	-	-	18.7±2.0
9	$2.8x10^6$/ml	4.0 hr.	-	-	-	8.4±1.2
		4.0 hr.	LPS (.01)	-	-	11.3±2.4
8	$3.5x10^6$/ml	4.0 hr.	-	-	-	7.0±1.0
		4.0 hr.	LPS (.2)	-	-	9.9±0.4
		4.0 hr.	LPS (.01)	-	-	12.4±3.7

a) Values in parentheses are concentrations of LPS (μg/ml).

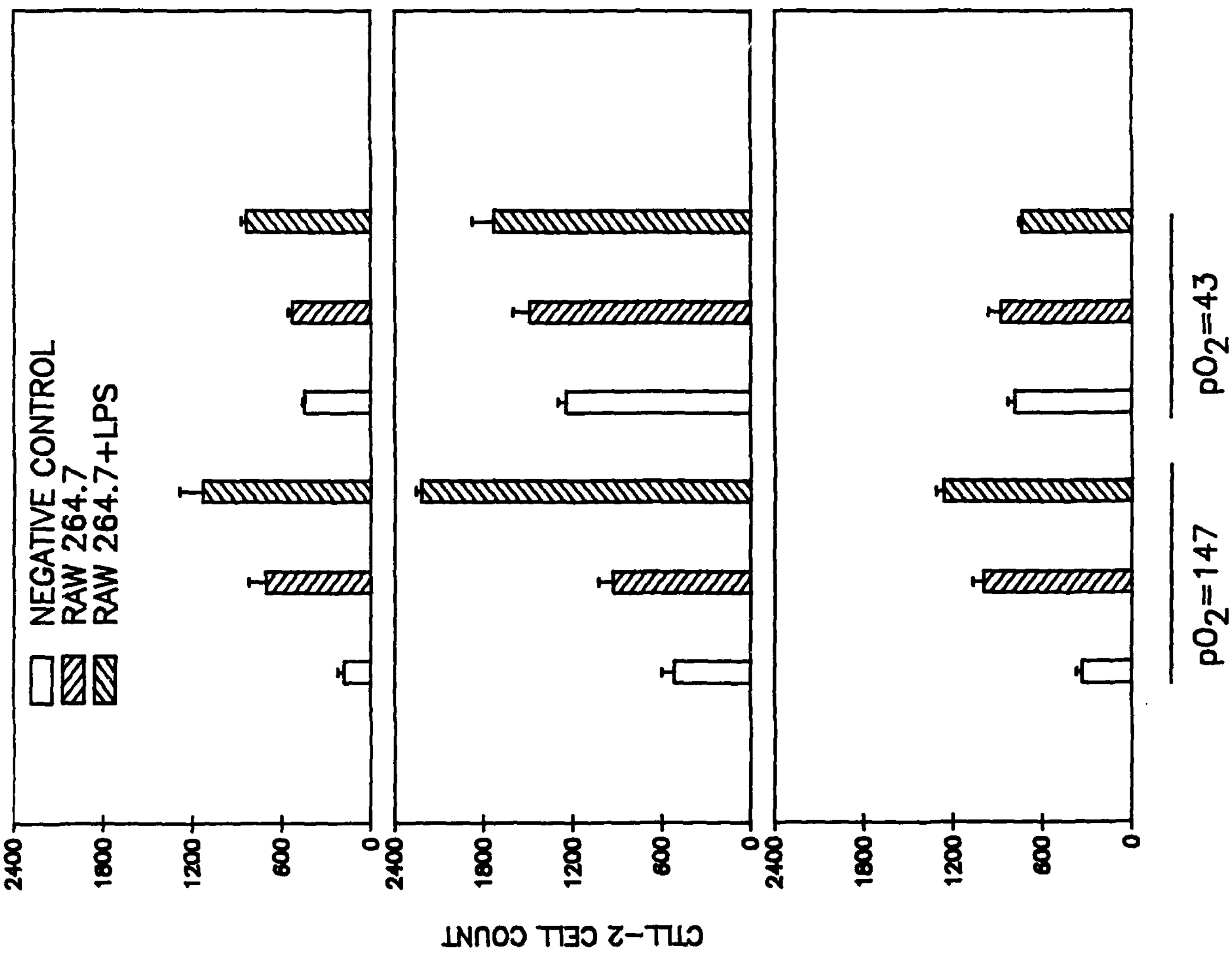
b) Thiol concentrations are means±standard deviation for triplicate samples.

the entire period for which thiol production is assessed to insure that absence of observed effect is not due to a rebound phenomenon following exposure. Unfortunately, the activation of macrophages for increased thiol production requires more than 3 hours of exposure to LPS (Exp. 8, 9, 12, & 13). Therefore, exposure to microwaves during the early hours of dense cultures would not allow assessment of the effects of microwave radiation on the LPS-induced increase in thiol production. An alternate approach based on the results of Exp. 14, 16, & 17 would be to incubate cultures with or without LPS for 20 hours to fully activate the LPS-treated cells, then centrifuge and concentrate the cells for short-term microwave exposure followed by harvest of supernate and analysis for thiols. The clear difference in thiol production between cells activated for this period of time and control cells has been demonstrated in a number of experiments (e.g., Tables 1 and 2) as has the ability to measure thiols produced in short-term, high-density cultures (Exp. 12). This method will be directly examined prior to microwave exposure experiments. Such experiments have not been conducted to date due to problems with the microwave transmsitter (J.L. Kiel, personal communication). The approaches outlined above should permit detection of any effects of microwave radiation on the early activation events stimulated by LPS which lead to increased thiol production as well as effects on thiol production per se, and these studies will be pursued when the microwave transmitter is operational.

Specific Goal 3 (Investigate the potential in vivo importance of thiol production by macrophages during lymphocyte responses by examining the behavior of the model system which has already been developed under low oxygen conditions such as those found in most tissues *in vivo*)

This goal prompted two detailed mechanistic studies of the mechanism(s) by which feeder cells and thiols such as 2-mercaptoethanol enhance lymphocyte growth and survival in culture (see manuscripts in Appendix). The results demonstrate that feeder cells and 2-mercaptoethanol probably function by providing antioxidant protection and be enhancing protein synthesis in cultured lymphoctyes. However, if the lymphocytes are cultured at physiological oxygen levels (pO_2=43) the dependence on antioxidant protection is decreased, and the effects of RAW 264.7 cells as feeder cells are diminished ("Involvement and Relative Importance of at Least Two Distinct Mechansisms in the Effects of 2-Mercaptoethanol on Murine Lymphocytes in Culture", Appendix; figure 2). In addition, others report that the increased cyst(e)ine uptake which apparently drives the increased protein synthesis we observed is probably not important *in vivo*, since freshly isolated lymphocytes have much higher levels of glutathione (an indicator of cysteine availability) than cultured lymphocytes (8). Collectively, these results suggest that thiol production by macrophages *in vivo* is probably not critical for normal lymphocyte survival and growth. It is possible, however, that the increased thiol production demonstrated upon macrophage activation could serve a protective role in inflammatory responses. We have recently found that human neutrophils do not produce significant levels of thiols in response to a variety of activation stimuli (results not shown). This suggests that macrophages and perhaps lymphocytes may produce thiols which protect surrounding tissues from the toxic forms of oxygen secreted by activated neutrophils. This hypothesis will be the subject of future investigations.

Figure 2. Effects of reduced oxygen concentration on RAW 264.7-CTLL-2 co-cultures. Co-cultures were performed as described previously (1) by incubating CTLL-2 cells with RAW 264.7 cells at initial concentrations of 10^5/ml each. After 24 hours at the indicated oxygen levels, the 24-well plates were swirled to suspend the CTLL-2 cells for counting. RAW 264.7 cells are adherent and are not dislodged by this procedure. Under normal oxygen levels (pO_2=147) CTLL-2 cells required addtional thiols for optimal growth and survival. Decreased oxygen levels allow improved growth in the absence of exogenous thiols and decrease the enhancement noted when RAW 264.7 cells are used as a source of thiols. LPS is used to stimulate increased thiol production by the RAW 264.7 cells in co-cultures and does not directly affect CTLL-2 cell survival or growth (1).


NEGATIVE CONTROL
RAW 264.7
RAW 264.7+LPS
CTLL-2 CELL COUNT
2400
1800
1200
600
0
pO_2=147
pO_2=43

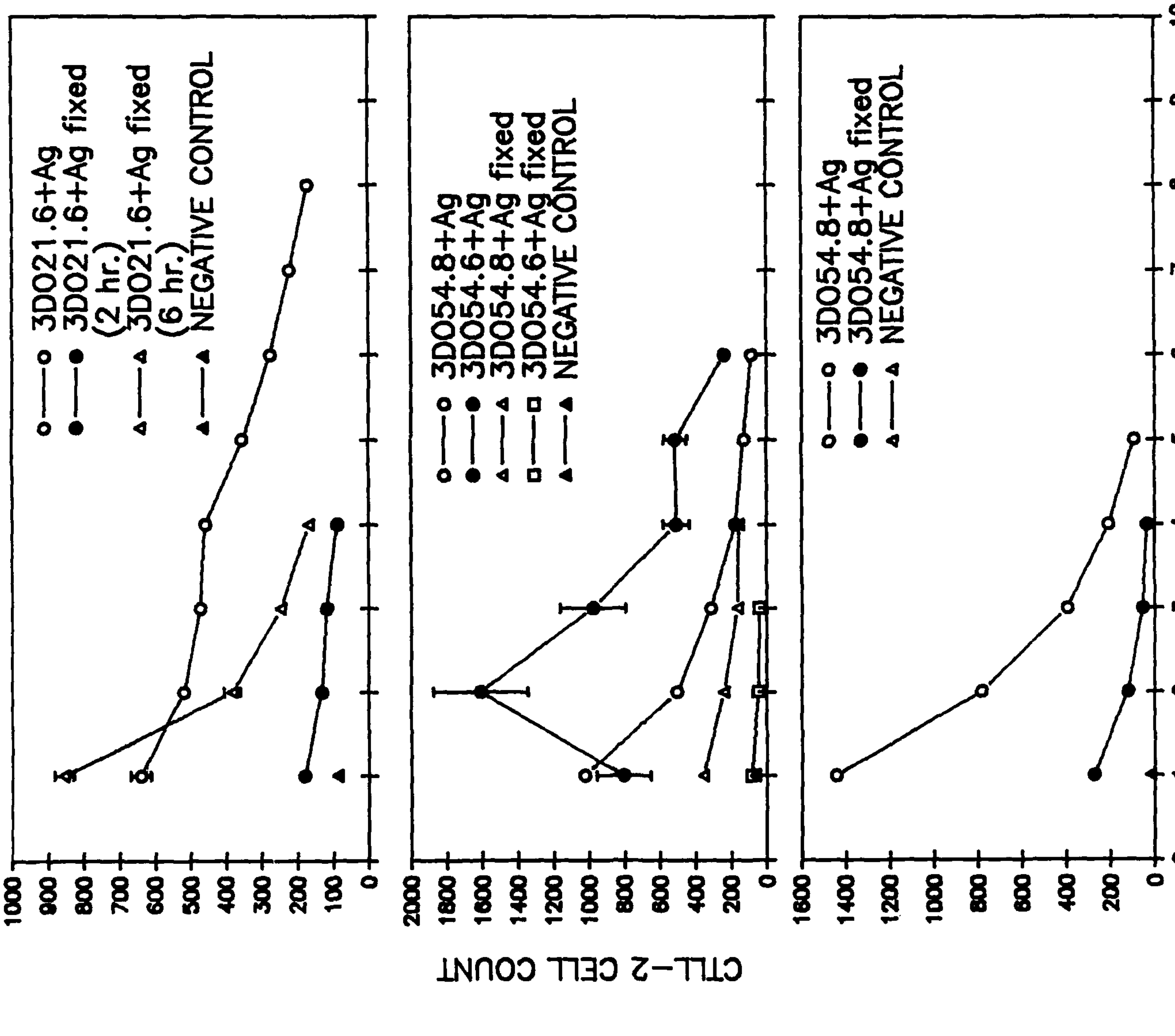
Specific Goal 4 (Begin developing a model system which would detect alterations of antigen processing and presentation as well as accessory cell-lymphocyte signaling and lymphokine production)

Helper T cell lines for these studies were obtained from Dr. Phillipa Marrack and the experiments were performed according to her published procedures (9). The results (Figure 3) indicate that this system is suitable for the study of the effects of microwave radiation or other physical and chemical agents on antigen processing and presentation. A particularly useful attribute of this system is that the antigen-presenting cells can be fixed with glutaraldehyde after exposure to antigen for 2 hours or more and still retain a portion of their ability to stimulate Interleukin 2 production by the T cell lines. Thus, it is possible to physically and temporally separate antigen processing/presentation and the T cell response to presented antigen. This will allow determination of whether the agent under investigation affects antigen processing/presentation, the T cell response to the presented antigen, or both. The relatively poor response observed with cells fixed after only 2 hours exposure to antigen (Figure 3) would need to be improved by manipulation of cell or antigen density before their use in microwave exposure experiments would be practical.

Other Studies

As the investigations described above progressed, my laboratory was also involved in collaborative studies with Drs. Janice and Howard Chambers (Deparments of Biological Sciences and Entomology, Mississippi State Univeristy) involving the effects of selected organophosphorous compounds on the immune system. Based on a report that one organophosphorous compound inhibits

Figure 3. A system for assessing the effects of physical and chemical agents on antigen processing and presenting and on T cell production of IL2 in response to presented antigen. Three T Helper Hybridomas (3DO21.6, 3DO54.8, and 3DO54.6) were kindly provided by Dr. Phillipa Marrack. These cells respond to the antigen ovalbumin in the context of $H\text{-}2^k$. The antigen presenting cells are A20.2J which were obtained from the American Type Culture Collection. Antigen at 1mg/ml was incubated with 10^5 A20.2J cells and 10^5 T cell hybridomas for 24 hours in a volume of 300µl, except in some experiments the A20.2J cells were pulsed with antigen for the indicated times, then fixed with glutaraldehyde and washed before adding them to the wells. Culture supernate was harvested and serially diluted in 100µl volumes. An equal volume of IL2-dependent CTLL-2 cells was added to each well and growth was assessed after 24 hours using an electronic particle counter. The production of IL2 by the T cell hybridomas indicates appropriate activation by the antigen presenting cells and is demonstrated by proliferation of the CTLL-2 cells in the presence of culture supernates from the T cell hybridomas. A20.2J cells and T cell hybridomas separately or both cell types together without antigen did not produce detectable IL2 (results not shown).



3DO21.6+Ag
3DO21.6+Ag fixed (2 hr.)
3DO21.6+Ag fixed (6 hr.)
NEGATIVE CONTROL
3DO54.8+Ag
3DO54.6+Ag
3DO54.8+Ag fixed
3DO54.6+Ag fixed
NEGATIVE CONTROL
3DO54.8+Ag
3DO54.8+Ag fixed
NEGATIVE CONTROL
CTLL-2 CELL COUNT
RECIPROCAL LOG2 DILUTION

splenocyte responses by stimulating inhibitory macrophages (10) and on our results indicating a potent inhibitory effect of other compounds on mitogen-induced splenocyte proliferation (11), the effects of selected organophosphorous compounds on thiol production by macrophages was examined. Since results reported above indicate that thiol production may be a sensitive indicator of macrophage activation, this seemed a reasonable approach. In addition, activated macrophages have been reported to inhibit lymphocyte responses (12, 13). The results indicate that the compounds tested enhance thiol production by freshly isolated, thioglycollate-elicited mouse macrophages (Table 5). This work should be followed by mixing experiments to determine the effect of macrophages treated with organophosphorous compounds on lymphocyte responses. An inhibitory effect would be consistent with a generalized mechanism of action for such compounds involving the generation of activated, inhibitory macrophages. To further understand the detailed mechanism(s) of such agents and to estimate their effects on the status of the immune system in vivo, it will be necessary to understand the relationships between macrophage activation status, various functions of activated macrophages such as catabolism of antigen and production of toxic forms of oxygen, and immunologically important functions such as antigen processing/presentation and kiling of microbes and tumors. These relationships are poorly documented at present, and a proposal had been submitted to the Air Force Office of Scientific Research which deals with this matter. Such information should be valuable in future investigations of chemical or physical agents which affect the immune system by altering macrophage activation status.

Table 5. Effects of Selected Organophosphorous Compounds on Thiol Production by RAW 264.7 cell and Thioglycollate-elicited mouse peritoneal macrophages.

Exp.	Cell Density (Initial)	Cell Type	Treatment	Thiols (µM)
5	$5x10^5$/ml	RAW 264.7	PSCP 100µM [a,b]	0
			50	3.8±0.3
			25	18.8±1.7
			12.5	25.4±0.6
			6.25	30.9±0.1
			MePxn 100	29.8±4.4
			50	23.8±6.8
			25	12.0±4.4
			12.5	12.6±4.9
			Pxn 100	37.4±13.4
			50	18.6±1.6
			25	18.3±1.6
			12.5	15.6±0.8
			Vehicle (ethanol)[c]	
			100	4.6±2.7
			50	6.8±1.6
			25	7.6±0.3
			12.5	8.4±1.4
18	$1x10^6$/ml	Macrophages	-	3.6±1.8
			PSCP 5.0 [d]	8.4±0.4
			1.0	7.6±0.2
19	$1x10^6$/ml	Macrophages	Vehicle	<2.4
			EPNxn 100	35.6±0.2
			50	15.5±0.9
			25	8.4±1.3
			MPNxn 100	25.7±1.3
			50	9.1±3.0
			25	3.2±0.4
			PPNxn 100	16.0±0.9
			50	4.1±0.9
			25	-0.9±1.3
			BPNxn 100	11.4±0.2
			50	4.5±1.5
			25	1.9±1.5
			APNxn 100	14.7±1.3
			50	4.5±1.6
			25	2.8±0.2

a) Abbreviations of test agents are as follows: PSCP=phenyl saligenin cyclic phosphate; MePxn=methyl paraoxon; Pxn=paraoxon; EPNxn=O-Ethyl O-p-nitrophenyl phenylphosphorothioate and in the other compounds the following replace ethyl: MPNxn=methyl;PPNxn=propyl;BPNxn=butyl;APNxn=amyl.

b) All treatments decreased or did not affect cell number. Thus, increases in thiol production on a per cell basis are greater than indicated here.

c) The concentration values indicate that the same amount of ethanol used for each concentration of test agents was used as a control in this experiment.

d) These samples contained polymyxin B at 5µg/ml to inhibit the activity of any LPS present (Fig. 1). LPS was not detectable by the Limulus Assay in test agents at the dilutions used.

ACKNOWLEDGEMENTS

The invaluable assistance of Dr. Johnathan Kiel and his colleagues at the USAF School of Aerospace Medicine and of Dr. Rodney Darrah and Ms. Sue Espy at UES are appreciated.

REFERENCES

1. Pruett, S.B. and Kiel, J.L. (1988) Quantitative aspects of the feeder cell phenomenon: Mechnanistic implications. Biochem. Biophys. Res. Comm. 150:1037.

2. Matusmoto, T., Takeshige, K. and Minakami, S. (1979) Inhibition of phagocytic metabolic changes of leukocytes by an intracellular calcium antagonist 8-(N,N-diethylamino)-octyl-3,4,5-trimethyoxybenzoate. Biochem. Biophys. Res. Comm. 88:974.

3. Justement, L.B., Aldrich, W.A., Wenger, G.D., O'Dorisio, M.S. and Zwilling, B.S. (1986) Modulation of cyclic AMP-dependent protein kinase isozyme expression associated with activation of a macrophage-like cell line. J. Immunol. 136:270.

4. Drapier, J.-C. and Hibbs, J.B. (1988) Differentiation of murine macrophages to express nonspecific cytotoxicity for tumor cells results in L-arginine-dependent inhibition of mitochondrial iron-sulfur enzymes in the macrophage effector cells. J. Immunol. 140:2829.

5. Bellavite, A. (1988) The superoxide-forming enzymatic system of phagocytes. Free Radical Biol. Med. 4:225.

6. Ralph, P. and Nakoinz, I. (1977) Antibody-dependent killing of erythrocyte and tumor targets by macrophage-related cell lines: enhancement by PPD and LPS. J. Immunol. 118:950.

7. Adams, D.O. and Hamilton, T.A. (1987) Molecular transductional mechanisms by which Interferon gamma and other signals regulate macrophage development. Immunol. Rev. 97:5.

8. Ishii, T., Sugita, Y. and Bannai, S. (1987) Regulation of glutathione levels in mouse spleen lymphocytes by transport of cysteine. J. Cell. Physiol. 133:330.

9. Shimonkevitz, R., Kappler, J., Marrack, P., and Grey, H. (1983) Antigen recognition by H-2 restricted T cells I. Cell-free antigen processing. J. Exp. Med. 158:303.

10. Rodgers, K.E., Imamura, T. and Devens, B.H. (1987) Investigations into the mechanism of immunosuppression cuased by acute treatment with O,O,S-Trimethyl Phosphorothioate: Generation of suppressive macrophages from treated animals. Toxicol. Appl. Pharmacol. 88:270.

11. Pruett, S.B. and Chambers, J.E. (1988) Effects of paraoxon, p-nitrophenol, phenyl saligenin cyclic phosphate, and phenol on the rat interleukin 2 system. Toxicol. Lett. 40:11.

12. Cluff, C.W. and Ziegler, H.K. (1987) An early response to lipopolysaccharide is the elicitation of macrophages specialized for antigen degradation with negative regulatory effects on the induction of specific immune responses. Infect. Immun. 55:1346.

13. Lee, J.-C., Gibson, C.W. and Eisenstein, T.K. (1985) Macrophage-mediated mitogenic suppression induced in mice of the C3H lineage by a vaccine strain of Salmonella typhimurium Cell. Immunol. 91:75.

FINAL REPORT NUMBER 111
REPORT NOT AVAILABLE AT THIS TIME
Dr. Wesley Tanaka
760-7MG-043