REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0772

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November, 1994	Final

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1993 Summer Research Extension Program Final Reports, Volume 1B, Armstrong Laboratory

5. FUNDING NUMBERS

6. AUTHORS
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research and Development Labs, Culver City, CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Contract Number: F4962-90-C-0076

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

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

14. SUBJECT TERMS
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, UNIVERSITIES

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1993

SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 1B

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

November 1994

PREFACE

This volume is part of a five-volume set that summarizes the research of participants in the 1993 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 1B of 5, presents the final reports of SREP participants at Armstrong Laboratory.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a 35 page management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1A	Armstrong Laboratory (part one)
1B	Armstrong Laboratory (part two)
2	Phillips Laboratory
3	Rome Laboratory
4A	Wright Laboratory (part one)
4B	Wright Laboratory (part two)
5	Arnold Engineering Development Center Frank J. Seiler Research Laboratory Wilford Hall Medical Center

1993 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1A

1993 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1B

1993 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2

1993 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2
cont'd

1993 SREP FINAL REPORTS

Rome Laboratory

VOLUME 3

1993 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4A

1993 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B

1993 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B
cont'd

Report #	Report Title Author's University	Report Author
33	Detection of Internal Defects in Multilayered Plates By Lamb Wave Acoustic Micro Universtiy of Arizona, Tucson, AZ	Dr. Kundu Tribikram Civil Engineering and WL/ML Engineering
34	Wavelet Analysis of Ultrasonic Signals for Non-Destructive Evaluation of Composi University of Dayton, Dayton, OH	Dr. Theresa Tuthill Electrical Engineering WL/ML
35	Stochastic Modeling of MBE Growth of Compoud Semiconductors University of Nevada, Las Vegas, NV	Dr. Ramasubrama Venkatasubraman Electrical and Computer WL/ML Engineering
36	Performance Evaluation And Improvement of a Resonant DC Link Inverter With A Lim North Dakota State University, Fargo, ND	Dr. Subbaraya Yuvarajan Electrical Engineering WL/PO
37	Three Component LDV Measurements in a Swirl Combustor North Carolina State University, Raleigh, NC	Dr. Richard Gould Mechanical and Aerospace WL/PO Engineering

1993 SREP FINAL REPORTS

VOLUME 5

1993 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The Summer Research Extension Program (SREP) is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations working at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for 75 SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Appendix 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative charges institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally $18,000 for a $20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1993 SREP PROGRAM

SELECTION DATA: A total of 719 faculty members (SFRP Associates) and 286 graduate students (GSRP associates) applied to participate in the 1992 Summer Research Program. From these applicants 185 SFRPs and 121 GSRPs were selected. The education level of those selected was as follows:

1992 SRP Associates, by Degree			
SFRP		GSRP	
PHD	MS	MS	BS
179	6	52	69

Of the participants in the 1992 Summer Research Program 90 percent of SFRPs and 25 percent of GSRPs submitted proposals for the SREP. Ninety proposals from SFRPs and ten from GSRPs were selected for funding, which equates to a selection rate of 54% of the SFRP proposals and of 34% for GSRP proposals.

1993 SREP: Proposals Submitted vs. Proposals Selected			
	Summer 1992 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	185	167	90
GSRP	121	29	10
TOTAL	306	196	100

The funding was provided as follows:

Contractual slots funded by AFOSR	75
Laboratory funded	14
Additional funding from AFOSR	11
Total	100

Six HBCU/MI associates from the 1992 summer program submitted SREP proposals; six were selected (none were lab-funded; all were funded by additional AFOSR funds).

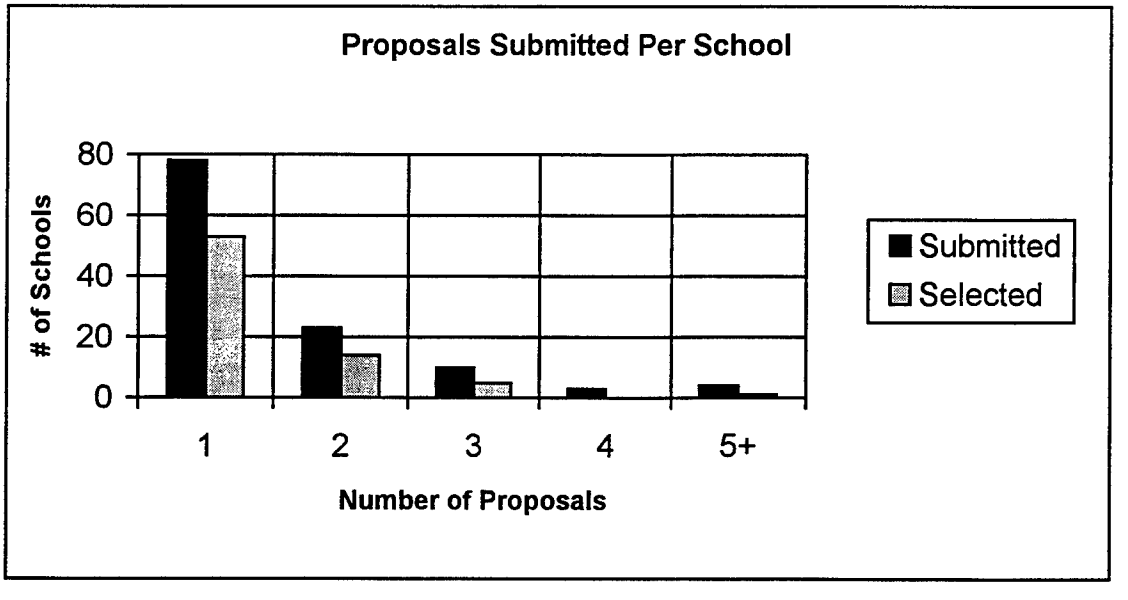
Proposals Submitted and Selected, by Laboratory		
	Applied	Selected
Air Force Civil Engineering Laboratory	9	4
Armstrong Laboratory	41	19
Arnold Engineering Development Center	12	4
Frank J. Seiler Research Laboratory	6	3
Phillips Laboratory	33	19
Rome Laboratory	31	13
Wilford Hall Medical Center	2	1
Wright Laboratory	62	37
TOTAL	196	100

Note: Phillips Laboratory funded 3 SREPs; Wright Laboratory funded 11; and AFOSR funded 11 beyond its contractual 75.

The 306 1992 Summer Research Program participants represented 135 institutions.

Institutions Represented on the 1992 SRP and 1993 SREP		
Number of schools represented in the Summer 92 Program	Number of schools represented in submitted proposals	Number of schools represented in Funded Proposals
135	118	73

Forty schools had more than one participant submitting proposals.



The selection rate for the 78 schools submitting 1 proposal (68%) was better than those submitting 2 proposals (61%), 3 proposals (50%), 4 proposals (0%) or 5+ proposals (25%). The 4 schools that submitted 5+ proposals accounted for 30 (15%) of the 196 proposals submitted.

Of the 196 proposals submitted, 159 offered institution cost sharing. Of the funded proposals which offered cost sharing, the minimum cost share was $1000.00, the maximum was $68,000.00 with an average cost share of $12,016.00.

Proposals and Institution Cost Sharing		
	Proposals Submitted	Proposals Funded
With cost sharing	159	82
Without cost sharing	37	18
Total	196	100

The SREP participants were residents of 41 different states. Number of states represented at each laboratory were:

States Represented, by Proposals Submitted/Selected per Laboratory		
	Proposals Submitted	Proposals Funded
Air Force Civil Engineering Laboratory	8	4
Armstrong Laboratory	21	13
Arnold Engineering Development Center	5	2
Frank J. Seiler Research Laboratory	5	3
Phillips Laboratory	16	14
Rome Laboratory	14	7
Wilford Hall Medical Center	2	1
Wright Laboratory	24	20

Eleven of the 1993 SREP Principal Investigators also participated in the 1992 SREP.

ADMINISTRATIVE EVALUATION: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions provided to them by RDL. Ninety seven final reports and two interim reports have been received and are included in this report. The subcontracts were funded by $1,991,623.00 of Air Force money. Institution cost sharing totaled $985,353.00.

TECHNICAL EVALUATION: The form used for the technical evaluation is provided as Appendix 2. ninety-two evaluation reports were received. Participants by laboratory versus evaluations submitted is shown below:

	Participants	Evaluations	Percent
Air Force Civil Engineering Laboratory	*	*	*
Armstrong Laboratory	23[1]	20	95.2
Arnold Engineering Development Center	4	4	100
Frank J. Seiler Research Laboratory	3	3	100
Phillips Laboratory	19[2]	18	100
Rome Laboratory	13	13	100
Wilford Hall Medical Center	1	1	100
Wright Laboratory	37	34	91.9
Total	100[3]	93	95.9

*AFCEL was combined with Wright Laboratory's Flight Dynamics Directorate and Armstrong Laboratories Environics Directorate in 1993. All four of AFCEL's SREP awards went to Armstrong Laboratories Environics Directorate, and their reports are included with Armstrong Lab.

Notes:

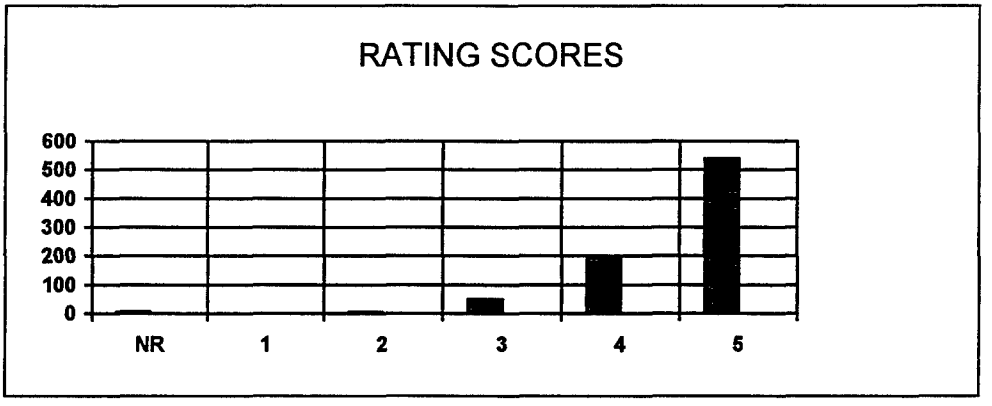
1: Research on two of the final reports was incomplete as of press time so there aren't any technical evaluations on them to process, yet. Percent complete is based upon 20/21=95.2%

2: One technical evaluation was not completed because one of the final reports was incomplete as of press time. Percent complete is based upon 18/18=100%

3: See notes 1 and 2 above. Percent complete is based upon 93/97=95.9%

The number of evaluations submitted for the 1993 SREP (95.9%) shows a marked improvement over the 1992 SREP submittals (65%).

PROGRAM EVALUATION: Each laboratory focal point evaluated ten areas (see Appendix 2) with a rating from one (lowest) to five (highest). The distribution of ratings was as follows:

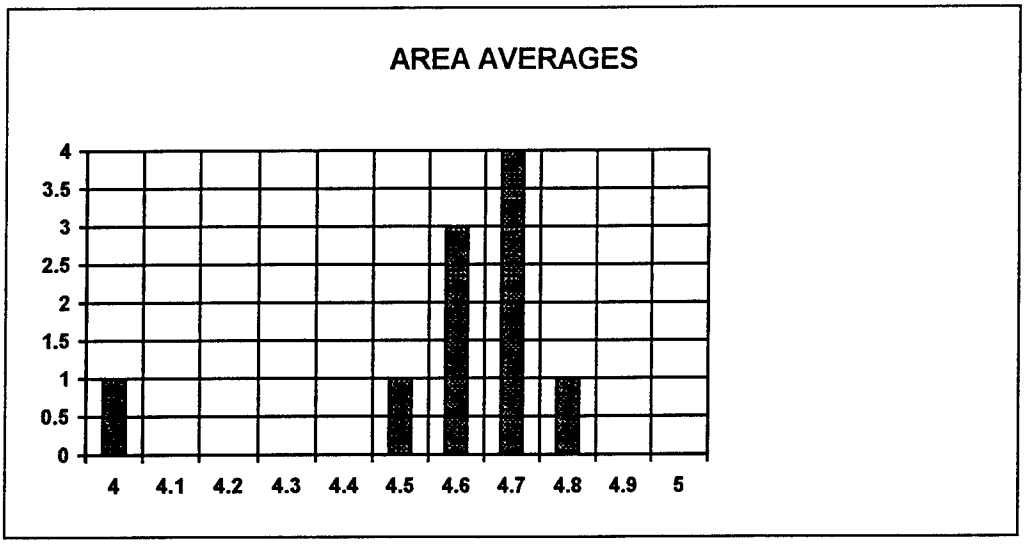


Rating	Not Rated	1	2	3	4	5
# Responses	7	1	7	62 (6%)	226 (25%)	617 (67%)

The 8 low ratings (one 1 and seven 2's) were for question 5 (one 2) "The USAF should continue to pursue the research in this SREP report" and question 10 (one 1 and six 2's) "The one-year period for complete SREP research is about right", in addition over 30% of the threes (20 of 62) were for question ten. The average rating by question was:

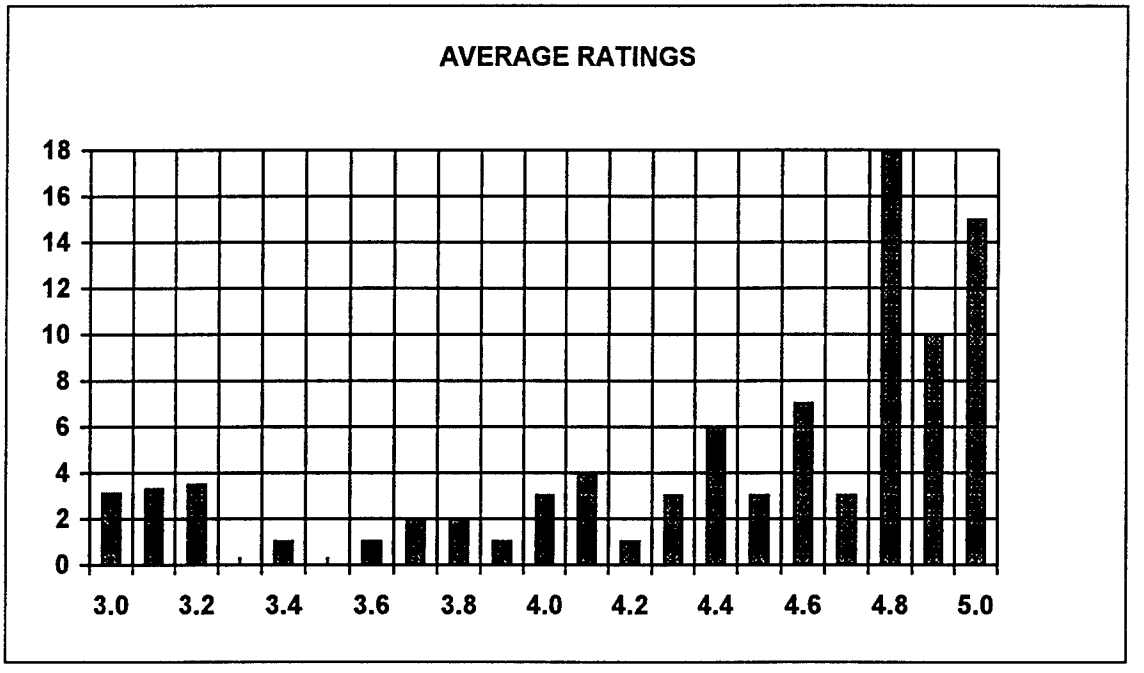
Question	1	2	3	4	5	6	7	8	9	10
Average	4.6	4.6	4.7	4.7	4.6	4.7	4.8	4.5	4.6	4.0

The distribution of the averages was:



Area 10 "the one-year period for complete SREP research is about right" had the lowest average rating (4.1). The overall average across all factors was 4.6 with a small sample standard deviation of 0.2. The average rating for area 10 (4.1) is approximately three sigma lower than the overall average (4.6) indicating that a significant number of the evaluators feel that a period of other than one year should be available for complete SREP research.

The average ratings ranged from 3.4 to 5.0. The overall average for those reports that were evaluated was 4.6. Since the distribution of the ratings is not a normal distribution the average of 4.6 is misleading. In fact over half of the reports received an average rating of 4.8 or higher. The distribution of the average report ratings is as shown:



It is clear from the high ratings that the laboratories place a high value on AFOSR's Summer Research Extension Programs.

3.0 SUBCONTRACTS SUMMARY

Table 1 provides a summary of the SREP subcontracts. The individual reports are published in volumes as shown:

Laboratory	Volume
Air Force Civil Engineering Laboratory	*
Armstrong Laboratory	1
Arnold Engineering Development Center	5
Frank J. Seiler Research Laboratory	5
Phillips Laboratory	2
Rome Laboratory	3
Wilford Hall Medical Center	5
Wright Laboratory	4A, 4B

*AFCEL was combined with Wright Laboratory's Flight Dynamics Directorate and Armstrong Laboratories Environics Directorate in 1993. All four of AFCEL's SREP awards went to Armstrong Laboratories Environics Directorate, and their reports are included with Armstrong Lab.

1993 SREP SUB-CONTRACT DATA

TABLE 1: SUBCONTRACTS SUMMARY

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Abbott , Ben Electrical Engineering Vanderbilt University, Nashville, TN	M.S.	AEDC/	01/01/93	12/31/93	$19619.00 $0.00
Alberts , Thomas Mechanical Engineering Old Dominion University, Norfolk, VA	PhD	FJSRL/	01/01/93	04/15/94	$20000.00 $8000.00
Avula , Xavier Mechanical & Aerospace Engineering University of Missouri, Rolla, MO	PhD	AL/AO	01/01/93	04/15/94	$20000.00 $1836.00
Balakrishan , S. Mechanical & Aerospace Engineering University of Missouri, Rolla, MO	PhD	WL/MN	12/01/92	12/14/93	$20000.00 $3996.00
Baumgarten , Joseph Mechanical Engineering Iowa State University, Ames, IA	PhD	PL/VT	01/01/93	04/01/94	$19916.00 $9083.00
Bayard , Jean-Pierre Electrical & Electronic Engineering California State University, Sacramento, CA	PhD	RL/ER	01/01/93	12/31/93	$20000.00 $7423.00
Bellem , Raymond Electrical & Computer Engineering University of Arizona, Tucson, AZ	PhD	PL/VT	01/01/93	02/28/94	$19956.00 $0.00
Biegl , Csaba Electrical Engineering Vanderbilt University, Nashville, TN	PhD	AEDC/	01/01/93	12/31/93	$19999.00 $0.00
Biggs , Albert Electrical Engineering University of Alabama, Huntsville, AL	PhD	PL/WS	01/01/93	12/31/93	$19975.00 $0.00
Burkey , Theodore Chemistry Memphis State University, Memphis, TN	PhD	WL/MN	01/01/93	12/31/93	$20000.00 $18648.00
Burkholder , Robert Electrical Engineering Ohio State University, Columbus, OH	PhD	WL/AA	01/01/93	12/31/93	$20000.00 $6727.00
Callens, Jr. , Eugene Mechanical and Industrial Engineer Louisiana Tech University, Ruston, LA	PhD	WL/MN	01/01/93	12/31/93	$20000.00 $5700.00
Chapman , Gary Mechnical Engineering University of California, Berkeley, CA	PhD	WL/MN	01/01/93	12/31/94	$20000.00 $0.00
Chen , Chien-In Electrical Engineering Wright State University, Dayton, OH	PhD	WL/EL	01/01/93	12/31/93	$20000.00 $32065.00
Chen , Jer-sen Computer Science & Engineering Wright State University, Dayton, OH	PhD	AL/CF	01/01/93	12/31/93	$20000.00 $31763.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Chen , Pinyuen Mathematics Syracuse University, Syracuse, NY	PhD	RL/IR	01/01/93	12/31/93	$20000.00 $0.00
Chow , Joe Industrial and Systems Engineering Florida International University, Miami, FL	PhD	WL/FI	01/01/93	01/14/94	$20000.00 $2500.00
Christensen , Thomas Physics University of Colorado, Colorado Springs, CO	PhD	FJSRL/	01/01/93	12/31/93	$20000.00 $5390.00
Collicott , Steven Aeronautics and Astronautical Engineering Purdue University, West Lafayette, IN	PhD	WL/FI	01/01/93	12/31/93	$20000.00 $13307.00
Cooke , Nancy Psychology New Mexico State University, Las Cruces, NM	PhD	AL/HR	01/01/93	12/31/93	$20000.00 $6178.00
Daley , Michael Electrical Engineering Memphis State, Memphis, TN	PhD	WHMC/	01/01/93	12/31/93	$20000.00 $18260.00
Davidson , Jennifer Electrical Engineering Iowa State University, Ames, IA	PhD	WL/MN	01/01/93	02/28/94	$19999.00 $0.00
Deivanayagam , Subramaniam Industrial Engineering Tennessee Technological University, Cookeville, TN	PhD	AL/HR	02/01/93	12/31/93	$20000.00 $12491.00
Elliott , David Engineering Arkansas Technology University, Russellville, AR	PhD	PL/RK	10/01/92	08/15/93	$20000.00 $50271.00
Erdman , Paul Physics and Astronomy University of Iowa, Iowa City, IA	M.S.	PL/RK	01/01/93	12/31/93	$20000.00 $26408.00
Ervin , Jamie Mechanical and Aerospace Engineering University of Dayton, Dayton, OH	PhD	WL/ML	01/01/93	12/31/93	$18632.00 $3000.00
Erwin , Daniel Aerospace Engineering University of Southern California, Los Angeles, CA	PhD	PL/RK	01/01/93	12/31/93	$19962.00 $12696.00
Ewert , Dan Electrical Engineering North Dakota State University, Fargo, ND	PhD	AL/AO	01/01/93	12/31/93	$20000.00 $2100.00
Farmer , Barry Materials Science and Engineering University of Virginia, Charlottesville, VA	PhD	WL/ML	01/01/93	02/28/94	$20000.00 $2000.00
Foo , Simon Electrical Engineering Florida State University, Tallahassee, FL	PhD	WL/MN	01/01/93	12/31/93	$19977.00 $0.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Friedman , Jeffrey Physics University of Puerto Rico, Mayaguez, PR	PhD	PL/GP	01/01/93	12/31/93	$20000.00 $10233.00
Gerstman , Bernard Physics Florida International University, Miami, FL	PhD	AL/OE	01/01/93	04/30/94	$19947.00 $2443.00
Gillman , Ann Chemistry Eastern Kentucky University, Richmond, KY	M.S.	WL/PO	01/01/93	12/31/93	$20000.00 $15618.00
Gimmestad , Gary Research Institute Georgia Institute of Technology, Atlanta, GA	PhD	PL/LI	01/01/93	12/31/93	$20000.00 $0.00
Gould , Richard Mechanical and Aerospace Engineering North Carolina State University, Raleigh, NC	PhD	WL/PO	01/01/93	12/31/93	$20000.00 $8004.00
Graham , Gary Mechanical Engineering Ohio University, Athens, OH	PhD	WL/FI	01/01/93	12/31/93	$20000.00 $5497.00
Gramoll , Kurt Aerospace Engineering Georgia Institute of Technology, Atlanta, GA	PhD	AEDC/	01/01/93	12/31/93	$19707.00 $14552.00
Graul , Susan Chemistry Carnegie Mellon University, Pittsburgh, PA	PhD	PL/WS	01/01/93	03/31/94	$20000.00 $0.00
Griffin , Steven Engineering University of Texas, San Antonio, TX	M.S.	PL/VT	01/01/93	12/31/93	$20000.00 $0.00
Grubbs , Elmer Electrical Engineering New Mexico Highlands University, Las Vegas, NM	PhD	WL/AA	01/01/93	12/31/93	$20000.00 $6747.00
Gupta , Pushpa Mathematics University of Maine, Orono, ME	PhD	AL/AO	01/01/93	12/31/93	$20000.00 $1472.00
Hall , Ian Materials Science University of Delaware, Newark, DE	PhD	WL/ML	01/01/93	12/31/93	$20000.00 $9580.00
Hedman , Paul Chemical Engineering Brigham Young University, Provo, UT	PhD	WL/PO	01/01/93	12/31/93	$19999.00 $7755.00
Henry , Robert Electrical & Computer Engineering University of Southwestern Louisiana, Lafayette, LA	PhD	RL/C3	12/01/92	05/31/93	$19883.00 $11404.00
Henson , James Electrical Engineering University of Nevada, Reno, NV	PhD	PL/WS	01/01/93	12/31/93	$19913.00 $9338.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Hoe , Benjamin Electrical Engineering Polytechnic University, Brooklyn, NY	M.S.	RL/C3	09/01/92	05/31/93	$19988.00 $7150.00
Hughes , Rod Psychology Bowling Green State University, Bowling Green, OH	M.S.	AL/CF	01/01/93	04/15/94	$20000.00 $20846.00
Hui , David Mechanical Engineering University of New Orleans, New Orleans, LA	PhD	WL/FI	01/01/93	12/31/93	$20000.00 $0.00
Humi , Mayer Mathematics Worcester Polytechnic Institut, Worcester, MA	PhD	PL/LI	01/01/93	12/31/93	$20000.00 $5000.00
Innocenti , Mario Aerospace Engineering Auburn University, Auburn, AL	PhD	WL/MN	01/01/93	02/28/94	$20000.00 $12536.00
Jean , Jack Computer Science & Engineering Wright State University, Dayton, OH	PhD	WL/AA	01/01/93	12/31/93	$20000.00 $34036.00
Jouny , Ismail Electrical Engineering Lafayette College, Easton, PA	PhD	WL/AA	01/01/93	12/31/93	$19381.00 $4500.00
Kaikhah , Khosrow Computer Science Southwest Texas State College, San Marcos, TX	PhD	RL/IR	01/01/93	12/31/93	$20000.00 $0.00
Kaw , Autar Mechanical Engineering University of South Florida, Tampa, FL	PhD	WL/ML	01/01/93	12/31/93	$20000.00 $22556.00
Kheyfets , Arkady Mathematics North Carolina State University, Raleigh, NC	PhD	PL/LI	01/01/93	12/31/93	$20000.00 $2500.00
Kitchart , Mark Mechanical Engineering North Carolina A & T State University, Greensboro, NC	M.S.	AL/EQ	01/01/93	12/31/93	$20000.00 $0.00
Koblasz , Arthur Civil Engineering Georgia Institute of Technology, Atlanta, GA	PhD	AL/AO	01/01/93	12/31/93	$19826.00 $0.00
Koivo , A. Electrical Engineering Purdue University, West Lafayette, IN	PhD	AL/CF	01/01/93	06/30/94	$20000.00 $0.00
Kundich , Robert Biomedical Engineering University of Tennessee, Memphis, TN	PhD	AL/CF	01/01/93	12/31/94	$20000.00 $23045.00
Kuo , Spencer Electrical Engineering Polytechnic University, Farmingdale, NY	PhD	PL/GP	01/01/93	04/30/94	$20000.00 $9731.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Liou , Juin Electrical and Computer Engineering Universtiy of Central Florida, Orlando, FL	PhD	WL/EL	01/01/93	12/31/93	$20000.00 $9073.00
Liou , Shy-Shenq Engineering San Francisco State Univesity, San Francisco, CA	PhD	WL/PO	01/01/93	12/31/93	$20000.00 $13387.00
Manoranjan , Valipuram Pure and Applied Mathematics Washington State University, Pullman, WA	PhD	AL/EQ	01/01/93	12/31/93	$19956.00 $10041.00
Marks , Dallas Electrical and Computer Engineering University of Cincinnati M.L., Cincinnati, OH	M.S.	WL/AA	10/01/92	06/30/93	$20000.00 $4731.00
Monsay , Evelyn Physics Le Moyne College, Syracuse, NY	PhD	RL/OC	01/01/93	12/31/93	$19634.00 $1510.00
Moor , William Industrial & Management Engineering Arizona State University, Tempe, AZ	PhD	AL/HR	01/01/93	12/31/93	$20000.00 $4833.00
Moore , Carlyle Physics Morehouse College, Atlanta, GA	PhD	AEDC/	01/01/93	12/31/93	$20000.00 $4880.00
Mulligan , B. Psychology University of Georgia Research, Athens, GA	PhD	AL/OE	01/01/93	04/15/94	$19998.00 $13936.00
Murphy , Richard Physics University of Missouri, Kansas City, MO	PhD	PL/WS	01/01/93	12/31/93	$20000.00 $13022.00
Nilan , Michael Information Studies Syracuse University, Syracuse, NY	PhD	RL/C3	01/01/93	12/31/93	$19998.00 $13016.00
Parrish , Allen Computer Science University of Alabama, Tuscaloosa, AL	PhD	RL/C3	01/01/93	12/31/93	$19919.00 $20599.00
Piersma , Bernard Chemistry Houghton College, Houghton, NY	PhD	FJSRL/	01/01/93	12/31/93	$20000.00 $4000.00
Potasek , Mary Applied Physics Columbia University, New York, NY	PhD	WL/ML	12/01/93	11/30/93	$20000.00 $7806.00
Qazi , Salahuddin Optical Communications SUNY/Institute of Technology, Utica, NY	PhD	RL/OC	01/01/93	12/31/93	$20000.00 $68000.00
Reardon , Kenneth Agricultural and Chemical Engineering Colorado State University, Fort Collins, CO	PhD	AL/EQ	01/01/93	01/31/94	$19996.00 $12561.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Reynolds , David Biomedical & Human Factors Wright State University, Dayton, OH	PhD	AL/CF	01/01/93	06/30/94	$20000.00 $14063.00
Robinson , Donald Chemistry Xavier University of Louisiana, New Orleans, LA	PhD	AL/OE	01/01/93	06/30/94	$20000.00 $12935.00
Rodriguez , Armando Electrical Engineering Arizona State University, Tempe, AZ	PhD	WL/MN	01/01/93	12/31/93	$20000.00 $0.00
Roe , Larry Mechanical Engineering Virginia Polytechnic Inst & State Coll., Blacksburg, VA	PhD	WL/PO	01/01/93	12/31/93	$20000.00 $11421.00
Romeu , Jorge Assistant Prof. of Mathematics SUNY College at Cortland, Cortland, NY	PhD	RL/OC	01/01/93	12/31/93	$19997.00 $7129.00
Roppel , Thaddeus Electrical Engineering Auburn University, Auburn, AL	PhD	WL/MN	01/01/93	12/31/93	$20000.00 $21133.00
Roznowski , Mary Psychology Ohio State University, Columbus, OH	PhD	AL/HR	01/01/93	03/31/94	$19953.00 $6086.00
Rudzinski , Walter Chemistry Southwest Texas State University, San Marcos, TX	PhD	AL/OE	01/01/93	12/31/93	$20000.00 $10120.00
Sargent , Robert Engineering and Computer Science Syracuse University, Syracuse, NY	PhD	RL/XP	01/01/93	12/31/93	$20000.00 $11931.00
Schonberg , William Civil and Environmental Engineering University of Alabama, Huntsville, AL	PhD	WL/MN	01/01/93	12/31/93	$19991.00 $5083.00
Shaw , Arnab Electrical Engineering Wright State University, Dayton, OH	PhD	WL/AA	01/01/93	12/31/93	$20000.00 $4766.00
Shively , Jon Engineering & Computer Science California State University, Northridge, CA	PhD	PL/VT	01/01/93	12/31/93	$20000.00 $9782.00
Slater , Robert Mechanical & Industrial Engineering University of Cincinnati, Cincinnati, OH	M.S.	WL/FI	01/01/93	12/31/93	$20000.00 $8257.00
Stenzel , Johanna Arts & Sciences University of Houston, Victoria, TX	PhD	PL/LI	01/01/93	12/31/93	$20000.00 $9056.00
Tan , Arjun Physics Alabama A & M University, Normal, AL	PhD	PL/WS	01/01/93	12/31/93	$20000.00 $1000.00

1993 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount Univ. Cost Share
Tetrick , Lois Industrial Relations Prog Wayne State University, Detroit, MI	PhD	AL/HR	01/01/93	12/31/93	$20000.00 $17872.00
Tew , Jeffery Industrial & Systems Engineering Virginia Polytechnic Institute, Blacksburg, VA	PhD	RL/IR	05/31/93	12/31/93	$16489.00 $4546.00
Tribikram , Kundu Civil Engineering and Engineering Universtiy of Arizona, Tucson, AZ	PhD	WL/ML	01/01/93	12/31/93	$20000.00 $9685.00
Tuthill , Theresa Electrical Engineering University of Dayton, Dayton, OH	PhD	WL/ML	01/01/93	12/31/93	$20000.00 $24002.00
Venkatasubraman , Ramasubrama Electrical and Computer Engineering University of Nevada, Las Vegas, NV	PhD	WL/ML	01/01/93	12/31/93	$20000.00 $18776.00
Wang , Xingwu Electrical Engineering Alfred University, Alfred, NY	PhD	AL/EQ	01/01/93	12/31/93	$20000.00 $10000.00
Whitefield , Philip Physics University of Missouri, Rolla, MO	PhD	PL/LI	01/01/93	03/01/94	$20000.00 $11040.00
Wightman , Colin Electrical Engineering New Mexico Institute of Mining, Socorro, NM	PhD	RL/IR	01/01/93	12/31/93	$20000.00 $1850.00
Womack , Michael Natural Science and Mathematics Macon College, Macon, GA	PhD	AL/OE	01/01/93	06/30/94	$19028.00 $6066.00
Yuvarajan , Subbaraya Electrical Engineering North Dakota State University, Fargo, ND	PhD	WL/PO	01/01/93	12/31/93	$19985.00 $22974.00

APPENDIX 1:

SAMPLE SREP SUBCONTRACT

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1993 SUMMER RESEARCH EXTENSION PROGRAM SUBCONTRACT 93-133

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

San Francisco State University
University Comptroller
San Francisco, CA 94132

REFERENCE: Summer Research Extension Program Proposal 93-133
Start Date: 01/01/93 End Date: 12/31/93
Proposal Amount: $20,000.00

(1) PRINCIPAL INVESTIGATOR: Dr. Shy Shenq P. Liou
Engineering
San Francisco State University
San Francisco, CA 94132

(2) UNITED STATES AFOSR CONTRACT NUMBER: F49620-90-C-09076

(3) CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA): 12.800
PROJECT TITLE: AIR FORCE DEFENSE RESEARCH SOURCES PROGRAM

(4) ATTACHMENTS 1 AND 2: SREP REPORT INSTRUCTIONS

*** SIGN SREP SUBCONTRACT AND RETURN TO RDL***

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-90-C-0076) to the United States Air Force to administer the Summer Research Programs (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Summer Research Extension Program (SREP). Approximately 75 SREP proposals annually will be selected by the Air Force for funding of up to $20,000; shared funding by the academic institution is encouraged. SREP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents such an agreement between RDL and the institution designated in Section 5 below.

2. RDL PAYMENTS: RDL will provide the following payments to SREP institutions:

- 90 percent of the negotiated SREP dollar amount at the start of the SREP Research period.
- the remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the SREP research.

3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

a. Assure that the research performed and the resources utilized adhere to those defined in the SREP proposal.
b. Provide the level and amounts of institutional support specified in the RIP proposal.
c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.

d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract, but no later than December 31, 1993. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1993.

e. Assure that the final report is submitted in accordance with Attachment 1.

f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade association meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.

g. Notify RDL of inventions or patents claimed as the result of this research as specified in Attachment 1.

h. RDL is required by the prime contract to flow down patent rights and technical data requirements in this subcontract. Attachment 2 to this subcontract contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

RDL Summer Research Program Office
5800 Uplander Way
Culver City, CA 90230-6608

5. By their signatures below, the parties agree to the provisions of this subcontract.

A.S. Sopher

Abe S. Sopher
RDL Contracts Manager

Signature of Institution Contracting Official

Typed/Printed Name

Date

Title

Institution

(Date/Phone)

ATTACHMENT 2
CONTRACT CLAUSES

This contract incorporates by reference the following clauses of the Federal Acquisition Regulations (FAR), with the same force and effect as if they were given in full text. Upon request, the Contracting Officer or RDL will make their full text available (FAR 52.252-2).

FAR CLAUSES	TITLE AND DATE
52.202-1	DEFINITIONS (SEP 1991)
52.203-1	OFFICIALS NOT TO BENEFIT (APR 1984)
52.203-3	GRATUITIES (APR 1984)
52.203-5	COVENANT AGAINST CONTINGENT FEES (APR 1984)
52.304-6	RESTRICTIONS ON SUBCONTRACTOR SALES TO THE GOVERNMENT (JUL 1985)
52.203-7	ANTI-KICKBACK PROCEDURES (OCT 1988)
52.203-12	LIMITATION ON PAYMENTS TO INFLUENCE CERTAIN FEDERAL TRANSACTIONS (JAN 1990)
52.204-2	SECURITY REQUIREMENTS (APR 1984)
52.209-6	PROTECTING THE GOVERNMENT'S INTEREST WHEN SUBCONTRACTING WITH CONTRACTORS DEBARRED, SUSPENDED, OR PROPOSED FOR DEBARMENT (NOV 1992)
52.212-8	DEFENSE PRIORITY AND ALLOCATION REQUIREMENTS (SEP 1990)
52.215-1	EXAMINATION OF RECORDS BY COMPTROLLER GENERAL (APR 1984)
52.215-2	AUDIT - NEGOTIATION (DEC 1989)
52.222-26	EQUAL OPPORTUNITY (APR 1984)
52.222-28	EQUAL OPPORTUNITY PREAWARD CLEARANCE OF SUBCONTRACTS (APR 1984)

52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS (APR 1984)
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS (APR 1984)
52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERAN AND VETERANS OF THE VIETNAM ERA (JAN 1988)
52.223-2	CLEAN AIR AND WATER (APR 1984)
52.232-6	DRUG-FREE WORKPLACE (JUL 1990)
52.224-1	PRIVACY ACT NOTIFICATION (APR 1984)
52.224-2	PRIVACY ACT (APR 1984)
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS (MAY 1989)
52.227-1	AUTHORIZATION AND CONSENT (APR 1984)
52.227-2	NOTICE AND ASSISTANCE REGARDING PATENT AND COPYRIGHT INFRINGEMENT (APR 1984)
52.227-10	FILING OF PATENT APPLICATIONS - CLASSIFIED SUBJECT MATTER (APR 1984)
52.227-11	PATENT RIGHTS - RETENTION BY THE CONTRACTOR (SHORT FORM) (JUN 1989)
52.228-6	INSURANCE - IMMUNITY FROM TORT LIABILITY (APR 1984)
52.228-7	INSURANCE - LIABILITY TO THIRD PERSONS (APR 1984)
52.230-5	DISCLOSURE AND CONSISTENCY OF COST ACCOUNTING PRACTICES (AUG 1992)
52.232-23	ASSIGNMENT OF CLAIMS (JAN 1986)
52.237-3	CONTINUITY OF SERVICES (JAN 1991)

52.246-25	LIMITATION OF LIABILITY - SERVICES (APR 1984)
52.249-6	TERMINATION (COST-REIMBURSEMENT) (MAY 1986)
52.249-14	EXCUSABLE DELAYS (APR 1984)
52.251-1	GOVERNMENT SUPPLY SOURCES (APR 1984)

APPENDIX 2:

SAMPLE TECHNICAL EVALUATION FORM

1993 SUMMER RESEARCH EXTENSION PROGRAM

RIP NO.: 93-0092
RIP ASSOCIATE: Dr. Gary T. Chapman

Provided are several evaluation statements followed by ratings of (1) through (5). A rating of (1) is the lowest and (5) is the highest. Circle the rating level number you best feel rates the statement. Document additional comments on the back of this evaluation form.

Mail or fax the completed form to :

RDL
Attn: 1993 SREP TECH EVALS
5800 Uplander Way
Culver City, CA 90230-6608
(FAX: 310 216-5940)

1.	This SREP report has a high level of technical merit.	1	2	3	4	5
2.	The SREP program is important to accomplishing the labs's mission	1	2	3	4	5
3.	This SREP report accomplished what the associate's proposal promised.	1	2	3	4	5
4.	This SREP report addresses area(s) important to the USAF	1	2	3	4	5
5.	The USAF should continue to pursue the research in this SREP report	1	2	3	4	5
6.	The USAF should maintain research relationships with this SREP associate	1	2	3	4	5
7.	The money spent on this SREP effort was well worth it	1	2	3	4	5
8.	This SREP report is well organized and well written	1	2	3	4	5
9.	I'll be eager to be a focal point for summer and SREP associates in the future.	1	2	3	4	5
10.	The one-year period for complete SREP research is about right	1	2	3	4	5

******USE THE BACK OF THIS FORM FOR ADDITIONAL COMMENTS******

LAB FOCAL POINT'S NAME (PRINT): _____

OFFICE SYMBOL: _____ PHONE: _____

THREE-DIMENSIONAL CALCULATION OF BLOOD FLOW IN A THICK-WALLED VESSEL USING THE FINITE ELEMENT METHOD: APPLICATION TO THE MODELING OF THE CARDIOVASCULAR SYSTEM IN ALTERED GRAVITY

Xavier J. R. Avula and Guruprasad Mysore
Department of Mechanical and Aerospace Engineering
and Engineering Mechanics
University of Missouri-Rolla
Rolla, MO 65401

Abstract

In this report, the feasibility of utilizing a three-dimensional finite element method for the calculation of blood flow in a thick-walled blood vessel is demonstrated. Using the linearized governing equations of motion, the deformation of a short segment of a blood vessel with variable wall thickness for one cycle of a heartbeat is calculated. The increased volume of the vessel segment due to its radial deformation allows the calculation of blcod pooling in the human arterio-venous system if the computations are extended to the entire human vascular tree. The gravitational effects can be easily incorporated in the developed method.

THREE-DIMENSIONAL CALCULATION OF BLOOD FLOW IN A THICK-WALLED VESSEL USING THE FINITE ELEMENT METHOD: APPLICATION TO THE MODELING OF THE CARDIOVASCULAR SYSTEM IN ALTERED GRAVITY

Xavier J. R. Avula and Guruprasad Mysore

INTRODUCTION

The function of humans in high performance aircraft can be impaired due to sustained accelerations on the cardiovascular system. The impaired performance can be attributed to inherent physiological limitations or subclinical cardiovascular disease exacerbated by high accelerations. To devise the protocol of corrective actions, mathematical modeling of the cardiovascular system can be useful in identifying the necessary parameters without indulging in extreme human experimentation. The mathematical modeling of the heart has been widely investigated. However, most investigators have addressed, as separate issues, either the deformation of the myocardium or the pattern of the blood flow. Addressing these issues simultaneously will be most beneficial for understanding the cardiovascular system in altered gravity. Since the heart and the blood vessels are topologically equivalent to each other, an investigation of blood flow through a thick-walled elastic tube is deemed reasonable. Over the years, there have been many theoretical and computational studies on flows in tubes depending upon the fluid properties (compressible or incompressible, Newtonian or non-Newtonian, etc.),and the tube properties such as its geometry and the degree of deformability. Because the applications are aimed at a realistic cardiovascular system which has complicated non-symmetric geometry and nonlinear material properties, a computational approach based on the finite element method (FEM) is followed.

In this report, the finite element approach for determining the three-dimensional blood flow in a thick-walled vessel for application in the modeling of the human cardiovascular system subjected to altered gravity is presented. The impairment of the cardiovascular function in abnormal gravitational fields, the dependence of the life-sustaining functions on the integrity of the cardiovascular system, and the prohibitive cost and risk involved in human experimentation in altered gravity provide the motivation for understanding the cardiovascular function through mathematical modeling and the associated computational approach [1,2,3,4]. This approachtakes advantage of available high performance computers, highly versatile three-dimensional finite element method, and realistic mechanical properties of the cardiovascular system.

PRESENT STATE OF KNOWLEDGE

There is no dearth of mathematical models of the cardiovascular system in the scientific literature. Womersley [5]

and Noordergraaf [6] presented mathematical analyses of some aspects of the cardiovascular system by using lumped parameter models. Taylor [7], and Kenner [8], and Attinger et al. [9] used distributive parameter models to analyze pressure-flow relationships in arteries and veins. It is generally believed that a lumped parameter model is superior to the distributive one in the evaluation of the overall cardiovascular system performance. Beneken and DeWit [10] characterized a large analog model of the entire human circulatory system in the form of approximately 40 equations. Rigorous mathematical formulations and extensive physiological information, including the factors affecting the contractility of the myocardium were introduced into the model. It has been demonstrated that the model provides quick solutions to parametric sensitivity tests. McLeod [11] proposed a physiological benchmark experiment (PHYSBE) to economize on programming efforts and to establish bases for comparison of various types of computer models of circulation. Sagawa [12, 13] described the overall circulatory regulation and the influence of the mechanical properties of the cardiovascular system on the control of circulation.

Complete models of the cardiovascular system under acceleration stress have been attempted by a few investigators. Avula and Oestreicher [14] presented a mechanical model consisting of linearized Navier-Stokes and finite elasticity equations to predict blood pooling under acceleration stress assuming a spherical model for the heart and axisymmetric thin tube model for blood vessels. Moore and Jaron [15], and Jaron et al. [16] have developed a non-linear multi-element model of the cardiovascular system which can calculate blood pressures and flows at any point in the cardiovascular system. It includes the effects of forces caused by acceleration as well as effects of several G-protection modes. The problem with multi-element models is that the capacitive and resistive elements introduce more uncertainties in the system model as opposed to direct mechanical models.

On the other end of the gravitational spectrum, with the advent of the space age, the effects of microgravity on the cardiovascular system has also become an important issue. Headaches and space motion sickness are believed to be caused by increased intracranial pressure as a result of exposure to microgravity leading to fluid shift. Because of the limited compliance of the upper body, one has to deal with the lower body to develop anti-microgravity devices. Many of these devices are still in the conceptual stage. Hargens et al. [17] have investigated the effects of microgravity on circulatory functions using experimental methods. However, mathematical models based on mechanical properties are desirable in view of costs and risks involved.

The stresses and deformations in the myocardium and the arterial wall were of concern to several investigators. Mirsky [18] calculated the shear and bending stresses in the ventricular wall using a thick-walled model of the left ventricle. Demiray [19], and Demiray and Vito [20] calculated the stresses in the arterial and

ventricular walls using the theory of finite deformations and experimentally determined strain energy function for soft biological tissue. Pao [21] et al. performed a finite element analysis of the left ventricle using axisymmetry of a spherical shell. McCulloch and Omens [22] analyzed the transmural distributions in canine ventricular myocardium using three-dimensional nonlinear finite element interpolating functions for a thick-walled cylinder model. Huyghe et al. [23] presented a porous medium finite element model of the passive left ventricle. However, the assumed axisymmetry is not realistic, and the problem addresses deformation in the duration of passive filling only. Ray and Ghista [24] presented a generalized direct computational algorithm to estimate the time dependent, in-vivo constitutive properties of the left ventricle, ascending aorta, and mitral valve leaflet. The time-dependent properties are important in constructing realistic models and their verification. Chen et al. [25] presented a finite element formulation for an active left ventricle model by modeling the strain in the myocardium by the biological activation as well as nonbiological stress. Moore et al. [26] have presented a theoretical foundation for a finite element computational model of the ventricular myocardium during sustained +Gz without really addressing the effect of the acceleration.

In a discussion on the cardiovascular system, reference to the flow of blood in compliant tubes is obvious. Porenta et al. [27] have developed a finite element model for blood flow in arteries with taper, branches, and stenoses assuming thin wall and axisymmetry. Rogers, et al. [28] presented an algorithm for the solution of the incompressible Navier-Stokes equations in three-dimensional generalized curvilinear coordinates. This algorithm is based on the method of artificial compressibility and uses a third order flux-difference splitting technique for the convective terms. They have applied this technique to calculate the flow through an artificial heart configuration with moving boundaries.

On the experimental side, much work on the effects of acceleration stress on the human body has been done at the United States Air Force School of Aerospace Medicine at Brooks Air Force Base, Texas. Burton et al. [29] subjected miniature swine to Gz acceleration to study its effect on the organism and extrapolated the results to humans. Parkhurst et al. [30] conducted experiments on human tolerance to high, sustained +Gz forces. Leverett et al. [31] investigated the physiologic responses to high sustained acceleration stress. Burton and McKenzie [32] determined the extent of heart pathology as a function of acceleration stress.

In the area of protective measures from gravitational stress, the Crew Technology Division at Brooks Air Force Base has done considerable research, and played an important role in the development of the G-suit and anti-G straining maneuvers. Burns [33] described some of the problems associated with high sustained +Gz; two of the major problems are maintenance of venous return to the heart and maintenance of blood pressure and blood flow to the brain.

The prevalence of heart disease in the general population has prompted researchers to take into consideration subclinical cardiovascular disease exacerbated by abnormal accelerations. Samn [34] has addressed the problem associated with the combined stress of subclinical cardiovascular diseases and acceleration. Particularly, the mitral valve prolapse among the pilots is the problem manifesting in lower tolerance for gravitational stress with susceptibility to developing mitral regurgitation leading to heart failure. This problem can be most effectively addressed through mathematical modeling if accurate geometry of the ventricles and heart valves are available. This is possible through modern imaging techniques.

It can be concluded from the information available in open literature that there is a great need for a comprehensive mathematical model of the cardiovascular system based on mechanics and realistic mechanical properties to study the effects of special environments such as high-g and microgravity. The interaction between the computational approach presented in this report, the treatment of nonlinear compliant materials such as myocardium and blood vessels, and the opportunities to develop electromechanical control devices to fight gravity or lack of it provides a major engineering challenge.

STATEMENT OF THE PROBLEM

The human body is well accustomed to the earth's force of gravity. However, recent aerospace developments have occasioned its exposure to the hazards of abnormal gravitational fields which are manifested in the form of vibration, impact, weightlessness, and rectilinear and angular accelerations beyond the levels of human tolerance. Abnormal accelerations on the human body, depending upon their severity and kind, are known to cause a variety of pathophysiologic effects such as headache, abdominal pain, impairment of vision, hemorrhage, fracture, and loss of consciousness [1-4]. All these effects influence the performance of pilots, compromising the high-performance capabilities of the new generation aircraft they operate, let alone the loss of aircraft, and possibly lives, by crash. In the context of space travel, cardiovascular adaptation to microgravity is of concern: shift of blood and tissue fluids from lower to upper body causing headward edema, reduced blood volume, and perhaps altered autonomic control of the circulation are significant issues. The cardiovascular system, being central to the homeostasis of the organism, is extremely susceptible to the hostilities of the changes in the environmental forces. The design of protective devices and procedures, and medical selection and physiological training protocols which are expected to provide acceleration tolerance and protection for the pilots during aircraft and spacecraft maneuvers must take into consideration the response of the cardiovascular system to the altered gravitational environment. Therefore, a thorough understanding of the cardiovascular system and its structure-function relationship in abnormal force environments is essential to any effort directed to overcome the gravitational trauma.

The prohibitiveness of actually subjecting the human body to abnormal, altered gravitational forces to gain knowledge of the cardiovascular system's response is obvious. The alternative is to develop a mathematical model and investigate its response. The need for mathematical models and the analysis of model features for prediction of system performance are well recognized in view of the cost and risk involved in testing the original system which, in this case, is the human subject. This report deals with the development of a finite element approach to the problem of three-dimensional blood flow through a highly deformable thick elastic tube which could be used in the modeling of the cardiovascular system in altered gravitational environment.

GOVERNING EQUATIONS

The cardiovascular system consists of several components, the inclusion of which will become too complicated to handle in a general model. While it is desirable to include the behavior of each cardiovascular component under acceleration stress in the total modeling effort, certain components can be lumped together to simplify the analysis and still preserve the character of the system.

The present study will utilize the Navier-Stokes equations for blood flow, and the large elastic deformation theory of thick-walled elastic chambers for the analysis of deformable blood vessels and the heart pump. The governing equations will couple the fluid flow, deformation of the blood vessels and the heart pump, and the material properties of the tissue in the construction of a realistic mathematical model.

The governing equations for the fluid are the momentum equation

$$\rho\,[\bar{u}_t + (\bar{u}\cdot\nabla)\,\bar{u}] = -\nabla p + \mu\nabla^2\bar{u} \tag{1}$$

and the continuity equation

$$\nabla\cdot\bar{u} = 0 \tag{2}$$

Here,

$\bar{u}$ = velocity vector
ρ = density of blood
p = pressure
μ = viscosity coefficient

The governing equations for the thick wall and boundary conditions are [36]:

$$\tau^{ij}\|_i + \rho_o F^j = \rho_o f^j$$
$$\tau^{ij} n_j = p^i \tag{3}$$

in which

τ^{ij} = stress tensor
ρ_o = wall density

F^j = body force vector
f^j = acceleration vector
p^i = pressure
n_j = unit normal

We assume the wall is hyperelastic and the stress tensor is derivable from a strain energy function $W = W(I_1, I_2, I_3)$:

$$\tau^{ij} = \frac{1}{2\sqrt{I_3}} \left(\frac{\partial W}{\partial \gamma_{ij}} + \frac{\partial W}{\partial \gamma_{ij}} \right) \tag{4}$$

where
W = strain energy function
I_1, I_2, I_3 = strain invariants
γ_{ij} = strain tensor

A common strain energy function for biological tissue is of the form

$$W = c\left[e^{b(I_1-3)}-1\right] + A\left[e^{a(\alpha-1)^2}-1\right] \tag{5}$$

Where α is the stretch ratio in the direction of the tissue fiber, and c, b, A and a are material parameters derived from test data. Different parameters represent different constitutive relations pertinent to the locality of the tissue such as myocardium or arterial wall, etc.

With proper treatment of boundary conditions at the solid-fluid interface, these equations can be solved by the finite element method [Oden, 1972]. A strain energy function appropriate for the biological tissue may be selected from available data bases.

THE FINITE ELEMENT METHOD

Because our final application is aimed at a realistic cardiovascular system which has a nonsymmetric complicated geometry, a three-dimensional finite element method has been used in the solution of the governing equations for the solid deformation.

The fluid flow in a three-dimensional thick-walled elastic tube segment and the stresses on an element isolated from the wall are depicted in Fig. 1. The system of equations (1) - (4) have been solved by the finite element method which is equally applicable to linear or nonlinear equations and can handle various types of boundary conditions rather easily in comparison with any attempt to solve these quations analytically.

Introducing the velocity-pressure formulation with approximations of the type

$$\begin{aligned} u_i(\bar{x}, t) &= \phi_N(\bar{x})\, u_{Ni}(t) \\ p(\bar{x}, t) &= \Psi_N(\bar{x})\, p_N(t) \end{aligned} \tag{6}$$

where N is the number of nodes in the element, and Φ and Ψ are the interpolation functions in the spatial domain. Equations (1) and (2) can be transformed into a system of algebraic equations for the solution of the nodal velocities. However, the deformation of the wall and the fluid pressure on the wall are inter-related.

The nodal displacements in the solid element of the wall can be approximated by

$$u_i\ (\bar{x}, t) = \beta_N\ (\bar{x})\ u_{N_i}\ (t) \tag{7}$$

where β represents the interpolation function. Then the strain tensor γ_{ij} becomes

$$\gamma_{ij} = \frac{1}{2}\ (\beta_{N,i} u_{N_j} + \beta_{N,j} u_{N_i} + \beta_{N,i} u_{N_K} \beta_{M_j} u_{M_K}) \tag{8}$$

Introduction of Eq (8) into the equation of motion (Eq (3)) yields the required pressure-displacement relationship in the form

$$f_k^N = \int_{\Omega^e} \tau^{ij}\ \Psi_{N',i}\ (\delta_{j_k} + \Psi_{M',j} u_{M_k})\ d\Omega^e \tag{9}$$

where f_k^N are the generalized nodal force components. Assembling Equs. (6) and (9) element by element yields a global system of equations which can be solved for the nodal displacements, velocities and pressure. $\bar{u}$ and u_i in Eqs. (1), (2) and (6) are the fluid velocities. u_i in Equs. (7), (8) and (9) are the nodal displacements of the solid element which will yield the boundary velocities at $r = R_1$ (inside boundary) upon differentiation. u_i of the fluid will be matched with the $\frac{\partial u_i}{\partial t}$ of the solid to satisfy the boundary conditions.

The displacements can be used to compute the strains and subsequently the stresses in the wall. Also, the deformed shape of the thick walled blood vessel and the heart which is its topological equivalent can be computed. These results may be used to develop a strategy for constructing a comprehensive mathematical model of the cardiovascular system in altered gravity.

SOLUTION FOR AN ARTERIAL SEGMENT

An arterial segment 0.25 inch long with wall thickness varying between 0.04 to 0.08 inch, both axially and circumferentially, is discretized into finite elements. The longitudinal and circumferential cross-sections of the segment are shown in Figs. 2 and 3 with nodes numbered as indicated.

A typical pressure profile generated during a heart beat is shown in Fig. 4 ([37]). This pressure profile is transformed into appropriate physical units and discretized for use in the finite element solution in the time domain. The transformed pressure profile is replotted in Fig. 5.

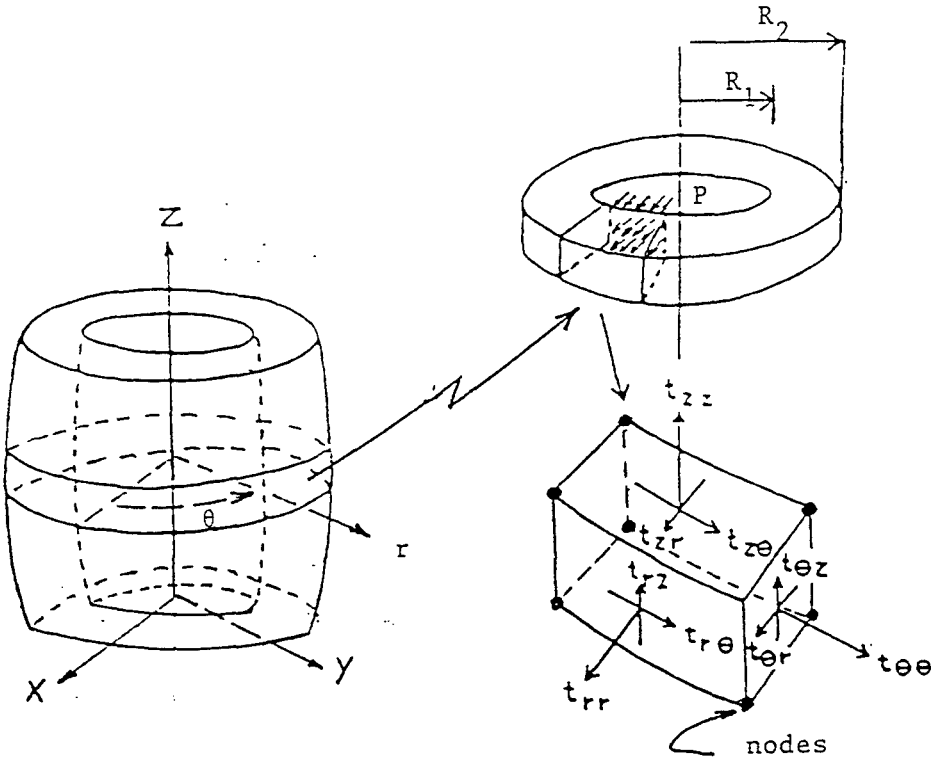


Fig. 1. Schematic drawings of a topologically equivalent thick-walled elastic tube.

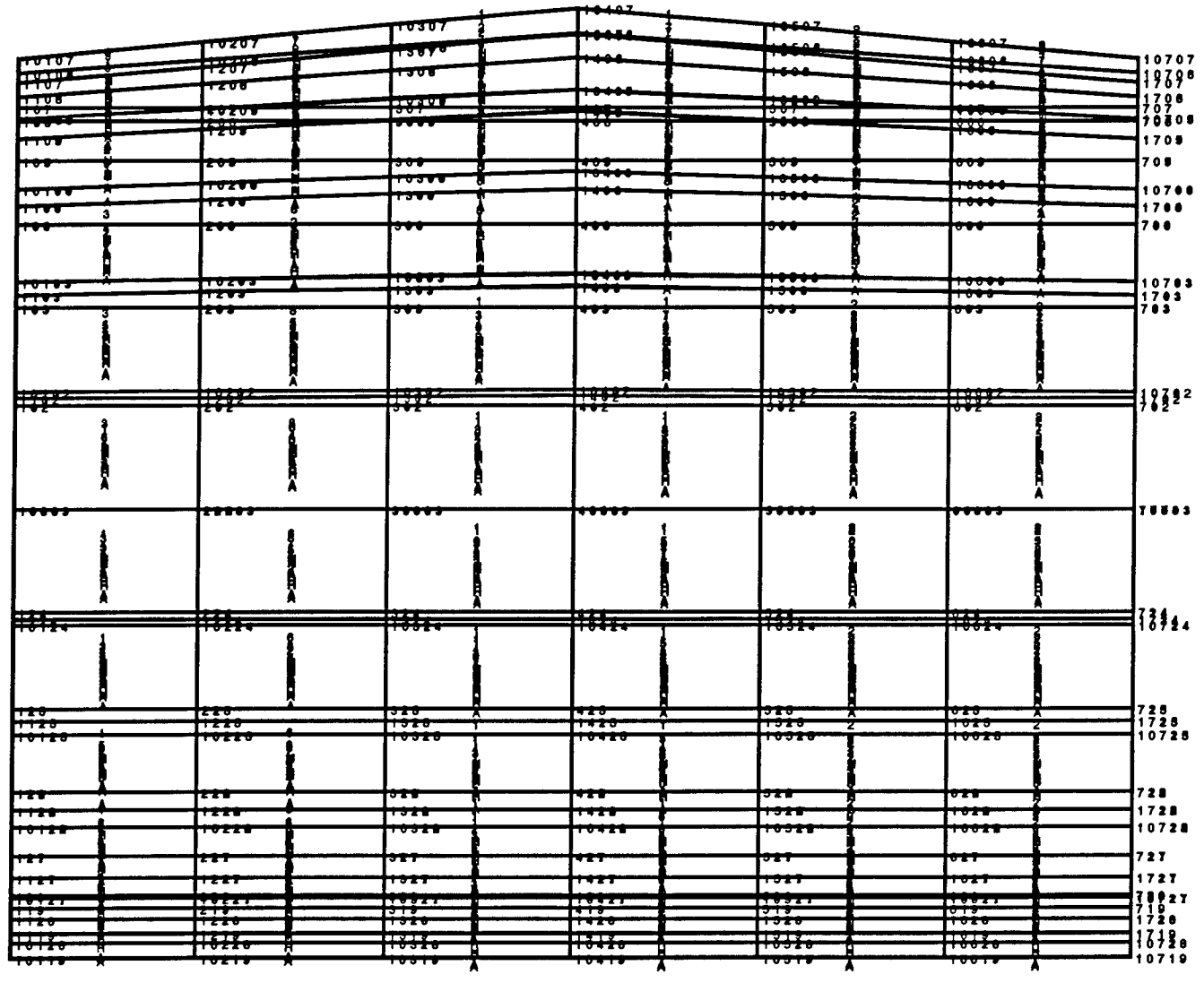


Fig. 2. Longitudinal cross-section of an arterial segment discretized finite elements and numbered nodes.

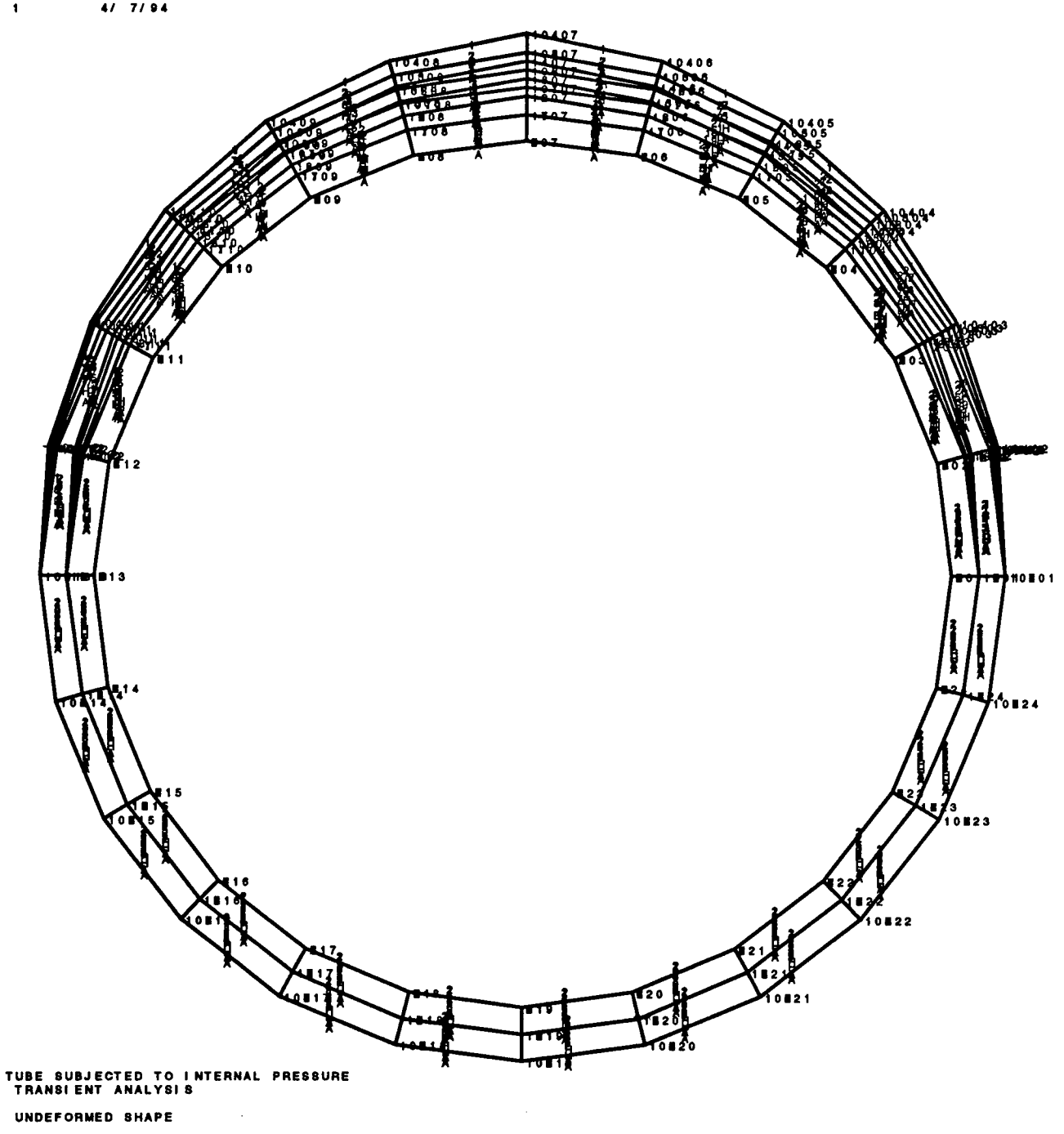


Fig. 3. Circumferential cross-section of an arterial segment discretized into finite elements with numbered nodes.

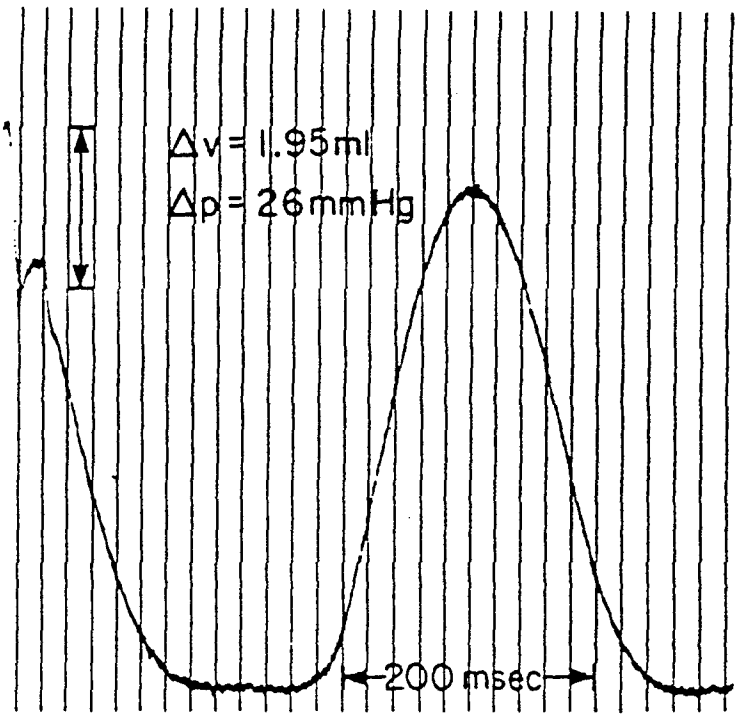


Fig. 4. A typical blood pressure profile generated during a heart beat.

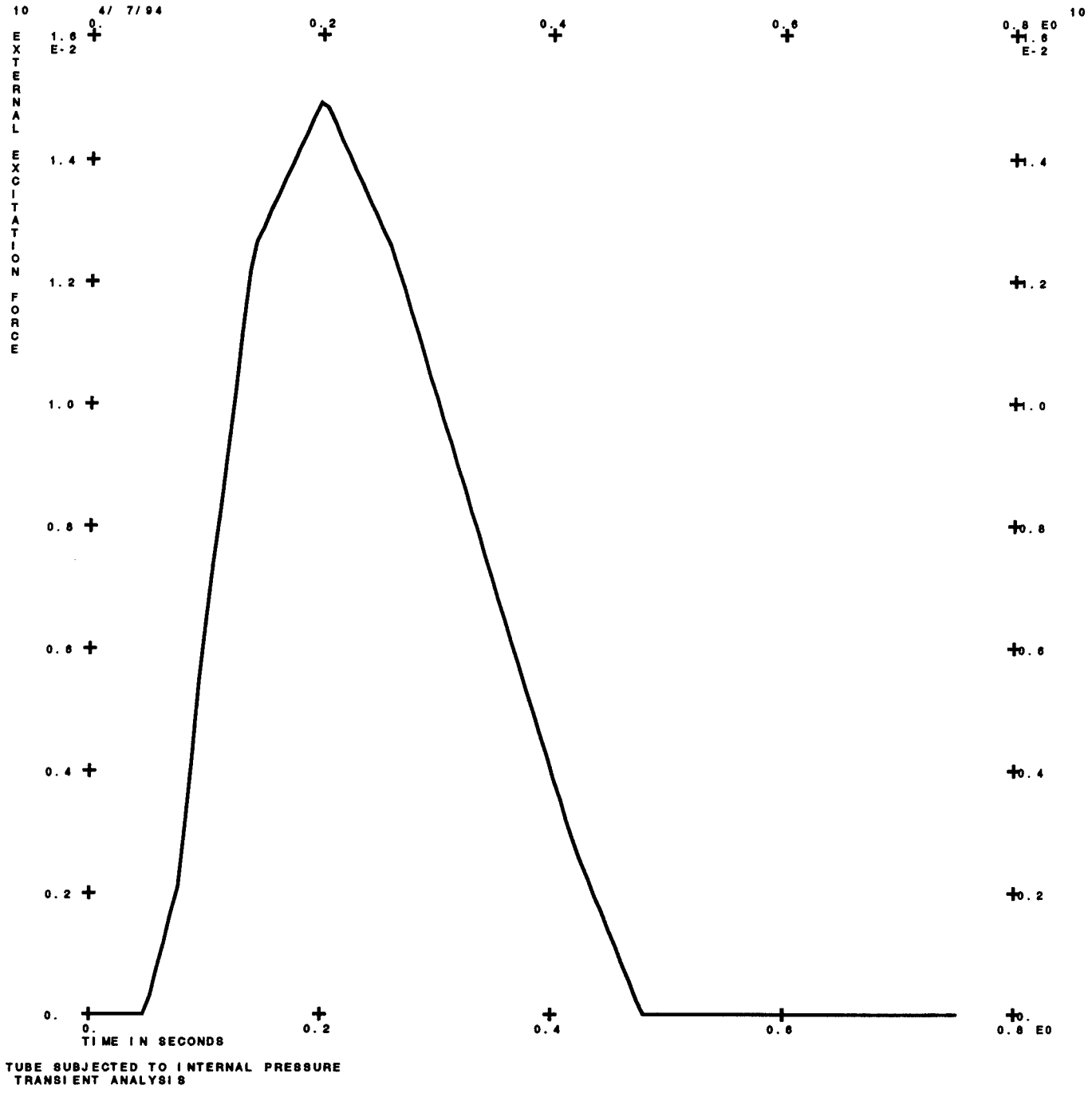


Fig. 5. Reconstructed pressure profile using appropriate units.

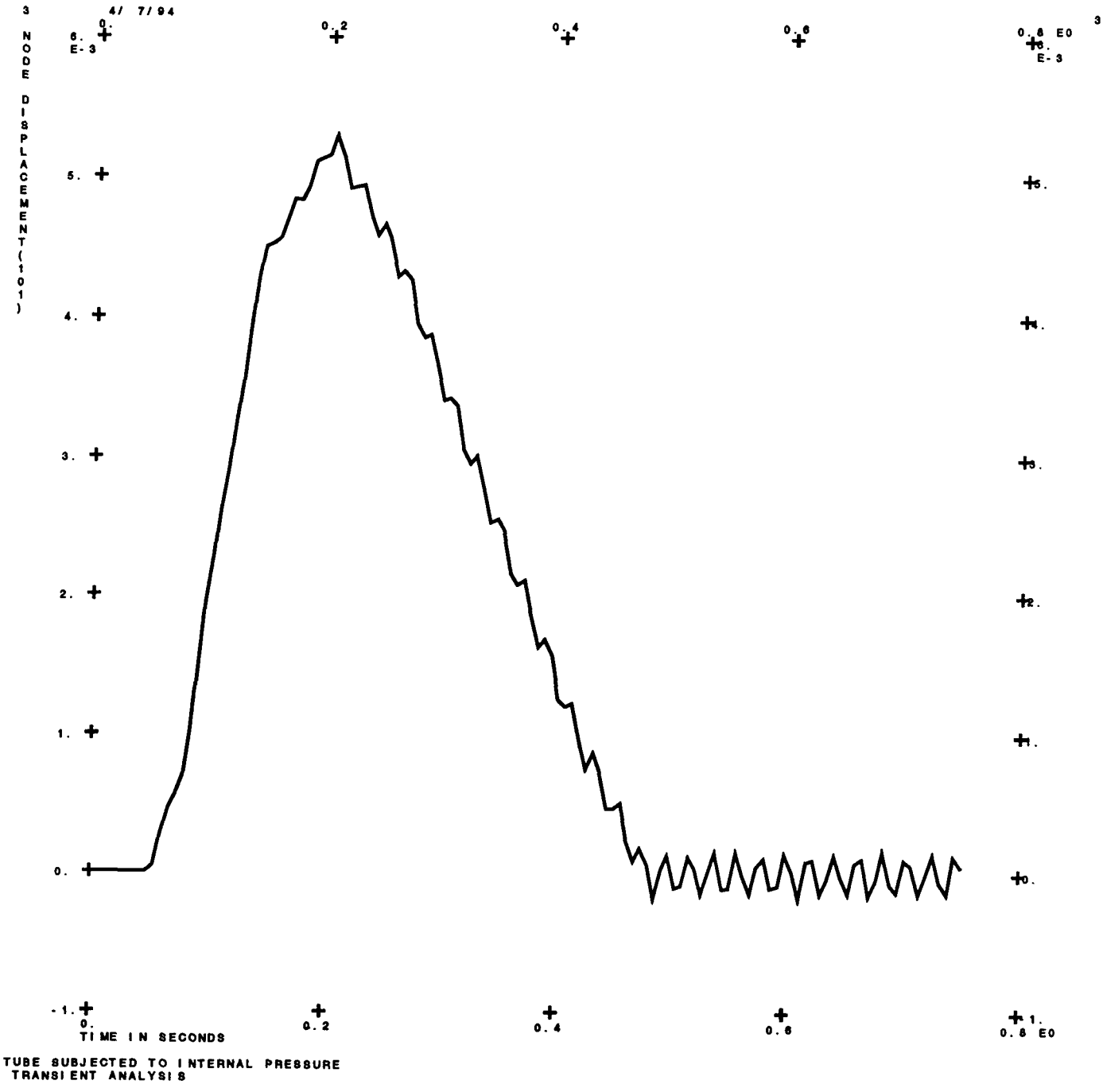


Fig. 6. Displacement of a node in the wall thickness as a response to changing pressure.

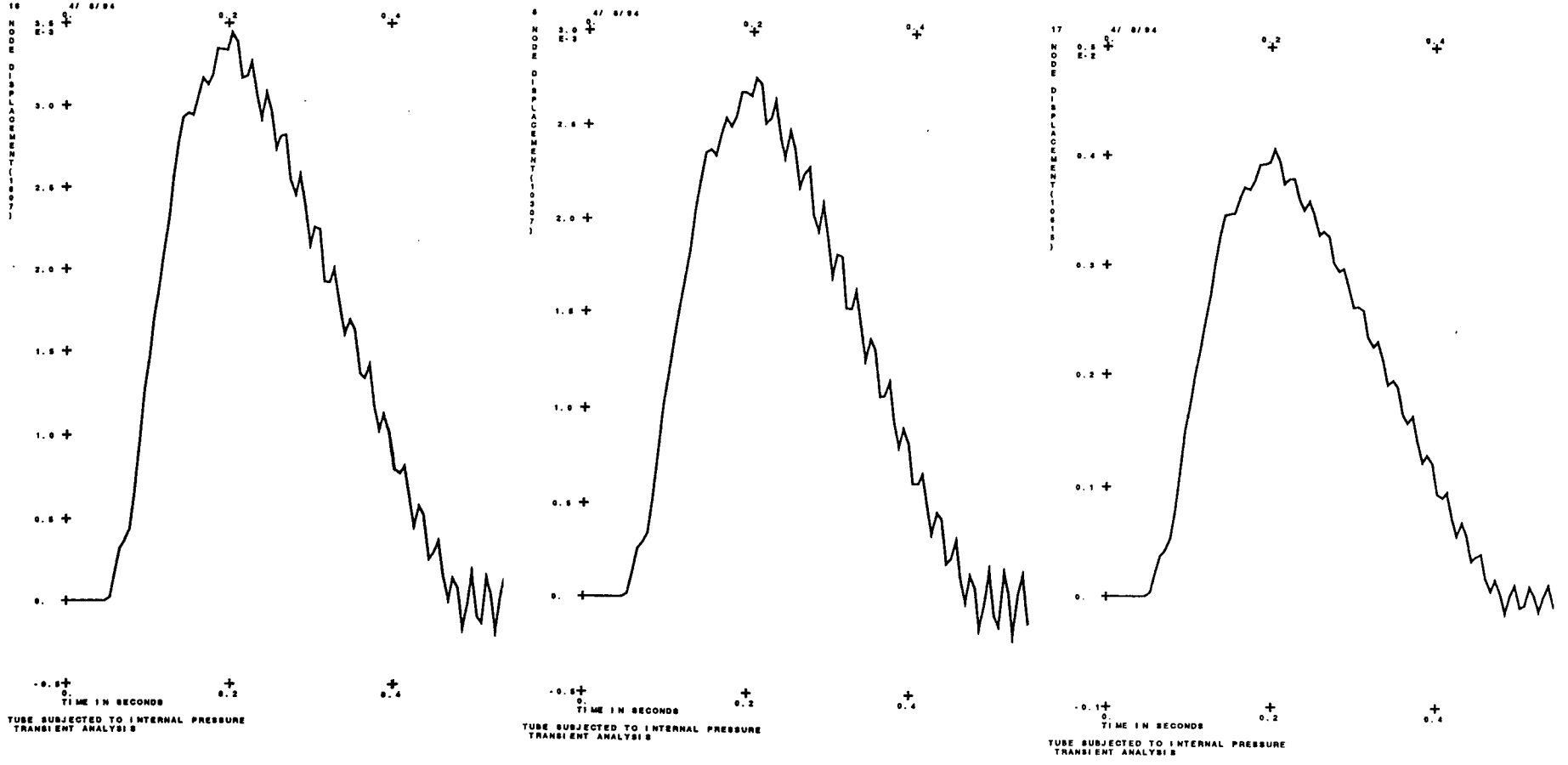


Fig. 7. A comparison of computed displacements of various nodes in the arterial wall thickness.

The response of the arterial wall to the pressure profile depicted in Fig. 5 is shown in Fig. 6 in terms of the displacement of the nodes.

In Fig. 7 is shown a comparison of time-dependent nodal displacements of different nodes.

The volume change in the arterial segment can be easily computed from the displacements of the nodes on the internal wall. The finite element procedure can be extended to the entire arterio-venous system to calculate the required volumetric changes. The linearization of the governing equations allowed the usage of MSC/NASTRAN finite element procedure in the computations.

CONCLUSIONS

The information from this research will help in the design of engineering devices to counteract the ill-effects of altered gravity encountered in high performance aircraft. An example is the Advanced Technology G-Suit which covers the entire legs and feet and is interconnected with an abdominal bladder [33]. Scientists recognize a continuing need to develop improved gravity-protective equipment and/or techniques for augmentation of the cardiovascular function for which a realistic cardiovascular system model is necessary. The finite element technique followed in this study can be applied to the entire cardiovascular system and the required parameters for the design of anti-G devices can be derived.

ACKNOWLEDGMENT

This work was performed under the United States Air Force Contract No. F49620-90-C-09076 awarded to the Research and Development Laboratories, Culver City, California, and Subcontract No. 93-168 awarded to the University of Missouri-Rolla. The authors gratefully acknowledge this support.

REFERENCES

[1] Hiatt, E. P., Meechan, J. P., and Galambos, R., Reports on Human Acceleration, Publication 901, National Academy of Sciences - National Research Council, Washington, DC, 1967.

[2] Space Sciences Board, NAS-NRC, Physiology in the Space Environment, Vol. 1 - Circulation, Publication 1485A, National Academy of Sciences, Washington, DC, 1968.

[3] Burton, R. R., Leverett. S. D., Jr., and Michaelson, E. D., "Man at High Sustained +Gz Acceleration: A Review", Aerospace Medicine, Vol. 45, pp. 1115-1136, 1974.

[4] Whinnery, J. E., "Medical Considerations for Human Exposure to Acceleration Induced Loss of Consciousness", Aviation, Space and Environmental Medicine, pp. 618-623, July 1991.

[5] Womersley, J. R., "Mathematical Analysis of the Arterial Circulation in a State of Oscillatory Motion", Wright Air

Development Center, Technical Report WADC-TR-56-164, 1958.
[6] Noordergraaf, A., "Hemodynamics", in Biological Engineering, Schwan, H., (Ed.), McGraw-Hill, New York, 1969.
[7] Taylor, M. D., "The Input Impedance of an Assembly of Randomly Branching Elastic Tubes", Biophysical Journal, Vol. 6, pp. 25-51, 1966.
[8] Kenner, T., "Flow and Pressure in Arteries", in Biomechanics: It's Foundations and Objectives, Fung, Y. C. et al. (Eds.) Prentice-Hall, Englewood Cliffs, New Jersey, 1972.
[9] Attinger, E. O., Anne, T., Mikami, T., and Sugawara, H., "Modeling of Pressure Flow Relationships in Veins", in Hemorheology, Copley, A., (Ed.), McGraw-Hill, New York, 1964.
[10] Beneken, J. E. W., and DeWit, B., "A Physical Approach to Hemodynamic Aspects of the Human Cardiovascular System", in Physical Bases of Circulatory Transport: Regulation and Exchange, Reeve, E. B. and Guyton, A. C., (Eds.), W. B. Saunders Company, Philadelphia, 1967.
[11] McLeod, J., "PHYSBE ... A Physiological Simulation Benchmark Experiment", Simulation, Vol. 7, pp. 324-329, December 1968.
[12] Sagawa, K., "Overall Circulatory Regulation", in Annual Reviews of Physiology, Vol. 21, pp. 295-330, Annual Reviews, Inc., Palo Alto, California, 1969.
[13] Sagawa, K., "The Circulation and Its Control I: Mechanical Properties of the Cardiovascular System", in Engineering Principles in Physiology, Vol.2, Brown, J. H. V. and Gann, D. S., (Eds.), Academic Press, New York, 1973.
[14] Avula, X. J. R. and Oestreicher, H. L., "Mathematical Model of the Cardiovascular System Under Acceleration Stress", Aviation, Space, and Environmental Medicine, pp. 279-286, January 1978.
[15] Moore, T. W., Jaron, D., "Cardiovascular Model for Studying Circulatory Impairment Under Acceleration", IEEE Transactions on Engineering in Medicine and Biology", Vol. 10, No. 1, pp. 37-40, 1991.
[16] Jaron, D., Moore, T. W., and Bai, J., "Cardiovascular Responses to Acceleration Stress: A Computer Simulation", IEEE Proceedings, Vol. 76, No. 6, pp. 700-707, 1988.
[17] Hargens, A. R., Watenpaugh, D. E., and Breit, G. A., "Control of Circulatory Functions in Altered Gravitational Fields", The Physiologist, Vol. 35, No. 1, Suppl., 1992.
[18] Mirsky, I., "Left Ventricular Stresses in the Intact Human Heart", Biophysical Journal, Vol. 9, pp. 189-208, 1969.
[19] Demiray, H., "Stresses in Ventricular Wall", Transactions of the ASME, J. of Applied Mechanics, pp. 194-197, June 1976.
[20] Demiray, H. and Vito, R. P., "Large Deformation Analysis of Soft Biomaterials", Developments in Theoretical and Applied Mechanics, Vol. 8, Proc. of the 8th Southeastern Conference on Theoretical and Applied Mechanics, pp. 515-522, 1976.
[21] Pao, Y. C., Ritman, E. L., and Wood, E. H., "Finite Element Analysis of Left Ventricular Myocardial Stresses", Journal of Biomechanics, Vol. 7, pp.409-417, 1974.
[22] McCulloch, A. D. and Omens, J. H., "Non-Homogeneous Analysis of Three-Dimensional Transmural Finite Deformation in Canine Ventricular Myocardium", Journal of Biomechanics, Vol. 24,

No. 7, pp. 539-548, 1991.
[23] Huyghe, J. M., van Campen, G. H., Arts, T., and Heethaar, R. M., "A Two-Phase Finite Element Model of the Diastolic Left Ventricle", Journal of Biomechanics, Vol. 24, No. 7, pp.527-538, 1991.
[24] Ray, G. and Ghista, D. N., "A Direct Computational Algorithm to Estimate In-Vivo Constitutive Properties of Some Cardiovascular Structures", Proceedings, International Conference on Finite Elements in Biomechanics, (B. R. Simon, Editor), Vol. 2, Tucson, Arizona, February 1980.
[25] Chen, C. J., Kwak, B. M., Rim, K., and Falsetti, H. L., "A Model for an Active Left Ventricle Deformation Formulation of a Nonlinear, Quasi-Steady Finite Element Analysis for Orthotropic, Three-Dimensional Myocardium", Proceedings, International Conference on Finite Elements in Biomechanics, (B. R. Simon, Editor), Vol. 2, Tucson, Arizona, February 1980.
[26] Moore, J. E., Tabarrok, B., and Haddow, J. B., "Mathematical Modelling of +Gz Acceleration Induced Stresses in the Human Ventricle Myocardium", The Physiologist, Vol. 35, No. 1, Suppl., 1992.
[27] Porenta, G., Young, D. F., and Rogge, T. R., "A Finite Element Model of Blood Flow in Arteries Including Taper, Branches, and Obstructions", Journal of Biomechanical Engineering, Vol. 108, No. 2, pp.161-167, 1986.
[28] Rogers, S. E., Kwak, D., and Kiris, C., "Steady and Unsteady Solutions of the Navier-Stokes Equations", AIAA Journal, Vol. 29, No. 4, pp. 603-610, 1991.
[29] Burton, R. R., "Positive (+Gz) Acceleration Tolerances of the Miniature Swine: Application as a Human Analog", Aerospace Medicine, Vol. 44, pp. 294-298, 1973.
[30] Parkhurst, M. J., Leverett, S. D., Jr., and Shurbrooks, S. J., Jr., "Human Tolerance to High, Sustained +Gz Acceleration", Aerospace Medicine, Vol. 43, pp.708-712, 1972.
[31] Leverett, S. D., Jr., Burton, R. R., Crossley, R. J., Michaelson, E. D., and Shurbrooks, S. J., Jr., "Physiologic Responses to High, Sustained +Gz Acceleration", USAF School of Aerospace Medicine, Technical Report 73-21, 1973.
[32] Burton, R. R., and McKenzie, W. F., "Heart Pathology Associated with Exposure to High Sustained Gz", Aviation, Space and Environmental Medicine, Vol. 46, No. 10, pp. 1251-1253, 1975.
[33] Burns, J. W., "High-G Cardiovascular Physiology", The Physiologist, Vol. 35, No. 1, Suppl., 1992.
[34] Samn, S., "The Role of Subclinical Cardiovascular Disease in High-G Flying: A Mathematical Modeling Approach", The Physiologist, Vol. 35, No. 1, Suppl., 1992.
[35] Oden, J. T., Finite Elements of Nonlinear Continua, McGraw-Hill, Inc., New York, New York, 1972.
[36] Green, A. E. and Adkins, J. E., Large Elastic Deformations and Nonlinear Continuum Mechanics, Clarendon Press, Oxford, 1960.
[37] Covell, J. W., "Mechanics of Contraction in the Intact Heart", In Biomechanics: Its Foundations and Objectives (Y.C. Fung, N. Perrone, M. Anliker, Eds.), Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972.

A STUDY OF THE CONTRAST DETECTION MODELING FOR HUMAN EYE AND ITS APPLICATION TO THE DETECTION AND RECOGNITION TASKS OF THERMAL IMAGERY

Dr. Jer-Sen Chen
Assistant Professor
Department of Computer Science and Eng.

Wright State University
3640 Colonel Glenn Highway
Dayton, OH 45435

Final Report for:
Summer Research Extension Program
Armstrong Laboratory

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

and

Wright State University

December 1993

A STUDY OF THE CONTRAST DETECTION MODELING FOR HUMAN EYE AND ITS APPLICATION TO THE DETECTION AND RECOGNITION TASKS OF THERMAL IMAGERY

Jer-Sen Chen
Wright State University
Department of Computer Science and Eng.

Abstract

This report presents a study on the effects of noise to the perception of contrast. The experiments conducted in this research consists of the contrast perception in various kinds of noise. We first investigate the white noise case, including both uniform and normal distributions, at different noise levels. It is followed by varying the frequency characteristics of the noise, including low pass filtered noise and $\frac{1}{f}$ noise. Contrast threshold models proposed by various researchers are then used to fit each experimental result to characterize and investigate the noise effects on contrast perception.

A STUDY OF THE CONTRAST DETECTION MODELING FOR HUMAN EYE AND ITS APPLICATION TO THE DETECTION AND RECOGNITION TASKS OF THERMAL IMAGERY

Jer-Sen Chen

Introduction

Human eye modeling is an important component in the performance modeling of the eletro-optical imaging systems. The evaluation of the interaction of human vision with the thermal imaging systems has long been studied and recognized as the essential factor in the prediction of the system performance. It is usually desired to characterize the model of the human eye with a simple analytical expression, in particular, the modulation transfer function (MTF). Once the analytical form of the human eye model is derived, it can then be integrated into other components of the eletro-optical system and proceed the performance modeling.

The NVL model [Ratches75] was developed for predicting the performance of United States Army thermal imaging system. It is used to predict the *minimum resolvable temperature* (MRT) for target detection and recognition. The model can adequately predict the MRT at mid-range spatial frequencies which corresponds to detecting modest sized targets at modest ranges. However, certain discrepancy exists between the laboratory measured MRT values and the NVL model predicted values at low or high spatial frequency ranges. The differences can be caused by a variety of reasons [Holst90]:

- The eye is not modeled appropriately at low or high spatial frequencies.
- There exists tremendous variability in observer response.
- The exact test procedure and data analysis methodology are not well documented.

The ability to accurately quantify the eye-brain detection process has been perhaps the major cause of the discrepancy between the NVL model predicted MRT and the laboratory measured values. The Kornfeld and Lawson model [Kornfeld71] used in NVL model for the eye model does

not include the inhibitory portion of the eye. This lack of complete eye modeling accounts in part for the discrepancy of the measured and predicted values at low spatial frequencies.

The other important factor is to account for the effect of noise on the modulation transfer function of the visual channel. The separation of the noise from the MTF of the eye is the customary method of analysis. Integration of noise into the modeling of the eye has recently received increasing attention from researchers. Many equivalent-noise models for contrast detection and discrimination tasks have also been developed.

Contrast Sensitivity Function in Noise

The J-shape curve is usually reported for the contrast sensitivity experiment of the human eye. The contrast sensitivity is commonly measured in the uniform fields free of noise. The shape of the curve actually varies especially in the presence of noise. Examples can be found in such experiments as [Pollehn70] and [vanMeeteren88].

Effects of the noise, in terms of coarseness, interfering with visual detection was studied by van Meeteren and Valeton [vanMeeteren88]. Similar results were derived in the cases of fine noise. When the coarseness of the noise increases, however, a more significant degree of degradation in the low spatial frequencies results. This paper presents a study on the effects of a wide range of noise to contrast perception. It is our hope that this investigation of noise effect could help explain and correct the discrepancy between the predicted and measured values of MRTD in the FLIR systems performance model [Scott90].

The experimental results and mathematical models proposed by previous researchers dealing with the presence of noise are first studied. The contrast threshold under the effect of noise is usually expressed in terms of multiplication or addition. In the addition model, the contrast threshold is expressed as the sum of internal and external noise. The other modeling formulates the effects noise in terms of multiplicative modification to contrast perception which is commonly used in computing minimum resolvable temperature in thermal imaging systems. Though there exists an equivalence between the models in terms of addition or multiplication, the dependence on spatial frequency is

modeled quite differently.

The modification of the J-shape contrast sensitivity function by the effects of noise is commonly expressed in terms of addition or multiplication. The noise effects on contrast threshold can also expressed in multiplicative form with respect to spatial frequency f_x, namely:

$$M_t^{'}(f_x) = M_t(f_x) \times N(f_x)$$

where $M_t^{'}(f_x)$ is the threshold contrast in noise, $M_t(f_x)$ is the threshold contrast without noise, and $N(f_x)$ accounts for the effect of noise.

This form is commonly used in night vision system where the contrast threshold is usually expressed in terms of minimum resolvable temperature (MRT). The MRT can be calculated from [Holst90]:

$$MRT = \frac{SNR \times NETD}{MTF_{sys} \times MTF_{eye}}$$

where $NETD$ is the system noise equivalent temperature difference which accounts for the noise level or the standard deviation. Holst and Pickard [Holst90] suggest that:

$$SNR(f_x) = 0.215(f_x)^{0.5509}$$

to account for the dependency on spatial frequency f_x. A similar result was reported by Driggers et al [Driggers90].

Internal Noise and External Noise

Noise effect in terms of addition is derived from the concept of internal noise and external noise. When no external noise is present, the threshold represents the energy necessary to overcome the existence of the internal noise. The contrast threshold in noise is therefore expressed as: (see [Daly89] and [Barten92])

$$M_t^{'2}(f_x) = M_t^2(f_x) + M_n^2(f_x)$$

where, $M_n(f_x)$ accounts for the noise effect. Note that the sum of squares in the above equation

indicates the independence or the orthogonality between internal noise and external noise. Also note that the equivalence between the models of multiplication and addition can be established by simply equating

$$N^2(f_x) = 1 + \frac{M_n^2(f_x)}{M_i^2(f_x)}$$

In the addition model, the main difference between Barten's model and Daly's model is that, while Daly simply uses a spatial frequency independent constant in his formulation for contrast threshold in noise, Barten uses the following to account for the variation of the effects of noise at different spatial frequencies:

$$M_n(f_x) = \sqrt{\frac{2\Phi_n}{XYT}}$$

X is not just the width (X_o) of the object, but is limited by a maximum angular size (X_e) and a maximum number of cycles (N_e) over which the eye can integrate information. Therefore:

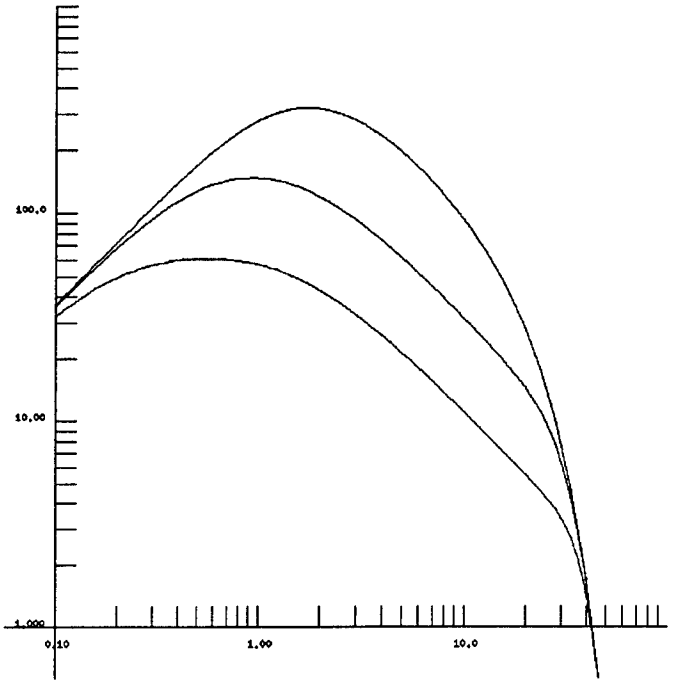
$$\frac{1}{X^2} = \frac{1}{X_o^2} + \frac{1}{X_e^2} + \frac{f_x^2}{N_e^2}$$

The values of $X_e = 12$ degrees and $N_e = 15$ cycles are used. There is a similar expression for Y, but the expression for T is independent of f_x, the spatial frequency, naturally [Barten92]. The value of $T = 0.1$ second is used.

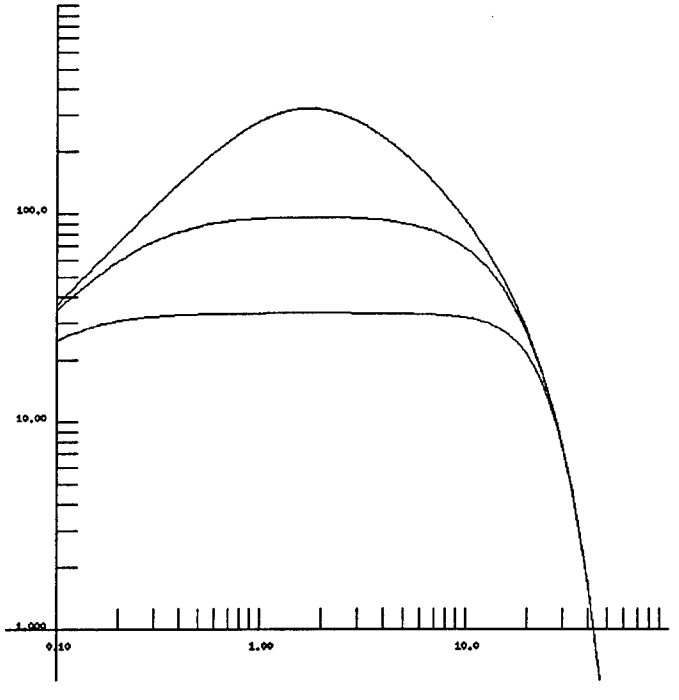
Figure 1 shows the difference between Barten and Daly models. Barten's model is shown in figure 1(a) while Daly's model in figure 1(b). In each subfigure, the top curve shows the CSF without noise, the curve in the middle shows the CSF with slight noise, and the bottom curve show the curve in higher degree of noise. Barten's model seems to be more consistent with the results by Pollehn and Roehrig [Pollehn70].

The Experiments

Our experiments on the effects of noise to contrast threshold are conducted to address various issues of noise modeling in contrast perception. The experiment starts with measuring the contrast



(a)



(b)

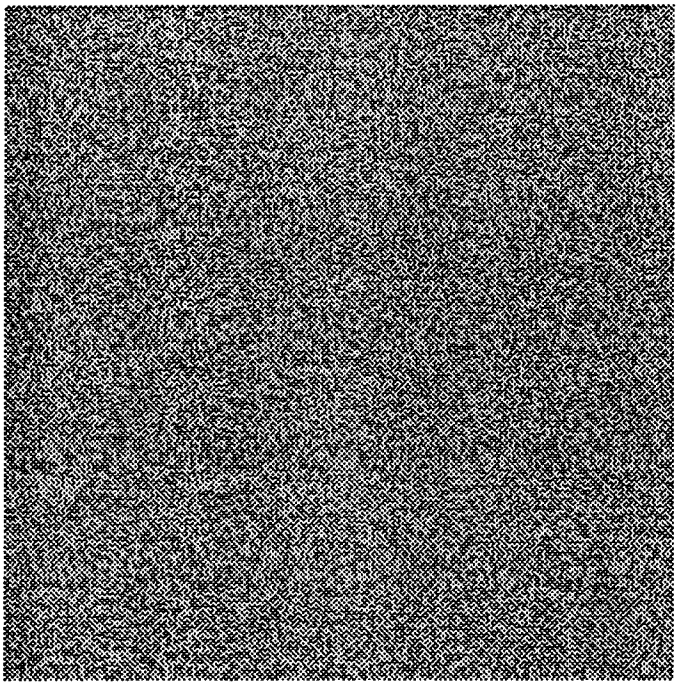
Figure 1: CSF in noise by Barten (a) and Daly (b)

thresholds without noise. We then add white noise at various noise levels to the sine wave grating to determine the contrast thresholds in noise. Each sample (pixel) of the white noise is independently generated by a random number generator with uniform distribution. The spectrum of the white noise is constant at all frequencies, and is characterized by the the noise level which is controlled by varying the standard deviation of the distribution of the random numbers. The results are then fitted with different mathematical models to account for the spatial frequency dependency of the contrast perception in the presence of white noise.

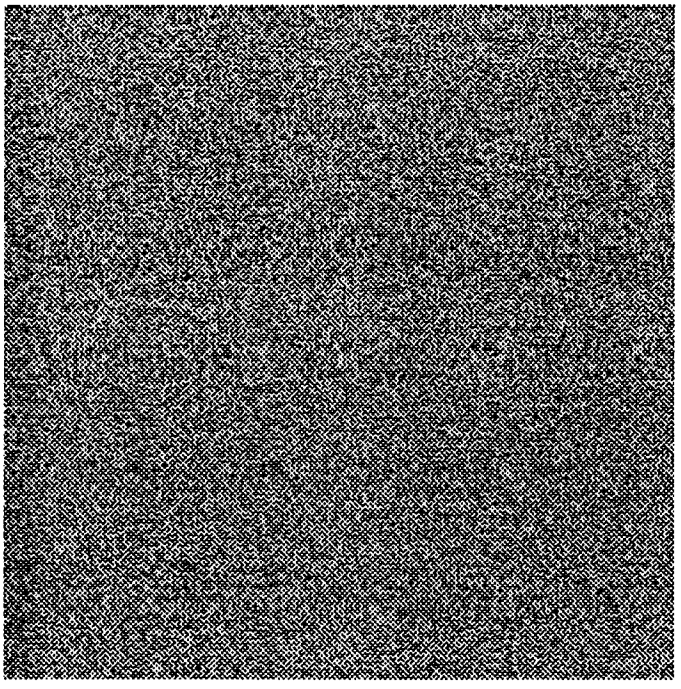
We also conduct the experiments of white noise with different type of distribution, namely the normal distribution. Each sample of the normally distributed white noise is generated by taking the average of 4 uniformly distributed random samples. Each sample is then scaled to generate desired standard deviation of the normal distribution. The results from two different distributed white noises with same standard deviation are then compared. Figure 2 shows the intensity profiles of the random white noise with two different distribution. Uniform distributed white noise is shown in figure 2(a), and normal distributed white noise is shown in figure 2(b). Both noises are set to equal standard deviation of 32 grayscale levels out of a full range of 256.

We then proceed to experiment on noise with different spatial spectral characteristics. In particular, the lowpass filtered noise is generated by passing the white noise through a digital filter cut off at desired frequencies. We also investigate the effects of $\frac{1}{f}$ noise. The $\frac{1}{f}$ noise is generated by using a midpoint displacement algorithm commonly used in fractal generation. Figure 3 shows the $\frac{1}{f}$ noise with standard deviation of 16 grayscale levels. The sine wave grating with different type of noise is shown in figure 4. Figure 4(a) is the signal with white noise, while figure 4(b) shows the sine wave added with $\frac{1}{f}$ noise. Note that the standard deviations of the two different types of noise are different. The standard deviation for white noise is 32 grayscale levels, while the $\frac{1}{f}$ noise is 16.

Sine wave gratings at nine different frequencies, ranging from 0.6 to 20 cycles per degree, are generated in each test to derive the contrast threshold at each frequency. At each frequency, a force choice method is used to ask the subject whether he or she can detect the presence of the sine wave grating. The contrast of the sine wave is then adjusted adaptively according to the response of the subject in order to derive the final contrast threshold at that spatial frequency.



(a)



(b)

Figure 2: Random white noise with (a) uniform (b) normal distribution

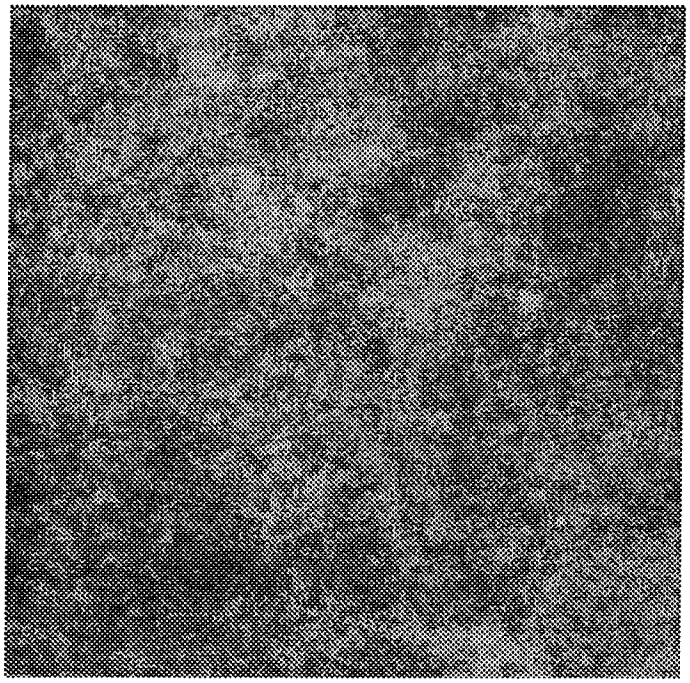
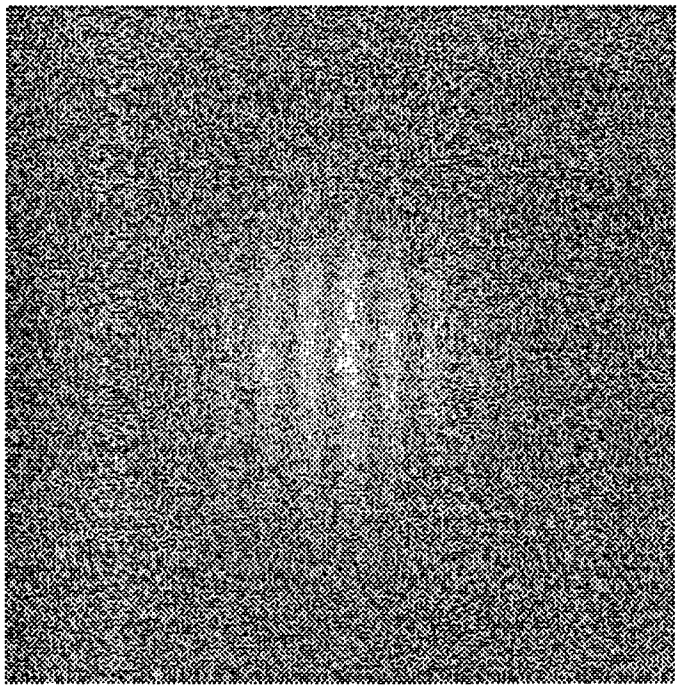
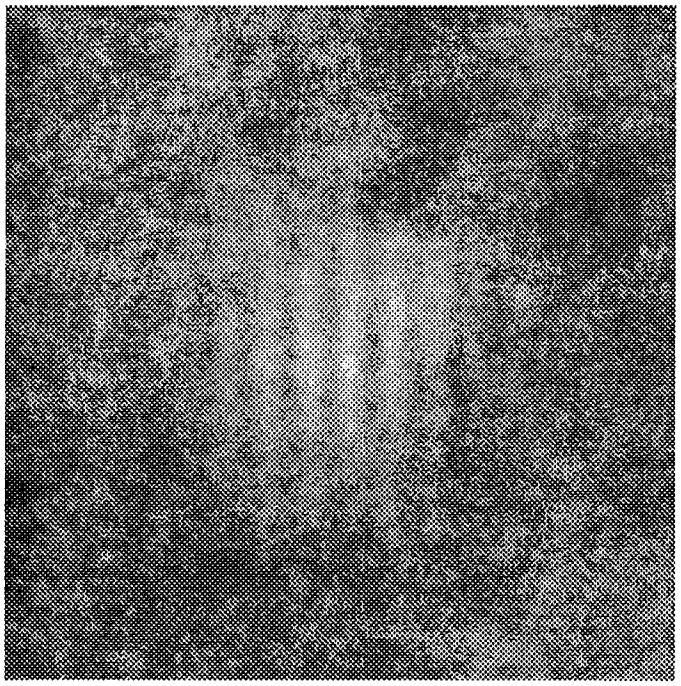


Figure 3: $\frac{1}{f}$ noise



(a)



(b)

Figure 4: Sine wave grating in (a) white noise, (b) $\frac{1}{f}$ noise.

There is a spatial masking on the sine wave grating but no temporal masking. That is the spatial boundary of the sine wave grating attenuates to zero through a Gaussian mask. Temporal edges, when the grating shows up as well as disappears from the screen, are not masked due to hardware limitation.

Each subject goes through 5 different sessions, namely,

- no noise,
- low level white noise,
- high level white noise,
- low pass filtered noise, and
- $\frac{1}{f}$ noise.

Each session last about 20 minutes.

In each case, the results are then compared with previous experiments, such as found in [Pollehn70] and [vanMeeteren88]. Furthermore, the results are fitted with different proposed formalism to verify the validity and the accuracy of the noise modeling.

Conclusions

The experiments were not completed at the time of the writing of the report to derive a population result. The conclusions made in this section are according to the previous experimental results and some preliminary tests we have so far conducted. A more detail and complete result will be derived once the population result becomes available.

The distribution of the random samples does not seem to have any impact on the contrast sensitivity as long as the standard deviation of the white noise is characterized. The limitations on the maximum obtainable grayscale levels introduce discrepancy between desired and actual random noises. For instance, for uniform distribution, a maximum standard deviation of 74 grayscale levels can be

obtained in a 256 total grayscale levels environment. For normal distribution, however, it takes a wider dynamic range to achieve the same standard deviation. In our simulation, we use a hard limiting function to limit all the random samples in the range of 256 grayscale levels. This, in general, introduces errors between the actual and the desired distributions, especially in the cases of high level normally distributed noises.

From the results by previous researchers and our experiments, the contrast sensitivity function of human eye in the presence of white noise proposed by Barten models better than the one proposed by Daly. In particular, the experimental results by Pollehn and Roehrig on the white noise at different noise levels do not exhibit any flat region of CSF in the presence of noise as suggested by Daly. Our preliminary experiments also show similar behavior.

Furthermore, the addition of random white noise introduces a shift in the spatial frequency where the maximum of the contrast sensitivity (or the minimum of the contrast threshold) occurs. In general, the frequency at which the maximum contrast sensitivity occurs decreases as the level of the random which noise increases. This particular behavior is also consistent with Barten's model.

In genearl, to account for the noise effect on the contrast sensitivity or threshold, the addition model:

$$M_t'^2(f_x) = M_t^2(f_x) + M_n^2(f_x)$$

provides better intuition on the visual thresholding process. In the case of white noise, where the spectrum is simply a constant at all spatial frequencies, the noise component M_n can not be accounted for with a constant. There is centain frequency masking effect which is modeled by Barten as:

$$M_n(f_x) = \sqrt{\frac{2\Phi_n}{XYT}}$$

and

$$\frac{1}{X^2} = \frac{1}{X_o^2} + \frac{1}{X_e^2} + \frac{f_x^2}{N_e^2}$$

Another dimension of the masking effect on the noise is the noise with frequency characteristics, either lowpass of $\frac{1}{f}$ affects the contrast sensitivity differently from white noise. van Meeteren and Valenton experiments on the noises at different lowpass frequencies, namely the fine noise which is white noise, the medium noise which is stopped at 20 cycles per degree, and the coarse noise which is stopped at 5 cycles per degree, provides another strong suggestion on the masking effect of the noise. That is the effect of the lowpassed noise on the contrast sensitivity is much less pronounced at the spatial frequencies beyond the cutt off frequency of the lowpass filter.

Acknowledgement

I would like to express my sincere gratitude to Dr. Brian H. Tsou, my laboratory focal point at AL/CFHV, for his advice and assistance throughout the research. I would also like to thank Dr. Scott S. Grigsby of Logicon Technical Services, Inc., who provides a great deal of help in setting up the experiments.

References

[Barten92] P.G.J. Barten, "Physical model for the Contrast Sensitivity of the Human Eye," SPIE Vol. 1666 Human Vision, Visual Processing, and Digital Display III, 1992, pp. 57-72.

[Daly89] S. Daly, "Application of a Noise Adaptive Contrast Sensitivity Function to Image Data Compression," SPIE Vol. 1077 Human Vision, Visual Processing, and Digital Display, 1989, pp. 217-227.

[Driggers90] R.G. Driggers, C.E. Halford, W. Pickard, and K. Williams, "Effects of Additive Noise on Four Bar Target Detection As a Function of Contrast and Spatial Frequency," Optical Engineering, Vol. 29, No. 2, 1990, pp. 110-113.

[Holst90] G.C. Holst and A.R. Taylor, "What Eye Model Should We Use for MRT Testing?" SPIE Vol. 1309, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing, 1990, pp. 67-75.

[Kornfeld71] G.H. Kornfeld and W.R. Lawson, "Visual-Perception Models," Journal of the Optical Society of America, Vol. 61, No. 6, 1971, pp. 811-820.

[Pollehn70] H. Pollehn and H. Roehrig, "Effect of Noise on the Modulation Transfer Function of the Visual Channel," Journal of the Optical Society of America, Vol. 60, No. 6, 1970, pp. 842-848.

[Ratches75] J. Ratches, et. al., "Night Vision Laboratory Static Performance Model for Thermal Viewing Systems," ECOM-7043, April 1975.

[Scott90] L.B. Scott and L.R. Condiff, "C2NVEO Advanced FLIR Systems Performance Model," SPIE Vol. 1309, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing, 1990, pp. 168-180.

[vanMeeteren88] A. van Meeteren and J.M. Valeton, Journal of the Optical Society of America A, Vol. 9, No. 5, 1988, pp. 438-444.

AN EVALUATIVE COMPARISON OF TECHNIQUES FOR MEASURING STUDENT SYSTEM KNOWLEDGE OF AVIONICS TROUBLESHOOTING

Nancy J. Cooke
Assistant Professor
Department of Psychology

Anna L. Rowe[1], Tracy L. Halgren, & Jeff A. Bauhs
Graduate Students
Department of Psychology

New Mexico State University
P.O. Box 30001/3452
Las Cruces, NM 88003

Final Report for:
Summer Research Extension Program
Armstrong Laboratory

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

and

New Mexico State University

December 1993

[1] Anna L. Rowe is a graduate student in the Psychology Department at Rice University, Houston, TX.

AN EVALUATIVE COMPARISON OF TECHNIQUES FOR MEASURING STUDENT SYSTEM KNOWLEDGE OF AVIONICS TROUBLESHOOTING

Nancy J. Cooke
Assistant Professor
Department of Psychology
New Mexico State University

Anna L. Rowe
Graduate Student
Department of Psychology
Rice University

Tracy L. Halgren & Jeff A. Bauhs
Graduate Students
Department of Psychology
New Mexico State University

Abstract

Intelligent tutors have the potential to enhance training in avionics troubleshooting by giving students more experience with specific problems. Part of the success of these tutors will depend on their ability to assess and diagnose the student's knowledge in order to direct pedagogical interventions. The broad goal of the research program described here is to develop a methodology for assessment and diagnosis of student knowledge in the domain of avionics troubleshooting. One step towards this goal involves the identification and evaluation of techniques for measuring students' system knowledge in the domain of avionics troubleshooting. This step is the focus of the research described in this report. Thus far, the results of this research program have (1) developed a means of identifying meaningful action patterns in on-line student-tutor interactions, (2) identified techniques for measuring students' system knowledge which predict troubleshooting performance, and (3) described how these techniques contribute to on-line student-tutor assessment.

AN EVALUATIVE COMPARISON OF TECHNIQUES FOR MEASURING STUDENT SYSTEM KNOWLEDGE OF AVIONICS TROUBLESHOOTING

Nancy J. Cooke, Anna L. Rowe, Tracy L. Halgren & Jeff A. Bauhs

INTRODUCTION

As tasks become more cognitively complex and demand more specialized skill, training issues are increasingly critical. The domain of avionics troubleshooting is a good example of such a task. Recent research (e.g., Nichols, Pokorny, Jones, Gott, & Alley, 1989) suggests that computerized intelligent tutoring systems may successfully supplement traditional training in complex tasks such as avionics troubleshooting where problems are ill-defined. Intelligent tutoring systems enable individuals to spend time learning a skill in a one-on-one environment in which a computer takes on the role of a human tutor. One goal of intelligent tutoring systems is to incorporate more individualized instruction based on a detailed assessment of student knowledge and diagnosis of cognitive strengths and weaknesses. Instructional intervention can then be directed at these strengths and weaknesses. The purpose of the work described here is to develop a methodology for the assessment and diagnosis of student knowledge in the domain of avionics troubleshooting.

The problem of assessment and diagnosis for intelligent tutoring systems has been approached in a number of ways. Many approaches focus on student actions to assess and diagnose student knowledge. For example, student actions may be examined for the identification of incorrect actions (e.g., Burton, 1982; Gitomer, 1992; Pokorny & Gott, 1992; Stevens, Collins, & Goldin, 1979), or student actions may be compared to a logically created ideal action set (e.g., Anderson, Boyle, & Reiser, 1985). One problem with these approaches is that the action data are impoverished relative to the much richer data obtained from cognitive methods such as verbal reports and structural analyses. These richer methods go beyond the student's actions to reveal the cognitive processes and knowledge underlying those actions. Such information is particularly critical for the identification of the knowledge base in complex, ill-specified domains like avionics troubleshooting. Furthermore, such information is important in the assessment and diagnosis of students' system knowledge. These issues are discussed in detail below.

Unlike the typical intelligent tutoring domains of algebra, geometry, and computer programming, avionics troubleshooting is ill-specified and is not associated with a well-documented body of knowledge. In these types of domains, before ideal models or buggy rules can be specified, it is first necessary to identify the specific knowledge necessary to perform the task. One cognitive task analysis (Hall, Gott, & Pokorny, 1990) of the avionics troubleshooting

domain has indicated that three types of knowledge are relevant for successful troubleshooting performance: (1) system (or how it works) knowledge, (2) strategic (or how-to-decide-what-to-do-and-when) knowledge, and (3) procedural (or how-to-do-it) knowledge (Gott, 1989). Evidence exists that suggests that system knowledge may be the most critical of these three (Gitomer, 1984; Glaser, et al., 1985; Hall, Gott, & Pokorny, 1990), although this point is not without controversy (Kieras, 1988, Rouse & Morris, 1986). Thus, the methodology developed in this research program focuses on the assessment and diagnosis of system knowledge.

Observations of student actions can reveal information about procedural and strategic knowledge, however, these observations are not likely to disclose system knowledge. Instead, relatively rich cognitive techniques have been used in the past to elicit this type of mental model knowledge. These techniques can be classified into four categories: (1) accuracy and time measures, (2) interviews, (3) process tracing/protocol analysis, and (4) structural analysis (Cooke & Rowe, 1993). Measurement methods drawn from each of these categories are likely to have advantages and disadvantages (Cooke, 1992), and no one method of measuring mental models has received universal acceptance. The different measurement approaches may each provide different sorts of information and have seldom been evaluated in terms of their respective reliabilities and validities. In short, the selection of a single optimal method for on-line student assessment of system knowledge is an uncertain enterprise at best. In this paper a pragmatic approach is taken in which optimal methods are minimally assumed to elicit system knowledge that is relevant to task performance.

Although rich cognitive methods seem better suited for measuring system knowledge than the action-oriented assessment approaches described earlier, they are not well-suited for on-line measurement. Instead, these methods typically involve the collection of "extra" data (e.g., verbal reports, similarity ratings) not typically collected during interactions with the tutor. Consequently, the use of these methods would entail interruption of the tutoring process to collect data in a task that would likely seem artificial to the student. What is needed is not only a sound method for measuring system knowledge, but one that can provide rich representations of this knowledge from student actions derived on-line. Such a methodology is the overall goal of our research program. The goal is to be able to map student actions (both errorful and correct) collected on-line onto a rich representation of student system knowledge. This representation can then be used to assess and diagnose student system knowledge and identify targets for intervention.

The Broad Plan: Mapping Student Actions onto System Knowledge

The overall goal of our research program involves making detailed inferences about a student's system knowledge from that student's actions. Four steps make such an inference

possible (Cooke & Rowe, 1993). The first step involves working backwards from the goal state--system knowledge, to the initial state--student actions. Interviews, process tracing, and structural analytic methods offer rich representations of system knowledge. However, it is necessary to know which of these methods provides the best measure of system knowledge in the domain of avionics troubleshooting (see Figure 1, Step 1). Therefore, the first step in reaching the overall goal involves identifying a valid method for eliciting and representing system knowledge required for avionics troubleshooting. Assuming that system knowledge is critical for performance, then a valid method of measuring this knowledge should reveal differences among subjects that correspond to performance differences.

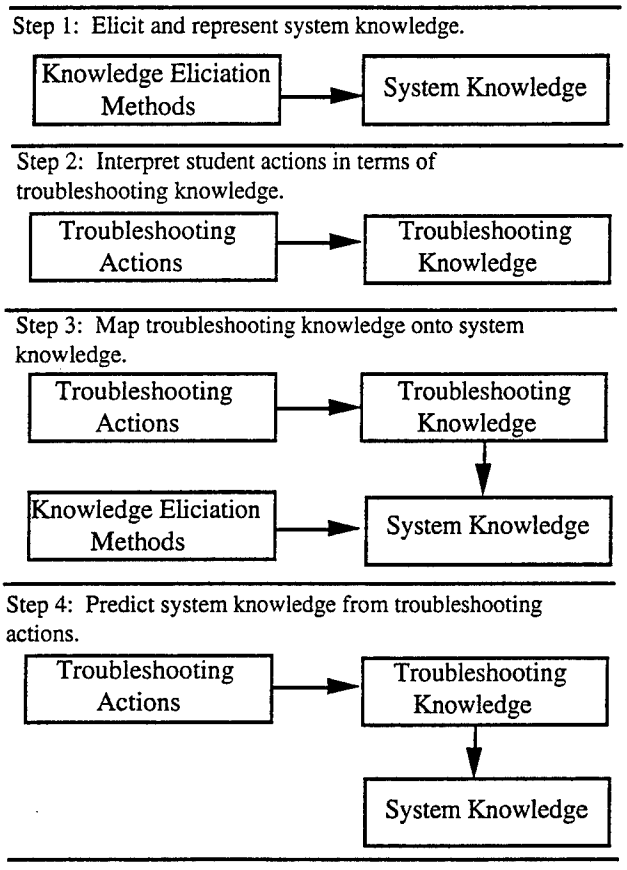


Figure 1. Steps involved in mapping student actions onto system knowledge.

As mentioned above these techniques require data collected off-line. Therefore, the next step involves determining how to derive this type of data from on-line interactions with the tutor. What is needed is a method designed to make use of action data collected on-line to derive representations of system knowledge. In other words a method is needed for identifying general relationships between student action patterns and patterns of system knowledge derived off-line,

so that later predictions can be made about system knowledge based on student actions. Thus, the identification of action patterns and the evaluation of the meaningfulness of these patterns is the second step (see Figure 1, Step 2). Along this line of reasoning, the third step entails mapping these action patterns onto patterns of system knowledge (see Figure 1, Step 3). Here, the goal is the identification of patterns of actions that correspond to distinct representations of system knowledge. Of course this step requires the elicitation of both actions and system knowledge from the same subjects. Assuming that Step 1 results in meaningful representations of system knowledge and assuming that system knowledge underlies actions (at least partially), then some correspondence should emerge. Finally, if this correspondence does emerge, then it would be possible to make predictions about system knowledge from troubleshooting actions collected on-line, thereby eliminating the extra data collection step (see Figure 1, Step 4). The four steps represented in Figure 1 comprise the long-term plan associated with the development of a new approach for assessing and diagnosing student system knowledge.

The contributions of this broad research plan are as follows:

1. A procedure for on-line assessment and diagnosis of student's system knowledge which involves mapping action patterns onto deficits or proficiencies in system knowledge.
2. A procedure which summarizes actions (errorful and correct) in terms of a rich representation of student knowledge that lends itself to qualitative analysis useful for diagnosis and intervention.
3. An assessment and diagnosis procedure that targets the complex domain of avionics troubleshooting.
4. A methodology that can be applied to the problem of eliciting knowledge from subject matter experts for tutor development.
5. A general test of the assumption that system knowledge underlies troubleshooting actions.

Previous Work

Research pertinent to the second step (identifying meaningful action patterns) was completed by Cooke and Rowe (1993). This particular step was carried out first because the data required were already available. Cooke and Rowe (1993) examined verbal troubleshooting action data collected by Nichols et al. (1989). The Pathfinder network scaling procedure (Schvaneveldt, 1990) was used to summarize the actions executed by subjects on a set of verbal troubleshooting tests. The results are promising in that they indicate that meaningful action patterns can be identified using this procedure. Specifically, the network patterns were predictive of troubleshooting performance (r (22)= .57), where performance was defined by a previous scoring method (Pokorny & Gott, 1992). Furthermore, using the network patterns,

Cooke and Rowe (1993) were able to differentiate tutor vs. no tutor groups in the Nichols et al. (1989) study. Finally, the networks revealed qualitative differences in the action sequences of high and low performers which were suggestive of potential intervention points. Thus, Cooke and Rowe (1993) identified a technique that captures meaningful action patterns during troubleshooting.

Current Work

The purpose of the research described in this report is the evaluation of different measures of system knowledge, primarily in terms of their ability to elicit system knowledge or mental model knowledge that is predictive of troubleshooting performance (i.e., the first step). The proposed research differs from other studies which have used multiple methods for measuring mental models in order to identify various cognitive factors underlying expertise (e.g., McCloskey, 1983; Gitomer, 1984) in that the primary interest here is a comparison of the measurement techniques themselves. In addition, this research represents an attempt to overcome some difficulties encountered with previous task analyses in the avionics troubleshooting domain by (1) embodying measurement techniques within a specific troubleshooting context, (2) employing new mental model measurement techniques including, pairwise comparisons, Pathfinder, and think aloud reports that have shown promising results in other domains, (3) using a performance-based criterion (i.e., verbal troubleshooting score) rather than relying on supervisor ratings, and (4) measuring performance on a continuous scale rather than a dichotomous scale in attempt to provide a more sensitive criterion.

The techniques were drawn from each of the major categories outlined above, excepting the time and accuracy measures category (i.e., interviews, process tracing/protocol analysis, and structural analysis). Specifically, subjects' mental models of an avionics system were measured using: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and think aloud while troubleshooting. All mental model measures took place within the context of a specific troubleshooting problem. In addition to completing the mental model measures, each subject worked to verbally troubleshoot the problem. Relating performance on each of the mental model measures to troubleshooting performance should offer insight into the strengths and weaknesses of each of the measures for accessing knowledge pertinent for performance.

METHOD

Problem Selection

All experimental materials were developed in the context of a particular troubleshooting problem in the flightline avionics or C Shop career field of the U.S. Air Force. The procedure

used to select this problem was designed for the selection of a moderately difficult troubleshooting problem. Such a problem would avoid the situation in which a technician might use a memorized procedure to solve the problem (i.e., the problem is too easy), and the situation in which a technician might be unable to solve the problem at all (i.e., the problem is too difficult). A moderately difficult troubleshooting problem presumably requires the invocation of a mental model for successful troubleshooting. This section describes the procedure used to select a moderately difficult problem. Much of the data used in the selection of this problem were gathered during Stages I-VIII of a PARI cognitive task analysis (Hall, Gott, & Pokorny, 1990) conducted by an Air Force research team in the C Shop career field (Hall, Pokorny, & Kane, 1993) at Eglin Air Force Base. The research team included two psychologists and two C Shop experts.

PARI (Precursor, Action, Result, Interpretation) is a cognitive task analysis methodology used by the Air Force as an integrated skill analysis/instructional development tool. The PARI data collection procedure consists of nine stages. In general, the first four stages serve to identify a sample of subject matter experts. These experts then assist the research team in identifying the general problem solving tasks encountered in the career field and the cognitive skills associated with successfully solving these tasks. The final five stages of PARI involve the collection of problem solving interview data from experts and novices and a set of follow-up reviews of the data. Hall, Pokorny, and Kane's (1993) PARI data were used in the selection of the moderately difficult troubleshooting problem. Data from Stage IX of this PARI analysis were not pertinent in the selection of a moderately difficult troubleshooting problem and thus were not utilized. The following paragraphs offer only a brief description of the PARI methodology; a more complete account of PARI can be found in Hall, Gott, and Pokorny (1990).

PARI-Stage I. The first stage of PARI is designed to identify subject matter experts who then go on to participate in the remaining PARI stages. In order to identify C Shop experts the Air Force research team conducted individual discussion sessions with technicians who had been identified as skilled by shop supervisors and were available for participation in the discussions. During the discussion session, the technician was asked to iteratively break down flightline avionics equipment systems (e.g., the Radar Warning Receiver or RWR system). First, the technician identified the subsystems of a particular system. The technician then broke the subsystems down to the component level, identifying the function of each component as it was named. This break-down continued until the technician believed that the components at the lowest level could not be further subdivided. In addition to iteratively breaking down the equipment systems, the technician addressed job training problems associated with the particular system.

Based on these discussions, the researchers, as a group, determined which of the sampled technicians qualified as experts. This determination was based on three aspects of each technician's discussion: 1) the quality of the verbalized equipment representations, 2) the identification of specific equipment component relations, and 3) the level of clarity in the technician's equipment descriptions. Selection was also based on availability to participate in the PARI sessions. Following the application of this process, the research team designated two of the sampled technicians as subject matter experts: an Air Force airman and an Air Force civil servant. These experts assisted the research team through the remaining stages of the cognitive task analysis.

PARI-Stage II. The second stage of PARI is designed to establish the training foci associated with the job in question. (PARI was developed to assist in the development of training that targets complex problems.) To form the C Shop training foci, the two experts worked to list and discuss the maintenance tasks (i.e., troubleshooting problems) they felt were difficult. These discussions were facilitated by a listing of cognitively complex maintenance tasks created by the Air Force research team from examinations of the results of occupational surveys and Air Force Specialty Training Standards. The Air Force research team used two related criteria to classify tasks as cognitively complex: the number of decisions required in performing the task and the stability of the task (or system) environment in which problem solving occurs. Tasks were considered cognitively complex if they required decision making (i.e., a procedure specifying step-by-step actions for solving the problem is not available) and if they occurred in a unstable environment (i.e., many factors must be considered in making decisions). Tasks meeting these two criteria were subsequently used by the research team to facilitate discussion with the two experts. The experts were asked specifically about any tasks they failed to name which had been identified as cognitively complex by the research team. If the experts agreed that the team's identified task was cognitively complex, it was included in the listing of complex tasks; if the experts did not view the task as cognitively complex, they discussed their reasoning for this assertion. The team and the expert then decided together whether or not to categorize the task as cognitively complex. Thus, the research team and the experts worked together to identify the cognitively complex tasks associated with the C Shop.

PARI-Stage III. The purpose of the third stage of PARI is the generation and consolidation of the problem types encountered by technicians working on the job. Using the cognitively complex tasks identified in Stage II, the experts worked to generate an exhaustive list of the equipment malfunctions (and their causes) that could initiate troubleshooting for these tasks. The experts independently specified fault instances in cause and effect language (e.g., "bad stimulus routing caused by a stuck relay") for each of the defined tasks. The experts then worked

together to consolidate the identified system causes and effects into meaningful categories (e.g., wiring faults). Fault instances were grouped together if they demanded similar knowledge and skills for solution. This grouping resulted in the following typology of problems which could initiate troubleshooting in the C Shop: set up procedure faults, switchology faults, cable faults, wiring faults, and electrical/component faults.

PARI-Stage IV. The fourth stage of PARI involves the development of representative troubleshooting problems for each of the problem categories specified in the problem typology. The typology serves to ensure that representative examples of all problem types are generated. The experts individually developed a representative troubleshooting problem for each of the five problem categories. For each of these problems the experts: (1) developed an overview or problem description which listed the fault location and the symptoms associated with the fault, (2) generated a problem statement that listed the system conditions and symptoms for presentation to other individuals for troubleshooting, and (3) listed the supporting technical documentation (e.g., test procedures, schematics) that would be required by others troubleshooting the problem. At the conclusion of PARI-Stage IV, each of the experts had designed one troubleshooting problem for each problem category.

PARI-Stage V. The purpose of the fifth stage of PARI is the anticipation of all possible solution paths for the developed problems. Thus, the experts worked individually to generate their own solutions to the problems they had developed. The PARI problem-solving structure was used to guide solution generation: Actions, Precursors to Actions, Results, and Interpretations of results were elicited. Following the recording of the initial solution, five "rehashes" were conducted in which the expert worked to 1) verify the initial solution, 2) generate alternative results and result interpretations for each step in the solution, 3) identify and evaluate alternative actions for each step in the solution, 4) name alternative appropriate equipment targets (precursors), given the previously executed steps, and 5) group the actions that seem to go together and to explain the basis for these groupings. At the conclusion of PARI-Stage V, the experts had each generated one solution for each of their respective problems.

PARI-Stage VI. During the sixth stage of PARI, experts naive to the problems are asked to generate solutions to those problems. Thus, two additional technicians were asked to produce solutions to the generated problems. These technicians had been identified by the two experts as skilled technicians, and both were at the 7-Level. The Air Force uses a three level classification system to designate expertise: all technicians start at the 3-level and move from the 5- to the 7-level after passing certain shop criteria (e.g., familiarity with certain equipment systems) as well as written tests. Thus, technicians at the 7-level have reached the highest level of this classification system.

The two technicians worked individually with one of the experts to solve the generated troubleshooting problems. The experts presented only the problems they had developed. Thus, the experts worked individually with each of the technicians. The expert began by presenting the problem statement to the technician (e.g., "During debrief, the crew chief reports an ASP 44"). The technician then worked to isolate the fault and repair the equipment through a series of iterative action-result steps. In each step the technician specified an action and the reason for taking that particular action. The expert responded by informing the technician of the action's effect on the equipment, and requested the technician's inference concerning equipment operation based on that effect. If the technician strayed from the active path during troubleshooting and did not return in a timely manner, the expert presenter provided coaching. The action-result cycle continued until the problem was solved. This procedure was followed for each problem in the problem set.

PARI-Stage VII. During the seventh stage of PARI, the expert problem developers review the problem set to determine the adequacy with which the developed problems assess the cognitive skills and knowledge required for skilled performance. The two expert developers reviewed the problems and ascertained that they were representative of problems encountered in the actual job environment. In addition, they determined that solutions to the problems required the types of cognitive skills and knowledge associated with skilled performance. In addition to making these judgments, one of the experts rank-ordered the troubleshooting problems in terms of their difficulty (see Table 1). His rank-ordering was based on his observations of the technicians troubleshooting the problems, in addition to his 12 years of experience in the C Shop. The two additional expert technicians who had worked to solve the developed problems were unable to solve the following three problems without coaching from the expert presenter: ICS Frequency Holes, ASP44, and CMD Safety Switch. Thus, these three problems were considered more than moderately difficult and were dropped from consideration, leaving the RWR-RF Loss problem as the most likely candidate for problem selection.

PARI-Stage VIII. During the eighth stage of PARI, less-skilled technicians are observed as they work to generate solutions for the developed problems. Thus, intermediate and novice technicians were asked to troubleshoot the remaining problems, using the PARI problem solving structure described above. One of the expert problem developers presented the problems he had developed to the technicians. The remaining problems were presented by an expert from the Air Force research team (the expert problem developer was unavailable). Observations of the troubleshooting behavior of these technicians revealed that novice technicians could isolate the fault associated with RWR--RF Loss problem quickly. In addition, they could easily solve the

Troubleshooting Problem Ranking
1. ICS Frequency Holes
2. RWR--RF Loss
3. ASP 44
4. CMD Safety Switch to Scope
5. RWR Wiring--Broken Wire
6. AAI--Cable at Radar
7. IFF Seat Switch
8. KY-58 Relay Panel
9. 51 & 52 Switches--AAI
10. Have Quick Word of Day

Table 1. Rank ordering of troubleshooting problems from most (1) to least (10) difficult offered by one of the expert problem developers.

problem if they had previous experience with a specific cable measurement tool, although this tool was only just being introduced to the C Shop and was not in common use. Because the less-skilled technicians could isolate the fault relatively quickly, the RWR--RF Loss problem was deemed inappropriate for selection as the moderately difficult troubleshooting problem. Observations of technicians solving the RWR Wiring--Broken Wire problem indicated that this problem revealed performance differences across technicians of different skill levels. Thus, this problem was selected for use as the troubleshooting problem in this study.

The RWR Wiring--Broken Wire Problem

The system important in troubleshooting the RWR Wiring--Broken Wire problem is the Radar Warning Receiver (RWR) system. The RWR is part of the F-15 Tactical Electronic Warfare System (TEWS) and is designed to detect, analyze, and identify threat radar signals. The RWR also controls countermeasure responses to those threats (e.g., the release of chafe).

The RWR System Components. The RWR system consists of eleven components (see Table 2; Gouley, 1992). A functional block diagram of the RWR system components is illustrated in Figure 2. In this section, the general functions of each of the RWR system components are discussed. First, the LRU2 (Line Replaceable Unit) provides the RWR system with its power requirements. Specifically, the LRU2 receives 115 VAC 3 Phase 400 HZ and +28 VDC power from the aircraft and converts it to the DC voltages needed by the RWR. In addition, the LRU2 stores the flight program required for normal operation of the RWR. Next, the LRU3, in conjunction with the low-band antenna, provides the aircraft with coverage in the low-band frequency range. In addition, the LRU3 processes high- and low-band receptions,

distributes data between components of the RWR, and interfaces the RWR with other avionics systems. The LRU6, in conjunction with the four high-band antennas (left wing tip, right wing tip, left fin, and right fin), provides the aircraft with omni-directional coverage in the high-band frequency range. The LRU9 provides all displays for the TEWS. It is located in the cockpit and provides the pilot with a view of the threat environment relative to the aircraft. The LRU10 is a control panel which allows the pilot to select on or off for the following systems: RWR, Internal Countermeasure Set (ICS), and Electronic Warfare Warning System (EWWS). The LRU11 is another control panel which allows the pilot to select RWR/ICS combat/training mode, mode of the TEWS pod, and ICS mode of operation.

Fault Location. The problem statement for the RWR Wiring--Broken Wire is listed in Table 3. This problem is caused by a shorted video cable between the LRU3 and the LRU9. Specifically, the video cable between LRU3-2J8 and LRU9-9J1 is shorted. Center pin 1 on this cable is shorted to (touching) the shield on the LRU3 end, resulting in a loss of data (see Figure 3). When functioning correctly, this video cable transmits data from the LRU3 to the LRU9. The transmitted data contains information about the type and placement of symbols on the LRU9's display. However, due to the short between pin 1 and the shield, the data to be transmitted to the LRU9 is lost, and the LRU9 is completely blank.

Component Name
LRU 2 (Power Supply)
LRU 3 (Low Band Receiver Processor)
LRU 6 (High Band Receiver)
LRU 9 (TEWS Display)
LRU 10 (TEWS control panel)
LRU 11 (TEWS immediate action control panel)
Left Fin Antenna
Right Fin Antenna
Left Wing Tip Antenna
Right Wing Tip Antenna
Low-Band Antenna

Table 2. RWR System Components

Based on the problem statement (Table 3), technicians troubleshooting this problem may examine the following areas as possible causes of the problem: LRU9, LRU3, LRU2, or aircraft wiring. Without properly diagnosing the symptoms, the technician may leave the active signal

path. For instance, the technician may perceive that the scope is blank because there is no power to the system. Such a perception may lead to an investigation of the circuit breaker panel, TEWS control panel, or the LRU2. A correct diagnosis, in addition to knowledge of the visual cues, can lead the technician to fault code 9911H1CA (available to technicians as part of the Technical Order or T.O.). This fault code lists the procedure to be followed given the following symptoms: no cross symbol or azimuth dots on TEWS display when power is applied to the RWR. The fault isolation (F.I.) tree for this fault code instructs the technician to check the data cable between the LRU3 and LRU9. Here, it is important that the technician has knowledge about the correct testing procedure of this cable. The F.I. only specifies to test for continuity across the center conductors from end-to-end for this cable. However, in order to detect the fault, the technician must thoroughly test the cable, which would include checking for continuity from center conductor to shield. Such a test would reveal that there is continuity here, indicating a short.

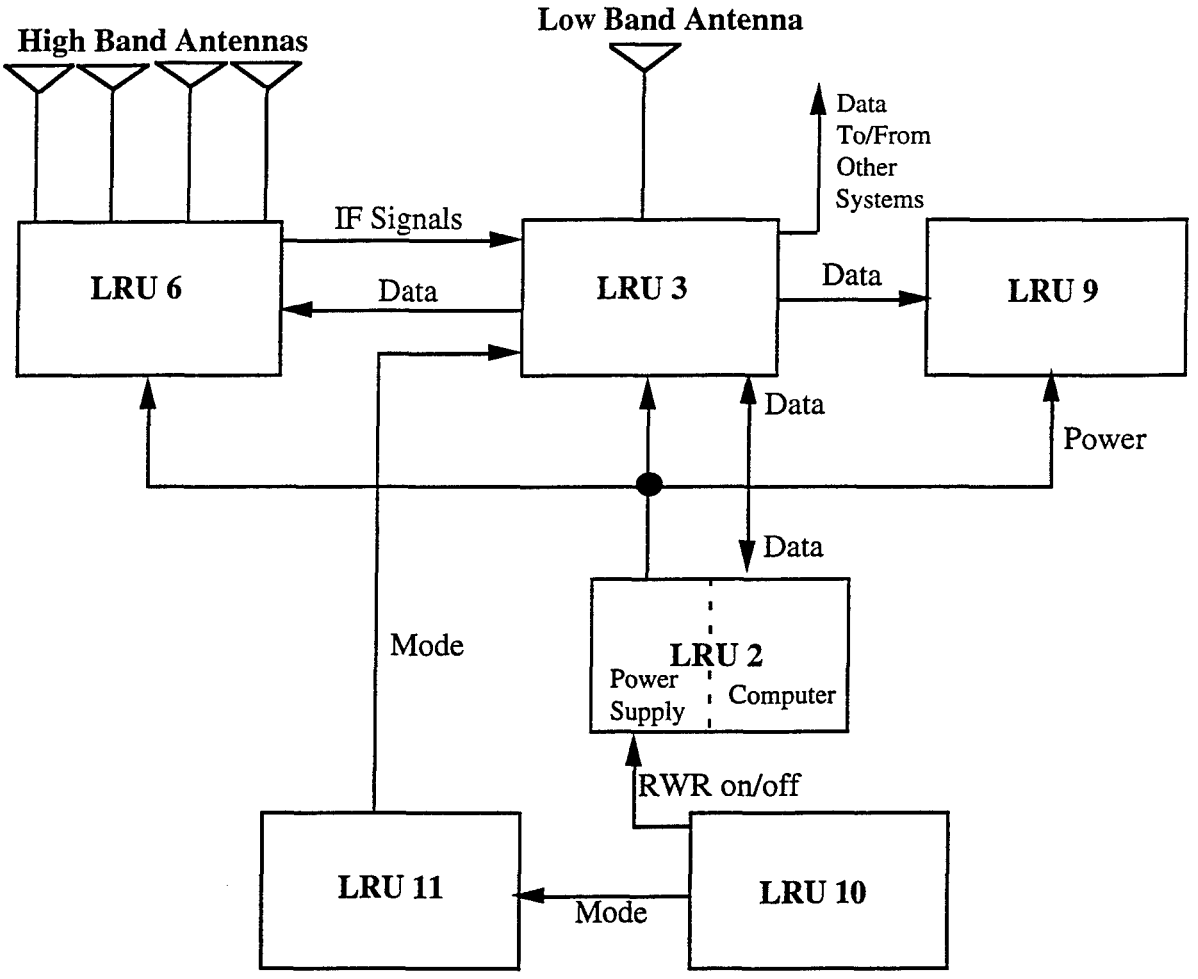


Figure 2. Functional block diagram of the RWR system.

"In debrief, the pilot reports that the RWR is inoperative, the BIT (built-in test) light is on, and the TEWS display is blank."

Table 3. Problem Statement for the RWR Wiring--Broken Wire Troubleshooting Problem.

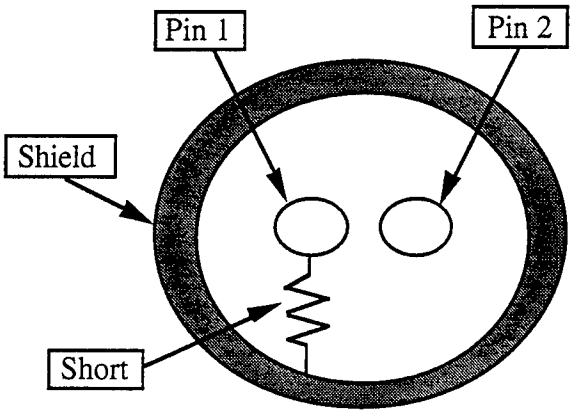


Figure 3. Cross-sectional view of the LRU3 end of the video cable between LRU3-2J8 and LRU9-9J1, in the context of the RWR Wiring--Broken Wire problem. Center pin 1 is shorted to the shield.

Subjects

Subjects were nineteen technicians in the flightline avionics (C Shop) career field of the Air Force. Each subject had been through a technical training school designed to prepare them for their specialty and had received subsequent on-the-job training. Technicians working in the C Shop are responsible for the identification, isolation, and repair of airborne avionics equipment systems, including the Radar Warning Receiver (RWR) and Identification--Friend or Foe (IFF). Technicians who could be spared from the shop to participate in the study were selected within the Air Force three-level system so that a range of proficiency levels could be achieved. Thus, six 3-levels, eight 5-levels, and five 7-levels participated.

Materials and Procedure

Technicians' mental models of an avionics system (i.e., the RWR) were measured using four techniques: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and think aloud while troubleshooting. Technicians' knowledge was also measured using a retrospective interview, but these data were collected for the purposes of the Air Force research team and will not be discussed here. All mental model measures took place within the context of the RWR Wiring--Broken Wire troubleshooting problem discussed earlier. (The RWR system is the avionics system important for troubleshooting this problem.) In

addition to completing all of the mental model measures, each technician worked to verbally troubleshoot the RWR Wiring--Broken Wire problem.

Laddering Structured Interview. Upon arrival to the testing session, the technician completed the laddering structured interview. This interview consisted of four steps. The problem statement for the troubleshooting problem was presented (see Table 3) to the technician. The first step of the interview consisted of asking the technician to identify the major system important in troubleshooting this problem. In Step 2 the technician was asked to name the major components of the identified system, in the context of the troubleshooting problem. The third step of the interview then commenced in which the technician was asked to name all of the major components of the identified system, regardless of problem context. In the fourth and final step the technician was asked to name the major systems with which the identified system interfaced, if any. Throughout these steps an index card with the identified component or system name written on it was prepared and placed before the technician, according to the arrangement specified by the technician. The index cards only served as memory aids for the technician and were discarded after the identified systems and components had been recorded.

Relatedness Ratings. Following the laddering structured interview, the technician completed familiarity ratings (i.e., 1 = highly familiar to 6 = unfamiliar) on the eleven RWR system components (see Table 2). The components were presented in a random order. The technician then completed relatedness ratings on all pairs of the same eleven RWR system components. The technician was told to make all relatedness ratings within the context of the RWR Wiring--Broken Wire troubleshooting problem. The technician was told to rate the items in terms of their functional relations. Furthermore, the technician was told that although two items can be functionally related in a number of different ways (e.g., by information flow through the system, by performing similar function), ratings should be made based on the first general impression of functional relatedness of the concepts, within the context of the troubleshooting problem.

Both the familiarity and relatedness ratings were completed on a computer; a HyperCard program was used to collect these data. The technician rated the relatedness of the component pairs by using a mouse to point and click on one of five sections on a bar, with the endpoints labeled "slightly related" and "highly related". If the technician wished to rate the component pairs as unrelated, a button labeled "unrelated" was available. A similar procedure was used to collect the familiarity ratings. The RWR Wiring--Broken Wire troubleshooting problem statement (see Table 3) was available for review at the top of the computer screen throughout the relatedness ratings task. Presentation of pairs was randomized across subjects, and order of items within pairs was counterbalanced.

Diagramming. After completing the ratings, the technician completed a diagramming task, using the component set just rated. Randomly ordered index cards with a component name printed on each were given to the technician, as was a set of directional and bi-directional white arrows. The technician was instructed to arrange and connect the components in a manner representing the actual function of the RWR system, in general. The technician specified directionality of relations with the unidirectional and bi-directional arrows. To illustrate the use of the arrows for representing functional relations, the technician was given an example diagram with an accompanying explanation from the domain of automobile engines (see Figure 4).

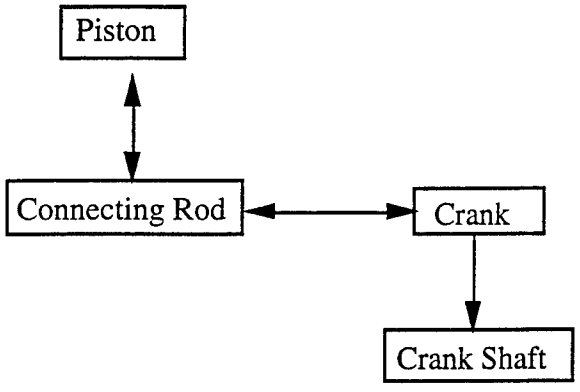


Figure 4. Example functional diagram from the domain of automobile engines.

After the technician completed the functional diagram representing the RWR system at a general level, the RWR Wiring--Broken Wire troubleshooting problem statement (see Table 3) was re-presented. The technician was then given a set of directional and bi-directional yellow arrows and was asked to use these yellow arrows to designate those components and/or connections most important in troubleshooting the RWR Wiring--Broken Wire problem. Finally, the technician was asked to explain, in his/her own words, both diagrams. These explanation data were collected to aid the examiner's understanding of the generated diagrams, and no further analyses were conducted on these data.

Think Aloud. The technician then proceeded to the troubleshooting/think aloud portion of the experiment. The technician was told that he/she would be verbally troubleshooting the problem used in each of the previous tasks, the goal being to isolate the fault and repair the equipment. The technician was instructed to think aloud continuously while working to solve the problem, verbally expressing all thoughts. Two practice think-aloud problems were then reviewed with the technician to ensure that the technician understood what was meant by thinking aloud (i.e., what is the result of multiplying 24 by 6, and how many windows are there in your house?). If the technician had difficulties during these practice problems (e.g., did not

speak), he/she was guided to think aloud. After successfully completing these practice think-aloud problems, the verbal troubleshooting session began.

The examiner re-presented the RWR Wiring--Broken Wire problem statement (see Table 3). All technical materials necessary for troubleshooting the problem were available. These materials included the C Shop Job Guide (J.G.) and the T.O. which contains fault isolation trees and schematic diagrams. The technician was instructed that the goal was to isolate the fault and repair the equipment through a series of iterative action-result steps. In addition, the technician was reminded to verbally express all thoughts while working to solve the problem. The technician began by specifying the first action he/she would take in troubleshooting the problem (e.g., check the ASP). The expert responded by informing the technician of the action's result (e.g., ASPs 5 and 49 latched). The technician then specified the next action, the expert gave the corresponding result, and so on. This action-result cycle continued until the problem was solved, the 45-minute time limit expired, or the technician gave up. All of the technician's, the expert's, and the examiner's responses were recorded with an audio tape recorder.

After completing the troubleshooting/think aloud task, the examiner and the technician reviewed the protocol action by action. The technician was asked to provide a retrospective report of: (1) why each action was taken, and (2) the information provided by the corresponding result of each action, in terms of the equipment in question. This retrospective interview was conducted for other purposes and will not be further discussed here.

Questionnaires. Following the troubleshooting action review, the technician completed two questionnaires. First, the technician completed a Likert-style questionnaire designed to allow an evaluation of the mental model measurement techniques. Specifically, the technician used a 6-point scale to rate each of the techniques on the following dimensions: difficulty (difficulty--easy), similarity to actual troubleshooting in the shop (different--similar), range of responses available (restricted--broad), realism relative to actual troubleshooting in the shop (artificial--realistic), and usefulness for measuring system knowledge (useless--useful). Space was available for additional written comments.

The technician was then given a questionnaire designed to allow a comparison of the measurement techniques. All pairs of tasks (measurement techniques and troubleshooting task) were presented on this questionnaire. The technician was asked to make pairwise judgments of the tasks by circling the task in the presented pair which best measured knowledge needed for actual troubleshooting of the RWR Wiring--Broken Wiring problem in the shop. Immediately after circling one of the tasks in the presented pair, the technician used an 8-point scale to rate the similarity of the knowledge measured by the circled task to knowledge needed for the actual troubleshooting of a problem in the shop. The endpoints of the scale were labeled "Not at all

similar" and "Extremely similar." Upon completion of this questionnaire, the technician was debriefed and excused. Separate from the testing session, a supervisor rating questionnaire was completed for each technician by his/her respective supervisor.

DATA ANALYSIS

Overview of Analyses

The data were analyzed to evaluate the various mental model measurement techniques in terms of their abilities to predict troubleshooting performance. The "bottom line" in real-world troubleshooting situations is performance. Thus, comparisons of the results obtained from the measurement techniques with the results from the troubleshooting task should provide a pragmatic means of assessing the validity of the techniques. This approach is not straightforward however, because the results of each of the techniques offer only information regarding the content of technicians' knowledge, whereas performance is generally indicated in terms of reference to some ideal or perfect score. Therefore, what is needed is an assessment of the technicians' knowledge. Such an assessment requires an ideal or "gold standard." In general, the gold standard for each measure was based on that measure applied to a group of four high performers (i.e., technicians who had the highest verbal troubleshooting scores). The group's combined results comprised the gold standard for that measure. In cases in which there was little agreement across high performers, the group was limited to those who agreed. The performance criterion was based on the outcome of subjects' performance on the verbal troubleshooting task. In the following section, the method used to generate a troubleshooting performance score for each technician (i.e., the criterion value) is described.

Troubleshooting Performance

Two subject matter experts independently scored the technicians' troubleshooting action protocols. These experts had participated in the troubleshooting problem development stage. In addition, one of the experts (Expert B) assisted the examiner in problem presentation when the technicians worked to troubleshoot the problem. The protocols given to the experts for scoring included the actions executed by the individual technicians, along with their corresponding results. The experts were instructed to score the protocols based only on the actions which were listed. They were told to use a 100-point scale (100 = perfect). In addition, the experts were asked to rank-order the protocols from best (1) to worst (19). Figures 5a and 5b represent the troubleshooting performance score distributions for Experts A and B, respectively.

An examination of the Figure 5 reveals that the two experts used the scale differently in scoring the troubleshooting protocols. Expert A's scores ranged from 0 to 100, whereas Expert B's scores ranged from 65 to 100. However, the scores awarded by the two experts were

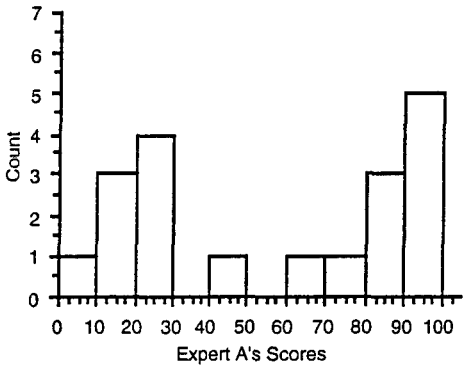


Figure 5a

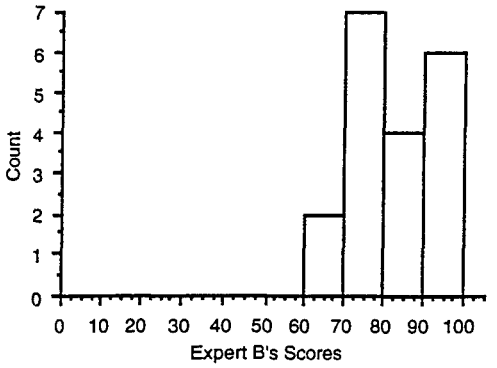


Figure 5b

Figure 5. Troubleshooting performance score distributions for Expert A (Figure 5a) and Expert B (Figure 5b).

significantly correlated[2] $r(17) = .883$, $p < .0001$, as were the rank-orderings given by the experts, Spearman's $r(17) = .847$, $p < .0005$. Thus, a single performance score was created for each technician by averaging the troubleshooting scores given by the two experts (see Figure 6). This score was then used as the performance score in all subsequent analyses. Four high performers emerged from this analysis who had performance scores one standard deviation greater than the mean (i.e., all scores 95 or above). These four technicians were used to create the gold standard for comparison in the remaining analyses.

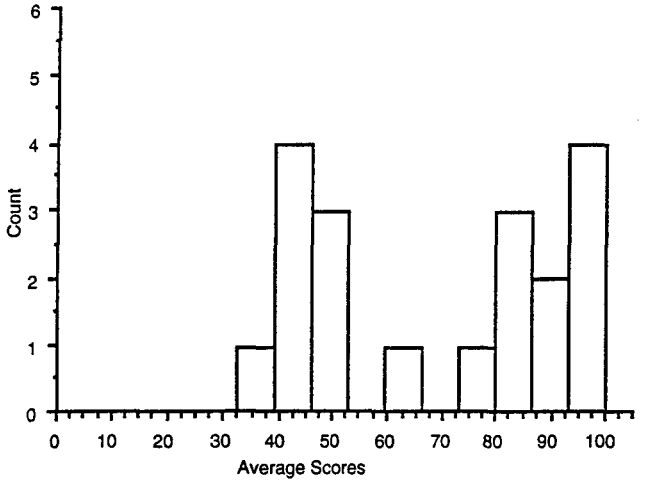


Figure 6. Frequency distribution of the average troubleshooting performance score awarded to each technician.

[2] All correlations reported in this paper are Pearson product moment correlations unless otherwise specified. In addition, because correlations between performance and knowledge were expected in particular directions, they were tested for significance using one-tailed probabilities.

RESULTS AND DISCUSSION

Laddering Structured Interview

To determine the level of inter-subject agreement among the four high performers, the proportion of shared items across lists for each pair of high performers was calculated for each step in the interview (see Table 4). That is, the ratio of shared items to the total number of different items listed was calculated for each step for all pairs of high performers. The resulting proportions indicate that the high performers agreed on the important components or systems for each step, particularly Steps 1 and 2. All high performers listed the RWR as the system important in troubleshooting the RWR Wiring--Broken Wire problem (Step 1). In addition, for Step 2 each of the high performers named at least three of the same components (i.e., the LRU2, LRU3, and LRU9) as important components in troubleshooting this problem.

		Technician Number		
	Step	6	8	14
5	Step 1	1.0	1.0	1.0
	Step 2	1.0	.75	.50
	Step 3	.67	.63	.71
	Step 4	.50	.67	1.0
6	Step 1		1.0	1.0
	Step 2		.67	.50
	Step 3		.63	.60
	Step 4		.40	.50
8	Step 1			1.0
	Step 2			.50
	Step 3			.67
	Step 4			.67

Table 4. The proportion of shared list items for pairs of high performers for each interview step.

A gold standard component/system list was created from the lists of these four high performers for each of the four steps (see Table 5). Items named by at least one of the high performers were included in the list. Knowledge indices for the 15 remaining technicians were then calculated in terms of the proportion of items shared with the gold standard list associated

with each step of the interview. The resulting knowledge indices for each step were correlated with the troubleshooting performance score, excepting Step 1 (see Table 6). At Step 1, all technicians named the RWR as the system important in troubleshooting the problem, indicating that all sampled technicians correctly interpreted the presented symptoms at a coarse level. Correlations conducted on the remaining three steps indicated that naming components important for troubleshooting the problem (Step 2) was predictive of troubleshooting performance, r (13) = .542, $p < .05$. This positive relationship indicates that good troubleshooters agreed with the high performers on the components important for troubleshooting the problem, whereas poor troubleshooters did not. The data resulting from Steps 3 and 4 (i.e., name all components regardless of context, and name interfacing systems) were not predictive of troubleshooting performance (see Table 6).

In addition to examining the proportion of shared components/systems, the number of items listed by technicians, but not by the high performers, was calculated for each interview step for each technician. These errors of commission made up a second set of knowledge indices which were then correlated with troubleshooting performance. Again, data from Step 1 were not included in these correlations because all technicians named the RWR as the important system in

Laddering Interview Steps			
Step 1	Step 2	Step 3	Step 4
RWR	LRU2	LRU6	EWWS
	LRU3	LRU10	ICMS
	LRU6	LRU11	Blanker
	LRU9	Left Wing Antenna	CC
	LRU10	Right Wing Antenna	CMD
	Aircraft Wiring	Left Fin Antenna	
		Right Fin Antenna	
		Low Band Antenna	
		Antenna Cables	
		ASP	

Table 5. Gold standard lists for each interview step. Note: The Step 3 list includes items named by individual technicians as major components of the identified system, regardless of problem context, in addition to those items named by that technician in Step 2.

troubleshooting the problem, and no errors of commission were made. Of the remaining three steps, only errors of commission occurring during Step 2 of the interview were predictive of

troubleshooting performance, r (13) = .505, $p < .05$. This positive correlation indicates that technicians who identified extra RWR system components important in troubleshooting the problem (not included in the gold standard list) were better troubleshooters than were technicians who did not. The correlations resulting from the remaining steps were not significant (See Table 6).

	Proportion of Shared Items	Errors of Commission
Step 2	.542	.505
Step 3	-.128	-.326
Step 4	-.200	-.296

Table 6. Correlations between troubleshooting performance and knowledge indices: 1) the proportion of shared items with the gold standard list and 2) errors of commission for Steps 2-4 of the laddering interview.

In general, the second step of the laddering structured interview is significantly related to troubleshooting performance. Those technicians who listed more components that were shared with the gold standard list performed the troubleshooting task better than those who listed fewer gold standard components. Unexpectedly, those technicians who also listed more components not on the gold standard list were also better troubleshooters than those who listed fewer. This latter result, although unexpected, corroborates earlier findings in this area. Specifically, Cooke and Rowe (1993) found that as students gained troubleshooting experience, they tended to execute a greater number of actions (and even more than high performers), although they did not seem to know when the actions should be applied. In a similar way, the best troubleshooters in this study believed that many components were relevant for troubleshooting the problem, including those which are actually relevant. Perhaps, early stages of the development of expertise can be characterized by a familiarity with many components and procedures, whereas the mapping of those components and procedures to a particular troubleshooting situation is a hallmark of later stages of expertise.

Finally, it is interesting that the laddering technique was predictive of troubleshooting performance only in the context of the troubleshooting problem. Lists of the general system components or interfacing components were not predictive. Interestingly, Gitomer (1984) also used a laddering structured interview in this domain, however he did not restrict the interview to a particular problem context. Instead, subjects were told to think of a specific LRU and to iteratively break this LRU down into its components. This lack of context may explain why

Gitomer did not observe a difference between skilled and less-skilled airmen (as defined by supervisor ratings) in the laddering technique.

Relatedness Ratings

Correlations of relatedness ratings for each pair of the four high performers were computed to determine degree of inter-subject agreement. These correlations are presented in Table 7. Note that the correlations are all high and statistically reliable ($p < .05$ with 54 degrees of freedom).

	Technician Number		
	6	8	14
5	.757	.769	.889
6		.662	.751
8			.899

Table 7. Intercorrelations of relatedness ratings for the four high performers.

In order to generate a graphical summary of the ratings, the data were submitted to the Pathfinder network scaling procedure, a descriptive multivariate statistical technique that represents pairwise proximity in a network form (Schvaneveldt, 1990). In the networks, concepts (or components in this case) are represented as nodes, and relations (functional relations in this case) are represented as links between nodes. This network representation can convey much more information about structure and relatedness than can be seen in the ratings themselves. For more detail on Pathfinder see Schvaneveldt (1990) or Cooke and Rowe (1993). The C statistic (Goldsmith & Davenport, 1990), a measure of shared links for matching nodes across two different networks, was calculated between each of the four high performers (Table 8). This measure ranges from 0 (low similarity) to 1 (high similarity) and can be viewed as a measure of association between two networks. The C values in Table 8 indicate that technician #8 shares fewer links with the other three high performers than the other three share with each other. Therefore the gold standard was computed in several ways: (1) averaging ratings of all four high performers, (2) using ratings of technician #8 only, and (3) averaging the ratings of all high performers, excluding #8. The resulting Pathfinder network based on the data from the second gold standard is presented in Figure 7a.

The knowledge index for each of the 15 technicians was based on the C value between the individual technician's network and each of the three gold standards described above. For comparison purposes Figure 7b presents the Pathfinder network based on the mean relatedness ratings of the eight lowest performers (i.e., troubleshooting score less than 50). Correlations were then computed between knowledge indices and troubleshooting scores to determine which,

	Technician Number		
	6	8	14
5	.887	.418	.868
6		.423	.759
8			.404

Table 8. C values for each pair of the four high performers' networks.

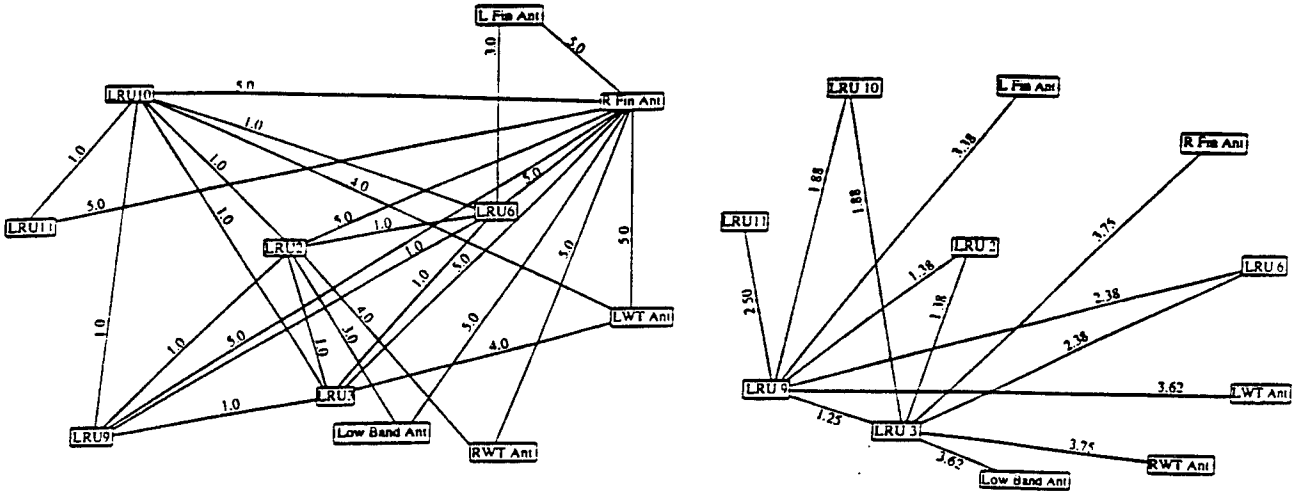


Figure 7a Figure 7b

Figure 7. Pathfinder networks (r = ∞, q = n-1) based on: a) the data from technician #8 - the second gold standard and b) the mean relatedness ratings given by the eight lowest performers (score < 50). Note: 1 = highly related, 6 = unrelated.

if any, of these three knowledge indices predicted performance. Correlations were respectively, .399, .527, and .425 for the first, second, and third gold standards described above. With 13 degrees of freedom the only predictive gold standard was the one based on technician #8 alone. A partial correlation (r (12) = .405, p = .06), although only marginally significant, suggested that technician #8 was predictive of performance even when the variance accounted for by the remaining three high performers (i.e., the third gold standard) was partialled out. On the other hand, the correlation between performance and the gold standard based on all four high performers (i.e., first gold standard) dropped from .399 to -.034 when the variance due to technician #8 was partialled out.

These results indicate that relatedness ratings, coupled with the Pathfinder network analysis procedure, are only marginally predictive of troubleshooting performance when all high

performers are used as the gold standard. More specifically, however, these results suggest that one of the high performers, technician #8, generated relatedness ratings that were significantly predictive of performance, independent of the remaining high performers. The C values for the high performers, in combination with the marginally significant partial correlation, indicate that there is little overlap between the ratings given by technician #8 and those given by the other three high performers. Interestingly, there are some other differences between technician #8 and the other high performers. That is, #8 is a 5-level, whereas the others are all 7-levels. Also, #8 has spent only four years in the C shop, whereas the others have spent between 6 and 8 years in the C shop. However, #8 performed as well as the other high performers in the troubleshooting task, receiving a score of 95 compared to the other three scores of 95, 97, and 100. Possible explanations for this pattern of results are discussed in the General Discussion section.

Diagramming Task

The following analyses are based on the general diagrams of the RWR system. The diagrams that were specific to the RWR Wiring--Broken Wire problem were not informative because the majority of the technicians deemed that only three to five system components were relevant to this problem. Each of the 19 technician's system diagrams was converted to an 11 by 11 asymmetric matrix, with ones representing the presence of a connection between components and zeros indicating no connection. To determine the level of diagram similarity among the four high performers the proportion of shared connections for each pair of diagrams was computed. This proportion was based on the number of links shared by the two diagrams divided by the total number of links in the union of the two diagrams. These proportions are presented in Table 9. In general, pairs of high performers shared about half of the links present in the two diagrams. However, closer inspection of Table 9 indicates that once again, the diagram of technician #8 shared the least with those of the other high performers (mean proportion of .42 for technician #8 compared to .68 for other pairs of high performers). For this reason, three gold standard diagrams were created parallel to the three gold standards for relatedness ratings: (1) a diagram based on all four high performers, (2) technician #8's diagram, and (3) a diagram based on all high performers excluding #8. The matrix representing the group diagrams (i.e., gold standards one and three) consisted of ones, indicating that the connection existed in at least one diagram, and zeros otherwise. Again these matrices were asymmetric.

Three diagramming knowledge indices were generated for each of the remaining 15 technicians by subtracting the technician's matrices from each of the three gold standard matrices and summing the absolute values of the differences. These indices should be zero if there was complete agreement with the gold standards. The correlations between these knowledge indices and trouble shooting performance were -.464, -.440, and .089 for gold standard one, two,

	Technician Number		
	6	8	14
5	.55	.48	.85
6		.39	.65
8			.40

Table 9. Proportion of shared connections for pairs of high performers' diagrams.

and three, respectively. The first two correlations are significant (p = .041 and .051, respectively) with 13 degrees of freedom. Note that a negative correlation is expected given that large knowledge indices indicate large diagram differences. The low correlation between gold standard three and performance, as well as a nonsignificant correlation of -.181 when technician # 8's diagram is partialled out of the first gold standard, suggests that the predictability of the high performers in this technique can be attributed to technician #8. In fact, when the three other high performers are partialled out of the correlation between technician #8 and performance, the resulting correlation is high and significant (r (12) = -.538, p= .02). Thus, as for relatedness ratings, technician #8 is more predictive of technician performance than the other three high performers. The diagram completed by technician # 8 is presented in Figure 8. Interestingly, using #8 as the gold standard, errors of omission (r (13) = -.508, p =.026) are better predictors of performance than errors of commission (r (13) = -.272, p=.167).

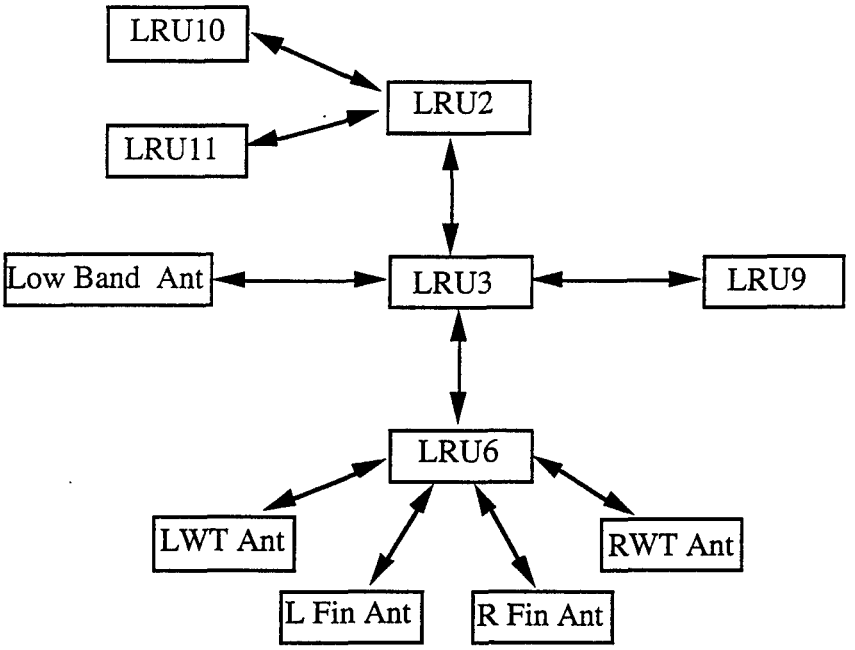


Figure 8. RWR system diagram created by technician # 8.

In general, the diagramming technique, at least in the context of the overall system, predicted troubleshooting performance well. Most of this predictability can be attributed to technician #8, the same technician whose relatedness ratings predicted performance. In terms of the diagramming technique, technician #8 differed from the other high performers in two major ways: (1) #8 used only bi-directional arrows, and (2) based on technicians' diagram explanations, #8 attempted to represent both information flow and power flow in his diagram, whereas the other high performers represented only information flow in their diagrams. It is also interesting that errors of omission should be predictive as opposed to commission. This result suggests that seeing "extra" system relations is not as problematic for troubleshooting as is failing to see one or more critical relations.

Think Aloud

A coding scheme for technicians' verbalizations in the think-aloud-while-troubleshooting technique was developed. The purpose of the coding scheme was to be able to classify verbalizations into discrete meaningful units that could be represented as nodes in a Pathfinder network. The main groups of verbalizations included: (1) action interpretation/explanation, (2) result interpretation, (3) component elimination, (4) elimination justification, (5) plan/prepare for test/check, and (6) Technical Order (T.O.) search/interpretation. The most abstract level of categorization was used for each category except for the action interpretation/explanation category. This category was broken down into sub-units because the type of action interpretation/explanation seemed important in distinguishing skill levels. For example, if a technician checked the fuses on the LRU2 (a power check) and verbalized an explanation other than a power explanation, this information would likely distinguish different skill levels and should be captured. The resulting coding scheme consisted of 22 verbalization units/nodes. Each technician's verbal protocol was coded by two raters. The inter-rater reliability achieved on the recoding of the action interpretation/explanation category was acceptable, with 92.6% agreement on 149 coded explanation/interpretation verbalizations. The raters discussed the 11 verbalizations on which they disagreed, and a compromise was met.

Also included in the coding scheme were meaningful troubleshooting actions that technicians executed. These were included in order to provide context for the verbalizations. Meaningful was defined as actions indicative of skill in troubleshooting the presented problem. For example, checking the LRU9 ETI meter (a power indicator) is indicative of troubleshooting skill for the RWR Wiring--Broken Wire problem because power to the LRU9 may be the cause of the problem. On the other hand, checking the LRU6 ETI meter is not because problems with the LRU6 are not indicated by the problem statement. The main groups of actions included: (1) debriefing questions, (2) equipment checks, (3) continuity tests, and (4) swaps. The most

abstract level of categorization that indicated troubleshooting skillfulness was used. Using this decision rule, an action unit was associated with either poor or good troubleshooting actions. The resulting coding scheme consisted of 75 action units/nodes. Two raters coded 5 of the action protocols together. They then coded the remaining 14 action protocols and achieved an acceptable level of reliability of 98.1% with 267 coded actions. The raters discussed the five actions on which they disagreed, and a compromise was met. The entire coding scheme (verbalizations and actions) included 97 events/nodes (22 verbalizations + 75 actions). The major coding categories and an example of each are presented in Table 10.

Verbalizations	
	• Action Interpretation / Explanation
	• "I want to see if those two are shorted to each other."
	• Result Interpretation
	• "That's the problem there; Pin 1 is shorted to the shield."
	• Component Elimination
	• "I wouldn't go to the circuit breaker panels."
	• Justification for Component Elimination
	• "Obviously he had some kind of power to it."
	• Plan / Prepare for Test or Check
	• "I'm going to look up to see what fault code they recommend."
	• T.O. Search / Interpretation
	• "It says if you've got a fault indicator 49 orange, which is the ASPS and TEWS display indicator is black and white, it gives you a first time failure, a second time failure, and a ground check-out failure."
Actions	
	• Debriefing Questions
	• Ask pilot: Did you have a flashing AI/SAM light?
	• Equipment Checks
	• Check the ASP
	• Continuity Tests
	• Check continuity between 65P-D002H (LRU3) and 65P-J009A (LRU9), pin 1 to shield.
	• Swaps
	• Swap the LRU3

Table 10. Major coding categories in the verbal protocol and action coding schemes. A representative example from each category is listed.

Transition probabilities for all event pairs (verbalizations and actions) were calculated for individual subjects by dividing the frequency with which specific event transitions (e.g., T.O. search/interpretation followed by continuity check between LRU2 and LRU3) occurred by the frequency with which the first event in the sequence occurred. For example, if T.O. search/interpretation occurred twice and was followed by continuity check between LRU2 and LRU3 on one of those occasions, then the transition probability would be 0.5. Note that these are first-order transitions only. Higher order transitions were not used because the immediate transitions were considered to be the most meaningful for this task. Also note that event transitions convey order information, and thus each event pair can be associated with two distinct transition probabilities.

Transition probabilities were calculated for each of the four high performers. Agreement among these technicians was assessed by correlating these probabilities for all pairs of high performers. These correlations are presented in Table 11. Note that all inter-technician correlations are low and are not statistically significant. However, because there is no single technician who seems different from the rest, the transition frequencies of all four technicians were combined as the gold standard. The gold standard transition matrix was then submitted to the Pathfinder network scaling algorithm, as were transition matrices of each of the individual technicians (Schvaneveldt, 1990). Similarity measures (i.e., C values as used for relatedness ratings) among pairs of networks for high performers corroborated the correlation results, with a mean inter-technician C of .06. This low agreement suggests that high performers conveyed very different thoughts in their verbalizations.

	Technician Number		
	6	8	14
5	.05	.15	.06
6		.06	.01
8			.03

Table 11. Correlations of transition probability matrices for pairs of high performers.

C values between the gold standard Pathfinder network and each individual technician served as a knowledge index. This index was not predictive of troubleshooting performance ($r(13) = -.026$). Note that this is not a matter of idiosyncrasies among the high performers in that gold standards based on any one of the four are not significantly predictive of performance (correlations of performance and knowledge indices based on individuals are -.325, -.005, -.133, and -.139). The highest observed correlation (between technician #5 and performance, $r = -.325$)

is in the opposite direction, indicating that the best troubleshooters verbalized in a manner least like that of technician #5.

An additional knowledge index was derived from a correlation of event frequencies (i.e., the frequency with which each discrete verbalization or action occurred) associated with an individual's protocol and event frequencies associated with the gold standard. Thus, this measure should be high to the extent that the technician exhibited the same verbalizations and performed the same actions as the high-performers, the same number of times. It should overlap with the Pathfinder network similarity measure in that they both take shared events into account. However, these two measures are divergent in that the Pathfinder measure includes information on event sequences, whereas the action frequency measure does not. Conversely, the frequency measure includes frequency of individual events, whereas Pathfinder does not. This measure (r (13) = -.394, p =.07) was slightly more predictive than the network similarity measure, though not significant. However, this correlation is negative, indicating that technicians with verbalization/action frequencies similar to that of the gold standard had lower troubleshooting scores.

In summary, the think aloud technique resulted in low levels of agreement among the four high performers and was not predictive of troubleshooting performance. The marginally significant negative correlation between the frequency knowledge index and performance suggests that the better technicians say different things at different rates than high performers. Specifically, high performers tended to say less than the other technicians. However, of the 15 technicians, those who performed well tended to read the Technical Orders and thus generate more verbal statements. Together these results seem to indicate that the think aloud technique does not assess knowledge that is critical for performance. Technicians may be unaware of much of the knowledge that underlies their troubleshooting performance and thus, when asked to think aloud, they verbalize thoughts that are independent of task performance.

Questionnaires

Each of the questions on the Likert-scale questionnaire was analyzed using a repeated measures ANOVA with the four techniques (i.e., laddering, ratings, diagramming, and think aloud) making up the four levels of the independent variable. Two of the five questions resulted in no significant technique effect. These questions had to do with similarity to actual troubleshooting and usefulness as a measure of system knowledge. The other three questions resulted in significant technique effect. For the task difficulty question, there was a significant effect of technique (F (3, 54) = 3.5, p = .02), that can be attributed to the fact that the laddering technique was rated as significantly easier than each of the other three techniques. Respective means for laddering, ratings, diagramming, and think aloud were 5.3, 4.2, 4.3, and 4.2. For the

question about restriction in range of responses, there was also a significant technique effect (F (3, 54) = 6.5, p = .003). The ratings technique was viewed as significantly more restrictive than laddering and think aloud techniques. Respective means for laddering, ratings, diagramming, and think aloud were 4.5, 3.6, 4.4, and 5.0. For the task artificiality question there was a significant technique effect (F (3, 54) = 3.42, p = .02), with the think aloud and laddering techniques receiving greater realism ratings than the other two tasks. Respective means for laddering, ratings, diagramming, and think aloud were 4.0, 3.4, 3.4, and 4.6. Also, the 19 technicians, on average, viewed system knowledge as extremely important (mean rating of 4.3 on a 5-point scale, SD = .65) for troubleshooting the RWR Wiring--Broken Wire problem.

On the second questionnaire technicians circled the technique in each pair of techniques which they felt best measured the knowledge necessary for troubleshooting. Percent responses indicated that technicians viewed the diagramming (39%) and think aloud (33%) techniques as better in this regard than the laddering (12%) and rating (16%) techniques.

Supervisor ratings were to have been collected for each of the 19 technicians. However, ratings for technicians from one of the participating fighter squadrons were unavailable. Six technicians are members of this squadron. Thus, ratings were collected for only 13 of the technicians. Due to the missing data, supervisor ratings were not analyzed.

GENERAL DISCUSSION

Summary and Recommendations for Measuring Mental Models

The purpose of this project was to identify one or more techniques suitable for measuring mental models of the type relevant to avionics troubleshooting. Measures were taken in the context of a specific troubleshooting problem. Care was taken in the selection of this problem to insure that it was (1) representative of the domain (i.e., C shop avionics), and (2) that it was of intermediate difficulty, presumably making the invocation of a system mental model necessary for successful troubleshooting. In addition to completing the mental model measures, each technician worked to verbally troubleshoot the problem. Comparing the results of the knowledge measures to troubleshooting performance should provide a pragmatic means of assessing the validity of the measures. Of the four techniques tested, all but the think-aloud technique were predictive of troubleshooting performance. Whether the think-aloud gold standard was constructed from individual high performers or groupings of these high performers, predictability remained low. On the other hand, the laddering (Step 2), relatedness ratings, and diagramming techniques were all predictive of troubleshooting performance. Also, when technician #8 was used as the gold standard in the ratings and diagramming techniques, predictability was optimal. An explanation for this is discussed below.

Partial correlations indicated that two of the techniques were each independently predictive of troubleshooting performance. Specifically, the relatedness ratings technique was predictive of performance independent of both the laddering technique (r (12) = .461, p = .042) and the diagramming technique (r (12) = .471, p = .039). Similarly, the laddering technique was predictive of performance independent of the ratings technique (r (12) = .480, p = .035) and to a lesser extent, the diagramming technique (r (12) = .421, p = .057). Interestingly, the diagramming technique was not predictive independent of the relatedness ratings (r (12) = -.362, p = .09) or the laddering (r (12) = -.246, p = .18) techniques.

Based on the results of this research, several recommendations can be made regarding system knowledge measurement techniques. First, because the relatedness ratings and laddering techniques are independent predictors of performance, it is recommended that these two methods be used to measure system knowledge in the domain of avionics troubleshooting. Interestingly, this recommendation is counter to the technicians' ratings which indicated that they believed the diagramming and think-aloud techniques were the best measures of system knowledge.

Second, based on the subjective ratings of technicians, as well as the procedural features of these two techniques, it is recommended that the laddering technique be used in cases in which both ratings and laddering cannot be used. Specifically, technicians thought that the ratings technique was the most restrictive in terms of response freedom, and that it lacked realism. Procedurally, relatedness ratings are also restrictive in the sense that presenting all pairs of a concept set quickly leads to an unmanageable number of pairs as the number of concepts increases, making the ratings technique nearly impossible to use with large concept sets. Even with a smaller number of concepts (around 20), the pairwise ratings task seems quite long and tedious to subjects. On the other hand, the laddering technique appears to be easier to implement. The laddering technique requires less background knowledge on the part of the researcher than does the ratings technique, especially when ratings are followed by Pathfinder analyses. In addition, technicians rated the laddering technique as realistic and easy to complete. However, there is a tradeoff to be considered. The ratings technique, while relatively more difficult to implement and analyze than the laddering technique, provides a graphical representation of knowledge that is much richer than the list produced using the laddering technique. Furthermore, the problems associated with the ratings technique are less important when the ultimate goal of our research program is considered: the on-line prediction of knowledge via action patterns. That is, once a correspondence between distinct representations of system knowledge and distinct action patterns has been identified, the knowledge elicitation step may be bypassed.

Findings on Measuring Mental Models

In the course of completing this research we noted several interesting observations relevant to mental model measurement and expertise in general. The first, and perhaps most important, observation dealt with defining the "gold standard." We found that a single technician, #8, contributed to most or all of the predictability associated with the high performers in the ratings and diagramming techniques. As mentioned earlier, technician #8 differed from the other high performers on two dimensions: years of experience and level obtained in the Air Force classification system. This technician may be a very good "intermediate" level technician rather than an "expert" level technician. Perhaps technician #8 was a better predictor than the other high performers, because they were too far removed from the entry-level technicians in terms of expertise, whereas #8 was not. Thus, although technician #8 and the other three high performers each performed well on the troubleshooting problem, their performance on the knowledge measures was different. Perhaps technician #8's knowledge was most like that of a good intermediate-level technician. Using experts as the gold standard assumes a linear relationship between the development of expert performance levels and the development of knowledge. Perhaps major qualitative changes in knowledge occur as expertise develops, making expert knowledge a poor predictor of novice performance.

Technician #8 was not predictive for two of the techniques. First, technician #8 was not predictive for the think-aloud technique. This lack of predictability may be more indicative of the think-aloud technique rather than using technician #8 as a gold standard. Verbalizing thoughts appeared to be problematic for the technicians. For example, some of the more experienced technicians did not verbalize, while other technicians simply read the T.O. out loud. In general, information that was verbalized was highly variable, and little of it seemed directly related to task performance. Second, technician #8 was not predictive for the laddering technique. Here, technician #8 did not perform differently from the other high performers. Perhaps the laddering technique taps into very basic system knowledge about existing components which may not evolve with expertise to the same extent as does knowledge about component interrelations which is tapped by the ratings and diagramming techniques.

A second interesting observation deals with the general context used in mental model measurement techniques. We found conflicting results regarding the use of a troubleshooting context in measuring system knowledge. For the laddering technique, we observed that providing a problem context was useful. Predicting troubleshooting performance was greatest when components relevant to the troubleshooting problem were listed (Step 2). However, listing all system components regardless of problem context (Step 3) was not predictive of troubleshooting performance. Providing a problem context in the diagramming technique

resulted in diagrams which were less predictive of performance than were diagrams constructed without restriction to the specific troubleshooting problem. This diagramming finding could be due to the nature of the troubleshooting problem. Specifically, there were very few diagram components that were relevant to the problem. It may also be that task order is responsible for these different problem-context effects. Context was helpful in the laddering technique where the context-specific questions preceded the general ones and not in the diagramming technique where the general questions came first.

Finally, this research appears to support a pattern of development in troubleshooting expertise observed by Cooke and Rowe (1993). They found that, after training, low performers exhibited a wide range of troubleshooting actions. A portion of these actions were executed by high performers, and a portion were not. These results suggest that prior to achieving expertise, but after some experience, technicians have an extensive array of knowledge in the form of executable actions, but they do not know when these actions apply. Results from the current study support this conclusion. The best technicians shared laddering components with the gold standard, but they also made more errors of commission. The fact that commission errors are predictive of performance indicates that the acquisition of this "extra" knowledge may, in fact, be a necessary stage in the development of expertise. Also, in the diagramming technique technicians with fewer omissions relative to the gold standard were more likely to be better troubleshooters. Again, the better troubleshooters may have a wide range of information, a subset of which is information considered critical by experts.

Interestingly, this phenomenon is also illustrated by a comparison of the network generated from technician #8's ratings and the aggregate network created from the eight lowest performers' ratings (see Figures 7a and 7b). First, the low performers focused exclusively on two system components: the LRU9 and the LRU3. Notice that all connections involve one or both of these components. These two components are also central in the network of technician #8, however, other appropriate components (e.g., the LRU2) are also central. Technician #8's network also includes many more links (including many weak ones, link weights > 3.0) than the low performers' network. Thus, with experience, it appears that technicians not only learn more executable actions, but technicians also appear to learn more interrelations among system components.

Future Directions

In this project two effective methods of measuring system knowledge were identified. Each of these measures produced results that corresponded to troubleshooting performance. In previous work (Cooke & Rowe, 1993) meaningful action patterns that corresponded to troubleshooting performance were also identified. Also, note that Cooke and Rowe's (1993)

results were replicated here with data collected for the RWR Wiring--Broken Wire troubleshooting problem. Event transition probabilities based on the coding scheme for actions only (see Think Aloud results) were generated and submitted to Pathfinder analyses. An action network from a subset of three high performers who agreed with each other was predictive of technician performance (r (13) = .415, p = .06). Thus, technicians' troubleshooting performance can be reliably predicted from relatedness rating and laddering measures as well as from action sequences.

Although the research completed thus far is valuable on its own (i.e., in terms of the identification of several techniques that can be used to predict performance off-line), the ultimate goal of our research program is a method for predicting knowledge during on-line tutor interactions. Thus, the next step toward this goal is to map action patterns onto knowledge measures. In other words, can patterns of actions be identified that correspond with specific mental models of the system? This mapping of actions to knowledge will enable technicians' mental models or deficits in system knowledge to be identified without requiring technicians to complete the laddering or ratings tasks. Instead, by relating action patterns to knowledge patterns, knowledge can then be predicted based only on action sequences.

Future work will involve collecting knowledge and action measures on the same group of subjects and identifying categories of action and knowledge patterns (i.e., types of mental models and associated actions). System or mental model knowledge can be predicted on-line to the extent that action categories map onto knowledge categories. Some analyses carried out on the data collected for this study are promising in this regard. Cluster analyses were performed on measures of inter-subject similarity for the action sequences as well as for the laddering (Step 2), ratings, and diagramming techniques. The action data resulted in one tight cluster of six subjects. If actions predict knowledge, then the technicians in this cluster who performed similar actions should also be clustered together in analyses of knowledge measures. Indeed, three of these six technicians were in the same cluster in the laddering (Step 2) and diagramming analyses, and five of the six were together in the ratings analysis. More work on matching actions with knowledge should not only reveal the extent of the match, but also the specific nature of the knowledge categories (or mental models) and action patterns that characterize the population of technicians. These issues and others remain for future research.

REFERENCES

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. Science, 228, 456-462.

Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 157-183). London: Academic Press.

Cooke, N.J. (1993). A Taxonomy and Evaluation of Knowledge Elicitation Techniques. Manuscript in preparation.

Cooke, N.J., & Rowe, A.L. (1993). An Approach to On-line Assessment and Diagnosis of Student Troubleshooting Knowledge. Air Force Technical Report, AL-TP-1993-0004. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Gitomer, D.H. (1984). A cognitive analysis of a complex troubleshooting task. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Gitomer, D.H. (1992). Foundations of HYDRIVE. Presentation given at the Basic Job Skills Progress Review, Brooks AFB, San Antonio, April 7-9.

Glaser, R., Lesgold, A., Lajoie, S., Eastman, R., Greenberg, L., Logan, D., Magone, M., Weiner, A., Wolf, R., & Yengo, L. (1985). Cognitive task analysis to enhance technical skills training and assessment. (Final Report to the Air Force Human Resources Laboratory on Contract No. F41689-8v3-C-0029.) Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization (pp. 75-87). Norwood, NJ: Ablex.

Gott, S. P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E. Z. Rothkopf (Ed.), Review of Research in Education, Vol 15, (pp. 97-169). Washington D.C.: American Educational Research Association.

Gouley, J.R.(1992) Integrated Avionics Communications, Navigation, and Penetration Aids System Specialist (F-15). Gunter Air Force Station, AL: Extension Course Institute Air University.

Hall, E.M., Gott, S.P., & Pokorny, R.A. (1990). A Procedural Guide to Cognitive Task Analysis: The PARI Methodology. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Hall, E.M., Pokorny, R.A., & Kane, R. (1993, in preparation). Identification of Instructional Goals: A Qualitative and Quantitative Analysis of Problem Solving Protocols. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Kieras, D.E. (1988). What mental model should be taught: Choosing instructional content for complex engineered systems. In J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 85-118). Hillsdale, NJ: Erlbaum.

McCloskey, M. (1983) Naive theories of motion. In D. Gentner & A.L. Stevens (Eds.), Mental Models (pp. 299-323). Hillsdale, NJ: Erlbaum.

Nichols, P.D., Pokorny, R., Jones, G., Gott, S.P., & Alley, W.E. (1989). Effective Instructional Processes Within an Intelligent Tutoring System. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division

Pokorny, R., & Gott, S. (1993, in preparation). The Evaluation of a Real-World Instructional System: Using Technical Experts as Raters. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.

Schvaneveldt, R. (Ed.), (1990). Pathfinder Associative Networks: Studies in Knowledge Organization . Norwood, NJ: Ablex.

Stevens, A., Collins, A., & Goldin, S.E. (1979). Misconceptions in student's understanding. International Journal of Man-Machine Studies, 11, 145-156.

An Experimental Investigation of Hand Torque Strength for Tightening Small Fasteners

S. Deivanayagam
Professor of Industrial Engineering
Tennessee Technological University
Cookeville, TN 38505

Final Report for:
Summer Extension Research Program
Armstrong Laboratory

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

December 1993

An Experimental Investigation of Hand Torque Strength for Tightening Small Fasteners

S. Deivanayagam
Professor of Industrial Engineering
Tennessee Technological University
Cookeville, TN 38505

Abstract

The human hand strength for tightening small fasteners was studied using an experimental approach. In many maintenance activities the technician often uses his/her hand to initially tighten small fasteners before using a wrench for final securing. Currently no information is available on the hand tightening strength for such activities. The experiment included small bolts, and flat key like fasteners. Three different orientations of the fasteners along with bare hands, gloved hands and small wrench tightening situations for a total of sixty different experimental conditions were studied. Young adults, twenty males and twenty females, participated in the study as volunteers and their torque strength capabilities under the sixty specific experimental conditions were collected. The results indicate significant effects due to the size and orientation of the fastener, the use of glove, and the gender of the person on the maximum torque applied. The data gathered will be useful for computerized man modelling of maintenance technicians.

An Experimental Investigation of Hand Strength for Tightening Small Fasteners

S. Deivanayagam

INTRODUCTION:

The operational effectiveness of a modern system very much depends on the maintainability of the system along with a number of other factors. Any system, no matter how well designed and made to perform, will require some maintenance and care during its useful life. How easy and how well the needed service can be provided will determine its operational readiness. Further, the maintainability will also impact the life-cycle cost of the system. Maintainability refers to the ease and the timely manner in which a system that is not ready to function can be put back into full operational condition. Reliability and maintainability are two major characteristics of a system determining its availability to perform its designed function.

Many modern weapon systems in practice have been found to have low maintainability resulting in lack of availability and high life-cycle cost. The technologically complex and highly sophisticated systems are notorious for such conditions. The primary reason for systems with low maintainability characteristics may be attributed to the insufficient considerations given to maintainability during the design stage of the system. Most often the lack of maintainability considerations is not due to any willful negligence on the part of the designer, but due to the lack of appropriate tools and techniques for systematic consideration of maintainability during the design stage.

The United States Air Force (USAF) has during the last decade experienced significant maintainability problems associated with its sophisticated weapon systems. With the objective of improving the state of affairs regarding the maintainability of its weapon systems the USAF has initiated several programs. One approach is to develop and use computerized man models representing the USAF technicians during early stages of system design so that the desirable maintainability characteristics of the system can be included before it is too late. CREW CHIEF and DEPTH are two such programs being developed by researchers at Armstrong Laboratory, Wright Patterson Air Force Base, Ohio for developing computerized man models of the USAF maintenance technician performing aircraft flight line maintenance activities (ref. 1). Both models require extensive data bases on human functional capabilities relating to various types of maintenance tasks. One such needed data base is on the human strength capabilities for tightening

and loosening - torquing - small fasteners directly with hands or using small wrenches under typical maintenance conditions. The models currently include only a limited amount of data regarding hand torquing. The author of this report spent the summer of 1992 at the Armstrong Laboratory as a Summer Faculty Researcher and conducted an experiment to collect some data in this regard (ref. 2) This report documents the extension research conducted at Tennessee Technological University (TTU) for the purpose of collecting more data to support the CREW CHIEF and DEPTH programs.

The results of the 1992 summer study indicated that while working with bare hands the torquing capability of the person was very much limited by the pain felt in the fingers and palm region due to the relatively small areas of contact on the soft tissues rather than by the inherent muscular strength capability. It is felt that the use of a glove may be beneficial to apply a higher torque by minimizing the feeling of pain. However, the use of gloves may itself have an effect on the person's ability to grip a small fastener by decreasing the dexterity. This research effort was undertaken to study and quantify the beneficial effects, if any, on the torquing capabilities of persons while tightening small fasteners as in typical maintenance activities.

OBJECTIVES:

This research was aimed at gathering human torque strength data while tightening small fasteners. Several conditions of typical maintenance task situations were simulated in the laboratory and human volunteer subjects performed the tasks while their torque strength values were collected. For the purposes of this study the following three general conditions were considered.

1. With bare hand, i.e., no other interface between the hand and the fastener.
2. With gloved hand, i.e., while wearing typical work gloves.
3. With small open-end wrenches, i.e., using a standard open-end wrench on the small fastener.

The above three general conditions were selected for the purpose of providing information for comparison and for the data base needed for the computer models. It was also decided to limit the scope of this to tightening actions by right handed subjects only. Torque strength for loosening actions and for left handed subjects, while important, were not considered so that the study would not become too extensive. For the purpose of this study the tightening action is defined as the application of clockwise torque on the fastener.

EXPERIMENTAL DESIGN:

Due the fact that there literally are dozens of different kinds of small fasteners it is almost impossible to include all of them in any study of limited resources. Therefore the following experimental design was developed in consultation with the USAF personnel for CREW CHIEF and DEPTH programs.

I. Hand-Fastener Interface -
- a. Without Gloves
- b. With Gloves
- c. With open-end wrenches
  - i. 3/8"
  - ii. 9/16"

The gloves used were regular leather work gloves, with rough exterior finish and no inside padding, available in most hardware stores. Five different sizes, shown in figure 1, were used to best fit the hand sizes of the individual subject. The open wrenches were the standard commercially available wrenches. Wrenches were used with bare hands (no gloves) only.

II. Type of Fastener -
- a. Bolt head
  - i. 3/8", six point
  - ii. 3/8", twelve point
  - iii. 9/16",six point
  - iv. 3/4", six point
- b. Flat Key
  - i. 1/2"long x 3/4"wide x 1/8"thick
  - ii. 1" long x 3/4"wide x 1/8"thick
  - iii. 2" long x 3/4"wide x 1/8" thick

The bolt heads used are parts of commercially available bolts used in aerospace systems as small fasteners. The six point and twelve point heads were selected to study any differences between the two on the torque strength. All four sizes were included in the 'with glove' and 'without glove' conditions. However only 3/8" and 9/16" six point sizes were used for the open-end wrench condition. The bolt heads for wrench conditions were milled out of solid high tension steel rods as it was found that the commercially available bolt shafts were not strong enough to withstand the high torque applied with the wrenches under certain conditions. The flat keys were the same ones used in the earlier study (ref 2.) conducted at Armstrong Laboratory. In addition to the three sizes mentioned, a larger key (3" long) was also used for benchmarking purposes only. Figure 2 shows the bolt heads and the flat keys used as fasteners for the non-wrench conditions of the study. The individual fasteners were welded to square metal plates to facilitate attachment to the torque dynamometer described later. Figure 3

Figure 1. Gloves used in the study

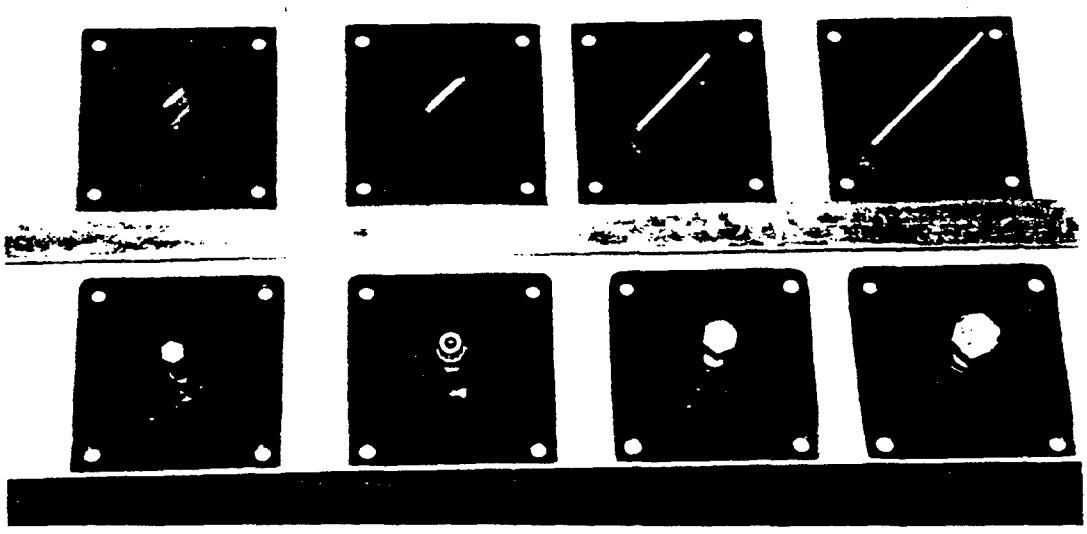


Figure 2. Bolt Heads (bottom) and Flat Keys (top)

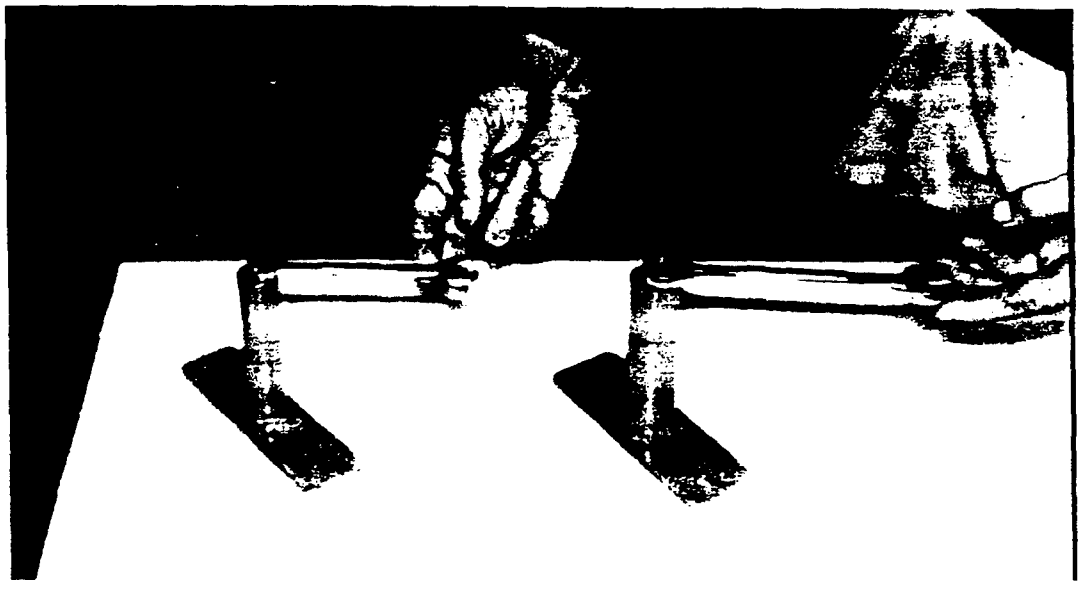


Figure 3. The 3/8" (left) and 9/16" (right) Bolt Heads with wrenches

shows the two bolt heads with wrenches attached as used for the wrench part of the study.

III. Orientation of the Fastener - The following three spatial orientations of the fasteners with respect to the subject were included.

a. Facing the subject
b. Transverse to the subject
c. Vertical, i.e., parallel to the subject

IV. Wrench Position - Four different wrench positions, at 90 degree increments, were considered for the study. This was in keeping with the earlier studies (ref. 2, 4) done in this regard.

Not all combinations of the above variables were included in the study for reasons of importance as well as the size of the effort. The following set of test exertions only were considered.

1. Bolt head without glove:
   4 sizes x 3 orientations ..... = 12 exertions
2. Bolt head with glove:
   4 sizes x 3 orientations ..... = 12 exertions
3. Flat key with glove:
   3 sizes x 3 orientations ..... = 9 exertions
4. Flat key without glove:
   3 sizes x 1 orientation ..... = 3 exertions
5. Wrench without glove:
   2 sizes x 3 orientations x 4 positions ..... = 24 exertions
   Total test exertions ..... = 60

In addition to the above sixty test exertions, two benchmark exertions were also performed by the subject at the beginning and end of each of the two test sessions for a total of 68 exertions per subject. The benchmark exertions were included to study any effects of fatigue, learning, etc. on the part of the subject. The bench mark exertions are the same as those considered in the earlier study (ref. 2). Table 1 shows the combinations of the experimental variables for each of the sixty test exertions.

SUBJECTS:

Twenty males and twenty females volunteered and participated in the study as subjects. They were all recruited from among the students at Tennessee Technological University. All of them considered themselves in very good health and were screened for their suitability for participation in the experiment with a questionnaire seeking information on their health

Table 1: List of Test Exertions

Bolts (24 exertions)		Orientations		
3/8" 6 pt	No Glove	Facing	Transverse	Vertical
	Glove	Facing	Transverse	Vertical
3/8" 12 pt	No Glove	Facing	Transverse	Vertical
	Glove	Facing	Transverse	Vertical
9/16" 6 pt	No Glove	Facing	Transverse	Vertical
	Glove	Facing	Transverse	Vertical
3/4" 6 pt	No Glove	Facing	Transverse	Vertical
	Glove	Facing	Transverse	Vertical

Flat Keys (12 exertions)		Orientations & Positions		
1/2"	No Glove	Facing 90°		
	Glove	Facing 90°	Transverse 0°	Vertical 0°
1"	No Glove	Facing 90°		
	Glove	Facing 90°	Transverse 0°	Vertical 0°
2"	No Glove	Facing 90°		
	Glove	Facing 90°	Transverse 0°	Vertical 0°

Open-End Wrenches (24 exertions)		Orientations & Positions		
3/8"	No Glove	Facing	Transverse	Vertical
		0°	0°	0°
		90°	90°	90°
		180°	180°	180°
		270°	270°	270°
9/16"	No Glove	Facing	Transverse	Vertical
		0°	0°	0°
		90°	90°	90°
		180°	180°	180°
		270°	270°	270°

conditions. The selection and participation of subjects in the experiment was approved by the Institutional Review Committee for the Protection of Human Subjects at Tennessee Technological University. The total time of participation per subject was around 4 to 5 hours over two sessions. On completion of the testing each was given a reward of twenty dollars. Several subjects received some academic credit in addition.

All subjects were right handed and ranged in age from 18 to 26 years. They were initially measured for a set of anthropometric and strength measurements following the standard procedures (refs. 4,5). The measurements were the same as in the earlier study (ref. 2). Some were required for the experimental set up and others were needed to describe the current subject sample for the purposes of CREW CHIEF and DEPTH computer models. Table 2 presents the statistical summary of the measurements of the subjects who participated in this study.

APPARATUS:

A computerized automated data acquisition apparatus consisting of strain gage load cells, amplifiers, A/D and D/A converters, and a microcomputer system was the primary data collection unit. Two load cells - Lebow Model 2192 - 200 for hand torquing and AL Design Model 500 for wrench torquing - were used to sense the applied torque. Two different sensor units were used to provide the needed range and sensitivity of the measurements. The torque dynamometer available in the Ergonomics Laboratory was modified so that quick changing of the torque sensors was possible. Figures 4, 5, and 6 show the hardware used for torque data acquisition. The torque dynamometer itself was mounted on a metal framework that made it possible to test for the three different orientations of the fasteners and also to accommodate different body sizes of the subjects. The anthropometric and strength measurements were made using standard instruments. Such instruments were used in all previous similar studies (refs. 3,4) conducted at Tennessee Technological University.

TESTING PROCEDURE:

The selected subject was initially briefed about the experiment and its objectives and his/her full cooperation was sought for the success of the research study. The subject was also asked to try the test and be familiar with the torque application procedure before actual data collection was begun.

The sixty test exertions were performed by each subject in two sessions with at least three days of rest between sessions. One session included all the 36 hand torquing exertions (with and without gloves) and the other session included all the 24 wrench torquing exertions (without gloves) .

Table 2: Statistical Summary of Subjects
(See References for Measurement Procedures)

Measurement			Males n = 20 Mean	Males n = 20 SD	Females n =20 Mean	Females n =20 SD
AGE	- AGE	yrs	22.3	1.6	21.0	2.1
WT	- Weight	lbs	175.8	46.8	135.4	25.3
STAT	- Stature	cm	177.8	6.4	162.9	7.5
GRHT	- Grip Height	cm	215.5	8.2	196.1	9.7
FGL	- Fwd. Grip Length	cm	46.4	4.5	40.4	4.0
HLN	- Hand Length	cm	19.3	1.0	17.2	1.2
HB	- Hand Breadth	cm	9.2	0.5	8.0	0.5
D1l	- Digit I Length	cm	11.6	0.6	10.6	0.6
D2L	- Digit II Length	cm	9.8	0.6	8.8	0.6
WCR	- Wrist Circum.	cm	17.8	1.3	15.7	1.2
FCR	- Forearm Circ.	cm	29.8	3.2	25.8	2.9
BCR	- Biceps Circum.	cm	33.3	3.6	29.3	5.2
GST	- Grip Strength	Kg	51.6	9.9	30.6	7.8
TEL	- 38 cm Lift	lbs	285.9	54.4	150.3	55.9
ELL	- Elbow Ht. Lift	lbs	99.2	16.7	53.6	18.2
PS	- One Hand Pull St.	lbs	116.9	33.8	74.7	24.8
SIXFTL	- 6' Lift Strength	lbs	83.5	19.8	*50.0*	*10.4*

*** Eight Female subjects were not able to lift the test apparatus with no weights added. (The test apparatus weighs 40 lbs.) The data given here represents the remaining twelve subjects who were able to perform the lift.

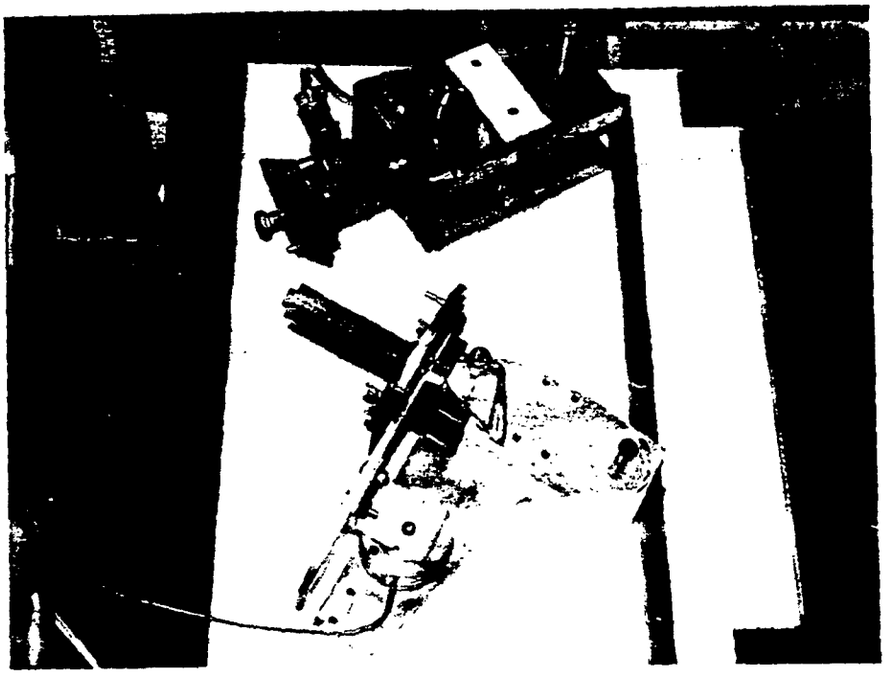


Figure 4. Torque Sensor Hardware Units
Top: Hand torquing unit Bottom: Wrench torquing unit

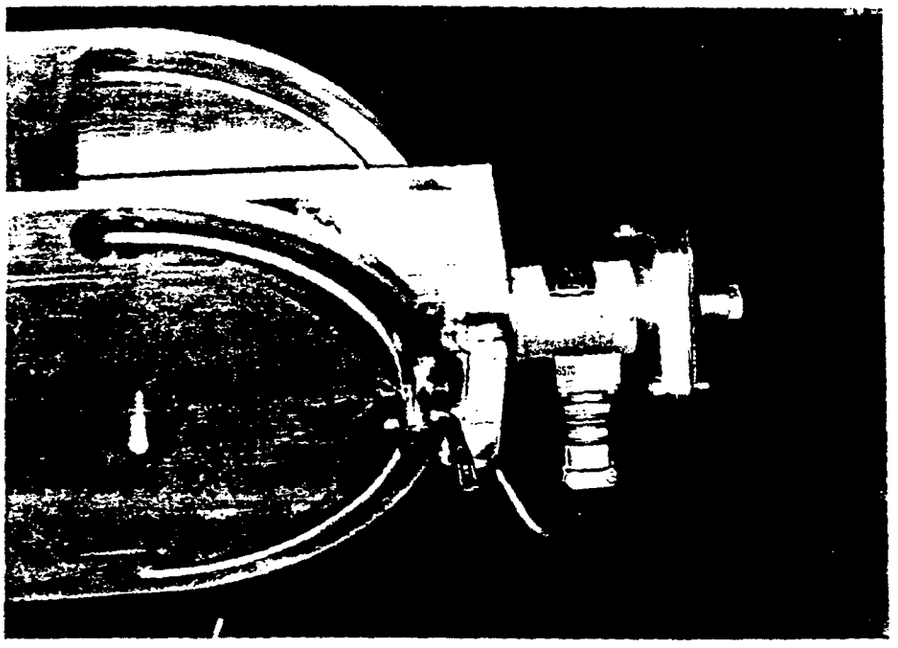


Figure 5. Hand Torque Sensor Unit installed in the Dynamometer

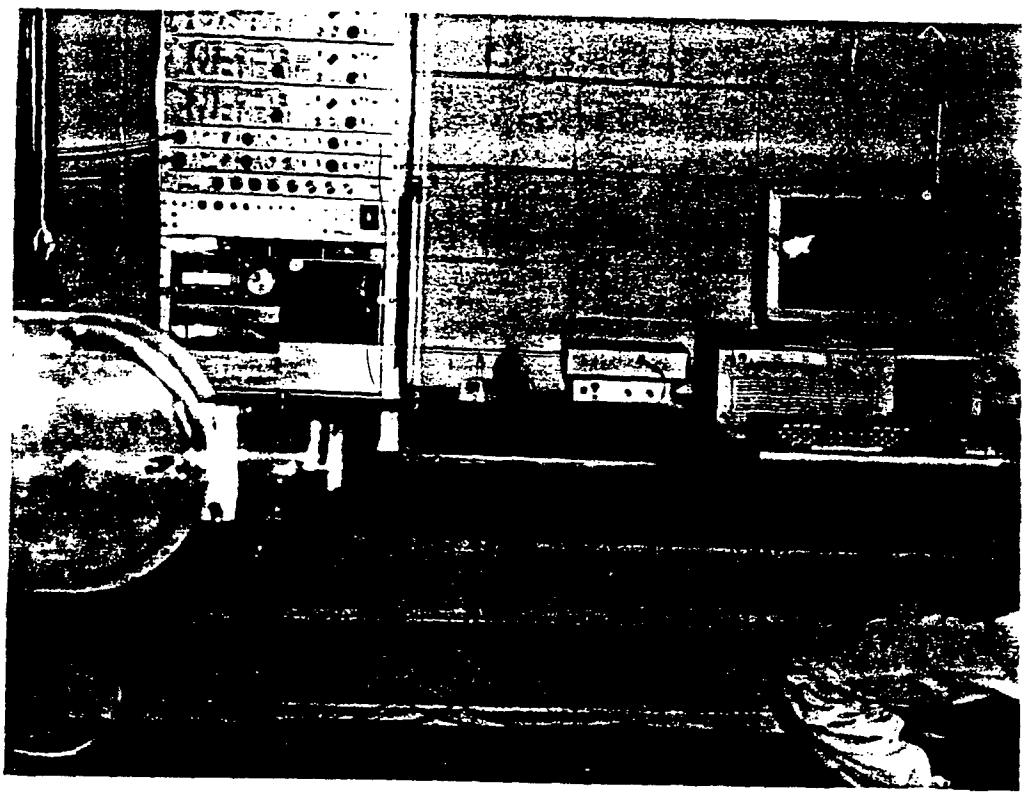


Figure 6. The Automated Data Acquisition System

Each testing session started with the two benchmark exertions and ended with the same benchmark exertions. The anthropometric and strength measurements were mostly made before the first test session. Half the males and half the females were randomly assigned to start with the hand torquing session and the other half started with the wrench torquing session. All test exertions within a session were performed in random order.

The torque dynamometer was adjusted vertically so that the center of the fastener was at a height equal to 50% of the individual subject's vertical reach height. The forward location of the fastener was controlled by requiring the subject to stand upright with the location of the right toe at a distance of 50% of his/her forward grip reach from the fastener. The subjects left foot could be placed at any comfortable location to the left and rear of the right foot for good balance and safety. The subject was then instructed on the proper gripping technique for the specific exertion and was asked to try it and be familiar with it. The gripping technique for each of the exertions were developed and standardized previously with the author and several other volunteers trying out. The standardized gripping techniques are briefly listed in table 3. The subjects were cautioned to use only upper body muscles for torquing the fastener and desist from bearing on the fastener with the body weight. They were also advised to prevent hand slipping while applying the torque. Many subjects found it difficult to prevent slipping while torquing the small bolt heads with hands. Figures 7 through 9 show three examples of typical exertion postures.

When the subject took the proper position and grip and was ready the experimenter entered a command to the computer to start collecting data and also cued the subject to apply the tightening torque. The subject had to develop the maximum torque within the first second without any jerky motion and maintain it for the following three seconds. The computer kept track of the time and torque applied and signalled with the beep at the end of four seconds for the subject to stop torquing the fastener and relax. The computer sampled the analog data at the rate of 10 per second and digitized the same on real time basis. The torque exertion should have met the following two criteria for acceptance.

1. The ratio of peak torque during the 0 to 1 second time period to the mean torque between the 1 to 4 second time period must be within the range 0.85 to 1.15.
2. No more than 8 sample points be outside the range of + or - 10% of the mean value during the time period 1 to 4 seconds.

At the end of exertion the computer displayed the above information along with a graph of time vs. torque for the experimenter to decide whether to accept the data or not. Figure 10 shows a sample of the display If not accepted, the exertion was repeated immediately after a rest break. The

Table 3: Guidelines for Gripping Fasteners

The following are guidelines for the subject to grip the fastener/ open-end wrench. Minor variations in gripping were permitted to suit the subjects hand size, comfort, etc.

BOLT HEAD (all sizes, clockwise, with and without gloves)

Facing	Wrist straight, palm down, bolt head gripped between thumb and side of pointer finger.
Transverse	Wrist straight, palm facing bolt head with thumb up, bolt head gripped between thumb and side of pointer finger.
Vertical	Wrist straight, elbow bent, arm level, palm down bolt head gripped between thumb and side of pointer finger.

FLAT KEY

Facing (clockwise)
0° - Palm facing down, thumb left top. (Benchmark 1 & 3, 3" Key)
90° - Palm facing left, thumb near right top.

Transverse (clockwise
0° - Palm facing plate, thumb near top.
90° - Palm facing plate, thumb far top.

Vertical (clockwise)
0° - Palm facing plate, thumb on left side at far end
90° - Palm down facing plate, thumb far left

Transverse (counter-clockwise)
90° - Palm down, thumb on top at near end,
index finger wrapped around far end

OPEN-END WRENCHES (all sizes, Clockwise)

Facing at 0° - Palm facing left, thumb over end of wrench.
Facing at 90° - Hand behind wrench, thumb extended on top.
Facing at 180° - Palm facing left, thumb around wrench.
Facing at 270° - Hand behind wrench, thumb over end.

Transverse at 0° - Palm facing left, thumb over end of wrench.
Transverse at 90° - Palm facing left, thumb over end of wrench.
Transverse at 180° - Palm facing left, thumb extended along wrench.
Transverse at 270° - Palm facing left, thumb extended along wrench.

Vertical at 0° - Palm up, thumb over end of wrench.
Vertical at 90° - Palm down, thumb extended along wrench.
Vertical at 180° - Palm up, thumb extended along wrench.
Vertical at 270° - Palm down, thumb across end of wrench.

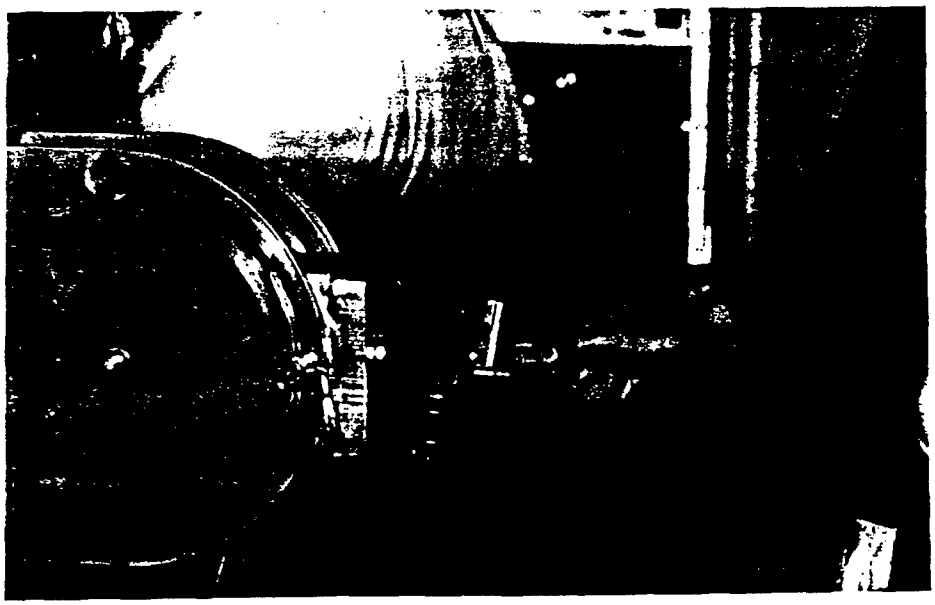


Figure 7. Test Exertion - Facing , Bolt Head, No Glove

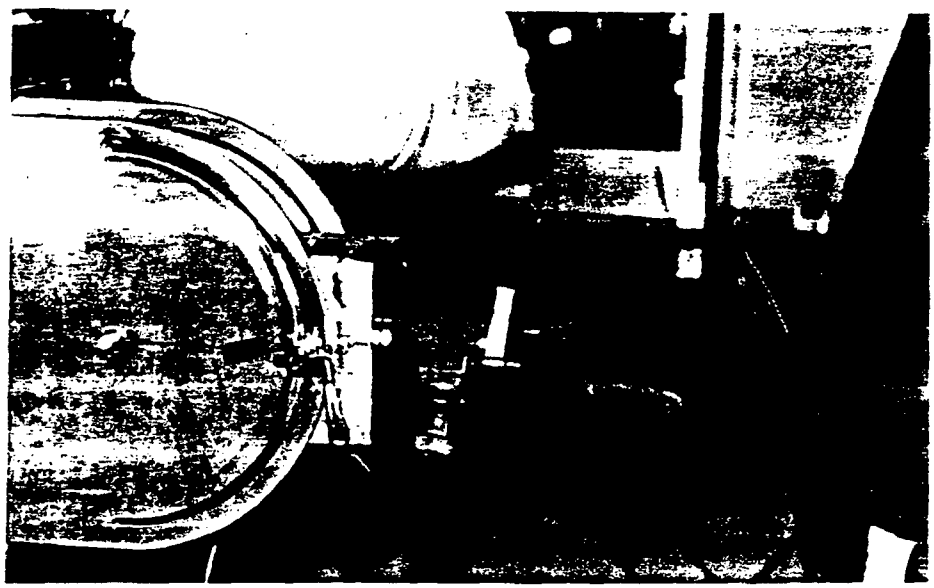


Figure 8. Test Exertion - Facing, Bolt Head, With Glove

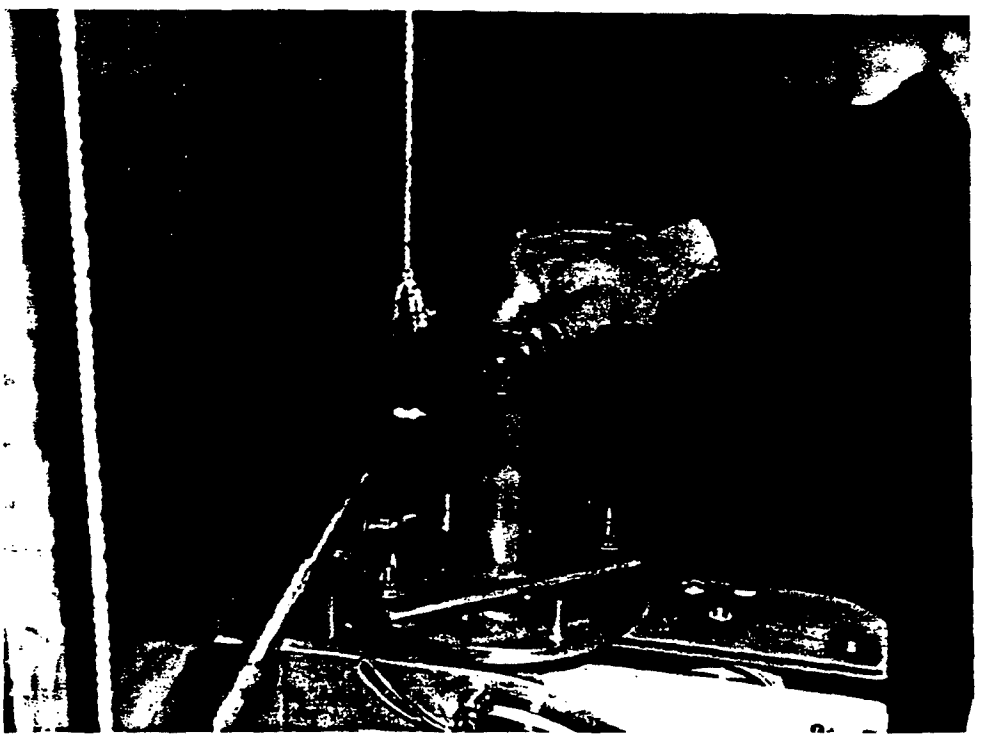


Figure 9. Test Exertion - Vertical, Bolt Head, With Wrench

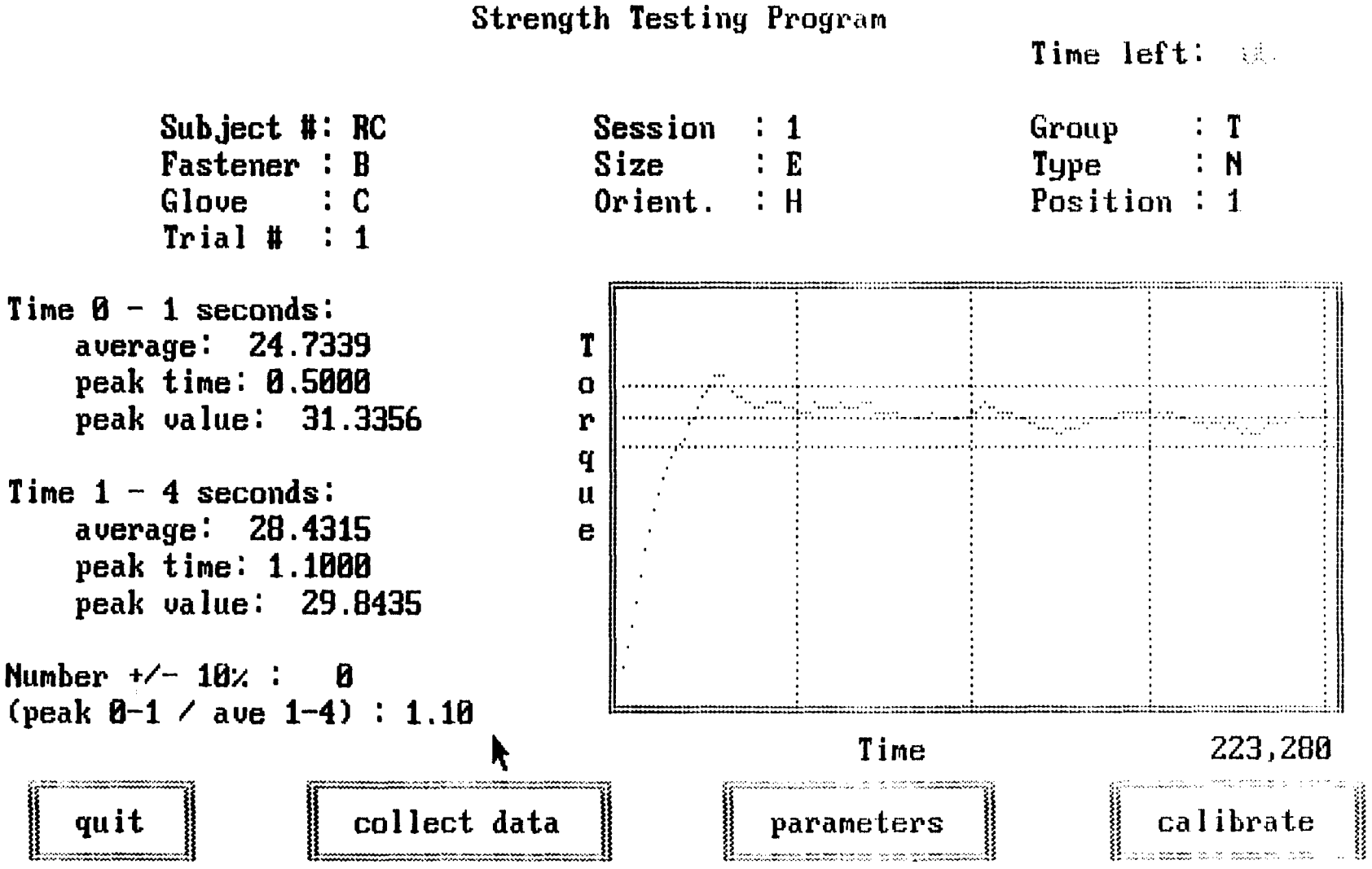


Figure 10. Example of the Display Screen of the Computer for a Typical Test Exertion.

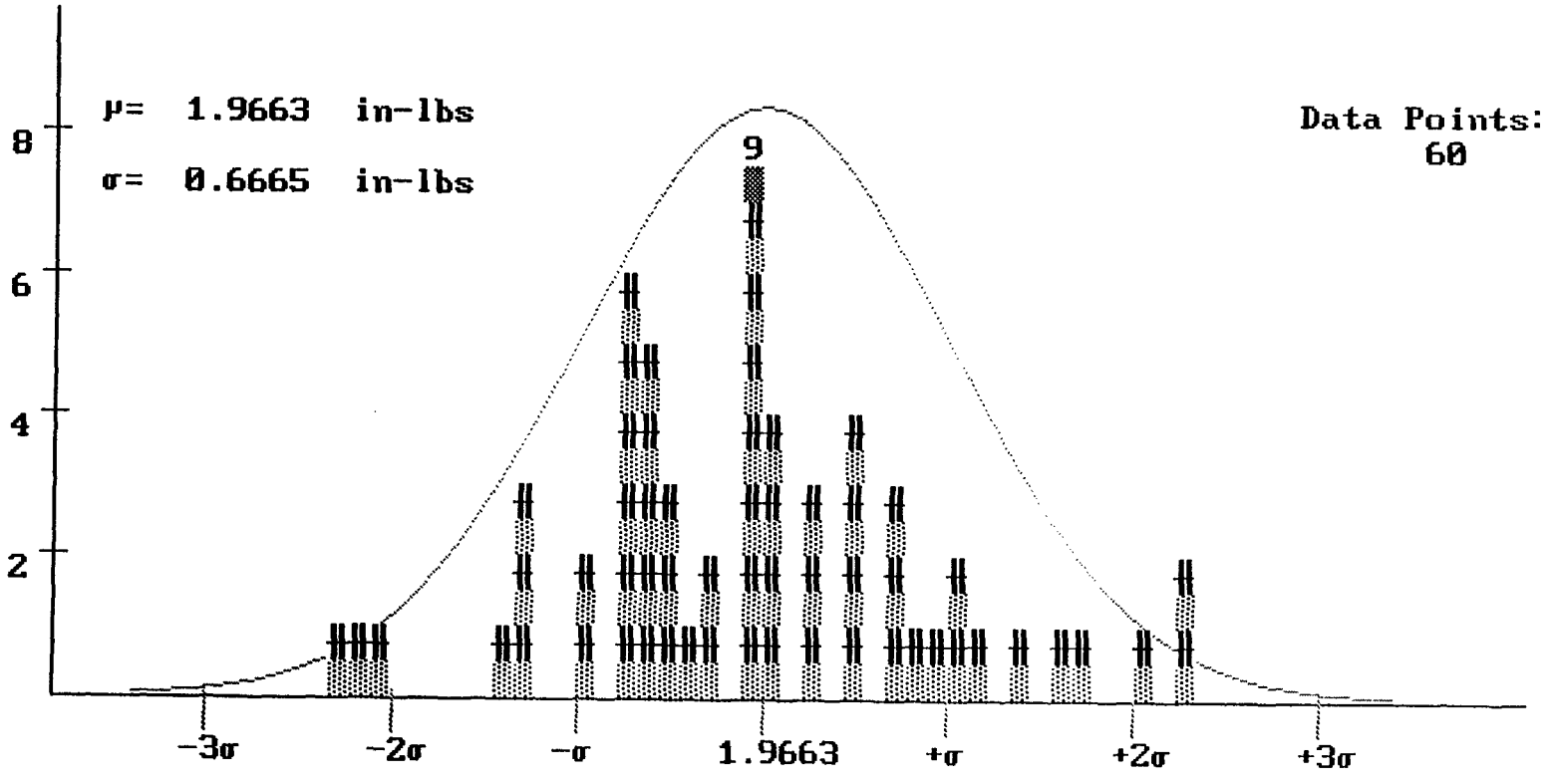


Figure 11. Frequency Distribution of Observed Torque Strength Values - Male Subjects, 6 Point, 3/8" Bolt Head, All Orientations, Without Glove.

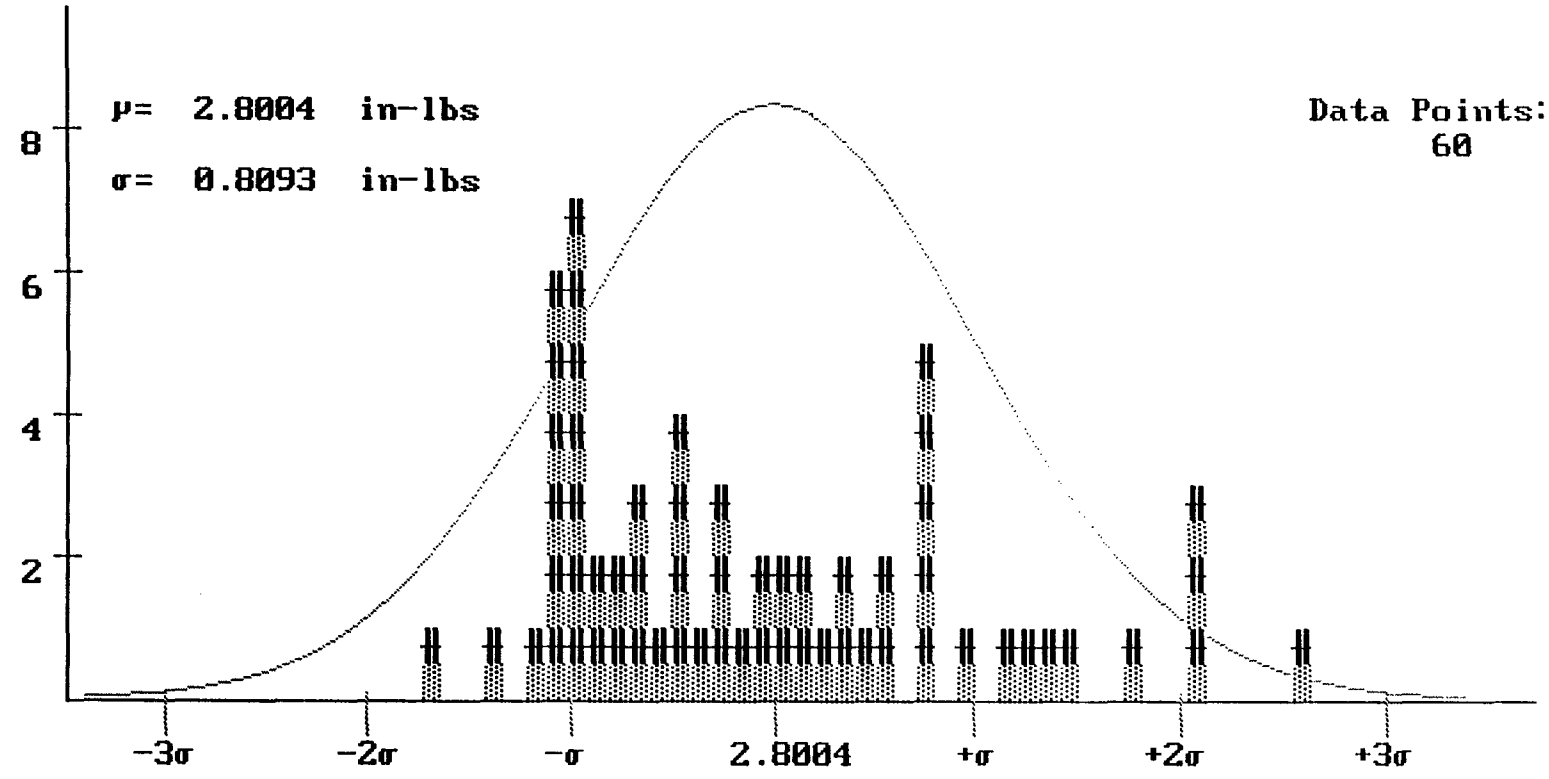


Figure 12. Frequency Distribution of Observed Torque Strength Values - Male Subjects, 6 point, 3/8" Bolt Head, All Orientations, With Glove.

subject always rested for a minimum of 2 minutes between exertions in order to minimize any effects of muscular fatigue and pain in the soft tissues. This time period was utilized often by the experimenter to reset the hardware for the next exertion. The subject was free to take any additional time for rest as he/she felt.

It was originally planned to use the data collection software developed earlier at Armstrong Laboratory for the CREW CHIEF program for this study. However, after considerable time spent trying to use the software, it was found that there were some serious compatibility problems with the present hardware. Therefore, a new software was developed for the purpose of data collection for the purposes of this study. This software has many user-friendly features incorporated in it.

RESULTS AND DISCUSSION:

The average value of the torque applied during the time period 1 to 4 seconds is considered to be the torque strength of the subject for the specific exertion. This is consistent with all previous studies conducted for similar purposes. The torque strength values of individual subjects were then pooled and the statistical averages were computed. Further ANOVA analyses were also performed on the data to test for the effects of the primary variables on the torque strength.

The raw data - torque strength - for males for the case of 3/8", 6 point bolt head without and with gloves are shown in figures 11 and 12 respectively. They are shown as frequency diagrams with a normal distribution overlapping for comparison purposes. Comparing the two figures it becomes clear that the glove seems to have an effect of increasing the torquing capability. Not only is the average value with glove (2.8 inch lbs.) higher than for the case of no glove (1.97 inch lbs.), the low end of the distribution for the 'with glove' case is truncated indicating that the subjects who might have had difficulty applying torque without slipping when no glove was worn were able to reduce the amount of slipping with the glove on. Further, the glove should have helped in reducing the pain sensation too. Similar observations were noted in other cases too. In the interest of space the discussion on raw data is limited to the above case only. The discussion in the following sections is based on the statistical summaries for each case as indicated.

Bolt Head: The averages and standard deviations for bolt heads are presented in table 4. The averages are also shown in figures 13 through 16 for different orientations. As can be expected the torque strength values increased as the size of the bolt head increased. Obviously the larger size helped in gripping (with less slipping) as well as increasing the surface

Table 4: Torque Strength on Bolt Heads (inch-Pounds)

Means

Size	Orientation	Males		Females	
		No Glove	Glove	No Glove	Glove
3/8" 6 pt	Facing	1.8	2.4	1.2	1.6
	Vertical	2.1	3.1	1.5	1.8
	Transverse	2.1	2.9	1.4	1.7
	Combined	2.0	2.8	1.3	1.7
3/8" 12 pt	Facing	2.5	2.7	1.4	1.7
	Vertical	3.3	4.0	1.9	2.2
	Transverse	3.0	3.7	1.8	1.9
	Combined	2.9	3.5	1.7	1.9
9/16" 6 pt	Facing	2.7	4.2	1.9	2.6
	Vertical	3.6	5.0	2.3	3.0
	Transverse	3.7	5.4	2.2	2.9
	Combined	3.3	4.9	2.1	2.8
3/4" 6 pt	Facing	4.6	6.0	2.8	3.3
	Vertical	5.6	7.4	3.4	4.1
	Transverse	5.7	7.4	3.4	3.9
	Combined	5.3	7.0	3.2	3.8

Standard Deviations

Size	Orientation	Males		Females	
		No Glove	Glove	No Glove	Glove
3/8" 6 pt	Facing	0.5	0.4	0.6	0.6
	Vertical	0.6	0.5	1.0	0.7
	Transverse	0.8	0.5	0.7	0.5
	Combined	0.7	0.8	0.4	0.6
3/8" 12 pt	Facing	0.7	0.4	0.5	0.5
	Vertical	1.0	0.6	1.1	0.9
	Transverse	1.0	0.5	1.2	0.6
	Combined	1.0	1.1	0.6	0.7
9/16" 6 pt	Facing	0.7	0.6	0.8	1.0
	Vertical	1.1	0.6	1.5	1.0
	Transverse	1.1	0.7	1.5	1.1
	Combined	1.1	1.4	0.7	1.1
3/4" 6 pt	Facing	1.4	1.0	1.7	1.2
	Vertical	1.8	1.0	1.8	1.4
	Transverse	1.4	1.0	1.6	1.2
	Combined	1.6	1.8	1.0	1.3

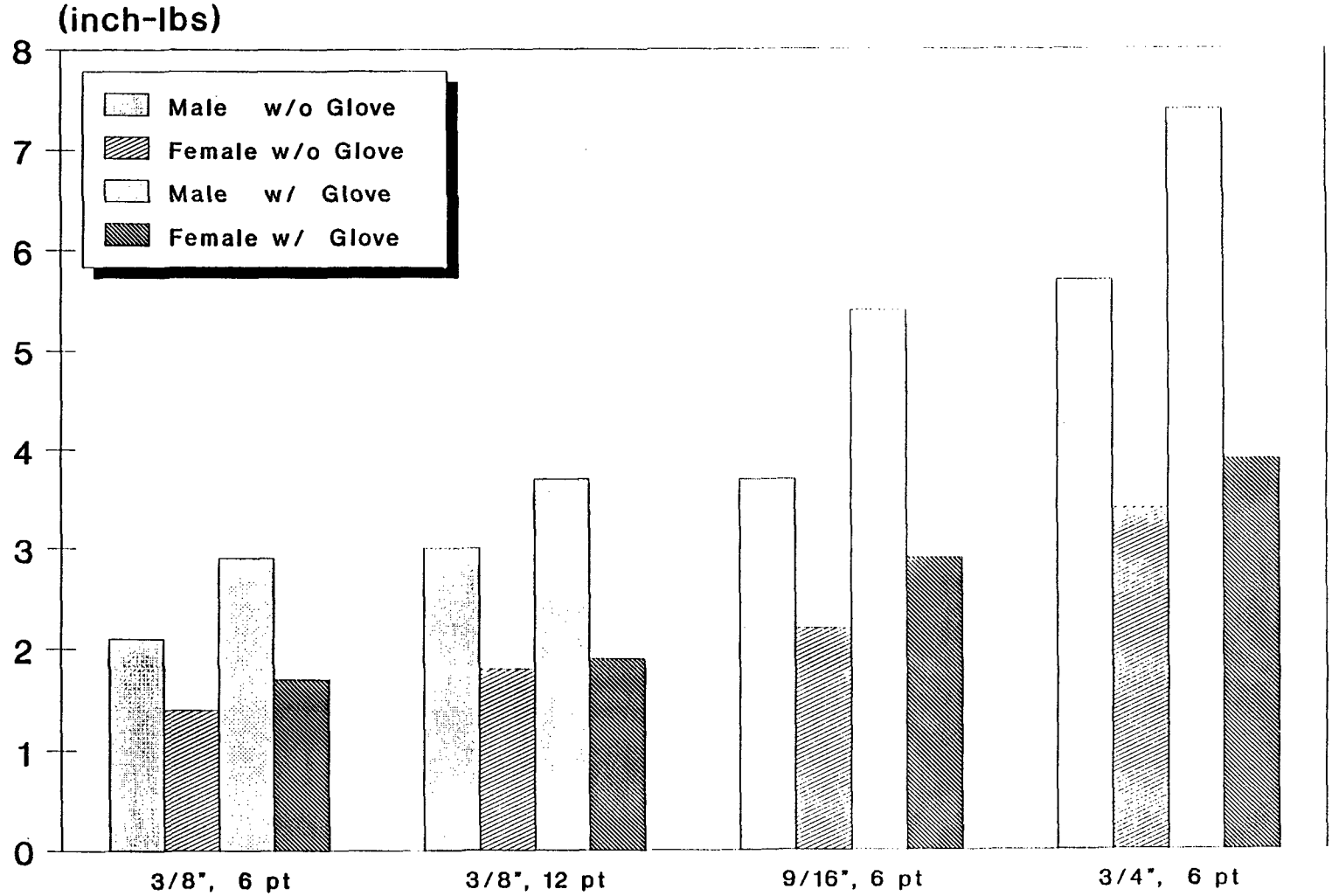


Figure 13: Torque Strength (in-lbs) for Bolt Heads in the Transverse Orientation

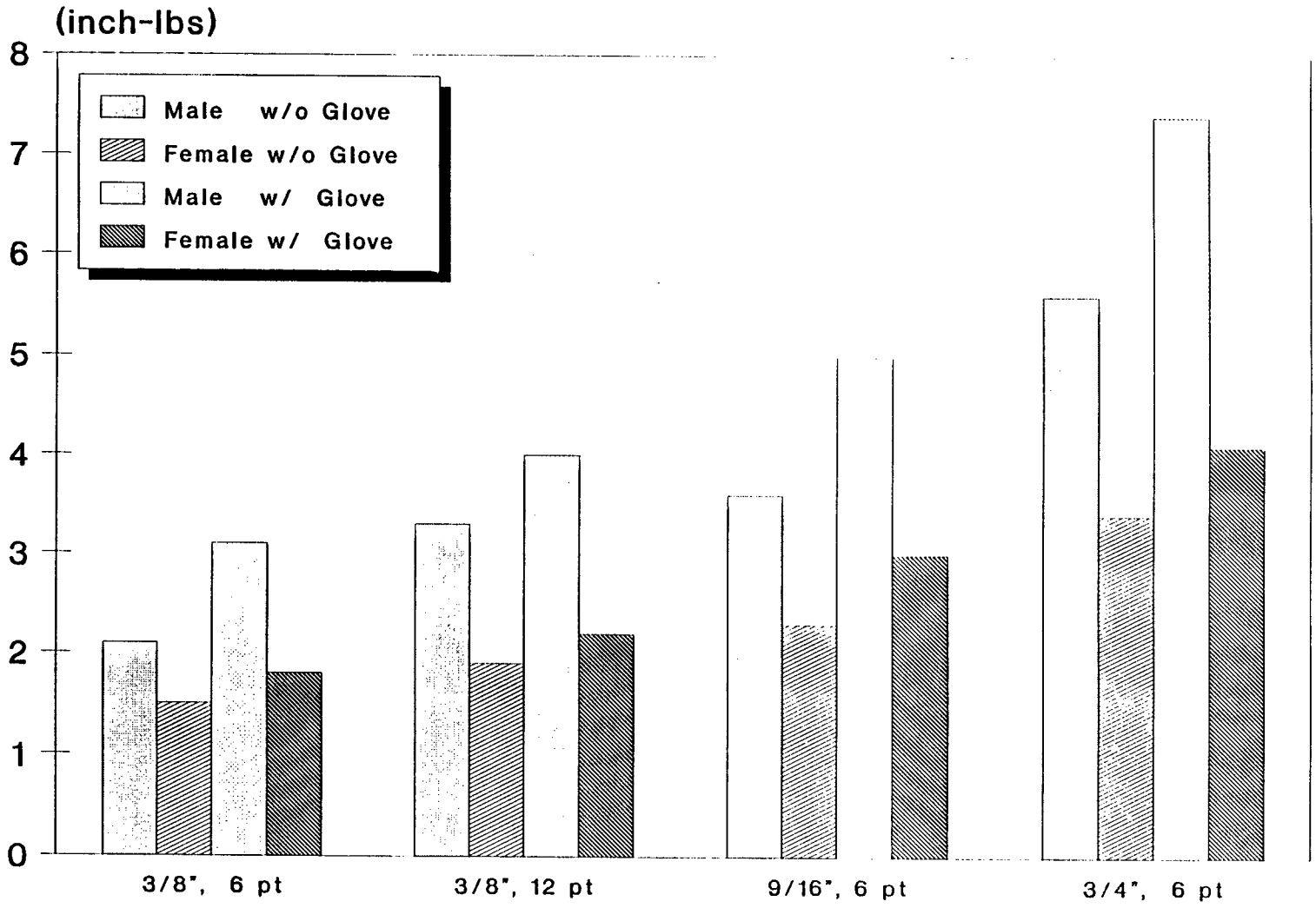


Figure 14: Torque Strength (in-lbs) for Bolt Heads in the Vertical Orientation

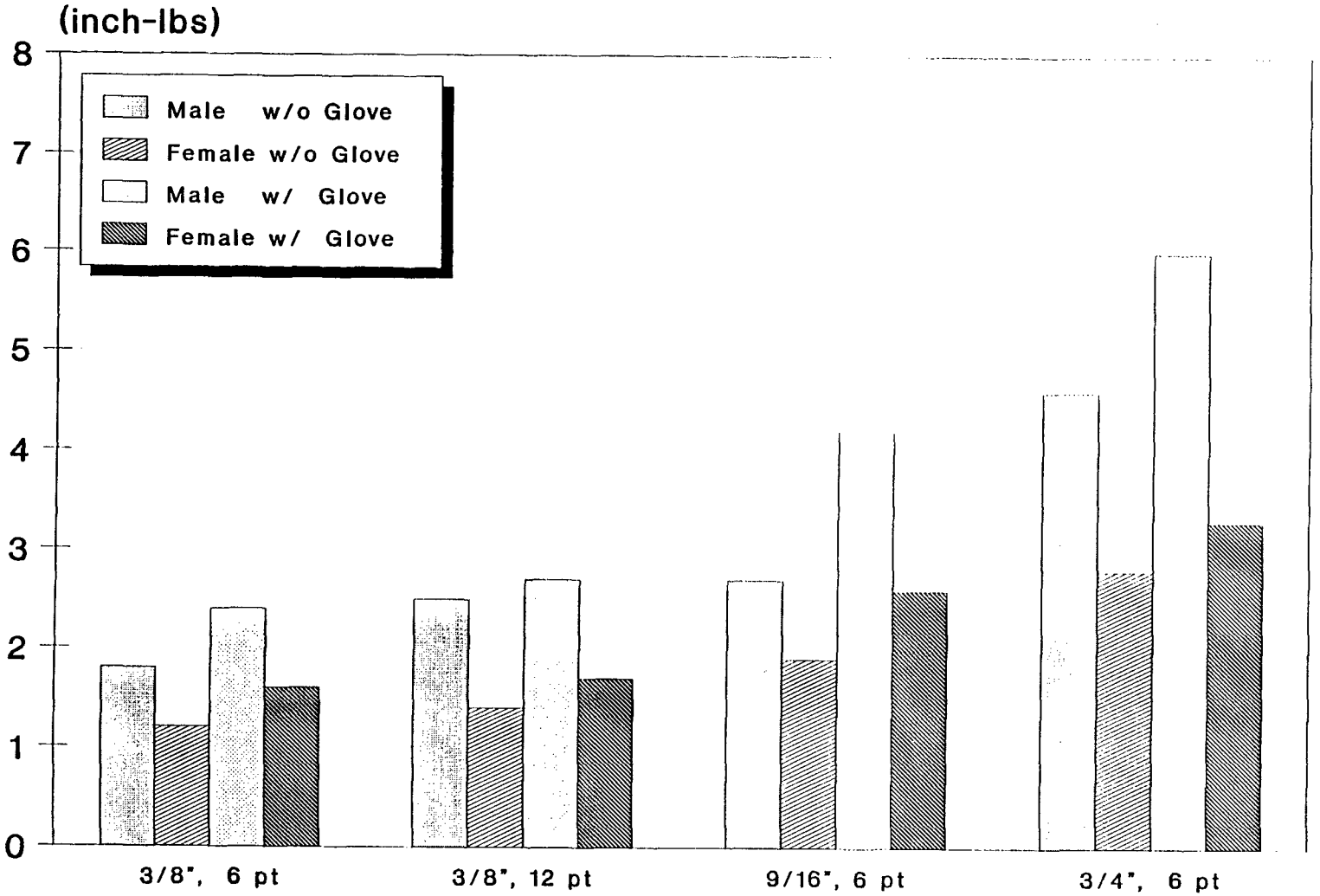


Figure 15: Torque Strength (in-lbs) for Bolt Heads in the Facing Orientation

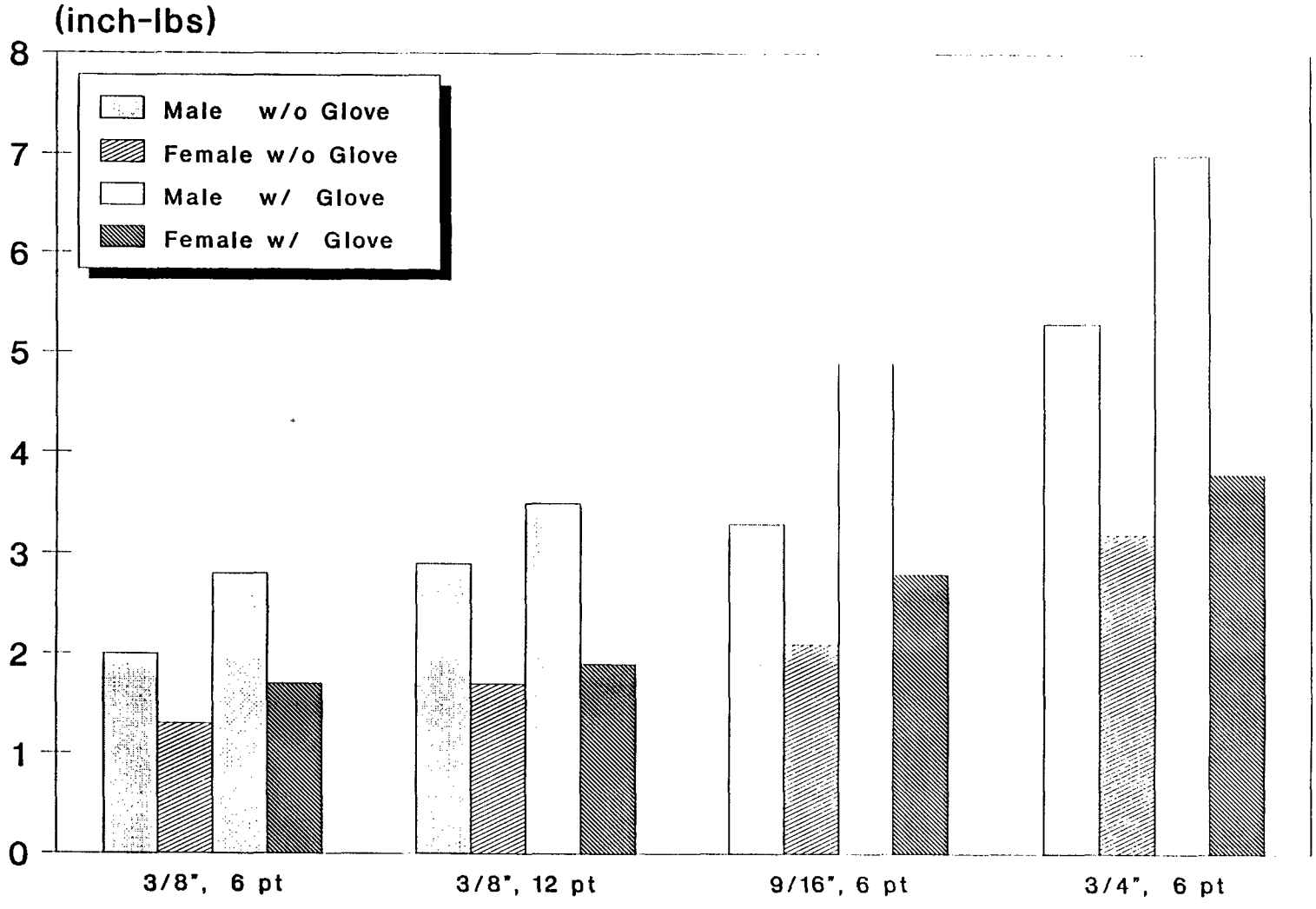


Figure 16: Torque Strength (in-lbs) for Bolt Heads With all Orientations Combined

area of contact resulting in higher torque values. The torque strength in the transverse and vertical orientations was very comparable. However in the case of facing orientation the torque strength values were always lower. This is probably due to the fact that the facing orientation, at the height set for the fastener, resulted in a hand/fore-arm posture that is not very advantageous biomechanically for torquing tasks. Use of glove resulted in a definite increase in the torquing capabilities irrespective of the size or the orientation of the bolt head. Subjects were also able to apply higher torque on the 12 point bolt head than the 6 point bolt head. This effect is probably due to the larger surface area of contact made possible by the 12 point head. The ANOVA results also indicated significant effects on the torque strength due to size, orientation, and glove, all at a level of 0.05 or less. Female subjects as a group had lesser absolute torque strength compared to male subjects. On the average, females were able to apply about 50 % to 70% of the torque applied by males. The significant effects of size, orientation, and glove are noticed in the case of females also. An important information gained is that the torque strength values for hand tightening of bolt heads will be in the range of a few inch pounds only.

Flat Keys: For the purpose of this study the flat key was considered to represent the generic case of small fasteners rather than any specific kind. The results are shown in table 5 and figure 17. The results indicate very similar effects due to size, glove, and gender as in the case of bolt heads. However, in the case of orientation, facing and vertical orientations resulted in lower torque values than transverse orientation. In the earlier study (ref. 2) comparing the transverse and facing orientations, it was found that the transverse orientation was definitely better and the present study seems to confirm that finding.

Open-end Wrenches: The final securing of small fasteners is generally accomplished using wrenches. Therefore, this study includes the case of torquing with standard size open end wrenches. The results are presented in table 6 and in figures 18 and 19. It can be seen that the torque strength values while using a wrench is many times higher than with bare hands, with or without glove. The leverage provided by the length of the wrench in combination with the ease of gripping is the major factor for this effect. The position and the orientation variables show an interaction effect because the gripping posture is very much influenced by these two. The gripping posture, in turn, influences the biomechanical advantage for the torquing action. For example, for the facing orientations, the 90 degree position offers the best condition for pushing down on a handle extending horizontally across the body. The same position of the wrench, for the transverse

Table 5: Torque Strength on Flat Keys (inch-Pounds)

Means

Size	Orientation	Males		Females	
		No Glove	Glove	No Glove	Glove
1/2"	Transverse at 0°		8.9		4.4
1"	Transverse at 0°		19.4		10.3
2"	Transverse at 0°		45.1		24.3
1/2"	Vertical at 0°		8.5		4.4
1"	Vertical at 0°		15.2		8.7
2"	Vertical at 0°		30.4		14.7
1/2"	Facing at 90°	5.3	6.9	3.8	4.4
1"	Facing at 90°	12.2	14.6	7.0	7.9
2"	Facing at 90°	28.7	34.0	13.6	16.1

Standard Deviations

Size	Orientation	Males		Females	
		No Glove	Glove	No Glove	Glove
1/2"	Transverse at 0°		4.1		1.7
1"	Transverse at 0°		6.8		5.2
2"	Transverse at 0°		18.7		12.6
1/2"	Vertical at 0°		2.3		1.8
1"	Vertical at 0°		3.9		3.7
2"	Vertical at 0°		9.2		7.3
1/2"	Facing at 90°	1.1	1.7	2.3	2.1
1"	Facing at 90°	3.1	4.3	2.7	3.1
2"	Facing at 90°	10.8	11.2	5.9	7.1

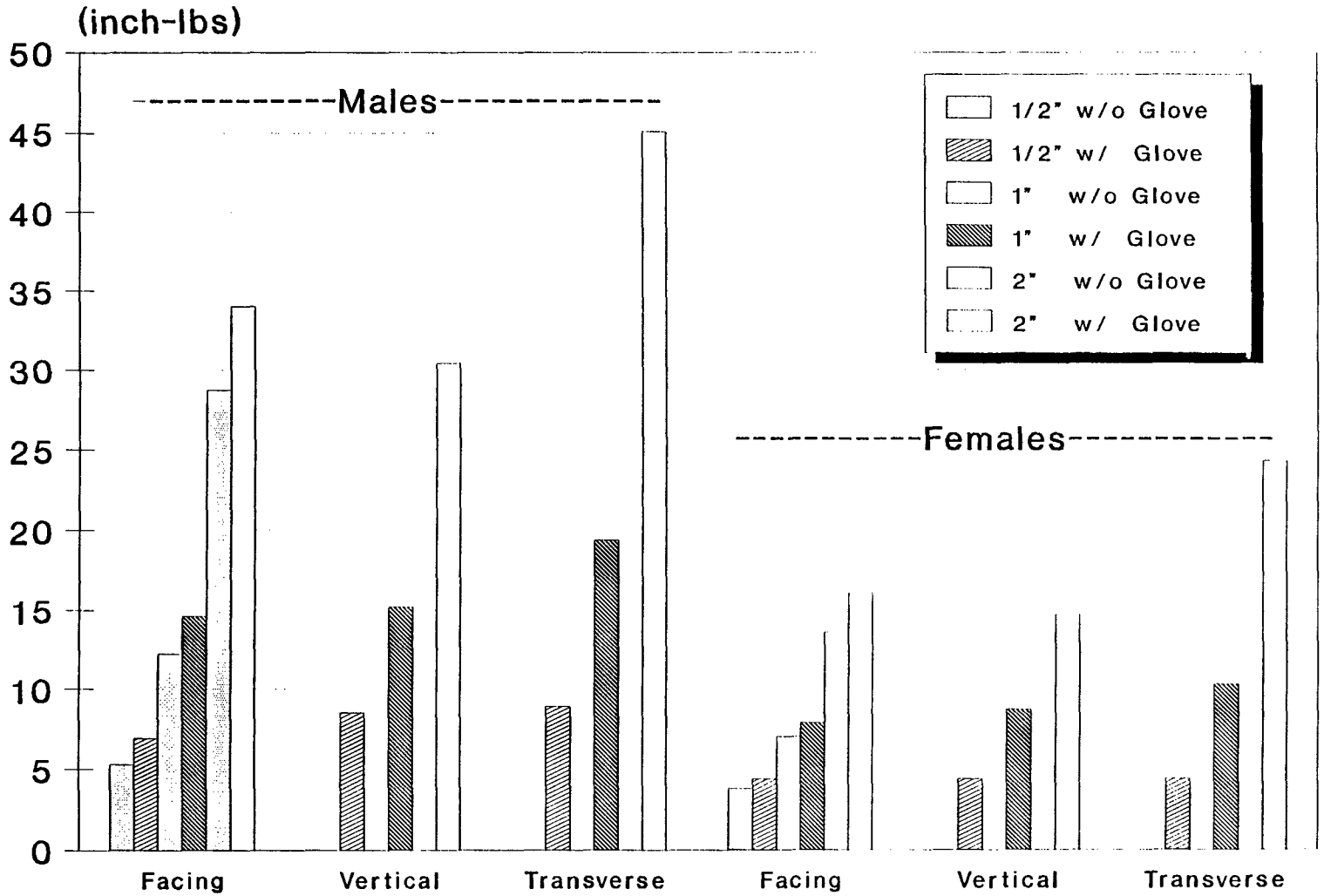


Figure 17: Torque Strength (in-lbs) for Flat Keys

Table 6: Torque Strength on Wrenches (inch-Pounds)

Means

3/8" Open-End Wrench

	Males			Females		
Position	Facing	Vertical	Transverse	Facing	Vertical	Transverse
0°	69.5	75.3	66.9	36.8	38.6	35.7
90°	113.3	125.8	87.2	47.8	59.1	48.5
180°	89.5	93.5	121.0	43.2	44.6	62.8
270°	92.2	55.1	111.6	42.0	31.1	59.4

9/16" Open-End Wrench

	Males			Females		
Position	Facing	Vertical	Transverse	Facing	Vertical	Transverse
0°	97.7	103.7	109.7	51.6	50.1	62.0
90°	188.8	166.5	137.6	90.1	78.9	79.4
180°	123.7	149.2	182.0	58.4	75.9	94.3
270°	141.0	92.0	180.5	59.3	45.8	86.9

Standard Deviations

3/8" Open-End Wrench

	Males			Females		
Position	Facing	Vertical	Transverse	Facing	Vertical	Transverse
0°	25.7	36.7	23.9	19.5	24.1	17.6
90°	56.5	55.4	32.7	25.2	33.0	24.5
180°	29.7	45.7	36.9	25.8	28.3	37.8
270°	35.0	23.6	46.2	21.3	17.8	33.8

9/16" Open-End Wrench

	Males			Females		
Position	Facing	Vertical	Transverse	Facing	Vertical	Transverse
0°	41.9	47.9	28.8	25.7	27.0	34.0
90°	51.1	61.5	40.1	51.1	42.1	40.8
180°	40.9	47.2	50.7	34.4	40.8	45.8
270°	52.8	38.8	61.4	35.7	24.2	47.7

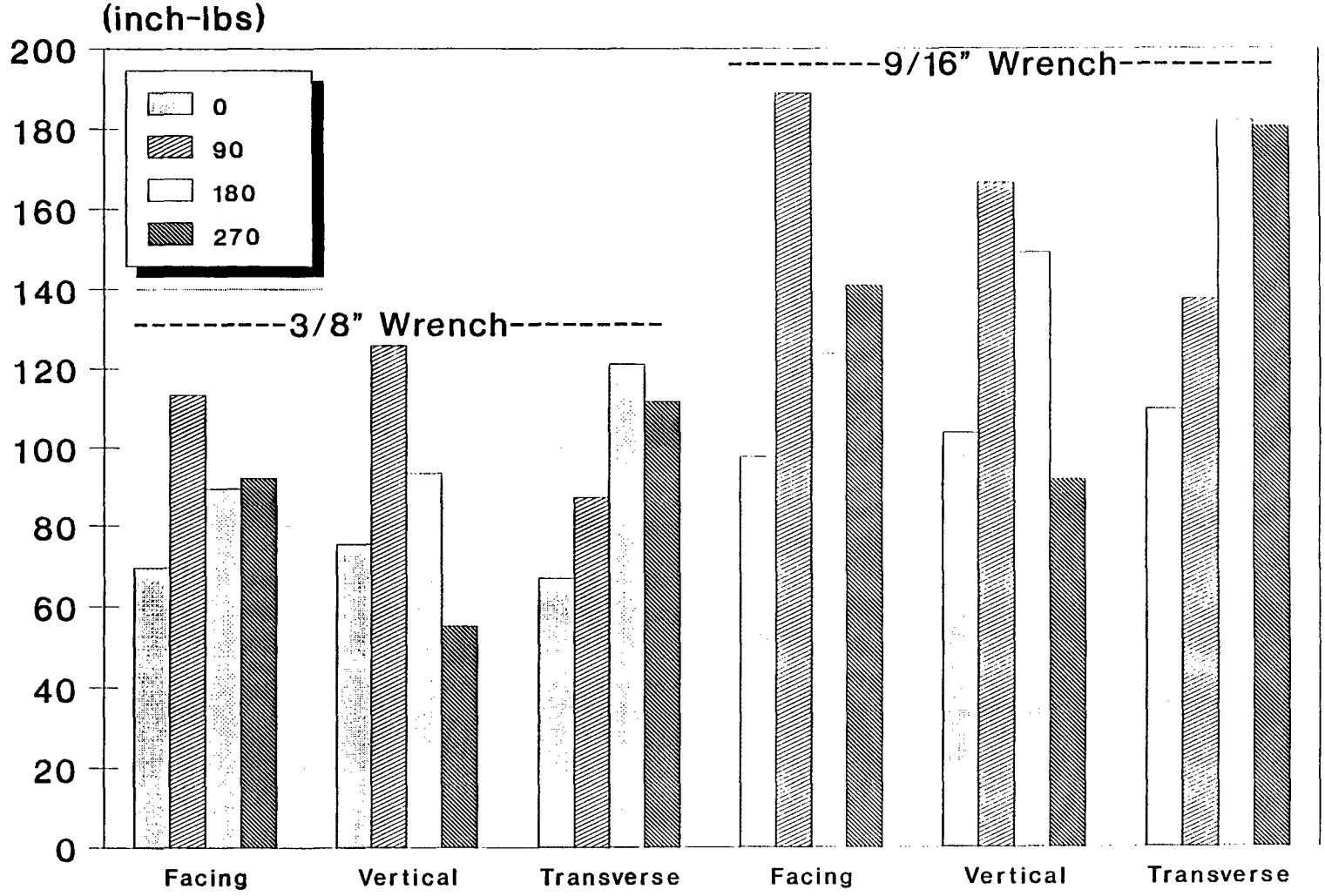


Figure 18: Torque Strength (in-lbs) of Male Subjects Using Wrenches

orientation, results in an extended gripping condition with the possibility of slipping and thus lowering the torque strength. At the same time, the 180 degree position with the wrench handle extending downwards offers the best condition for pulling up in the transverse orientation than for pushing to the left in the facing orientation.

The mean and the standard deviation of torque values for the benchmark exertions are shown in table 7. Benchmark 1 and 3 are identical exertions performed at the beginning and the end of each session respectively. Similarly benchmarks 2 and 4 are identical. From the table it can be observed that the values of torque strength exertions changed very little from the beginning to the end of session. Thus it appears that the testing procedure followed did not result in any significant level of fatigue for the subjects.

CONCLUSION:

The following conclusions can be made from the results and the experiences gained from this experimental study.

1. The tightening torque that can be applied with hand, on small fasteners such as bolt heads and keys, is generally low, in the order of few inch pounds and will vary depending on several factors.
2. The larger the gripping size of the fastener, the higher the hand torquing capability.
3. Wearing a work glove will result in higher torque applied on the fastener.
4. The orientation of the fastener with respect to the person has an effect on the maximum torque applied on the fastener.
5. Generally females have lower torque strength than males for the tasks considered in this study in the range of 50 to 70 percent.
6. The maximum torque that can be applied on a 12 point bolt head is considerably more than that can be applied on a 6 point bolt head.
7. The use of a small open-end wrench will increase the applied torque by several times compared to hand tightening. The effects of size, orientation, and gender of the person have significant effects. The position of the wrench, i.e., the direction in which the wrench extends, with respect to the subject also has an effect. However, this effect varies with the orientation of the fastener. The effects of the position and the orientation can be attributed to biomechanical factors.

ACKNOWLEDGEMENTS:

The author wishes to express his thanks to Dr. Joe W. McDaniel of USAF Armstrong Laboratory for his strong support and many useful suggestions for this effort. Sincere thanks to Mr. Edward Boyle, also of USAF Armstrong Laboratory, for his encouragement and help during the various phases of this

Table 7: Torque Strength on Benchmark Exertions
3" Flat Key (inch-Pounds)

Means

		Male	Female
Benchmark 1	Bolt & Key Session	36.5	20.1
	Wrench Session	34.8	22.7
Benchmark 2	Bolt & Key Session	52.3	30.1
	Wrench Session	50.7	28.6
Benchmark 3	Bolt & Key Session	35.8	18.9
	Wrench Session	35.8	20.9
Benchmark 4	Bolt & Key Session	46.4	25.9
	Wrench Session	49.4	29.4

Standard Deviations

Benchmark 1	Bolt & Key Session	8.9	7.3
	Wrench Session	10.3	9.3
Benchmark 2	Bolt & Key Session	20.2	13.6
	Wrench Session	15.4	13.1
Benchmark 3	Bolt & Key Session	10.8	7.3
	Wrench Session	10.1	9.5
Benchmark 4	Bolt & Key Session	16.5	12.7
	Wrench Session	17.6	14.2

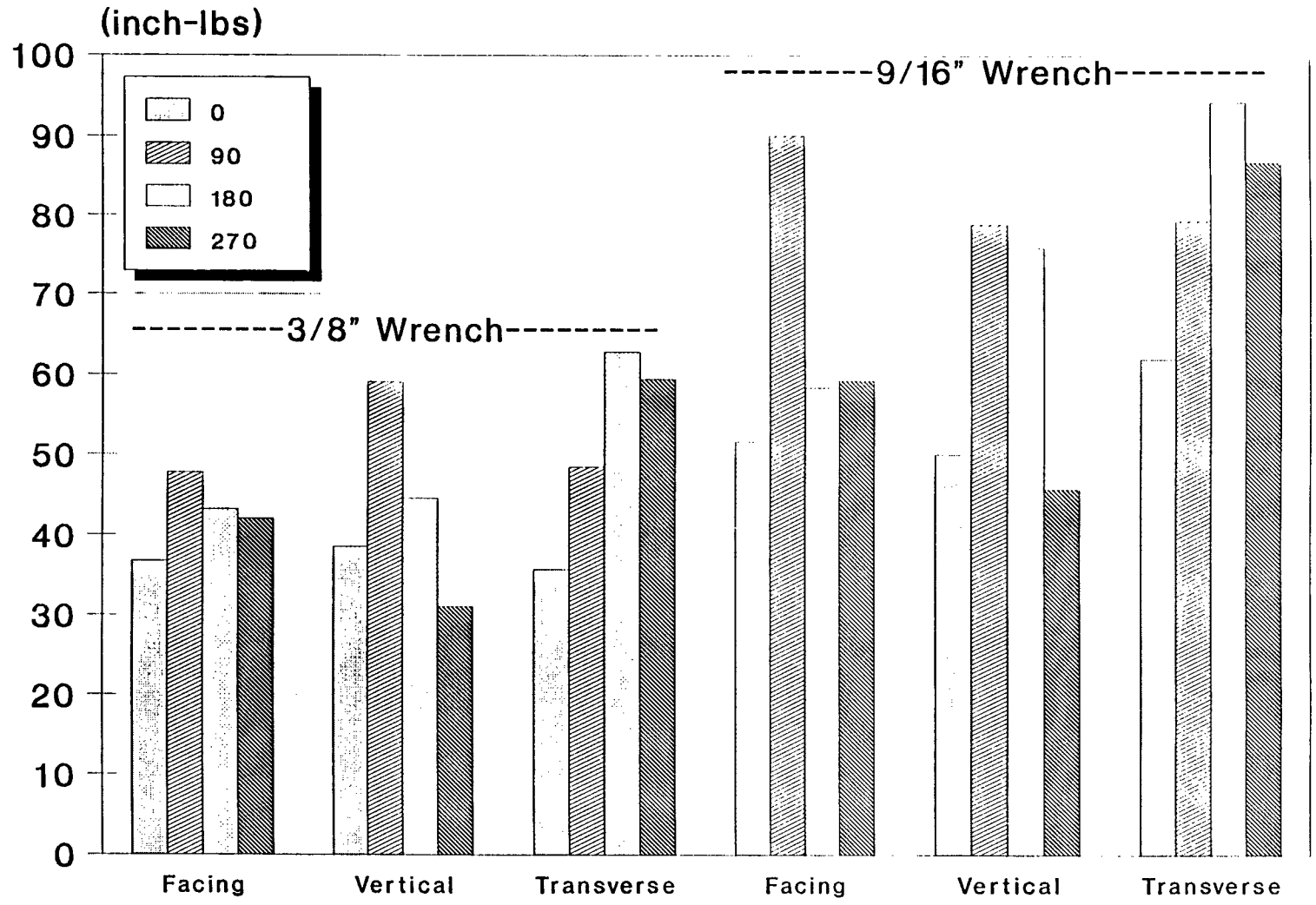


Figure 19: Torque Strength (in-lbs) of Female Subjects Using Wrenches

study. Mr. Russell Carter, graduate student in the Department of Industrial Engineering at Tennessee Technological University, spent many hours testing the subjects, analyzing the data, and helped in several other ways. His effort for the success of this project is gratefully acknowledged. Thanks to Mr. David Ress, another graduate student, who made the very significant contribution of developing the software for data acquisition. Thanks are also due to Mr. Robert Finegan and Mr. Bill Mitchell, Laboratory Technicians at Tennessee Technological University for their assistance in building and maintaining the experimental hardware. Thanks to the volunteer subjects who made the study possible. Finally, thanks are due to AFSOR and RDL and the TTU administration for their financial support for this research effort.

REFERENCES:

1. Boyle, E., Ianni, J. Easterly, Harper,S., and Korna, M., Human Centered Technology for Maintainability: Workshop Proceedings, Air Force Systems Command. Report # AL-TP-1991-0010, June 1991.
2. Deivanayagam, S., Hand Torque Strength for Small Fasteners, Report to AFOSR for Summer Research Program, August 1992.
3. Deivanayagam, S., and Weaver, T., Effects of Handle Length and Bolt Orientation on Torque Strength Applied During Simulated Maintenance Tasks, Trends in Ergonomics/Human Factors V, 1988, pp 827-833.
4. Gibbons, L.E., Summary of Ergonomics Research for the CREW CHIEF model Development (U), Armstrong Aerospace Medical Research Laboratory, Report # AAMRL-TR-90-038, December 1989.
5. Griener, T.M., Hand Anthropometry of U.S. Army Personnel, Technical Report Natick/TR-92/011, US Army Natick RDE Center, MA 01760, December 1991.
6. McDaniel, J.W., and Askern, W.B., Computer-Aided Design Models to Support Ergonomics, Ergonomics International 85, Proceedings of IEA Congress, September 1985, pp. 442-444.

DETERMINATION OF TOTAL PERIPHERAL RESISTANCE, ARTERIAL AND VENOUS COMPLIANCE DURING CLOSELY-COUPLED CHANGES IN Gz FROM 0-9 Gz.

Daniel L. Ewert
Assistant Professor
Department of Electrical Engineering

North Dakota State University
Fargo, North Dakota 58105

Final Report for:
Research Initiation Program
Armstrong Laboratory

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

and

North Dakota State University

December, 1993

DETERMINATION OF TOTAL PERIPHERAL RESISTANCE, ARTERIAL AND VENOUS COMPLIANCE DURING CLOSELY-COUPLED CHANGES IN Gz FROM 0-9 Gz.

Daniel L. Ewert
Assistant Professor
Department of Electrical Engineering
North Dakota State University

Abstract

A new mathematical technique was developed to estimate total peripheral resistance (TPR), systemic arterial compliance (SAC) and venous compliance (VC) during transient conditions throughout a continuum of Gz stress from 0 to 9 Gz. Due to technical problems VC could not be estimated. The mathematical technique was verified using an electrical circuit analog. Physiological estimates were obtained for TPR and SAC for centrifuge runs of 2, 3, 5, 7, 9 Gz and for parabolic flight where the Gz stress varied from 0 to 2 Gz. The estimates of TPR and SAC showed definite trends as Gz varied. Calculations of TPR and SAC were performed on Gz-suited and unsuited baboons. Also, differences were found in TPR and SAC between different states of hydration. It appears that this approach can detect small changes in TPR and SAC which may help in further understanding of the basic cardiovascular coping mechanisms during variations in Gz.

INTRODUCTION

Background

It is not well understood how closely-coupled changes in the +Gz field affect the physiology of the cardiovascular system - specifically, total peripheral resistance (TPR), systemic arterial compliance (SAC) and venous compliance (VC). Furthermore, the use of anti-g suits and straining maneuvers complicate understanding the relation between changes in +Gz stress and TPR, SAC and VC. Key to minimizing the biological barrier in high performance aircraft is fully understanding these relations.

The main obstacle in understanding effects of closely-coupled changes in +Gz stress is that the familiar steady-state analysis will not work. For example, the usual calculation of TPR by dividing the mean pressure by the mean flow produces faulty results during transient +Gz events. Other investigators have developed mathematical techniques to deal with estimating TPR and SAC. However, these approaches have limitations which will be discussed in detail later.

Purpose

The principal objective of the proposed research is to validate and apply the newly developed technique to predict TPR, SAC and VC during closely-coupled changes in +Gz across a continuum from 0 to +9 Gz.

Scope

First, a review of the tasks completed during the 1992 AFOSR Summer Research Program will be discussed. Included in the development of the

technique are:

1) Literature Summary

2) Model description

3) Mathematical Development

Second, a description of the tasks performed under the AFOSR Research Initiation Grant will be given. These tasks include:

1) Validation of the Technique

2) Application of the Technique.

TASKS COMPLETED (1992 AFOSR Summer Research Program)

Literature Summary

Other investigators have developed methods to estimate TPR and SAC during transient conditions. Perhaps Toorop et. al (1987) best characterized unsteady TPR and SAC by using a Newton fitting procedure. This is an optimization procedure which is computationally time-consuming and, if the data are noisy, can become "stuck" in local minima which yield erroneous values of TPR and Ca. Liu et. al (1986) developed a method which relies on the shape of the diastolic portion of the aortic wave form. With this information, and assuming steady conditions, the arterial compliance can be calculated. Due to these approaches, these techniques are not reliable during transient (unsteady) events such as short-term centrifuge runs and parabolic flight profiles.

Model Description

Shown in Figure 1 is the electrical circuit model of the circulatory system. From this model, equations can be developed to predict TPR, SAC and VC.

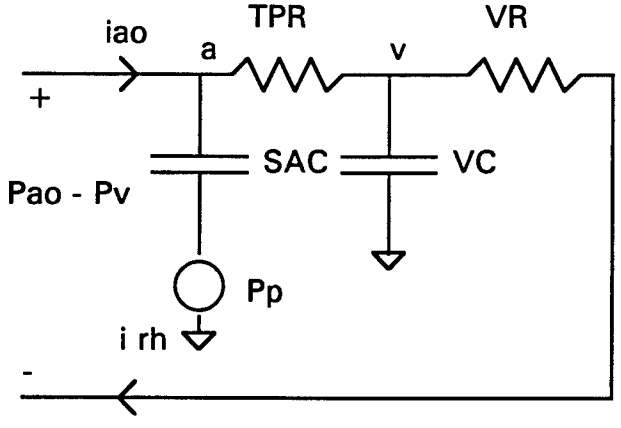


Figure 1: Circuit Analog of Systemic Circulation

Definitions:

i_{ao} = aortic flow

SAC = "Lumped" arterial capacitance

TPR = "Lumped" arterial resistance

VC = "Lumped" venous capacitance

VR = "Lumped" venous resistance

i_{rh} = Right heart flow

P_p = Pleural pressure

Mathematical Development

Assumptions:

(1) SAC, TPR, VC, VR, are constant over one beat

(2) $R_a >> R_v$ ($P_{Ra} \gg P_v$)

(3) Total blood volume = constant

Applying Kirchoff's current law at node A_a yields the following equation.

$$i_{ao} = SAC \frac{d(P_{ao} - P_p)}{dt} + \frac{P_{ao} - P_v}{TPR}$$

i_{Ao}, P_{ao}, P_p and P_v are experimentally measured.

SAC and TPR are unknowns.

It should be noted that if the beat is steady, the net flow into the capacitor will be zero and the equation collapses to the clinical definition of TPR :

$$TPR = \frac{P_{ao} - P_v}{i_{ao}}$$

Applying Kirchoff's current law at node V yields the following equation:

$$i_{RH} = \frac{P_{\infty} - P_v}{TPR} - VC \frac{dP_v}{dt}$$

Next, integrate both equations over two intervals

$$\int_{t_0}^{t_1} i_{\infty} dt = SAC\left[\left(P_{\infty}(t_1) - P_p(t_1)\right) - \left(P_{\infty}(t_0) - P_p(t_0)\right)\right] + \frac{1}{TPR}\int_{t_0}^{t_1}\left(P_{\infty} - P_v\right) dt$$

$$\int_{t_1}^{t_2} i_{\infty} dt = SAC\left[\left(P_{\infty}(t_2) - P_p(t_2)\right) - \left(P_{\infty}(t_1) - P_p(t_1)\right)\right] + \frac{1}{TPR}\int_{t_1}^{t_2}\left(P_{\infty} - P_v\right) dt$$

$$\int_{t_0}^{t_1} i_{RH}\, dt = \frac{1}{VR}\int_{t_0}^{t_1}\left(P_{\infty} - P_v\right) dt - VC\left[P_v(t_1) - P_v(t_0)\right]$$

$$\int_{t_1}^{t_2} i_{RH}\, dt = \frac{1}{VR}\int_{t_1}^{t_2}\left(P_{\infty} - P_v\right) dt - V\,C\left[P_v(t_2) - P_v(t_1)\right]$$

Here the equations become

$$SV_{RH} = V_{VC} + V_{VR}$$

Where:

SV_{RH} = Stroke Volume of Right heart

V_{VC} = Volume in venous compliance

V_{VR} = Volume through arterial resistance

Thus for arterial side there are two equations and two unknowns and for the venous side one obtains two equations and two unknowns.

After investigating the many choices of intervals, it was found that to get reliable estimates of SAC, the intervals which produce maximum aortic pressure changes were best, as long as the maximum pressure occured in the first 30 - 45% of the beat length. Thus t_o was selected at the beginning of the beat, t_2 was selected at the end of the beat and t_1 was selected at the time of maximum pressure (or about 40% of the beat).

TASKS COMPLETED (AFOSR Research Initiation Program)

Validation

It was important to verify that the analysis technique produces results which reflect real changes in known values of TPR and SAC. This is difficult to do in-vivo, but a simple method to accomplish this was to use an electric circuit analog with variable resistors and capacitors as was shown in Figure 1.

A voltage proportional to aortic pressure was input to the circuit and the concomitant current (flow) was measured for a series of TPR and SAC values. The analysis technique was applied to the pressure and flow waveforms to estimate TPR and SAC on a per-beat basis. A beat is defined as one full pressure and flow oscillation. One can see from Figures 2 - 4 that the predicted values closely matched the actual values for a variety of conditions. Figure 2 shows that given an "medium" value of TPR (R) and SAC (C), the technique will closely predict R and C. The larger circles represent the actual values and the small circles represent estimated values. The input voltage (pressure) was varied both sinusoidally and by a slowly rising and then slowly falling ramp. Even with large variations in input voltage and current, the technique was able to estimate charge (volume). Figure 3 demonstrates how the technique performed when TPR (R) and SAC (C) are changing. Here the excitation voltage was a sinusoid with a positive offset so that no negative voltages were experienced (0 - +20V). Once again R, C, and charge were closely predicted by the technique. Figure 4 shows what happens when the changes in R and C are in opposite directions and the circuit is excited by a sinusoidal voltage (0 - +20V). R, C and charge from the

Figure 2

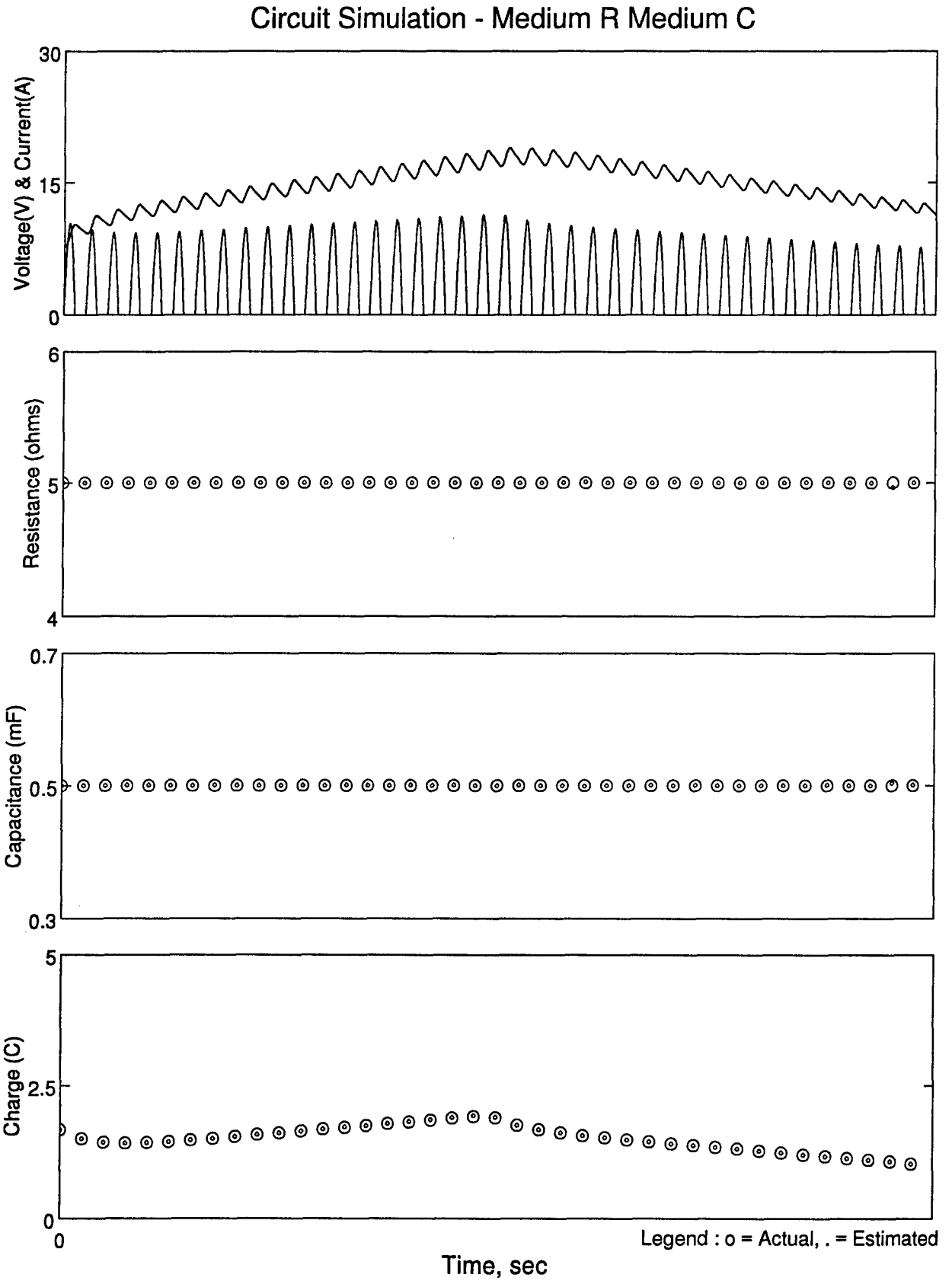

Circuit Simulation - Medium R Medium C
Voltage(V) & Current(A)
30
15
0
Resistance (ohms)
6
5
4
Capacitance (mF)
0.7
0.5
0.3
Charge (C)
5
2.5
0
0
Legend : o = Actual, . = Estimated
Time, sec

Figure 3

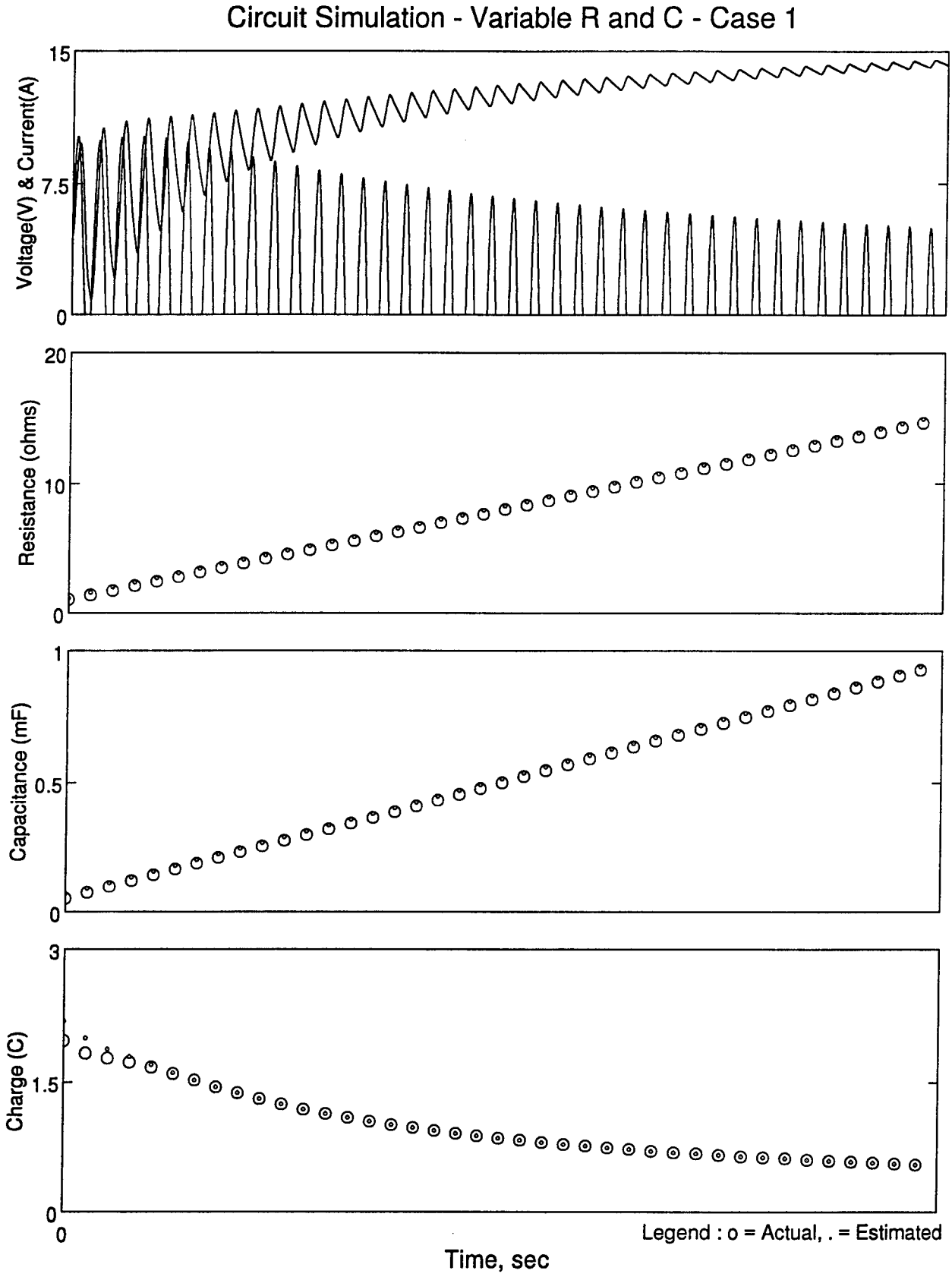
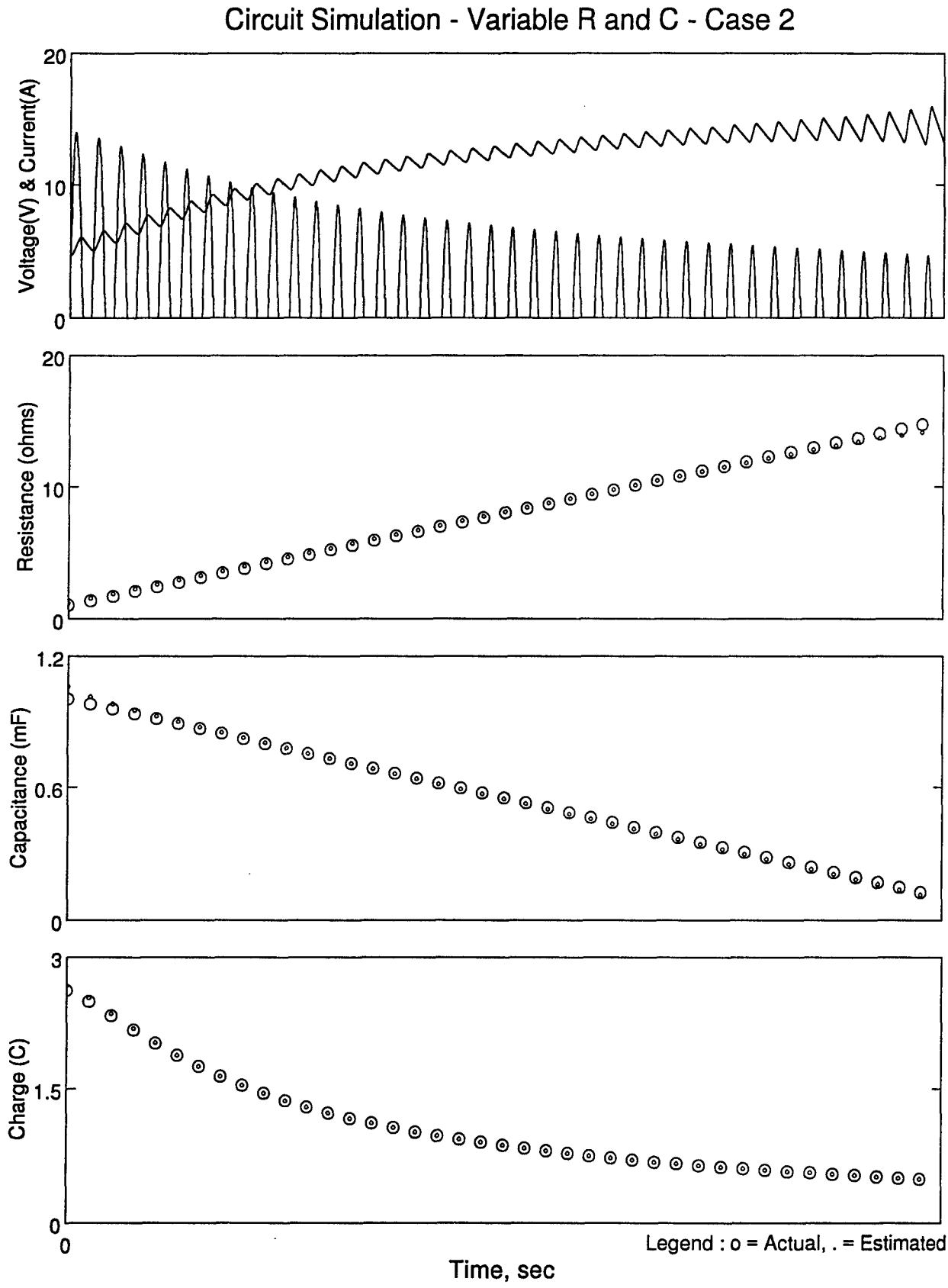

Circuit Simulation - Variable R and C - Case 1
Voltage(V) & Current(A)
15
7.5
0
Resistance (ohms)
20
10
0
Capacitance (mF)
1
0.5
0
Charge (C)
3
1.5
0
0
Legend : o = Actual, . = Estimated
Time, sec

Figure 4


Circuit Simulation - Variable R and C - Case 2
Voltage(V) & Current(A)
20
10
0
Resistance (ohms)
20
10
0
Capacitance (mF)
1.2
0.6
0
Charge (C)
3
1.5
0
0
Legend : o = Actual, . = Estimated
Time, sec

voltage and current waveforms were predicted accurately by the technique.

The susceptibility of the technique to added noise (zero-mean) is shown in Figure 5. Here the error in % is defined as

$$\text{error}(\%) = \left| \frac{\text{predicted} - \text{actual}}{\text{actual}} \right|$$

Sigma was defined as

$$\text{sigma of added noise} = \frac{\text{std dev of noise}}{\text{maximum of input data}}$$

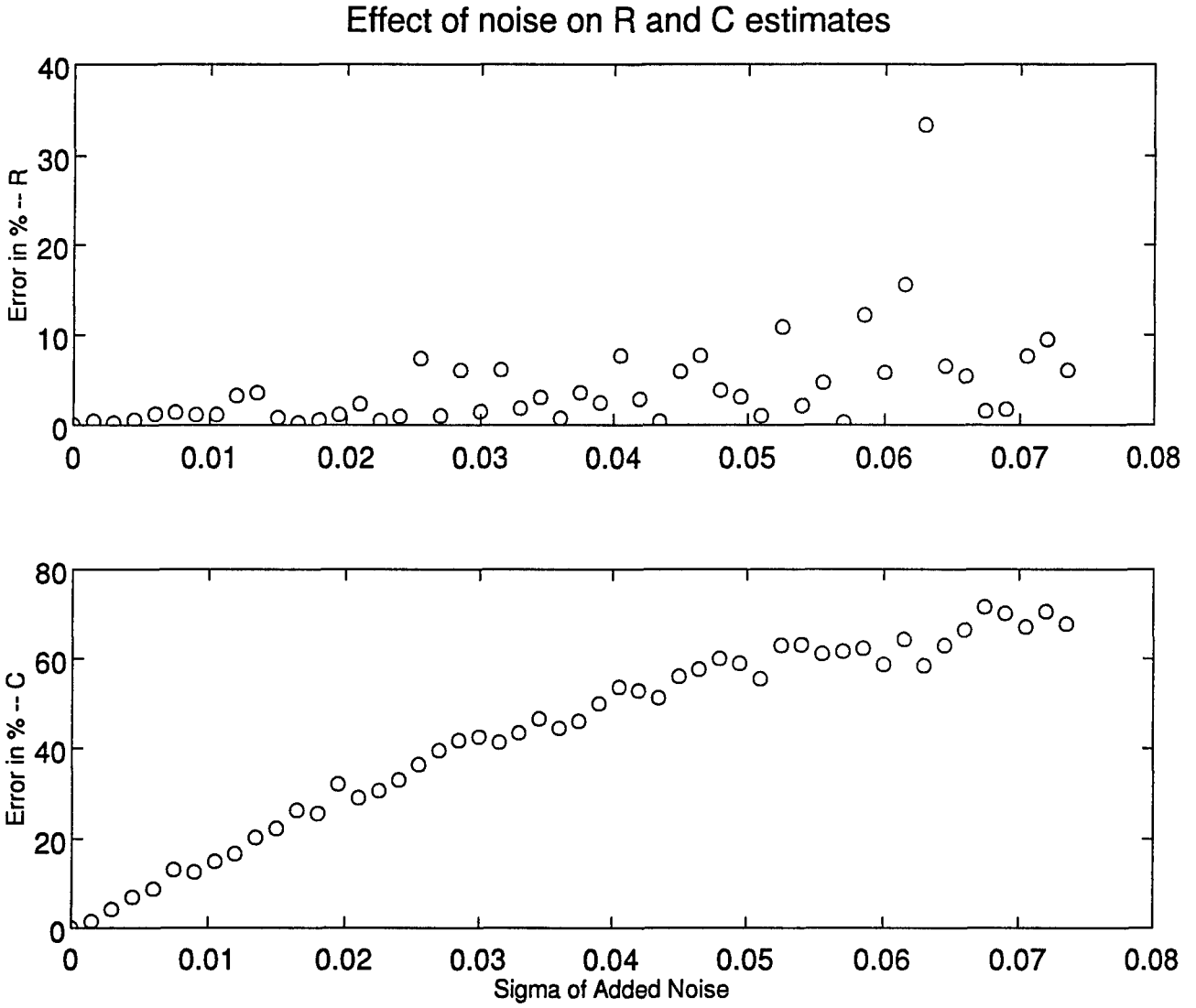
The maximum of input data is the maximum of voltage and current respectively.

It is no surprise that the SAC calculation was more sensitive to noise because it relied only on the endpoints in pressure whereas TPR is based on the integration of pressure.

It was expected that the agreement between the actual and predicted values would be close because the electrical circuit follows the physical laws used to estimate the values. However, it is not known whether or not the 2-element model is a good approximation of the physiological circulatory system. Since a lumped model is an abstraction of the real system, one cannot anatomically locate a "real" TPR and SAC. Thus one cannot directly check the validity of the 2-element model. Therefore, an indirect method must be used to determine the effectiveness of the 2-element model.

Most other investigators have input a known (measured) aortic pressure waveform to an electrical circuit model and an aortic flow can be calculated

Figure 5


Effect of noise on R and C estimates
Error in % -- R
0
10
20
30
40
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Error in % -- C
0
20
40
60
80
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Sigma of Added Noise

by using estimated TPR and SAC. The generated flow should closely match the physiological flow if the model is good. It is our opinion that flow comparison is not a good indicator, because most 2 and 3 - element windkessel models do not account for wave reflections known to occur in the circulatory system. Rather we proposed that to indirectly validate the 2-element model, volumes should be used instead of flows. Recall, that the original equations were integrated to yield

$$SV = V_{SAC} + V_{TPR}.$$

If the aortic waveform (measured) is integrated the actual SV can be determined. Given an aortic pressure waveform and the estimated parameters (TPR, Ca), V_C and V_r can be estimated.

$$V_{TPR} = \frac{1}{TPR}\int (P_{ao} - P_v)dt$$

$$V_{SAC} = SAC(P_2 - P_1)$$

Where:

$P_2 = P_{ao} - P_p$ (at the end of a beat)

$P_1 = P_{ao} - P_p$ (at the beginning of a beat)

Adding V_C and V_r produces an estimate of SV which was compared to the integral of the aortic flow (measured) and a comparison between actual and estimated stroke volume was used as a mark of "goodness" in applying the technique to physiological data.

Application of the Technique

Centrifuge

A short duration $+G_z$ (2, 3, 5, 7, and 9) run was conducted on baboons. Due to instrumentation problems (pulmonary flow) the venous system parameters could not be checked. The results are shown in Figures 6 and 7. In Figure 6 (no G-suit) TPR increased as +Gz stress was applied and SAC fell at onset of $+G_z$. This response is what one would expect. The resistance begins to increase probably due to the baroreceptor reflex. It appears that the greater the Gz stress the larger the increase in resistance. After the high $+G_z$ episode the resistance begins to decrease, probably due to reduced baroreceptor reflex activity. SAC decreases may be due to stretching of the main arterial blood vessels during Gz exposure - making them more rigid. As pressure increased in the arterial system the capacitance increased. In the bottom graph, stroke volumes were compared and the result showed good agreement. Figure 7 depicts a similar Gz profile except that the subject has a g-suit. Here aortic blood pressure increases dramatically when the g-suit is activated. It appears that TPR rises sharply with the onset of the g-suit activation (the activation of the g-suit coincides with the rapid increase in aortic pressure). Once again the capacitance fell, meaning the arterial system became more rigid - just as one would expect when the legs and abdomen are squeezed. The agreement between estimated and actual stroke volume was quite good. Sometimes estimates of TPR, SAC and stroke volume are missing because of poor signal quality during portions of the experiment.

Figure 6

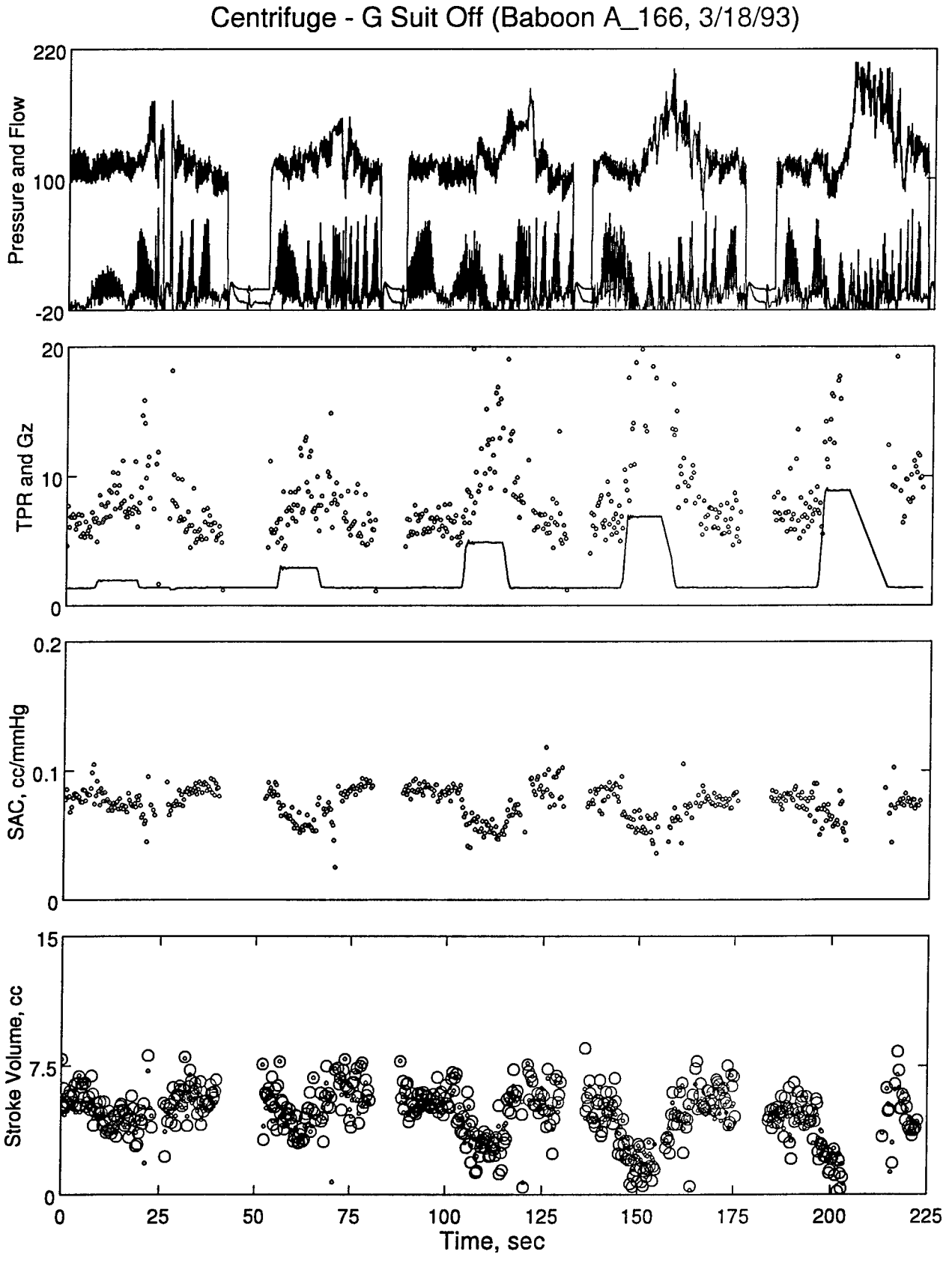
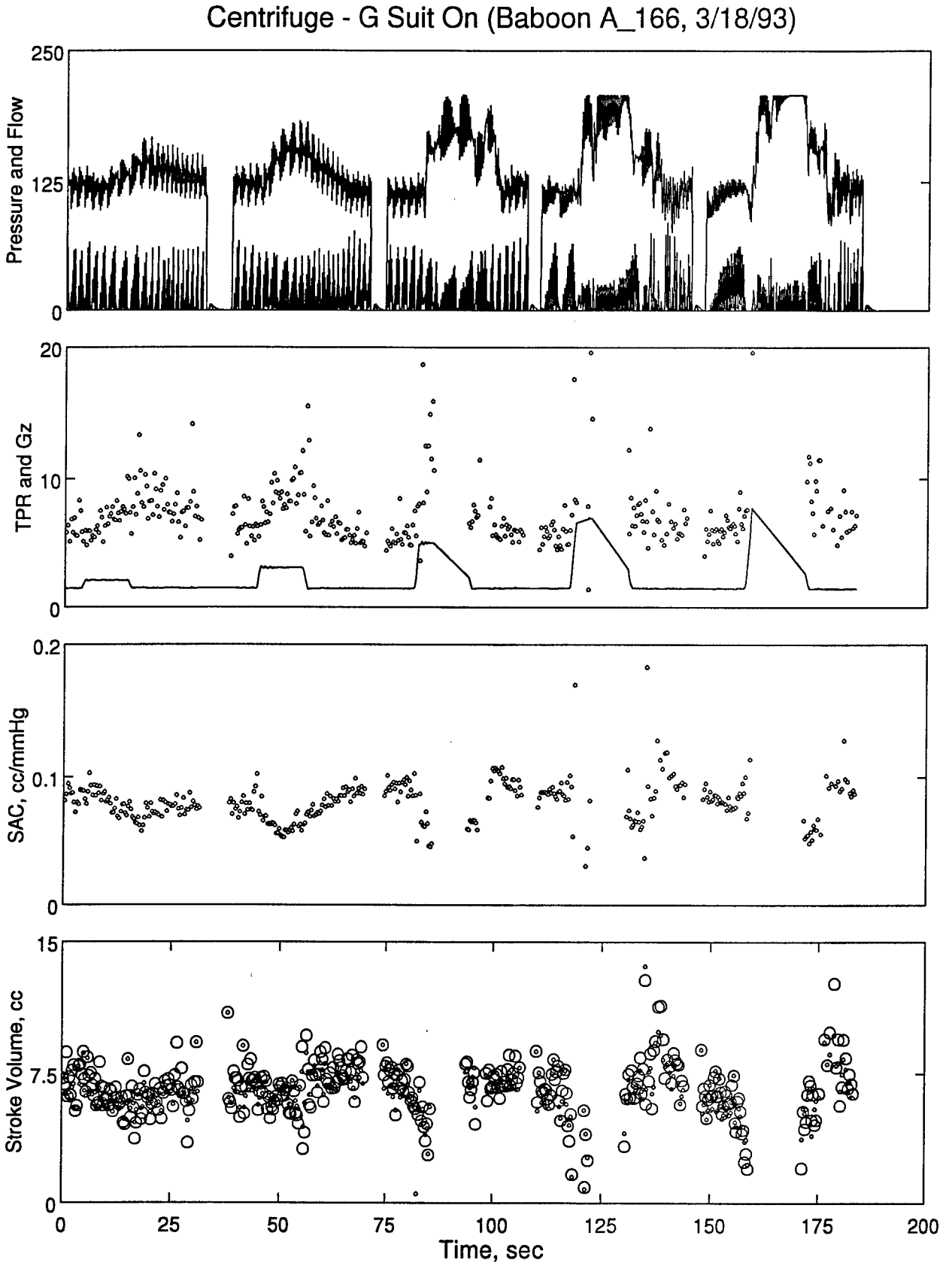

Centrifuge - G Suit Off (Baboon A_166, 3/18/93)
Pressure and Flow
220
100
-20
TPR and Gz
20
10
0
SAC, cc/mmHg
0.2
0.1
0
Stroke Volume, cc
15
7.5
0
0
25
50
75
100
125
150
175
200
225
Time, sec

Figure 7


Centrifuge - G Suit On (Baboon A_166, 3/18/93)
Pressure and Flow
250
125
0
TPR and Gz
20
10
0
SAC, cc/mmHg
0.2
0.1
0
Stroke Volume, cc
15
7.5
0
0
25
50
75
100
125
150
175
200
Time, sec

Parabolic Flight

Figures 8 - 9 show the effect of 0 - 2 +Gz stress on TPR and SAC. Figure 8 shows the response of a volume-depleted animal and Figure 9 shows the response of the same animal to a volume-loaded condition. As one would expect, the stroke volume in the volume-depleted condition was less than that of the volume-loaded condition. Also, there was much more variation on the stroke volume in the volume-depleted state. TPR varied directly with Gz in the volume-depleted state and in the volume-loaded state little variation in TPR was found. TPR was higher (on average) in the volume-depleted state - probably because when volume is lower a higher TPR is necessary to maintain pressure. Capacitance was higher in the volume-depleted state. Note the acute change in capacitance during the first few seconds of the volume-loaded condition. This change is probably due to the addition of volume into the animal just previous to this set of parabolas. Figure 10 shows the response of the same animal on another day (2 days prior) in a "normal volume condition". These responses appear similar to the volume-depleted responses. Later it was learned that the animal was not euvolemic but was very close to a dehydrated state by measurement of blood volume using Evan's blue dye.

The technique was implemented in MATLAB (@ Mathworks) and perform TPR, SAC and stroke volume calculations for 200 beats took about 1 - 2 minutes on a 486 50MHz PC. Previously, on a similar machine at the Laboratory for Aerospace Cardiovascular Research using another software package, the same analysis would have taken days.

Figure 8

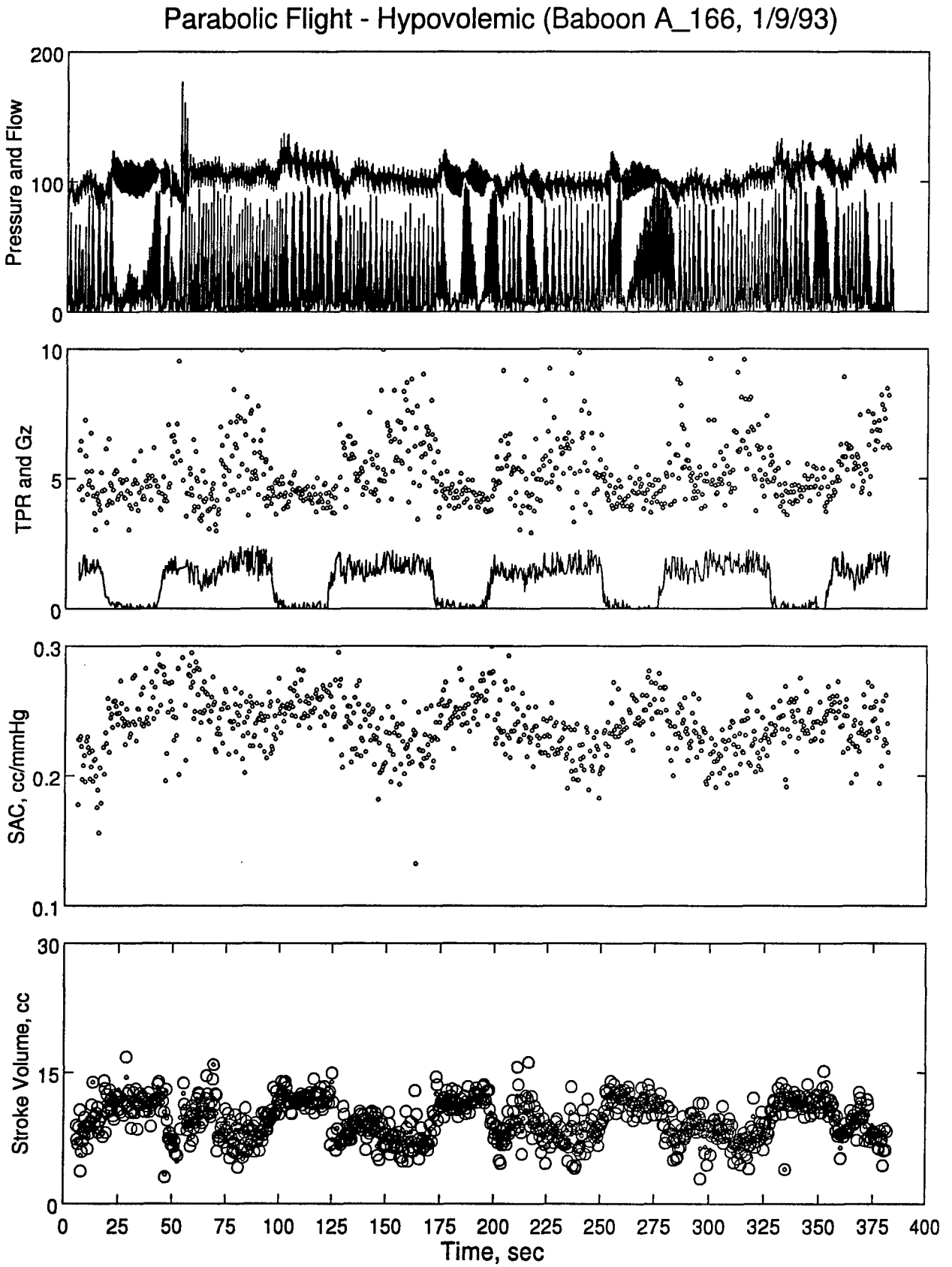

Parabolic Flight - Hypovolemic (Baboon A_166, 1/9/93)
Pressure and Flow
200
100
0
TPR and Gz
10
5
0
SAC, cc/mmHg
0.3
0.2
0.1
Stroke Volume, cc
30
15
0
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400
Time, sec

Figure 9

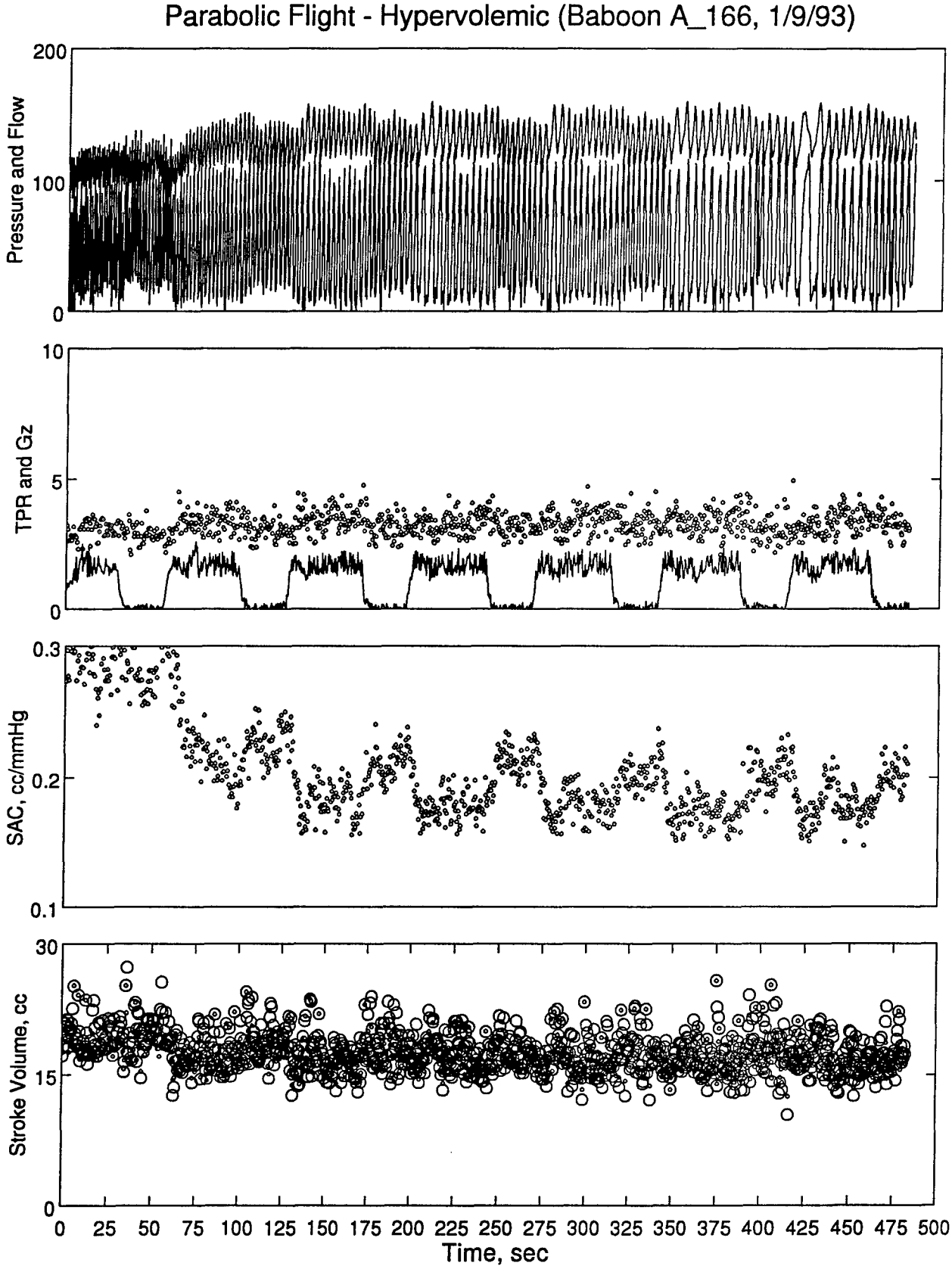
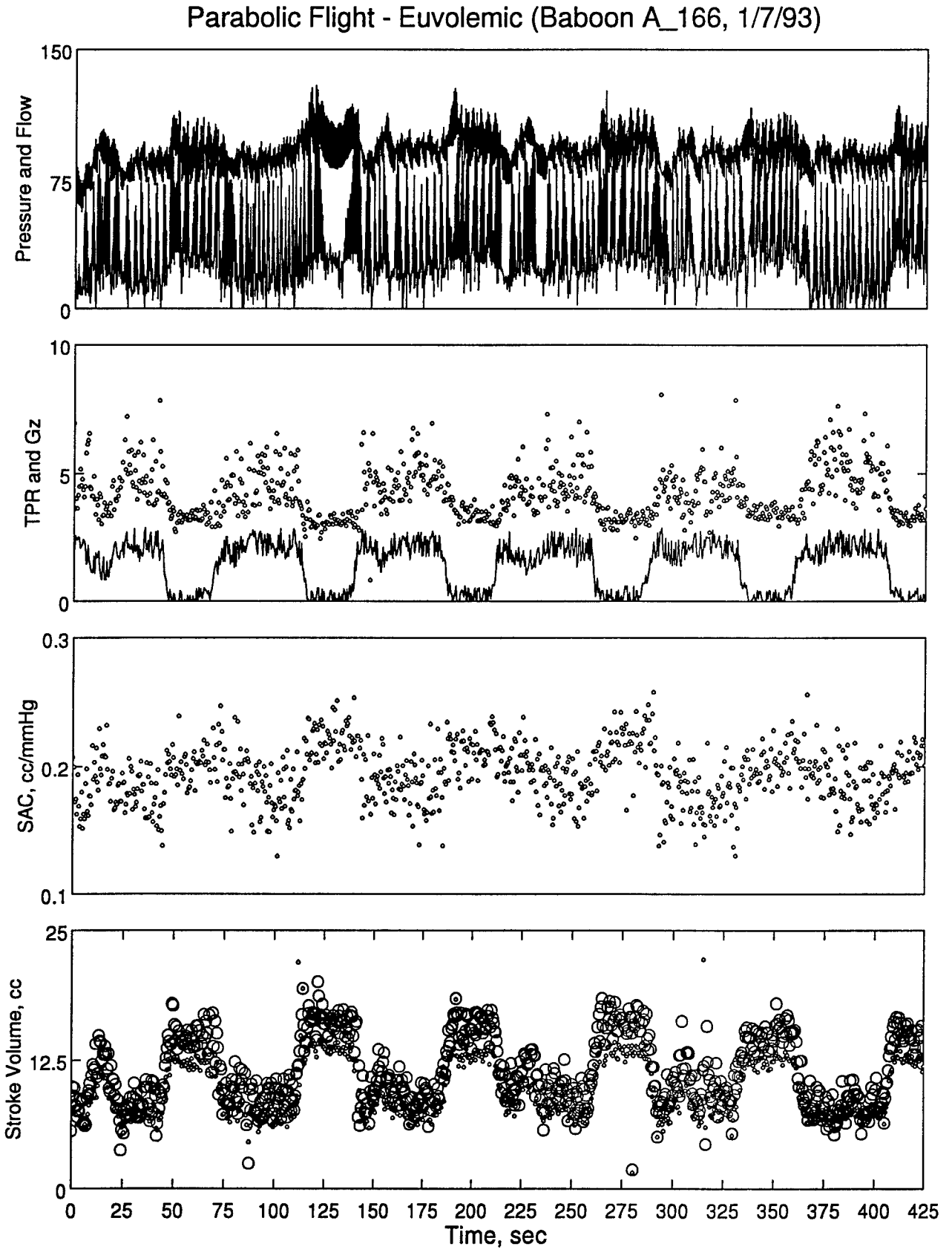

Parabolic Flight - Hypervolemic (Baboon A_166, 1/9/93)
Pressure and Flow
200
100
0
TPR and Gz
10
5
0
SAC, cc/mmHg
0.3
0.2
0.1
Stroke Volume, cc
30
15
0
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
Time, sec

Figure 10


Parabolic Flight - Euvolemic (Baboon A_166, 1/7/93)
Pressure and Flow
150
75
0
TPR and Gz
10
5
0
SAC, cc/mmHg
0.3
0.2
0.1
Stroke Volume, cc
25
12.5
0
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425
Time, sec

CONCLUSION

The purpose of the study was to validate and apply a new mathematical technique to estimate TPR, SAC and VC. A VC estimation was not possible due to technical problems in measuring pulmonary artery flow. Using a electrical circuit analog the technique was validated.

Physiological estimates were obtained for TPR, SAC and stroke volume, and initial results look promising in that they show changes in their values that can be explained in physiological principles. It is important to be reminded that no "gold standard" exists for TPR and SAC, but if the two element model is a good model of the physiological system, then the method will be a useful tool. To know if the two element model is a good model of the physiological system, indirect means must be employed to determine the validity of the technique. To this end, it was thought that if the model predicted stroke volume, this would be one test to provide indirect evidence to support the validity of technique. For these examples it appears that stroke volume as predicted by the model matches the actual stroke volume. Other information that is reassuring is that changes that occur in TPR, SAC and stroke volume do make sense from known physiological principles. However, since a "lumped" TPR and SAC do not exist, it is impossible to know for sure that the technique is valid. More experimental conditions must be examined to be confident that the technique yields appropriate results.

The method gives reasonable results and one should be optimistic about its future possibilities. One such possibility includes developing a TPR and SAC meter so that if aortic pressure and aortic flow signals are connected to a "box", the output will be near real time TPR and SAC measurements. This way, the results of the experiment will be available while the experiment is in progress.

Acknowledgments:

This author would like to thank the team from the Laboratory for Aerospace Cardiovascular Research for their support. Furthermore, the major intellectual contributions of Dr. R.D. Swope, Dr. V. Phaniraj, Dr. R. Crisman , Dr. R. D. Latham and Capt. D. A. Self were greatly appreciated.

REFERENCES

Liu, Z., Brin, K. P., and F. C. P. Yin, "Estimation of Total Arterial Compliance: an Improved Method and Evaluation of Current Methods", American Journal of Physiology, 251 (Heart, Circ. Physiol. 20), H588-H600, 1986.

Toorup, G. P., Westerhof, N., and G. Elzinga, "Beat-to-Beat Estimation of Peripheral Resistance and Arterial Compliance during Pressure Transients", American Journal of Physiology, 252, (Heart, Circ. Physiol. 21), H1275-H1283, 1987.

A COMPUTATIONAL THERMAL MODEL AND THEORETICAL THERMODYNAMIC MODEL OF LASER INDUCED RETINAL DAMAGE

Bernard S. Gerstman
Associate Professor
Department of Physics

Florida International University
University Park
Miami, FL 33199

Final Report for:
Research Initiation Program
Armstrong Laboratory

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

and

Florida International University

April 1994

A COMPUTATIONAL THERMAL MODEL AND THEORETICAL THERMODYNAMIC MODEL OF LASER INDUCED RETINAL DAMAGE

Bernard S. Gerstman
Associate Professor
Department of Physics
Florida International University

Abstract

The immediate thermodynamic effects of absorption of a laser pulse in the retina were studied. A computer model was developed in order to determine the spatial and temporal profiles of temperature rises. The absorption occurs in a retinal pigment epithelium cell modeled as an aqueous environment with absorption occurring at small spherical sites with absorption coefficients representative of melanosomes. Laser pulse characteristics are user controlled input variables. The computational model outputs temperature rises in the surrounding aqueous medium during and after the pulse. For short pulses, energy is deposited sufficiently quickly for the temperature to rise high enough for vaporization to occur. The growth of the bubbles produced is determined by the laser pulse input parameters and properties of the absorbing melansomes and the surrounding medium. The growth of a bubble that forms around a model melanosome is also studied. These temperature rises and pressure effects are expected to be important factors in laser induced damage to retinal cells.

A COMPUTATIONAL THERMAL MODEL AND THEORETICAL THERMODYNAMIC MODEL OF LASER INDUCED RETINAL DAMAGE

Bernard S. Gerstman

INTRODUCTION

An important ramification of the continual progress in laser development is the increased danger of injury to the visual system. This danger of injury is present in military and civilian use and can be either accidental or intentional. The potential for injury, and the importance of understanding how it occurs, continues to grow as a result of the steady advancement in laser technology which provides increasing pulse energies and shorter pulse durations.

The research described here was undertaken in order to understand the primary effects of the laser energy immediately after absorption in the retina. The main absorption site for intense laser pulses is expected to be the melanosomes in the retinal pigment epithelium (RPE). A computer model has been developed in order to determine the temperature rise of the melanin and the surrounding cellular material due to absorption of the laser pulse. This model incorporates melanosomes as small spherical absorption sites that are 1 micron (10^{-6}m) in radius and have absorption characteristics of melanin, embedded in an homogeneous non-absorbing medium with the thermal characteristics of water to represent the cell.

In addition to the temperature rise, damage can be caused by bubble formation. Significant heat conduction away from a melanin granule requires timescales of the order of microseconds. For pulses of shorter duration than this, temperature rises are expected to be high enough to cause vaporization of the surrounding medium. Expressions for the expected growth in bubbles surrounding the melanosomes are used to determine the maximum sizes. The temperature rises and bubble formation are expected to be two immediate consequences of the laser absorption that

lead to retinal damage.

This work was done in collaboration with Dr. M. Rogers, Lt. Colonel United States Air Force, Dr. R. Thompson, Captain, United States Air Force, and Dr. S. Jacques. The implementation of the thermal model for temperature rise and the associated computer code was carried out by Dr. Thompson.

OVERVIEW OF DAMAGE MECHANISMS

Laser induced functional impairment to the visual system affecting task performance will occur if there is damage to photoreceptors or their associated nerve transmission pathways. Damage to these biological cells can result from a variety of causes. Some of the possibilities which were not investigated in this work were chemical or phototoxic effects in which absorption of photons causes electronic excitations which either rearrange chemical bonds to produce noxious chemical reactions or break chemical bonds which causes molecules in biochemical pathways to become inactive. It is also possible that the electric fields in extremely intense laser pulses are strong enough to electrically disrupt cellular membranes.

This is a report on investigations of the initial thermal and vaporization effects from the absorption of laser pulses. Absorption of the laser energy by a melanosome causes a rise in temperature of the melanosome and leads to two important mechanisms for cellular damage: conduction of heat from the melanosome into the cellular material, and formation and expansion of a bubble around the heated melanosome. Both the heating of the cellular material and the bubble expansion will lead to damage to cellular components and disruption of the normal cellular functioning.

This report will describe the computer model that has been developed for determining the rise in temperature of the cellular material

surrounding a melanosome that has been irradiated by a laser pulse. This computational model allows a user to specify as input the characteristics of a laser pulse. The output is a profile of the temperature rise of the surrounding material as a function of position and time. For laser pulses shorter than the characteristic heat conduction time away from the melanosome (microseconds) it is expected that the heat absorbed by a melanin granule will cause the temperature to rise above 100 C. This will lead to bubble formation around the melanosome. The creation and expansion of this bubble will be an additional source of damage to the cell. This report also describes the calculation of the characteristics of the bubble that will result from absorption of a sub microsecond laser pulse by a melanin granule.

THERMAL COMPUTATIONAL MODEL

Previous Computational Models

Computer programs exist [1] that calculate temperature rises in the retina due to absorption of laser pulses. These models have a common feature which limits their usefulness and which we attempt to improve upon with our melanosome absorption model.

The important feature which limits the usefulness of early models is their approach in modeling the absorption characteristics of the retina. While absorption of visible light occurs in the photoreceptors, and is the primary event in the visual process, the photoreceptors absorb only 5% of incident visible light. Absorption of light by the photoreceptors can lead to damage to the photoreceptors through photochemical and phototoxic mechanisms such as permanent bleaching, but this requires long pulses ($\tau >$ 1 sec) with short wavelengths and high corneal irradiance (5 J/cm^2 at 325

nm, 20 J/cm² at 442 nm, 3000 J/cm² at 632 nm) [2]. These energies are all well above thresholds for damage, and imply that other absorption sites can result in retinal damage at lower irradiances.

The most important site for retinal damage is the retinal pigment epithelium which absorbs 50% of incident visible radiation [3], an order of magnitude more energy than the photoreceptors. This strong absorption by the melanosomes makes the RPE the likely location for the source of temperature rises and bubble formation that can lead to damage to the retina at threshold levels of irradiance [4,5]. Evidence that near threshold damage is centered in the RPE was observed by Gueneau, et. al. [6]. Additional absorption can also occur in the choroid, but is less likely to cause damage since it is located posterior to the RPE and thus receives less incident light than the RPE, and the flow of blood supplies a mechanism for heat removal.

The retina is composed of different layers of cells. Previous computer models modeled the retina as composed of layers but with each layer treated as homogeneous with uniform absorption properties. Rather than having the absorption of light in the RPE occurring at the individual melanosomes, the same amount of energy absorption was spread out over the entire monolayer of RPE cells. Though the same amount of total energy is deposited in the same overall volume, the dispersing of the absorption leads to less concentration of heat than occurs when the absorption is concentrated in the melanosomes. The dramatic effects of concentrated energy are therefore not fully represented, which we attempt to correct with our model.

Details of Computer Model with Localized Melanosome Absorption

The basic component of our computer model is the localized absorption of light in the melanosomes. The melanosomes are represented as spheres

of radius R_a with absorption coefficient α (both of these properties are user controlled). The spheres, and the surrounding cellular medium are given the thermal characteristics of water; heat capacity c=4.19 x 10^3 J/kg•°C, density ρ=10^3 kg/m^3, thermal conductivity κ=0.57 J/m•s•°C. The heat equation, with the absorbed energy represented by a source term, can then be solved using Laplace transforms [7,8] to get the temperature rise as a function of time and radial distance from the spherical granule. This approach is used in the computer program to investigate the effects of temperature rises in the RPE layer [9]. The computer model allows the user to input the number of melanosomes per RPE cell. The importance of this feature is that it allows a systematic investigation of thermal effects as a function of melanosome number density and thus a comparison between individuals or between species. Additionally, the number of melanosomes can be varied from one run to the next around an average that is appropriate for a given species. This variability in melanosome number density introduces variability in whether or not damage will occur for constant laser pulse parameters, and thus allows the determination of an ED_{50} and comparison with probit analysis curves.

Initial runs have been carried out to determine the effectiveness of the program. This was done for comparison with the work of Takata [1] with laser parameters for which the Takata homogeneous absorption model and the melanosome absorption model are expected to give similar results. This occurs for pulse durations long enough for heat to diffuse from melanosomes throughout the cell during the pulse and thus resemble temperature profiles expected from homogeneous absorption. Heat conduction over cellular distances of 10 μm occurs on time scales from 10^{-6}-10^{-3} seconds. Therefore, a much longer pulse duration of τ_p=0.1 second was chosen. A pulse of this duration has a corneal ED_{50}=0.00273 J/cm².

Takata used an Arrhenius damage integral approach, first used in this

context by Henriques [10]. This assumes that damage builds up with an Arrhenius type rate and can be expressed as

$$\Omega = \int^{t} e^{c_1 - c_2/T} dt \qquad (1)$$

The damage endpoint is defined to be $\Omega \equiv 1$. Takata used constants c_1=149 and c_2=50,000°K. The integration is over a period of time in which the temperature, in °K, is elevated above normal body temperature of 310°K. Since significant heat conduction can occur in a millisecond, a 0.1 second pulse causes a temperature rise that is effectively a square wave of duration 0.1 second. The initial temperature rise and final temperature drop occur in time scales that are short compared to the pulse duration. Using this pulse duration with the values of c_1 and c_2 in Eq. (2), Takata found that a temperature rise of 30.3°C was necessary to achieve Ω=1.

To test the melanosome model, the same ED_{50} was input for a 0.1 second pulse and the temperature rise of the cell was computed. The model cell was chosen to have dimensions representative of an RPE cell of 20μm x 20μm x 10μm. The melanosomes were treated as spheres of radius R_a=1μm with an absorption coefficient of 1800 cm^{-1}, characteristic for light of λ=532 nm [11, 12]. The laser pulse input was given a corneal irradiance of the known ED_{50}=0.00273 J/cm². The retinal irradiance was determined by multiplying by a factor of 10^5 and the beam had a gaussian profile with $1/e^2$ radius of 15μm. Finally, a melanosome number density of 100 melanosomes/cell was chosen as an order of magnitude estimate. Using these parameters, the computer model found that the center of the cell attained a temperature rise of 32°C, elevating it to a temperature of 342°K for most of the pulse. This is in excellent agreement with the

results 30.3 °C of the Takata model, confirming the similarity of the two at long pulse durations.

This model can now be used for investigations in regimes where the Takata model is not expected to be valid; specifically, temperature profiles for pulses in the millisecond range in which the volume within a few μm's of a melanosome is expected to reach significantly higher temperatures than parts of the cell that are not near a melanosome. In this situation, the average cellular temperature given by the Takata homogeneous absorption model will not reflect the hot spots near the melanosomes. The present computer model will exhibit these localized hot spots, which may produce damage at irradiances that are lower than the irradiances necessary to raise the average temperature of the entire cell above a critical value. Additionally, the user defined melanosome number density allows variability in the cellular energy absorption. This can be used to investigate difference between different species, as well as allowing for the determination of probable damage levels, i.e. ED_{50}'s, in a single species by varying the melanosome number density around a species specific average, e.g. n_{ave}=100/cell with a standard deviation of 15/cell.

This computer program will be extremely valuable for pulse lengths long enough that conduction is effective in preventing the melanosome from reaching 100°C. For pulse lengths less than approximately 10^{-5} seconds, the experimentally measured ED_{50}'s lead to temperature increases that will cause vaporization of the cellular medium around the melanosomes. The computer model described above has no mechanism for phase changes and therefore does not give a full description in the short pulse regime. The remainder of this report describes the investigations into bubble formation and growth for ultrashort laser pulses.

BUBBLE FORMATION AND GROWTH

Thermodynamic Conditions of Growth

We now investigate bubble generation resulting from laser pulses with durations of less than a microsecond. During the laser absorption and bubble growth process, heat loss is unimportant. The melanosomes are approximately $1\mu m=10^{-6}m$ in radius with bubbles developing around them, and the cellular material surrounding the melanosome is treated as water. Using the thermal properties of water, we find that the approximate speed for heat conduction is on the order of:

$$v(\text{thermal}) \approx \kappa/\rho cL \approx 0.1 \text{ m/s} \qquad (2)$$

where $c=4.19 \times 10^3$ J/kg•°C, density $\rho=10^3$ kg/m^3, and thermal conductivity $\kappa=0.57$ J/m•s•°C as used earlier, and L=1 μm is used as a characteristic size. During a laser pulse of $\tau_p < 10^{-6}$ seconds, heat conduction occurs over a negligible distance from the melanosome and thus negligible heat loss occurs during absorption and subsequent bubble growth.

Bubble Growth

The adiabatic nature of the process allows the calculation of the important parameters characterizing the growth of a bubble. Under adiabatic conditions the relationship PV^γ=constant holds and therefore

$$\frac{V}{V_o} = \left(\frac{P_o}{P}\right)^{1/\gamma} \qquad (3)$$

where V_0 and P_0 are the initial volume and pressure of the bubble when it starts its adiabatic expansion at the end of the laser pulse, and γ is the ratio of the specific heat of the vapor at constant pressure to the specific heat at constant volume. Equation (3) leads to a relationship

between the initial radius of the bubble surrounding the melanosome, and the maximum radius. The radius of the melanosome itself will be denoted by R_a. The radius and pressure within the bubble will be r and P, and their values immediately after laser absorption will be r_o and P_o. The volume of the vapor is therefore

$$V(r) = 4\pi(r^3-R_a^3)/3 \qquad (4).$$

Combined with Eq. (3), this gives

$$r^3 = R_a^3 + (r_o^3-R_a^3)(P_o/P)^{1/\gamma} \qquad (5)$$

which can be rewritten in terms of dimensionless quantities as

$$(r/R_a)^3 = 1 + [(r_o/R_a)^3-1](P_o/P)^{1/\gamma} \qquad (6)$$

Equation (6) is an expression for the radius of a bubble as a function of the pressure of the vapor within the bubble.

Following the work by Cleary [13] the initial pressure P_o will be taken to be the critical pressure of water of 218 atmospheres. The reasoning behind this is as follows. The kinetics of vaporization is a non-equilibrium process, but eventually equilibrium will be reached between the vapor phase within the bubble and the liquid phase surrounding it. Since there are two distinct phases, the temperature and pressure cannot go above the critical values, which for water are T_c=374 C and P_c=218 atmospheres (221 bars) with a critical density of ρ_c=0.315 g/cm^3 [14]. Since the rate of energy input is faster than either the expansion of the bubble or the heat conduction rate, the critical conditions will at some time be reached if enough energy is absorbed by the melanosome to

produce vaporization. Therefore, bubble formation can be treated as a process in which the laser energy absorbed by the melanosome creates a saturated vapor with T_o=374°C and P_o=218 atmospheres and whose initial radius r_o is determined by the total energy absorbed by the melanosome. Bubble growth then occurs in an adiabatic expansion that is rapid compared to heat loss. (Note that the final volume of the bubble depends on the product of P_oV_o and that the energy of the laser pulse is used to both vaporize cellular fluid in creating V_o, and increase the initial pressure P_o. Thus, if the energy of the laser does not raise the pressure to a P_o=218 atmospheres, more cellular fluid can be vaporized leading to a larger V_o. This tends to limit the variation in the product P_oV_o and thus the final volume (and radius) are not extremely sensitive to the actual value used for P_o.)

In order to justify the adiabatic treatment of the bubble growth, the velocity of expansion must be much greater than the rate of heat loss. The characteristics of the expansion can be determined by following the treatment of Lamb [15] for the rate of expansion of the bubble radius

$$\dot{r}^2 = \frac{2c_o^2}{3(\gamma-1)}\left[\left(\frac{r_o-R_a}{r-R_a}\right)^3 - \left(\frac{r_o-R_a}{r-R_a}\right)^{3\gamma}\right] \qquad (7)$$

where $c_o=\sqrt{P_o/\rho}$ and ρ is the density of the liquid. The value of r at which the speed of expansion is a maximum is obtained from Eq. (7) by taking the derivative of $\dot{r}$ with respect to r, and gives

$$(r-R_a) = \gamma^{1/(3\gamma-3)}\ (r_o-R_a) \qquad (8)$$

The maximum rate of expansion that occurs at this value of r is [16]

$$\dot{r}_{max}^2 = \frac{2}{3} c_o^2 \gamma^{\gamma/(1-\gamma)} \qquad (9)$$

Finally, the time at which a bubble reaches a radius r during its growth phase can be gotten from Eq. (7)

$$t = \frac{r_o - R_a}{c_o} (2Z)^{1/2} \left(1 + \frac{2}{3} Z + \frac{Z^2}{5}\right) \qquad (10)$$

where $Z=(r-r_o)/(r_o-R_a)$.

If we insert $\gamma=4/3$ [13,15] into Eq. (9), the maximum speed of bubble growth is $\dot{r}_{max} = .46c_o$. Using $P_o=218$ atmospheres and $\rho=1g/cm^3$ gives $c_o \approx 150$ m/s and $\dot{r}_{max} \approx 70$ m/s, which is more than two orders of magnitude larger than the characteristic thermal conduction rate of Eq. (2). Thus, expansion occurs on a timescale that is much shorter than heat loss and this justifies the use of an adiabatic treatment during expansion. This can also be seen from Eq. (10). If we use representative values of $R_a=10^{-6}$m, $r_o=2R_a$, the time it takes a bubble to grow to $2r_o$ is on the order of 10^{-7} seconds.

Bubble Size as a Function of Laser Intensity

In studying cellular damage, we are most interested in the maximum size that the bubble reaches, r_m. We now show how the size of a bubble depends on laser intensity and melanosome properties (radius and absorption coefficient).

Using Eq. (6) to get the maximum bubble size we obtain

$$(r_m/R_a)^3 = 1+[(r_o/R_a)^3-1](P_o/P_{min})^{1/\gamma} \qquad (11)$$

The minimum pressure, at which the bubble stops expanding, is taken to be

the ambient pressure of one atmosphere and this will be the same for all bubbles. In actuality, the outward momentum of the liquid cellular medium may cause an overshoot in which the bubble's vapor has a pressure of less than one atmosphere. However, this will have a small effect on r_m since the bubble radius has a dependence on pressure of $r_m \propto P_{min}^{-1/3\gamma} \approx P_{min}^{-1/4}$. For example, an overshoot in which the pressure drops to ½ atmosphere will result in a final bubble radius that is only 19% larger than if the pressure goes no lower than one atmosphere.

Using P_o=218 atmospheres, the ratio P_o/P_{min} is equal to 218 for all laser pulses with sufficient intensity to raise the melanosome to T_c=374°C. Therefore, Eq. (3) tells us that the maximum volume of the vapor in the bubble is $218^{3/4}$=56.7 times larger than V_o, with the melanosome occupying constant volume inside the bubble. This will be true even if the melanosome breaks apart during the process, as long as the pieces remain inside the bubble. In order to calculate r_m from Eq. (11) therefore requires only an expression for r_o/R_a, which depends on the energy absorbed by the melanosome.

The energy required to raise one gram of water from body temperature of 37 C at 1 atmosphere of pressure to the critical point at 374 C at 218 atmospheres will be denoted by q (the value of q is approximately 2770 J/g as will be shown later). At the end of the laser absorption process, the energy E absorbed by a melanosome in an ultrashort pulse has created a vaporized volume V_o containing saturated steam and raised the temperature of the melanosome to the same 374°C. The initial volume of the steam will be

$$V_o = \frac{4\pi}{3}(r_o^3 - R_a^3) = \frac{E - E_m}{q} \times \frac{1}{\rho_c} \qquad (12)$$

where E_m is the energy required to raise a melanosome from 37°C to 374°C and is equal to $E_m=c_m\rho_m(4\pi/3\ R_a^3)\Delta T$. The specific heat of melanin is $c_m=2.51$ J/g·°C and its density is $\rho_m=1.35$ g/cm^3 [17]. For $R_a=10^{-6}$m, this gives a value of $E_m=4.9 \times 10^{-9}$ J.

The calculation for r_o continues by evaluating the energy absorbed by the melanosome. For a path length of d, through a material with an absorption coefficient α, the fraction of light absorbed is $(1-e^{-\alpha d})$. If I_o is the intensity of the laser in J/cm², then the energy incident on a spherical melanosome is $E_o=\pi a^2 I_o$, where $a\equiv R_a$ is the radius of the melanosome. With d=2a and 2aα between 0.2 and 0.4, as to be expected for visible light incident on a melanosome with $a=10^{-6}$m and α in the range between 1000 cm^{-1} and 2000 cm^{-1} [11,12], the fraction of light absorbed in a sphere is approximately $(1-e^{-2a\alpha})/1.5$ (See Appendix I for the derivation leading to the factor 1.5). This gives Eq. (13) for the absorbed energy

$$E = \pi a^2 I_o(1-e^{-2a\alpha})/1.5 \qquad (13)$$

Using α=1800 cm^{-1}, and $a=10^{-6}$m, the fraction absorbed is approximately 0.21 and the energy absorbed is $E=.21(I_o\pi a^2)$.

We can use these values to get an approximate number for the intensity necessary to produce a bubble in the manner described above. Equation (12) leads to an initial volume that is vaporized of

$$V_o=\frac{4\pi}{3}(r_o^3-10^{-12}cm^3)=\frac{6.5\times10^{-9}cm^2 I_o-4.9\times10^{-9}J}{q}\times\frac{1}{.315g/cm^3} \qquad (14)$$

Equation (14) represents the scenario in which any energy above $E_m=4.9$ x

10^{-9} J absorbed by a 1μm melanosome is used to produce an initial bubble with the vapor at the critical point. In order for at least 4.9 x 10^{-9}J to be absorbed, Eq. (14) shows that the intensity hitting the melanosome must be at least

$$I_o = 4.9 \times 10^{-9} J / 6.5 \times 10^{-9} cm^2 = 0.76 \ J/cm^2 \qquad (15)$$

Two comments must be made concerning this I_o:

1)The value of I_o will vary because it depends on several factors; shape (factor of .21), wavelength (α), and size (volume of melanosome to be heated depends on a^3 but absorbed energy depends on $a^2(1-e^{-2\alpha a})$). The value of I_o=0.76 J/cm^2 is for a spherical melanosome with a=1 μm and α=1800 cm^{-1}.

2)An intensity of less than I_o does not mean that no bubble is formed but instead that the initial conditions of the bubble are less than those of a vapor with the critical values of T_o=374°C and P_o=218 atmospheres, and initial bubble formation cannot be treated in the manner leading to Eqs. (14) and (15). This treatment is appropriate for this report which is interested in MVL's because ED_{50} measurements below 10^{-6} seconds give a retinal intensity of approximately 1 J/cm^2, and thus above the lower bound on I_o of Eq. (15).

Continuing with our analysis, Eq. (11) shows that r_m/R_a depends on $[(r_o/R_a)^3-1]$. Using Eq. (13) for the energy absorbed, and the expression for E_m following Eq. (12) (with R_a given in cm) for the energy needed to raise the temperature of a melanosome from 37°C to 374°C, Eq. (12) becomes

$$\left(\frac{r_o}{R_a}\right)^3 - 1 = \frac{3}{4R_a^3} \frac{\frac{R_a^2 I_o (1-e^{-2a\alpha})}{1.5} - \frac{4}{3} c_m \rho_m R_a^3 \Delta T}{q \rho_c} \qquad (16)$$

Combining the numerical factors gives

$$\left(\frac{r_o}{R_a}\right)^3 - 1 = \frac{.5 R_a^2 I_o (1-e^{-2a\alpha}) - c_m \rho_m R_a^3 \Delta T}{q\, \rho_c\, R_a^3} \qquad (17)$$

Using this expression in Eq. (11) gives

$$\left(\frac{r_m}{R_a}\right)^3 = 1 + \left(\frac{.5 R_a^2 I_o (1-e^{-2a\alpha}) - c_m \rho_m R_a^3 \Delta T}{q\, \rho_c\, R_a^3}\right)\left(\frac{P_o}{P_{min}}\right)^{1/\gamma} \qquad (18)$$

with R_a in cm, α in cm^{-1}, and I_o the retinal intensity in J/cm^2.

We note an intriguing consequence of Eq. (18). For low absorption by a melanosome, $2a\alpha \ll 1$, i.e. either low absorptivity α, or for very small melanosomes, we have $(1-e^{-2a\alpha}) \rightarrow 2a\alpha \equiv 2R_a\alpha$. Under these conditions, Eq. (18) becomes

$$\left(\frac{r_m}{R_a}\right)^3 = 1 + \left(\frac{\alpha I_o - c_m \rho_m \Delta T}{q\, \rho_c}\right)\left(\frac{P_o}{P_{min}}\right)^{1/\gamma} \qquad (19)$$

Thus, for low absorption, r_m/R_a becomes independent of R_a, the size of the melanosome, and depends primarily on the laser intensity I_o and melanin properties that are common to all melanosomes.

The value for q, the number of joules of energy required to raise 1 gram of water from 37°C at 1 atmosphere of pressure to the critical point at 374°C and 218 atmospheres is determined from thermodynamic considerations

$$\Delta E = \Delta H - \Delta(PV) \qquad (20)$$

where the energy change q is represented in Eq. (20) by E, and H is the enthalpy of the process. Since the change is energy is a state function, we can use any path in P-T space to evaluate ΔE. We use a path in which first, at constant pressure of 1 atmosphere, water is raised from body temperature to the critical temperature. Since dH=dQ+VdP, during this constant pressure part of the process $\Delta H=dQ$ (heat) even though volume changes do occur. The heat required to raise water from body temperature of 37°C to 100°C is 263 J/g. The water is then transformed to vapor at 1 atmosphere pressure which requires heat of 2262 J/g. Heat is then used to raise the temperature of the steam from 100°C (373°K) to the critical temperature of 374°C (647°K). The heat capacity for steam at a constant pressure of 1 atmosphere increases over this temperature range and we use the following expression for c_p [18]:

$$c_p = a + (b \times 10^{-3})T + (c \times 10^{-6})T^2 \qquad (21)$$

with a=1.67 J/g·°K, b=0.59 J/g·°K², and c=0.019 J/g·°K³. Performing the integral of $\int c_p dT$ from 373°K to 647°K gives a heat of 541 J/g. Thus, the enthalpy change in raising the temperature of H_2O from 37°C to 374°C, all at 1 atmosphere of pressure, is 3066 J/g.

The remaining changes in energy needed for evaluating Eq. (20) are: ΔH due to the step in which the temperature remains constant at 647°K and the pressure is increased isothermally from 1 atmosphere to 218 atmospheres, as well as the evaluation of $\Delta(PV)$ in Eq. (20) for the entire process. Since the product PV is also a state function, we can ignore intermediate steps and evaluate $\Delta(PV)=P_cV_c-P_oV_o$ with P_c=218 atmospheres =

221 x 10^5N/m², V_c=3.17 cm^3/g, P_o= 1 atmosphere = 1.013 x 10^5N/m², V_o=1 cm^3/g. This gives $\Delta(PV)$=70 J/g. The ΔH for the isothermal compression of the steam at 647°C from P_o=1 atmosphere to P_c=218 atmospheres can be evaluated by rewriting Eq. (20) as $\Delta H=\Delta E+\Delta(PV)$. If the steam behaved as an ideal gas during the isothermal compression then we would have $\Delta H=0$ since E=E(T) for an ideal gas and PV=RT, so $\Delta E=0$ and $\Delta(PV)=0$. In actuality the steam may behave as an ideal gas at P_i=1 atmosphere but does not behave as an ideal gas as the critical pressure is reached. Treating steam as an ideal gas at 1 atmosphere and using ρ_c=.315 g/cm^3 allows the evaluation of $\Delta(PV)=P_cV_c-P_iV_i$ for this step: P_cV_c=70 J/g, and P_iV_i(ideal)=RT=299 J/g. This gives $\Delta(PV)$=-230 J/g. The evaluation of ΔE is not as straightforward. During this isothermal compression process, work is done on the steam which tends to increase its energy. However, in order for the temperature to remain constant, this energy must be either lost to heat ΔH, or used in bond modifications. If the steam behaved as an ideal gas then its internal energy, which is purely kinetic for an ideal gas, would not change and this $\Delta E=0$ would imply $\Delta H=\Delta(PV)$. In a non-ideal gas however, there are bonds between molecules and it is possible to increase the energy of the system without a corresponding temperature increase. Thus, in compressing the steam isothermally, some of the energy put in as work can stay in the system and need not be lost as heat. Nevertheless, since we are compressing a vapor in which the interactions between molecules are much weaker than in liquids, and since a phase transformation does not occur during the compression, we will assume that the change in internal energy is small since T remains constant and set $\Delta E=0$ for this isothermal compression. The actual value could be determined from the virial coefficients of steam under the appropriate

conditions of temperature and pressure. With $\Delta E=0$, we have $\Delta H=\Delta(PV)=-230$ J/g for the isothermal compression.

Adding up all the contributions to ΔH we get $\Delta H=3066$ J/g - 230 J/g = 2836 J/g. The overall $\Delta(PV)=70$ J/g. Inserting these numbers into Eq. (20) gives $q=\Delta E\approx 2770$ J/g for insertion into Eq. (18), with an uncertainty due to setting $\Delta E=0$ for the isothermal compression as explained in the previous paragraph.

RESULTS

We now have values for all the factors necessary for calculating the expected r_m due to absorption of energy from submicrosecond laser pulses by melanosomes in the retinal pigment epithelium. Equation (18) shows explicitly how r_m depends on the pulse intensity I_o, and on melanosome properties; α, R_a, and q. We first get a representative value for r_m by using representative values for all the parameters in Eq. (18) as listed earlier in this report: $a\equiv R_a=10^{-6}m=10^{-4}cm$, $I_o=1 J/cm^2$, $\alpha=1800\ cm^{-1}$, $c_m=2.51$ J/g·°C, $\rho_m=1.35\ g/cm^3$, $\Delta T=374°C-37°C=337°C$, $\rho_c=0.315\ g/cm^3$, $P_o=218$ atmospheres, $P_{min}=1$ atmosphere, $\gamma=4/3$, and q=2770 J/g. Inserting these numbers into Eq. (18) gives a result of

$$For\ R_a=1\,\mu m: \qquad \left(\frac{r_m}{R_a}\right)^3 \approx 25, \qquad r_m\approx 2.9R_a=2.9\,\mu m \qquad (22)$$

The following Tables show how the maximum bubble radius, r_m, varies as a function of different parameters. Tables I show the variation of r_m with the melanosome properties $a(R_a)$ and absorptivity α, for different laser intensities I_o, at a q=2770 J/g. Tables II have the same variations in α, a, and I_o, but using a significantly lower q of 2000 J/g, and Tables

III are similar but with a high q of 3500 J/g. The variation of q from 2000 J/g to 3500 J/g leads to variations in r_m of only 20% or smaller.

Entries of 0.0 do not mean that no bubble is formed but instead signify that the present model is not applicable because not enough energy was absorbed by the melanosome to raise the temperature of the melanosome to T_c=374°C. Future work is required to model bubble dynamics in this situation.

Table I. Maximum bubble radius r_m, calculated from Eq. (18), as a function of melanosome radius 'a', melanosome absorptivity 'α', and laser retinal intensity 'I_o'. Other parameters: c_m=2.51 J/g·°C, ρ_m=1.35 g/cm^3, ΔT=337°C, ρ_c=0.315 g/cm$_3$, P_o/P_{min}=218, γ=4/3.
q=2770 J/g.

Table I.1) I_o(J/cm²): 0.50

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.000	0.000	0.000	0.000	0.000
	0.60	0.000	0.000	0.000	0.000	0.000	0.000
	0.80	0.000	0.000	0.000	0.000	0.000	0.000
a(μm)	1.00	0.000	0.000	0.000	0.000	0.000	0.000
	1.20	0.000	0.000	0.000	0.000	0.000	0.000
	1.40	0.000	0.000	0.000	0.000	0.000	0.000
	1.60	0.000	0.000	0.000	0.000	0.000	0.000

Table I.2) I_o(J/cm²): 1.00

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.419	0.937	1.160	1.318	1.443
	0.60	0.000	0.000	1.316	1.666	1.905	2.092
	0.80	0.000	0.000	1.620	2.117	2.442	2.691
a(μm)	1.00	0.000	0.000	1.832	2.508	2.926	3.237
	1.20	0.000	0.000	1.915	2.832	3.352	3.730
	1.40	0.000	0.000	1.781	3.076	3.715	4.165
	1.60	0.000	0.000	0.000	3.223	4.008	4.537

Table I.3) I_o(J/cm²): 1.50

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.094	1.349	1.530	1.673	1.794	1.899
	0.60	1.591	1.977	2.246	2.458	2.634	2.787
	0.80	2.052	2.572	2.929	3.207	3.437	3.633
a(μm)	1.00	2.474	3.135	3.579	3.921	4.202	4.440
	1.20	2.855	3.663	4.195	4.600	4.929	5.206
	1.40	3.189	4.156	4.777	5.243	5.619	5.933
	1.60	3.472	4.611	5.322	5.850	6.271	6.620

Table I.4) I_o(J/cm²): 2.00

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.490	1.691	1.850	1.984	2.100	2.203
	0.60	2.200	2.497	2.731	2.927	3.095	3.244
	0.80	2.886	3.277	3.583	3.837	4.054	4.244
a(μm)	1.00	3.547	4.031	4.406	4.715	4.978	5.206
	1.20	4.183	4.758	5.200	5.561	5.866	6.130
	1.40	4.795	5.459	5.966	6.376	6.721	7.017
	1.60	5.381	6.134	6.702	7.159	7.541	7.867

Table II. Maximum bubble radius r_m, calculated from Eq. (18), as a function of melanosome radius 'a', melanosome absorptivity 'α', and laser retinal intensity 'I_o'. Other parameters: c_m=2.51 J/g·°C, ρ_m=1.35 g/cm^3, ΔT=337°C, ρ_c=0.315 g/cm$_3$, P_o/P_{min}=218, γ=4/3.
q=2000 J/g.

Table II.1) I_o(J/cm^2): 0.50

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.000	0.000	0.000	0.000	0.000
	0.60	0.000	0.000	0.000	0.000	0.000	0.000
	0.80	0.000	0.000	0.000	0.000	0.000	0.000
a(μm)	1.00	0.000	0.000	0.000	0.000	0.000	0.000
	1.20	0.000	0.000	0.000	0.000	0.000	0.000
	1.40	0.000	0.000	0.000	0.000	0.000	0.000
	1.60	0.000	0.000	0.000	0.000	0.000	0.000

Table II.2) I_o(J/cm^2): 1.00

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.426	1.037	1.288	1.465	1.605
	0.60	0.000	0.000	1.454	1.849	2.117	2.326
	0.80	0.000	0.000	1.786	2.348	2.713	2.992
a(μm)	1.00	0.000	0.000	2.011	2.780	3.249	3.599
	1.20	0.000	0.000	2.085	3.134	3.720	4.145
	1.40	0.000	0.000	1.891	3.398	4.120	4.626
	1.60	0.000	0.000	0.000	3.551	4.441	5.037

Table II.3) I_o(J/cm²): 1.50

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.214	1.500	1.702	1.863	1.998	2.115
	0.60	1.765	2.198	2.499	2.736	2.933	3.103
	0.80	2.275	2.859	3.259	3.570	3.827	4.046
a(μm)	1.00	2.741	3.484	3.981	4.364	4.678	4.944
	1.20	3.160	4.070	4.666	5.119	5.487	5.797
	1.40	3.527	4.616	5.312	5.834	6.254	6.605
	1.60	3.835	5.120	5.918	6.509	6.979	7.369

Table II.4) I_o(J/cm²): 2.00

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.658	1.883	2.061	2.210	2.340	2.454
	0.60	2.448	2.780	3.042	3.260	3.448	3.613
	0.80	3.210	3.648	3.990	4.273	4.516	4.728
a(μm)	1.00	3.945	4.487	4.906	5.251	5.544	5.800
	1.20	4.653	5.296	5.790	6.193	6.534	6.828
	1.40	5.332	6.076	6.642	7.100	7.485	7.816
	1.60	5.984	6.826	7.461	7.972	8.398	8.762

Table III. Maximum bubble radius r_m, calculated from Eq. (18), as a function of melanosome radius 'a', melanosome absorptivity 'α', and laser retinal intensity 'I_o'. Other parameters: c_m=2.51 J/g·°C, ρ_m=1.35 g/cm^3, ΔT=337°C, ρ_c=0.315 g/cm$_3$, P_o/P_{min}=218, γ=4/3.
q=3500 J/g.

Table III.1) I_o(J/cm^2): 0.50

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.000	0.000	0.000	0.000	0.000
	0.60	0.000	0.000	0.000	0.000	0.000	0.000
	0.80	0.000	0.000	0.000	0.000	0.000	0.000
a(μm)	1.00	0.000	0.000	0.000	0.000	0.000	0.000
	1.20	0.000	0.000	0.000	0.000	0.000	0.000
	1.40	0.000	0.000	0.000	0.000	0.000	0.000
	1.60	0.000	0.000	0.000	0.000	0.000	0.000

Table III.2) I_o(J/cm^2): 1.00

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.415	0.873	1.077	1.222	1.337
	0.60	0.000	0.000	1.227	1.547	1.767	1.939
	0.80	0.000	0.000	1.514	1.968	2.266	2.494
a(μm)	1.00	0.000	0.000	1.718	2.333	2.716	3.002
	1.20	0.000	0.000	1.809	2.637	3.113	3.460
	1.40	0.000	0.000	1.715	2.868	3.452	3.865
	1.60	0.000	0.000	0.000	3.013	3.728	4.213

Table III.3) I_o(J/cm²): 1.50

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.017	1.251	1.417	1.550	1.661	1.758
	0.60	1.479	1.833	2.081	2.276	2.439	2.580
	0.80	1.908	2.386	2.714	2.971	3.183	3.364
a(μm)	1.00	2.302	2.908	3.317	3.633	3.891	4.111
	1.20	2.658	3.399	3.888	4.262	4.565	4.821
	1.40	2.972	3.857	4.428	4.858	5.205	5.494
	1.60	3.239	4.281	4.935	5.421	5.809	6.131

Table III.4) I_o(J/cm²): 2.00

	r_m	α(1/cm) 1000.	1200.	1400.	1600.	1800.	2000.
	0.40	1.381	1.566	1.713	1.837	1.944	2.039
	0.60	2.039	2.313	2.529	2.709	2.865	3.002
	0.80	2.674	3.035	3.318	3.552	3.753	3.928
a(μm)	1.00	3.287	3.733	4.080	4.365	4.608	4.819
	1.20	3.878	4.408	4.815	5.149	5.430	5.674
	1.40	4.445	5.057	5.524	5.903	6.221	6.495
	1.60	4.989	5.682	6.207	6.629	6.981	7.282

CONCLUSIONS

The calculation leading to the result of Eq. (15) is strong support for both the validity of the model used in this report and of the importance of bubble formation in causing minimal visible lesions (MVL) from ultrashort laser pulses incident on the eye. Eq. (15) shows that the present model requires a retinal intensity of 0.76 J/cm² for bubble formation to occur in an "average" melanosome with radius of 1 μm and α=1800 cm^{-1}. This is slightly below the experimental ED_{50} for ultrashort pulses which is found to be approximately 1 J/cm², with an uncertainty from use of 10^5 as the ratio of retinal intensity to corneal intensity. This close agreement supports the idea that bubble formation is a necessary step in causing MVL in ultrashort pulses and that this model is a reasonable theoretical treatment for calculating the size of bubbles as a function of relevant parameters.

The importance of various parameters for bubble growth can be ascertained from Eq. (18) and the information in Tables I-III.

1)The maximum bubble volume varies with 1/q, and therefore r_m varies as $q^{-1/3}$. This is weak dependence in which r_m only varies by 20% or less when q is changed from 2000 J/g to 3500 J/g.

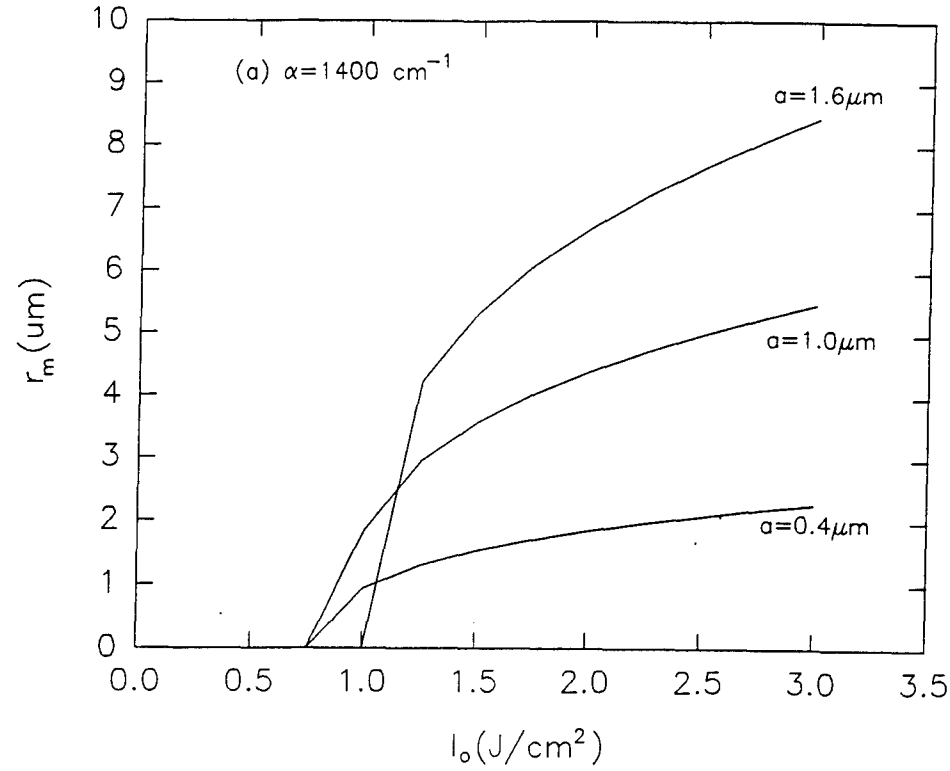
2)The dependence of r_m on the ratio of the pressures at the beginning and end of the bubble expansion is $(P_o/P_{min})^{-1/3\gamma}$. With γ=4/3, this gives a weak dependence of $(P_o/P_{min})^{-1/4}$. Thus a major change in the ratio of P_o/P_{min}, such as by a factor of two, leads to a change in r_m by a factor of only 1.19.

From Eq. (18) and the results presented in the Tables, it is clear that the melanosome radius a, absorptivity α, and the laser retinal intensity I_o are important factors in determining r_m. However, the dependence of r_m on these parameters is complex due to the non-linear way in which these parameters influence r_m. Thus, how strong a dependence r_m

has on any one of these depends on the specific values of the other two. A few general trends are discernible:

3)r_m increases monotonically with α and I_o, but not linearly, as seen in Fig. (1). The dependence is better categorized as a threshold dependence, as expected from Eq. (18). The complicated, threshold type dependence is exhibited in Table IV in which an I_o=0.756 J/cm² is used which is just above the threshold intensity for formation of a bubble for a melanosome with a=1 μm and α=1800 cm^{-1}, as explained earlier (see Eq. (15)).

4)Because of its appearance in several terms in Eq. (18), the dependence on a(R_a) is much more complicated, and not always monotonic. Even the ratio r_m/R_a is not monotonic and for certain values of α and I_o, increasing R_a leads to *decreasing* r_m. This occurs around threshold values for I_o; see the columns with α=1200 and 1400 in Tables I.2, II.2, III.2 and Table IV. In the situation in which $\alpha a \ll 1$, Eq. (18) reduces to Eq. (19). The ratio r_m/R_a then becomes independent of R_a, and in addition α and I_o take on straightforward threshold type dependencies. However, this approximation becomes valid only for $\alpha a < 0.05$, which makes this regime of limited relevance since melanosomes usually have a>1 μm and α>1000 cm^{-1}. Thus, the full treatment of Eq. (18) is generally required.



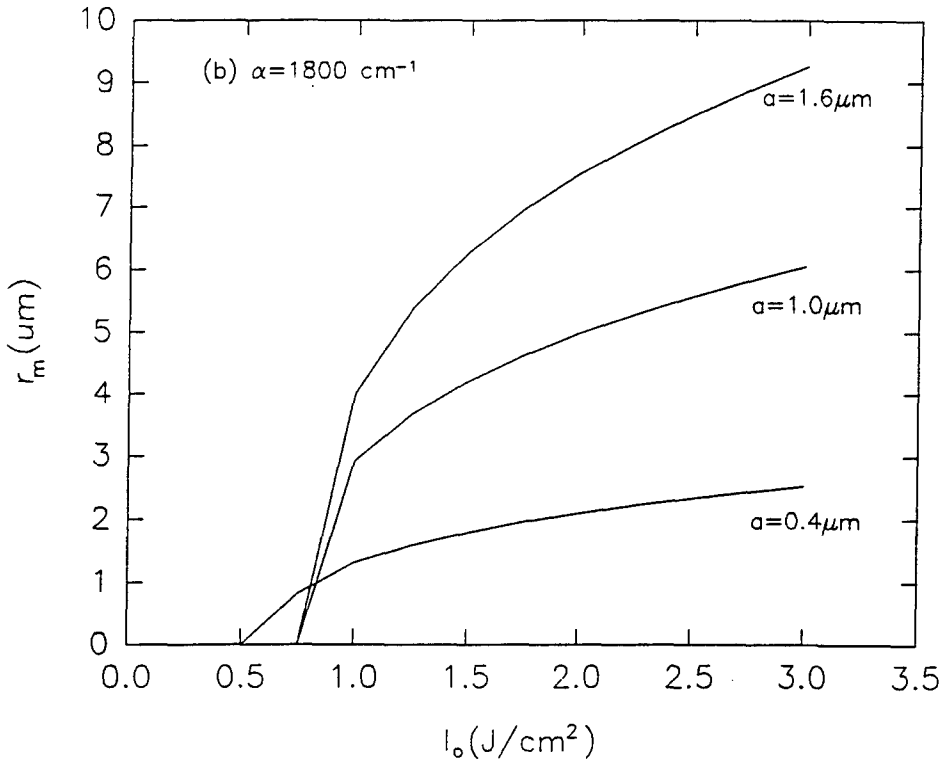


Figure 1. Maximum bubble radius as a function of retinal intensity, showing a threshold dependence. Figs. (1a) and (1b) differ in the melanosome's absorption coefficient.

Table IV. Maximum bubble radius r_m, calculated from Eq. (18), as a function of melanosome radius 'a', and melanosome absorptivity 'α'. Laser retinal intensity I_o=0.756 J/cm² is chosen to exhibit threshold type dependence. Other parameters: c_m=2.51 J/g·°C, ρ_m=1.35 g/cm^3, ΔT=337°C, ρ_c=0.315 g/cm$_3$, P_o/P_{min}=218, γ=4/3, q=2770 J/g.

		α(1/cm)					
	r_m	1000.	1200.	1400.	1600.	1800.	2000.
	0.40	0.000	0.000	0.000	0.000	0.837	1.041
	0.60	0.000	0.000	0.000	0.000	1.110	1.451
	0.80	0.000	0.000	0.000	0.000	1.230	1.773
a(μm)	1.00	0.000	0.000	0.000	0.000	1.018	1.981
	1.20	0.000	0.000	0.000	0.000	0.000	2.029
	1.40	0.000	0.000	0.000	0.000	0.000	1.772
	1.60	0.000	0.000	0.000	0.000	0.000	0.000

This paper presents a theoretical approach for calculating maximum bubble size formed in retinal pigment epithelium cells due to ultrashort laser pulses. The agreement between the threshold energy required by the model for the formation of bubbles with the experimental ED_{50} shows the relevance of the model for damage assessment. A full understanding of the damage process requires additional work. This work will be directed towards understanding the mechanisms by which bubbles actually cause cellular damage. These mechanisms will obviously derive from the physical expansion of the bubble. The main question is how the physical expansion couples to the biochemistry of a cell in a destructive fashion. Does the expansion destroy enough cellular proteins to cause MVL immediately or is the damage initially minor but enough to prevent the cell's biochemical pathways from repairing the initially damaged areas as well as disrupting transport channels sufficiently so that the cell is not able to continue carrying out its usual functions at the necessary rates?

In order to distinguish between mechanisms such as these, or others, experiments must be done that look at immediate effects of the laser pulse. Immediate effects, in this case, are as soon as the bubble has finished expansion which is on the sub millisecond time scale. This requires use of experimental methods such as optical techniques that look for changes in protein absorption characteristics on these time scales, similar to pump-probe techniques.

Finally, this paper has presented theoretical techniques for calculating the effects of short laser pulses ($\tau<10^{-6}$ seconds) in which heat conduction away from the melanosome is initially unimportant and the absorbed energy is channeled into vaporization and bubble formation. The first part of this report reported on a computational method for predicting temperature rises produced by relatively long laser pulses ($\tau>10^{-5}$ seconds) for which heat is conducted away fast enough so that

ED_{50}'s imply no vaporization. For laser pulses of duration of approximately 10^{-6} seconds, the theoretical treatment will be complicated by the presence of both vaporization and conduction. This middle regime remains to be investigated in detail.

APPENDIX

For light of uniform intensity $I_0(J/cm^2)$ hitting a sphere of radius 'a' with uniform absorption coefficient α, the fraction of light absorbed is calculated as follows. The total energy hitting the sphere is $E_0=I_0\pi a^2$. For a path length of d, the fraction of light that passes through is $I_T(d)/I_0= e^{-\alpha d}$. The average of this fraction over a sphere gives the energy E_T that is transmitted and not absorbed and is given by

$$\frac{E_T}{E_o}=\frac{1}{\pi a^2}\int_o^a e^{-2\alpha\sqrt{a^2-r^2}}2\pi r dr \qquad (A.1)$$

where r for a light ray is the distance of closest approach to the center, and $2\sqrt{a^2-r^2}$ is the path length of the light ray through the melanosome. Figure (A.1) is a diagram of the process.

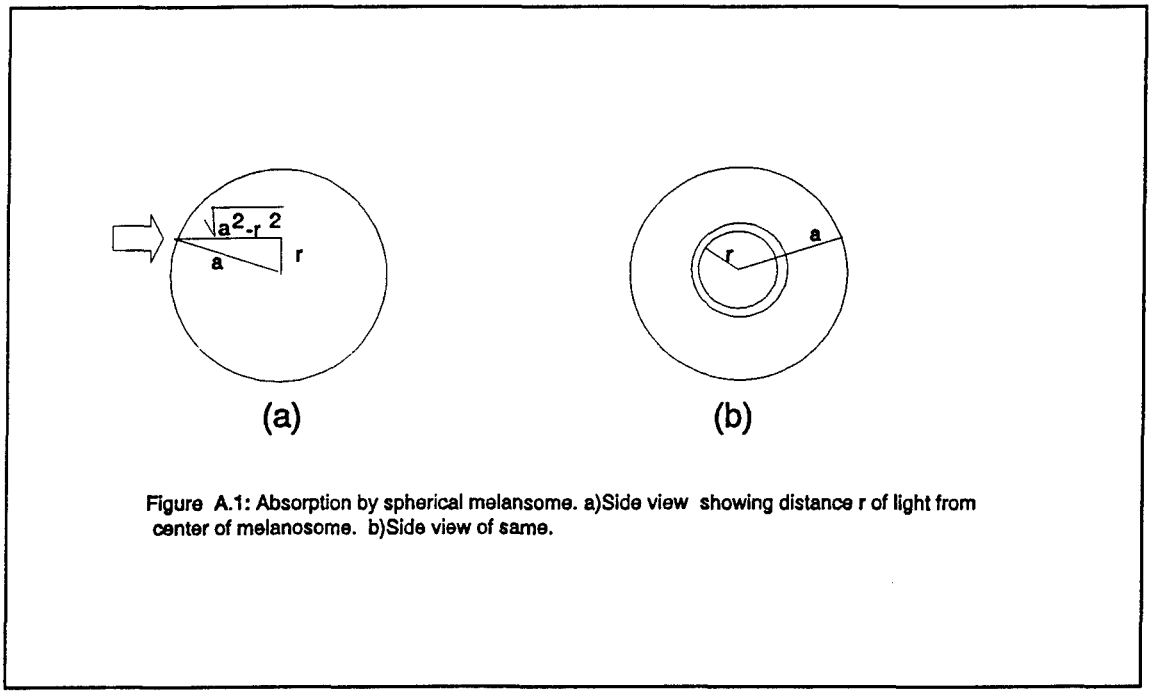


Figure A.1: Absorption by spherical melansome. a)Side view showing distance r of light from center of melanosome. b)Side view of same.

The integration can be performed with a change of variable of $u=2\alpha\sqrt{a^2-r^2}$ and gives

$$\frac{E_T}{E_o} = \frac{1}{2\alpha^2 a^2}\left[1 - e^{-2\alpha a}(1+2\alpha a)\right] \qquad (A.2)$$

The energy absorbed is

$$E = E_o - E_T = \pi a^2 I_o\left[1 - \frac{1}{2\alpha^2 a^2}\left(1 - e^{-2\alpha a}(1+2\alpha a)\right)\right] = C\pi a^2 I_o \qquad (A.3)$$

For a sphere that has the properties of a melanosome with a~1 μm and a visible light α~1800 cm^{-1}, the fraction of energy absorbed is $E=0.21E_o=0.21(\pi a^2 I_o)$.

References

1)A. N. Takata, L. Goldfinch, J. K. Hinds, L. P. Kuan, N. Thomopoulis, A. Weigandt, *Thermal Model of Laser-Induced Eye Damage*, USAFSAM Technical Report J-TR-74-6324, 1974.

2)D. H. Sliney, First International Symposium on Laser Biological Effects and Exposure Limits, Ed. by L. Court, A. Duchene, and D. Courant, Paris, Nov. 1986.

3)W. T. Ham, H. A. Mueller, A. I. Goldman, B. E. Newnam, L. M. Holland, T. Kuwabara, Science, 185, 362, 1974.

4)M. L. Wolbarsht, A. L. Walsh, and G. George, Applied Optics, 20, 2184, 1981.

5)R. Cubeddu, F. Docchio, R. Ramponi, and M. Boulton, IEEE Journal of Quantum Electronics, 26, 2218, 1990.

6)G. Gueneau, V. Baille, D. Courant, M. Dubos, and L. Court, First International Symposium on Laser Biological Effects and Exposure Limits, Ed. by L. Court, A. Duchene, and D. Courant, Paris, Nov. 1986.

7)H.S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, Oxford University Press, London, 1948.

8)H. Goldenberg and C. J. Tranter, British Journal of Applied Physics, 3, 296, 1952.

9)C. R. Thompson, Melanin Granule Model for Heating of Tissue by Laser, SPIE Proceedings, Vol. 2134, OE/LASE, 1994.

10)F. C. Henriques, Archives of Pathology, 43, 489, 1947.

11)F. L'Esperance, *Ophthalmic Lasers*, C. V. Mosby Company, St. Louis, 1989.

12)A. Vogel and R. Birngruber, Lasers and Light in Ophthalmology, 5(1), 9, 1992.

13)S. F. Cleary, Laser Applications in Medicine and Biology, Ed. M. L. Wolbarsht, Plenum Press, New York, 1977.

14)K. Ražnjević, Handbook of Thermodynamic Tables and Charts, Hemisphere Publishing Corporation, Washington, McGraw Hill Book Company, New York, 1976.

15)H. Lamb, *Hydrodynamics*, Dover, New York, 1945.

16)Footnote: This equation is the corrected version of Eq. (12), p. 123 of Lamb, and Eq. (8) in Cleary.

17)J. R. Hayes and M. L. Wolbarsht, Aerospace Medicine, 39, 474, 1968.

18)Chemical Rubber Company Handbook of Chemistry and Physics, Edited by R. C. Weast, CRC Press Inc., Boca Raton, Florida, 70th Edition, 1989-1990.

A COMPARISON OF VARIOUS ESTIMATORS OF HALF-LIFE IN THE AIR FORCE HEALTH STUDY

Pushpa L. Gupta
Professor

Department of Mathematics and Statistics

University of Maine
5752 Neville Hall
Orono, Maine 04469-5752

Telephone: 207-581-3914

Final Report for:

Research Initiation Program
Armstrong Laboratory

Sponsored by:

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

December 1993

A COMPARISON OF VARIOUS ESTIMATORS OF HALF-LIFE IN THE AIR FORCE HEALTH STUDY

Pushpa L. Gupta
Professor
Department of Mathematics and Statistics
University of Maine

ABSTRACT

There are various methods available for estimating the half-life of 2, 3, 7, 8 tetrachlorodibenzo-p-dioxin (TCDD) in Ranch Hand Vietnam veterans. In this project multivariate extensions with covariates of the unconditional model introduced by Gupta (1990, 1991), are studied and applied to the extended Ranch Hand data set of 235 subjects. Residual and quantile plots, showing the fit of these extended unconditional models, are provided. Tables containing the point and interval estimates of half-life based on these models are presented and some comparisons based on the widths of these intervals are made.

Finally the analytical expressions of the efficiencies of the estimators of half-life based on mixed effects model of Tripathi (1990) and Michalek et. al. (1992), derived by Gupta (1992), are updated by using second order approximations.

A COMPARISON OF VARIOUS ESTIMATIORS OF HALF-LIFE IN THE AIR FORCE HEALTH STUDY

Pushpa L. Gupta

1. INTRODUCTION

Half-life estimation of half-life of TCDD in Ranch Hand Vietnam veterans has been of considerable interest to the U.S. Air Force scientists. For the past several years, epidemiology division of the USAF School of Aerospace Medicine at Brooks Air Force Base has been conducting studies to estimate the median half-life of the TCDD concentration level in Ranch Hand Veterans.

There are five procedures available for estimating the half-life of TCDD in Ranch Hand veterans: (1) a non-parametric estimate of Phillips (1989), (2) a non-parametric estimate of Pirkle et. al. (1989), (3) A parametric estimate of Michalek et. al. (1989), (4) an estimate based on mixed-linear model with covariates of Tripathi (1989, 1990) and Michalek et. al. (1992), (5) a parametric estimate based on conditional and unconditional model (full model) and on extensions of the full model accommodating multiple measurements and covariates, Gupta (1990, 1991).

In this project multivariate extensions with covariates of the unconditional model introduced by Gupta (1990, 1991) are studied and applied to the extended Ranch Hand data set of 235 subjects. Residual and quantile plots, showing the fit of these extended unconditional models, are provided. Tables containing the point and interval estimates of half-life based on these models are presented and some comparisons based on the widths of these intervals are made.

Finally the analytical expressions of the efficiencies of the estimators of half-life, based on mixed effects model of Tripathi (1990) and Michalek et. al. (1992), derived by Gupta (1992), are updated by using second order approximations. These are contained in sections 6 and 7. Section 2 contains the unconditional model with two measurements together with the half-life estimator and its variance. Section 3 contains three multivariate extensions, of the unconditional model of section 2, together with expressions of estimators of half-life and variances of the estimates. Section 4 contains covariate extensions of section 2 and section 5 contains covariate extensions of section 3 and the table of half-life estimates based on these models and graphs of (1) decay rate regression, (2) observed decay rate vs. body fat, and (3) decay rate Q - Q plot showing the fit of these models. In what follows the distribution function and the density function of a standard normal are denoted by $\Phi(\cdot)$ and $\phi(\cdot)$ respectively.

2. UNCONDITIONAL MODEL

Assuming that the first order kinetics model

$$C_T = C_0 e^{-\lambda T} \tag{2.1}$$

holds, where C_T is the TCDD concentration T years after exposure, C_0 is the initial exposure (unknown) and λ is the unknown decay rate, which we are assuming to be random. From (2.1), the TCDD half-life, based on two measurements C_1 and C_2 taken T_1 and T_2 years after exposure, is

$$Y = \frac{\Delta T \ln 2}{\ln(C_1/C_2)} = d/U, \tag{2.2}$$

where $\Delta T = T_2 - T_1 > 0$, $U = \ln(C_1/C_2)$, and $d = \Delta T \ln 2$. Since

$\ln C_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$ (see Wolfe et. al. 1988),

$\ln(C_1/C_2) \sim N(\mu, \sigma^2)$, where $\mu = \mu_1 - \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$,

$\sigma_{12} = \mathrm{Cov}(\ln C_1, \ln C_2)$. Then the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\,\sigma}\,\frac{d}{y^2}\exp\left[-((d/y)-\mu)^2/2\sigma^2\right], & y \neq 0 \\ 0, & y = 0 \end{cases} \qquad (2.3)$$

The median half-life M is given by

$$M = \frac{d}{\mu + \sigma\Phi^{-1}(a + \Phi(-\mu/\sigma))}, \qquad (2.4)$$

where $a = \pm .5$ according as $M \gtrless 0$.

Confidence Interval for M based on MLE's

The nonconstant part of the full log-likelihood is ℓ_F, where

$$\ell_F(\mu, \sigma^2 \,|\, y_i,\ i = 1,2,\ldots,N) = -N \ln \sigma - \sum_{i=1}^{N} ((d_i/y_i)-\mu)^2/2\sigma^2, \qquad (2.5)$$

where $d_i = \Delta T_i \ln 2$, $i = 1,2,\ldots,N$.

The maximum likelihood estimators (MLE's) of μ and σ^2 together with their expectations, variances, and covariances are given by

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} d_i/Y_i,\ E(\hat{\mu}) = \mu,\ \mathrm{Var}(\hat{\mu}) = \sigma^2/N$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} ((d_i/Y_i) - \hat{\mu})^2,\ E(\hat{\sigma}^2) = \left(\frac{N-1}{N}\right)\sigma^2,\ \mathrm{Var}\left(\hat{\sigma}^2\right) = \frac{2(N-1)\sigma^4}{N^2},$$

$$\mathrm{Cov}(\hat{\mu}, \hat{\sigma}^2) = 0. \qquad (2.6)$$

Thus the estimator of the median half-life M is

$$\hat{M} = \frac{d}{\hat{\mu} + \hat{\sigma}\Phi^{-1}(a + \Phi(-\hat{\mu}/\hat{\sigma}))} = g(\hat{\mu},\hat{\sigma}^2) \qquad (2.7)$$

and

$$\mathrm{Var}(\hat{M}) = \mathrm{Var}(g(\hat{\mu},\hat{\sigma}^2)) \doteq \left(\frac{\partial g}{\partial \hat{\mu}}\right)^2 \mathrm{Var}(\hat{\mu}) + \left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)^2 \mathrm{Var}(\hat{\sigma}^2) + 2\left(\frac{\partial g}{\partial \hat{\mu}}\right)\left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)\mathrm{Cov}\left(\hat{\mu},\hat{\sigma}^2\right)$$

$$= \left(\frac{\partial g}{\partial \hat{\mu}}\right)^2 \frac{\sigma^2}{N} + \left(\frac{\partial g}{\partial \hat{\sigma}^2}\right)^2 \frac{2(N-1)\sigma^4}{N^2}, \qquad (2.8)$$

see Elandt-Johnson and Johnson (1980), where the partial derivatives are evaluated at $(E(\hat{\mu}),E(\hat{\sigma}^2)) = (\mu,\frac{\sigma^2(N-1)}{N})$. The partials of $g(\cdot)$ are given below:

$$\frac{\partial g}{\partial \hat{\mu}} = -\frac{g^2}{d}\left[1 - \frac{\phi(-\hat{\mu}/\hat{\sigma})}{\phi[(d - g\hat{\mu})/g\hat{\sigma}]}\right],$$

$$\frac{\partial g}{\partial \hat{\sigma}^2} = -\frac{g}{2\hat{\sigma}^2}\left[1 + \frac{\hat{\mu}}{g}\frac{\partial g}{\partial \hat{\mu}}\right], \qquad (2.9)$$

where g is given by (2.7).

Using the large sample properties of $\hat{M}$, $(100)(1-\alpha)\%$ approximate confidence interval (CI) for M is $\hat{M} \pm \hat{\sigma}_{\hat{M}} z_{\alpha/2}$, where $\hat{\sigma}_{\hat{M}}$ is an estimate of $\sigma_{\hat{M}}$.

Confidence Interval for M based on LSE's

The least-squares estimates (LSE) of μ and σ^2 together with their expected values, variances, and covariance are given by

$$\hat{\mu}_{LS} = \hat{\mu} = \text{MLE of } \mu$$

$$\hat{\sigma}^2_{LS} = \frac{N}{N-1}\hat{\sigma}^2, \text{ where } \hat{\sigma}^2 \text{ is the MLE of } \sigma^2.$$

$$E\left(\hat{\sigma}^2_{LS}\right) = \sigma^2, \quad \mathrm{Var}\left(\hat{\sigma}^2_{LS}\right) = \frac{2}{N-1}\sigma^4.$$

The estimator of the median half-life is

$$\hat{M}_{LS} = \frac{d}{\hat{\mu}_{LS} + \hat{\sigma}_{LS}\Phi^{-1}(a + \Phi(-\hat{\mu}_{LS}/\hat{\sigma}_{LS}))} = h(\hat{\mu}_{LS}, \hat{\sigma}_{LS}) \qquad (2.10)$$

and

$$\begin{aligned} \mathrm{Var}(\hat{M}) = \mathrm{Var}(h(\hat{\mu},\hat{\sigma}^2)) &\doteq \left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right)^2 \mathrm{var}\left(\hat{\mu}_{LS}\right) + \left(\frac{\partial h}{\partial \hat{\sigma}^2_{LS}}\right)^2 \mathrm{var}\left(\hat{\sigma}^2_{LS}\right) \\ &\quad + 2\left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right)\left(\frac{\partial h}{\partial \hat{\sigma}^2_{LS}}\right)\mathrm{Cov}(\hat{\mu}_{LS}, \sigma^2_{LS}) \\ &= \left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right)^2 \frac{\sigma^2}{N} + \left(\frac{\partial h}{\partial \hat{\sigma}^2_{LS}}\right)^2 \frac{2\sigma^4}{N-1}, \qquad (2.11) \end{aligned}$$

where the partials of $h(\cdot)$ are evaluated at $(E(\hat{\mu}_{LS}), E(\hat{\sigma}^2)) = (\mu, \sigma^2)$ and are given by:

$$\left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right) = -\frac{h^2}{d}\left[1 - \frac{\phi(-\mu_{LS}/\hat{\sigma}_{LS})}{\phi[(d-h\hat{\mu}_{LS})/h\hat{\sigma}_{LS}]}\right]$$

$$\left(\frac{\partial h}{\partial \hat{\sigma}_{LS}}\right) = -\frac{h}{2\hat{\sigma}^2_{LS}}\left[1 + \frac{\hat{\mu}_{LS}}{h}\frac{\partial h}{\partial \hat{\mu}_{LS}}\right],$$

where h is given by (2.10). The approximate $(1 - \alpha)100\%$ confidence interval of M based on the least-square estimates is $\hat{M}_{LS} \mp \hat{\sigma}_{\hat{M}_{LS}} z_{\alpha/2}$, where $\hat{\sigma}_{\hat{M}_{LS}}$ is an estimate of $\sigma_{\hat{M}_{LS}}$.

Remark: It can be verified that the estimate of M based on least squares is less than the estimate of M based on maximum likelihood.

3. MULTIVARIATE EXTENSIONS OF THE UNCONDITIONAL MODEL (FULL MODEL)

In this section we present three extensions of the unconditional model of section 2 when more than two measurements are available. Let C_{ij} be the jth TCDD concentration of the ith individual at time T_{ij}.

Let $\Delta T_{ij} = T_{ij+1} - T_{ij}$, and $\Delta_1 T_{ij} = T_{ij} - T_{i1}$, i = 1, 2, ..., N, j = 1, 2, ..., J. Under the assumption of first-order kinetics, the half-life is given by $\frac{\ln 2}{\lambda}$. The value of λ can be obtained as follows:

Model I: From equation (2.1) we have

$$\ln(C_{ij}/C_{ij+1}) = \lambda \Delta T_{ij}, \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, J-1.$$

Hence

$$\lambda = \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}} . \tag{3.1}$$

Summing (3.1) with respect to j, we have

$$\lambda = \frac{1}{J-1} \sum_{j=1}^{J-1} \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}} . \tag{3.2}$$

Then the half-life for the ith subject is

$$Y_i = \frac{(J-1)\ln 2}{\sum_{j=1}^{J-1} \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}}} = \frac{d_i}{U_i} , \tag{3.3}$$

where $d_i = (J-1)\ln 2$, and $U_i = \sum_{j=1}^{J-1} \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}}$.

U_i, being a linear combination of normal random variables, is $N(\mu,\sigma^2)$. We will assume that $U_1, U_2, \ldots, U_N$ are independent and identically distributed (IID) random variables with the common $N(\mu,\sigma^2)$ distribution. As a result the observed random variables Y_i, i = 1,2,...,N, are also IID with some common distribution. We will assume that the half-life of TCDD is a random variable denoted by Y = d/U, where d = (J-1)ln 2, and

$U = \sum_{j=1}^{J-1} \ln(C_j/C_{j+1})/\Delta T_j$, with the common distribution given by (2.3)

in section 2. The MLE $\hat{M}$ of M, the median of interest, and $\hat{\sigma}_{\hat{M}}$ the estimate of the standard deviation of $\hat{M}$ can be obtained as in (2.7) and (2.8), where $\hat{\mu}$ and $\hat{\sigma}^2$ are given by (2.6) with $d_i = (J-1)\ln 2$, and

$$d_i/Y_i = U_i = \sum_{j=1}^{J-1} \ln(C_{ij}/C_{ij+1})/\Delta T_{ij}, \quad i = 1,2,\ldots,N.$$

Model II: From equation (2.1) one can have

$$\ln(C_{i1}/C_{ij+1}) = \lambda \Delta_1 T_{ij}, \quad i = 1, 2,\ldots,N, \quad j = 1,2,\ldots,J-1. \qquad (3.4)$$

Summing (3.4) with respect to j, we have

$$\lambda = \frac{\sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1})}{\sum_{j=1}^{J-1} \Delta_1 T_{ij}}. \qquad (3.5)$$

Then the half-life for the ith subject is

$$Y_i = \frac{(\ln 2) \sum_{j=1}^{J-1} \Delta_1 T_{ij}}{\sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1})} = \frac{d_i}{U_i}, \qquad (3.6)$$

where $d_i = \ln 2 \sum_{j=1}^{J-1} \Delta_1 T_{ij}$, $\Delta_1 T_{ij} = T_{ij} - T_{i1}$, and $U_i = \sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1})$. As in model I, U_i, being a linear combination of normal random variables, is normal with mean μ and variance σ^2.

Model III: From equation (3.4) we have

$$\lambda = \frac{\ln(C_{i1}/C_{ij+1})}{\Delta_1 T_{ij}}, \quad i = 1, 2,\ldots,N, \quad j = 1,2,\ldots,J-1. \qquad (3.7)$$

Summing (3.7) with respect to j, we have

$$\lambda = \frac{1}{J-1} \sum_{j=1}^{J-1} \frac{\ln(C_{i1}/C_{ij+1})}{\Delta_1 T_{ij}} \quad . \tag{3.8}$$

Then the half-life for the ith subject is

$$Y_i = \frac{(J-1)\ln 2}{\sum_{j=1}^{J-1} \frac{\ln(C_{i1}/C_{ij+1})}{\Delta_1 T_{ij}}} = \frac{d_i}{U_i} \quad , \tag{3.9}$$

where $d_i = (J-1) \ln 2$, and $U_i = \sum_{j=1}^{J-1} \frac{\ln(C_{i1}/C_{ij+1})}{\Delta_1 T_{ij}}$.

Here also, as in Model I and II, U_i, being a linear combination of normal random variables, is $N(\mu,\sigma^2)$. It may be remarked that the MLE's and confidence intervals of the TCDD half-life under model II and III can be obtained as under model I by replacing the d_i's and Y_i's of Model I in the estimates by the d_i's and Y_i's of Model II and Model III respectively.

4. Unconditional Model with Covariates (Two Measurements)

The TCDD half-life of the ith individual, using (2.11), can be given as

$$Y_i = \frac{\Delta T_i \ln 2}{\ln(C_{i1}/C_{i2})} = \frac{d_i}{U_i} \quad , \; i = 1, 2, \ldots, N, \tag{4.1}$$

where $d_i = \Delta T_i \ln 2 > 0$, $U_i = \ln(C_{i1}/C_{i2})$. Let

$$\ln(C_{ij}) = \beta_{oj} + \sum_{k=1}^{r} \beta_k x_{ij,k} + \varepsilon_{ij}, \text{ where } \ln C_{ij} \sim N(\mu_j, \sigma_j^2) \quad ,$$

$$\varepsilon_{ij} \sim N(0,\sigma_j^2), \; i = 1, 2, \ldots, N; \; j = 1, 2. \quad \text{Then} \tag{4.2}$$

$$U_i = \beta_{01} - \beta_{02} + \sum_{k=1}^{r} \beta_k(x_{i1,k} - x_{i2,k}) + \varepsilon_{i1} - \varepsilon_{i2}$$

$$= \beta_0 + \sum_{k=1}^{r} \beta_1 z_{ik} + \varepsilon_i,$$

where $\beta_0 = \beta_{01} - \beta_{02}$, $z_{ik} = x_{i1,k} - x_{i2,k}$, $\varepsilon_i = \varepsilon_{i1} - \varepsilon_{i2}$,
$i = 1, 2, \ldots, N$, $k = 1, 2, \ldots, r$. (4.3)

$$\mu|_{z_{i1},z_{i2}\ldots,z_{ir}} = E[\ln(C_{i1}|C_{i2})] = \beta_0 + \sum_{k=1}^{r} \beta_k z_{ik}.$$

$$\sigma^2|_{z_{i1},z_{i2},\ldots,z_{ir}} = \mathrm{Var}[\ln(C_{i1}|C_{i2})] = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12},$$

where $\sigma_{12} = \mathrm{Cov}[\ln C_{i1}, \ln C_{i2}]$. (4.4)

It will be assumed that $U_1, U_2, \ldots, U_N$ are IID $N(\mu,\sigma^2)$ random variables. As a result, the observed random variables Y_i, $i = 1, 2, \ldots, N$, are also IID with some common distribution. We will assume that the half-life of TCDD is a random variable Y with this common distribution given by

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{d}{y^2} \exp[-(\frac{d}{y} - \beta_0 - \sum_{k=1}^{r} \beta_k z_k)^2/2\sigma^2], \quad y \neq 0, \tag{4.5}$$

where the mean is modeled by $\mu = \beta_0 + \sum_{k=1}^{r} \beta_k z_k$.

Then the likelihood of the sample $Y_1, Y_2, \ldots, Y_N$ is

$$L(\beta,\sigma^2) = \prod_{i=1}^{N} \frac{d_i}{\sqrt{2\pi\ \sigma y_i^2}} \exp[-(\frac{d_i}{y_i} - \beta_0 - \sum_{k=1}^{r} \beta_k z_{ik})^2/2\sigma^2]. \tag{4.6}$$

The non-constant part of the full log-likelihood is

$$\ell_F(\underset{\sim}{\beta},\sigma^2) = -\frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(\frac{d_i}{y_i} - \beta_0 - \sum_{i=1}^{r} \beta_k z_{ik}\right)^2. \tag{4.7}$$

Thus the likelihood equations are

$$\frac{\partial}{\partial\sigma^2}\ell_F(\beta,\sigma^2) = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}\left(\frac{d_i}{y_i} - \beta_0 - \sum_{k=1}^{r}\beta_k z_{ik}\right)^2 = 0 \qquad (4.8)$$

$$\frac{\partial}{\partial\beta_k}\ell_F(\beta,\sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(\frac{d_i}{y_i} - \beta_0 - \sum_{k=1}^{r}\beta_k z_{ik}\right)z_{ik} = 0,\ k = 0,1,\ldots,r, \qquad (4.9)$$

where $z_{i0} = 1$. Solving (4.8) and (4.9), we have

$$\hat{\sigma}^2 = \frac{1}{N}(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})'(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}}),$$

$$\underset{\sim}{\hat{\beta}} = (Z'Z)^{-1}Z'\underset{\sim}{U}.$$

Also $\hat{\mu} = [1\ \bar{z}_1\ \bar{z}_2\ \ldots\ \bar{z}_r]\underset{\sim}{\hat{\beta}} = \underset{\sim}{a}\underset{\sim}{\hat{\beta}}$, where $\underset{\sim}{a}$ is a vector of column means from the design matrix Z given below. (4.10)

The quantities in the above expressions are defined as follows:

$$U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}_{N\times 1}, \quad U_i = \ln(C_{i1}/C_{i2}),\ i = 1,2,\ldots,N.$$

$$\underset{\sim}{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_r \end{bmatrix}_{(r+1)\times 1}, \quad Z = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & z_{N1} & z_{N2} & \cdots & z_{Nr} \end{bmatrix}_{N\times(r+1)}$$

$$\bar{z}_k = \sum_{i=1}^{N} z_{ik}/N,\ z_{ik} = x_{i1,k} - x_{i2,k},\ i = 1,2,\ldots,N,\ k = 1,2,\ldots,r.$$

The expectations, variances, and covariance of $\hat{\mu}$ and $\hat{\sigma}^2$ are given as follows:

$$E(\hat{\mu}) = \mu$$

$$\mathrm{Var}(\hat{\mu}) = \underset{\sim}{a}\,\mathrm{Var}(\underset{\sim}{\hat{\beta}})\underset{\sim}{a}'$$

$$= \sigma^2\,\underset{\sim}{a}\,(Z'Z)^{-1}\underset{\sim}{a}'$$

Since $(\underset{\sim}{U} - z\underset{\sim}{\hat{\beta}})'(\underset{\sim}{U} - z\underset{\sim}{\hat{\beta}})/\sigma^2 \sim \chi^2_{N-r-1}$,

$$E(\hat{\sigma}^2) = \left(\frac{N - (r+1)}{N}\right)\sigma^2$$

and

$$\mathrm{Var}(\hat{\sigma}^2) = \frac{2(N - (r + 1))}{N^2}\,\sigma^4 .$$

Also $\mathrm{Cov}(\hat{\mu},\hat{\sigma}^2) = \underset{\sim}{a}\,\mathrm{Cov}(\hat{\underset{\sim}{\beta}},\hat{\sigma}^2) = 0$, since $\hat{\underset{\sim}{\beta}}$ and $\hat{\sigma}^2$ are independent, see (Graybill 1976, page 176).

The MLE of the median half-life is given by

$$\hat{M} = \frac{d}{\hat{\mu} + \hat{\sigma}\Phi^{-1}(a + \Phi(-\hat{\mu}/\hat{\sigma}))} ,$$

where $a = \pm .5$ according as $M \gtrless 0$, $d = \sum_{i=1}^{N} \frac{d_i}{N}$.

$$\mathrm{Var}(\hat{M}) = \sigma^2_{\hat{M}} \doteq \left(\frac{\partial \hat{M}}{\partial \hat{\mu}}\right)^2 \mathrm{var}(\hat{\mu}) + \left(\frac{\partial \hat{M}}{\partial \sigma^2}\right)^2 \mathrm{var}(\hat{\sigma}^2) + 2\left(\frac{\partial \hat{M}}{\partial \hat{\mu}}\right)\left(\frac{\partial \hat{M}}{\partial \hat{\sigma}^2}\right)\mathrm{Cov}\left(\hat{\mu},\hat{\sigma}^2\right),$$

where the partials are given by (2.9) and are evaluated at $(E(\hat{\mu}),\ E(\hat{\sigma}^2))$.

Remark. If least squares estimates of the parameters are used in computing the $\hat{M}$ and its variance, then

$$\hat{\mu}_{LS} = \hat{\mu} = \text{MLE of } \mu$$

$$\hat{\sigma}^2_{LS} = \frac{N}{N - (r+1)}\hat{\sigma}^2, \text{ where } \hat{\sigma}^2 \text{ is the MLE of } \sigma^2.$$

$$E(\hat{\mu}_{LS}) = \mu,\ \mathrm{Var}(\hat{\mu}) = \underset{\sim}{a}\,\mathrm{Var}(\hat{\underset{\sim}{\beta}})\underset{\sim}{a}' = \sigma^2\underset{\sim}{a}(z'z)^{-1}\underset{\sim}{a}$$

$$E(\hat{\sigma}^2_{LS}) = \sigma^2,\ \mathrm{Var}(\hat{\sigma}^2_{LS}) = \frac{2\sigma^4}{N - (r+1)}$$

$$\mathrm{Cov}(\hat{\mu}_{LS},\hat{\sigma}^2_{LS}) = 0.$$

$$\therefore\ \hat{M}_{LS} = \frac{d}{\hat{\mu}_{LS} + \hat{\sigma}_{LS}\Phi^{-1}\left(a + \Phi(-\hat{\mu}_{LS}/\hat{\sigma}_{LS})\right)} = h\left(\hat{\mu}_{LS},\hat{\sigma}^2_{LS}\right)$$

and

$$\mathrm{Var}(\hat{\mu}_{LS}) = \mathrm{Var}\left[h\left(\hat{\mu}_{LS}, \hat{\sigma}^2_{LS}\right)\right] \doteq \left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right)^2 \mathrm{Var}(\hat{\mu}) + \left(\frac{\partial h}{\partial \hat{\sigma}^2_{LS}}\right)^2 \mathrm{Var}(\hat{\sigma}^2_{LS}) + 2\left(\frac{\partial h}{\partial \hat{\mu}_{LS}}\right)\left(\frac{\partial h}{\partial \hat{\sigma}^2_{LS}}\right)\mathrm{Cov}\left(\hat{\mu}_{LS}, \hat{\sigma}^2_{LS}\right).$$

As before the partials, which are given by (2.9), are evaluated at $(E(\hat{\mu}_{LS}), E(\hat{\sigma}^2_{LS}))$.

5. Multivariate Extensions of the Unconditional Models with Covariates

In this section we estimate the half-life and find the CI by modeling the mean of U_i under models I, II, and III presented in section 3. Here C_{ij} denotes, as before, the jth measurement of the ith individual at time T_{ij}. $\Delta T_{ij} = T_{i,j+1} - T_{ij}$, $j = 1, 2, \ldots, J-1$, $i = 1, 2, \ldots, N$.

Model I: From (3.3), the half-life of the ith subject is

$$Y_i = \frac{(J-1)\ln 2}{\sum_{j=1}^{J-1} \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}}} = \frac{d_i}{U_i}, \quad i = 1, 2, \ldots, N. \tag{5.1}$$

Let $\ln C_{ij} = \beta_{oj} + \sum_{k=1}^{r} \beta_k x_{ij,k} + \epsilon_{ij}$, $\epsilon_{ij} \sim N(0, \sigma_j^2)$, $i = 1, 2, \ldots, N$,

$j = 1, 2, \ldots, J-1.$ (5.2)

Then

$$U_i = \sum_{j=1}^{J-1} \frac{\ln(C_{ij}/C_{ij+1})}{\Delta T_{ij}}$$

$$= \sum_{j=1}^{J-1} \frac{\beta_{oj} - \beta_{oj+1})}{\Delta T_{ij}} + \sum_{j=1}^{J-1} \sum_{k=1}^{r} \beta_k \frac{(x_{ij,k} - x_{ij+1,k})}{\Delta T_{ij}} + \sum_{j=1}^{J-1} \frac{\left(\epsilon_{ij} - \epsilon_{ij+1}\right)}{\Delta T_{ij}}$$

$$= \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta T_{ij}} + \sum_{k=1}^{r} \beta_k \sum_{j=1}^{J-1} \frac{(x_{ij,k} - x_{ij+1,k})}{\Delta T_{ij}} + \sum_{j=1}^{J-1} \frac{\epsilon_{ij} - \epsilon_{ij+1}}{\Delta T_{ij}}$$

$$= \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta T_{ij}} + \sum_{k=1}^{r} \beta_k z_{ik} + \epsilon_i,$$

where $z_{ik} = \sum_{j=1}^{J-1} \frac{x_{ij,k} - x_{ij+1,k}}{\Delta T_{ij}}$, $Var(\varepsilon_i) = \sigma^2$, $i = 1, 2, \ldots, N$,

$k = 1, 2, \ldots, r.$ (5.3)

Notice that (5.3) is similar to (4.3). In what follows the following notations will be used.

$$\underset{\sim}{U} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}_{N\times 1}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0J-1} \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}_{(r+J-1)\times 1}$$

$$Z = \begin{bmatrix} \frac{1}{\Delta T_{11}} & \frac{1}{\Delta T_{12}} & \cdots & \frac{1}{\Delta T_{1J-1}} & z_{11} & z_{12}\cdots z_{1r} \\ \frac{1}{\Delta_{21}} & \frac{1}{\Delta T_{22}} & & \frac{1}{\Delta T_{2J-1}} & z_{21} & z_{22}\cdots z_{2r} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \frac{1}{\Delta T_{N1}} & \frac{1}{\Delta T_{N2}} & & \frac{1}{\Delta T_{N,J-1}} & z_{N1} & z_{N2}\cdots z_{Nr} \end{bmatrix}_{N\times(r+J-1)}$$

where $z_{ik} = \sum_{j=1}^{J-1} \frac{x_{ij,k} - x_{ij+1k}}{\Delta T_{ij}}$, $i = 1,2,\ldots,N$; $k = 1,2,\ldots,r.$ (5.4)

The pdf of $Y = \frac{(J-1)\ln 2}{\sum_{j=1}^{J-1} \frac{1}{\Delta T_j} \ln(c_j/c_{j+1})} = \frac{d}{U}$

is $f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \frac{d}{y^2} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{d}{y} - \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta T_j} - \sum_{k=1}^{r} \beta_k z_k\right)^2\right)$, $y \neq 0$. (5.5)

The log-likelihood of the sample $Y_1, Y_2 \ldots Y_N$ is

$$L(\underset{\sim}{\beta}, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \frac{d_i}{y_i^2} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{d_i}{y_i} - \sum_{j=1}^{J-1} \frac{\beta_{oj}}{\Delta T_{ij}} - \sum_{k=1}^{r} \beta_k z_{ik}\right)^2\right) \quad (5.6)$$

The nonconstant part of the log-likelihood is given by

$$\ell_F(\underset{\sim}{\beta}, \sigma^2) = -\frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(\frac{d_i}{y_i} - \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta T_{ij}} - \sum_{k=1}^{r} \beta_k z_{ik}\right)^2. \quad (5.7)$$

The likelihood equations are

$$\frac{\partial \ell_F(\underset{\sim}{\beta},\sigma^2)}{\partial \sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}\left(\frac{d_i}{y_i} - \sum_{j=1}^{J-1}\frac{\beta'_{oj}}{\Delta T_{ij}} - \sum_{k=1}^{r}\beta_k z_{ik}\right)^2 = 0 \qquad (5.8)$$

$$\begin{cases} \dfrac{\partial \ell_F(\underset{\sim}{\beta},\sigma^2)}{\partial \beta'_{o\ell}} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{N}\left(\frac{d_i}{y_i} - \sum_{j=1}^{J-1}\frac{\beta'_{oj}}{\Delta T_{ij}} - \sum_{k=1}^{r}\beta_k Z_{ik}\right)\frac{1}{\Delta T_{i\ell}} = 0, \ \ell = 1,2,\ldots,J-1, \\ \dfrac{\partial \ell_F(\underset{\sim}{\beta},\sigma^2)}{\partial \beta_k} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{N}\left(\frac{d_i}{y_i} - \sum_{j=1}^{J-1}\frac{\beta'_{oj}}{\Delta T_{ij}} - \sum_{k=1}^{r}\beta_k z_{ik}\right)z_{ik} = 0, \ k = 1,2,\ldots,r. \end{cases} \qquad (5.9)$$

Solving (5.8) and (5.9), we have

$$\hat{\sigma}^2 = \frac{1}{N}(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})$$

$$\underset{\sim}{\hat{\beta}} = (Z'Z)^{-1}Z'\underset{\sim}{U}.$$

The overall mean μ of U can be estimated by

$$\hat{\mu} = \left[\sum_{i=1}^{N}\frac{1}{N\Delta T_{i1}} \ \sum_{i=1}^{N}\frac{1}{N\Delta T_{i2}} \ldots \sum_{1}^{N}\frac{1}{N\Delta T_{iJ-1}} \ \sum_{i=1}^{N}\frac{z_{i1}}{N} \ldots \sum_{i=1}^{N}\frac{z_{ir}}{N}\right]\underset{\sim}{\hat{\beta}}_{1\times(r+j-1)} \qquad (5.10)$$

$$= \underset{\sim}{a}\underset{\sim}{\hat{\beta}}, \text{ where } \underset{\sim}{a} \text{ is a vector of column means of Z in (5.4).}$$

The expectations, variances, and covariance of $\hat{\mu}$ and $\hat{\sigma}^2$ are given as follows:

$E(\hat{\mu}) = \mu$, $Var(\hat{\mu}) = \underset{\sim}{a}\ Var(\underset{\sim}{\hat{\beta}})\underset{\sim}{a}' = \underset{\sim}{a}(Z'Z)^{-1}\underset{\sim}{a}'$.

Since $(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})'(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})/\sigma^2 \sim \chi^2_{N-(r+J-1)}$,

$$E(\hat{\sigma}^2) = (N - (r + J - 1))\sigma^2/N ,$$

$$Var(\hat{\sigma}^2) = 2(N - (r + J - 1))\sigma^4/N^2$$

$$Cov(\hat{\mu}, \hat{\sigma}^2) = \underset{\sim}{a}\ Cov(\underset{\sim}{\hat{\beta}}, \hat{\sigma}^2) = 0.$$

The point estimate and the interval estimate of M can be obtained as before with appropriate changes.

Remark. For least squares estimates, the following changes are to be made.

$$\hat{\mu}_{LS} = \hat{\mu}_{ML} = \hat{\mu} \text{ and}$$

$$\hat{\sigma}^2_{LS} = \frac{N}{N-(r+J-1)}\,\hat{\sigma}^2_{ML}.$$

Hence $E(\sigma^2_{LS}) = \sigma^2$ and

$$Var(\hat{\sigma}^2_{LS}) = \frac{2\sigma^4}{N-(r+J-1)}.$$

Model II: Following the notation of section 3, from equation (3.6), we have the half-life of the ith individual as

$$Y_i = \frac{\ln 2 \sum_{j=1}^{J-1} \Delta_1 T_{ij}}{\sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1})} = \frac{d_i}{U_i}, \text{ where}$$

$$d_i = \ln 2 \sum_{j=1}^{J-1} \Delta_1 T_{ij},\ \Delta_1 T_{ij} = T_{ij} - T_{i1},\ i = 1,2,\ldots,N, j = 1, 2,\ldots,J.$$

Here also as in (5.2), let $\ln C_{ij} = \beta_{oj} + \sum_{k=1}^{r} \beta_k x_{ij,k} + \epsilon_{ij}$,

where $\epsilon_{ij} \sim N(0,\sigma_j^2)$, $i = 1, 2,\ldots,N$, $j = 1, 2,\ldots,J$. (5.13)

Thus $U_i = \sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1})$

$$= \sum_{j=1}^{J-1} (\beta_{o1} - \beta_{oj+1}) + \sum_{j=1}^{J-1} \sum_{k=1}^{r} \beta_k (x_{i1,k} - x_{ij+1,k}) + \sum_{j=1}^{J-1} (\epsilon_{i1} - \epsilon_{ij+1})$$

$$= \beta_o + \sum_{k=1}^{r} \beta_k \sum_{j=1}^{J-1} (x_{i1,k} - x_{ij+1,k}) + \sum_{j=1}^{J-1} (\epsilon_{i1} - \epsilon_{ij+1})$$

$$= \beta_o + \sum_{k=1}^{r} \beta_k z_{ik} + \epsilon_i,$$

where $\beta_o = \sum_{j=1}^{J-1} (\beta_{01} - \beta_{oj+1})$, $z_{ik} = \sum_{j=1}^{J-1} (x_{i1,k} - x_{ij+1,k})$,

$$\epsilon_i = \sum_{j=1}^{J-1} (\epsilon_{i1} - \epsilon_{ij+1}),\ i = 1, 2,\ldots,N,\ j = 1, 2,\ldots,J-1.$$

If we follow exactly as in model I, with $\sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta T_{ij}}$ replaced by β_o, z_{ik} in model I replaced by z_{ik} in model II, we will get

$$\hat{\sigma}^2 = \frac{1}{N} (\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})'(\underset{\sim}{U} - Z\underset{\sim}{\hat{\beta}})$$

$$\underset{\sim}{\hat{\beta}} = (Z'Z)^{-1}Z'\underset{\sim}{U}.$$

The overall mean of U can be estimated by

$$\hat{\mu} = \left[1 \quad \sum_{i=1}^{N} \frac{z_{i1}}{N} \quad \sum_{i=1}^{N} \frac{z_{i2}}{N} \quad \ldots \quad \sum_{i=1}^{N} \frac{z_{ir}}{N}\right] \underset{\sim}{\hat{\beta}}$$

where

$$Z = \begin{bmatrix} 1 & z_{11} & z_{12} & \ldots & z_{1r} \\ 1 & z_{21} & z_{22} & \ldots & z_{2r} \\ \vdots & \vdots & \vdots & & \\ 1 & z_{N1} & z_{N2} & \ldots & z_{Nr} \end{bmatrix}_{N\times(r+J-1)}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}_{(r+1)\times 1}, \quad d_i = \ln 2 \sum_{j=1}^{J-1} \Delta_1 T_{ij},$$

$$U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}_{N\times 1}, \quad U_i = \sum_{j=1}^{J-1} \ln(C_{i1}/C_{ij+1}).$$

The estimates of M are then obtained as before.

Model III: Here the half-life of the ith individual from (3.9) is

$$Y_i = \frac{(J-1)\ln 2}{\sum_{j=1}^{J-1} \frac{1}{\Delta_1 T_{ij}} (\ln C_{i1}/C_{ij+1})} = \frac{d_i}{U_i}, \quad i = 1, 2, \ldots, N.$$

Following Model I or Model II, let

$$\ln C_{ij} = \beta_{oj} + \sum_{k=1}^{r} \beta_k x_{ij,k} + \varepsilon_{ij}, \quad i = 1,2,\ldots N, \quad j = 1,2,\ldots,J-1.$$

Then $U_i = \sum_{j=1}^{J-1} \frac{1}{\Delta_1 T_{ij}} \ln(C_{i1}/C_{ij+1})$

$$= \sum_{j=1}^{J-1} \frac{(\beta_{o1} - \beta_{oj+1})}{\Delta_1 T_{ij}} + \sum_{j=1}^{J-1} \sum_{k=1}^{r} \beta_k \frac{(x_{i1,k} - x_{ij+1,k})}{\Delta_1 T_{ij}} + \sum_{j=1}^{J-1} \frac{(\varepsilon_{i1} - \varepsilon_{ij+1})}{\Delta_1 T_{ij}}$$

$$= \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta_1 T_{ij}} + \sum_{k=1}^{r} \beta_k \sum_{j=1}^{J-1} \frac{(x_{i1,k} - x_{ij+1,k})}{\Delta_1 T_{ij}} + \varepsilon_i,$$

$$= \sum_{j=1}^{J-1} \frac{\beta'_{oj}}{\Delta_1 T_{ij}} + \sum_{k=1}^{r} \beta_k z_{ik} + \varepsilon_i,$$

where $z_{ik} = \sum_{j=1}^{J-1} \frac{(x_{i1,k} - x_{ij+1,k})}{\Delta_1 T_{ij}}$, $i = 1, 2, \ldots N$, $k = 1,2,\ldots,r$. We now proceed as in model I with the following changes in equation (5.4).

Here $U_i = \sum_{j=1}^{J-1} \frac{1}{\Delta_1 T_{ij}} \ln(C_{i1}/C_{ij+1})$, $z_{ik} = \sum_{j=1}^{J-1} \frac{x_{i1,k} - x_{ij+1,k}}{\Delta_1 T_{ij}}$, and ΔT_{ij} replaced by $T_{ij} - T_{i1}$.

All these extended models with covariates are applied to the Ranch Hand data of 235 subjects and 1982, 1987 and 1992 measurements with age and percent body fat as covariates. The results are summarized in the following table:

TABLE

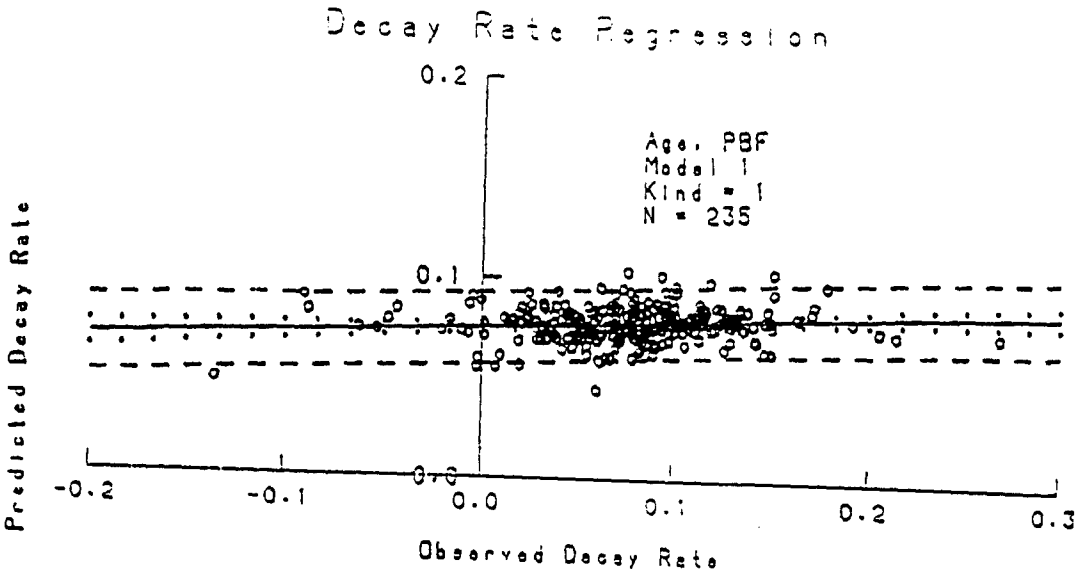
Estimates of TCDD Half-Life Based on Multivariate Extensions with Age and Percent Body Fat as Covariates

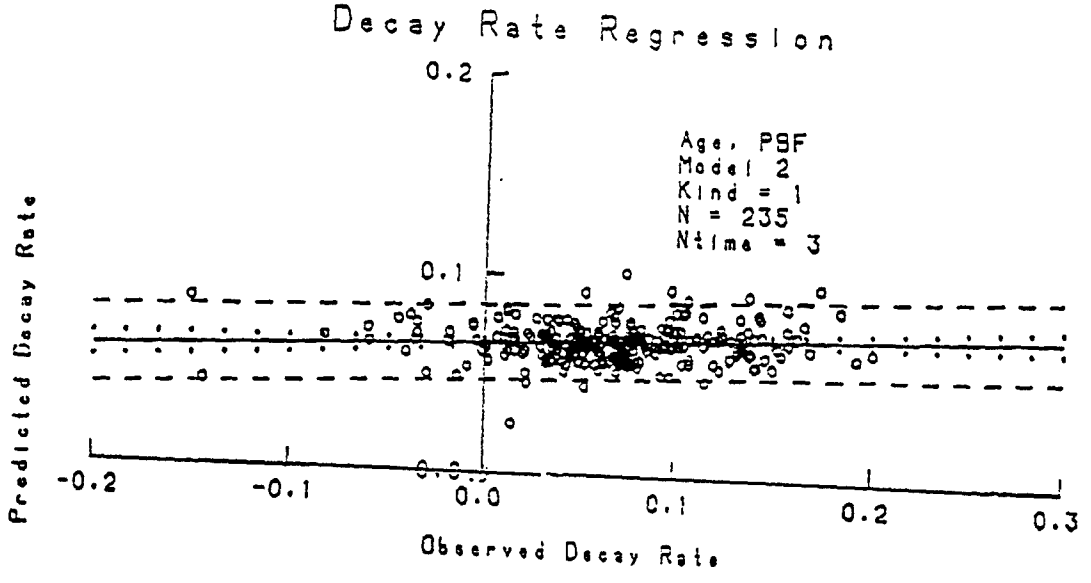
MODEL	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}$	$\hat{\sigma}$
1	.31273	-.012272	-.025315	.077195	.051081
					.050645
2	.42809	-.034688	-.029114	.067756	.05386
					.05340
3	-.22539	-.088611	-.033328	.062439	.058811
					.058308

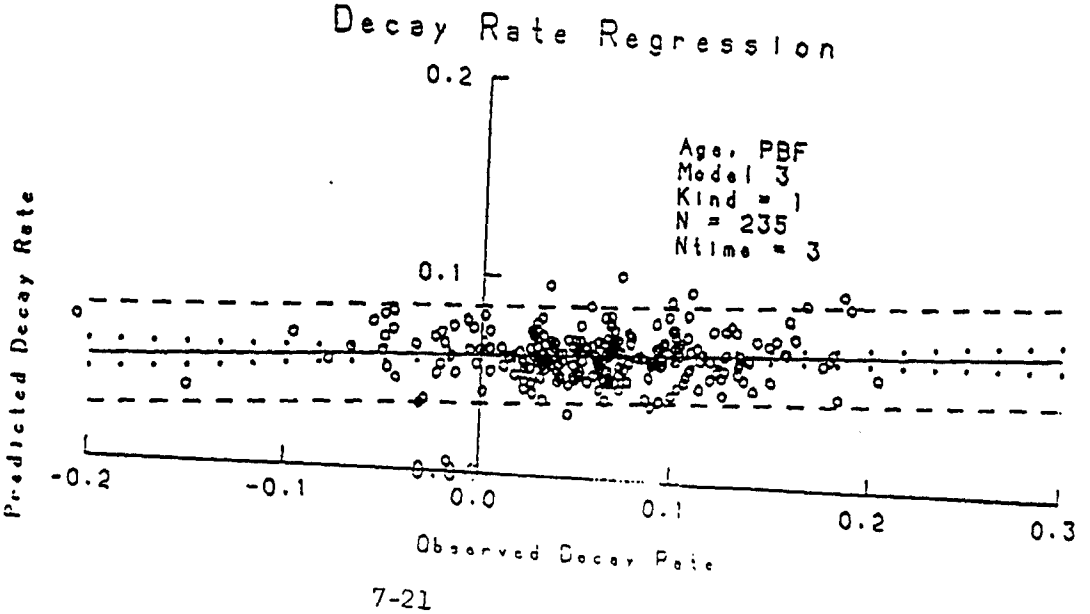
MODEL	$\hat{M}$	$\hat{\sigma}_{\hat{\mu}}$	95% C.I. for M	Width of C.I.
1	8.0973	.25630	(7.5949, 8.5996)	1.0047
	8.1242	.25595	(7.6225, 8.6258)	1.0033
2	8.4544	.29318	(7.8797, 9.0290)	1.1493
	8.4949	.29344	(7.9198, 9.0700)	1.1502
3	8.2341	.30612	(7.6341, 8.8341)	1.2000
	8.2842	.30700	(7.6825, 8.8859)	1.2034

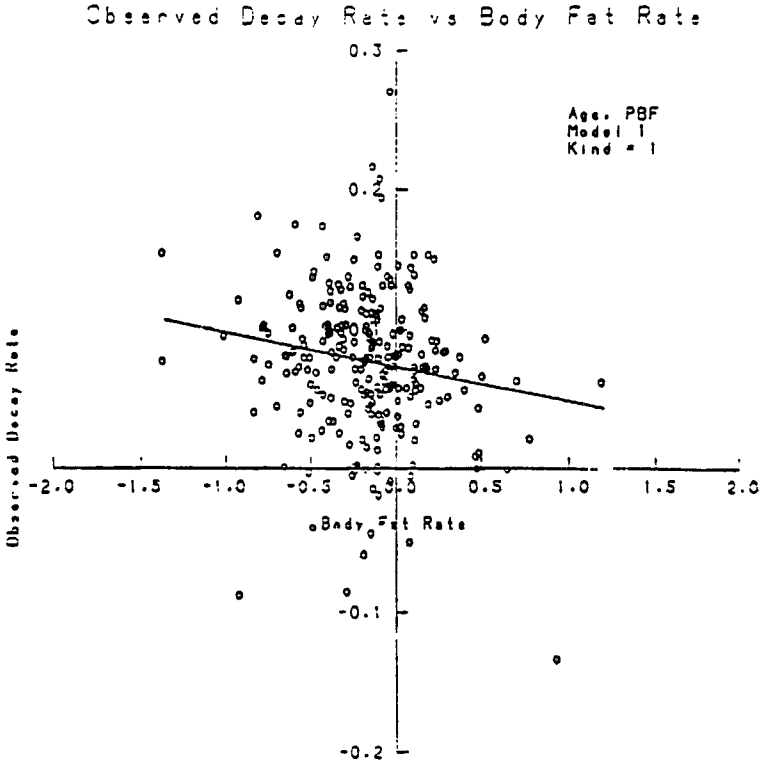
Upper (lower) numbers have been calculated using MLE (LSE).

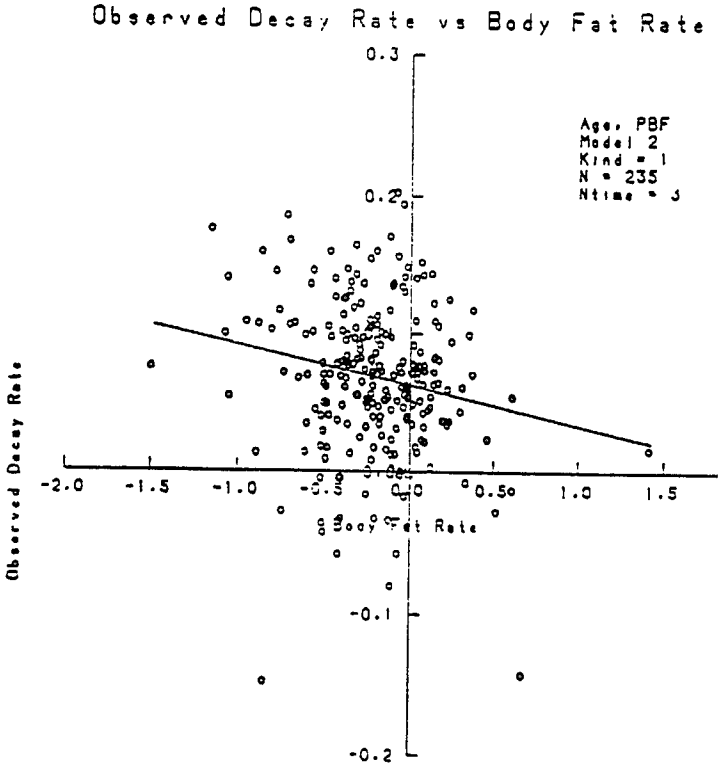
The fitting of the models is displayed by the graphs numbered 1-9 showing (1) decay rate regression, (2) observed decay rate vs. body fat rate and (3) decay rate Q - Q plot. It is observed that the models presented in this section fit the Ranch Hand data very well.

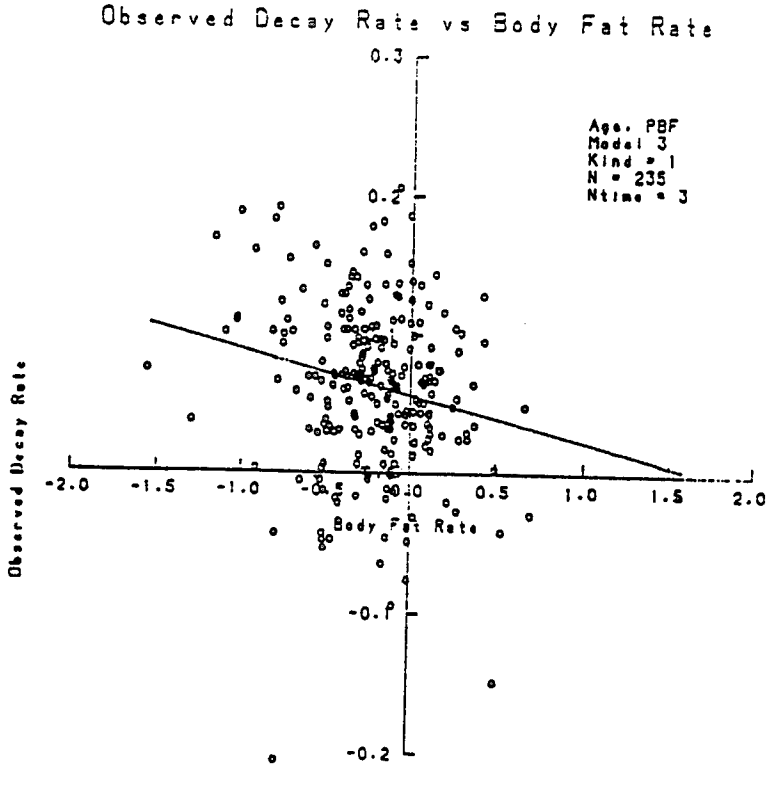

Decay Rate Regression
Age, PBF
Model 1
Kind = 1
N = 235
Predicted Decay Rate
Observed Decay Rate
0.2
0.1
0.0
-0.2
-0.1
0.0
0.1
0.2
0.3

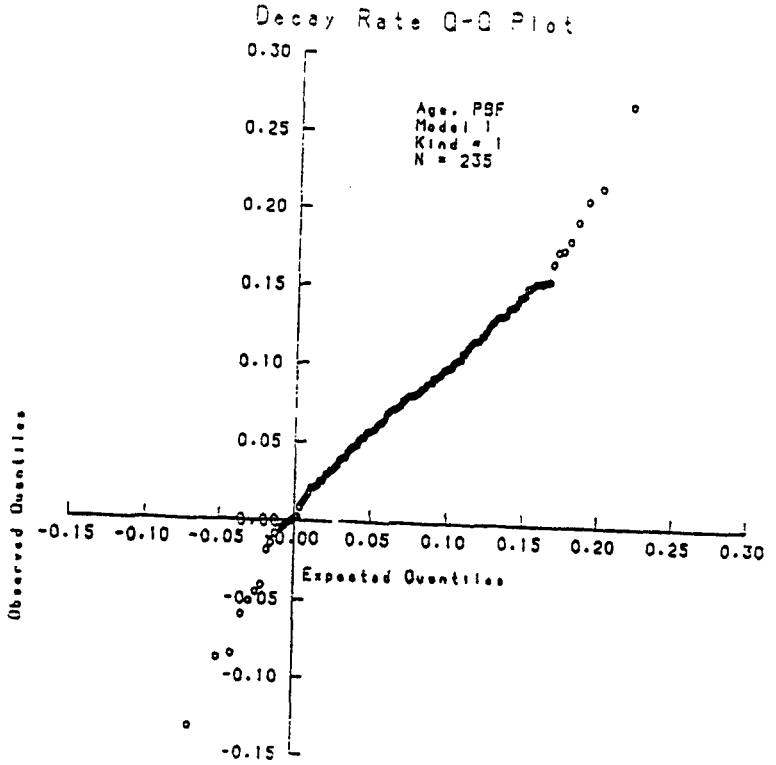

Decay Rate Regression
Age, PBF
Model 2
Kind = 1
N = 235
Ntime = 3
Predicted Decay Rate
Observed Decay Rate
0.2
0.1
0.0
-0.2
-0.1
0.0
0.1
0.2
0.3

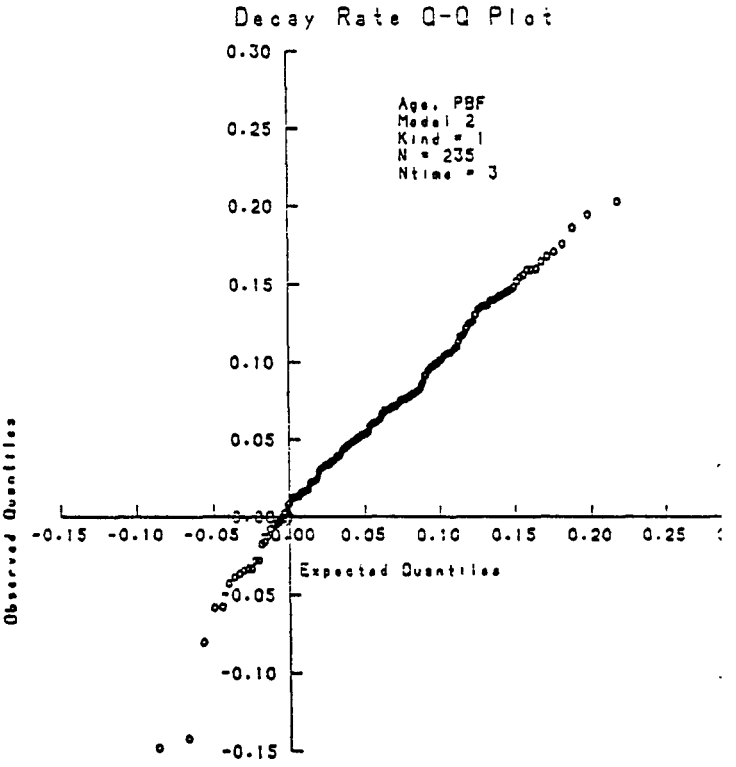

Decay Rate Regression
Age, PBF
Model 3
Kind = 1
N = 235
Ntime = 3
Predicted Decay Rate
Observed Decay Rate
0.2
0.1
0.0
-0.2
-0.1
0.0
0.1
0.2
0.3

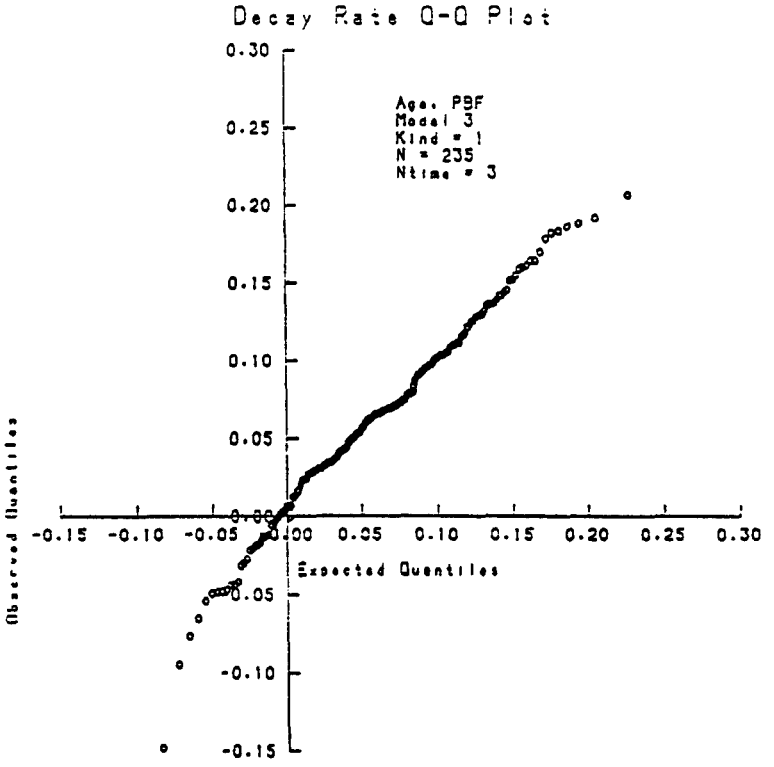

Observed Decay Rate vs Body Fat Rate
Age, PBF
Model 1
Kind = 1
Observed Decay Rate
Body Fat Rate
0.3
0.2
-0.1
-0.2
-2.0
-1.5
-1.0
-0.5
0.5
1.0
1.5
2.0


Observed Decay Rate vs Body Fat Rate
Age, PBF
Model 2
Kind = 1
N = 235
Observed Decay Rate
Body Fat Rate
0.3
0.2
-0.1
-0.2
-2.0
-1.5
-1.0
0.5
1.0
1.5


Observed Decay Rate vs Body Fat Rate
Age, PBF
Model 3
Kind = 1
N = 235
Ntime = 3
Observed Decay Rate
Body Fat Rate
0.3
0.2
-0.2
-2.0
-1.5
-1.0
0.5
1.0
1.5
2.0


Decay Rate Q-Q Plot
Age, PBF
Model 1
Kind = 1
N = 235
Observed Quantiles
Expected Quantiles
0.30
0.25
0.20
0.15
0.10
0.05
-0.05
-0.10
-0.15
-0.15 -0.10 -0.05 0.05 0.10 0.15 0.20 0.25 0.30


Decay Rate Q-Q Plot
Age, PBF
Model 2
Kind = 1
N = 235
Ntime = 3
Observed Quantiles
Expected Quantiles
0.30
0.25
0.20
0.15
0.10
0.05
-0.05
-0.10
-0.15
-0.15 -0.10 -0.05 0.05 0.10 0.15 0.20 0.25


Decay Rate Q-Q Plot
Age, PBF
Model 3
Kind = 1
N = 235
Ntime = 3
Observed Quantiles
Expected Quantiles
0.30
0.25
0.20
0.15
0.10
0.05
-0.05
-0.10
-0.15
-0.15 -0.10 -0.05 0.05 0.10 0.15 0.20 0.25 0.30

6. VARIANCE-COVARIANCE MATRIX OF THE ESTIMATORS OF THE PARAMETERS (SINGLE COVARIATE)

In this section, we shall obtain the variance-covariance matrix of the estimators of the parameters in the mixed-effects linear model, which is introduced below.

Let the jth measurement on the ith individual be given by

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 t_{ij} + \varepsilon_{ij},\ j = 1, 2.$$

Then

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix}$$

$$= X_i\beta + Zi\alpha_i + \varepsilon_i,\ i = 1, 2, \ldots, N, \qquad (6.1)$$

where X_i and Z_i are known 2×2 and $2 \chi 1$ design matrics, β is a vector of 2 fixed population parameters to be estimated, α_i is random effect, and ε_i is a random error vector associated with the ith individual, such that

$$\varepsilon_i \sim N(0_{2\times1}, \sigma^2 I_2)$$

$$\alpha_i \sim N(0, \sigma_\alpha^2 = \sigma^2\theta).$$

It follows that

$$Y_i|\alpha_i \sim N(X_i\beta + Z_i\alpha_i, \sigma^2 I_2)$$

and

$$Y_i \sim N(X_i\beta, \Sigma_i),$$

where $\Sigma_i = \sigma^2 V_i(\theta)$, $V_i(\theta) = I_2 + \theta\perp_2$ for all i. I_2 denotes 2×2 identity matrix and $\perp_2$ is a 2×2 matrix of ones. Y_i and Y_j are independent for $i \neq j$.

Now if we let $Y' = [Y_1', Y_2', \ldots, Y_N']$, $X' = [X_1', X_2', \ldots, x_N']$,

$Z = \text{diag}(Z_1, Z_2, \ldots, Z_N)$, $\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_N)$, $\varepsilon' = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_N')$, $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_N)$, then we can write the full model as

$$Y = X\beta + Z\alpha + \varepsilon, \tag{6.2}$$

where

$$Y \mid \alpha \sim N(X\beta + Z\alpha, \sigma^2 I_{2N})$$

and

$$Y \sim N(X\beta, \Sigma),$$

where

$$\Sigma = \sigma^2 \left(I_{2N} + \theta \perp_{2N}\right).$$

Since $Y_i \sim N(X_i\beta, \sigma^2 V_i)$, $i = 1, 2, \ldots, N$, and independent, the full-likelihood, without the constant term, is given by

$$\ell_F = -N \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{N} \log \left| V_i(\theta) \right| - \frac{1}{2\sigma^2} \sum_{i=1}^{N} r_i^T V_i^{-1}(\theta) r_i, \tag{6.3}$$

where $r_i = y_i - X_i\beta$. Since $V_i(\theta)$ is dependent on θ, in order to save space we will drop the argument θ and write V_i instead of $V_i(\theta)$.

A general criticism of the ML estimators for the variance components σ and θ, obtained from (6.3), is that they are biased downward because they do not take into account the loss of degrees of freedom from the estimation of β. The restricted maximum likelihood (RML) method corrects for this by defining estimates of the variance components as the maximizers of the log-likelihood based on $2N - 2$ linearly independent contrasts, where $2N$ is the total number of observations from all individuals. Thus our estimates of θ, σ and β are based on restricted log-likelihood (see Lindstrom and Bates,

1988), ℓ_R (ignoring the constant term) given below,

$$\ell_R(\beta,\sigma,\theta|y) = -\frac{1}{2}\log\left|\sigma^{-2}\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right| + \ell_F(\beta,\sigma,\theta|y)$$

$$= -(N-1)\log\sigma^2 - \frac{1}{2}\log\left|\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right| - \frac{1}{2\sigma^2}\sum_{i=1}^{N} r_i^T V_i^{-1} r_i$$

$$-\frac{1}{2}\sum_{i=1}^{N}\log|V_i| \qquad (6.4)$$

To simplify calculations we obtain first the RML estimate of σ^2, which is

$$\hat{\sigma}^2_{RML} = \frac{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i}{2N-2} \qquad (6.5)$$

and substitute $\hat{\sigma}^2_{RML}$ in equation (6.4) to get the profile log-likelihood of β and θ, as

$$p_R(\beta,\theta|y) = -(N-1)\log\sum_{1}^{N} r_i^T V_i^{-1} r_i - \frac{1}{2}\log\left|\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right| - \frac{1}{2}\sum_{i=1}^{N}\log|V_i|. \qquad (6.6)$$

The derivatives of the profile log-likelihood with respect to β and θ are given below as:

$$\frac{\partial p_R(\beta,\theta|y)}{\partial\beta} = 2(N-1)\frac{\sum_{i=1}^{N} X_i^T V_i^{-1} r_i}{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i} = 2(N-1)\frac{H_2(V^{-1})}{H_1(V^{-1})}. \qquad (6.7)$$

$$\frac{\partial p_R(\beta,\theta|y)}{\partial\theta} = (N-1)\frac{\sum_{i=1}^{N} r_i^T A r_i}{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i} + \frac{1}{2}\,\text{trace}\left\{\left(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i^T A X_i\right)\right\}$$

$$- \frac{1}{2} \sum_{i=1}^{N} \operatorname{trace}\left(V_i^{-1} \frac{\partial V_i}{\partial \theta}\right)$$

$$= \frac{(N-1)}{(2\theta+1)^2} \frac{\sum_{i=1}^{N} r_i^T(\perp_2) r_i}{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i} + \frac{1}{2} \operatorname{trace}\left(\left(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^T(\perp_2) X_i\right)\right)/(2\theta+1)^2$$

$$- \frac{N}{2} \operatorname{trace}\left(\frac{1}{(2\theta+1)}(\perp_2)\right)$$

$$= \frac{(N-1)}{(2\theta+1)^2} \frac{H_1(\perp_2)}{H_1(V^{-1})} + \frac{1}{2(2\theta+1)^2} \operatorname{trace}\left((H(V^{-1}))^{-1} H(\perp_2)\right) - \frac{N}{2\theta+1}. \quad (6.8)$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T \partial\beta} = 2(N-1)\left[- \frac{\sum_{i=1}^{N} X_i^T V_i^{-1} X_i}{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i} + \frac{2\left(\sum_{i=1}^{N} X_i^T V_i^{-1} r_i\right)\left(\sum_{i=1}^{N} X_i^T V_i^{-1} r_i\right)^T}{\left(\sum_{i=1}^{N} r_i^T V_i^{-1} r_i\right)^2}\right]$$

$$= 2(N-1)\left[- \frac{H(V^{-1})}{H_1(V^{-1})} + \frac{2H_2(V^{-1})(H_2(V^{-1}))^T}{(H_1(V^{-1}))^2}\right]. \quad (6.9)$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta^2} = \frac{(N-1)}{(2\theta+1)^3}\left[- \frac{4H_1(\perp_2)}{H_1(V^{-1})} + \frac{1}{(2\theta+1)}\left(\frac{H_1(\perp_2)}{H_1(V^{-1})}\right)^2\right]$$

$$+ \frac{1}{2(2\theta+1)^3} \operatorname{trace}\left(\frac{(H(V^{-1}))^{-1} H(\perp_2)(H(V^{-1}))H(\perp_2)}{(2\theta+1)} - 4(H(V^{-1}))^{-1} H(\perp_2)\right)$$

$$+ \frac{2N}{(2\theta+1)^2}. \quad (6.10)$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta\partial\beta} = \frac{2(N-1)}{(2\theta+1)^2}\left[- \frac{H_2(\perp_2)}{H_1(V^{-1})} + \frac{H_2(V^{-1})H_1(\perp_2)}{(H_1(V^{-1}))^2}\right], \quad (6.11)$$

where $V_i = V = \begin{bmatrix} \theta+1 & \theta \\ \theta & \theta+1 \end{bmatrix}$, $|V_i| = 2\theta + 1$, $V_i^{-1} = \frac{1}{2\theta+1}\begin{bmatrix} \theta+1 & -\theta \\ -\theta & \theta+1 \end{bmatrix}$,

$$\perp_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad A = V_i^{-1} \frac{\partial V_i}{\partial\theta} V_i^{-1} = \frac{\perp 2}{(2\theta+1)^2}, \quad V_i^{-1} \frac{\partial V_i}{\partial\theta} = \frac{\perp_2}{(2\theta+1)},$$

and $H(S) = \sum_{i=1}^{N} X_i^T S X_i$, $H_1(S) = \sum_{i=1}^{N} r_i^T S r_i$, and $H_2(S) = \sum_{i=1}^{N} X_i^T S r_i$.

For details on partials, see appendix I. Since

$\sigma^{-2} H_1(V^{-1}) \sim \chi^2_{2N}$ and $H_2(V^{-1}) \sim N(0_{2\times 1}, \sigma^2 H(V^{-1}))$ (see Appendix II),

$$E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T\partial\beta}\right) = 2(N-1)\left[-H(V^{-1})E\left(\frac{1}{H_1(V^{-1})}\right) + 2E\left(\frac{H_2(V^{-1})(H_2(V^{-1}))^T}{(H_1(V^{-1}))^2}\right)\right]$$

$$= 2(N-1)\left[-\frac{H(V^{-1})}{2(N-1)\sigma^2} + 2E\left(\frac{H_2(V^{-1})(H_2(V^{-1}))^T}{(H_1(V^{-1}))^2}\right)\right]$$

$$\doteq 2(N-1)\left[-\frac{H(V^{-1})}{2(N-1)\sigma^2} + 2\frac{H(V^{-1})}{4N^2\sigma^2}\right], \text{ see Appendix III,}$$

$$= -\left(\frac{N^2+1-N}{N^2}\right)\frac{H(V^{-1})}{\sigma^2}$$

$$= -\left(\frac{N^2+1-N}{N^2+N}\right)\frac{1}{(2\theta+1)\sigma^2}\begin{bmatrix} 2N & \sum_{i=1}^{N}(t_{i1}+t_{i2}) \\ \sum_{i=1}^{N}(t_{i1}+t_{i2}) & \sum_{i=1}^{N}(\theta+1)(t_{i1}-t_{i2})^2+2t_{i1}t_{i2} \end{bmatrix}. \quad (6.12)$$

Since $\frac{\sigma^{-2}}{2(2\theta+1)} H_1(\perp_2) \sim \chi^2_N$ and $\sigma^{-2}H_1(V^{-1}) \sim \chi^2_{2N}$,

$$\frac{\sigma^{-2}H_1(\perp_2)}{2(2\theta+1)\sigma^{-2}H_1(V^{-1})} = \frac{1}{2(2\theta+1)}\frac{H_1(\perp_2)}{H_1(V^{-1})} \sim \text{Beta}(N/2, N/2) \text{ (appendix II).}$$

Therefore

$$E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta^2}\right) = \frac{(N-1)}{(2\theta+1)^3}\left[-4E\left(\frac{H_1(\perp_2)}{H_1(V^{-1})}\right) + \frac{1}{2\theta+1}E\left(\frac{H_1(\perp_2)}{H_1(V^{-1})}\right)^2\right]$$

$$+ \frac{2N}{(2\theta+1)^2} + k$$

$$= \frac{N-1}{(2\theta+1)^3}\left[-4(2\theta+1) + 4(2\theta+1)\frac{(N+2)}{4(N+1)}\right] + \frac{2N}{(2\theta+1)^2} + k$$

$$= - \frac{(N^2 - 3N - 2)}{(N + 1)(2\theta + 1)^2} + k, \qquad (6.13)$$

where

$$k = \frac{1}{2(2\theta + 1)^3} \text{ trace } \left\{ \frac{(H(V^{-1}))^{-1} H(\perp_2)(H(V^{-1}))^{-1} H(\perp_2)}{2\theta + 1} - 4(H(V^{-1}))^{-1} H(\perp_2) \right\}$$

$$E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial\theta\partial\beta}\right) = \frac{2(N-1)}{(2\theta+1)^2}\left[- E\left(\frac{H_2(\perp_2)}{H_1(V^{-1})}\right) + E\left(\frac{H_2(V^{-1})H_1(\perp_2)}{(H_1(V^{-1}))^2}\right)\right]$$

$$\doteq \frac{2(N-1)}{(2\theta+1)^2}\left[0_{2\times 1} + 0_{2\times 1}\right] = 0_{2\times 1},$$

see appendix III. (6.14)

Therefore, the information matrix I of $\hat{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$ is of the form

$$I = \begin{bmatrix} a & 0 & 0 \\ 0 & b & d \\ 0 & d & c \end{bmatrix},$$

where

$$a = - E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial\theta^2}\right)$$

$$= \frac{N^2 - 3N - 2}{(N + 1)(2\theta + 1)^2} - \frac{1}{2(2\theta + 1)^3} \text{ trace } \left\{ \frac{(H(V^{-1}))^{-1} H(\perp_2)(H(V^{-1}))^{-1} H(\perp_2)}{2\theta + 1} \right\}$$

$$+ \frac{1}{2(2\theta + 1)^3} \text{ trace } (4(H(V^{-1}))^{-1} H(\perp_2)),$$

$$b = - E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial\beta_0^2}\right) = \frac{2(N^2 + 1 - N)}{\sigma^2(2\theta + 1)N}, \quad \text{see (6.12)}$$

$$d = - E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial\beta_0\partial\beta_1}\right) = \frac{\left(\sum_{i=1}^{N} t_{i1} + t_{i2}\right)(N^2 + 1 - N)}{\sigma^2(2\theta + 1)N^2}$$

$$c = - E\left(\frac{\partial^2 p_R(\beta, \theta | y)}{\partial\beta_1^2}\right) = \frac{\sum_{i=1}^{N} \{(\theta+1)(t_{i1} - t_{i2})^2 + 2t_{i1}t_{i2}\}(N^2+1-N)}{\sigma^2(2\theta + 1)N^2}$$

Hence the variance-covariance matrix of $\hat{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$ is

$$I^{-1} = \begin{bmatrix} \frac{1}{a} & 0 & 0 \\ 0 & \frac{c}{bc-d^2} & \frac{-d}{bc-d^2} \\ 0 & \frac{-d}{bc-d^2} & \frac{b}{bc-d^2} \end{bmatrix} . \qquad (6.15)$$

Hence $\text{Var}(\hat{\theta}) = \frac{1}{a}$, $\text{Var}(\hat{\beta}_0) = \frac{c}{bc-d^2}$, $\text{Var}(\hat{\beta}_1) = \frac{b}{bc-d^2}$,

$$\text{Cov}(\hat{\beta}_0,\hat{\beta}_1) = \frac{-d}{bc-d^2} \quad \text{and} \quad \text{Cov}(\hat{\theta},\hat{\beta}_0) = 0 = \text{Cov}(\hat{\theta},\hat{\beta}_1).$$

The half-life of the population for the model (6.1) is given by

$$t_{1/2} = \frac{\ln 2}{\lambda} = -\frac{\ln 2}{\beta_1}.$$

We are interested in the $\text{Var}(\hat{t}_{1/2})$ and hence in the $\text{Var}(\hat{\beta}_1)$, which is given by

$$\text{Var}(\hat{\beta}_1) = \frac{b}{bc - d^2}$$

$$= \frac{\frac{2N(N^2 + 1 - N)}{\sigma^2(2\theta+1)N^2}}{\left(\frac{N^2 + 1-N}{\sigma^2(2\theta+1)N^2}\right)^2 \left\{ 2N\{ \sum_{i=1}^{N} (\theta+1)(t_{i1}-t_{i2})^2+2t_{i1}t_{i2} - \left\{ \sum_{i=1}^{N} (t_{it}+t_{i2}) \right\}^2 \right\}}$$

$$= \frac{2N^3\sigma^2(2\theta + 1)}{(N^2+1-N)\left\{ 2N\left\{ \sum_{i=1}^{N} (\theta+1)(t_{i1} - t_{i2})^2+2t_{i1}t_{i2} \right\} - \left(\sum_{i=1}^{N} t_{i1}+t_{i2} \right)^2 \right\}}. \qquad (6.16)$$

7. VARIANCE-COVARIANCE MATRIX OF THE ESTIMATORS OF THE PARAMETERS OF THE GENERAL MODEL

In this section, we shall obtain the variance-covariance matrix of the estimators of the parameters in the mixed effects model for the repeated measures. Here we will consider N subjects, (p + 1)

covariates, and J measurements on each subject whereas in section 6, p = 1, J = 2. Let the jth measurement on the ith individual be given by

$$Y_{ij} = \beta_0 + \alpha_i + \sum_{\ell=1}^{p} \beta_\ell x_{ij\ell} + \varepsilon_{ij}, \quad j = 1, 2, \ldots, J$$

Then

$$Y_i = X_i\beta + \alpha_i Z_i + \varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{7.1}$$

where X_i and Z_i are $J \times (p + 1)$ and $J \times 1$ design matrices, β is a $(p + 1) \times 1$ vector of fixed population parameters to be estimated, α_i and ε_i are random such that

$$\varepsilon_i \sim N\left(0_{J\times 1}, \sigma^2 I_J\right)$$

$$\alpha_i \sim N\left(0, \sigma_\alpha^2 = \theta\sigma^2\right).$$

It follows that $Y_i \big| \alpha_i \sim N\left(X_i\beta + Z_i\alpha_i, \sigma^2 I_J\right)$ and

$$Y_i \sim N(X_i\beta, \Sigma_i),$$

where $\Sigma_i = \sigma^2 V_i(\theta) = \sigma^2(I_J + \theta Z_i Z_i^T) = \sigma^2(I_J + \theta 1_J)$ and Y_i is independent of Y_j, $i \neq j$. As in section 6, we can represent the full model as

$$Y = X\beta + Z\alpha + \varepsilon, \tag{7.2}$$

where $Y' = (Y_1', Y_2', \ldots, Y_N')$, $X' = [X_1', X_2', \ldots, X_N']$
$Z = \text{diag}(Z_1, Z_2, \ldots, Z_N)$, $\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_N)$, $\varepsilon' = (\varepsilon_1', \varepsilon_2, \ldots, \varepsilon_N')$ and

$$Y \big| \alpha \sim N(X\beta + Z\alpha, \sigma^2 I_{JN})$$

and

$$Y \sim N(X\beta, \Sigma),$$

where $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_N) = \sigma^2 V$, $V = I_{JN} + \theta ZZ' = I_{JN} + \theta 1_{JN}, 1_{JN}$ is a matrix of ones.

Since $Y_i \sim N(X_i\beta, \sigma^2 V_i(\theta))$, $i = 1, 2, \ldots, N$ and independent, the non-constant part of the full log-likelihood is given by

$$\ell_F = -\frac{NJ}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^{N} \log \left| V_i(\theta) \right| - \frac{1}{2\sigma^2} \sum_{i=1}^{N} r_i^T V_i^{-1}(\theta) r_i, \qquad (7.3)$$

where $r_i = y_i - X_i\beta$. In order to save space $V_i(\theta)$ is written as V_i. Since the ML estimates are biased downward, our estimates of θ, σ^2, and β are based on the restricted log-likelihood given below

$$\ell_R(\beta, \theta, \sigma | y) = -\frac{[NJ - (p+1)]}{2} \log \sigma^2 - \frac{1}{2} \log \left| \sum_{i=1}^{N} X_i^T V_i^{-1} X_i \right|$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{N} r_i^T V_i^{-1} r_i - \frac{1}{2} \sum_{i=1}^{N} \log \left| V_i \right|. \qquad (7.4)$$

The RML estimate of σ^2 is given by

$$\hat{\sigma}^2_{RML} = \frac{\sum_{i=1}^{N} r_i^T V_i^{-1} r_i}{NJ-(p+1)}. \qquad (7.5)$$

Substituting (7.5) in equation (7.4), we get the restricted profile log-likelihood of β and θ as follows:

$$p_R(\beta, \theta | y) = -\frac{NJ - (p+1)}{2} \log \sum_{i=1}^{N} r_i^T V_i^{-1} r_i - \frac{1}{2} \log \left| \sum_{i=1}^{N} X_i^T V_i^{-1} X_i \right|$$

$$- \frac{1}{2} \sum_{i=1}^{N} \log \left| V_i \right|. \qquad (7.6)$$

The derivatives of the restricted profile log-likelihood with respect to β and θ, after simplification, are given as

$$\frac{\partial p_R(\beta,\theta|y)}{\partial\beta} = [NJ-(p+1)]\,\frac{H_2(V^{-1})}{H_1(V^{-1})}. \tag{7.7}$$

$$\frac{\partial p_R(\beta,\theta|y)}{\partial\theta} = \left[\frac{NJ-(p+1)}{2(J\theta+1)^2}\right]\frac{H_1(\perp_J)}{H_1(V^{-1})} + \frac{1}{2(J\theta+1)^2}\,\text{trace}\,((H(V^{-1}))^{-1}H(\perp_J)) - \frac{NJ}{2(J\theta+1)}. \tag{7.8}$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T\partial\beta} = [NJ-(p+1)]\left[\frac{H(V^{-1})}{H_1(V^{-1})} + 2\,\frac{H_2(V^{-1})\left(H_2(V^{-1})\right)^T}{(H_1(V^{-1}))^2}\right]. \tag{7.9}$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta^2} = \frac{NJ-(p+1)}{2(J\theta+1)^3}\left[-\frac{2JH_1(\perp_J)}{H_1(V^{-1})} + \frac{1}{(J\theta+1)}\left[\frac{H_1(\perp_J)}{H_1(V^{-1})}\right]^2\right] + \frac{1}{2(J\theta+1)^4}\,\text{trace}\left((H(V^{-1}))^{-1}H(\perp_J)\times(H(V^{-1}))^{-1}H(I_J)\right) - \frac{J}{(J\theta+1)^3}\,\text{trace}\left((H(V^{-1}))^{-1}H(\perp_J)\right) + \frac{N}{2}\left(\frac{J}{J\theta+1}\right)^2. \tag{7.10}$$

$$\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta\partial\beta} = \frac{NJ-(p+1)}{(J\theta+1)^2}\left[-\frac{H_2(\perp_J)}{H_1(V^{-1})} + \frac{H_2(V^{-1})H_1(\perp_J)}{(H_1(V^{-1}))^2}\right], \tag{7.11}$$

where $V_i(\theta) = V$ has $1+\theta$ in the diagonal and θ off diagonal, $|V_i| = J\theta + 1$, $V_i^{-1} = [v_{i,jk}]$, $v_{i,jk} = \frac{-\theta}{1+J\theta}$, $j \neq k$, $v_{i,jj} = 1 - \frac{\theta}{1+J\theta}$, $\perp_J$, a matrix of ones, $A = V_i^{-1}\frac{\partial V_i}{\partial\theta}V_i^{-1} = \frac{\perp_J}{(J\theta+1)^2}$,

$V_i^{-1}\frac{\partial V_i}{\partial\theta} = \frac{\perp_J}{J\theta+1}$ and $H(S) = \sum_{i=1}^{N} X_i^T S X_i$, $H_1(S) = \sum_{i=1}^{N} r_i^T S r_i$, and $H_2(S) = \sum_{i=1}^{N} X_i^T S r_i$.

All the equations from 6.3-6.11, in section 6, are special cases of equations 7.3-7.11, in section 7, repsectively. Since

$\frac{H_1(V^{-1})}{\sigma^2} \sim \chi^2_{JN}$ and $H_2(V^{-1}) \sim N\left((0_{(p+1)\times 1}, \sigma^2 H(V^{-1})\right)$,

$$E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T\partial\beta}\right) = [NJ - (p+1)]\left[-H(V^{-1})E\left(\frac{1}{H_1(V^{-1})}\right) + 2E\left(\frac{H_2(V^{-1})H_2(V^{-1})^T}{[H_1(V^{-1})]^2}\right)\right]$$

$$\doteq [NJ - (p+1)]\left[\frac{-H(V^{-1})}{(NJ-2)\sigma^2} + \frac{2\sigma^2 H(V^{-1})}{\sigma^4(JN)^2}\right]$$

$$= -\frac{[NJ - (p+1)](J^2N^2-2JN+4)}{(JN)^2(JN-2)}\frac{H(V^{-1})}{\sigma^2}. \qquad (7.12)$$

Since

$$\frac{H_1(\perp_J)}{J(J\theta+1)\sigma^2} \sim \chi^2_N, \quad \frac{H_1(V^{-1})}{\sigma^2} \sim \chi^2_{JN}, \quad \frac{1}{J(J\theta+1)}\frac{H_1(\perp_J)}{H_1(V^{-1})} \sim \text{beta}(N/2, (J-1)N/2)$$

(see appendix II),

$$E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta^2}\right) = \frac{NJ - (p+1)}{2(J\theta+1)^3}\left[-2JE\left(\frac{H_1(\perp_J)}{H_1(V^{-1})}\right) + \frac{1}{J\theta+1}E\left(\frac{H_1(\perp_J)}{H_1(V^{-1})}\right)^2\right]$$

$$+ \frac{1}{2(J\theta+1)^4}\text{trace}\left((H(V^{-1}))^{-1}H(\perp_J)\times(H(V^{-1}))^{-1}H(\perp_J)\right)$$

$$- \frac{J}{(J\theta+1)^3}\text{trace}\left((H(V^{-1}))^{-1}H(\perp_J)\right) + \frac{N}{2}\left(\frac{J}{J\theta+1}\right)^2$$

$$= \frac{[NJ - (p+1)]J}{2(J\theta+1)^2}\left[-2 + \frac{N+2}{NJ+2}\right] + \frac{N}{2}\left(\frac{J}{J\theta+1}\right)^2 + k$$

$$= \frac{J}{2(J\theta+1)^2}\left[p + 1 - \frac{[NJ - (p+1)][NJ - N]}{NJ+2}\right] + k,$$

where $k = \frac{1}{2(J\theta+1)^4}\text{trace}\left([H(V^{-1})]^{-1}H(\perp_J)\times[H(V^{-1})]^{-1}H(\perp_J)\right)$

$$- \frac{J}{(J\theta+1)^3}\text{trace}\left((H(V^{-1}))^{-1}H(\perp_J)\right). \qquad (7.13)$$

Notice that (6.13) is a special case of (7.13) with J = 2, p = 1.

$$E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta\partial\beta}\right) = \frac{[NJ - (p+1)]}{(J\theta + 1)^2}\left[-E\,\frac{H_2(\perp_J)}{H_1(V^{-1})} + E\left(\frac{H_2(V^{-1})H_1(\perp_J)}{(H_1(V^{-1}))^2}\right)\right]$$

$$\doteq \frac{[NJ - (p+1)]}{(J\theta + 1)^2}\left[0_{J\times 1} + 0_{J\times 1}\right], \text{ see Appendix III,}$$

$$= 0_{J\times 1}. \qquad (7.14)$$

Therefore, the information matrix I of $\hat{\theta}$, $\hat{\beta}$ is of the form

$$I = -\left[\begin{array}{c|c} E\left(\dfrac{\partial^2 p_R(\beta,\theta y)}{\partial\theta^2}\right) & 0_{1\times(p+1)} \\ \hline 0_{(p+1)\times 1} & E\left(\dfrac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T\partial\beta}\right) \end{array}\right], \text{ where } E\left(\frac{\partial^2 p_R(\beta,\theta|y)}{\partial\theta^2}\right) \text{ and}$$

$E\left(\dfrac{\partial^2 p_R(\beta,\theta|y)}{\partial\beta^T\partial\beta}\right)$ are given by (7.13) and (7.12). Variance-covariance matrix of $\hat{\theta}$ and $\hat{\beta}$ is given by the inverse of the information matrix I given above.

8. CONCLUSIONS AND RECOMMENDATIONS

In this project multivariate extensions of the unconditional model with covariates are studied and applied on the extended Ranch Hand data set with 235 subjects. Residual and quantile plots, showing the fit of these extended unconditional models are provided. Tables containing the point and interval estimates of half-life are given and some comparisons based on the widths of these intervals are made.

In addition, the means and variances of the estimators obtained by Gupta (1992) have been updated by using second order approximation.

Considering the fit of the models and the widths of the confidence intervals, it is hoped that this work will be a significant contribution to the Air Force Health Study.

Due to the shortage of time, Michalek et. al.'s model, which was simplified and extended by Gupta (1990), could not be applied on new data set of 235 subjects. This and other issues will be considered in a future study if funding is available.

ACKNOWLEDGEMENTS

I would like to thank the Armstrong Laboratory and the Air Force Office of Scientific Research for sponsoring this research. I would also like to thank the RDL for their efficiency in administrative aspect of the program.

My special thanks are due to Dr. Joel Michalek for providing me the background and some insight into this problem. I am also very thankful to Mr. Thomas White for generous computational help.

REFERENCES

Elandt-Johnson, R. C. and Johnson, N. L. (1980). Survival Models and Data Analysis. John Wiley and Sons.

Graybill, F. A. (1976). Theory and Application of the Linear Model. Duxbury Press.

Gupta, P. L. (1990). Dioxin Half-life Estimation in Veterans of Project Ranch Hand, UES Report.

_____ (1991). An investigation of Dioxin Half-life Estimation in Veterans of Project Ranch Hand. UES Mini Grant Report.

_____ (1992). Estimation of Dioxin Half-life in the Air Force Health Study. RDL Report.

Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. Journal of the American Statistical Association, 83, 1014-1022.

Michalek, J. E., Mihalko, D., White, T., and Patterson, Jr., D. G. (1989). Maximum Likelihood Estimation of Half-Life Based on Two Measurements per Subject (unpublished manuscript).

Michalek, J. E., Tripathi, R. C., Caudill, S. P. and Pirkle, J. L. (1992). Investigation of TCDD Half-Life Heterogeneity in Veterans of Operation Ranch Hand. Journal of Toxicology and Environmental Health, 35, 29-38.

Phillips, D. L. (1989). Propogation of Error and Bias in Half-Life Estimates Based on Two Measurements. Archives of Environmental Toxicology, 18, 508-514.

Pirkle, J. L., Wolfe, W. H., Patterson, D. C. Needham, L. L., Michalek, J. E., Miner, J. C., Peterson, M. R., and Phillips, D. G. (1989). Estimates of the Half-Life of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin in Vietnam Veterans of Operation Ranch Hand. Journal of Toxicology and Environmental Health, 27, 165-171.

Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology. Brooks/Cole.

Tripathi, R. C. (1989). An Investigation of Dioxin Half-Life Heterogeneity in Humans Based on Two Measurements per Subject. UES Report.

_____ (1990). In Investigation of Dioxin Half-Life Estimation in Humans Based on Two or More Measurements per Subject. UES Report.

Wolfe, W. H., Michalek, J. E., Miner, J. C., Peterson, M. R., Pirkle, J. L., Patterson Jr., D. G., and Needham, L. L. (1988). Serum 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin levels in Air Force Health Study

Participants-Preliminary Report. Morbidity and Mortality Weekly Report, 37, 309-311.

SAS Institute, Inc. (1990). SAS/STAT Users Guide, Version 6, 4th ed., Vol. 2, Cary, N. C.: SAS Institute, Inc.

APPENDIX I

We will follow the following convention for vector derivatives: if x is a scalar, Y is a $p \times 1$ vector, Z is a $q \times 1$ vector, and A is a $p \times q$ matrix, then $\frac{\partial x}{\partial Y}$ and $\left(\frac{\partial x}{\partial Y^T}\right)^T$ are $p \times 1$ vectors and $\frac{\partial Y}{\partial Z^T}$, $\frac{\partial A}{\partial x}$, $\frac{\partial^2 x}{\partial Z^T \partial Y}$ and $\frac{\partial x}{\partial A}$ are all $p \times q$ matrices. The derivatives of $r_i^T V_i^{-1} r_i$, $\log\left|V_i\right|$, and $\log\left|\sum_{i=1}^{N} X_i^T V_i^T X_i\right|$, which appear in the profile likelihood, are given below.

1. $\frac{\partial\left|V_i\right|}{\partial\theta_j} = \left|V_i\right| \operatorname{trace}\left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j}\right)$, V_i is any square matrix.

2. $\frac{\partial(V_i^{-1})}{\partial\theta_j} = - V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1}$, V_i is any square matrix.

3. $\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial r_i} = \left(V_i^{-1} + (V_i^{-1})^T\right) r_i = 2V_i^{-1} r_i$, V_i^{-1} symmetric.

4. $\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial r_i^T} = \left[\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial r_i}\right]^T$

5. $\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial\beta} = \frac{\partial r_i}{\partial\beta} \left(\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial r_i}\right)$

$= - 2X_i^T V_i^{-1} r_i$, when V_i^{-1} is symmetric.

6. $\frac{\partial(r_i^T V_i^{-1} r_i)}{\partial\theta_j} = - r_i^T V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} r_i = - r_i^T A_{ij} r_i$.

7. $\frac{\partial^2 r_i^T V_i^{-1} r_i}{\partial\beta^T \partial\beta} = \frac{\partial}{\partial\beta^T} (- 2X_i^T V_i^{-1} r_i)$

$= 2X_i^T V_i^{-1} X_i$.

8. $$\frac{\partial^2 r_i^T V_i^{-1} r_i}{\partial\theta_j \partial\beta} = \frac{\partial}{\partial\theta_j}\left(- X_i^T (V_i^{-1} + (V_i^{-1})^T) r_i\right),$$

$$= - X_i^T \left(\frac{\partial V_i^{-1}}{\partial\theta_j} + \frac{\partial (V_i^{-1})^T}{\partial\theta_j}\right) r_i, \quad V_i^{-1} \text{ is not symmetric}$$

$$= - X_i^T \left(- V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} - \left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1}\right)^T\right) r_i$$

$$= X_i^T \left(A_{ij} + A_{ij}^T\right) r_i, \text{ if } V_i^{-1} \text{ is not symmetric.}$$

If V_i^{-1} is symmetric, then

$$\frac{\partial^2 r_i^T V_i^{-1} r_i}{\partial\theta_j \partial\beta} = 2 X_i^T A_{ij} r_i.$$

9. $$\frac{\partial^2 r_i^T V_i^{-1} r_i}{\partial\theta_k \partial\theta_j} = \frac{\partial}{\partial\theta_k}\left(\frac{\partial}{\partial\theta_j} r_i^T V_i^{-1} r_i\right)$$

$$= \frac{\partial}{\partial\theta_k}\left(- r_i^T V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} r_i\right)$$

$$= r_i^T \left(V_i^{-1} \frac{\partial V_i}{\partial\theta_k} V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} - V_i^{-1} \frac{\partial^2 V_i}{\partial\theta_k \partial\theta_j} V_i^{-1} + V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} \frac{\partial V_i}{\partial\theta_k} V_i^{-1}\right) r_i$$

$$= r_i^T V_i^{-1} \left(\frac{\partial V_i}{\partial\theta_k} V_i^{-1} \frac{\partial V_i}{\partial\theta_j} - \frac{\partial^2 V_i}{\partial\theta_k \partial\theta_j} + \frac{\partial V_i}{\partial\theta_j} V_i^{-1} \frac{\partial V_i}{\partial\theta_k}\right) V_i^{-1} r_i$$

10. $$\frac{\partial \log|V_i|}{\partial\theta_j} = \frac{1}{|V_i|} \frac{\partial |V_i|}{\partial\theta_j} = \frac{|V_i|}{|V_i|} \operatorname{trace}\left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j}\right) = \operatorname{trace}\left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j}\right)$$

11. $$\frac{\partial^2 \log|V_i|}{\partial\theta_k \partial\theta_j} = \frac{\partial}{\partial\theta_k} \operatorname{trace}\left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j}\right)$$

$$= \operatorname{trace}\left(\frac{\partial}{\partial\theta_k}\left(V_i^{-1} \frac{\partial V_i}{\partial\theta_j}\right)\right)$$

$$= \operatorname{trace}\left(\frac{\partial V_i^{-1}}{\partial\theta_k} \frac{\partial V_i}{\partial\theta_j} + V_i^{-1} \frac{\partial^2 V_i}{\partial\theta_k \partial\theta_j}\right)$$

$$= \operatorname{trace}\left(-V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} \frac{\partial V_i}{\partial\theta_k} + V_i^{-1} \frac{\partial^2 V_i}{\partial\theta_k \partial\theta_j}\right)$$

12. $$\frac{\partial}{\partial\theta_j}\log\Big|\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\Big| = \operatorname{trace}\Big(\Big(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\Big)^{-1} \frac{\partial}{\partial\theta_j} \sum_{i=1}^{N} X_i^T V_i^{-1} X_i\Big)$$

$$= \operatorname{trace}\Big(\Big(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\Big)^{-1} \sum_{i=1}^{N} -X_i^T V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1} X_i\Big)$$

$$= -\sum_{i=1}^{N} \operatorname{trace}\big(H^{-1} X_i^T A_{ij} X_i\big), \text{ where } H = \sum_{i=1}^{N} X_i^T V_i^{-1} X_i.$$

13. $$\frac{\partial^2 \log\Big|\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\Big|}{\partial\theta_k \partial\theta_j} = \frac{\partial}{\partial\theta_k}\Big(-\sum_{i=1}^{N} \operatorname{trace}\big(H^{-1} X_i^T A_{ij} X_i\big)\Big)$$

$$= -\sum_{i=1}^{N} \operatorname{trace}\Big(-H^{-1}\frac{\partial H}{\partial\theta_k} H^{-1} X_i^T A_{ij} X_i + H^{-1} X_i^T \frac{\partial A_{ij}}{\partial\theta_k} X_i\Big)$$

$$= -\sum_{i=1}^{N} \operatorname{trace}\Big(-H^{-1}\Big(\sum_{i=1}^{N} -X_i^T A_{ik} X_i\Big) H^{-1} X_i^T A_{ij} X_i + H^{-1} X_i^T \frac{\partial A_{ij}}{\partial\theta_k} X_i\Big)$$

$$= -\operatorname{trace}\Big(H^{-1}\Big(\sum_{i=1}^{N} X_i^T A_{ik} X_i\Big) H^{-1}\Big(\sum_{i=1}^{N} X_i^T A_{ij} X_i\Big) + H^{-1} \sum_{i=1}^{N} X_i^T \frac{\partial A_{ij}}{\partial\theta_k} X_i\Big)$$

where

$$A_{ij} = V_i^{-1} \frac{\partial V_i}{\partial\theta_j} V_i^{-1},$$

$$\frac{\partial A_{ij}}{\partial\theta_k} = -V_i^{-1}\Big(\frac{\partial V_i}{\partial\theta_k} V_i^{-1} \frac{\partial V_i}{\partial\theta_j} - \frac{\partial^2 V_i}{\partial\theta_k \partial\theta_j} + \frac{\partial V_i}{\partial\theta_j} V_i^{-1} \frac{\partial V_i}{\partial\theta_k}\Big) V_i^{-1},$$

and $H = \sum_{i=1}^{N} X_i^T V_i^{-1} X_i$.

The above reesults are derived in more general setting than required for our work. For our calculations $\theta_k = \theta_j = \theta$.

APPENDIX II

Here we will obtain the distributions of $\frac{H_1(V^{-1})}{\sigma^2}$, $\frac{H_1(\perp_J)}{\sigma^2 J(J\theta+1)}$,

and $\dfrac{\dfrac{H_1(\underline{\underline{1}}_J)}{\sigma^2 J(J\theta+1)}}{\dfrac{H_1(V^{-1})}{\sigma^2}}$ with the help of the following theorem:

If $Y \sim N_p(\mu,\Sigma)$, then the quadratic form $Y'AY \sim \chi^2(r,\delta)$, where $\delta = \mu'A\mu$ and the rank(A) = r, iff $A\Sigma A = A$ or $A\Sigma$ is idempotent (see Timm (1975), page 132).

Consider

1. $\dfrac{H_1(V^{-1})}{\sigma^2} = \sum_{i=1}^{N} \dfrac{(Y_i - X_i\beta)^T V^{-1}(Y_i - X_i\beta)}{\sigma^2} \sim \chi^2_{JN}$, where

$Y_i - X_i\beta \sim N(0_{J\times 1}, \sigma^2 V)$. In case J = 2, $\dfrac{H_1(V^{-1})}{\sigma^2} \sim \chi^2_{2N}$.

2. $\dfrac{H_1(\underline{\underline{1}}_J)}{J(J\theta+1)\sigma^2} \sim \chi^2_N$, where $\underline{\underline{1}}_j$ is a matrix of ones. For J = 2,

$\dfrac{H_1(\underline{\underline{1}}_2)}{2(2\theta+1)\sigma^2} \sim \chi^2_N$.

3. $\dfrac{\dfrac{H_1(\underline{\underline{1}}_J)}{J(J\theta+1)\sigma^2}}{\dfrac{H_1(V^{-1})}{\sigma^2}} = \dfrac{Q_e}{Q_t} = \dfrac{Q_e}{Q_e + Q_t - Q_e}$,

where $Q_t - Q_e = \dfrac{1}{\sigma^2}\sum_{i=1}^{N}(Y_i - X_i\beta)^T\left(V^{-1} - \dfrac{\underline{\underline{1}}_J}{J(J\theta+1)}\right)(Y_i - X_i\beta)$

$= \dfrac{1}{\sigma^2 J}\sum_{i=1}^{N}(Y_i - X_i\beta)^T W(Y_i - X_i\beta)$,

where $W = [w_{ik}]$ is a J × J matrix with $w_{ik} = -1$, $i \neq k$, and $W_{ik} = J - 1$ for i = k. By using the above theorem again $Q_t - Q_e \sim \chi^2_{(J-1)N}$. Also $Q_t - Q_e$ and Q_e are independent χ^2 variables.

Therefore

$$\frac{H_1(\perp_J)}{J(J\theta+1)H_1(V^{-1})} = \frac{Q_e}{Q_t} = \frac{Q_e}{Q_e + Q_t - Q_e} = \frac{\chi^2_N}{\chi^2_N + \chi^2_{(J-1)N}} \sim \text{Beta}\left(\frac{N}{2}, \frac{(J-1)N}{2}\right) = \beta(a,b), \text{ say.}$$

Consequently

$$E\left(\frac{Q_e}{Q_t}\right) = \frac{a}{a+b} = \frac{1}{J} = \frac{1}{2} \text{ for } J = 2.$$

$$\text{Var}\left(\frac{Q_e}{Q_t}\right) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{2(J-1)}{J^2(JN+2)} = \frac{1}{4(N+1)}$$

and

$$E\left[\left(\frac{Q_e}{Q_t}\right)^2\right] = \text{Var}\left(\frac{Q_e}{Q_t}\right) + \left[E\left(\frac{Q_e}{Q_t}\right)\right]^2 = \frac{N+2}{J(JN+2)}.$$

Finally, we would like to find the $E(H_2(V^{-1})H_1(\perp_J))$, where

$$H_2(V^{-1}) = \sum_{i=1}^{N} X_i^T V^{-1}(Y_i - X_i\beta)$$

$$= \sum_{i=1}^{N}\left[\sum_{\ell=1}^{J} x_{i,\ell r}a_{i\ell} - \frac{\theta}{1+J\theta}\left(\sum_{\ell=1}^{J} x_{i,\ell r}\right)\left(\sum_{k=1}^{J} a_{ik}\right)\right]_{(p+1)\times 1},$$

where $r = 0, 1, 2, \ldots, p$, and $x_{i,\ell 0} = 1$, $\ell = 1, 2, \ldots, J$,

and

$$H_1(\perp_J) = \sum_{i=1}^{N}(Y_i - X_i\beta)^T(\perp_J)(Y_i - X_i\beta) = \sum_{i=1}^{N}\left(\sum_{k=1}^{J} a_{ik}\right)^2,$$

where $a_{ik} = y_{ik} - \sum_{m=0}^{p}\beta_m x_{i,km}$, $x_{i,k0} = 1$, $k = 1, 2, \ldots, J$.

$$E(H_2(V^{-1})(H_1(\perp_J))$$

$$= \sum_{i=1}^{N}\sum_{s=1}^{N}\left[\sum_{\ell=1}^{J} x_{i,\ell r}\, E\left(a_{i\ell}\left(\sum_{k=1}^{J} a_{sk}\right)^2\right) - \frac{\theta}{1+J\theta}\left(\sum_{\ell=1}^{J} x_{i,\ell r}\right)E\left(\left(\sum_{k=1}^{J} a_{ik}\right)\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)\right]_{(p+1)\times 1}$$

$$= [0]_{(p+1)\times 1},$$

since

$$E\left(\left(\sum_{k=1}^{J} a_{ik}\right)\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)$$

$= E\left(\left(\sum_{k=1}^{J} a_{ik}\right)E\left(\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)\right)$, $i \neq s$, since $\sum_{k=1}^{J} a_{ik}$ and $\left(\sum_{k=1}^{J} a_{sk}\right)^2$ are independent

$= 0$, since $E\left(\sum_{k=1}^{J} a_{ik}\right) = 0$

$E\left(\left(\sum_{k=1}^{J} a_{ik}\right)^3\right)$, $i = s$

$= 0$, odd central moments of a normal variable are zero.

Also $E\left(a_{i\ell}\left(\sum_{k=1}^{J} a_{sk}\right)^2\right) = 0$, $\ell = 1, 2, \ldots, J$, since

$$0 = E\left(\left(\sum_{\ell=1}^{J} a_{i\ell}\right)\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)$$

$$= \sum_{\ell=1}^{J} E\left(\left(a_{i\ell}\right)\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)$$

$= J\, E\left(\left(a_{i\ell}\right)\left(\sum_{k=1}^{J} a_{sk}\right)^2\right)$, since $a_{i1}, a_{i2}, \ldots, a_{ij}$ are identically distributed.

Appendix III

In this section we will show the following by using the delta method:

a. $E\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right) \doteq 0$

$$E\left(\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)^T\right) \doteq \frac{\sigma^2}{(JN)^2} H(V^{-1}).$$

b. $E\left(\frac{H_2(\perp_J)}{\sigma^{-2}H_1(V^{-1})}\right) \doteq 0$

c. $E\left(\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)\left(\frac{\sigma^{-2}H_1(\perp_J)}{J(J\theta + 1)}/\sigma^{-2}H_1(V^{-1})\right)\right) = 0$

$$\therefore \text{Cov}\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})} \cdot \frac{\sigma^{-2}H_1(\quad)}{J(J\theta+1)\sigma^{-2}H_1(V^{-1})}\right) = 0$$

Here $H(S) = \sum_{i=1}^{N} X_i^T S X_i$, $H_1(S) = \sum_{i=1}^{N} r_i^T S r_i$, and $H_2(S) = \sum_{i=1}^{N} X_i^T S r_i$.

Let $g(W) = \frac{W_1}{W_2}$, where $E(W_1) = \xi_1$, $E(W_2) = \xi_2$.

Then

$$E(g(W)) \doteq g(\xi) + \frac{1}{2}\left[\frac{\partial^2 g}{\partial \xi_1^2}\text{Var}(W_1) + \frac{\partial^2 g}{\partial \xi_2^2}\text{Var}(W_2) + 2\frac{\partial^2 g}{\partial \xi_1 \partial \xi_2}\text{Cov}(W_1, W_2)\right] \quad (1)$$

where $\frac{\partial g(W)}{\partial W_1} = \frac{1}{W_2}$, $\frac{\partial g(W)}{\partial W_2} = -\frac{W_1}{W_2^2}$

$$\frac{\partial^2 g(W)}{\partial W_1^2} = 0, \quad \frac{\partial^2 g(W)}{\partial W_2^2} = \frac{2W_1}{W_2^3}$$

$$\frac{\partial^2 g}{\partial W_1 \partial W_2} = -\frac{1}{W_2^2}.$$

Then (1) becomes

$$E(g(W)) \doteq \frac{\xi_1}{\xi_2} + \frac{1}{2}\left[\frac{2\xi_1}{\xi_2^3}\text{Var}(W_2) - \frac{2}{\xi_2^2}\text{Cov}(W_1, W_2)\right] \quad (2)$$

and

$$\text{Var}(g(W)) = E((g(W) - g(\xi))^2)$$

$$\doteq \left(\frac{\partial g}{\partial \xi_1}\right)^2 \text{Var}(W_1) + \left(\frac{\partial g}{\partial \xi_2}\right)^2 \text{Var}(W_2) + 2\left(\frac{\partial g}{\partial \xi_1}\right)\left(\frac{\partial g}{\partial \xi_2}\right)\text{Cov}(W_1, W_2)$$

$$= \left(\frac{1}{\xi_2}\right)^2 \text{Var}(W_1) + \left(\frac{-\xi_1}{\xi_2}\right)^2 \text{Var}(W_2) - \frac{2\xi_1}{\xi_2^3}\text{Cov}(W_1, W_2) \quad (3)$$

$\text{Cov}(r_i, r_j^T Q r_j) = 0$, where r_i, r_j are independent normal with mean zero and Q is any $N \times N$ matrix. (4)

a. Let $g(W) = \frac{W_1}{W_2}$, where $W_1 = H_2(V^{-1}) \sim N\left(0_{(p+1)\times 1}, \sigma^2 H(V^{-1})\right)$, and

$W_2 = \sigma^{-2}H_1(V^{-1}) \sim \chi^2_{JN}$. Therefore $\xi_1 = E(W_1) = 0_{(p+1)\times 1}$,

$\text{Var}(W_1) = \sigma^2 H(V^{-1})$, $\xi_2 = E(W_2) = JN$, $\text{Var}(W_2) = 2JN$.

From equation (2), we have

$$E\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1}X)}\right) \doteq -\frac{1}{(JN)^2}\,\mathrm{Cov}(H_2(V^{-1}),\ \sigma^{-2}H_1(V^{-1}))$$

$$= -\frac{1}{(JN)^2}\,\mathrm{Cov}\Big(\sum_{i=1}^{N} X_i^T V^{-1} r_i, \sigma^{-2}\sum_{j=1}^{N} r_j^T V^{-1} r_j\Big)$$

$$= -\frac{\sigma^{-2}}{(JN)^2}\sum_{i=1}^{N}\sum_{j=1}^{N} X_i^T V^{-1}\mathrm{Cov}\left(r_i, r_j^T V^{-1} r_j\right)$$

$$= 0_{(p+1)\times 1}, \text{ see equation (4).} \tag{5}$$

From equation (3) we have

$$\mathrm{Var}\left(\frac{H_2(V^{-1})}{\sigma^2 H_1(V^{-1})}\right) \doteq \frac{\sigma^2}{(JN)^2}\,H(V^{-1}) \tag{6}$$

Therefore

$$E\left(\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)^T\right) = \mathrm{Var}\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right) + \left(E\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)\right)^2$$

$$= \frac{\sigma^2}{(JN)^2}\,H(V^{-1}) \tag{7}$$

b. Let $W_1 = H_2(\perp_J)$, $W_2 = \sigma^{-2}H_1(V^{-1})$. Then $\xi_1 = E(W_1) = 0_{J\times 1}$, $\xi_2 = E(W_2) = JN$. Then from equation (2), we have

$$E\left(\frac{H_2(\perp J)}{\sigma^{-2}H_1(V^{-1})}\right) \doteq -\frac{1}{(JN)^2}\,\mathrm{Cov}(H_2(\perp_J),\ \sigma^2 H_1(V^{-1}))$$

$$= -\frac{1}{(JN)^2}\,\mathrm{Cov}\Big(\sum_{i=1}^{N} X_i^T \perp_J r_i, \sigma^{-2}\sum_{j=1}^{N} r_j^T V^{-1} r_j\Big)$$

$$-\frac{\sigma^{-2}}{(JN)^2}\sum_{i=1}^{N}\sum_{J=1}^{N} X_i^T \perp_J \mathrm{Cov}(r_i, r_j^T r^{-1} r_j)$$

$$= 0. \tag{8}$$

c. Notice that $H_2(V^{-1}) = W_1 \sim N\left(0_{(p+1)\times 1},\ \sigma^2 H(r^{-1})\right)$, $\sigma^{-2}H_1(V^{-1})$

$$= W_2 \sim \chi^2 JN,$$

and $\dfrac{\sigma^{-2}H_1(\perp_J)}{J(J\theta+1)} = W_3 \sim \chi^2_N$, see appendix II. Then

$E(W_1) = \xi_1 = 0_{(p+1)\times 1}$, $E(W_2) = \xi_2 = JN$, $E(W_3) = N$

$$\mathrm{Var}(W_1) = \sigma^2 H(V^{-1}),\ \mathrm{Var}(W_2) = 2JN,\ \mathrm{Var}(W_3) = 2N.$$

Let $g(W) = \dfrac{W_1 W_3}{W_2^2}$, then by the delta method

$$E\big((g(W)\big) \doteq$$

$$\frac{\xi_1\xi_3}{\xi_2^2} + \frac{1}{2}\left[\sum_{i=1}^{3}\frac{\partial^2 g}{\partial \xi_i^2}\,\mathrm{Var}(W_i) + \sum_{\substack{i<j\\ i,j=1,2,3}}\sum \frac{\partial^2 g}{\partial \xi_i \partial \xi_j}\,\mathrm{Cov}(W_i,W_j)\right], \qquad (9)$$

where

$$\frac{\partial g}{\partial W_1} = \frac{W_3}{W_2^2},\ \frac{\partial g}{\partial W_2} = \frac{-2W_1W_3}{W_2^3},\ \frac{\partial g}{\partial W_3} = \frac{W_1}{W_2^2}$$

$$\frac{\partial^2 g}{\partial W_1^2} = 0,\ \frac{\partial^2 g}{\partial W_2^2} = \frac{6W_1W_3}{W_2^4},\ \frac{\partial^2 g}{\partial W_3^2} = 0$$

$$\frac{\partial^2 g}{\partial W_2 W_1} = -\frac{2W_3}{W_2^3},\ \frac{\partial^2 g}{\partial W_3 \partial W_1} = \frac{1}{W_2^2},\ \frac{\partial^2 g}{\partial W_2 \partial W_3} = \frac{-2W_1}{W_2^3}$$

Therefore (9) becomes

$$E\left(\frac{W_1W_3}{W_2^2}\right) = E\left(\left(\frac{H_2(V^{-1})}{\sigma^{-2}H_1(V^{-1})}\right)\left(\frac{\sigma^{-2}H_1(\underline{\ }_J)}{J(J\theta+1)\sigma^{-2}H_1(V^{-1})}\right)\right) \doteq \frac{2}{2}\left[-\frac{2\xi_3}{\xi_2^3}\,(W_1,W_2) + \frac{1}{\xi_2^2}\,\mathrm{Cov}(W_1,W_3)\right],$$

where $\mathrm{Cov}(W_1,W_2) = \mathrm{Cov}(H_2(V^{-1}),\ \sigma^{-2}H_1(V^{-1})) = 0$, from equation (4)

and $\mathrm{Cov}(W_1,W_3) = \mathrm{Cov}\left(H_2(V^{-1}),\ \dfrac{\sigma^{-2}H_1(\underline{\ }_J)}{J(J\theta+1)}\right)$

$$= \frac{\sigma^{-2}}{J(J\theta+1)}\,\mathrm{Cov}\Big(\sum_{i=1}^{N} X_i^T V^{-1} r_i,\ \sum_{j=1}^{N} r_j^T \underline{\ }_J r_j\Big)$$

$$= \frac{\sigma^{-2}}{J(J\theta+1)}\sum_{j=1}^{N}\sum_{i=1}^{N} X_i^T V^{-1}\mathrm{Cov}(r_i, r_j^T \underline{\ }_J r_j)$$

$$= 0 \quad \text{from equation (4).}$$

Therefore $E\left(\dfrac{W_1W_3}{W_2^2}\right) = E\left(\left(\dfrac{H_2(V^{-1})}{\sigma^2 H_1(V^{-1})}\right)\left(\dfrac{\sigma^{-2}H_1(\underline{\ }_J)}{J(J\theta+1)\sigma^{-2}H_1(V^{-1})}\right)\right) \doteq 0. \qquad (10)$

Exogenous Melatonin Facilitates Daytime Sleep Without Decreasing Post-Sleep Performance

Rod J. Hughes
Research Associate
Department of Psychology

Bowling Green State University
Bowling Green, OH 43403

Final Report for:
Research Initiation Program
Armstrong Laboratory

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

and

Bowling Green State University

April 1994

EXOGENOUS MELATONIN FACILITATES DAYTIME SLEEP WITHOUT DECREASING POST-SLEEP PERFORMANCE

Rod J. Hughes
Research Associate
Department of Psychology
Bowling Green State University

Abstract

The present investigation used a randomized, placebo-controlled, double blind design to assess the hypnotic efficacy of exogenous melatonin. Daytime melatonin administration decreases body temperature, increases fatigue (e.g., French, et al., 1994) and decreases performance (e.g., Lieberman, 1984). The present investigation assessed the efficacy of four doses of melatonin (0 mg-placebo, 1 mg, 10 mg, and 40 mg) given at 1000 hours, on inducing and sustaining a polygraphically recorded daytime nap (4 hr). Various measures of sleep were used to compare the efficacy of each dose. Key among these measures were latency to sleep, sleep efficiency, and sleep architecture. Tympanic temperature was assessed before and after the nap. This study also assessed for potential negative side effects such as carry-over fatigue. This was assessed by a performance assessment battery consisting of various psychomotor and cognitive tasks. It was predicted that melatonin would facilitate daytime sleep without the negative side-effects associated with current benzodiazepines. Preliminary results ($n = 8$) suggest that melatonin can significantly improve sleep for a moderate duration daytime nap. Compared to placebo, melatonin, especially the higher doses, was efficacious in both initiating and sustaining a moderate duration daytime nap. Melatonin also caused dose-dependent decreases in temperature. Further, melatonin yielded no significant decrements in performance 7-9 hr after administration. This investigation shows that melatonin can significantly facilitate daytime sleep without significant side effects.

EXOGENOUS MELATONIN FACILITATES DAYTIME SLEEP WITHOUT DECREASING POST-SLEEP PERFORMANCE

Rod J. Hughes

Introduction

Advances in technology have enabled humans to work and travel across the 24-hr day and to establish work-rest (wake-sleep) behavioral patterns that are independent of the solar light-dark cycle. A price is paid for this freedom. While behavioral patterns can be changed quickly, underlying endogenous physiological patterns cannot (Wever, 1979). The result can be an "out-of-phase" relationship between behavioral rhythms and physiological rhythms. Nighttime performance is degraded because physiological rhythms such as alertness, psychomotor performance, and cognitive performance are not at peak levels. The fatigue associated with trying to work at night is exacerbated by trying to sleep during the day out-of-phase with endogenous sleep rhythms (Naitoh, et al., 1990). This combination often leads to serious negative consequences such as decreased performance, and increases in accidents (e.g., Akerstedt, 1988). Operating under such desynchrony for extended periods can also pose health risks including gastrointestinal disorders and increased risk of cardiovascular disease (Moore-Ede & Richardson, 1985).

To improve sleep during the rest period and enhance alertness during the work period, hypnotic drugs, of which the benzodiazepines are the most popular, are prescribed. These drugs, however, sometimes have serious negative side effects and the sleep that they induce is pharmacological (Pascually, et al., 1992) as opposed to physiological (natural). Aside from affecting the form of sleep, benzodiazepines produce excessive grogginess that often persists beyond the sleep episode. Further, benzodiazepines have the potential to produce anterograde amnesia (French, et al., 1990; Hommer, 1991). Finally, some benzodiazepines like Temazepam have intermediate elimination half-lives (10-12 hours) that could cause accumulation and carry-over effects with chronic administration. A need

for more natural hypnotics exists. The present investigation assessed the hypnotic efficacy of a natural substance that has not shown any of the negative side effects associated with the benzodiazepines. This substance is the pineal hormone melatonin. The present investigation used a placebo-controlled, double blind design to test the effects of exogenous melatonin administration on polygraphically recorded daytime sleep as well as on subjective, behavioral and physiological measures of fatigue. Further, this investigation assessed for potential side effects typically associated with the benzodiazepines such as anterograde amnesia and carry-over fatigue.

ENDOGENOUS MELATONIN

Endogenous melatonin is produced primarily in the pineal gland via the metabolism of tryptophan (Reiter, 1991c). The pineal gland is a tiny organ located near the center of the brain. It receives input from the visual system (see Reiter, 1991b) via a complex series of neural pathways that begins with the photoreceptors of the retina, passes through the superchiasmatic nuclei (SCN) via the retinohypothalamic tract and eventually ends in sympathetic innervation of the pinealocytes (the endocrine elements of the pineal gland). The catecholamine norepinephrine (NE) is the primary neurotransmitter responsible for transmitting the light/dark message from the postganglionic neurons to the pinealocytes. Within the pinealocytes, the amino-acid tryptophan is picked up from the blood stream and converted into melatonin (see Figure 1 for a diagram of the steps involved in this conversion). From the pineal gland melatonin is released directly into the blood stream. Because of its high lipid solubility melatonin carries its message to all cells in the body. In fact, melatonin potentially enters "every part of every cell in the body" (Reiter, 1991a). Endogenous melatonin's passive message of darkness is then interpreted by each organ, ultimately mediating the effects of photoperiod throughout the body (Reiter, 1991b).

The pineal gland produces and releases melatonin primarily at night (Reiter, 1990). This endogenous secretion pattern is not masked by the sleep-wake schedule, however, release of melatonin is inhibited by sufficiently bright light (Lewy, et al., 1980). In humans the nocturnal elevation of endogenous melatonin is associated with the decrease in noctornal body temperature as well as with the onset and duration of nighttime sleep. For example, extending the dark phase of the circadian cycle extends the duration of melatonin secretion as well as the duration of nighttime sleep (Wehr,1991). Suppressing the nocturnal secretion of endogenous melatonin with bright artificial light attenuates the nocturnal decrease in body temperature and increases subjective and objective measures of alertness including performance (Badia 1991; French, 1990).

EXOGENOUS MELATONIN

Exogenous Melatonin and Phase Shifting Circadian Rhythms

Exogenous melatonin (xMT) has dual effects on human physiology. First, the chronic administration of melatonin will shift rhythms according to a phase response curve. Melatonin administered in the evening phase-advances circadian rhythms. Melatonin presented in the morning phase-delays rhythms (Lewy, et al., 1992). However, relative to the phase-shifting effects of bright light, these effects are small and occur only after multiple days of administration.

Clinically, melatonin is useful in phase setting free-running circadian rhythms in blind subjects (Arendt, et al., 1988; Palm et al., 1991; Sack, et al., 1987) as well as in facilitating adjustment to other circadian desynchronies such as jet-lag (Arendt, et al., 1986,1987; Petrie, et al., 1989; Samuel, et al., 1991), and delayed sleep phase syndrome (Dahlitz, et al., 1991).

Aside from its circadian phase setting and phase shifting effects, melatonin also has immediate sedating qualities. Giving exogenous melatonin during the day, when

endogenous levels are very low, decreases body temperature and decreases subjective measures of alertness (e.g., French, Hughes, Whitmore, Neville & Reiter, 1993; Lieberman, et al., 1984). Daytime melatonin administration also decreases psychomotor performance (e.g., Lieberman, et al., 1984). Further, the daytime administration of exogenous melatonin reduces sleep onset latencies (Anton-Tay, et al., 1971; Cramer, et al., 1974; Dollins, et. al., 1993; Hughes, et al., 1994). The present investigation focused on these acute effects of exogenous melatonin administration.

Immediate Physiological Effects of Exogenous Melatonin

Exogenous melatonin administered in the daytime elevates plasma levels of melatonin several fold. In previous research we showed that the administration of 1 mg, 10 mg and 100 mg at 0915 hr elevated plasma levels of melatonin several fold. Figure 2 depicts the mean serum melatonin levels yielded by these doses (Hughes, 1992). Daytime melatonin administration increases prolactin and slightly increases levels of Growth Hormone (Waldhauser, et al., 1987). Large doses slow the frequency of EEG and induce mild sedation (Anton-Tay, et al., 1971). Exogenous melatonin administration also decreases body temperature. In fact, recent research shows that different doses of melatonin given at 0915 hr significantly decreases oral temperature in a dose-dependent manner (French et al., 1993).

Exogenous Melatonin and Fatigue

Evidence supporting the relationship between melatonin and fatigue comes from several sources. Lieberman, et al., (1984) showed that the administration of 80 mg of melatonin hourly for 3 hr produced significant tranquilizing effects. Exogenous melatonin significantly increased subjective measures fatigue and also caused psychomotor performance decrements in the form of longer reaction times (Lieberman, et al., 1984).

Others have reported similar increases in subjective measures of fatigue for varying doses of exogenous melatonin (e.g., Arendt, et al., 1984; Nickelson et al., 1989).

As noted, we have shown that, compared to placebo, 10 mg and 100 mg of melatonin given at 0915 hr yielded dose dependent increases in salivary and serum melatonin levels. Increased melatonin levels caused corresponding and long lasting decreases in oral temperature and subjective measures of alertness (French, et al., 1993).

Exogenous Melatonin and Sleep

There has been some assessment of the hypnotic efficacy of melatonin on daytime sleep. Nickelson et al., (1989) showed that subchronic doses (7 days) of 50 mg at 0900 hr decreased subjective reports of sleep onset latency. Using an insomnia paradigm Waldhauser et al., (1990) reported that large doses of melatonin significantly improved several measures of sleep quality. Melatonin at 2100 hr decreased sleep latency and decreased the number of awakenings. Melatonin also increased sleep efficiency. Anton-Tay et al., (1971) reported that acute intravenous administration of melatonin at 1600 hr improved polygraphically scored sleep latency. Cramer et al., (1974) replicated this effect for the acute oral administration of a 50 mg dose at 1600 hr and at 2130 hr. It is interesting to note that Vollrath and colleagues reported similar sleep onset improvements of non-polygraphically scored sleep for much smaller doses given intranasally. They administered 1.7 mg of MT intranasally between the times of 0900 and 1000 hr (Vollrath, et al., 1981). Using a behavioral measure of sleep (a "dead-man's" switch), Dollins reported melatonin induced reductions in sleep latency and increases in the amount of sleep in a short daytime nap (Dollins, et al., 1993). Previous investigations have primarily tested the effects of melatonin on initiating daytime sleep and on the amount of sleep in a short duration daytime nap. The efficacy of exogenous melatonin on sustaining a long duration daytime sleep episode has yet to be reported.

In pilot research, we administered 0 mg and 100 mg of melatonin at 0900 hr to six subjects who were subsequently restricted to bed from 1000 hr to 1600 hr. Polygraphic recording of their sleep revealed that 100 mg of melatonin shortened subjects sleep latency by an average of 45%, and increased sleep efficiency (see Figure 3). These data also suggest that melatonin increases the percentage of stage 2 sleep and decreases the percentage of stage 3 sleep. The extraordinary amount of time subjects in the placebo condition slept (sleep efficiency = .88) was likely due to the fact that we did not directly control for the amount of sleep subjects received the night prior to the experiment. Nevertheless the results of our pilot study suggest that melatonin will be efficacious in promoting daytime sleep. The present investigation used a placebo controlled, double-blind design to assess the efficacy of exogenous melatonin in initiating and sustaining a long duration daytime sleep episode. We assessed the effects of four doses of melatonin (0 mg, 1 mg, 10 mg, and 40 mg) given at 1000 hr on tympanic temperature, and sleep in a four hr nap. We also assessed subjective mood ratings, and performance both prior to and after the nap.

Method

SUBJECTS

Eight healthy male volunteers between the ages of 18 and 28 participated in the entire study. Subjects participated for a total of four days each separated by a 2-4 day washout period. All subjects underwent an extensive screening process prior to participation. Subjects were screened for chronic smoking, alcohol use, heavy caffeine use and unstable work-rest schedules. Finally, subjects underwent a physical examination to insure that they were in good health at the time of the study.

MATERIALS

Melatonin and Placebo

The drug conditions were 0 mg (placebo), 1 mg, 10 mg and 40 mg of crystalline melatonin (Sigma Chemical, St. Louis, MO). All doses were mixed in gelatin capsules with 400 mg of Psyllium (6.0 grams of pure Psyllium fiber per teaspoon, Colon Pure, General Nutrition Centers). The order of dose administration was counterbalanced across subjects according to a Latin Square design.

Sleep Assessment

Sleep was recorded using the standardized electrode montage: EEG brain sites C3 and O1 referenced to the mastoids (A1,A2), eye movements (EOG) and chin muscle tension (EMG). Sleep was scored into 30 second epochs according to standardized scoring procedures (Rechtscheffen & Kales, 1968).

Performance Assessment

The UTCPAB consists of various psychomotor and cognitive tasks that have proven to show performance decrements associated with fatigue (Gillooly, et al., 1990; Thorne, et al., 1985). The performance tasks assessed various types of performance including *perceptual input*; Procedural Memory Tasks (Basic and Coded), *central processing*; Memory Recall Task, Continuous Recognition Task, and *information integration/manipulation*; Mathematical Processing Task.

Subjective Measures of Fatigue

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

Stanford Sleepiness Scale) have shown to be sensitive to fatigue related to circadian rhythms, sleep deprivation and melatonin administration.

Temperature

Tympanic temperature was assessed with a FirstTemp tympanic temperature probe (Intelligent Medical Systems, Carlsbad, CA).

PROCEDURE

Subjects arrived in the laboratory at 2300 hr the night before the study, at which time they read and signed the informed voluntary consent forms. From 2400 hr until 0700 hr subjects slept in the laboratory. At 0730 hr electrode leads for physiological recording were attached. While in the laboratory subjects sat upright in a comfortable chair in constant, normal, light (100 lux). During the sleep assessment, subjects were restricted to bed and the lights were turned down (20 lux). See Figure 4 for a schematic diagram of the testing day.

The first testing block began at 0900 hr. Testing blocks were given at 0900, 1030, 1630, and 1730 hr. At 1000 hr subjects ingested a pill consisting of 0 mg, 1 mg, 10 mg, or 40 mg of melatonin. Subjects were in bed for calibration of the polygraph by 1145 hr and the beginning of the sleep assessment phase was at 1200 hr. Subjects remained in bed until 1600 hr for a total time in bed of four hours. At 1630 hr subjects took the first of two post sleep performance assessment batteries (1630 and 1730 hr).

Results

MELATONIN AND SLEEP

Differences in sleep were assessed using a one-way, repeated measures, ANOVA. As shown in Figure 5 melatonin significantly shortened sleep latency [$F(3,21) = 4.87$, $p < .05$].

Tukey post hoc tests revealed that, compared to placebo, 10 mg of melatonin significantly reduced sleep onset latency ($p < .05$). Melatonin significantly increased sleep efficiency (Figure 6) [$F(3,21) = 3.27$, $p < .05$]. Post hoc analyses revealed that 10 mg produced significantly more sleep than placebo ($p < .05$). Table 1 contains the average percentage of each sleep stage by dose level. Melatonin significantly increased the percentage of Stage 2 sleep [$F(3,21) = 7.88$, $p < .01$]. Post hoc tests revealed that these increases over placebo (see Figure 7) were statistically significant for 1 mg ($p < .05$) and 10 mg ($p < .05$). Increases in the percentage of Stage 2 sleep were accompanied by significant decreases in the percentage of Stage 3-4 sleep [$F(3,21) = 6.2$, $p < .01$]. As depicted in Figure 8, all doses of melatonin significantly decreased the percentage of Stage 3-4 sleep ($p < .05$).

MELATONIN AND BODY TEMPERATURE

Two way (Drug X Time of Day) repeated measures ANOVAs were performed on the temperature difference scores. Dose-dependent, decreases in temperature (Figure 9) were revealed by a significant main effect for dose [$F(3,21) = 9.48$, $p < .01$]. Compared to placebo, all doses of melatonin significantly reduced temperature ($p < .05$).

MELATONIN AND SUBJECTIVE FATIGUE

As depicted in Figure 10 subjective feelings of fatigue followed body temperature. Thus, small dose-dependent increases in subjective fatigue were reported, especially prior to the nap.

MELATONIN AND PERFORMANCE

Results of 4 (dose) X 3 (time) repeated measures ANOVAs indicated that 1 mg of melatonin significantly improved performance on the Mathematical Processing task. As indicated by a main effect for dose [$F(3,21) = 3.15$, $p < .05$], 1 mg significantly shortened

the average reaction times to correct stimuli (See Figure 11). For most tasks the higher doses of melatonin impaired performance. For example, the 40 mg dose decreased performance on the Continuous Recognition Task (See Figures 13 & 14). For throughput, the dose main effect approached statistical significance [$F(3,21) = 2.59$, $p = .08$]. The Memory Recall Task yielded similar results (Figures 15 & 16). Compared to placebo, melatonin decreased throughput performance 60 min after administration. After the nap, however, these decrements had diminished. This change was reflected in a dose by time interaction that approached significance [$F(6,42) = 2.02$, $p = .08$]. Melatonin had little affect on performance following the nap. For these times, there were no significant differences between doses.

Discussion

Compared to placebo, melatonin caused dose-dependent and long lasting decreases in temperature. Melatonin was efficacious in both initiating and sustaining daytime sleep. Melatonin also affected sleep architecture by increasing Stage 2 and decreasing Stage 3-4 sleep.

For many of the tasks there was a trend for larger doses of melatonin to decrease performance assessed 60 min after administration. The 1 mg dose actually improved performance for the Mathematical Processing task. This is likely due to the mild sedation effect of such a low dose. This sedation could have improved performance by decreasing tension or by facilitating subjects' attentional resources.

Our previous research showed that subjective effects of melatonin administration peak around 3 to 4 hr after administration (French, et al., 1994). Here too, we found dose-dependent increases in subjective fatigue following melatonin administration. However, this investigation was primarily designed to assess sleep during melatonin's peak effects (i.e., 3 to 4 hr after administration). Consequently, melatonin levels peaked while subjects were attempting to sleep. For this reason, we did not expect to find large differences in performance assessed 60 min after administration. Further, because melatonin is a natural

substance with a relatively short half-life, we did not expect to find carry-over fatigue induced decrements in performance tested after the nap. The fact that melatonin facilitates sleep without decreasing post-nap performance suggests that the mechanisms mediating melatonin's hypnotic effects are physiological rather than pharmacological.

The mechanism by which exogenous melatonin affects sleep has yet to be determined. However, it is unlikely that melatonin, like other hypnotics, decreases CNS activity by increasing GABA. Rather, melatonin's hypnotic effects are likely secondary to its thermoregulatory effects. Sleep is directly associated with endogenous melatonin and with body temperature. As noted, extending the dark phase of a circadian cycle extends the duration of melatonin secretion as well as the duration of nighttime sleep (Wehr, 1991) . Furthermore, sleep tendency increases as body temperature decreases (e.g., Richardson, et al., 1982). Finally, sleep duration is determined by the phase of the temperature rhythm at bedtime, in that we tend to wake up on the increasing phase of the temperature rhythm (Czeisler, et al., 1980). Thus, sleep tendency and sleep duration are inversely related to body temperature. Given these relationships and given that endogenous melatonin mediates body temperature, it is likely that the hypnotic effects of exogenous melatonin are a result of decreasing body temperature.

We did not measure body temperature during sleep. Therefore, the direct relationship between the two was not assessed. Also, a 4 hr nap was not long enough to accurately assess the effects of melatonin on sleep duration. Indeed, at the end of the nap nearly all of the subjects in the melatonin conditions were still asleep (see Table 2). This could explain why the 10 mg dose was at least as efficacious as the 40 mg dose. On the other hand, the dose-dependent effects of exogenous melatonin may asymptote around 10 mg.

This investigation showed that 10 mg of melatonin is as good as 40 mg at promoting sleep in a four hour nap. Thus, this investigation brings us closer to identifying the correct dose for promoting sleep at this time of day. Further research is required to determine if this relationship will hold for longer naps or for naps at different times across the diurnal phase. In conclusion, melatonin was efficacious in initiating and sustaining daytime sleep without the side effects associated with benzodiazepines. Melatonin may be the ideal dose

for promoting the sleep of individuals forced to function in desynchronous environments/operations. Improving the sleep of these individuals will likely lead to increased productivity, decreased accidents and eventually to decreased long-term health risks.

Figure 1. Conversion of Tryptophan to Melatonin.

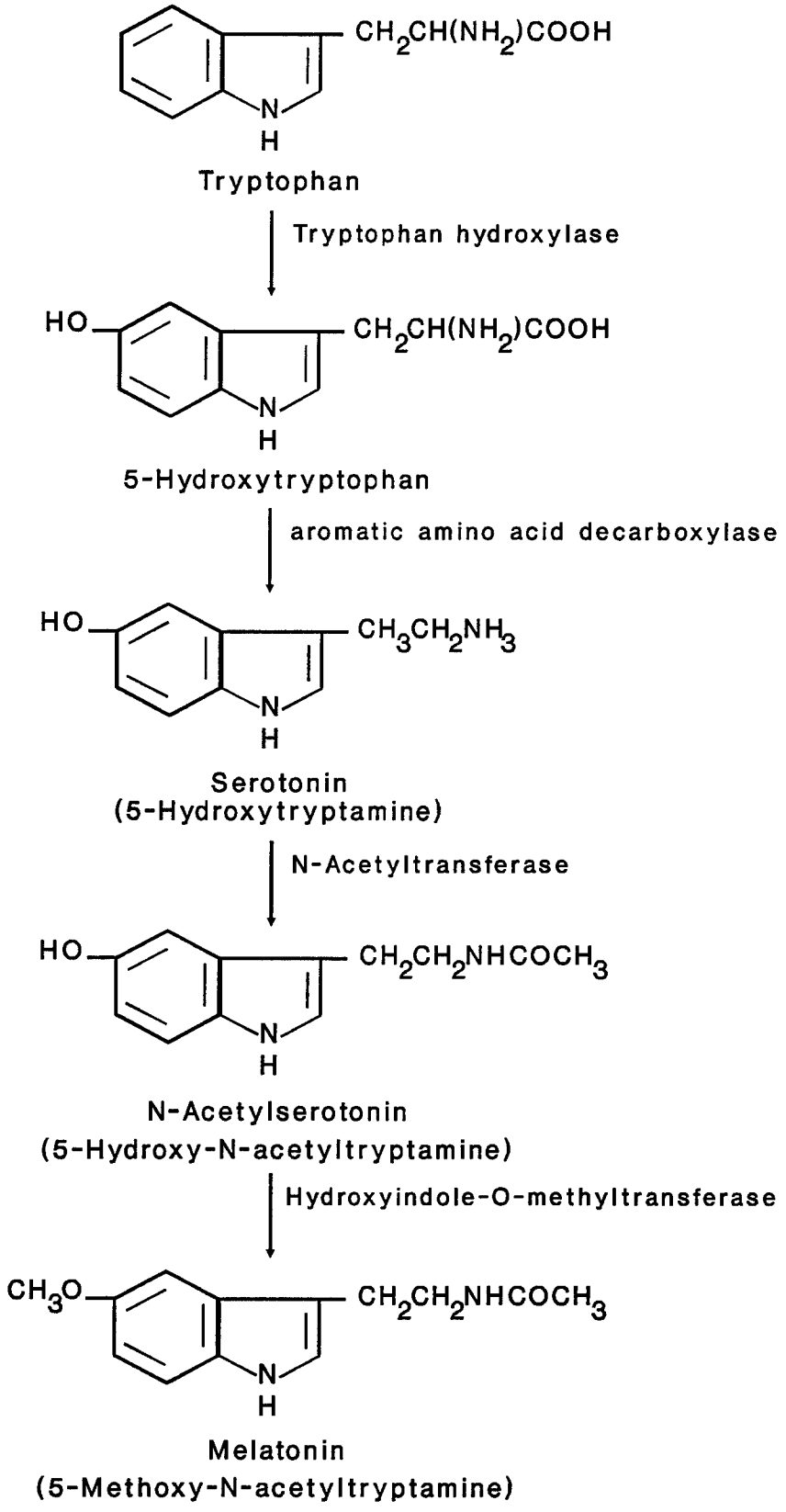


Figure 2. Melatonin induced increases in serum melatonin levels for 1 mg, 10 mg, and 100 mg of exogenous melatonin give at 0915 hr. This figure was adapted from Hughes (1992).

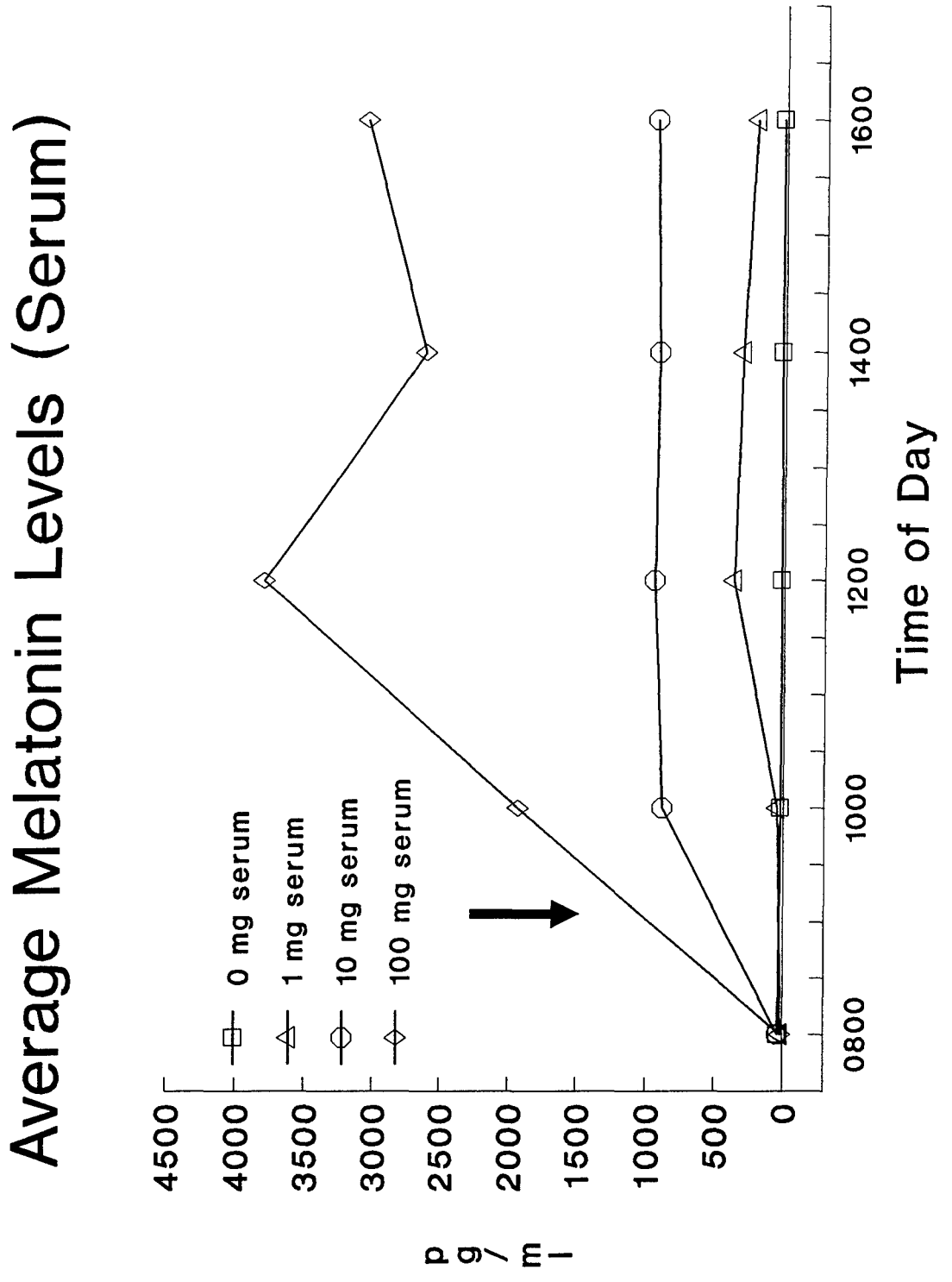


Figure 3. Results of pilot work (n=6) on the effects of melatonin (100 mg) and placebo on sleep latency, sleep efficiency and the percentage of stages 2 and 3-4 sleep.

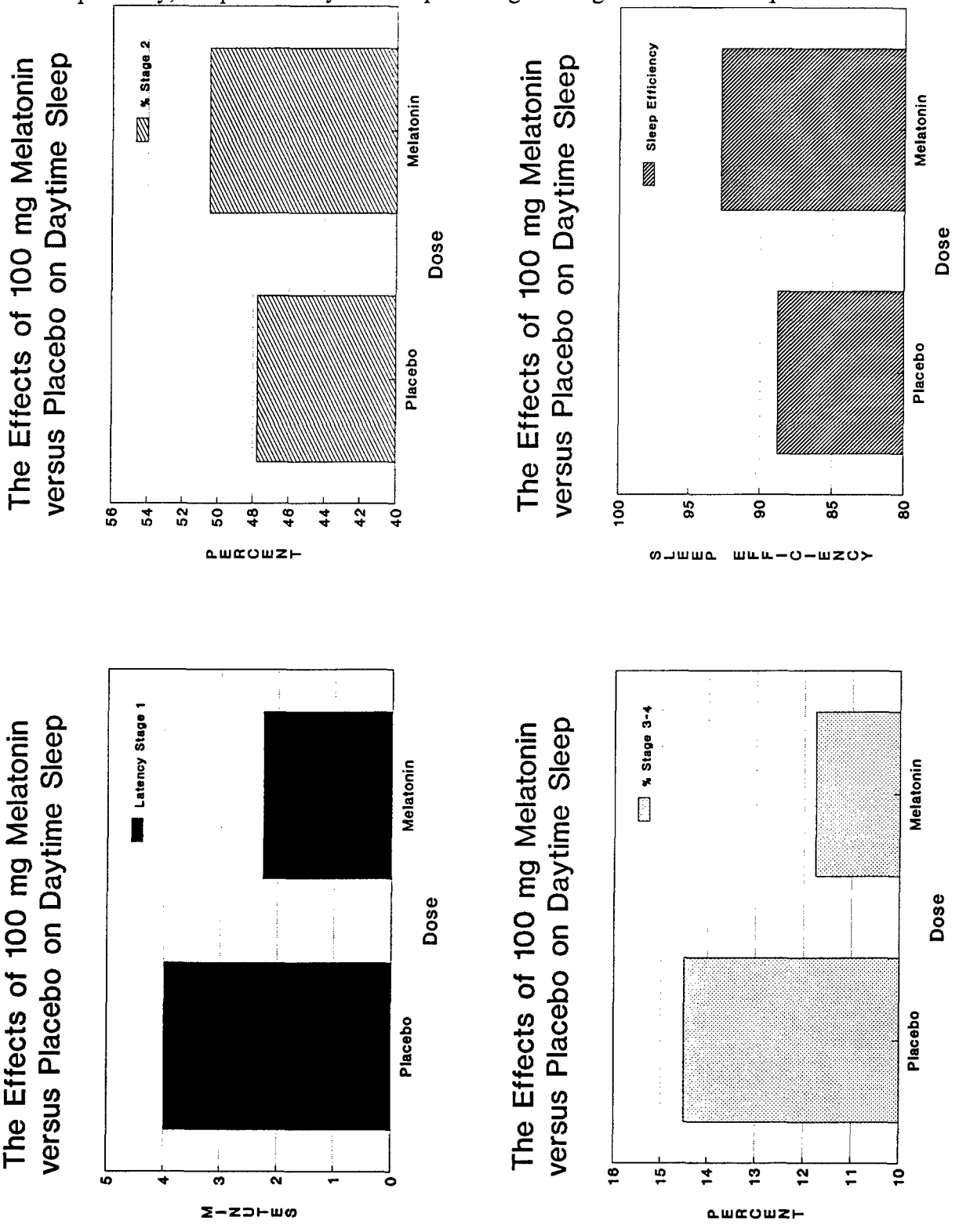


Figure 4. Pictorial time line of the testing sessions.

Time	Activity
0700	Wake up - paper work - breakfast - begin hook up
0730	"hook up" (Electrode placement)
0800	
0830	
0900	Test 0
0930	
1000	Pill administration 0 mg, 1 mg, 40 mg, or 80 mg
1030	Test 1
1100	
1130	In bed, calibration
1200	Lights out - Begin sleep assessment phase of experiment
1230	
1300	
1330	
1400	
1430	
1500	
1530	
1600	Lights on - End sleep assessment phase. Get out of bed.
1630	Test 2
1700	
1730	Test 3
1800	
1830	
1900	
1930	
2000	End experiment.

Procedural Time Line of the Testing Day. Subjects will be in the laboratory from 0700 hours to 2000 hours (13 hours).

Figure 5. Melatonin and Sleep Latency

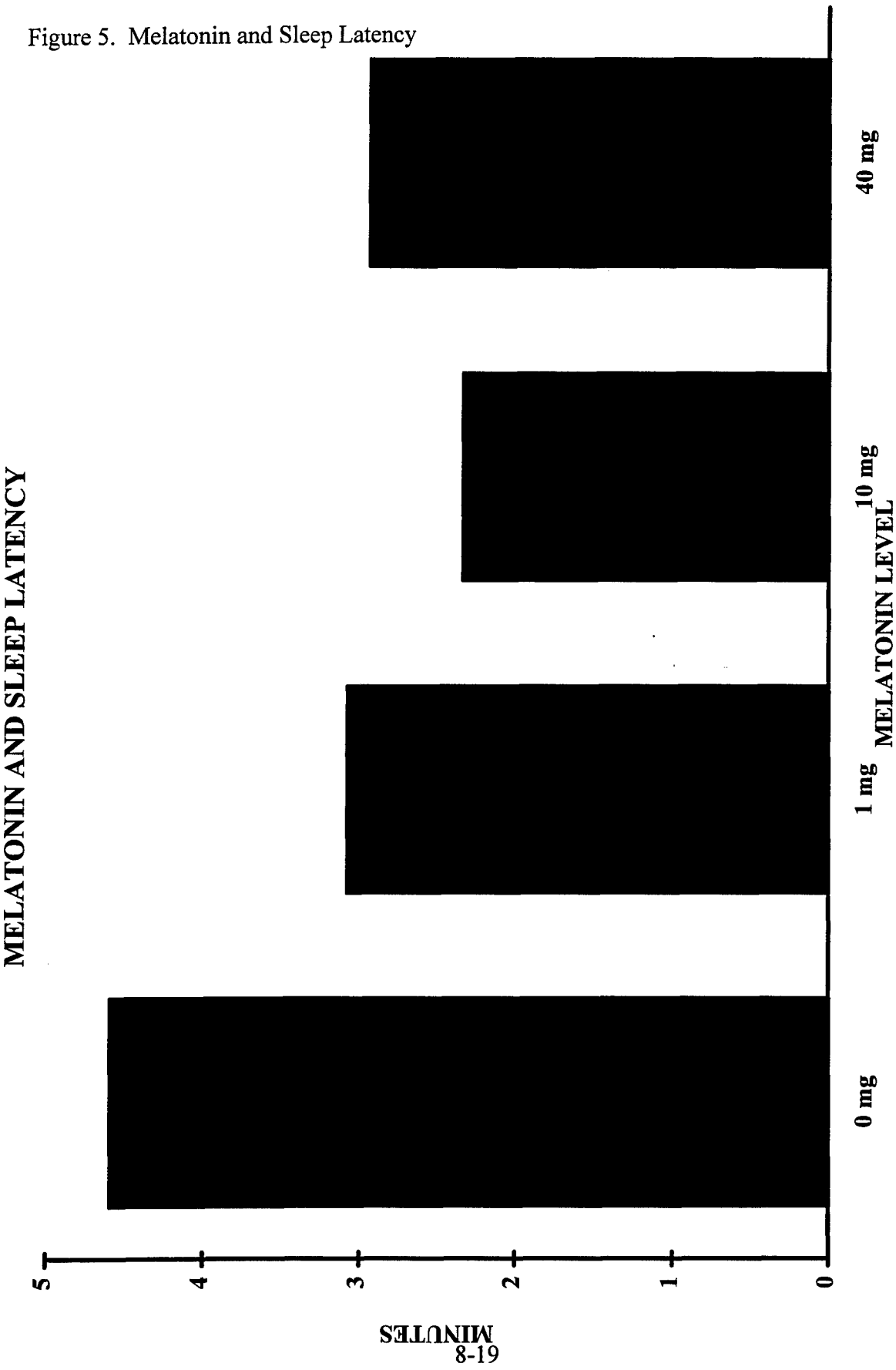


Figure 6. Melatonin and Sleep Efficiency

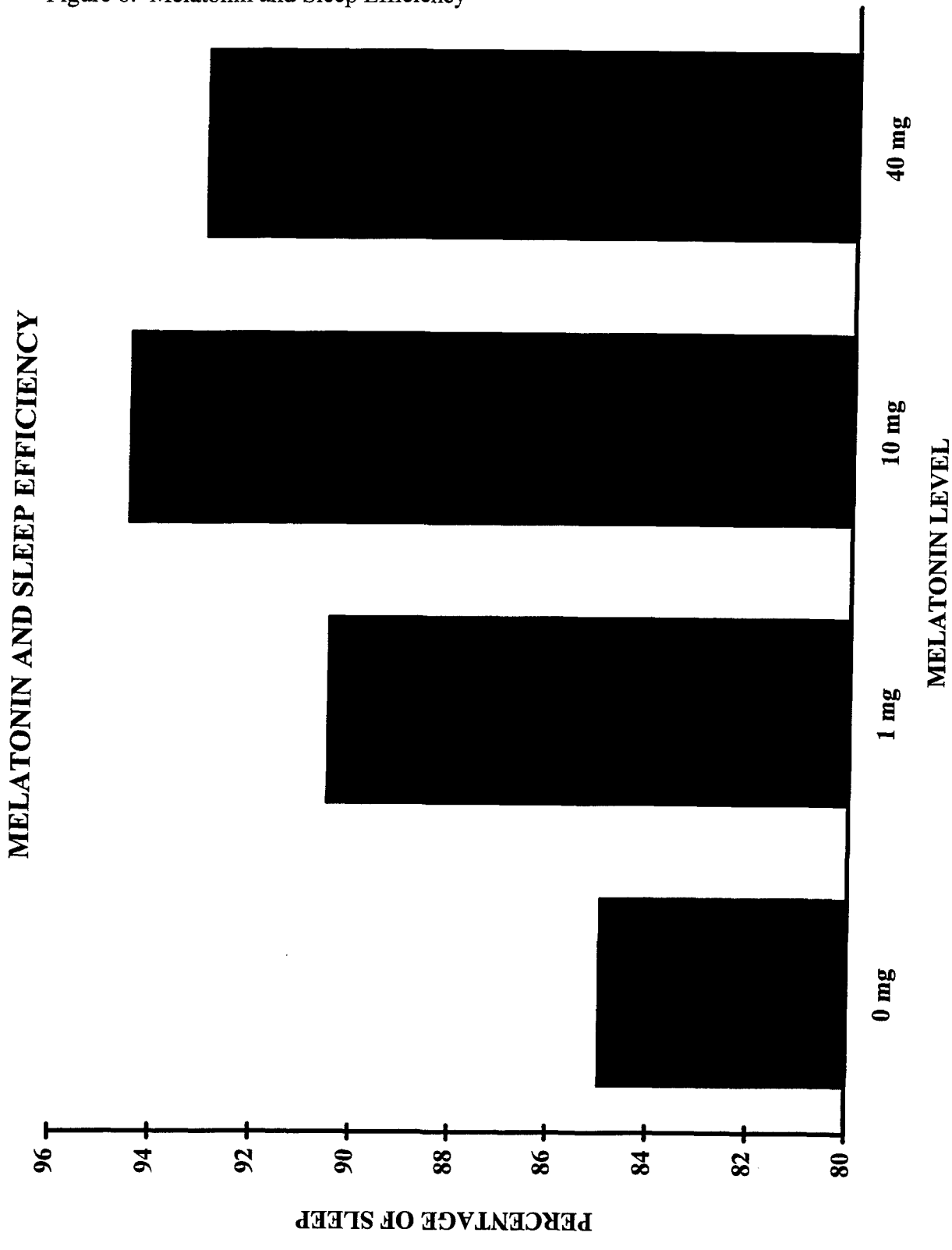


Figure 7. Melatonin and Percent Stage 2 Sleep

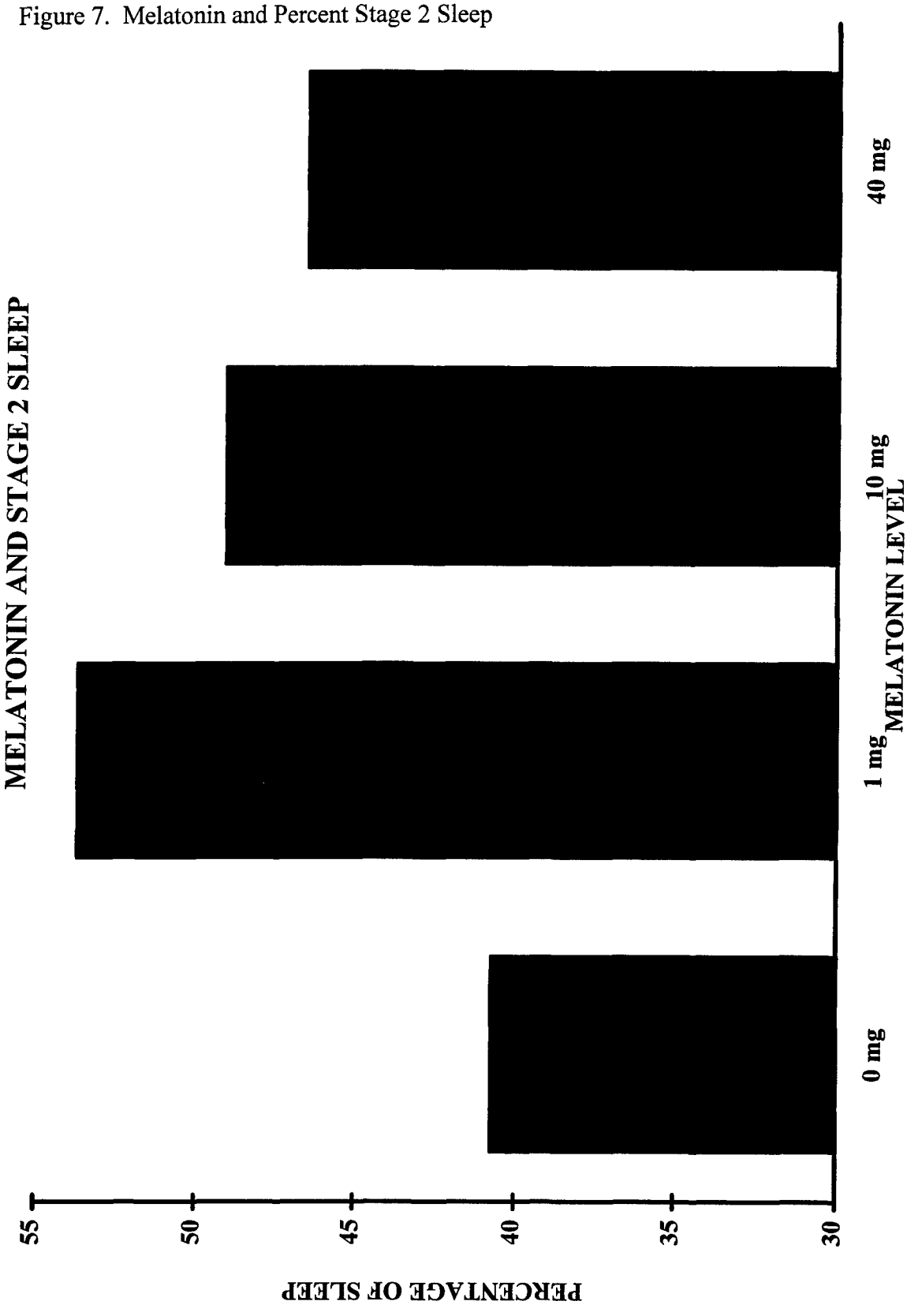


Figure 8. Melatonin and Percent Stage 3-4 Sleep

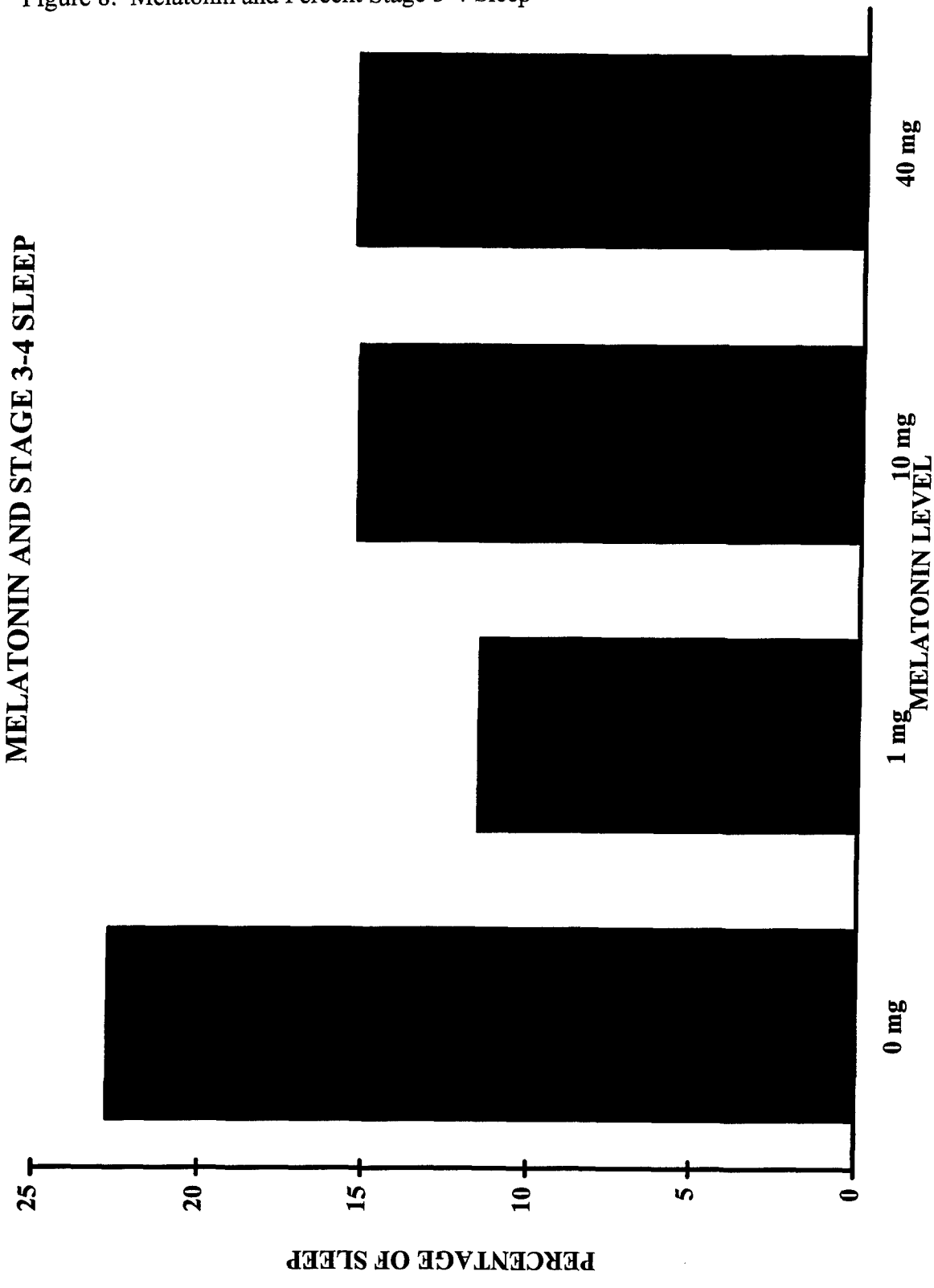


Figure 9. Melatonin and Body Temperature

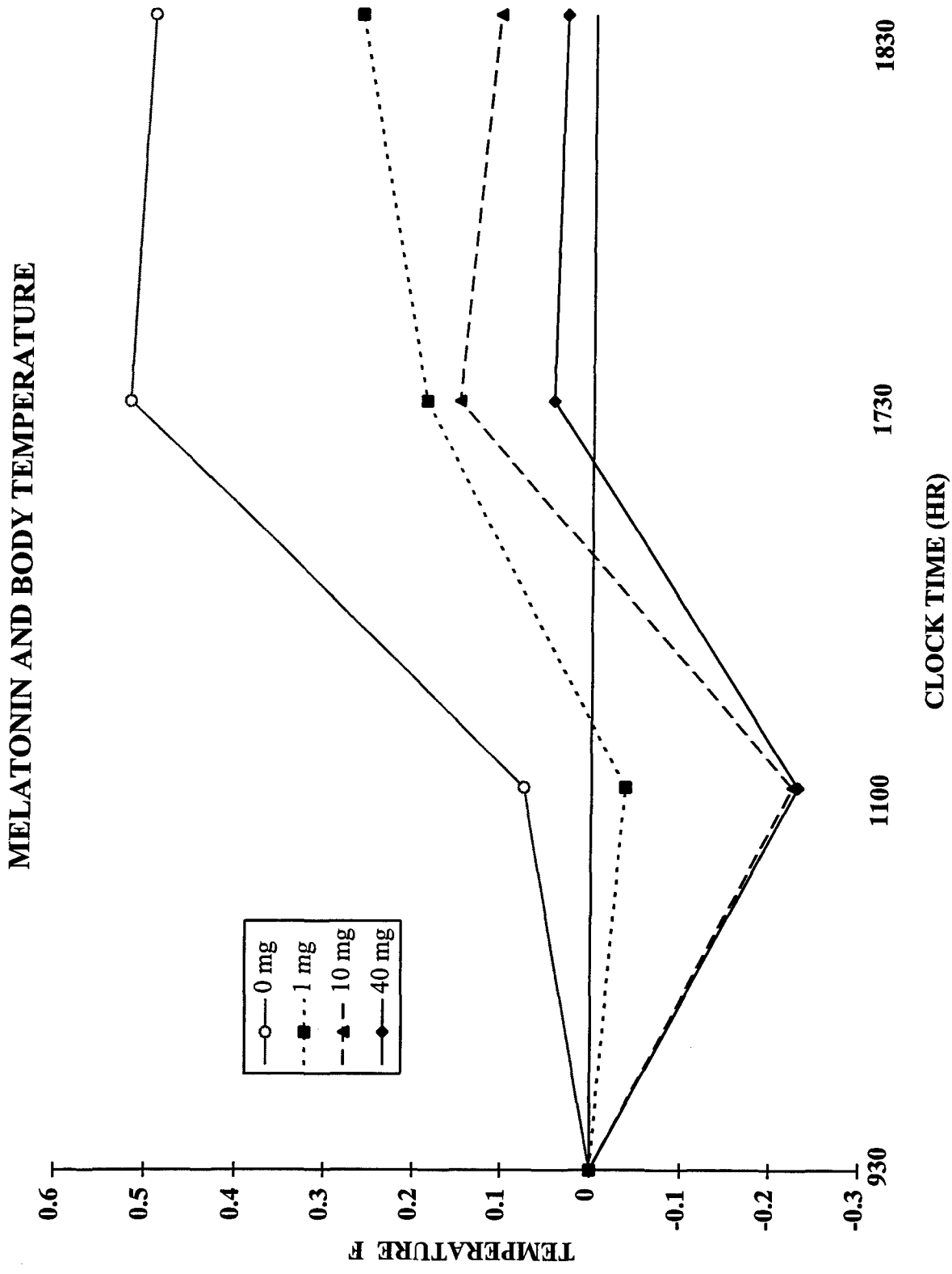


Figure 10. Melatonin and Subjective Fatigue

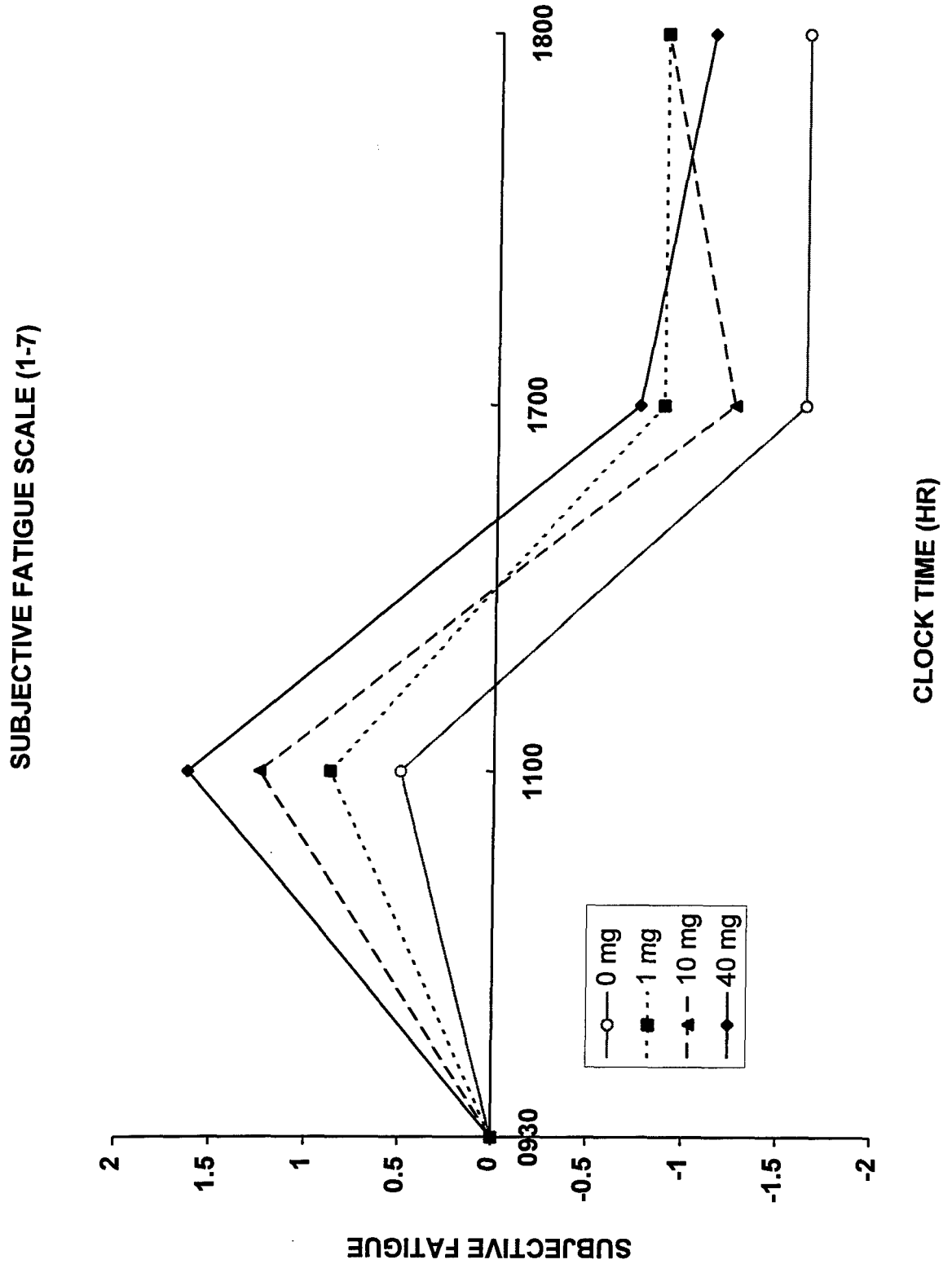


Figure 11. Mathematical Processing Task: Mean Correct Reaction Time.
Data are presented as differences from baseline (0930 hr).

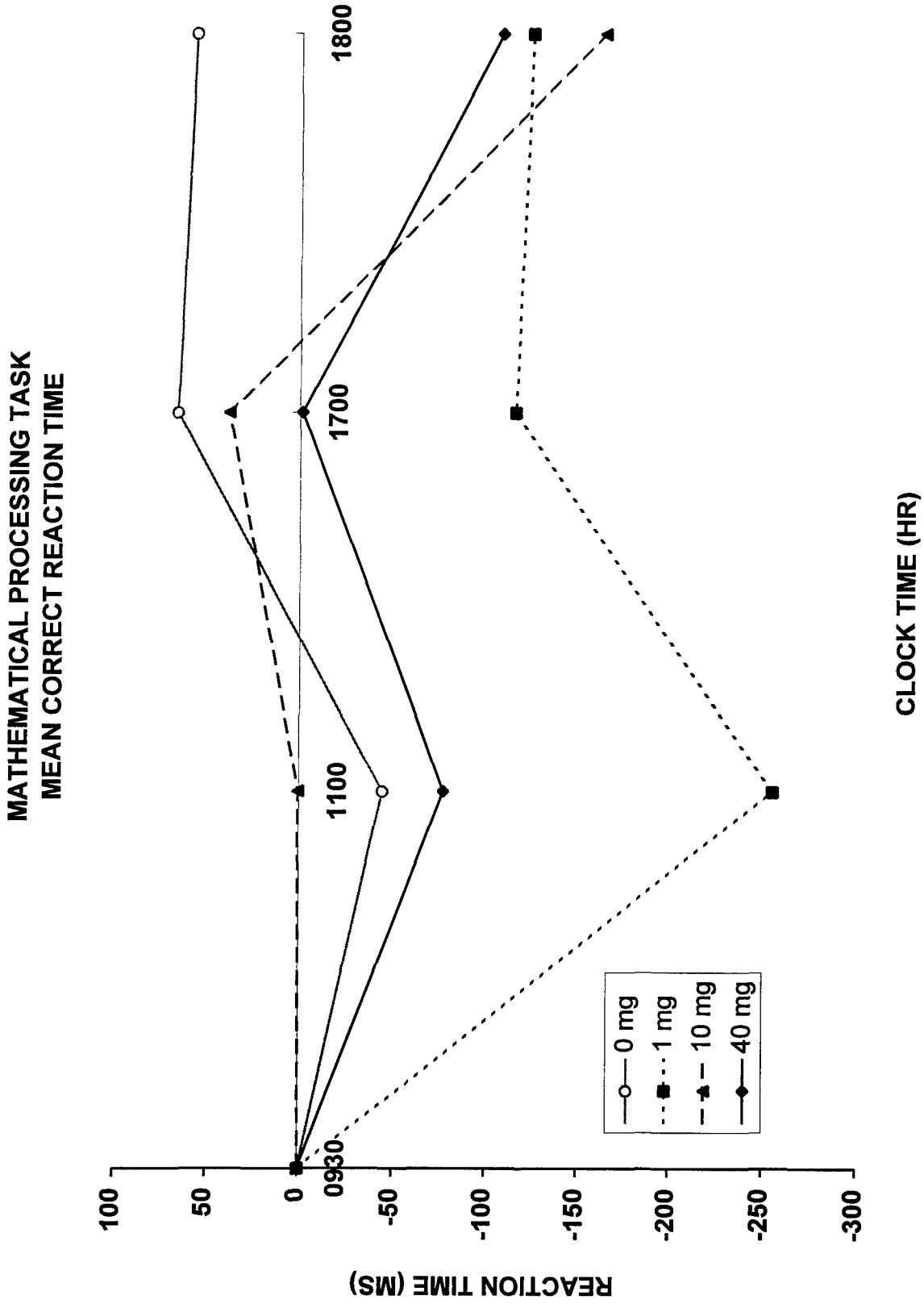


Figure 12. Mathematical Processing Task: Throughput.
Data are presented as differences from baseline (0930 hr).

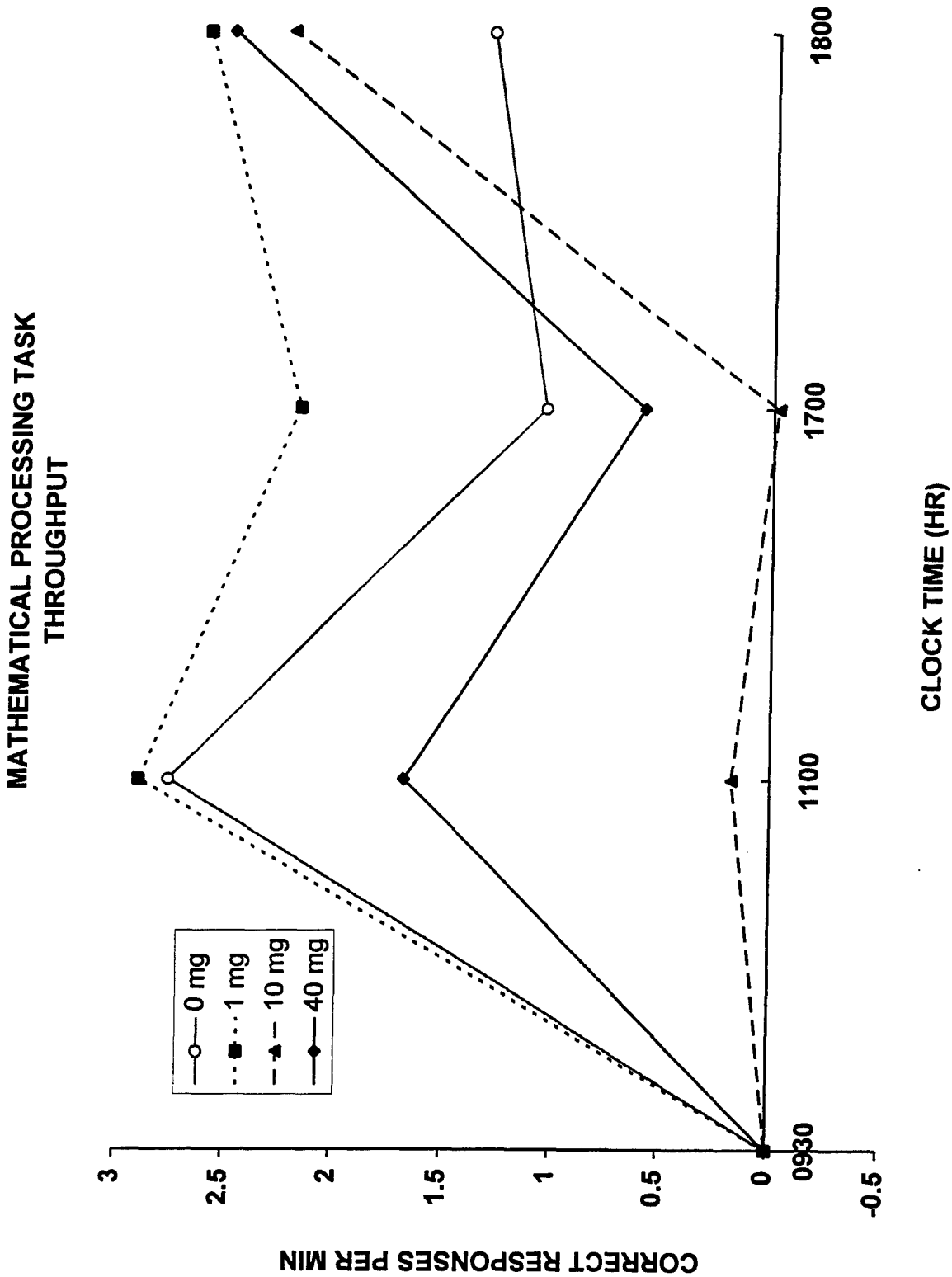


Figure 13. Continuous Recognition Task: Mean Correct Reaction Time.
Data are presented as differences from baseline (0930 hr).

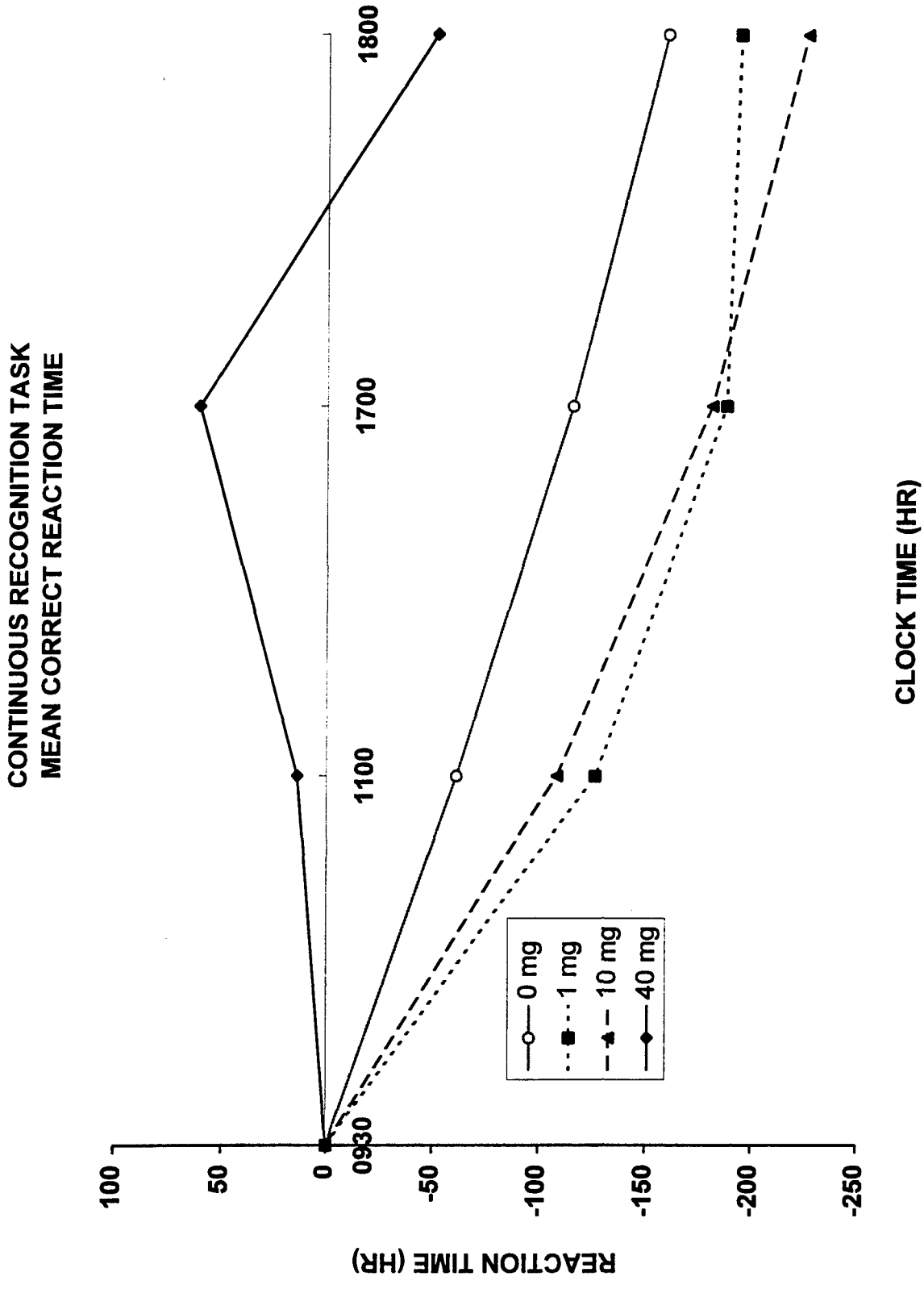


Figure 14. Continuous Recognition Task: Throughput.
Data are presented as differences from baseline (0930 hr).

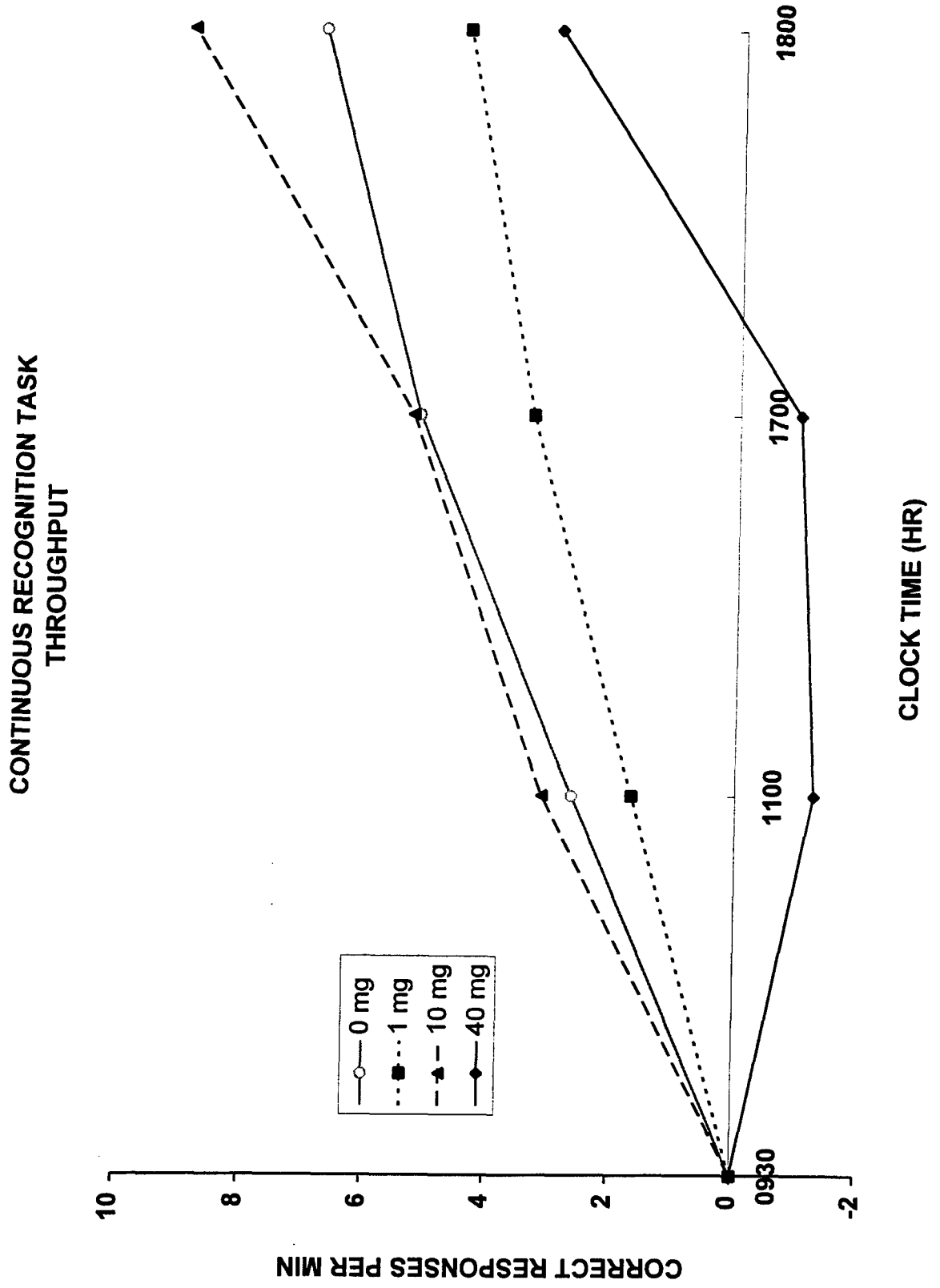


Figure 15. Memory Recall Task: Mean Correct Reaction Time.
Data are presented as differences from baseline (0930 hr).

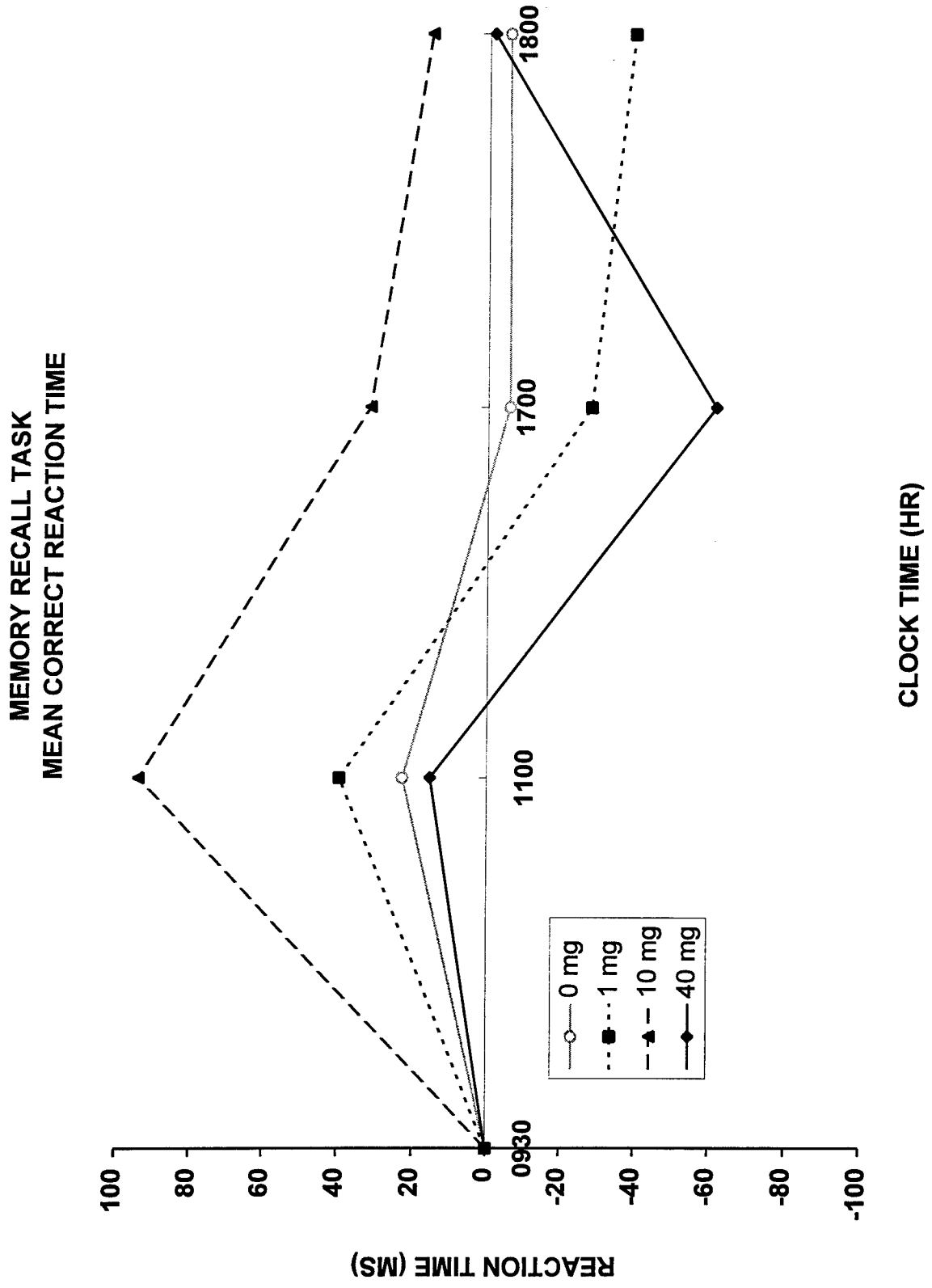


Figure 16. Memory Recall Task: Throughput.
Data are presented as differences from baseline (0930 hr).

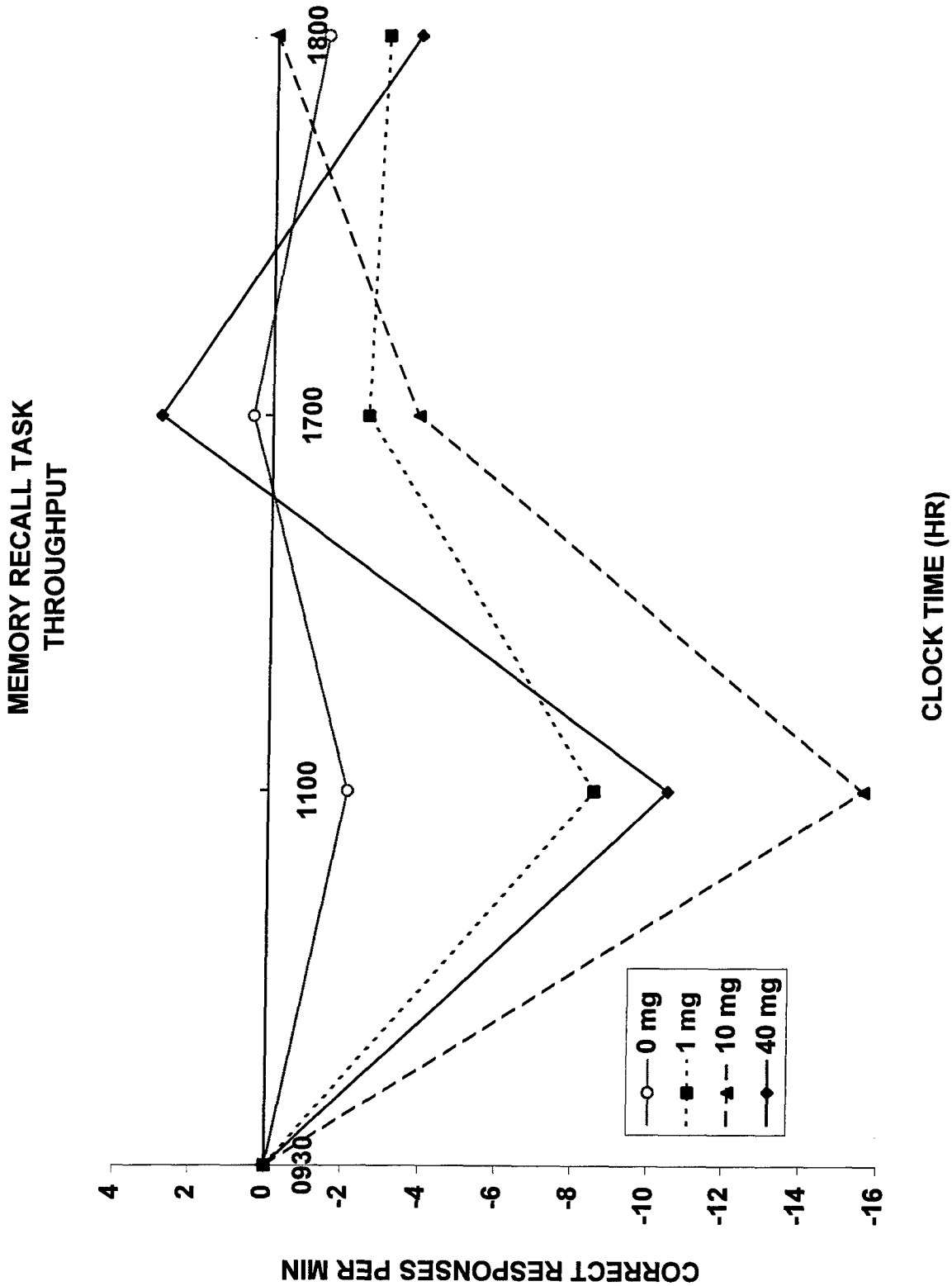


Figure 17. Procedural Memory - Basic Task: Mean Correct Reaction Time.
Data are presented as differences from baseline (0930 hr).

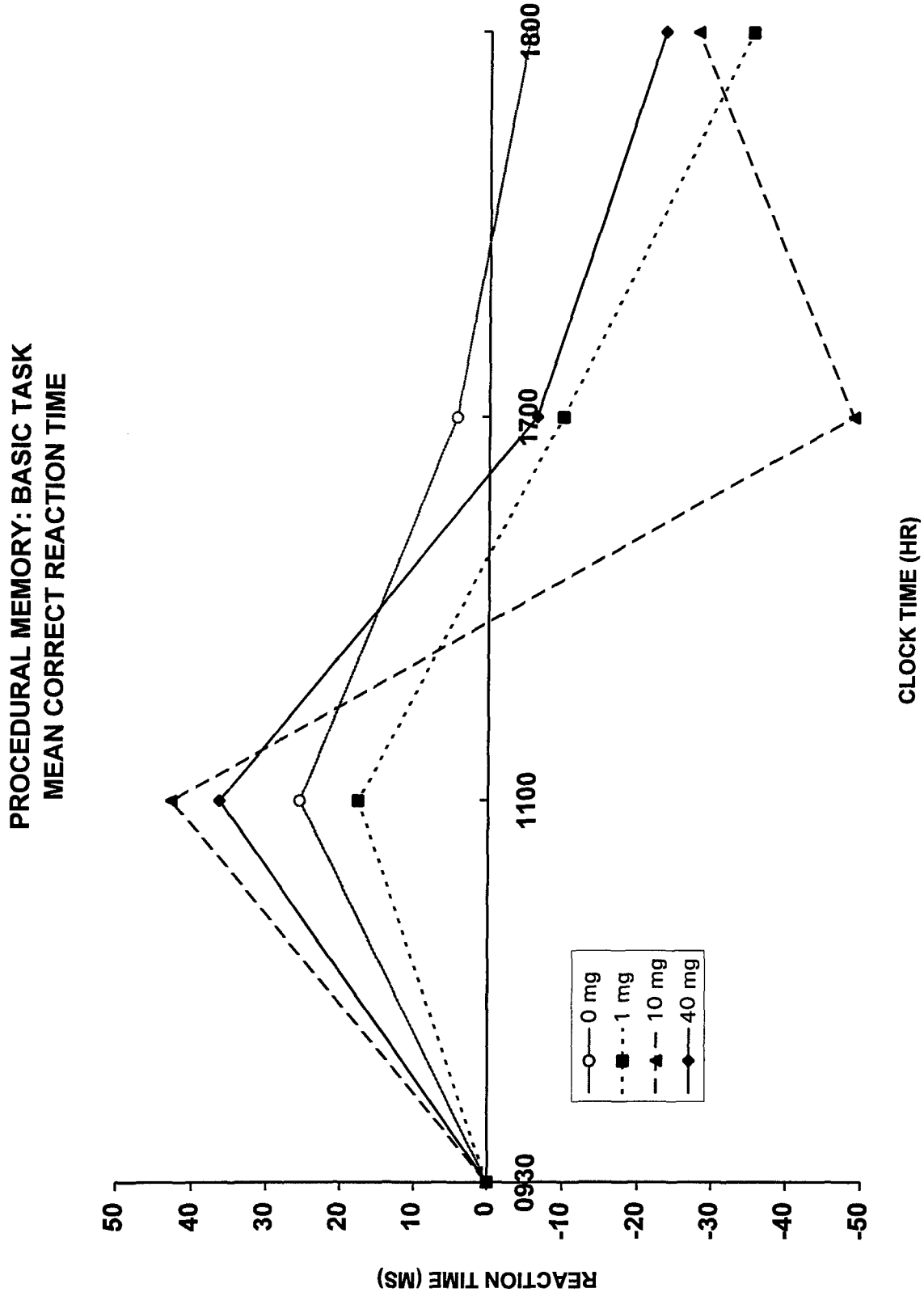


Figure 18. Procedural Memory - Basic Task: Throughput.
Data are presented as differences from baseline (0930 hr).

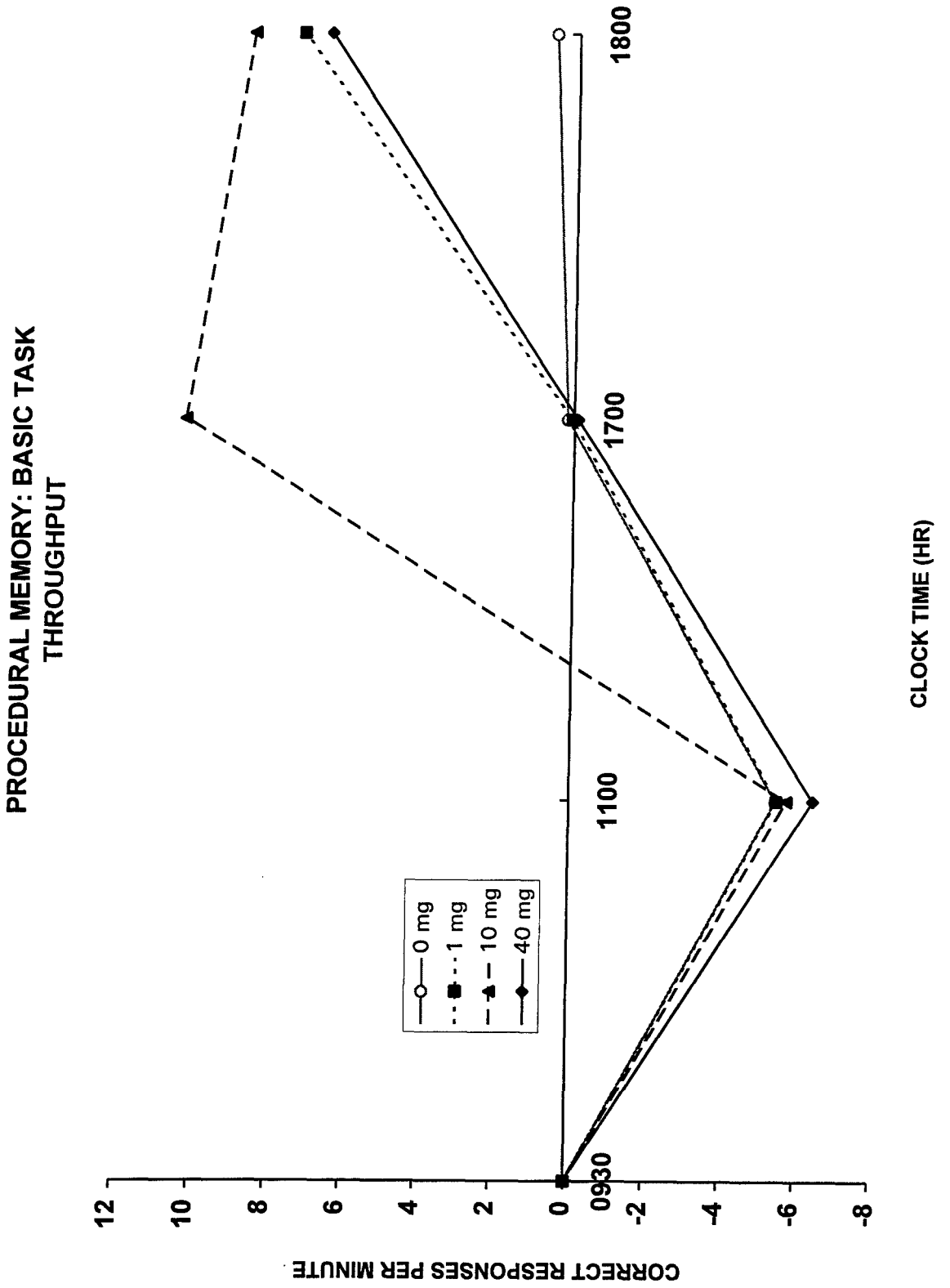


Figure 19. Procedural Memory - Coded Task: Mean Correct Reaction Time.
Data are presented as differences from baseline (0930 hr).

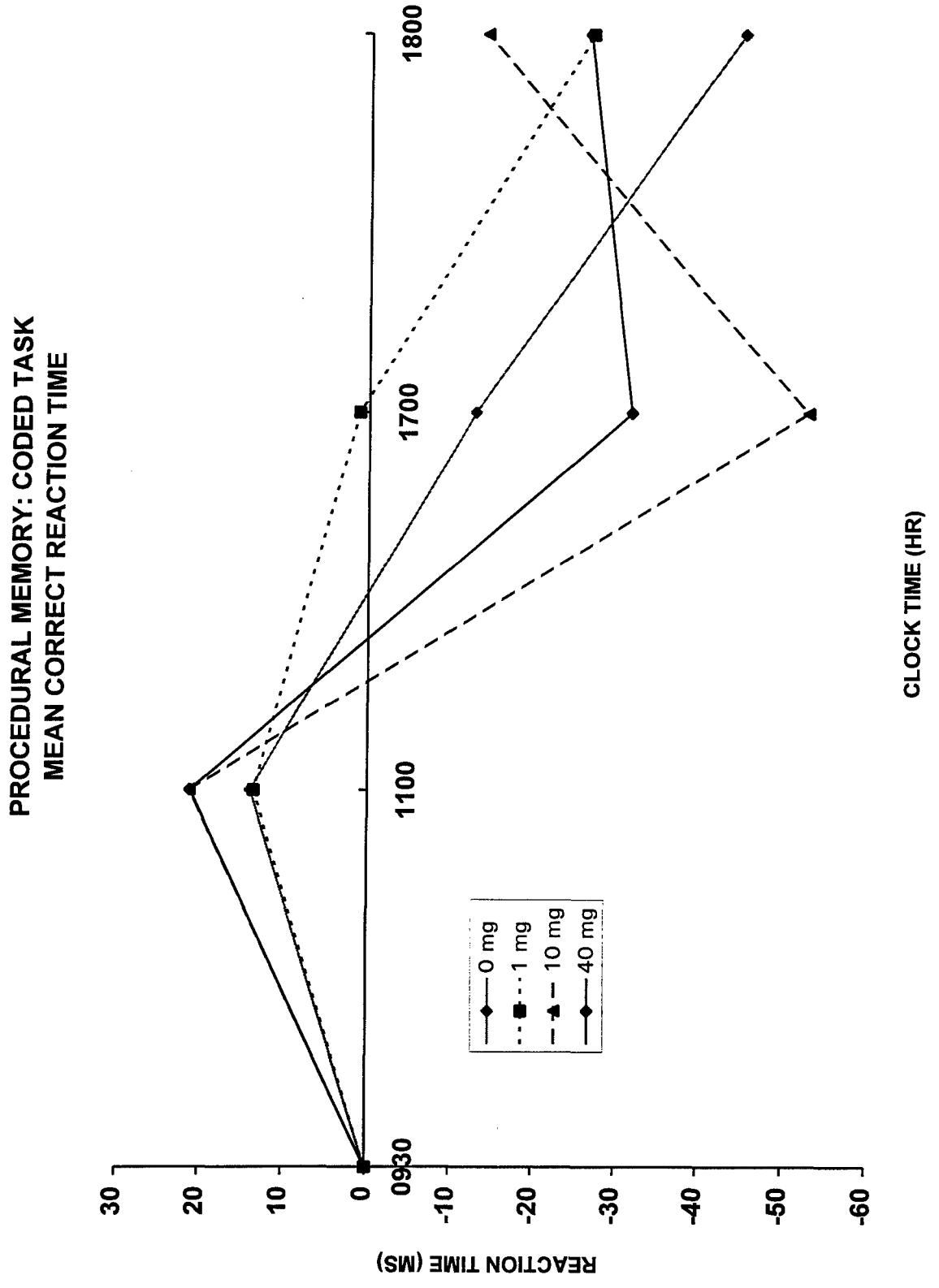


Figure 20. Procedural Memory - Coded Task: Throughput.
Data are presented as differences from baseline (0930 hr).

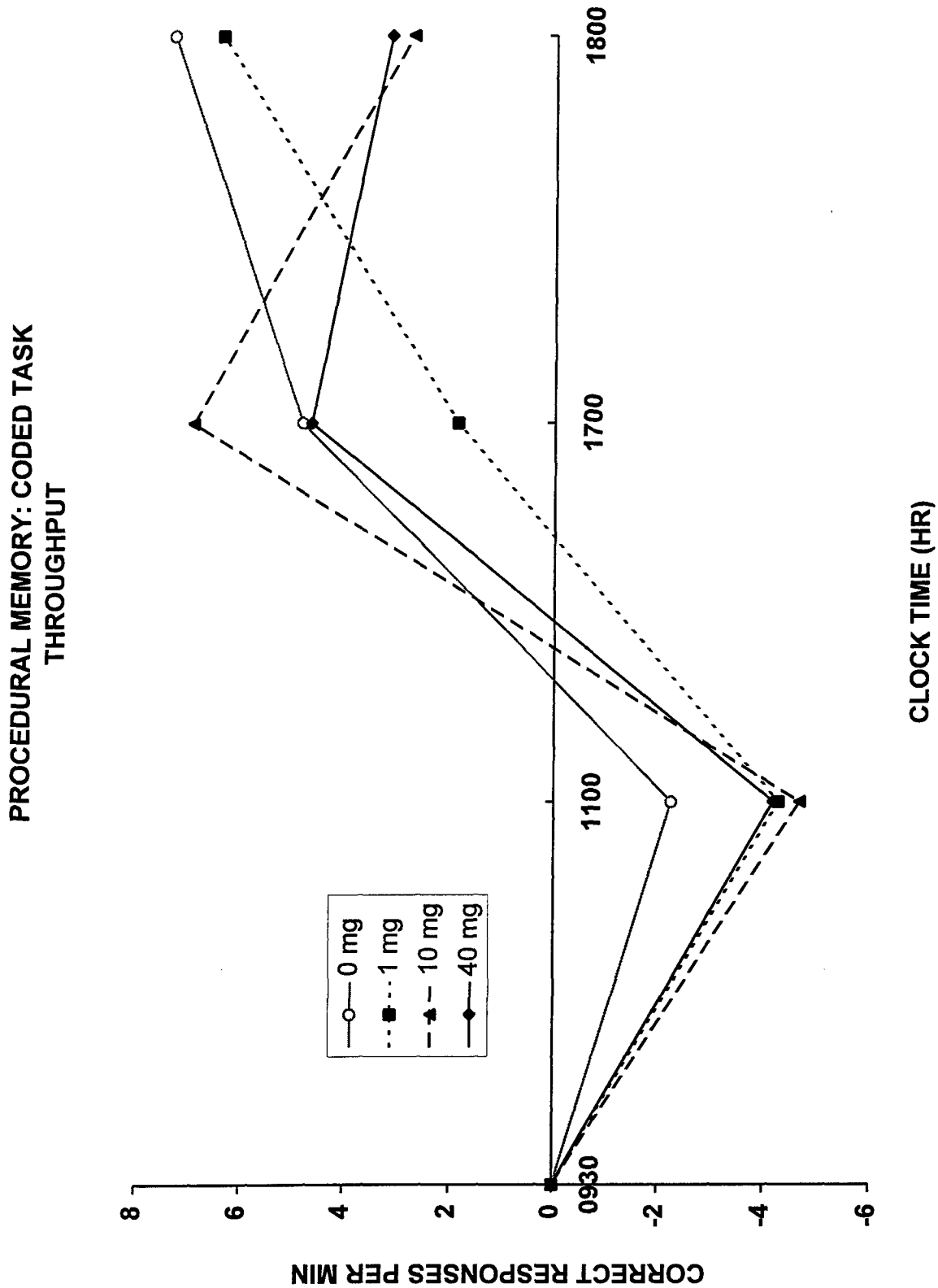


Table 1. Average Percentage of Sleep Stage by Dose.

Table 1: Average Percentage of Sleep Stage by Dose

Sleep Stage	0 mg	1 mg	10 mg	40 mg
Stage 1	**13.03%**	**12.64%**	**13.05%**	**13.06%**
Stage 2	**40.78%**	**53.71%***	**49.09%***	**46.53%**
Stage 3-4	**22.8%**	**11.61%***	**15.36%***	**15.5%***
Stage REM	**23.4%**	**22.04%**	**22.51%**	**24.94%**

*Significantly different from placebo $p < .05$

Table 2. Proportion of subjects still asleep at the end of the nap (4 hr) by dose.

Table 2. Proportion of subjects still asleep at 4 hr.

0 mg	1 mg	10 mg	40 mg
.5	.83	1.0	.83

References

Akerstedt, T. (1988). Sleepiness as a consequence of shift work. Sleep, 11(1), 17-34.

Aldhous, M., Franey, C., Wright, J. & Arendt, J. (1985). Plasma concentrations of melatonin in man following oral absorption of different preparations. British Journal of Clinical Pharmacology, 19, 517-521.

Anton-Tay, F., Diaz, J. L. & Fernandez-Guardiola, A. (1971). On the effect of melatonin upon human brain: Its possible therapeutic implications. Life Sciences, 10, 841-850.

Arendt, J., Borbely, A. A., Franey, C. & Wright, J. (1984). The effect of chronic, small doses of melatonin given in the late afternoon on fatigue in man: a preliminary study. Neuroscience Letters, 45, 317-321.

Arendt, J., Aldhous, M. & Marks, V. (1986). Allevieation of jet-lag by melatonin: preliminary results of a controlled double-blind trial. British Medical Journal, 292, 1170.

Arendt, J., Aldhous, M., Marks, M., Folkard, S., English, J., Marks, V. & Arendt, J. H. (1987). Some effects of jet-lag and its treatment by melatonin. Ergonimics, 30, 1379-1393.

Arendt, J., Aldhous, M. & Wright, J. (1988). Synchronization of a disturbed sleep-wake cycle in a blind man by melatonin. Lancet, 1 772-773.

Badia, P., Meyers, B. & Murphy, P. Melatonin and thermoregulation. In R.J. Reiter & H-S. Yu (Eds) Melatonin: Biosynthesis, Physiological Effects, and Clinical Applications. (1992). CRC Press, Boca Raton, FL.

Cramer, H., Rudolph, J. & Consbruch, U. (1974). On the effects of melatonin on sleep and behaviour in man. In Serotonin: new vistas, biochemestry and behavioral and clinical studies (E. Costa Ed.) Advances in Biochemical Psychopharmacology Volume 2, pp. 187-191. Raven Press, New York.

Czeisler, C., Weitaman, E., Moore-Ede, M., Zimmerman, J. & Knauer, R. (1980). Human sleep: Its duration and organization depend on its circadian phase. Science, 210, 1264-1267.

Dahlitz, M., Alvarez, B., Vignau, J., English, J., Arendt, J. & Parkes, J. D. (1991). Delayed sleep phase syndrome response to melatonin. Lancet, 337(8750), 1121-1124.

Dollins, A., et al., (November, 1993). Induced daytime melatonin levels comparable to normal nocturnal levels affect human mood and performance. (abstract) 23rd annual meeting of the Society of Neuroscience. Washington, D.C.

French, J., Boll, P. & Storm, W. A. (1986). Temazepam and performance following a sleep cycle shift. Annual Review of Chronopharmacology, 7, 41-44.

French, J., Hughes, R., Whitmore, J., Neville, K., & Reiter, R. J. (1993) Effects of diurnal melatonin on oral temperature and subjective fatigue. Research to be presented at the 1994 meeting of the Aerospace Medical Association.

Gillooly, P. B., Smolensky, M. H., Albright, D. L., Barholomew, H. & Thorne, D. R. (1990). Circadian variation in human performance evaluated by the Walter Reed Performance Assessment Battery. Chronobiology International, 7(2), 143-153.

Hommer D.W. (1992). Benzodiazepines and cognitive and psychomotor effects. In P.R. Roy-Byrne & D.S. Cowley (Eds) Benzodiazepines in Clinical Practice: Risks and Benefits. American Psychiatric Press, Washington, DC. pp 111-130.

Hughes, R. J. The effects of two doses of exogenous melatonin on temperature and subjective fatigue. Final report submitted to the Air Force Office of Scientific Research, September, 1992.

Lewy, A. J., Wehr, T. A., Goodwin, F. K., Newsome, D. A. & Markey, S. P. (1980). Light suppresses melatonin secretion in humans. Science, 210, 169-171.

Lewy, A. J., Ahmed, S., Latham Jackson, J. M. & Sack R. L. (1992). Melatonin shifts human circadian rhythms according to a phase-response curve. Chronobiology International, 9(5), 380-392.

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

Moore-Ede, M. C. & Richardson, G. S. (1985). Medical implications of shift-work. Annual Review of Medicine, 36, 607-617.

Naitoh, P., Kelly, T. L. & Englund, C. (1990). Health effects of sleep deprivation. In Occupational Medicine: State of the Art Reviews, 5(2). Philadelphia, Hanley & Belfus, pp 209-207.

Naitoh, P., Kelly, T. L. & Englund, C. (1990). Health effects of sleep deprivation. In Occupational Medicine: State of the Art Reviews, 5(2). Philadelphia, Hanley & Belfus, pp 209-207.

Nickelson, T., Demisch, L., Demisch, K., Radermacher, B. & Schoffling, K. (1989). Influence of subchronic intake of melatonin at various times of the day on fatigue and hormonal levels: a placebo-controlled, double-blind trial. Journal of Pineal Research, 6(4), 325-334.

Palm, L., Blennow, G. & Wetterberg, L. (1991). Correction of non-24-hour sleep/wake cycle by melatonin in a blind retarded boy. Ann Neurol, 29(3), 336-339.

Pascually, R. (1991). Benzodiazepines and Sleep. In P.R. Roy-Byrne & D.S. Cowley (Eds) Benzodiazepines in Clinical Practice: Risks and Benefits. American Psychiatric Press, Washington, DC. pp 93-107.

Petrie, K., Conaglen, J. V., Thompson, L. & Chamberlan, C. (1989). Effects of melatonin on jet-lag after dual flights. British Medical Journal, 198, 705-707.

Reiter, R. J. (1990). Pineal rhythmicity: Neural, behavioral, and endocrine consequences. In (M. Shafii and S.L. Schafii, Eds) Biological Rhythms, Mood Disorders, Light Therapy, and the Pineal Gland. American Psychiatric Press, Washington.

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

Reiter, R. J. (1991b). Neurendocrine effects of light. International Journal of Biometeorology, 35, 169-175

Reiter, R. J. (1991c). Pineal melatonin: Cell biology of its synthesis and of its physiological interactions. Endocrine Reviews, 12(2), 151-180.

Reiter, R. J. (1992). The ageing pineal gland and its physiological consequences. BioEssays, 14(3), 169-175.

Richardson, G., Carskadon, E., Orav, E. & Dement, W. (1982). Circadian variation of sleep tendency in elderly and young adult subjects. Sleep, 5, S82-S94.

Samel, A., Wegmann, H-M., Vejvoda, M., MaaB, H., Gundel, A. & Schutz, M. (1991). Influence of melatonin treatment on human circadian rhythmicity before and after a simulated 9-Hr time shift. Journal of Biological Research, 323, 235-248.

Storm, W. G. & Parke, R. C. (1986). FB-111A aircrew use of temazepam during surge operations AGARD-CP-415, 12-1 - 12-11.

Thorne, D. R., Genser, S. G., Sing, H. C. & Hegge, F. W. (1985). The Walter Reed Performance Assessment Battery. Neurobehavioral Toxicology and Teratology, 7, 415-418.

Vollrath, L., Semm, P. & Gammel, G. (1981). Sleep induction by intranasal application of melatonin. Advances in the Biosciences, 29, 327-329.

Waldhauser, F., Lieberman, H., Lynch, H., Waldhauser, M., Herkner, K., Frisch, H., Vierhapper, H., Waldhausl, W., Schemper, M., Wurtman, R. & Crowley, W. (1987). A pharmacological dose of melatonin increases PRL levels in males without altering those of GH, LH, FSH, TSH, Testosterone or Cortisol. Neuroendocrinology, 46, 125-130.

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

Wever, R. A. (1979). The circadian system of man: results of experiments under temporal isolation. Springer Verlag, New York.

WHITE-NOISE ANALYSIS OF CAROTID BARORECEPTOR FUNCTION IN BABOONS

Arthur Koblasz, PhD
Associate Professor
School of Civil Engineering

Georgia Institute of Technology
Atlanta, GA 30332-0355

Final Report
for
Summer Extension Program

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

January 1994

WHITE-NOISE ANALYSIS OF CAROTID BARORECEPTOR FUNCTION IN BABOONS

Arthur J. Koblasz
Associate Professor
School of Civil Engineering
Georgia Institute of Technology

Abstract

A white-noise protocol was evaluated for characterizing carotid baroreceptor function in three adult male baboons. The white-noise (pseudo-random binary) stimulus was created by varying the pressure in the right carotid sinus. The pseudo-random stimulus was sustained for a period of 3 minutes at 3 different mean levels-- 40 mm Hg, 70 mm Hg and 100 mm Hg. The baroreceptor response at each mean level was indicated by continuously measuring the pressure in both the right atrium and the aorta. First-order Wiener Kernels were calculated from this stimulus-response data (see Appendix A for analytical details). The first-order Wiener Kernels were then used to predict the pressure changes in the right atrium and aorta which would result from a pulse of pressure at the right carotid sinus.

The white-noise protocol reveals that the carotid baroreceptors are still effecting right atrium pressure more than 50 seconds after the pressure in the right carotid has been pulse modulated. If our predicted responses are correct, then the time-constants for g-induced cardiovascular changes can be much longer than expected. Also, the delayed response is in the wrong direction-- a positive pulse at the carotid sinus causes vasoconstriction.

These unexpected results could be the result of reduced pO2 at the right carotid sinus, which is a common situation in prolonged air combat. Other investigators have reported that the carotid baroreceptor response is entirely eliminated under similar conditions. In future experiments, we will investigate the effect of pO2 on carotid baroreceptor transduction speed. We will also use aperiodic pulse stimuli, i.e. random intervals between the pulses, and will continue to measure the response to each pulse stimulus for 100 seconds after each stimulus.

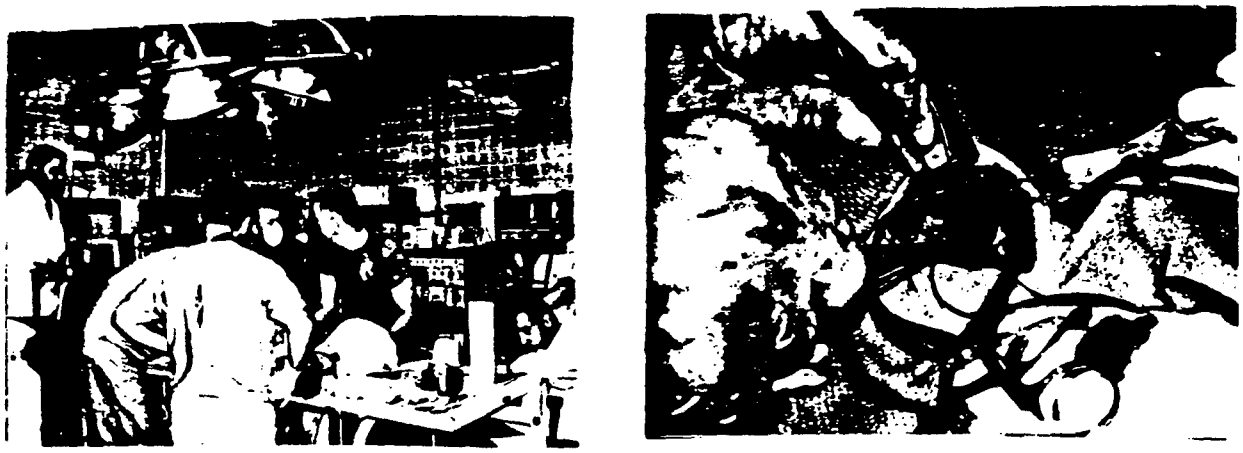


Figure 1.

Background

The white-noise method for characterizing nonlinear biological systems has been successfully demonstrated in catfish retina studies (Marmarelis and Naka, 1973), in human ERG studies (Koblasz, et al, 1980) and in a variety of other vertebrate studies (Marmarelis, 1976). Similar nonlinear characterization schemes have been demonstrated using random square wave stimuli (Yasui and Koblasz, 1984) and random pulse sequences (Fricker and Sanders, 1975).

Recently, a white-noise protocol has been demonstrated in a study of aortic baroreceptor function using in vivo rabbit preparations (Masaru, et. al., 1990). The aortic pressure was modulated by electrically stimulating the right ventricle using pacing electrodes triggered at a constant frequency of 400 beats per minute. The pacing was sustained for variable durations of greater than 1 second per burst, and the interval between bursts was varied to produce aortic pressure fluctuations with a fairly flat power spectrum over the frequency range of DC to 1/2 Hz. (In our protocol, we will block the normal pulsatile flow into the left carotid artery and will then modulate the left carotid pressure to produce a stimulus power spectrum which is very flat over the frequency range of DC to 5 Hz.)

In Masaru's protocol, the aortic pressure was measured using a high-fidelity micromanometer (Millar MPC-500), and the baroreceptor response was measured using Ag/AgCl bipolar electrodes positioned at the distal end of the (desheathed) left aortic depressor nerve. Linear transfer functions were calculated using these stimulus/response data. The resulting math model characterized the combination of aortic wall mechanics followed by neural transduction and encoding mechanisms.

EXPERIMENTAL PROTOCOL

Three adult male baboons were anesthetized with Ketamine (30mg IM) followed by periodic doses of a-Chloralose (50mg/kg for the first dose and 20 mg/hr thereafter). Pancuronium Bromide (.1mg/kg IV) was also given to each animal to reduce muscle activity. A 3-French Millar pressure transducer (single-tipped) was inserted into the right femoral artery to record systemic pressure changes resulting from the pressure stimuli applied to the right carotid sinus. EKG was also recorded to indicate changes in heartrate. Figure 1 presents a collection of photographs which were taken during the animal surgery.

In the first set of experiments, the variations in carotid pressure were produced by randomly varying the flow rate of IV fluid injected into the common carotid artery. The flow rate was modulated using a stepping motor connected to a linear hydraulic valve, as depicted in Figure 2. At periodic intervals, equivalent volumes of blood were removed from the right femoral vein. Unfortunately, the relationship between the volume of blood injected into the carotid artery and the carotid sinus pressure is much more complicated than we had expected. Futhermore, it is difficult to maintain a constant total blood volume during a prolonged experiment.

In the second set of experiments, the right carotid sinus was isolated by ligating all incoming and outgoing vessels approximately 1 cm above and below the carotid sinus. A small catheter was then inserted into the carotid sinus, along with a 2-French Millar (single-tipped) pressure transducer. The catheter was connected to a Skinner three-way hydraulic valve (Type B14), which allowed the catheter to be switched between two reservoirs of warmed physiological saline solution. The reservoirs were positioned at different heights above the carotid sinus; hence, switching between the two reservoirs produced step changes in pressure inside the carotid sinus. Since the carotid sinus was completely isolated, the pressure shifts occurred nearly instantaneously with no significant flow of saline into or out of the sinus.

The circuit shown in Figure 3 was used to create a pseudo-random binary (two-level) signal and was also used to generate periodic step and pulse signals-- the latter being more conventional stimuli. The pseudo-random binary signal was used as a control input to the three-way hydraulic valve, thus producing pseudo-random binary pressure changes inside the carotid sinus. The more conventional step and pulse stimuli were also generated to validate the pseudo-random data (see report for Summer 1992).

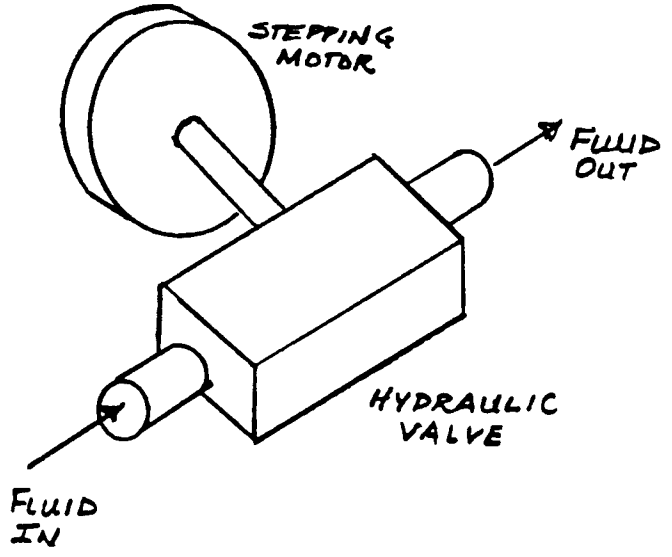


FIGURE 2
Linear hydraulic valve

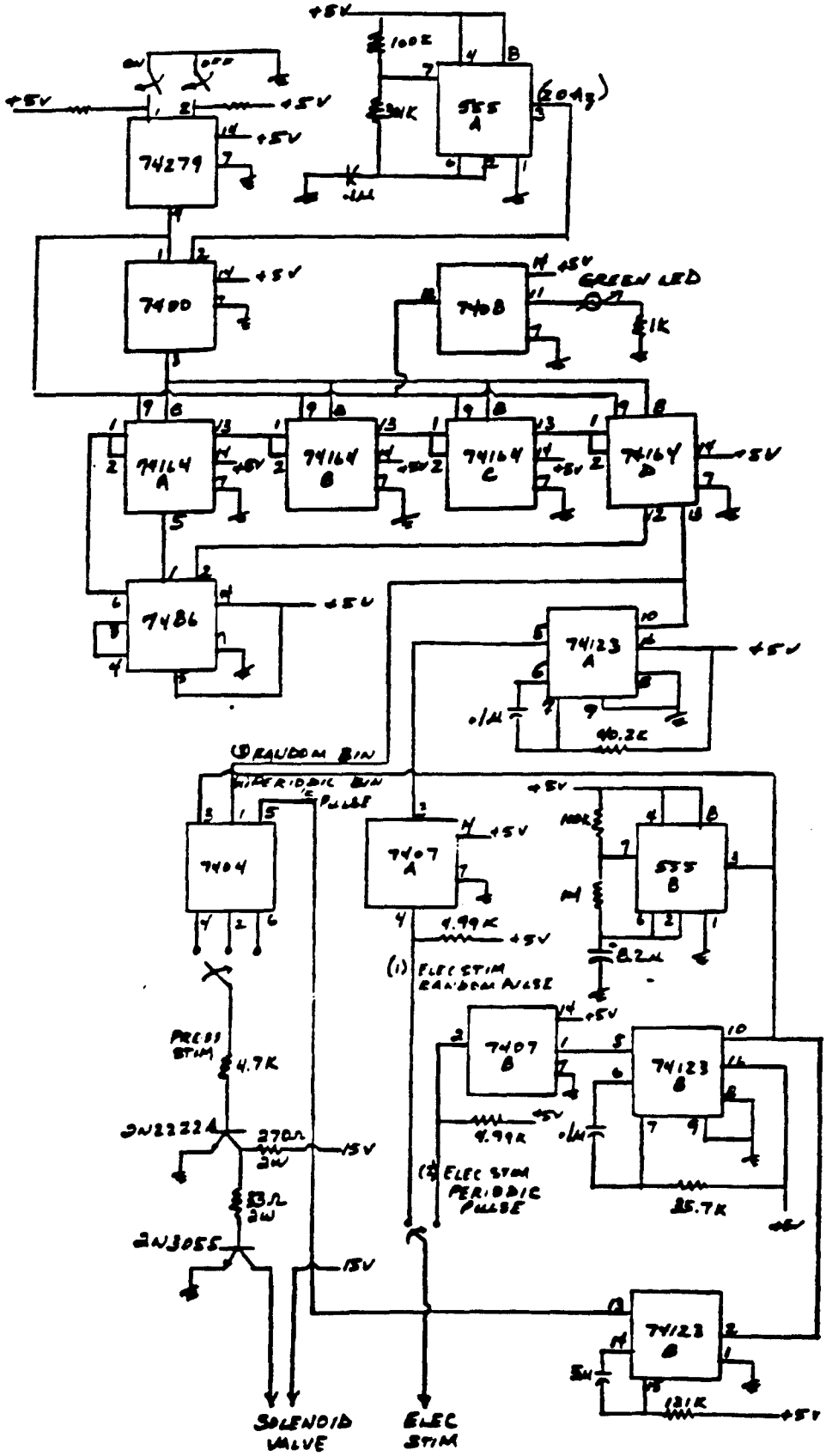


Figure 3
Electrical Circuit

During each period of data acquisition, the pseudo-random binary stimulus was sustained for three minutes . In the first run, the two reservoirs were positioned to provide a mean carotid pressure of 40mm Hg, and the random binary stimulus fluctuated about the mean by ± 20mm Hg. In the second run, the mean was raised to 70mm Hg, and the binary fluctuations were first set at ± 20mm Hg and then increased to ± 50mm Hg. In the last run, the mean was set at 100mm Hg and the binary fluctuations were ± 20mm Hg.

RESULTS

Figure 4 shows a plot of the carotid pressure (i.e. stimulus) during the first run on the third animal, when the mean was set at 40mm Hg, and the random binary fluctuations were + 20mm Hg. Figure 5 is a log-log plot of the power spectrum of this signal. The pressure stimulus appears to be flat over the bandwidth, DC to .75Hz.

Figure 6 presents a plot of the pressure measured in the right atrium (i.e. response) during the same period of time as Figure 4. The low frequency periodic component (approximately .2 Hz) is related to respiration, and the high frequency component (approximately 3 Hz) is the typical systolic/diastolic waveform. Figure 7 is a plot of the EKG interval during this same period.

Figure 8 displays the first-order Wiener kernels (see Appendix A for formula) for each of the animals at the same stimulus mean level of 70mm Hg. These transfer functions were calculated by defining the pressure in the carotid sinus as the stimulus and the pressure in the right atrium as the response. The response data was low-pass filtered (corner=1.5 Hz) to remove the systolic/diastolic fluctuations-- without attenuating the lower frequency, respiratory-dependent fluctuations. Figure 9 shows the comparable plots of the first-order kernels when the aortic pressure is considered to be the response.

Under certain conditions (see Appendix A for details), the above first-order Wiener kernels predict the response to a pulse stimulus occurring at the origin of each plot. In all of the kernels presented in Figures 8 and 9, the predicted responses occur more than 50 seconds after the hypothetical pulse stimulus. This unexpected long time-constant is undoubtedly the result of reduced pO2 in the carotid sinus. It is also surprising that the predicted pulse responses are in the wrong direction; a positive pulse at the carotid sinus causes vasoconstriction. Since the predicted negative pulse response is nearly the same, the carotid baroreceptors may be only responding to the negative-going edge of the stimulus.

Figure 10 shows how the first-order Wiener kernel changes when the mean pressure level in the carotid sinus is changed. At the higher mean levels, the latencies appear to be smaller. Hence, the longest time-constants occur at low pO2 and low mean carotid pressures, which is a common situation in prolonged high-G combat. Figure 11 presents these same Wiener kernels in a different perspective-- the peak values are shown for each respiratory period.

The obvious next step in the analytical progression is to calculate the average responses to the pulse and step stimuli which were used-- comparing the measured pulse/step responses with the predicted responses obtained from the white-noise data. Unfortunately, we were not expecting latencies as long as the above white-noise data indicates; therefore, the pulse and step stimuli were presented with short intervals-- an order of magnitude smaller than we needed. In future experiments it will be important to use aperiodic pulse and step stimuli with much longer intervals between stimuli.

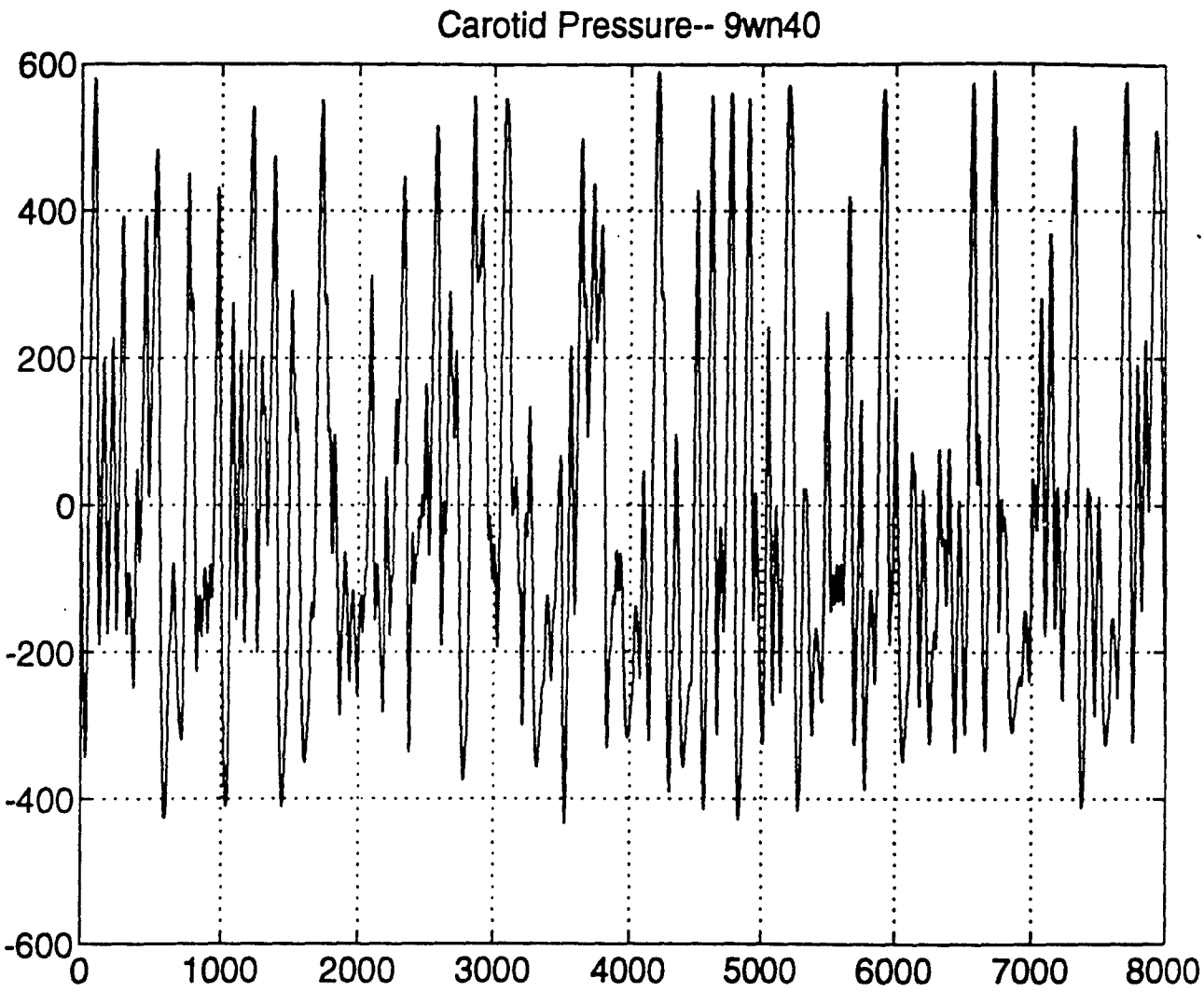


Figure 4
Carotid pressure measured during white-noise protocol

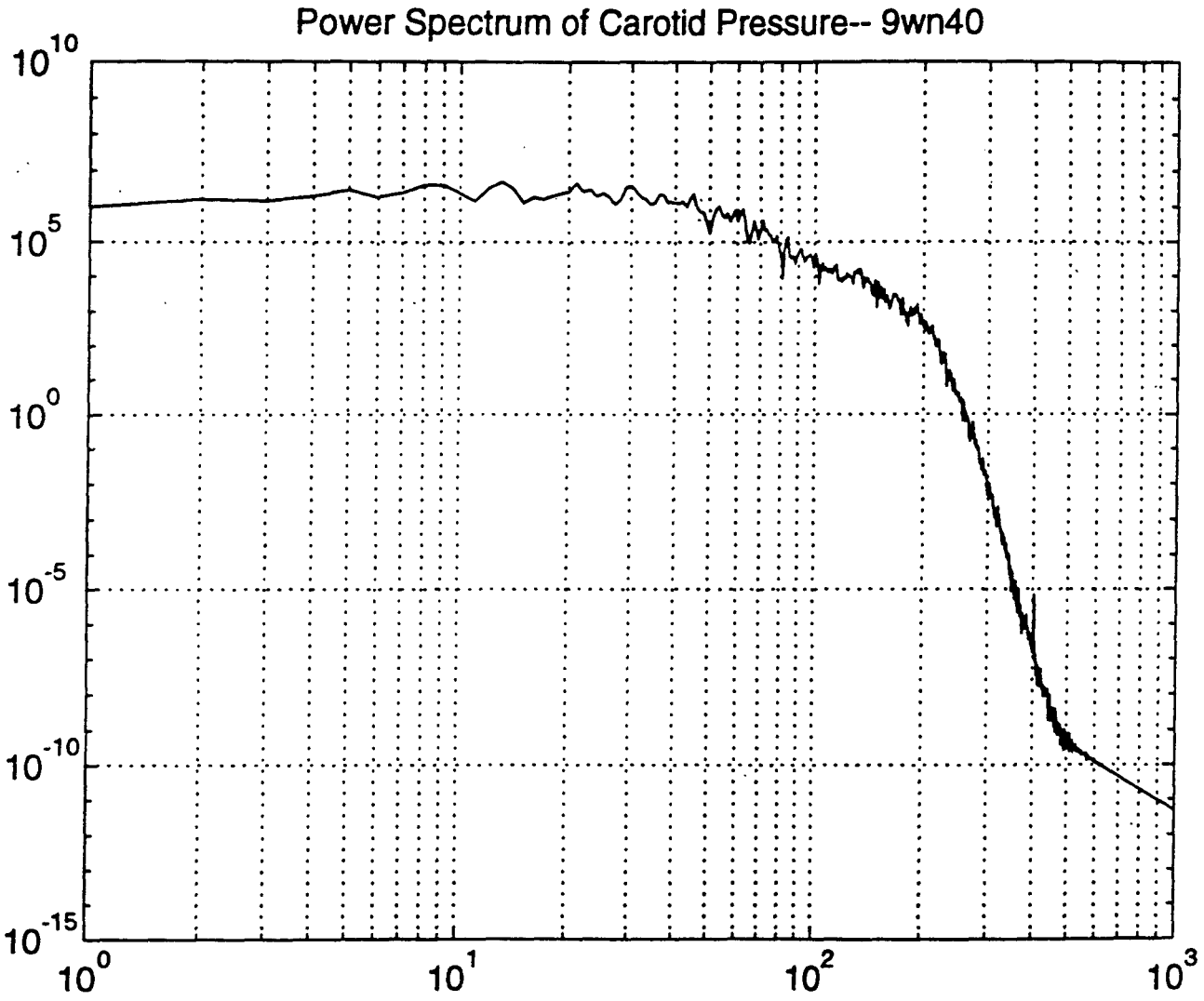


Figure 5
Power spectrum of white-noise stimulus

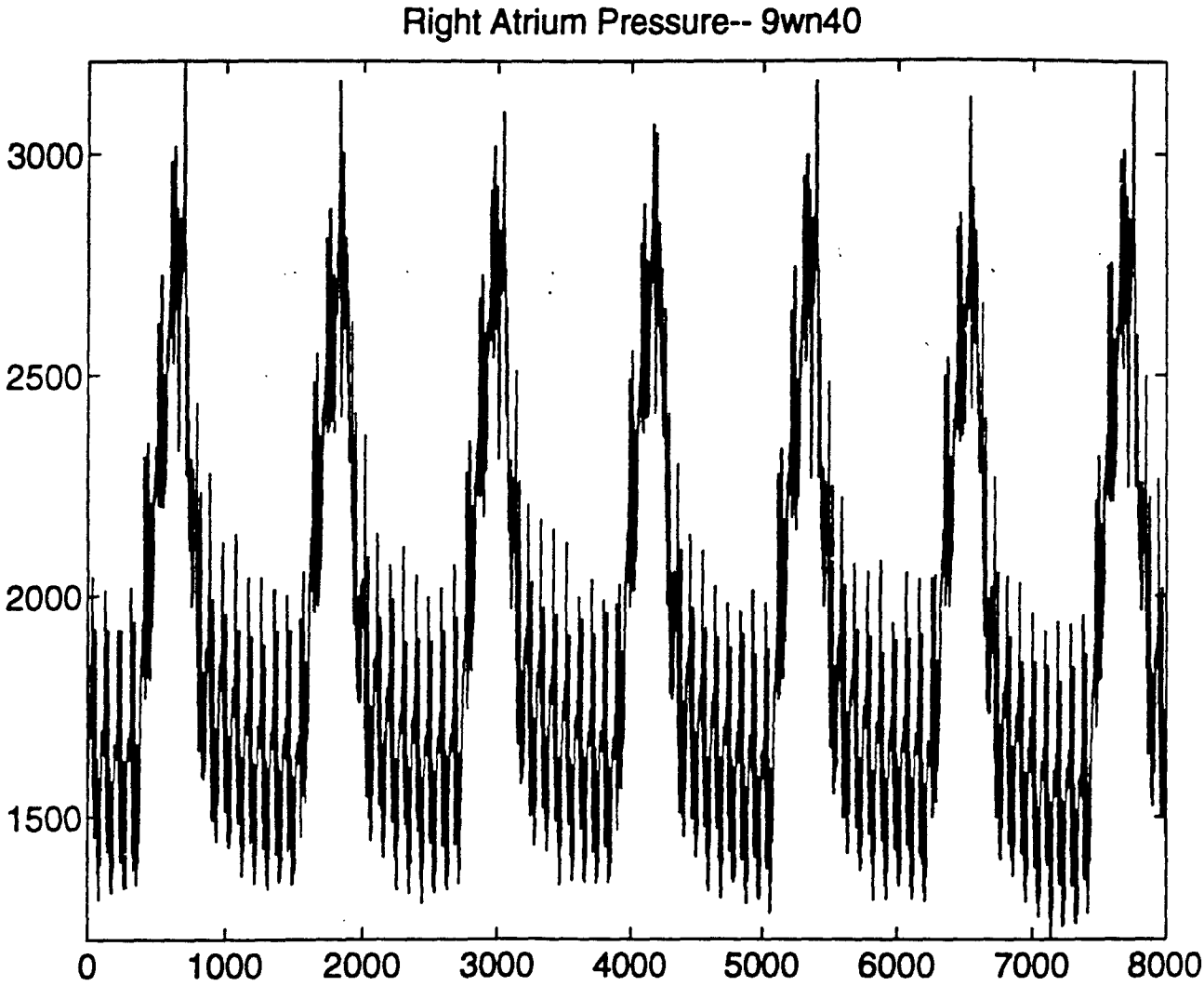


Figure 6
Pressure measured in Right Atrium during white-noise experiment

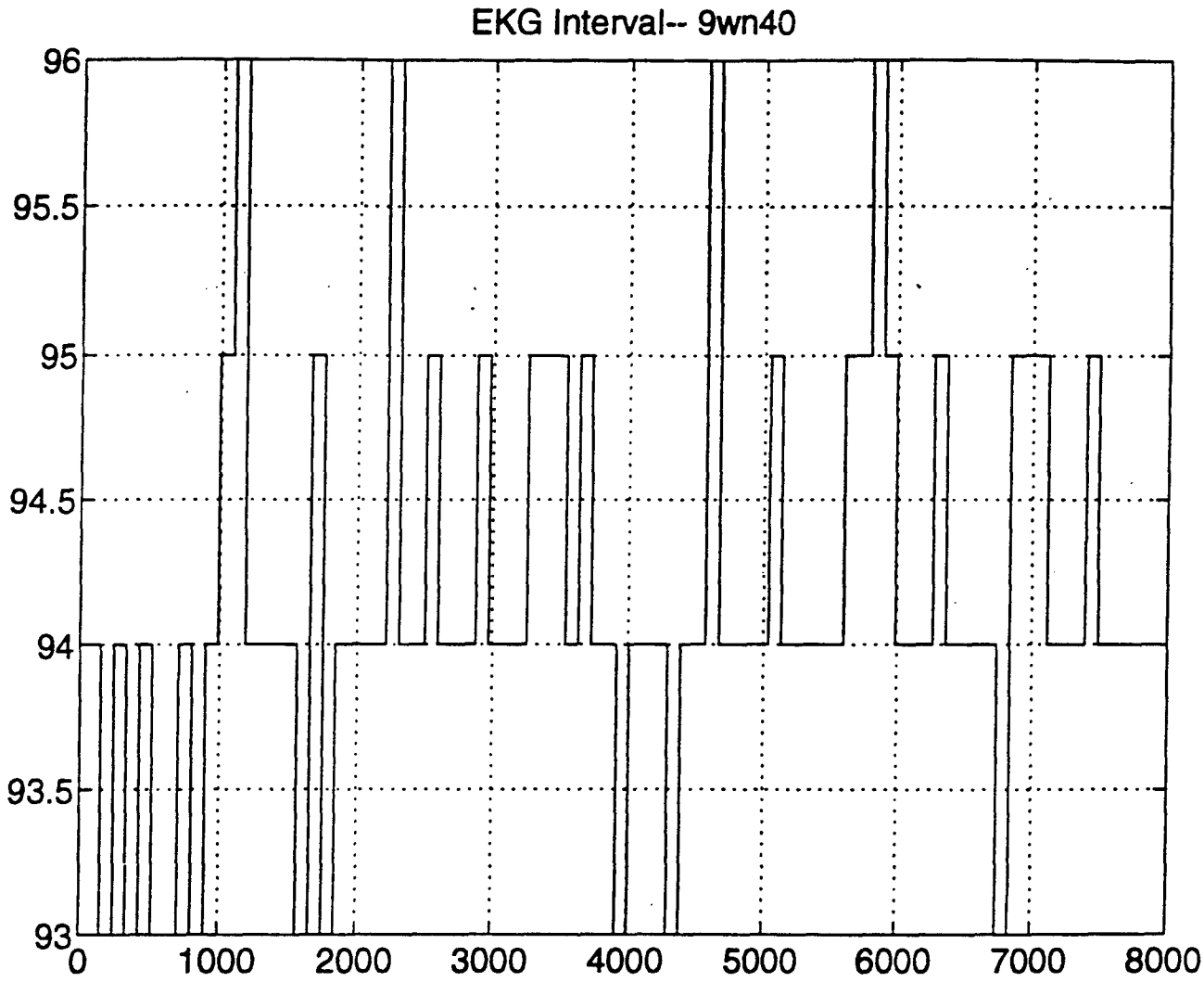


Figure 7
EKG interval measured during white-noise experiment

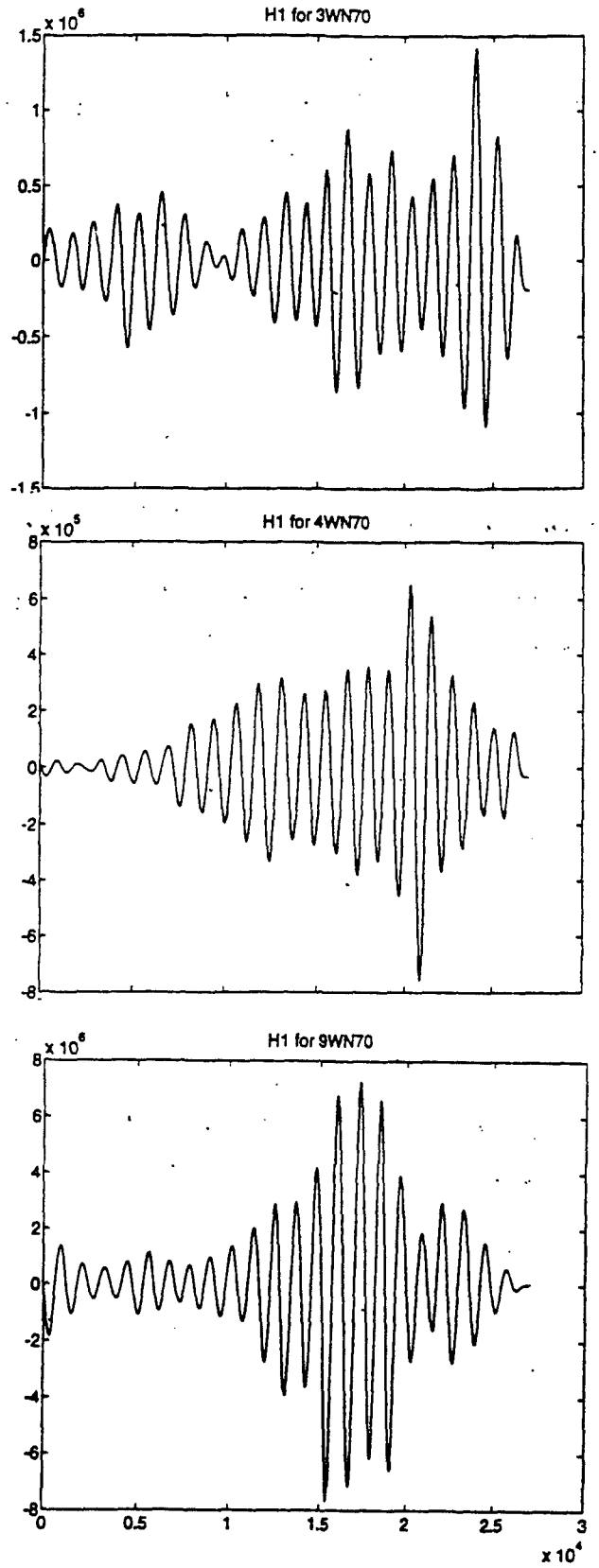


Figure 8
First-order Wiener kernels for 3 different baboons
at mean stimulus level of 70 mm Hg--
response measured at Right Atrium

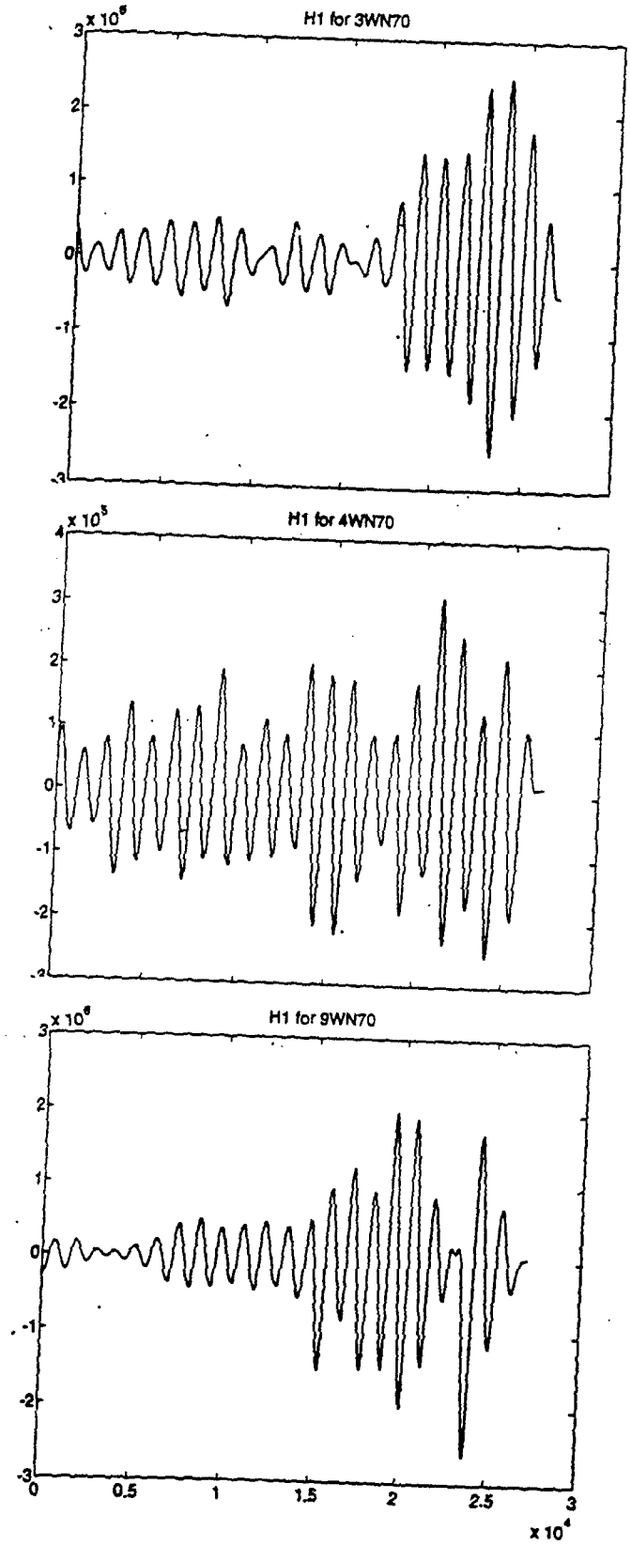


Figure 9
First-order Wiener kernels for 3 different baboons
at mean stimulus level of 70 mm Hg--
response measured at Aorta

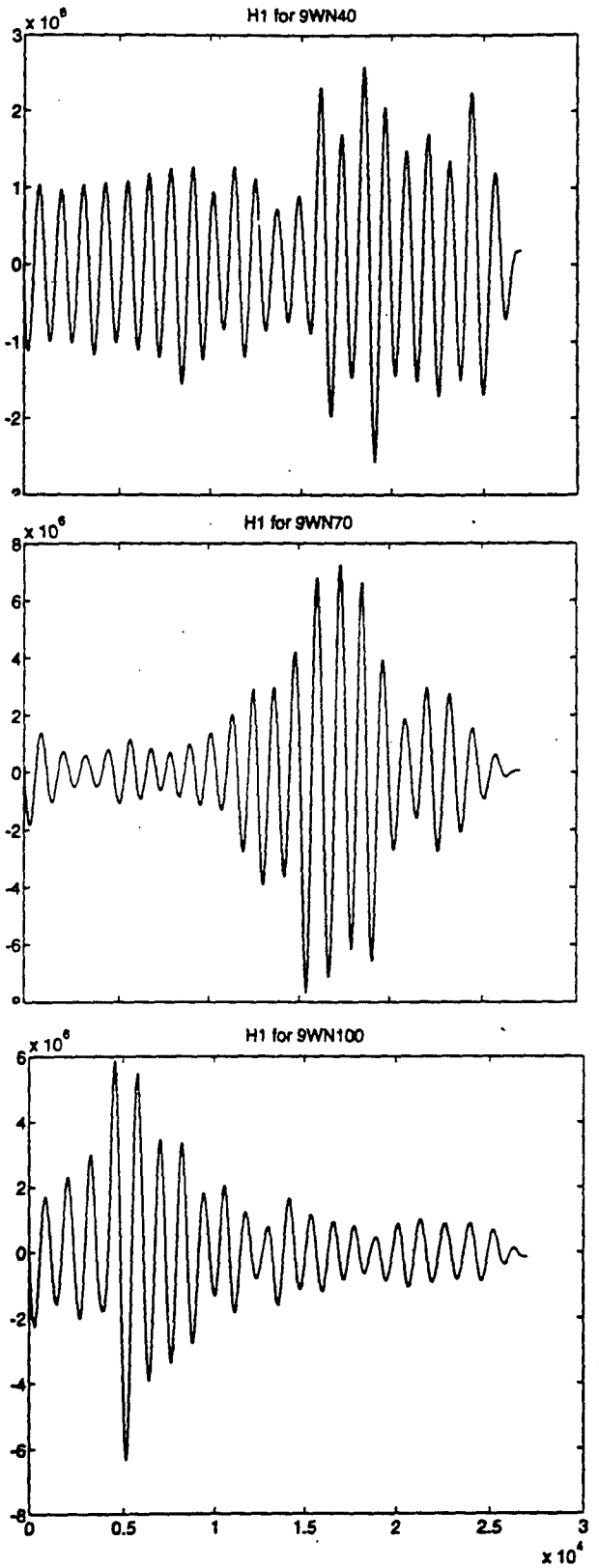


Figure 10
First-order Wiener kernels at 3 different mean levels--
40 mm Hg, 70 mm Hg and 100 mm Hg

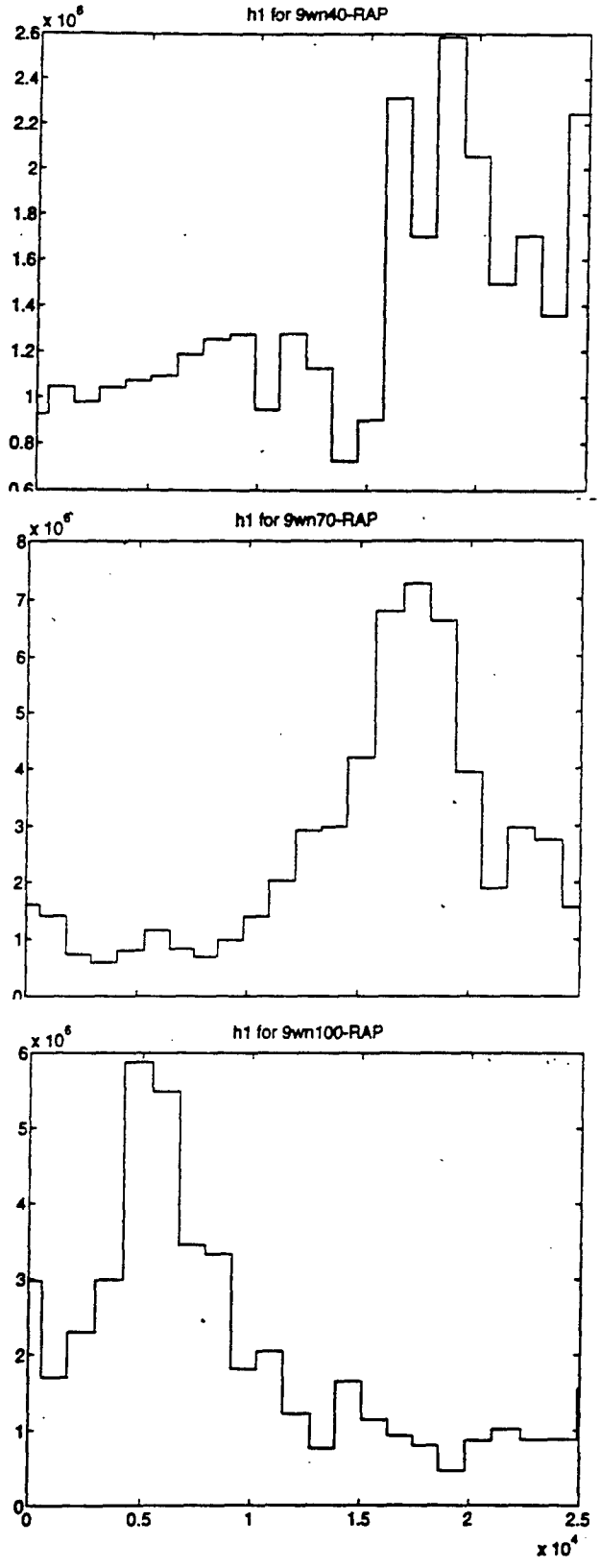


Figure 11
First-order Wiener kernels at 40 mm Hg, 70 mm Hg and 100 mm Hg--
plotting peak pressures during each respiratory period

References

Andresen, M J Krauhs and A Brown. 1978. Relationship of aortic wall and baroreceptor properties during development in normotensive and spontaneous hypertensive rats. Circulation Research. 43:728-738.

Brown, A. 1980. Receptors under pressure. An update on baroreceptors. Circulation Research. 46:1-10.

Fricker, S and J Sanders. 1975. A new method of cone electroretinography: the rapid random flash response. Investigative Ophthalmology. 14:131-137.

Koblasz, A, J Rae, M Correia and M Ni. 1980. Wiener kernels and frequency response functions of the human retina. IEEE Biomedical Engineering. 27:68-75.

Marmarelis, P and K Naka. 1973. Nonlinear analysis and synthesis of receptive field responses in the catfish retina. J Neurophysiology. 36:619-653.

Marmarelis, V. 1976. Identification of nonlinear systems through quasi-white test signals. Caltech PhD Thesis.

Masaru, S, T Imaizumi, K Sunagawa, Y Hirooka, K Todaka, A Takeshita and M Kakamura. 1990. American Journal of Physiology. H887-H895.

Yasui, S and A Koblasz. 1984. Transmission of array data by multi-variate convolution and cross-correlation using white-noise reference signal. International Journal of Systems Science. 15:525-541.

APPENDIX

White-noise analysis of biological systems

Arthur Koblasz, Ph.D.

Departments of Ophthalmology & Otolaryngology, The University of Texas Medical Branch, Galveston, Texas 77550

The linear and nonlinear operations of a biological system can be represented by a set of functions called Wiener kernels. This type of analysis is becoming increasingly important in the field of biological systems analysis. This paper reviews the theoretical and practical aspects of testing a biological system with white-noise and provides a guide for interpreting Wiener kernels which result from such studies. A study of the human electroretinogram is presented as an example.

Introduction

Researchers have been studying biological systems for centuries. Usually the unknown system is analysed by controlling an input to the system and observing the response. Both linear and nonlinear analyses are possible using test inputs of pulses, square waves, or sinewaves. A linear system can be entirely characterized by its unit-impulse response. Furthermore, once the unit-impulse response of a linear system is known, the response to any input can be predicted using the convolution integral. A nonlinear system can also be studied using a stimulus which is a sequence of impulses. However, the impulse response is likely to be a nonlinear function of the impulse amplitude. Furthermore, the response to one input stimulus may change the characteristics of the system such that the next impulse stimulus elicits a different response.

Recently, a new testing procedure which utilises band limited white-noise inputs [1, 2, 3] has become popular. For time-invariant nonlinear systems, Wiener [1] proposed the following functional expansion to define the response y(t) for a zero-mean, Gaussian, white-noise input x(t):

$$y(t) = \sum_{n=0}^{\infty} G_n(h_n;x)$$

where the first three terms of the expansion are

$$G_0 = h_0$$

$$G_1 = \int_0^{\infty} h_1(\tau)x(t-\tau)d\tau$$

$$G_2 = \int_0^{\infty}\int_0^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 - P\int_0^{\infty} h_2(\tau_1,\tau_1)d\tau_1$$

The term G, (hi; x) is like the convolution integral of linear system theory. The functions $h_n(\tau_1,\cdots,\tau_n)$ called Wiener kernels and P refers to the constant power spectral density of the white-noise stimulus. Once the Wiener kernels are known, the response y(t) to any input x(t) can be predicted. Hence, the characterization problem becomes one of estimating the Wiener kernels for the unknown system.

Lee and Schetzen [2] derived the following equation to estimate the Wiener kernels in time of multiple-order cross-correlations between the white-noise stimulus and the measured response:

$$h_n(\tau_1,\tau_2,\cdots,\tau_n) = \frac{1}{n!P^n} E\{[y(t) - \sum_{k=1}^{N} G_{n-2k}(h_{n-2k};x)]x(t-\tau_1)\cdots x(t-\tau_n)\} \quad (3)$$

where E denotes expected value and N is the largest integer not exceeding n/2.

That is,

$$h_0 = E[x(t)]$$

$$h_1(\tau) = \frac{1}{P} E[y(t)x(t-\tau)]$$

$$h_2(\tau_1,\tau_2) = \frac{1}{2P^2} E\{[y(t)-h_0]x(t-\tau_1)x(t-\tau_2)\} \quad (4)$$

For example, the double cross-correlation which defines $h_2(\tau_1,\tau_2)$ is obtained by simply multiplying the response y at each time t by the preceding stimulus amplitude at time t-T1, and then again multiplying this product by the stimulus amplitude at time t-T2. The average value of this double product for all t indicates the value of the second-order kernel $h_2(\tau_1,\tau_2)$ at $\tau_1 = T_1$ and $\tau_2 = T_2$.

Defining the response in terms of the Wiener expansion becomes theoretical very simple. However, the characterization is computationally practical only when the higher-order kernels are relatively small, in other words, when the Wiener expansion can be truncated after some low-order term. Although the expansion is mean-square convergent for white-noise inputs, it may not converge for deterministic inputs. Hence, it is best to interpret the Wiener kernels in the context of a white-noise stimulus.

The general interpretation of the Wiener kernels is discussed in reference [4]. If the Wiener expansion can be truncated after the second-order term, then the kernels can be more simply interpreted in terms of impulse responses as described in reference [3]. That is, suppose an impulse of amplitude m_1 occurs at time T_1, as an input $x(t) = m_1\delta(t-T_1)$, where $\delta(t-T_1)$ is a dirac delta function.

Then from equations (1) and (2), the predicted response is

$$y_1(t) = C_1 + m_1h_1(t-T_1) + m_1^2h_2(t-T_1, t-T_1) \quad (5)$$

where the constant

$$C_1 = h_0 - m_1^2\int_0^{\infty} h_2(\tau,\tau)d\tau \quad (6)$$

Notice if the system were linear, the second- and higher-order kernels would be zero. Then changes in the magnitude of the impulse stimulus would only produce changes in the amplitude of the predicted response, i.e., $y_1(t) = C + m_1h_1(t-T_1)$. However, for a nonlinear system, because of the different scale factors for the first- and second-order kernels, the shape of the impulse response can be a function of the amplitude of the input impulse.

If we next let the input be an impulse at time T_2 of amplitude m_2, then

$$y_2(t) = C_2 + m_2h_1(t-T_2) + m_2^2h_2(t-T_2, t-T_2) \quad (7)$$

Finally, if the input is a combination of the pulse at T_1 and the pulse at T_2, then the double-impulse predicted response is

$$y_3(t) = y_1(t) + y_2(t) + 2m_1m_2h_2(t-T_1, t-T_2) \quad (8)$$

The term $2m_1m_2h_2(t-T_1, t-T_2)$ is a measure of the error in assuming the superposition of the single-impulse responses $y_1(t)$ and $y_2(t)$ referred to above. This error is usually dependent on the spacing between the two pulses (T_2-T_1).

Hence, we could determine the second-order Wiener kernel for this special case by simply measuring the single-impulse responses for different amplitude impulses and the double-impulse responses for different spacing between impulses. The second-order kernel could be then tabulated as in Figure 1a or drawn as a contour map by connecting points of equal value as in Figure 1b. Finally, the axes of the contour map could be changed as in Figure 1c, which is the format used in subsequent figures. On the other hand, it is much simpler to estimate the Wiener kernels using the cross-correlation formula of equation (3).

Portions of this work were supported by U.S. Public Health Grant EY 01526.

Pitfalls in the practical application

There are several subtle assumptions associated with the Wiener expansion which at times make the practical application of the theory very difficult. First, it is impossible to generate a perfectly Gaussian and white stimulus. The bandwidth of the white-noise signal must be finite and is frequently not exactly flat for any range of frequencies. Furthermore, the stimulus characteristics may change significantly between experiments; therefore, the same random stimulus should be repeated for all experiments in order to insure that differences in the kernels are not related to differences in stimulus characteristics. Then it is appropriate to consider the characterization problem as depicted in Figure 2. If the bandwidth of the white-noise stimulus encloses the characteristic bandwidth [3] of the unknown system and if the bandpass characteristics are known, then the exact transfer characteristics of the unknown system can be deconvolved from the kernels defining the total system, which is obvious from references [5] and [6]. Fortunately, the deconvolution is not necessary for empirical studies, in other words, when it is sufficient to identify structural changes in the kernels for the total system in Figure 2 resulting from functional changes of the unknown biological system.

Another pitfall in the practical application of the theory involves the assumption of time-invariance. Many biological systems undergo systematic changes in characteristics with respect to time (for example, retinal adaptation), which are actually deterministic and time-invariant. Yet for testing periods comparable in time to the time constant of the biological changes, the system will be seen to be time-variant. In such cases the kernels estimated by equation (3) will reflect the time average of the system characteristics during the testing period, which can be useful information. However, the Wiener expansion will be invalid.

Another mistake possible in implementing the theory is related to the common assumption that third- and higher-order kernels are insignificant. The functional interpretation of the kernels is highly dependent upon the amplitudes of the higher-order terms in the Wiener expansion [4]. A simple test for the significance of higher-order terms in the Wiener expansion is to use the lower-order terms to predict the response to the same white-noise stimulus originally used for testing the system [3]. If the predicted response is nearly the same as the measured response, then it is safe to assume higher-order terms are insignificant for white-noise inputs.

This test for nonlinearities is not possible when the signal-to-noise ratio is very small, which is frequently the case for biological measurements. Another (partial) test more appropriate for noisy systems is to measure the response to discrete sinewave stimuli with a range of frequencies covering the bandwidth of the white-noise stimulus. The amplitude of the n_{th}- harmonic of the sinewave response indicates the magnitude of the main diagonal of the n-dimensional Fourier Transform of the n_{th}-order kernel, but does not assess the amplitudes of off-diagonal components (deduced from equation (6) of reference [5]).

Finally, the signal-to-noise ratio itself can be a major problem in implementing the theory. The white-noise stimulus tests the system with relatively small modulations about the mean of the stimulus. This permits testing at physiological levels, but since the stimulus is small, the response is also small; consequently, the signal-to-noise ratio of the measured response can become very small. However, if the noise is uncorrelated with the stimulus, very long experiments can be performed to effectively average out the uncorrelated noise [4]. Alternatively, a different type of random stimulus which has a

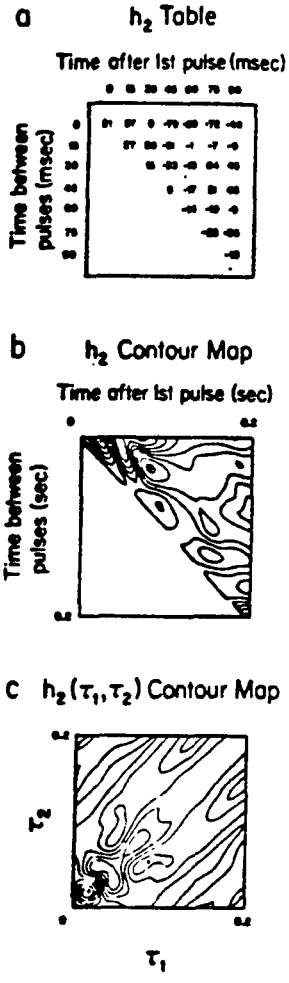


Figure 1. (a) table representing second-order Wiener kernel derived from two inpulse experiments, i.e., the measured response is the superposition of single pulse responses plus values shown in the table along a horizontal line specified by the spacing between the pulses; (b) contour map obtained by connected points of equal value in matrix similar to that presented in Figure 1(a); (c) contour map resulting when axes of Figure 1(b) are changed.

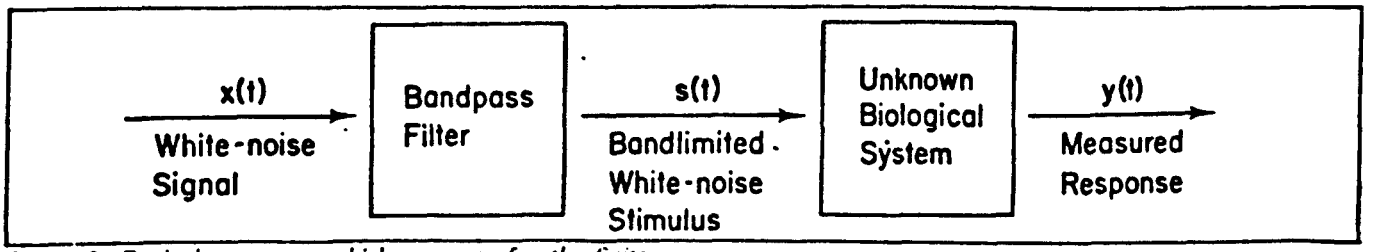


Figure 2. Equivalent system which accounts for the finite bandwidth of the stimulus.

non-Gaussian amplitude distribution, such as a pseudo-random binary sequence (PRBS), can be used. In general, such alternative stimuli produce larger response signals, but the resulting kernels are more difficult to interpret.

Example application

We have analyzed the human electroretinogram (ERG) response to white-noise modulations in light intensity [7]. The retinal response to a light stimulus causes local changes in ion concentrations which are propagated to the front corneal surface of the eye via the electrolytic media inside the eye. An electrode positioned near the front corneal surface detects a spatial average of this retinal response and is called the ERG. The characteristics of the ERG are known to change systematically for certain retinal diseases. For example, the scotopic components of the ERG flash response are significantly attenuated by retinitis pigmentosa [7]. It is possible that the Wiener kernels will empirically reflect these functional changes more precisely. The following experiment produced first- and second-order Wiener kernels for assumed normal subjects with the same retinal adaptation states.

Equipment

There are several ways to generate a Gaussian, bandlimited, white-noise voltage signal. We used a design which applied the random noise signal generated by a reverse-biased zener diode to a digital circuit designed to produce a random-period binary signal [7]. When this is appropriately high- and low-pass filtered, it produces the desired continuous random signal [8, 9].

Figure 3 shows the optical bench used to provide white-noise stimuli of different mean levels. The pen motor rotated a 3M louver filter. The transmission through this filter was proportional to the sine of the angle between the plane of the filter and the incident light. We used maximum louver excursions of only six degrees, and within this range the transmission can be assumed proportional to the angle with an error less than .2%; therefore, white-noise current signals driving the pen motor produced the desired modulations in light intensity. The modulated light was next passed through neutral density filters to permit step changes in the mean level of the random modulations. The light was then optically condensed onto one end of a coherent fibre-optic bundle and transmitted to the Maxwellian optical assembly also shown in Figure 3. This assembly of lenses created a 45-degree Maxwellian view [7] of the light stimulus with a superimposed cross-hair target located virtually at infinity.

The ERG response was measured by the cup electrode [10] mounted at the top of the Maxwellian optical assembly. The cup was first filled with an artificial tear solution, then each subject submerged his eye into the solution with an orientation similar to that required when looking into a conventional microscope. An annular-shaped Ag/AgCl electrode suspended in the solution detected the ERG signal, and a second cup electrode measured the ERG of the unstimulated eye for use as a reference signal.

During an experiment, the stimulus and response signals were sampled at 5 ms intervals (in other words, 100 Hz Nyquist folding frequency) by a 12-bit A/D converter, and multiplexed onto a digital tape. Later the data sets were digitally filtered and analyzed using a PDP-11/45 computer system.

Experimental protocol

A normal adult male subject was used for the white-noise ERG test. The subject was initially exposed to ambient room illumination of approximately 3 cd/m² for a period of 15 minutes. After this preliminary adaptation period, the subject submerged each eye into the solution contained in the cup electrodes and adjusted himself to attain a 45-degree Maxwellian view of the white-noise stimulus. As discussed above, the stimulus was super-imposed onto a cross-hair target located virtually at infinity and projected to the right eye; the ERG response was measured from the stimulated right eye with respect to a reference signal obtained from the unstimulated left eye. The mean of the random stimulus was set at 85 trolands (td.) for the first three minutes, then immediately changed to 850 td. for a second three-minute period, and finally changed to 8500 td. for the last three minutes.

Results and Discussion

Figure 4 presents the mean first-order Wiener kernal ± one standard deviation of the mean, for an ensemble of four 30-

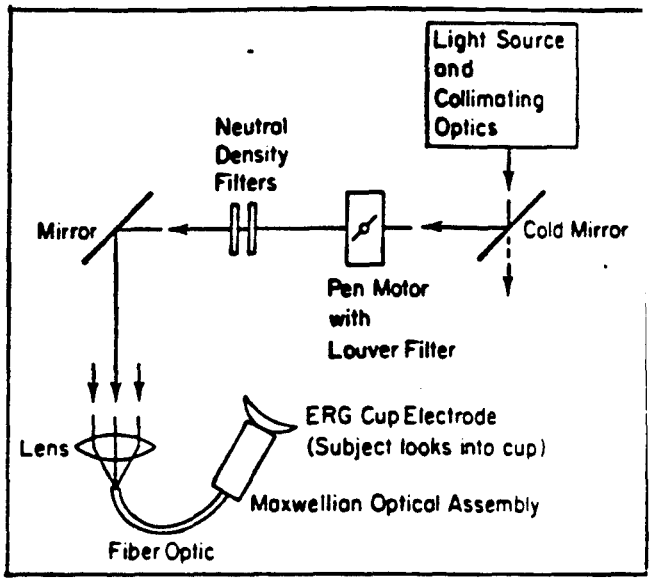


Figure 3. Electro-optical bench and ERG cup electrode.

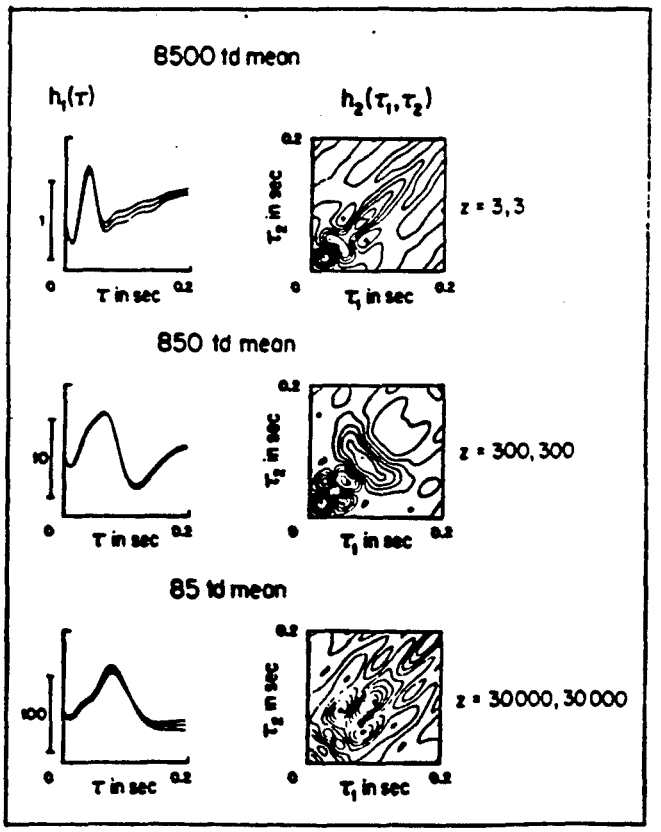


Figure 4. Mean $h_1(T)$ and mean ± standard deviation of the mean along with mean $h_2(T_1, T_2)$ for different average levels of white-noise light stimuli; the numbers alongside the kernel estimates indicate the relative amplitudes of the kernels.

second segments of data at each intensity level. A contour plot of the mean second-order kernel for each level is also shown for this subject.

The standard deviation of the mean of each first-order kernel estimate is relatively small, signifying that only small variations occur between kernels estimated for the same stimulus conditions. Yet both the first- and second-order kernels change drastically for step changes in the stimulus mean. Hence, we can infer that the kernels are reliable and in some way reflect actual ERG characteristics. However, the exact transfer characteristics of the ERG are undoubtedly distorted as discussed in the context of Figure 2.

It is possible to relate some of the *observed* changes in the kernels with *changes* in retinal function. For example, the maximum amplitudes of the kernels decrease with increasing stimulus mean, reflecting a general decrease in retinal sensitivity. The latencies of the distinct components also seem to decrease with increasing stimulus mean. Hopefully, such changes in the kernels can be empirically linked with retinal disease processes.

Hence, the first- and second-order Wiener kernels can accurately characterize both linear and nonlinear aspects of the ERG system. Furthermore, the representation is in a concise and meaningful form. Presumably many biological systems can be studied using the analytical procedures described for the ERG, and the results may offer better resolution of the mechanisms which control the system.

ACKNOWLEDGEMENTS

The experimental work presented in this manuscript was performed in the laboratories of Dr. Derek H. Fender at the California Institute of Technology. I am very grateful for the guidance and help availed the by Dr. Fender and his laboratory assistants. I am also very grateful f the many theoretical insights generously rendered by Dr. Syozo Yas of the University of Texas Medical Branch.

REFERENCES

[1] Wiener, N. *Nonlinear Problems in Random Theory* (Wiley, Ne York, 1958)

[2] Lee, Y. W. and Schetzen, M. (1965). Measurement of the Wien kernels of a nonlinear system by crosscorrelation, *Int. J. Contr* Vol. 2, pp. 237-254.

[3] Marmarelis, P. and Naka, K. I. (1972). White-noise analysis of neuron chain: An application of the Wiener theory, *Scienc* Vol. 175, pp. 1276-1278.

[4] Yasui, S. Stochastic functional Fourier series, Volterra series an nonlinear systems analysis, *IEEE Trans on Automatic Contro* Accepted for publication 1978.

[5] French A. and Butz, E. (1973). Measuring the Wiener kernels of nonlinear system using the Fast Fourier Transform algorithm *J. Control*, Vol. 3, pp. 529-539.

[6] Chan R. *Spatial Dynamics of Vertebrate Retinal Neurons.* (Ph.D Thesis, California Institute of Technology, 1976).

[7] Koblasz, A. *Nonlinear Analysis of the Human Electroretinogram* (Ph. D. Thesis, California Institute of Technology, 1976).

[8] Mayo, J. E., Pawula, R. F. and Rice, S. O. (1973). On a non-linea problem involving RC noise, *IEEE Trans. on Info. Theory*, Vo 19, pp. 128-135.

[9] Pawula, R. and Tsai, A. (1969). Theoretical and experimenta results for the distribution of a certain nonlinear functional of th Orstein- Uhlenbeck process, *IEEE Trans. on Info. Theory*, Vo 15, pp. 532-535.

[10] Koblasz, A. Cup Electrode for Human ERG, *Investigativ Ophthalmology*, Vol. 17, pp. 575-577.

Digital ultrasonic imaging with microprocessor manipulation

P. T. Ostro, B.Sc.

Dept. of Medical Physics, Queen's Medical Centre, University Hospital, Nottingham.

L. Harrison, B.Eng., Ph.D.

Dept. of Electrical and Electronic Engineering, University of Nottingham, Nottingham NG7 2RD

The author describes the development of a digital scan converter system which accepts ultrasonic B-scans in "oscilloscope" format and displays them in video format. The system involves the use of a microprocesor, the funciton of which is outlined in the paper. A brief comparison with a system employing a scan converter tube is given.

Introduction

The use of ultrasound in medical imaging is now wide-spread. The technique depends on a pulse-echo principle, whereby a pulse of high frequency sound is emitted into the body and any echoes that are produced are detected and displayed. The simplest method of viewing the echoes consists of an amplitude versus time display, and is called an A-scan. The familiar, two-dimensional B-scan is obtained by combining the A-scan with information about the position and orientation of the emitting transducer at the time the pulse was generated. The necessary spatial information is collected and processed to draw a line on an oscilloscope which mimics the path travelled by the ultrasound within the body. The intensity of this line is controlled by the amplitude of the A-scan.

Most B-mode scanners rely on an operator to manually scan the transducer over the patient. Since this is a process which takes a long time when compared to the persistance of a normal cathode ray tube (C.R.T.), resort must be made to a storage type tube in order to build up the image. The earliest versions of these were capable only of displaying two tones — black and white. The A-scan was therefore electronically thresholded and the images produced were called 'bi-stable'.

ADAPTIVE CONTROL ARCHITECTURE FOR TELEOPERATED FREFLEX SYSTEM

Antti J. Koivo
Professor

and

Manish Goel
Graduate Student

School of Electrical Engineering
Purdue University
1285 Electrical Engineering Building
West Lafayette, IN 47907-1285

Final Report for:
Summer Research Extension Program

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

January 1993 - May 1994

ADAPTIVE CONTROL ARCHITECTURE FOR TELEOPERATED FREFLEX SYSTEM

A.J. Koivo
Professor

M. Goel
Graduate Student

School of Electrical Engineering
Purdue University

Abstract

The need for achieving human functionality in hazardous and/or remote environments with robotic manipulators has given rise to teleoperated systems. In this class of manipulator systems, a human operator controls through a master manipulator a slave manipulator at a remote site. The system would allow the operator to comfortably perform a remote task, and free him/her of any risks or envisioned difficulties that would be associated with the direct manipulation of he/she were at the remote site.

The realism of interaction with a remote environment is achieved by using bilateral teleoperation, in which along with the forward transmission of velocity commands from the master to the slave, the forces sensed by the slave in performing the task are reflected to the operator. Incorporating force reflection is a step towards achieving *telepresence,* the feel of being present and directly performing the task at the remote site.

An adaptive control scheme is proposed for bilateral teleoperation. A teleoperator system is modeled, and an ideal response is explicitly defined. If the dynamical model is not known, and/or the environmental conditions are unpredictable or timevarying, the performance of the system can be improved by using an adaptive controller. In this work, an adaptive two-port network is used as the basis to design an adaptive controller for the teleoperated system. Computer simulations are used to demonstrate the performance of the designed system.

1. INTRODUCTION

1.1 Background

The need for achieving human functionality in hazardous and/or remote environments with robotic manipulators has given rise to teleoperated systems. In this class of manipulator systems. a human operator controls through a master manipulator a slave manipulator at a remote site. The system would allow the operator to comfortably perform a remote task, and free him/her of any risks or anticipated difficulties that would be associated with the direct manipulation if he/she were at the remote site. Teleoperation has grown by leaps and bounds since its inception, and has developed into a very promising branch of robotics. Teleoperator systems are used in performing risky tasks in space or outside airplanes, in the handling of radioactive materials, in mining, construction, fire-fighting and in deep sea applications. Some of the other more demanding applications of teleoperators include microsurgery, in which a teleoperated robotic manipulator can assist a surgeon. With advances in the field, applications are bound to multiply. One of the primary goals of teleoperation is to achieve telepresence for the operator, the feel of being present while performing a task from a remote site.

1.2 Bilateral Teleoperation

Teleoperation is greatly enhanced, if along with the forward transmission of commands from the master manipulator to the slave manipulator a feedback from the slave to the master is incorporated in the system. Kinesthetic and vision information are typical in the feedback employed in such applications. Reflecting slave contact

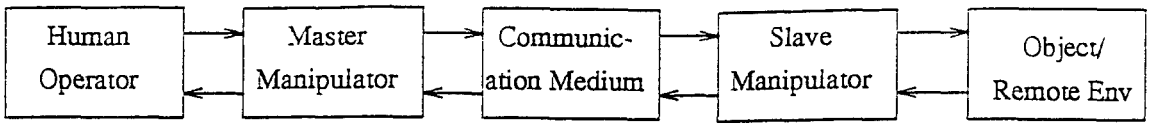


Figure 1.1 A Bilateral Teleoperator System

forces to the operator usually improves the task performance, for example, by decreasing the task completion time [10], and increasing the realism of interaction with the environment. When teleoperation utilizes sensory feedback involving exchange of information at two distinct ports, the operation is said to be bilateral. Figure 1.1 shows the block diagram of a bilateral teleoperator system.

1.3 Definition of an Ideal Teleoperator System

The definition of an ideal teleoperator system is application specific. In some applications the teleoperator is required to provide a direct extension of the human operator's manipulation capabilities to a remote site. In some others, the operator's forces and/or position may be scaled to perform a remote task. Several definitions of an ideal teleoperator performance have been presented in literature. The ideal system can be considered as the one [4, 8] that provideds a response to the human operator as if he/she were linked to the remote task by a weightless, virtual rod of infinite stiffness. Alternatively, the ideal response may be defined [22] to be such that the position and force of the slave arm are identical to those of the master arm, irrespective of the object dynamics. Viewing the system from a human factors perspective, it is suggested in [21] that an ideal system is such that it reduces the operator's fatigue in exerting a constant force by incorporating variable force feedback gain. A function is proposed in [11] to specify the relation between the dynamic behaviors of the master and slave manipulators; it allows position and/or force scaling. It is commented in [15] that a teleoperator with adjustable impedance characteristics should be used to specify the ideal behavior of a teleoperated system.

In this thesis, the ideal teleoperator is considered to be the one that allows a direct extension of the operator's manipulation capability to the remote site. The ideal

system is thus such that it provides him/her with a response he would obtain if he were directly performing manipulations at the remote site. In terms of physical quantities, an ideal system can be described as the one in which the slave arm perfectly tracks the operator commanded velocities, while the master manipulator reflects to the operator the same forces that the slave senses in its interaction with the environment. It is further desirable that these responses be independent of the object/environment characteristics.

1.4 Thesis Objective

This thesis focuses on the problem of developing a control-scheme to achieve an effective position/force coupling between the master and slave manipulators. To realize one of the foregoing ideal responses in varying and unpredictable environments, an adaptive control scheme is proposed. Simulation results are also provided to demonstrate the validity of the proposed control scheme.

1.5 Organization

The thesis is organized in the following manner. Chapter 2 presents the dynamical model of a teleoperator system, and discusses the two-port representation for analysis of the behavior of the teleoperated system. Chapter 3 describes the design of a controller that realizes a defined ideal response. Chapter 4 presents the results of computer simulations to demonstrate the performance of the proposed controller scheme. Finally, Chapter 5 summarizes the work highlighting contributions, and providing suggestions for future work.

2. DYNAMICAL MODEL OF TELEOPERATOR SYSTEM

2.1 Description of Teleoperator System

Teleoperator systems generally consist of master and slave manipulators with multiple degrees of freedom (DOF). The master and slave manipulators may be kinematically identical. In this case, a joint to joint coordination can be achieved. Or, they may be kinematically different; then the commanded motion needs to be specified for the end-effector in Cartesian space coordinates. Kinematically identical master and slave arms are typically used in applications involving anthropomorphic arms. In these cases, a human operator commands the master manipulator which functions as an exoskeleton recording the motion of the human arm. This information is transmitted over the medium to the computer of the slave manipulator which tries to follow the motion of the master arm.

Figure 2.1 shows a typical multiple DOF teleoperated system. The joint velocities ($\dot{\theta}_M$) commanded by the operator to position and orient the end-effector are measured. These joint velocities are transformed to the end effector's velocity vector of the master manipulator expressed in the Cartesian base coordinate system ($\dot{x}_m$) by using the master manipulator's Jacobian matrix $J_M(\theta_M)$. This Cartesian velocity vector is transmitted through the communication medium to the slave and represents the desired velocity to be tracked by the slave manipulator. Also, the force sensed at the master arm's end effector expressed in the Cartesian system (f_m) is transmitted to the computer of the slave manipulator. At the slave's end, the velocity and force vectors are transformed to the joint space of the slave manipulator. Indeed, the Cartesian velocity vector is transformed to the joint space ($\dot{\theta}_{S_d}$) of the slave arm by using the inverse of the slave arm's Jacobian $J_S^{-1}(\theta_S)$, which is assumed to exist. The measured

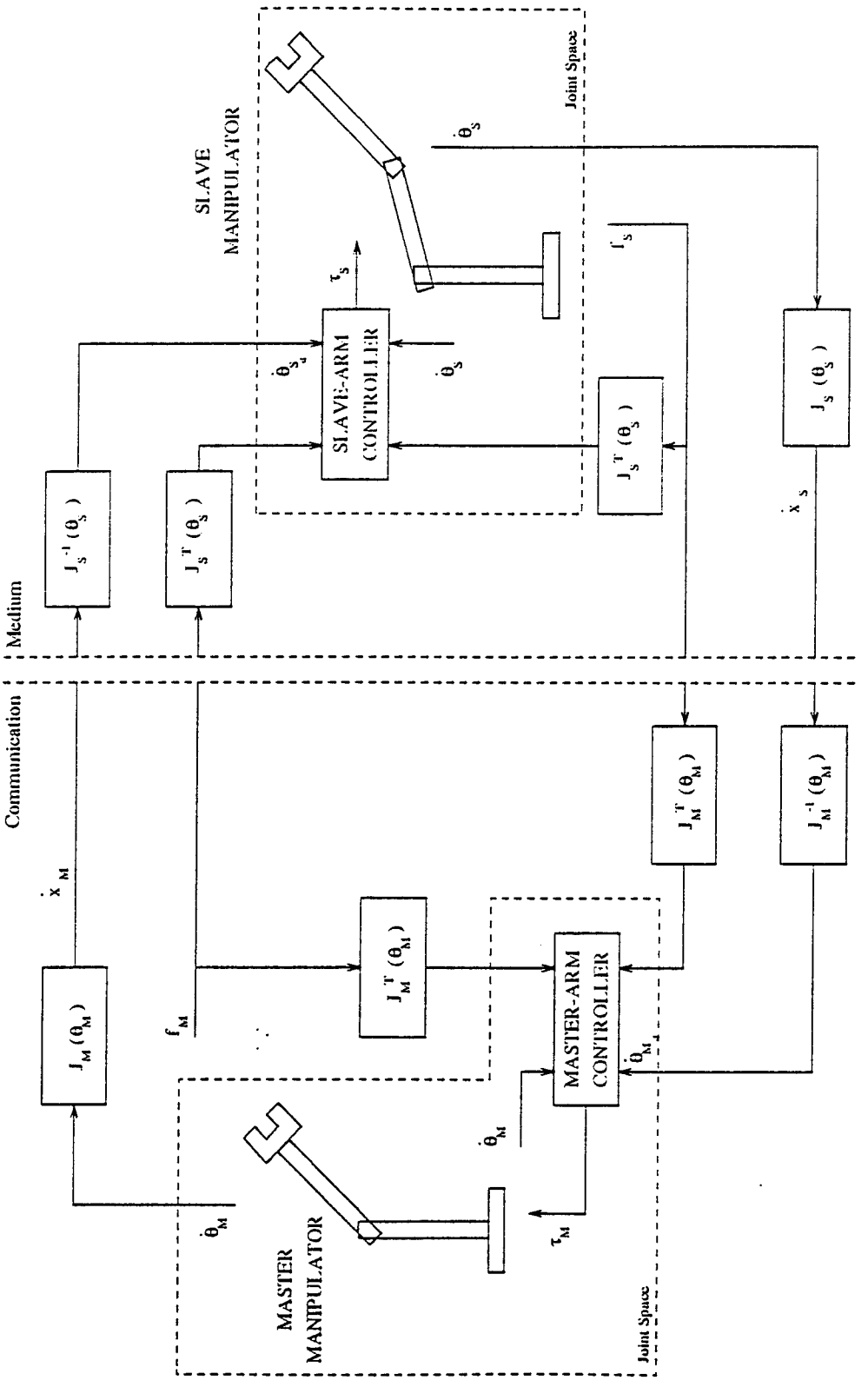


Figure 2.1 A Multi-dimensional Teleoperator System

end effector force of the master arm is obtained in the slave manipulator's joint space by using $J_S^{-T}(\theta_S)$, the inverse of the transposed Jacobian matrix of the slave arm. After transmitting the slave arm's Cartesian velocity and force over the medium to the computer of master arm. a similar procedure is applied to determine the foregoing variables of the slave arm in the master arm computer. The velocities and forces of the master manipulator serve as the desired trajectory of the slave manipulator. A controller is to be designed so that the tracking of the desired trajectory by the slave arm is achieved. A transmission delay may be introduced by the communication medium, but is not considered in this discussion.

2.2 A One-Dimensional Teleoperator System Model

Although teleoperator systems are typically multi-dimensional it has become a common practice to study the dynamics of the system in one dimension [9, 22]. The results can then be generalized to multi-dimensional manipulators [2, 8, 9]. Indeed, the extension to the multiple dimensional manipulators could be achieved by applying the one-dimensional control scheme to each of the joints of the master and slave manipulators in the framework discussed in section 2.1, under the common assumption of independent joint dynamics.

Consider a one-dimensional teleoperator system model shown in figure 2.2.

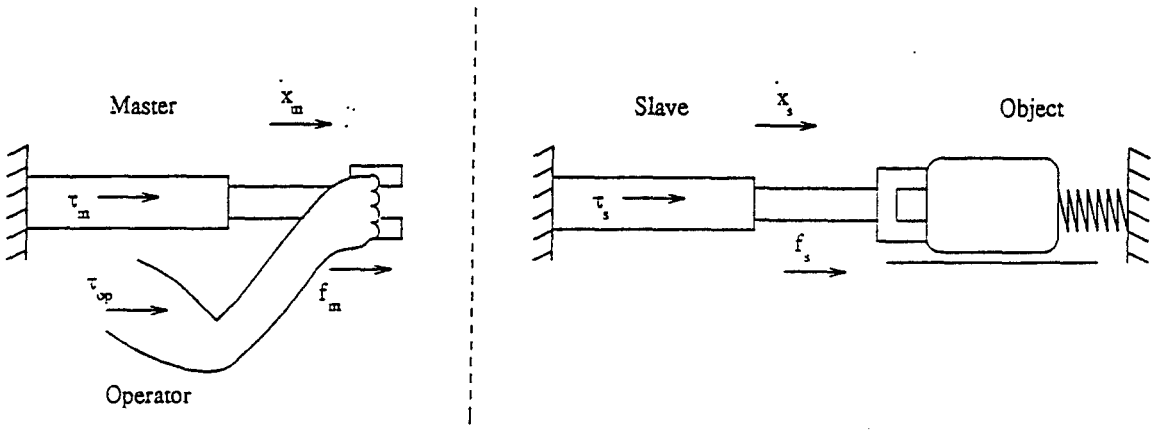


Figure 2.2 A One-dimensional Teleoperator System

The dynamics of the operator may be approximated as:

$$m_{op}\ddot{x}_m + b_{op}\dot{x}_m + c_{op}x_m = \tau_{op} - f_m \qquad [2.1]$$

where m_{op}, b_{op} and c_{op} refer to the mass, viscous coefficient and stiffness of the operator respectively, and τ_{op} is the force generated by the operator. It is assumed that the operator grasps the master arm firmly and there is no slippage in the grip during the operation.

Although it is usually assumed that the operators arm parameters (b_{op} and c_{op}) remain constant, it should be noted that these parameters may change during the operation [15, 22].

The dynamics of the master and slave arms in this case may be represented as follows:

$$m_m\ddot{x}_m + b_m\dot{x}_m = \tau_m + f_m \qquad [2.2]$$

$$m_s\ddot{x}_s + b_s\dot{x}_s = \tau_s - f_s \qquad [2.3]$$

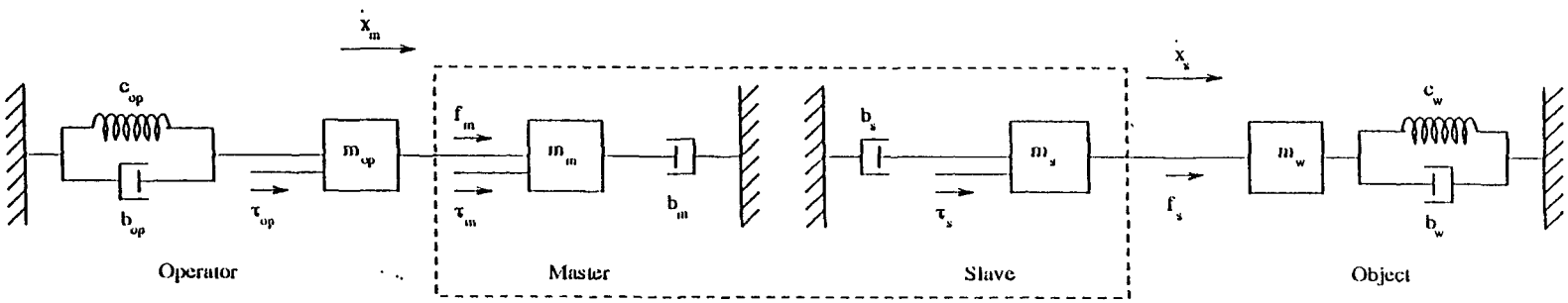
where x_m and x_s denote the displacements of the master and slave arms respectively, and m_m, b_m and m_s, b_s, are the masses and coefficients of viscous friction of the master and slave manipulators, respectively. f_m denotes the generalized force that the operator applies to the master arm, and f_s the force that the slave arm feels in it's interaction with the object/environment. τ_m and τ_s represent the joint actuator torques of the master and slave, respectively.

The dynamics of the slave-arm and its interaction with the object/environment are modeled as:

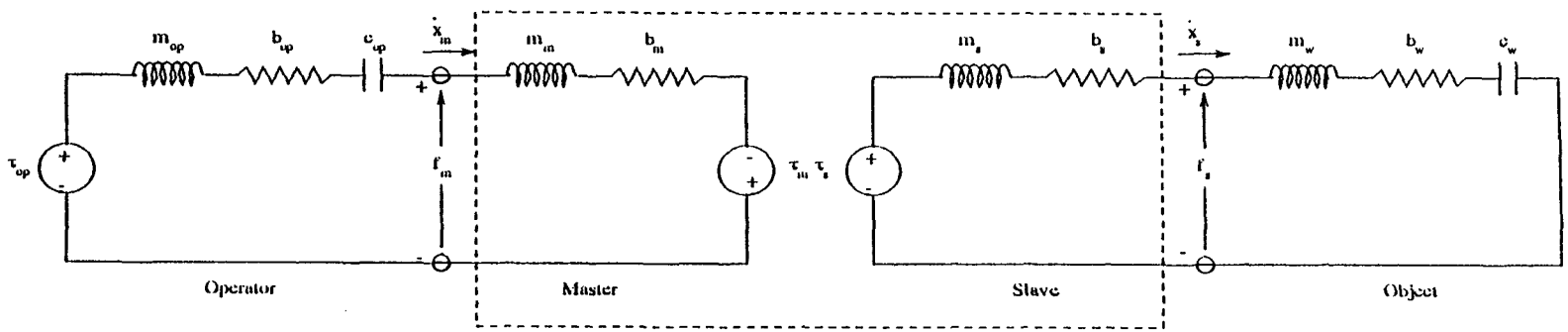
$$m_w\ddot{x}_s + b_w\dot{x}_s + c_w x_s = f_s \qquad [2.4]$$

where m_w, b_w and c_w, denote the mass, viscous coefficient and stiffness of the object, respectively. The displacement of the object is assumed to be equal to that of the slave arm, based on the assumption that the object is firmly grasped by the slave arm, and does not separate from the slave-arm during its motion.

Figure 2.3(a) is obtained on the basis of equations 2.1 to 2.4, and shows the mass-damper-spring representation of a one DOF system.



(a) Mass-Spring-Damper Representation of Teleoperator System



(b) Circuit Representation of Teleoperator System

Figure 2.3 Mechanical and Circuit Representations of Teleoperator System

2.3 Circuit Representation of Teleoperator System

The similarity in the mathematical models of mechanical and electrical systems, gives rise to an analogy in their representation. Using the force-voltage, velocity-current analogy between the systems, the master-slave system discussed in the previous section can be represented by an analogous electrical circuit. The forces in the mechanical system are analogous to voltages, and the velocities to the currents in the electrical system. Based on this analogy, the parameters of mass, viscousity and stiffness correspond to inductance, resistance and capacitance, respectively. The master slave system can thus be represented by an analogous electrical circuit as shown in figure 2.3(b).

2.4 A Two-port Model of Teleoperation

Two-port models are often used in the analysis of electrical circuits involving energy exchange at two distinct ports. Teleoperator systems lend themselves very well to the (analogous) two-port network formulation. The operator-master and the slave-environment ports are shown in figure 2.4. They denote the two ports of energy/information exchange in a teleoperator system. The two-port network model representation of a teleoperator system is conveniently used in the analysis. Using an impedance matrix teleoperator systems with adjustable characteristics are studied in [15]. A hybrid two-port network model is employed in [9].

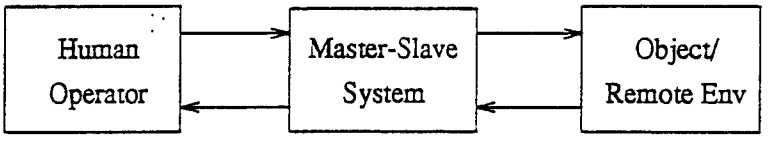


Figure 2.4 A Two-port Block Representation of a Teleoperator System

Figure 2.5 shows a two-port network, where I_1 an I_2 denote the currents at the input and output ports, respectively, while V_1 and V_2 denote the voltages at the respective ports. The driving source of the network is represented by V_{in} and the impedance of the source by Z_{in}. The voltages and currents of this network can be

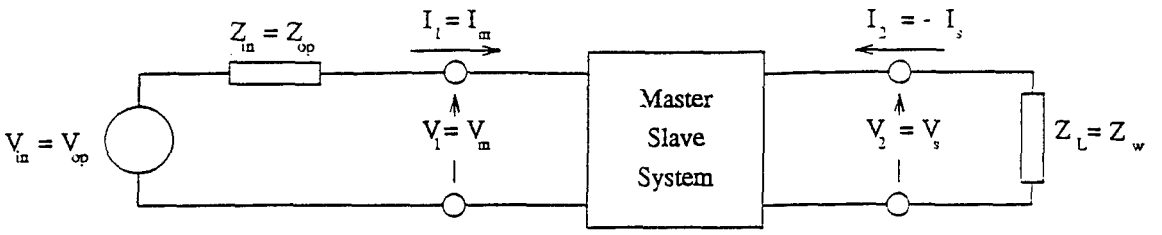


Figure 2.5 A Two-port Network

related as $V = ZI$ or by denoting $V = [V_1\ V_2]^T$ and $I = [I_1\ I_2]^T$, as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad [2.5]$$

where the impedance matrix Z is

$$Z = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \qquad [2.6]$$

and $V = [V_1\ V_2]^T$ and $I = [I_1\ I_2]^T$.

The analogous circuit representation of the teleoperator system in figure 2.4 can now be described as the two-port network shown in figure 2.5. The following notations are used in figure 2.5: $V_{op} = \tau_{op}, Z_{op} = m_{op}s + b_{op} + c_{op}1/s, Z_w = m_w s + b_w + c_w 1/s, V_m = f_m, V_s = f_s, I_m = \dot{x}_m\ and\ I_s = \dot{x}_s$.

2.5 Discussion

The dynamical model of a one DOF teleoperator is presented. Based on the similarity between the equations describing a mechanical system and a circuit. the two-port network representation of the teleoperator is adopted. This model lends itself to the analysis of the system behavior and the design of the controller.

3. CONTROLLER DESIGN

3.1 Specifications of Controller Design for Ideal Performance

A control scheme is now developed to realize the ideal performance of a teleoperator system. The ideal system performance is required to satisfy the following conditions:

- the velocity of the slave arm is equal to that of the master arm;
- The generalized force exerted by the end-effector of the slave arm on the environment is equal to the generalized force that the operator applies to the master arm.

These conditions should be achieved while satisfying the system equations.

In terms of the variables of the dynamical model represented by equation 1.1 to 1.4, the above conditions can be expressed as

$$\dot{x}_s = \dot{x}_m \qquad [3.1]$$

$$f_m = f_s \qquad [3.2]$$

Equations 3.1 and 3.2 must be satisfied for all values of the contact impedance Z_w, which models the contact between the end-effector of the slave arm and the environment.

The conditions of equations 3.1 and 3.2 along with equations 1.1 to 1.4 specify the constraints that must be satisfied during the time when the end-effector of the slave arm is in contact with the environment.

3.2 General Control Scheme

The feedback controller of the system generates the inputs to the actuators of the master and slave arms. The controller used in this study is chosen as [22]

$$\tau_m = \begin{bmatrix} P_m & Q_m \end{bmatrix} \begin{bmatrix} \dot{x}_m \\ f_m \end{bmatrix} - \begin{bmatrix} R_m & S_m \end{bmatrix} \begin{bmatrix} \dot{x}_s \\ f_s \end{bmatrix} \qquad [3.3]$$

$$\tau_s = \begin{bmatrix} P_s & Q_s \end{bmatrix} \begin{bmatrix} \dot{x}_m \\ f_m \end{bmatrix} - \begin{bmatrix} R_s & S_s \end{bmatrix} \begin{bmatrix} \dot{x}_s \\ f_s \end{bmatrix} \qquad [3.4]$$

where $P_m = P_{m_1}s + P_{m_2} + P_{m_3}\frac{1}{s}$; $R_m = R_{m_1}s + R_{m_2} + R_{m_3}\frac{1}{s}$; $P_s = P_{s_1}s + P_{s_2} + P_{s_3}\frac{1}{s}$; $R_s = R_{s_1}s + R_{s_2} + R_{s_3}\frac{1}{s}$.

The feedback gains P_{m_1}, P_{m_2} and P_{m_3}; R_{s_1}, R_{s_2} and R_{s_3}; P_{m_1}, P_{m_2} and P_{m_3}; and R_{s_1}, R_{s_2} and R_{s_3} are constants to be determined by the design. The variable s is a differential operator, i.e., $s = d/dt$.

In calculating the actuator inputs, the master and slave velocities $\dot{x}_m$ and $\dot{x}_s$and the master and slave forces, f_m and f_s are assumed to be available from measurements. The positions and accelerations can be obtained by appropriate software operations.

To develop a control scheme, it is assumed that the time delay in the communication medium between the master and slave is negligible. Further, it is assumed that the position and force transmission scales (i.e. the ratio of the positions or forces of the slave arm to those of the master arm) are unity. The scaling of the force and position in the teleoperator system would be an extension of the above approach.

3.3 Controller Design For The Ideal Response

The general control scheme outlined in section 3.2 is next specified by determining the gains so that the conditions for the ideal response will be met.

Substituting equations 3.3 and 3.4 into equations 2.2 and 2.3 gives

$$\begin{bmatrix} Z_m - P_m & -R_m \\ -P_s & -(Z_s + R_s) \end{bmatrix} \begin{bmatrix} I_m \\ -I_s \end{bmatrix} = \begin{bmatrix} 1 + Q_m & -S_m \\ Q_s & -(1 + S_s) \end{bmatrix} \begin{bmatrix} V_m \\ V_s \end{bmatrix} \qquad [3.5]$$

where Z_m=$m_m s$+b_m, Z_s=$m_s s$−b_s, I_m=$\dot{x}_m$, I_s=$\dot{x}_s$, V_m=f_m and V_s=f_s

The impedances in equation 3.5 are given by [22]:

$$z_{11} = \frac{(1+S_s)(Z_m - P_m) + S_m P_s}{(1+S_s)(1+Q_m) - S_m Q_s} \triangleq \frac{N_{11}}{D_Z} \quad [3.6]$$

$$z_{12} = \frac{-(1+S_s)R_m + S_m(Z_s + R_s)}{(1+S_s)(1+Q_m) - S_m Q_s} \triangleq \frac{N_{12}}{D_Z} \quad [3.7]$$

$$z_{21} = \frac{(1+Q_m)P_s + Q_s(Z_m - P_m)}{(1+S_s)(1+Q_m) - S_m Q_s} \triangleq \frac{N_{21}}{D_Z} \quad [3.8]$$

$$z_{22} = \frac{(1+Q_m)(Z_s + R_m) - Q_s R_m}{(1+S_s)(1+Q_m) - S_m Q_s} \triangleq \frac{N_{22}}{D_Z} \quad [3.9]$$

$$|Z| = \frac{(Z_m - P_m)(Z_s + R_s) + P_s R_m}{(1+S_s)(1+Q_m) - S_m Q_s} \triangleq \frac{D_Y}{D_Z} \quad [3.10]$$

where $|Z|$ is the determinant of the impedance matrix Z.

Based on the definition of ideal response given in equations 3.1 and 3.2, necessary conditions for the controller design are obtained.

If Z_L denotes the impedance at the contact of the slave hand with the environment, then V_s=$Z_L I_s$. Then, equation 3.5 can be solved for I_m/I_s=$\dot{x}_m/\dot{x}_s$to obtain:

$$\frac{\dot{x}_m}{\dot{x}_s} = \frac{(N_{22} + D_Z Z_L)}{N_{21}} \quad [3.11]$$

Similarly, solving equation 3.5 for V_m/V_s=f_m/f_s, one obtains

$$\frac{f_m}{f_s} = \frac{(N_{11} Z_L + D_Y)}{(N_{21} Z_L)} \quad [3.12]$$

It is desirable that the controller gains are independent of the task impedance Z_L. Further, the ideal operating conditions of the teleoperator system require that $\dot{x}_m/\dot{x}_s = 1$ and $f_m/f_s = 1$ in equations 3.11 and 3.12. These conditions are satisfied if the following conditions are satisfied [22]

$$D_Z = 0 \quad [3.13]$$

$$N_{22} = N_{21} \neq 0 \quad [3.14]$$

$$D_Y = 0 \quad [3.15]$$

$$N_{21} = N_{11} \neq 0 \quad [3.16]$$

It should be noted that when D_Z=0, the impedance matrix Z is undefined. To circumvent this ambiguity, a chain matrix is introduced to obtain the following model

$$\begin{bmatrix} V_m \\ I_m \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \begin{bmatrix} V_s \\ -I_s \end{bmatrix} \qquad [3.17]$$

where the coefficient matrix is denoted by K. The chain matrix K is often used in network theory, when the output of a two-port network is connected to the input of another two-port network. For the system in equation 3.5, the chain matrix is

$$K = \begin{bmatrix} \frac{N_{11}}{N_{21}} & \frac{D_Y}{N_{21}} \\ \frac{D_Z}{N_{21}} & \frac{N_{22}}{N_{21}} \end{bmatrix} \qquad [3.18]$$

It can now be checked that the four foregoing conditions for the ideal response of the teleoperator system result in

$$K = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad [3.19]$$

From equation 3.17 and 3.19, it is seen that the ideal response can be achieved when equations 3.13–3.16 are satisfied. Indeed, equations 3.13–3.16 determine the constant controller gains.

3.4 An Adaptive Controller

To achieve the ideal response, equations 3.13–3.16 must be satisfied. These equations contain sixteen variables in twelve equations. Therefore, some of the gains can be arbitrarily chosen. They can be determined so as to the best possible performance of the system.

If however the dynamical model of the system is not accurately known, for example due to unknown masses of the payload or due to other unpredictable disturbances acting on the system, constant controller gains may not result in an acceptable performance. The need for frequent gain adjustments of the controller to comply with varying operating conditions can be overcome by employing an adaptive controller [13].

Because teleoperated systems are frequently operated in varying and unpredictable environments, an adaptive control scheme appears as an attractive choice to enhance the performance of the system.

An adaptive scheme functions so as to (i) recognize changes in the operating conditions through measurements, (ii) evaluate the system performance based on measurements and (iii) makes adjustments in the controller parameters to conform with existing operating conditions [13].

In the model described for the teleoperated system, the estimated parameters in the two-port representation of the system reflect the existing operating conditions. When the external environmental conditions change, so do the numerical values in the impedance matrix. These values will be estimated by an on-line estimation scheme on the basis of measurements. They then determine the controller gains, which change with the impedance matrix and thus the environmental conditions. By incorporating an additional parameter estimation scheme in the system design, the problem of unknown impedances of the master and slave arms can also be overcome.

3.4.1 Master and Slave Arm Impedance Estimation

The mechanical impedances of the master and the slave arms are required to be known for the design of the controller. Since these parameters are not usually known, they need to be estimated on-line, which is discussed next.

Equations 2.2 and 2.3 can be rewritten as

$$\begin{bmatrix} \tau_m + f_m \\ \tau_s - f_s \end{bmatrix} = \begin{bmatrix} \ddot{x}_m & \dot{x}_m & 0 & 0 \\ 0 & 0 & \ddot{x}_s & \dot{x}_s \end{bmatrix} \alpha \qquad [3.20]$$

$$\alpha = \begin{bmatrix} m_m & b_m & m_s & b_s \end{bmatrix}^T \qquad [3.21]$$

where the superscript T signifies transposition.

The problem is now to determine an estimate $\hat{\alpha}$ for the value of the parameter vector α. When the values of $\dot{x}_m$, $\dot{x}_s$, f_m, f_s τ_m and τ_s are known, equation 3.20 forms the basis of an algorithm that recursively calculates $\hat{\alpha}$. The well-known least squared

errors method [13] will be used here. The accelerations $\ddot{x}_m$ and $\ddot{x}_s$ in equation 3.20 can be approximated by using Euler first-order backward finite differences on the measured values $\dot{x}_m$ and $\dot{x}_s$. Assuming a sufficiently short sampling period T, the accelerations $\ddot{x}_m$ and $\ddot{x}_s$ can be calculated as:

$$\ddot{x}_m(k) \approx \frac{\dot{x}_m(k) - \dot{x}_m(k-1)}{T} \quad [3.22]$$

$$\ddot{x}_s(k) \approx \frac{\dot{x}_s(k) - \dot{x}_s(k-1)}{T} \quad [3.23]$$

where the sampling period has been dropped in the arguments, and $k = 1, 2, \ldots$

Equation 3.20 can be rewritten for the k_{th} sampling instant as

$$y(k) = \phi^T(k)\alpha(k) + e(k) \quad [3.24]$$

where $e(k) = [e_1(k)\, e_2(k)]^T$ represents model errors and

$$y(k) = \begin{bmatrix} \tau_m(k) + f_m(k) \\ \tau_s(k) - f_s(k) \end{bmatrix}$$

$$\phi^T(k) = \begin{bmatrix} \ddot{x}_m(k) & \dot{x}_m(k) & 0 & 0 \\ 0 & 0 & \ddot{x}_s(k) & \dot{x}_s(k) \end{bmatrix}$$

The least squared error estimate of the parameter vector α is obtained as [13]

$$\hat{\alpha}(k+1) = \hat{\alpha}(k) + K(k)\left[y(k) - \phi^T(k)\hat{\alpha}(k)\right] \quad [3.25]$$

$$K(k) = P(k)\phi(k)\left[\phi^T(k)P(k)\phi(k) + I\right]^{-1} \quad [3.26]$$

$$P(k) = P(k-1) - P(k-1)\phi(k)\left[\phi^T(k)P(k-1)\phi(k) + I\right]^{-1}\phi^T(k)P(k-1) \quad [3.27]$$

where matrix $P(k)$ is symmetric and positive definite.

Equations 3.25–3.27 determine an estimate $\hat{\alpha}$ at each sampling instant on the basis of the latest available measurements.

The estimates of the master and slave arm parameters are then incorporated in the equations that specify the controller gains.

3.4.2 Determination of Controller Gains Based on Two-port Parameters

Necessary conditions for the ideal response are specified in terms of the controller gains by equations 3.13–3.16. They can be used to solve for the controller gains appearing in equations 3.6–3.10.

Substituting V_m=f_m, V_s=f_s, I_m=$\dot{x}_m$ and I_s=$\dot{x}_s$, equation 2.5 is rewritten as

$$\begin{bmatrix} f_m \\ f_s \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} \dot{x}_m \\ -\dot{x}_s \end{bmatrix} \qquad [3.28]$$

where the elements $z_{ij}, i, j = 1, 2$ in equation 3.28 are assumed to be of the following forms:

$$z_{11} = a_{11}s + b_{11} + c_{11}\frac{1}{s} \qquad [3.29]$$

$$z_{12} = a_{12}s + b_{12} + c_{12}\frac{1}{s} \qquad [3.30]$$

$$z_{21} = a_{21}s + b_{21} + c_{21}\frac{1}{s} \qquad [3.31]$$

$$z_{22} = a_{22}s + b_{22} + c_{22}\frac{1}{s} \qquad [3.32]$$

Parameters a_{ij}, b_{ij}, cij are the unknown parameters to be estimated.

The substitution of equations 3.29–3.32 into equation 3.28 gives

$$\begin{bmatrix} f_m \\ f_s \end{bmatrix} = \begin{bmatrix} \ddot{x}_m & \dot{x}_m & x_m & -\ddot{x}_s & -\dot{x}_s & -x_s & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddot{x}_m & \dot{x}_m & x_m & -\ddot{x}_s & -\dot{x}_s & -x_s \end{bmatrix} \beta \qquad [3.33]$$

where

$$\beta = \begin{bmatrix} a_{11} & b_{11} & c_{11} & a_{12} & b_{12} & c_{12} & a_{21} & b_{21} & c_{21} & a_{22} & b_{22} & c_{22} \end{bmatrix}^T \qquad [3.34]$$

If the values of $\ddot{x}_m$, $\dot{x}_m$, x_m, $\ddot{x}_s$, $\dot{x}_s$, x_s, f_m and f_s are known, equation 3.33 replaces equation 3.24 in the estimation algorithm. Then, one can calculate recursively the estimate $\hat{\beta}$ using equations 3.25–3.27. It follows that the estimate $\hat{z}_{ij}$, $i, j = 1, 2$ of the impedance matrix Z is then updated at each sampling instant. when additional measurements become available.

The conditions of the ideal response specified by the four equations 3.13–3.16 contain sixteen unknown controller gains. By equating the coefficients of the equal powers of s, one obtains twelve equations which contain the sixteen unknowns. Thus, there are more unknowns than equations and the problem is therefore underspecified. In order to obtain a well-posed problem, additional equations for this system need to be obtained, for example by introducing additional constraints. Although these constraints are not necessary for the ideal response, they are introduced for convenience. Further, since they do affect the conditions of ideal response. they provide us with sufficient solutions for implementing the ideal teleoperator system response.

Equations 3.6–3.10 specify the relations that exist between the controller gains and the elements of the impedance matrix. Incorporating these relations in the design of the controller would lead to gain selection corresponding to existing operating conditions as specified by the estimated values $\hat{z}$. Therefore, one may choose $N_{21}=M\hat{z}_{21}$, where M is a constant determined by experimentation. $N_{21}=M\hat{z}_{21}$ implies

$$(1+Q_m)P_s + Q_s(\hat{Z}_m - P_m) = M(\hat{a}_{21}s + \hat{b}_{21} + \hat{c}_{21}\frac{1}{s}) \qquad [3.35]$$

Further, the four conditions of ideal response represented by equations 3.13-3.16 yield the following equations

$$(1+S_s)(1+Q_m) - S_m Q_s = 0 \qquad [3.36]$$

$$(1+Q_m)(\hat{Z}_s + R_m) - Q_s R_m = M(\hat{a}_{21}s + \hat{b}_{21} + \hat{c}_{21}\frac{1}{s}) \qquad [3.37]$$

$$(\hat{Z}_m - P_m)(\hat{Z}_s + R_s) + P_s R_m = 0 \qquad [3.38]$$

$$(1+S_s)(\hat{Z}_m - P_m) + S_m P_s = M(\hat{a}_{21}s + \hat{b}_{21} + \hat{c}_{21}\frac{1}{s}) \qquad [3.39]$$

For convenience, let $N_{12} = N_{21}$. It implies that

$$-(1+S_s)R_m + S_m(\hat{Z}_s + R_s) = M(\hat{a}_{21}s + \hat{b}_{21} + \hat{c}_{21}\frac{1}{s}) \qquad [3.40]$$

Equations 3.35–3.40, are to be solved for the sixteen controller gains viz. P_{m_1}, P_{m_2}, P_{m_3}, Q_m, R_{m_1}, R_{m_2}, R_{m_3}, S_m, P_{s_1}, P_{s_2}, P_{s_3}, Q_s, R_{s_1}, R_{s_2}, R_{s_3} and S_s. The

detailed steps are presented in the appendix. Solving for the gains, the following results are obtained

$$P_{m_1} = \frac{-M\hat{a}_{21} + \hat{m}_s S_m + R_{s_1} S_m + \hat{m}_m (S_s + 1)}{(S_s + 1)} \quad [3.41]$$

$$P_{m_2} = \frac{-M\hat{b}_{21} + \hat{b}_s S_m + R_{s_2} S_m + \hat{b}_m (S_s + 1)}{(S_s + 1)} \quad [3.42]$$

$$P_{m_3} = \frac{-M\hat{c}_{21} + R_{s_3} S_m}{(S_s + 1)} \quad [3.43]$$

$$P_{s_1} = \hat{m}_s + R_{s_1} \quad [3.44]$$

$$P_{s_2} = \hat{b}_s + R_{s_2} \quad [3.45]$$

$$P_{s_3} = R_{s_3} \quad [3.46]$$

$$R_{m_1} = \frac{-M\hat{a}_{21} + \hat{m}_s S_m + R_{s_1} S_m}{(S_s + 1)} \quad [3.47]$$

$$R_{m_2} = \frac{-M\hat{b}_{21} + \hat{b}_s S_m + R_{s_2} S_m}{(S_s + 1)} \quad [3.48]$$

$$R_{m_3} = \frac{-M\hat{c}_{21} + R_{s_3} S_m}{(S_s + 1)} \quad [3.49]$$

$$Q_s = S_s + 1 \quad [3.50]$$

$$Q_m = S_m - 1 \quad [3.51]$$

$$S_s \neq -1 \quad [3.52]$$

while gains R_{s_1}, R_{s_2}, R_{s_3} and S_m are arbitrarily selected. Equations 3.41–3.52 thus specify the controller gains.

In implementing a controller with constant gains, constant values are used in place of $\hat{a}_{21}$, $\hat{b}_{21}$ and $\hat{c}_{21}$, in the derived expressions for the gains.

3.5 Discussion

In this chapter a controller has been developed to achieve the defined ideal response. Moreover, the controller design is based on an adaptive scheme to adjust the gains depending on the operating conditions, and to circumvent the problem of unknown master and slave arm impedances.

4. SIMULATION

The proposed adaptive control algorithm was tested by simulating the teleoperator system. The simulation results will next be presented. The teleoperator system shown in figure 2.2 is used in the study. The dynamical model given in equations 2.1 to 2.4 is simulated using Matlab to demonstrate the performance of the system.

In the simulations, the operator manipulates a remote object with the aid of the master-slave manipulator system. It consists of two identical one dimensional manipulators. It is assumed that the operator holds the master arm with a firm grip, and does not loosen it, or let the grip slip; similarly it is assumed that the slave arm grasps the object firmly.

In the simulations, the operator generates a sinusoidal force of $\tau_{op} = 5 - 5\cos(4\pi t)$. The operator's arm impedance parameters (b_{op} and c_{op}) remain constant during the operation. A sampling period T=1 ms is used in the simulation.

4.1 System Parameters Used in Simulations

The parameters of the operator are: mass m_{op}=2.0 kg, viscous coefficient b_{op}=2.0 N-s/m and stiffness coefficient c_{op}=10.0 N/m [22].

The master and slave arms are assumed to be identical. The parameters in their models are: mass m_m=m_s=2.0 kg and viscous coefficient b_m=b_s=0.1 N-s/m.

It should be noted that these numerical values of the master and slave arm parameters are used in the simulations of the respective arm dynamics. The estimated values of these parameters are used in the controller design.

The contact model between the object/environment and the end-effector of the slave arm is described by the following parameters: mass m_w=0.5 kg, viscous coefficient b_w=1.1 N-s/m and stiffness coefficient c_w=10.0 N/m.

4.2 Determination of Controller Gains

The controller gains are determined on the basis of equations 3.41 to 3.52. In simulating the system with the constant gain controller, the expressions of $\hat{a}_{21}$, $\hat{b}_{21}$ and $\hat{c}_{21}$ are replaced by 1.0. This choice was seen to obtain the best response for the constant gain controller. The five controller gains, viz. R_{s_1}, R_{s_2}, R_{s_3}, S_m and S_s, are chosen after some experimentation. They assume the following values. R_{s_1}=-2.0, R_{s_2}=30.0, R_{s_3}=180.0, S_m=1.0 and S_s=0.0.

Based on the values of R_{s_1}, R_{s_2}, R_{s_3}, S_m and S_s, gains P_{s_3}, Q_s and Q_m are calculated from equations 3.46, 3.50 and 3.51 to obtain: P_{s_3}=180.0, Q_s=1.0 and Q_m=0.0.

4.3 Numerical Results

In the first simulation, no disturbances, or variations in the object/environment parameters were introduced. The objective is to demonstrate the performances of the constant gain controller and the adaptive controller. In the next simulation, disturbances in the form of variations in the object/environment parameters are introduced. The simulation results are then used to compare the performance of the adaptively controlled system to that of the system with the constant gain controller.

4.3.1 Simulation 1

The performances of the constant gain controller and the adaptive control scheme in the absence of external disturbances are demonstrated. The initial estimates for the master and slave arms are: mass $\hat{m}_m(0) = \hat{m}_s(0) = 1.0$ kg and viscous coefficient $\hat{b}_m(0) = \hat{b}_s(0) = 0.5$ N-s/m.

The simulation was carried on for a period of 4 seconds and the following results were obtained.

The position responses x_m and x_s of the master and slave arms with the constant gain controller are shown in figures 4.1 and 4.2. They are compared with the corresponding ideal response x_i, the response obtained when the operator directly manipulates the object.

Figures 4.3 and 4.4 shows the force responses. The force f_m that the operator applies to the master arm. and the force f_s that the slave arm exerts on the object are compared with the force f_i, that the operator would have exerted on the object in direct manipulation.

Figures 4.5 and 4.6 show the position responses with the adaptive controller, and figures 4.7 and 4.8 show the force responses.

Figure 4.9 shows the adaptive controller gains P_{m_1}, P_{m_2}, P_{m_3}, P_{s_1}, P_{s_2}, R_{m_1}, R_{m_2} and R_{m_3}, that specify τ_m and τ_s.

Figures 4.10 to 4.13 show the estimated parameters of z_{11}, z_{12}, z_{21}, z_{22}, the components of the impedance matrix of the two port model.

The estimated impedance parameters of the master and slave arms are shown in figure 4.14.

4.3.2 Simulation 2

The performance of the system with the adaptive controller is compared to that of the system with the constant gain controller, when the systems was subjected to disturbances/variations in the environment. The simulation was carried on with essentially the same parameters and inputs as simulation 1 with the only difference being the variation of some object parameters. The object parameters were varied at two points in this simulation. At time = 0.4 seconds, the parameters of the contact model are suddenly changed to m_w= 0.8 kg, b_w= 1.11 N-s/m and c_w= 10.5 N/m for a period of 20 ms, after which they are again given the initial values. Then at time = 0.52 seconds, they are changed to m_w= 0.7 kg, b_w= 1.15 N-s/m and c_w= 11.0 N-s.

and left at that value for the rest of the simulation. This simulation is carried on for a period of 0.7 seconds. The change in object parameters could be considered to be a result of a change in the surface characteristics itself if the object is assumed to be moved on that surface. For the controller with constant gains, the values $\hat{a}_{21}$, $\hat{b}_{21}$and $\hat{c}_{21}$ appearing in the expressions for the gains (equations 3.41–3.52) are set to 1 for the entire simulation.

Figures 4.15 and 4.16 show the position responses of the master and slave arms, respectively, when the controller gains are constant. They are compared with the desired ideal response. Also shown shown in the figures are the errors of the respective actual responses.

Figures 4.17 and 4.18 show the force responses of the master and slave arms, respectively for the controller with constant gains.

Figures 4.19 and 4.20 show the position responses of the master and slave arms, respectively, for the adaptive controller compared with the desired ideal response.

Figures 4.21 and 4.22 show the force responses of the respective arms for the adaptive controller.

4.4 Discussion

The first simulation compares the performance of the constant gain controller to that of the adaptive controller when the system dynamics are free of variations or disturbances. The system performance with the adaptive controller is seen to be slightly better that with the constant gain controller. This improvement is observed with the adaptive controller since its gains adjust along the various sections of the trajectory corresponding to the varying two-port model.

The second simulation presents results which can be used to compare the performance of the adaptive controller to that of the controller with constant gains when disturbances are introduced in the object/environment dynamics. The first disturbance introduced at time = 0.4 seconds results in oscillations in the force response of

the nonadaptive controller. The response settles down after 20 ms when the disturbance is removed. The adaptive controller however adjusts to the introduced variation in the dynamical model and keeps the tracking error small. The relatively larger errors in the force response at time = 0.4, 0.42 and 0.52 seconds in the case of the adaptive controller correspond to the learning period of the adaptive scheme. These errors however, reduce to low values after a few samples, when the controller has adapted to the introduced disturbances. Again with the second disturbance introduced at time = 0.52 seconds, the force responses of the master and slave arms for the nonadaptive controller show large errors which apparently increases. The adaptive controller recognizes these changes and the controller gains are adjusted to account for the changes in the dynamics of the system. Comparing the errors obtained for the position and force tracking of the two schemes, the improvement in system performance with the adaptive controller can be seen.

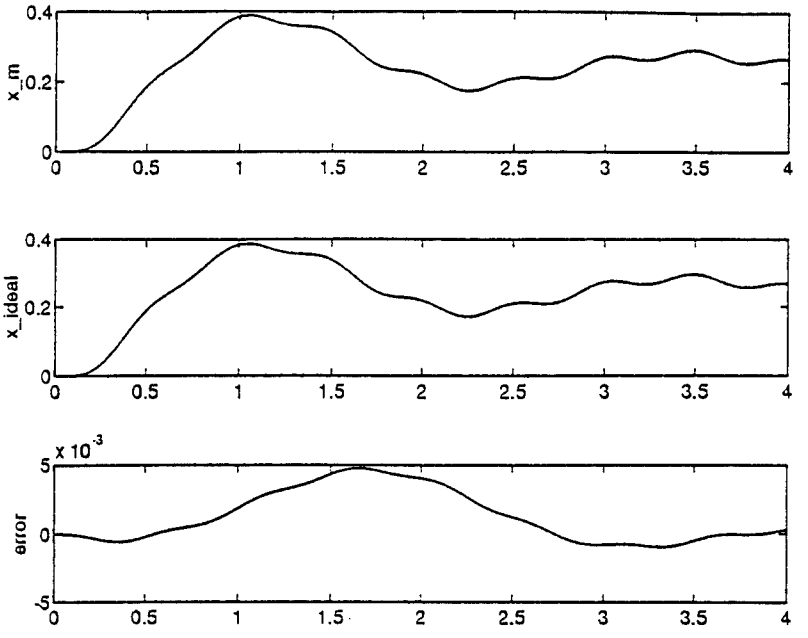


Figure 4.1 Nonadaptive Controller: Master Arm Position Response

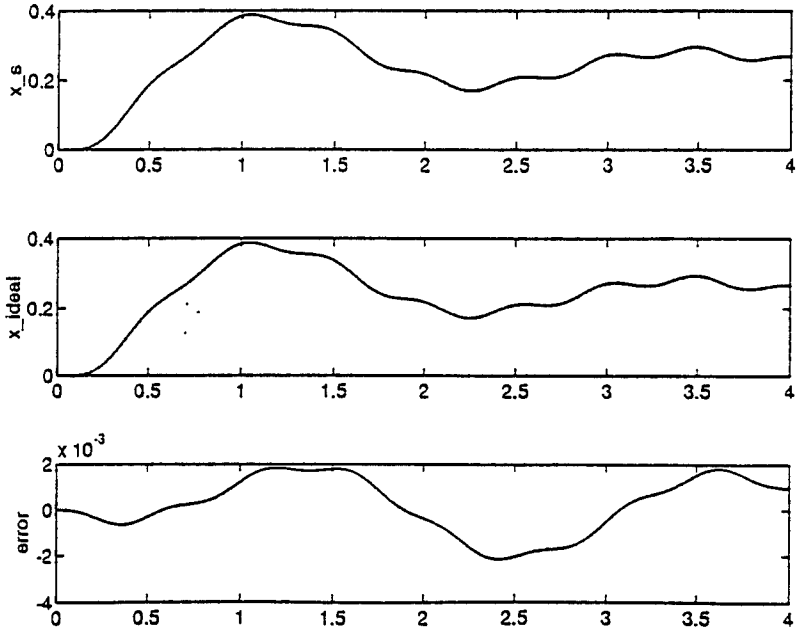


Figure 4.2 Nonadaptive Controller: Slave Arm Position Response

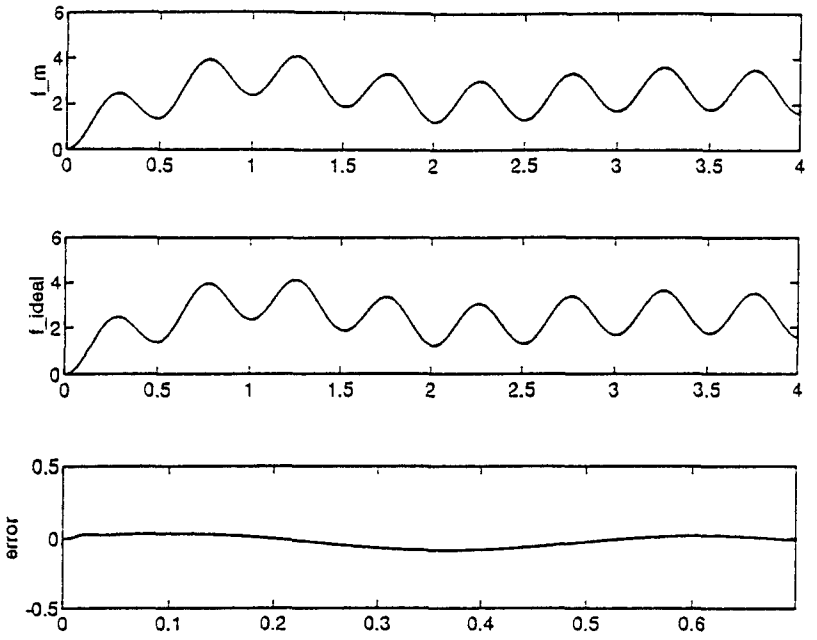


Figure 4.3 Nonadaptive Controller: Master Arm Force Response

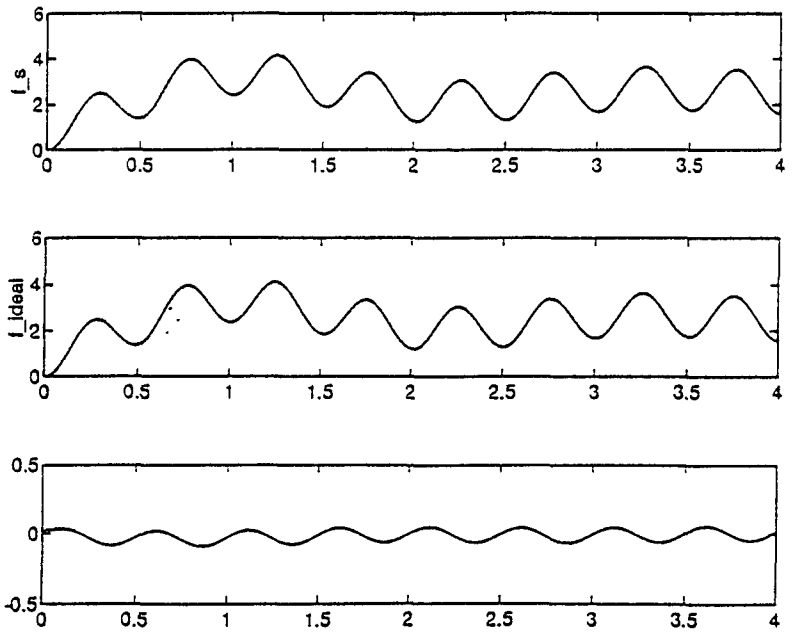


Figure 4.4 Nonadaptive Controller: Slave Arm Force Response

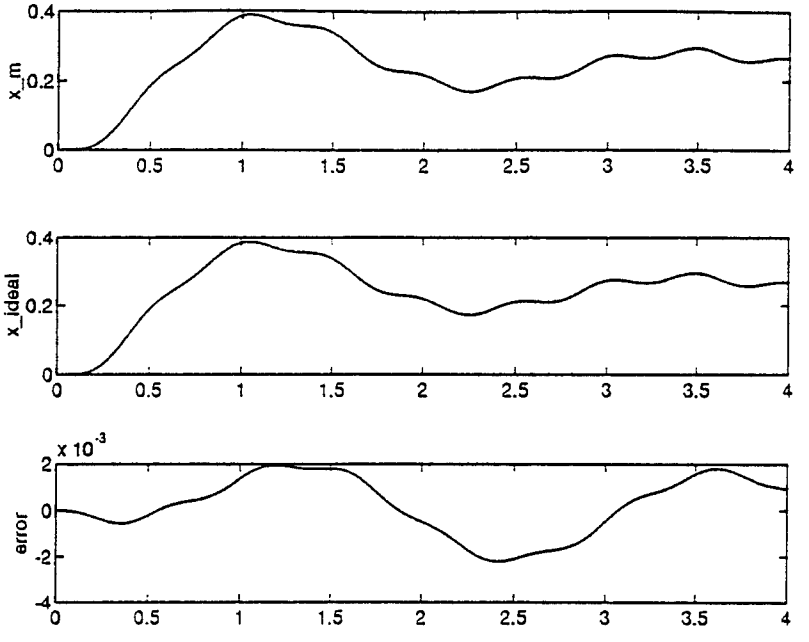


Figure 4.5 Adaptive Controller: Master Arm Position Response

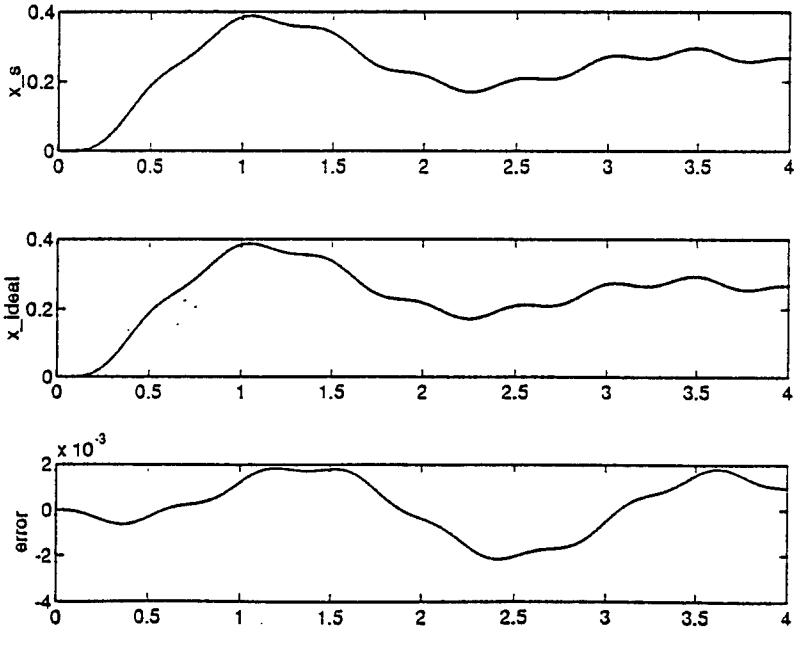


Figure 4.6 Adaptive Controller: Slave Arm Position Response

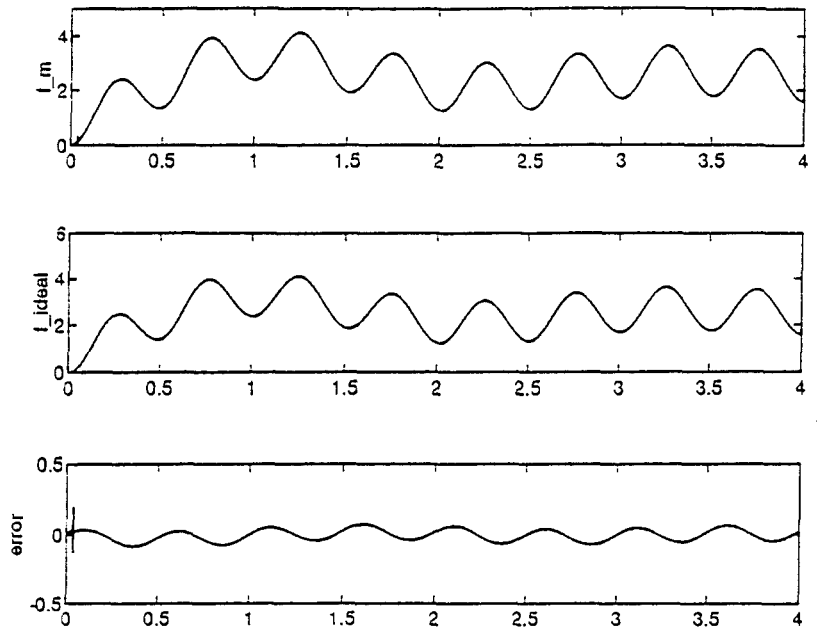


Figure 4.7 Adaptive Controller: Master Arm Force Response

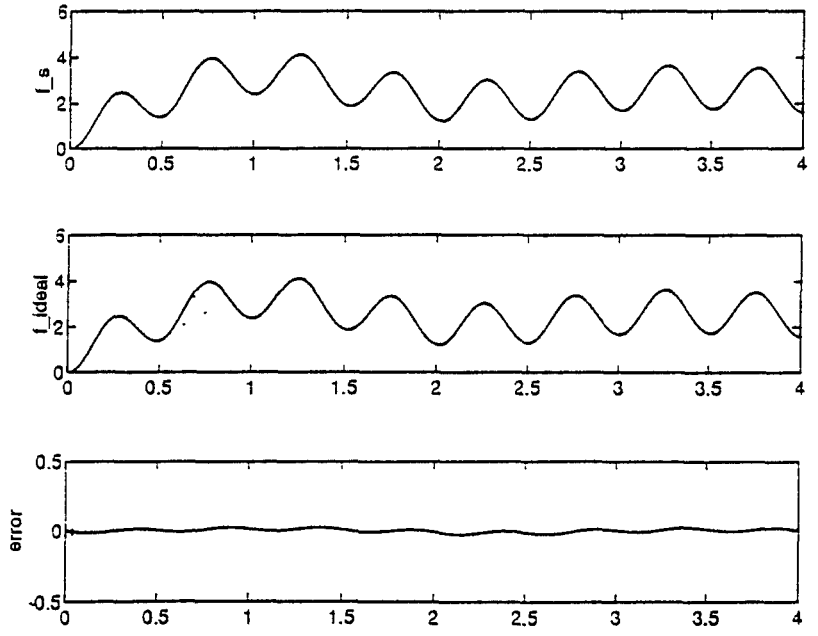


Figure 4.8 Adaptive Controller: Slave Arm Force Response

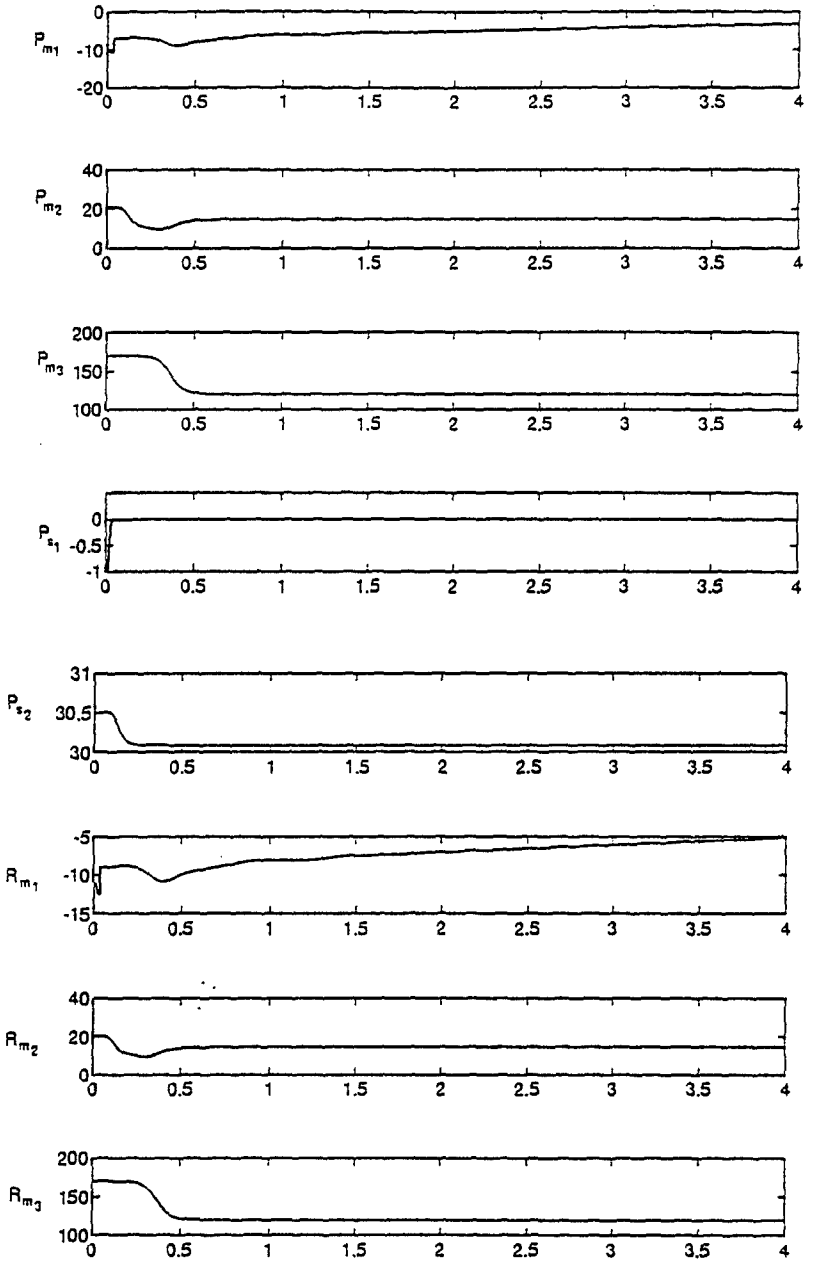


Figure 4.9 Adaptive Controller Gains

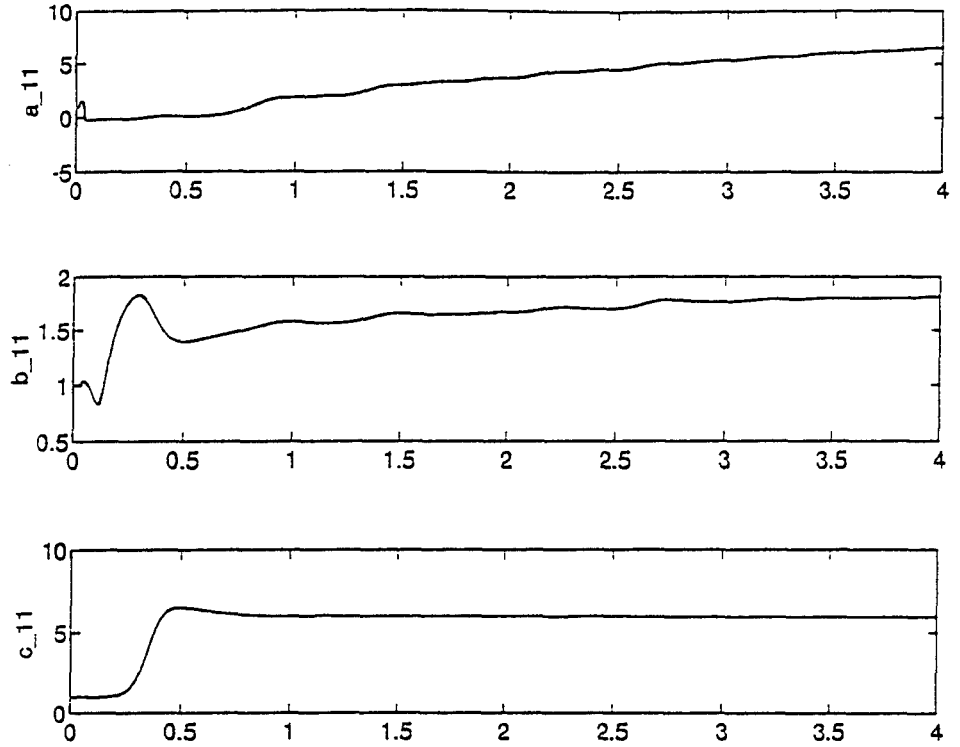


Figure 4.10 Estimate of Impedance Component z_{11} (z_{11}=$a_{11}s$+b_{11}+c_{11}/s)

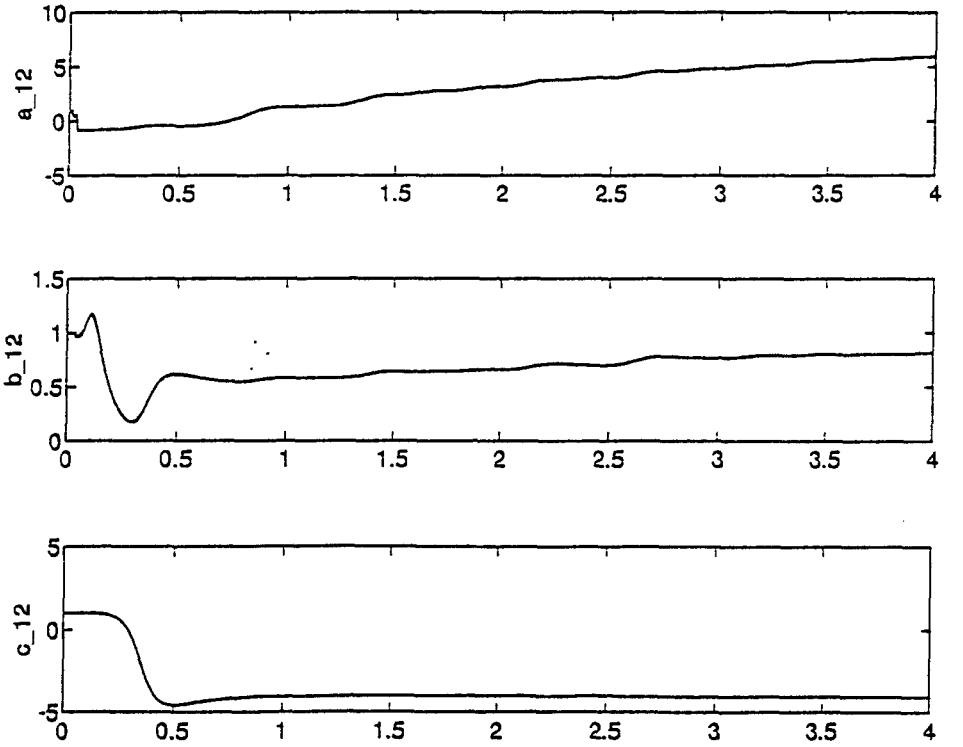


Figure 4.11 Estimate of Impedance Component z_{12} (z_{12}=$a_{12}s$+b_{12}+c_{12}/s)

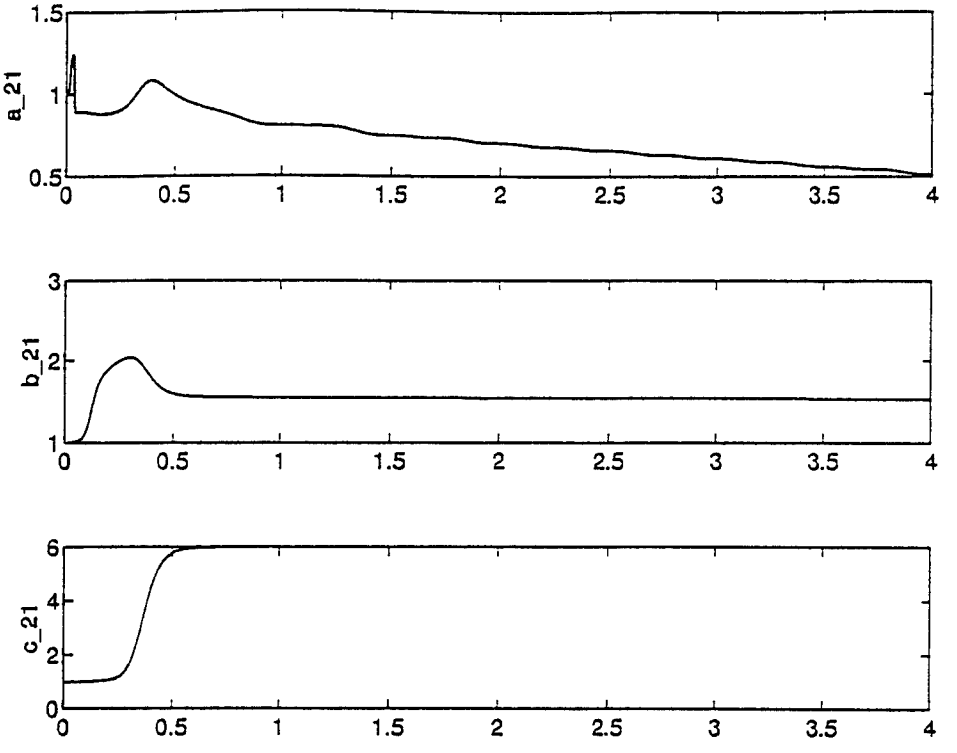


Figure 4.12 Estimate of Impedance Component z_{21} (z_{21}=$a_{21}s$+b_{21}+c_{21}/s)

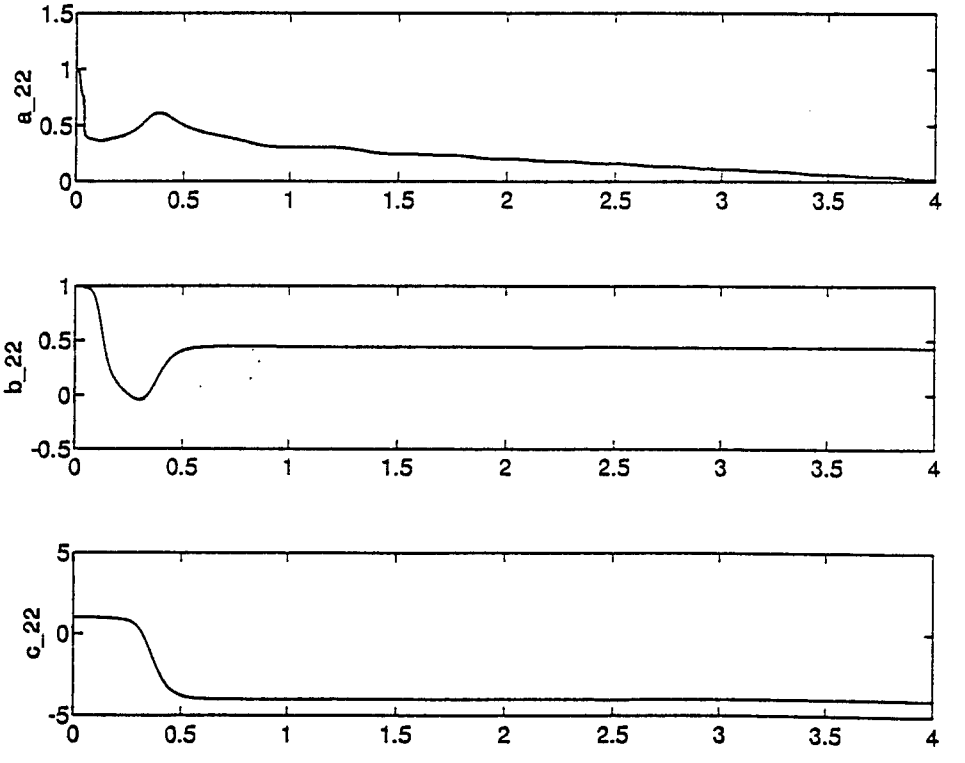


Figure 4.13 Estimate of Impedance Component z_{22} (z_{22}=$a_{22}s$+b_{22}+c_{22}/s)

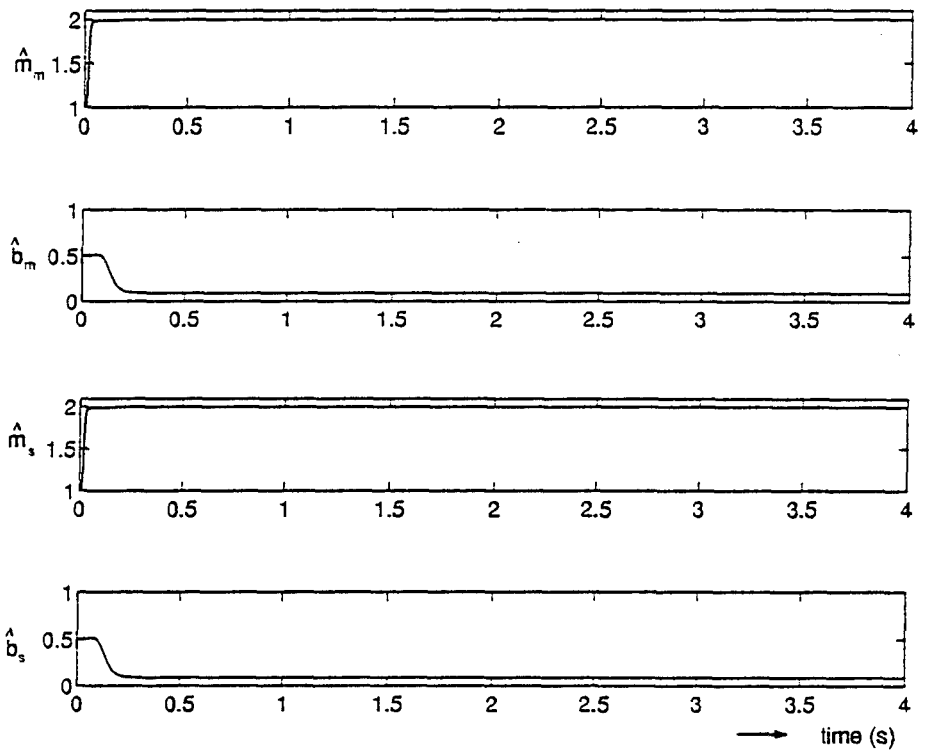


Figure 4.14 Master and Slave Arm Impedance Estimates

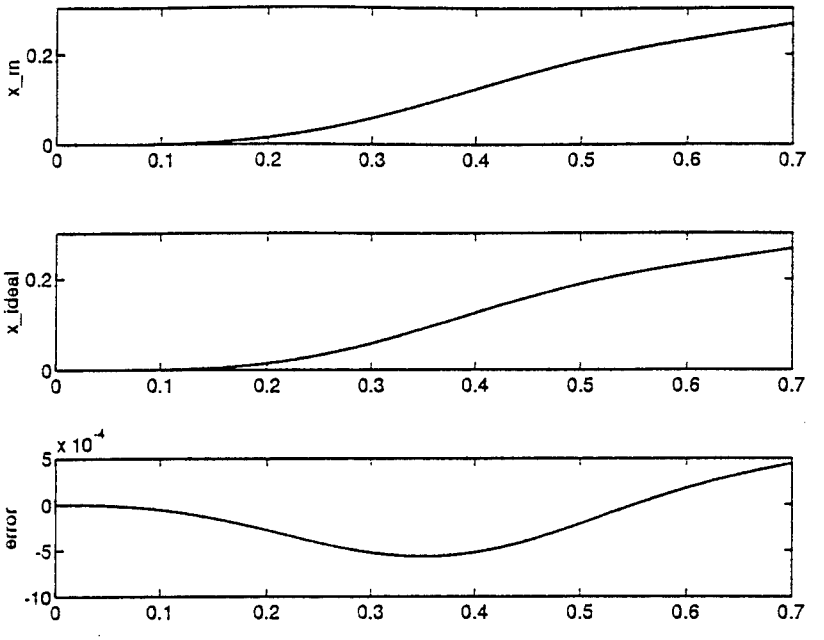


Figure 4.15 Nonadaptive Controller: Master Arm Position Response

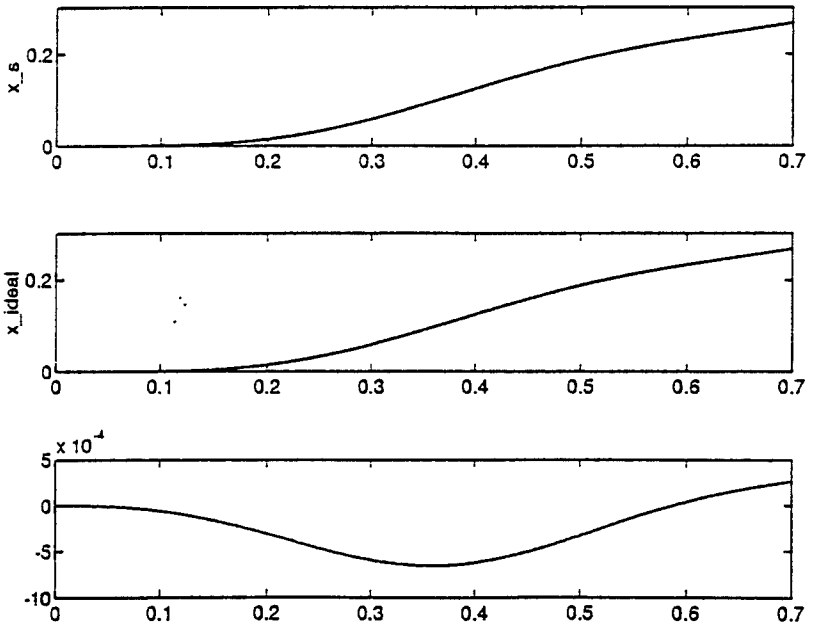


Figure 4.16 Nonadaptive Controller: Slave Arm Position Response

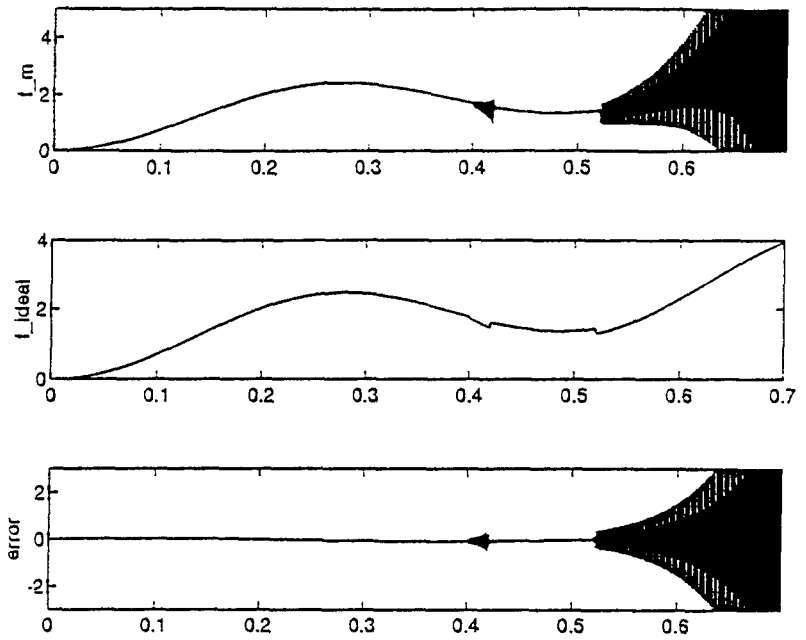


Figure 4.17 Nonadaptive Controller: Master Arm Force Response

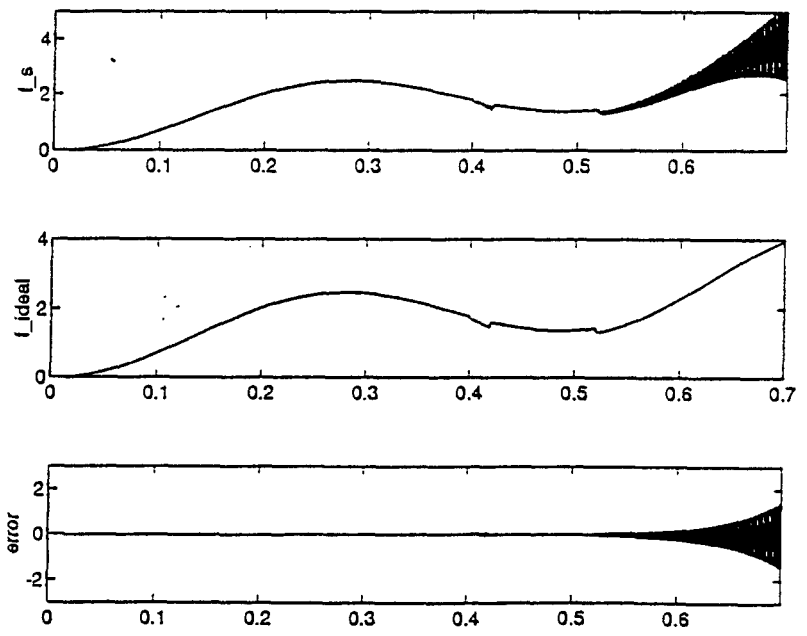


Figure 4.18 Nonadaptive Controller: Slave Arm Force Response

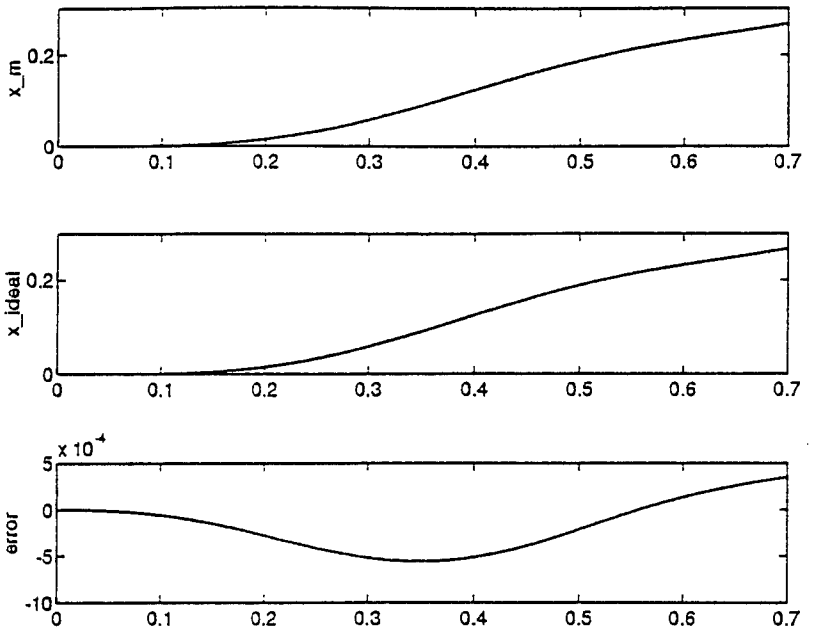


Figure 4.19 Adaptive Controller: Master Arm Position Response

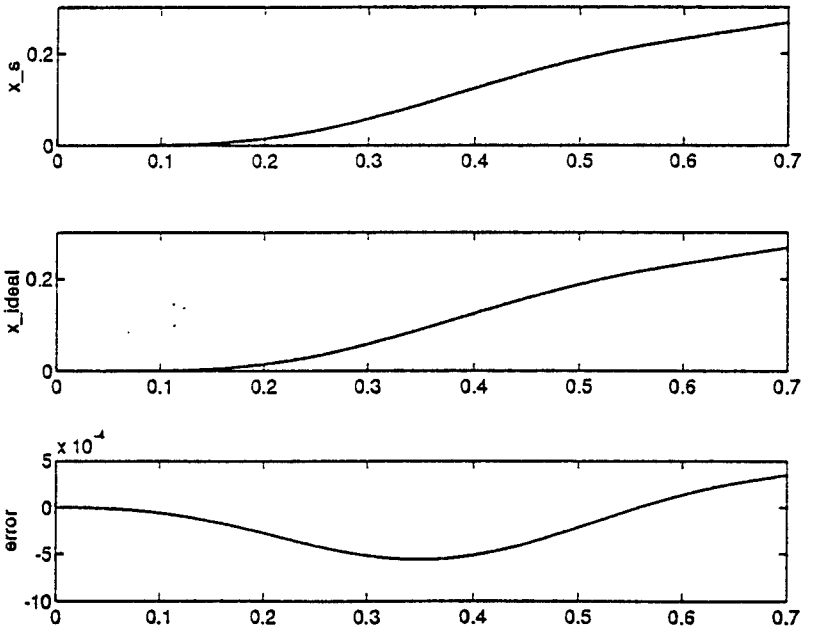


Figure 4.20 Adaptive Controller: Slave Arm Position Response

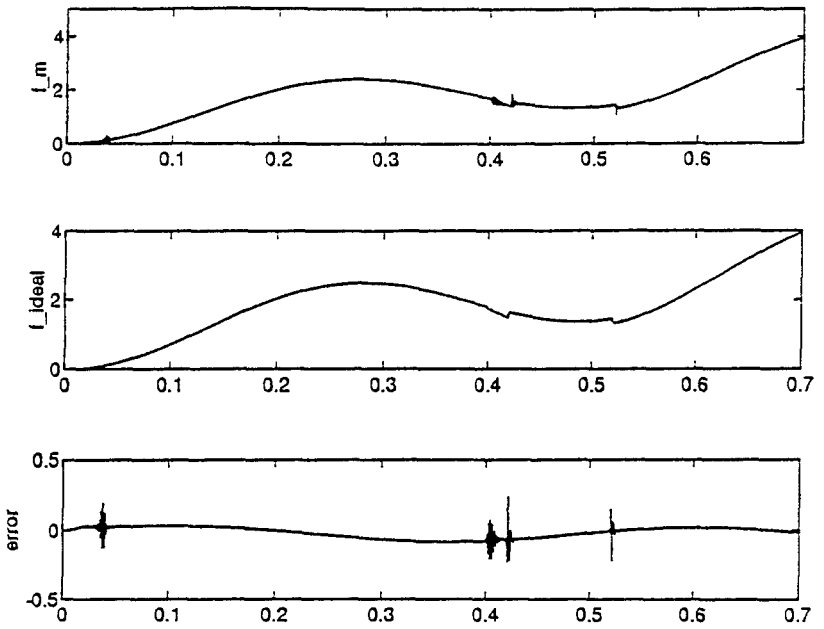


Figure 4.21 Adaptive Controller: Master Arm Force Response

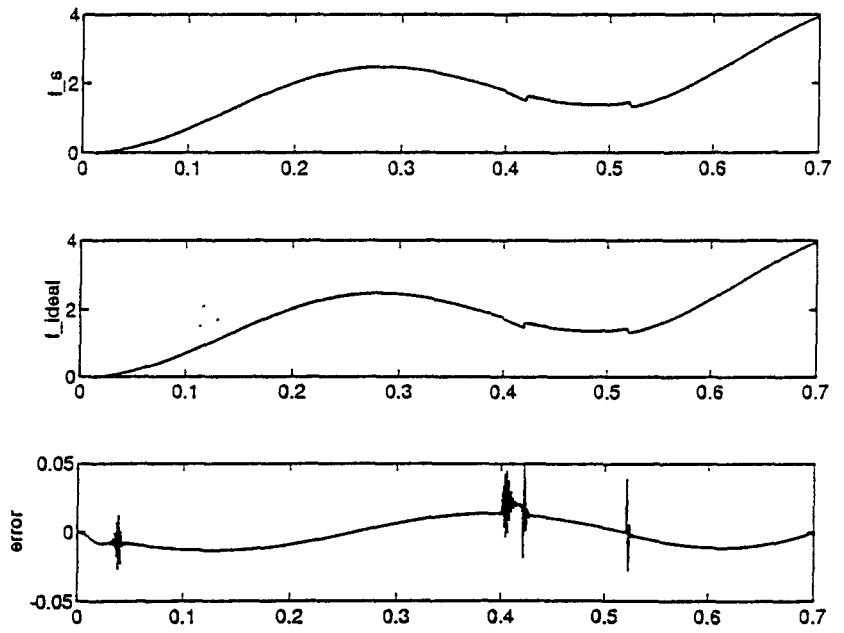


Figure 4.22 Adaptive Controller: Slave Arm Force Response

5. CONCLUSION

An adaptive control control scheme has been proposed to achieve an ideal response in a bilateral teleoperator system.

A teleoperator system with multiple DOF was discussed. For convenience. a model of a one degree-of-freedom system was chosen for the analysis. The dynamics of the entire system, including the operator and object/environment were presented. An ideal response of the system was explicitly defined. It was expressed in the framework of a two-port representation of the teleoperator system. A control scheme was developed to achieve the defined ideal response. An adaptive controller was then designed. Results of computer simulations were presented to demonstrate the performance of the proposed control scheme. The benefits of an adaptive control scheme were discussed.

The designed controller was shown by simulation results to enhance the system performance in varying and/or unpredictable environments. The scheme took into account existing operating conditions which were indicated by the estimated parameters of the two-port representation of the system. These parameters were assumed to be constant over a sampling period but could change from one sampling period to the next. They were used to specify the controller gains. When disturbances in the form of varying environment parameters were present, the performance of the control scheme was tested by simulations. The adaptive control scheme indicates an improved performance when compared to the control scheme with constant gains.

Although no stability analysis has been presented in this work, it is believed that the concept of gain selection for stability, as discussed in [22] could be extended to

the scheme proposed here. An extension of this work to include stability criterion in gain selection is suggested for further research.

In applications of teleoperator systems to flight, space, or undersea operations, one of the primary concerns has been the inherent communication delay. Even small delays in bilaterally controlled systems can potentially make the system unstable. A preliminary study was performed on developing control schemes for teleoperated systems with time delay. A literature survey was conducted to study existing schemes to cope with the time delay problem [1, 3, 5, 6, 7, 12, 14, 16, 17, 18, 19, 20].

An approach which seems quite promising in simulations for small delays is to use predictive control. In this approach, the desired trajectory for the slave manipulator is predicted at the remote site for planning. Every sample received from the master manipulator is used to continually update the predictor model. A task/environment impedance model is maintained at the master side that is used in conjunction with the operator commanded velocities to obtain the reflected force. This approach could allow a continuous loop to exist between the master and slave manipulators, making the existence of time delay almost transparent to the operator. This would allow him/her to operate comfortably without being confused or strained. Although this approach seems reasonable for small time delays further work needs to be done on this approach before it can be considered to be acceptable.