AFRL-SR-BL-TR-98-

0782

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including t ...athering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

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

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1994 Graduate Student Research Program Final Reports, Volume 7, 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: F49620-93-C-0063

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Graduate Student Research Program (USAF- GSRP) is designed to introduce university, college, and technical institute graduate students to Air Force research. This is accomplished by the graduate students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. 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, SUMMER, 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 -- 1994

GRADUATE STUDENT RESEARCH PROGRAM FINAL REPORTS

VOLUME 7

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.

December 1994

DTIC QUALITY INSPECTED 4

19981204 045

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

This document is one of a set of 16 volumes describing the 1994 AFOSR Summer Research Program. The following volumes comprise the set:

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

GSRP FINAL REPORT TABLE OF CONTENTS i-xiv

1. INTRODUCTION 1

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM 2

3. RECRUITING AND SELECTION 3

4. SITE VISITS 4

5. HBCU/MI PARTICIPATION 4

6. SRP FUNDING SOURCES 5

7. COMPENSATION FOR PARTICIPANTS 5

8. CONTENTS OF THE 1994 REPORT 6

APPENDICIES:

A. PROGRAM STATISTICAL SUMMARY A-1

B. SRP EVALUATION RESPONSES B-1

GSRP FINAL REPORTS

SRP Final Report Table of Contents

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
MS Jennifer M Ball	Wright State University , Dayton, , OH Relation Between Detection and Intelligibility in	AL/CFBA	7- 1
MR. Richard G Best	Southwest Missouri State Univ. , Springfield, , MO The Effects of Socialization on Vocational Aspirat	AL/HRMJ	7- 2
MR. Daniel A Brown	Grand Canyon University , Phoenix, , AZ Toward Modeling Higher Level Control Systems: Inc	AL/HRAU	7- 3
MS. Susan T Chitwood	Bowling Green State University , Bowling Green, , OH Further Explorations in Epistemological Space	AL/CFHP	7- 4
John E Cisneros	California State University , Los Angeles , CA Aurally Directed Search: A Comparison Between Syn	AL/CFBA	7- 5
Candace E Clary	Duke University , Durham , NC Intra-Ocular Laser Surgical Probe (ILSP) for Vitre	AL/OEO	7- 6
Robert M Colbert	Villanova University , Villanova , PA Finite Element Modeling of Manikin Necks for the A	AL/CFBV	7- 7
MR. Mark C Delgado	University of Georgia , Athens, , GA Determination of the Oxidative Redox Capacity of A	AL/EQC	7- 8
MR. Steven J Essler	North Dakota State University , Fargo, , ND Estimation of Four Arterial Vascular Parameters fo	AL/AOCIY	7- 9
Lawrence R Gottlob	Arizona State University , Tempe , AZ Accuracy Curves in a Location-Cuing Paradigm for V	AL/HRAT	7- 10
MS. Jennifer L Greenis	Michigan State University , East Lansing, , MI ITS Evaluation: A Review of the Past and Recommen	AL/HRTE	7- 11

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
MS. Patricia M Harn	University of Washington , Seattle , WA Testing R-Wise: Reading and Writing in a Supporti	AL/HRTI	7- 12
Jason E Hill	University of Scranton , Scranton , PA Rapid Bacterial DNA Fingerprintiing by the Polymer	AL/AOEL	7- 13
Mr. Rod J Hughes	Oregon Health Sciences University , Portland , OR Melatonin, Body Temperature and Sleep in Humans:	AL/CFTO	7- 14
Mr Roman G Longoria	Rice University , Houston , TX The Measurement of Work Experience: Issues and Im	AL/HRTE	7- 15
Uyen A Luong	Trinity University , San Antonio, , TX Millimeter Wave-Induced Hypertension Does Not Invo	AL/OER	7- 16
MR. Scott A Macbeth	University of Dayton , Dayton, , OH Using Electronic Brainstroming Tools to Visually	AL/HRG	7- 17
MR. Nicholas F Muto	University of Scranton , Scranton, , PA Rapid Bacterial DNA Fingerprinting by the Polymera	AL/AOEL	7- 18
Kevin P Nau	Kent State University , Kent , OH Proposal for the Establishment of a Comprehensive	AL/CFTF	7- 19
MR. Eric O Riise	Cal State Univ/Chico , Chico, , CA A Study of Interaction in Distance Learning	ALL/HRT	7- 20
Ms Heather E Roberts	Virginia Tech , Blacksburg , VA Gender and RAcial Equity of the Air Force Officer	AL/HRMM	7- 21
MR. Dale F Rucker	West Virginia University , Morgantown, , WV Improved Numerical Modeling of Groundwater Flow an	AL\EQC	7- 23

Author	University/Institution Report Title	Armstrong Laboratory Directorate	Vol-Page
Arthur M Ryan	Wright State University , Dayton , OH The Workload Assessment Monitor: Progress Towards	AL/CFHP	7- 24
Mark J Schroeder	North Dakota State University , Fargo , ND Arterial Elastance in the Maximization of External	AL/AOCIY	7- 25
MS. Katherine M Specht	Ohio State University , Columbus, , OH Tactile Perception in a Virtual Environment	AL/CFBA	7- 26
MR. Joseph M Stauffer	University of Iowa , Iowa City , IA Predicting Pilot Training Success with Logistic or	AL/HRMA	7- 27
DR. Scott G Stavrou	University of Central Arkansas , Conway, , AR Analysis of the Absorption and Metabolism of Trich	AL/OET	7- 28
MS. Virginia K Stromquist	Michigan Technological Inst. , Houghton, , MI Solid Phase Microextraction as a Method for Quanti	AL/EQC	7- 29
MR. Eric S Wieser	Trinity University , San Antonio, , TX Millimeter Wave-Induced Hypotension Does Not Invol	AL/OER	7- 30
MR. Gregory S Zinkel	Georgia Institute Technology , Atlanta, , GA A Study of the Use Predictive Modeling for Dynamic	AL/AO	7- 31

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
MR. William W Brocklehurst	University of Cincinnati , Cincinnati , OH Effect of Dissolved Gases on the Discharge Coeffic	PL/RKFA	8- 1
MR. Stephen E Clarke	Utah State University , Logan , UT REPORT NOT AVAILABLE AT PRESS TIME	PL/VTRP	8- 2
Peyman Ensaf	University of Denver , Denver , CO Influence of Model Complexity and Aeroelastic Cont	PL/WSA	8- 3
MR. Frank S Gulczinski	University of Michigan , Ann Arbor, , MI Interior Spectroscopic Investigation of Plasma Com	PL/RKCO	8- 4
MR. Derik C Herpfer	University of Cincinnati , Cincinnati , OH Drop Sizing of a Like-Impinging Element Injector I	PL/RKFA	8- 5
MR. Phillip N Hutton	Old Dominion University , Norfolk, , VA Effectiveness of Thermionic Heat Pipe Module	PL/VTPN	8- 6
MR. Robert J Leiweke	Ohio State University , Columbus , OH A Single Temperature/Material Ablation Algorithm f	PL/WSP	8- 7
MR. John I Lipp	Michigan Technological Univ , Houghton , MI Estimation of Tilts Extended Images in the Presenc	PL/LIAE	8- 8
MR. Stephen A Luker	University of Alabama , Tuscaloosa , AL Determining Cloud Coverages for Input to Thermal C	PL/GPAA	8- 9
MR. Daniel T Moriarty	MIT , Cambridge , MA Laboratory Experiments with the Versatile Toroidal	PL/GPSG	8- 10
MR. Tim C Newell	University of North Texas , Denton , TX Synchronization Using Control Chaotic Diode Resona	PL/LIDN	8- 11

SRP Final Report Table of Contents

Author	University/Institution Report Title	Phillips Laboratory Directorate	Vol-Page
Jeffrey W Nicholson	University of New Mexico , Albuquerque , NM Relaxation Processes in Gain Switched Iodine Laser	PL/LIDB	8- 12
MR. Sean R Olin	Boston University , Boston, , MA An Investigation of Flight Characteristics of the	PL/SX	8- 13
MS. Janet M Petroski	Cal State Univ/Northridge , Northridge , CA Thermoluminescence of Simple Species in Solod Mole	PL/RKFE	8- 14
MR. Aaron J Ridley	University of Michigan , Ann Arbor, , MI The Dynamic Convection Reversal Boundary	PL/GPIA	8- 15
MR. Richard M Salasovich	University of Cincinnati , Cincinnati , OH Fabrication and Mechanical Testing of Mixed-Matrix	PL/VTSC	8- 16
MR. Kevin L Scales	University of New Mexico , Albuquerque , NM A Study of Numeical Methods in Atomosperic Light P	PL/LIMI	8- 17
Greg T Sharp	University of New Mexico , Albuquerque , NM Further Studies of High Temperature Cs-Ba Tacitron	PL/WSP	8- 18
MR. Joseph M Sorci	Massachusetts Inst. Technology , Cambridge, , MA A Study of Low Frequency Weak Turbulence in a Holl	PL/GPID	8- 19
Jonathan Stohs	University of New Mexico , Albuquerque , NM Radiation Exposure of Photonic Devices	PL/VTET	8- 20
MR. Jose Suarez	Florida Inst. of Technology , Patrick AFB, , FL Meteoriod & Orbital Debris Collision Hazard Analys	PL/WSC	8- 21
MR. Tony F von Sadovszky	University of Nevada, Ren , Reno, , nv PICLL: A Portable Parallel 3D PIC Code Impleme	PL/WSP	8- 22

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
MR. John C Bertot	Syracuse University , Syracuse, , NY Transferring Technology Via the Internet	RL/XPP	9- 1
MR. Jerry M Couretas	University of Arizona , Tucson, , AZ Analysis of Extraction and Aggregation Techniques	RL/IRDO	9- 2
MR. Frederick L Crabbe	Univ California/Los Angeles , Los Angeles, , CA Using Self-Organization to Develop Vector Represen	RL/IRDO	9- 3
MS. Julie Hsu	Louisiana State University , Baton Rouge, , LA An Assignment Based Approach to Parallel-Machine S	RL/C3CA	9- 4
Andrew J Laffely	Univ. of Maine , Orono , ME Automatic Extraction of Drainage Network from Di	RL/IRRP	9- 5
MR. Daniel K Lee	Southern Illinois University , Carbondale , IL Mutual Coupling Effect of Square Microstrip Patch	RL/ERA	9- 6
Slawomir J Marcinkowski	Syracuse University , Syracuse , NY Transferring Technology Via the Internet	RL/XPP	9- 7
MR. Sean S O'Keefe	Cornell University , Ithaca, , NY Integration of Optoelectronic Devices with Microwa	RL/OCBP	9- 8
MR Robert L Popp	University of Connecticut , Storrs , CT Multisensor-Multitarget Data Fusion Using an S-Dim	RL/OCTM	9- 9
MR. Steven J Pratt	Syracuse University , Syracuse, , NY Analysis and Comparison of the Performance/Life Co	RL/C3AB	9- 10
Francis X Reichmeyer	Syracuse University , Syracuse , NY Local Area ATM Network Interfaces	RL/C3AB	9- 11

Author	University/Institution Report Title	Rome Laboratory Directorate	Vol-Page
MR. David H Sackett	Rochester Institute of Technol , Rochestter , NY Self-Sustained Pulsation and High Speed Optical Ne	RL/OCPA	9- 12
ME. Paul E Shames	Univiersity of Cal/San Diego , San Diego, , CA A Study of High Speed Polarization Rotators for Us	RL/IRAP	9- 13
Terrence W Towe	Univ of Arkansas-Fayetteville , Fayetteville , AR REPORT NOT AVAILABLE AT PRESS TIME	RL/ERX	9- 14
Okechukwu C Ugweje	Florida Atlantic University , Boca Raton, , FL A Program Plan for Transmitting High-Data-Rate ATM	RL/C3BA	9- 15
MR. Martin A Villarica	Syracuse University , Syracuse, , NY A Study of the Application of Fractuals and Kineti	RL/ERDR	9- 16
Stanley J Wenndt	Colorado State University , Fort Collins , CO An Investigation of Cepstrum Based Speaker Identif	RL/IRAA	9- 17

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
MS. Terri L Alexander	University of Central Florida , Orlando, , FL Design of Spectroscopic Material-Characterization	WL/MNGS	10- 1
MR. Joseph L Binford, III	University of Dayton , Dayton, , OH Thermal Stability Apparatus Design and Error Analy	WL/MLPO	10- 2
MR Jonathan A Bishop	University of Oklahoma , Norman, , OK Influence of Model Complexity and Aeroelastic Cons	WL/FIBR	10- 3
MR. Steven P Burns	Purdue University , West Lafayette, , IN The Use of Pressure Sensitive Paints on Rotating M	WL/POTF	10- 4
MR. Lance H Carter	University of Texas/Austin , Austin , TX Gain-Scheduled Bank-to-Turn Autopilot Design Using	WL/MNAG	10- 5
MR. Peng Chen	University of North Texas , Denton , TX Synthesis & Characterization of Lanthanum Phosphat	WL/MLLM	10- 6
MS. Lora A Cintavey	University of Cincinnati , Cincinnati , OH Processing and Characterization of Nonlinear Optic	WL/MLBP	10- 7
MR. William K Cope	University of Illinois/Urbana , Urbana, , IL Assessment of Gasp for the Simulation of Scramjet	WL/POPT	10- 8
Mike J Cutbirth	Oklahoma State University , Stillwater , OK A Study of the Heat Transfer for the High Flux Hea	WL/POOS	10- 9
Craig M Files	University of Idaho , Moscow , ID Using a Search Heuristic in an NP-Complete Problem	WL/AART-	10- 10
MR Edward M Friel	University of Dayton , Dayton , OH Direction Finding in the Presence of a Near Field	WL/AARM	10- 11

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
MR. Keith D Grinstead	Purdue University , West Lafayette, , IN Obtaininng the Correction Factors for Two-Photon I	WL/POSF	10- 12
MR. Jason J Hugenroth	Louisiana State University , Baton Rouge, , LA A Research Plan for Evaluating Wave Gun as a Low-L	WL/MNAA	10- 13
MR. Andrew Kager	University of Central Florida , Orlando, , FL A Numerical Study of the Effect of Base and Collec	WL/ELRA	10- 14
MR. John C Lewis	University of Kentucky , Lexington , KY A Theory for the Testing of Materials Under Combin	WL/MNM	10- 15
MR. Kenneth P Luke	Wright State University , Dayton, , OH Three-Dimensional Modeling Using a Calibrated Came	WL/AARF-	10- 16
John D Mai	Univ of Calif-Los Angeles , Los Angeles , CA Preliminary Characterization and Calibration of Mi	WL/FIME	10- 17
MR Mark A Manzardo	Univ of Alabama-Huntsville , Huntsville , AL High Speed Imaging Infrared Polarimetry	WL/MNGS	10- 18
MR. David B Maring	University of FloridaFL , Gainesville, , FL Fabrication and White-Light Characterization of An	WL/MNG	10- 19
MR. Joseph R Miramonti	University of Missouri/Columbi , Columbia, a , MO Scanning Image Alegbra Networks for Vehicle Identi	WL/MNGA	10- 20
MS. Jennifer S Naylor	Auburn University , Auburn, , AL Automation Control Issues in the Development of an	WL/MNAG	10- 21
MR. Ned F O'Brien	University of Dayton , Dayton, , OH A Study of RF Fiber Optic Communication Link Techn	WL/AAAI-	10- 22

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
MR. Tae W Park	Univ of Illinois/Chicago , Chicago, , IL A Numerical Study of DropleVortx Interactions in a	WL/POSF	10- 23
MS Margaret F Pinnell	University of Dayton , Dayton , OH A Parametric Study of the Factors Affecting the Op	WL/MLBM	10- 24
MR. Seth M Pinsky	Oregon Grad Inst. Sci & Tech , Portland, , OR S-Parameter Measurements on a GaAsFET Variable-Gai	WL/ELM	10- 25
MR. David A Ress	Tennessee Technological Univ , Cookeville, , TN Modifications To The Thinker Discovery System	WL/MLIM	10- 26
Mohammed A Samad	University of New Orleans , New Orleans , LA A Study of Delamination Damage and Energy Exchange	WL/FIVS	10- 27
MR. Robert W Slater III	University of Cincinnati , Cincinnati , OH Low-Velocity Impact of Moisture-Conditioned Lamina	WL/FIBE	10- 28
Ed P Socci	University of Virginia , Charlottesville , VA X-ray Diffraction Study of Siloxane/Cholestrol Bas	WL/MLPO	10- 29
MR. Edward A Thompson	University of Cincinnati , Cincinnati, , OH Annealed Fuzzy Control for a Self-Tuning Piezolect	WL/FIBG	10- 30
MR. Brent A Veltkamp	Michigan State University , East Lansing, , MI Pixel Plane Design for SIMD Graphic Processor	WL/ELED	10- 31
MR. Christopher C Vogt	University of Cal San Diego , San Diego, , CA The Combinatorics of Function Decomposition and Ap	WL/AART-	10- 32
MR. Ralph J Volino	University of Minnesota , Minneapolis, , MN Documentation of Boundary Layer Characteristics fo	WL/POTT	10- 33

Author	University/Institution Report Title	Wright Laboratory Directorate	Vol-Page
MR. Jeffrey A Walrath	University of Cincinnati , Cincinnati, , OH Integration of Champ Firm Macro Library with DSS	WL/AAAT-	10- 34

Author	University/Institution Report Title	Arnold Engineering Development Center Directorate	Vol-Page
MR. William A Alford	Univ. Tennessee Space Inst. , Tullahoma, , TN Design of Testing and Debugging Software for C31 N	Sverdrup	11- 1
MR. Brian C DeAngelis	University of Illinois/Urbana- , Urbana , IL Performance and Validation Studies of the Kiva-II	SVERDRU	11- 2
MR. Harold D Helsley	Univ Tennessee Space Inst , Tullahoma, , TN Development of a Monitor for a Multi-Processor Net	Sverdrup	11- 3
MR. Christopher W Humphres	University of Alabama , Tuscaloosa, , AL Parallelization of Chimera Utilizing PVM	Calspan	11- 4
MR. Curtis S Mashburn	Universtity of Tennessee Space , Tullahoma, , TN Prediction of the Performance of a 7-Stage Axial-F	Sverdrup	11- 5
MR. William S Meredith	Univ of Tennesse Sp. Inst. , Tullhoma, , TN REPORT NOT AVAILABLE AT PRESS TIME	Calspan	11- 6
MR. Michael S Moore	Vanderbilt University , Nashville , TN A Model Based Real Time Image Processing System	SVERDRU	11- 7
MR David T Pratt	Univ of Tennessee Space Inst , Tullahoma , TN Analysis and Comparison of the Performance/Life Co	Sverdrup	11- 8
MR. David B Underhill	Univ of Tenn Space Inst , Tullahoma , TN Application of Vorticity Confinement to a Delta Wi	Calspan	11- 9

SRP Final Report Table of Contents

Author	University/Institution Report Title	Frank J Seiler Research Laboratory Directorate	Vol-Page
MR. Christian S Bahn	Colorado School of Mines , Golden, , CO A Theoretical Study of Lithium and Molten Salt Gra	FJSRL/NC	11- 10
MR. Antonio M Ferreira	Memphis State University , Memphis, , TN Theoretical Investigations of the NLO Properties o	FJSRL/NC	11- 11
MS Joan Fuller	University of Alabama , Tuscaloosa , AL Investigations of Carbon Materials in Alkali Metal	FJSRL/CD	11- 12

Author	University/Institution Report Title	Wilford Hall Medical Center Directorate	Vol-Page
MR Ramachandra P Tummala	University of Miami , Coral Gables , FL Effects of Temperature on Various Hematological P	WHMC/RD	11- 13

1. INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

The numbers of projected summer research participants in each of the three categories are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest; and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1994 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After completion of the 1990 contract, RDL won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

Table 1: SRP Participation, by Year

YEAR	Number of Participants			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years; in one case a laboratory did not fund any additional associates. However, the table shows that, overall, the number of participating associates increased this year because two laboratories funded more associates than they had in previous years.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 44-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous (over 600 annually) individual requesters.

Due to a delay in awarding the new contract, RDL was not able to place advertisements in any of the following publications in which the SRP is normally advertised: *Black Issues in Higher Education, Chemical & Engineering News, IEEE Spectrum* and *Physics Today*.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1994 are shown in this table.

Table 2: 1994 Applicants and Participants

PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	600	192	30
(HBCU/MI)	(90)	(16)	(7)
GSRP	322	117	11
(HBCU/MI)	(11)	(6)	(0)
HSAP	562	133	14
TOTAL	1484	442	55

4. SITE VISITS

During June and July of 1994, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

In previous years, an RDL program representative visited from seven to ten different HBCU/MIs to promote interest in the SRP among the faculty and graduate students. Due to the late contract award date (January 1994) no time was available to visit HBCU/MIs this past year.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, seven HBCU/MI SFRPs declined after they were selected. The following table records HBCU/MI participation in this program.

Table 3: SRP HBCU/MI Participation, by Year

YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6

6. SRP FUNDING SOURCES

Funding sources for the 1994 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1994 SRP selected participants are shown here.

Table 4: 1994 SRP Associate Funding

FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	150	98*1	121*2
USAF Laboratory Funds	37	19	12
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	5	0	0
TOTAL	192	117	133

*1 - 100 were selected, but two canceled too late to be replaced.
*2 - 125 were selected, but four canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

Table 5: 1994 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994
Faculty Members	$690	$718	$740	$740
Graduate Student (Master's Degree)	$425	$442	$455	$455
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391
High School Student (First Year)	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $37/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 58% (178 out of 309) claimed travel reimbursements at an average round-trip cost of $860.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-one percent (78 out of 192) of faculty associates took orientation trips at an average cost of $498. Many faculty associates noted on their evaluation forms that due to the late notice of acceptance into the 1994 SRP (caused by the late award in January 1994 of the contract) there wasn't enough time to attend an orientation visit prior to their tour start date. In 1993, 58 % of SFRP associates took orientation visits at an average cost of $685.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1994 REPORT

The complete set of reports for the 1994 SRP includes this program management report augmented by fifteen volumes of final research reports by the 1994 associates as indicated below:

Table 6: 1994 SRP Final Report Volume Assignments

	VOLUME		
LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, FJSRL, WHMC	6	11	16

AEDC = Arnold Engineering Development Center
FJSRL = Frank J. Seiler Research Laboratory
WHMC = Wilford Hall Medical Center

APPENDIX A -- PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP and GSRP associates represent 158 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 46 states plus Washington D.C. and Puerto Rico. Selectees represent 40 states.

GSRP - Applicants came from 46 states and Puerto Rico. Selectees represent 34 states.

HSAP - Applicants came from fifteen states. Selectees represent ten states.

C. Academic Disciplines Represented

The academic disciplines of the combined 192 SFRP associates are as follows:

Electrical Engineering	22.4%
Mechanical Engineering	14.0%
Physics: General, Nuclear & Plasma	12.2%
Chemistry & Chemical Engineering	11.2%
Mathematics & Statistics	8.1%
Psychology	7.0%
Computer Science	6.4%
Aerospace & Aeronautical Engineering	4.8%
Engineering Science	2.7%
Biology & Inorganic Chemistry	2.2%
Physics: Electro-Optics & Photonics	2.2%
Communication	1.6%
Industrial & Civil Engineering	1.6%
Physiology	1.1%
Polymer Science	1.1%
Education	0.5%
Pharmaceutics	0.5%
Veterinary Medicine	0.5%
TOTAL	100%

Table A-1. Total Participants

Number of Participants	
SFRP	192
GSRP	117
HSAP	133
TOTAL	442

Table A-2. Degrees Represented

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	189	0	189
Master's	3	47	50
Bachelor's	0	70	70
TOTAL	192	117	309

Table A-3. SFRP Academic Titles

Academic Titles	
Assistant Professor	74
Associate Professor	63
Professor	44
Instructor	5
Chairman	1
Visiting Professor	1
Visiting Assoc. Prof.	1
Research Associate	3
TOTAL	192

Table A-4. Source of Learning About SRP

SOURCE	SFRP		GSRP	
	Applicants	Selectees	Applicants	Selectees
Applied/participated in prior years	26%	37%	10%	13%
Colleague familiar with SRP	19%	17%	12%	12%
Brochure mailed to institution	32%	18%	19%	12%
Contact with Air Force laboratory	15%	24%	9%	12%
Faculty Advisor (GSRPs Only)	--	--	39%	43%
Other source	8%	4%	11%	8%
TOTAL	100%	100%	100%	100%

Table A-5. Ethnic Background of Applicants and Selectees

	SFRP		GSRP		HSAP	
	Applicants	Selectees	Applicants	Selectees	Applicants	Selectees
American Indian or Native Alaskan	0.2%	0%	1%	0%	0.4%	0%
Asian/Pacific Islander	30%	20%	6%	8%	7%	10%
Black	4%	1.5%	3%	3%	7%	2%
Hispanic	3%	1.9%	4%	4.5%	11%	8%
Caucasian	51%	63%	77%	77%	70%	75%
Preferred not to answer	12%	14%	9%	7%	4%	5%
TOTAL	100%	100%	100%	100%	99%	100%

Table A-6. Percentages of Selectees receiving their 1st, 2nd, or 3rd Choices of Directorate

	1st Choice	2nd Choice	3rd Choice	Other Than Their Choice
SFRP	70%	7%	3%	20%
GSRP	76%	2%	2%	20%

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	116
USAF Laboratory Focal Points	109
USAF Laboratory HSAP Mentors	54

All groups indicate near-unanimous enthusiasm for the SRP experience.

Typical comments from 1994 SRP associates are:

> "[The SRP was an] excellent opportunity to work in state-of-the-art facility with top-notch people."
>
> "[The SRP experience] enabled exposure to interesting scientific application problems; enhancement of knowledge and insight into 'real-world' problems."
>
> "[The SRP] was a great opportunity for resourceful and independent faculty [members] from small colleges to obtain research credentials."
>
> "The laboratory personnel I worked with are tremendous, both personally and scientifically. I cannot emphasize how wonderful they are."
>
> "The one-on-one relationship with my mentor and the hands on research experience improved [my] understanding of physics in addition to improving my library research skills. Very valuable for [both] college and career!"

Typical comments from laboratory focal points and mentors are:

"This program [AFOSR - SFRP] has been a 'God Send' for us. Ties established with summer faculty have proven invaluable."

"Program was excellent from our perspective. So much was accomplished that new options became viable "

"This program managed to get around most of the red tape and 'BS' associated with most Air Force programs. Good Job!"

'Great program for high school students to be introduced to the research environment. Highly educational for others [at laboratory]."

'This is an excellent program to introduce students to technology and give them a feel for [science/engineering] career fields. I view any return benefit to the government to be 'icing on the cake' and have usually benefitted."

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below (Note: basically the same as in previous years.)

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Laboratory sponsor seminar presentations of work conducted by associates, and/or organized social functions for associates to collectively meet and share SRP experiences.

C. Laboratory focal points collectively suggest more AFOSR allocated associate positions, so that more people may share in the experience.

D. Associates collectively suggest higher stipends for SRP associates.

E. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same arguement was used to raise the faculty and graduate student participation time a few years ago.)

2. 1994 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 109 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	10	30	50	0	20	50	40	0	10	40	60	0	0
AL	44	34	50	6	9	54	34	12	0	56	31	12	0
FJSRL	3	33	33	33	0	67	33	0	0	33	67	0	0
PL	14	28	43	28	0	57	21	21	0	71	28	0	0
RL	3	33	67	0	0	67	0	33	0	100	0	0	0
WHMC	1	0	0	100	0	0	100	0	0	0	100	0	0
WL	46	15	61	24	0	56	30	13	0	76	17	6	0
Total	**121**	**25%**	**43%**	**27%**	**4%**	**50%**	**37%**	**11%**	**1%**	**54%**	**43%**	**3%**	**0%**

LFP Evaluation Summary. The summarized repsonses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5. a. Benefits of associate's work to laboratory:
    b. Benefits of associate's work to Air Force:
6. a. Enhancement of research qualifications for LFP and staff:
    b. Enhancement of research qualifications for SFRP associate:
    c. Enhancement of research qualifications for GSRP associate:
7. a. Enhancement of knowledge for LFP and staff:
    b. Enhancement of knowledge for SFRP associate:
    c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10. a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
    b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Laboratory Focal Point Reponses to above questions

	AEDC	***AL***	***FJSRL***	***PL***	***RL***	***WHMC***	***WL***
= Evals Recv'd	10	32	3	14	3	1	46
Question =							
2	90 %	62 %	100 %	64 %	100 %	100 %	83 %
2a	**3.5**	**3.5**	4.7	4.4	4.0	4.0	**3.7**
2b	4.0	3.8	4.0	4.3	4.3	4.0	3.9
3	4.2	3.6	4.3	3.8	4.7	4.0	4.0
4	3.8	3.9	4.0	4.2	4.3	NO ENTRY	4.0
5a	4.1	4.4	4.7	4.9	4.3	3.0	4.6
5b	4.0	4.2	4.7	4.7	4.3	3.0	4.5
6a	**3.6**	4.1	**3.7**	4.5	4.3	**3.0**	4.1
6b	3.6	4.0	4.0	4.4	4.7	3.0	4.2
6c	3.3	4.2	4.0	4.5	4.5	3.0	4.2
7a	3.9	4.3	4.0	4.6	4.0	3.0	4.2
7b	4.1	4.3	4.3	4.6	4.7	3.0	4.3
7c	3.3	4.1	4.5	4.5	4.5	5.0	4.3
8	4.2	4.3	5.0	4.9	4.3	5.0	4.7
9	3.8	4.1	4.7	5.0	4.7	5.0	4.6
10a	4.6	4.5	5.0	4.9	4.7	5.0	4.7
10b	4.3	4.2	5.0	4.3	5.0	5.0	4.5
11	4.1	4.5	4.3	4.9	4.7	4.0	4.4
12	4.1	3.9	4.0	4.4	4.7	3.0	4.1
13a	**3.8**	**2.9**	4.0	4.0	4.7	**3.0**	**3.6**
13b	**3.8**	**2.9**	4.0	4.3	4.7	**3.0**	**3.8**
14	4.5	4.4	5.0	4.9	4.7	4.0	4.5

3. 1994 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 275 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

1. The match between the laboratories research and your field:		4.6
2. Your working relationship with your LFP:		4.8
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.5
5. Lab readiness for you: LFP, task, plan:		4.3
6. Lab readiness for you: equipment, supplies, facilities:		4.1
7. Lab resources:		4.3
8. Lab research and administrative support:		4.5
9. Adequacy of brochure and associate handbook:		4.3
10. RDL communications with you:		4.3
11. Overall payment procedures:		**3.8**
12. Overall assessment of the SRP:		4.7
13. a. Would you apply again?	Yes:	85%
b. Will you continue this or related research?	Yes:	95%
14. Was length of your tour satisfactory?	Yes:	86%
15. Percentage of associates who engaged in:		
a. Seminar presentation:		52%
b. Technical meetings:		32%
c. Social functions:		03%
d. Other		01%

16. Percentage of associates who experienced difficulties in:

a. Finding housing: 12%
b. Check Cashing: 03%

17. Where did you stay during your SRP tour?

a. At Home: 20%
b. With Friend: 06%
c. On Local Economy: 47%
d. Base Quarters: 10%

THIS SECTION FACULTY ONLY:

18. Were graduate students working with you? Yes: 23%

19. Would you bring graduate students next year? Yes: 56%

20. Value of orientation visit:

Essential: 29%
Convenient: 20%
Not Worth Cost: 01%
Not Used: 34%

THIS SECTION GRADUATE STUDENTS ONLY:

21. Who did you work with:

University Professor: 18%
Laboratory Scientist: 54%

4. 1994 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

The summarized results listed below are from the 54 mentor evaluations received.

1. Mentor apprentice preferences:

Table B-3. Air Force Mentor Responses

		How Many Apprentices Would You Prefer To Get ?			
		HSAP Apprentices Preferred			
Laboratory	***# Evals Recv'd***	***0***	**1**	**2**	**3+**
AEDC	6	0	100	0	0
AL	17	29	47	6	18
PL	9	22	78	0	0
RL	4	25	75	0	0
WL	18	22	55	17	6
Total	**54**	**20%**	**71%**	**5%**	**5%**

Mentors were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

2. Mentors involved in SRP apprentice application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Laboratory's preparation for apprentice:
4. Mentor's preparation for apprentice:
5. Length of research tour:
6. Benefits of apprentice's work to U.S. Air force:
7. Enhancement of academic qualifications for apprentice:
8. Enhancement of research skills for apprentice:
9. Value of U.S. Air Force/high school links:
10. Mentor's working relationship with apprentice:
11. Expenditure of mentor's time worthwhile:
12. Quality of program literature for apprentice:
13. a. Quality of RDL's communications with mentors:
    b. Quality of RDL's communication with apprentices:
14. Overall assessment of SRP:

	AEDC	*AL*	*PL*	*RL*	*WL*
# Evals Recv'd	**6**	**17**	**9**	**4**	**18**
Question #					
2	100 %	76 %	56 %	75 %	61 %
2a	4.2	4.0	**3.1**	**3.7**	**3.5**
2b	4.0	4.5	4.0	4.0	3.8
3	4.3	**3.8**	**3.9**	**3.8**	**3.8**
4	4.5	**3.7**	**3.4**	4.2	**3.9**
5	**3.5**	4.1	**3.1**	**3.7**	**3.6**
6	4.3	3.9	4.0	4.0	4.2
7	4.0	4.4	4.3	4.2	3.9
8	4.7	4.4	4.4	4.2	4.0
9	4.7	4.2	3.7	4.5	4.0
10	4.7	4.5	4.4	4.5	4.2
11	4.8	4.3	4.0	4.5	4.1
12	4.2	4.1	4.1	4.8	3.4
13a	**3.5**	**3.9**	**3.7**	4.0	**3.1**
13b	4.0	4.1	3.4	4.0	3.5
14	4.3	4.5	3.8	4.5	4.1

5. 1994 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 116 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

1. Match of lab research to you interest:		3.9
2. Apprentices working relationship with their mentor and other lab scientists:		4.6
3. Enhancement of your academic qualifications:		4.4
4. Enhancement of your research qualifications:		4.1
5. Lab readiness for you: mentor, task, work plan		**3.7**
6. Lab readiness for you: equipment supplies facilities		4.3
7. Lab resources: availability		4.3
8. Lab research and administrative support:		4.4
9. Adequacy of RDL's apprentice handbook and administrative materials:		4.0
10. Responsiveness of RDL's communications:		**3.5**
11. Overall payment procedures:		**3.3**
12. Overall assessment of SRP value to you:		4.5
13. Would you apply again next year?	Yes:	88%
14. Was length of SRP tour satisfactory?	Yes:	78%
15. Percentages of apprentices who engaged in:		
a. Seminar presentation:		48%
b. Technical meetings:		23%
c. Social functions:		18%

RELATION BETWEEN DETECTION AND INTELLIGIBILITY IN FREE-FIELD MASKING

Robert H. Gilkey
Assistant Professor
and
Jennifer M. Ball
Graduate Research Assistant
Department of Psychology

Wright State University
Dayton, OH 45435

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Laboratory

October 1994

RELATION BETWEEN DETECTION AND INTELLIGIBILITY IN FREE-FIELD MASKING

Robert H. Gilkey
Assistant Professor
and
Jennifer M. Ball
Graduate Research Assistant
Department of Psychology

ABSTRACT

Experimental and theoretical studies are investigating spatial hearing by measuring signal detectability and speech intelligibility in the free field. The research emphasizes the impact of interfering auditory stimulation on spatial hearing performance. Studies that examine the detectibility of signals as a function of their spatial relation to a masker will be used to predict the intelligibility of masked speech. The frequency-dependent role of specific acoustic cues for mediating detection and recognition performance will be addressed. This research will have direct relevance for basic science by delineating the acoustic cues and potential mechanisms underlying spatial hearing phenomena. The results will also have relevance to the design of auditory displays and virtual realities by specifying how the spatial distribution of sounds influences the ability of listeners to detect and understand auditory signals.

RELATION BETWEEN DETECTION AND INTELLIGIBILITY IN FREE-FIELD MASKING

Robert H. Gilkey
and
Jennifer M. Ball

INTRODUCTION

The overall goal of our program of research is to determine the acoustic cues that underlie the spatial hearing abilities of human listeners. The work described here directly compares performance in detection and speech intelligibility tasks, to determine whether intelligibility results can be predicted from detection data. Experimental conditions will include both horizontal and vertical separations between signal and masker. The results of these experiments will help to establish a standard for predicting and evaluating the detectability and intelligibility of signals in auditory displays and virtual environments.

Cherry (1953) coined the term "cocktail-party" effect to describe the ability of a listener to "hear out" a particular sound in the presence of other competing sounds, a situation that might be encountered while trying to listen to a particular conversation at a cocktail party. Cherry believed that the spatial distribution of the sounds was a critical factor underlying this effect. That is, the signal (the message to which the listener is trying to attend) will be easier to hear when it emanates from a spatial location that is different than those of the maskers (the interfering sounds that the listener is trying to ignore). This relation between the spatial parameters of the stimuli and the ability to hear a particular stimulus has been of great interest and

importance to both basic and applied scientists. Basic scientists have routinely employed masking tasks to answer questions about how the auditory system analyzes and represents information; in the same way, masking experiments can provide important information about how the auditory system analyzes the essentially non-spatial peripheral representation of auditory information into a three-dimensional perceptual representation of auditory space. Applied scientists have sought to realize performance gains by introducing spatial information into auditory displays.

Although relatively few studies have directly examined the influence of the spatial distribution of the sounds on the ability to detect and understand auditory information, there is an extensive literature of headphone-based studies that have examined "analogous" stimulus situations (see Durlach and Colburn, 1978, and Colburn and Durlach, 1978, for reviews). This research has emphasized the role of interaural differences in determining the observers' ability to perceive auditory signals. For example, the detectability of a low-frequency signal can be increased by as much as 15 dB when the interaural parameters of the signal are different from the interaural parameters of the masker. This change in detectibility, relative to the case where the interaural parameters of both the signal and the masker are the same, is known as the Binaural Masking Level Difference (BMLD). Although the importance of these interaural cues in mediating the cocktail-party effect has often been touted, there are relatively few studies that have directly examined the relation between these BMLD experiments and the performance of subjects in a free-field masking task.

Whereas most of the headphone-based literature has focused on detection tasks, the small free-field masking literature has mainly focused on the intelligibility of the speech signals as a function of the spatial separation between the signal and the masker. Plomp (1976) investigated the intelligibility of speech presented from a single speaker directly in front of the listener as a function of the spatial location of a noise or speech masker. He found that the intelligibility threshold for the speech could be decreased by as much as 5 to 6 dB by spatially separating the signal from the masker. Although he found an advantage for two-eared listening of about 2.5 dB across all of his conditions, the advantage was not systematically related to the signal and masker separation. Bronkhorst and Plomp (1988) had subjects listen to binaural recordings made through the KEMAR manikin. In their experiments, the signal was presented from a speaker directly in front of the manikin and the masker could originate from various locations within the horizontal plane, surrounding the manikin in azimuth. They were able to use signal processing techniques to systematically manipulate the interaural information available to the listener. They found maximum increases in intelligibility of about 10 dB when the signal and masker were separated by 90°. By systematically manipulating the interaural parameter of the signal, they showed that 7-8 dB of the increase resulted from the head-shadow effect, whereas, only 2-3 dB of the increase resulted from interaural time differences. Further analysis showed that most of the head-shadow effect resulted from having an ear placed where the signal-to-noise ratio was favorable, and not from the interaural level differences per se. Zurek (1992) reviewed and modeled the data from a number of intelligibility studies and concluded that about 3

dB of the average 5-dB "binaural advantage" (the increase in intelligibility when listening with two ears instead of only one ear) observed in these studies, resulted because one of the ears was positioned where the effective signal-to-noise ratio was favorable. Only about 2 dB of the observed binaural advantage resulted from binaural interaction (i.e., the use of interaural time differences and interaural level differences).

The few studies that have investigated the detectability of masked signals in the free field have not indicated a large role for binaural interaction either. Doll, Hanna, and Russotti (1992) investigated the detectability of an amplitude-modulated 500-Hz tone presented from a speaker that was centered between two symmetrically placed (with respect to the median plane) noise sources. They found that the detectability of the signal increased by only about 3 dB as the noise sources were separated from the signal in azimuth (separations in elevation were not considered).

Saberi, Dostal, Sadralodabai, Bull, and Perrott (1991) considered both horizontal and vertical separation between the signal and a single masker. They found that the detectability of a broadband click-train signal increased by as much as 15-18 dB when it was separated from a Gaussian noise in azimuth. The detectability of the signal could be increased by as much as 6 dB when the signal and masker were vertically separated within the median plane. The changes in detectability with separations in azimuth could have been mediated by a variety of potential acoustic cues, including changes in interaural parameters. On the other hand, the changes in detectability with vertical separations are unlikely to have

been based on changes in interaural parameters, because the interaural differences for all locations in the median plane are minimal.

Good and Gilkey (1992) and Gilkey and Good (1994) extended the findings of Saberi et al. (1991) by band-limiting both the signal and the masker to lie within low- (below 1.4 kHz), mid- (1.2 to 6.8 kHz), or high- (above 3.5 kHz) frequency regions. These frequency regions were chosen because work on sound localization indicated that the effectiveness of interaural time cues is greatest in the low-frequency region, that the effectiveness of interaural level differences is greatest in the mid-frequency region and perhaps the high-frequency region, and that the effectiveness of spectral modulations introduced by the pinnae are greatest in the high-frequency region. They found that in all conditions the changes in detectability with spatial separations were as large or larger in the high-frequency region as they were in the mid-frequency region or the low-frequency region. Traditional models of binaural masking, based on interaural differences, did not predict the increases in detectability observed with vertical separations within the median plane. Moreover, these models seem inadequate to explain the effects of stimulus frequency, because the increase in the magnitude of the interaural level difference with increasing frequency was not great enough to predict the observed improvement in performance between mid-frequency and high-frequency conditions.

Gilkey, Good, and Ball (1994) compared the effects of spatial separations for "real" and "virtual" sounds, in order to determine the relative importance of monaural and binaural cues for detection. The virtual sounds were generated by

passing the source waveforms through head-related transfer functions, which reproduced the direction-specific filtering of the head and pinnae that would be present in a real sound field. Because the stimuli were presented through headphones, monaural and binaural presentations could be compared by merely turning off one channel. Although there was some evidence suggesting a small role for interaural cues at low frequencies, in most cases the best monaural performance was as good as binaural performance, suggesting that the increases in detectability observed in the free field, by Gilkey and Good (1994) and others, could have been mediated by monaural changes in the effective signal-to-noise ratio, rather than by changes in interaural information.

Overall, the results of these detection studies indicate that reductions in masking on the order of 8 to 18 dB can be observed in free-field masking situations when the signal and the masker are spatially separated. Both horizontal and vertical separations can lead to substantial masking reductions. The pattern of results from these experiments emphasizes the importance of high-frequency monaural information.

Although one might expect that speech intelligibility scores could be predicted from detection performance, few studies have measured both detection and intelligibility thresholds on the same subjects. In general, the results from studies in this literature have been limited in two ways: 1) Only a relatively limited set of signal and masker spatial configurations have been examined, specifically those involving spatial separations within the horizontal plane; 2) The results from intelligibility studies have not been directly compared to those from detection

studies; moreover, the frequency range of the speech signals has typically not been manipulated in a way that would allow detailed consideration of the relation between the detectability of individual acoustic cues and the intelligibility of the speech signals.

The research reported here is examining the relation between detection and intelligibility results in the free field and determining the degree to which intelligibility depends on the detectability of cues in specific spectral regions.

METHOD

Much of our effort this summer has been focused on stimulus preparation and programming for the planned experiment.

The experiment will be conducted at the Auditory Localization Facility of the Armstrong Laboratory at Wright-Patterson Air Force Base. Available at this facility is a large anechoic chamber, which houses a 4.3-m diameter geodesic sphere. Mounted on the surface of the sphere are 277 Bose 4.5-inch speakers. This is a unique facility that allows the experimenter considerable control over the spatial distribution of sound sources when conducting free-field masking or sound localization research. During the experiment, the subject is seated with his/her head in the center of the sphere. Directly in front of the subject, mounted on the surface of the sphere, is a monochrome video monitor, which is used to display the response alternatives. The subject chooses among the words using a hand-held, 6-button response box.

The intelligibility of masked speech presented in the free field is being measured using the Modified Rhyme Technique (House, Williams, Hecker, and

Kryter, 1965). Six different talkers (3 males and 3 females) recorded three tokens of each word of the six 50-word lists suggested by House et al. The words on the list were selected to "contain representatives from the major classes of speech sounds". The recordings were made through a high-quality microphone onto digital audio tape at a sampling rate of 44.1 kHz, while the talker was seated in a quite room. The recordings were transferred to a SPARC workstation, where individual speech tokens were isolated using the ESPS/waves+ software package and adjusted to have equal RMS energy. A clear token of each word will be selected from the three recorded tokens and the final list will be tested to assure 100% intelligibility in the quiet. (Additional recordings will be made, as necessary, to assure 100% intelligibility for each list.) We will also examine detection performance with click-train signals, similar to those examined by Gilkey and Good (1994), for selected signal and masker locations.

The masker is a "speech-spectrum" noise, designed to match the long-term average spectrum of the speech tokens. The duration of the masker was chosen so that the noise would begin 50 ms before, and end 50 ms after, the longest speech token.

We will be examining performance with broadband stimuli (i.e., no additional filtering), and with stimuli constrained to lie within low-, mid-, or high-frequency regions. When the signal is bandlimited to a low-, mid-, or high-frequency band, it will be filtered through a 1.33-octave filter centered at 590 Hz, 2860 Hz, or 8270 Hz, respectively. When the masker is band-limited to a low-, mid-, or high-frequency

band, it will be passed through a 2.0-octave filter centered at 590 Hz, 2860 Hz, or 8270 Hz, respectively.

We will examine intelligibility for signal and masker locations comparable to those that Gilkey and Good (1994) examined in their study of masked detection. Specifically, maskers will be presented from directly in front of the subject {0° azimuth, 0° elevation}, directly above the subject {0° azimuth, 90° elevation}, and directly to the subject's right {90° azimuth, 0° elevation}. Both horizontal and vertical separations between the signal and the masker will be examined.

Throughout each trial, a closed set of six words is shown on the video display. The six possible words differ by only a single consonant, which either occurred in the initial or final position for each of the six words. A speech token is presented from the signal speaker 300 ms after the display is turned on. Simultaneously, the masker is presented from the same or from a different speaker. A 3-s response interval follows the stimulus presentation. The subjects respond by pressing one of six buttons on the response box to indicate the word they believe was presented. During the response interval, the word that the subject selects will be highlighted and the subject may change his/her response; the last response made during the response interval is recorded. Trial-by-trial performance feedback will be not provided.

EXPECTED RESULTS

The results will be analyzed and compared to the results of the experiments of Gilkey and Good (1994) and Gilkey, Good, and Ball (1994) in order to determine, for each frequency region, the agreement between the detectability of click-train

signals and the intelligibility of speech. The responses to individual speech sounds will be analyzed to determine which phonemic distinctions become more discriminable when the speech signal is spatially separated from the masker.

Plomp (1976) and Bronkhorst and Plomp (1988) measured the intelligibility of broadband speech stimuli that were presented from directly in front of the subject. We anticipate that we will observe comparable results under comparable conditions. However, when the signal is to the side, we expect to realize larger gains, particularly when the signal and masker are on opposite sides of the head, because of the substantial head-shadow effect under these conditions. We expect to observe modest increases in intelligibility with separations in elevation. Gilkey, Good, and Ball (1994) showed that the increase in detectability with elevation occurs largely at high frequencies. Therefore, we anticipate that any observed increases in intelligibility will be for speech sounds with significant high-frequency energy (e.g., fricatives and stop constants).

By comparing the results with bandlimited speech to those for broadband speech and to the detection results from this and previous studies, we should be able to determine the frequency specific changes in the audibility of the speech information when the signal and the masker are separated. Because the subjects task is to choose the correct word from a closed-set of six words, we expect the subjects to be able to eliminate some incorrect words (i.e., increase the probability of a correct response), even when the effective or actual frequency range of the speech signal has been severely restricted. For example, when the decisions of the subjects are based on high-frequency information only, we anticipate that they will

be able to distinguish stops and fricatives from other speech sounds and will often be able to distinguish them for each other, but may have difficulty distinguishing among fricatives and among stops. When decisions are based on mid-frequency information only, they should be able to distinguish among stops (e.g. based on place of articulation) and among fricatives. When decisions are based on low-frequency information only, it should be possible to distinguish among stops (based on voicing). However, it should be difficult to distinguish among fricatives, although affricates may be distinguishable from other classes of speech sounds.

This study will have important implications for basic science, in that it explicitly attempts to relate the results from detection and intelligibility studies. It will also have important implications for applied science, by specifying the intelligibility of speech signals that can be expected in auditory displays, as a function of both the effective bandwidth of the communication channel and the spatial separation between the signal and the masker.

APPENDIX: OTHER RESEARCH ACTIVITIES

Considerable effort was expended during the period of RDL support on the preparation of an edited book and on the preparation of two chapters describing our research.

In September 1993, Timothy R. Anderson and Robert H. Gilkey organized the Conference on Binaural and Spatial Hearing at Wright-Patterson Air Force Base. This was a major international conference, with 36 presentations by basic and applied scientist and more than a hundred conference attendees. Conference speakers agreed to submit chapters for a book loosely based on the conference.

The book nears completion and we plan on submitting it to the publisher before the end of the year.

We have also been preparing two chapters for the book. The first, by Good, Gilkey, and Ball, describes the results of our free-field experiments on masked detection and on masked localization. The second chapter, by Janko, Anderson, and Gilkey, describes our work on the modeling of human sound localization.

REFERENCES

Bronkhorst, A.W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. Journal of the Acoustical Society of America, 83, 1508-1516.

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. Journal of the Acoustical Society of America, 25, 975-979.

Colburn, H.S., & Durlach, N.I. (1978). Models of binaural interaction. In E.C. Carterette & M.P. Friedman (Eds.). Handbook of Perception, (pp 467-518). New York: Academic Press.

Doll, T.J., Hanna, T.E., & Russotti, J.S. (1992). Masking in three dimensional auditory displays. Human Factors, 34, 255-265.

Durlach, N.I., & Colburn, H.S. (1978). Binaural phenomena. In E.C. Carterette & M.P. Friedman (Eds.). Handbook of Perception, (pp 365-466). New York: Academic Press.

Gilkey, R.H., & Good, M.D. (1994). Effects of frequency on free-field masking. Human Factors, in press.

Gilkey, R.H., Good, M.D., & Ball, J.M. (1994). A comparison of 'free-field' masking for real and for virtual sounds. Manuscript in revision for the Journal of the Acoustical Society of America.

Good, M.D., & Gilkey, R.H. (1992). Masking between spatially separated sounds. Proceedings of the 36th Annual Meeting of Human Factors Society, 1, 253-257.

House, A.S., Williams, C.E., Hecker, H.L., & Kryter, K.D. (1965). Articulation-testing methods: consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America, 37, 158-165.

Plomp, R. (1976). Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). Acustica, 34, 200-211.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., & Perrott, D.R. (1991). Free field release from masking. Journal of the Acoustical Society of America, 90, 1355-1370.

Zurek, P.M. (1992). Binaural advantages and directional effects in speech intelligibility. In G.A. Studebaker & I. Hochberg (Eds.). Acoustical Factors Affecting Hearing Aid Performance, (pp. 255-276). Boston: Allyn and Bacon.

THE EFFECTS OF SOCIALIZATION ON VOCATIONAL ASPIRATIONS OF MIDDLE SCHOOL CHILDREN

Richard G. Best
Graduate Student
Department of Psychology

Southwest Missouri State University
901 S. National
Springfield, MO 65804

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

THE EFFECTS OF SOCIALIZATION ON VOCATIONAL ASPIRATIONS OF MIDDLE SCHOOL CHILDREN

Richard G. Best
Graduate Student
Department of Psychology
Southwest Missouri State University

Abstract

According to labor forecasts for the period from 1990 to the year 2000, the demography of the work force will change such that women will constitute 60% of new entry level workers, while minority groups will account for approximately one third of the total labor force (Offerman & Gowing, 1990). Research indicates that childhood socialization may serve to perpetuate stereotypes restricting vocational development for members of these groups. Consequently, occupational preparation for tomorrow's replacement work force is typically at a disadvantage. In response to this concern, the Armstrong Laboratory has developed plans for an automated, career counseling and exploration system (**ACCESS**) designed to enhance vocational knowledge of middle school children. An assessment component matching the student's interests and abilities with consonant job prerequisites will be integrated into the system. The purpose of the current research was to provide a theoretical foundation investigating anticipatory vocational socialization, defined by Jablin (1987) as vocational development prior to organizational entry. While Jablin considered vocational socialization as a two phase process, consisting of 1) vocational choice/socialization; and 2) organizational choice/entry, the current framework concentrates primarily on vocational choice socialization. The resulting product from this research was a survey designed to identify vocational knowledge of middle school children influenced by socialized expectancies and limitations imposed by social policy (see Figure 1).

THE EFFECTS OF SOCIALIZATION ON VOCATIONAL ASPIRATIONS OF MIDDLE SCHOOL CHILDREN

Richard G. Best

Introduction

Future labor trends predict that two of histories' most underrepresented segments of society will combine to constitute nearly three-fourths of tomorrow's replacement work force. Although the labor force is expected to grow at a slower rate than at any time since the 1930's, it will also reflect a massive influx of women and minorities between the years 1990 and 2000. Specifically, 60% of all new entry level positions will be occupied by women while one third of the total labor force will be comprised of minorities (Offerman & Gowing, 1990). Slower increases in work force size can result in a plummeting economic growth rate, unless organizations enhance worker productivity. Furthermore, tremendous expansion of middle-aged workers, remaining from the baby-boom era, will increase competition for scarce, high level organizational positions. New positions will therefore require enhanced skill levels to maximize productivity and higher educational attainment will be necessary to compete for high level positions (Offerman & Gowing, 1990). In view of these projections, the Armstrong Laboratory, at Brooks Air Force Base, is proposing a next-generation automated career counseling and exploration system to facilitate the transition of women and minorities into the work force. The purpose of the present study is to provide a theoretical foundation upon which to base such a system, focusing on socialization influences on occupational preparation for women and minorities.

Discussion of the Process

Socialization is an intricate, lifelong process in which one acquires beliefs, loyalties, moral and ethical values, and perspectives of the human world and its institutions (Borow, 1984).

Anticipatory vocational socialization, defined as conditioning and preparation to occupy organizational positions, is a developmental process spanning individual maturation from childhood to young adulthood, and beyond (Jablin, 1987). According to Van Maanen (1975, p. 82), as occupational information is acquired from various sources, it is compared with the individual's self concept, "weighing the factors and alternatives involved in choosing an occupation and finally making a series of conscious choices which determine the direction of his/her career (cited in Jablin, 1987). Sources of vocational information are derived primarily from family members, educational institutions, peer groups, and the media (Jablin, 1987). In providing a conceptual framework for the career counseling system, this report will examine the effects of anticipatory vocational socialization, with its various subcultures, on self concept in determining achievement motivation as a precursor to occupational choice.

SELF CONCEPT

Self concept, ostensibly determined early in the socialization process, is defined by personal attributes, such as competence, control orientation, and personal values and interests, and by self esteem appraisals regarding these attributes (Rathus & Nevid, 1992). Socialization interacts with gender differences, race and ethnicity, and class status, to shape our self concept, or our impressions of ourselves. Studies investigating self esteem patterns among fifth and sixth graders indicate a positive correlation with parenting styles. For example, high self esteem is associated with strict, highly involved, (but not harsh or cruel), parenting strategies (Coopersmith, 1967). The current conceptualization posits that self concept will ultimately moderate achievement motivation in view of self efficacy expectations and control orientation.

Self Efficacy Expectancy Influences on Self Esteem

Presumably, self efficacy expectancies, which influence our

motivation to engage in challenges as well as our perseverance to accomplish them, precede the self appraisal process. Expectancy Theory argues that self efficacy expectancies determine the degree of effort that we will allocate towards accomplishing challenging tasks according to two precepts: 1) self confidence in our ability to accomplish the task; and 2) an expectation that the outcome will be of value to us. Furthermore, self esteem may be enhanced by engaging in tasks that promise intrinsically rewarding results, such as those that are consonant with our values and interests (Rathus & Nevid, 1992). Conversely, confidence to pursue challenging and rewarding tasks may be handicapped by the socialization process, such as through gender typing, and by societal policies reflecting racial/ethnic discrimination, and social class distinctions, thereby debilitating opportunities for enhancing self esteem.

Control Orientation Effects on Achievement Motivation

A construct related to self efficacy is control orientation, or locus of control, which refers to a bipolar source of behavioral attribution. Specifically, recognition of behavioral outcomes may be attributed to internal sources contingent upon an individual's behavior, or consequences may be credited to external sources, such as luck, chance, or fate. Rotter (1966) defined internal-external locus of control as "...the degree to which the individual perceives that a reward follows from, or is contingent upon, his own behavior or attributes versus the degree to which he feels the reward is controlled by forces outside of himself and may occur independently of his own actions". Outcomes not perceived as dependent upon an individual's behavior, or resulting from luck, chance, or fate are said to be construed from an external orientation, whereas consequences believed to be conditionally related to individual behavior are said to be perceived from an internal orientation. Locus of control is believed to influence achievement motivation in the following manner: Internal achievement motivation, identified by Borow (1984) as the belief

that one's goals are attainable in view of higher degrees of environmental coping and progressive mastery; conversely, external achievement motivation derives from a background characterized by bewildering inconsistencies in parental reaction that impose irregular, unpredictable social consequences, which presumably encourage learned helplessness. Governed by this sense of futility, external individuals lack confidence to manipulate the environment (Borow 1984). Supporting the assertion that locus of control is implicated in achievement motivation, Trice & Gilbert (1990) report results indicating that 60% of fourth graders identified as external were characterized by a lack of career ambition, or aspiring to a fantasy career. On the other hand, 93% of internally classified students revealed realistic vocational goals. Anticipatory vocational socialization, incorporating the various, distinct subunits, is presumed to indirectly influence achievement motivation through self concept.

Family Subunit

Immediate family members constitute the reference group that exerts the most pervasive and durable influence on social development in children. Family settings inculcate children with gender-typing, elemental rules of conduct, influences of social interaction, and other task-oriented organizing activities (Goldstein & Oldham, 1979). For example, fathers frequently discuss work activities, specifically issues concerning social interaction, in the home, exposing children at early ages to socialized expectancies. Families also provide the vehicle for the transmission of attitudes and values to other settings, which ultimately influence career planning. Leifer & Lesser (1976 p. 38), proposed that "parents were the primary determiners of occupational choices for adolescents and young adults" (cited in Jablin, 1987). The system of rewards and denials, imposed either consciously or unconsciously throughout the child's socialization, marks the beginning of a sense of competence and limitations. Children who witness productive, achieving behavior in the home, and whose own

developing habits of success have been reinforced, will likely transfer that behavior to social settings such as schools and occupations (Super, 1957).

Gender Differences

Females, although purportedly aspiring to marry and have children at earlier ages than males, also develop salient vocational concerns sooner. For example, there is a significant increase in females' aspirations to work outside the home, attempting to balance both work and family roles. Regarding issues of vocational preference, females report more willingness to accommodate their careers around family obligations. Furthermore, women generally deem income and status as secondary incentives, aspiring to occupations allowing the fulfillment of personal values and interests, whereas males tend to seek careers characterized by high status (Flanagan, 1993).

Social Class Distinctions

By early adolescence, socio-economic status (SES) becomes a prominent indicator of occupational preference. Parents' occupational and educational experiences can also influence child-rearing strategies as part of the socialization process. With regard to social status, family practices mediate the reproduction of social stratification across generations. Featherman (1980) reviewed sociological work on the status attainment model proposing that SES indicators, such as parent's educational level, father's occupational status, and source and level of family income are transmitted intergenerationally. Personal experiences of parents from higher socio-economic backgrounds generally reflect greater autonomy, more intellectual complexity, and increased self-direction, therefore these same attributes are also valued in their children. Conversely, parental experience with routinization, and lack of autonomy tends to encourage more restrictive values emphasizing conformity (Flanagan, 1993). Furthermore, a positive relationship exists between the parents' occupational and educational attainments and the achievements of their children

(Flanagan, 1993). Occupational inheritance, the social induction of offspring into the parent's occupation, though less prevalent today, may also exert influence on the child's occupational choice. According to observational learning, parental role models may reinforce the child's desire to follow in their footsteps.

Social Class Interactions with Gender and Ethnicity

The increased participation of women and minorities in the future work force warrants concerted attention as to the effects of socialization on their vocational preparation. Gender and social class may interact to constrain adolescent career aspirations, influencing the pursuit of only those occupations perceived as attainable. According to Kirchner & Vondracek (1973), gender stereotypes regarding occupational roles may develop as early as age three (cited in Jablin, 1987). Support for their assertion was reported in research conducted by Gettys & Cann (1981) in which 78% of two and three year old children identified male dolls with construction work, while only 23% of the children paired the male doll with the teaching profession (cited in Rathus & Nevid, 1992). Furthermore, studies during the Great Depression suggested that response to a family's financial crisis was gender-typed such that adolescent males were encouraged to pursue "odd jobs", while female adolescents were expected to assist with the domestic duties. More recent research analyzing stress on the relationship between adolescents and their parents (resulting from adolescents' quest for greater participation in family decisions) revealed compelling evidence for gender differences in response to financial hardship. Controlling for parents' educational levels, financial strains were significantly related to expectations for the daughter to seek vocational training or full-time employment following high school (Flanagan, 1993). Vocational occupations are generally considered both lower in prestige and predominantly more sex-typed than other professions. Yet females from working-class families tend to focus myopically on vocational occupations, regardless of ability. Conversely, females from higher SES backgrounds are entering previously male-dominated, high status occupations, validating the

assertion that working class families generally adhere to more traditional, sex-typed attitudes.

Working-class families purportedly espouse traditional gender stereotypes because their aspirations are circumscribed by limited options and greater pressure to begin work in order to earn an income (Flanagan, 1993).

The fact that tomorrow's work force is today's minority children suggests possible socio-political realignment. Because more than half of all minorities are being raised in poverty, ill-served by education, legislators, corporate executives, and educators must focus on poverty as an issue that affects national productivity, not merely social concerns. According to Offerman & Gowing (1990), a growing concern for current literacy levels among today's youth suggests that recruiting qualified applicants for increasingly higher skilled positions will become more and more difficult.

Educational Subunit

Responsibility for preparing children for their roles in society is being increasingly delegated to schools. According to Jablin (1987), educational institutions are considered the most significant source of vocational information. Children entering the school system are challenged with meeting societal demands, dealing with impersonal standards, facing competition from their peers, and being routinely judged by others. School settings allow children to establish competencies and learn that their work is differentially valued (Borow, 1984). Typically, however, vocational information conveyed in the school system focuses on superficial qualities of different occupations rather than on specific role characteristics. While vocational socialization in schools does not specifically address occupational content, it does identify communication styles that conform to implicit interaction norms.

Gender-Typed Restrictions Educational Opportunities

Gender stereotyping has historically worked to the disadvantage of

females with regard to education. Throughout most of history, sexist attitudes stereotyping women as emotional, irrational and naturally disposed to child-rearing and homemaking also regarded females as unsuited for education. Twentieth century opinion professes equality in scholastic aptitude for both girls and boys, although differential expectations persist (Rathus & Nevid, 1992). According to Meece et al. (1982), by junior high, boys consider themselves more competent in math than girls do, in spite of equivalent grades. Math connotes greater utility for junior high boys thereby fostering a more positive perception than for their female counterparts who report higher incidence of math anxiety (Meece et al., 1982; Tobias & Weissbrod, 1980). In the absence of an established genetic link for these differential propensities, accountability rests, for the most part, on socialization. Sherman (1983) maintains that females are deterred from taking math courses, considered part of the "male domain", because they imply masculine traits such as ambition, independence, self-confidence and spatial ability (cited in Rathus & Nevid, 1992). Although generally favorable traits, they are stereotyped as distinctly masculine attributes. Teachers may unwittingly sustain masculine stereotypes for high achievement in math and science by dissuading females from these pursuits because they are inconsistent with the feminine role. According to Meece et al. (1982), teachers maintain higher expectations for males in math courses and therefore allocate more time for instructing and interacting with them. Teacher expectancy effects can prove debilitating for female students whose self-confidence may be particularly vulnerable due to a lack of prior experience with math and science courses. Differential perceptions of self-competence in theses areas reflect pervasive gender stereotyping (Kahle et al., 1993).

Social Class Restrictions on Educational Opportunities

According to Gunn (1964), by the time students reach junior high school, they have already begun to assimilate occupational ranking by social status. Forecasts offered by the Bureau of Labor

Statistic's Occupational Outlook Handbook (1992) indicate that, due to labor supply-and-demand ratios, high status occupations will require increased levels of academic achievement, inferring a greater necessity for educational equity across all groups.

According to Flanagan (1993), educational values may partially explain differential motivation for academic achievement by social class. The American work ethic and education are advocated as mechanisms for enhancing social mobility. Considering the expected utility of education in terms of occupational rewards, there appears a strong positive relationship between social class and academic achievement. Class boundaries, constructed by social homogenization of communities, restrict mobility within the social stratification impairing achievement motivation as a source of enhancing self identity. Discrepant property values promote disparate allocation of resources because local property taxes constitute the primary funding mechanism for schools. In a summary submitted by the National Assessment of Vocational Education (Wirt, Murasicin, Goodwin, & Meyer, 1989), it was revealed that school districts comprised of predominantly disadvantaged students reported 40% fewer vocational courses than districts inhabiting the wealthiest populations (cited in Flanagan, 1993). Consequently, vocational education is not as readily available to those who would benefit most from it.

Tracking policies, intended to assign students to instructional levels commensurate with their abilities, often legitimate inequality by promoting aptitude differences reflecting social origin. Upper class families often encourage prominent class placement and course objectives for their children, both predictive of academic achievement. Students assigned to lower class rankings are at risk for internalizing inaccurate beliefs about individual deficits, such as impaired abilities, and lack of motivation, when compared with their classmates (Flanagan, 1993). Injurious implications of low-ability group placement include enhanced discrepancies in academic achievement between low and high ability groups, precipitated by biased teacher expectancies. Students

presumably of lower ability generally receive less attention and detailed feedback from their teachers, and have less stringent demands required of them (Eccles & Wigfield, 1985).
In a 1988 survey of eighth graders, the National Educational Longitudinal Study (NELS) identified SES as the single best predictor of grades and test scores, contributing to an inverse relationship between social class and high school completion. Furthermore, very few low SES eighth graders achieve advanced levels in reading or math (Flanagan, 1993).

Gender & Social Class Restrictions on Educational Opportunities

Financial concerns often prove even more limiting for female adolescents than for male adolescents in reference to secondary education. Flanagan (1993) asserts that, when financial pressures are an issue, parents are more likely to endorse a college education for their son rather than their daughter, rationalizing that it will warrant a better return on investment. Socialization often compels the daughter to initiate the deferment of her college education, exacerbating the notion that investing in a son's education is financially more rewarding. Social class may, therefore, exert stronger influence on educational and career aspirations of female adolescents, whereas academic ability and achievement are better predictors of males' aspirations (Flanagan, 1993).

Ethnicity & Social Class Restrictions on Educational Opportunities

Nelson-Le Gall (1991) argued that during the past two to three decades, there have been gains in academic achievement among African-American students as evidenced by an increase in average educational attainment level from 8th grade completion to high school graduation (cited in Pollard, 1993). Gaps in scholastic performance still exist, however, and manifest themselves as early as second grade. According to Pollard (1993), African-American students are still underrepresented in college attendance rates and many continue to be excluded from the school setting. Restricted

academic performance of African-American students has been attributed to psychological factors, such as impaired self concept and lack of motivation. African-American students may be less inclined to participate in a curriculum that virtually ignores their cultural heritage. Furthermore, decreased teacher expectations due to tracking or exclusion policies reassert the declining motivation principle (Pollard, 1993). Racial stratification in American society cultivates low social status for many African-Americans by prohibiting them from equal participation in social and economic institutions. Inequities in school resources deny African-American students, especially those of impoverished backgrounds, access to adequate educational resources, resulting in lower academic achievement. Inadequate educational opportunities for African-American students promotes negative self-perceptions, decreased motivation, and lowered levels of academic achievement (Pollard, 1993).

Peer Subunit

Peer groups are thought to define social interaction by prescribing socially acceptable rules of conduct while also delineating the consequences of inappropriate behavior. The adolescent is entering a stage of self-assertion and individuality by distancing him/herself from parental authority. Demos & Demos (1973) argue that current views of adolescent psychological development emerged during the late 19th century. Thus far, however, social scientists and counselors have not systematically investigated the influence of peer groups on adolescent thinking about occupational and life values (Borow, 1984). Research that is available, however, suggests a positive relationship among career aspirations of adolescent peer groups (Jablin, 1985). Speculatively, peers may serve to confirm or disconfirm the desirability of different occupations. Tangri (1972) argued that the more that adolescents discuss stereotypic, gender-typed occupations, the more likely they are to conform to socialized expectancies. The specific nature of the effects of peer interaction and socialization on adolescent

perception of communication in occupational contexts requires further investigation.

Media Subunit

Media portrayals of stereotypic masculine and feminine roles sustain gender-typed depictions that may persist into adulthood (Christenson & Roberts, 1983). Considering the inordinate amount of time children spend in front of the television, this becomes a much more salient concern. Studies researching this phenomenon generally agree that the amount of time devoted to watching television by elementary and secondary school students in the United States significantly outweighs the time allotted to homework assignments. American fifth grade school children devote 256 minutes weekly to homework, compared to 368 minutes by their Japanese counterparts (Garfinkel, 1983). Television's stereotypic caricatures of various occupations may debilitate authentic occupational awareness (Borow, 1984).

Proposed Statistical Analyses

In view of the of the diversity of school districts and the availability of resources at the Armstrong Laboratory, pilot test sites for the Armstrong Laboratory Future Work Oriented Survey are being targeted to the San Antonio area during the 1994 Fall school term. Proposed statistical testing of the instrument will include: (1) Factor analytic techniques to determine if items group according to a priori constructs; (2) Reliability estimates to determine internal consistency. Content and construct validity statistics will also be performed to verify item accuracy. Results from the analyses will be used to refine constructs comprising the framework investigating vocational socialization. Finally, once empirically determined constructs for the conceptual framework have been identified, they will be subjected to model testing using structural equation analysis techniques.

Conclusion

In response to concerns regarding impoverished occupational preparation imparted to women and minorities during the vocational socialization process, the Armstrong Laboratory is developing an automated career counseling and exploration system for students (**ACCESS**). This program is intended to encourage occupational preparation by addressing middle school students' irrationally-based perceptions regarding vocational opportunities due to gender-typing and racial discrimination. Additionally, by enhancing occupational preparedness, self efficacy expectancies may be enhanced, thereby fostering greater confidence to embark on a wider array of career choices. This project will be based on a conceptual framework focusing on the effects of vocational socialization on self concept. The process of vocational socialization derives from various sources which, through their interactions with gender differences, racial or ethnic origins, and social class, impact our self concept. Self impressions are influenced by appraisals of personal traits, such as self competence and internal versus external locus of control orientation. Furthermore, self evaluations can be positively influenced by engaging in tasks that are consonant with our values and interests. Favorable self esteem appraisals and confidence in one's ability to succeed in intrinsically rewarding tasks will result in internal achievement motivation. The current framework is based on the premise that internally oriented achievement motivation will enhance confidence to embark on vocational opportunities in spite of irrationally imposed perceptions and discriminatory social policy. Perhaps as women and minorities continue to claim their share of the labor market, the myths disseminated during socialization will disappear.

Figure 1

Below is a sample of the survey constructed to identify vocational awareness of middle school children and perceptual constraints they may infer from the socialization process. Included are sample items from constructs identifying: (1) the socialization process (incorporating the various subcultures); (2) self concept (including self efficacy, locus of control, and values and interests); (3) perceptions of equality; and (4) achievement motivation.

THE ARMSTRONG LABORATORY FUTURE WORK ORIENTED SURVEY

Please use the following scale to indicate your agreement or disagreement with each of the statements listed below. If you do not know the answer, or it doesn't apply to you, mark **G**:

A	B	C	D	E	F	G
Strongly Agree	Agree	Slightly Agree	Disagree	Slightly Disagree	Strongly Disagree	Don't Know

SOCIALIZATION INFLUENCES

Family

1. When I grow up, I want to be just like my father.
2. When I grow up, I want to be just like my mother.
3. I will need to get a job rather than go to college so I can help support my family.
4. It is fair for the father to stay home and take care of the household while the mother gets a job to support the family.

Educational

5. Boys are naturally better in math than girls.
6. If there is only enough money to send one child to college, then parents should send a son rather than a daughter.
7. I have experienced prejudice from other students in school.
8. Women make better teachers than men do.
9. It is difficult for minority students (for example African Americans, Hispanics, Asians, and Native Americans) to get a quality education in America.
10. Taking courses in math and science will help me prepare for my future job.
11. I'll have to go to college to get a good job.
12. I'm taking courses now that will help me prepare for my future job.
13. A high school education is all that I need to be successful.

Figure 1 continued

Peer

14. I'd rather talk to my friends about my future plans than to my parents or teachers.
15. My teachers encourage me to go to college.

Media

16. Television should include more women in parts usually played by men.
17. It is important to do homework first, before doing other enjoyable things.
18. You can't believe everything you see on television.
19. Jobs I see on television are accurate descriptions of real life jobs.

VALUES AND INTERESTS

20. I would like to be a math teacher some day.
21. I want a job where I can work independently.
22. I want a career that will allow me to express my values and explore my interests.
23. I enjoy working with computers.
24. Being satisfied with my future job will be more important to me than making a lot of money.
25. Helping others is more important to me than making money.

SELF-EFFICACY

26. I am confident about my strengths and can overcome my weaknesses.
27. I know that I will find a good job when I grow up.
28. I could do a good job raising a family **and** have a successful career.
29. It will be difficult for me to achieve my future goals.
30. Learning new skills is difficult for me.
31. I can handle unexpected problems easily.
32. I succeed at most things I try.
33. I experience many failures in life.

* Note: Items 29-33 are modified items taken from The Expectancy for Success Scale developed by Hale & Fibel (1978)

LOCUS OF CONTROL

34. Success comes from working hard, luck has little to do with it.
35. Getting a good job depends on being in the right place at the right time.
36. Planning for the future makes things turn out better.

37. It is useless for me to try my best in school because most of the other students are smarter than me.
38. I have control over the things that happen to me.
39. Whether or not I do my homework has much to do with what kinds of grades I get.
40. I can change what might happen tomorrow by what I do today.

* Note: Items 34-40 are modified versions of items from Rotter's Internal-External Locus of Control Scale (Rotter 1966) and The Nowicki-Strickland Locus of Control Scale for Children (1973).

PERCEIVED EQUALITY

41. Women are just as qualified as men to be President of the United States.
42. Even though we are all different, we should be treated equally.
43. Members of minority groups, (for example African Americans, Hispanics, Asians, and Native Americans), have to work harder to succeed than caucasians do.
44. Everyone has an equal chance of succeeding today.
45. There aren't that many opportunities for someone like me.
46. Caucasians usually get the best opportunities.

ACHIEVEMENT MOTIVATION

47. I try to get the highest grade in all of my classes.
48. Graduating from college is a goal of mine.
49. When I grow up, I just want a job that pays the bills.
50. Planning for my future **now** will help me achieve my goals.
51. I would rather raise a family than have a career.
52. I would willingly quit a job to raise a family.
53. It is too soon for me to worry about planning for my career.
54. I want to be president of a big company when I grow up.
55. It is important to me to be the best at everything I do.

References

Borow, H. "Occupational Socialization: Acquiring a Sense of Work." In N.C. Gysbers and Associates (Eds.), Designing Careers. San Francisco, Washington, London: Jossey-Bass, 1984.

Bureau of Labor Statistics, U.S. Department of Labor. "Tomorrow's Jobs." Reported in Occupational Outlook Handbook. May, 1992, pp. 8-10.

Career Counseling: Walsh, W.B., & Osipow, S.H. (Ed). Laurence Erlbaum; Hillsdale, New York 1990.

Chapman, W. & Katz, M.R. (1983). Career information systems in secondary schools: A survey and assessment. The Vocational Guidance Quarterly, 165-177.

Christenson, P.G., & Roberts, D.F. "The role of television in the formation of children's social attitudes." In M.J.A. Howe (Ed.), Learning from television: Psychological and educational research (pp. 79-99). New York, N.Y. Academic Press, 1983.

Coopersmith, S. The antecedents of self-esteem. San Francisco: W.H. Freeman, 1967.

Demos, J., & Demos, V. Adolescence in Historical Perspective." In M. Gordon (Ed.), The American Family in Social-Historical Perspective. New York:St. Martin's Press, 1973.

Eccles, J.S., & Wigfield, A. "Teacher expectations and student motivation." In J. Dusek (Ed.), Teacher Expectancies. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc., 1985.

Featherman, D.L. "Schooling and Occupational Careers: Constancy and Change in Worldly Success." In O.G. Brim and J. Kagan (Eds.), Constancy and Change in Human Development. Cambridge Mass.: Harvard University Press, 1980.

Flanagan, C. (1993). Gender and social class: Intersecting issues in women's achievement. Educational Psychologist, 28(4), 357-378.

Garfinkel, P. "The Best 'Jewish Mother' in the World." Psychology Today, September 1983, pp. 56-60.

Goldstein, B., & Oldham, J. Children at work: A study of socialization. New Brunswick, N.J.: Transaction, 1979.

Gunn, B. (1964). Children's conceptions of occupational prestige. Personnel and Guidance Journal, 42, 558-563.

Jablin, F.M., Putman, L.L., Roberts, K.H., & Porter, L.W. (1987). An interdisciplinary perspective: Organizational entry, assimilation, and exit. Handbook of Organizational Communication. Sage: Newbury Park CA.

Kahle, J.B., Parker, L.H., Rennie, L.J., & Riley, D. (1993). Gender differences in science education: Building a model. Educational Psychologist, 28(4), 379-404.

Meece, J.L., Parsons, J.E., Kaczala, C.M., Goff, S.B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. Psychological Bulletin, 91, 324-348.

Offerman, L.R., & Gowing, M.K. (1990). Organizations of the future: Changes and challenges. American Psychologist, Vol. 45, No. 2, 95-108.

Orr, V. (1983). Scientific illiteracy in the high-tech age. Air Force Magazine, 73-76.

Osipow, S.H. (1987). Counseling psychology: Theory, research and practice in career counseling. Annual Review of Psychology, 38, 257-278.

Pollard, D.S. (1993). Gender, achievement, and African-American students' perceptions of their school experience. Educational Psychologist, 28(4), 341-356.

Rathus, S.A., & Nevid, J.S. Adjustment and Growth: The Challenges of Life. New York, N.Y.: Harcourt Brace College Publishers, 1992.

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80. (1 Whole No. 609).

Super, D.E. The Psychology of Careers. New York: Harper & Row, 1957.

Tangri, S. (1972). Determinants of occupational role innovation among college women. Journal of Social Issues, 28, 177-199.

Tobias, S., & Weissbrod, C. (1980). Anxiety and mathematics: An update. Harvard Educational Review, 50, 63-70.

Trice, A.D., & Gilbert, T.M., (1990). Locus of control and career aspirations of fourth grade students. Psychological reports, 67 (3, Pt2), 1345-1346.

Walsh, W.B., & Osipow, S.H. (1990). Career Counseling: Contemporary Topics in Vocational Psychology. In ed. Hillsdale, New Jersey: Hove and London.

TOWARD MODELING HIGHER LEVEL CONTROL SYSTEMS: INCLUDING MEMORY'S PLACE IN LEARNING

Daniel A. Brown
Master's Student
Department Of Education

Grand Canyon University
3300 West Camelback Road
Phoenix, AZ 85017

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

TOWARD MODELING HIGHER LEVEL CONTROL SYSTEMS: INCLUDING MEMORY'S PLACE IN LEARNING

Daniel A. Brown
Master's Student
Department Of Education
Grand Canyon University

Abstract

For the past few summers Dr. Thomas Hancock and Dr. Richard Thurman have been investigating Perceptual Control Theory (PCT). It has been the desire of Dr. Hancock to use the variables from a drill program Dr. Thurman designed to produce a more precise cognitive model of a learner. This study looks into possibilities for such a model.

Through self-reports during the drill program, investigation into current non-PCT models of memory processing and the creation of flow chart models a PCT model was initiated. In the self reports phase it was found that the drill subject had a high level of awareness of strategies used to learn the material. As well, a link between certitude rating and the speed and clarity of items retrieved from memory was observed.

After the self-reports had been completed three flow chart models were attempted. These flowcharts use a non-traditional PCT approach of looking at memory. Memory is viewed as a network which is being operated on by a hierarchy of control systems which brings perceptions from memory into a match with perceptions of the subject's environment. The models of Stephen Grossberg and John Anderson had a key role in the design of these flowchart models.

As well as a description of the trials made in this study, some suggestions toward future studies are made.

TOWARD MODELING HIGHER LEVEL CONTROL SYSTEMS: INCLUDING MEMORY'S PLACE IN LEARNING

Daniel Brown

Introduction

During the past few summers Dr. Thomas Hancock has been working as a summer faculty associate at Armstrong Laboratories. Dr. Hancock in association with Dr. Richard Thurman of Armstrong Labs, has been investigating Perceptual Control Theory (PCT) (Powers, 1973). It has been their feeling that PCT describes the way that humans function. Dr. Hancock has been particularly interested in doing mathematical modeling at the higher levels of the control theory hierarchy.

Dr. Thurman designed a CBT program related to identifying radar signals which could be used for training (Hancock, Thurman & Hubbard, 1993a). The program was designed with the thought that the data recorded for the program's variables could lead to a more precise cognitive model of a learner and eventually to an understanding of what subjects are controlling for as they do the drill. In fact when the study was done it was found that three primary groups of subjects were identified. It appeared that one group was controlling for learning, the next group controlled to get just the correct answer and the third group controlled to simply finish the drill. The primary factor that correlated to these groups was the amount of time that each student spent using feedback time.

As well as finding these three groups of subjects, the study also found the other variables to be highly predictive for future correct answers. The variables that this drill recorded were response time, feedback time, certitude rating and certitude time. By looking at the values recorded the future correctness of subjects could be accurately predicted.

Even with all of these strong predictors found from this drill it still does not give a control theory model of what occurs during the drill. The ultimate goal of these studies has been to create such a PCT model. This paper summarizes the current work done along these lines during this summer at Armstrong laboratories. The goal of this study has been to use the information gained from the studies done by Dr. Hancock and Dr. Thurman and

use them to create, ultimately, a working mathematical model of higher level control systems based in PCT.

Methodology

This study occurred in three major phases. First a series of self-reports were done while practicing the drill from Dr. Hancock's study. After the self-reporting was done attempts were made to create flowcharts of memory processing based on Power's Control Theory model. After several flowcharts were attempted, the need for a better understanding of information processing models became apparent. This led to phase three where the information processing models of Anderson (1973) and Grossberg (1986) were investigated with the desire to apply them to the flow chart models from stage two

Subject

The subject for this study was the researcher, a graduate student from Grand Canyon University. Because the focus of this study is to understand subjects one at a time, only an individual subject was needed.

Phase I: Radar drill runs with self-report

Materials

A Macintosh Quadra 610 computer was the primary tool of the study. The radar drill used in Stage I is a Hypercard based program designed by Dr. Thurman at Armstrong laboratories. All self-reporting was done by hand by the subject.

The radar drill consists of a series of windows presented to the subject on the computer screen. The first window displays a wave form in one of three positions on the screen. The wave form can be a spiked, rounded or squared while the position can be at the top, middle or bottom of the display. A low, medium or high tone is played with each display. This allows for twenty-seven different combinations of position, wave form and tone. Each combination has a distinct name and these names are listed at the right side of the display. The objective of the drill was to match each display combination with its corresponding name.

The program ran through a series of windows in the following cycle (see Hancock, Thurman, & Hubbard, 1993b):

Frame 1. The first window appears and the subject chooses the name which corresponds to the display combination.
--Response time and correctness are recorded here.
Frame 2. If the answer is correct a second window appears showing the subject that the answer was correct. If the answer was wrong a different

second window appears showing both the answer chosen as well as the correct answer. In both cases the subject is able to get learner feedback for each name by clicking on the name to the right of the window.
--At this point feedback study times are recorded.
Frame 3. Once the subject has completed their study of the feedback they click on the "next" button and a final window appears. The subject makes a certitude rating based on how certain they are that they will get the same display name correct the next time that display appears.
--The time between the clicking of the "next" button and the certitude choice is recorded as the certitude rating time. The certitude rating is recorded here as well.
Return to frame 1. The original window appears again with a different display combination. This cycle continues until the end of the run is reached.

Procedure

For the self-report stage of the study the subject completed four sessions of the radar drill. Each of these sessions consisted of two runs through the drill for a total of eight runs. A run consists of beginning the drill and continuing until the program states End of Day. Sessions 1 through 3 were done prior to phase II & III of the study and session 4 was done on one day after the first attempts at phase III had been attempted. For each session only one run was self reported. The other runs were left unreported so that, if necessary, the data from unreported runs could be compared to reported runs.

The self reports were made by the subject by stopping the radar drill program and writing down the given self report on a separate sheet of paper. The transition between stopping the radar drill and writing the self-report took from 0.9 seconds to 1.45 seconds. In each self report the subject recorded the one item in working memory. The radar drill could be interrupted at any frame to make a self report without damaging the accuracy of the data taken by the radar drill. These procedures were followed to preserve the validity of the self-report (Ericsson & Simon, 1980) while maintaining the integrity of the radar drill data.

The goal of the self-reports was to gain information and insights which would help to create a control theory based flow chart model of the cognitive processes of the subject during the drill. To accomplish this the subject monitored and reported the following items:

1. Certitude. Is this a controlled variable and how does it relate to cognition in this radar drill?

2. What strategies are being used to match a name with its corresponding display?

3. Does response time for initial response relate to cognitive strategies or certitude?

4. Can the subject recognize what aspects of his cognitive processes are controlled variables?

In session 4 the main focus of the self reports was to observe how the subject stored and retrieved information from memory.

Each self-report was written, and for some, an explanation was made to help clarify the self report. The explanation, shown in italics in the results section, was written by the subject immediately following the self-report. These were simply made for clarification and should not be considered as part of the valid self-report.

Phase II: Investigation into models of information processing

As the flow charts were being designed and discussed the consistent problem was in deciding what role memory played in the cognitive processing in the radar drill. As well, the basis of the self report was on the strategies of learning the information. These strategies all related to how to store and retrieve the information more quickly and accurately. Therefore, it was concluded that, if a model was going to be made, then a better understanding of current information processing models was needed. So the second phase of the project was to attempt a model of memory processing within the control system framework. To accomplish this an investigation into the current established models of Stephen Grossberg (1986) and Anderson (1973) were investigated.

The objective was to look at their work which already had mathematical models of memory and to apply these to the flowcharts in phase III. The majority of this time was spent on studying Grossberg's work as it had a more thorough mathematical and neurological basis. After the models had been reviewed, attempts were made to apply the theory from Grossberg's model to the flow charts which had been designed.

Phase III: Control system flow charts

The next stage of the study was to create several flow charts for potential control system models of what was actually happening cognitively during the drill. The information from the self-reporting and from the

investigation into information processing models was used in the construction and analysis of the different flow charts. The variables that were used as inputs and outputs for the flow chart models were taken from Thurman's computer drill (mainly response time and certitude).

The flow charts focus primarily on the processing of information into and from memory. This was done because the subject's self reports were focused most often on his strategies of memory storage and retrieval. It was felt that memory storage and retrieval was one common aspect which occurred at each frame of the radar drill. As well, the investigation into Grossberg's work had given some helpful insights into how such a model could be made.

The basic design of each of the flow charts was to have a hierarchy of control systems (Marken, 1989). Each control system would have an input, a reference standard, a comparator and an output. Multiple levels of control were chosen because of the complex system that was being modeled. Although there were a few variations to the Control Theory format, we tried to stay with a PCT structure.

Results

As described in the methodology, the process of this study had three stages: first, drilling with self-report; second, to investigate current mathematical models of the encoding and retrieval of information to memory; and third, making flow charts of potential control systems to describe the experiment cognitively.

Phase I: Radar drill runs with self-report

Listed below are the self-reports from the radar drill runs. The normal print shows the self-report that was written under the guidelines of Ericsson & Simon (1980). All sentences written in italics are clarifying remarks which were written immediately following the corresponding self-report. These were written at the time of the drill yet are for clarification only.

Session 1: Run 1 self-reports:

- I am trying to memorize the three aspects that go with each name.
- Memorizing the aspects has worked (fairly well) for four of the signals but I can't memorize any more.
- I am trying to come up with a mnemonic device for each trio of aspects.*(for middle position, spiked shape and medium tone I would say, "Spike is in the middle of the yard barking.")*
- I am having trouble creating a mnemonic for Flap Track. I seem to be confusing between mnemonics as I try to create it.
- The mnemonics that are clear pictures to me are coming easily while some of the others are jumbled. This seems to be related to how well the mnemonic matches the features of the display.

· I need to come up with a new strategy for identifying because I can't create enough distinct mnemonics for each display.
· End Session.

Session 2: Run 1 self-reports:
· I am beginning to pair up displays I don't know with displays that have the same position and shape that I do know.
· There are still some problems remembering displays I have only seen a few times but this strategy seems to be working better.
· The displays I have paired with displays I know come to mind quicker and I feel more certain about them.
· From my "feel", response time seems to be faster on items related to mnemonics I know well. *(This was not later confirmed by the data from the drill program as the self report comments were not correlated to individual frames in the drill.)*
· End Session.

Session 3: Run 1 self reports:
· I seem to be grouping displays in groups of three.
· The groups have the same shape and position and have different sounds to distinguish each one.
· There is name for one member of each group that comes to mind first, then I remember the other display name. *(Dog house comes to mind first when I see Pat Hand, then I distinguish that it is the wrong sound and realize the display is Pat Hand.)*
· If I don't have a strong mnemonic for an individual display then it seems like my response time and certitude are related to the strength of the related display's mnemonic. *Strength here refers to how quickly and clearly the mnemonic was retrieved in comparison to the other mnemonics devices.* *(As above, this was not correlated to data recorded by the drill program.)*
· At this point my time spent on feedback is mostly related to getting the tone right. *I have a hard tome distinguishing between medium and low tone at times.*
· End Session.

Session 4: Run 2. Done after first attempts at CS models.
· The speed of retrieval and degree of certitude seem to be related to the strength of the mnemonic.
· At this point there are several displays for which the name just pops into my head, but I'm finding myself checking this against the grouping before I give my response.
· An exception to the last observation seems to be when the name which "pops" in my head is the primary display for a group of three.*(i.e.. Band stand is the primary display for its group of three so when I see Band Stand I get the name quickly and don't check it against it's grouping. When I see back net, a member of Band Stand's trio, I check it against its group before answering.)*
· The general memory retrieval process I am aware of follows these steps*:

Step 1: Is the name immediately in STM?
If yes give answer. If no goto step 2.

Step 2: Can the position, shape and the primary name of the corresponding trio be identified?
If yes goto step 3. If no goto step 4

Step 3: Can the correct name be retrieved from among the identified trio?
If yes give answer. If no goto step 4

Step 4: Go through the list name by name and eliminate those names which are surely not the right answer until you are reasonably certain you have found the right one.

(This is not a clear cut process as the steps weren't always this distinct, yet in the later trials this process seemed to be followed the majority of the time.)

*These steps were written out toward the beginning of this run and were verified over the course of the run.

- End Session.

From the self reports it was noted that certitude seems to be related to the subjects sense of the "strength" of a given memory trace in the memory net, where strength refers primarily to the speed and clarity of retrieval. The records from all sessions of the drill show that the certitude rating was not held constant during the drill. Based on this it seems that certitude rating is not a controlled variable. What it does seem to do is serve as an indicator of the subject's sense of the strength of the memory net for the item retrieved.

Throughout the self-reports a series of reports on current strategies are given. This series of strategies seem to be at PCT's program level. First simple memorization was tried and as the information to be memorized became cumbersome more complex strategies of memorization were implemented. The general strategy for encoding and retrieval is summarized in the 4-step process from session 4 of the drill. Although the particular strategy for learning this information was new to the subject, the general strategies used were already established by the subject in the past to learn basic information.

The subject was unable to identify what variables were being controlled cognitively. Even though the subject was trying to maintain as orientation to learn all of the display combinations, he was not able to maintain by self report if this was in fact a controlled variable. One item that was apparent was that the mnemonic devices that were related to a clear and quick retrieval had stronger certitude ratings and quicker response times. This applies more to the extreme cases. The display combinations that had very weak associations took a noticeably long time to retrieve while those which were very strong were retrieved within 1-2 seconds. The subject was not able to distinguish between the retrieval times for items which had similar association strengths.

Phase II: Investigation into models of information processing

The self-reports from phase I were focused primarily on the strategies used to encode into and retrieve from memory the drill information. Because of this it was decided that the main focus of the flow charts would be to model the mechanics of memory storage and retrieval. Initial attempts at creating flowcharts models were frustrated by an inability to design a memory component into the control system. More information was needed about how information is processed in memory. In order to obtain more information in this area the information processing models of J. R. Anderson (1976) and Stephen Grossberg (1982,1986) were studied.

First Anderson's model was investigated. His model views memory as a system of nodes that are connected in a network. Each node represents some piece of information and the connections in the network represent the connections between these pieces of information. The means of memory retrieval in Anderson's model is related to what he calls spreading activation. This is where a stimulus perturbs the memory net and this results in the spread of this perturbation throughout the memory network until an steady state is reached. In this view each memory is represented by a pattern in the memory network.

Anderson's model helped to shed some light for the study in that the idea of a memory network was helpful. The ides of perturbations to a memory network as a way to activate changes in the memory network was also an important help. Although his model was helpful, it sometimes contradicts what is known about neurological functioning (Grossberg, 1986). For example, Anderson's theory states that the amount of activation arriving at a node decreases with an increase in the number of links traversed. Instead neurological evidence shows that activations do not act passively nor do they decrease across nerve pathways (Kluffler & Nicholls, 1976). As well, Anderson's model isn't nearly as rigorous mathematically as Grossberg's. For this reason it was decided that the majority of the time would be focused in Grossberg's model.

Grossberg views the memory as a series of interconnected nodes as well but there are several differences. The mathematics which Grossberg applies is based of a series of differential equations which are used to describe the strength of both short and long term memory traces which exist between nodes.

These memory traces come into existence through perturbations initiated at nodes in the memory network. The strength of the signal then falls of as a function of time (Grossberg 1983). This is true for both long and short term memory.

Grossberg's model is deeply rooted in both neurological studies and memory processing studies (Grossberg, 1983). This was particularly important because PCT as well is based, at least at the lower levels on what knowledge is now available about the nervous system and the brain (Powers, 1973). This being the case it seemed as though Grossberg's model would fit more closely with PCT than others.

The primary help that was derived from Grossberg's model were that his view of memory led to the idea of a scope of scanning mechanism (SSM) for the flow chart models. This part the control system flowchart models would measure activation activity in the memory net as well as cause the necessary perturbations to occur to get a proper configuration in the memory net. In his model long and short term memory interact and affect each other. Encoding happens primarily when short term memory is strongly perturbed and results in changes to the long term memory network. In retrieval, long term memory nodes are activated and this results in the activation of related short term memory signals.

Grossberg's model also views memory encoding and retrieval as two parts of the same process instead of two completely different processes. It also seems to view storage and retrieval as occurring concurrently. If this is the case then instead of having a control system for both encoding and retrieval of information one control system can be used for both.

Another advantage of Grossberg's model is that he has supplied a significant amount of mathematical modeling in his theory. This makes his work conducive to adapt to a testable control system memory model. His equations could be used or drawn from when constructing a mathematical model of memory in control system models. This could be helpful as a reference because PCT has only a limited study on memory and its role in human learning.

Phase III: Control system flow charts

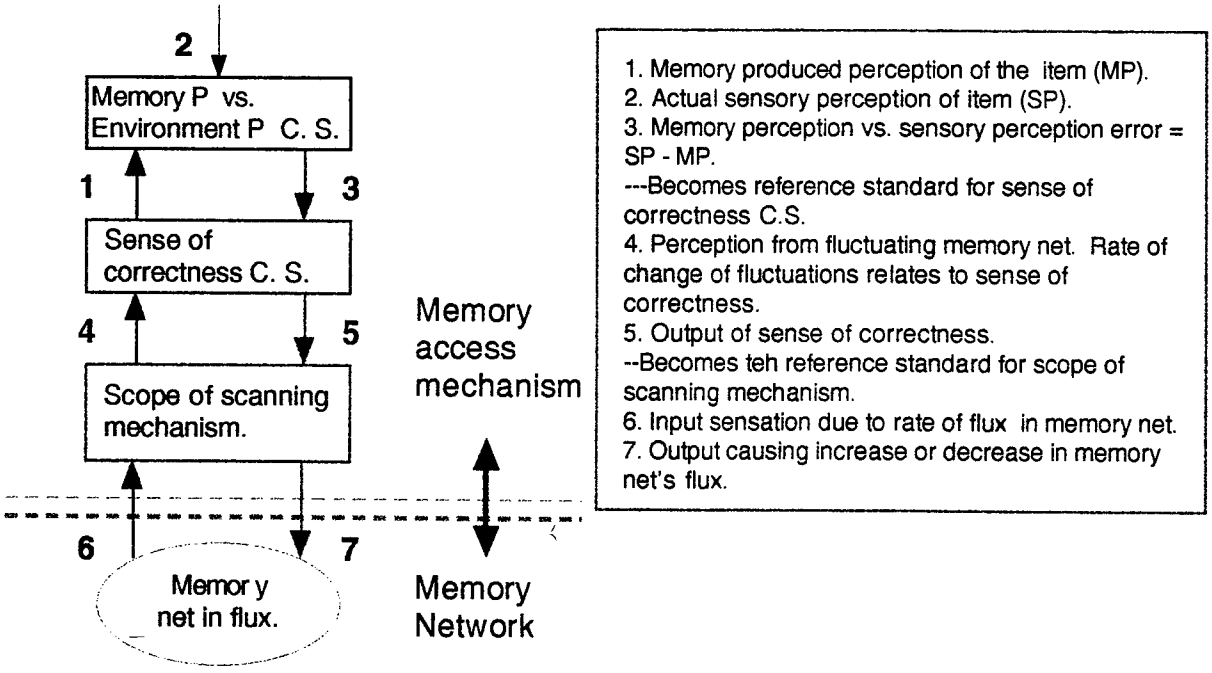
Having gone through the first six runs on the radar drill and reviewing the models of Grossberg and Anderson several flowchart models were attempted.

The primary goal of these flow charts was to create a design that would model the processes of memory storage and retrieval during the radar drill.

Flowchart 1

In this first attempt at a flow chart model, multiple levels of control were attempted. In this certitude and response time were being modeled as inputs for the control system during memory retrieval. The model, shown in figure 1, attempts to represent the mechanics of memory retrieval which occurs at the stage of the first window of the radar drill (identifying the name of a display combination). In addition, memory is represented here as a system of nodes. The memory net in flux views memory as it is in Grossberg's model. This means that the memory net is a series of nodes in constant flux. The strength of these connections between nodes is related to the ability and speed of retrieval for any given piece or combination of information in the memory network

Figure 1: Version 1 of memory control system.



Because of the complexity of memory processing this model has three levels. The highest level in this model is the perception control system (CS). Below this level is the sense of correctness CS and at the lowest level of control is the scope of scanning mechanism. The three levels of control

systems interact with the memory net. The memory net is thought of as being in constant flux and as producing a perceptual signal for given displays in the radar drill.

The key aspect of this model is the highest level of control which is the memory perception/environment perception control system. At this highest level the control system's input is a perceptual signal produced by the memory and the reference signal is the perceptual signal from the actual environment. The two are compared and the error between the two is the output of the control system.

This model has perceptions of the environment at the highest level of the control system, in fact it is the primary reference standard for the model. This appears to be upside down with respect to the traditional control theory paradigm. The important thing about this model is that it views the memory net as the environment. The purpose of this model is to produce memories which match the current perceptions from the radar drill. Thus the memory net produces imitation environmental perceptions and these are compared to the actual environmental perceptions at the highest level of control.

Attempts were made to make the model with the memory signal at the highest level and environmental perceptions at the bottom. No way could be found in which the environmental perceptions could actually result in changing what is in memory. The result of many trials was to put the environmental perceptions at the top as the reference standard.

At this highest level of control in the model a perception from the memory network is received as an input (see figure 1). This input is compared to the actual perceptual signal the subject is perceiving form the drill program. Any error results in a non-zero output which becomes the reference standard for the sense of correctness CS.

The input for this sense of correctness CS comes from the rate of change of the flux in the memory network. The more fluctuation per unit time the lower the sense of correctness which might be measured indirectly by the subject's certitude rating. The input is compared to the reference standard and the error that results becomes the reference standard for the SSM.

At the SSM level an input from the memory net gives the current rate of scanning. If the current rate of scanning doesn't match the reference level then the system outputs a signal to drive a change in the current memory

scanning level. The response time is the amount time it takes between the initial perception of the radar drill frame until the error in the perception CS is zero.

The hierarchy of this model quickly comes to problems. First, sense of correctness may not have a relationship to errors in memory vs. environment if the person is not controlling to be correct. It is also hard to see a relationship between how the error from the memory vs. environment CS can become a viable reference standard for sense of correctness and then scope of scanning.

This model does however provide a new way of looking at the retrieval of memory in a control theory format. In fact, the main benefit of this model was that it served as a springboard for the second memory model flowchart.

Flowchart 2

The next flowchart looks at memory storage and retrieval as a series of control systems (figure 2). It is based on flow chart 1 and tries to describe a possible mechanism by which memory can be stored and retrieved. In this model, memory is viewed as an active network which outputs a pattern of impulses which simulate the perceptions from the current frame of the drill. At the highest level of this model, the memory produced perception is compared to the perception of the item to be learned.

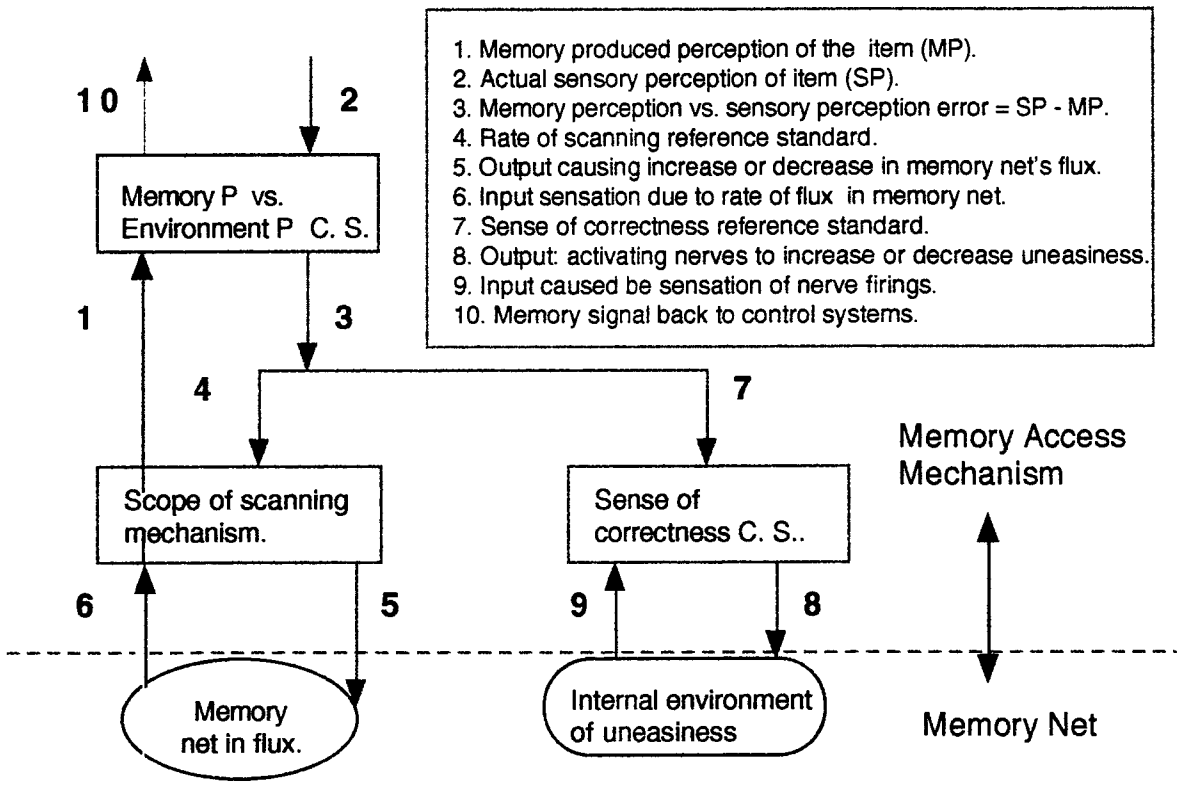
The driving component of this model is the control system which compares the perception from environment to the perception from memory. In this control system an input signal is received from memory. This input signal is a signal coming from memory which appears to the control system to be a normal perception. The input is compared to the environment perception of the frame in the drill program. The environment perception acts as the reference standard for the control system as it is the signal which the memory is trying to duplicate.

If there is a match between the two signals then a zero error exists, otherwise a memory/environment error occurs. This output error acts as a reference standard for the other two control systems. The first of these is referred to as the scope of scanning mechanism. This is the control system which directly drives the fluctuation in the memory net.

This SSM has an input of intensities coming from the memory net fluctuations. The intensities of the fluctuations are compared to the desired

reference level of fluctuations and an error results in an output which causes a change in the fluctuations of the memory net. This is the control system which would cause the perturbations to the memory net and cause changes in long or short term memory. In addition, this could be the place where reorganization occurs in a control system paradigm.

Figure 2: Version 2 of memory control system.



The third control system in this model relates the sense of correctness to the error in the memory signal comparator. The error from the memory becomes the reference standard for the sense of correctness control system. The input for this CS comes from the subject's internal environment of uneasiness (perhaps measured indirectly by certitude rating). An error from this CS results in a change in the subject's internal environment of uneasiness. This part of the model assumes that the subject has some type of standard which requires a sense of correctness as well as having a match in memory. This control system would not have the same relationship if this assumption was not true for an individual.

As in the first flowchart model, environmental perception is the highest reference level in this CS. This is in contrast with Powers' models which

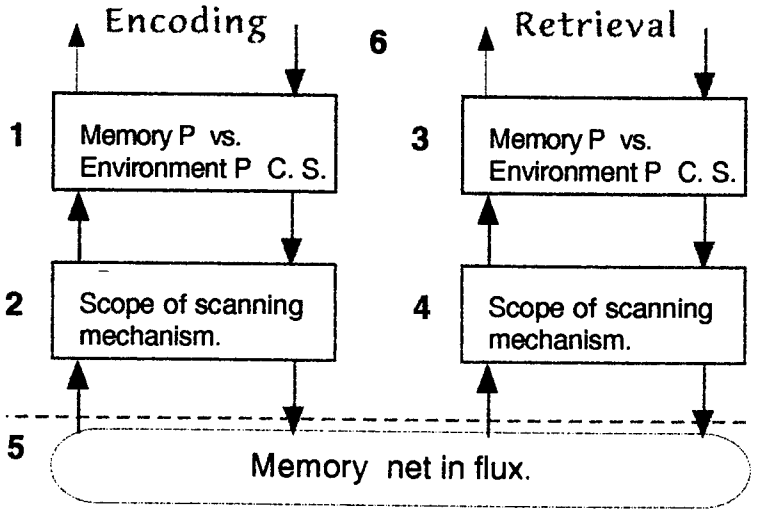
always have environmental perceptions coming in as inputs from the lower levels (Powers, 1973). The reason for this contrast is that this model sees memory as something being operated on by the higher level systems with the goal that what is currently in the memory net will be the same as the perceptions of the environment or what is being learned. Here memory becomes the environment for the memory model control system.

This model gives a description of the mechanism of memory in the activation and stimulus of the memory net. As well it takes care of the problem of where the sense of certainty could fit in. The next step would be to look at a flowchart model which would take into account both the encoding and retrieval of memory. This is needed because Short term memory is perturbed during encoding and long-term memory is perturbed during retrieval. Flowchart 3 addresses this point.

Flowchart 3

This model addresses for both encoding and retrieval. It is based directly on flowchart 2 and the control systems at each level have similar functions. The addition to this flowchart is that it has one hierarchy of control systems for the encoding of memory and another for retrieval (figure 3). This is were Grossberg's model would be key. For encoding the (SSM) would cause a change in the memory net. In retrieval the SSM would cause only a scanning of the current memory net. These systems would be able to function simultaneously.

Figure 3: Version 3 of memory control system.



1. Memory perception vs. actual perception comparitor for memory encoding.
2. Scope of scanning device which causes changes in the memory net during learning.
3.Memory perception vs. actual perception comparitor for memory retrieval.
4.Scope of scanning device which only causes further scanning during retrieval.
5. Memory net in flux & changing due to the influenc of these control systems.
6. Memory signals outputted to other control systems.

The hierarchy of control systems to the left of figure 3 represent the encoding process of memory. For this side the scope of scanning mechanism has the capacity to cause restructuring of the memory net so that it outputs the correct memory signal. The other hierarchy of control systems represents the retrieval mechanism of memory. Here the scope of scanning mechanism only has the capacity to cause further scanning of existing memory. In this model encoding and retrieval could act concurrently or perhaps a higher level control system would cause one or the other to operate.

Discussion

This study has looked at several aspects of memory with a PCT framework. Although the ultimate goal of a working PCT memory model is still far off, the information in this study may prove to be a viable start into such a model. As well, the study has led onto some suggestions for future studies based on the work of Dr. Hancock and Dr. Thurman as well as this study.

Phase I

In the self report phase of the study it was found that nearly all of the self reports were related to strategies used to encode and retrieve the information from the drill program. The primary key to memorizing the information was the use of mnemonic devices. Certitude, speed of retrieval (possibly measured by response time) and clarity of the memory were all related in the self reports. This indicates that the measures of the drill program match the items that the subject was aware of during the drill. One modeling strategy would be to investigate the strategies used by subjects in such drill programs. The strategies would differ by individual subjects so some nominal measurements for the model would need to be used to discover and define the strategies used by each subject. Finding such nominal measurements for such a study would be key.

The problem with these types of measurements is that they are descriptive. They can be used to indicate how well a subject's control systems are operating and even give some insight into what is being controlled, yet these measures can't be used as input variables for a mathematical models of memory. Such input variables can be hard to find as there is no way, in a drill such as this, to directly measure inputs into higher level control systems. Still some variables need to be found which will be more useful and measurable as inputs.

Phase II

Because of the heavy focus on encoding and retrieval strategies from the self reports, a closer look at memory was made. In attempting some primary flowcharts to the cognitive processes used for the radar drill, we concluded that memory was the key component of the process. This conclusion was reached based on the realization that not one frame of the drill program is run without some type of memory retrieval or encoding. In addition memory may be a key aspect to creating any models of higher level control systems.

The memory network models of both Anderson and Grossberg were useful in the design on the flowchart models of memory. The advantage of Grossberg's model is that it has a rigorous mathematical model of memory that could be easily adapted once a reasonable flowchart model was encountered. Both models view that memory is a network of nodes that are connected and restructured do to perturbations of the nodes in the network. These models, along with the insights gained from the self report, led to the construction of the flow charts that are used in this study. Grossberg's model would come into play as the model of the memory met in flux in the flowcharts presented in this study.

Again the problem of having variables to measure the inputs of such a model are not currently available. To make a functioning model of memory there is a need for direct of indirect measures of he perturbations to short and long term memory nodes. Finding such measures would be an important help to future memory modeling attempts.

Phase III

The flowcharts attempted in this model have two features worth noting. First is the view of memory as the environment for the encoding and retrieval mechanism. The whole goal of such a model would be to cause the memory network to match the perceptions that the subject has at any level of control outside the memory model. For example, at the category level, a frame on the drill may present the category 'high tone'. This becomes the reference signal for the memory model. The model will cause perturbations to the memory net until a configuration of the memory net is reached that outputs a matching category perception. This view of the memory net as the environment results in the inverted hierarchy of control with environmental perceptions as the control standard at the highest level. This allows the memory to be affected and changed by the environmental perceptions. Because people control their

perceptions from the environment this model still has the person as the ultimate control over memory yet allows for the altering of the memory network.

The second feature of the models is the memory net in flux aspect of the model. The memory encoding and retrieval device perturbs the memory net. The response and output as a result of these perturbations would be calculated. Grossberg's model would be used at this level to show what the affects of the memory net are and would produce the outputs from the memory net back to be compared to the present-time perception. Thus we would have a clear specification of the input to and the output from the memory net.

In conclusion it would be good to identify the similarities and differences between this view of memory and the traditional one used in Powers' PCT model. Powers views memory as existing at each level of hierarchy (Powers , 1973). This study's flow chart model strictly keeps the perceptions coming from memory the same as the perceptions coming from the environment, therefore, it could do this at any level of perception as well. A major difference is that the model presented in the study presents memory as being constantly accessed while in Power's model memory is accessed through gates which can be either open or closed. These gates are opened or closed as needed by the related control system. Although Powers' model is clear and founded on proven models, the mechanism by which memory is accessed and how the gates are opened or closed is unclear. What memory's role is in the functioning of a given control system is unclear as well.

The models presented in this study offer one suggestion as to what a possible mechanism of memory access may be. Further investigations into the mechanisms and place of memory in human cognition are definitely needed.

References

Anderson, J. R. (1976). Language, memory, and thought. Hillside, NJ, Erlbaum.

Ericsson, K. A. & Simon, H. A. (1980). Verbal report as valid data. Psychological Review, 8, 215-251.

Grossberg, S. (1982). Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns. Competition and Cooperation in Neural Networks. New York, Springer-Verlag, 448-485.

Grossberg, Stephen (1986). Pattern Recognition by Humans and Machines. Patterns of Humans and Machines. Volume 1, Speech Perception. Orlando, Academic press, 187-285.

Hancock, T. E., Thurman, R. A., & Hubbard, D. C. (1993). Performance predicted by student processing models. Presented to Division 21 of the American Psychological Association Annual Convention, Toronto.

Hancock, T. E., Thurman, R. A., & Hubbard, D. C. (1993). The use of post-response information: Explanations according to the discrepancy construct of control theory. Presented to Division 21 of the American Psychological Association Annual Convention, Toronto.

Marken, R. S., & Powers, W. T. (1989). Levels of intention in behavior. Volitional Action, Elsevier Science Publishers, B. V. , 409-430.

Kluffler, D. H.. & Nicholls, J. G. (1976). From Neuron to Brain. Sunderland, MA, Sinauer.

Powers, W. T. (1978). Quantitative analysis of purposeful systems: Some spadework at the foundations of scientific psychology. Psychological Review, 85, 417-435.

Powers, W. T. (1973). Behavior: The control of perception. Chicago: Aldine.

FURTHER EXPLORATIONS IN EPISTEMOLOGICAL SPACE

Susan T. Chitwood
Research Assistant
Department of Psychology
&
Ryan D. Tweney
Professor
Department of Psychology

Bowling Green State University
Bowling Green, OH 43403

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Wright Patterson Air Force Base, Dayton, Ohio

and

Armstrong Laboratory

September, 1994

FURTHER EXPLORATIONS IN EPISTEMOLOGICAL SPACE

Susan T. Chitwood & Ryan D. Tweney
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

Abstract

An earlier claim made by Chitwood & Tweney (1993) was tested, specifically, that the epistemological environment encountered by Michael Faraday (1891-1867) in the course of his successful investigation of the properties of electromagnetic induction, was sufficiently "chaotic" such that Faraday must have used something equivalent to a confirmation heuristic. That investigation utilized standard back-propagation neural networks, which, when trained with vectors representing components of 54 of Faraday's experiments, were unable to learn the vectors representing the outcomes of those experiments. The present study explored four variations of network types and configurations in efforts to disconfirm the prior claim. Despite numerous manipulations of network parameters, none of our network instantiations led to successful learning when all 54 experimental representations were included in a training set, although when the "null" outcomes (those in which Faraday did not achieve an effect) were excluded from the training set the networks could learn the remaining (i.e. "confirmatory") outcomes quite well. Converging results from each network instantiation suggest that Neural Networks are promising tools for the exploration of the epistemological properties of scientific work. In particular, examinations of the results of our final network manipulations which utilized a Kohonen self-organizing algorithm, while incomplete, support our stance that "epistemological space" is a very important notion towards a cognitively based understanding of successful scientific endeavors. Other areas of expertise (e.g. piloting) may also be appropriate for Neural Network examination.

FURTHER EXPLORATIONS IN EPISTEMOLOGICAL SPACE

Susan T. Chitwood & Ryan D. Tweney

The present paper describes four studies which follow from an earlier series described by Chitwood & Tweney (1993). In that paper, a method was developed for using neural networks to evaluate the epistemological space of 54 experiments conducted by Michael Faraday (1791-1867) in the course of exploring his most notable discovery, that of electromagnetic induction, in 1831. In brief, the diary records kept by Faraday were encoded by assigning numerical values to attributes for each of the experiments in the series. Values were assigned for aspects of the equipment used, the physical setup of the experiment, the actions carried out by Faraday, and the results of each experiment. Thus, each of the 54 experiments was represented by a single vector having 40 components, and, for each experiment, a second vector having six components to describe the outcome of the experiment.

The setup vectors were presented as input to a neural network having 40 input nodes, three hidden units, and six output nodes. The six-component results vectors were used as training outcomes using backpropagation as the method of learning. Thus, each input was presented in turn as input, and the output calculated and matched against the intended target vector. Discrepancies between the outputs and the target were used to compute "delta" values which, in turn, were used to modify the weights of the neural net. This procedure continued until the discrepancy between target and result was minimal.

In the earlier study, we found that neural networks were unable to learn to predict the outcomes of the experiments when the entire set of 54 experiments were used, though smaller sets, representing 12 to 15 experiments, could be learned quite easily. We were able to show that the reason for the difficulty resided in the fact that many of the experiments recorded by Faraday were "null" outcome experiments that did not produce the intended result. Since many of these were highly similar to successful experiments, we were, in effect, presenting the network with an epistemologically chaotic environment, in which highly similar setups sometimes led to an expected result and sometimes did not. We confirmed that this was the source of trouble by separating

the experiments into a group of 27 successful outcomes and 27 unsuccessful outcomes. The neural networks were able to learn the set of successful studies but continued to do poorly with the unsuccessful ones. This outcome was used to support a claim that Faraday must have used something equivalent to a confirmation heuristic to make sense out of his results; had he treated each outcome as representing evidence to be equally weighted in reaching a conclusion (as a neural network must do), then he could not have reached the conclusions that he did.

The present investigation was conducted in order to further test this claim. Since our major result depends upon a failure of a neural net to learn a particular set of input patterns, and this failure is used to support inferences about the epistemological chaos of the space within which Faraday worked, it is important to show that in fact learning is not possible under the circumstances we have described. It is, of course, difficult to base a claim on a failure to find learning. Though a number of different parameters of network performance and learning were manipulated in attempts to find an optimal configuration (e.g., the number of hidden units was manipulated, the learning coefficients, and so on), there is no assurance that no such combination of variables could lead to successful learning.

The problem was approached in several ways in the present series of studies. First, we replicated the initial findings using somewhat different parameters, and including one additional data analysis to confirm our earlier outcome. Second, a more powerful learning algorithm was implemented,one in which each weight in the network was allowed to change separately from the others (rather than each weight change being dependent on all of the others). This permitted the learning procedure to "tailor" individual units to particular input combinations. Third, following an earlier study carried out by R. Chadwick (personal communication, May, 1994), additional input nodes were added to each experiment, representing Faraday's "expectations" for the outcome of each experiment, in order to determine if this would disambiguate the chaotic character of the epistemological space. Fourth, a self-organizing procedure was used to determine whether the inferred overall organization of the experiments (successful vs. unsuccessful) could be found in an unsupervised learning situation in which the network had to "discover" any underlying patterns.

Study 1 - The Basic Paradigm: A Replication

Study 1 Methodology: The beginning step was a replication of results obtained by Chitwood & Tweney (1994). To accomplish this, we built a three-layer neural network consisting of 40 input units, three hidden units, and six output units. Three sets of input files were used for training, one consisting of all 54 experiments, one consisting of 27 experiments in which Faraday's expected outcomes were achieved (the "Confirmatory" set), and one containing the 27 experiments in which his expected outcomes were not attained (the "Null" set).

All networks were trained using a sigmoid transfer function for the hidden units and a sine function for the output layer. Standard backpropagation learning procedures were used. Learning rates and momenta were varied as a function of trial number for the complete set, and differed for the hidden layer and the output layer connections. The exact schedule used is given in Chitwood & Tweney (1994). For the Confirmatory and Null sets, learning rate and momentum were constant across trials, but varied with layer, as in the earlier study. In all three cases, epoch size was set equal to number of experiments in the input set. All inputs and outputs were scaled to lie within the range -1.0 to +1.0. The Confirmatory and Null sets were run for 50,000 trials each, and the Complete set for 100,000 trials.

Study 1 Results: Outcomes were analyzed by assessing deviation scores for each experiment. Thus, each output from the network, after training, was matched against the target, and the sum of the absolute value of the node by node discrepancy (a "Deviation" score) was computed. In general, the Complete set was not learned to any greater extent than in the previous study, with the lowest overall Root Mean Square Error (RMS Error) being still quite high. As before, the Confirmatory set was learned to very low levels of RMS Error, and the Null set was hardly learned at all. (RMS Scores are not reported here, since they are much less informative than the more specific deviation scores.) Deviation scores were plotted for all three sets and are shown in Figure 1. Note that, by plotting the Confirmatory and Null sets separately on the same axes for the Complete set, it is possible to see whether the Complete training set led to differential learning of these two types. It is clear from Figure 1 that this is not the case: as before, combining the two kinds of experiments appears to create an "unlearnable" situation for the network.

Study 2 - Backpropagation: Another Approach

Perhaps the difficulty encountered by the networks in learning the entire set of 54 inputs and the null set is based on idiosyncrasies encountered during the course of learning. Thus, we have argued in the earlier paper that the set of 54 experiments taken as a whole is epistemologically chaotic in character, but perhaps what we are claiming to be chaotic is in fact merely difficult because of the need to explicitly tailor each response to a particular input. If so, then perhaps the appropriate outcome of each experiment could be learned if the specific weights of the network were individually modified. To explore this possibility, we constructed networks in which training was carried out using a "Delta-Bar-Delta" algorithm.

The usual backpropagation algorithm, as described by Rumelhart & McClelland (1986), relies upon a "steepest descent" algorithm in which the backpropagated weight changes are chosen in such a way that they point in the direction of steepest descent, i.e., at each step, weight changes are based on the overall direction that will maximally reduce the discrepancy between the output and the target. Such a procedure is very powerful but can also be very slow, particularly if the local error surface is canyon-like, i.e., possesses a deep and long chasm. Steepest descent algorithms then can get trapped in a kind of "zigzag" in which the weight changes jump from one side of the canyon to the other, rather than angling across the canyon wall to get to the minimum.

Jacobs (1988) described an algorithm, the Delta-Bar-Delta procedure, that can avoid this difficulty (see also Masters, 1993). In effect, the algorithm looks at the past history of each weight change to determine if the sign of the change has been oscillating. If so, then the learning rate for that weight is decreased. If the sign has remained the same over several trials, on the other hand, then the learning rate is increased for that weight. In this fashion, weight changes that are trapped in zigzags are reduced in influence and those that are not are enhanced. The result is a network that can learn much more quickly in certain situations. Note too that each weight change is calculated based on local information only, rather than on the global properties of the error surface. In a situation such as ours, where we expect inconsistency between the effects of confirming and null information, the Delta-Bar-Delta procedure may make it possible for a network to learn all of the relationships in the input set.

Study 2 Methodology: Three layer networks were built for each of the three data sets, having the same architecture as before (i.e., 40-3-6, fully connected). In this case, a sigmoid activation function was used for both the hidden and the output layers. Learning rates, of course, are defined differently for the Delta-Bar-delta procedure. Here, an initial learning rate of 0.30 was used for all weights. This was subject to three other parameters that adjusted the overall learning rate, as described above, depending on local circumstances. The three relevant parameters are a convex weighting factor (0.70 here), a constant learning rate increment (0.07) and a constant learning rate decrement (0.40). Epoch sizes equaled 16 for the Confirmatory and Null sets and 27 for the Complete sets. Training was continued until the RMS Error appeared to stabilize, a total of 20,000 trials for each set.

Study 2 Results: Results were assessed using deviation scores, which are shown in Figure 2 . As before, learning was not satisfactory for either the Complete nor for the Null sets. The best overall performance occurred with the Confirmatory set, though here also performance is low, compared to the prior study outcome. Note however that the Delta-Bar-Delta procedure shows marked improvement in one respect over the results of the earlier study: by comparing Figure 1 and Figure 2, it is clear that the performance of the Complete set network on the confirmatory and null experiments considered separately is comparable in character to the performance of the separate networks trained on these subsets. Note in particular the long flat plateau in the center of the Null set items. In spite of this, however, it is still apparent that the Delta-Bar-Delta procedure has not, in general, led to a markedly different picture from that established earlier.

Study 3 - Adding Cognitive Expectations

R. Chadwick (personal communication, May, 1994) suggested that the reason why the complete training set of 54 experiments could not be learned was that the network had no way to know which domain was the locus of the expected effect for each experiment. By adding an "expectation" about what a particular experiment was intended to show, perhaps the learning problem faced by the networks could be overcome. In effect, Chadwick proposed adding a cognitive dimension to the input training vectors. He tested this notion by building networks which incorporated five additional nodes, one for each domain of expected effects (current, spark, chemical action, heat, and magnetism). Training vectors were coded to have one of these five nodes

activated depending on the intention of the experiment. Chadwick reported that all three training sets could be easily learned using standard backpropagation procedures.

Study 3 Methodology: A possible objection to Chadwick's procedure, however, is that adding one node per expected outcome in effect trivializes the learning task, that is, five of the six output nodes could be perfectly mapped from only five of the input nodes. If so, then the remaining output node could be predicted based on one, some, or all of the remaining 40 input nodes. One way around this objection is to code the expectations in a distributed fashion, to avoid the one-to-one mapping between input and output nodes. Accordingly, we added three nodes to the basic set of 40, coding current as 101, chemical action as 011, spark as 110, heat as 010, and magnetism as 100. Both standard backpropagation, as in the first study above, and a Delta-Bar-Delta procedure, as in the second study, were implemented. For standard backpropagation, parameters for the Complete set were identical to the earlier study. Confirmation and Null sets were identical except for a gradually reduced learning schedule across trials. In the second method, all parameters were as in the earlier study above.

Study 3 Results: Results for standard backpropagation are shown in Figure 3, and the results for the Delta-Bar-Delta procedure are shown in Figure 4. Using the standard backpropagation procedure with expectations led to improved performance in all cases; however the Complete set was still not learned to satisfactory levels. The Delta-Bar-Delta procedure with expectations produced no improvement in performance for any of the three sets. Thus, we concluded that adding a cognitive dimension could marginally improve performance, though the fundamental difficulty with the Complete set and the Null set remained.

To summarize the results of the first three studies, it is apparent that the fundamental difficulty described by Chitwood & Tweney (1994) is present regardless of the specific learning remains viable. It is interesting to note, comparing the solutions across the three studies, that certain specific experiments are problematic for the neural networks in all cases. For example, Experiment #21 in the Null series is always anomalous -- either it is the only one learned (as in the first study), or it is among the worst method used; similarly, Experiment #4 in the Null set was always poorly learned. Interestingly, these two experiments are the only two in the entire series in which Faraday was predicting the occurrence of a spark. It is not surprising that

the networks would have difficulty with these experiments -- the experimental setups are close to many of those that sought currents, and yet we are asking the network to recognize them as exceptional in their outputs. This kind of inconsistency is exactly what makes the entire set of experiments "chaotic" in our terms. In all, then, the initial conclusion, that the Complete set is somehow "epistemologically chaotic" remains quite viable.

Study 4 - Self-Organizing Networks

Kohonen (1988) has described a procedure by which a network can learn to categorize input data sets based upon higher-order similarities among elements. Such a procedure, one of a class of unsupervised training methods, permits the network to discover structure without bias toward an a priori classification imposed by the investigator. In the present case, the results of the experiments conducted by Faraday determine the training set used in our earlier supervised learning procedures -- these are not exactly arbitrarily a priori! Nevertheless, it is important to ask whether the input data set alone, i.e., the similarities among the 54 experimental setups, implies any hierarchical structure that is relevant to our claim.

In addition, it is worth asking whether a self-organizing network can restructure the input patterns in such a way that learning of the input output relationships can be facilitated for later backpropagation algorithms. In effect, if a self-organizing net imposes a simplified structure upon the input nodes, can this structure facilitate learning of the outcomes of the experiments in a supervised learning context? And, if so, does the imposed structure make conceptual sense in terms of our overall claims?

Study 4 Methodology: The present study utilized a network having three layers, an input layer of 40 elements, as before, a hidden layer consisting of a 10 (rows) by 5 (columns) array of units (a so-called "Kohonen Layer,") and an output layer (as before) of six output units. Learning occurred in two distinct phases. In the first, unsupervised, phase, the input units were successively activated by each vector in the training set and the units in the Kohonen Layer competed for the connection, the winner being the one unit with the smallest distance between itself and the input vector. Distance is defined here using the 40 components of the input vector to define one point, and the 40 connection weights between each input component and the hidden unit to define a second point. Distance is then taken as the Euclidean distance between these two points. The nature of the training procedure relies upon a "Winner Take All" strategy. At the end of training each input vector was

thus associated with one, and only one, hidden unit, though hidden units could end up representing zero, one, or many input vectors. In addition, to minimize the number of Kohonen units that never win, a "conscience" mechanism was used in which any unit with a recent history of winning was given slightly less chance to win on subsequent trials. This avoids having all of the inputs represented by one or a few hidden units. The pattern of representation thus places input experiments into sets of similar experiments. Once this training was complete, the second learning phase was carried out. The resulting hidden units from the Kohonen layer were used in a standard backpropagation procedure in which the target vectors (the results of each experiment) were used to modify weights between the hidden layer and the output layer. Note that this procedure allows one the advantage of interpreting the meaning of each hidden unit in terms of its implications for the resulting solution. Furthermore, since the Kohonen layer is constrained to use only one unit per input vector, the learning task for the supervised phase is simpler than that in the usual backpropagation -- in effect, the Kohonen layer is removing some of the inconsistencies among input experiments before they can affect the learning of appropriate weights to the output layer.

Each of the three training sets, Complete, Confirmatory, and Null, was used to train five identical 10 x 5 Kohonen nets for 5000 trials each, each of which was back-propagated as described above in a second stage of training, for an additional 23000 trials minimum. Backpropagation was carried out using three hidden units between the Kohonen layer and the output layer.

Study 4 Results: The Kohonen solutions were examined to determine which experiments were captured by the same hidden unit during the unsupervised phase of training. In effect, two experiments which fall under the same Kohonen unit can be considered as very similar to each other. Further, the closeness of experiments in "Kohonen Space" is a measure of relative similarity. The results of the backpropagation phase of training was assessed as in the earlier studies, using the Deviation score to assess how accurately each experiment's outcome was represented by the network. Deviation scores for all five trials of each network are shown in Figures 5, 6, and 7. Generally, results are consistent across trials, though some exceptions are apparent. Since each Kohonen net starts from randomly determined initialization values, it is not surprising that the solutions might vary. Note, however, that the key case, that of the Complete set of 54 experiments is not learned well by any of the

Kohonen nets, and, further, that the pattern of solutions is consistent with the earlier studies -- a remarkable fact, given the difference in the present procedure.

Examination of the specific nodes that code particular experiments within the Kohonen layer was carried out by inspection. Of course, the same nodes do not capture the same inputs on each trial, making this a laborious task. While this analysis is incomplete at present, it is evident that some expected commonalities do show up in the Kohonen layer, while others do not.

Discussion

Closer examination of the results of the Kohonen solutions promises to be very revealing as a way of identifying those aspects of the input data set that are especially problematic for the networks to learn, and hence suggestive of specific heuristics that Faraday may have needed to make sense of his data. Further, the examination of the Kohonen results may suggest possible manipulations of the data set to determine what changes can be made to make the Complete set learnable with ease by a neural network. We have begun to explore the properties of the experiments (as coded for our networks) via statistical tools such as hierarchical cluster analyses and multi-dimensional scaling; the results, though preliminary, suggest that our earlier conclusions are strongly supported by these "standard" tools, and therefore that, a neural network approach to the understanding of scientific activity is more than viable. To start, we plan to move in the direction of creating "Virtual Faradays" as a way of conducting "experiments" to test specific hypotheses about his scientific strategies. If these prove successful, our line of research should transfer to more general areas beyond scientific experiment, for example, to expertise or to the examination of individual differences in laboratory tasks.

References

Chitwood, S. T. & Tweney, R. D. (1993). Why confirmation " bias"?: The epistemological environment of scientific experiment. Unpublished manuscript, Bowling Green State University, Department of Psychology, Bowling Green, OH.

Jacobs, R.A. (1988), Increased rates of convergence through learning rate adaptation. Neural Networks, 1, 295-307.

Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78 (9), 1464-1480.

Masters, T. (1993). Practical neural network recipes in C++. Boston: Academic Press, Inc.

Rumelhart, D. E. & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1. Cambridge, MA: MIT Press.

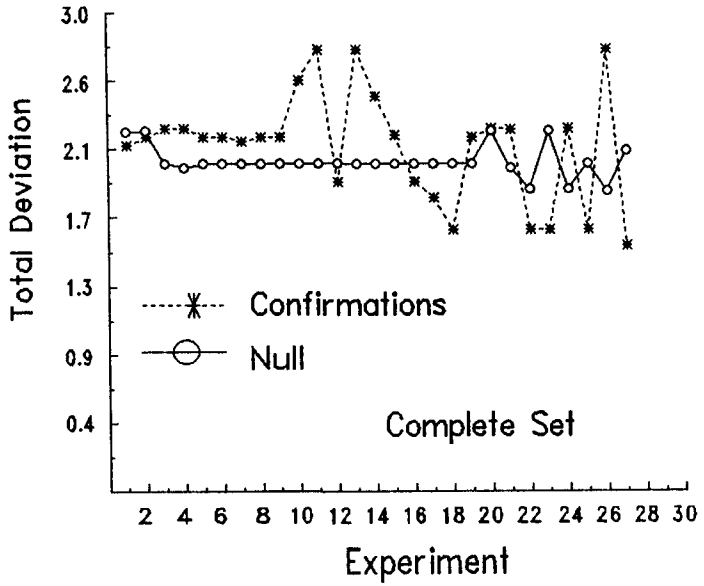
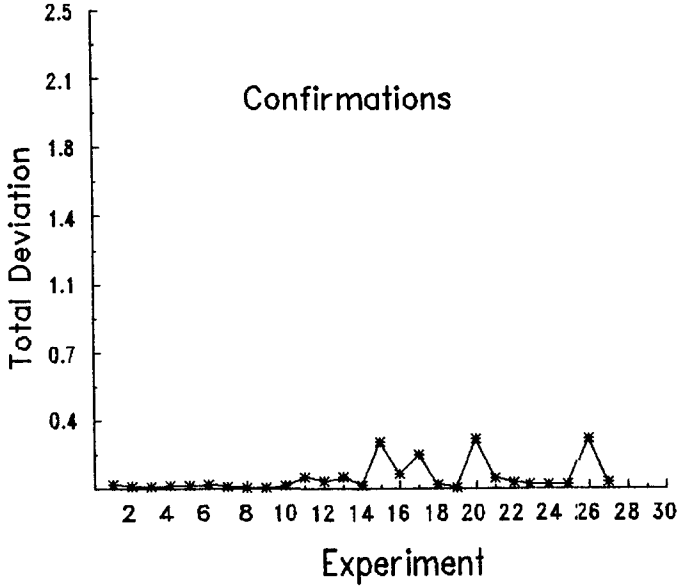
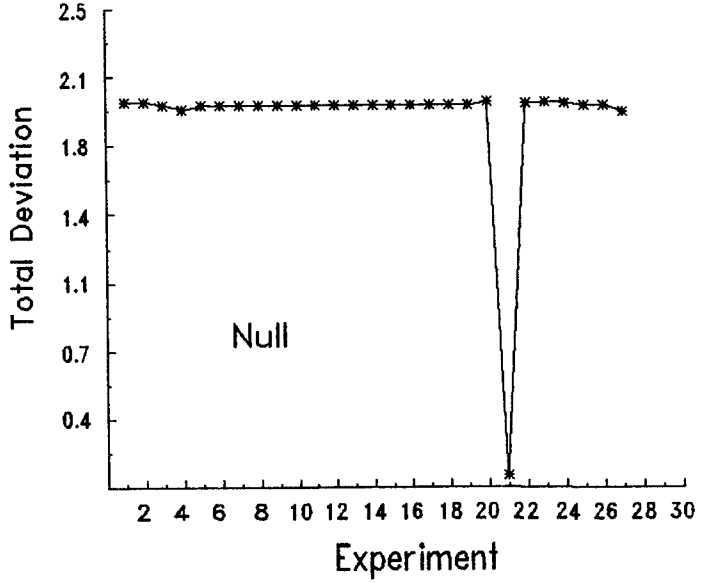


Figure 1
Study 1
The Basic Paradigm





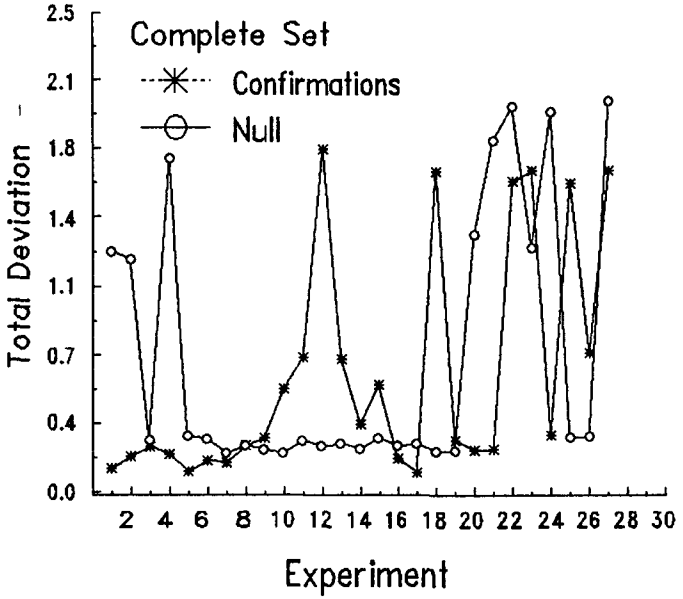
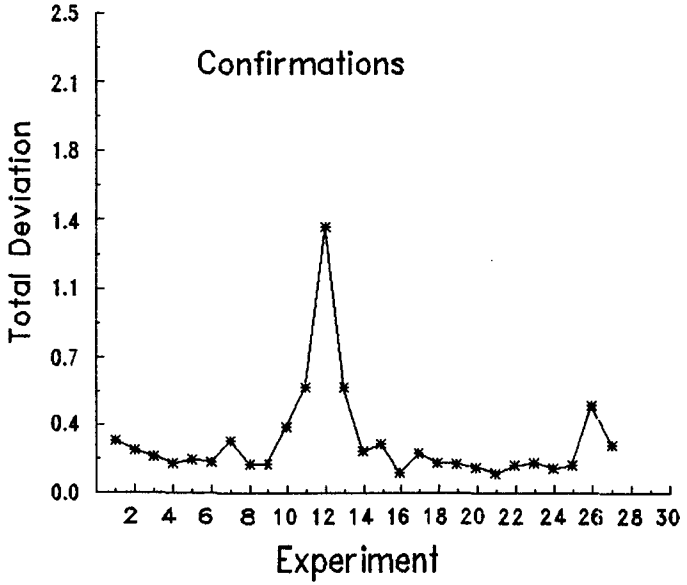
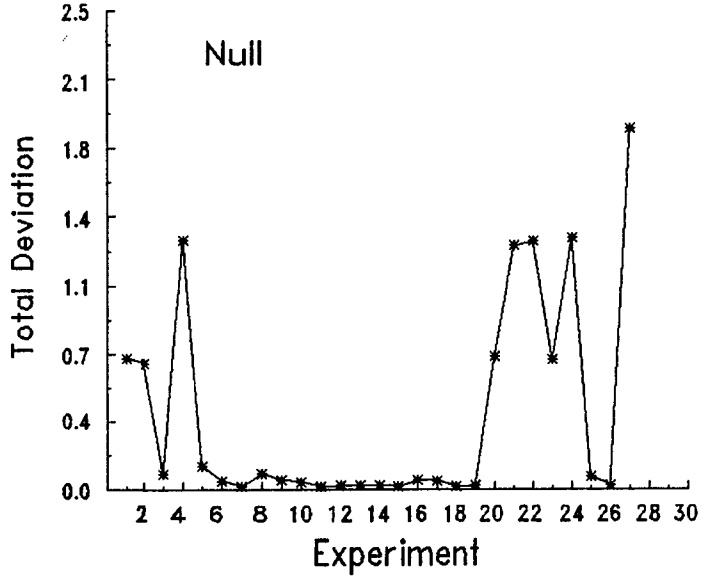


Figure 2
Study 2
Backprop: Another Approach





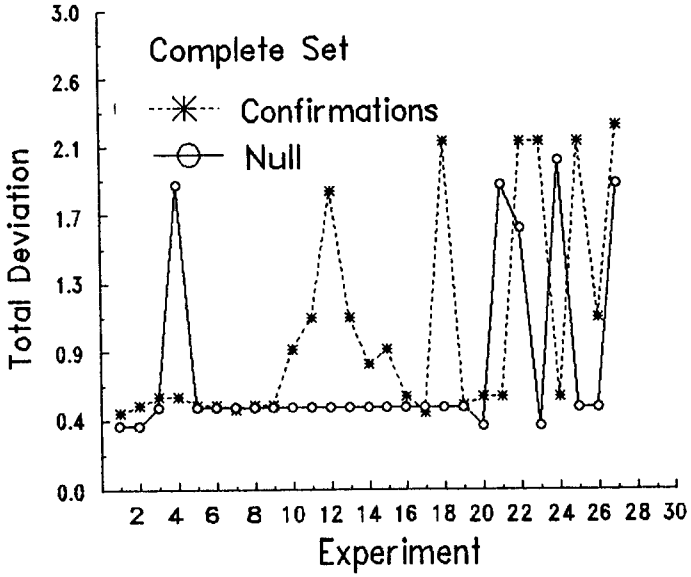
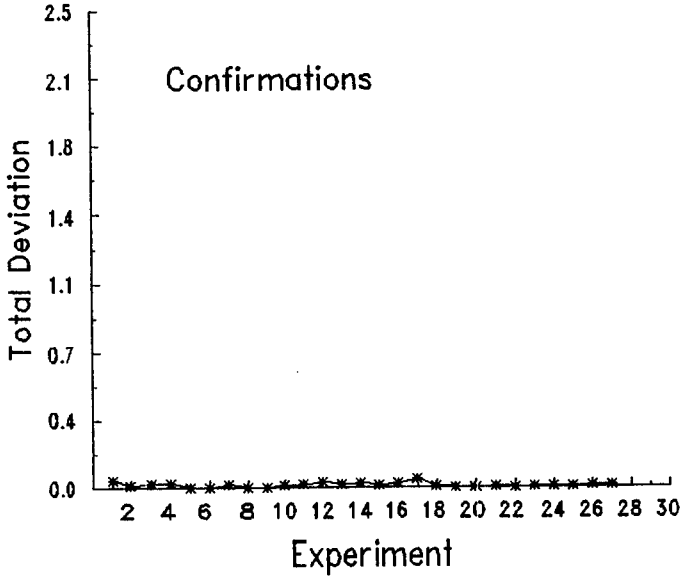
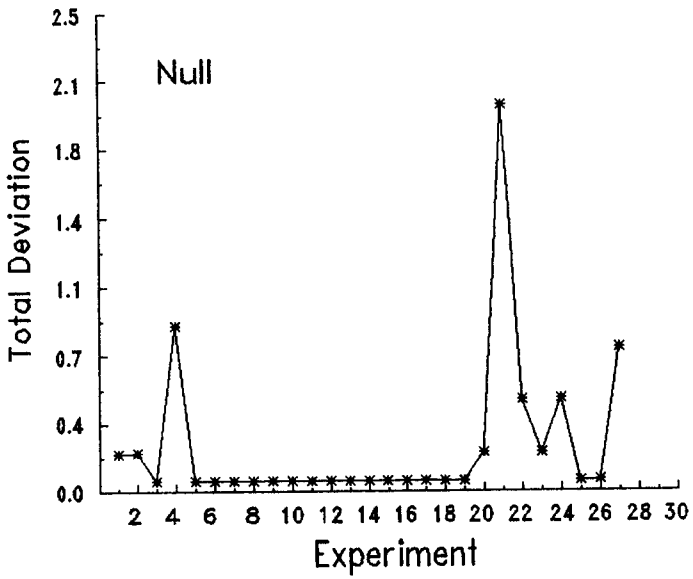


Figure 3
Study 3
Adding Cognitive Expectations
Normal Backprop





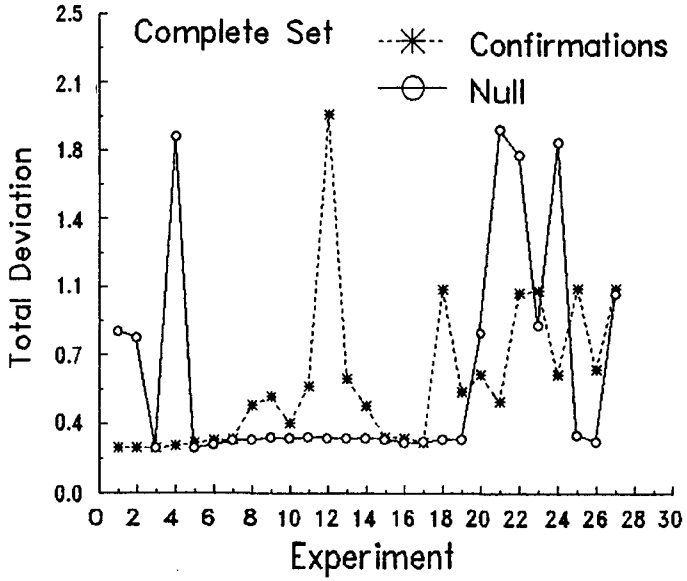
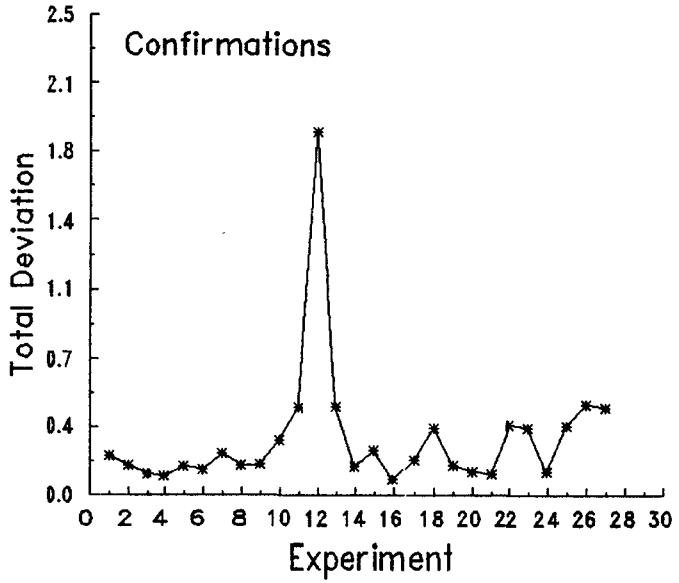
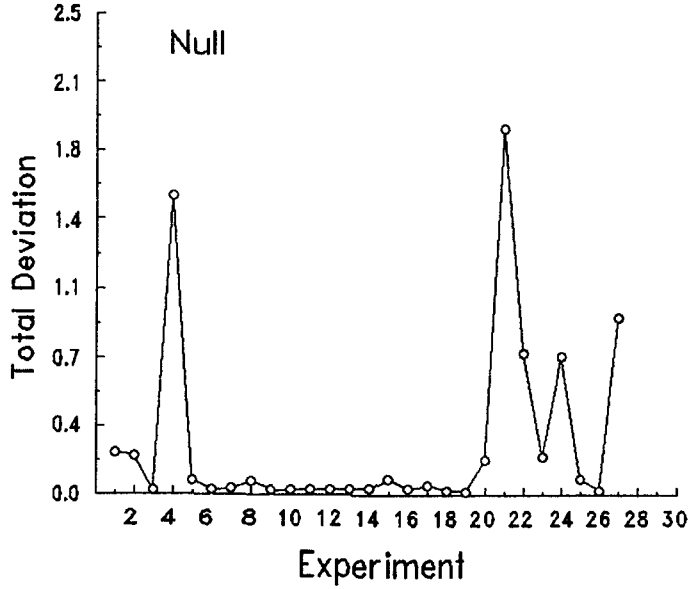


Figure 4
Study 3
Adding Cognitive Expectations
Delta–Bar–Delta





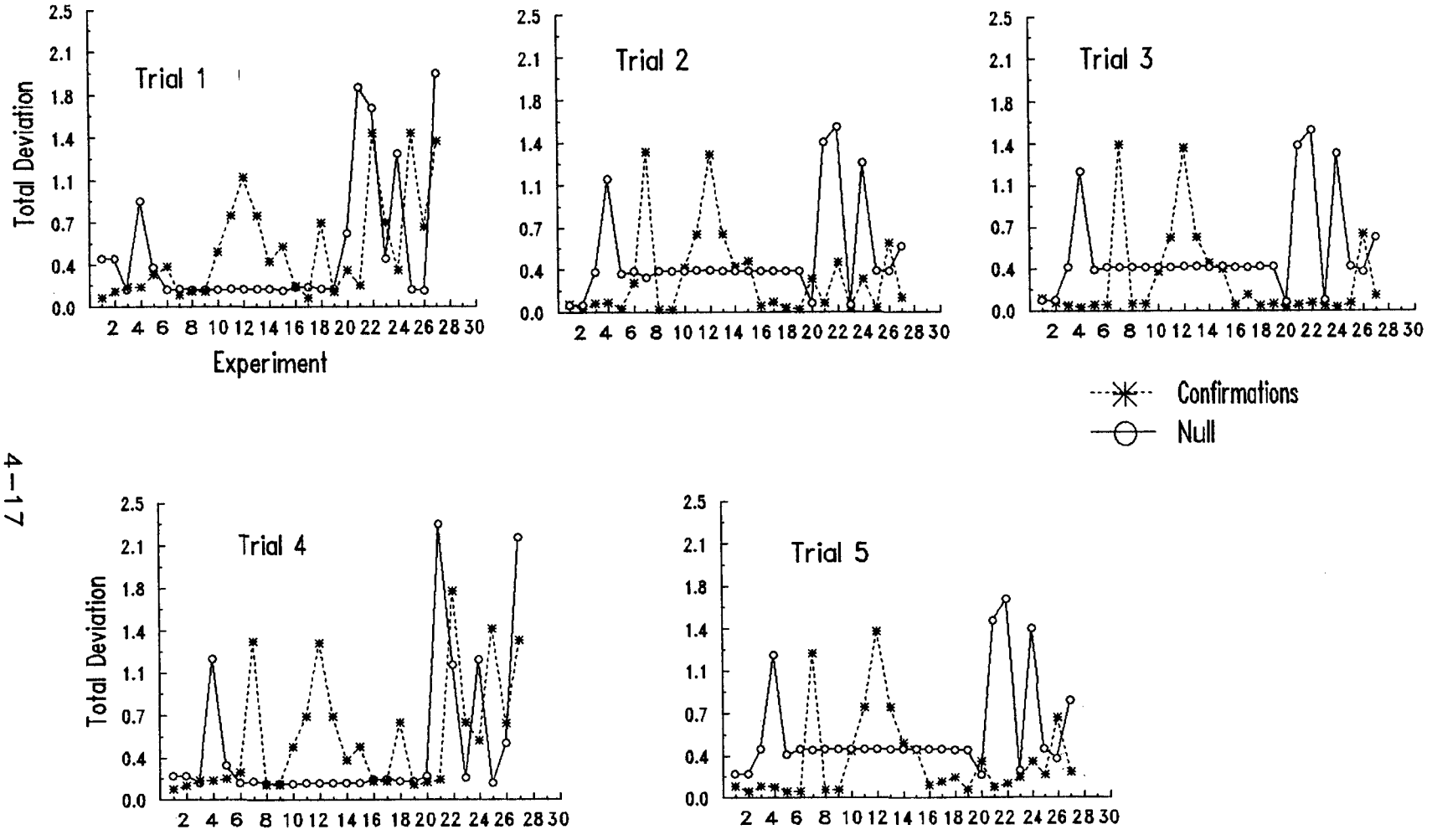


Figure 5
Study 4
Self-Organizing Networks
Complete Set

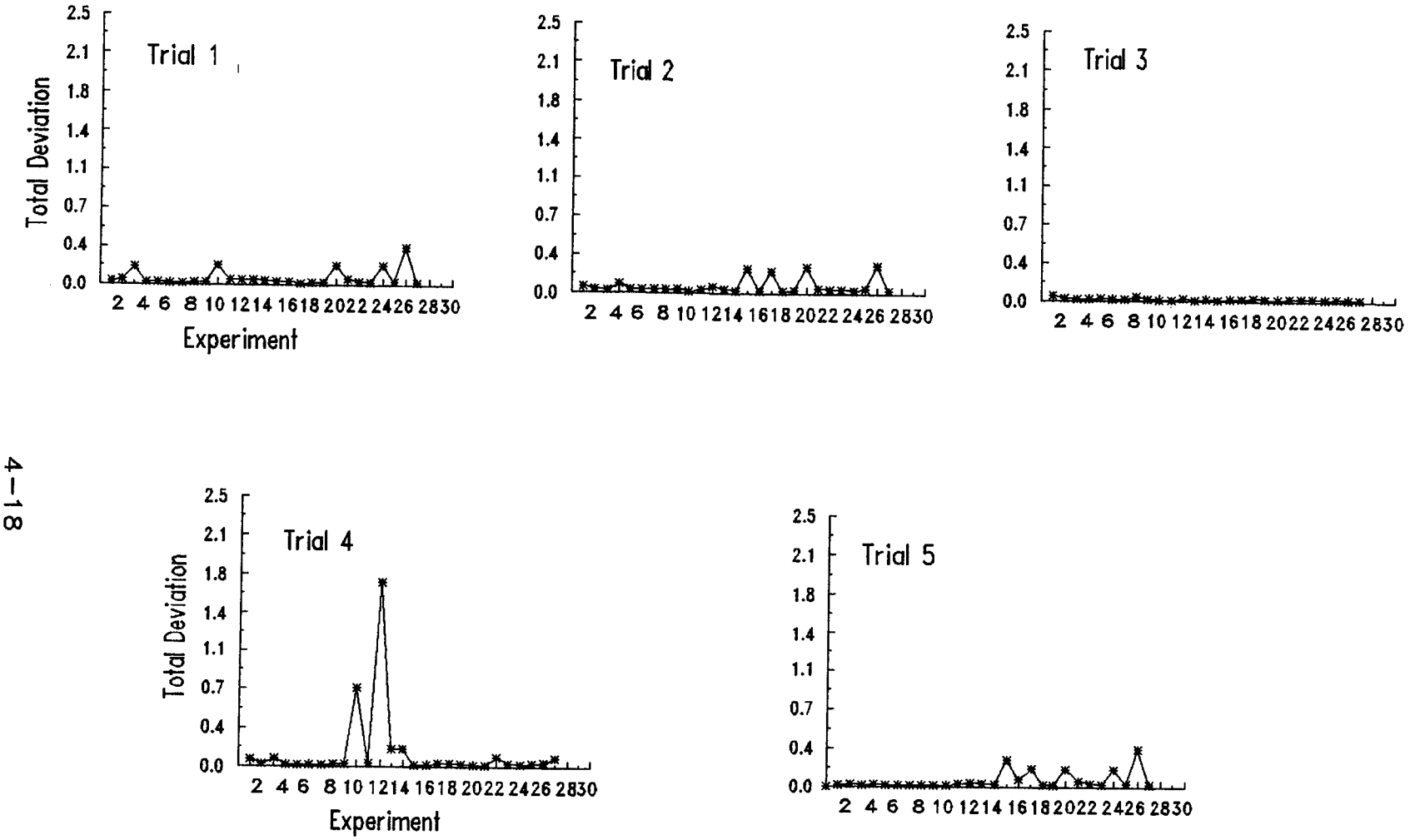


Figure 6
Study 4: Self-Organizing Networks
Confirmations

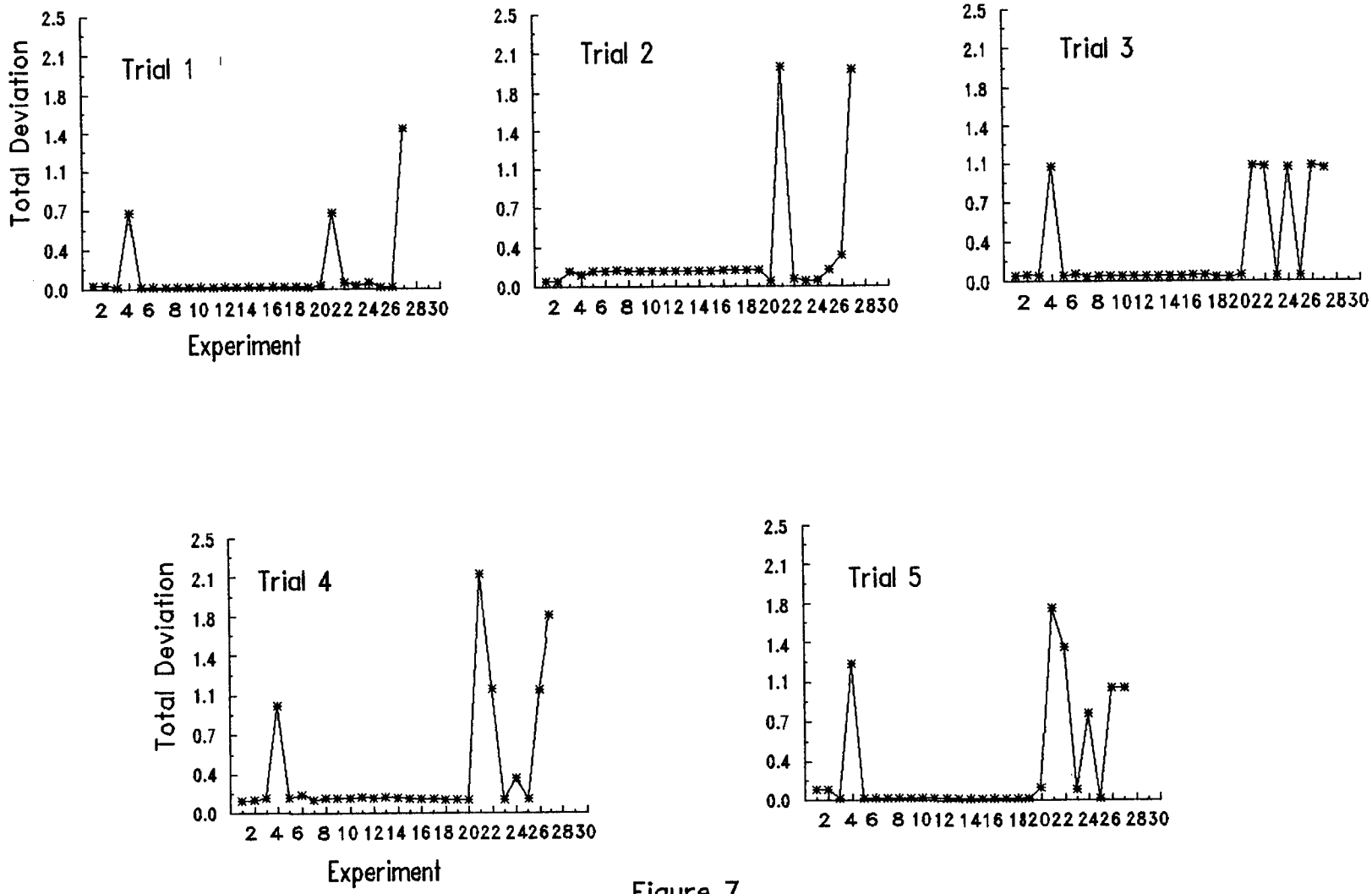


Figure 7
Study 4: Self-Organizing Networks
Null Results

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

John E. Cisneros

Graduate Student

Department of Psychology

University of California, Riverside

Riverside, CA 92521

&

David R. Perrott

California State University at Los Angeles

5151 State University Drive

Los Angeles, CA 90032

Final Report for:

Graduate Summer Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1994

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

John E. Cisneros

Graduate Student

Department of Psychology

University of California at Riverside

Abstract

The present report describes the first two experiments of an extensive series that are currently being conducted on the application of spatial information derived from auditory signals upon visual processing: more specifically, this research investigates the impact of acoustic information upon a human subject's ability to locate and identify visual targets (maintenance of situational awareness). The results of the first experiment confirmed earlier reports (Perrott, Saberi, Brown and Strybel, 1990) that aurally directed visual search was substantially more efficient than unaided search even when the field to be scanned extended a full 360 degrees in azimuth and nearly a full 180 degrees in elevation. This baseline experiment was repeated with audio signals presented over earphones (a 3-D synthesized sound field). Performance in the latter situation was essentially identical to that encountered in the free field (i.e., natural environment), especially for visual targets initially located in the frontal hemi-field. These results indicate:(1) that free field listening environments can be generated in obviously non-free field situations (such as a cockpit of an airplane) with little loss in the utility of the derived spatial input and (2) that such information can substantially improve the human subject's ability to process visual information.

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

John E. Cisneros

INTRODUCTION

A. Overview

In the design of aircraft cockpits, as is also true of most modern human work stations, it is fair to say that most of the information provided to the operator comes via the visual modality. And there are a number of excellent reasons in support of such a strategy. As Neisser (1967) pointed out, the visual channel can be thought of as a parallel processor without peer (at least with regards to other human sensory systems). Whether there is but one item present against a homogeneous background that forms the "field of view" or literally thousands of visual elements randomly arrayed across the same space, at one level the viewer can be said to have an awareness of all the items at the same time. But any attempt to extract more than the most rudimentary information from the array, say the shape of the red figures in the field, and it becomes immediately obvious that the visual system performs as a limited capacity processor restricted to but a small portion of the field of view. In effect, though an operator might see 20 illuminated dials displayed in front of him at the same time (parallel capacity), extraction of information from each dial can only be performed sequentially (serial capacity).

Technological advances over the last several decades have made it possible to greatly increase the quantity (and quality) of information that can be made available to the pilot. But while more information was expected to markedly increase the operator's situational awareness and therefore performance, such expectations have not always been fully satisfied. The problem, of course, stems from the fact that processing of most visual information requires the pilot to attend to one small region of the viewing space until the data is extracted. Thus more information, no matter how useful, requires more processing time if each potential information source is to be utilized. It is probably no coincidence that the abrupt increase in

information available to human operators parallels quite well with the rapid and sustained growth in research dealing with what has been termed visual search, (see Wolf, 1994 for a review of this literature), a paradigm in which the subject is asked to find one item (a target) among a set of other items (distracters). Of course, in the context of a pilot operating an aircraft, the notion that one of many data sources should be examined (i.e., is a target) may not be evident without inspection.

By the 1980's, the emphasis had shifted from providing the pilot more information to helping him or her to cope with the information already available. The idea that there in fact may be too much information available under some conditions via the visual channel (visual overload) prompted considerable attention to how to deal with this "new" problem. A number of approaches were tried with varying degrees of success, a reconsideration of how the visual information is displayed, for example (e.g., heads-up-display and helmet mounted displays).

One approach to the problem of "information overload" that has been gaining increased attention begins with the counter-intuitive notion that the operator of a complex system might optimize the utilization of the existing array of information with the addition of yet more information. Numerous researchers recognized that the bottleneck suffered by the system operator was not too much information but rather too little information as to which information source was most relevant at any given moment. The potential of providing the missing information via the auditory channel was quickly identified (.e.g., Doll, et. al., 1986).

By the mid-1980's the Human Engineering Division of the Air Force Aerospace Medical Research Laboratory at Wright Patterson Air Force Base began a program of research to develop a system for the synthesis of a "free-field" listening environment (McKinley, 1988, Ericson and McKinley, 1989 and McKinley, Ericson and D'Angelo, 1994). The plan was to provide auditory spatial information to human operators under conditions that were not normally conducive to such an attempt (e.g., the cockpit of an

aircraft) by recreating the information normally available to the listener in a free-field via earphones. It was felt that the successful simulation of a 3-D auditory array could improve operator performance by increasing the operator's "operational awareness". Exactly how much advantage could be attained by providing auditory spatial information was unknown when this program was initiated.

During the same period, the Psychoacoustic Laboratory at California State University, Los Angeles also started a long term project directed at examining what use human subjects could make of spatial information derived from the auditory modality. For example, while there was an extensive literature that existed regarding the ability of humans to utilize spatial information from the visual modality to direct behavior (i.e., ear-hand coordination, Woodworth & Schlosberg, 1954), no systematic attempt had been made with spatial information derived from the auditory modality (i.e., ear-hand coordination). Indeed, what work had been reported tended to focus upon the inverse issue, that is whether movements by listeners would alter auditory localization performance.

It eventually became obvious to the participants in these laboratories that the two independent research programs were at least complimentary in nature. For example, some of the early findings obtained in the free-field (e.g., Perrott, 1988a and 1988b) suggested that a substantial improvement in visual information processing could be achieved if spatial input from the auditory modality were made available but, outside of the laboratory, the listening conditions in most "modern" environments were seldom conducive to localizing sounds. It seemed clear that the 3-D sound system that the Air Force was developing could greatly expand the number of situations and tasks that could benefit from the application of this data. Similarly, while the Air Force research effort (now the Bioacoustic Branch of the Armstrong Aerospace Medical Research Laboratory) had made considerable progress in the development of their 3-D auditory system, a substantial program of research was still needed to determine just how effective their system was in the "simulation" of free-field localization cues and, equally important, what advantage could be achieved if such information was made available to the human operator.

The experiments that are reported in this paper represent the initial attempt to combine the efforts

of these two laboratories toward a common goal: An assessment of the impact of auditory spatial information upon the information processing capacity of a human subject. In the first experiment, the effects of the presence or absence of auditory spatial information upon a visual search task was examined in an anechoic environment (i.e., under ideal free-field listening conditions). This would be the referent conditions since almost all naturally occurring auditory spatial cues would be available to the subjects in this situation. The second experiment, using the same subjects, was identical to the first except that the auditory spatial information was delivered through earphones using the 3-D sound system developed by the Air Force. The rational for the particular experimental paradigm employed in both experiments is developed in the following sections.

B. Auditory Psychomotor Coordination

As noted above, the Psychoacoustic Laboratory at California State University, Los Angeles began a long term program of research concerned with the utilization of the spatial information acquired via the auditory modality. In the initial series of experiments (see Perrott, Ambarsoom and Tucker, 1987), the focus was upon auditory psychomotor coordination or more specifically, the ability of subjects to regulate spatially organized behavior based upon spatial information from the auditory modality. The first case that was considered, having a subject turn his head so as to "face" a sound source, was stimulated by the fact that such a shift in orientation in response to the onset of a sound had been frequently observed in man and animals (e.g., Pavlov, 1927; and Sokolov, 1967). As by way of example, in a seminal series of publications, Konishi and his co-workers measured localization performance in the barn owl, an organism that utilizes sounds in its nightly search for prey. Observation of the head position, in this specie, proved to be both a reliable and sensitive measure of localization capacity.

However, the attempt to extend this approach to human subjects (Perrott, et. al., 1987) was not particularly successful. Human subjects pointed their nose at sound sources with considerably less accuracy than they could discriminate the position of a sound source using a non-motor response. It

eventually became evident why college sophomores do not behave like owls. The barn owl must turn its face toward a sound source because the eyes are essentially immobile. College sophomores, on the other hand, are not so constrained. An analysis of video tapes obtained while the subjects were localizing a hidden sound source revealed that concurrent movements of the head and eyes were typical when a subject was asked to "face" a sound source.

Eventually it became clear that, in spite of the instructions, "straight ahead" for these subjects was, within broad limits, defined by where their eyes were directed. While it was clear that one should be able to train a subject to point his nose at a sound source, the fact that shifts in gaze toward a source seemed far more natural and, upon reflection, a far more useful response for a human to make, forced a reconsideration of our general approach. Prompted by these observation, a series of experiments were initiated to evaluate just how well subjects could point their eyes at sound sources.

C. Visual Search

While the visual modality has exceptional capacity to resolve the distribution of light, the fine resolution required to correctly identify, say the letter "A" on this page, extends, at most, only a few degrees from the line of gaze (the image must fall on or near the fovea). Reading or any other activity that requires a relatively high degree of acuity requires that the subject frequently make adjustments in the position of her eyes so that the energy to be evaluated falls in the central visual field. The eye (saccadic eye movements) and head movements that are encountered when human subjects attempt to fixate upon a new visual target can be quite fast. Indeed, shifts in the line of gaze in excess of 700 degrees per second may be encountered as the subject moves to a new fixation point. In effect, little time is required to change from one fixation point to another in the immediate field, though the time required to initiate a head and/or eye saccade is an entirely different matter (latencies on the order of several hundred ms are commonly encountered, see Perrott, et. al., 1990).

In a series of papers (Perrott, et. el., 1987; Perrott, 1988b; and Perrott and Saberi, 1988) the

following speculation was offered: To understand modern human auditory localization performance, which appears to be relatively good compared to some species that have been studied, it might be useful to consider this capacity from an evolutionary or ecological perspective. In humans, the forward placement of the eyes is essential for an extensive binocular field of view (and excellent depth perception) however, the resulting binocular capacity in this relatively large headed animal came at some potential expense: At all times, more than half of the immediate environment is out of the field of view. Modern humans, along with other species that have to observe the world form a "narrow window", probably spend an inordinate amount of time just scanning the immediate environment. Of course, under natural conditions (unlike the modern environment), our ancestors had to be concerned with predators and potential prey as they moved around the environment. The absence of visual capacity in the rear hemi-field is a significant limitation. It was in this context that we argued that the ability to resolve the location of a sound source has a critical role, at least for humans.

While most objects in the world are normally quiet, those that emit sounds are frequently significant. For a specie with a restricted view, a re-orientation of the gaze toward the source of sound would make sense. Indeed, both humans and dogs, two species that have similar visual constraints, the orientation toward the source of a sound is so common as to have been labeled a "reflex" (see Sokolov, 1967 for a discussion of the orientation reflex). When considered in this context, the auditory spatial channel could be said to have a significant role in the determining of <u>what</u> information the visual spatial channel processes.

The capacity of auditory spatial channel to "control" the visual modality is particularly evident if the sound is intense or unexpected or just novel. Such events seem to demand "attention" and can readily disrupt otherwise significant ongoing behavior. Thus there were numerous reasons to expect that spatial information from the auditory channel, when used to "point" the visual channel to relevant events should be successful.

The results from several experiments concerned with the impact of spatially correlated sounds that

simply indicated "where" visual information could be obtained (Perrott, 1988a; Perrott, 1988b; Perrott, et. al., 1990; and Perrott, et. al., 1991) clearly supported this expectation. Two functions were identified and will be discussed here. This first involves the ability of a subject to discern that a change (any change) has taken place in the environment. Aside from the obvious advantage that auditory events can be detected regardless of the relative orientation of the subject at the moment the event occurs, it is also well known that with similar stimulus levels, simple reaction times to sounds are faster than those obtained to lights. A latency advantage on the order of 20-40 ms (e.g. Woodworth and Schlosberg. 1954), a relatively small effect, has been reported for auditory stimuli when compared to the latencies obtained for visual events located in the central visual field. However, when more peripheral locations (out to 80 degrees from the fovea) are considered, the auditory advantage can expand to a hundred ms or more when the subjects are only required to report that an event has occurred (Sadralodabai, Cisneros and Perrott. 1994).

The second function is concerned with the accuracy with which shifts in "gaze" can be directed toward visual targets. As noted earlier, such shifts can be completed very quickly. But, due to the long latency required to organize such a movement, errors in the movement (i.e., the shift fails to bring the target on to the fovea) are particularly costly since an additional interval is required to organize the additional movement (an intersaccade latency). Under conditions in which a single visual event is presented within the visual field, the reduction in the time to localize and identify which of two targets is present can be a great as 500-700 ms if a sound source is present from the same location as the visual target (the sound contains no information as to which target is present, only where it is located). Such large effects were evident for targets initially located within the visual field, between 60-80 degrees from the fovea (Perrott, et. a;., 1990). It was the magnitude of this latter effect that led to the hypothesis that localization accuracy was substantially enhanced by the presence of spatial information from the auditory modality. More recent research has tended to confirm this explanation. In the periphery, in excess of 25 degrees from the fovea, auditory localization performance is moderately superior to that observed in the visual modality (Perrott, et. al., 1993) and the best localization performance in the periphery is observed

when both auditory and visual spatial information are available concurrently (Perrott, 1993).

All of the previous research that has been performed involving visual search in the presence of spatial information from the auditory modality has been confined largely to the frontal hemi-field (within 120 degrees of the initial line of gaze) and generally to only limited variations in the possible elevation of the visual target relative to the original fixation point maintained by the subject. In the first experiment, the time required to localize and identify which of two visual targets was presented on each trial, was determined for events broadly distributed across both the front and rear hemi-fields.

I. EXPERIMENT 1. LOCALIZATION AND IDENTIFICATION OF VISUAL TARGETS: EFFECTS OF AUDITORY SPATIAL CUEING IN THE FREE FIELD

A. Methods

Five subjects, ages 20-25, participated in all aspects of the experimental program. Four of the five were drawn from the experimental subject pool maintained by the Armstrong Laboratory and the fifth was on of the authors (J.C.). None of the subjects reported any history of visual or auditory abnormality.

The tests were conducted with the subject's head located at the center of a spherical array of loudspeakers (radius of 7 feet). while a total of 272 loudspeakers would be required to provide sources spaced evenly at approximately 15 degree intervals across the entire space, the 10 locations directly below the subject were not used in these experiments. The speakers, in turn, were mounted on the inside of a geodesic sphere, the surface of which was covered with acoustic foam to reduce reflections. And finally, the whole apparatus in which the subject was seated was mounted inside a large anechoic chamber.

Mounted on the front of each of the 262 loudspeakers was a four element array of L.E.D.'s distributed to form a diamond. During testing, either the center two elements of the array from one speaker would be activated on a given trial to form a vertical "line" or the lateral two L.E.D.'s would be active forming a horizontal "line". The relatively small size of the figures that were generated by this arrangement was done to ensure that the subjects would have to employ the central visual field to

discriminate which was present. Similarly, the use of two active L.E.D.'s in both configurations avoided any systematic difference in brightness that might allow the subjects to identify whether the vertical or horizontal configuration was present without having to bring the target into the central visual field. The L.E.D. displays produced a moderate amount of illumination (16 foot-candles) that was readily apparent in the otherwise dark chamber. Under the circumstances employed, the visual targets, when activated, provided a high contrast source that could be readily detected.

Under all conditions tested, the primary task was a two-alternative, forced choice paradigm in which the subject had to indicate, via push button, which of two visual targets was present on that trial (a vertical or horizontal "line"). Similarly, all tests were conducted with a high degree of spatial uncertainty as to "where" the next event would occur. Within a session, all 262 locations from which a visual target might be generated had an equal likelihood of being selected on a given trial. Since the visual target could be readily identified once the image had been brought within the central visual field (few identification errors were expected), the primary problem faced by the subject was "finding" the target and making the appropriate shift in the line of gaze.

Two experimental conditions were evaluated across successive blocks of trials (five sessions per subject per condition). In the Spatially Uncorrelated Sound Condition, all trials began with the onset of a broad band noise presented from a speaker at a fixed location. The sound was used to indicate that a visual target was now active and that the subject should begin his search. The sound did not provide any information regarding "where" the target was located. In the second condition, Spatially Correlated Sound Condition, the same general configuration was employed except that sound now came from the same location as the visual target (one of 262 speakers-target locations). All other aspects of the experimental sessions were identical.

The subject was instructed to remain seated and face forward (zero degrees azimuth and zero degrees elevation was defined as that point directly in front of the subject's nose) during the three second inter-trial interval (i.e., the head was approximately centered in the speaker array since we did not wish to

restrict or alter the subject's natural movements of head and/or body with a bite bar or any other such device) until the auditory cue was sounded. At the onset of the acoustic signal, they were instructed to locate and identify as quickly as possible whether the L.E.D. array was in the horizontal or vertical configuration. In actuality, they were encouraged to move their eyes, head and even torso in whatever manner seemed natural while they searched for the visual target. And finally, having located the target, they were to indicate which array was present on the trial by pushing one of two hand help buttons. Their response was used to terminate both the visual target and the acoustic cue and begin a new inter-trial interval. They were then required to return to their initial position (facing forward) to await the next trial. Targets were presented from all locations within each session (262 trials) using a randomization without a replacement technique. Both the elapsed time between the onset of the visual target and the subject's response (reaction time) and whether or not his response was correct was recorded for each of the 262 target locations. Each subject completed all 5 replications of a given experimental condition before continuing on with the next task (fixed-block design), though the order that the subjects performed each experimental condition was randomized.

B. Results

The latencies obtained when the subjects were required to locate and identify which of two visual targets were present as a function of the location of the target relative to the subject's initial line of gaze, without benefit of spatial information from the auditory modality were as follows: For most of the frontal field, latencies fell within the range of 1000-1500 ms. A rapid increase in latencies was encountered beyond roughly 80 degrees azimuth and, to a lesser extent, above or below 50 degrees elevation (an increase of approximately 1500 ms). These results were not unexpected since the targets were now at or just beyond the limits of the subject's field of view. For visual targets located in the rear hemi-field, latencies in excess of 3000-5000 ms were common (3-4 times those observed for targets in the central visual field).

The advantage obtained when the locus of the visual target is marked by a sound source at the

same location was clearly evident. For most targets located in the frontal hemi-field, latencies are generally less than 1000 ms. But probably the most remarkable effect is evident in the rear hemi-field. Latencies were consistently below 1500 ms for almost all locations examined. In effect, in the presence of spatial information from the auditory system, subjects can locate and identify visual targets behind them with the same efficiency as they could locate "uncued" targets in the front.

If one only considers the relative distance of the target from the initial fixation point in terms of azimuth, most of the advantage created by informing the subject with a spatially correlated sound regarding where to find the visual target is evident in the rear hemi-field. the reduction in search time is in excess of 2500 ms for most of the latter target locations.

In contrast, if one only considers the relative elevation of the target, the largest reductions in search time occur with events located 50 degrees below the initial line of gaze (a saving of 2000 ms or more). But a reduction on the order of 1000 ms, a substantial improvement, was clearly apparent across the remaining elevations. There seems little question that the subjects were able to utilize the auditory spatial information to resolve the relative elevation of the target.

II. EXPERIMENT 2. LOCALIZATION AND IDENTIFICATION OF VISUAL TARGETS: EFFECTS OF AUDITORY SPATIAL CUEING WITH A 3-D VIRTUAL SOUND SYSTEM

A. Methods

All aspects of this experiment are the same as the proceeding experiment except that the spatially correlated sound employed to direct the subject to the visual target was generated by the Air Force version of a 3-D virtual sound display yoked to a head-tracking device. An extensive technical discussion of this system can be found in McKinley (1988).

B. Results

A summary of the performance obtained when the localization of the targets were cued using the virtual sound system is as follows: In general, performance is very similar to that encountered when actual

sound sources were used for targets located within 130 degrees of the initial line of gaze. For most of the frontal hemi-field, reaction times are less than 1000 ms and for targets located just beyond the limits of the initial visual field (80-130 degrees) the range increases to 1500 ms. At greater azimuths, the virtual sound is both less effective than the non-simulated and more effective than the spatially uncorrelated sound condition.

The reduction in the latencies generated by the virtual sound system relative to the uncued condition is remarkably similar to that described earlier when the sounds when the sounds were presented in the free field. Greatest improvement is evident in the rear hemi-field (azimuths greater than 90 degrees) and at the lower elevations (below 50 degrees).

Some "cost" was encountered by our subjects when the auditory spatial cue was delivered by the 3-D virtual sound display (re. natural free field listening conditions). In general, latencies are longer when the cues are presented over earphones however, the performance is exceptionally good (i.e.similar to that of "real" sources) within 90 degrees of the initial fixation point. In terms of azimuth, localizing targets in the rear hemi-field required an additional 500 ms to be accomplished. And in terms of variations in elevation, the latencies are somewhat longer (several hundred ms) when auditory cues are delivered via headphones.

III. Discussion

The argument that we would like to make is that the spatial channel of the auditory system evolved, in humans at least, to serve the ocular motor system responsible for shifts in gaze. While the results of the first experiment are in complete agreement with this proposition, we recognize that this "evolutionary" hypothesis is not directly testable. What we can say is that the localization and identification of visual targets can be completed far more quickly when spatial information from the auditory modality is provided. The improvement obtained for events in the frontal hemi-field replicate out earlier observations (e.g., Perrott, et. al., 1990) and the impact of this information for events in the rear hemi-field seems to provide a reasonable extension of this earlier research.

We also believe that the improvement in visual target acquisition obtained here represents a minimum description of the advantage available. First, as noted earlier, signaling the subject to begin the search for a target using a spatially uncorrelated auditory event (the control condition used here) is more effective than if the visual signal had been used. indeed, without the spatially uncorrelated sound , latencies would have been considerably longer in the peripheral regions of the frontal hemi-field (Sadralodabai, Cisneros and Perrott, 1994) and, of course, the task would have been nearly unmanageable for the subjects for events outside of the initial field of view. And second, all tests were conducted under low illumination and without the presence of "visual distracters". Both of the latter aspects would make the task of detecting and localizing the visual target using only spatial information from the visual channel considerably easier. In the low ambient light available, the visual target array stood out as a singular , well defined figure and not merely one item from an array of potential visual targets. As has been demonstrated in previous research (Perrott, et. al., 1991), the advantage created by providing spatially correlated sounds increases substantially as a direct function of the number of alternative visual figures (distracters) present in the field. In summary, we believe that the current test provides a conservative estimate of the value of auditory localization cues in directing gaze.

The results from the second experiment are partially encouraging. Before we attempt to identify the limitations of the Air Force's system used to "simulate" free field auditory spatial information, let us start with a description of the system's successes. First, when compared to the control condition (spatially uncorrelated sound), the simulation of the auditory spatial information markedly improved visual search performance, sometimes by several seconds, regardless of the initial location of the visual target. And second, even when compared to spatially correlated sounds in the free field, for events in the forward hemi-field (extending approximately 80 degrees laterally in both directions and 40 degrees vertically) search latencies were essentially the same in these two conditions. Of course, this is the region of greatest concern. Whether one provides information using a heads-up display (HUD) or the more common instrument panel and cockpit windows, most of the critical signals will tend to be located in the frontal

hemi-field.

As noted earlier, the "simulation" was systematically less effective for targets located more than 110 degrees from the initial fixation point. Part of the explanation may lie in the rate at which the system can update information to the subject. With the large movements required to orient toward a source say at 180 degrees azimuth (head and torso), "average" velocities in excess of 300 degrees per second did occur and "peak" velocities well above even this rate would be common. Thus, unlike the condition encountered with real sources, during particularly rapid movements, incorrect information regarding the "current" locus of the target would be given to the subject.

The performance obtained as a function of the relative elevation of the visual targets was also degraded relative to that obtained in the free field listening condition. The fact that individual head-related transfer functions were not employed would seem to be a reasonable explanation for this "failure".

In conclusion, the application of spatial information from the auditory modality does generate significant advantage for human subjects attempting to locate and identify visual targets, for all regions of the subject's immediate field. Moreover, the simulation of a 3-D auditory display seems to be a practical method by which such information can be made available to human operators regardless of the characteristics of the environment in which they are located. And more specifically, the 3-D auditory space simulation system developed at the Armstrong Laboratory over the last decade can readily be used in that capacity with significant benefit.

REFERENCES

Doll, T., Gerth, J., Engelman, W. and Folks, D. (1986) Development of simulated directional audio for cockpit applications (USAF Report AAMRL-TR-86-014). Wright Patterson Air Force Base, OH: Armstrong Aerospace Medical Laboratory.

Ericson, M. and McKinley, R. (1989) Auditory localization cue synthesis and human performance. Proceedings of the National Aerospace and Electronics Conference. pgs. 718-725.

McKinley, R., Ericson, M. and D'Angelo, W. (1994) 3-dimensional auditory displays: Development, applications and performance. Aviation, Space, and Environmental Medicine, 65, A31-A38.

McKinley, R. (1988) Concept and design of an auditory localization cue synthesizer. [Thesis] Air Force Institute of Technology; AFIT/GE/ENG/88D-29.

Neisser, I. (1967) Cognitive Psychology. New York: Appleton, Century, Crofts.

Pavlov, I. (1927) Conditioned Reflexes. (G.V. Anrep, Trans.). New York: Oxford University Press.

Perrott, D. (1988a) Auditory psychomotor coordination: Auditory spatial information can facilitate the localization of visual targets. Proceedings of the Localization Symposium. (CHABA), Washington D.C.

Perrott, D. (1988b) Auditory psychomotor coordination. Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 81-85). Santa Monica, CA: Human Factors Society.

Perrott, D., Ambarsoom, H. and Tucker, J. (1987) Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions. J. Acoustical Society of America, 82, 1637-1645.

Perrott, D. and Saberi, K. (1990). Functional integration of auditory and visual sensory processes. Paper presented at the annual meeting of the American Psychological Society, New Orleans.

Perrott, D., Saberi, K., Brown, K. and Strybel, T. (1990) Auditory psychomotor coordination and visual search behavior. Perception and Psychophysics, 48, 214-226.

Perrott, D., Sadralodabai, t., Saberi, K. and Strybel, T. (1991) Aurally aided search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors, 33, 389-400.

Perrott, D. (1993) Two modalities and one world. Proceeding of the International Meeting of the Audio Engineering Society. Copenhagen, Denmark.

Perrott, D., Costantino, B. and Cisneros, J. (1993) Auditory and visual localization performance in a sequential discrimination task. J. Acoustical Society of America, 93, 2134-2138.

Sadralodabai, T. and Perrott D. (1990) Effects of spatially correlated auditory information on eye movements. Paper presented at the Fall meeting of the Acoustical Society of America.

Sokolov, Y. (1967) Perception and the Conditioned Reflex (S.W/. Wayenfeld, trans.) New York: Pergamon.

Wolf, J. (1994) Guided search 2.0: A revised model of visual search. Psychonomic Bulletin and Review, 1, 202-238.

Woodworth, R. and Schlosberg, H. (1954) Experimental Psychology. New York: Holt, Rhinehart and Winston.

INTRA-OCULAR LASER SURGICAL PROBE (ILSP) FOR VITREOUS MICRO-SURGERY

Candace E. Clary
Graduate Researcher
Department of Biomedical Engineering

University of Alabama at Birmingham
256 Business-Engineering Complex
1150 Tenth Avenue South
Birmingham, AL 35294-4461

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

INTRA-OCULAR LASER SURGICAL PROBE (ILSP) FOR VITREOUS MICRO-SURGERY

Candace E. Clary
Graduate Researcher
Department of Biomedical Engineering
University of Alabama at Birmingham

Abstract

The surgical treatment of many vitreoretinal diseases, involves removal of membranes or vitreous strands overlying the surface of the retina. Instruments currently utilized in this procedure have disadvantages inherent in their mechanical nature. A model laser surgical probe has been designed and built (patent pending) by Lt. Daniel X. Hammer (AL/OEO, Brooks AFB, TX) and Cynthia A. Toth, M.D. (Department of Ophthalmology, Duke University, Durham, NC) to perform cutting of fibrovascular membranes within the vitreous cavity of the eye using laser induced breakdown (LIB).[1] *LIB is a process by which atoms are ionized and a plasma of quasi-free electrons and ions is created. The probe consists of a multimode optical fiber, for maneuverability and light delivery, with a gradient index (GRIN) lens attached for micro-focusing. The problematic areas in the development of the probe were examined. These include delivery of high-energy Q-switched 5 ns pulses of 1064 nm Nd:YAG laser through an optical fiber and micro-focusing light energy from fiber to achieve LIB.*

INTRA-OCULAR LASER SURGICAL PROBE (ILSP) FOR VITREOUS MICRO-SURGERY

Candace E. Clary

I. Introduction

The treatment of many proliferative vitreoretinal diseases requires cutting of membrane strands overlying the retina. Proliferative vitreoretinal diseases can cause ocular vessels to shrink and close. This elimination of a tissue's nutritive source has two effects. It fosters the growth of abnormal new vessels and scar tissue, and causes some of the tissue served previously by closed vessels to die. New vessels and scar tissue form and grow along the surface of the retina and attach to the back surface of the vitreous gel. The gel pulls on the attached vessels and scar tissue, which, in turn, pulls on, and lifts up the retina. The treatment for this condition involves reattaching the retina and removing the abnormal vessels and scar tissue from its surface. Removing the vitreous and the scar tissue is a delicate process which requires a surgeon to lift and peel strands away from the retina. In severe cases, the procedure may take several hours.

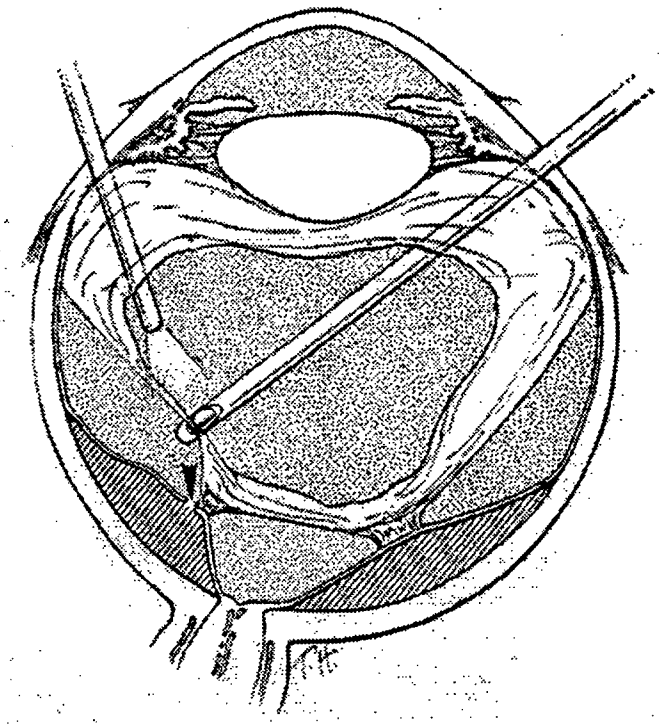


Figure 1. Retinal break due to traction caused by mechanical vitreous suction cutter during vitreoretinal surgery.[2]

Portions of vitreous micro-surgery which involve cutting are currently performed with mechanical instruments. These mechanical instruments cause difficulties for two important reasons. First, they can cause unnecessary damage to the tissue being cut, through shearing forces to tissue collaterally surrounding the instrument. Further damage can also occur through traction forces imposed on

tissue connected at some distance away from the instrument, as indicated by the arrow in Figure 1. Second, mechanical instruments must surround the tissue in order for cutting to be performed. The ILSP is designed to replace some of these mechanical devices, to eliminate both of these problems, and to provide a cleaner, sharper cut.

Lasers are already utilized in ocular and vitreous surgeries. However, they are generally used as heat sources for procedures, such as spot-welding, where the dominant physical mechanism is thermal. The proposed instrument, discussed in this paper, uses laser-induced breakdown (LIB) to cut diseased tissue. LIB is the ionization of molecules to form a plasma spark due to high laser irradiance. The mechanisms for cutting are the ionization of the material, as well as mechanical, from a shock wave produced during the breakdown event, and thermal, from effects caused by the flash.

The laser light would be delivered into the eye through an optical fiber, which would be inserted into the eye in the traditional manner (Figure 2), and microfocused with a gradient index (GRIN) lens. A GRIN probe for LIB that does not utilize an optical fiber has already been built and tested at the Armstrong Laboratory at Brooks AFB, TX (AL/OEO). This probe uses collimated Nd:YAG pulses of 5 ns focused through the GRIN lens to achieve LIB and is otherwise the same as the proposed

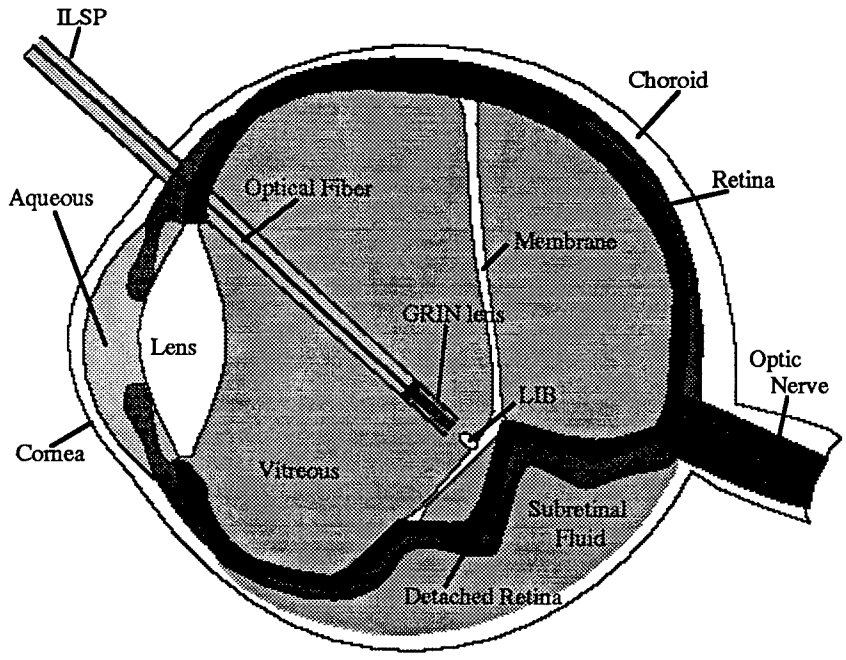


Figure 2. Proposed ILSP instrument used as cutting device using laser-induced breakdown for vitreoretinal surgery.

optical instrument, except that it does not employ an optical fiber. For simplicity, the fiberless version will be referred to as "Probe 1" in this paper. Although Probe 1 was very successful in achieving breakdown, delivery of energy through an optical fiber is necessary for the probe to be functional in surgery.

This paper examines the current state of development of the fiber optic endoprobe, assesses the problematic areas, and investigates possible solutions and direction for future work.

2. Theory

2.1 Laser-Induced Breakdown

One concern with the proposed instrument involves the viability of causing LIB in the eye during surgical procedures. Since the ILSP would be the first application of LIB deep within the posterior portion of the eye, the implications of the secondary effects caused by the bubbles and shock wave produced during the breakdown event are unknown.

Plasma expansion occurs during LIB in a liquid. It is initiated by avalanche ionization and continues as superheating, vaporization, and thermal expansion. Plasma collapse then occurs, during which a spark is produced due to electron recombination and photon emission. In addition, hypersonic shock waves propagate through the media, and a cavitation bubble forms and collapses, often several times, before producing small, visible gas bubbles which stream from the site. The whole of the breakdown event has the potential to cause thermal as well as mechanical damage. The physical characteristics of the breakdown event, the dominant damage mechanism, the energy required to start the cascade, and the energy required to perpetuate it, varies slightly with pulse duration and spot size. They are a few factors which must be accounted for in the probe design.

Based on previously reported nanosecond breakdown thresholds[5], theoretical computer determination of breakdown energy[6], examination of the energy required for breakdown for Probe 1, and spot size-analysis of the fiber probe with a .23 pitch lens[7], the energy predicted to initiate breakdown in tap water with 5 ns pulses for the current version of the probe is $3.0 \pm .5$ mJ.

Currently, the two most critical aspects in the development of the GRIN endoprobe are delivery of high-power laser through the fiber and microfocusing for laser-induced breakdown.

2.2 Fiber Transmission

The complex problems associated with Q-switched Nd:YAG transmission through multimode fibers have been studied for several years by researchers. There are four aspects which seem to play particularly important roles in determining the success of transmission. They are fiber preparation techniques, focusing conditions, connectors, and beam profile.

The best method of fiber face-preparation is still widely disputed. One of the limitations of power delivered into the fiber is an undesired LIB event at the front face of the fiber. The intentions of this probe are to utilize LIB for cutting at the *end* of the probe. Any LIB event before the laser light reaches that point, i..e. at the front face of the fiber, is undesired. Since LIB is impurity-dependent, that is, the probability of breakdown occurring increases with increased impurities, a greater amount of power can be transmitted through a fiber with a smoother surface. These impurities can take the form of both material deposits from moisture and polishing materials as well as voids and inclusions on the surface. Pascal Rol et al. contend that for high power transmission, a properly cleaved surface is better than one that is mechanically polished.[8] In theory, if it was possible to obtain a perfect cleave, the fiber face would then have a finish as fine as the composition of the fiber itself.

Robert Setchell, of Sandia National Laboratories in Albuquerque, NM, has studied the problem of Q-switched Nd:YAG transmission for several years and claims that since cleaved fibers showed a wide range in finishes, a polish is preferred because of consistency.[9] Setchell found that the choice of polishing material is as important as the technique itself. He is partial to a cerium oxide solution because he found that aluminum oxide tends to leave subsurface deposits.

However, Setchell also claims that laser absorption by residual contaminants from mechanical polishing is not the predominant cause of front-face breakdown. His conclusion is based on results obtained in comparing cleaved fibers and mechanically-polished fibers. For his study, the cleaved fibers damaged at lower energies and were found to have a higher front-face hydrogen content. It is possible that front-face damage could be blamed on ambient water rather than on polishing residue and it is reasonable to speculate that cleaved fiber faces are more absorbing of ambient water than are polished fiber faces.

In addition to surface finish, the focusing of the beam and the mechanical alignment also play an important role in front face damage. For fibers less than 600μm in diameter, the beam should be focused to a spot with a diameter approximately 70% of the core's diameter. A spot that is too small risks damaging the face by increasing the irradiance of the spot. A spot that is too large risks damaging the face by heating at the core-cladding interface. In addition, this spot must be formed by a lens system chosen with a focal distance which minimizes losses into the cladding caused by larger entrance angles than the numerical aperture allows. On the other hand, an entrance angle that is too small causes damage as well. The general rule is for the entrance angle to be between 30% and 90% of the fiber's specified numerical aperture.[10] Using achromatic or aplanatic lenses designed to reduce spherical aberrations could help more closely match the proper theoretically-determined lens to a real one.

In transmitting high power laser pulses, the temperature-tolerance of the system must be considered. Conventional epoxy connectors are, therefore, not recommended for this application. A few manufacturers now market "high-power" connectors which use a different mechanism for attachment and allow for an air or sapphire space between the fiber and connector. However, these generally must be factory installed, which increases cost and sacrifices time and availability.

For many laser applications, it is generally accepted that a Gaussian beam profile is the preferred one due to its superior focusing characteristics. That may be true strictly for focusing, but the dominant limiting factor of power transmission for this application is exceeding the required irradiance for a breakdown event to occur. Therefore, the beam profile which minimizes the irradiance at all areas on the fiber face is more advantageous for delivering high power into a fiber. It is generally accepted that the laser beam should be multimode in order to facilitate the highest power transmission.

Other than focusing into the fiber, there is not an advantage to using a laser with a Gaussian spatial distribution. The beam profile of the laser does not affect the focusing ability at the other end of the probe since transmission through a multimode fiber mixes the modes anyway.

Figure 3 shows the general irradiance pattern across the beam's cross-section given a Gaussian

beam profile. For a Gaussian profile, the irradiance at the center is increased drastically, whereas, for a multimode one, the same amount of energy might be more distributed, reducing the irradiance at each area of the cross-section and maximizing power transmitted without fiber damage.

It has been found that coupling high power into fibers via vacuum can reduce the probability of breakdown at the front face of the fiber by removing physical material (impurities) at that site, thereby increasing power transmission .[11]

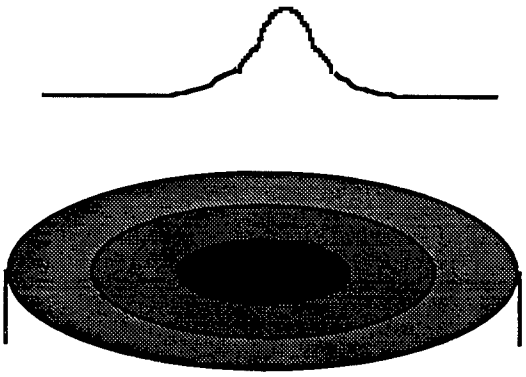


Figure 3. A Gaussian beam profile and its corresponding energy distribution across a circular fiber face. Darker color corresponds to higher radiance. Focusing a Gaussian beam increases the irradiance in the center dramatically.

The next consideration in fiber transmission of high power Nd:YAG is the type of fiber used. For Q-switched 1064 nm Nd:YAG, the preferred fiber is one of a pure fused silica core with a fluorine-doped silica cladding. Although the composition of silica glass is the same, from manufacturer to manufacturer, the fibers are not. Some of the variations were investigated and the results are reported further in the paper.

2.3 Microfocusing

In a conventional optical system, each optical component has a constant index of refraction. The behavior of the system is determined by the combined curvature, thickness, and index of each component. In a gradient index system, however, an element has an index which varies continuously within the material. In general, using gradient index elements in a system contributes to a cost reduction, weight reduction, and increased reliability of the system.[12] Also, the manufacturing process of using dopands to create the continuous gradient index, allows for a GRIN lens to be physically constructed much smaller than a traditional lens with good accuracy.

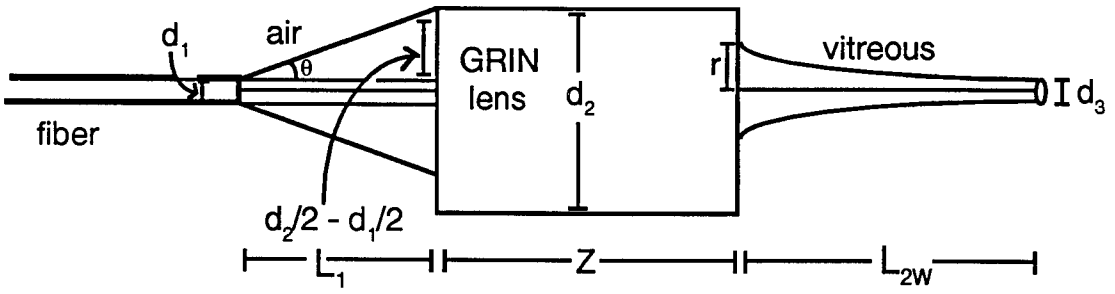
An instrument which uses laser-induced breakdown as a damage mechanism for ocular surgery has previously been constructed by Pascal Rol, et al., of the Institute of Biomedical Engineering and Medical Informatics, the Swiss Federal Institute of Technology, and the University of Zurich all in Zurich, Switzerland, and the University Eye Clinic in Berne, Switzerland.[3] That instrument utilizes a hemispherically-shaped fiber tip to collimate rays and then supplies enough power through the fiber to achieve the required irradiance for the breakdown event. The shaped fiber tip functions in a manner similar to that of a traditional lens in that it uses the curvature (with maximum effect achieved with a hemispherical shape) of the "lens" with a constant refractive index difference ($n \approx 1.46$ for pure silica fibers) to provide the "focusing" effect. However, because the refractive index difference is greatly reduced in water or vitreous fluid ($n \approx 1.33$) versus air ($n \approx 1.0$), the focusing effect of such a lens in vitreous or water is limited.

Instead of using a graded thickness of material of constant index of refraction like a traditional lens, a GRIN lens has graded index of refraction of material of constant thickness. Because of this, a GRIN lens is much more accommodating in mediums other than air than are traditional lenses (and thus, a shaped fiber tip). In addition, though the use of a GRIN lens would introduce some coupling losses not present using a shaped fiber tip, the fact that the GRIN lens causes rays to focus well in water or vitreous makes it an invaluable part of this design. Since the shaped fiber tip is intended only to collimate rays[4], the irradiance at the breakdown event is also the irradiance behind it, on the retina.

The focusing effect of the GRIN lens proves useful in two ways: (1) It allows less energy to be used to achieve the required irradiance due to the smaller spot size, thereby reducing patient exposure; and (2) it reduces the risk of damage to the retina by increasing the spot size projected onto it, due to the highly diverging rays after focusing, thereby decreasing the irradiance.

The gradient index (GRIN) lens in the ophthalmic probe has a refractive index which continuously decreases radially outward from the optical axis. Light travels as a spiral (forming a cylindrical helix) through the lens due to its changes in velocity in materials of higher and lower indices of refraction.[13] The focusing characteristics of the lens are determined by the material properties of the lens as well as its pitch.[14] The pitch of the lens is defined in terms of the number of cycles of the helix will

exist within a lens at a given time. It is often thought of in terms of the two-dimensional projection of the helix, which is a sinusoid. A 1.00 pitch lens allows a whole wavelength of that sinusoid, while a .5 allows half and .25 a quarter of a wavelength. Therefore, in order to achieve the desired focusing effect, a lens must be chosen according to the source.



FIBER SPECS		LENS SPECS			PHYSICAL DIMENSIONS (mm)		
Exit NA	Diameter (μm)	Pitch	Type	Diameter (mm)	Spot Diameter	L1	L2
.22	100	.29	W	1.8	.0461	3.8636	0.5562
			H	1.8	.0356	3.8636	0.2230
		.23	W	1.8	.0599	3.8636	1.7405
			H	1.8	.0440	3.8636	1.0874
.22	200	.29	W	1.8	.0977	3.6364	0.6192
			H	1.8	.0755	3.6364	0.2613
		.23	W	1.8	.1287	3.6364	1.8284
			H	1.8	.0942	3.6364	1.1356
.22	400	.29	W	1.8	.2219	3.1818	0.7701
			H	1.8	.1712	3.1818	0.3543
		.23	W	1.8	.3016	3.1818	2.0460
			H	1.8	.2196	3.1818	1.2548

Figure 4. Physical dimensions of configuration used for ray tracing analysis. d_1 is the core diameter of the fiber, d_2 the diameter of the GRIN lens, and d_3 the diameter of the resulting spot. The front-face working distance is represented by L_1 and the rear-face working distance in water or vitreous by L_{2W}. The lens depth, Z, is a function of the pitch of the lens as well as the material properties of the lens. Tabulated calculations of fiber / lens combinations and their associated spot sizes and front- and rear-face working distances. A numerical aperture of .22 is typical for many 200μm fibers including those of Polymicro Technologies, Inc. and Fiberguide Industries (both tapered and not tapered).

For the GRIN probe previously tested, the collimated source caused the tightest focus to be achieved with a .23 pitch lens. For a fiber source, however, the rays are diverging. A .29 lens would produce the smallest spot size in this case. Figure 4 tabulates the theoretically predicted spot sizes and working distances for lenses of .23 and .29 pitch given a fiber source of a particular diameter and numerical

aperture. These values were determined from mathematical guidelines in the Nippon Sheet Glass (NSG) product guide along with geometrical optics and should be considered only as a first order approximation. A more exact solution could be found using Fourier analysis.

NSG manufactures three standard types of GRIN lenses, S, W, and H, which have different material properties. In general, the values in Figure 4 are for standard NSG lenses. For the purposes of product development, however, it is important to note that these are the standard lenses for .23 pitch, but the only NSG standard lens for a .29 pitch lens is the W-1.8.

The physical dimensions of the microfocusing section of the probe are also shown in Figure 4 in the diagram above the table. The trends illustrated by a graphical comparison of the performance of five standard NSG lenses (S-1.0, S-1.8, S-2.0, W-1.8, W-2.0, and H-1.8) show that for an application similar to the fiber probe, the H lens seems to produce the smallest spot size, followed by the W and then the S. The trends also seem to agree with concepts from Fourier optics in that the larger lenses produce smaller spot sizes. The same trend also explains the necessity for an air space between the fiber and the lens. The air space allows for the full front face of the lens to be filled by the rays emerging from the smaller fiber, maximizing the effect of using a larger lens. Therefore, it is reasonable to assume that an even smaller spot would be produced with a .29 pitch H-2.0 GRIN lens. In addition, NSG manufactures a plano-convex GRIN lens, which is intended to reduce spherical aberration. Although it is marketed for coupling *into* fibers, it would useful for microfocusing because less spherical aberration would reduce the spot size even further.

There are a few additional concerns in optimizing lens types, spot sizes, and conditions for laser-induced breakdown. First, the lenses which produce the smallest spots also tend to have the shortest rear-face working distance. The .29 pitch lenses and the H type lenses, while producing the smaller spot sizes, also have shorter working distances. The issue then becomes the risk of lens damage. If breakdown occurs too close to the rear face of the GRIN lens, the shock wave produced in the event could cause mechanical damage to the lens itself. Since the precise relationship between the size of plasmas and spatial extent of shock waves compared to the power input has yet to be determined, the rear-face working

distance which puts the lens at risk is also unknown. Generally, smaller spots produce smaller plasma volumes, so the shorter focal lengths which produce the smaller spots may ultimately have no effect in terms of damaging the lens. However, it is still a consideration in the optimization of the system and product development.

Second, with nanosecond pulses, even though a certain irradiance causes breakdown for one spot size, the irradiance value required to initiate the event is increased as spots get smaller.[15] Therefore, the expected *dramatic* decrease in energy required for breakdown with smaller spots is lessened slightly in effect. It is expected, however, that the increase in required irradiance is probably not consequential compared to the advantages of obtaining a cleaner cut with less energy due to smaller plasma sizes.

The third consideration with the microfocusing aspect involves the pattern of lens diameter to resulting spot size. It has already been established that the larger lenses produce a smaller spot. However, the specifications for instruments used in ophthalmic surgery generally dictate a 1 mm maximum diameter. This is one other consideration in determining the trade-offs for product development.

As previously discussed, Rol achieved LIB through his probe by shaping the fiber tip. There are four effective methods for shaping fiber tips. They are using a Bunsen micro-burner[16] an electrical arc[17], a micro furnace[18], and a CO_2 or similar laser[19]. The idea of utilizing a built-in collimator via a hemispherically-shaped fiber (which yields the minimum radius of curvature) was investigated as a means to achieve a smaller spot size through a GRIN lens. The finished focusing effects should then be smaller, and tighter, identical to those observed with Probe 1. Two observations emerged from this investigation. First, the concept of a GRIN lens using pitch to determine its focusing effects would, ideally, eliminate the need for collimated rays. The desired effects could be engineered by choosing the right lens. In practice, however, completely customized lenses are not always desired for convenience and economics. Second, although Rol chose to shape his fibers with an electrical arc due to its superior reproducibility over other methods, the shaping of fiber tips themselves is not very reproducible compared to combining a flat-cleaved fiber with a GRIN lens. It is difficult to form a perfect hemisphere and then to repeat it. Therefore, the rays emerging from the shaped fiber are not always collimated, as illustrated in Figure 5, and the degree

of focusing also varies. The gain in focusing from a shaped fiber with a .23 pitch GRIN lens over a flat fiber with a .23 GRIN lens is not significant enough to outweigh the irreproducibility and inconsistency inherent in the shaped fibers themselves.

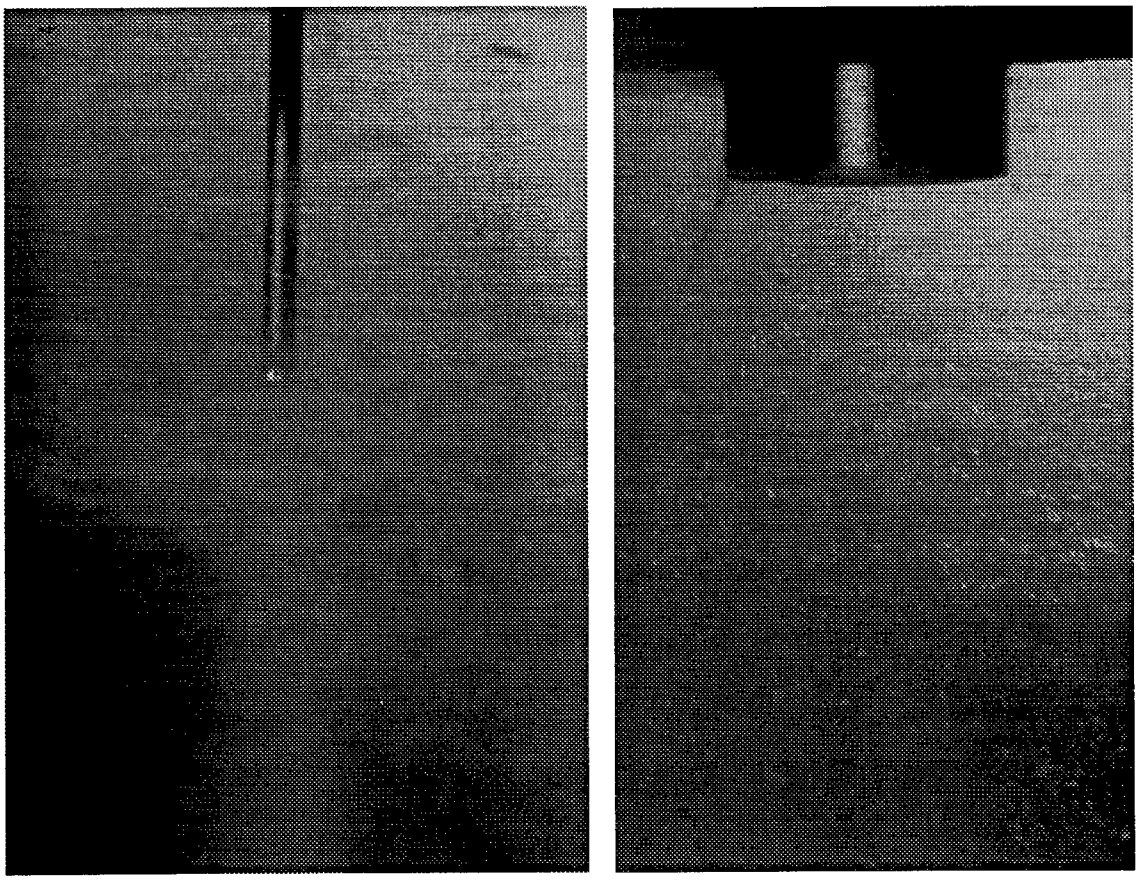


Figure 5. Photographs of focusing effects of He-Ne laser source in water through hemispherically-shaped fiber (l) and GRIN lens with flat-tipped fiber (r). The hemispherical fiber was shaped with a CO_2 laser. The flat fiber was placed approximately 3.6 mm from the front of the W-1.8 GRIN lens. The photo shows the rear of the GRIN lens (top) and rays focusing through it. The He-Ne rays were made visible with a milky water solution.

3. Materials and Methods

The Spectra-Physics Quanta-Ray GCR-3RA Nd:YAG laser used in this experiment produces 5 ns pulses at 1064 nm at variable (single shot to 10 Hz) pulse repetition frequency. The beam profile is approximately 75% Gaussian. The energy output of the laser was controlled

with two half-wave plates and two polarization cubes. The energy detectors used were J4 Molectron detectors output to a Molectron JD2000 Joulemeter Ratiometer. Figure 6 is a diagram of the experimental set-up. The beam-splitter and detector A were used to determine the amount of input energy before the focusing lenses and detector B to determine the transmitted energy, with ratio of the two determining the transmission percentage.

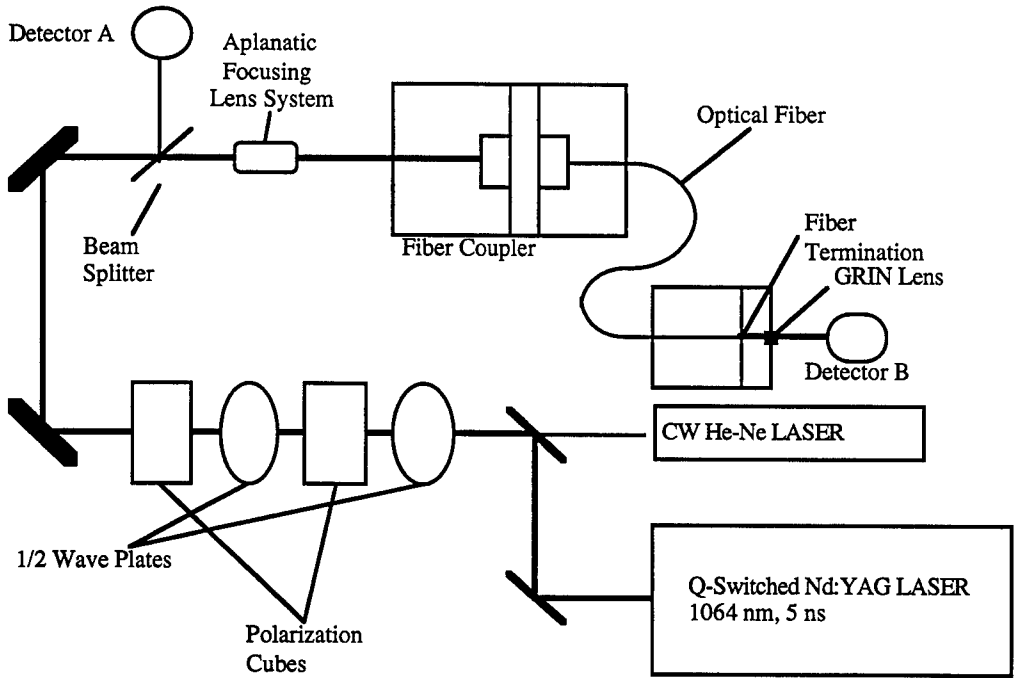


Figure 6. ILSP experimental set-up. Energy input from Q-switched Nd:YAG source controlled by two 1/2 wave plates and two polarization cubes. He-Ne laser used for alignment. Surgical instrument in operating room would consist of a source, composed of the laser source, the focusing lens system, and the fiber coupler. The optical fiber would be enclosed in a protective tubing and would plug into the source unit. The fiber termination point and GRIN lens would be enclosed in a stainless steel casing and the termination of the GRIN lens would be the point of the probe termination.

The focusing lens system used was a CVI 50 mm laser aplanat in series with an aplanatic meniscus lens. The combined focal length is 33 mm. The fibers tested were all of pure fused silica core with a fluorine-doped silica cladding. The fibers were prepared using a mechanical polishing technique with aluminum oxide for the beginning stages and ferric oxide for the final stage. The connectors used were standard SMA epoxy connectors.

4. Results and Observations

Most optical fiber manufacturers report the damage threshold of their fibers in terms of tests performed with CW lasers. Many have not determined a damage threshold for pulsed laser energy. The idea of a threshold for a certain fiber is misleading when referring to pulsed energy. This is due to the many variables which cause unknown and indeterminable increased irradiance in a confined area on the fiber face, leading to damage. With CW lasers, the dominant damage mechanism is thermal, which causes a more consistent energy level to be associated with damage. The damage threshold from pulsed lasers is highly dependent on factors such as the spatial beam profile, fiber preparation techniques, and launching conditions. Therefore, a theoretical damage threshold provided by manufacturers could be unreachable, unapproachable, and generally not useful for practical purposes.

Impurities present within the fiber act as a catalyst for damage. Therefore, though the composition of fused silica fibers may be the same, the manufacturing process could play a large role in the success of the fiber for high power transmission. The process which produces a more pure fiber should allow for more power transmitted through fibers it produces.

A sample of fibers from various manufacturers were tested for maximum energy transmitted for 5 ns pulses. It has been speculated that since the fibers are all of pure fused silica core with identical composition, each of the different fibers of a given size should perform equally given a similar laser beam profile, preparation, and launching conditions. However, because fiber damage caused by laser-induced breakdown is highly impurity dependent, it is expected that, though the formulated make-up of the fibers should remain the same, those of purest composition would allow for higher power transmission. Figure 7 reports the results obtained in testing the fibers. The values given are those for the methods, conditions, set-up, and materials specified previously in this paper. Actual values would vary with changing conditions.

Manufacturer	Product #	Max. Tx Energy
Ceram Optec	UV200/240N	2.0 mJ
Polymicro Technologies	FIP200220240	2.488 mJ (max)
		2.514 mJ (max)
		2.7 mJ (damage??)
		2.0 mJ (max)
Fiberguide	AFS200/220N	1.2 mJ
Fiberguide	AFT200TO100Y-1.1	1.5 mJ
3M	FT-200-UMT	1.13 mJ

Figure 7. Damage energies at which fibers tested with experimental set-up, equipment, and techniques previously described in this paper. Each of these values describes subsurface damage. All of these fibers (except Fiberguide AFT200TO100Y-1.1) were 200μm diameter fused silica core with fluorine-doped silica cladding. Fiberguide AFT200TO100Y-1.1 was tapered from 200μm at the entrance to 100μm at the exit.

The fibers manufactured by Polymicro Technologies, Inc. obviously yielded the best results for the equipment and methods used in this experiment. The "max" in parentheses next to some values indicates that the fiber initially did not damage at that value, but the limitations due to focusing effects caused at the front face of the fiber created a ceiling value for transmission. Eventually, the fiber did damage at that value, but the value given should not be interpreted as the "damage" threshold of the fiber.

Although the Polymicro Technologies fibers seemed to be the most durable among those tested, there is a tremendous advantage to utilizing tapered fibers, as was observed by the results tabulated in Figure 7. The tapered fibers were manufactured by Fiberguide Industries. The Fiberguide AFT200TO100Y-1.1 fiber in the table above is one which has a core 200μm diameter at its entrance face and a core 100μm diameter at its exit face. The advantage of using a tapered fiber for this applications is being able to distribute more energy over the front face and couple more power into the fiber while providing for a smaller exit diameter, higher irradiance, and sharper tissue cut. The amount of power coupled through the *tapered* Fiberguide fiber, although about equal to the amount coupled through the 200μm *straight* Fiberguide fiber, yields an irradiance at the other end of the fiber that is four times greater. Another possibility is to use a tapered 400μm to 100μm fiber for even greater coupling efficiency and higher resulting irradiance at the end of the probe.

During the testing of these fibers, it became apparent that the primary concern may not be the damage threshold of the fibers, as previously thought, but the amount of power *available* for coupling into

the fiber. The reason for this is due to the spatial distribution of the laser beam used. As stated before, the ideal beam profile for high power fiber transmission is multimode. The coupling power gained by reducing the irradiance on the fiber face with a multimode beam outweighs the superior focusing effects for coupling produced by a Gaussian beam. In addition, in a multimode fiber, a Gaussian beam provides no advantage at the focusing end of the probe due to the mixing of modes.

In focusing a Gaussian beam, the irradiance is higher at the center of the focus and lends to a breakdown event, which is visualized as a flash, at the front face of the fiber at a lower energy than would be required for a multimode beam. The flash occurs at the position of the fiber face, not at the center of the focal area, due to the impurity-dependence of LIB. For nanosecond pulses, as the energy input increases, the size of the plasma and its accompanying flash increases with it monotonically.[20] The larger plasma and shock wave increase the damage risk to the fiber and decrease the amount of energy available for fiber transmission. Given identical conditions, from shot to shot, the power transmitted is drastically reduced on flashing shots because a large amount of energy is dissipated in the breakdown event itself. The energy contributing to the breakdown event is no longer available for transmission through the fiber. At low power inputs, as the fiber is backed away from the focal area, the power transmission is reduced significantly. At higher input energies, however, without a uniform beam distribution, backing the fiber away from the focal area can actually increase the power transmission by decreasing the probability of a flash occurring. The amount of energy dissipated in the breakdown event eventually creates a ceiling on the amount of energy available for fiber transmission. This observation emphasizes the need for minimizing the flash which occurs at the front fiber face and, therefore, the importance of using a multimode beam profile.

5. Conclusions and Suggestions for Future Development

An optical cutting instrument for vitreous surgery would be advantageous over some mechanical instruments currently used in the procedure. One which employs laser-induced breakdown by microfocusing laser light from a fiber with a gradient index lens would be advantageous over other optical instruments in terms of safety, versatility, and efficiency.

There are a few possibilities which must be considered for the future development of the product. Currently, the most important consideration is the spatial profile of the beam, since this factor is the primary limit at this time to energy available for transmission. Other possibilities include: using a Fiberguide tapered 400μm to 100μm fiber to optimize power transmission and microfocusing, air- or sapphire-spaced connectors to eliminate use of epoxy, altering mechanical polishing methods by trying different materials, using a vacuum for coupling into the fibers, using H type and .29 pitch GRIN lenses, and using GRIN plano-convex lenses for reduced spherical aberration in microfocusing.

6. Acknowledgements

This work was Sponsored by the Air Force Office of Scientific Research Bolling Air Force Base, DC and Armstrong Laboratory. I would especially like to thank 1Lt Daniel Hammer of AL/OEO, Brooks AFB, TX and Dr. Cynthia Toth, Duke University Eye Center, Durham, NC. Also, thanks to the U.S. Army Medical Research Detachment Laser Lab for the use of their CO_2 laser. In addition, I would like to thank all those at the Optical Radiation Division for their time and resources, including Paul Kennedy, Dennis Maier, John Mileski, Gary Noojin, Pat Roach, Ben Rockwell, and Randy Thompson.

References

1. C.A. Toth, D.X. Hammer, K. Slawinski, and G.D. Noojin, "Gradient Index (GRIN) Lens Multimode Fiber Probe for Laser Induced Breakdown in the Eye," *Proceedings of SPIE*, Vol. 2126, pp. 291-304, 1994.

2. T.A. Rice, C.P. Wilkinson, and R.G. Michaels, *Retinal Detachment*, Mosby, St. Louis, 1990.

3. P. Rol, P. Niederer, F. Fankhauser, M. Arigoni, and E. De Haller, "Q-switched Pulses and Optical Breakdown Generation Though Optical Fibers," *Laser and Light in Ophthalmology*, Vol. 3, No. 3, pp. 213-219, 1990.

4. P. Rol, F. Fankhauser, S. Kwasniewska, and P. Niederer, "A Comparison of Ophthalmic Fiber Optic Microfocusing Systems," *Lasers and Light in Ophthalmology*, Vol. 2, No. 2, pp. 115-124, 1988.

5. F. Docchio, C.a. Sacchi, and J. Marshall, "Experimental Investigation of Optical Breakdown Thresholds in Ocular Media Under Single Pulse Irradiation with Different Pulse Duration," *Lasers in Ophthalmology*, Vol. 1, No. 2, p. 83, 1986.

6. Computer program created by P. Kennedy, Brooks AFB, TX.

7. Toth, et al., 1994.

8. P. Rol, P, Niederer, and F. Fankhauser, "High-Power Laser Transmission Through Opical Fibers: Applications to Ophthalmology," *Laser Applications in Medicine and Biology*, Plenum Press, NY, 1991.

9. R.E. Setchell, "Damage Studies in High-Power Fiber Transmission Systems," *Proceedings of SPIE*, Vol. 2114, 1994.

10. Fiberguide Industries Application Note: "Optical Fiber for High Power Pulsed Laser Delivery."

11. Conversation with Dr. Wolfgang Neuberger, President, CeramOptec GmbH, Siemensstrasse 12, Bonn, D-53121, Germany, June 10, 1994.

12. D.T. Moore, *Gradient Index Optical Systems*, The Institute of Optics, University of Rochester, Rochester, NY 14637, 1988.

13. D.T. Moore, *Selected Papers on Gradient-Index Optics*, SPIE Milestone Series, SPIE Optical Engineering Press, Bellingham, WA, 1993.

14. Nippon Sheet Glass America, Inc., Selfoc® Product Guide Application Notes, Somerset, NJ, 1993.

15. Conversation with Dr. Paul Kennedy, Research Physicist, Optical Radiation Division, Brooks AFB, TX, August 19, 1994.

16. D. Kato, "Light Coupling From a Stripe-Geometry GaAs Diode Laser into an Optical Fiber with Spherical End," *J. Applie Physics*, Vol. 44, No. 6, pp. 2756-2758, 1973.

17. H. Ward, "Molding of Laser Energy by Shaped Optic Fiber Tips," *Lasers in Surgery and Medicine*, Vol. 7, pp. 405-413, 1987.

18. V. Russo, .C. Righini, S. Sottini, and S. Trigari, "Microlens-Ended Fibers -- a New Fabrication Technique," *Proceedings of SPIE*, Vol. 405, pp. 21-27, 1983.

19. U.C. Paek and A.L. Weaver, "Formation of a Spherical Lens at Optical Fiber Ends With a CO_2 Laser," *Applied Optics*, Vol. 14, No. 2, pp. 294-298, 1975.

20. F. Docchio, P. Regondi, M.R.C. Capon, and J. Mellerio, "Study of the Temporal and Spatial Dynamics of Plasmas Induced in Liquids by Nanosecond Nd:YAG Laser Pulses. 2: Plasma Luminescence and Shielding," *Applied Optics*, Vol. 27, No. 17, pp. 3669-3674, 1988.

FINITE ELEMENT MODELING OF MANIKIN NECKS FOR THE ATB MODEL

Robert Colbert
Department of Mechanical Engineering

Villanova University
Villanova, PA 19085

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

FINITE ELEMENT MODELING OF MANIKIN NECKS FOR THE ATB MODEL

Robert Colbert
Graduate Student
Department of Mechanical Engineering
Villanova University

Abstract

The Articulated Total Body (ATB) is a rigid body dynamic model used at Armstrong Laboratory to predict the response of the human body in different environments such as automobile collisions and pilot ejections. In some situations the rigid model does not accurately predict the response. There are segments of the human body and other structures in the simulation that exhibit large deformation effects. A new version of the ATB couples the rigid body behavior with the deformation of the individual segments. The displacements due to deformation can be determined by using finite element modal analysis with models of those segments of interest. This report concentrates on the modeling of manikin necks which have shown large deformation in certain environments. The results of the modal analysis of these necks are to be used in the validation of the new version of the ATB. Hybrid II and Hybrid III manikin neck models and modal solutions are presented here.

FINITE ELEMENT MODELING OF MANIKIN NECKS FOR THE ATB MODEL

Robert Colbert

Introduction

The Articulated Total Body (ATB) is a rigid body dynamic model of the human body. The ATB is used at the Armstrong Aerospace Medical Research Lab (AAMRL) to predict the mechanical response of the human body in various situations such as automobile collisions and pilot ejections. There are, however, limitations with this rigid model. Some segments, such as the neck, undergo significant deformation in certain test environments. The new version of the ATB computer model couples the deformation of individual segments with their overall rigid body behavior (Ashrafiuon, 1993). Linear, small-angle approximations are assumed for this deformation. With this assumption, the displacements experienced by a segment due to deformation can be determined by using modal analysis. The finite element method will be employed to develop linear, elastic models of the deformable segments in order to determine the normal modes of vibration. This information will then be used by the ATB during simulation.

In the various dynamic environments for which the ATB is used, there are segments of the human body and other structures in the simulations (airbags, seat cushions, etc.) that can not be modeled as simply rigid. Researchers at AAMRL have determined that an accurate response of the human neck is the first priority to properly determine the dynamic threshold of human subjects. Before finite element models of the human neck can be developed, validation of the new version of the ATB is required. The validation compares simulation results of standard static and dynamic tests

to experimental data. These tests use Hybrid II, Hybrid III, and other prototype manikin necks exclusively (Spittle et al., 1992). This report presents the results of finite element modal analysis of manikin necks. It is to be used in conjunction with the validation report as presented by Ashrafiuon (1994).

ANSYS version 5.0a by Swanson Analysis Systems Inc. is the finite element package used in this project because its accuracy, performance, and dependability are well recognized. A Sun Workstation available at Wright Patterson Air Force Base was used as the computing platform for ANSYS.

ANSYS Solution Method

Modal analysis is used to determine the natural frequencies and mode shapes of a structure. The new version of the ATB requires this information. Since modal analysis is linear, any nonlinearities present in the model are ignored. The material models used for all manikin necks in this report are simple: linear, elastic, and isotropic. This linear material model therefore requires only three parameters to characterize its behavior: Young's Modulus (E), Poisson's ratio (ν), and density (ρ). The manikin necks are comprised of only three different materials. The properties of each are summarized in Table 1. The properties of the butyl rubber have some uncertainty, especially Young's Modulus. The properties vary significantly depending on temperature, humidity, batch, recovery time, and the engineering application, static or dynamic. Average properties under dynamic excitation were used in all cases.

Table 1: Manikin Necks Material Property Summary

Material	Young's Modulus (psi)	Poisson Ratio	Density ($lb\text{-}sec^2 / in^4$)
Steel	30×10^6	0.3	7.3548×10^{-4}
Aluminum	10×10^6	0.33	2.54×10^{-4}
Butyl Rubber	1200	0.49	8.8894×10^{-5}

The basic equation solved in a typical undamped modal analysis is the classical eigenvalue problem:

$$[K][\phi_i] = \omega_i^2[M][\phi_i]$$

where [K] is the stiffness matrix, [M] is the mass matrix, ω_i is the natural frequency of mode i, and ϕ_i is the mode shape of mode i. ANSYS offers several methods to solve this equation. This report uses the Householder-Bisection-Inverse iteration algorithm (ANSYS User's Guide, Vol IV). This method is a reduced method because it only uses a subset of the total number of degrees of freedom (DOF) and therefore is an approximate method. ANSYS uses a matrix reduction technique to reduce the size of the mass and stiffness matrices to facilitate a quick solution. This reduction technique uses selected master degrees of freedom from the entire structure to approximate the solution. The number of master DOF to use should at least be twice the desired number of modes. In this report, only the first six modes of the manikin necks are extracted to reduce the size and complexity of the subsequent ATB simulation. This is justified by using the Ritz approximation (Wilson, et al. 1982).

The choice of master DOF is somewhat arbitrary. The total number chosen is varied to test the accuracy of the solution; this will be discussed below. In all cases, the automatic generation of

master DOF is used to obtain the solution. ANSYS selects the master DOF by choosing those DOF with high stiffness-to-mass ratios in relation to neighboring nodes. The automatic selection of master DOF provides excellent results with a large savings in CPU time.

Hybrid II

The Hybrid II neck is a symmetric, cylindrical butyl rubber mold with steel end plates. A 0.5 in diameter hole runs through the length of the structure. The neck is 4.87 inches in length with a radius of approximately 1.5 inches. The finite element model uses a mapped mesh of 8-node solid elements with three DOF per node. At the ends of the neck, a circular insert was included to provide joint connection nodes for use in the ATB simulation. This insert requires the use of a free mesh of 10-node tetrahedrons also with three DOF per node. Each of these inserts is only 0.25 inches in length. Their inclusion in the model has negligible effects on the results. Figure 1 shows a finite element mesh with a total of 1986 elements and 2862 nodes.

For the simulations for which this model was developed, the Hybrid II neck is bolted at the base in three places. The nodes corresponding to these bolt locations are constrained in all directions. For the solution, a total of 100 and 1000 master DOF, out of a possible 8586 (3 x 2862 nodes) DOF, are automatically selected to compare the accuracy of the results. The 100 DOF solution took approximately 20-30 minutes to complete while the 1000 DOF case required approximately 3 hours. The difference in the first natural frequency is negligible: 40.63 Hz for the 100 DOF and 40.41 Hz for the 1000 DOF case. The solutions are within 1% of each other while the CPU time for the 1000 DOF solution is a factor of six greater. Clearly, since only the first six modes are of interest, the 100 DOF solution will suffice.

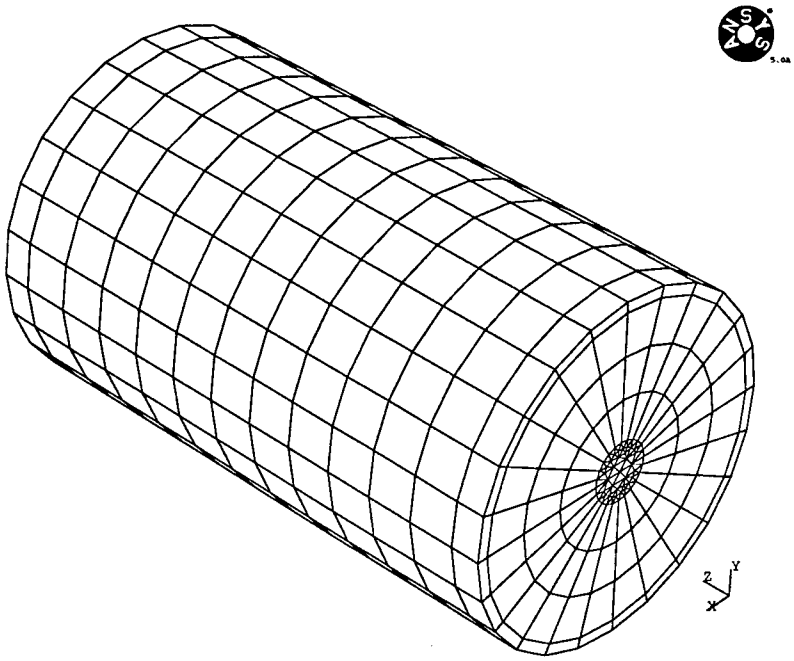


Figure 1: Finite Element Mesh of Hybrid II Neck

In addition to determining the sensitivity of the solution to the number of master DOF selected, a check of mesh density is performed. For one solution, the mesh density was approximately doubled and similar results are found. For a mesh with 1986 elements as compared to a mesh with 3914 elements the differences of the first natural frequencies are approximately 1%. Both meshes have a total number of master DOF of 100. The computational time for the finer mesh was approximately twice as long as the coarse mesh of 1986 elements. It may have been possible to even further reduce the mesh density while still obtaining reasonable results but it was decided that the mesh of 1986 elements and 2862 nodes is adequate. A summary of the sensitivity information is shown in Table 2.

Table 2: Summary of Solution Sensitivity for Hybrid II

Mesh	Nodes	Elements	Master DOF	ω_1 (Hz)	CPU Time (min)
1	2862	1986	1000	40.41	180
1	2862	1986	100	40.63	25
2	4954	3914	100	40.29	55

The solution for the Hybrid II neck is presented in Figures 2-7. Since the Hybrid II neck is bolted at one end, the structure resembles a cantilever beam and its mode shapes can be explained accordingly. The first two modes shapes are the first bending modes at approximately 40.7 Hz. Since the Hybrid II is symmetric, the first two mode shapes should be identical and orthogonal. Notice in Figures 2 and 3 that mode shapes are not aligned with the coordinate system but are rotated approximately 45^0 degrees. The slight numerical difference between the first two frequencies can be attributed to solving the reduced problem instead of using the full system matrices. Figure 4 shows the first torsion mode while the fourth mode in Figure 5 is an axial compression mode. Finally, Figures 6 and 7 show the second set of bending modes with approximately equal frequencies; higher modes typically exhibit larger numerical error in lumped mass approximations. A solution summary is given in Table 3.

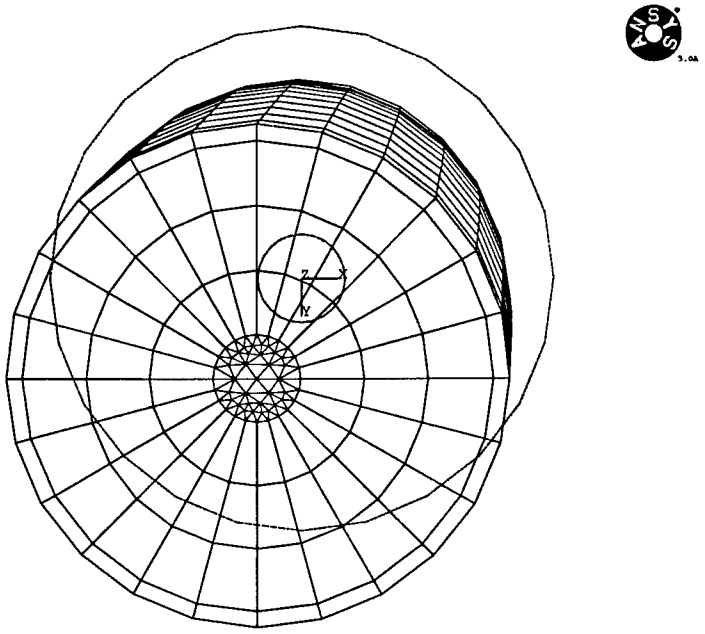


Figure 2: Hybrid II First Mode Shape (40.63 Hz)

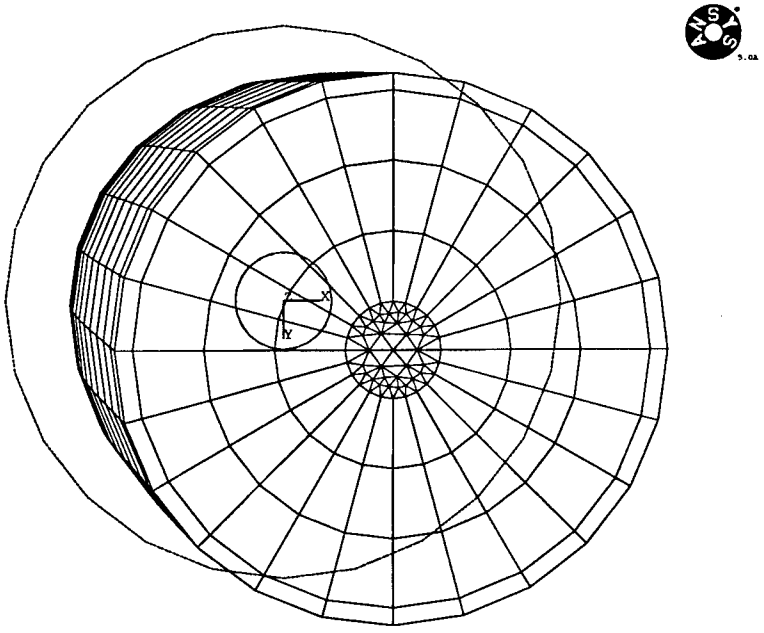


Figure 3: Hybrid II Second Mode Shape (40.84 Hz)

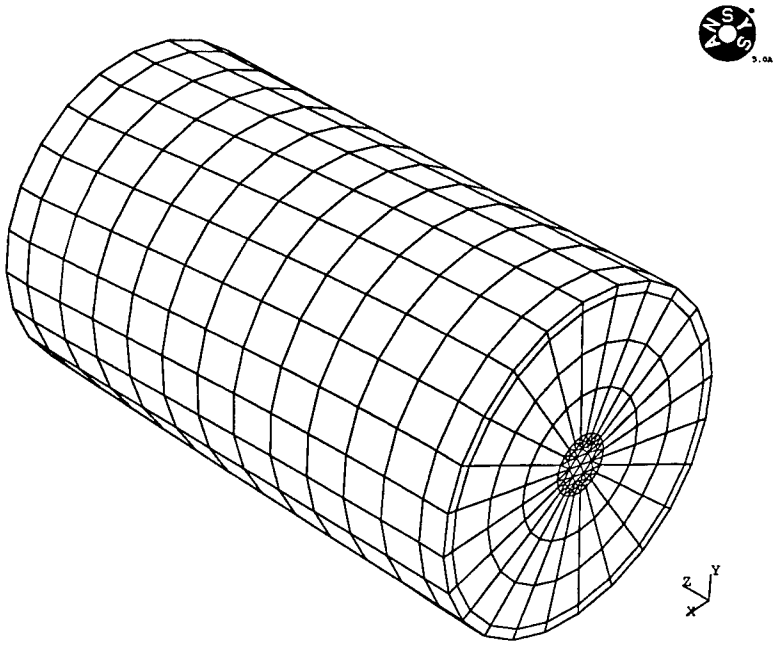


Figure 4: Hybrid II Third Mode Shape (90.23 Hz)

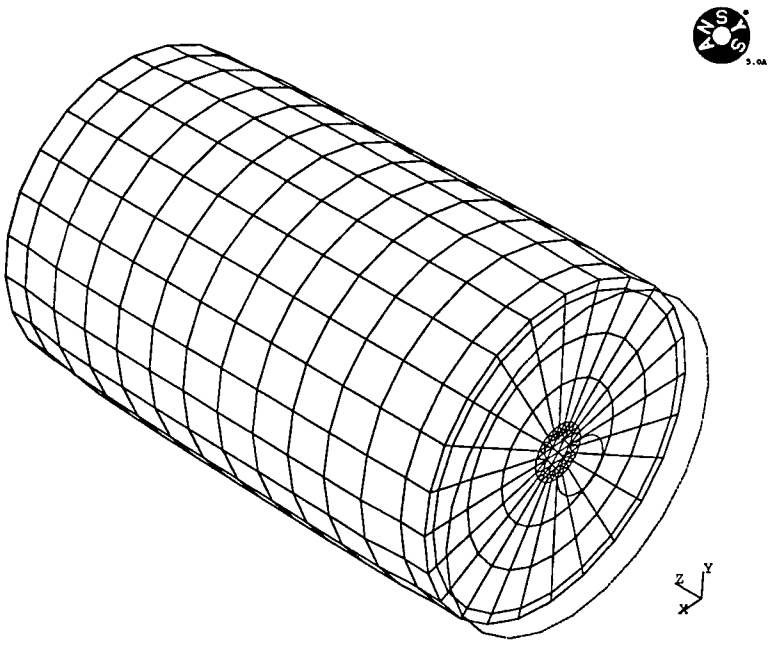


Figure 5: Hybrid II Fourth Mode Shape (157.70 Hz)

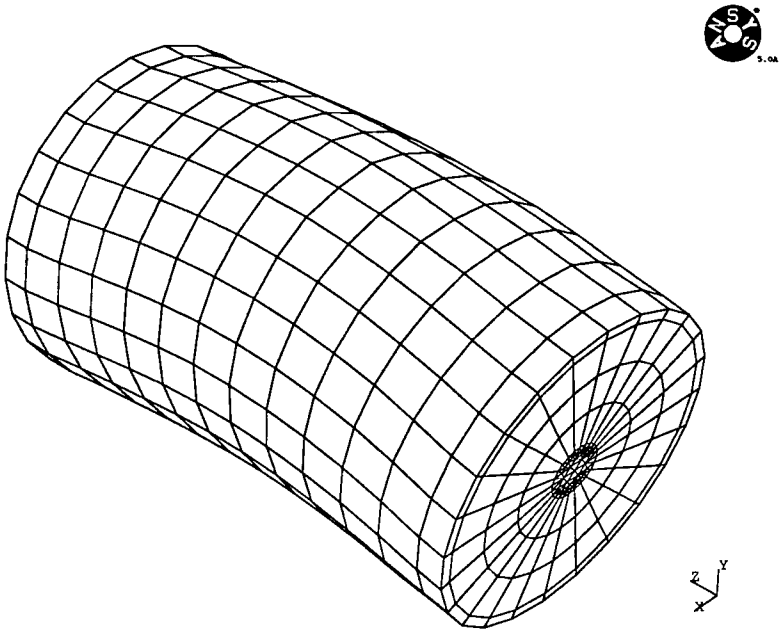


Figure 6: Hybrid II Fifth Mode Shape (185.21 Hz)

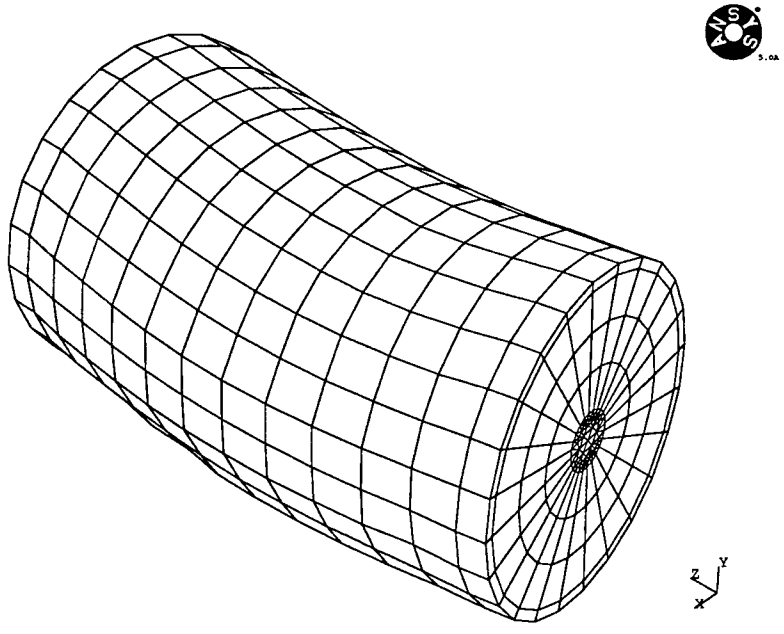


Figure 7: Hybrid II Sixth Mode Shape (190.00 Hz)

Table 3: Hybrid II Modal Analysis Solution Summary

Mode	Frequency (Hz)	Description
1	40.63	First Bending
2	40.84	First Bending
3	90.23	First Torsion
4	157.70	Axial Compression
5	185.21	Second Bending
6	190.00	Second Bending

Hybrid III

The Hybrid III neck is also made with butyl rubber but is very different than the Hybrid II. The Hybrid III is segmented with three 3.4 inch diameter aluminum plates in between the rubber sections to simulate the vertebral disks. The center rubber sections are 2.7 inches in diameter and are offset 0.2 inches towards the front of the neck to provide a different response in flexion and extension bending. In addition, slices are made in the rubber material towards the front to more closely simulate the asymmetrical bending characteristics of the human neck. There are also aluminum end plates to facilitate assembly with a manikin. Finally, a steel cable runs through a 0.625 inch diameter hole in the neck. The cable is torqued to limit excessively large rotations in the neck. The total length of the neck is 5.66 inches.

Meshing of the Hybrid III is not as straightforward as the Hybrid II. Since modal analysis is linear, it is not possible to model the slits or the cable. To model the slits in the neck would require using a contact surface which implies a "force" as a function of time. Since the classical eigenvalue problem does not have a force vector, this is not possible. The cable is a nonlinear effect that is

activated at a certain neck rotation, which again is not possible with this linear analysis. Both of these effects have to be addressed in the ATB simulation.

This mesh also uses both 8-node solids and 10-node tetrahedrals. Inserts are also used here for the joint connection nodes and again their effect is negligible. The Hybrid III neck is bolted at one end in the experimental tests. There are four bolts at the base and the nodes corresponding to these bolts are constrained in all directions. This model can also be explained in terms of the modal shapes of a cantilever beam. A typical mesh is shown is Figures 8 and 9; it has 2386 elements and 3427 nodes. A similar sensitivity analysis is performed for the Hybrid III neck and is summarized in Table 4. The mesh in Figures 8 and 9 have 100 DOF.

Table 4: Hybrid III Solution Sensitivity

Mesh	Nodes	Elements	Master DOF	ω_1 (Hz)	CPU Time (min)
1	3427	2386	1000	35.88	210
1	3427	2386	100	36.14	30
2	5633	4428	100	36.02	55

The results of the modal analysis for the Hybrid III are presented in Figures 10-15 and are summarized in Table 5. Since this model is asymmetric, the first two modes should not be identical, however the difference in their natural frequencies can probably be attributed to numerical error (36.14 Hz and 36.35 Hz). These first two modes are the first bending modes of a cantilever beam. Although it is not shown in Figure 11, the second mode exhibits some twist probably to the asymmetric geometry. The differences in flexion and extension of the physical neck is not apparent in the finite element model; this too would have to adjusted in the ATB simulation. The third mode

is the first torsion mode excited at a frequency of 57.19 Hz. The Hybrid III is different than the Hybrid II results in that the second set of bending modes occur at the fourth and fifth modes at frequencies of 140.97 Hz and 144.28 Hz. Finally the sixth mode shape is the second torsion mode at 184.31 Hz.

Table 5: Hybrid III Modal Analysis Solution Summary

Mode	Frequency (Hz)	Description
1	36.14	First Bending
2	36.35	First Bending
3	57.19	First Torsion
4	140.97	Second Bending
5	144.28	Second Bending
6	184.31	Second Torsion

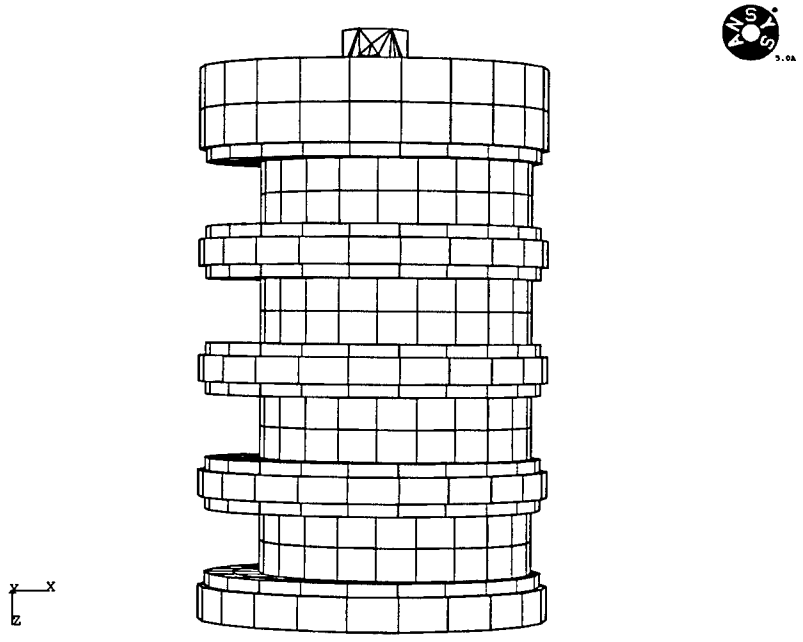


Figure 8: Hybrid III Finite Element Mesh

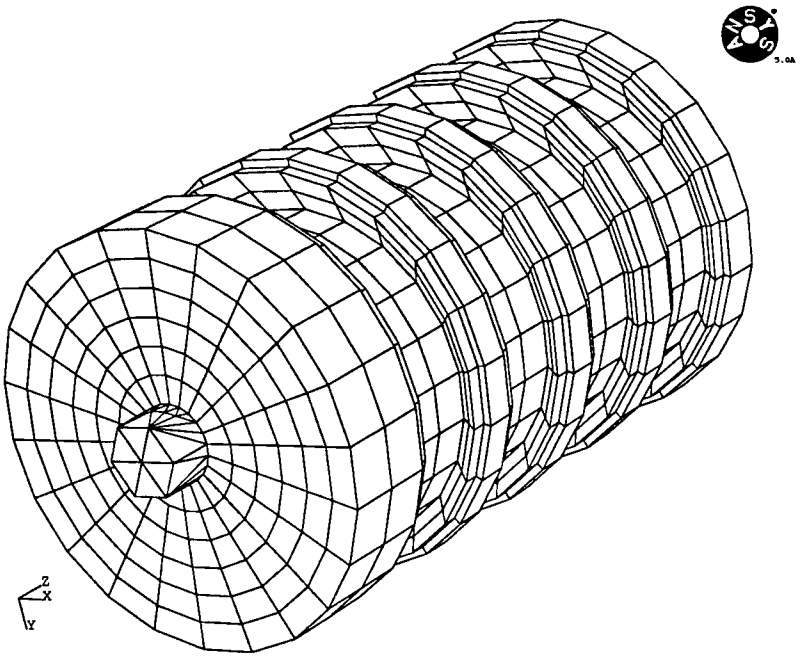


Figure 9: Hybrid III Finite Element Mesh

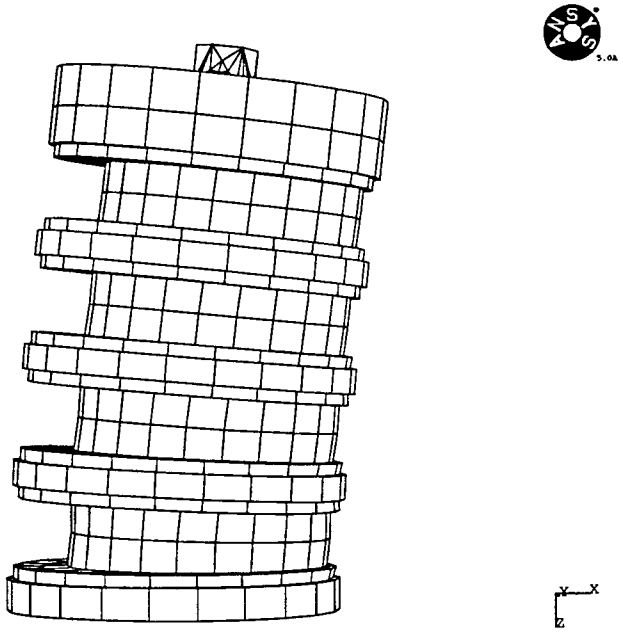


Figure 10: Hybrid III First Mode Shape (36.14 Hz)

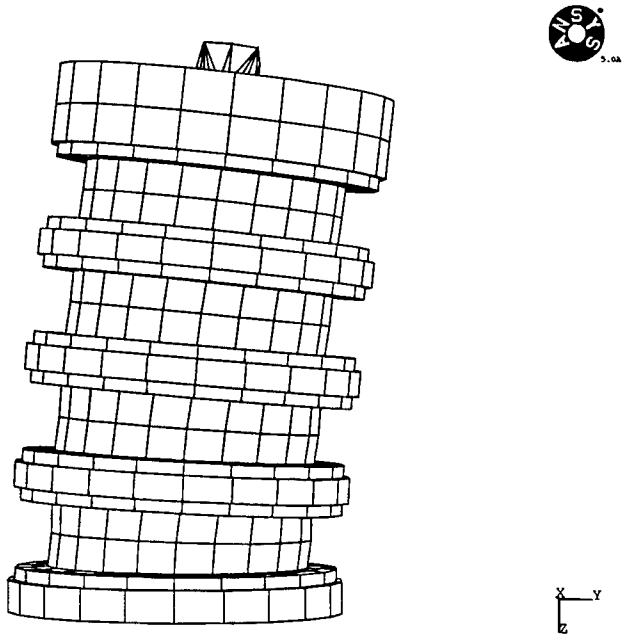


Figure 11: Hybrid III Second Mode Shape (36.35 Hz)

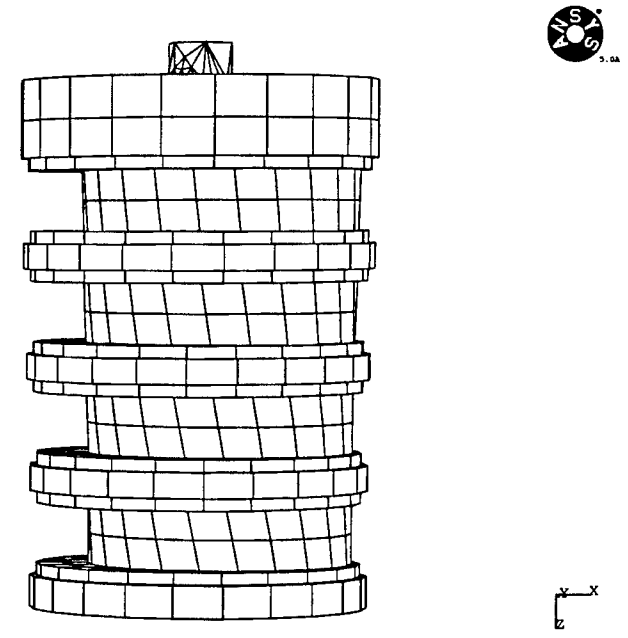


Figure 12: Hybrid III Third Mode Shape (57.19 Hz)

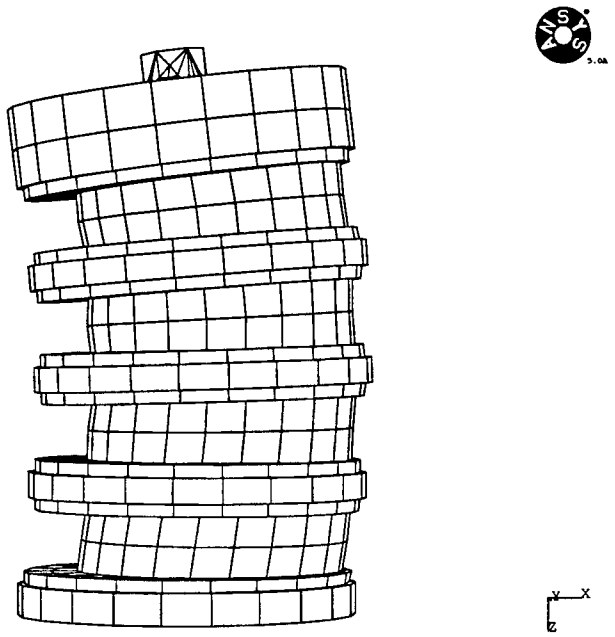


Figure 13: Hybrid III Fourth Mode Shape (140.97 Hz)

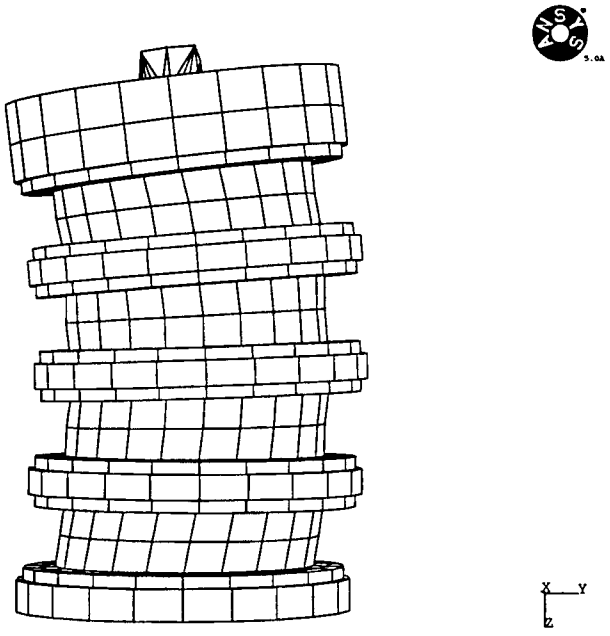


Figure 14: Hybrid III Fifth Mode Shape (144.28 Hz)

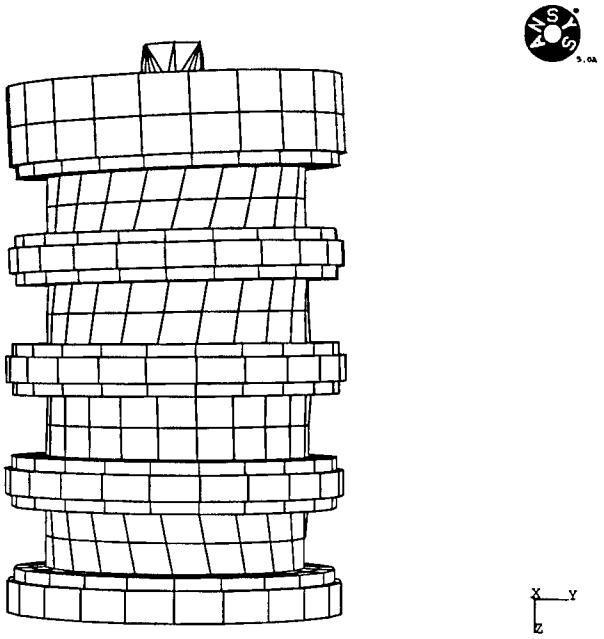


Figure 15: Hybrid III Sixth Mode Shape (184.37 Hz)

Prototype Necks

Researchers at AAMRL are in the process of evaluating several new necks to replace the Hybrid III. One neck in particular was designed in-house which simulates the skeletal structure of the human neck. This prototype uses aluminum segments with two other polymer materials that are difficult to obtain reliable properties for modeling. Standard tests have been completed on this neck. A preliminary finite element model was developed, but time constraints limited its completeness. In addition, some characteristics of this prototype neck are not well understood and require more study.

Conclusion

Finite element models of the Hybrid II and Hybrid III manikin necks have been developed. Modal analysis was completed on these models and the first six modes have been extracted for inclusion in the validation of the new ATB model. Solution sensitivity to mesh density and the number of master degrees of freedom has been examined. The manikin neck models developed exhibit an acceptable accuracy with a relatively inexpensive use of computer resources. Another prototype neck was studied, however time constraints permitted only a cursory examination and a limited model.

REFERENCES

ANSYS User's Manual For Revision 5.0, 1992, Volume I: Procedures. Swanson Analysis Systems, Inc.

ANSYS User's Manual For Revision 5.0, 1992, Volume IV: Theory. Swanson Analysis Systems, Inc.

Ashrafiuon, H., 1994, "ATB Simulation of Deformable Manikin Neck Models", The 1994 AFOSR Summer Faculty Research Program, Armstrong Laboratory, Final Report.

Ashrafiuon, H., 1993, "Modeling of Flexible Bodies for the ATB Model", The 1993 AFOSR Summer Faculty Research Program, Armstrong Laboratory, Final Report.

Spittle, E.K., Shipley, B.W., and Kaleps, I., 1992, "Hybrid II and Hybrid III Dummy Neck Properties for Computer Model", Report No. AL-TR-1992-0049, National Technical Information Service, Springfield, VA.

Wilson, E.L., Yuan, M. W., and Dickens, J. M., 1982, "Dynamic Analysis by Direct Superposition of Ritz Vectors", Earthquake Engineering and Structural Dynamics, Vol. 10, pp. 813-821.

ESTIMATION OF FOUR ARTERIAL VASCULAR PARAMETERS FOR TRANSIENT AND STEADY BEATS

Steven J. Essler
Graduate Research Assistant
Department of Electrical Engineering

North Dakota State University
EE Building, Room 101
Fargo, North Dakota 58105

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

September 1994

ESTIMATION OF FOUR ARTERIAL VASCULAR PARAMETERS FOR TRANSIENT AND STEADY BEATS

Steven J. Essler
Graduate Research Assistant
Department of Electrical Engineering
North Dakota State University

Abstract

A faster method was developed to estimate four arterial vascular parameters under steady and transient beat conditions. A four element electrical circuit was used as a model for the arterial vascular system. Mathematical development for the impedance of this model was reduced to its real and imaginary components. Fast fourier transforms (FFT) were used on simulated aortic pressure (AoP) and aortic flow (AoF) data to obtain arterial impedance at various frequencies. Using a numerical routine, the four parameters could then be estimated from a best fit solution using both the mathematical equations for impedance and the FFT impedance.

Estimations were done for steady and transient beats. Transient beats were estimated assuming that the transient behavior was a linear ramp due to a low amplitude, low frequency, baseline shift. Results show that this technique can estimate the four parameters accurately for both steady and transient beats. At the same time, this FFT algorithm proved to be a much faster way to estimate these parameters.

ESTIMATION OF FOUR ARTERIAL VASCULAR PARAMETERS FOR TRANSIENT AND STEADY BEATS

Steven J. Essler

INTRODUCTION

Purpose:

The purpose of this research was to find a faster way to predict the parameters of a four element arterial vascular model for both steady and transient beats.

Background:

It is not well understood how changes in the parameters of the arterial vascular system are affected by gravitational force in the Z-direction (Gz) from 0 to 9 Gz. By modeling the arterial system and studying the effects of Gz upon the model one can then begin to study the effects that +Gz forces has upon the actual physiological system. The arterial vascular system has been modeled with two parameters since the eighteenth century by Stephen Hales and is called the Windkessel model (Milnor, 1990). This model uses a resistor (total peripheral resistance) and a capacitor (systemic arterial compliance) in parallel. Even though this model is simple, the lumped elements neglect all the variations in dimensions and physical properties within the arterial tree, and cannot be used to study pulse wave velocity or the transformation of pulse waves as they travel through the system (Milnor). Better yet is the three element model which is widely used (Toorup et al., 1987) and (Nichols et al., 1990). In this model a characteristic impedance is added in series with the windkessel model to better mimic the arterial input impedance at low and high frequencies. However, a four element model was used in this research. An inductor was placed in parallel with the characteristic impedance of the three element model to take into account the affects of the blood inertia. Since the aortic pressure did not rise fast enough, an inertial component was placed in parallel with the characteristic impedance (Schroeder, 1994). The four element model produces a simulated aortic pressure response nearly identical to the physiological aortic pressure and gave a more realistic input impedance magnitude and phase response as compared to the two or three element models. Also, the dicrotic notch in AoP was present in the waveform due to an inertia component. Therefore, a

four element model was used and includes blood inertia (Lp), characteristic impedance (Rp), arterial resistance (Ra) and an arterial compliance capacitance (Ca).

Previous methods for the estimation of this four parameter model have been done for steady beats (Schroeder). However, the method used was very slow and time consuming. Also some estimations were done for transient pressure beats (Yin et al.,1989) and (Toorop et al.). The methods used there were not fourier techniques and they were only done for two and three element models. Using a fast fourier transform technique would be much faster to estimate these parameters. Fourier techniques are nothing new to find input impedance to the arterial system (Nichols et al.), but it has never been done for the four parameter model. Data was obtained from Gz-suited and unsuited baboons to estimate this four parameter model, but some of the data being obtained has a transient baseline shift rendering this data useless for most common kind of parameter estimations.

During some experiments at the Brooks Air Force Base, it was noticed that some of the signals can have a baseline shift superimposed upon the signal. This is particularly noticeable in centrifuge runs due to the high +Gz forces obtained. Not only for baseline shifts, but during pressure transients like atrial fibrillation there is a need to be able to estimate beat-to-beat transients. During such transients the pressure at the onset and end of a cardiac cycle usually differ. This pressure difference necessitates a modification of usual methods for estimating these hemodynamic parameters (Yin, et al.) A linear assumption will be used in modifying a transient beat to be able to estimate the four parameter model, which will be discussed in detail later.

Thus, the projection of this research was to develop this new method to estimate the parameters faster and for transient beat data as well. The quickly obtained estimates can than be better used in the studies of +Gz forces and the effects that is put upon the cardiovascular system.

Scope:

First, the arterial model will be described. Second, mathematical development of the impedance equations and transient behaviors will be derived. Third, simulation results for both steady and transient

beats are given. Lastly, the conclusions and future developments will be discussed.

ARTERIAL MODEL DESCRIPTION

Shown in Figure 1 is the electrical circuit model of the arterial vascular system that was used. From this model, the impedance equations can be developed to predict Lp, Rp, Ca and Ra.

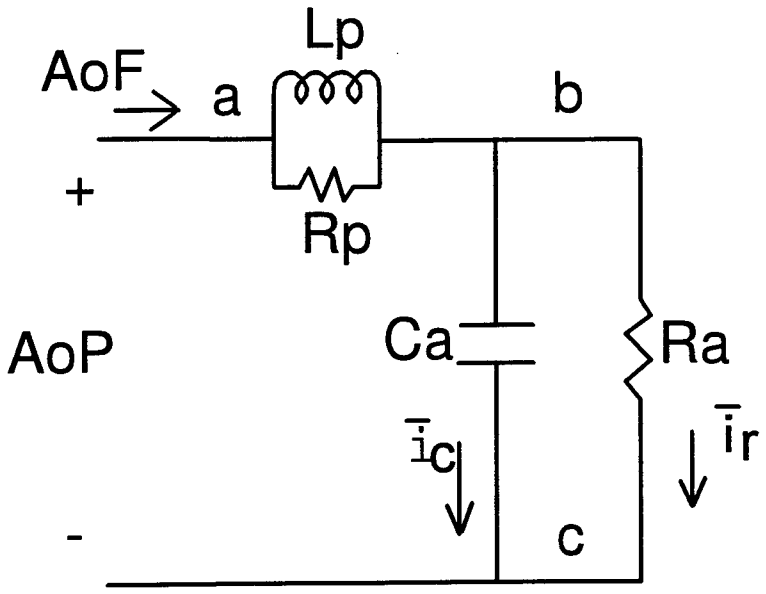


Figure 1: Arterial Vascular Model

Circuit definitions:

AoF	= aortic flow	(mL/sec)
AoP	= aortic pressure	(mmHg)
Lp	= blood inertia	(mmHg*s^2/mL)
Rp	= characteristic impedance	(mmHg*s/mL)
Ra	= arterial resistance	(mmHg*s/mL)
Ca	= arterial capacitance	(mL/mmHg)

MATHEMATICAL DEVELOPMENT

Assumptions:

(1) Ca, arterial capacitance, is constant over one beat

(2) The baseline shift of AoP is linear

(3) The system is a linear time-invariant system

The input impedance equations can be developed using Laplace transforms. For an easier visual of the equations, let L = Lp, Z = Rp, C = Ca and R = Ra. Starting with the impedance between nodes a and b, the impedance is:

$$Z_1(s) = (R/sC) / (R + 1/sC) \qquad \text{(Eq. 1)}$$

Similarly the impedance between nodes b and c is:

$$Z_2(s) = sZL / (sL + Z) \qquad \text{(Eq. 2)}$$

So that the total impedance between nodes a and c is $Z_{total}(s) = Z_1(s) + Z_2(s)$

$$Z_{total}(s) = (R/sC) / (R + 1/sC) + sZL / (sL + Z) \qquad \text{(Eq. 3)}$$

After algebraic manipulation, letting $s = j\omega$ to get to the frequency domain from the Laplace transform, and multiplying numerator and denominator by the complex conjugate one can arrive at:

$$Z_{total}(j\omega) = \{[Z\omega^4 + \omega^2(1/RC^2 + Z/R^2C^2) + Z^2/RC^2L^2] + j[\omega^3(Z^2/L - 1/C) + \omega Z^2/LC(1/R^2C - 1/L)] \; / \; [\omega^4 + \omega^2(1/R^2C^2 + Z^2/L^2) + Z^2/R^2C^2L^2]\}$$

(Eq. 4)

Now let $K_4 = \omega^4$, $K_3 = \omega^3$, $K_2 = \omega^2$ and $K_1 = \omega$. And further reducing $Z_{total}(j\omega)$ into its real and imaginary parts one can obtain the following:

$$Im(Z_{total}(j\omega)) = \{[K_3 R^2L(Z^2C^2 - CL) + K_1Z^2(L - R^2C)] \; / \; [K_4(RCL)^2 + K_2(L^2 + Z^2R^2C^2) + Z^2]\}$$

(Eq. 5)

and

$$Re(Z_{total}(j\omega)) = \{[K_4 ZR^2C^2L^2 + K_2L^2(R + Z) + Z^2R] \; / \; [K_4 R^2C^2L^2 + K_2(L^2 + Z^2R^2C^2) + Z^2]\}$$

(Eq. 6)

The basic algorithm to solve for these parameters was implemented using Matlab© software because of its capabilities and ease of handling large data matrices. The algorithm performs a fast fourier transform (FFT) on AoP and AoF respectively. The ratio of the transformed AoP over AoF is then taken to get the impedance in the frequency domain. From this, the real and imaginary components are extracted and are used as the comparison for a sum of a least squared error algorithm against the imaginary and real impedance Eq. 5 and Eq. 6. Initial guesses are required to start the algorithm. Simulated data (AoP and AoF) were used from previous work (Schroeder) which predicts the same parameters but at a much slower rate.

To apply the technique to transient beats, a different approach was used. As can be seen in Figure 2, the AoP waveform for one beat has imposed on it a transient baseline shift from beginning to ending of the beat. The assumption that this transient behavior is linear is justified by the fact that the baseline shift appeared to be a linear transition of low magnitude. The probable frequency of the transient appeared much lower than the frequency of the heart rate, thus the portion of one cycle of the transient signal is only a small fraction of a cycle as compared to one cycle of the heart beat. Also the assumption of this system being a linear time-invariant system allows the application of the superposition theorem. With this

theorem, the transient triangle that is formed from the linear assumption of the baseline shift (see Fig. 2) can be subtracted from the original signal, AoP. This is required for performing an FFT on the signal, as an FFT requires that the beginning and ending point for one cycle to be at the same magnitude, i.e., they have the same start and stop points in the cycle.

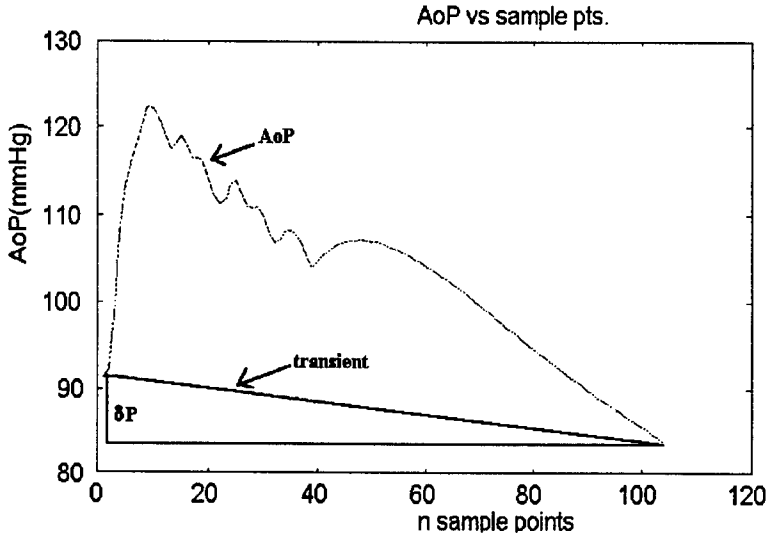


Figure 2: AoP vs sample points

The problem with removing this transient triangle from AoP is that the DC term (at zero hertz) from the FFT is smaller than if it wasn't removed, thus resulting in error in the parameter estimation for Ra. To correct for this, compensation was made for the change in pressure, δP, that is seen in Fig. 2. The δP results from a discharge of Ca in the model. This, in turn, changes the current through the resistor Ra, since Ra is the discharge path for Ca. Hence, this change of current through Ra leads to a false estimation of Ra. Having knowledge of δP, the average current into Ca can be calculated using the equation for compliance (Milnor) , Eq. 7.

$$Ca = \delta V / \delta P \qquad \text{(Eq. 7)}$$

where Ca = compliance, δV = change in volume, δP = change in pressure

The developed algorithm initially estimates Ca to solve for δV, thus finding the change in volume.

The average current through the arterial capacitance, Ca, is then calculated using Eq. 8.

$$\bar{i}_c = \delta V / T_{HR} \qquad \text{(Eq. 8)}$$

where $\bar{i}_c$ = average capacitance current, δV = change in volume, T_{HR} = heart rate (sec.)

From this, the average value of flow can be found through the resistor, Ra, by Eq. 9.

$$\bar{i}_r = \bar{i}_h - \bar{i}_c \qquad \text{(Eq. 9)}$$

Where $\bar{i}_r$ (see Fig. 1) is the average current through arterial resistance, Ra, $\bar{i}_h$ is the average current through the resistor obtained from the FFT of AoF at zero Hertz and $\bar{i}_c$ = average capacitance current, (see Fig. 1) which may be positive or negative depending on the polarity of the transient baseline shift. The newly corrected resistance value is then calculated using the DC FFT component divided by this new value of $\bar{i}_r$ found in Eq. 9. And finally, the algorithm is reiterated using this new resistance value as an initial guess for Ra along with the other previously estimated parameters Lp, Rp and Ca as initial guesses to arrive at the final four parameter estimation.

SIMULATION RESULTS

The algorithm developed above was used and implemented using Matlab© software with simulated AoP and AoF waveforms as inputs which were obtained from Schroeder's work. The resulting parameter estimations are then compared with Schroeder's estimations. First this was done for a steady beat and secondly for a transient beat using the first five harmonics obtained from the FFT algorithm. The first five harmonics were used because 95% of the energy in most pressure and flow signals is contained in the first five harmonics (Milnor, 1990). Also parameter estimations were made without any compensation for transient behavior to show how badly the parameters are estimated. The results of that are also shown and compared against Schroeder's estimations.

Steady beat:

A sample steady beat of AoP and AoF can be seen in Fig. 3. These AoP and AoF signals were used as inputs to the FFT algorithm derived above. Tabulated results and comparisons to Schroeder's work can be

found in Table 1. The resulting time factor involved in calculating these parameters was the same for both the steady and transient beats using the FFT algorithm. The software was run on a 55 Mhz IBM compatible personal computer and the FFT algorithm took approximately 6 seconds to estimate the four parameters. Whereas, Schroeder's estimation routine took approximately 30 minutes and 45 seconds to estimate those same four parameters. Thus, the FFT algorithm performed the estimation on the order of 300 times faster than the Schroeder estimation.

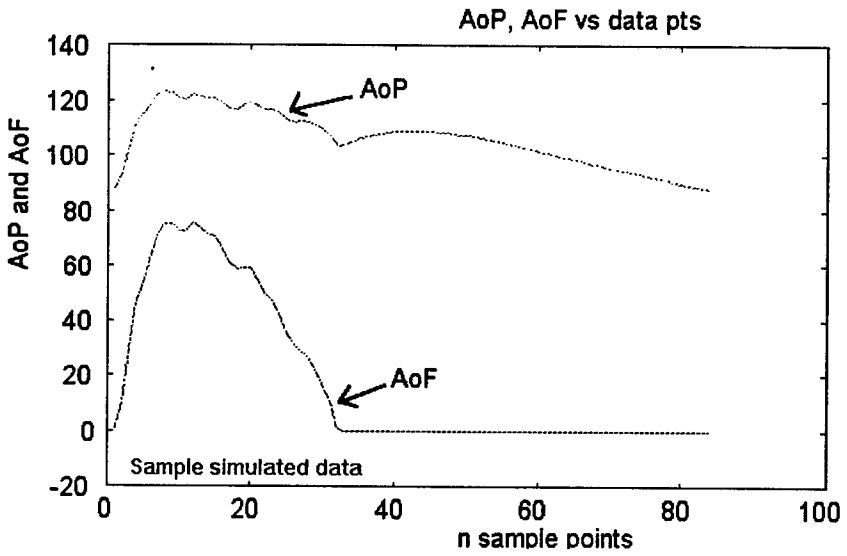


Figure 3: Simulated AoP and AoF vs 'n' data points

Parameter	FFT Algorithm	Schroeder estimate	% difference
Lp	0.02	0.02	0.00
Rp	0.11	0.11	0.00
Ca	0.59	0.58	1.72
Ra	5.82	5.91	1.52

Table 1: Results for a steady beat

Where % difference with respect to Schroeder's estimations were calculated as follows:

$$\% \text{ diff} = [\ |\text{Sch} - \text{fft1}|\ /\ \text{Sch}\ ] * 100\% \qquad \text{(Eq. 10)}$$

Sch = Schroeder estimate, fft1 = estimate for FFT algorithm developed in this paper

Transient beat:

To obtain a transient aortic pressure waveform, two parameters were changed in the generation of the simulated data from Schroeder's work. Specifically, Ca was changed to 1.2*previous Ca, and also the new heart rate was changed to 0.8*previous heart rate. This allowed a transient to be obtained before the simulation arrived at its steady state condition. In Fig. 4, the simulated transient AoP and AoF can be seen. A more detailed look at the transient AoP can be found in Fig. 5. In this example, the transient δP is 7.63 mmHg which is approximately 18% of the peak to peak amplitude of AoP. With these AoP and AoF signals as inputs to the FFT algorithm, the results in Table 2 were obtained. Also comparison against Schroeder's estimations and estimations without any transient corrections in the algorithm can be seen in Table 2.

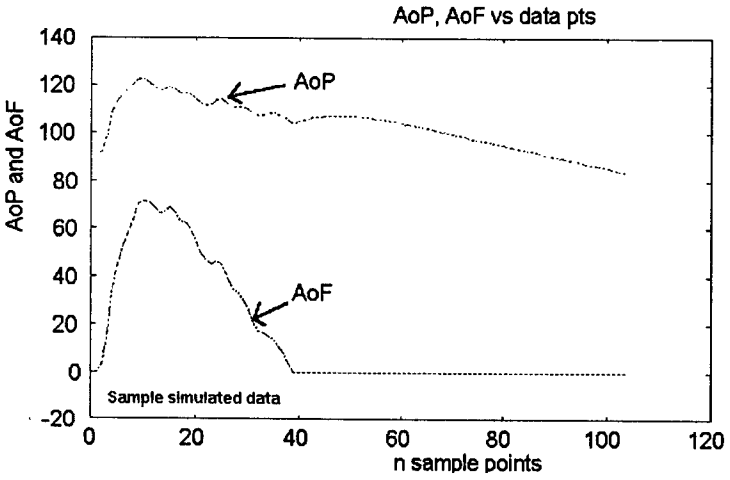


Figure 4: Simulated transient AoP and AoF vs 'n' data points

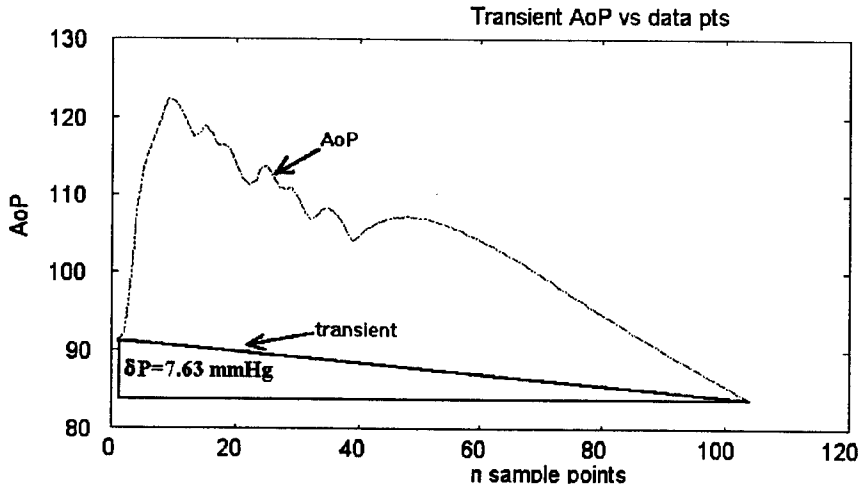


Figure 5: Simulated transient AoP vs 'n' data points

Parameter	FFT algorithm estimation	FFT algorithm w/o transient correction	Schroeder estimation	% difference FFT vs. Schroeder	% difference w/o correction vs. Schroeder
Lp	0.02	0.02	0.02	0.00	0.00
Rp	0.13	0.11	0.13	0.00	15.38
Ca	0.58	0.74	0.58	0.00	27.59
Ra	6.11	6.84	5.91	3.38	15.74

Table 2: Results for a transient beat

CONCLUSIONS

This FFT method proved to estimate steady beats and transient beats with a significant time savings from previous methods. The significant time savings was on the order of magnitude 300 times faster using this FFT method. It was determined that the transient behavior of the aortic pressure could be compensated for and the overall results using the FFT technique showed little differences as compared with Schroeder's work. Whereas, if the transient was not compensated for, significant error was found in the parameter estimations. With a faster technique for the four parameter estimations and having estimation capabilities for transient beats as well, the analysis of the arterial vascular system under micro-gravity studies will be made easier. Having knowledge of the parameters quickly could lead to development of counter measures for use in +Gz and micro-gravity space flights.

Acknowledgments:

The author would like to thank Research Development Laboratories, Dr. Dan Ewert and the employees of Systems Research Laboratories at Brooks Air Force Base for the guidance and opportunity to perform this research.

REFERENCES

Milnor, W., 1990, Cardiovascular Physiology, Oxford University Press.

Nichols W., O'Rourke M., 1990, McDonald's Blood Flow in Arteries, Edward Arnold, 3rd ed.

Schroeder, M, 1994, Optimal Ventricular-Arterial Coupling With Respect to Arterial Elastance, Masters Thesis, North Dakota State University.

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

Yin, Frank C.P., Liu Z., "Estimating Arterial Resistance and Compliance During Transient Conditions in Humans", Am. J. Physiol. 257 (Heart Circ. Physiol. 26): H190-H197, 1989.

ACCURACY CURVES IN A LOCATION-CUING PARADIGM FOR VISUAL ATTENTION

Lawrence R. Gottlob
Department of Psychology

Arizona State University
Tempe, AZ 85201

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

ACCURACY CURVES IN A LOCATION-CUING PARADIGM FOR VISUAL ATTENTION

Lawrence R. Gottlob
Department of Psychology
Arizona State University

Abstract

The aim of the study was to investigate the allocation of visual attention in order to differentiate between two general classes of mechanisms: (a) switching attention across locations on different trials, and (b) sharing attention across multiple locations within a trial. A location-cuing method was used to investigate the time-course of attention growth at valid and invalid locations, as a function of cue probability. It was proposed that the accuracy curves produced would be diagnostic of whether a switching or sharing strategy was used to allocate attention over the visual field. The pattern for valid curves differed from the pattern of invalid curves. However, the data did not show a clear effect of cue probability and could not be analyzed for switching versus sharing.

ACCURACY CURVES IN A LOCATION-CUING PARADIGM FOR VISUAL ATTENTION

Lawrence R. Gottlob

The movement of attention in the visual field, in the absence of eye movements, is well established in the literature (Allport, 1992; Cheal & Lyon, 1989; Jonides, 1980; Posner, 1980). The location-cuing procedure is a method whereby a given location is cued in the visual field, and detection or discrimination performance is compared across various conditions. The main finding has been that performance is better at cued than at uncued locations.

There have been a few popular metaphors for the movement or allocation of attention. Two of the most cited are the spotlight (Posner, 1980; Tsal, 1983) and the zoom lens (Eriksen & St. James, 1986). However, neither the spotlight nor the zoom lens metaphor could accommodate findings that attention concentration may not necessarily be restricted to contiguous areas of a display (Driver & Baylis, 1989).

In both spotlight and zoom lens metaphors, the focus of attention is changed using a switching mechanism. Both conceive of attention as an increase in processing rate of a small area, with regions outside this area processed at a background level that is not adjustable (Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Jonides, 1983). When one particular location of the visual field is analyzed, it can be conceived of as being in one of two possible states: attended, or not attended. With a switching mechanism, the attention system varies the "amount" of processing resources allocated to a given point in the visual field (over a number of trials) by varying the percentage of trials in which that point is included in the area of increased processing.

In contrast to a switching mechanism which has a fixed strength focus and background, with a sharing mechanism the system has the ability to vary the proportion of resources allocated to different locations in an area of the visual field within a trial (e.g., Jonides, 1980; LaBerge & Brown, 1986). With a sharing mechanism, it cannot be said that any given location is clearly inside or outside the focus of attention. Processing of the various locations proceeds in parallel, with the total amount of processing restricted by capacity limitations. The rates of processing at different locations can take on one of a large number of

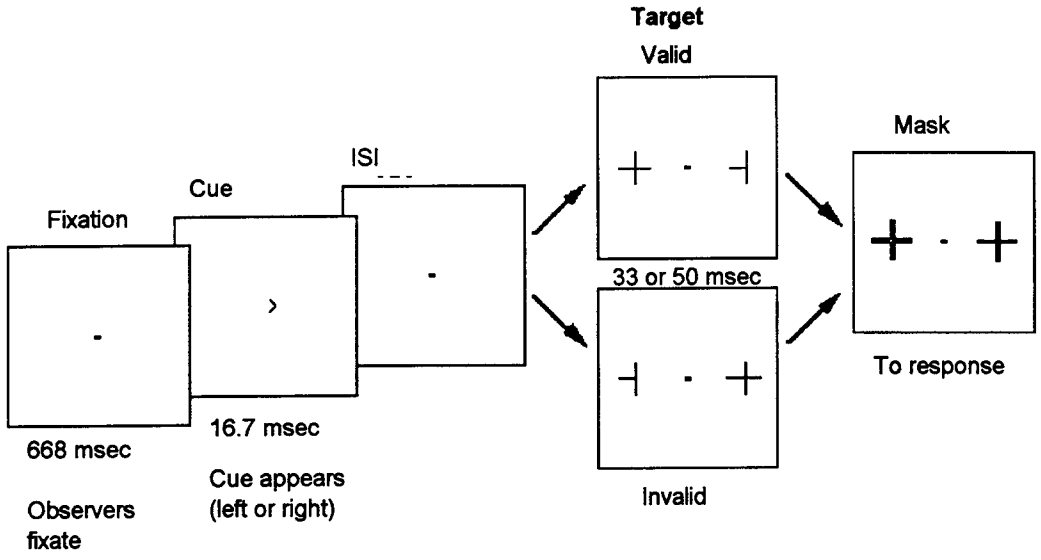
discrete values, or a value in a continuous range, in contrast to a switching mechanism, where the rate of processing of a given area is restricted to one of two discrete values. One formulation which is consistent with an attention sharing mechanism is the gradient-filter metaphor (Cheal, Lyon, & Gottlob, 1994). In such a gradient, there is an area of increased processing (focal area), the strength and spatial extent of which can be altered to some extent, as can the strength of regions outside the focal area.

Gottlob, Cheal, and Lyon (1994) investigated the two possible kinds of mechanisms (switching and sharing) for the allocation of visual attention. The technique used in the second experiment of Gottlob et al. (1994) was to examine the shapes of attention operating characteristic curves (AOC) as percentage of valid cues was varied across condition. An AOC curve is analogous to an ROC curve, and is produced by plotting accuracies (proportions correct) from two tasks against each other in an X-Y plot (Sperling, 1984). The AOCs in Gottlob et al. (1994) consisted of X-Y plots with valid accuracy on the y-axis and invalid accuracy on the x-axis. One point was plotted for each attention allocation condition. Switching strategies where the system is restricted to two possible states will produce linear AOCs with a slope of negative one. Sharing strategies with conservation of overall attentional resources will produce AOCs that are concave toward the origin. Based on model fits, the sharing strategy was supported but not unequivocally. It appeared that observers were sharing, but that conservation of overall resources was violated.

The present experiment was performed as a follow-up to the experiments in Gottlob et al. (1994). It was suggested by the results of the first experiment in Gottlob et al. (1994), where the characteristic shapes of curves were proposed as diagnostic of switching or sharing. The rationale for the present experiment will be explained in reference to a schematic of the method (Figure 1). Observers are presented with a central arrow cue that points to one of two locations. Two types of trial are presented: (1) valid trials where the cued location and the target location are the same, and (2) invalid trials where the two locations differ. As stated above, generally the cued (valid) location shows better performance than the uncued (invalid) location. Observers are exposed to three blocked conditions: (1) 100% where all cues are valid, (2) 75% where 75% of trials contain valid cues and 25% contain invalid cues, and (3) 50%

where half the trials are valid and half are invalid. In addition, the amount of time between cue onset and target onset is manipulated; the cue-target onset asynchrony (SOA) ranges from 16 to 300 msec.

Figure 1. Order of events in a location-cuing trial (adapted from Cheal et al., 1991). Observers first fixate on a central bar. The cue appears, followed by target presentation after a variable cue-target onset asynchrony (SOA). ISI is SOA minus 16.7 msec cue duration. Valid and invalid trials are included.



It has been found that the SOA-accuracy curves differ across both valid-invalid and probability (100%, 75%, 50%) conditions (Cheal et al., 1991; Gottlob et al., 1994). Valid curves, with central arrow cues, rise to asymptote at about 250-300 msec SOA, while invalid curves are generally flat or slightly decreasing with SOA. In addition, probability affects the shape of accuracy-SOA curves; for valid curves, 75% curves lie above 50% curves, while for invalid curves, invalid 75% curves lie below invalid 50% curves.

The technique used in the present experiment was based on the observation that valid and invalid accuracy-SOA curves have different shapes, and thus, might be diagnostic of switching vs. sharing

strategies. A switching strategy would include a certain number of trials where two events co-occur (one emitted by the observer and one emitted by the computer monitor): (1) the observer does not use the cue and orients attention to the uncued location, and (2) the trial is invalid. These types of trials will be termed "false valid" trials because the SOA curves should have the same pattern as valid trials, according to a switching model. The converse type of trial, where the trial is valid but the observer does not use the cue, will be called "false invalid" and the SOA curves should resemble those for invalid trials. According to a switching model, invalid trials should contain a certain proportion of false valid trials, and valid curves should contain a certain proportion of false invalid trials. There are two possible tests for switching: (1) look for evidence of false valid trials among the invalid trials, and (2) look for evidence of false invalid trials among the valid trials.

There were some assumptions to be made about the shapes of the curves. It is possible to estimate a "pure" valid SOA curve by using a 100% valid condition, but a pure invalid curve can only be modeled from theory, since it is not possible to run a pure invalid condition. (Observers could turn a 0% valid cue into a 100% valid reversed cue.) It was assumed that pure invalid trials are flat with SOA, for the following reason: The attention framework from Cheal et al. (1991) suggests that on invalid trials, attention is not allocated to the target location until a non-target appears to the cued location and the target appears at the uncued location. Since the system receives no benefit from the cue, the SOA curve would be expected to be flat. Therefore, the ideal pure invalid SOA curve was modeled as a line with slope 0. In summary, ideal valid (and false invalid) curves were hypothesized to have a distinct rising signature, while ideal invalid (and false valid) curves were modeled with zero slope.

Predicted results

Three probability conditions were presented: 100%, 75%, and 50%. Accuracy was measured over 6 SOAs between 0 and 300 msec. Accuracy-SOA curves were generated for valid and invalid trials A plausible technique for testing switching models is to detect the presence of false valid trials in the invalid curves, by looking at the way the invalid curve profile changes as a function of cue validity. Another would be to detect the presence of false invalid trials in the valid curve pattern. For switching in

a two-location display, using as a parameter 1-*c*, where *c* is the proportion of trials the cue is used, the proportion of false valid trials would be the following:

$$FV = (1-c)(1-p)$$

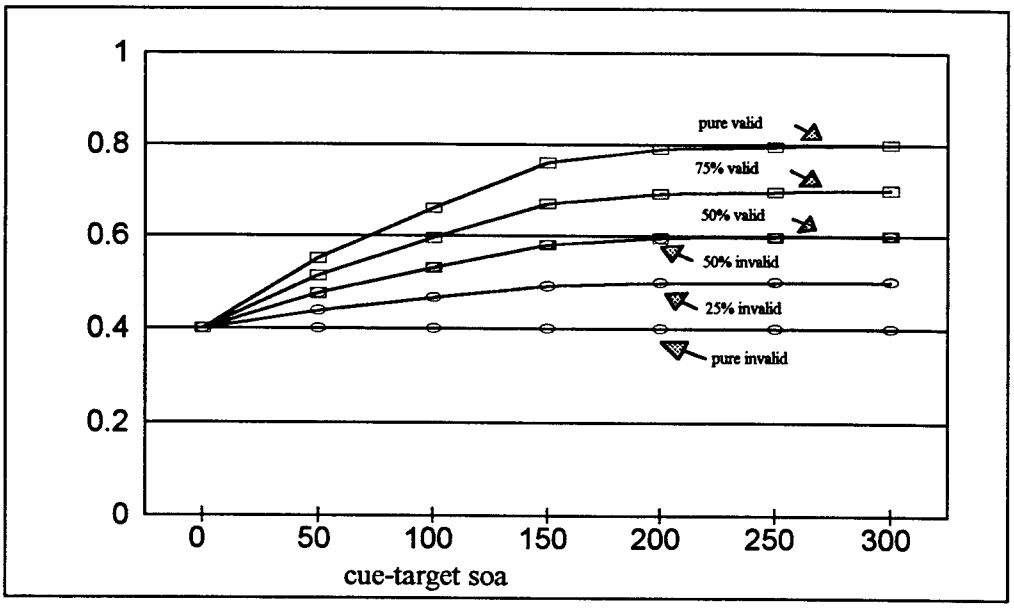
$$FV|(notP) = 1-c$$

where *p* is the proportion of valid trials and P is the event "valid trial". Assuming that subjects attempt to match probabilities, *c* will be expected to differ across conditions, and since invalid curves would be a weighted average of pure invalid and false valid curves, the shapes of the invalid curves are predicted to differ with a switching mechanism. Similarly, the proportion of false invalid curves among valid curves would be expected to be *c* (see Figure 2).

A possible technique for analysis would be to assume that data are a mixture of two types of trials, corresponding to events *C* and *notC*. Valid trials would be assumed to contain 1-c false invalid trials, and invalid trials would be assumed to contain c false valid trials. The false invalid signature will be assumed to be a line with zero slope and an intercept to be determined by the 0 SOA value for the valid and invalid curves (if they are close in value). The false valid signature will be modeled as the 100% valid curve. The basic technique will be to take the data and "subtract out" the false valid/invalid curves, and look at the resultant curves. For example, given conditions of 75% and 50%, and the assumption of probability matching, it could be assumed that valid curves contain c false invalid trials, where $c = p$. Those trials could be subtracted out from the data. A result consistent with switching would be that the resultant curves would be identical to each other and identical to the 100% valid curve. Similarly, resultant invalid curves (after subtraction of false valid curves) could be compared to each other and to the hypothesized pure invalid curve. The null hypotheses will be that the resultant curves do not differ, and rejection of the null hypothesis would indicate that observers do not use a switching strategy.

Specific predictions are as follows: (1) valid curves will lie above invalid curves and the shapes of valid and invalid curves will differ; (2) the valid curves will be ordered 100%, 75%, and 50%, with the 100% highest; (3) the invalid curves will be ordered 50%, 75% with the 50% higher. If the accuracy-SOA curves have the predicted pattern, then they will be analyzed for switching/sharing.

Figure 2. Hypothesized accuracy-SOA curves resulting from a probability-matching (switching) strategy.



Method

Observers

Three naive female observers between the ages of 17 and 25, with normal vision, were paid for their participation.

Apparatus

Stimuli were displayed on an IBM-XT with an EGA color monitor running at 60 mHz. An adjustable head and chin rest fixed the eye-to-screen distance at approximately 37 cm. Eye movement was monitored with a video camera. Responses were recorded on the numeric keypad of a standard IBM keyboard.

Stimuli

Targets consisted of "T"s (0.75 °) in one of four orientations (pointing right, left, up, or down); target luminance was 80 cd/m^2. Targets could appear in one of two locations 6° from a central fixation point; their possible locations were at 3 and 9 o'clock. Stimuli were presented as white pixels on a dark gray background.

Procedure

The order of events in each condition was identical (see Figure 1). A fixation bar (0.15°) appeared for 668 msec, followed by a 16.7 msec central arrow cue pointing to either the left or right location. The cue was followed by a target at one location, and a "plus" sign at the other location, after a variable SOA. SOAs used were (rounded to the nearest msec) 16, 50, 100, 150, 200, 250, and 300 msec. Following the presentation of the target, a mask (an outline of all possible targets) was presented at both locations. The observer's task was to indicate, by pressing the corresponding arrow on the numeric keypad, the direction in which the leg of the "T" target was pointing. The observer was instructed that accuracy, and not speed, was required.

There were three blocked conditions that consisted of different percentages (probabilities) of valid cues; observers were informed of what condition was being run. In the 100% condition, the cue always indicated the correct position of the target. In the 75% condition, the cue was valid (indicating the correct target location) on 75% of the trials (75% condition valid), but on 25% of the trials the target would appear at the uncued location. The 50% condition was similar to the 75% condition except for the probabilities. In order to avoid confusing terminology, the valid conditions will be referred to as 75% valid and 50% valid, but the invalid conditions will be referred to as "75% condition invalid" (which consists of 25% of the trials in the 75% blocks) and "50% condition invalid" (which consists of 50% of the trials in the 50% blocks).

Each observer was run for 2 to 4 sessions in the 100% condition for training purposes. Target durations were 33 or 50 msec depending on observer skill. There were an equal number of sessions at the

three probabilities, and equal numbers of trials for each SOA. Cue and stimulus location, direction of target, and target duration were randomized within blocks. Total number of trials per observer per condition per SOA ranged from a maximum of 448 trials per point in the 100% condition to a minimum of 112 trials per point in the invalid condition of the 75% blocks. Proportion of correct responses was calculated for each SOA for valid and invalid conditions, for both individual observers and combined observers.

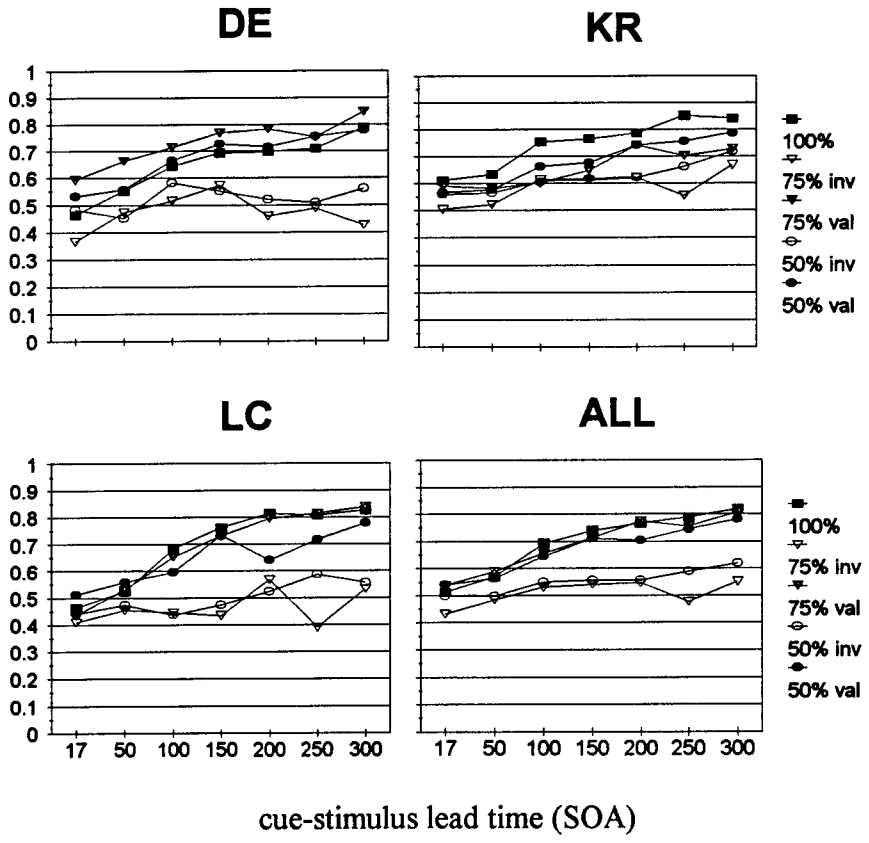
Results

Accuracy by SOA is illustrated in Figure 3 for each condition: 100%, 75% valid, 75% condition invalid, 50% valid, and 50% condition invalid. Data are illustrated in Figure 3 for individual observers DE, KR, and LC, and for observers combined (ALL). One finding that replicates earlier studies is that valid curves lie above invalid curves. It was predicted that 100% accuracy would be above 75% and 50% valid curves; however, only subject KR showed that pattern. None of the predictions concerning the relationships between 75% and 50% valid curves, and 75% and 50% invalid curves, were supported.

Discussion

The data do not support the predictions and do not replicate the findings of Gottlob et al. (1994). It appears that observers do not change their strategies as a function of probability of valid cues. It may be that not enough trials were run to get reliable accuracy-SOA curves. Since no allocation strategies are apparent, performance cannot be analyzed for switching vs. sharing. Since this experiment is one in a series that is exploring attentional allocation, more investigation is needed to determine the reason for the lack of evidence for varying the allocation of attention.

Figure 3. Data for individual observers and for observers combined. Vertical axis is accuracy (percentage correct).



References

Allport, A. (1992). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance* XIV (pp. 183-218). Cambridge: MIT Press.

Cheal, M. L., & Lyon, D. R. (1989). Attention effects on form discrimination at different eccentricities. *Quarterly Journal of Experimental Psychology: A. Human Experimental Psychology, 41A,* 719-746.

Cheal, M. L., Lyon, D. R., & Gottlob, L. R. (1994). A framework for understanding the allocation of attention in a location-precued discrimination. *Quarterly Journal of Experimental Psychology: A. Human Experimental Psychology,* 47A, 699-739..

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

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 448-456.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225-240.

Eriksen, C. W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583-597.

Gottlob, L. R., Cheal, M. L., & Lyon, D. R. (1994). Cue validity manipulation in a location-cuing task: sharing versus switching models for attention. Manuscript submitted for publication.

Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology, 34,* 103-112.

Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society, 21,* 247- 250.

LaBerge, D. & Brown, V. (1986). Variations in the size of the visual field in which targets are presented: An attentional range effect. *Perception & Psychophysics, 40,* 188-200.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Sperling, G. W. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds), *Varieties of attention* (pp. 103-181). London: Academic Press.

ITS EVALUATION: A REVIEW OF THE PAST AND RECOMMENDATIONS FOR THE FUTURE

Jennifer L. Greenis
Graduate Student
Department of Industrial and Organizational Psychology

Michigan State University
Department of Psychology
129 Psychology Research Building
East Lansing, MI 48824-1117

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

ITS EVALUATION: A REVIEW OF THE PAST AND RECOMMENDATIONS FOR THE FUTURE

Jennifer L. Greenis
Graduate Student
Department of Industrial and Organizational Psychology
Michigan State University

Abstract

Scientists and educators are expecting great things from Intelligent Tutoring Systems (ITSs). Over the past twenty years, many intelligent tutors have been developed. Unfortunately, evaluations of these tutors are scarce and have not supported many of the expectations expressed. Thus, researchers need to critically examine the quality of ITSs, as well as the research evaluating these systems, to learn how system development and evaluation can be improved. The purpose of this paper is to review recent ITS evaluations, examine how ITS evaluation has changed over time, and make recommendations to guide future research. First, the architecture and boundaries of ITSs are defined, and potential instructional, outcome, and administrative benefits are delineated. ITS evaluations are then reviewed and analyzed with respect to their research questions, methodologies, and criteria for effectiveness. Empirical evidence in support of ITS efficacy is summarized. Lastly, recommendations are given to guide future ITS evaluation. In the future, ITS researchers should (1) evaluate factors which contribute to the efficacy of an ITS (2) utilize process measures, as well as wider range of outcome measures, (3) evaluate transfer validity in addition to training validity, (4) consider ITSs from a systems perspective, and (5) utilize a systematic approach for evaluation and development.

ITS EVALUATION: A REVIEW OF THE PAST AND RECOMMENDATIONS FOR THE FUTURE

Jennifer L. Greenis

Introduction

Scientists and educators are expecting great things from Intelligent Tutoring Systems (ITSs). These systems offer many potential advantages over traditional forms of instruction such as self-paced instruction, immediate and individualized feedback based on cognitive models of learning, interactive learning, efficient (re)training capability, automatic testing and scoring, and learner progress summaries. Thus, their potential for improving learning and transfer of learning to work contexts is immense. Over the past twenty years, many intelligent tutors have been developed. Unfortunately, evaluations of these tutors are scarce (Shute & Psotka, 1994) and have not supported many of our expectations. It is critical that researchers examine the quality of ITSs and the research evaluating these systems to learn how system development and evaluation can be improved. To address this need, the paper will (1) delineate the architecture, boundaries, and potential benefits of ITSs, (2) review and assess the quality of ITS evaluations, and (3) make recommendations for future ITS development and evaluation.

What is an Intelligent Tutoring System?

An ITS is a training system which attempts to detect, respond to, and anticipate the individualized needs of each learner while executing a teaching plan. It should distinguish between content and form of instruction, so that it can generate different presentations of each subject matter unit as needed and in the form which is most beneficial for each learner (Ohlsson, 1986). Much of the theoretical foundation for ITS development has been based on literature concerning mastery learning, aptitude-treatment interaction, cognitive psychology, artificial intelligence, and computer-aided instruction (CAI) (Regian & Shute, 1992; Shute & Psotka, 1994). The first two areas of research, mastery learning and aptitude-treatment interaction, are driven by the belief that individually tailored instruction is superior to group-oriented instruction. This is a key assumption of ITSs. Cognitive psychology examines issues of representation and organization of knowledge types in human memory and benefits the development of ITS by addressing the nature of errors. Furthermore, the introduction of artificial intelligence technology into the field of CAI prompted development of intelligent computer-aided instruction (ICAI) (Regian & Shute, 1992), also referred to as ITS (Sleeman & Brown, 1982).

Boundaries of ITS: What is "Intelligence"?. So, what is an ITS and what is not? In other words, how can we differentiate between intelligent and non-intelligent computer-aided instruction? Regian and Shute (1992) conceptualize a continuum of computer-based training ranging from CAI to ICAI. It is important *not* to view this continuum as a progression from worse-to-better instruction (Shute & Psotka, 1994). The instructional method chosen should correspond to the learning situation and curriculum; in many learning situations, the use of an ITS would be overkill and a more appropriate method should be chosen. Virtually all computer-based training systems along this continuum are self-paced and individualized through branching routines. However, more "intelligent" computer-based training systems utilize a more powerful approach to individualization which seeks to encode knowledge, rather than decisions. CAI and ICAI can also be distinguished by the degree to which their instruction is tailored to the student's cognitive model (Regian, 1991; Regian & Shute, 1992; Shute & Psotka, 1994). In summary, two components make up the intelligence in an ITS: diagnosis and remediation. That is, an ITS must be able to (1) accurately diagnose students' knowledge structures, skills, and learning styles using principles (not pre-programmed responses) to select appropriate responses and (2) adapt instruction accordingly (Shute & Psotka, 1994; Sleeman & Brown, 1982).

System Architecture. Perhaps the biggest strength of an intelligent tutor is its ability to provide better training by providing feedback on an individual basis. Feedback is a crucial element of instruction because it (1) allows trainees to identify skilled performance and correct errors before they become habitual, (2) facilitates acquisition and transfer of knowledge and skills, and (3) increases trainee motivation (Kozlowski, Ford, & Smith, 1993). Individualized, adaptive instruction (e.g., specific feedback) is typically achieved through four interacting components, namely, the student model, expert model, instructional module, and interface (Kline, 1988; Nwana, 1991; Regian, 1991). This constitutes the general architecture of an ITS, however, the presence and magnitude of these components varies extensively among systems (Nwana, 1991).

The purpose of the student model is to assess the learner's knowledge state and hypothesize the learner's conceptions and strategies (Gisolfi, Balzano, & Dattolo, 1993). Ideally, the student model should provide an internal representation of the student detailing the student's level of experience, learning style, cognitive limitations and strengths, verbal and spatial abilities, and mental model of the system (Norcio & Stanley, 1989). The student model is obtained through cognitive diagnosis, the dynamic evaluation of a learner's cognitive state using principles rather than preprogrammed responses to individualize instruction (Regian & Shute, 1992). The expert model contains the knowledge to be taught (Kline, 1988). Once the system has diagnosed how the student learns and thinks, the obtained student model can be compared to the expert model in order to determine what the student does and does not know (Kline, 1988). Based on this comparison, appropriate feedback can be provided. The instructional module explicates the most appropriate teaching strategies and determines the timing and presentation of the learning material. This module dictates the degree

of advice, support, explanation, and control given to the learner (Nwana, 1991), with the end-goal being a match between the student's cognitive processes and the domain knowledge (represented by the student and expert models) (Farquhar & Orey, 1993). The fourth component, the user interface, provides the methods by which the student interacts with the ITS to solve domain problems (Burger & DeSoi, 1992; Regian, 1991). It should promote clear communication between the system and the student (Farquhar & Orey, 1993). Output interface methods include computer-generated graphics and text and speech synthesizers, while possible input devices include a mouse, keyboard, and joystick (Regian, 1991).

Potential Benefits. What makes an ITS distinctive from other methods of delivery? In other words, what does an ITS enable the learner to do or do better (from a learning, instructional, or training perspective) than other media or methods of instruction? These questions are addressed from an ideal perspective by discussing potential instructional benefits, (learning) outcome benefits, and administrative benefits. Following this section, ITS evaluations are reviewed so that the degree to which these benefits have been realized can be examined.

The term **instructional benefits** is used to describe advantages derived from the instructional environment of an ITS (e.g., individualized feedback, self-pacing, and learner engagement). Overall, ITSs provide a rich practice environment characterized by self-paced learning, high quality interaction, and ample opportunity to practice. ITSs have more fidelity and can visually show examples that other media can not demonstrate as well. The quantity and quality of the feedback provided is superior. Based on the student, expert, and teaching modules, feedback is individualized (to the student's learning style, knowledge level, etc.), immediate, relevant, expert, and process-oriented. Another important instructional benefit is that ITSs "engage the learner in the learning process" (by eliciting more practice, providing continuous feedback, challenging the learner to think, etc.). Learner engagement has been defined as the "percentage of time devoted to learning in which the student is actually on-task and engaged with the instructional materials and activities being presented" (Borich, 1989, p. 4). While traditional methods often cast the learner in a passive role (e.g., listening or watching), the interactional style of these systems provides a more active learning environment which stimulates learner engagement. Furthermore, by interacting one-on-one with the system, individuals can regulate their own learning pace. This results in more efficient and superior learning. Since instructors are often challenged with extremely diverse classes, individualized feedback and self-paced learning allows for better and easier handling of variance in student abilities and interests (e.g., instructors can concentrate on the more "needy" learners).

As a result of these instructional benefits, learners should experience many **outcome benefits**. This term is used to describe cognitive, skill-based, and affective benefits resulting from the use of ITSs. More specifically, the rich practice environment of ITSs (described above) should result in a higher motivation to learn, more enjoyable learning for many individuals, improved learning and transfer (on-the-job performance), and overall, increased organizational productivity. Lastly, ITSs have several **administrative benefits**. ITSs are

very efficient; they can adapt to each learner's pace and effectively teach the same amount of material in less time than other instructional methods. ITS efficiency has implications for quicker job re-training in a society presently characterized by down-sizing and frequent job turnover. ITSs are advantageous to use because they are mobile and replicable. When a good ITS is designed, it can be replicated and distributed where needed. This is not the case for human instructors (at least until the genetic engineers find a way). Furthermore, these systems are consistent and unbiased. Unlike humans, they do not have "bad days" or make personal judgements (e.g., "You are a slacker"). Lastly, they permit performance data to be collected unobtrusively. This is extremely advantageous to instructors tracking student progress, as well as researchers gaging the effectiveness of these systems.

Traditional methods of instruction (e.g., lecture, programmed instruction, and on-the-job training) may also provide some of these benefits. Educators should consider the benefits of each instructional method and select one (or a combination) which will provide the best match between the curriculum, training objectives, and available resources. The true potential for ITS will be in areas where "uncertainty is high and information is combinatorially explosive" (Burns & Parlett, 1991; p. 6). Given these learning situations in combination with limited resources and time constraints, intelligent tutoring systems hold the most promise for increased rate and quality of knowledge and skill acquisition and transfer.

A Review of ITS Evaluations

Now that the potential benefits of ITSs have been discussed, the extent to which these benefits have been evidenced when evaluating ITSs should be addressed. Given the relative newness of intelligent tutors, evaluation is necessary to determine their "success" at enhancing training and on-the-job performance and to guide future design. Evaluations should investigate the efficacy of an ITS (Shute and Regian, 1993) to assess if the system teaches what it was intended to teach, to what degree, in comparison to what, and at what cost. Both learning and transfer outcomes should be evaluated so that training and transfer validities can be assessed (Goldstein, 1993). To justify future ITS development, researchers need to demonstrate that these tutors are either (1) more effective at learning and transfer than traditional instruction, or (2) equally effective but more efficient in terms of cost and time. This section opens with a brief discussion of early ITS evaluation studies. Following, a review of more recent ITS evaluations is conducted delineating the purpose of each system, evaluation methods (i.e., sample, independent variables, and dependent variables), and results (see Table 1). Generalizations are made concerning the foci, methodologies, and effectiveness criteria characteristic of ITS evaluation; subsequently, conclusions about the effectiveness of ITSs are made based on evaluation results.

Early ITS Evaluation. "By the mid-1980's, the development of tutors greatly exceeded their evaluations" (Shute & Psotka, 1994). This is not surprising, given that early ITS projects focused on the technical aspects of the system rather than on instructional features. When Fletcher (1988) conducted a review

of nine ICAI systems developed by the military, he found virtually no information concerning the effectiveness of the training systems. Another review of early ITS was conducted by Legree and Gillis (1991). ITSs reviewed include: (1) Proust, a Pascal programming tutor, (2) West, a mathematics game tutor, (3) Pixie, an algebra tutor, (4) MACH III, a troubleshooting tutor for radar mechanics, (5) LISP, a LISP programming tutor, and (6) Smithtown, a discovery world that teaches scientific inquiry skills in the context of microeconomics. The review revealed mixed support for the potential of intelligent tutoring systems. Group differences were not demonstrated when comparing intelligent tutors to human tutorial and control conditions for the Proust and West (Center for the Study of Evaluation, 1986) and Pixie tutors (Sleeman, Kelly, Martinak, Ward, & Moore, 1988 and 1989). However, these evaluation studies are criticized for their use of limited instructional interventions (two problems for Proust, one hour for West, and five class periods for Pixie) and small sample sizes (on average, 10 subjects per group). These factors serve to undermine the studies' statistical power and may account for the non-favorable evaluations (Legree & Gillis, 1991). Evaluations of the LISP, MACH III, and Smithtown systems did yield favorable evaluations (Anderson, Boyle, & Reiser, 1985; Kurland, Granville, & MacLaughlin, 1990; Raghavan & Katz, 1989). Despite similarly small samples, these studies involved more extensive ITS interventions (32 hours for MACH III, one semester at a university for the LISP tutor, and five hours for Smithtown). As a result, subjects trained on ITSs demonstrated better performance on knowledge-based tests (approximately one standard deviation) and quicker learning times.

Based on the review, Legree and Gillis made several recommendations for future ITS evaluation. First, evaluations should include three experimental conditions (traditional classroom control, human tutorial, and computer tutorial) in order to allow for performance data comparisons. Second, researchers should evaluate extensive systems (i.e., systems expected to have a large impact on performance by virtue of its wide scope) and utilize much larger sample sizes (i.e., more than 34 subjects per group). Furthermore, researchers should describe in detail the measures used to evaluate effectiveness. Information on reliability, validity, effect size, and variance estimates is critical, especially when results indicate no significant differences between groups.

Recent ITS Evaluation. Shute and Psotka (1994) recently conducted a comprehensive review of the ITS literature. Their paper provides an excellent overview of the foundations, components, and history of ITS, as well as future issues for research and development. Evaluations of six intelligent tutors were described. These tutors include the Geometry tutor (tutor for geometry proofs), Sherlock (avionics troubleshooting tutor), Bridge (Pascal programming tutor), *Stat Lady* (statistics tutor, probability lesson), and Smithtown and the LISP tutor (mentioned above). They conclude that: (1) there is a need for more systematic, controlled ITS evaluations and a standard approach for system design and assessment, and (2) intelligent tutors do "accelerate learning with, at the very least, no degradation in outcome performance compared to appropriate control groups" (p. 35).

However, specific recommendations for future evaluation could not be made since the quality of the studies was not assessed.

In order to appraise the quality of ITS evaluations and make recommendations for future evaluation, the author of this paper reviewed ten studies which assessed a total of seven intelligent tutors. This review is selective, focusing primarily on recently published studies (i.e., research published after 1990). Early evaluations of the LISP tutor and Smithtown were included to assess changes in ITS evaluation over time. To facilitate study comparisons and generalizations, key information is provided in table format (studies will not be addressed individually within the body of the paper). Table 1 specifies the purpose of each system and the evaluators, characteristics of each study (i.e., the sample size and independent and dependent variables), and the results. Systems reviewed include the LISP tutor, Sherlock, Smithtown, and *Stat Lady* (mentioned above); Loader (teaches console-operations in the context of a railroad yard); a Flight Engineering tutor (teaches flight engineering skills); and an Electricity tutor (teaches principles of electricity).

ITS Evaluation: Some Generalizations. From Table 1, conclusions about the characteristics and quality of ITS evaluation (in terms of the research foci, methodologies, and criteria for effectiveness utilized) can be drawn. After these generalizations are stated, conclusions concerning the efficacy of ITSs are made. The first conclusion is that the focus of ITS evaluation (i.e., research questions addressed) has changed over the past decade. Early evaluation investigated whether an ITS was "effective" (i.e., demonstration studies) or more effective than another instructional method (i.e., benchmark studies). The primary (and often only) independent variable manipulated was the instructional method (referred to as "tutor type" in Table 1). In contrast, recent studies have moved beyond simple demonstrations of effectiveness to investigate the effects of different ITS environments and individual characteristics on ITS effectiveness. Amount of instruction (number of problems given), type of feedback, and provision (or absence) of a dynamic learning model have been manipulated to determine the effectiveness of various ITS environments. Individual characteristics such as working memory capacity, general knowledge, cognitive ability, and exploratory behavior have also been studied, illustrating that aptitude-treatment interactions are currently an important focus of ITS evaluation. By focusing on these types of evaluation questions, researchers are likely to learn more about what makes an ITS effective and the circumstances under which an ITS is more effective. The second conclusion drawn is that researchers are beginning to use much larger sample sizes. As pointed out by Legree and Gillis (1990) and as seen in Table 1, early evaluations (e.g., the LISP tutor and Smithtown) used sample sizes of approximately 30 subjects (i.e., about 10 subjects per group). Recent studies, however, have utilized much larger samples (ranging from 168 to 311 subjects in Table 1); thus, they have more statistical power to detect significant differences (e.g., between instructional methods). Also, most studies which evaluate the effects of instructional method on learning

outcomes and efficiency now include three conditions (e.g., ITS, traditional lecture, and control) as recommended by Legree and Gillis (1991).

Furthermore, the types of measures used to evaluate system efficacy have not changed much over time (in complexity and diversity). Researchers typically focus on cognitive and skill-based outcome criteria, as opposed to affective outcome criteria and process criteria, and only utilize a small portion of the potential, relevant measures. Conclusions of system efficacy are often based on simple learning outcome and efficiency measures (i.e., pre- and post-knowledge tests, the number of learner errors, and completion time). Unfortunately, these measures only "scratch the surface" in determining the true effectiveness of ITSs. Some recent studies have developed and utilized more specific outcome tests (i.e., measures of declarative knowledge and procedural skill) (Shute, 1992; Shute, 1993b) based on Anderson's (1983) stages of cognitive skill acquisition, and one study investigated affective outcomes (i.e., the enjoyability and perceived helpfulness of an ITS) in addition to learning outcomes (Shute & Gawlick-Grendell, in press). In general, these measures are used to assess training validity; evaluation of transfer validity is rare or non-existent. Shute (1992, 1993b) evaluated learners' abilities to generalize knowledge and skills beyond what was explicitly instructed by an electricity tutor (i.e., transfer of knowledge and skills to new problems). This is a step in the right direction; however, assessment of transfer to relevant work and/or educational environments is even more critical when considering the efficacy of an intelligent tutor (or any instructional method). If an avionics troubleshooting tutor improves training performance, but not on-the-job performance, then the tutor has ultimately failed.

Several additional generalizations concerning ITS evaluation can be drawn from the literature (but not from Table 1 since space limitations precluded this information). First, ITS evaluations are often conducted in laboratory settings. Typical subjects are Air Force/military recruits and individuals hired from temporary employment agencies. Studies conducted in educational environments (e.g., high school or university) to evaluate effects of ITS on student learning are the exception. Another conclusion is that recent studies evaluate more "extensive" ITSs (i.e., systems expected to have a large impact on performance as a result of more robust intervention). Subjects received 20 hours of instruction from Sherlock (Nichols et al., in prep), up to 30 hours from Bridge (Shute, 1991), and 45 hours from the Electricity tutor (Shute, 1993b). Also, between-subjects designs are consistently used to compare instructional methods and different ITS environments so no subject is exposed to more than one method or environment. Lastly, researchers are beginning to provide more thorough descriptions of their ITS evaluation studies. For example, instead of stating that a "knowledge test" was used to evaluate changes in knowledge acquisition (Anderson, Boyle, & Reiser, 1985), researchers now describe the content of the test (e.g., declarative knowledge) and how the knowledge or skill was measured (e.g., 15 item multiple choice test given on the computer) (Shute, 1992; Shute & Gawlick-Grendell, 1994). However, reporting scale reliabilities, validities, and effect sizes has not become a common practice; researchers should include these statistics so that their effects can be assessed and conclusions about ITS effectiveness can be drawn across studies. Overall, the quality of ITS evaluation is improving. Researchers have begun to utilize larger sample

sizes and more complex outcome criteria, evaluate more extensive systems, and ask more specific, relevant research questions.

Now that the characteristics and quality of ITS evaluation have been described, empirical support of ITS efficacy will be addressed. Evaluation results are mixed, but there is some support for outcome benefits and administrative advantages of ITSs. Instructional benefits of ITSs, on the other hand, have been neglected in formal evaluations (they are typically assumed), so justified conclusions to whether ITSs provide better feedback, more feedback, more learning engagement, etc. can not be made as of yet. In support of learning outcome benefits, studies generally demonstrate that ITSs lead to knowledge and skill acquisition equivalent to, and sometimes better than, other instructional methods. With regard to administrative advantages (e.g., learning efficiency), studies demonstrate that ITSs are often more efficient (i.e., can bring learners to mastery quicker) than other instructional methods. One of the most important conclusions is that the effectiveness of ITSs (and different environments within ITSs) varies with individuals and learning outcomes. Studies investigating aptitude-treatment interactions have found that some individuals (e.g., learners characterized by high cognitive ability and exploratory behaviors) learn better from intelligent computer-based training while others (e.g., individuals characterized by low cognitive ability and less exploration) learn better from human-based training. There is also evidence that ITSs are more effective for teaching particular skills or knowledge (e.g., declarative knowledge), while human tutors are better at imparting other skills or knowledge (e.g., procedural skills). This implies that learning may be maximized by utilizing ITSs in combination with different instructional methods (e.g., human tutors). More research is needed to investigate aptitude-treatment interactions and "outcome-treatment interactions."

Despite support for the efficacy of ITSs, current systems are limited in the form and extent of domain knowledge possible (Acker, Lester, Souther, and Porter, 1991). That is, the typical system currently focuses on a single task and covers only a small portion of the domain knowledge. This limits a system's ability to generate coherent explanations which, in return, limits the impact on learning and transfer. Although early ITSs were criticized for their lack of good theoretical foundation and the limited extent to which they provided specific feedback, allowed exploratory behavior, and adapted to different learners (Sleeman, 1984), current ITSs have addressed and overcome many of these limitations.

Recommendations for Future ITS Evaluation

Based on the above review, it is clear that the quality of ITS evaluations is improving, yet there is still room for improvement. Since poorly planned or conducted evaluations may lead to false conclusions concerning the effectiveness of ITSs, researchers must address the quality of ITS studies. Based on the current state of ITS evaluation, five recommendations to improve ITS evaluation are made. Future ITS research should (1) evaluate factors which make an ITS effective or ineffective, (2) employ process criteria, as well as more diverse outcome criteria, (3) evaluate transfer of learning, (4) consider ITSs from a systems perspective, and (5) utilize a

systematic approach for evaluation and development. Improving the quality and quantity of ITS evaluations will lead to improvements in the design of ITSs, and thereby magnify the instructional, outcome, and administrative benefits achieved.

1. Studies should evaluate WHY an ITS is effective (or ineffective). Over ten years ago, Sleeman and Brown made a call for research to evaluate not only whether an ITS is effective, but also "in what ways it is effective and why, and in what ways it is ineffective and why" (p. 9). Today, their call remains for the most part unanswered. Although outcome benefits (e.g., knowledge and skill acquisition) and testing benefits (e.g., learning efficiency) have been evaluated, instructional benefits have typically been assumed rather than formally evaluated. Process measures (i.e., criteria that help determine the source of the instructional effect) should be used to evaluate the instructional benefits of ITSs in order to determine factors affecting the ITS effectiveness (or ineffectiveness). That is, research should investigate the extent to which ITSs promote instructional benefits such as quality feedback, learner engagement, self-pacing, sustained attention, learning guidance, etc.; this is especially important when a system is evaluated unfavorably. If it is found that an ITS fails to sustain learners' attention, then its teaching strategies (instructional module) should be modified. Similarly, if a system fails to provide accurate individualized feedback, then the expert and student models should be reassessed. By evaluating these factors in addition to learning outcomes, researchers can discover WHY a system is effective or ineffective and then utilize this knowledge to improve ITS development.

Several studies reviewed did evaluate the effects of different types of feedback and different amounts of instruction given by ITSs. These will shed light on how ITSs can be designed more effectively. Currently, an evaluation of *Stat Lady* is in progress to test the efficacy (learning outcomes and efficiency) and utility (cost-benefit) of the student modeling approach. This experiment compares an intelligent (i.e., has a student model) version of *Stat Lady* to a non-intelligent version. The "intelligent" Stat Lady has a student model which contains models of symbolic knowledge, procedural skill, and conceptual knowledge (Shute, 1994). Studies like this are greatly encouraged. Nevertheless, studies comparing different instructional attributes *across* instructional methods (instead of within) are needed. This would allow researchers to determine ways in which ITSs may be more effective and ways in which other instructional methods may be more effective, so that optimal instruction may be designed (possible utilizing a combination of various methods).

2. Studies should include process measures, as well as a wider range of outcome measures. In order to determine more about the efficacy of ITSs, researchers should employ more process measures and more diverse outcome measures (e.g., affective measures). Regian and Shute (1992) suggest a number of outcome measures to use when evaluating the effectiveness of intelligent tutors. Some of these include declarative and procedural knowledge, performance latency and accuracy, near and far transfer, skill retention and decay, automatic skill,

and higher-order knowledge. They stress the importance of selecting dependent measures which reflect the goals of the ITS and the evaluation study, and using multiple dependent measures. From the literature review, it is clear that only a handful of these potential outcome criteria have been used to evaluate intelligent tutors, and the measures typically used (e.g., multiple choice tests) may not adequately capture the process of knowledge and skill acquisition.

Kraiger, Ford, and Salas (1993) propose cognitive, skill-based, and affective learning outcomes relevant to training and recommend potential evaluation measures. Categories of cognitive learning outcomes include verbal knowledge, knowledge organization, and cognitive strategies; potential evaluation methods include recognition and recall tests, structural assessment, and probed protocol analysis. Categories of skill-based learning outcomes include compilation and automaticity; some potential measures include structured situational interviews and secondary task performance. Affective learning outcomes include attitudinal and motivation categories which can be measured with self-reports. The quality of ITS evaluation would be greatly improved by utilization of their classification scheme of learning outcomes.

Royer, Cisero, and Carlo (1993) also present a classification scheme of cognitive skill assessment procedures based on Anderson's (1982) ACT* theory of cognitive skill development and the work of Glaser, Lesgold, and Lajoie (1985) on dimensions of cognitive skill. They stress that researchers should assess cognitive skills based on the stage of skill development (i.e., declarative stage, knowledge compilation stage, and procedural stage), not whether or not skills were acquired. Many possible measures for each dimension of cognition (i.e., knowledge organization and structure, depth of problem representation, quality of mental models, efficiency of procedures, automaticity of performance, and metacognitive skill for learning) are proposed. Future evaluations of ITSs should take advantage of these classification schemes and potential measures when selecting outcome criteria for effectiveness. As noted by the work of Kraiger, Ford, & Salas (1993), affective outcome measures should be included in addition to cognitive and skill-based measures. Moreover, future evaluation should utilize process measures. Although often overlooked when evaluating ITSs, process criteria are just as critical as outcome criteria, if not more important, since strict reliance on outcome measures makes it extremely difficult to determine why an instructional method is successful or unsuccessful (Campbell, 1988).

3. Studies should evaluate the transfer of learning following training. Furthermore, researchers should evaluate the transfer validity of ITSs in addition to evaluating training validity. Positive transfer of training is defined as the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job. Transfer implies that learned behavior is generalized to the job context and maintained over time on the job. Researchers need to determine how ITSs improve on-the-job performance (or do they?). This is the ultimate test of effectiveness; if a system contributes to learning, but the learning does not transfer to the desired situations, then the system has failed (Baldwin & Ford, 1988). The lack of transfer evaluation is not surprising since researchers have only recently begun to evaluate ITSs. Thus, their first concern is that these

systems improve learning during training. Results have supported the training validity of ITSs; it is now time to address transfer evaluation issues.

4. Studies should evaluate ITSs from a systems perspective. Future research should consider ITSs from a systems perspective. Training researchers now attend to pre- and post-training environments as important determinants of training effectiveness and view trainees as active participants in the system who interact with the environment before, during, and after training (Tannenbaum & Yukl, 1992). Thus, researchers need to consider the role of the instructor evaluating and developing ITSs. ITS should allow instructors more opportunities to intervene in the learning process (Burns & Parlett, 1991). The notion of teacher-proof CAI systems is impractical and unbeneficial, even for ITSs. Designers, operators and researchers of ITSs should incorporate human teachers into the evaluation and decision-making process (Hativa, & Lesgold, 1991). Given this view, studies which evaluate instruction based on a combination of human and intelligent computer tutors would be beneficial.

5. A systematic approach to ITS evaluation should be adopted. Lastly, in order to assess the merits of ITSs, a systematic approach to evaluation is needed. Several approaches have been recently offered (Shute & Regian, 1993; and Steuk & Fleming, 1990). Shute and Regian (1993) recommend seven steps to guide ITS evaluation. However, these principles apply to any evaluation study conducted, not just to ITS evaluation. They address how to conduct a good ITS evaluation, but do not address what issues should be evaluated. Steuk and Fleming (1989) present a taxonomy of issues (i.e., research, methodological, and life cycle issues) important to consider when evaluating ITSs. In addition to these approaches, ITS evaluation and development could benefit from a "process framework" which addresses the question, "Why is a cognitive tutor (more) effective?" In other words, if a cognitive tutor is better than another system or traditional instruction at teaching certain learning skills, then what aspects of the tutor are responsible for this increased effectiveness? As emphasized by Legree and Gillis (1991) researchers should report reliability, validity, effect size, and variance estimates so that more systematic evaluation and development can be conducted.

In conclusion, numerous studies comparing ITSs to traditional instruction have demonstrated that ITSs are as effective while less time-consuming. Nonetheless, these systems remain costly to develop, evaluate, and implement, and researchers have not consistently demonstrated that they are *more* effective than traditional instruction for advancing knowledge and skill acquisition or transfer. Recommendations to guide future ITS evaluation have been given. A final caution is that the effectiveness of ITSs should never be assumed; a system's effectiveness should only be determined after a carefully planned evaluation study has been conducted. As the quality of ITS evaluation improves, so will the quality of ITSs. It is only a matter of time before the instructional, learning outcome, and administrative benefits of ITSs are realized to their fullest potential.

References

Acker, L., Lester, J., Souther, A., & Porter, B. (1991). Generating coherent explanations to answer students' questions. In H. Burns, J.W. Parlett, and C.L. Redfield (Eds.), Intelligent Tutoring Systems: Evolutions in Design (pp. 151-176). Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1983). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

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

Baldwin, T.T., & Ford, J.K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63-105.

Borich, G.D. (1989, Aug.). Air Force Instructor Evaluation Enhancement: Effective Teaching Behaviors and Assessment Procedures. Tech. Rep. AFHRL-TP-88-55. Training Systems Division, Brooks AFB, TX.

Burger, M.L., & DeSoi, J.F. (1992). The cognitive apprenticeship analogue: a strategy for using ITS technology for the delivery of instruction and as a research tool for the study of teaching and learning. International Journal of Man-Machine Studies, 36, 775-795.

Burns, H., & Parlett, J.W. (1991). The evolution of intelligent tutoring systems: Dimensions of design. In H. Burns, J.W. Parlett, and C.L. Redfield (Eds.), Intelligent Tutoring Systems: Evolutions in Design (pp. 1-11). Hillsdale, NJ: Erlbaum.

Campbell, J.P. (1988). Training design for performance improvement. In J.P. Campbell, R.J. Campbell and associates (Eds.), Productivity in Organizations: New Perspectives from Industrial and Organizational Psychology (pp. 177-215). San Fransisco: Jossey-Bass Publishers.

Center for the Study of Evaluation (1986). Intelligent computer aided instruction (ICAI): Formative evaluation of two systems (ARI Research Note 86-29). Alexandria, VA: U.S. Army Research Institute. (DTIC No. AD-A167 910).

Farquhar, J.D., & Orey, M. (1993). Intelligent tutoring systems: Toward knowledge representation. Manuscript submitted for publication.

Farquhar, J.D., & Regian, J.W. (1993, May). The effects of a dynamic graphical model during simulation-based training of console operation skill. Proceedings of 1993 Conference on Intelligent Computer-Aided Training and Virtual Environment Technology, Houston, TX.

Fletcher, J.D. (1988). Intelligent tutoring systems in the military. In S.J. Andriole and G.W. Hopple (Eds.), Defense Applications of Artificial Intelligence: Progress and Prospects (pp. 33-59). Lexington, MA: Lexington Books.

Gisolfi, A., Balzano, W., & Dattolo, A. (1993, Jan.). Enhancing the learning process with expert systems. Educational Technology, xxxiii(1), 25-32.

Glaser, R., Lesgold, A., & Lajoie, S. (1985). Toward a cognitive theory for the measurement of achievement. In R.R. Ronning, J. Glover, J.C. Conoley, & J.C. Witt (Eds.), The influence of cognitive psychology on testing and measurement (pp. 41-85). Hillsdale, NJ: Erlbaum.

Goldstein, I.L. (1993). Training in Organizations (Third Ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Hativa, N., & Lesgold, A. (1991). The computer as a tutor- can it adapt to the individual learner? Instructional Science, 20, 49-78.

Kline, K.B. (1988). Intelligent systems for human resources. Aviation, Space, and Environmental Medicine, 59(11 sect. 2), 65-68.

Kozlowski, S.W.J., Ford, J.K., & Smith, E.M. (1993, May). Training concepts, principles, and guidelines for the acquisition, transfer, and enhancement of team tactical decision making skills I: A conceptual framework and literature review (Report No. IPPSR/ORG/PSY 93-1.1). Naval Training Systems Center, Orlando, FL.

Kurland, L.C., Granville, R.A., & MacLaughlin, D.M. (1990). Design, development, and implementation of an intelligent tutoring system (ITS) for training radar of mechanics to troubleshoot. Unpublished manuscript, Bolt, Beranek, and Newman, Systems and Technologies Corporation, Cambridge, MA.

Legree, P.J., & Gillis, P.D. (1991, Spring). Product effectiveness evaluation criteria for intelligent tutoring systems, Journal of Computer-Based Instruction, 18(2), 57-62.

Lesgold, A.M., Bonar, J., Ivill, J., & Bowen, A. (1989). An intelligent tutoring system for electronics troubleshooting: DC-circuit understanding. In L. Resnick (Ed.), Knowing and Learning: Issues for the Cognitive Psychology of Instruction. Hillsdale, NJ: Erlbaum.

Nichols, P., Pokorny, R., Jones, G., Gott, S.P., & Alley, W.E. (in preparation). Evaluation of an avionics troubleshooting tutoring system. Technical Report, Armstrong Laboratory, Human Resources Directorate, Brooks AFB, TX.

Norcio, A.F., & Stanley, J. (1989, April). Adaptive human-computer interfaces: A literature survey and perspectives. Transactions on Systems, Man, and Cybernetics, 19(2), 399-408.

Nwana, H.S. (1991). User modelling and user adapted interaction in an intelligent tutoring system. User Modeling and User-Adapted Interaction, 1, 1-32.

Ohlsson, S. (1986). Some principles of intelligent tutoring. Instructional Science, 14, 293-326.

Raghavan, K., & Katz, A. (1989). Smithtown: An intelligent tutoring system. Technical Horizons in Education Journal, 17(1), 50-54.

Regian, J.W. (1991). Representing and teaching high performance tasks within intelligent tutoring systems. In H. Burns, J.W. Parlett, and C.L. Redfield (Eds.), Intelligent Tutoring Systems: Evolutions in Design (pp. 225-241). Hillsdale, NJ: Erlbaum.

Regian, J.W., & Shute, V.J. (1992). Automated instruction as an approach to individualization. In J.W. Regian & V.J. Shute (Eds.), Cognitive Approaches to Automated Instruction (pp. 1-13). Hillsdale, NJ: Erlbaum.

Royer, J.M., Cisero C.A., & Carlo, M.S. (1993, Summer). Techniques and procedures for assessing cognitive

skills. Review of Educational Research, 63(2), 201-243.

Shute, V.J. (1991). Who is likely to acquire programming skills? Journal of Educational Computing Research, 7, 1-24.

Shute, V.J. (1992). Aptitude-treatment interactions and cognitive skill diagnosis. In J.W. Regian & V.J. Shute (Eds.), Cognitive approaches to automated instruction (pp. 15-47). Hillsdale, NJ: Erlbaum.

Shute, V.J. (1993a). A macroadaptive approach to tutoring. Journal of Artificial Intelligence and Education, 4(1), 61-93.

Shute, V.J. (1993b). A comparison of learning environments: All that glitters. In S.P. Lajoie & S.J. Derry (Eds.), Computers as Cognitive Tools (pp. 47-74), Hillsdale, NJ: Erlbaum.

Shute, V.J. (1994). Regarding the "I" in ITS: Student modeling. In T. Ottmann & I. Tomek (Eds.), Proceedings of Educational Multi-media and Hyper-media (pp. 50-57). Charlottesville, VA: Association for the Advancement of Computing in Education.

Shute, V.J., & Gawlick-Grendell, L.A. (in press). What does the computer contribute to learning? Computers in Education: An International Journal.

Shute, V.J., Gawlick-Grendell, L.A., & Young, R. (1993, April). An Experiential System for Learning Probability: *Stat Lady*. Paper presented at the American Educational Research Association, Atlanta, GA.

Shute, V.J., & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. Interactive Learning Environments, 1, 51-76.

Shute, V.J., & Psotka, J. (1994, May). Intelligent tutoring systems: Past, present, and future (Tech. Rep. AL/HR-TP-1994-0005). Armstrong Laboratory, Human Resources Directorate, Brooks AFB, TX.

Shute, V.J., & Regian, J.W. (1993). Principles for evaluating intelligent tutoring systems. In a special evaluation issue of: Journal of Artificial Intelligence & Education, 4(3), 245-271.

Sleeman, D.H., & Brown, J.S. (1982). Intelligent Tutoring Systems. London: Academic Press.

Sleeman, D., Kelly, A.E., Martinak, R., Ward, R.D., & Moore, J.L. (1988). Diagnosis and remediation in the context of intelligent tutoring systems (ARI Research Note 88-66). Alexandria, VA: Army Research Institute. (DTIC No. AD-A199 024).

Sleeman, D., Kelly, A.E., Martinak, R., Ward, R.D., & Moore, J.L. (1989). Studies in the diagnosis and remediation of high school algebra students. Cognitive Science, 13, 551-568.

Steuk, K., & Fleming, J.L. (1990). Intelligent Tutoring Systems: A Taxonomy of Evaluation Issues. Technical Paper ADHEL-TP-89-79. Training Systems Division, Brooks AFB, TX.

Tannenbaum, S., & Yukl, G. (1992). Training and development in work organizations. In M.R. Rosenzweig & L.W. Porter (Eds.), Annual Review of Psychology (Vol. 43). Palo Alto, CA: Annual Reviews, Inc.

Table 1: Review of ITS Evaluations

ITS	Evaluation Study	Method SS = total number of subjects	Results
LISP tutor: LISP programming skills	Anderson, Boyle, & Reiser (1985): STUDY 1	Sample: 30 SS Independent Variables: 1. Tutor type: computer tutor, human tutor, and learn on own Dependent Variables: 1. LISP knowledge test 2. Time	1) Learning outcomes: No sign. differences between all 3 groups on knowledge test 2) Learning efficiency: (time on tutor) Computer and human tutor SS not sign. different in completion times but faster than control SS (11.4, 15, 26.5 hrs)
	Anderson, Boyle, & Reiser (1985): STUDY 2	Sample: 20 SS Independent Variables: 1. Tutor type: computer tutor vs. learn on own Dependent Variables: 1. LISP knowledge test 2. Time	1) Learning outcomes: Computer tutor SS scored 43% better on knowledge test 2) Learning efficiency: (time on tutor) Computer tutor SS spent 30% less time doing lesson problems
Sherlock: Avionics troubleshooting	Nichols, Pokorny, Jones, Gott, & Alley (in prep)	Sample: 32 SS Independent Variables: 1. Tutor type: computer tutor vs. control (on-the-job training) Dependent Variables: 1. Paper & pencil/verbal knowledge test	1) Learning outcomes: Computer tutor SS performed sign. better than control SS and equivalent to experienced technicians on both verbal and paper & pencil knowledge tests

Electricity Tutor (MHO): Principles of electricity (i.e., Ohm's and Kirchoff's Laws)	Shute (1992)	Sample: 282 SS Independent Variables: 1. Environment: rule-inductive vs. rule-applied feedback 2. Cognitive ability (computerized battery of tests) Dependent Variables: 1. Learning outcome measures: a. declarative knowledge b. conceptual understanding of laws c. procedural skill (compute problems) d. generalization	1) Learning outcomes: a. Environment No sign. difference for rule-applied and rule-inductive SS for outcome measures b. Aptitude-Treatment Interaction: 1. Declarative knowledge: high-ability SS learned more if assigned to inductive environment; low-ability SS learned more in applied 2. Procedural knowledge: high-ability SS developed more skill applied environment; low-ability SS performed poorly regardless of environment 3. Generalization same as for procedural knowledge
	Shute (1993b)	Sample: 309 SS Independent Variables: 1. Environment: rule-inductive vs. rule-applied feedback 2. Exploratory behavior (proportion of time spent exploring on tutor) Dependent Variables: 1. Learning outcomes: a. declarative knowledge b. conceptual understanding of laws c. procedural skill (compute problems) d. generalization 2. Learning efficiency: a. speed (time on tutor) b. accuracy (negatively correlated with number of problems)	1) Learning Outcomes: a. Environment No sign. differences on outcome tests for rule-applied and rule-inductive environments b. Aptitude-Treatment Interaction More exploratory learners performed better in inductive environment; less exploratory learners performed better in applied environment 2) Learning Efficiency: a. Environment Rule-applied SS took sign. less time to complete the curriculum and fewer problems to reach mastery b. Aptitude-Treatment Interaction More exploratory learners performed more efficiently in inductive environment; less exploratory learners performed more efficiently in applied environment

Stat Lady: Probability lesson	Shute, Gawlick-Grendell, & Young (1993)	Sample: 168 SS Independent Variables: 1. Tutor type: computer tutor, lecture, & control 2. Aptitude (SAT scores) Dependent Variables: 1. Learning outcome (one test; two parts) a. declarative knowledge b. procedural skill	1) Learning outcomes: a. Tutor type 1. Declarative knowledge Computer tutor SS performed sign. better than other groups 2. Procedural skill Lecture SS performed sign. better than other groups b. Aptitude-Treatment Interaction Aptitude x tutor type 1. High-aptitude SS learned sign. more from computer tutor on learning outcomes 2. No sign. difference for low-aptitude SS
	Shute & Gawlick-Grendell (in press)	Sample: 311 SS Independent Variables: 1. Tutor type: computer tutor, paper & pencil workbook, & control 2. Cognitive ability (computerized battery of tests) Dependent Variables: 1. Learning outcome (one test; 2 parts) a. declarative knowledge b. procedural skill 2. Affective measure (enjoyability and helpfulness)	1) Learning outcomes: a. Tutor type No sign. difference between computer tutor and workbook SS for combined learning measure, but both groups outperformed control 1. Declarative knowledge Computer tutor SS acquired more knowledge than workbook SS 2. Procedural skill No sign. difference between computer tutor and workbook SS b. Aptitude-Treatment Interaction Tutor type x aptitude 1. High aptitude SS performed better with computer tutor 2. No sign. difference for low aptitude SS 2) Affective measures: Instruction by computer tutor perceived as more enjoyable and beneficial than by the workbook

Smithtown: Scientific inquiry skills in context of micro-economics (i.e., laws of supply and demand)	Shute & Glaser (1990)	Sample: 30 SS Independent Variables: 1. Tutor type: computer tutor, classroom, & control Dependent Variables: 1. Economic knowledge test	1) Learning outcomes: Computer tutor & classroom SS showed equivalent improvements and exceeded control on knowledge test 2) Learning efficiency: (time on tutor) Computer tutor SS learned same material in less than half the time than classroom SS (11 vs. 5 hours) (Note: learning times were imposed by the method)
Loader: Artificial task; execute console-operations to accomplish goals (railroad yard)	Farquhar & Regian (1993)	Sample: 29 SS Independent Variables: 1. Model: no model vs. dynamic model (model during practice trials only) Dependent Variables: 1. Post-training performance (on computer) a. completion time b. errors	1) Learning efficiency: a. Model 1. Speed (completion time) Dynamic model SS performed sign. faster than no model SS 2. Accuracy (learning errors): Dynamic model SS made sign. fewer errors than no model SS
Flight Engineering Tutor: Flight engineering skills	Shute (1993a)	Sample: 178 SS Independent Variables: 1. Environment (# problems): abbreviated (3 probs.) vs. extended (12 probs.) 2. Aptitude a. Working memory (4-Term Order Test) b. General knowledge (geography test) Dependent Variables: 1. Learning outcome measures a. Basic graph knowledge b. Complex graph knowledge & skill c. Tutor-specific knowledge & skill	1) Learning outcomes: a. Environment No sign. differences between groups on outcome measures b. Aptitude-Treatment Interaction Working memory x general knowledge x environment 1. SS with more general knowledge but low working memory capacities performed sign. better in abbreviated environment on outcome measures 2. SS with less general knowledge and high working memory capacities performed sign. better in extended environment on outcome measures

TESTING R-WISE: READING AND WRITING IN A SUPPORTIVE ENVIRONMENT

Patricia M. Harn
Graduate Student
Department of Technical Communication
University of Washington
Mail Stop FH-40
Seattle, WA 98195

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

September 1994

TESTING R-WISE: READING AND WRITING IN A SUPPORTIVE ENVIRONMENT

Patricia M. Harn
Graduate Student
Department of Technical Communication
University of Washington

ABSTRACT

R-WISE (Reading and Writing in a Supportive Environment) is part of a seven-year Air Force effort--the Fundamental Skills Training Project--to transition the latest innovations in computer-aided instruction to the public schools. This "intelligent" critical literacy skills tutor has been under development since 1991. Currently, two different versions of R-WISE are being tested in ten schools throughout the nation. Offering two different versions--the Lean and the Rich--is necessary to assess how much intelligent advice is optimal for a given student aptitude and teacher style. Results of the 1994-1995 field evaluation will be available in the fall of 1995. My own role as a technical writer has been to create several student and teacher R-WISE user's manuals.

TESTING R-WISE: READING AND WRITING IN A SUPPORTIVE ENVIRONMENT

Patricia M. Harn

INTRODUCTION

This summer, while a GSRP with Armstrong Laboratory at Brooks Air Force Base, I served as technical writer, creating user manuals. The specific project I documented was R-WISE (Reading and Writing in a Supportive Environment). R-WISE is part of a seven-year Air Force effort--the Fundamental Skills Training Project--to transition the latest innovations in computer-aided instruction to the public schools.

R-WISE has been under development since 1991. During the 1992-1993 school year, a prototype of the intelligent writing tutor was introduced at MacArthur High School in San Antonio, Texas. Numerous revisions were made to the software based upon student and teacher comments. Then, in the fall of 1993, the first non-prototype version was shipped out to various schools across the nation for testing.

Preliminary results of the 1993-1994 pilot test show a statistically significant increase of 7 percent in writing performance, as measured by a comparison between pre- and post-test writing samples (Carlson & Crevoisier, 1994). A second San Antonio high school served as the control group. The test designers devised a means of scoring the approximately 2,200 sample papers on a 1 to 6 scale, using two evaluators for each student paper. Inter-rater reliability was .79.

The R-WISE tutor is Air Force researcher Dr. Pat Carlson's adaptation of Bereiter & Scardamalia's (1987) seminal work in composition theory. This text, *The Psychology of*

Written Composition, introduces the notion of exploring both *content space* (information about the writing content) and *rhetoric space* (information about the writing process itself) to improve the results of student composition. The authors also make several important distinctions between the writing habits of novices and experts. Novice writers practice a form of writing called *knowledge telling*, characterized mainly by a lack of planning and an egocentric perspective. Expert writers, however, are adept in the art of *knowledge transformation*, possessing a wealth of planning techniques and familiarity with a wide repertoire of text possibilities. The goal of R-WISE, therefore, is to strengthen higher-order thinking skills by prompting the student with strategic planning techniques and robust problem-solving behaviors.

Until recently, computer-assisted writing instruction has undertaken little more than the checking of spelling and grammar, with limited results deriving from this limited set of goals. Now all that is needed is a tutor that assists in developing the higher order thinking skills used among expert writers. R-WISE makes this long-awaited advance to actual instruction in the writing process. This instruction comes in the form of a rudimentary computer-based trainer (CBT) and an expert system. Expert system technology is the branch of artificial intelligence that seeks to emulate the knowledge of a human expert in a specific field or "domain" (Bielawski & Lewand, 1991). R-WISE's expert system was constructed from the knowledge of several experienced English instructors.

R-WISE is comprised of three intelligent tools covering each phase of the writing process: Cubing (for pre-writing), Idea Board (for paragraph outlining and drafting), and Re-Vision (for editing on three levels). Though the student is able to produce a complete text in any of the tools, each tool highlights a different phase of the writing process. Thus, the tools are both interrelated and independent, with the intelligent help provided only at the area of emphasis for each tool (see Figure 1, bold boxes).

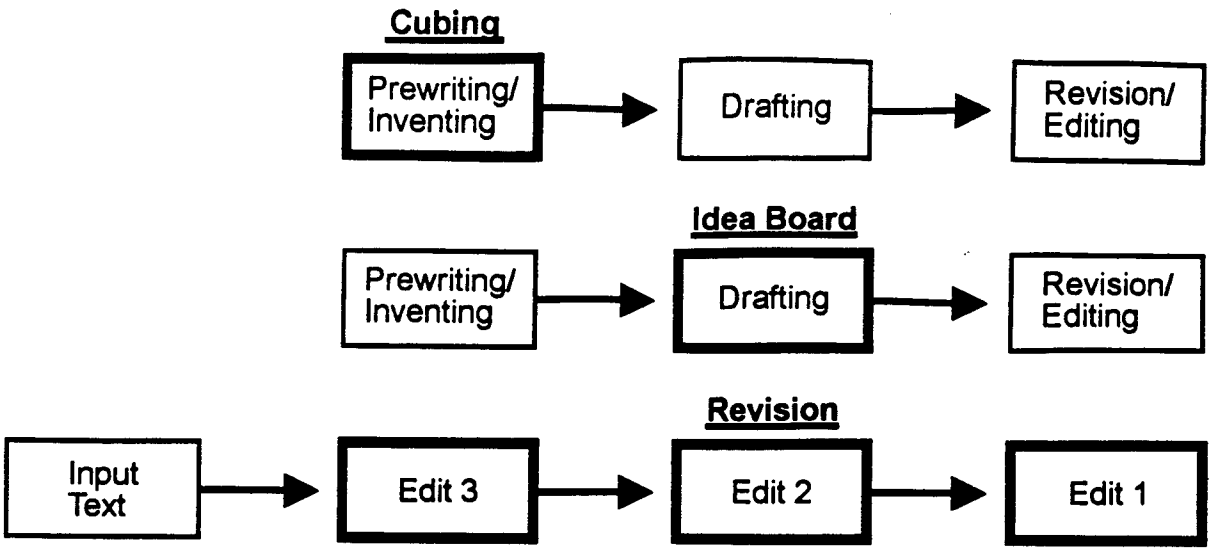


Figure 1. Access to the intelligent advice

Within each of the three writing tools, there are three different levels of intelligent help: Think About It, Get a Hint, and Get Advice. The first level, Think About It, attempts to diagnose the writing problem. Thus, Think About It serves as a gateway to the other two. The next level, Get a Hint, is a more in-depth tutor that offers a mini-review of a specific writing skill based upon the diagnosis in Think About It. The final level of help, Get Advice, is action-oriented, offering suggestions for a specific set of actions the writer can take to immediately improve his or her text. Again, the advice is based upon the initial diagnosis in Think About It.

PROBLEM STATEMENT

Though the introduction of R-WISE in the 1993-1994 school year produced a statistically significant gain in student writing performance, further improvement is sought. Based upon instructor feedback from the ten test schools, one clear problem has emerged: getting students to fully utilize the three levels of help in the three different

tools. The proposed solution to this problem has been to create and test two different versions of the R-WISE software, the Lean version and the Rich version. In the Lean version, the one currently being tested in the fall of 1994, pared-down intelligent help is provided automatically, without student solicitation. In the Rich version, which will be tested early in 1995, the help is provided only upon student request. Because the student must take the initiative to solicit help in the Rich version, this help will be more extensive and detailed--in a word, more "rich." Thus, the current research centers on testing the relative effectiveness of the two different versions.

METHODS

This year, the two different versions of R-WISE will be subjected to a field evaluation at ten different sites nation-wide. The ten schools involved represent a broad cross-section of student abilities and socio-economic backgrounds, ranging, for example, from a high-performing community college in upstate New York to a group of low-performing high school students from an Indian reservation near Albuquerque, New Mexico. Approximately 3000 students and 50 teachers are participating in the test. The research, headed by Dr. Pat Carlson, is what is called an Aptitude Treatment Interaction study. Dr. Carlson hopes to investigate the inter-relationships of teaching style, student aptitude, and version of the R-WISE tutor (Lean or Rich). The purpose is to assess the effectiveness of both versions of the tutor, given varying student aptitudes and teaching approaches. The research design is a 2 X 2 factorial, as illustrated in Figure 2:

	Semester 1	Semester 2
Teacher 1	Lean v.	Lean v.
Teacher 1	Lean v.	Rich v.

Figure 2. 1994-1995 research design

As shown in Figure 2, the teacher or teaching style can be held constant, with half the students moving to the Rich version of R-WISE for the second semester, and the other half staying on the Lean version. Students will be randomly assigned to either version. As in the 1993-1994 pilot study, the effectiveness of the two versions will be determined by comparing pre- and post-test student essays to see if the improvement in writing is statistically significant as compared to the control group. In addition, the students will be given a reading test and an OLSAT (Otis-Lennon School Ability Test).

RESULTS

The outcome of this study will be reported in the summer of 1995.

CONCLUSIONS

The only conclusions that can be drawn at this time are those deriving from the results of the 1993-1994 pilot test, which shows a statistically significant improvement in student writing.

As stated in the Introduction, my role for both this summer and last has been to write the student and teacher user manuals for the R-WISE software. This summer, I created the following items:

- Completely revised last year's *R-WISE User's Manual* in order to create two separate documents, one for the Lean version and one for the Rich. Both of these booklets are approximately 75 pages in length.
- Wrote an 18-page quick reference manual for the Authoring Tool. The Authoring Tool is the software used by the teachers to set up student lessons on the computer.

- Created three "quick cards" for various parts of the R-WISE program and the Authoring Tool. Quick cards are small brochures that distill a program's most salient features.
- Wrote a 36-page quick reference manual for the Lean version of R-WISE. A quick reference manual is something of a cross between a quick card and a full-sized user's manual.
- Created two full-graphic R-WISE promotional posters in response to teacher request.
- Created two posters outlining the differences between the three levels of intelligent help available in the three tools.
- Examined and critiqued, in writing, the user interface for the R-WISE software.

I also oversaw the printing and distribution of all these written materials, attended part of the teacher training seminar, and helped the software engineers with various projects when necessary. At this point, I could go on about the importance of technical documentation to the new software user, but allow me to provide an anecdotal demonstration of its importance instead. Several times, the teachers, who are responsible for administrating the R-WISE project on a day-to-day basis, requested more copies of a booklet than we had given them. This occurred despite the fact that we at Brooks Air Force base took pains to make sure they had been given enough. By this, we can reasonably suspect that the R-WISE written materials are being needed and used.

REFERENCES

Bereiter, C. & Scardamalia, M. (1987). *The Psychology of Written Composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bielawski, L. & Lewand, R. (1991). *Intelligent Systems Design: Integrating Expert Systems, Hypermedia, and Database Technologies.* New York: John Wiley & Sons.

Carlson, P. & Crevoisier, M. (June 25-30, 1994). R-WISE: A computerized environment for tutoring critical literacy. In T. Ottmann & I. Tomek (Eds.), *Educational Multimedia and Hypermedia 1994* (pp. 111-116). Proceedings, World Conference on Educational Multimedia and Hypermedia: Vancouver, BC, Canada.

RAPID BACTERIAL DNA FINGERPRINTING BY THE POLYMERASE CHAIN REACTION

Jason E. Hill, B.S. Biology
Nicholas F. Muto, B.S. Biology

University of Scranton
Scranton, PA 18510

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

RAPID BACTERIAL DNA FINGERPRINTING BY THE POLYMERASE CHAIN REACTION

Jason E. Hill, B.S.
Nicholas F. Muto, B.S.
Department of Biology
University of Scranton

Abstract

Typing of bacterial strains by polymerase chain reaction fingerprinting was studied. Bacterial strains were grown overnight and the DNA isolated by the CTAB method. This study utilized REP and ERIC primers, which target dispersed repetitive sequences, for gram negative bacteria (especially *E. coli*, *Salmonella*, and *Pseudomonas*). Primers were derived from repetitive sequences in *M. pneumoniae* and used with the gram positive organism *S. aureus*. Differential fingerprints were obtained by PCR showing which strains were derived from the same bacterial clone.

RAPID BACTERIAL DNA FINGERPRINTING

BY THE POLYMERASE CHAIN REACTION

Jason E. Hill
Nicholas F. Muto

Introduction

Bacterial typing is a complex and lengthy task. The many classical ways of typing bacteria generally belong to the realm of microbiology. These classical methods are not one hundred percent accurate. Difficulties in accurately typing bacterial infections are further compounded by the fastidiousness of certain organisms. DNA fingerprinting, a tool of the molecular microbiologist, is a powerful way of accurately typing bacteria. It obviates the lengthy process of culturing fastidious organisms and is widely applicable (when an outbreak of a bacterial infection occurs, it does not necessarily follow that every victim is infected with the same bacterial clone).

Most bacteria contain sequences that are repeated throughout their genome. These repetitive elements do not occupy the same positions in all clones. They may be separated by variable amounts of DNA in different clones. DNA fingerprinting utilizes this variable amount of DNA to type bacteria. The sequences of the repetitive elements are known and primers have been developed complimentary to parts of the repetitive sequences. PCR is used to amplify the regions of DNA between the primers. When the fingerprints are analyzed on an electrophoretic gel the different sizes of amplified DNA give a unique fingerprint for every clone. This procedure is both rapid and accurate.

Methodology

Bacterial cultures were provided by our focal lab. The DNA was isolated by the procedure described by Versalovic, et al. 1994, p. 17, but with alterations. Briefly, 1.5 ml cultures were grown overnight in BHI broth. The cells were pelleted at 3,000 rpm for 10 min. in a microcentrifuge, then washed once with 1M NaCl and once with TE buffer. The cells were resuspended in TE buffer and lysed with 10 ul lysozyme (5 mg/ml) for gram negative bacteria and 75 units of mutanolysin for gram positive bacteria (12 units of lysostaphin for *S. aureus*). The cells were incubated for 30 min. at 37°C, then 30 ul of 10% SDS and 3 ul proteinase K (20 mg/ml) were added and the cells were incubated for 1 hr at 37°C. To this 100 ul of 5M NaCl was added followed by 80 ul of 1% CTAB/1M NaCl solution. The samples were incubated for 10 min. at 65°C. The samples were extracted once with an equal volume of chloroform, once with phenol:chloroform, and finally one more time with chloroform. The DNA was precipitated with an equal volume of isopropanol and resuspended in sterile water. The DNA was incorporated into a 25 ul PCR using the following reagent concentrations: 1 XPCR buffer #11 (Opti-Prime Kit from Stratagene), 15 mM bovine serum albumin (Opti-Prime Kit from Stratagene), 300 uM dNTP mix (by Boehringer Mannheim), 1 uM of two opposing oligonucleotide primers (Wenzel & Herrman, p. 8338) (Table 1), 100 ng DNA, and 1.75 U/ul Taq polymerase (Perkin Elmer/Cetus). The reaction was brought up to 25 ul with sterile water. This PCR cocktail was used for gram positive bacteria. The primer sequences were taken from an article describing repetitive sequences in *M. pneumoniae* and were synthesized by the Midland Certified Reagent Company. Cycling conditions were as follows: Initial denaturation at 95°C for 3 min., then 30 cycles

of denaturation at 94°C-1 min., annealing at 43°C-1 min., extension at 72°C-2 min., and a final extension at 72°C for 5 min. The PCR samples were visualized on an agarose gel (1.2% Seakem GTG agarose in 1X TAE buffer containing ethidium bromide) and photographed. Fingerprint bands were sized using a 100 bp ladder and a 1 kb ladder from Gibco BRL. Fingerprint bands were compared based on size and intensity. For gram negative bacteria the PCR cocktail was different: 1X PCR buffer (Versalovic, et al. 1994, p. 21), 10% DMSO, 1.25 mM each dNTP, 50 pmoles of two opposing primers (ERIC) (Table 1), and 2 units of Taq polymerase. The cycling conditions were: 95°C for 7 min., 30 cycles of 52°C-1 min., 65°C-8 min., 94°C-1 min., and 65°C for 16 min.

Results

Figure 1 shows five *S. aureus* strains amplified with two different sets of primers, P1-M2 and RW3, and RW2A and RW3A. Strains 61, 1816, and 1844 have similar fingerprints with both sets of primers. Strains 1824 and 1830 are clearly different from the other strains, demonstrated again by both primer sets.

Figure 2 shows ten strains amplified with *M. pneumoniae* primers RW2 and RW3. This figure is an example of the difficulty we had with DNA concentrations. Some of the fingerprints are distinct, others are faint, and some do not appear at all.

For gram negative bacteria, ERIC-PCR was utilized to generate fingerprints of their genomes. The most complex and distinct genomic fingerprints were obtained from samples of *E. coli* (Fig 3). *Pseudomonas aeruginosa* (Fig 4), as well as some *Salmonella* samples (Fig 5) also yielded fairly decent amplification products.

Conclusion

Four primers were developed from *M. pneumoniae* repetitive sequences. Of all possible combinations P1-M2, RW3 and RW2, RW3 amplified the most strains. Consistency of the PCR was the biggest stumbling block. A PCR that produced excellent fingerprints one time would produce either poor or no fingerprints when run a second time under the same conditions. There are many nuances of multiplex PCR that need to be considered compared to normal PCR. One explanation is that we generally had low yields of DNA from gram positive organisms, especially *S. aureus*, due to the tenacity of the cell wall to resist complete lysis. Also, annealing temperature seemed to be critical even to within one degree although we did not have time to fully explore the effects of altering the annealing temperature. With the research accomplished at Brooks AFB, our home laboratory at the University of Scranton in Scranton, PA should develop a consistent fingerprinting protocol for *S. aureus* which will then be used at Brooks AFB in the near future.

Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences occur throughout the genomes of many enteric gram negative bacteria. By performing PCR with a primer set found within these sequences and amplifying the regions between the sequences, unique, strain specific fingerprints are obtained allowing for the typing of these organisms. Various samples of the bacteria *E. coli*, *P. aeruginosa*, and *Salmonella* were typed with this method. As the work we have begun is continued at the University of Scranton, we will eventually be able to examine entire outbreaks of infection by a certain bacterium and determine the source of each case involved. This is a powerful and invaluable use of the polymerase chain reaction.

Electrophoresis Profile

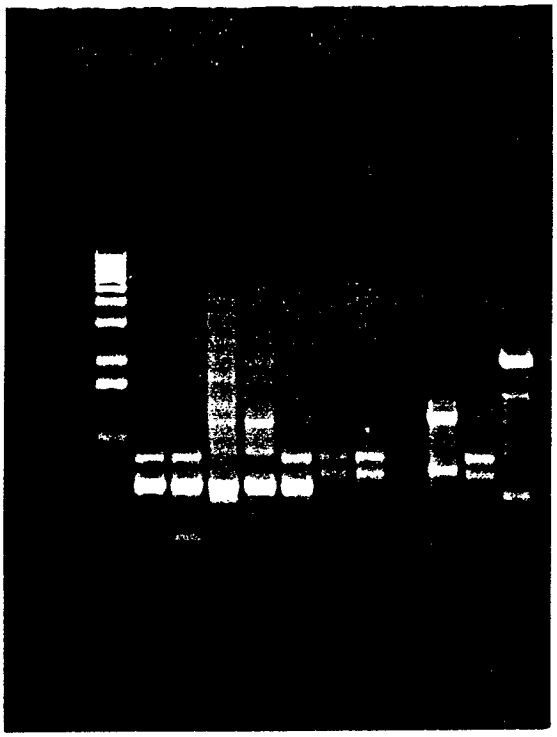
Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 8/8/94

Voltage 65 V
Final time
Initial mAMP
Initial time
Total time

WELL #	SAMPLE	
1	1 Kb ladder	
2	S. aureus 61	Primers P1-M2 and RW3
3	1816	
4	1824	
5	1830	
6	1844	
7	61	Primers RW2A and RW3A
8	1816	
9	1824	
10	1830	
11	1844	
12	100 bp ladder	

Figure 1



Electrophoresis Profile

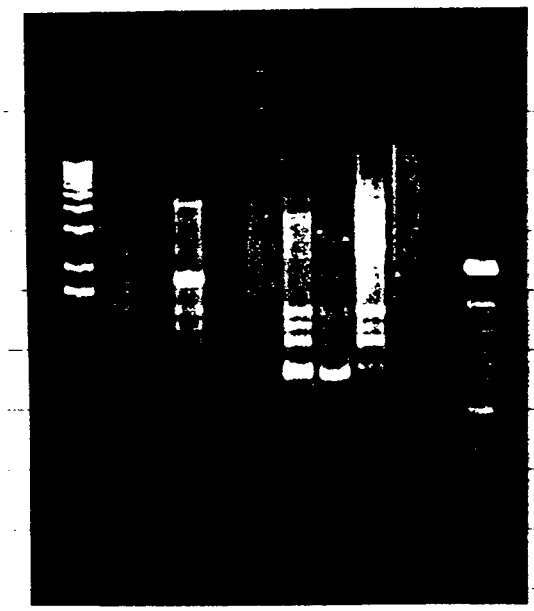
Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 7/28/94

Voltage 65 V
Final time
Initial mAMP
Initial time
Total time

WELL #	SAMPLE
1	1 Kb ladder
2	S. aureus 2289
3	2290
4	2291
5	2292
6	2293
7	1844
8	2344
9	2345
10	61
11	negative control
12	100 bp ladder

Figure 2



Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 7/22/94
Voltage 50
Final time
Initial mAMP
Initial time
Total time 3:00

WELL #	① SAMPLE	WELL #	② SAMPLE
1	1kb LADDER	13	1kb
2	2446	14	1001
3	996	15	994
4	2287	16	BT93 502
5	2296	17	BT92 2172
6	3067	18	995
7	2295	19	997
8	1002	20	1000
9	STYLATO 0157:H7	21	1003
10	BT93 499	22	999
11	1008	23	2432
12	1kb	24	3118

ERIC 1R/ ERIC 2 PRIMERS

10 µL/well

ALL SAMPLES E. coli

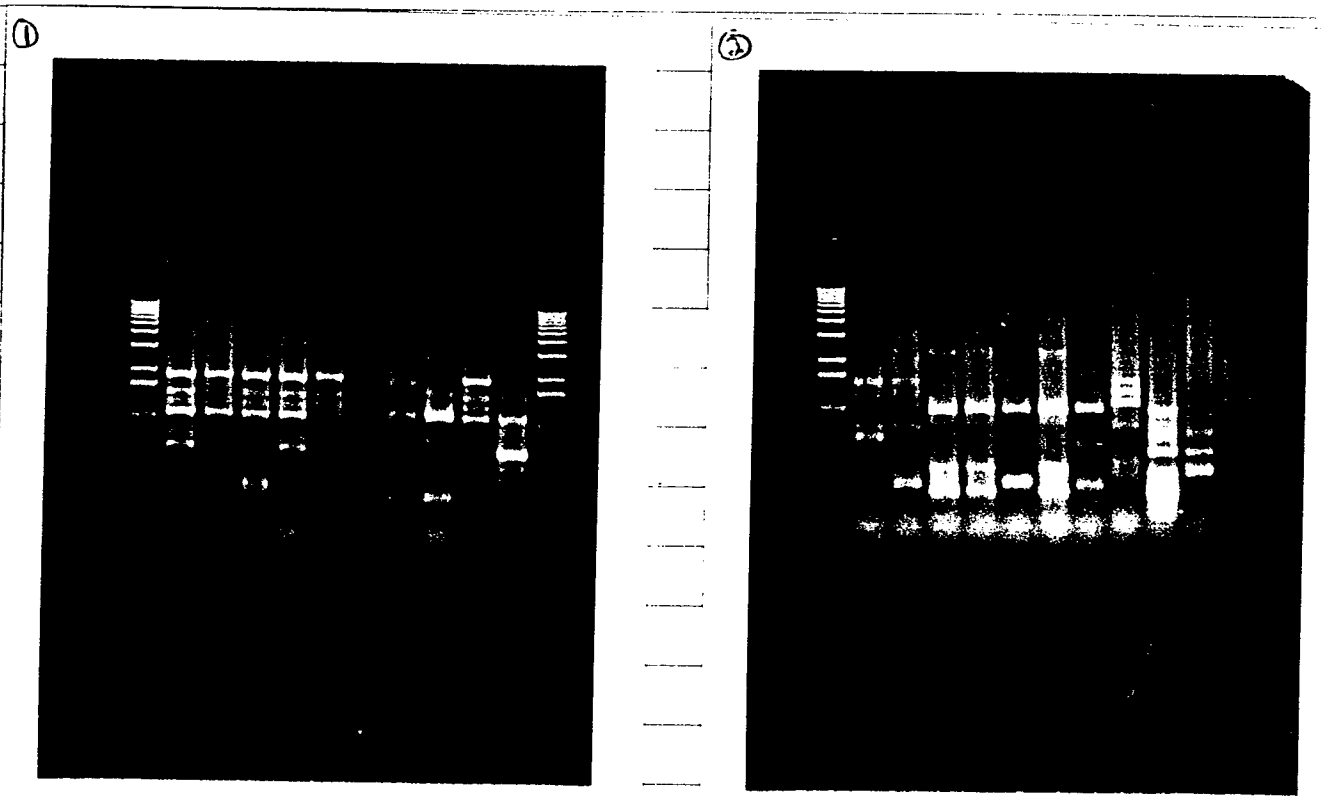


FIGURE 3

Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE), TBE
Electrophoresis apparatus (S), M, L

Date 7/26/94

Voltage 50
Final time
Initial mAMP
Initial time
Total time 2:30

WELL #	SAMPLE
1	1kb LADDER
2	P.aeruginosa 2573
3	" 2659
4	" 2575
5	" 2691
6	" 3005
7	" 3002
8	E. coli BT9Y 2252
9	BT9Y 1711
10	BT9Y 3070
11	NEG. CONTROL
12	1kb LADDER

ERIC 1R/ ERIC 2 PRIMERS

10µL/well

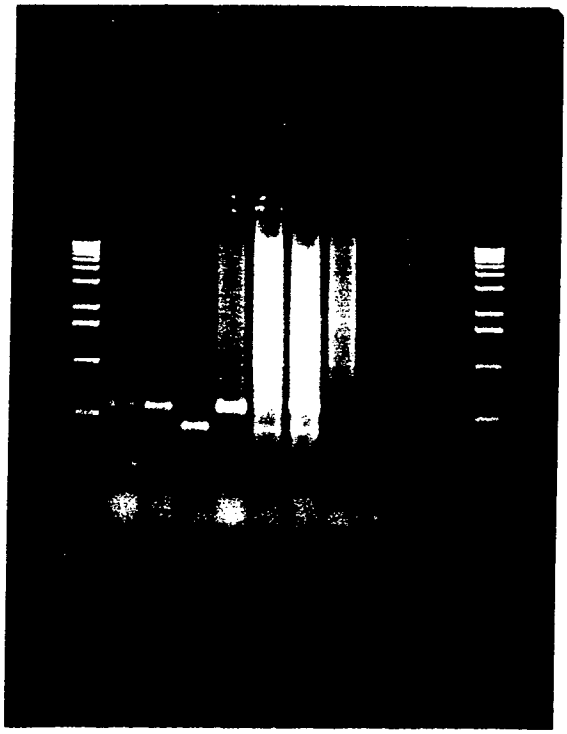


FIGURE 4

Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE) TBE
Electrophoresis apparatus (S), M, L

Date 3/4/94

Voltage 60
Final time
Initial mAMP
Initial time 3:30
Total time 2:00

WELL #	SAMPLE
1	1kb LADDER
2	P. aeruginosa 3002
3	" 3005
4	Salmonella 2798
5	" 2799
6	" 3072
7	" 2757
8	UNKNOWN 3003
9	E. coli BT92 3072
10	" BT9Y 2252
11	" 2797-2
12	" 2573-1
13	NEG. CONTROL
14	100 bp LADDER

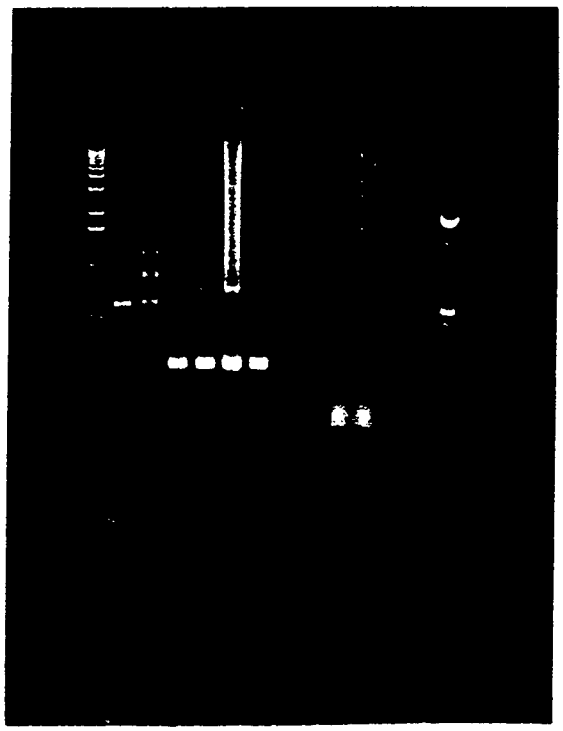


FIGURE 5

FIGURES/TABLES

Table 1: Composition of Oligonucleotide Primers

ERIC 1R	3'-CACTTAGGGGTCCTCGAATGTA-5'
ERIC 2	5'-AAGTAAGTGACTGGGGTGAGCG-3'
RW2	5'-TCTTTACGCGTTACGTATT-3'
RW3	5'-CTCAAAACAACGACACCGG-3'
P1-M2	5'-CCCCCACCACTAAGCACAC-3'

BIBLIOGRAPHY

Versalovic, J., Koeuth, T. and Lupski, J.R., 1991. Distribution of Repetitive DNA Sequences in Eubacteria and Application to Fingerprinting of Bacterial Genomes. *Nucleic Acids Research* 19(24),6823-6831.

Versalovic, J., Schneider, M., deBruijn, F.J., and Lupski, J.R. Genomic Fingerprinting of Bacteria Using Repetitive Sequence Based PCR (rep-PCR) *Methods in Molecular and Cellular Biology*, IN Press.

Wenzel, R., and Herrmann, R., 1988. Repetitive DNA Sequences in *Mycoplasma pneumoniae*. *Nucleic Acids Research* 16(17), 8337-50.

MELATONIN, BODY TEMPERATURE AND SLEEP IN HUMANS: A REVIEW OF A NEW HYPNOTIC DRUG

Rod J. Hughes
Post Doctoral Fellow
Department of Psychiatry

Oregon Health Sciences University
3181 SW Sam Jackson Park Road
Portland OR 97201

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

MELATONIN, BODY TEMPERATURE AND SLEEP IN HUMANS: A REVIEW OF A NEW HYPNOTIC DRUG

Rod J. Hughes

Introduction

Recent humanitarian missions to Rwanda and long range tactical missions before and during the Gulf War exemplify the evolving role of the Air Force as it executes its global responsibilities. Global Power and Global Reach efforts require not only that air crew fly extended missions crossing many time zones but also that air crew perform at peak levels at their destination. Additionally, military downsizing threatens to place even more pressure on human personnel. To complete their missions, air crew are often required to establish work-rest (wake-sleep) behavioral patterns that are independent of the solar light-dark cycle. However, while behavioral patterns can be changed quickly, underlying endogenous physiological patterns cannot. The result can be an out-of-phase relationship between behavioral rhythms and physiological rhythms. This out of phase relationship is termed circadian desynchrony. Circadian desynchrony occurs in sustained operations and in operations in which personnel are required to function out-of-phase with their endogenous circadian rhythms (e.g., shift work and rapid transmeridian travel). In these operations, nighttime performance is impaired because physiological rhythms such as alertness, psychomotor abilities and cognitive abilities are not at peak levels (e.g., Åkerstedt, 1988). 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. Daytime sleep is more difficult to initiate, is fragmented by multiple awakenings and is truncated (Naitoh, Kelly & Englund, 1990). The combination of working at night and sleeping during the day often leads to serious negative consequences such as impaired performance (e.g., Keran, Smith, Duchon, Robinson & Trites, 1991; Leung, & Becker, 1992) and increases in accidents (e.g., Mitler, et al., 1988; Novak, Smolensky, Fairchild & Reves, 1990). Operating under such desynchrony for many years is also associated with health risks including gastrointestinal disorders and increased risk of cardiovascular disease (Moore-Ede & Richardson, 1985; Naitoh, Kelly & Englund, 1990).

To improve sleep for these individuals, physicians frequently prescribe hypnotic drugs, of which the benzodiazepines are the most popular. The hypnotic effects of benzodiazepines are a result of increasing the neural inhibition mediated by γ-aminobutyric acid (GABA). Thus, their mechanism of action is through CNS deactivation. The direct effects of benzodiazepine administration include, sedation, drowsiness, decreased anxiety and muscle relaxation (Rall, 1990). Benzodiazepines also facilitate the onset and maintenance of sleep. Specifically, benzodiazepines shorten sleep latencies, reduce the number of awakenings, reduce the overall amount of wake time and increase the total amount sleep time (see Borbely, Åkerstedt, Benoit, Holsboer & Oswald, 1991; Pascually, 1991). The sleep induced by these drugs, however, is different from "natural" (physiological) sleep. For instance, benzodiazepines often cause increases in one marker of stage 2 sleep: sleep spindles. The sleep architecture of benzodiazepine induced sleep is also different from normal sleep. For example, although, the effect of benzodiazepines on stage 1 sleep is variable, time spent in stage 1 is typically reduced. All benzodiazepines

dramatically increase the amount of time spent in stage 2 sleep. Further, benzodiazepines are associated with significant reductions of stage 3 and stage 4 sleep (Borbely, et al., 1991). In fact, with multiple administrations these important stages of sleep may be inhibited completely (Pascually, 1991). Most benzodiazepines also increase REM latency and decrease the overall amount of time spent in REM sleep, while increasing the number of REM intervals (Pascually, 1991; Rall, 1990).

The elimination half-life of the CNS effects of different types benzodiazepines ranges from 2 - 100 hours (Rall, 1990). Therefore, the hypnotic effects, including psychomotor performance impairments, of these drugs may extend for many hours (e.g., Hommer, 1991; Roth, Hartse, Saab, Piccione & Kramer, 1980), sometimes carrying over into the next day (Carskadon, Seidel, Greenblatt & Dement, 1982; Pascually, 1991). The benzodiazepines also have many side effects, including: ataxia, blurred vision, confusion, dry mouth, dysarthria, epigastric distress, euphoria, lightheadedness, nausea, torpor, vomiting, and weakness (Rall, 1990). Anterograde amnesia is another common side effect of the benzodiazepines (Curran, 1986; French, et al., 1990). This side effect is typically reported in travelers who take benzodiazepines to help reduce symptoms of jet-lag associated with rapid transmeridian travel. These individuals can get off the plane and check into their hotel, without storing any new memories since falling asleep on the plane (e.g., Morris & Estes, 1987). The primary determinant of the potency of these amnestic effects is the drugs affinity for the benzodiazepine receptor (Hommer, 1991). Thus, drugs like Triazolam, which have high affinity for the benzodiazepine receptor, yield the largest amnestic effects (FDA, 1985). In short, while the benzodiazepines are effective hypnotics, they also present numerous side effects.

Temazepam is currently approved for use in Air Force personnel, under the supervision of a flight surgeon. This benzodiazepine has an intermediate elimination half-life of 10-12 hr. Although, Temazepam is effective in promoting and sustaining sleep, it also has the side-effects typical of the benzodiazepines, such as pharmacological sleep architecture and anterograde amnesia (FDA, 1985; French, et al., 1990). Furthermore, the half-life of Temazepam's hypnotic effects may make its use unsuitable for some missions. The benzodiazepines are clearly better than their hypnotic predecessors, the barbiturates. However, a need for more natural, alternative, hypnotics exists.

Recently, the pineal hormone melatonin has been suggested as a possible alternative hypnotic (see Dawson & Encel, 1993 for a review). This non-toxic substance (Barchas, DaCosta, & Spector, 1967) has been administered to many hundreds of people without side effects. In addition, the hypnotic effects of exogenous melatonin (xMT) may be physiological, perhaps secondary to its influence on the circadian system including body temperature.

Daytime melatonin administration has previously shown to attenuate the daytime rise in body temperature (Dollins, et al., 1993; Dollins, Zhdanova, Wurtman, Lynch & Deng, 1994; Hughes, 1992; Hughes, Badia, French, Santiago & Plenzler, 1994; Lieberman, Waldhauser, Garfield, Lynch & Wurtman, 1984) and to facilitate the initiation of daytime sleep (Anton-Tay, Diaz & Fernandez-Guardiola, 1971; Cramer, Bohme, Kendel & Donnadieu, 1976; Cramer, Rudolph, Consbruch & Kendel, 1974; Dollins, et al., 1994; Hughes, et al., 1994; Vollrath, Semm & Gammel, 1981). Therefore, melatonin may potentially be a safe and natural alternative to currently prescribed hypnotic drugs. The present paper provides a review of the effects of melatonin on body

temperature and sleep. A description of how melatonin, body temperature and sleep relate to each other is presented. Finally, evidence that melatonin can facilitate sleep is reviewed.

Introduction to Melatonin

Endogenous melatonin is produced primarily in the pineal gland through the metabolism of dietary tryptophan. Melatonin is synthesized and secreted mainly at night (2100 - 0800 hr) and under low levels of illumination. The timing of melatonin production is determined by the output of the primary circadian oscillator. Evidence implicates the suprachiasmatic nuclei (SCN), located in the anterior hypothalamus, as the internal oscillator controlling mammalian circadian rhythms. For instance, the SCN's 24 hr pattern of firing continues in vitro (Green & Gillette, 1982). Additionally, when the SCN is transplanted in vivo the recipient's circadian rhythm period is determined by the donor's original period (Ralph, Foster, Davis & Menaker, 1990). Thus, in mammals, the SCN is responsible for generating approximately 24 hr oscillations and for entraining endogenous circadian rhythms (including melatonin) with the external photoperiod.

The circadian system begins with the retina. Light stimulating the photoreceptors of the retina is communicated to the SCN directly via the retinohypothalamic tract and indirectly via the geniculohypothalamic tract. The SCN transduces changes in photoperiod into internal electro-chemical rhythmicity. Information about the photoperiod leaves the SCN via a series of efferent projections through the upper thoracic spinal cord and on through the superior cervical ganglia, eventually terminating in sympathetic innervation of the pinealocytes (the endocrine elements of the pineal gland) (Moore & Cord, 1985; Reiter, 1991). The catecholamine, norepinephrine, is the primary neurotransmitter responsible for communication from the postganglionic neurons to the pinealocytes. Within the pinealocytes, the amino-acid tryptophan is picked up from the blood stream and eventually metabolized into melatonin (see Figure 1 for the steps involved in this conversion) which is released directly into the cerebral spinal fluid and the blood stream. Melatonin is highly lipophilic and easily crosses the blood-brain barrier (Reiter, 1991a). Binding sites and receptors for melatonin are pervasive and have been reported centrally (Stankov, Fraschini & Reiter, 1991) and peripherally (Stankov & Reiter, 1990). Because of its high diffusibility, however, melatonin can also enter cellular and sub-cellular components (Reiter, et al., 1993).

The Role of Melatonin

The phylogentic age of the melatonin molecule is very old since it is found in plants and in at least one unicellular organism: the dinoflagellate (Poeggeler, Balzer, Hardeland & Lerchl, 1991). For these organisms, as well as for higher order species, research is rapidly establishing melatonin as one of the most important, endogenous free-radical scavengers (Reiter, et al., 1993; Tan, Chen, Poeggeler Manchester & Reiter, 1993) protecting cellular and intracellular molecules from oxidative damage (Reiter, et al., 1994). Besides being an effective antioxidant, melatonin is also an important component of the overall circadian system.

In mammals, melatonin is thought to be the chemical messenger of the circadian system, responsible for synchronizing the entire circadian structure (Armstrong, 1989). In effect, melatonin is the chemical expression of darkness communicating the nighttime message throughout the entire body (Reiter, 1991b). In addition to this role, the presence of high affinity binding sites for melatonin in human fetal and adult SCN (Reppert, Weaver,

Rivkees & Stopa, 1988) suggests that melatonin affects the biological clock directly. Indeed, melatonin administration to rat SCN, in vitro, phase shifts the SCN's electrical activity (McArthur, Gillette & Prosser, 1991). In animals, melatonin injections can entrain free-running circadian rhythms (Redman, Armstrong & Ng, 1993). In humans, exogenous melatonin administration phase shifts circadian rhythms according to an established phase-response curve (PRC) (Lewy, Ahmed, Latham Jackson, & Sack, 1992). Melatonin administered in the late afternoon and early evening phase-advances circadian rhythms and melatonin administered in the early morning phase-delays rhythms (Lewy, et al., 1992). Daily administration of xMT can also synchronize human circadian rhythms and has been efficacious in the treatment of various circadian rhythm related disorders. For instance, daily melatonin administration is efficacious in synchronizing free-running circadian rhythms in the totally blind (Arendt, Aldhous & Wright, 1988; Palm, Blennow & Wetterberg, 1991; Sack, Lewy, Blood, Stevenson & Keith, 1991). Daily melatonin administration is also efficacious in the treatment of circadian rhythm related sleep-wake disorders such as insomnia (MacFarlane, Cleghorn, Brown & Striner, 1991) and delayed sleep-phase syndrome (Dahlitz, et al., 1991). In addition, by phase-shifting or synchronizing circadian rhythms daily melatonin administration facilitates adjustment to other circadian desynchronies such as jet-lag (Arendt, et al., 1987; Petrie, Conaglen, Thompson & Chamberlan, 1989; Petrie, Dawson, Thompson & Brook, 1993; Samuel, et al., 1991) and shift work (Folkard, Arendt & Clark, 1993; Sack, et al., 1994). It appears that melatonin has direct effects on the SCN, serving as part of a feedback mechanism for the circadian system. Thus, melatonin has at least three primary roles: 1. as free radical scavenger, 2. as the body's chemical messenger of "darkness" or "time to sleep" and 3. as an integral component of at least one feedback mechanism in the circadian system.

Circadian Rhythm of Melatonin

Melatonin is synthesized and released primarily at night, with nighttime levels exceeding daytime levels by several times (Weitzman, et al., 1978). While daytime levels of melatonin in serum are typically less than 10 pg/ml, nighttime levels can peak as high as 50 to 150 pg/ml (see Waldhauser & Dietzel, 1992). In dim light, the onset of melatonin's nighttime secretion episode can be quantified as the time when physiological levels of melatonin exceed 10 pg/ml or the dim light melatonin onset (DLMO) (Lewy & Sack, 1989). The DLMO is a stable marker of circadian phase (Lewy & Sack, 1989). The nighttime secretion of melatonin typically begins around 2000-2200 hr (Lewy, 1989) rises throughout the night, peaks around 0300-0500 hr and declines precipitously thereafter (Waldhauser & Dietzel, 1992). The entire episode of elevated melatonin levels typically lasts for 9-11 hr (Waldhauser & Dietzel, 1992).

Nighttime secretion of melatonin is not masked by the sleep-wake schedule (Morris, Lack & Barrett, 1991). The nighttime biosynthesis and secretion of melatonin is, however, inhibited by sufficiently bright light (Lewy, Wehr, Goodwin, Newsome & Markey, 1980). This is achieved via sympathetic innervation from the SCN (see above). The physiological and behavioral effects of this suppression will be discussed in detail later. The melatonin rhythm is also influenced by circannual-like changes in photoperiod (Wehr, 1992). Although in natural settings humans appear to be insensitive to seasonal variations in photoperiod (Illnerová, Zvolský & Vanecek, 1985; Kennaway, & Royles, 1986; Kennaway, & Van Dorp, 1991), in controlled laboratory settings the melatonin

circadian rhythm appears to be sensitive to such changes. For instance, simulated winter (e.g., light:dark = 8:16 hr) and summer (e.g., light:dark = 16:8 hr) light-dark (L-D) cycles yield nighttime melatonin secretion periods corresponding to the dark phase of the photoperiod (Buresová, Dvoráková, Zvolský & Illnerová, 1992; Wehr, 1992; Wehr, et al., 1993). In natural settings, similar seasonal variations may be masked by the use of artificial light.

Circadian Rhythm of Temperature

Daily oscillations in human body temperature have been shown for centuries (see Marotte & Timbal, 1981). In entrained individuals, the body temperature rhythm begins around 0900 hr at about 37.2° C, rises slowly across the day and peaks around 2000 hr at about 37.4° C. Temperature then falls across the night to trough around 0400 hr at about 36.5° C (Refinetti & Menaker, 1992) at which time it begins to rise rapidly again until morning (0900 hr). The shape of this daily oscillation differs from the melatonin rhythm, in that the body temperature rhythm approximates a cosine wave. However, nighttime changes of the two rhythms almost mirror each other, and daytime differences may diminish as methods for quantifying low levels of melatonin become more sensitive.

Relationship Between Melatonin and Body Temperature

There is considerable evidence for a direct, inverse, relationship between melatonin and body temperature. Temporally, physiological levels of serum melatonin are inversely related to body temperature (see Reiter, 1990). In fact, melatonin levels are thought to mediate circadian changes in body temperature (for a review see Badia, Myers, & Murphy, 1992). Evidence for this relationship comes from empirical investigations of melatonin and body temperature as well as from correlational research on populations with abnormal melatonin levels. For instance, patients suffering from anorexia nervosa and bulimia nervosa exhibit elevated 24 hr melatonin levels and lower body temperature levels (Ferrari, Fraschini & Brambilla, 1990). Depressed patients have lower nighttime melatonin levels (Wetterberg, 1978; Wetterberg, Beck-Friis, Kjellman & Ljunggren, 1984) and elevated nighttime body temperature (Avery, Wildschiotz, & Rafaelson, 1982; Schulz & Lund, 1983). Furthermore, normal age-related reductions in the amplitude of the melatonin rhythm (Iguchi, Kato, & Ibayashi, 1982; Sharma, et al., 1989; Touitou, et al., 1984; Waldhauser, Ehrhart, & Förster, 1993) are correlated with changes in the temperature rhythm, including: a shorter period (Weitzman, Moline, Czeisler, & Zimmerman, 1982), an advance in phase (Campbell, Gillin, Kripke, Erikson & Clopton, 1989; Czeisler, et al., 1992; Richardson, Carskadon, Orav, & Dement, 1982; Weitzman, Moline, Czeisler & Zimmerman, 1982; Zepelin & McDonald, 1987) and a reduction in amplitude (Czeisler, et al., 1992; Nakazawa, et al., 1991; Touitou, et al., 1986; Van Coevorden, et al., 1991; Vitiello, et al., 1986; Weitzman, et al., 1982; Van Coevorden, et al., 1991).

Correlational evidence for the influence of melatonin on body temperature is supported by empirical research. For instance, suppressing the nighttime secretion of endogenous melatonin with bright light attenuates the nighttime decrease in body temperature (Badia, et al., 1991; French, et al., 1991). Similar effects of bright lights on body temperature are not found during the daytime (Badia, et al., 1991), when endogenous melatonin levels are low. The latter finding suggests that this effect is mediated by the suppression of endogenous melatonin (Badia, et al., 1991). Indeed, the effects of nighttime bright lights on body temperature are reversed with simultaneous

melatonin administration (Cagnacci, Soldani & Yen, 1993; Sack, et al., 1992; Strassman, Quallis, Lisansky & Peake, 1991). These investigations will be discussed in more detail later.

The most compelling evidence for melatonin's affect on body temperature comes from the administration of exogenous melatonin (xMT) during the daytime when endogenous melatonin levels are very low and temperature is rising. The first report of the effect of exogenous melatonin on body temperature was by Carman, Post, Buswell and Goodwin, (1976) who reported significant reductions in oral temperature following oral administrations of very large doses of exogenous melatonin (approximately 1 g daily). Other dose levels have been shown to lower daytime body temperature (Cagnacci, Elliott & Yen, 1992; Dollins, et al., 1993; French, et al., 1993; Hughes, 1992; Hughes, et al., 1994; Lieberman, et al., 1984) even small doses that yield serum melatonin levels that approximate nighttime physiological levels (300 μg) (Dollins et al., 1994). In summary, there is a consistent, inverse, relationship between melatonin and body temperature, both of which have well documented relationships with the sleep-wake rhythm.

The Circadian Rhythm of Sleep

Historically, the sleep-wake rhythm has been assessed using three methodological designs (see Webb, 1994): (a) *displacement designs* in which the sleep phase is moved to a different time (e.g., shift work and rapid transmeridian travel), (b) *desynchronous designs* in which external time cues like the L-D cycle are either modified or are set to a cycle length greater than or less than 24 hr, and (c) *time-free designs* in which subjects are placed in an environment absent of all time-cues (for an historical review see Webb, 1994). The specific circadian rhythms of the sleep-wake system are presented with their relationships with melatonin and body temperature.

Melatonin, Body Temperature and Sleep

Temporally, the circadian rhythm of sleepiness is inversely associated with body temperature and directly associated with melatonin. For instance, subjective feelings of sleepiness are negatively correlated with body temperature levels and positively correlated with melatonin levels (Åkerstedt, 1988; Åkerstedt, Gillberg & Wetterberg, 1982). Further, cognitive and psychomotor performance are directly related to body temperature levels and inversely associated with melatonin (Åkerstedt, Gillberg & Wetterberg, 1982; see Campbell, 1992 and Smith, 1992 for reviews). Finally, sleep tendency is positively related to melatonin and inversely related to temperature (Richardson, Carskadon, Orav, & Dement, 1982; Walsh, & Sugarman, 1988).

Under entrained conditions, we tend to fall asleep at night when melatonin is rising (Birkeland, 1982) and body temperature is falling (Campbell, & Broughton, 1994) and awaken in the morning when melatonin is falling and body temperature is rising (Czeisler, Weitzman, Moore-Ede, Zimmerman & Knauer, 1980). More specifically, sleep is best initiated soon after body temperature has achieved its maximum rate of decline (the steepest slope) (Campbell, & Broughton, 1994). Also, the amount of sleep in the first hour after sleep onset is negatively correlated with the proximity between sleep onset and the maximum rate of temperature decline (Campbell, & Broughton, 1994). Once asleep, nighttime surges in melatonin output are associated with nighttime awakenings (Birkeland, 1982), suggesting that in addition to facilitating the initiation of sleep, melatonin may also be associated with the restoration of sleep following nighttime awakenings.

In support of this role, the amount of nighttime sleep is positively correlated with the amount of nighttime melatonin secretion (Morris, Lack, & Barrett, 1990; Wehr, 1991). For instance, the duration of nighttime sleep is directly related to the amplitude of the melatonin rhythm (Morris, Lack & Barrett, 1990) and the temperature rhythm (Wever, 1988). Duration of nighttime sleep is also directly associated with the duration of nighttime melatonin secretion. For instance, extending the dark phase of the circadian cycle extends the duration of melatonin secretion as well as the duration of nighttime sleep (e.g., Wehr, et al., 1993). Compared to L-D cycles containing 8 hr of darkness, L-D cycles containing 16 hr of darkness yield substantially longer (3 hr) nighttime melatonin secretion episodes (Buresová, Dvoráková, Zvolský & Illnerová, 1992). In response to 8 hr or 14 hr of darkness, Wehr reported two investigations demonstrating similar changes in the duration of nighttime melatonin secretion (Wehr, 1992; Wehr, et al., 1993). Physiologically, extending the dark phase of the L-D cycle extends the duration of nighttime melatonin secretion, extends the duration of lower nighttime body temperature, increases prolactin secretion and delays the early morning rise in cortisol secretion (Wehr, et al., 1993). These changes are associated with longer durations of nighttime sleep (Wehr, 1992; Wehr, et al., 1993). Thus, the nighttime elevation of endogenous melatonin is temporally associated with the decrease in nighttime body temperature as well as with the onset and duration of nighttime sleep.

Correlational evidence in populations with abnormal circadian rhythms

In addition to the various circadian rhythm sleep disorders (see International Classification of Sleep Disorders or ICSD, 1990), individuals suffering from insomnia have disturbed melatonin rhythms (MacFarlane, Cleghorn & Brown, 1984). Populations with abnormal melatonin and temperature rhythms also suffer from disturbed sleep rhythms. For instance, sleep disturbances are prevalent in individuals suffering from anorexia nervosa and bulimia nervosa. Additionally, mothers of preterm infants exhibit a reduction in the amplitude of melatonin rhythm which is associated with a reduction in the amplitude of the sleep rhythm (McMillen, Mulvogue, Kok, Deayton, Nowak, & Adamson, 1993). In these individuals, however, environmental influences certainly exacerbate sleep disruptions.

Affective disorders

As noted, depressed patients have lower nighttime melatonin levels (Claustrat, Chazot, Brun, Jordan, & Sassolas, 1984; Brown, et al., 1985; Wetterberg, 1978; Wetterberg, Beck-Friis, Kjellman & Ljunggren, 1984) and elevated nighttime body temperatures (Avery, Wildschiotz, & Rafaelson, 1982; Schulz & Lund, 1983). Depressed patients also exhibit various sleep disturbances, including, early morning awakenings, lower sleep efficiency, decreased slow wave sleep, decreased REM latency and an increased first REM period (see Gillin, Mendelson & Kupfer, 1988; Kupfer, Frank, Jarrett, Reynolds & Thase, 1988 for reviews).

Ontogenetic relationships among melatonin, temperature and sleep

Age-related reductions in nighttime melatonin secretion (Iguchi, Kato, & Ibayashi, 1982; Sharma, et al., 1989; Touitou, et al., 1984; Waldhauser, et al., 1993) are associated with disruptions in the body temperature rhythm, most notably an advance in phase (Campbell, Gillin, Kripke, Erikson & Clopton, 1989; Czeisler, et al., 1992; Richardson, Carskadon, Orav, & Dement, 1982; Weitzman, Moline, Czeisler & Zimmerman, 1982; Zepelin &

McDonald, 1987) and a reduction in amplitude (Czeisler, et al., 1992; Nakazawa, et al., 1991; Touitou, et al., 1986; Van Coevorden, et al., 1991; Vitiello, et al., 1986; Weitzman, et al., 1982). These changes in the melatonin and body temperature rhythms are associated with significant disturbances in the circadian rhythm of sleep. For instance, age-related phase advances of bedtimes and awakening times are well documented (Czeisler, et al., 1992; Miles & Dement, 1980; Tune, 1969). This advance of the sleep-wake rhythm is coupled with an age-related advance in the subjective peak of alertness (Czeisler, et al., 1992). Age-related reductions in the amplitude of the sleep rhythm are revealed by increases in the number of nighttime awakenings (Carskadon, Brown, & Dement, 1982; Feinberg, 1969; Smith, Karacan, & Yang, 1977 and increased napping during the day (Lewis, 1969; Tune, 1969; Webb & Swinburne, 1971). Ontogenetic changes in the distribution of sleep stages (sleep architecture) are also apparent. Compared to younger subjects, older subjects spend a lower percentage of time in stage REM sleep and slow wave sleep and a higher percentage of time in stage 1 sleep and stage 2 sleep (Czeisler, et al., 1992). Finally, there appears to be an age-related advance in the distribution of REM sleep across the night (Van Coevorden, et al., 1991; Weitzman, et al., 1982). In summary, ontogenetic disruptions in the circadian rhythms of melatonin and body temperature are associated with disruptions in the sleep-wake rhythm.

Blind People

In totally blind people, endogenous temperature and melatonin rhythms are often out-of-phase with the L-D cycle (Lewy & Newsome, 1983; Miles, Raynal & Wilson, 1977; Sack, Lewy, Blood, Keith, & Nakagawa; in press; Tzischinsky, Skene, Epstein & Lavie, 1991). Since, their propensity for sleep remains in phase with the melatonin rhythm (Nakagawa, Sack & Lewy, 1992) and since in free running conditions the duration of a sleep episode is determined by the phase of the temperature and melatonin rhythms at sleep onset (e.g., Czeisler, et al., 1980), it is no wonder that sleep disturbances are prevalent in totally blind individuals (Arendt, Aldhous & Wright, 1988; Folkard, Arendt, Aldhous & Kennet, 1990; Martens, & Endlich, Hildebrandt, & Moog, 1990; Miles & Wilson, 1977; Okawa, et al., 1987; Palm, Blennow & Wetterberg, 1991; Sack, Lewy, Blood, Stevenson & Keith, 1991; Sack, Lewy & Hoban, 1987; Tzischinsky, Skene, Epstein & Lavie, 1991).

In summary, many disturbances in the sleep-wake cycle are associated with disturbances in the melatonin and temperature rhythms. As already noted, daily treatment of exogenous melatonin administration has been used to shift or set the phase of individuals suffering from circadian rhythm related sleep and wake disorders.

Investigations of the Relationships among Melatonin, Temperature and Sleep

Time free designs

Close relationships among the melatonin, temperature and sleep rhythms in entrained conditions and in populations with abnormal circadian rhythms are supported by empirical investigations of normal subjects placed in time free designs. Circadian rhythms of subjects placed in time free conditions eventually desynchronize or free run. Following this desynchrony, the sleep-wake rhythm free-runs out of phase with the temperature rhythm (Wever, 1979) and presumably the melatonin rhythm. Thus, subjects may initiate sleep at various phases of the temperature rhythm. In these conditions, subjects tend to initiate sleep more frequently when temperature is

declining (Åkerstedt, 1988; Buysse, Monk, Reynolds, Mesiano, Houck & Kupfer, 1993; Zulley & Campbell, 1985; Zulley, Wever, Aschoff, 1981) and melatonin is peaking. Further, the duration of sleep is determined by the phase of the temperature rhythm at bedtime, in that individuals tend to wake up on the rising phase of the temperature rhythm (Czeisler, et al., 1980; Gillberg & Åkerstedt, 1982; Strogatz, Kronauer & Czeisler, 1986; Zulley, Wever & Aschoff, 1981) when melatonin is declining.

Desynchronous designs

Research using desynchronous designs supports relationships among melatonin, body temperature and sleep. For instance, when subjects are placed in L-D cycles outside of the circadian range of entrainment, subjective alertness remains in phase with the body temperature rhythm (Monk, Moline, Fookson & Peetz, 1991). Employing a 20 min. day (13 min. of wake time and 7 min. of sleep opportunity), Nakagawa, Sack and Lewy (1992) compared the amount of sleep in 7 min. nap opportunities (sleep propensity) with circadian rhythms of melatonin, temperature and cortisol. In a case study of a totally blind man, these authors reported that sleep propensity or sleepiness was directly related to melatonin levels and negatively associated with body temperature (Nakagawa, Sack, & Lewy 1992). The distribution of sleep propensity and melatonin rhythms reported by Nakagawa, Sack, & Lewy (1992) are inversely correlated to the pattern of performance and alertness found in similar designs using normal subjects (Lavie, Gopher & Wollman, 1987).

Displacement designs

Circadian rhythm desynchrony associated with flying across multiple time zones (jet-lag) and with working across the 24 hr day (shift work) is associated with decreased performance at night (subjective night in the case of transmeridian travel), which is exacerbated by poor sleep during the day (or subjective day). Given that sleep duration is determined by the phase of the temperature rhythm at bedtime (Czeisler, et al., 1980; Gillberg & Åkerstedt, 1982; Strogatz, Kronauer & Czeisler, 1986; Zulley, Wever & Aschoff, 1981), it is not surprising that the most consistent finding of shift-work is that trying to sleep during one's subjective day results in a shorter sleep length (Åkerstedt, 1983; Dahlgren, 1981; Moore-Ede, & Richardson, 1985; Naitoh, et al., 1990; Walsh, Tepas, & Moss, 1981). This effect is also reported for rapid transmeridian travel (e.g., Gander, & Graeber, 1987; Graeber, 1988; Winget, DeRoshia, Markley, & Holley, 1984).

Empirical evidence for relationships among melatonin, body temperature and sleep

β-blockers

Investigations of β-adrenergic antagonists at doses large enough to suppress endogenous melatonin secretion have also been shown to attenuate the nocturnal decline of core body temperature (Cagnacci, Elliot & Yen, 1992). These effects are associated with significant sleep disturbances including: increases in the number of nocturnal awakenings and overall reduced sleep (Dimenas, Kerr & MacDonald, 1990).

Bright Lights

In review, elevated nighttime melatonin levels are associated with declining body temperature, with increased subjective and objective measures of sleep and sleepiness and with performance impairments. Inhibiting the nighttime biosynthesis and secretion of melatonin with bright light attenuates the nighttime decline of body temperature (Badia, et al., 1991; French, et al., 1991). These effects are associated with increases in subjective measures of alertness and activity (Badia, et al., 1991; French, et al., 1991) improvements in psychomotor and cognitive performance (Badia, et al., 1991; Campbell & Dawson, 1990; French, et al., 1991) greater power in the EEG beta frequency band (Badia, et al., 1991) and a decrease in the latency and efficiency of short duration (7 min.) naps (Sack, et al., 1992). Thus, attenuating the nighttime expression of endogenous melatonin with bright light administration attenuates the nighttime decline in body temperature and alertness and attenuates the nighttime rise in sleepiness and sleep.

Daytime photic stimulation does not yield similar effects for body temperature (Badia, et al., 1991), subjective alertness (Badia, et al., 1991) or sleep latencies (Murphy, et al., 1991). That the effects of bright light are not found during the day, when endogenous melatonin levels are very low, suggests that the immediate effects of bright lights are mediated by a reduction in endogenous melatonin levels. Indeed, strong evidence for this relationship has been revealed by reversing the effects of nighttime bright light administration with simultaneous melatonin administration.

Simultaneous administration of exogenous melatonin and bright artificial light

As noted, the effects of bright lights (2,500 lux) on body temperature have been reversed with the simultaneous infusion of exogenous melatonin (0.05 μg/min.) from 2300 - 0500 hr (Strassman, Quallis, Lisansky, & Peake, 1991). Cagnacci, Soldani and Yen showed that, for women, two doses of oral melatonin (1 mg and .75 mg) given at 2030 hr and at 2300 hr reversed the temperature effects of light (3,000 lux) presented from 2100 - 0100 hr (Cagnacci, Soldani, & Yen, 1993). Hourly, oral doses of melatonin (0.5 mg) from 1900 - 0700 hr reverse the effects of all night bright light (2,500 lux) on body temperature and sleep propensity (the amount of sleep in hourly, 7 min., naps (Sack, et al., 1992). Thus, antagonizing nighttime melatonin secretion with bright lights increases alertness and decreases sleep and sleepiness. That these effects are reversed by the reintroduction of melatonin provides indirect support for melatonin's involvement in the temperature and sleep-wake rhythms.

Melatonin administration: body temperature and sleep

The most compelling evidence for melatonin's involvement in regulating sleep and sleepiness comes from the direct administration of exogenous melatonin. Lerner and Case (1960) first reported mild sedative effects of a 100-200 mg dose of exogenous melatonin given intravenously. In subsequent investigations, xMT proved to be a "potent inducer of sleep" (Lerner & Nordlund, 1978).

Exogenous melatonin injections induce sleep in chickens (Barchas, DaCosta & Spector, 1967; Hishikawa, Cramer & Kuhlo, 1969). In cats, direct bilateral administration of melatonin to the preoptic region induces sleep (Marczynski, Yamaguchi, Ling & Grodzinska, 1964). After xMT administration, sleep was preceded by slowing and synchronization of subcortical EEG activity as well as by an increase in the amplitude of subcortical EEG activity (Marczynski, Yamaguchi, Ling & Grodzinska, 1964). Similar, but reduced effects were found with

administration to the nucleus centralis medialis (Marczynski, Yamaguchi, Ling & Grodzinska, 1964). No sleep inducing effects were found after administration to the brain stem reticular formation.

Exogenous melatonin and nighttime sleep in humans

The *daytime* administration of high levels of exogenous melatonin does not improve *nighttime* sleep. Melatonin (250 mg) given orally 4 times a day (at unspecified times) for 6 days did not improve nighttime sleep in 3 young healthy men and 3 young healthy women (Fernandez-Guradiola & Anton-Tay, 1974). This sub-chronic (i.e., less than 3 weeks) administration slightly increased sleep latency and REM sleep latency. Daytime xMT decreased nighttime stage 4 sleep and increased stage 2 sleep. Finally, daytime xMT increased number of nighttime awakenings and increased REM density (Fernandez-Guradiola & Anton-Tay, 1974). Thus, administration of large doses of xMT during the daytime does not improve nighttime sleep.

Administering xMT at night, after nighttime melatonin secretion has been initiated, does not appear to facilitate normal sleep. For instance, James, Mendelson, Sack, Rosenthal, and Wehr (1987) administered 0 mg, 1 mg or 5 mg of oral xMT at 2245 hr to 7 men and 3 women (ages 21-40). Subjects slept between 2300 and 0700 hr. The only parameter of sleep affected by xMT was REM latency which was increased by the 5 mg dose only (James, et al., 1987). Using a much higher dose across 5 nights, Ferini-Strambi, et al. (1992) failed to find any significant effect of melatonin on nighttime sleep. In 6 healthy young males, giving 100 mg of xMT orally to subjects at 2230 hr for 5 consecutive nights did not improve polygraphically recorded sleep from 2300 - 0700 hr (Ferini-Strambi, et al., 1992). Thus, melatonin given at night does not appear to facilitate nighttime sleep. This could be due to the presence of elevated endogenous melatonin levels or to a ceiling effect on nighttime sleep or both.

Waldhauser, et al. (1990) used an insomnia paradigm to address the ceiling effect issue. This paradigm involved the all night (2230 - 0600 hr) presentation of tape recorded street noise (68-90 dB). Under these conditions, 80 mg of melatonin given orally at 2100 hr improved several measures of sleep. Compared to placebo, melatonin decreased sleep latency, decreased the number of awakenings and increased sleep efficiency (Waldhauser, Saletu, & Trinchard-Lugan, 1990). Melatonin also decreased the percentage of stage 1 sleep, increased the percentage of stage 2 sleep and decreased the mean REM interval (Waldhauser, et al., 1990). Two explanations for Waldhauser's results are presented. First, these subjects may have been given xMT prior to at the nighttime endogenous melatonin onset (DLMO) and that xMT administered close to the DLMO but not long after may improve sleep latency. Cramer, et al. (1974) reported shorter sleep latencies for subjects given oral melatonin at 2130 hr. 50 mg or 0 mg of xMT was injected at 2130 hr to 15 healthy subjects. Melatonin did not affect quantitative analysis of EEG recorded 10 min. after administration (Cramer, et al., 1974). Melatonin did reduce sleep latency (Cramer, et al., 1974) without significantly affecting any other measure of nighttime sleep. Thus, it is hypothesized that melatonin serves as a signal to the sleep-wake system that it is time to sleep. Administering exogenous melatonin before the endogenous signal has been sent can facilitate the initiation of sleep (Anton-Tay, Diaz & Fernandez-Guardiola, 1971; Cramer, et al., 1974; 1976; Dollins, et al., 1994; Vollrath, Semm & Gammel, 1981; Waldhauser, et al., 1990) while administering exogenous melatonin after the endogenous signal has been sent may have little effect (Ferini-Strambini, et al., 1992; James, et al., 1987). Given the results of these

investigations, the time of the hypothetical endogenous signal may be close to the DLMO. It should be noted, however, that rather than being a passive message, the "signal" may reflect a threshold of circulating melatonin above which sleep is actively facilitated and below which sleep is not. Second, Waldhauser's reported changes in sleep architecture are consistent with higher auditory thresholds. In sleep, lower body temperature levels (and presumably higher melatonin levels) are associated with increased auditory thresholds (Lammers, Badia, Hughes & Harsh, 1991). During nighttime sleep, auditory threshold is inversely associated with body temperature (i.e., follows an inverse U function). So, melatonin may be related to auditory threshold. In Waldhauser's investigation, exogenous melatonin administration may have reduced the adverse effects of the noise by increasing auditory thresholds, particularly in the middle part of the night. It should be noted that melatonin did not significantly affect a terminal awakening threshold tested in the morning (Waldhauser, et al., 1990). Nevertheless, the sleep architecture results do not appear to reflect pharmacological sleep. Since sleep in the placebo group was fragmented by the noise, likely resulting in more stage 1, less stage 2 sleep and shorter REM intervals, this condition may not have been the proper control condition for the comparison of "natural" sleep. The sleep architecture of melatonin should be compared with normal nighttime sleep. With this comparison, melatonin has already shown to have little to no effect on sleep architecture (Cramer, et al., 1974; Ferini-Strambini, et al., 1992; James, et al., 1987). For now, it is hypothesized that the relatively small effects of xMT at night are likely due to the preexisting presence of high levels of endogenous melatonin which are already affecting the sleep-wake system. That xMT may have hypnotic effects only in the absence of high levels of endogenous melatonin provides further support for the role of endogenous melatonin in the regulation of the sleep-wake cycle.

Melatonin and daytime sleep in humans

Anton-Tay et al. (1971) reported the first investigation assessing melatonin's sleep inducing effects in humans. Melatonin (between 0.25 mg/kg and 1.25 mg/kg n = 5, and 1.25 mg/kg n = 6) was administered intravenously at 1600 hr. Subjects were required to lay down and perform several psychomotor and cognitive tasks, including a time estimation task. Compared to placebo (solvent injections) melatonin had large hypnotic effects seen initially in parieto-occipital EEG deactivation. Melatonin also increased time interval estimates of successive light pulses and slightly increased reaction time. Fifteen to twenty minutes after melatonin administration all subjects fell asleep and were allowed to sleep for 45 min. Following the nap, melatonin increased the percentage and amplitude of alpha frequency (Anton-Tay, et al., 1971). Cramer et al. (1974) reported that the daytime (1600) intravenous administration of 50 mg of melatonin facilitated polygraphically recorded sleep latency. In another investigation, Cramer, Bohme, Kendel & Donnadieu (1976), administered 50 mg of xMT, intravenously, between 1600 - 1700 hr to healthy young males. Here, melatonin decreased both the latency to sleep and the latency to slow wave sleep (Cramer, et al., 1976). Vollrath and colleagues reported improvements of non-polygraphically scored sleep latencies for 1.7 mg of melatonin administered intranasally from 0900 - 1000 hr (Vollrath, et al., 1981). Using a crossover design, Lieberman, et al. (1984) administered oral doses of xMT (80 mg each) or placebo at 1200, 1300, and 1400 hr to 14 male volunteers (ages 18-45). Compared to placebo, xMT decreased oral temperature, increased subjective self-ratings of fatigue, sleepiness and confusion (Lieberman, et al., 1984). Melatonin also decreased

subjective measures of activity and vigor, slowed reaction times and decreased errors of commission on a four choice reaction time task (Lieberman, et al., 1984). Dollins, and colleagues (1993) replicated these effects, on 20 healthy males, using the administration of a single dose at 1145 hr. For this administration, all doses of melatonin (10 mg, 20 mg, 40 mg, or 80 mg) decreased oral temperature, increased subjective measures of fatigue and sleepiness and decreased subjective self-ratings of vigor and activity (Dollins, et al., 1993). Melatonin also slowed psychomotor reaction times on a four choice task and impaired accuracy on a long-duration (60 min.) auditory vigilance task (Dollins, et al., 1993). This range of doses administered in the morning (0915 hr) yields similar effects. In another placebo controlled, double-blind design, Hughes administered xMT (0 mg, 10 mg & 100 mg) to young healthy males and measured oral temperature and subjective mood ratings. Melatonin decreased oral temperature, in a dose-dependent manner, increased subjective feelings of fatigue and sleepiness and decreased subjective feelings of activity and vigor (Hughes, 1992). In a placebo-controlled, double blind assessment of an indirect, behavioral, measure of sleep (a microswitch), Dollins et al. (1994) reported the effects of four relatively small doses of melatonin (0 mg, 0.1 mg, 0.3 mg, 1 mg and 10 mg). Melatonin decreased oral temperature, increased subjective fatigue and sleepiness and decreased accuracy on an auditory vigilance task (Dollins, et al., 1994). All but the smallest dose of xMT shortened sleep latency in a short duration (30 min.) daytime nap (Dollins, et al., 1994). Hughes et al. (1994) used a placebo controlled, double-blind, crossover design to assess the hypnotic efficacy of 3 doses of xMT (1mg, 10 mg, & 40 mg) administered at 1000 hr. Compared to placebo, all doses of melatonin shortened sleep latency. Melatonin, especially the higher doses, increased sleep duration in a four sleep episode (1200 - 1600 hr). In fact, all subjects in these two conditions were still asleep at the end of the fourth hour. Melatonin also reduced stage 3/4 sleep and increased stage 2 sleep. This may reflect benzodiazepine like affects on sleep architecture, however, melatonin did not appear to yield a reduction in delta frequency amplitude associated with benzodiazepines. Additionally, melatonin did not increase sleep spindles and did not appear decreased alpha frequency amplitude (Hughes, et al., 1994). In fact, the sleep architecture of melatonin induced naps may have been more like physiological nighttime sleep than the placebo nap (Hughes, et al., 1994). Melatonin did not yield anterograde amnesia tested before and after the nap. Further, melatonin did not impair performance tested after the nap (Hughes, et al., 1994). In conclusion, melatonin is a safe, naturally occurring hormone that when administered exogenously is efficacious in initiating and sustaining sleep. The sleep induced by melatonin may be more physiological and thus more restorative than benzodiazepine induced sleep. Finally, melatonin does not yield the negative side effects associated with the benzodiazepines. Therefore, melatonin may prove to be a safe and efficacious alternative to currently prescribed hypnotics.

References

Åkerstedt, T. (1983). A questionnaire study of sleep/wake disturbances and irregular work hours. *Sleep Research, 12*, 358.

Åkerstedt, T., Gillberg, M., & Wetterberg, L. (1982). The circadian covariation of fatigue and urinary melatonin. *Biological Psychiatry, 17*, 547-554.

Åkerstedt, T. (1988). Sleepiness as a consequence of shift work. Sleep, 11(1), 17-34.

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., Aldhous, M. & Marks, V. (1986). Alleviation 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. *Ergonomics, 30*, 1379-1393.

Arendt, J., Aldhous, M., & Wright, J. (1988). Synchronization of a disturbed sleep-wake cycle in a blind man by melatonin treatment. *The Lancet, i.*, 772-773.

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.

Armstrong, S. M., (1989). Melatonin: The internal zeitgeber of mammals? *Pineal Research Reviews, 7*, 157-202.

Avery, D., Wildschiotz, G., & Rafaelson, O. J. (1982). Nighttime temperature in affective disorders. *J. of Affective Disorders, 4*, 61.

Badia, P., Myers, B., Boecker, M., Culpepper, J., & Harsh, J. (1991). Bright light effects on body temperature, alertness, EEG and behavior. *Physiology and Behavior, 50*, 583-588.

Badia, P., Myers, B., & Murphy, P. (1992). Melatonin and thermoregulation. In H. Yu & R. Reiter (eds.) *Melatonin: Biosynthesis, physiological effects, and clinical applications* (pp. 349-364). Boca Raton, FL: CRC.

Barchas, J., DaCosta, F., & Spector, S. (1967). Acute pharmacology of melatonin. *Nature, 214*, 919-920.

Birkeland, A. J. (1982). Plasma melatonin levels and nighttime transitions between sleep and wakefulness. *Neuroendocrinology, 34*, 126-131.

Brown, R., Kocsis, J. H., Caroff, S., Amsterdam, J., Winokur, A., Stokes, P. E., & Frazer, A. (1985). Differences in nocturnal melatonin secretion between melancholic and depressed patients and control subjects. *American Journal of Psychiatry, 142*, 811-816.

Borbély, A. A., Åkerstedt, T., Benoit, O., Holsboer, F. & Oswald, I. (1991). Hypnotics and sleep physiology: a consensus report. *Eur. Arch. Psychiatry Clin. Neuroscience, 241*, 13-21.

Cagnacci, A., Elliott, J. & Yen, S. (1992). Melatonin: a major regulator of the circadian rhythm of core temperature in humans. *Journal of Clinical Endocrinology, 74*, 447-452.

Cagnacci, A., Soldani, R. & Yen, S. (1993). The effect of light on core body temperature is mediated by melatonin in women. *Journal of Clinical Endocrinology and Metabolism, 76*, 1036-1038.

Campbell, S. S. (1992). Effects of Sleep and Circadian Rhythms on Performance. In D. M. Jones & A. P. Smith (eds.), *Handbook of Human Performance, 3*. New York and London: Academic Press, pp. 195-216.

Campbell, S. S., & Broughton, R. J. (1994). Rapid decline in body temperature before sleep: Fluffing the physiological pillow? *Chronobiology International, 11*(2), 126-131.

Campbell, S. S., & Dawson, D. (1990). Enhancement of nighttime alertness and performance with bright ambient light. *Physiology and Behavior, 48*, 317-320.

Campbell, S. S., Gillin, J. C., Kripke, D. F., Erikson, P., & Clopton, P. (1989). Gender differences in the circadian temperature rhythms of healthy elderly subjects: relationships to sleep quality. *Sleep, 12*, 529-536.

Carman, J. S., Post, R. M., Buswell, R., & Goodwin, F. K. (1976). Negative effects of melatonin on depression. *American Journal of Psychiatry, 113*(10), 1181-1186.

Carskadon, M. A., Brown, E. D., & Dement, W. C. (1982). Sleep fragmentation in the elderly, relationship to daytime sleep tendency. *Neurobiology of Aging, 3*, 321-327.

Carskadon, M. A., Seidel, W. F., Greenblatt, D. J., & Dement, W. C. (1982). Daytime carryover of Triazolam and Flurazepam in elderly insomniacs. *Sleep, 5*(4), 361-371.

Claustrat, B., Chazot, G., Brun, J., Jordan, D., & Sassolas, G. (1984). A chronological study of melatonin and cortisol seeretion in depressed subjects: Plasma melatonin, a biochemical marker in major depression. *Biological Psychiatry, 19*, 1215-1228.

Cramer, H., Rudolph, J. & Consbruch, U., Kendel K. (1974). On the effects of melatonin on sleep and behavior in man. In *Serotonin: new vistas, biochemistry and behavioral and clinical studies* (E. Costa Ed.) Advances in Biochemical Psychopharmacology Volume 2, pp. 187-191. Raven Press, New York.

Cramer, H., Bohme, W., Kendel, K., & Donnadieu, M. (1976). Liberation of growth hormone and melanocyte-stimulating hormone in natural and melatonin induced human sleep. *Arzneim. - Forsch. (Drug Research), 26*, 1076-1078.

Curran, H. V. (1986). Tranquilizing memories: a review of the effects of benzodiazepines on human memory. *Biological Psychology, 23*, 179-231.

Czeisler, C. A., Dumont, M., Duffy, J. F., Steinberg, J. D., Richardson, G. S., Brown, E. N., Sánchez, R., Ríos, C. D., & Ronda J. M. (1992). Association of sleep-wake habits in older people with changes in output of circadian pacemaker. *The Lancet, 340*, 933-936.

Czeisler, C. A., Zimmerman, J. C., Ronda, J. M., Moore-Ede, M. C., & Weitzman, E. D. (1980). Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep, 2*, 329-346.

Dahlgren, K., (1981). Adjustments of circadian rhythms and EEG sleep functions to day and night sleep among permanent night workers and rotating shift workers. *Psychophysiology, 18*, 981-391.

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.

Dawson, D., & Encel, N. (1993). Melatonin and sleep in humans. *Journal of Pineal Research, 15*, 1-12.

Dimenas, E., Kerr, D., & MacDonald, I. (1990). Beta-adrenoceptor blockade and CNS-related symptoms: a randomized, double-blind, placebo-controlled comparison of metaprolol CR/ZOK, atenolol and propranolol LA in healthy subjects. *J. Clin. Pharmacol. 30*(2 Suppl), S103-S107.

Dollins, A. B., Lynch, H. J., Wurtman, R. J., Deng, M. H., Kischka, K. U., Gleason, R. E., & Lieberman, H. R. (1993). Effect of pharmacological daytime doses of melatonin on human mood and performance. *Psychopharmacology, 112*, 490-496.

Dollins, A. B., Zhdanova, I. V., Wurtman, R. J., Lynch, H. J., & Deng, M. H. (1994). Effect of inducing nighttime serum melatonin concentrations in daytime sleep, mood, body temperature, and performance. *Proceedings of the National Academy of Sciences, 91*, 1824-1828.

Feinberg, I. (1969). Effects of age on human sleep patterns. In A. Kales (Ed.), *Sleep: Physiology and Pathology* (pp. 39-52). Philadelphia: Lippincott.

Ferrari, E., Fraschini, F., & Brambilla, F. (1990). Hormonal circadian rhythms in eating disorders. *Biological Psychiatry, 27*, 1007.

Ferini-Strambi, L., Zucconi, M., Biella, G., Oldani, A., Stankov, B., Fraschini, F., & Smirne, S. (1992). *Sleep Research.*

Folkard, S., Arendt, J., Aldhous, M., & Kennet, H. (1990). Melatonin stabilizes sleep onset time in a blind man without entrainment of cortisol or temperature rhythms. *Neuroscience Letters, 113*, 193-198.

Food and Drug Administration: (1985). FDA, adverse experiences spontaneous reporting system: comparison of Restoril, Dalmane and Halcione. Washington, DC, US Government Printing Office.

French, J., Boll, P., Storm, W. F. & Dowd, P. (1989). Temazepam and performance following a sleep cycle shift. *Annual Review of Chronopharmacology, 7*, 41-44.

French, J., Hannon, P., & Brainard, G. C. (1991). Photic effects on sustained performance. *Proceedings of Space Operations, Applications and Research,* Houston, TX.of bright illuminance on body temperature and human performance. *Annual Review of Chronopharmacology, 7*, 37-40.

Gander, P., & Graeber, R. C. (1987). Sleep in pilots flying short-haul commercial schedules. *Erg., 30*, 1365-1377.

Gillberg, M., & Åkerstedt, T. (1982). Body temperature and sleep at different times of the day. *Sleep, 5*, 378-388.

Gillin, J. C., Mendelson, W. B., & Kupfer, D. J. (1988). The sleep disturbances of depression: Clues to pathophysiology with special reference to the circadian rapid eye movement rhythm. In D. J. Kupfer, T. H. Monk, & J. D. Barchas (Eds.), *Biological Rhythms and Mental Disorders* (pp. 27-54). New York: Guilford.

Graeber, R. C. (1988). Aircrew fatigue and circadian rhythmicity. In E.L. Weiner, D.C Nagel (eds.): *Human Factors in Aviation.* New York, Academic Press.

Green, D. J., & Gillette, R. (1982). Circadian rhythm of firing rate recorded from single cells in the rate suprachiasmatic slice. *Brain Research, 245*, 198-200.

Greenblatt, D. J., & Shader, R. I. (1974). *Benzodiazepines in Clinical Practice.* New York, Raven Press.

Hishikawa, Y., Cramer, H., & Kuhlo, W. (1969). Natural and melatonin-induced sleep in young chickens - A behavioral and electrographic study. *Experimental Brain Research, 7*, 84-95.

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. (1992). The effects of two doses of exogenous melatonin on temperature and subjective fatigue. Final report submitted to the Air Force Office of Scientific Research.

Hughes, R., Badia, P., French, J., Santiago, L., & Plenzler, S. (1994). Melatonin induced changes in body temperature and daytime sleep. *Sleep Research, 23*, 496.

ICSD - International Classification of Sleep Disorders: Diagnostic and coding manual. Diagnostic Classification Steering Committee, Thorpy, M. J., Chairman. Rochester, Minnesota: Amer. Sleep Disorders Association, 1990.

Iguchi, H., Kato, K., & Ibayashi, H. (1982). Age-dependent reduction in serum melatonin concentrations in healthy human subjects. *J. of Clin. Endocrinol. Metab.*, *55*, 27.

James, S. P., Mendelson, W. B., Sack, D. A., Rosenthal, N. E., & Wehr, T. A. (1987). The effect of melatonin on normal sleep. *Neuropsychopharmacology*, *1*(1), 41-44.

Kay, D. C., Blackburn, A. B., Buckingham, J. A., & Karacan, I. (1976). Human pharmacology of sleep. In R. L., Williams, & I., Karacan, (Eds.), *Pharmacology of Sleep*, pp. 83-210, New York, Wiley & Sons.

Kennaway, D. J., & Royles, P. (1986). Circadian rhythms of 6-sulphatoxymelatonin, cortisol and electrolyte excretion at the summer and winter solstices in normal men and women. *Acta Endocrinologia*, *113*, 450-456.

Kennaway, D. J., & Van Dorp, C. F. (1991). Free-running rhythms of melatonin, cortisol, electrolytes and sleep in humans in Antarctica.

Keran, C. M., Smith, T. J., Duchon, J. C., Roginson, D., & Trites, D. (1991). Shift-related effects of psychophysiologic performance in underground workers. In. W. Karwowski, & J.W. Yates (Eds.), *Advances in Industrial Ergonomics and Safety III*, Taylor & Francis.

Kupfer, D. J., Frank, E., Jarrett, D. B., Reynolds III, C. F., & Thase, M. E. (1988). Interrelationship of electroencephalographic sleep chronobiology and depression. In D. J. Kupfer, T. H. Monk, & J. D. Barchas (Eds.), *Biological Rhythms and Mental Disorders* (pp. 1-26). New York: Guilford.

Lavie, P., Gopher, D., & Wollman, M. (1987). Thirty-six hour correspondence between performance and sleepiness cycles. *Psychophysiology*, *24*(4), 430-438.

Lavie, P., & Shulamit, S. (1989). Twenty-four-hour structure of sleepiness in morning and evening persons investigated by the ultrashort sleep-wake cycle. *Sleep*, *12*, 522-528.

Lammers, W., Badia, P., Hughes, R., & Harsh, J. (1991). Temperature, time-of-night testing, and responsiveness to stimuli presented while sleeping. *Psychophysiology*, *28*. 463-467.

Lerner, A. B., & Case, J. C. (1960). Melatonin. *Fed. Proc.*, *19*, 590-592.

Lerner, A. B., & Nordlund, J. J. (1978). Melatonin: Clinical pharmacology. *J. of Neural. Trans., Suppl. 13*, 339-347.

Lewis, S. A. (1969). Sleep patterns during afternoon naps in the young and elderly. *Brit. J. of Psychiatry*, *115*, 107-108.

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.

Lewy, A. J., & Newsome, D. A. (1983). Different types of melatonin circadian secretory rhythms in some blind subjects. *J. of Clin. Enodocrin. Metab.*, *56*, 1103-1107.

Lewy, A. J., & Sack, R. L. (1989). The dim light melatonin onset (DLMO) as a marker for circadian phase position. *Chronobiology International*, *6*, 93-102.

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.

Leung, L, & Becker, C. E. (1992). Sleep deprivation and house staff performance: Update 1984-1991. *Journal of Occupational Medicine*, *34*(12), 1153-1160.

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

Marotte, H. & Timbal, J. (1981). Circadian rhythm of temperature in man: comparative study with two experiment protocols. *Chronobiologia*, *8*, 97-100.

Martens, H., Endlich, H., Hildebrandt, G., & Moog, R. (1990). Sleep/wake distribution in blind subjects with and without sleep complaints. *Sleep Research*, *19*, 398.

Marczynski, T. J., Yamaguchi, N., Ling, G. M. & Grodzinska, L. (1964). Sleep induced by the administration of melatonin (5-Methoxy-N-acetyltryptamine) to the hypothalamus in unrestricted cats. *Experientia*, *20*, 435-437.

McArthur, A. J., Gillette M. U., & Prosser, R. A. (1991). Melatonin directly resets the rat suprachiasmatic circadian clock in vitro. *Brain Research*, *565*, 158-161.

McMillen, I. C., Mulvogue, H. M., Kok, J. S. M., Deayton, J. M., Nowak, R., & Adamson, T. M. (1993). Circadian rhythms in sleep and wakefulness and in salivary melatonin and cortisol concentrations in mothers of term and preterm infants. *Sleep*, *16*, 624-631.

Miles, L. E., & Dement, W. C. (1980). Sleep and aging. *Sleep*, *3*, 119-220.

Miles, L. E. M., & Raynal, D. M., & Wilson, M. A. (1977). Blind man living in normal society has circadian rhythm of 24.9 hours. *Science, 198*, 421-423.
Miles, L. E. M., & Wilson, M. A. (1977). High incidence of cyclic sleep/wake disorders in the blind. *Sleep Research, 6*, 192.
Mitler, M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F., & Graeber, R. C. (1988). Catastrophes, sleep and public policy: Consensus report. *Sleep, 11*(1), 100-109.
Monk, T. H., Moline, M. L., Fookson, J. E. & Peetz, S. M. (1991). Circadian determinants of subjective alertness. *Journal of Biological Rhythms, 4*(4), 393-404.
Moore-Ede, M. C., & Richardson, G. S. (1985). Medical implications of shiftwork. *Ann. Rev. of Med., 36*, 607-617.
Morris, H. H., & Estes, M. L. (1987). Traveler's amnesia: transient global amnesia secondary to triazolam. *Journal of the American Medical Association, 258*, 945-946.
Morris, M., Lack, L., & Barrett, J. (1990). The effect of sleep/wake state on nocturnal melatonin secretion. *Journal of Pineal Research, 9*, 133-138.
Murphy, P., Myers, B., Badia, P., & Harsh, J. (1991). The effects of bright light on daytime sleep latencies. *Sleep Research, 20*, 465.
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.
Nakagawa, H., Sack, R. L., & Lewy, A. J. (1992). Sleep propensity free-runs with the temperature, melatonin and cortisol rhythms in a totally blind person. *Sleep, 15*(4), 330-336.
Nakazawa, Y., Nonaka, K., Nishida, N., Hayashida, N., Miyahara, Y., Kotorii, T., & Matsuaoka, K. (1991). Comparison of body temperature rhythms between healthy elderly and healthy young adults. *Japanese Journal of Psychiatry, 45*, 37-43.
Novak, R. D., Smolensky, M. H., Fairchild, E. J., & Reves, R. R. (1990). Shiftwork and industrial injuries at a chemical plant in southwest Texas. *Chronobiology International, 7*(2), 155-164.
Okawa, M., Nanami, T., Wada, S., Shimizu, T., Hishikawa, Y., Sasaki, H., Nagamine, H., & Takahashi, K. (1987). Four congenitally blind children with circadian sleep-wake rhythm disorder. *Sleep, 10*, 101-110.
Palm, L., Blennow, G., & Wetterberg, L. (1991). Correction of non-24-hour sleep/wake cycle by melatonin in a blind retarded boy. *Annals of Neurology, 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.
Perez, W. A., Masline, P. J., Ramsey, E. G. & Urban, K. E. (1987). *Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology (U).* (Report No. AAMRL-TR-97-007). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory, Human Systems Division, US Air Force.
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.
Petrie, K., Dawson, A. G., Thompson, L., & Brook, R. (1993). A double-blind trial of melatonin as a treatment for jet-lag in international cabin crews. *Biological Psychiatry, 33*, 526-530.
Poeggeler, B., Balzer, I., Hardeland, R. & Lerchl, A. (1991). Pineal hormone melatonin oscillates also in the dinoflagellate *Gonyaulax polyedra. Naturwissenschaften, 78*, 268-269.
Rall, T. W. (1990). Hypnotics and sedatives; Ethanol. In A.G., Gilman, T.W. Rall, A.S. Nies, & P. Taylor (eds.) *The Pharmacological Basis of Therapeutics*, Chapter 17, pp. 345-382.
Ralph, M. R., Foster, R. G., Davis, F. C., & Menaker, M. (1990). Transplanted suprachiasmatic nucleus determines circadian period. *Science, 247*, 975-978.
Rechtschaffen, A., & Kales, A. (Eds.). (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Washington, DC: US Government Printing Office.
Redman, J., Armstrong, S. & Ng, K. T. (1993). Free-running activity rhythms in the rat: Entrainment by melatonin. *Science, 219*, 1089-1091.
Refinetti, R., & Menaker, M. (1992). The circadian rhythm of body temperature. *Physio. and Behav., 51*, 613-637.
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 Sci., 6*, 223-227.
Reiter, R. J. (1991b). Melatonin: the chemical expression of darkness. *Mol. & Cel. Endocr., 79*, C153-C158.

Reiter, R. J. (1991c). Neurendocrine effects of light. *International Journal of Biometeorology, 35*, 169-175
Reiter, R. J. (1991d). Pineal melatonin: Cell biology of its synthesis and of its physiological interactions. *Endocrine Reviews, 12*(2), 151-180.
Reiter, R. J. (1992). The aging pineal gland and its physiological consequences. *BioEssays, 14*(3), 169-175.
Reiter, R. J., Poeggeler, B., Tan, D.-X., Chen, L.-D., Manchester, L.C., Guerrero, J. M. (1993). Antioxidant capacity of melatonin: a novel action not requiring a receptor. *Neuroendocrinol. Letters, 15*, 103-116.
Reppert, S. M., Weaver, D. R., Rivkees, S. A., & Stopa, E. G. (1988). Putative melatonin receptors in a human biological clock. *Science, 242*, 78-81.
Richardson, G., Carskadon, E., Orav, E. & Dement, W. (1982). Circadian variation of sleep tendency in elderly and young adult subjects. *Sleep, 5*, S82-S94.
Roth, T., Hartse, K. M., Saab, P. G., Piccione, P. M., & Kramer, M. (1980). The effects of Flurazepam, Lorazepam, and Triazolam on sleep and memory. *Psychopharmacology, 70*, 231-237.
Saarela, S., & Reiter, R. J. (1993). Function of melatonin in thermoregulatory processes. *Life Sci., 54*, 295-311.
Sack, R. L., Lewy, A. J., Blood, M. L., Keith, L. D., & Nakagawa, H. (in press). Circadian rhythm abnormalities in totally blind people; incidence and clinical significance. *Journal of Clinical Endocrinology and Metabolism.*
Sack, R. L., Lewy, A. J., Blood, M. L., Stevenson, J., & Keith, L. D. (1991). Melatonin administration to blind people: Phase advances and entrainment. *Journal of Biological Rhythms, 6*(3), 249-261.
Sack, R. L., Lewy, A. J. & Hoban, T. M. (1987). Free-running melatonin rhythm in blind people: Phase shifts with melatonin and Triazolam administration. In L. Rensing, U. van der Heiden, & M. C. Mackey, (eds.), *Temporal Discovery in Human Oscillatory Systems*, Springer, Heidelberg. 129-224.
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.
Schulz, H. & Lund, R. (1983). Sleep onset REM episodes are associated with circadian parameters of body temperature. A study in depressed patients and normal controls. *Biological Psychiatry, 14*, 595-600.
Sharma, M., Palacios-Bois, J., Schwartz, G., Iskandar, H., Thakur, M., Quirion, R., & Nair, N. P. V. (1989). Circadian rhythms of melatonin and cortisone in aging. *Biological Psychiatry, 25*, 305.
Smith, A. P. (1992). Time of day and performance. In D. M. Jones & A. P. Smith (eds.), *Handbook of Human Performance, 3*. New York and London: Academic Press, pp. 217-235.
Smith, J. R., Karacan, I., & Yang, M. (1977). Ontogeny of delta activity during human sleep. *Electroencephalography and Clinical Neurophysiolgy, 43*, 229-237.
Stankov, B., Fraschini, F., & Reiter, R. J. (1991). Melatonin binding sites in the central nervous system. *Brain Research Reviews, 16*, 245-256.
Stankov, B., & Reiter, R. J. (1990). Melatonin receptors: current status, facts and hypotheses. *Life Sciences, 46*, 971-982.
Storm, W. G. & Parke, R. C. (1986). FB-111A aircrew use of temazepam during surge operations AGARD-CP-415, 12-1 - 12-11.
Strogatz, S., et al. (1986). Circadian regulation dominates homeostatic control of sleep length and prior wake length in humans. Sleep, 9(2), 353-364.
Sugden, D. (1983). Psychopharmacological effects of melatonin in mouse and rat. *Journal of Pharmacology and Experimental Therapeutics, 227*(3), 587-591.
Tan, D. -X., Chen, L.-D., Poeggeler, B., Manchester, L.C., & Reiter, R. J. (1993). Melatonin: a potent, endogenous hydroxyl scavenger. *Endocr. J., 1*, 57-60.
Touitou, Y., Fevre, M., Bogdan, A., Reinberg, A., de prins, J., Beck, H., & Touitou, C. (1984). Patterns of plasma melatonin with aging and mental condition: stability of nyctohemeral rhythms and differences in seasonal variation. *Acta Endocrinologia, 106*, 145.
Touitou, Y., Reinberg, A., Bogdan, A., Auzéby, A., Beck, H., & Touitou, C. (1986). Age-related changes in both circadian and seasonal rhythms of rectal temperature with special reference to senile dementia of Alzheimer type. *Gerontology, 32*, 110-118.
Tune, G. (1969). The influence of age and temperament on the adult human sleep-wakefulness pattern. *British Journal of Psychology, 60*, 431-441.
Tzischinsky, O., Skene, D., Epstein, R., & Lavie, P. (1991). Circadian rhythms in 6-Sulphatoxymelatonin and nighttime sleep in blind children. *Chronobiology International, 8*(3), 168-175.

Van Coevorden, A., Mockel, J., Laurent, E., Kerkhofs, M., L'Hermite-Balériaux, M., Decoster, C., Néve, P., & Van Cauter, E. (1991). Neuroendocrine rhythms and sleep in aging men. *American Journal of Physiology, 260, (Endocrinol. Metab. 23)* E651-E661.

Vitiello, M. V., Smallwood, R., G., Avery, D. H., Pascually, R. A., Martin, D. C., Prinz, P. N. (1986). Circadian temperature rhythms in young adult and aged men. *Neurobiology of Aging, 7*, 97-100.

Vollrath, L., Semm, P. & Gammel, G. (1981). Sleep induction by intranasal application of melatonin. *Advances in the Biosciences, 29*, 327-329.

Waldhauser, F., & Dietzel, M. (1992). Daily and annual rhythms in human melatonin secretion: Role in puberty control. *Annals New York Academy of Sciences*, 205-213.

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.

Waldhauser, F., Ehrhart, B., & Förster, (1993). Clinical aspects of the melatonin action: impact of development, aging and puberty, involvement of melatonin in psychiatric disease and importance of neuroimmunoendocrine interactions. *Experientia, 49*, 671-681.

Waldhauser, F., Waldhauser, M., Lieberman, H. R., Deng, M., Lynch, H. J., & Wurtman, R. J. (1984). Bioavailability of oral melatonin in humans. *Neuroendocrinology, 39*, 307-313.

Walsh, J. K., Tepas, D. I., & Moss, P. D. The EEG sleep of night and rotating shift workers. In L.C., Johnson, D.I., Tepas, W.P., Colquhoun, & M.J., Colligan, (eds.). *Biological Rhythms, Sleep and Shift Work.* New York, Spectrum, 1981, PP. 371-381.

Walsh, J. K., & Sugarman, J. L. (1988). Staying awake at night: Relation to physiological sleep tendency. *Sleep Research*, 18, *400.*

Webb, W. B. (1994). Sleep as a biological rhythm: a historical review. *Sleep, 17*(2), 188-194.

Webb, W. B., & Swinburne, H. (1971). An observational study of sleep in the aged. *Perception and Motor Skills, 32*, 895-898.

Weitzman, E. D., Moline, M. L., Czeisler, C. A., & Zimmerman, J. C. (1982). Chronobiology of aging: Temperature, sleep-wake rhythms and entrainment. *Neurobiology of Aging, 3*, 299-309.

Wetterberg, L., (1978). Melatonin in humans: physiological and clinical studies. *J. of Neur. Trans. Supp., 13*, 289.

Wetterberg, L., Beck-Friis, J., Kjellman B. F., & Ljunggren, J. -G. (1984). Circadian rhythms in melatonin and cortisol secretion in depression. In E. Usdin, et al. (Eds.), *Frontiers in Biochemical and Pharmacological Research in Depression*, New York, Raven Press.

Wever, R. A. (1988). Order and disorder in circadian rhythmicity: Possible relations to mental disorders. In D. J. Kupfer, T. H. Monk, & J. D. Barchas (Eds.), *Biological Rhythms and Mental Disorders* (pp. 253-346). New York: Guilford.

Winget, C. M., DeRoshia, C. W., Markley, C. L., & Holley, D. C. (1984). A review of human physiological and performance changes associated with desynchronosis of biological rhythms. *Aviat. Space Environmental Medicine, 55*(12), 1085-1093.

Zepelin, H., & McDonald, C. S. (1987). Age differences in autonomic variables during sleep. *Journal of Gerontology, 42*, 142-146.

Zulley, J. & Campbell, S. S. (1985). Napping behavior during "spontaneous internal desynchronization": sleep remains in synchrony with body temperature. Human Neurobiology, 4, 123-126.

Zulley, J., et al. (1981). The dependence of onset and duration of sleep on the circadian rhythm of rectal temperature. Pflugers Arch, 391, 314-318.

The Measurement of Work Experience: Issues and Implications

Roman G. Longoria
Rice University
Department of Psychology
P.O. Box 1892
Houston, Texas 77251

Final Report for:
Graduate Student Research Program
Armstrong Laboratories

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

September 1994

The Measurement of Work Experience: Issues and Implications

Roman G. Longoria
Rice University
Department of Psychology

Abstract

There has been a wealth of research regarding the effects of work experience on various outcomes. Most of this research has used tenure as the means to quantify experience so that the more organizational or position tenure an incumbent has, the more experience implied. More recent research has provided a framework for classifying work experience in terms of what is done, how often, and for how long. Specific links should be made between the criterion and the predictor, so if one wanted to make specific predictions of the relationship between experience and specific task performance, then one should quantify work experience at the task level. The type of data needed to quantify specific task experience can be obtained through a technique similar to the task or job analysis. However, before a more data-driven quantification of work experience can be implemented, systematic investigations are needed to determine the factors affecting the validity and reliability of this measurement technique. The present paper addresses these issues, reviews the literature regarding the accuracy of work experience ratings, and discusses how this research can apply to personnel selection and the evaluation of training programs.

The Measurement of Work Experience: Issues and Implications

Roman G. Longoria

The relationship between varieties of work experiences and subsequent outcomes has been studied for quite some time. For instance, a recent review and meta-analysis of the relationship between work experience and job performance revealed a consistent, positive relationship (Quinones, Ford, & Teachout, 1994). However, the study also noted inconsistencies in the measurement of work experiences across studies. Most of the studies included in their meta-analysis used tenure, or time in position or in the organization, as the means of quantifying work experience.

The Quinones, et al. meta-analysis (1994) provides a framework that classifies work experience into three dimensions: amount, type, and time. It also incorporates these three dimensions into three levels of analysis: task, job, and organizational level. At the task level, amount would represent the number of times a task was performed, type of task would represent its difficulty or complexity, and time on a task would measure the number of minutes, days, months , or years spent performing the task. For the job level, amount would detail the number of different tasks (breadth) a job entails, type would describe the job in terms of complexity or prestige, and time would denote tenure in that position. Amount of experience at the organizational level would be a measure of the number of jobs held, while type at this level would describe the variety of organizations, such as research and development or manufacturing. Finally, time at the organizational level would be a simple measure of organizational tenure. This classification scheme provides the beginning of a systemic view of the work environment and how each component (i.e., task, job, organization) may provide its own experiential stimuli. What effects work experience may have certainly depends on how experience is measured (Quinones, et al., 1994).

The Measurement of Work Experience

What is experience? A laypersons' viewpoint would equate experience with age. We consider the elderly to be more experienced. But what is it about being around for a longer period of time that makes a person more experienced? If we consider work experience to be exposure to a variety of stimuli (amount, type), then time would be a necessary but not sufficient manner in which to quantify a job incumbent's level of experience. In other words, knowing how long an incumbent has been doing the job

is not as useful as knowing exactly was done during that time. But time in position or organizational tenure is a common method of quantifying work experience in previous studies (Quinones, et al., 1994).

The intention of using information about an employee's past work experience to predict future performance should be based on two assumptions. First, one must specifically define the construct of the performance criteria (e.g., management skills). Second, one should establish what specific previous experience would be predictive of the criteria (i.e., previous management duties). But the issue arises regarding the best way to measure one's experience with management duties. In industry, historical sources of experiential data has come from an incumbent's personnel records, supervisor interviews, or the use of retrospective self-reports. For instance, an applicant for a management position may be asked "Have you had any management experience?" or "For how long did you serve in a management position, and what did you do?". However, little is known regarding the validity and reliability of data collected in these manners. Cornelius and Lynes (1980) contend that human judges are not particularly effective decision makers when information on many dimensions is available. But there is data to support that incumbents can provide an accurate self-assessment of their own KSAs (Levine, Florry, & Ash, 1977). In addition, Russel, Mattson, Devlin, and Atwater (1990) suggest that biographical information has the potential for improving the prediction of criteria, and Pannone (1984) supports the superiority of using specific biographical data regarding previous work experiences as screening devices rather than using broad screening criteria such as the attainment of a certain education level and/or years of work experience.

If one wanted to use an employee's work experience to predict specific performance at the task level, then one would want to be as specific in the measurement of previous task experience. If job performance is going to measured using a multi-dimensional approach, then "the most rational way to conduct criterion development begins with the job analysis" (Borman, 1991). Providing a multi-dimensional approach to the measurement of work experience may enhance the predictor-criterion linkage. To utilize the Quinones et al., (1994) work experience classification framework at the task or job level, one would want to quantify what tasks were conducted, how often, and for how long. The cornerstone to the measurement of this data is the job analysis, and this would satisfy the "need to link the job analysis with the criterion measure" (Schmitt, 1987, in Cranny & Doherty, 1988, pp. 312-322).

Job Analysis and Accuracy

A job analysis is a detailed dissection of a given job in terms of the specific tasks conducted in the job, the procedures through which these tasks are conducted, and other necessary work behaviors

such as supervision or decision making. It also includes a description of the context under which the components of the job are performed such as the work environment, interaction with equipment, working conditions, and methods of compensation, as well as a description of the KSAs needed for the performance of the job (McCormick, 1976, in Harvey, 1991). Since a job analysis is developed to be a measure of what is done in a job, how a job is performed, and under what conditions the job is performed, it is imperative that the analysis be done in as objective manner as possible. "Describing observable should be the sole goal of the job analysis" (Harvey, 1991, p. 75).

Objectivity of the job analysis is important for a variety of reasons. First, there is a legal necessity for establishing an objective description of a job to be used in making personnel decisions. Second, if the job analysis is also going to be used for research purposes, then it is paramount that the description provided be an accurate representation of what the job entails. If the job analysis includes a detailed, specific, and objective description of types of tasks, procedures, and other work behaviors, then it can provide an accurate representation of the work experiences common for the incumbents of that job. Thus work experience measured at the task level, using a job analysis technique, may be more predictive of specific task performance criterions.

There are a variety of sources from which job analysis or other work experience ratings can be gathered. These include the independent analysts, supervisor ratings, and self-reports (Arvey, Davis, McGowen, & Dipboye, 1982; Cranny & Doherty, 1988; Anderson, Warner, & Spencer, 1984; Landy & Vasey, 1991; Pannone, 1984). In addition to the sources of job analysis ratings, there are a variety of questioning techniques used in gathering data. Borman, Dorsey, and Ackerman (1992) used the Job Activities Checklist to measure the relative time spent on activities. Job incumbents would make ratings on a particular activity based on a 6-point scale, anchored with "no time spent on this activity" to "much more time than on other activities". Landy and Vasey (1991) provide an example of a more absolute task frequency measure. They had job incumbents rate how often they conducted a task using a 7-point scale ranging from task "not performed at all" to "performed at least once a day". Another method of measurement can be found in Pannone (1984) where job applicants were required to rate their previous work experience with regard to specific tasks on a 4-point scale ranging from "previous job did not require me to perform this task" to "I supervised others performing this task". Asher (1972) found that "hard", verifiable biographical items were better predictors of performance than "soft", unverifiable items, and Shaffer, Saunders, and Owens (1986) reported that more objective biographical data items were more consistently reported than subjective items. It is clear that there are multiple manners in

which to collect job analysis data; however, little is known about the conditions under which different question formats are more valid and more effective in performance prediction.

If the job analysis is going to be used as a measure of a work experience, then an investigation into the biases and errors to which the job analysis is susceptible may uncover and explain within job variance in rating scores due to a job incumbent's sub-group membership or other characteristics. In other words, do incumbents of the same job category report different types and levels of work experience, and is this discrepancy attributable to sub-group membership and/or individual differences? To answer these questions, a review was conducted on the literature regarding job analysis accuracy.

Biases and Errors in Job Analysis Ratings

There is an abundance of research investigating variables which influence the accuracy of job analysis ratings. Theoretically, one can separate the sources of inaccuracy into rater bias and rating error. A bias is a systematic rating tendency that may or may not be in error, whereas rating errors can be either systematic or random. This distinction is important because rater biases, if accounted for, could be controlled or corrected by some means developed to increase the rater's objectivity. Although variance assumed to be the product of a systematic rater bias may indeed contain error, it can not be assumed that all bias is error.

In the job analysis discrepancies among raters typically have been attributed to error variance, but it seems equally probable that differences in description of jobs are due to biases that distort the perceptions of job analysts during the job analysis process or that the discrepancies represent differences in the allocation of tasks which make up the job (Arvey, et al., 1982; Taylor, 1978, in Mullins & Kimbrough, 1988; Cranny & Doherty, 1988; Schmitt & Cohen, 1989). One can not just discard the "unreliable" items because this would lead to an incomplete analysis of the job because these items may uncover ways in which "members of the various sub-groups may be treated differently on the job or ways in which members of different sub-groups perceive their jobs differently" (Schmitt & Cohen, 1989, p.103).

Research has investigated whether incumbents in the same job may differ in job analysis ratings depending on particular sub-groups to which they are classified, such as racial (White, African-American, Hispanic), gender (male, female), and experience level (0-2 years, 2-4 years, etc.). A review of the literature indicates that most findings regarding this issue are not robust, and that there is conflicting data across studies. In addition, when between-group differences were found regarding task ratings, it is not clear whether those differences stemmed from biases affecting job perceptions, or

because of actual, systematic differences in task assignments. The following discussion focuses on these differences as well as the accuracy of job analysis. In addition, voids in the literature are discussed regarding issues that need to be addressed empirically.

Experience Level. One of the most studied topics in job analysis accuracy is the effect of incumbent 'experience' on job analysis ratings. The measure of incumbent experience varies across these studies with the most common definition being organizational tenure (Cornelius & Lynes, 1980; Mullins & Kimbrough, 1988; Landy & Vasey, 1991; Borman, et al., 1992). Examinations of the relationship between employee experience have provided mixed results. Smith and Hakel (1979) had job incumbents and supervisors rate jobs using the Position Analysis Questionnaire (PAQ) (McCormick, Jeanneret, & Mecham, 1972). They reported that as job levels increase, so do the reliability coefficients of job analysis ratings. This implies either that higher level jobs are easier to measure or that those in the higher positions are better able to analyze their job. Contrary to these findings, Cornelius (1980) reported that differences in tenure could not predict the reliabilities of job analysis ratings, either when using test-retest, incumbent-supervisor, or incumbent-analysts similarities.

Schmitt and Cohen (1989) reported that managers with more than one year of experience provided virtually identical information on the time spent on tasks as well as their difficulty. Landy and Vasey (1991), using an analysis of variance (ANOVA) approach comparing mean task frequency ratings for each experience level group, reported that as levels of experience change, so too does the frequency of the tasks conducted. This finding was replicated by Borman, et al. (1992) with a different subject population. These data can be explained in a variety of manners. It is probable that as one's tenure in a job increases, changes occur in task assignment which would indicate a true difference in jobs between experience levels. For instance, Landy and Vasey (1991) suggests that higher level officers choose the tasks they prefer while less senior officers are relegated to perform the less preferred tasks. This 'pecking order' would suggests that as seniority increases, the tasks which make up the job actually change. Borman et al.'s (1992) subject pool consisted of stockbrokers. They suggest that as tenure increase in this job, the frequency of certain tasks change. This is not surprising, however, since one would expect that more senior stockbrokers would spend less time on tasks such as seeking referrals and seeking information from co-workers, and would spend more time on tasks such as helping other stockbrokers, dealing with clients in non-business settings, and handling general administrative paperwork. It may be that the conduction of tasks and the tasks' frequencies vary across levels of experience because of a 'pecking order' or because of professional necessity; however, an unexplored hypothesis is that the perceptions of the job and encoding and recall of job activities and their

characteristics vary as time goes on. Each explanation may be true to some extent, but future research is needed to examine sources of this variance.

Gender. Arvey, Passino, and Lounsbury (1977) examined the effects of gender on job analysis ratings. Using the PAQ, raters conducted a job analysis based on information provided through a narrative and visual slides. They were interested in looking at the effects of the gender of the job incumbent as well as the gender of the analysts on the ratings for 32 PAQ dimensions established for the job in question. Gender of the job incumbent did not influence job analysis ratings, while the gender of the analysts did have a marginal effect. There was a tendency for females to consistently give lower scores on the PAQ job analysis dimensions than males, although only one dimension actually reached significance. The authors concluded that "we should be alert to situations in which particular jobs or job families are consistently analyzed by either male or female analysts only" (p. 415). But this conclusion may have been premature. Subsequent research has failed to replicate the gender differences of job analysts ratings (Arvey, et al., 1982, Schwab & Grams, 1985; Grams & Schwab, 1985).

Schmitt and Cohen (1989) examined differences in job analysis ratings between male and female incumbents of middle-manager Civil Service jobs. They found gender differences on the frequency ratings of several tasks: "men reported spending more time talking, meeting, and consulting with people outside of the organization, whereas women . . . report spending more time interpreting and formulation policy for people within the organization" (p. 98). However, the study notes that it was unclear how to attribute this discrepancy, but the most likely scenario was that the actual job that the men and women in this study did consisted of sex-stereotyped tasks, and that the job incumbents either placed more emphasis in jobs that they preferred or the jobs in which they held higher perceived confidence. But another explanation is possible. The jobs that the men and women held may not have been the same. It could well be that the frequency ratings reported correspond to the assignment of job tasks and, in actuality, represent an underlying mechanism which influences task allocation. In other words, the types of tasks assigned to men and women and the frequency under which those tasks are conducted may differ as a function of their gender because of actual self-selection or supervisor allocation of job tasks. Landy and Vasey (1991) hinted at this when they found gender differences in frequency ratings among tasks for police officers. They noted that "the differences between male and female officers are just as likely to be the result of experience differences as gender difference. The most reasonable conclusion to draw is that gender differences are confounded by experiences differences" (p. 42). But this is yet an untested hypothesis.

Performance Level. Another question that interested researchers was whether or not an incumbent's job performance was related to job analysis ratings. Conley and Sackett (1987) examined the relative accuracy of job analysis ratings given from high and low performers. In their study, subjects (police officers) were split into high and low performance groups based on supervisor evaluations. Each group generated task inventories as well as a description of the KSAs needed to perform the job. The tasks were rated on importance, time spent, criticality, and difficulty. No significant differences were found between the groups. This indicates that differences in performance ratings are not related to differing concepts of what tasks a given job entails or the KSAs needed to perform the job adequately. These findings were consistent with Wexley and Silverman (1978).

The Conley and Sackett study (1987), however, may not provide a complete depiction of the relationship between performance levels and job analysis ratings. First of all, the authors note that the job studied (police work) provided extensive training and structure, and whether or not these results are replicated in jobs which include little or no formal training or that are void of rigid structure is yet to be tested. Another question also remains. Performance in this study was rated on a subjective, relative scale. High and low performers were rated in comparison to cohort groups. So a low performer was low compared to other officers in his/her own group. Although the authors provided some evidence that there was comparability between the relative ranking and a more objective measure (number of cases processed and number of arrests), this analysis was based on only 12 subjects and should not take the place of a more empirical examination. Future research should examine whether or not more objective performance ratings at the task level correspond to job analysis ratings of those tasks. In other words, do job incumbents rate tasks differently as a function of their individual performance on those tasks? It could be that differences in job analysis ratings are an accurate representation of how job incumbents select or are assigned tasks based on performance.

One factor that may interact with task performance and task ratings is the self-efficacy of the incumbent. Schmitt and Cohen (1989) made the statement that perhaps job incumbents, in their description of job tasks, may focus on the tasks for which they have higher perceived competence. Although, there is evidence to support the notion that people tend to seek self-assessment through task choice, whether diagnosing high or low ability, and that success and failure play conceptually similar roles in determining persistence, until level of ability can be ascertained (Strube & Roemmele, 1985; Trope & Ben-Yair, 1982). However, it could be that higher performance early on would lead to an incumbent being assigned more opportunities to perform the tasks for which proficiency is shown, thus job incumbents would emphasize or actually choose the tasks for which they hold higher self-efficacy.

This would explain higher variance in job analysis ratings for the less experienced. As tenure increases the variance in performance may decline for a variety of reasons, including a ceiling effect, changes in task selection or assignment based on seniority, or changes of perceptions of the relative importance of specific tasks. At this point this is still conjecture, but future research may uncover the relationships between an incumbent's level of performance and the selection, assignment, and perception of tasks, as well as how these factors influence job analysis ratings.

Job Stereotypes. Members of different sub-groups may have differing stereotypes about the jobs they possess. These stereotypes may change as tenure increases and more objective information about the job becomes available, or it may be that different groups of incumbents carry with them stereotypes about their jobs and these stereotypes continuously impact task selection or perceptions. The stereotypes held by supervisors or analysts may also impact the job analysis ratings, directly or indirectly. These stereotypes may affect perceptions of task frequency or importance ratings, or may actually influence task allocation procedures. Research has examined these issues but, again, has provided mixed results.

Arvey, et al. (1977) found gender differences in job analysis ratings. They reported differences between male and female analysts. Female analysts tended to give consistently lower scores on PAQ job analysis dimensions than the male analysts. The gender of the job incumbent, however, was inconsequential. But the authors suggests that perhaps a gender of incumbent effect may have been found if the job depicted had been a sex-stereotyped position (e.g., physician, clerical worker). In other words, if the incumbent was depicted as filling a sex stereotyped job, perhaps then the job analyst would have displayed a bias towards using a gender stereotype in the job analysis.

Smith and Hakel (1979) had job incumbents, supervisors, and job analysts use the PAQ to rate the incumbents' jobs. In addition they gave two groups of students either the title of the job or the title of the job along with a narrative description of the job. They then had the students rate the job using the same PAQ measure. One of the issues they were interested in examining was whether or not these different judge categories had stereotypes, and thus inherent judging biases, that influenced the job analysis ratings; thus "convergence (in job analysis ratings between judge categories) would be expected because of shared stereotypes in addition to, or perhaps instead of, differentially accurate knowledge of the job" (p. 686). They found that the reliability coefficients of the job analysis ratings between students either with or without a job description, the incumbents, supervisors, or analysts were not significantly different from each other. They concluded that there was no evidence that any of these groups were more or less accurate in job analysis ratings.

Harvey and Lozada-Larsen (1988) investigated the *shared-stereotype hypothesis*. This hypothesis holds that "raters in the different groups held common stereotypes of the work performed on each of the jobs; similar ratings would be expected because the groups used essentially the same information, regardless of accuracy" (p. 457). Job analysts were placed into one of three groups: those with only a job title, those with a only a job description, and those with both a job title and a job description. Raters in both the job-title-and-description and the job-description-only groups were more accurate than those with only a job title. The "presence of task summaries improved the accuracy relative to using only a job title" (p. 459). This contradicts Smith and Hakel (1979) and the shared-stereotype hypothesis. However, future research should investigate if these results would be replicated when the jobs in question "have more popular titles" (p. 461) or have inherent gender-stereotypes associated with them.

Landy and Vasey (1991) discussed the role that stereotypes may have in the explanation of their data. Again, they found gender differences in job analysis frequency ratings in police officers. It could have been that there was an actual difference in task assignments based upon the influence of stereotypes held by the dispatcher. But since the dispatching of calls was controlled by computer assignment, this possibility was rejected. However, the authors suggest that it was still possible that the stereotypes held by the officers themselves may account for biases in their self-report of duties, where the actual tasks and frequencies are equivalent between groups but the reporting through a job analysis differs because of the differing stereotypes individuals have regarding their jobs. However, the authors contend that if job incumbent stereotypes were a factor in the job analysis then "it would be an enormously complex task to respond both in a way that produced internally consistent components and differential patterns across components" (p. 46).

Job and Task Characteristics. When jobs allow the worker the discretion to place emphasis on a variety of job activities, individuals may develop different perceptions of the job's demands (Conley & Sackett, 1987). There may also be differing formal and informal job requirements for incumbents that would lead to incumbents performing the job differently in order to meet those requirements or to optimize other benefits (Green & Stutzman, 1986). Different aspects of the job could receive changing emphasis based upon incumbent perceptions of the tasks characteristics, such as task importance, criticality, complexity or difficulty. In addition, it may be that task characteristics moderate the relationship between the factors studied and job analysis ratings. For example, Mullins and Kimbrough (1988), in trying to explain the mixed results of the relationship between job performance and job analysis ratings, suggested that perhaps job complexity or criticality interact. This is certainly an area of

research that has not been tapped, although there has been some supposition: "The tasks the police officer perform and the job dimensions relevant to policing are much greater and much more variable than those of a dormitory supervisor" and similarly "the nature of a policeperson's job is more critical than that of a dormitory supervisor . . . The incompetent police officer may be avoiding crucial tasks, and may be in fact, doing a different job . . . In contrast dormitory supervisors are all doing the same tasks and the difference in their performance is a matter of degree only (excellent to poor)" (Mullins & Kimbrough, 1988, p.662). However, their data could not provide an avenue to test these hypotheses. In addition, this idea would contradict finding reported by Conley and Sackett (1987) who found that there was no difference between high and low performers and job ratings regarding the importance, criticality, or difficulty of the tasks.

Other Individual Differences. The influence of other individual characteristics on job analysis ratings have been studied or discussed in the literature. Differences in characteristics such as interests in the job, perceived importance of the task, and self-interests have been hypothesized to influence how one rates his/her job. These differences may indicate actual differences in the task or differential perceptions of what tasks make up the job. Arvey, et al. (1982) found marginal support for the notion that a job incumbent's interest in the job, as perceived by an analyst, has some effect on the job analysis ratings. However, they report that although statistically significant, the degree of job interest is not particularly important.

Other researchers have looked at the perceived importance of the task as an avenue for job analysis ratings. Mullins and Kimbrough (1988) had respondents analyze their job by rating its dimensions using a 5-point index ranging from: 1) extremely unimportant to patrolperson success, to 5) extremely important to patrolperson success. The raters also rank ordered the job dimensions in terms of relative importance. However, other researchers disagree on the validity of using importance ratings in the job analysis. Cranny and Doherty (1988) argue against forming behavior dimensions by intercorrelating behavior items based upon importance ratings and then factor analyzing the resulting correlation matrix in an attempt to form behavior categories. They argue that job analysis categories should be based upon functional differences or hypothesized common skills or abilities. Landy and Vasey (1991) also agree that job analyses based upon importance ratings are influenced by divergent ideas of what constitutes importance and task frequency is less susceptible to the same type of bias.

Another individual characteristic that may influence job analysis ratings is self-interest. Self-interest may be a source of bias by which incumbents or even supervisors exaggerate certain job element scores in order to make the job appear more important than it is (Smith & Hakel, 1979). It could

also be the result of social desirability (Anderson, et al., 1984). Either way, self-interest bias, or job inflation, has implications for the evaluation of work experience when using the job analyses as a source of data.

Several researchers have studied job inflation in attempts to develop methods of detecting "fakers" or to account for an individual's tendency to inflate a job (Anderson, et al., 1984; Green & Stutzman, 1986; Pannone, 1984). One approach to detecting job inflation is to include task statements which are known, a priori, to be either unrelated or bogus to the job under investigation. One can then examine which individuals are more likely to inflate jobs, and what effect this inflation may have on the overall accuracy of measurement of work experience. Pannone (1984) looked at the effects job inflation has on the relationship between biographical questionnaires and written job-knowledge tests. This biographical questionnaire tapped into previous work experience and included bogus tasks so that job inflation could be detected. When all subjects were included in this analysis, the correlation was .42; however, when one looks only at the data from those who did not inflate their biographical questionnaire, this relationship increased to .55. The correlation for only those incumbents who did inflate their biographical questionnaire was .26. Others agree that the presence of an inflation bias decreases the accuracy, and therefore the usefulness, of the job analysis or other self-examination metrics (Green & Stutzman, 1986; Anderson, et al., 1984).

Green and Stutzman (1986) reported that those subjects who inflated their jobs could either have just been careless in making their responses, might have had difficulty reading or understanding the tasks statements, or may have wanted to project an inflated image of their respective jobs. However, in Anderson, et al. (1984) the median reliability of the inflation scales, consisting of bogus tasks, was .86. The authors argue that the high internal consistency of the inflation scale would not have emerged if the bogus task items were merely tricky or confusing and/or measured different constructs. Shaffer, et al. (1986) state that inaccuracy of biographical data may come from a distorted self-perception which is different than an intention to provide fake responses. Whatever the source of this inflation bias, it is apparent that more research is needed to examine the effectiveness of job analyses which include methods of detecting job inflation (Pannone, 1984) as well as methods for correcting for the bias (Anderson, et al., 1984).

Lastly, the influence of general cognitive abilities on the job analysis has had limited study in the literature. Conley and Sackett (1987) reported no relationship between scores on the Fleishman Abilities Scale and job analysis accuracy. However others have speculated that accuracy in job assessment is connected to the intellectual skills necessary for providing reliable information (Smith & Hakel, 1979).

Cornelius and Lynes (1980) argue that education level may be a surrogate for cognitive abilities. They reported a significant positive relationship between education level and the extent of agreement between incumbent job ratings and job analysts ratings. However, others have found little or no relationship between education level and job analysis accuracy (Mullins & Kimbrough, 1988; Landy & Vasey, 1991).

Accuracy Measured

A controversial question has arisen from the review of this job analysis literature: how to measure job analysis accuracy. There seem to be several distinct questions that arise when trying to address this issue. First, is one job analysis instrument more accurate than another at tapping the true nature of the respondent's actual job? The answer to this probably lies in the psychometric properties of the job analysis instrument, and is not within the confines of the present paper. This paper is not so much interested in how to conduct a job analysis, but rather what information should be considered when interpreting job analysis ratings across individuals, especially when using the ratings as a measure of previous work experience.

The next question regards the manner used to determine which respondents, or groups of respondents, are more accurate in their job analysis scores. This is a tricky question because one has to determine whether a person is unable to accurately analyze his/her job, or if the procedure used has tapped into a true difference between the respondents' job activities and others to which the comparison is made. A third question involves the procedure used to make between group comparisons. This is not a measure of accuracy, but instead is an attempt to understand between group differences in the job analysis ratings.

One manner in which accuracy has been measured is a derivative of Cronbach's (1955) ratings components which is based on deviations from means. "The rater whose overall average is close to the (group's) overall average true score will tend to be more accurate than one whose average rating is far from the true score average" (Murphy, Garcia, Kerkar, Martin, & Balzer, 1982, p.321) For instance, Green and Stutzman (1986) measured job analysis accuracy in terms of "the distance between an employee's response on the task statements and his/her unit's 'centriod' on these statements" (p. 551). They suggest that since different people create different job analysis responses, screening the job analysts may be in order, and the number of respondents should be more than three to ensure higher reliability. Other researchers have adopted this method as the basis for accuracy measurements (Conley & Sackett, 1987; Harvey & Lozada-Larsen, 1988; Murphy, et al., 1982).

But using the centriod techniques has its limitations. The centriod is defined as the mean rating aggregated from the entire group. This mean is a theoretical true score and estimate of the actual job characteristics. However, there is little empirical evidence to support the notion that this theoretical true score is an accurate representation of the actual true score. It may be just as likely that higher variance from group means in job analysis ratings between groups is a reflection of actual within-group differences in job tasks and their frequencies.

Other statistics used to gauge job analysis accuracy are reliability coefficients. Smith and Hakel (1979) used correlation coefficients (transformed into z scores) to assess differences in accuracy between five groups: 1) incumbents; 2) supervisors; 3) job analysts; 4) students with only a job title; 5) students with a job title and a job specification. They found no significant differences in reliabilities between these groups (r's: .59, .63, .63, .51, .49 respectively). But the fact there was no significant difference may be explained by the large amount of within-group variance of the reliability coefficients. Other researchers have used reliability coefficients in a similar manner (Shaffer, et al., 1986; Cornelius & Lynes, 1980). Similarly, Pannone (1984) used reliability coefficients to separately correlate biographical data (measuring work experience) and years of tenure with work sample test. He found that the biographical data had a stronger correlation with scores on the work sample test than did a simple measure of tenure (.42 vs. .13, respectively). But by using the reliability coefficients of each group to assess differential accuracy, one still does not address the underlying issue brought forth by the present paper. Inaccuracy implies deviations from a true score. Lower reliabilities may represent actual differences within a job category. Besides, given the low correlations between incumbent and supervisor regarding job analysis ratings, it is difficult to determine who is right and who is wrong (Harvey, 1991). The lack of agreement between an incumbent and an observer may not necessarily indicate that the incumbent has replied incorrectly, it may reflect the observer's inability to provide accurate information (Shaffer, et al., 1986).

The contradiction between Smith and Hakel (1979) and Harvey and Lozada-Larsen (1988) may be, in part, attributable to variations in methods used to assess accuracy. As described above, Smith and Hakel (1979) used correlation coefficients to measure accuracy and found no difference in accuracy between the job title-job description group and the job title only group. Harvey and Lozada-Larsen (1988) found that raters in both the job-title-and-description and the job-description-only groups were more accurate than those with only a job title. However, the method they used was based on deviations from sub-group means (i.e., Cronbach, 1955).

However, differences in methods of measuring job analysis accuracy can not account for all of the contradictions between studies. Smith and Hakel (1979) reported that as job levels increases, so do the reliability coefficients of job analysis ratings. Cornelius (1980) used a similar method and reported that differences in tenure could not predict the reliabilities of job analysis ratings. In addition, Arvey, et al. (1977) and Arvey, et al. (1982) used similar methods but found contradictory results regarding gender differences and job analysis ratings.

The third question did not involve a measure of job analysis accuracy, but a means through which one could understand between group differences in the responses to job analysis. The ANOVA approach appears to be the most common. Arvey, et al. (1977) used a factorial design to examine the effects of differences in the gender of the job incumbent and of the job analysts on job analysis biases (discussed above). Other researchers have accepted the ANOVA paradigm (Arvey, et al., 1982; Borman, et. al, 1992; Landy & Vasey, 1991; Mullins & Kimbrough, 1988; Schmitt & Cohen, 1989).

Implications for Selection and Training

If employers could successfully measure applicants' or employees' previous work experience, then future performance could be more accurately predicted (Quinones, et al., 1994). Measuring experience at the task level would allow for a stronger linkage between this predictor and the criterion of task performance. It would provide an objective, systematic analysis of an applicant's qualifications, and would also allow for an analysis of task areas in which employees are deficient and could stand more training.

In order to fully and successfully evaluate the effectiveness of a training course, one would have to evaluate subjects' performance on the various KSAs being trained. Although it is important to measure immediate learning after a training course, an important dimension of training effectiveness is the extent to which individuals transfer what has been learned from the training environment to the work environment (Goldstein, 1993). But this is a particularly difficult endeavor because of the variety of confounding variables which may exist in the transfer environment. One such variable is the opportunity trainees have to perform each trained task while on the job (Ford, Quinones, Sego, & Sorra, 1992). Since it has been shown that there are individual differences in opportunities to perform tasks between trainees of the same training course, it can be argued that work experience data at the task level should be taken into account in any subsequent performance measurement meant to provide an evaluation of a training course's effectiveness (Ford, et al., 1992; Quinones, et al., in press). In addition it would be a useful result if these types of data could inform the organization on what activities job incumbents should spend

more time on to increase performance (Borman, et al., 1992). For instance, with supervisory experience especially important for supervisory proficiency, training that includes opportunities to gain actual experience with the target supervisory tasks would seem to be important (Borman, Hanson, Oppler, Palukos, & White ,1993)

Summary

There are a variety of directions in which work experience research can take. There is a need for a theoretical framework which would integrate variables concerning the availability of work experiences, the effect these experiences have upon the individual, and the relationship these experiences have with a variety of outcome measures. There are also needs to understand why individuals differ in the report of work experience even though there is no actual difference in experience, as well as to understand why the same job may provide different experiences across individuals. But before these questions can be answered, a more fundamental issue must be addressed. A systematic method must be developed to quantify work experiences as well as to determine the criteria against which to assess the construct validity of these measures (Quinones, et al., 1994). The present paper suggests that a job analysis technique would provide data on job tasks and behaviors in terms of "type, amount, and time" (Quinones, et al., 1994). The result would be an accurate and systematic measurement of previous work experience which would have implications for employee selection, placement, training, as well as the direction of future research.

References

Anderson, C.D., Warner, J.L. & Spencer, C.C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. Journal of Applied Psychology, 69 (4), 574-580.

Arvey, R.D., Davis, G.A., McGowen, S.L. & Dipboye, R.L. (1982). Potential sources of bias in job analytic processes. Academy of Management Journal, 25 (3), 618-629.

Arvey, R.D., Passino, E.M., & Lounsbury, J.W. (1977). Job analysis results as influenced by sex of incumbent and sex of analyst. Journal of Applied Psychology, 62 (4), 411-416.

Asher, J.J. (1972). The biographical item: Can it be improved? Personnel Psychology, 25, 251-269.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M.D. Dunnette & L.M. Hough (Eds.), Handbook of industrial and organizational psychology (pp. 271-326). Palo Alto, CA: Consulting Psychologist Press, Inc.

Borman, W.C., Dorsey, D. & Ackerman, L. (1992). Time-spent responses as time allocation strategies: Relations with sales performance in a stockbroker sample. Personnel Psychology, 45, 763-777.

Borman, W.C., Hanson, M.A., Oppler, S.H., Palukos, E.D., & White, L.A. (1993). Role of early supervisory experience in supervisor performance. Journal of Applied Psychology, 78 (3), 443-449.

Conley, P.R. & Sackett P.R. (1987). Effects of using high- versus low- performing job incumbents as sources of job-analysis information. Journal of Applied Psychology, 72 (3), 434-437.

Cornelius, E.T., Lynes, K.S. (1980). A comparison of holistic and decomposed judgment strategies in job analyses by job incumbents. Journal of Applied Psychology, 65 (2), 155-163.

Cranny, C.J. & Doherty, M.E. (1988). Importance ratings in job analysis: Note on the misinterpretation of factor analysis. Journal of Applied Psychology, 73 (2), 320-322.

Cronbach, L.J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 52, 177-193.

Ford, K.J., Quinones, M.A., Sego, D.J., & Sorra, J.S. (1992). Factors affecting the opportunity to perform trained tasks on the job. Personnel Psychology, 45, 511-527.

Goldstein, I.L. (1993). Training in Organizations (3rd ed.). Pacific Grove, CA.: Brooks/Cole.

Grams, R. & Schwab, D.P. (1985). An investigation of systematic gender-related error in job evaluation. Academy of Management Journal, 28, 279-290.

Green, S.B. & Stutzman, T. (1986). The evaluation of methods to select respondents to structured job-analysis questionnaires. Personnel Psychology, 39, 543-56

Harvey, R.J. (1991) Job analysis. In M.D. Dunnette & L.M. Hough (Eds.), Handbook of industrial and organizational psychology (pp. 71-164). Palo Alto, CA: Consulting Psychologist Press, Inc.

Harvey, R.J. & Lozada-Larsen, S.R. (1988). Influence of amount of job descriptive information on job analysis rating accuracy. Journal of Applied Psychology, 73 (3), 457-461.

Landy, F.J. & Vasey, J. (1991). Job analysis: The composition of SME samples. Personnel Psychology, 44, 27-50.

Levine, E.L., Florry, A., Ash, R.A. (1977). Self-assessment in personnel selection. Journal of Applied Psychology, 62 (4), 428-435.

McCormick, E.J., Jeanneret, P.R., & Mecham, R.C. (1972). A study of job characteristics and job dimensions based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology, 56, 347-368.

Mullins, W.C. & Kimbrough, W.W. (1988). Group composition as a determinant of job analysis outcomes. Journal of Applied Psychology, 73 (4), 657-664.

Murphy, K.R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W.K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 67 (3), 320-325.

Pannone, R.D. (1984). Predicting test performance: A content valid approach to screening applicants. Personnel Psychology, 37, 507-514.

Quinones, M.A., Ford, K.J., Sego, D.J., Smith, E.M. (in press). The effects of individual and transfer environment characteristics on the opportunity to perform trained tasks. Training Research Journal.

Quinones, M.A., Ford, J.K. & Teachout, M. (1994, July). Experience and performance: A conceptual and meta-analytic review. Presented at the 23rd Congress of Applied Psychology, Madrid, Spain.

Russell, C.J., Mattson, J., Devlin, S.E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. Journal of Applied Psychology, 75 (5), 569-580.

Schmitt, N. & Cohen, S.A. (1989). Internal analyses of task ratings by job incumbents. Journal of Applied Psychology, 74 (1), 96-104.

Schwab, D.P. & Grams, R. (1985). Sex-related errors in job evaluations: A "real-world" test. Journal of Applied Psychology, 70, 533-539.

Shaffer, G.S., Saunders, V., Owens, W.A. (1986). Additional evidence for the accuracy of biographical data: Long-term retests and observer ratings. Personnel Psychology, 39, 791-809.

Smith, J.E. & Hakel, M.D. (1979). Convergent among data sources, response bias, and reliability and validity of a structured job analysis questionnaire. Personnel Psychology, 32, 677-692.

Strube, M.J. & Roemmele, L.A. (1985). Self-enhancement, self-assessment, and self-evaluative task choice. Journal of Applied Psychology, 49 (4), 981-993.

Trope, Y. & Ben-Yair, E. (1982). Task construction and persistence as means for self-assessment of abilities. Journal of Applied Psychology, 42 (4), 637-645.

Wexley, K.N. & Silverman, S.B. (1978). An examination of differences between managerial effectiveness and response patterns on a structured job analysis questionnaire. Journal of Applied Psychology, 63, 646-649.

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser
Research Associate
Department of Biology

Trinity University
715 Stadium Dr.
San Antonio, TX 78212-7200

Final Report for:
Summer Graduate Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser
Associate Researchers
Department of Biology
Trinity University

Abstract

In ketamine-anesthetized rats, sustained whole-body exposure to 35-GHz millimeter wave radiofrequency radiation (RFR) produces hyperthermia, visceral vasodilation, and subsequent hypotension resulting in death of the subject (Physiologist 34:246, 1991). This study sought to determine whether this phenomenon (i.e., eradication of compensatory splanchnic vasoconstriction precipitating hypotension) is caused by vasodilatory factor(s) present in the circulating blood during circulatory failure. In search of evidence for a humoral visceral vasodilator, we performed a blood transfusion experiment. Two groups of rats (n=10 for each group) were used for the protocol. In the experimental group, one rat (donor rat) was exposed to RFR until mean arterial pressure (MAP) fell to 75 mmHg (arbitrarily assigned point of shock induction from previous work). At this point, 5 ml of blood were withdrawn from the hypotensive rat via the left carotid artery. This blood was subsequently infused into the recipient rat via the right jugular vein while an equal volume of blood was withdrawn simultaneously from the right femoral artery. MAP was monitored on the recipient rat for a 5 minute control period prior to transfusion and during the entire transfusion. In the control group, the same procedure was employed without exposing the donor subject to RFR. Therefore, in the control paradigm, the donor subject was normotensive when the blood was withdrawn. Immediately following transfusion in both groups, we observed an initial decrease in MAP followed by a similar increase returning MAP to control period levels. The recipient rats in the experimental paradigm did demonstrate a more pronounced decline in MAP post-transfusion as compared to the recipient rats in the control group (20.4 mmHg to 9.3 mmHg, respectively); however, those differences in mean maximum decrease in MAP were not shown to be significant (p=0.051). Therefore, we conclude that the vasodilatory factor(s) is not a humoral agent.

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser

Introduction

In humans and other mammals, maintenance of homeostasis is vital to survival. Homeostasis involves the regulation of physiological variables within a very narrow range. One of the many regulated variables is internal body temperature. Although all mammals possess thermoregulatory mechanisms that maintain their respective optimal internal temperature, prolonged extreme temperature changes can result in failure of the thermoregulatory system.

A primary mechanism of heat loss during thermal stress is through dilation of the cutaneous vasculature. In mild to moderate heat stress, arterial blood pressure is maintained at normal levels despite the marked cutaneous vasodilation by both an increase in cardiac output and a redistribution of blood flow from the viscera to the skin. That is, cutaneous vasodilation is normally accompanied by a compensatory vasoconstriction in visceral vascular beds that is primarily mediated by increases in sympathetic nervous system activity (Rowell, 1986; Kregel and Gisolfi, 1989).

Severe hyperthermia, however, may result in heat stroke, a condition characterized by a precipitous fall in arterial blood pressure. Heat stroke may, in turn, lead to a state of circulatory shock, in which tissue hypoperfusion occurs. Although the mechanism(s) responsible for this circulatory dysfunction is still in question, it now appears that a significant loss of peripheral vascular tone occurs in vascular beds that were previously constricted. Adolph (1923/4) first suggested that circulatory failure contributes to heat-induced circulatory shock. Subsequently, Daily and Harrison (1948) demonstrated that the hypotension and decreased cardiac output attendant to severe hyperthermia in humans were the result of peripheral pooling of blood. Kielblock et al. (1982) later proposed that fatal heat-induced shock resulted from cardiac failure due to a marked decline in vascular resistance after the loss of compensatory

vasoconstriction.

Kregel et al. (1988) directly measured the sequence and nature of vascular responses to environmental heat stress in conscious and anesthetized rats. In these heat-stressed rats, mean arterial pressure (MAP) increased until core temperature reached 41.5°C, at which point MAP fell precipitously. Mesenteric vascular resistance increased during the early stages of heat but declined sharply before the sudden fall in MAP. Thus, a selective loss of compensatory splanchnic vasoconstriction appears to trigger the circulatory collapse associated with severe hyperthermia. The sudden splanchnic vasodilation, combined with continued cutaneous vasodilation, produces hypotension by decreasing both total peripheral vascular resistance and venous return; the latter ultimately results in decreased cardiac output.

Visceral vasodilation preceding shock induction has been demonstrated during millimeter wave (MMW) irradiation, as it does during environmental heat-induced shock. In our model of heat stress (i.e. MMW exposure), using ketamine-anesthetized rats, mesenteric blood flow decreased during the early stages of MMW irradiation but then dramatically increased immediately prior to the onset of hypotension (Frei, *et al*, in preparation). Therefore, our model of heat-induced shock induction is analogous to that produced by environmental heating because, in both cases, eradication of compensatory splanchnic vasoconstriction precipitates hypotension.

There are several known possible endogenous vasodilators including opiates, catecholamines, nitric oxide, cytokines, arachidonic acid metabolites, bradykinin, histamine and some other small humoral peptides. Kregel *et al.* (1990) ruled out opiates, splanchnic sympathetic neurotransmitters and catecholamines as possibilities, since blockade of each of these potential mediators failed to prevent visceral vasodilation. In the MMW-induced heat stress model, nitric oxide, a potent gaseous vasodilator implicated in several other forms of circulatory failure, does not appear to be responsible for the noted hypotension. Chronic nitric oxide synthetase blockade studies concluded that nitric oxide was not the vasodilator (Wieser *et al.*, 1994). Although

several of these vasodilator possibilities have been extensively studied, the primary factor(s) involved have yet to be identified.

In order to narrow down the remaining possible vasodilator candidates, the present study, employing the MMW-induced heat stress model, sets out to determine if the factor(s) is present in the circulating blood during circulatory failure.

MATERIALS AND METHODS

Animals and Surgical Preparation

Forty male Sprague-Dawley rats (Charles River Laboratories), weighing between 328 and 402 g (368 ± 5g) were used in this study. Animals were housed in polycarbonate cages and provided food and water *ad libitum*. The rats were maintained on a 12 h/12 h, light/dark cycle (lights on at 0600) in a climatically controlled environment (ambient temperature of 24.0 ± 0.5°C).

Immediately prior to experimentation, two rats were anesthetized with ketamine HCl (150 mg/kg, I.M.). Administration of ketamine at this dose level provides prolonged anesthesia in Sprague-Dawley rats (Smith *et al.*, 1980; Jauchem *et al.*, 1984). Supplemental ketamine injections were administered throughout the duration of the experiment to ensure proper anesthetized conditions for the subjects.

Donor Subject

The larger of the two rats was designated as the donor subject. A catheter (Teflon, 28 gauge i.d.) was placed into the aorta via the left carotid artery for measurement of mean arterial blood pressure and later used for blood withdrawal. After surgery, the rat was placed on a holder consisting of seven 0.5-cm (O.D.) Plexiglas rods mounted in a semicircular pattern on 4 X 6 cm Plexiglas plates (0.5 cm thick). The electrocardiogram (ECG), mean arterial pressure (MAP), respiration and temperatures at five locations were continuously monitored using a Gould TA 2000 recorder. A Lead II ECG was used to monitor the subject with subcutaneous nylon-covered flurocarbon leads in the right arm, right leg and left leg (ground). The arterial catheter was attached to a pre-calibrated blood

pressure transducer (P10EZ, Statham) which was interfaced with a pressure processor (Gould 13-4615-52). Respiratory rate was monitored by a pneumatic transduction method employing a piezoelectric pressure transducer (Model 320-0102-B, Narco Biosystems). Heart rate (HR) was determined from ECG readings. Temperature was recorded from five sites: (1) colonic (T_c) (5-6 cm post-sphincter), (2) left subcutaneous (T_{sl})(lateral, midthoracic, side facing the source of radiation), (3) right subcutaneous (T_{sr})(lateral, midthoracic, side away from radiation source), (4) right tympanic (T_t), and (5) tail (T_{ta}). Tail temperature was measured subcutaneously from the dorsal surface approximately 2 cm from the base of the tail. All of the above recorded variables were monitored by a Unisys computer system via a software program specifically developed for physiological measurements (Berger *et al.*, 1991).

Recipient Subject

The smaller rat was designated the recipient subject, and catheters were inserted into three different locations: aorta via the left carotid artery, right jugular vein, and the left femoral artery. The left carotid artery was used to measure mean arterial pressure; the right jugular vein was used for the infusion of blood while the femoral artery served as a means of blood withdrawal.

During the surgical procedures on both rats, T_c was measured using an electrothermia monitor (Vitek, model 101) and was maintained at a temperature of 37.5 $\pm$ 0.5°C.

Exposure Conditions and Equipment

Experimental donor rats were individually exposed to 35-GHz continuous wave radiofrequency radiation (RFR) at an incident power density resulting in a whole body average specific absorption rate of 13 W/kg. The animals were aligned in the E orientation (long axis parallel to the electric field) during the exposure time. Prior to exposure, physiological control readings were recorded for a five minute period. The control period was subsequently followed by 35-GHz RFR. Irradiation was continued until mean arterial pressure deceased to 75 mmHg

(arbitrarily defined from previous work as the point of shock induction), at which point the RFR was turned off and the animal was prepared for blood withdrawal.

RF fields were generated by an Applied Electromagnetics Millimeter Wave Exposure System and were transmitted by a model 3-28-725 standard-gain horn antenna (Macom Millimeter Products, Inc.). Irradiation was performed under far-field conditions (animals positioned 110 cm from the antenna). The incident power density (75mW/cm^2) of the RFR fields was determined with an electromagnetic radiation monitor (Model 8600, Narda Microwave Corporation), employing a Model 8623D probe. During exposures, generator power output was monitored continuously with a Model 432B Hewlett Packard power meter. Irradiation was conducted in an Eccosorb RF-shielded anechoic chamber (Rantec, Emerson Electric Co.) at Brooks Air Force Base, Texas. The chamber temperature and relative humidity were maintained at 27.0±0.5° and 20±5% RH, respectively.

Transfusion Procedures

Immediately following shock induction in the irradiated rat, 5 ml of blood were withdrawn via the left carotid artery. The withdrawal was performed using a Harvard Apparatus 44 pump (model 55-1144) at a rate of 1 ml/min. The blood was collected in a heparinized syringe.The collection of 5 ml of blood from the donor rat in conjunction with the shock induction resulted in the death of this subject shortly after the withdrawal was complete.

During the withdrawal of blood from the donor subject, control readings of MAP and respiratory rate were obtained on the recipient rat for five minutes via the same recording apparatus as described for the donor rat. Also, T_c was monitored via an electrothermia monitor (Vitek, model 101).

The syringe containing the blood withdrawn from the donor was subsequently placed on a Razel Syringe pump (model 4-99..M)and connected to the catheter in the right jugular vein of the recipient rat. An empty heparinized plastic syringe was mounted onto the Harvard Apparatus 44 pump (model 55-1144) and connected to the catheter in the left femoral artery. Withdrawal of blood from

the left femoral artery and infusion of the blood from the donor rat occurred concurrently at a rate of 1 ml/min in order to maintain a constant blood volume. During the transfusion of blood into the recipient rat, the MAP and T_c were continuously recorded. These parameters were monitored for thirty minutes after the completion of the transfusion procedure.

The recipient rat was euthanized with a overdose of ketamine HCl at the end of the experiment.

The rats were divided into two groups: (1) a control group (n=10) in which transfusion occurred between two non-irradiated rats and (2) an experimental group (n=10) where the transfusion occurred between an irradiated and a non-irradiated rat.

For the control group, the donor rat was monitored for the same physiological parameters as the donor subject in the experimental group; however, no radiation was applied. Similar to the experimental group, five minutes of control readings for the donor rat were attained with T_c between 37.0 ± 0.5°C. These parameters were recorded for an additional thirty minutes, approximately the amount of time required for shock induction in the irradiated rats from the experimental group. At the end of the thirty minutes, the transfusion procedure was performed as described above.

Data Analysis

Preliminary statistical comparisons of MAP in the recipient rat between control and experimental group were performed at twelve different time intervals: control (mean of MAP values 2 min prior to transfusion), pre-transfusion (MAP immediately prior to transfusion), 0 min (the last MAP value during the transfusion), 0.5, 1, 2, 3, 4, 5, 10, 20, and 30 minutes post-transfusion. Statistical comparisons of each time period were accomplished by a two-way analysis of variance (ANOVA) with repeated measures.

Statistical comparison of mean maximum decrease in MAP in the recipient rat following transfusion were performed comparing control and experimental groups. The mean maximum decrease in MAP was calculated by taking the difference between

a mean from 2.5 minutes of MAP values at the end of the transfusion and the lowest MAP reading post-transfusion.

RESULTS

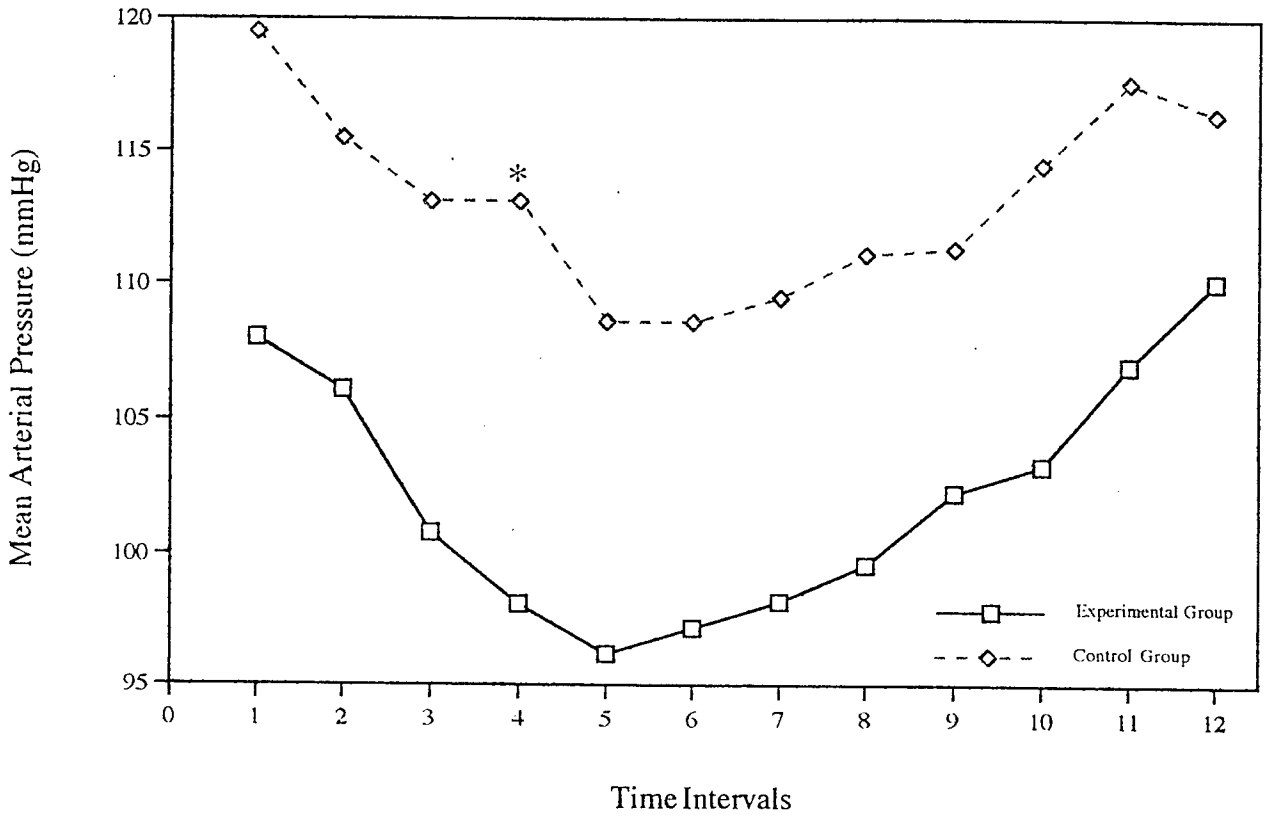
Table 1 shows that the time from the control period until the beginning of the transfusion used in the experimental paradigm is similar to the allotted 30 minute time prior to transfusion for the control paradigm. In the exposed animals, the mean T_c and T_{sl} reached 40.1°C and 45.0°C, respectively, prior to transfusion. The control group's mean T_c and T_{sl} remained constant during the 30 minutes prior to transfusion at 36.9°C and 35.3°C, respectively.

Table 1. Donor rat's parameters (mean values with n=10) prior to cross-circulation.

	Experimental	Control
Time to shock (min)	32.54	30.00
Tc (°C)	40.13	36.86
Tls (°C)	45.00	35.25

Figure 1 graphs the MAP over time for both the control and experimental group plotting 12 time intervals: control period, pretransfusion, post transfusion, 0.5 min, 1 min, 2 min, 3 min, 4 min, 5 min, 10 min, 20 min, and 30 min. The upper line represents data from the control group, while the lower line shows values for the experimental group. There were no significant differences of MAP values between control and experimental groups except at 0.5 min post-transfusion. Both groups show similar trends in MAP changes (i.e., initial decrease followed by a slow increase in MAP) with the lowest MAP value occurring at 1 min post-transfusion.

Figure 1. MAP graph versus time for both experimental and control groups.



* Significant difference between control and experimental values ($p \leq 0.05$)

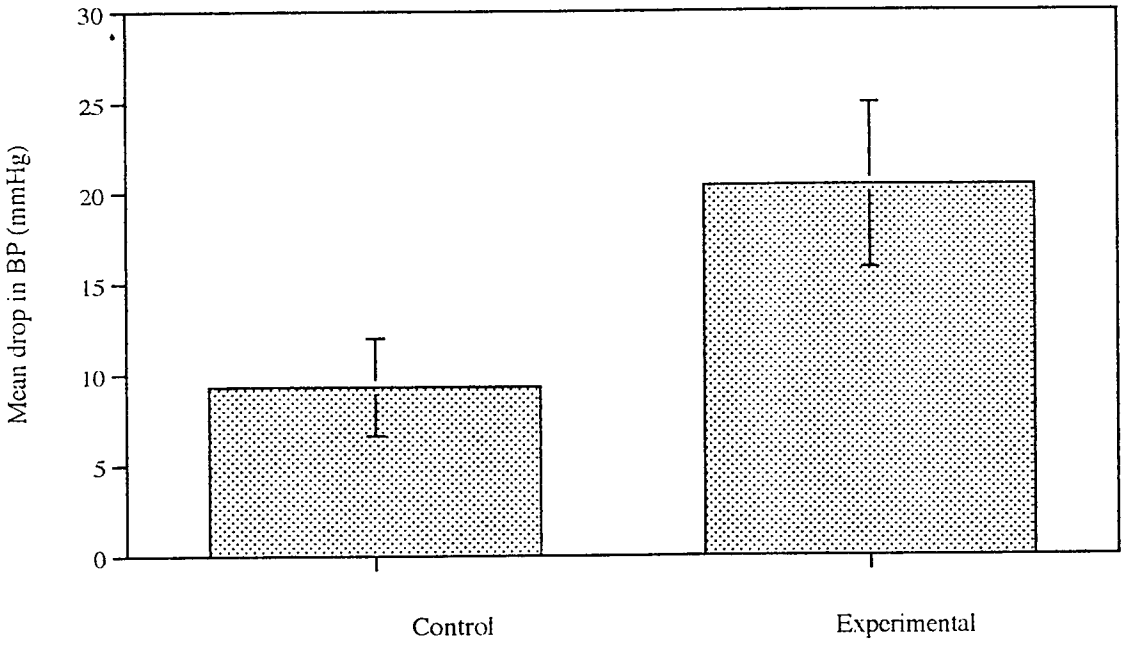
LEGEND FOR TIME INTERVALS

1- Control period	7- 3 min post-transfusion
2- Pre-transfusion	8- 4 min post-transfusion
3- 0 min	9- 5 min post-transfusion
4- 0.5 min post-transfusion	10- 10 min post-transfusion
5- 1 min post-transfusion	11- 20 min post-transfusion
6- 2 min post-transfusion	12- 30 min post-transfusion

Figure 2 is a bar graph showing the mean maximum changes in MAP for both groups. This change was calculated by obtaining a mean BP value from 2.5 min at the end of the transfusion and the minimum BP following transfusion. The difference between these values represents the maximum change in MAP. The mean of these maxima is represented in the bar graph.

Although there was no significant difference (p=0.051) in the maximum change in MAP between the control and experimental group, the experimental group shows a greater change in MAP than the control group. The mean maximum change in MAP of the control and experimental groups were 9.3 mmHg and 20.4 mmHg, respectively. As the p-value indicates, the values between the two groups were borderline to being significantly different.

Figure 2. Control and Experimental Groups Mean Maximum Change After Transfusion



DISCUSSION

Our results suggest that the vasodilator(s) responsible for the MMW heat-induced hypotension is either not humoral in nature or not detectable via our transfusion protocol. Figure 1, showing the MAP over time, depicts similar trends in both the control and experimental groups (i.e., initial decrease followed by a similar increase in MAP after the transfusion). This suggests that an artifact of the protocol may be partially responsible for the drop in MAP during and immediately following transfusion.

However, it appears that some of the drop in MAP may not be entirely due to protocol technique. There was a noted significant difference in MAP values at 0.5 minutes following transfusion between the control and experimental group, with the experimental group experiencing a greater drop in MAP immediately following transfusion. Also, Figure 2 shows that the experimental group had a greater mean maximum decrease in MAP than the control group. Although this difference was not significant, there was a trend in greater mean maximum decrease in MAP in the experimental group that just failed to reach significance (p-value=0.051). These findings of a greater drop in MAP following transfusion for the experimental group suggest the presence of some blood-borne vasodilator. Therefore, we do not completely discount the possible existence of a humoral vasodilatory factor(s).

REFERENCES

Adolph, E.F., and W.B. Fulton. The effects of exposure to high temperatures upon the circulation in man. *Am. J. Physiol.* 67:573-588, 1923/4.

Berger, R.E., M.R. Frei, and J.R. Jauchem. Computer-based monitoring of thermoregulatory responses of anesthetized rats exposed to radiofrequency radiation. *Physiologist* 34:246.

Daily, W. M. and T. R. Harrison. A study of the mechanism and treatment of experimental heat pyrexia. *Am. J. Med. Sci.* 215:42-55, 1948.

Frei, M.R., R.E. Berger, and J.R. Jauchem. Visceral vasodilation precedes circulatory shock during millimeter wave irradiation. In preparation.

Gisolfi, C.V., Matthes, R.D., Kregel, K.C., and Oppliger, R. Splanchnic sympathetic nerve activity and circulating catecholamines in the hyperthermic rat. *J. Appl. Physiol.* 70:1821-1826, 1991.

Hecker, M., M.L. Foegh, and P.W. Ramwell. The Eicosanoids: Prostaglandins, Thromboxanes, Leukotrienes, and Related Compounds. In: *Basic & Clinical Pharmacology*. Connecticut: Appleton& Lange, 1992, p. 263-277.

Kielblock, A.J., N.B. Strydom, F.J. Burger, P.J. Pretorius and M. Manjoo. Cardiovascular origins of heatstroke pathophysiology: an anesthetized rat model. *Aviat. Space Environ. Med.* 53:171-178, 1982.

Kregal, K.C. and Gisolfi, C.V. Circulatory responses to heat after celiac ganglionectomy or adrenal demedullation. *J. Appl. Physiol.* 66:1359-1363, 1989.

Kregal, K.C., P.T. Wall, and C.V. Gisolfi. Peripheral vascular responses to hyperthermia in the rat. *J. Appl. Physiol.* 64:2582-2588, 1988.

Kregal, K.C. and Gisolfi, C.V. Circulatory responses to vasoconstrictor agents during passive heating in the rat. *J. Appl. Physiol.* 68:1220-1227, 1990.

Parratt, J.R. Neurohumoral agents and their release in shock. In: *Handbook of Shock and Trauma*. edited by B.M. Altura, A.M. Lefer, and W. Schumer. New York: Raven Press, 1983, p. 311-336.

Rowell, L.B. *Human Circulation: Regulation during Physical Stress*. New York: Oxford University Press, 1986.

Schauer, A. Initial liberation of biogenic amines and effect of further mediators following application of endotoxins. In: *Gram-Negative Bacterial Infections*, edited by B. Urbaschek, R. Urbaschek, and E. Netter. Springer, Berlin, 1975, p. 315-322.

Wieser, E.S., Luong, A.U., Ryan, K.L., Frei, M.R., Berger, R.E., and Jauchem, J.R. (Abstract) Chronic nitric oxide (NO) inhibition: Effects on hypotension induced by 35-GHz radiofrequency radiation (RFR) exposure. *The FASEB J.* 2:A593, 1994.

USING ELECTRONIC BRAINSTORMING TOOLS TO VISUALLY REPRESENT THE IDEAS OF OTHERS: A PROPOSAL FOR RESEARCH

Kenneth A. Graetz

Assistant Professor

and

Scott Macbeth

Graduate Research Assistant

Department of Psychology

The University of Dayton

300 College Park

Dayton, OH 45469-1430

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

and

Armstrong Laboratory

September, 1994

USING ELECTRONIC BRAINSTORMING TOOLS TO VISUALLY REPRESENT THE IDEAS OF OTHERS: A PROPOSAL FOR RESEARCH

Kenneth A. Graetz

Assistant Professor

and

Scott MacBeth

Graduate Research Assistant

Department of Psychology

The University of Dayton

The manner in which electronic brainstorming tools visually represent ideas may have important consequences for ideational performance. Existing information displays differ with respect to (a) the degree to which users control their own access to group information, (b) the visual representation of the information on the screen, and (c) the emphasis on group versus individual productivity. An explanation for the apparent lack of creativity of electronically assisted, interacting groups is presented based on the distinction between blind versus heuristical search processes. It is argued that, while existing brainstorming tools eliminate or reduce the detrimental effects of various situational factors, the cognitive algorithm typically used by brainstormers in interacting groups, the trailblazing heuristic, still prohibits the exploration of previously activated ideational categories. Three computer brainstorming studies, involving manipulations of motivational orientation and information display, are proposed in order to explore the effects of this heuristic search process on ideational performance. The results are expected to enhance the development of effective brainstorming software.

RAPID BACTERIAL DNA FINGERPRINTING BY THE POLYMERASE CHAIN REACTION

Jason E. Hill, B.S. Biology
Nicholas F. Muto, B.S. Biology

University of Scranton
Scranton, PA 18510

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

RAPID BACTERIAL DNA FINGERPRINTING BY THE POLYMERASE CHAIN REACTION

Jason E. Hill, B.S.
Nicholas F. Muto, B.S.
Department of Biology
University of Scranton

Abstract

Typing of bacterial strains by polymerase chain reaction fingerprinting was studied. Bacterial strains were grown overnight and the DNA isolated by the CTAB method. This study utilized REP and ERIC primers, which target dispersed repetitive sequences, for gram negative bacteria (especially *E. coli*, *Salmonella*, and *Pseudomonas*). Primers were derived from repetitive sequences in *M. pneumoniae* and used with the gram positive organism *S. aureus*. Differential fingerprints were obtained by PCR showing which strains were derived from the same bacterial clone.

RAPID BACTERIAL DNA FINGERPRINTING

BY THE POLYMERASE CHAIN REACTION

Jason E. Hill
Nicholas F. Muto

Introduction

Bacterial typing is a complex and lengthy task. The many classical ways of typing bacteria generally belong to the realm of microbiology. These classical methods are not one hundred percent accurate. Difficulties in accurately typing bacterial infections are further compounded by the fastidiousness of certain organisms. DNA fingerprinting, a tool of the molecular microbiologist, is a powerful way of accurately typing bacteria. It obviates the lengthy process of culturing fastidious organisms and is widely applicable (when an outbreak of a bacterial infection occurs, it does not necessarily follow that every victim is infected with the same bacterial clone).

Most bacteria contain sequences that are repeated throughout their genome. These repetitive elements do not occupy the same positions in all clones. They may be separated by variable amounts of DNA in different clones. DNA fingerprinting utilizes this variable amount of DNA to type bacteria. The sequences of the repetitive elements are known and primers have been developed complimentary to parts of the repetitive sequences. PCR is used to amplify the regions of DNA between the primers. When the fingerprints are analyzed on an electrophoretic gel the different sizes of amplified DNA give a unique fingerprint for every clone. This procedure is both rapid and accurate.

Methodology

Bacterial cultures were provided by our focal lab. The DNA was isolated by the procedure described by Versalovic, et al. 1994, p. 17, but with alterations. Briefly, 1.5 ml cultures were grown overnight in BHI broth. The cells were pelleted at 3,000 rpm for 10 min. in a microcentrifuge, then washed once with 1M NaCl and once with TE buffer. The cells were resuspended in TE buffer and lysed with 10 ul lysozyme (5 mg/ml) for gram negative bacteria and 75 units of mutanolysin for gram positive bacteria (12 units of lysostaphin for *S. aureus*). The cells were incubated for 30 min. at 37°C, then 30 ul of 10% SDS and 3 ul proteinase K (20 mg/ml) were added and the cells were incubated for 1 hr at 37°C. To this 100 ul of 5M NaCl was added followed by 80 ul of 1% CTAB/1M NaCl solution. The samples were incubated for 10 min. at 65°C. The samples were extracted once with an equal volume of chloroform, once with phenol:chloroform, and finally one more time with chloroform. The DNA was precipitated with an equal volume of isopropanol and resuspended in sterile water. The DNA was incorporated into a 25 ul PCR using the following reagent concentrations: 1 XPCR buffer #11 (Opti-Prime Kit from Stratagene), 15 mM bovine serum albumin (Opti-Prime Kit from Stratagene), 300 uM dNTP mix (by Boehringer Mannheim), 1 uM of two opposing oligonucleotide primers (Wenzel & Herrman, p. 8338) (Table 1), 100 ng DNA, and 1.75 U/ul Taq polymerase (Perkin Elmer/Cetus). The reaction was brought up to 25 ul with sterile water. This PCR cocktail was used for gram positive bacteria. The primer sequences were taken from an article describing repetitive sequences in *M. pneumoniae* and were synthesized by the Midland Certified Reagent Company. Cycling conditions were as follows: Initial denaturation at 95°C for 3 min., then 30 cycles

of denaturation at 94°C-1 min., annealing at 43°C-1 min., extension at 72°C-2 min., and a final extension at 72°C for 5 min. The PCR samples were visualized on an agarose gel (1.2% Seakem GTG agarose in 1X TAE buffer containing ethidium bromide) and photographed. Fingerprint bands were sized using a 100 bp ladder and a 1 kb ladder from Gibco BRL. Fingerprint bands were compared based on size and intensity. For gram negative bacteria the PCR cocktail was different: 1X PCR buffer (Versalovic, et al. 1994, p. 21), 10% DMSO, 1.25 mM each dNTP, 50 pmoles of two opposing primers (ERIC) (Table 1), and 2 units of Taq polymerase. The cycling conditions were: 95°C for 7 min., 30 cycles of 52°C-1 min., 65°C-8 min., 94°C-1 min., and 65°C for 16 min.

Results

Figure 1 shows five *S. aureus* strains amplified with two different sets of primers, P1-M2 and RW3, and RW2A and RW3A. Strains 61, 1816, and 1844 have similar fingerprints with both sets of primers. Strains 1824 and 1830 are clearly different from the other strains, demonstrated again by both primer sets.

Figure 2 shows ten strains amplified with *M. pneumoniae* primers RW2 and RW3. This figure is an example of the difficulty we had with DNA concentrations. Some of the fingerprints are distinct, others are faint, and some do not appear at all.

For gram negative bacteria, ERIC-PCR was utilized to generate fingerprints of their genomes. The most complex and distinct genomic fingerprints were obtained from samples of *E. coli* (Fig 3). *Pseudomonas aeruginosa* (Fig 4), as well as some *Salmonella* samples (Fig 5) also yielded fairly decent amplification products.

Conclusion

Four primers were developed from *M. pneumoniae* repetitive sequences. Of all possible combinations P1-M2, RW3 and RW2, RW3 amplified the most strains. Consistency of the PCR was the biggest stumbling block. A PCR that produced excellent fingerprints one time would produce either poor or no fingerprints when run a second time under the same conditions. There are many nuances of multiplex PCR that need to be considered compared to normal PCR. One explanation is that we generally had low yields of DNA from gram positive organisms, especially *S. aureus*, due to the tenacity of the cell wall to resist complete lysis. Also, annealing temperature seemed to be critical even to within one degree although we did not have time to fully explore the effects of altering the annealing temperature. With the research accomplished at Brooks AFB, our home laboratory at the University of Scranton in Scranton, PA should develop a consistent fingerprinting protocol for *S. aureus* which will then be used at Brooks AFB in the near future.

Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences occur throughout the genomes of many enteric gram negative bacteria. By performing PCR with a primer set found within these sequences and amplifying the regions between the sequences, unique, strain specific fingerprints are obtained allowing for the typing of these organisms. Various samples of the bacteria *E. coli*, *P. aeruginosa*, and *Salmonella* were typed with this method. As the work we have begun is continued at the University of Scranton, we will eventually be able to examine entire outbreaks of infection by a certain bacterium and determine the source of each case involved. This is a powerful and invaluable use of the polymerase chain reaction.

Electrophoresis Profile

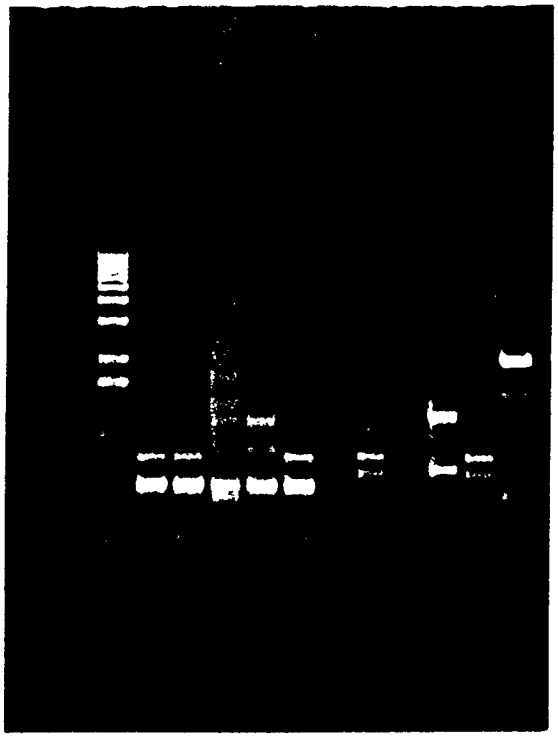
Agarose NS-GTG, SP-GTG, (8K-GTG)
Percent gel 1.2
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 8/8/94

Voltage 65 V
Final time
Initial mAMP
Initial time
Total time

WELL #	SAMPLE	
1	1 Kb ladder	
2	S. aureus 61	Primers P1-M2 and RW3
3	1816	
4	1824	
5	1830	
6	1844	
7	61	Primers RW2A and RW3A
8	1816	
9	1824	
10	1830	
11	1844	
12	100 bp ladder	

Figure 1



Electrophoresis Profile

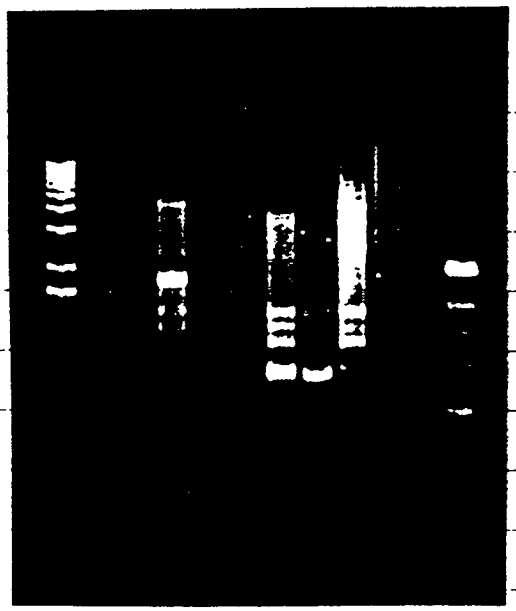
Agarose NS-GTG, SP-GTG, (8K-GTG)
Percent gel 1.2
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 7/28/94

Voltage 65 V
Final time
Initial mAMP
Initial time
Total time

WELL #	SAMPLE
1	1 Kb ladder
2	S. aureus 2289
3	2290
4	2291
5	2292
6	2293
7	1844
8	2344
9	2345
10	61
11	negative control
12	100 bp ladder

Figure 2



Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE) TBE
Electrophoresis apparatus (S), M, L

Date 7/22/94

Voltage 50
Final time
Initial mAMP
Initial time
Total time 3:00

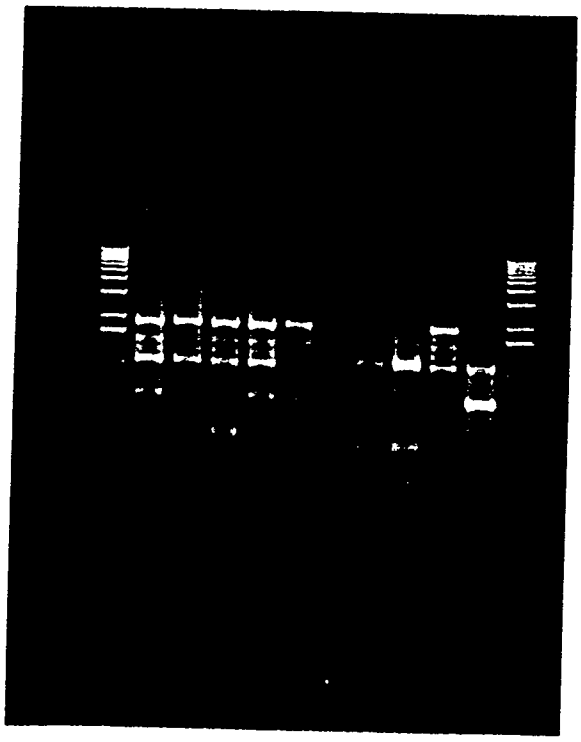
WELL #	① SAMPLE	②
1	1kb LADDER	13 1kb
2	2446	14 1001
3	996	15 994
4	2287	16 BT93 502
5	2296	17 BT92 2172
6	3067	18 995
7	2295	19 997
8	1002	20 1000
9	STRAIN 0157:H7	21 1003
10	BT93 499	22 999
11	1008	23 2432
12	1kb	24 3118

ERIC 1R/
ERIC 2
PRIMERS

10 µL/well

ALL SAMPLES
E. coli

①



②

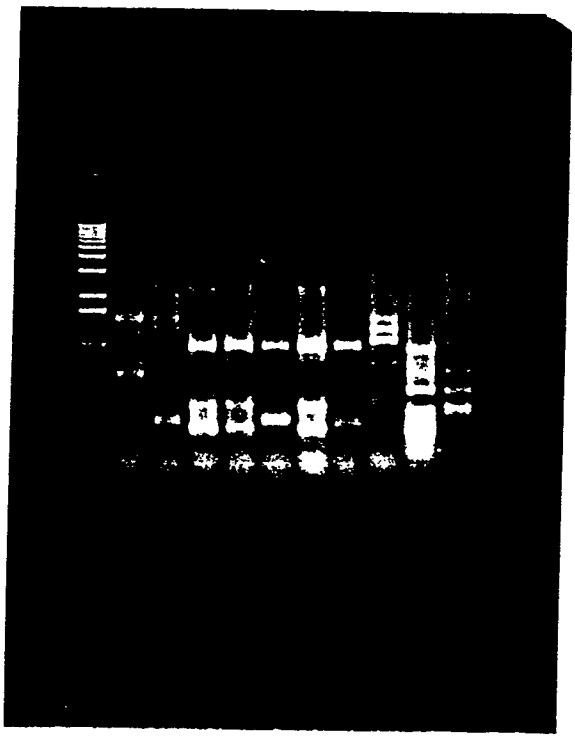


FIGURE 3

Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE), TBE
Electrophoresis apparatus (S), M, L

Date 7/26/94

Voltage 50
Final time
Initial mAMP
Initial time
Total time 2:30

WELL #	SAMPLE
1	1kb LADDER
2	P. cepacia 2573
3	" 2659
4	" 2575
5	" 2691
6	" 3005
7	" 3002
8	E. coli BT94 2252
9	BT94 1711
10	BT94 3070
11	NEG. CONTROL
12	1kb LADDER

ERIC 1R/ ERIC 2 PRIMERS

10pL/well



FIGURE 4

Electrophoresis Profile

Agarose NS-GTG, SP-GTG, (SK-GTG)
Percent gel 1.2%
Buffer (TAE) TBE
Electrophoresis apparatus (S) M, L

Date 3/4/94

Voltage 60
Final time
Initial mAMP
Initial time 3:10
Total time 2:00

WELL #	SAMPLE
1	1kb LADDER
2	P. cepacia 3002
3	" 3005
4	Salmonella 2982
5	" 2980
6	" 3072
7	" 2989
8	UNKNOWN 3003
9	E. coli BT92 3070
10	BT94 2252
11	" 2939-2
12	" 2539-1
13	NEG. CONTROL
14	100 bp LADDER

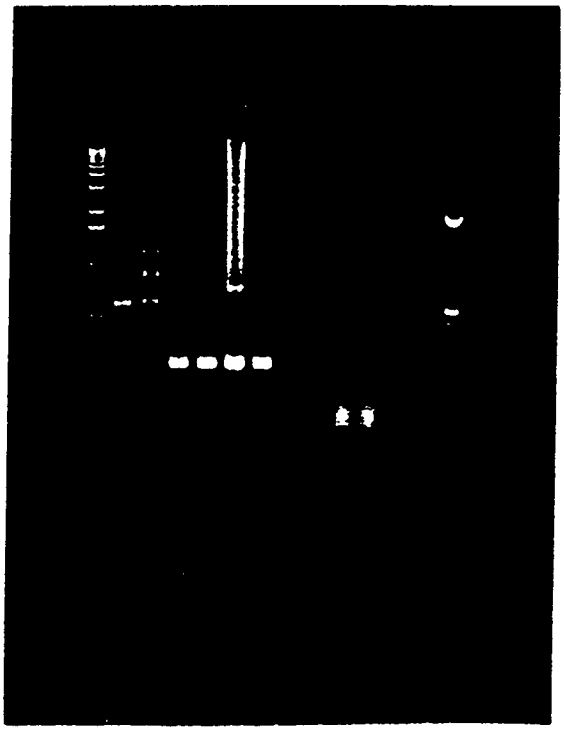


FIGURE 5

FIGURES/TABLES

Table 1: Composition of Oligonucleotide Primers

ERIC 1R	3'-CACTTAGGGGTCCTCGAATGTA-5'
ERIC 2	5'-AAGTAAGTGACTGGGGTGAGCG-3'
RW2	5'-TCTTTACGCGTTACGTATT-3'
RW3	5'-CTCAAAACAACGACACCGG-3'
P1-M2	5'-CCCCCACCACTAAGCACAC-3'

BIBLIOGRAPHY

Versalovic, J., Koeuth, T. and Lupski, J.R., 1991. Distribution of Repetitive DNA Sequences in Eubacteria and Application to Fingerprinting of Bacterial Genomes. *Nucleic Acids Research* 19(24),6823-6831.

Versalovic, J., Schneider, M., deBruijn, F.J., and Lupski, J.R. Genomic Fingerprinting of Bacteria Using Repetitive Sequence Based PCR (rep-PCR) *Methods in Molecular and Cellular Biology*, IN Press.

Wenzel, R., and Herrmann, R., 1988. Repetitive DNA Sequences in *Mycoplasma pneumoniae*. *Nucleic Acids Research* 16(17), 8337-50.

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Professor
Eric Riise
Graduate Student
Department of Communication Design

California State University, Chico
Chico, CA 95928

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Professor
Eric Riise
Graduate Student
Department of Communication Design
California State University, Chico

Abstract

Interaction in distance learning was studied. A survey of the literature found that most studies were lacking in rigor and the methodologies were weak in regards to interaction. To answer the many questions about interaction effects in distance learning, a better definition of the variable interaction is needed. This paper lays out a taxonomy of interaction for evaluation and research.

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Eric Riise

Introduction

In the past distance learning has been largely used to bring education and training programs to learners who would otherwise not have access to the classes offered (*U.S. Office of Technology Assessment, 1989).* Courses and programs have been offered primarily for adult learners interested in college credit or vocational and professional improvement. Distance learning networks are operated for the most part by colleges and universities or by corporations and government agencies.

There is ample evidence in the literature that, in most cases, distance learning appears to be as effective as face-to-face instruction in the classroom (Moore, 1989). A comprehensive review of the current research regarding the use of dynamic video media in instruction conducted by Wetzel, et al, (1994) found, "The general conclusion from this evolving field is that it is possible to have no decrement or only a small decrement at a remote site compared to the performance of students at the live transmitting site, (p. 20)". They examined studies of both preproduced telecourses (i.e., non-interactive) and interactive teletraining in terms of effectiveness, acceptance and costs. They found the primary attraction of distance learning for students is convenience: proximity to where they work and live and flexibility in personal scheduling and work requirements.

In a comprehensive review of the current research regarding the use of dynamic video, the Texas Higher Education Coordinating Board (1986) reviewed the course results of four college level telecourses and found student achievement was comparable to conventional on-campus classes. They also examined grade distributions of Texas community colleges offering telecourses and found they did not differ significantly from traditional classroom grade spreads. A review of the literature by Miller, et al, (1993) could not identify a single study that has shown distance learning diminished content learning. Some studies found advantages of distance classes over traditional classroom instruction (e.g., Barron, 1987, Weingard, 1984, and Keane and Cary, 1990).

Most distant learners report they are satisfied with their remote instruction and some reportedly preferred the distance learning mode. However, the limited number of studies in the

affective domain and the lack of rigor in the methodologies do not permit reliable conclusions about preference.

That these effects hold true for many subjects and a variety of media and delivery means, indicates that learner motivation may be an overarching factor in the learning process *for these students.* Indeed, it has been asserted that motivation is the single most important factor for student learning and when motivation is high, it is difficult to prevent learning (Main, 1992).

However, distance learning is entering a new phase. The transformation from analog to digital communication technology is creating a new environment for multimedia interactive instructional media and telecommunication networks. Distance learning is no longer being viewed as simply a means to provide access for those unable to meet in the classroom, but as a viable alternative to classroom instruction as a primary mode of instruction . The promise of two-way interactive video, voice and data available in every home and office via the information superhighway has fired the imagination of educators and non-educators alike in the potential for providing elementary and secondary schooling in classrooms without walls. Decisions are now being addressed on the basis of how cost-effective distance education is compared with traditional classroom instruction. The sentiment is reflected in a recent comment by a corporate officer that the "lean budgets of today's economy drive alternative training and educational delivery systems. Traditional stand-up instruction does not stand up to the scrutiny of the cost conscious business manager." (Grant, 1994) Universities are experimenting with delivery of instruction to students in their dormitory rooms or homes through local area networks or public data services such as America Online. Public telephone and cable distribution systems are under study as well, as a means of providing instruction without assembling students in a classroom.

Whether the success of distance learning with adult learners will work equally well for all students is still a question. Most applications to date have been involved with academically advanced high school students and independent adult learners--individuals who presumably already possess strong study skills, high motivation and discipline (*U.S. Office of Technology Assessment, 1989).* The Congressional Office of Technical Assessment (OTA) has concluded that research of distance learning for elementary and secondary application would be most usefully concentrated on practical questions about the educational experience such as learner outcomes of various teaching techniques and instructional design approaches.

With distance learning being considered as a replacement for traditional classroom, the designers and developers of distance learning instruction can no longer depend upon the intrinsic

motivation of self-selection. Like the public school teacher in the traditional classroom, the students will present an array of interest in the subject and education for the distance learning instructor. The changing nature of the distant learner from adult volunteer to adolescent required attendee presents new requirements for the instructional designer and teacher.

This study examines interactivity as a function of the instructional design and presentation of distance learning lessons. The complex interplay of interaction in distance learning is not well understood at this time (Haynes and Diehorn, 1992). One reason is the relative dearth of studies examining interaction in distance learning education. Others include the poor controls used in the research that has been conducted and a reliance on self-selected groups exposed to the distant learning and traditional classroom conditions. Finally, there is the relatively simplistic manner in which interaction has been defined in the studies. It is also likely that studies showing poor performance for distance learning situations are less likely to be submitted for publication or published when they are submitted.

Background

Intuitively we know that interaction is important in the instructional process. We strive for interaction in the traditional classroom. The concept of small teacher-student ratios is based on the belief that the smaller class size permits a richer interaction. The ultimate learning environment is considered to be one-on-one where the instruction can be individualized to the student's perceived needs and learning style. It is axiomatic that proximity in interpersonal communication enriches interaction. Wetzel, et al, (1994) determined that, "Increasing the degree of fidelity or interactivity of video teletraining to that with live instruction generally increases effectiveness and satisfaction "(p. 21). But the empirical evidence is weak and the studies cited are generally lacking in methodological rigor. Klinger and Connet (1992) state, "...telecourses must include a strong element of interaction to be truly effective as a learning method. Interaction is *essential* for the student to remain interested and steered forward for success" (p. 88). Their conclusions, however, are based on experience rather than empirical studies. How, then, are we to explain the results of the many studies which indicate there is little difference in learning between students in the traditional classroom and students at distant learning sites?

It is difficult to tell from the literature. There are very few studies that have examined interactivity as an independent variable and those that purport to have studied its effects generally looked at interaction only in terms of frequency. For example, Van Haalen and Miller (1994) reported on interactivity as a predictor of student success in a satellite learning program, but interactivity was measured on the basis of telephone logs recording only the number of calls from

students to the teacher both during and after class for the school year. No attempt was made to capture the length of the interaction or its topical relevance. Most interaction reports are observational and associated more with learner attitudes about the delivery mode than with achievement. Rupinski (1991), for example, found that student preferences for traditional classroom training can be reduced by making conditions at the remote site (including two-way video) more like a "live" classroom.

From studies where interaction is included as a variable, the effects of interaction of learning outcomes is ambiguous. In a study comparing instruction by audiotape, videotape and telelecture, Beare (1989) found the lack of individual opportunity to interact with the instructor regularly did not significantly reduce student scores on course examinations. In a one-way video course with two-way audio, Van Haalen and Miller (1994) reported interaction effects were not linear but, rather, a polynomial curve in the form of an inverted U. At each end of the interactive continuum, student performance (in terms of course grades) was poor. They only measured student initiated interactions with the instructor, however, and not interactions designed into the instructor's presentation as student-student discussion activities. It is possible that in this situation, the students with the most questions are those with a need for additional information to keep up with the instruction. Conversely, students who never ask questions may be reluctant to expose a lack of knowledge.

A problem with the descriptive studies is the lack of a control group. How do the students in the distant learning class differ from those in the residential classroom? Zigerell (1986) gives a hint with his survey of telecourse students in community college courses. Most of the students had not taken a telecourse before (65 percent). Of those, 69 percent were women, and carried less than 10 semester units. About half worked at least 40 hours a week. Only 17 percent said they were enrolled because they preferred distance learning. This is quite different from most residential college students.

In one of the more carefully designed comparative studies, (Simpson, et al, 1991) found the most critical condition for success in interactive teletraining is the ability of students to see the instructor and have two-way audio communication . Two-way video appeared to have little effect, but any degradation in audio quality caused negative comments. Not surprising since most instruction is still language-based. In comparing final examination scores, the decrease between student performance at the originating site and the remote sites was less than three percent in any of the instructional modes. The value of Simpson's studies are that they compared complete courses-

-not just modules covering a few hours of instruction. Stoloff (1991) found distant learners became more indifferent to differences between teletraining and traditional classroom methods of instruction over time, but instructors still favored their classroom.

There is a need in distance learning research to adopt an expanded view of effective teacher-student communication. It involves integrating a variety of communication forms and channels that include verbal communication, vocal communication--the volume, rate, tone, pitch and inflection--mediated messages, body language and situational messages--manipulation of distance, time and number of participants (Hennings, 1975).

Instruction, Communication and Interaction

Teaching is primarily a communication art. If we accept the interdependent relationship between source and receiver in the communication process described by Berlo (1960), then teaching should emphasize interaction between instructor and learners. We learn by taking an active role in the process (Hefzallah, 1990). Buckminster Fuller asserts the instructional environment "is an interacting situation in which the continuity of experience and the relating of experience are critically important." (1966, p 16). Hefzallah summarized the connection succinctly, "to teach is to communicate, to communicate is to interact, to interact is to learn (1990, p 38)."

Socrates knew the value of interaction in learning. Students were required to discover knowledge through a series of questions and answers. By contrast, the Sophists were the first lecturers. They knew everything and were ready to explain it (Highet, 1957). But, here's the rub. While the Sophists grew rich, dressed in royal purple and traveled by sedan chair, Socrates' sandals were worn and his tunic undyed. His discovery learning was not cost-effective. To make intelligent decisions in designing distance learning systems and lessons, we need information about the trade-offs between effectiveness and efficiency in the amount and quality of interaction in the instructional process.

The seminal studies by Chu and Schramm (1979) established that children learn efficiently from instructional television and from instructional radio "given favorable conditions" (p.vi and 1). The favorable conditions generally refer to the similarities in presentation where students in both groups hear the same lecture, see the same visual and read the same printed materials. Most of the studies supportive of these conclusions contrasted students in traditional classes with those at remote locations. For the most part, the studies reviewed by Chu and Schramm used a mass media model for the instruction, i.e., the transmission was largely one-way with feedback limited and delayed. In this mode, the student is largely passive, at least in terms of real-time interpersonal

interaction. This is the conduit theory of communication applied to distance learning described by Clark (1983) as an analogy in which "media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition," (p. 445). To extend Clark's analogy, however, if the truck can't be refrigerated to carry fresh fruit, that does affect our nutrition. So, the capability of the delivery system is a factor in the instruction presentation and possibly student learning.

In his review of the distance learning literature, Kozma (1991) acknowledged the mass media conduit theory, but concluded the interpersonal theory of communication with its rich and immediate feedback was more appropriate for the interactive nature of teaching. Distance learning should attempt to replicate the "live" classroom through "virtual medium" (Kozma, 1992, p. 182).

It's not only the quantity of interaction that affects the learning . The quality of interaction is also a factor. The interminable prompt to press ENTER that was so common in early computer aided instruction although interactive was a numbing experience. Dale states "education has to choose creative interaction of the learners over rote imitative reaction (1978, p 24)." This is what designers of distance learning education face: determining the amount and nature of interaction that is most effective and efficient for achieving the learning objectives of the class.

Two issues are identified by Miller, et al (1993) as being important in considering how well distance learning duplicates the learning environment of the traditional classroom. The first is whether interaction is reduced among distance learners. Even though the technical capacity for such interaction is available, students may be inhibited from participating interactively by a variety of reasons such as awkwardness in interrupting or unfamiliarity with equipment. The second issue concerns the degree of student engagement in traditional and distance learning conditions, i.e., do distant students tend to be less attentive in a distance learning environment? Nadel (1988) in a comparative study of distance learning modes concluded that students learn from any medium, in school or out, whether they intend to or not, providing the content of the medium leads them to *pay attention* (emphasis added). That is a very large proviso and corresponds to Kozma's (1991) concept of involvement which is manifested behaviorally as participation and can be measured by interactions.

A Taxonomy of Interaction for Evaluation and Research

To answer the many questions about interaction effects in distance learning, a better definition of the variable interaction is needed. Interaction in distance learning is obviously a complex behavior. What is needed is a model for examining its many dimensions. Boak and Kirby

(1989) developed the System for Audio Teleconferencing Analysis (SATA) instrument for analyzing classroom interaction that affords some direction for researchers. Their schema has three categories: who initiates the interaction (student or instructor); the direction of the interaction (an individual student, the class as a whole, or instructor); and the context of the interaction (procedural, content specific, or social).

This schema can be expanded to examine more thoroughly the interaction process in the distant learning classroom. Six categories of interactivity have been identified by this researcher as relevant for distance learning research. These may not be comprehensive but provide a beginning point in developing a taxonomy of distance learning classroom interaction. They are **AMOUNT, TYPE, TIMELINESS, METHOD, SPONTANEITY,** and **QUALITY**. Each of these components are compound variables themselves with several levels.

1. There are two dimensions in measuring the **amount** of interaction--the frequency of occurrence and the length of the dialog. Frequency is perhaps the most commonly captured data in distance learning studies involving interaction. It is most often examined in terms of how often student feedback occurs, i.e., the mean occurrence per student per class period. Frequency can also be measured by how it is spaced during the presentation (clustering by instructional activity.) The length of each interaction is relatively straight forward and is of greatest interest when related to **type, method** and **quality.**

2. The **type** of interaction refers to the participants. In a distance learning class this would include instructor-student exchanges, student-student interactions, and student involvement with the lesson materials. The instructor-student interaction can be further categorized as to whether it is instructor initiated or student initiated. Student-lesson materials interaction may be either required or student choice. Each of these levels can be further subdivided as occurring within the class period or outside the class period. The measurement would be frequency of occurrence. The cells would appear as follows:

Fig. 1 Type of interaction

Initiated by		Student-Student	Student-lesson Materials		
Instructor	Student		Required	Voluntary	
					Within Class
					Outside Class

Measuring student-to-student interaction will require some means for observing or recording activity at the distant learning site(s). Haynes and Dillon (1992) found distance students in a library science course interacted much less with the instructor and much more with each other during class even though they complained at times that it interfered with attending to the instructor. The results of the study did not indicate a significant difference in student performance between exam scores of distant and on-site learners which would seem to show interaction type has little effect on distance learning. There are some methodological problems with the study, however, that make generalization of the findings difficult.

3. **Timeliness** is a measurement of the immediacy of the communication feedback. It is the amount of time between the attempt to interact begins until the message is received by the addressee. It is an indicator of the efficiency of the communication system. It presumes a two-way communication system. In those situations where the instruction is preproduced and delivered on schedule or on demand, there is no interaction in the class. Broadcast television and cable delivered instruction fall into this category. It's the mass media model of communication. An example might be Ken Burns' *Civil War* television series aired over PBS. It was certainly educational, but a passive one-way delivery when viewed on PBS. However, when video tapes of that series are used by a teacher as instructional material in a distance learning history class where feedback is expected of students discussing the programs, the instruction becomes interactive. Timeliness is a continuous variable that ranges from zero in the real-time interactions of a traditional classroom to several days or even weeks for a correspondence course administered through the post office.

4. The **method** of interaction refers to the manner in which the communication message is encoded. Voice is the most common method of interaction in the traditional classroom. Satellite transmissions of one-way video with a two-way telephone audio channel have been the system of

choice for most distance education and training systems. However, there has been considerable interest in text-based interaction systems using computer-based data network delivery. With the conversion of analog to digital communication and the interest in establishing high capacity public switched digital infrastructure, there is an expanding effort to determine how this information superhighway can be exploited for education and training use. Already, compression technologies allow two-way, real-time digital video and audio transmission over conventional twisted pair phone lines through digital switches albeit without full motion or the fidelity of analog television. The **method** of interaction, then, should be addressed in studies of interaction effects. In addition to voice and text, interaction may occur through visual non-verbal gestures, response pads, graphic display, and photos. There are obviously many combinations and sub-levels possible with the various methods that need to be considered in developing the measurement methodology and instrument especially when newer multimedia workstations are used in the delivery of instruction.

5. The **spontaneity** dimension of interaction refers to whether the feedback is a planned event embedded in the lesson plan as part of a learning activity or a spur of the moment exchange triggered by the presentation. It may be important to determine whether *ad hoc* interactions are one-on-one or part of a group discussion. **Spontaneity** can be cross-tabbed with **amount, type** and other variables of the interaction schema.

6. The **quality** of interaction is the most difficult dimension of interaction to define operationally. The possible levels are almost infinite. Many of the other categories have quality implications and a case could be made that this is an overarching variable that subsumes all the components of interaction. For purposes of this taxonomy, **quality** is defined in five dimensions: *intensity, relevance, depth, formality,* and *opportunity*. *Intensity* reflects the emotional involvement of the participants in the interaction. The levels are routine (which includes repetitive, procedural and expected responses); interested (exploratory, explanatory, and expansive), and emotionally involved (excitement, fear, enjoyment, attachment, and anger). It is difficult to distinguish the intensity of a communication exchange, but trained observers can discriminate among the categories.

The components of *relevance* are classified as professionally related, involve the lesson content (subject matter) or have personal relevance. *Depth* is a continuum ranging from the trivial to substantive. The *formality* of the interaction is classified as formal or informal. *Opportunity* is the ability to interact when desired. It could be a function of class size, the technical capability of the system, or the instructional design of the lesson that accommodates interactions. Real-time two-way audio and video is expensive and the cost increases in direct relation to the number of distant

learning sites. It is important in emulating a traditional classroom, but the value decreases as the number of students in the class increases. The *opportunity* for interaction is inversely correlated with the class enrollment. The effect of class size is as true for the student in the traditional classroom as it is for the distant learner. Everyone remembers those large lecture halls where a professor addresses a class of hundreds. A satellite broadcast may enlarge the class to thousands of students. The chance of interacting with the instructor dwindles no matter how sophisticated the communication system. While the concept of the President appearing on a national talk show to interact with the public is politically appealing, the opportunity of any particular citizen to actually ask the president a question (never mind a give and take dialog) approaches the probability of winning the lottery. The idea that the caller who does get to ask a question represents some number of other viewers or listeners may have some validity, but is accomplished more economically by use of studio questioners. It is worthwhile measuring timeliness, however, even when the size of the class makes opportunity difficult. A study by Fulford and Zhang (1993) suggests the perception of overall interaction is a stronger predictor of student satisfaction than personal interaction. Although the class size was only 123 students in five locations, two of which had one-way video and two-way audio and three with two-way video and audio. The perception of overall interaction (self report) and satisfaction with the class had a strong correlation despite the actual number of personal interactions. The strength of "vicarious" interaction effect did diminish from the first of the three sessions to the last. We shouldn't be too surprised by these findings. The appeal of game shows and talk shows is largely the interaction between host and guests or contestant.

This taxonomy of interaction variables provides a framework for research and evaluation of the effects of interaction in distance learning. It may need modification and elaboration as new questions arise, but it allows the research to proceed more systematically in order that findings may be grouped for meta-analysis and meaningful comparisons made among studies. The next step is to develop operational definitions and measurement instruments for each variable that can be tested for accuracy and validity. The goal is to establish a body of literature from which theoretical concepts and generalizations can be made as to the efficacy of interaction activities that will be useful to system designers and instructional developers of distance learning instruction. The need for better methodology in distance learning studies is apparent. Research to date indicates there is little difference in achievement attributable to delivery technique. Intuitively that does not seem right even though studies have consistently reported performance of standardized tests to be similar, regardless of medium used (Salomon and Clark, 1977 and Ritchie and Newby, 1989).

Relating Interaction with Other Variables

Interaction always occurs within a context. The utility of organizing the dimensions of interaction variables lies in finding how they relate to other components of distance learning. There are numerous factors that may be affected by, or have an affect on, interaction in distance learning. Generally these factors can be classified as those concerned with the course and those dealing with its delivery, i.e., the communication technology. The communication technology is continually changing and especially at this watershed stage of conversion from analog to digital communication. Not only are the media becoming amorphous with digital multimedia, but also the industry infrastructure is in a state of flux as telephone, cable, and television companies seek acquisitions, alliances and mergers that will position them as players in the digital, interactive, multimedia future of the information superhighway. New products, new systems and new capabilities will demand continuing research for its effects on distance learning interaction.

Instructional strategies and activities involve all the components of instructional design with the added complexity of distance delivery (Wagner, 1990). There is a large body of literature available on instructional process, but despite the scrutiny of what goes on in the classroom, teaching remains very much an art form. Distance learning may depend even more on instructor charisma and style than the traditional classroom in which case instructor characteristics are important to examine in terms of their effect on interaction. It is axiomatic that the difference between a good teacher and a great teacher is the ability to motivate their students to learn (Main, 1992). Interpersonal communication skills are more critical when students are not physically present in the classroom. The technology of distance learning changes the dynamics of instruction. Beaudoin (1990) suggests distance education revolves around a learner-centered system with instructor skills focused on facilitating learning and organizing instructional resources.

Inserting technology in the instructional process requires greater attention to lesson design and instructional preparation. This factor needs to be more carefully examined and controlled in distance learning research. Miller (1989) argues that curriculum issues are more important than the delivery technology. Farr and Shaeffer (1993) provide a discussion on media selection variables for distance learning application.

Course variables include such things as the subject matter, student characteristics, instructional strategies and activities, media selection and instructor characteristics. Subject matter can be characterized in terms of type (cognitive, psychomotor, affective), depth or complexity (basic skills, advanced studies), application (practical, theoretical), level of proficiency (familiarity, mastery, automation) and domain (history, language skills, electronics, etc.). This listing is not

comprehensive. Each category is a compound variable and the dimensions provided are certainly not exhaustive. There may be other taxonomies of instructional techniques and subjects that may be more useful for hypotheses generation for a particular distance learning situation.

Student characteristics involve age, gender, motivation, prior knowledge and experience. An important consideration is whether enrollment is voluntary or required. Self-selection of distance learners and traditional classroom students contaminates many of the field studies reported in the literature. This may not be an important factor when distance learning is only used as an outreach for students unable or unwilling to attend residence courses. The relevant question is, do the students learn? How does interaction affect the learning for *these students.* It is when distance learning is being considered as an alternative for traditional classroom education and training that attention needs to be given to any differences between the comparison groups. To make generalizations about interaction effects for this use of distance learning, the differences in characteristics of students who select distance learning and those in the traditional class setting must be controlled.

Summary

The successful expansion of distance learning as an alternative to the traditional classroom is dependent upon the improvement of instructional design to approximate the richness of the interaction that occurs face-to-face. The technology for fully interactive distance learning is not the hurdle. The problem is how to elicit active participation by the learner whose interest in the particular subject and education, in general, is minimal at best. Interactivity seems to hold promise. We need to find the best techniques for achieving it in a cost effective manner. Hopefully, this taxonomy will serve as a useful tool in finding some answers.

References

Barron, D. D. (1978). "Faculty and Student Perceptions of Distance Education ," *Journal of Education for Library and Information Science,* Vol. 27, p. 257-271.

Beare, P. L. (1989). "The Comparative Effectiveness of Videotape, Audiotape, and Telelecture in Delivering Continuing Teacher Education," *The Journal of Distance Education,* Vol. 3, No. 2, 58-66.

Beaudoin, M. (1990). "The Instructor's Changing Role in Distance Education," *The American Journal of Distance Education,* Vol. 4, No. 2, p. 21-30.

Berlo, David K. (1960). *The Process of Communication--An Introduction to Theory and Practice,* Holt, Rinehart and Winston, Inc., New York, NY.

Boak, C. and D. Kirby (1989). "Teaching by Teleconference: What Goes On," *Proceedings of the 8th Annual Conference of the Canadian Association for the Study of Adult Education,* University of Ottawa, Cornwall, Ontario, Canada, p. 26-32.

Chu, Godwin C. and Wilbur Schramm. *Learning from Television, (4th Ed.),* National Association of Educational Broadcasters, Washington, D.C.

Clark, R. (1983). "Reconsidering Research on Learning from Media," *Review of Educational Research,* Vol. 53, p. 445-459.

Dale, Edgar (1978). *The Good Mind,* Phi Delta Kappa Educational Foundation, Fastback 105, Bloomington, IN.

Farr, C. W. and J. M. Shaeffer (1993). "Matching Media, Methods, and Objectives in Distance Education," *Educational Technology Journal,* July, p. 52-55.

Fulford, Catherine P. and Shuquiang Zhang (1993). "Perceptions of Interaction: The Critical Predictor in Distance Education," *The American Journal of Distance Education,* Vol. 7, No. 3.

Fuller, Buckminster (1966). "What I Have Learned, How Little I Know," *Saturday Review,* Vol. 49, No. 46, Nov. 12, p 6.

Grant, F. L.(1994). Personal conversation with the author regarding the feasibility of integration of corporate training and technical information services into virtual classroom and library resources at the Idaho National Energy Laboratory.

Haynes, Kathleen J. M. and Connie Dillon (1992). "Distance Education: Learning Outcomes, Interaction, and Attitude, *Journal of Education for Library and Information Science,* Vol. 33, p. 35-46.

Hefzallah, Ibrahim M. (1990). "The Learning Environment," Chap 2 of *The New Learning and Telecommunications Technologies,* edited by Ibrahim M. Hefzallah, Charles C. Thomas, Pub., Springfield, IL.

Hennings, Dorothy G. (1975). *Mastering Classroom Communication--What Interaction Analysis Tells the Teacher, Goodyear Publishing Co., Pacific Palisades, CA.*

Highet, Gilbert (1957). *The Art of Teaching,* Vintage Books, New York, NY.

Keene, D. S. and J. S. Cary (1990). "Effectiveness of Distance Education Approach to U.S. Army Reserve Component Training," *The American Journal of Distance Education,* Vol. 4, No. 2, p. 14-20.

Klinger, Toby H. and Michael R. Connet (1992). "Designing Distance Learning Courses for Critical Thinking," *T.H.E. Journal,* October.

Kozma, R. B. (1991). "Learning with Media," *Journal of Educational Research,* Vol. 61, p. 179-211.

Main, Robert G. (1992). *Integrating the Affective Domain in the Instructional Design Process,* U.S. Air Force Technical Report, Armstrong Laboratory, Brooks Air Force Base, TX.

Miller, J., W. C. McKenna and P. Ramsey (1993). "An Evaluation of Student Content Learning and Affective Perceptions of a Two-Way Interactive Video Learning Experience," *Educational Technology,* June, p. 51-55.

Moore, Michael, 1989. *Effects of Distance Learning: A Summary of the Literature,* OTA contractor report, U.S. Dept. of Commerce, National Technical Information Service (NTIS), Springfield, VA.

Ritchie, H. and T. J. Newby (1989). "Classroom Lecture/Discussion vs. Live Televised Instruction! A Comparison of Effects on Student Performance Attitude and Interaction, *The American Journal of Distance Education,* Vol. 3 No. 3, p. 36-45.

Rupinski, T. E. (1991). *Analysis of Video Teletraining Utilization, Effectiveness, and Acceptance,* CRM Research Memorandum 91-36, Center for Naval Analysis, Alexandria, VA.

Salomon, G. and R. E. Clark (1977). "Reexamining the Methodology of Research of Media and Technology in Education;" *Review of Educational Research,* Vol. 47, No. 1, p. 99-120.

Simpson, Henry, H. Lauren Pugh and Steven W. Parchman (1991). *Empirical Comparisons of Alternative Video Teletraining Technologies,* (NPRDC-TR-92-3), Navy Personnel Research and Development Center, San Diego, CA.

Stoloff, P. H. (1991). *Cost-effectiveness of U.S. Navy Video Teletraining System Alternatives,* CRM Research Memorandum 91-195, Center for Naval Analysis, Alexandria, VA.

Texas Higher Education Coordinating Board (1986). *Instructional Television: A Research Review and Status Report,* ERIC No. ED 276 422, Texas College and University System, Division of Universities and Research, Austin TX.

U.S. Office of Technology Assessment (1989). *Linking for Learning, A New Course for Education, ..* OTA-SET-4 30, U.S. Government Printing Office, Washington, D.C.

Van Haalen, T. and Miller, G. (1994). "Interactivity as a Predictor of Student Success in Satellite Learning Programs," *DEOSNEWS,* The Distance Education On-line Symposium, Vol. 4, No. 6, ACSDE@PSUVM.PSU.EDU.

Wagner, E. (1990). "Instructional Design and Development: Contingency Management for Distance Education," In *Contemporary Issues in American Distance Education,* Michael Moore (Ed.), Pergammon Press, p. 298-311.

Wetzel, C.D., Radke, P.H. and Stern, H.W. (1994). Instructional effectiveness of Video Media. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Weingand, D. E. (1984). "Telecommunications Delivery of Education: A Comparison with the Traditional Classroom," *Journal of Education for Library and Information Science,* Vol. 25, p. 3-12.

Zigerell, J. J. (1986). *A Guide to Telecourses and Their Uses.* ERIC No. ED 280 439, Coast Community College District, Cost Mesa, CA.

GENDER AND RACIAL EQUITY OF THE AIR FORCE OFFICER QUALIFYING TEST (AFOQT) IN OFFICER TRAINING SCHOOL SELECTION DECISIONS

Heather E. Roberts
Graduate Student
Department of Psychology

Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061-0436

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

GENDER AND RACIAL EQUITY OF THE AIR FORCE OFFICER QUALIFYING TEST (AFOQT) IN OFFICER TRAINING SCHOOL SELECTION DECISIONS

Heather E. Roberts
Graduate Student
Department of Psychology
Virginia Polytechnic Institute and State University

Abstract

The present study investigated the relationship between performance on the Air Force Officer Qualifying Test (AFOQT) and performance in Officer Training School (OTS) for race and gender. All composites were shown to be valid predictors of OTS performance for all subgroups. Minority subgroups had lower mean scores on the aptitude composites. Regressions of Final Course Grade and Officer Training Effectiveness Reports (OTER) on aptitude composites were compared for gender and racial subgroups to assess the predictive equity of three AFOQT composite scores (Academic Aptitude, Verbal, and Quantitative) for OTS. Results were consistent with the literature in education, industry, and prior studies conducted in the military. Predominant findings showed evidence of level bias in the prediction of Final Course Grades for both gender and racial subgroups. However, in all cases of level bias, minority subgroup performance was overpredicted, resulting in a higher selection rate for female and black cadets into OTS. Implications of the results and suggestions for future research are discussed.

GENDER AND RACIAL EQUITY OF THE AIR FORCE OFFICER QUALIFYING TEST (AFOQT) IN OFFICER TRAINING SCHOOL SELECTION DECISIONS

Heather E. Roberts

Introduction

Equal opportunity and fair employment practices continue to receive considerable attention in education, industry, and military organizations. Although standardized examinations are quite commonplace, their use in selection and placement decisions remains controversial. Specifically, racial and gender equity on standardized aptitude examinations is a topic that has been debated at great length by individuals involved in the development of tests and the selection of applicants for training or employment (Cascio, 1992; Daula, Smith, & Nord, 1990; Helms, 1992; Hunter, Schmidt, & Hunter, 1979; Jensen, 1980; Linn, 1978; Rothstein & McDaniel, 1991; Russell, 1994; Wunder & Sparks, 1991). The basic assumption of the classical model of selection is that scores on employment tests are linearly related to measures of job performance. When identifiable subgroups of the population (e.g., men, women, racial subgroups) are compared, differences in average scores on ability tests are typically found (Hartigan & Wigdor, 1989; Linn, 1982; Schmidt, 1988). The controversy revolves around whether test scores obtained by various subgroups are accurate indicators of their "true" score, or whether bias exists in the aptitude measure to discriminate against subgroup populations.

Cleary's (1968) psychometric model is the most widely used model in the evaluation of test fairness. Cleary distinguishes between test bias and test fairness in her definition: "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup" (Cleary, 1968, p. 115). This definition is currently accepted by both the Uniform Guidelines (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice, 1978) and the Society for Industrial and Organizational Psychology (SIOP, 1987), and is the definition usually relied upon in empirical studies of group differences in validity (Shepard, 1982). According to Wunder and Sparks (1991), "Fairness in personnel selection occurs when uniform application of standards, procedures, rules, and policies has the same result for each individual, without regard to classification by sex, race, ethnic group, national status, religion, age, handicap, or other legally protected status" (p. 114).

The problems of assessing the predictive validity of a test and assessing its fairness with regard to how the test will be used are directly related (Cleary, Humphreys, Kendrick, & Wesman, 1975). Specifically, if the inference drawn from a test score is made with the smallest feasible random error and there is no constant error as a result of subgroup membership, then the test can be considered fair for that particular use. Differential prediction involves the over- or underprediction of subgroup performance when a common regression equation is used. Underprediction for members of a subgroup produces a lower computed probability of success or lower

criterion performance at a given aptitude score compared with another subgroup. This serves to screen out subgroup candidates who would perform successfully if given the opportunity. Overprediction for members of a subgroup produces a higher computed probability of success or higher criterion performance at a given aptitude score compared with another subgroup. Overprediction of the performance of a protected group when a common regression line is used indicates bias in the measure, but is not evidence for test unfairness. Indeed, the literature indicates that test bias rarely occurs and, when it does, it tends to work to the favor of the minority subgroups by overpredicting their performance on criterion measures.

There is little convincing evidence that well-constructed tests, when properly administered and interpreted, are more valid predictors for the majority subgroup than for other subgroups in the population (Hunter, Schmidt, & Rauschenberger, 1975; Rothstein & McDaniel, 1991; Wigdor & Garner, 1982). A panel established by the National Academy of Sciences concluded that the cultural disadvantages among subgroups that lead to depressed test score performance also serve to lower job performance (National Academy of Sciences, 1982). It is important to recognize that subgroup mean differences are unrelated to test unfairness as currently defined. Although minority subgroups will tend to have lower average scores on ability measures, this is not evidence for the unfairness of a test. In fact, cumulative results from many studies of standardized ability tests indicate that they are fair to minority subgroup members under the Cleary model of test fairness (Schmidt & Hunter, 1980; Bartlett, Bobko, Mosier, & Hannon, 1978). Although the majority subgroup will most often have higher scores on the test and criterion measures, use of the test for selection will not be considered unfair if an examination of the two regression equations reveals that the minority intercept is lower than the majority intercept. This will result in an overprediction of minority test scores, i.e., indicating that the test was fair to minority subgroups, and the use of a common regression line will lead to more minority applicants being hired.

Research in the Military Sector

Increased attention has turned toward the use of standardized cognitive tests for military selection and classification. The rapid entry of women and black cadets into the military has left many relevant testing issues unaddressed. For both officers and enlistees, education and aptitude serve as the primary selection criteria. Officers are required to have a college education and take the aptitude measure designated by their military service (Brown, 1987). Unlike the Armed Services Vocational Aptitude Battery (ASVAB) used by all four military services for selection and classification of enlisted personnel, there is no common aptitude measure for military officers and relatively little documentation of the tests that are used by the various branches for this purpose (Brown, 1987; Cowan & Sperl, 1989).

Several studies have been conducted in the military setting to evaluate whether aptitude tests currently used in military branches to select and classify enlisted personnel demonstrate differential prediction for various subgroups. Early studies of predictive equity were limited by the small minority subgroup membership in most military occupations. The earliest use of regression comparisons to evaluate the fairness of a standardized test

for different subgroups was by Gordon (1953). She concluded that the same minimum test scores on the ASVAB could be used without bias for both black and white subgroups as a selection standard for Air Force technical training schools. The past 20 years has seen a surge of interest in the sensitivity and fairness of standardized tests, as well as an increase in the use of regression comparisons in addition to validity coefficients. Most studies have generally supported the fairness of the ASVAB for race (see Welsh, Kucinkas, & Curran, 1990, for a review; McLaughlin, Rossmeissl, Wise, Brandt, & Wang, 1984; Wilbourne, Valentine, & Ree, 1984; Wise, Welsh, Grafton, Foley, Earles, Sawin, & Divgi, 1992). For example, McLaughlin et al. (1984) examined the differences between racial subgroup specific and common regression lines in a large study of Army recruits, and found results indicating few or no differences among groups in the regions of the minimum aptitude qualifying scores. However, Wise et al. (1992) detected small but significant differences indicating a greater sensitivity in the ASVAB for whites than for blacks.

Studies of the predictive equity of military selection tests for male and female military personnel continue to be mixed. Wilbourne et al. (1984) did not find statistically significant differences in the mean validities of male and female airmen for final technical school grades. McLaughlin et al. (1984) reported small differences in validity patterns for males and females in Army enlisted samples. Additionally, they reported that female performance was consistently underpredicted in clerical or administrative jobs, but consistently overpredicted for male-dominated jobs. Wise et al. (1992) reported that the ASVAB technical composites were more sensitive predictors for females than for males. In a study of the sex equity of an Air Force pilot selection test, Sawin (1990) concluded that the Pilot and Navigator-Technical composites of the standardized officer test were equitable predictors of male and female performance in Undergraduate Pilot Training (UPT). In another study of the sex equity of pilot candidate measures, Siem and Sawin (1990) reported that there were overall differences on average test scores and ratings for males and females; however, those overall differences were not associated with inequity of prediction of UPT training outcomes. Carretta (1990) reported results that were also consistent with the above conclusions. He found that male and female pilot candidates commissioned through Reserve Officer Training Corps (ROTC) and Officer Training School (OTS) did not differ significantly in UPT performance when they were matched on level of pre-UPT performance composite.

The Present Study

During the 1980s, approximately 20,000 officers were commissioned annually in the Air Force, Army, Navy, and Marine Corps (Brown, 1987; U. S. Bureau of the Census, 1992). It is evident that aptitude tests are strongly relied upon as tools to identify those candidates who are likely to succeed as United States military officers. Nevertheless, an examination of the literature and a search in the 50-year bibliography of the selection and classification of United States military officers (Cowan & Sperl, 1989) confirmed there were only a few studies across all branches that investigated the fairness of officer aptitude tests for various subgroups, and these were concentrated in the Air Force (Carretta, 1990; Mathews, 1977; Sawin, 1990; Siem & Sawin, 1990). Most

studies have investigated the AFOQT composites for equitable selection of pilots. Only one study conducted more than 15 years ago has investigated the equity of the AFOQT composites for OTS and selection (Mathews, 1977). Mathews (1977) found that OTS performance of blacks was overpredicted by AFOQT composites. The present study attempted to readdress and extend that study by evaluating both the racial and gender equity of a more recent version of the Air Force's aptitude examination for selection into OTS.

During the 1980s, the Air Force commissioned approximately 6,500 new officers each year. Of these, approximately 30%, or 2,000, entered through the OTS precommissioning program (Air Force Association, 1989). The OTS is one of three sources or programs used by the Air Force to meet manning requirements for officer jobs. The other sources -- Air Force Academy and ROTC -- select and train officers in conjunction with their undergraduate education. The OTS, however, is specifically designed for candidates who have already obtained their undergraduate degree, but lack formal training in an officership curriculum.

The aptitude test that is used in the selection and classification of many officer candidates for the United States Air Force is the Air Force Officer Qualifying Test (AFOQT). The AFOQT is one of several selection criteria used for the Officer Training School and Air Force Reserve Officer Training Corps (AFROTC) commissioning program (Cowan, Barrett, & Wegner, 1989, 1990). Before a board reviews an applicant's record for admission into OTS, the record is prescreened to insure that the applicant meets or exceeds minimum qualifying scores. Eligible applicants for OTS are then evaluated by a selection board composed of senior officers. Each board member rates applicants subjectively, based on a "whole-person concept," after reviewing a variety of factors, including AFOQT scores, education (type of degree, grade point average, and coursework), employment and military experience, awards and achievements, letters of recommendation, and recruiter ratings of potential (Department of the Air Force, 1990). These evaluations produce an order-of-merit list of all applicants from which cadet selections are made.

Although the AFOQT has undergone two revisions since Mathews (1977) investigated the racial equity of the test, there has not been a recent investigation evaluating the fairness of the AFOQT for OTS performance on the basis of gender and racial subgroup membership. Thus, the purpose of the present study was to examine the predictive equity of the AFOQT for selection into Air Force Officer Training School for both gender and racial subgroups.

Method

Sample

The sample was 13,559 (12,166 males and 1,393 females) Air Force officer cadets who entered OTS between 1982 and 1988 and tested on the AFOQT Form O, and for whom an OTS training record existed. The data consisted of a restricted sample of cadets who had undergone a 2-stage officer selection process. In the first stage, their test scores were reviewed to insure they met or exceeded the minimum qualifying scores on the Verbal and Quantitative composites of the AFOQT. Those who met the first stage requirements entered the

second stage in which their applications for officer pre-commissioning training were reviewed by an OTS selection board using a "whole person" selection approach. Approximately four percent (n =511) of the sample were black cadets. Average age for all cadets was 24 years.

Procedure

All of the data for this study came from archival data bases maintained by the Air Force Armstrong Laboratory.

Predictor Variables. The AFOQT, race, and gender were used as predictors in the present study. The AFOQT is a paper-and-pencil aptitude test battery used to select civilian or prior-service applicants for officer pre-commissioning programs and to classify commissioned officers into aircrew job specialties (pilot vs. navigator training). Form O consists of 16 subtests that assess five ability domains: verbal, quantitative, academic aptitude, pilot, and navigator-technical (Rogers, Roach, & Wegner, 1986). The subscore composites used to select candidates for officer precommissioning training were verbal, quantitative, and academic ability measured in the percentile metric. The AFOQT Verbal composite includes subtests for vocabulary, English usage, and verbal analogies. The AFOQT Quantitative composite includes subtests for mathematical reasoning, and ability to understand graphs and tables. The AFOQT Academic Aptitude composite, previously called the Officer Quality composite, is obtained by combining the Verbal and Quantitative composite scores. Rogers, Roach, and Wegner (1986), using a formula developed by Wherry and Gaylord (1943), reported reliability coefficients for the Academic Aptitude, Verbal, and Quantitative composites of .96, .94, and .92, respectively.

OTS Performance Criteria. Officer Training School criteria available for assessment were a Final Course Grade for OTS graduates and an Officer Training Effectiveness Report (OTER). Final Course Grade was a numerical score from 75 to 99, averaged from OTS coursework scores (i.e., five written examinations administered during training) (Cowan et al., 1990). OTER data resulted from a performance review, conducted by instructors, of the cadets' overall accomplishments. The review was conducted at the eleventh week of training. Instructors used a four-point scale to rate a cadet's performance ranging from unsatisfactory to outstanding. The basic assumption of any study of test bias is that the criterion which the test is designed to predict is unbiased.

Analytic Strategy

Means and standard deviations were computed separately for the three aptitude composite predictors (Academic Aptitude, Verbal, Quantitative) of the AFOQT and the criterion measures of performance (Final Course Grade, OTER) for gender and racial subgroups.

Uncorrected zero-order correlations between composite predictor variables and performance criteria were computed by subgroup and tested for significance. However, personnel screening procedures that produce a restricted sample of eligible candidates will reduce the variance in the factors of consideration and operate to depress the magnitude of obtained validity coefficients (Gulliksen, 1950). Subjects for the current study

represented a restricted sample of OTS cadets due to selection on the AFOQT; therefore, corrected correlation coefficients are reported as well.

Subgroup validities provide important information about the utility of the test for predicting performance of various subgroups; however, they should not be relied upon as the sole evidence for determining whether the test discriminates against minority subgroups. The standard errors of estimate (SEE) were also compared using the F-statistic (Reynolds, 1982) to determine if the accuracy of prediction was different for gender and racial subgroups. Further, the focal test to evaluate the predictive equity of the AFOQT composites for the performance criteria was the Cleary regression model. Despite the proliferation of models that are available to investigate the fairness of tests, the Cleary model of selection fairness is still more widely accepted and practiced than any other (Schmidt & Hunter, 1982; Shepard, 1982). Specifically, to test for bias effects, the general linear model (GLM) approach (Pedhazaur, 1982) was implemented to test for both slope and level bias. Slope bias is evident when the difference in the predicted performance scores for subgroup members varies at different levels of predictor scores. Level bias exists when different subgroups have parallel regression lines (same slopes) but the intercepts are different.

Each of the three AFOQT composite scores of interest (Academic Aptitude, Verbal, Quantitative) was separately tested for each criterion variable to investigate the predictive equity of the composites. The testing of linear models involves comparing a "full model" to a "restricted model," which contains a subset of the variables from the full model, to evaluate whether there is a loss in predictive efficiency. Analysis of the extent of loss in predictive efficiency is determined with an F-statistic (Ward & Jennings, 1973).

Table 1 defines the linear models that were tested in the present study. The starting model (Model A) for each analysis contained three predictor variables: (1) subgroup membership using binary coding; i.e., 1=male, 0=female, (2) the AFOQT composite score for the male subgroup, and (3) the AFOQT composite score for the female subgroup. First, each composite was tested for slope bias for males and females using the F statistic (Model A vs. Model B). When evidence of slope bias (Model A) was found, the analysis sequence was terminated. If no slope bias was present, each composite was tested for the over- or underprediction of performance, or level bias (Model B), relative to a common regression line (Model C). The magnitude and direction of subgroup performance differences, if any, were then examined using predicted scores. Identical analyses were conducted for the black and white subgroup samples.

Results

Means and standard deviations (SD) for the predictor and criterion variables are shown in Table 2. Females scored approximately one-third of one SD higher than males on the Verbal composite, and scored approximately one-third of one SD lower than males on the Quantitative composite, resulting in roughly equivalent average scores on the Academic Aptitude composite. Black OTS entrants averaged slightly more than one-half of one SD lower than whites on Academic Aptitude and Quantitative composites and just under one-half

of one SD lower than whites on the Verbal composite. The subgroups achieved comparable performance levels in OTS; the means and standard deviations of both the Final Course Grade and OTER criteria were similar for males and females and for blacks and whites.

Table 3 presents the uncorrected and corrected zero-order correlations between predictor and criterion variables in the present study. All composites were valid predictors of grades and ratings for all subgroups. OTS Final Course Grade correlated from .20 to .39 with the aptitude measures. The r's for all three test composites were comparable in magnitude for male and female cadets and for black and white cadets with one exception; the Quantitative composite coefficient was .05 higher for black cadets than for white cadets. Compared to those for the Final Course Grade criterion, the coefficients were considerably smaller for OTER appraisals for all subgroups, ranging from .04 to .16, and the size of the validity differences between subgroups was generally larger. Black subgroup coefficients were substantially higher (.08 or more) than white subgroup coefficients on the Academic Aptitude and Quantitative composites.

When the correlations were corrected for the effects of range restriction, all correlation coefficients increased in magnitude. Generally, the increase was at least .15 for each composite for male, female, and white subgroups in predicting Final Course Grade. The increase in validity coefficients was slightly lower for the black subgroup, about .07 to .10 across composites. The increase in prediction of OTER performance was somewhat lower across all subgroups, generally increasing .07 for male, white, and black subgroups, and increasing .09 to .13 for females.

The standard errors of estimate (SEE) which reflect the accuracy associated with the subgroup-specific correlations are shown in Table 5. About 68 percent of the members of the subgroup would be expected to obtain performance levels which fall within plus or minus one SEE of the value predicted from the regression of the criterion on the test composite. The SEEs in predicting Final Course Grade and OTER evaluations from the aptitude composites were not significantly different for male and female subgroups. Similar results were obtained on the OTER criterion for blacks and whites. The exception to this favorable pattern concerned prediction of the Final Course Grade measure for the racial subgroups, where the test composites were less accurate for blacks than for whites. Although the observed SEE differences were small (less than one-half of one grade point), they were statistically significant at the .001 level for the relatively large black and white subgroup sample sizes employed in the study.[1]

The first set of linear model comparisons was performed for the male and female cadet subgroups to investigate whether there was slope or level bias for the Academic Aptitude predictor. A summary of the results

[1] The use of the F-statistic in comparing linear models to test hypotheses associated with the Cleary model of test bias assumes homoscedasticity; that is, that the conditional variances of the criterion for each level of the predictor are equal for both subgroups. Differences in SEEs suggest non-homogeneity, a factor which may distort results of slope and intercept tests. In the present study, significant differences were obtained for only 1 of the 4 conditions analyzed. Further, the sampling distribution of F is known to be robust to other than gross violations.

is presented in Table 5. The starting model (Model A) was significantly predictive of final grade average (R-square = .131, F[3, 13555] = 678.55, p < .001). Elimination of the sex-by-composite interaction terms (Model B) did not significantly decrease predictive effectiveness (R-square change = .000, F [1, 13555] = 2.08, NS). Thus, no evidence of slope bias was detected for the Academic Aptitude composite. Elimination of the sex membership term (Model C) did significantly decrease the model multiple correlation (R-square change = .003, F [1, 13556] = 38.86, p < .001). Therefore, there was evidence of level bias for Academic Aptitude scores in predicting Final Course Grade. The model comparisons for the Verbal and Quantitative composites followed a similar pattern for Final Course Grades, indicating there was no evidence of slope bias but there was evidence of level bias (See Table 5).

Neither slope nor level bias was detected in the Academic Aptitude measure for predicting OTER evaluations. However, there was modest evidence of differences in intercepts of the Verbal composite (p < .05) and evidence of slope bias was indicated by the Quantitative composite (p < .05). In each case where level bias was present, minority subgroup performance was over-predicted. This means that minority subgroup performance was less than would be expected from their test scores.

Identical regression analyses were conducted for the black and white cadet subgroups to determine whether there was slope or level bias for the Academic Aptitude predictor on Final Course Grade (See Table 6). Model A was again significantly predictive of final grade average (R-square = .13, F[3, 12960] = 635.23, p < .001). Elimination of the race-by-composite interaction terms (Model B) did not significantly decrease predictive accuracy (R-square change = .00, F [1, 12960] = .66, NS), indicating there was no slope bias for the Academic Aptitude composite variable. Eliminating the race term from the model (Model C) significantly decreased the model multiple correlation (R-square change = .00, F [1, 12961] = 7.38, p < .01), providing evidence of level bias. The models for the Verbal and Quantitative composites followed a similar pattern for OTS Final Course Grades, again showing no indication of slope bias but affirmative evidence for level bias at the .001 significance level.

No evidence for slope or level bias was present in either the Academic Aptitude or Verbal composite measures for predicting OTERs. However, similar to the male and female analyses, the Quantitative composite showed evidence of slope bias.

Table 7 gives the magnitude and direction of level bias for the subgroup samples. The predicted performance scores indicate the amount of overprediction or underprediction relative to a common regression line of the criterion for members of each subgroup with equivalent aptitudes. Predicted performance scores from total group composite means are shown to demonstrate differences in expected performance from a common regression line. It can be noted from Table 7 that there was an overprediction of criterion performance for those composites that showed evidence of level bias for the black subgroup and the female subgroup. For both gender and racial effects, predicted criterion scores using the common regression line were much closer to the

predicted criterion scores of the majority group. For instance, male Final Course Grades were underpredicted .05 points using the common regression line for Academic Aptitude, while female Final Course Grades were overpredicted .52 points using the common regression line. Similarly, white Final Course Grades were underpredicted only .02 points using the common regression line for Academic Aptitude, while black Final Course Grades were overpredicted .38 points. This pattern was consistent for all composites and criterion variables. Figures 1 and 2 present an illustration of level bias for gender and race for the regression of Final Course Grade on the Academic Aptitude composite. Figures 3 and 4 illustrate the slope bias for gender and race for the regression of OTER on the Quantitative composite.

Discussion

Comprehensive reviews of the literature have provided strong evidence that black subgroups tend to have lower scores on standardized tests of cognitive ability. The literature generally seems to find that in cases where there is evidence of test bias, it is usually in the form of level differences with the overprediction of minority subgroup performance. Results of the present study were consistent with prior studies of standardized tests in education and industry (Cleary, 1968; Feild, Bayley, & Bayley, 1977; National Research Council, 1989). More importantly, the results were consistent with a similar study of an earlier version of the AFOQT (Mathews, 1977). From data collected during the 1970s, Mathews (1977) concluded that OTS performance of blacks was overpredicted by AFOQT composites. Fifteen years later we are drawing the same conclusions with data collected during the 1980s on a revised test for officer selection. Test discrimination against blacks or females is not supported by the data presented in the present study since minority subgroup performance in OTS was again overpredicted. As newer versions of the test are developed, researchers should continue to document the equity of the officer selection measures for gender and racial subgroups. Future studies should be aided by an increase in minority subgroup representation for officer candidacy.

The Air Force is currently using a common regression line for selection practices. Although the present data show that some error has occurred by using a common line, the actual intercept differences between subgroups are negligible. In other words, although the presence of level differences indicates that the test is biased, the practical implications of the magnitude of the differences should be addressed. In this study, sample sizes were relatively large so that that even statistically significant findings should be evaluated in the larger context of practical significance. This is due to the fact that as sample sizes increase, smaller and smaller real differences are likely to be judged statistically significant. The gender differences found were small, ranging from a low of one-third of a criterion point to a high of one criterion point difference for Final Course Grade, and less than one-twentieth of one criterion point in predicting OTER outcomes. Racial differences were also of small magnitude, with subgroup differences of only one-tenth of one standard deviation (approximately one-half of one criterion point) for Final Course Grade, and even smaller differences for the prediction of OTERs (less than one-twentieth of one criterion point).

Overall, the results do not indicate that the use of the AFOQT in a consistent manner by the selection board will result in unfair selection practices. In contrast, both female and black subgroups will have higher expected performance scores than would be predicted using separate regression lines. While statistically significant differences were obtained, the practical differences in the ability to predict performance outcomes in OTS are not sufficient to warrant recommended changes in the composites at this time. Nevertheless, there are limitations of the current study and recommendations for future research that should be addressed. First, the current study addressed the effects of range restriction by correcting to a random sample of 3,000 applicants; in future research more accurate estimates of true validities could be obtained by correcting to the individual subgroups instead. In addition, the present study only utilized final grades and subjective instructor evaluations in pre-commissioning training. Similar studies should be conducted that utilize later indices of officer performance, such as grades in technical schools and objective measures of job performance. Future research could also investigate the equity of the AFOQT composites for groups other than OTS candidates. For instance, a study of the AFOQT composites for predicting training performance outcomes among ROTC cadets is needed. In addition, a study is needed to investigate the equity of the AFOQT for minority subgroups applying for navigator positions as the number of minority candidates in Undergraduate Navigator Training (UNT) increases.

A final recommendation for future research is the consideration of item-level analyses. There are several techniques one may employ to investigate the AFOQT at the item level. First, test developers should evaluate candidate AFOQT test items prior to operational use using modern test theory (sometimes referred to as latent trait theory or item response theory) to investigate the parameters of the individual items. A comparison of the item characteristic curves would enable the researcher to evaluate the difficulty, discrimination, and guessing parameters of the item to determine if those parameters differ between subgroups. In addition, after criterion data become available, one may want to conduct Cleary analyses at the item level. Specifically, Cleary analyses, similar to those performed at the test level in the current study, could be conducted for every individual item to determine whether the item is biased. The inclusion of an assessment of differential item functioning would contribute substantively to the AFOQT literature and would expose individual items that may be contributing to differences in the test scores of minority and majority subgroups.

In conclusion, the present study found that the use of a common regression line in a consistent manner by individuals involved in officer selection will <u>not</u> result in selection practices that discriminate unfairly against female or black cadets. In contrast, the use of AFOQT test scores will result in an expectation of higher performance than is actually realized. This is congruous with many studies that have been concerned with the equity of standardized tests for racial and gender subgroups in both civilian and military environments. At this time, the results do not suggest that there is a need for immediate changes in the AFOQT composites used for officer selection. However, future research should always consider new measures that reduce adverse impact and remediate overall differences in aptitude measures among subgroups.

References

Air Force Association. (1989, May). An Air Force almanac. Air Force Magazine, pp.45-60.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannon, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. Personnel Psychology, 31, 233-241.

Brown, D. C. (1987). Military officers: Commissioning sources and selection criteria (HumRRO FR- PRD-87-42). Alexandria, VA: Human Resources Research Organization.

Carretta, T. R. (1990, April). Gender differences in USAF pilot training performance. Paper presented at the 12th Psychology in the Department of Defense Symposium, Colorado Springs, CO.

Cascio, W. F. (1992). Applied psychology in personnel management (third edition). Englewood Cliffs, NJ: Prentice-Hall.

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 5, 114-124.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.

Cowan, D. K., Barrett, L. E., & Wegner, T. G. (1990). Air Force Reserve Officer Training Corps selection system validation (AFHRL-TR-88-54). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cowan, D. K., Barrett, L. E., & Wegner, T. G. (1990). Air Force Officer Training School selection system validation (AFHRL-TR-89-65). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cowan, D. K., & Sperl, T. C. (1988). Selection and classification of United States military officers: A fifty-year bibliography (1937-1986) (AFHRL-TP-88-45). Brooks AFB, TX: Air Force Human Resources Laboratory.

Daula, T., Smith, D. A., & Nord, R. (1990). Inequality in the military: Fact or fiction? American Sociological Review, 55, 714-718.

Department of the Air Force. (1990, May 31). Officer Training School (OTS) precommissioning program, USAF (AF Regulation 53-27). Washington, DC: Headquarters US Air Force.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice. (1978). Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures. Federal Register, 43, 38290-38315.

Feild, H., Bayley, G. A., & Bayley, S. (1977). Employment test validation for minority and nonminority production workers. Personnel psychology, 30, 37-46.

Gordon, M. A. (1953). A study in the applicability of the same minimum qualifying scores for technical schools to white males, WAF and negro males (AFHRL-TR-53-54). Lackland AFB, TX: Human Resources Research Center.

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley & Sons, Inc.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery. Washington, DC: National Academy Press.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? American Psychologist, 47, 1083-1101.

Hunter , J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 86, 721-735.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Linn, R. (1978). Single group validity, differential validity, and differential prediction. Journal of Applied Psychology, 63, 507-517.

Linn, R. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), Ability testing: Uses, consequences, and controversies. Washington, DC: National Academy Press.

Mathews, J. J. (1977). Racial equity in selection in Air Force officer training school and undergraduate flying training (AFHRL-TR-77-22). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., & Wang, M. (1984). Validation of current alternative Armed Services Vocational Aptitude Battery (ASVAB) area composites, based on training and Skill Qualification Test (SQT) information in fiscal year 1981 and 1982 (ARI-TR-651, AD-A156 807). Alexandria, VA: Army Research Institute.

National Academy of Sciences (1982). Ability testing: Uses, consequences, and controversies (vol. 1). Washington, DC: National Academy Press.

National Research Council. (1989). Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery. Washington, DC: National Academy Press.

Pedhazur, E. J. (1982). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Reynolds, C. E. (1982). Methods for detecting construct and predictive bias. In R. A. Bert (Ed.), Handbook of Methods for Detecting Task Bias (pp. 199-227). Baltimore, MD: Johns Hopkins University Press.

Rogers, D. L., Roach, B. W., & Wegner, T. G. (1986). Air Force Officer Qualifying Test Form O: Development and standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Rothstein, H. R., & McDaniel, M. A. (1991). Mental abilities testing in industry. In J. W. Jones, B. D. Steffy, & D. W. Bray (Eds.), Applying Psychology in Business (pp. 236-242). Lexington, MA: Lexington Books.

Russell, T. L. (1994). Models of fairness. In J. P. Campbell (Ed.) & T. L. Russell, Building a joint-service classification research roadmap: Methodological issues in selection and classification (AL/HR-TR-1994-0013). Alexandria, VA: Human Resources Research Organization.

Sawin, L. L. (1990, August). The sex equity of an Air Force pilot selection test. Paper presented at the annual convention of the American Psychological Association, Boston, MA.

Schmidt, F. L., & Hunter, J. E. (1980). The future of criterion-related validity. Personnel Psychology, 33, 41-60.

Schmidt, F. L., & Hunter, J. E. (1982). Two pitfalls in assessing fairness of selection tests using the regression model. Personnel Psychology, 35, 601-607.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), Handbook of Methods for Detecting Test Bias (pp. 9-30). Baltimore, MD: The Johns Hopkins University Press.

Siem, F. M., & Sawin, L. L. (1990). Comparison of male and female USAF pilot candidates. Proceedings of the AGARD Symposium on Recruiting, Selection, Training, and Military Operations of Female Aircrew, Tours, France.

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

U. S. Bureau of the Census (1992). Statistical abstract of the United States: 1992 (112th ed.). Washington, DC: Author.

Ward, J. H., & Jennings, E. (1973). Introduction to linear models. Englewood Cliffs, NJ: Prentice- Hall.

Welsh, J. R., Kucinkas, S. K., & Curran, L. T. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Wherry, R. J., & Gaylord, R. H. (1943). The concept of test and item reliability in relation to factor pattern. Psychometrika, 8, 247-264.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). Ability testing: Uses, consequences, and controversies. Part 1. Report of the Committee. Washington, DC: National Academy Press.

Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D. R. (1992). Sensitivity and fairness of the Armed Services Vocational Aptitude battery (ASVAB) technical composites (DMDC Technical Report 92-002). Monterey, CA: Personnel Testing Division, Defense Manpower Data Center.

Wilbourne, J. M., Valentine, L. D., Jr., & Ree, M. J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, and 10 to Air Force technical school final grades (AFHRL-TR-84-8, AD-A144 213). Brooks AFB, TX: Manpower and Personnel Division, Air force Human Resources Laboratory.

Wunder, S., & Sparks, C. P. (1991). Fairness in personnel selection. In J. W. Jones, B. D. Steffy, & D. W. Bray (Eds.), Applying Psychology in Business (pp. 114-114-120). Lexington, MA: Lexington Books.

Table 1. Linear Models for Gender and Racial Comparisons

Sex Models

Model A (Starting Model - Slope Bias)
Predicted Y = Unit vector + Male vector + Female Test vector + Male Test vector

Model B (Level Bias)
Predicted Y = Unit vector + Test vector + Male vector

Model C (Homogeneity of Regression)
Predicted Y = Unit vector + Test vector

Race Models

Model A (Starting Model - Slope Bias)
Predicted Y = Unit vector + White vector + Black Test vector + White Test vector

Model B (Level bias)
Predicted Y = Unit vector + Test vector + White vector

Model C (Homogeneity of Regression)
Predicted Y = Unit vector + Test vector

Note: Test vector represented one of three AFOQT composites, Academic Aptitude, Verbal, or Quantitative. Male vector and White vector were binary categorical variables.

Table 2. Means and Standard Deviations for Percentile Score Predictor and Criterion Variables

Predictor Variables	Male (n=12,166)	Female (n=1,393)	White (n=12,453)	Black (n=511)
AFOQT - AA				
Mean	66.47	66.07	67.30	53.14
SD	21.71	20.48	21.19	24.14
AFOQT - V				
Mean	67.37	74.97	68.89	58.72
SD	22.47	20.70	22.08	24.93
AFOQT - Q				
Mean	62.82	53.71	62.69	47.67
SD	22.99	23.16	22.86	25.17
Criterion Variables	**Male (n=12,166)**	**Female (n=1,393)**	**White (n=12,453)**	**Black (n=511)**
Final Grade				
Mean	91.93	91.33	91.93	90.72
SD	3.47	3.56	3.45	3.95
OTER				
Mean	3.81	3.77	3.81	3.70
SD	1.45	1.46	1.45	1.42

Table 3. Uncorrected (and Corrected) Zero-Order Correlations of AFOQT Composite Variables with Performance Criteria

	Male (n=12,166)		Female (n=1,393)	
	Final Grade	OTER	Final Grade	OTER
AFOQT - AA	.36*** (.50)	.09*** (.16)	.37*** (.54)	.09*** (.21)
AFOQT - V	.39*** (.53)	.11*** (.18)	.39*** (.55)	.06** (.19)
AFOQT - Q	.21*** (.37)	.04*** (.11)	.22*** (.41)	.10*** (.19)

	White (n=12,453)		Black (n=511)	
	Final Grade	OTER	Final Grade	OTER
AFOQT - AA	.35*** (.50)	.08*** (.16)	.38*** (.45)	.16*** (.22)
AFOQT - V	.38*** (.52)	.11*** (.17)	.38*** (.46)	.13** (.19)
AFOQT - Q	.20*** (.37)	.04*** (.11)	.25*** (.35)	.14*** (.21)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 4 Standard Error of Estimates (SEE) of AFOQT Composite Variables with Performance

AFOQT Composite	Final Grade			OTER		
	Male SEE	Female SEE	Diff.in SEE	Male SEE	Female SEE	Diff.in SEE
AFOQT - AA	3.24	3.32	.08	1.44	1.45	.01
AFOQT - V	3.19	3.28	.10	1.44	1.45	.02
AFOQT - Q	3.39	3.48	.09	1.45	1.45	.00

AFOQT Composite	Final Grade			OTER		
	White SEE	Black SEE	Diff.in SEE	White SEE	Black SEE	Diff.in SEE
AFOQT - AA	3.24	3.66	.43 ***	1.44	1.41	.04
AFOQT - V	3.66	3.19	.47 ***	1.41	1.44	.03
AFOQT - Q	3.83	3.38	.45 ***	1.41	1.45	.04

Note: N = 12,453 for White sample; N = 511 for Black sample
N = 12,166 for Male sample; N = 1,393 for Female sample
* $p < .05$
** $p < .01$
*** $p < .001$

Table 5. Regression Analyses Results for OTS Male and Female Cadets

Criterion	AFOQT Predictor	Multiple Correlation Squared (R-sq) Model A	Model B	Model C	Tests for Homogeneity Slope Bias F(A vs B)	Level Bias F(B vs C)
Final Course Grade	AFOQT - AA	.131	.130	.128	2.08	38.86 ***
	AFOQT - V	.157	.157	.148	1.90	136.43 ***
	AFOQT - Q	.046	.046	.046	0.17	10.04 **
OTER	AFOQT - AA	.008	.008	.008	0.14	0.82
	AFOQT - V	.012	.012	.012	2.66	5.22 *
	AFOQT - Q	.002	.002	.002	5.05 *	

* $p < .05$
** $p < .01$
*** $p < .001$

Table 6. Regression Analyses Results for OTS White and Black Cadets

Criterion	AFOQT Predictor	Multiple Correlation Squared (R-sq) Model A	Model B	Model C	Tests for Homogeneity Slope Bias F(A vs B)	Level Bias F(B vs C)
Final Course Grade	AFOQT - AA	.128	.128	.128	0.66	7.38 **
	AFOQT - V	.150	.150	.149	0.03	17.36 ***
	AFOQT - Q	.047	.047	.045	2.22	23.27 ***
OTER	AFOQT - AA	.008	.008	.008	1.83	.24
	AFOQT - V	.012	.012	.012	0.03	.47
	AFOQT - Q	.002	.002	.002	4.98 *	

* $p < .05$
** $p < .01$
*** $p < .001$

Table 7. Magnitude and Direction of Level Bias - Predicted Performance Scores for Subgroups

		Gender Effects			Racial Effects		
		Prediction for Males (n=12,166)	Prediction with Common Regression Line	Prediction for Females (n=1,393)	Prediction for Whites (n=12,453)	Prediction with Common Regression Line	Prediction for Blacks (n=511)
Final Course Grade	AFOQT - AA	91.92	91.87	91.35	91.90	91.88	91.50
	AFOQT - V	91.97	91.87	90.91	91.91	91.88	91.30
	AFOQT - Q	91.90	91.87	91.59	91.91	91.88	91.17
OTER	AFOQT - AA	3.81	3.80	3.77	3.81	3.81	3.78
	AFOQT - V	3.81	3.80	3.72	3.81	3.81	3.77

Note. Mean AFOQTcomposite scores for the total subsamples were used to obtain mean predicted criterion scores. Combined male and female Academic Aptitude, Verbal, and Quantitative composite mean scores were 66.43, 68.15, and 61.89, respectively. Combined black and white Academic Aptitude, Verbal, and Quantitative composite mean scores were 66.74, 68.49, and 62.10, respectively.

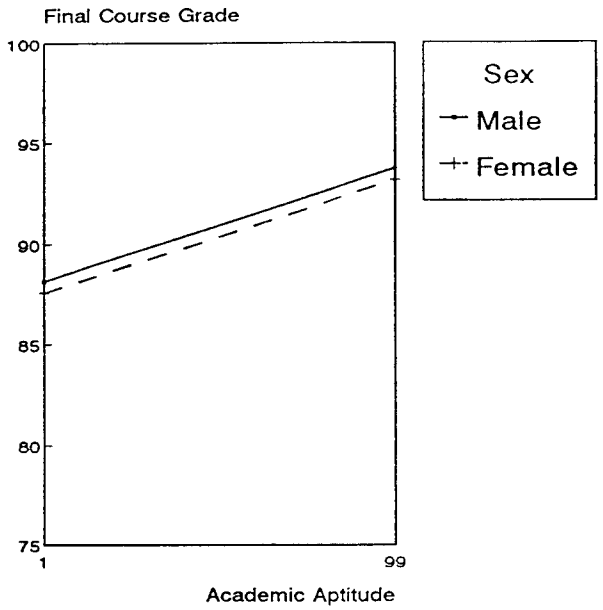


Figure 1. Level Bias for Academic Aptitude Composite and Final Course Grade
(Y' = 87.52 + 0.572(Sex)+ 0.06(Test)

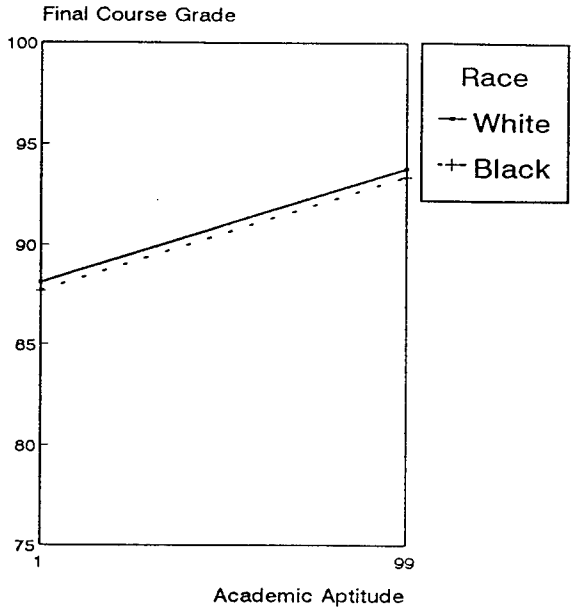


Figure 2. Level Bias for Academic Aptitude Composite and Final Course Grade
(Y'= 87.66 + 0.40(Race) + 0.06(Test)

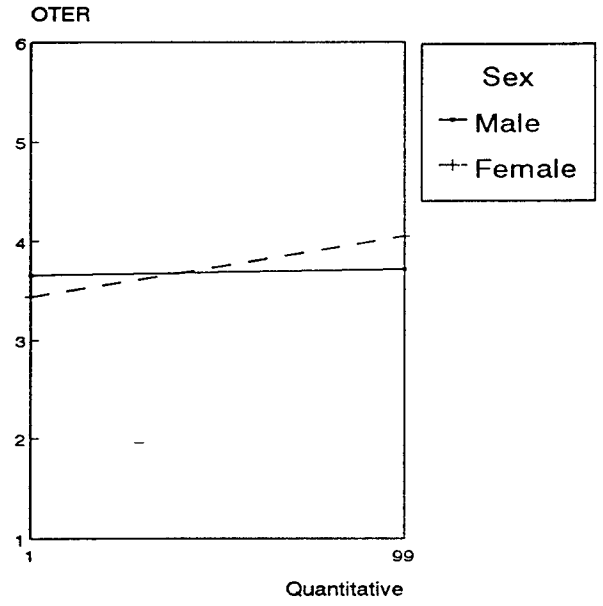


Figure 3. Slope Bias for Quantitative Composite and OTER
(Y' = 3.43 + 0.232(Sex) + 0.001(MLTest) + 0.006(FMTest)

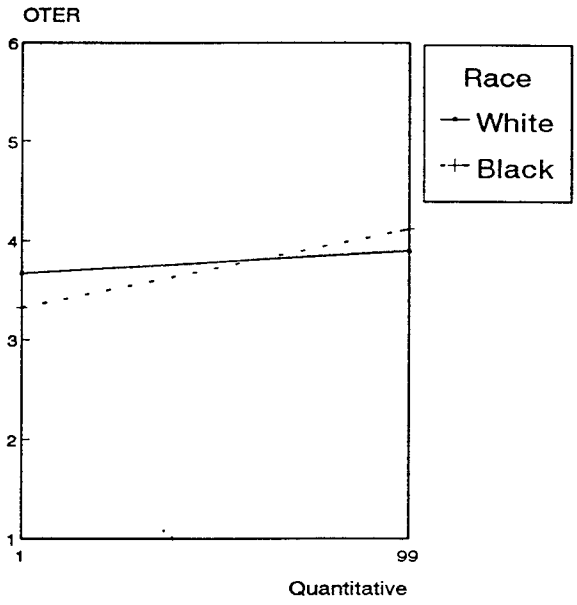


Figure 4. Slope Bias for Quantitative Composite and OTER
(Y' = 3.31 + 0.360(Race) + 0.000(WTTest) + 0.008(BKTest)

IMPROVED NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray
Associate Professor

Dale F. Rucker
Graduate Research Assistant

Department of Civil and Environmental Engineering
West Virginia University
Morgantown, WV 26506-6103

Final report for:
Summer Faculty Research Program
Armstrong Laboratory

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

September 1994

IMPROVED NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray
Associate Professor

Dale F. Rucker
Graduate Research Assistant

Department of Civil and Environmental Engineering
West Virginia University

Abstract

Public domain computer programs were used to attempt an improved model of the tritium plume observed during Macrodispersion Experiment 2 (MADE-2), a field scale natural gradient experiment conducted at Columbus Air Force Base, Mississippi. The program Geo-EAS used head and hydraulic conductivity data at a relatively small number of irregularly spaced test locations to estimate corresponding values at the more numerous nodes of a computational grid having 66 rows, 21 columns, and 9 layers. The finite difference program MODFLOW was used to simulate the flow of groundwater through a 330 m x 105 m computational domain. The recent BCF2 subroutine package, which permits rewetting of cells, allowed the vertical discretization to be more accurate than in previous studies. Solutions for the 468 day experiment were obtained using a Sun Sparcstation 2 for several choices of convergence and storage parameters. The simulations had small mass balance errors and were consistent with continuous head observations. The smallest storage coefficients gave the best agreement. One persistent feature of the predicted head field was a tendency for the head to decline toward the northwest. This suggests that the plume should bend toward the northwest, but the observations show a bend toward the northeast. This discrepancy is probably due to inaccurate head boundary conditions resulting from a lack of piezometers in the northern part of the computational domain. The flow model is about as accurate as the data permit.

Tritium plume simulations used the mixed Lagrangian-Eulerian finite difference program MT3D to solve the contaminant transport equation using the MODFLOW-predicted flow field. Thirteen runs were made using various advection algorithms and dispersivities, but none was successful. Numerical instabilities or grossly unrealistic predictions ended every run by simulation day 141. Further work is needed to obtain a satisfactory plume prediction.

IMPROVED NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray

Dale F. Rucker

INTRODUCTION

Faced with the need to remediate groundwater pollution at many of its bases, the Air Force has undertaken an extensive program of research on subsurface contaminant transport. The Macrodispersion Experiment 2 (MADE-2), conducted together with the Electric Power Research Institute and the Tennessee Valley Authority, was a key element of this effort. MADE-2 was a field-scale natural gradient experiment performed in 1990-91 at Columbus Air Force Base in Columbus, Mississippi. A MADE-2 database has been prepared by Boggs and others (1993a) and analyses have been published by Boggs and others (1993b) and by Stauffer and others (1994).

The MADE-2 test site was an area about 300 m x 200 m with about 2 m of relief. It was covered primarily by weeds and brush, and contained no streams or ponds. The 10 m to 15 m thick upper layer of soil was a shallow alluvial terrace containing an unconfined aquifer. This was bounded below by an aquitard of marine silt and clay (Boggs, Young, Benton, and Chung; 1990). The aquifer soil was classified as poorly sorted to well sorted sandy gravel and gravelly sand with minor amounts of silt and clay. The aquifer was found to consist of irregular lenses and layers having typical horizontal dimensions on the order of 8 m and typical vertical dimensions on the order of 1 m.

The heterogeneity of the MADE-2 site was much greater than that of other reported natural gradient macrodispersion experiments. Measurements using the borehole flowmeter method showed hydraulic conductivity variations of up to four orders of magnitude in individual profiles. Rehfelt, Boggs, and Gelhar (1992) found that the variance of the natural logarithm of the hydraulic conductivity was at least an order of magnitude larger at Columbus than at Borden, Twin Lakes, or Cape Cod. The horizontal and vertical correlation scales for hydraulic conductivity were also larger by factors of 1.75 or more.

MADE-2 focused on the fate and transport of dissolved organic chemicals of the types found in jet fuels and solvents. A volume of 9.7 m^3 of tracer solution was injected at a constant rate for 48.5 hours through 5 wells spaced 1 m apart. The

solution contained tritiated water (an essentially passive tracer), benzene, naphthalene, p-xylene, and o-dichlorobenzene. The spread of the plume in three dimensions was monitored for 15 months by analyzing water samples drawn from up to 328 multilevel sampling wells (at up to 30 depths per well) and 56 BarCad positive displacement samplers. Five comprehensive sets of water samples (called snapshots) were obtained at intervals of about 100 days. Plots of concentration contours in horizontal planes showed that the tritium plume spread in an essentially linear fashion with a tendency to bend toward the northeast. The vertical structure along the plume axis was complex.

Boggs and others (1993b), based on numerical integration of the tritium concentrations, found ratios of observed mass to injected mass in the first four snapshots of 1.52, 1.05, 0.98, and 0.77, respectively. The 52% overestimate in the initial snapshot was attributed to preferential sampling from more permeable zones and to vertical interconnections between sampling points. The 23% underestimate in snapshot 4 was partially due to the motion of the plume's leading edge past the farthest downstream sampling points. Snaphot 5 was not intended to define the entire plume.

Our objective in the 1994 Summer Research Program was to obtain improved simulations of the MADE-2 tritium plume using public domain computer codes for groundwater flow and contaminant transport. The present work is an extension of the senior author's previous efforts as an AFOSR Summer Faculty Fellow (Gray, 1992; 1993).

FLOW MODELING

In accord with most groundwater studies, in the present work the effects of density variations are assumed to be negligible, so that the flow equation can be solved without knowing the concentration field. The resulting velocity field is input to the transport equation, which is then solved for the concentrations. These calculations were performed using computer programs MODFLOW for the flow problem and MT3D for the transport problem. Many other programs were used to prepare input files or to analyze results. Unless noted otherwise, these were written by the authors of this report in FORTRAN 77.

MODFLOW (McDonald and Harbaugh, 1988) is a U. S. Geological Survey (USGS) public domain FORTRAN 77 program for the solution of the groundwater flow equation. The program's name refers to its modular structure which facilitates the insertion of new subroutine packages to handle specific tasks. The version used here, MODFLOW/mt, was obtained from Dr. Chunmiao Zheng, the author of MT3D, and

incorporated several new subroutine packages which are described below. Flexibility, robustness, clarity of coding, and outstanding documentation all contributed to the selection of MODFLOW for this project.

The basic MODFLOW program solves a block centered finite difference approximation to the groundwater flow equation on a variable cell size, three dimensional rectangular grid. MODFLOW allows for anisotropy so long as the grid axes are aligned with the principal directions of hydraulic conductivity. It can solve either steady or transient cases and provides options for recharge, wells, and other hydrologic features. Both confined and unconfined aquifers can be modeled. The original block centered flow package (BCF1) allowed the dewatering of layers during periods of water table decline, but could not handle rewetting due to a rising water table. This was an important limitation in modeling MADE-2 due to the pronounced water table fluctuations which were observed. The version used here incorporated BCF2 (McDonald, Harbaugh, Orr, and Ackerman; 1991), a newer package which allows rewetting. The present MODFLOW also incorporated PCG2 (Hill, 1990), a preconditioned conjugate gradient solver; LKMT18, which generates output files in a format suitable for input to MT3D; and STR1, a stream interaction package which was not used.

The user of MODFLOW must input the grid geometry, boundary and initial conditions, values related to the principal hydraulic conductivities for each cell, storage coefficients for each cell, and source parameters.

The definition of a suitable computational grid is the first step in applying MODFLOW. In view of the heterogeneity of the site and the nature of the plume, a uniform three dimensional grid was selected. As in Gray (1993), the grid consists of 9 layers, each containing 66 rows and 21 columns of 5 m x 5 m cells. The horizontal grid is identical to that of Gray (1993) with the 105 m and 330 m sides parallel to the x and y axes of the MADE-2 coordinate system, respectively. The origin of the MADE-2 coordinate system is at the center of the cell which contains all 5 injection wells (row 61, column 11). In terms of MADE-2 coordinates, the domain extends from -52.5 m to +52.5 m in the x direction and from -27.5 m to +302.5 m in the y direction.

One of the most critical steps in the development of a numerical model is geostatistical analysis, the process by which a relatively small number of irregularly spaced observations of some variable are used to assign values at the relatively large number of regularly spaced computational nodes. Gray (1993) used the commercial program SURFER for this task. In the present study the public domain software package Geo-EAS Version 1.2.1 (Englund and Sparks, 1991)

was employed. Geo-EAS is a menu driven personal computer program developed by the Environmental Protection Agency (EPA) primarily to perform two dimensional kriging. Geo-EAS allows the user to closely control most aspects of the kriging process, including the selection of linear, spherical, exponential, or Gaussian variograms. The program can also calculate descriptive statistics and produce two dimensional contour plots. In comparison with SURFER Version 4, Geo-EAS is less polished, has inferior graphics, and has more glitches, e.g. the Gaussian variogram doesn't always work. On the other hand, Geo-EAS is more flexible and is much less of a black box. In this study all kriging was done using Geo-EAS, but most of the final contour plots were made using SURFER Version 4.

Geological logs from 32 locations scattered over and near the site were analyzed to determine the vertical boundaries of the aquifer. Program XLTOGE was written to reformat the measured ground surface and aquitard top elevations for input to Geo-EAS. These data were kriged using a linear variogram for the ground surface elevation and a spherical variogram for the aquifer bottom elevation. The ground surface elevation was estimated to vary from 64.68 m to 65.99 m, and the aquifer bottom was estimated to range from 49.90 m to 55.51 m MSL.

The rewetting capability of the BCF2 package allowed for a more efficient vertical grid spacing that had been used previously. In Gray (1993), the computational domain was bounded below by an impermeable plane at 51.0 m, and the lower 8 layers were each 1 m thick. The top layer, with a base at 59.0 m, had an upper boundary which fluctuated with the water table. As the observed water table reached its peak in May 1991, cells in the top layer were up to 6.1 m thick. This was undesirable from the standpoint of accuracy, but was necessary because BCF1 required the lower boundary of the top layer to be low enough to guarantee against dewatering.

In the present grid the base of the upper layer is at 63.0 m, so that its saturated thickness should never exceed 2.1 m. The next seven layers are each 1 m thick. The top of the lowest layer is at 56.0 m, and its impermeable bottom varies to match the top of the aquitard. The thickness of the lowest layer ranges from 0.49 m to 6.10 m with a mean of 3.31 m. In terms of MODFLOW classification, layer 1 is unconfined, layers 2 through 7 are fully convertible (LAYCON = 3), and layers 8 and 9 are confined.

There were 82 piezometers scattered irregularly over and near the computational domain. Heads were recorded continuously in 16 piezometers. There were also 17 manual piezometer surveys conducted at intervals of about one month and typically covering 45 piezometers. The continuous and survey observations showed good

agreement. From the first observations, about 1 week before injection, until about 180 days after injection, heads declined smoothly less than 1 m. After that date heads underwent larger and more erratic changes. These results showed that a transient model was essential.

The piezometric heads from the monthly surveys were needed to establish the initial head at each node, as well as the head at each boundary node as a function of time. Using SURFER, Gray (1993) kriged using all of the available heads, pooling all depths and including piezometers which were far from the computational domain. The results were assigned as initial and boundary conditions to all layers, i.e. there was no variation of head with depth. The numerical solutions obtained with these conditions showed heads which dropped toward the northwest corner of the grid, suggesting that the plume should bend toward the northwest. As the observations showed the plume bending toward the northeast, it was important to be more careful in translating the observed heads into initial and boundary conditions.

The commercial spreadsheet Quattro Pro for Windows was used to examine the distribution of the piezometer screen midpoint elevations. It was noticed that most were close to either 60.5 m or 56.0 m. Geo-EAS was used to reject piezometers which were not close to these elevations or were too far outside the computational domain. The pizometers selected for kriging consisted of an upper set of 15 whose screen midpoints ranged from 59.76 m to 61.22 m with a mean of 60.55 m, and a lower set of 23 whose elevations were between 55.51 m and 56.71 m with a mean of 55.95 m. Figure 1 shows that the coverage of the (plan) north end of the computational domain was sparse at both levels.

MADETOGE was written to segregate the monthly piezometer survey data into upper and lower piezometer files. These files were kriged with linear variograms using Geo-EAS. Figure 2 shows the results for the upper and lower piezometer sets for the survey of March 8, 1991. In almost every survey the heads at both levels decline toward the northwest. The upper level heads were assigned to layers 1 through 4, and the lower level heads to layers 8 and 9. Heads were specified for layers 5, 6, and 7 by linear interpolation. Program BASMAKER wrote the MODFLOW Basic package input file which included the initial heads at every node. Program GHBMAKER created the input file for the MODFLOW General Head Boundary package. The function of this package was to maintain specified heads at every boundary node (Dirichlet boundary conditions).

The net recharge was the difference between precipitation and evapotranspiration. Daily precipitation and temperature data were measured at

the CAFB weather station, less than 2 km from the test site. Daily pan evaporation data from State University, about 35 km distant, were supplied by State Climatologist Dr. C. L. Wax. Missing evaporation data were estimated from the daily maximum temperatures using the empirical equation of Pote and Wax (1986). Based on the recommendation of Dr. Wax, a pan coefficient of 0.8 was used to estimate the evapotranspiration.

The 17 piezometer surveys and the two day injection period were used to define 18 stress periods during which all boundary conditions and water sources were constant. These were the same periods used by Gray (1993). Except for the injection period, the stress periods were approximately centered on the survey dates. The recharge rates were the averages of the daily values. Table 1 defines the stress periods used in MODFLOW. The injection occurred at a rate of 4.85 m^3/day on simulation days 15 and 16 at row 61, column 11, and layer 7. A constant time step of 2 days was used in all the MODFLOW simulations.

Table 1. Stress periods and recharge rates used in MADE-2 simulations.

stress period	starting date	starting sim. day number	period length [days]	head survey date	survey sim. day number	recharge rate [m/day]
1	June 12	1	14	June 19	8	-0.00313
2 *	June 26	15	2	"	"	-0.00478
3	June 28	17	36	July 23	42	-0.00148
4	Aug. 3	53	28	Aug. 13	63	-0.00409
5	Aug. 31	81	32	Sept. 17	98	-0.00286
6	Oct. 2	113	26	Oct. 15	126	-0.00107
7	Oct. 28	139	24	Nov. 7	149	+0.00071
8	Nov. 21	163	32	Dec. 5	177	+0.00942
9	Dec. 23	195	32	Jan. 8	211	+0.00387
10	Jan. 24	227	30	Feb. 8	242	+0.00809
11	Feb. 23	257	28	Mar. 8	270	+0.00114
12	Mar. 23	285	30	Apr. 4	297	+0.00794
13	Apr. 22	315	24	May 10	333	+0.01022
14	May 16	339	18	May 20	343	+0.00357
15	June 3	357	24	June 13	367	+0.00046
16	June 27	381	34	July 9	393	-0.00273
17	July 31	415	32	Aug. 19	434	-0.00159
18	Sept. 1	447	22	Sept. 11	457	-0.00384
last day	Sept. 22	468				

* injection period

Vertical profiles of horizontal hydraulic conductivity were measured in 67 wells scattered in and around the computational domain. The data were measured over successive 15 cm layers using a borehole flowmeter. The gaps where the well screens were jointed were filled in with the values immediately above and below. The height profiled and the layer boundaries varied from well to well.

KAVG94 was written to relate these profiles to the grid layers. The tops of the profiles varied from 57.62 m to 62.68 m. The program extended each profile up to 64.0 m using the conductivity at the top of the profile. The lowest points varied from 51.88 m to 56.22 m. Profiles were extended down to 56.0 m or the next lower integer elevation using the conductivity at the lowest point. The extended profiles were averaged arithmetically over each MODFLOW layer to generate horizontal conductivities. With the assumption that each 15 cm slice of material was isotropic, the extended profiles were averaged harmonically between the midpoints of the MODFLOW layers to generate vertical leakances. Leakance is the vertical conductivity divided by the thickness between adjacent nodes. Due to the variable thickness of layer 9, the leakance between layers 8 and 9 was calculated for the interval from 56.5 m to 55.5 m rather than to the actual midpoint of the lowest cells. Exceptions occurred at wells K-2, K-26, and K-28 where the profiles ended at 56.0 m.

The next task was to interpolate and extrapolate the averaged profiles horizontally so as to obtain the horizontal conductivity and vertical leakance at each node of the computational grid. The averaged profiles were log transformed using KA2LOG, kriged with Geo-EAS, and transformed back by DLOGFILE. The log transformation was necessary to avoid negative values in the kriging process. Spherical or exponential variograms were used. Program BCF2MAKER was written to format the conductivity values for input to the MODFLOW BCF2 package.

During execution, MODFLOW calculates the transmissivity of the cells which are partially saturated by multiplying the horizontal conductivity of the cell by its saturated depth. Since the horizontal hydraulic conductivity represents an average over the entire cell thickness, this is correct only if the cell is truly homogeneous. The vertical leakance is treated as a constant as long as a cell contains water, even though it represents an average over the full region between nodes. This is not correct either.

Little was known about the storage coefficients. A specific yield of 0.1 was measured in a single traditional pump test (AT-2) (Boggs, Young, Benton, and Chung; 1990). No measurements of specific storage were made, so a confined

storage coefficient base value of 0.0001 was assumed, based on textbook values for specific storage in sand and sandy gravel (Anderson and Woessner, 1992). In view of the great uncertainty of these parameters, simulations were run with higher and lower values in order to investigate the sensitivity of the results. In each simulation, the storage coefficients were constant throughout the grid. In reality, great variability is expected; but there was no defensible way to account for this on the basis of the available data.

The 468 day experiment was simulated on a Sun Sparcstation 2 using the PCG2 solver. In spite of the rather severe vertical motion of the water table, MODFLOW performed reliably. Table 2 lists the differences among the five cases which were computed.

Table 2. MODFLOW simulation summary.

Case	RELAX	WETDRY [meters]	specific yield	confined storage coef.	run time [min.]	final volume error
1	0.98	-0.1	0.1	0.0001	60	-0.25%
2	1.00	-0.1	0.1	0.0001	unknown	-0.24%
3	0.98	-0.01	0.1	0.0001	72	-0.25%
4	0.98	-0.1	0.2	0.0005	94	-1.52%
5	0.98	-0.1	0.05	0.00005	58	-0.23%

Taking Case 1 as the base case, Case 2 tests the effect of increasing RELAX, a convergence parameter in the PCG2 solver package. This variation left the solution virtually unchanged. Case 3 examines the effect of reducing WETDRY, a parameter in the BCF2 package which controls cell rewetting. The negative sign indicates that the rewetting of cell x depends on the head in the cell below. The absolute value of WETDRY is the amount by which the head in the cell below must exceed the bottom elevation of cell x before it rewets. Case 3 results were virtually identical with Case 1. A positive value of WETDRY makes rewetting depend on the heads in the four horizontally adjacent cells. Runs with positive values of WETDRY invariably failed to converge.

Cases 4 and 5 varied the storage coefficient values. It can be seen that increasing the storage coefficients increases the volumetric discrepancy and the run time. The effects on the nature of the solution are discussed further below, but they have not yet been fully assessed.

Figure 3 presents the Case 1 head contours on simulation day 270 (March 8, 1991) in layers 4 and 9. Compared with the kriged distributions for the upper and lower piezometers on the same day shown in Figure 2, it can be seen that the head distributions are both qualitatively and quantitatively similar. In both the

predicted and observed cases, the flow is downward. The tendency for the heads to decline toward the northwest is evident in this figure and throughout the simulation.

In order to obtain a numerical measure of agreement, the simulated heads were compared to the continuous head observations. Program WELLGRPH was written to extract from the MODFLOW binary output file the head time series for those cells which contained continuously monitored piezometers. The continuous piezometer records show erratic day to day variations which cannot be predicted by a model whose boundary conditions change only 16 times in 468 days. To provide a reasonable basis of comparison, the daily observed heads were averaged over each stress period by program HYDROGRA. Figure 4 compares the Case 1 predictions to the observed (averaged) heads at two piezometers with the same horizontal position. The simulated results adjust rapidly to the boundary conditions for each stress period. The model results are better at the upper level (P55a) than at the lower level (P55b), where the model overpredicts markedly in stress periods 9, 11, and 13.

The averaged observations were subtracted from the unaveraged MODFLOW heads and the maximum, minimum, and root mean square (rms) differences were summarized in Table 3. Case 5, with the smallest storage coefficients, gives the best overall accuracy. Case 4 has the greatest excursions from the observations, yet its rms deviation is smaller than Case 1. Although the ability of the model to reproduce the observations is imperfect, it is hard to see how the model could be improved given the limitations of the data base.

Table 3. Deviation of MODFLOW heads from continuous observations [meters].

	min.	min.	min.	max.	max.	max.	rms	rms	rms
Well	Case 1	Case 4	Case 5	Case 1	Case 4	Case 5	Case 1	Case 4	Case 5
P53a	-0.65	-1.32	-0.57	0.74	0.51	0.14	0.329	0.228	0.194
P54a	-0.53	-0.84	-0.37	0.39	0.58	0.30	0.143	0.165	0.136
P54b	-0.42	-0.78	-0.17	0.43	0.52	0.43	0.147	0.159	0.143
P55a	-0.53	-0.80	-0.37	0.44	0.50	0.44	0.199	0.204	0.199
P55b	-0.12	-0.44	+0.01	1.01	1.01	1.01	0.374	0.374	0.374
P60a	-0.51	-0.51	-1.51	0.30	0.38	0.30	0.188	0.188	0.188
P61a	-0.40	-0.40	-0.40	0.36	0.36	0.36	0.188	0.188	0.188
P61b	-0.39	-0.39	-0.39	0.23	0.23	0.23	0.154	0.154	0.154
average	-0.44	-0.69	-0.35	0.49	0.51	0.40	0.215	0.208	0.197

TRANSPORT MODELING

MT3D is a public domain program developed for the EPA to solve the three dimensional groundwater transport equation for dissolved contaminants (Zheng, 1990). MT3D is coded in Fortran 77 and uses the same modular structure as MODFLOW. In fact, MT3D accepts as input the head and flux distributions computed by MODFLOW (or similar flow models). MT3D then predicts the concentration field of a single contaminant which undergoes advection, dispersion, and chemical reactions. The program provides for various types of point and area sources and sinks including wells, recharge, and flows through the domain boundaries. MT3D Version 1.80 was used in this study.

Because of the computational difficulties of numerical dispersion and oscillation in advection-dominated flows, MT3D incorporates four options for calculating the advection term. The Method of Characteristics (MOC) tracks a large number of imaginary tracer particles forward in time. The Modified Method of Characteristics (MMOC) tracks particles located at the cell nodes backward in time. The MMOC requires much less computation than the MOC, but it is not as effective in eliminating artificial dispersion, especially near sharp fronts. The Hybrid Method of Characteristics (HMOC) uses the MOC near sharp concentration gradients and the MMOC in the remainder of the domain. An Eulerian Upstream Differencing (UD) option is provided for problems in which advection does not dominate.

The dispersion terms are computed using a fully explicit Eulerian central difference method. For isotropic media, the dispersion coefficients are based on longitudinal and transverse dispersivities. For more complex situations, there is an option which distinguishes horizontal and vertical transverse dispersivities. The explicit formulation reduces the memory needed, but requires limits on the time step to assure numerical stability. Consequently each flow model time step may be automatically subdivided into several transport steps in order to maintain numerical stability in MT3D.

MT3D allows both equilibrium sorption and first order irreversible rate reactions. Equilibrium sorption reactions transfer contaminant between the dissolved phase and the solid phase (which is sorbed to the soil matrix) at time scales much shorter than those of the flow. These reactions may be described by linear isotherms or nonlinear isotherms of the Freundlich or Langmuir types. In first order irreversible rate reactions the rate of mass loss is linearly proportional to the mass present. This class includes radioactive decay and

certain types of biodegradation.

MT3D requires information beyond that needed for and calculated by MODFLOW. A porosity is needed for each cell in order to calculate seepage velocities, yet porosities were measured in only four core holes. The 84 samples had a mean porosity of 0.32, and this value was used for every cell. Based on the MADE-2 observations and an assumed two dimensional analytical model for the plume, Boggs and others (1993b) estimated the longitudinal dispersivity to be 10 m and the transverse horizontal dispersivity to be less than 2.2 m. The base values of dispersivity used were 10 m in the longitudinal direction, 1 m in the horizontal transverse direction, and 0.1 m in the vertical transverse direction. For the purpose of calculating concentrations, every wetted layer was assumed to have a uniform thickness of 1 m, although the actual thickness varied for the top and bottom layers.

MT3D was applied only to the tritium plume. The molecular diffusion coefficient of tritium in water, calculated using the Wilke-Chang method, was multiplied by an assumed tortuosity of 0.25 to yield the value of 2.16×10^{-4} m^2/day for the molecular diffusion coefficient of tritium in a saturated porous medium. The injected fluid had a tritium concentration of 0.0555 Ci/m^3; and the natural background, including recharge and boundary inflows, was set to zero. Water leaving the domain carried the concentration of the cell it last occupied. Sorption does not affect tritiated water, but tritium decays with a 12.26 year half-life.

The transport simulations attempted, all based on the Case 1 MODFLOW head solution, are summarized in Table 4. None are remotely satisfactory. No run extended beyond simulation day 141 because by that time each had experienced a numerical failure or had been terminated because the solution was unreasonable. In general, the run times were inconveniently long. The mass discrepancies appear either unacceptably large (MOC, MMOC, and HMOC) or remarkably tiny (UD), but the meaning of this parameter is not clear. Runs 7 (HMOC) and 8 (UD) predicted nearly identical plumes even though the mass discrepancies were very different.

Run 3 produced a widely spread plume even though the dispersion package was turned off. This appears to be a numerical shortcoming of the MMOC method because no-dispersion runs 5 (MOC) and 6 (HMOC) predicted unrealistically small spreads. All of the no-dispersion runs were free from negative concentrations. Run 11 was a repetition of Run 9 using double precision arithmetic; the results were identical. In Runs 9 and 11 the dispersivities in the longitudinal,

transverse horizontal, and transverse vertical directions were 4.0 m, 0.4 m, and 0.4 m, respectively. Runs 12 (UD) and 14 (HMOC) used dispersivities in the longitudinal, transverse horizontal, and transverse vertical directions of 1.0 m, 0.1 m, and 0.1 m, respectively. In Run 13 (UD) the dispersivities were all 0, but molecular diffusion was active. In general, the smaller the dispersivities, the more realistic the plume appeared.

Table 4. Summary of MT3D simulations.

Run	advection method	dispersion	long. dispersivity [m]	last sim. day	run time [hours]	mass discrep.	plume characteristics
1	HMOC	yes	10.0	30.2*	15.75	+7.93%	wide spread, some < 0
2	MMOC	yes	10.0	5.0*	1.75	n.a.	injection not started
3	MMOC	no	n.a.	129.4	10.4	+82%	wide spread
4	HMOC	no	n.a.	20.4*	0.72	+19.2%	not recorded
5	MOC	no	n.a.	62.1*	3.5	-13.1%	confined to 7 cells
6	HMOC	no	n.a.	140.9	17.38	+17.2%	confined to 8 cells
7	HMOC	yes	10.0	44.6*	47.05	+4.55%	wide spread lots < 0
8	UD	yes	10.0	61.2	16.6	-0.0001%	wide spread, lots < 0
9	UD	yes	4.0	90.4	<21.4	+0.0001%	wide spread, lots < 0
11	UD **	yes	4.0	90.4	<29	+0.0001%	identical to case 9
12	UD	yes	1.0	128	<5.37	+0.0002%	realistic, lots <0
13	UD	yes	0.0	138.3	<8	+0.0003%	realistic, few <0
14	HMOC	yes	1.0	105.9*	<12.6	+12.3%	realistic, lots<0

* run terminated by user. ** double precision.

CONCLUSIONS

1. Geo-EAS Version 1.2.1 is technically superior to SURFER Version 4. It provides a satisfactory tool for exploratory data analysis and two dimensional kriging. SURFER has better graphic capabilities.

2. Three dimensional groundwater flow simulations using MODFLOW are practical and consistent. The rewetting capability of the BCF2 package improves the accuracy of simulations in which the water table fluctuates as much as in MADE-2.

3. Although the flow model has not been subjected to grid refinement or extensive parametric variation studies, the comparison between the simulated and observed heads is satisfactory. Given the existing data, there is little prospect for significant improvement.

4. The simulated head distribution suggests that the plume should bend toward the northwest. The observations show the plume bending toward the northeast. This discrepancy is probably due to inaccurate head boundary conditions caused by a lack of piezometers near the northern end of the grid.

5. We were unsuccessful in our attempts to simulate the spread of the tritium plume using MT3D. Further efforts to achieve complete, accurate simulations of the tritium plume should be made.

ACKNOWLEDGEMENTS

We thank Dr. Tom Stauffer, Dr. Howard Mayfield, and Mr. Chris Antworth of Armstrong Laboratory; and Dr. Kirk Hatfield of the University of Florida for their advice and assistance.

REFERENCES

M. P. Anderson and W. W. Woessner, 1992. Applied Groundwater Modeling, Academic Press, New York, p. 41.

J. M. Boggs, S. C. Young, D. J. Benton, and Y. C. Chung, 1990. Hydrogeological Characterization of the MADE Site, EPRI Topical Report EN-6915, Electric Power Research Institute, Palo Alto, California.

J. M. Boggs, S. C. Young, L. Beard, L. W. Gelhar, K. H. Rehfeldt, and E. E. Adams, 1992. Field Study of Dispersion in a Heterogeneous Aquifer 1. Overview and Site Description, Water Resources Research, 28, 3281-3291.

J. M. Boggs, L. M. Beard, S. E. Long, W. G. MacIntyre, C. P. Antworth, and T. B. Stauffer, 1993a. Database for the Second Macrodispersion Experiment (MADE-2), Electric Power Research Institute, TR-102072.

J. M. Boggs, L. M. Beard, W. R. Waldrop, T. B. Stauffer, W. G. MacIntyre, and C. P. Antworth, 1993b. Transport of Tritium and Four Organic Compounds During a Natural-Gradient Experiment (MADE-2), EPRI draft report.

E. Englund and A. Sparks, 1991. Geo-EAS 1.2.1 Geostatistical Environmental Software User's Guide, U. S. Environmental Protection Agency EPA 600/8-91/008.

D. D. Gray, 1992. Preliminary Numerical Model of Groundwater Flow at the MADE2 Site, U. S. Air Force Summer Faculty Research Program (SFRP) Reports, Volume 6, pp. 11-1 through 11-19, Air Force Office of Scientific Research, Bolling AFB, Washington, DC.

D. D. Gray, 1993. Numerical Modeling of Groundwater Flow and Transport at the MADE-2 Site, U. S. Air Force Summer Research Program Final Reports, Volume 2, pp. 21-1 through 21-20, Air Force Office of Scientific Research, Bolling AFB, Washington, DC. NTIS ADA 278 693.

M. C. Hill, 1990. Preconditioned Conjugate Gradient 2 (PCG2), A Computer Program for solving Ground-Water Flow Equations, U. S. Geological Survey, Water-Resources Investigations Report 90-4048.

M. G. McDonald and A. W. Harbaugh, 1988. Techniques of Water-Resources Investigations of the United States Geological Survey, Chapter A1, A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, U. S. Government Printing Office, Washington, D. C.

M. G. McDonald, A. W. Harbaugh, B. R. Orr, and D. J. Ackerman, 1991. A Method of Converting No-Flow Cells to Variable-Head Cells for the U. S. Geological Survey Modular Finite-Difference Ground-Water Flow Model, U. S. Geological Survey Open-File Report 91-536.

J. W. Pote and C. L. Wax, 1986. Climatological Aspects of Irrigation Design Criteria in Mississippi, Technical Bulletin 138, Mississippi Agricultural and Forestry Experiment Station, Mississippi State University, Mississippi State, Mississippi.

K. R. Rehfeldt, J. M. Boggs, and L. W. Gelhar, 1992. Field Study of Dispersion in a Heterogeneous Aquifer 3. Geostatistical Analysis of Hydraulic Conductivity, Water Resources Research, 28, 3309-3324.

T. B. Stauffer, C. P. Antworth, R. G. Young, W. G. MacIntyre, J. M. Boggs, and L. M. Beard, 1994. Degradation of Aromatic Hydrocarbons in an Aquifer During a Field Experiment Demonstrating the Feasibility of Remediation by Natural Attenuation, U. S. Air Force, Armstrong Laboratory, AL/EQ-TR-1993-0007.

C. Zheng, 1990. MT3D, A Modular Three-Dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems, prepared for U. S. EPA Robert S. Kerr Environmental Research Laboratory, Ada, Oklahoma, by S. S. Papadopulos & Associates, Inc., Rockville, Maryland.

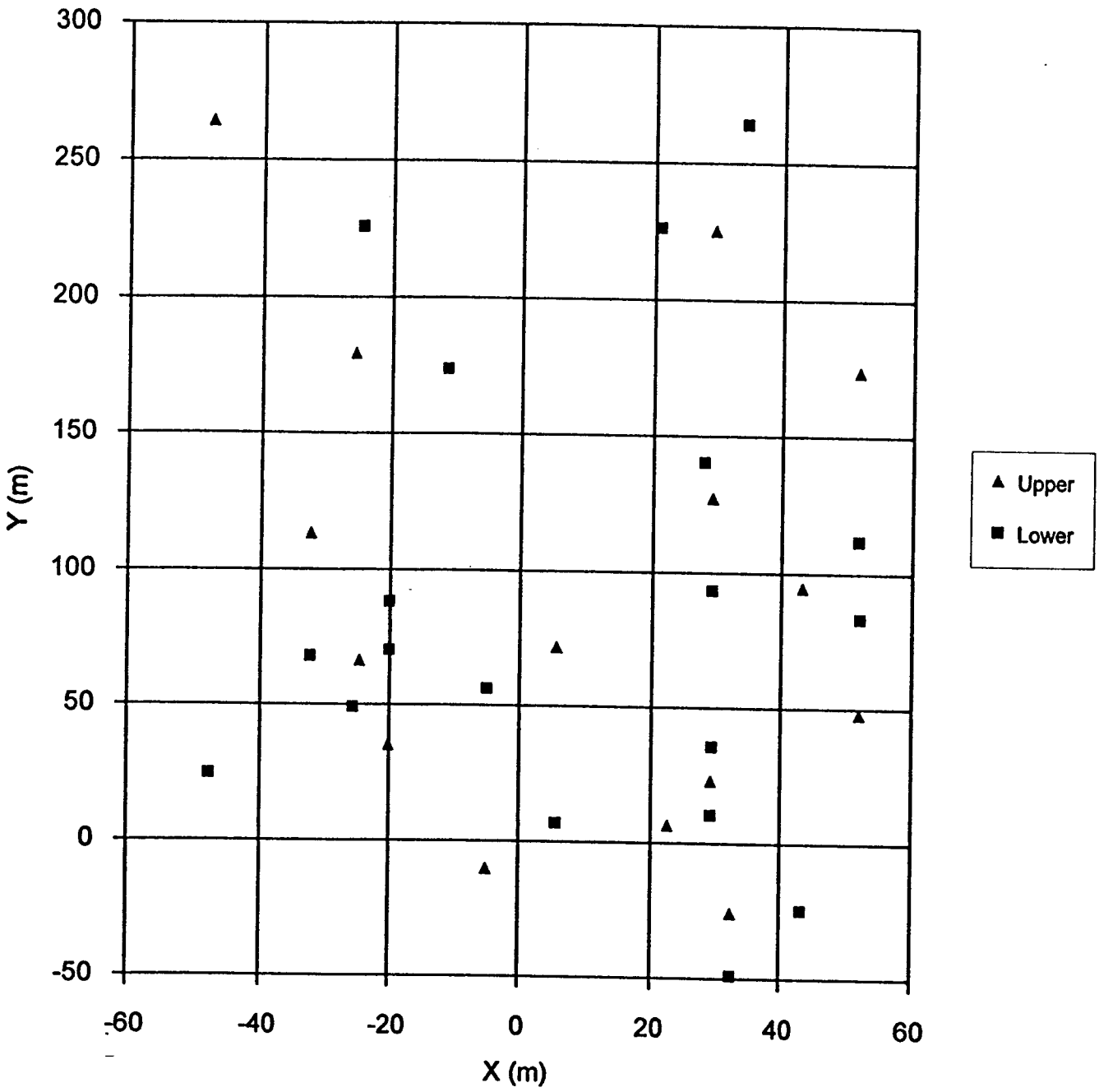


Figure 1. Locations of upper (squares) and lower (triangles) piezometers used to establish initial and boundary conditions. Four locations are common to both sets.

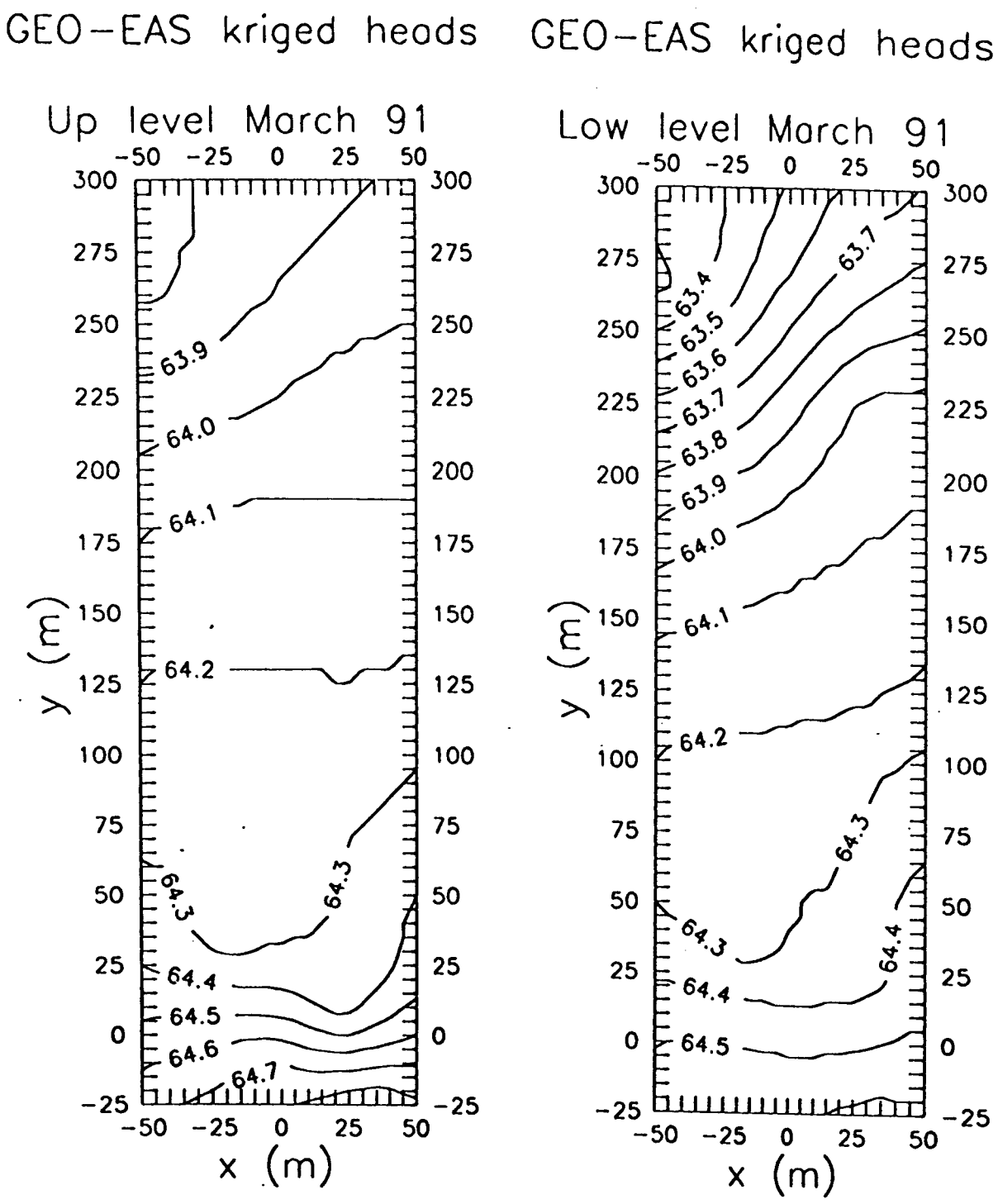


Figure 2. Upper (left) and lower (right) kriged head distributions for simulation day 270 (March 8, 1991). Heads are in meters.

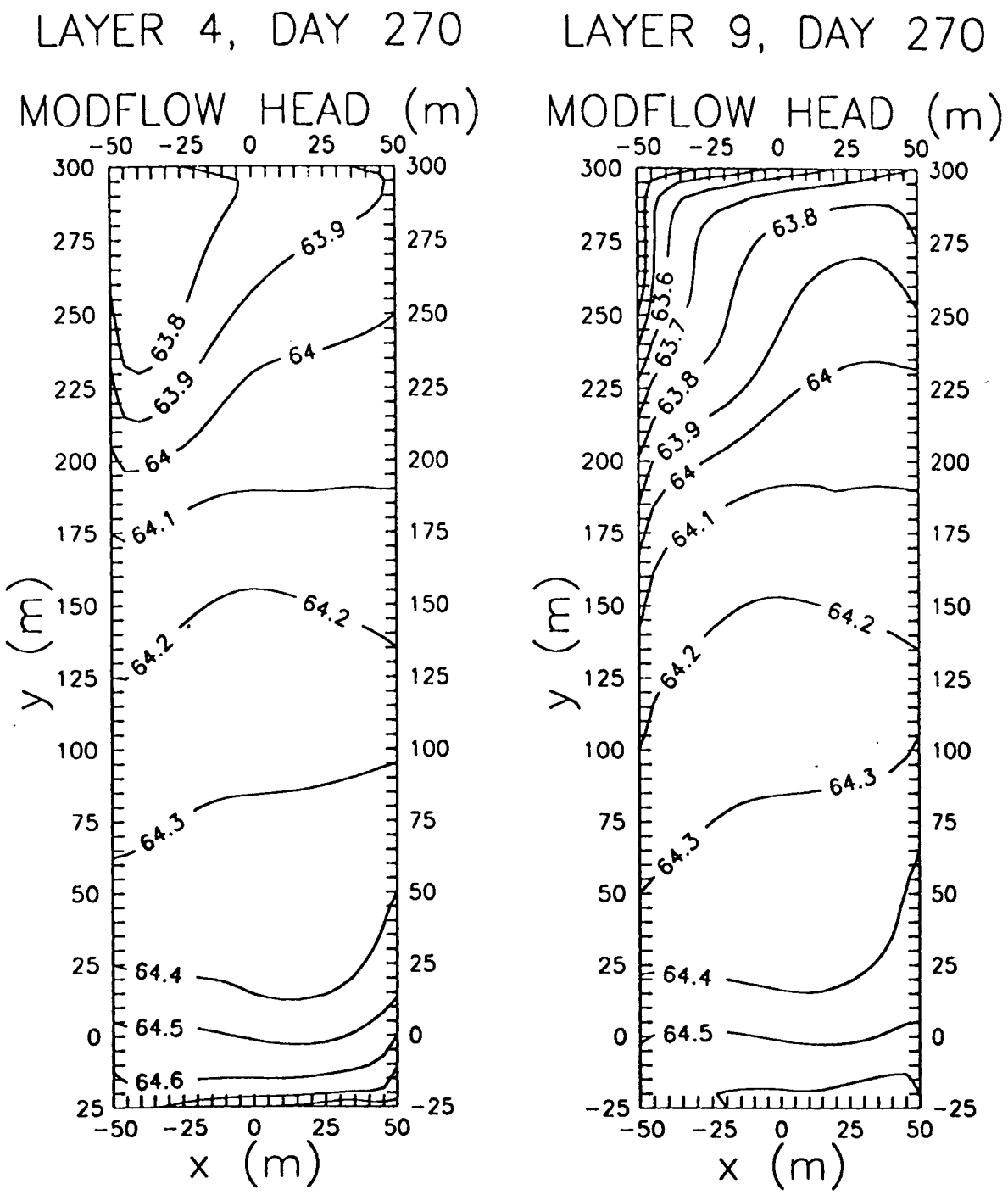


Figure 3. MODFLOW Case 1 simulated heads for layers 4 (left) and 9 (right) for simulation day 270 (March 8, 1991). Heads are in meters.

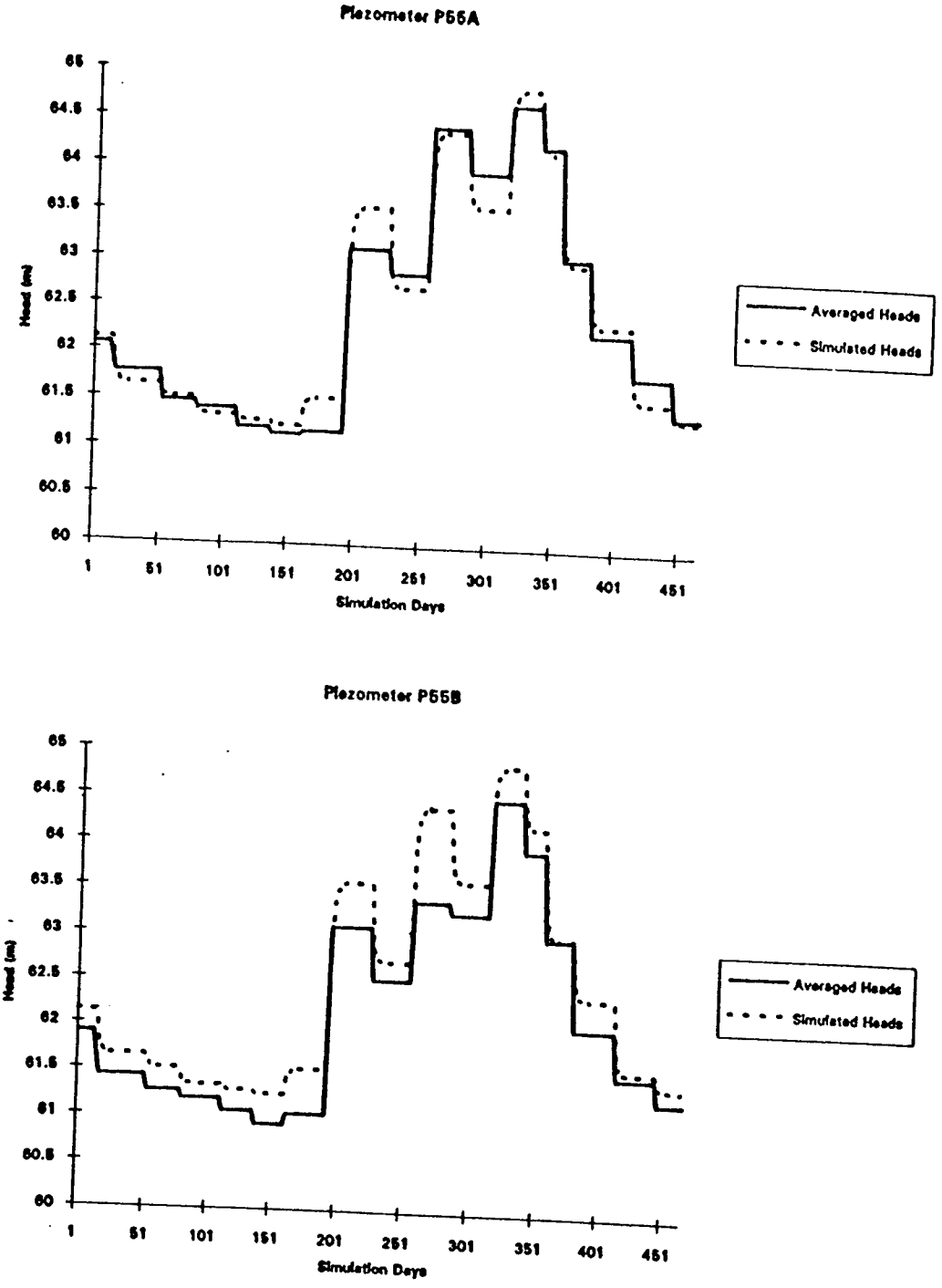


Figure 4. Comparison of MODFLOW Case 1 simulated heads with observed heads averaged over stress periods at piezometers P55a (top) and P55b (bottom). Heads are in meters.

THE WORKLOAD ASSESSMENT MONITOR: PROGRESS TOWARDS ON-LINE CLASSIFICATION OF MENTAL WORKLOAD IN HUMAN SUBJECTS

Arthur M. Ryan
Graduate Student
Department of Psychology

Wright State University
3640 Colonel Glenn Hwy.
Dayton, OH 45435

Final Report for:
Summer Graduate Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

THE WORKLOAD ASSESSMENT MONITOR: PROGRESS TOWARDS ON-LINE CLASSIFICATION OF MENTAL WORKLOAD IN HUMAN SUBJECTS

Arthur M. Ryan
Graduate Student
Department of Psychology
Wright State University

Abstract

The primary goals of the current project were to test the WAM system in a multi-task environment which represented three different MWL conditions, determine which EEG sites were sensitive to the different multi-task scenarios, and to refine the EEG bands to be used as input to the classifier. Based on visual inspection of the performance results it was determined that three levels of MWL had been established using the MATB software. Clearly, the WAM classifier was not sensitive to differences in MWL as the average classification score for each MATB task scenario was approximately 2. The EEG spectral bands used as input features to the WAM classifier in the present study where not the same as suggested by the PCA analysis. The PCA analysis indicates that the Alpha band should be used as an input feature but does not provide support that the Theta band helps discriminated MWL. Instead the results of the PCA indicated that in addition to the Alpha band, a high frequency band may prove to be a better input feature to the WAM classifier than the Theta band as indicated by the significant factor scores in the high frequency region. Analysis of the MATB data indicates the workload classes are not separable using the current feature inputs. These results suggest that the means for all three MATB scenarios may have been very close and the covariances were large resulting in a large degree of overlap using the current frequency domain input features (i.e., Alpha and Theta bands).

THE WORKLOAD ASSESSMENT MONITOR: PROGRESS TOWARDS ON-LINE CLASSIFICATION OF MENTAL WORKLOAD IN HUMAN SUBJECTS

Arthur M. Ryan

INTRODUCTION

As human-machine systems have become more complex, the workload of the human operator has changed from physical to mental. Gopher and Donchin (1986) suggest that mental workload (MWL) may be viewed as the difference between the capacities of the information processing system required for task performance and the capacity available at any given time. While physical workload is obvious and can be measured in terms of energy expenditure, a metric for MWL which offers the lucidity and reliability of physical workload is not yet available. However, as human-machine systems continue to be automated, and with the concept of "smart" systems which could in real-time dynamically reallocate system functions based on a current assessment of operator workload, the need for sensitive, diagnostic, and non-intrusive workload measures becomes apparent. Although a large number of workload assessment procedures have been proposed, most can be classified into one of three major categories. Performance-based measures derive an index of workload from some aspect of operator behavior. Subjective measures require operators to judge and report their own experience of the imposed workload. Physiological measures infer the level of workload from some aspect of the operator's physiological response to task demands (O'Donnell & Eggemeier, 1986). Since physiological measures are continuous and non-intrusive they have the potential to provide on-line evaluation of MWL.

The mission of the Workload Assessment Monitor (WAM) project is to

develop an on-line real-time system to monitor workload levels. Specifically, seven electroencephalogram (EEG) sites, the electro-occulogram (EOG), the electrocardiogram (ECG), and respiration are monitored and workload level is assessed using a statistical classifier. Selection of input features to the classifier can be a complicated process.

The primary goals of the current project were to test the WAM system in a multi-task environment which represented three different MWL conditions, determine which EEG sites were sensitive to the different multi-task scenarios, and to refine the EEG bands to be used as input to the classifier.

MULTI-TASK SCENARIO DEVELOPMENT

Method

The Multi-Attribute Task Battery (MATB) software was chosen to simulate the multi-task environment. The MATB consists of monitoring, tracking, communication, and process control tasks. Each task has its own "window" area on the monitor and are graphically depicted as they appear to the subject in Figure 1.

Monitoring. The monitoring task consists of warning lights and probability monitoring. Two lights are located in the upper portion of this window. One is a green light and the subjects must respond when the light is extinguished. The second is a red light to which the subject must respond when it occasionally turns on. Probability monitoring involves four vertical dials with arrows which normally fluctuate about the midpoint. The subjects's task was to respond when these fluctuations deviate significantly from center. All three monitoring tasks were present through the simulation. The monitoring tasks can be manipulated by the frequency and temporal overlap of events. Reaction time was the dependent measure for all monitoring tasks.

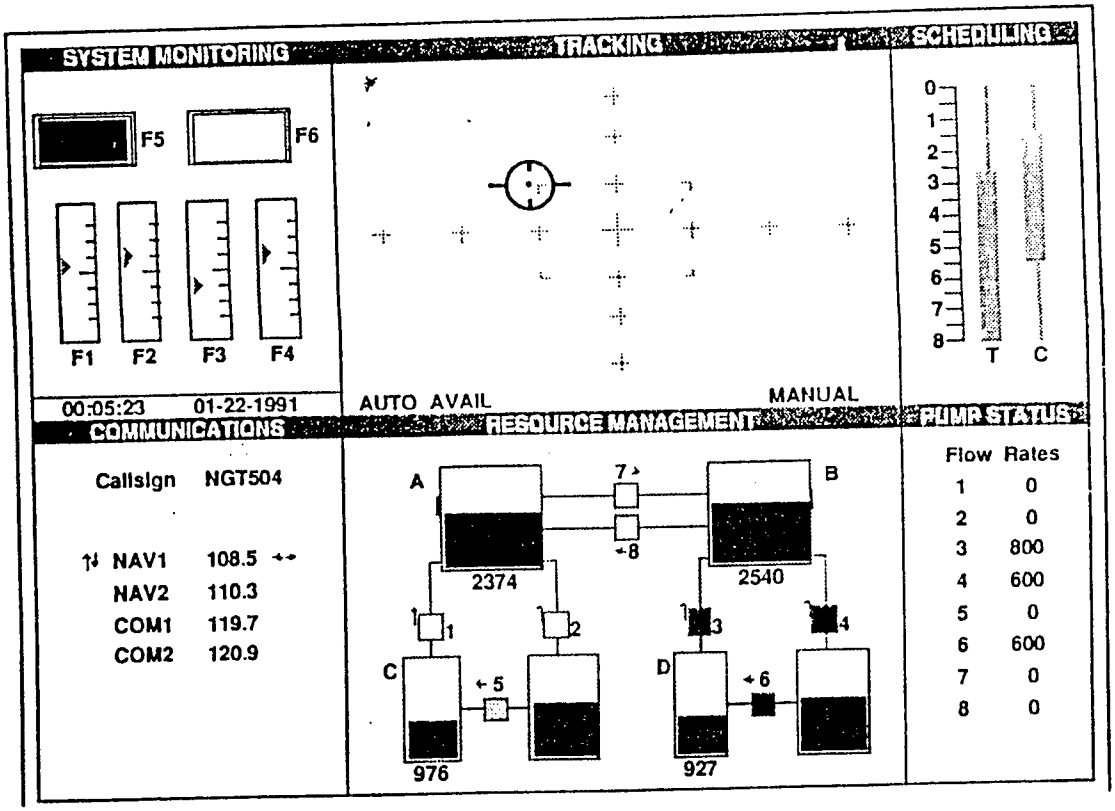


Figure 1. The Multi-Attribute Task Battery

Tracking. The tracking task was two dimensional and compensatory in nature. The subject's task was to stay on course by keeping the moving curser centered within the dotted lines which form a rectangle. This was achieved by deflections of the joystick. Root mean square error (RMSE) was the dependent measure for the tracking task. The MATB software provides three levels of tracking difficulty. Additionally, the tracking gain can be manipulated. Smaller RMSE indicates better tracking performance.

Communication. The communication task consisted of a series of audio messages. These messages began with a six-digit "call sign", repeated once, and a command to change the frequency of one of the four channels listed on

the screen. The subject had to discriminate their call sign, "NGT504", from distractor call signs and had to change the frequency for the required channel as quickly as possible. The communication task can be manipulated by the frequency and temporal overlap of events. Reaction time was the dependent measure.

Resource Management. The resource management window provides a diagram of the resource management system. The six rectangular regions were tanks which hold fuel. The green levels within the tanks represent the amount of fuel in each tank. The lines which connect the tanks were pumps which can transfer fuel from one tank to another in the direction indicated by the corresponding arrow. The numbers under the tanks represent the amount of fuel in gallons in each tank. The process of transferring fuel was accomplished by activating pumps. At the onset of each trial tanks A and B contain approximately 2000 gallons of fuel each and tanks C and D contain approximately 1000 gallons of fuel each. Subjects were required to maintain the level of fuel in both tanks A and B at 2500 gallons each. The resource management task can be manipulated by changing the flow rates of the pumps, pump failures, and the rate fuel empties from tanks A and B. Deviation from 2500 gallons for tanks A and B was the dependent measure.

Task scenarios. The MATB software allowed the experimenter to develop different task scenarios which may represent different levels of MWL. Three task scenarios were developed and were intended to represent low, medium, and high workload.

The low scenario included two vertical dial deviations from center, one red light, one red light, one communication task, one communication distractor, and one pump failure. The tracking was set to low with a gain of

35. Tanks A and B both emptied at 800 gallons per minute. Pumps 1 and 3 transferred fuel at the rate of 800 gallons per minute. Pumps 2, 4, 5, and 6 transferred fuel at a rate of 600 gallons per minute. While pumps 7 and 8 transferred fuel at a rate of 400 gallons per minute. The flow rates for the low scenario are considered default flow rates.

The medium scenario included three vertical dial deviations from center, three red lights, three green lights, two communication tasks, two communication distractors, and two pump failures. The tracking was set to medium with a tracking gain of 35. Tank A emptied at 1200 gallons per minute while tank B emptied at 800 gallons per minute. Pumps 1 - 8 transferred fuel at default rates plus 200 gallons per minute.

The high scenario included five vertical dial deviations from center, six red lights, six green lights, five communication tasks, three communications distractors, and four pump failures. The tracking task was set to high with a tracking gain of 25. Tank A emptied at 1800 gallons per minute while tank B emptied at 800 gallons per minute. Pumps 1 - 8 transferred fuel at default rates plus 300 gallons per minute. All scenarios lasted 3 minutes and events where spaced such that minimal temporal overlap occurred.

Procedure

Three subjects participated in this phase of the project. After sufficient practice with each task scenario (to stabilize performance), each subject performed each task scenario twice.

Results

Monitoring tasks. For lights the mean reaction time for the low scenario was 1.81 sec, for the medium scenario 1.93 sec, and for the high scenario 2.90 sec. For dials the mean reaction time for the low scenario was 6.06 sec, for

the medium scenario 6.71 sec, and for the high scenario 10.16 sec.

Communication task. For the communication task the mean reaction time for the low scenario was 3.13 sec, for the medium scenario 3.70 sec, and for the high scenario 4.60 sec.

Tracking task. For the tracking task the RMSE for the low scenario was 16.47, for the medium scenario 39.15, and for the high scenario 79.21.

Resource management. For the resource management task the average deviation from 2500 gallons for both tanks A and B for the low scenario was 116 gallons, for the medium scenario 204 gallons, and for the high scenario 532 gallons.

DETERMINE SENSITIVE EEG SITES

Method

In order to determine which seven EEG sites were most sensitive to the differences in MLW as determined by performance on the different task scenarios, differences in amplitude (uV) at each of 21 EEG sites across task scenarios were evaluated.

Procedure

One subject performed each task scenario twice and EEG was collected from 21 EEG sites using the Bio-Logic system.

Results

Visual inspection using the Bio-Logic brain mapping software indicated that the F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, and T6 EEG sites were most sensitive to the three task scenarios.

TESTING WAM USING THE MATB SCENARIOS

Method

The WAM uses physiological data as feature input to a Bayes Quadratic Classifier in order to classify MWL. Specifically, seven EEG sites (F3, Fz,

F4, C3, Cz, C4, P3, Pz, P4, and T6), EOG, ECG, and respiration were monitored on-line while performing the different MATB task scenarios, and workload level was assessed using the statistical classifier.

The WAM system uses the energy in the EEG spectral bands Alpha (8 - 12 Hz) and Theta (4 - 8 Hz) as input features to the classifier. These bands were chosen from experience and previous workload studies. Additionally, respiration rate, heart rate, and eyeblink rate data are also included resulting in a sample vector with 17 features (7 EEG sites X 2 EEG spectral bands per electrode site and the 3 rate features).

Procedure.

Training (130 sec from each MATB scenario) and test (360 sec from each MATB scenario) data were collect from one subject. The EEG was sampled using 128 samples per second and FFT's of each second data segment were computed and stored. A total of 14 electrodes were connected to the subject. Including 3 for the ECG, 2 for the ears (one reference, one ground), 2 for vertical EOG, and 7 for the EEG sites. Note that during the training phase, the physiological samples from each MATB scenario were collected and used to estimate the statistics (class means and covariances) required by the classifier. During the testing phase, the different MATB scenarios were performed and the physiological samples were input to the classifier. Using the estimated statistics from the training phase, the classifier outputs (once every second) an estimate of MWL based on an estimate of the "unknown" physiological sample. The classifier outputs a 1 for low workload, a 2 for medium workload, or a 3 for high workload. Performance data and the classifier output data were stored in order to evaluate the performance of the classifier as well as the performance of the subject.

Results

Monitoring tasks. For lights the mean reaction time for the low scenario was 2.14 sec, for the medium scenario 1.57 sec, and for the high scenario 1.87 sec. For dials the mean reaction time for the low scenario was 5.95 sec, for the medium scenario 6.60 sec, and for the high scenario 5.53 sec.

Communication task. For the communication task the mean reaction time for the low scenario was 1.48 sec, for the medium scenario 3.66 sec, and for the high scenario 3.16 sec.

Tracking task. For the tracking task the RMSE for the low scenario was 29.58, for the medium scenario 69.34, and for the high scenario 92.82.

Resource management. For the resource management task the average deviation from 2500 gallons for both tanks A and B for the low scenario was 133.01 gallons, for the medium scenario 141.79 gallons, and for the high scenario 229.6 gallons.

WAM classifier. The WAM classifier rated the low workload MATB scenario on average as 1.923, the medium workload MATB scenario on average as 1.965, and on average rated the high workload MATB scenario as 1.908.

REFINEMENT OF EEG BANDS

Method

The WAM system uses the energy in the EEG spectral bands Alpha (8 - 12 Hz) and Theta (4 - 8 Hz) as input features to the classifier. These bands were chosen from experience and previous workload studies. However, other bands may prove to more effective input features to the classifier. In the past features have been selected based on a Principle Components Analysis (PCA). The PCA method was used to determine if different EEG spectral band might prove to be more effective input features for the current MATB task

scenarios.

The PCA method is based on the idea that a distribution can be represented by a coordinate system with principal axes along dimensions which account for the largest variance. The principal axes are the eigenvectors of the correlation matrix of the distribution. EEG spectral bands (input features for the WAM classifier) can be determined by selecting only those eigenvectors associated with the largest eigenvalues of the correlation matrix. The largest eigenvalues represent the largest variance.

Procedure

The procedure used for determining the new EEG spectral bands using PCA with varimax rotation started by computing the correlation matrix over all task scenarios and the seven electrode sites. The eigenvalues and eigenvectors of the correlation matrix were determined. The 6 eigenvectors associated with the 6 largest eigenvalues were retained. Finally, the retained eigenvectors were rotated using the varimax rotation procedure. This method was implemented using the WAM EEG data from the training session. After keeping and rotating the eigenvectors associated with the 6 largest eigenvalues, factor scores depicted in Figure 2 were obtained.

Results

Note that factors 1, 3, and 4 are statistically significant (i.e., the magnitude of the factor score is greater that 0.6).

GENERAL DISCUSSION

The primary goals of the current project were to test the WAM system in a multi-task environment which represented three different MWL conditions, determine which EEG sites were sensitive to the different multi-task scenarios, and to refine the EEG bands to be used as input to the classifier.

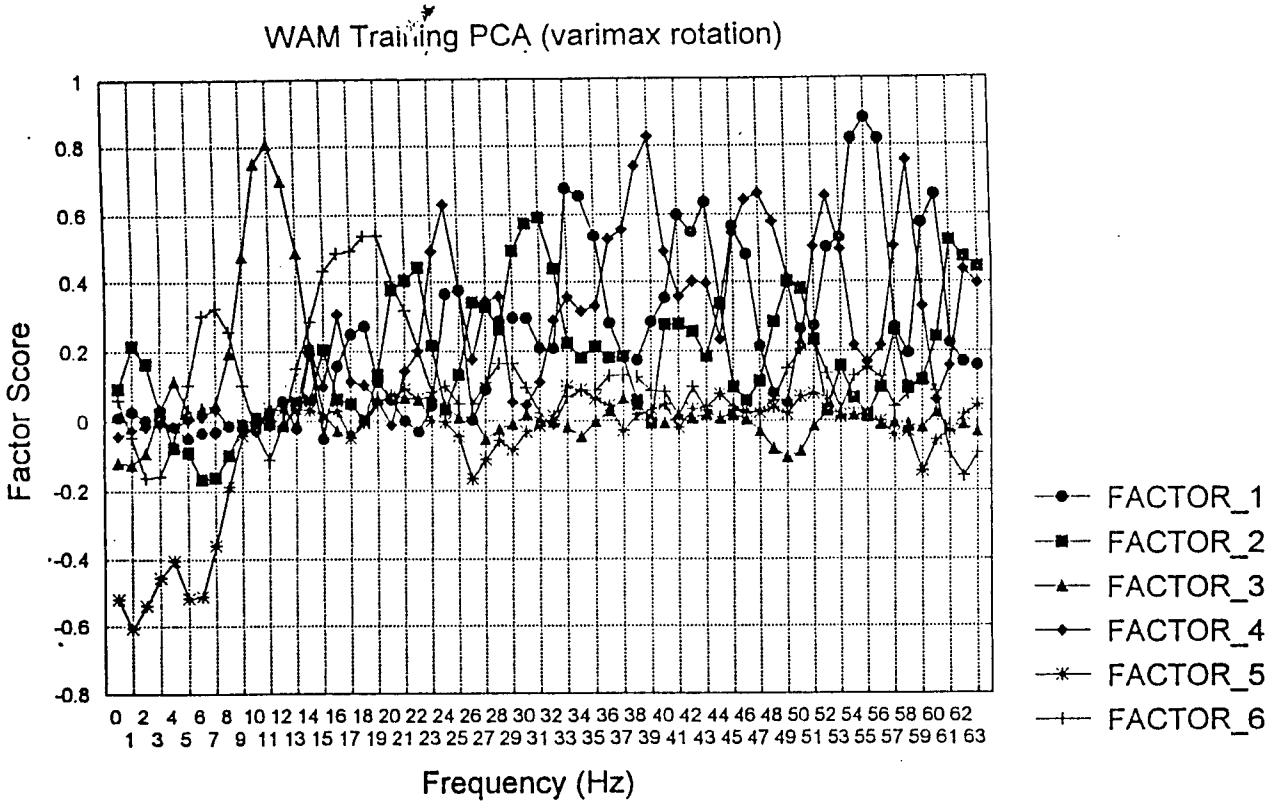


Figure 2. Principal Components Analysis Results

Based on visual inspection of the performance results it was determined that three levels of MWL had been established using the MATB software. While the performance on the tracking and resource management tasks indicate clear differences in MWL for the three MATB scenarios, the data from the monitoring and communication tasks did not indicated differences in MWL. Clearly, the WAM classifier was not sensitive to differences in MWL as the average classification score for each MATB task scenario was approximately 2.

The EEG spectral bands used as input features to the WAM classifier in the present study where not the same as suggested by the PCA. The PCA

analysis indicates that the Alpha band should be used as an input feature but does not provide support that the Theta band helps discriminated MWL. Instead the results of the PCA indicated that in addition to the Alpha band, a high frequency band may prove to be a better input feature to the WAM classifier than the Theta band as indicated by the significant factor scores in the high frequency region.

Conclusion

Analysis of the MATB data indicates the workload classes are not separable using the current feature inputs. These results suggest that the means for all three MATB scenarios may have been very close and the covariances were large resulting in a large degree of overlap using the current frequency domain input features.

Recommendations

If MATB software is used to develop task scenarios which represent different levels of MWL, it is recommended that the scenarios be modified such that significant performance differences across all task scenarios are obtained for all tasks included within each scenario.

In future experiments it is recommended that the input features to the WAM classifier be "customized" for each individual subject. The optimal EEG sites and EEG spectral bands to be used as input features may vary across subjects. While visual inspection using the Bio-Logic brain mapping software indicated that the F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, and T6 EEG sites were most sensitive to the three task scenarios these sites may differ for different subjects. Additionally, the results form the PCA analysis indicate that limiting the number of EEG spectral bands as input features to 2 may not be warranted since several factor scores were found to be significant.

Finally, since measures from different classes (i.e., performance, subjective, and physiological) are often found to dissociate from each other, it is clear that multiple measures must be observed together to adequately capture MWL. Thus, it is recommended that input features to the WAM classifier not be limited to physiological measures.

REFERENCES

Gopher, D., & Donchin, E. (1986). Worrkload: an examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), Handbook of perception and human performance, Volume 2 (pp. 41-1-41-49). New York: John Wiley.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), Handbook of perception and human performance, Volume 2 (pp. 42-1-42-49). New York: John Wiley.

ARTERIAL ELASTANCE IN THE MAXIMIZATION OF EXTERNAL WORK TRANSFER

Mark J. Schroeder
Graduate Student
Department of Electrical Engineering

North Dakota State University
Fargo, ND 58105

Final Report for:
Graduate Student Research Program

Armstrong Laboratory

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

August, 1994

ARTERIAL ELASTANCE IN THE MAXIMIZATION OF EXTERNAL WORK TRANSFER

Mark J. Schroeder
Graduate Student
Department of Electrical Engineering
North Dakota State University

Abstract

The following work is an extension of the maximization of external work (EW) with respect to effective arterial elastance (eff. Ea) performed by Sunagawa et al. (9). A comparison of eff. Ea and physiological arterial elastance (Ea) is presented in order to clarify the main differences. Constrained maximization of EW with respect to Ea is then performed. This was accomplished by developing a ventricular-arterial coupling model to 1) estimate cardiovascular (CV) parameters from physiological data, 2) simulate the CV data and calculate EW, and 3) simulate CV data at various values of arterial capacitance and calculate EW in order to compare to the EW calculated in step 2. The model consists of a four element arterial model, a two element left ventricular model and a three element aortic valve model. The model provides freedom to change arterial capacitance while constraining mean arterial pressure (MAP) and cardiac output (CO) to within 2% by changing the heart rate. Results indicate that as arterial capacitance increases, EW asymptotically approaches a maximum slightly above the operating point and EW decreases as arterial capacitance decreases. It's concluded that Ea is maintained to provide near maximal EW; however, regulating arterial pressure and flow throughout a beat seem to take precedence.

ARTERIAL ELASTANCE IN THE MAXIMIZATION OF EXTERNAL WORK TRANSFER

Mark J. Schroeder

Introduction

A better understanding of ventricular-arterial coupling may help provide better preventive measures and improved clinical care for CV patients and may lead to better equipment for astronauts and pilots in-flight. Of the various methods used to investigate ventricular-arterial coupling, one focuses on the maximization of EW with respect to eff. Ea (9). EW is a popular method for viewing coupling since it involves both aortic pressure and flow, two important components of coupling.

The widely used effective arterial elastance (eff. Ea) described by Sunagawa et al. (8,9) is used as an index to describe the elastic state of the arterial system. This index was devised to describe the coupling of the arterial system to the ventricle in order to predict the equilibrium stroke volume. Sunagawa et al. (8,9) have shown this to be a good method for both this purpose and for determining the arterial system's elastic state that provides maximal external work (EW) transfer. They have shown that maximal EW transfer occurs when eff. Ea is equal to the end-systolic elastance (Ees). However, Sunagawa's eff. Ea differs from the physiological meaning of arterial elastance (Ea). Actually, "effective arterial elastance changes more with changes in physical arterial resistance than with changes in physical arterial compliance" (8). Therefore, in order to extend Sunagawa's work of maximizing external work with respect to eff. Ea, both types of elastance will be described and contrasted. Then, a computer simulation will be used to test the effects that changing physiological Ea has on EW transfer.

Methods

Four clinically normal male baboons (*Papio anubis)*, 5-7 years old, 18-28 kg, were used for this study. All baboons were cared for under the provisions of the *Guide for the Care and Use of Laboratory Animals* (NIH No. 80-23). All subjects were housed in large primate cages, maintained on 12-h circadian cycles and allowed food and water *ad libitum*. Diet consisted of

Purina monkey chow biscuits and fresh fruit. The animals were instrumented under general anesthesia as described previously (2).This study was reviewed by the Armstrong Laboratory Animal Care and Use Committee and was found to be in accordance with all federal guidelines and Air force regulations that govern the use of non-human primates in biomedical research.

A Millar micromanometer 3-F double-tip catheter (Millar Instruments Inc, Houston, TX) was used for simultaneously measuring left ventricular and aortic pressures. Millar transducers received excitation through Gould bridge amplifiers (Model 13-4615-30, Gould Inc., Houston, TX). The catheter was dynamically calibrated in steps of 50 from 0- 200 mm Hg. The catheter was inserted via a vascular access port. A Zepeda EMF flow probe measured aortic blood flow.

Centrifuge

Prior to each centrifuge protocol, the baboons were sedated with Ketamine (5-10 mg/kg), and placed in special confinement chairs. The chair was bolted to the animal end of the Brooks AFB centrifuge arm. A series of rapid onset/offset (5 G/s) episodes were then performed to differing levels of gravitational stress ranging from 2 to 9 +Gz (peak G duration = 10s). The time between each G run was variable and was based on a return of heart rate to near-baseline (within 5%) values (6).

KC-135

Parabolic flights were performed using a modified KC-135A aircraft operated by NASA JSC's Reduced Gravity Program. The aircraft was flown from Ellington Field, Houston, TX to Kelly AFB, San Antonio, TX from which our flights were staged. A parabolic flight consists of approximately 40 in-flight "parabolas" in which 0G and 2G conditions are reached intermittently for approximately 30 seconds each. The flight protocol met NASA JSC committee approvals and flight safety requirements. All necessary instrumentation from our laboratory was mounted and cabled in specially designed safety racks for mounting in the KC-135. An independent power conditioner was flown as well (BEST, Model 3.1 kVA).

The physiological signals were continuously monitored and recorded on stripchart and VHS tape for both the centrifuge and parabolic flight.

Background

Sunagawa et al. (8,9) proposed to describe coupling of the arterial system with the left ventricle on the pressure-volume plane by assuming both to be elastic chambers (Figure 1a). By characterizing the properties of the arterial system in terms of the end-systolic pressure and ejected volume, they could graphically determine the equilibrium stroke volume. Figure 1b is a graphic representation of this ventricular-arterial coupling method plotted on the left ventricular pressure-volume plane.

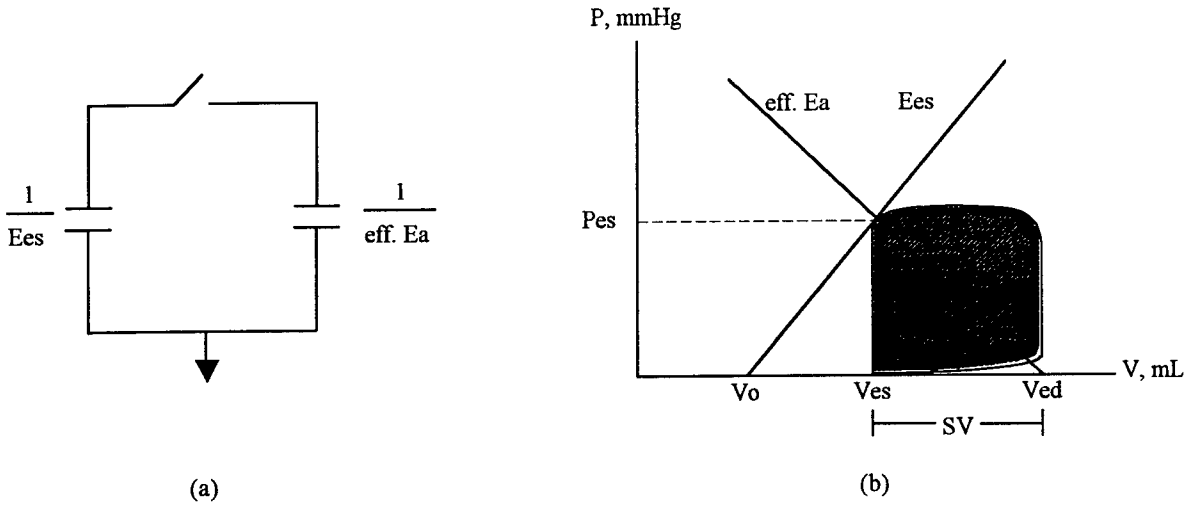


Figure 1. a) Sunagawa's proposed coupling model of the left ventricle and arterial system. Ees is end-systolic elastance of the left ventricle and eff. Ea is effective arterial elastance. b) Ventricular-arterial coupling method plotted on the left ventricular pressure-volume plane. Pes is the end-systolic pressure, EW is external work, Vo is the volume left in the left ventricle if ejecting to the atmosphere, SV is stroke volume, and Ves and Ved are the volumes of the left ventricle at end-systole and end-diastole, respectively. (From Sunagawa, Maughan, and Sagawa 1985)

Sunagawa et al. (8) defined eff. Ea to be:

$$\text{eff.Ea} = \frac{\text{Pes}}{\text{SV}} \approx \frac{R_T}{T} \qquad \text{Eq. 1}$$

where R_T is the total arterial resistance and T is the cycle length of the cardiac contraction. They showed that EW transfer is maximized when eff. Ea = Ees (9). However, since eff. Ea is an index which actually changes more with changes in total peripheral resistance (TPR), it is still unclear how *physiological* Ea affects EW transfer.

Eff. Ea vs. Ea

The difference between the physiological meaning of arterial elastance and Sunagawa's effective arterial elastance can be shown graphically in Figure 2. The curve on the left describes

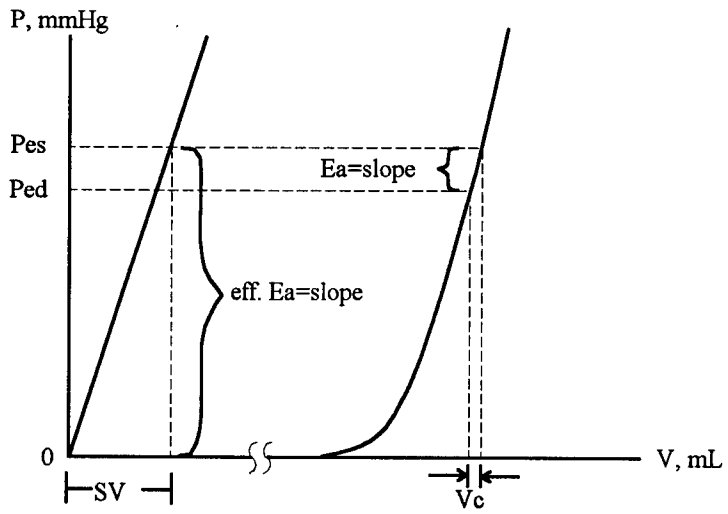


Figure 2. Ea and eff. Ea. Ped is the end-diastolic pressure of the previous beat and Vc is the volume which entered the capacitance (aorta). The left and right portions of the abscissa are not necessarily scaled the same.

Sunagawa's eff. Ea and adheres to Eq. 1. The curve on the right represents the familiar systemic arterial volume-pressure curve (3), which is used to describe Ea as:

$$Ea = \frac{Pes - Ped}{Vc} \qquad \text{(Eq. 2)}$$

The elastance of the arterial wall can be described as the change in pressure due to a change in volume (change in pressure/change in volume). Equation 2 adheres to this definition since Vc is the change in blood volume of the arterial system (capacitance) and Pes-Ped is the pressure change due to that volume. On the other hand, Equation 1 uses the entire aortic pressure, Pes, and the entire SV. However, to estimate arterial elastance for a pressure change from zero to Pes, one would need to use the entire volume that created the large pressure change, not just the stroke volume. This is why effective arterial elastance is an index and not the actual physiological arterial elastance.

Cardiovascular Model

In order to study the effects that changing Ea has on EW, a ventricular-arterial coupling model was devised (Figure 3). The left ventricular model, by Berger and Li (1), consists of a time-dependent source, Elv(t), and a pressure-dependent source resistance, Rlv(P). The aortic valve model consists of an inductor, Ls, a resistor, Rs, and a diode. The diode represents the opening and closing of the aortic valve and Rs and Ls represent the blood inertia and valve resistance seen between the aortic and left ventricular pressure transducers. The arterial model is made up of a parallel inductor, Lp, and resistor, Rp, in series with a parallel capacitor, C, and resistor, R (all of which are assumed constant throughout a beat). R and C (C = 1/Ea) represent total peripheral resistance and systemic arterial compliance (SAC), respectively; Lp accounts for blood inertia, and Rp represents the high frequency impedance. Plv(t) is left ventricular pressure, Q(t) is aortic blood flow, and Pa(t) is aortic pressure.

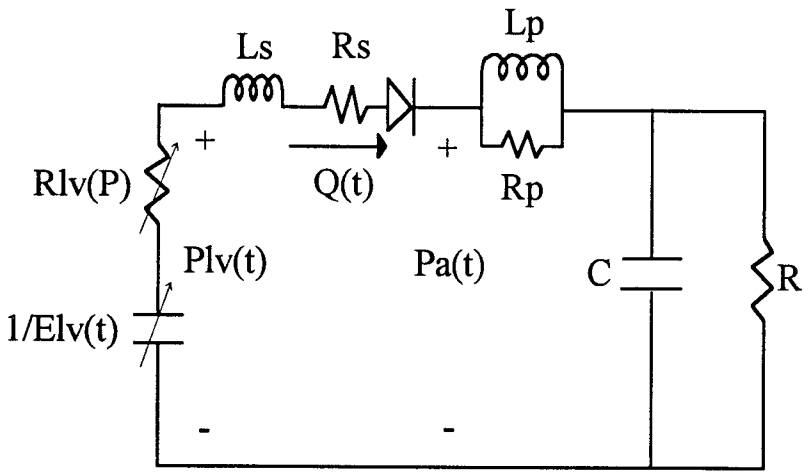


Figure 3. Electrical circuit model of ventricular-arterial coupling.

Parameter Estimation

Parameters R and C in Figure 3 were first estimated using the two-element model technique described by Self et al. (6). In general, this technique involves simultaneously solving the following two equations:

$$\int_{t0}^{t1} Q(t)dt = C[Pa(t1) - Pa(t0)] + \frac{1}{R}\int_{t0}^{t1} Pa(t)dt \qquad \text{(Eq. 3a)}$$

$$\int_{t1}^{t2} Q(t)dt = C[Pa(t2) - Pa(t1)] + \frac{1}{R}\int_{t1}^{t2} Pa(t)dt \qquad \text{(Eq. 3b)}$$

where Q is blood flow, Pa is aortic pressure, R is TPR, and C is SAC. The points of integration are described in Figure 4, where t1 occurs at peak aortic pressure.

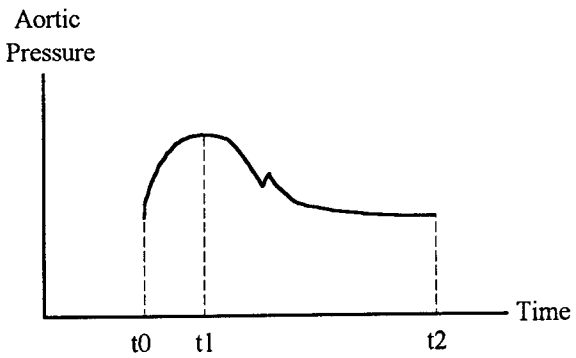


Figure 4. Aortic pressure.

However, sometimes unrealistic results for C were encountered, depending on where peak aortic pressure occurred. To improve the estimation technique, t1 was forced to be at least 25% of the entire beat length from the point t0. If the peak aortic pressure occurred after 25% of the beat length, t1 was chosen at the peak aortic pressure point. This method provided more consistent results for C, and R was only slightly affected. This technique was used throughout this work to estimate R and to provide an initial estimate for C.

Rp and Lp were given initial starting values. Using the actual flow as input and solving the ordinary differential equations of the arterial model, aortic pressure was simulated. A Matlab™ optimization routine was used to minimize the sum-squared error between the physiological aortic pressure and the simulated aortic pressure by changing Rp, Lp, and C (wave reflection was ignored). R was left as previously estimated since the difference in R estimates between the two-element and four-element models was negligible. In contrast, C was affected more than R. Thus Rp, Lp, and C were given initial starting values, the optimization procedure was carried out and the estimates for Rp, Lp, R, and C were stored. Estimations of the remaining model parameters are described in the appendix.

Computer Simulation

Once the model parameters were estimated, they were fed into a computer simulation that produced data similar to the physiological data. The computer simulation consisted of solving a

set of four linked first-order ordinary differential equations (shown in the appendix) which describe the cardiovascular model in Figure 3. This was done within the Matlab™ environment using a Matlab™ m-file that solves ordinary differential equations using the Runge-Kutta method. The estimated Elv(t) served as input to the circuit and produced an output of one beat. A continuous loop was created so that simulated beats could run to a steady-state condition defined by two successive end-diastolic aortic pressures with less than 0.1 mmHg difference.

The last complete simulated beat that contained a complete systolic phase followed by a complete diastolic phase - the steady-state beat - was then used to calculate EW. The method used to calculate EW was to determine the integral of left ventricular pressure with respect to ejected volume. This was performed using Eq. 4, where Plv is left ventricular pressure, V is ejected volume, and 'a' and 'b' are the points of beginning ejection and end ejection, respectively. The left ventricular filling pressure was ignored in the calculation of EW due to its inability to be calculated; however, its effect on the total EW was small and would have been similar for each beat. The simulation process was then repeated at different capacitances. EW was calculated for each steady-state simulated beat in order to compare it to the first simulated result.

$$EW = \sum_{n=a}^{n=b} \int_{Vn}^{Vn+1} Plv \, dV \qquad \text{(Eq. 4)}$$

Constraints were set according to O'Rourke et al. stating that ventricular-arterial coupling concerns "steady flow of blood through the body's capillaries" during systole and "perfusion of the heart as an organ" during diastole (5). In other words, cardiac output (CO) and mean arterial pressure (MAP) must be maintained to provide effective ventricular-arterial coupling. Therefore, mean arterial pressure (MAP) and cardiac output (CO) of simulated beats at new capacitances were required to remain within a 2% tolerance of the original simulated beat with normalized capacitance (Cn) equal to one, where Cn = new Ca / actual Ca. The objective was both to maintain physiological conditions and to keep as many variables as possible constant. This was done by scaling the sampling rate of the input, Elv, to provide a new heart rate when necessary.

For example, if Cn = .4 and MAP dropped below tolerance, then heart rate was increased to raise MAP. This was repeated until the tolerance was satisfied; then a new capacitance would be run. If, however, either or both MAP and CO tolerances were unable to be met, then no more capacitances were run. Cn was changed in increments and decrements of .2 from Cn = 1. This was continued until either one of the tolerances could not be met or a set limit of Cn=.2 or Cn = 3 was reached.

Results

Results were obtained from a total of 28 beats from 4 animals under various gravitational conditions. Two beats were analyzed from each animal in each of the following G-conditions: 0Gz, 1.4Gz, ~2Gz, and 3Gz. However, one animal had no flow signal during the parabolic flight; therefore, the corresponding 0Gz and 2Gz beats were unobtainable.

Figure 5 shows the normalized EW mean and standard deviation of all beats at each value of normalized capacitance. The plot indicates a slight increasing trend in EW as capacitance increases from a small value. For the constraints given, it also shows that the operating capacitance of the cardiovascular system is nearly optimal as far as maximization of EW is concerned. More importantly though, this work provides evidence that ventricular-arterial coupling does not maximize external work at some unique value of compliance; but implies that EW asymptotically approaches a maximum as compliance approaches infinity. Intuitively, this seems to make sense since when C is large, energy isn't expended on creating pressure.

Figure 6 shows typical PV loops for a beat from actual data and for simulations at various capacitances. As expected, a low capacitance (highly elastic) results in a large pulse pressure with less EW and a large capacitance results in a small pulse pressure with more volume ejected and more EW output.

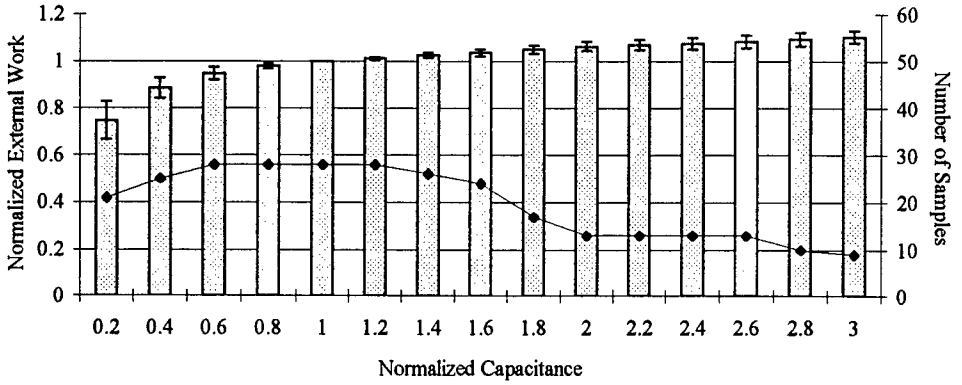


Figure 5: The bars indicate the normalized EW at values of normalized arterial capacitance from .2 to 3 in increments of .2. The diamond symbol / line shows the number of samples at the particular capacitance value. The error bars indicate the standard deviation of the samples.

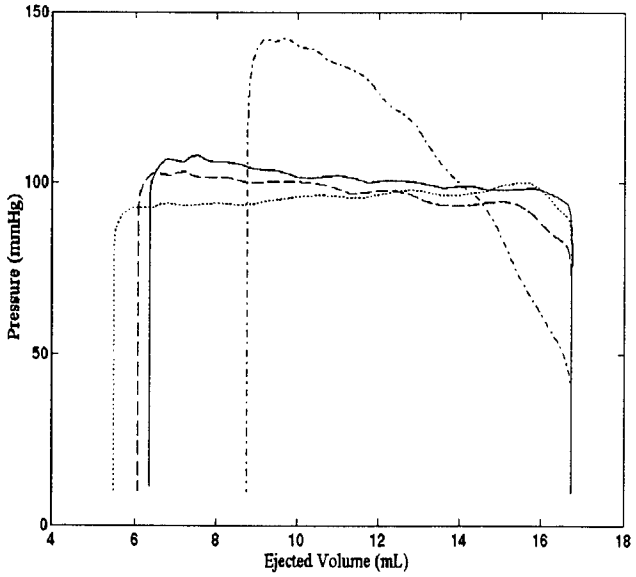


Figure 6. PV loops for various capacitances. The solid line (-) represents physiological data. The remaining PV loops are from the corresponding simulated data. For Cn = 1 (--), EW = 1032 mmHg*ml. For Cn = .2 (-·-·), EW = 875 mmHg*ml. For Cn = 2.2 (···), EW = 1072 mmHg*ml.

Discussion

The previous sections have shown just how eff. Ea and Ea differ. Since EW transfer tends to increase as capacitance increases, it can be discerned that EW transfer is not maximized at the unique point at which Ea = Ees. Also, since eff. Ea changes more with changes in TPR, there seems to be no logical connection between eff. Ea and Ea. However, Sunagawa's analytical ventricular-arterial matching utilizes a PV loop with a constant ejection pressure. In essence, this assumes an infinite capacitance (which we have shown does maximize EW transfer) and gives way to the other main arterial component, TPR. This may partially explain why eff. Ea is more sensitive to changes in TPR than to changes in Ea.

More importantly, however, the technique for evaluating eff. Ea, seems to be inherently more sensitive to changes in TPR than to changes in Ea. The following discussion may help describe why effective arterial elastance can change more with a change in arterial resistance than with a change in arterial compliance. First, if the arterial system were purely capacitive (infinite resistance), the PV loop would look similar to the solid PV loop in Figure 7.

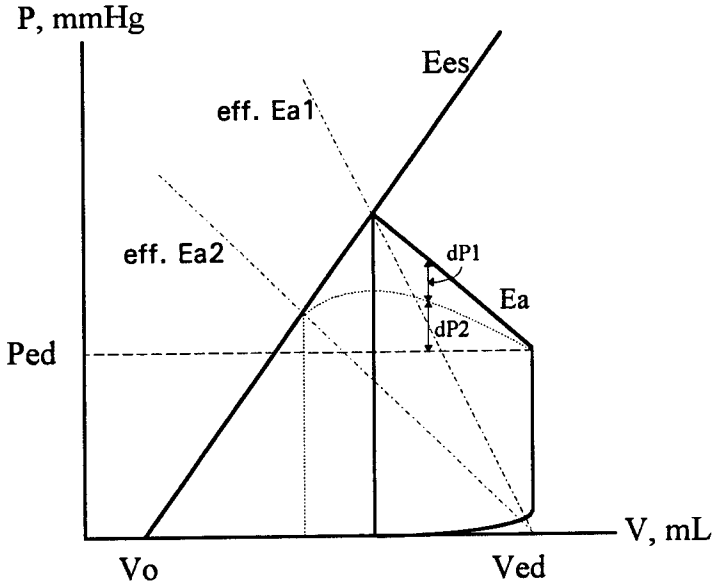


Figure 7. Hypothetical changes in eff. Ea as TPR changes from infinite to finite.

This can be explained by assuming the elastance of the arterial system to be constant; and, therefore, aortic and left ventricular pressures would increase linearly with respect to an increase in arterial volume. In other words, an increase in arterial volume will cause pressure to increase proportionately to Ea (Ea = change of pressure / change of volume). This results in the arterial elastance index of eff. Ea1. However, if arterial resistance is finite, some of the blood volume in the capacitor can flow through the arterioles, decreasing the pressure in the capacitor (aorta) and the left ventricle. This new, more realistic, pressure-volume loop is represented by the dotted PV loop. One can see that even though the actual arterial elastance has not changed, the arterial elastance index changed dramatically, from eff. Ea1 to eff. Ea2. The pressure differences dP1 and dP2 can be equated to Ea*Vr and Ea*Vc, respectively, where Vr and Vc represent the volume through the resistor and in the capacitor, respectively. The volume moved through the finite resistance causes a drop in pressure, dP1, from the pressure that would have existed had all the volume remained in the capacitor. This helps explain the difference between physiological and effective arterial elastance.

Since results indicate that EW is only slightly sensitive to arterial capacitance, one might ask what role the arterial capacitance plays in the cardiovascular system. Recalling that a higher capacitance will allow for an increase in stroke volume without a large increase in pressure, this situation might be ideal during normal conditions of rest or in a microgravity situation. Also, since a lower capacitance creates a higher pressure and a decreased stroke volume, this may be ideal in a prolonged upright or high gravitational condition. However, this usually seems to occur at the expense of external work.

The variable most affected by changing capacitance was pulse pressure. The question now turns to the importance of pulse pressure in the cardiovascular system. First, consider the effect of no capacitance in the arterial system. The main effect of this would be a large pulse pressure during ejection and no pressure during diastole due to the lack of a discharging capacitor. In this case, there would not be constant perfusion of the organs; and, more importantly, there would be no blood flow to the coronary arteries during diastole. Remembering that coupling during diastole concerns perfusion of the heart, this would be a life threatening situation.

Second, a large arterial capacitance would result in pooling of blood in the direction of gravity in the arterial system. Therefore, a person standing upright would experience severe pooling of blood in the feet and lower legs. This situation would cause low blood pressure in the head, potentially causing the person to pass out due to a lack of brain perfusion. Also, if the arterial system has a greater capacity for storing blood, then the venous side would have an equally less amount of blood from which to draw. If the venous reserve drops too much, the cardiovascular system will lose some of its control and buffering mechanisms which could result in the inability to maintain basal metabolic flow requirements of the body.

These two cases provide reasons for SAC to exist in the systemic arterial vasculature and for it to be tightly regulated. However, the cardiovascular system has other features used to aid in control and regulation of aortic pressure and blood flow. These include heart rate, TPR, strength of contraction of the left ventricle, and the sensory mechanisms that help control these variables. Acting together, these variables confound the issue of the importance of capacitance in the cardiovascular system. However, they do provide the body with a highly regulated system that is not only necessary to maintain body metabolism, but readily adjusts to most situations.

Though MAP and CO were used as constraints, using external work to evaluate optimal ventricular-arterial coupling does not seem to be a good method. External work transfer on a per beat basis gives no consideration to how the arterial system responds after ejection. For example, if external work is maximized, but the coronary arteries or brain are not being perfused, then ventricular-arterial coupling is not being maximized. Therefore, due to the importance of the diastolic phase, external work alone should not be used to define optimal ventricular-arterial coupling.

Appendix

Model state equations

The following shows the cardiovascular model's state equations in matrix form where x_1 is left ventricular effective volume, x_2 is blood flow, x_3 is the aortic pressure after the diode, and x_4 is the pressure seen between the parallel inductor/resistor and the parallel capacitor/resistor.

The first matrix equation was used when blood flow was greater than zero. The second matrix equation was used when blood flow was equal to zero. These equations were solved with a Matlab™ m-file using the Runge-Kutta method.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ \frac{Elv}{Ls} & \frac{-(Rlv+Rs)}{Ls} & \frac{-1}{Ls} & 0 \\ \frac{Rp \cdot Elv}{Ls} & \frac{1}{C} - \frac{Rp(Rlv+Rs)}{Ls} & -\left(\frac{Rp}{Ls}+\frac{Rp}{Lp}\right) & \left(\frac{Rp}{Lp}-\frac{1}{R \cdot C}\right) \\ 0 & \frac{1}{C} & 0 & \frac{-1}{R \cdot C} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{-Rp}{Lp} & \frac{Rp}{Lp} - \frac{1}{R \cdot C} \\ 0 & 0 & 0 & \frac{-1}{R \cdot C} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \qquad \text{(Eq. 5)}$$

Model Parameter Estimation

The isovolumic maximum pressure (Pmax) was predicted using a method by Sunagawa et al. (10) which uses the left ventricular isovolumic pressure curves to fit a sinusoid. A line drawn from Pmax tangent to the upper left corner of the PV loop was used to estimate end-systolic elastance (Ees) (11). The left ventricular effective volume was then estimated by Pmax/Ees.

Elv(t) was estimated using an equation derived by Berger and Li (1):

$$Elv(t) = \frac{Plv(t)}{(EDV - \int Q(t)dt - Vo)(1 - k \cdot Q(t))} \qquad \text{(Eq. 6)}$$

where Plv(t) is left ventricular pressure, EDV is end-diastolic volume, Q(t) is flow rate out of the left ventricle, Vo is the volume at which the ventricle cannot create a pressure, and k is a constant

that relates pressure to the left ventricular systolic resistance. Alternately, EDV-Vo is equal to effective volume, which was estimated as described above. Shroff et al. (7) have assumed k to be 0.0015 s/mL, but a method was developed to estimate k and is described next.

The source resistance, Rlv(P), is linearly related to the isovolumic pressure, Po(t) (7). Here, isovolumic pressure is Elv(t)*Veff(t), where Veff(t) is the effective volume at time t, rather than Elv(t)*Veff(0) as described in the work by Sunagawa et al. (10). Shroff et al. (7) described left ventricular pressure as Po(t)*(1-k*Q(t)), where Q(t) is the flow rate and k is a constant of proportionality that relates the isovolumic pressure to the source resistance. Therefore, one can solve for the constant

$$k = \frac{Po(t) - Plv(t)}{Po(t) * Q(t)} \qquad \text{(Eq. 7)}$$

Since Plv(t) and Q(t) are known from the physiological data, only Po(t) needs to be determined. Po(t) equals Elv(t)*Veff(t); unfortunately, k is needed to estimate the ejecting Elv(t). Therefore, ejecting Elv(t) was estimated by using the non-ejecting Elv(t). This was done by dividing the non-ejecting isovolumic pressure curve (obtained while estimating Pmax) by Veff(0). The non-ejecting Elv(t) was multiplied by Veff(t) to obtain an estimate for Po(t). Solving for k produced an estimate at each value of time. A statistical method, the mean of the shorth (4), was used to estimate k.

Ls and Rs were initially estimated by solving Eq. (8) in the same manner that R and C were initially estimated (two equations, two unknowns), where i is blood flow velocity, Pa is aortic pressure, and Ls and Rs are a series inductance and resistance, respectively. Then, an optimization routine was used to refine the estimates.

$$Plv - Pa = Ls \cdot \frac{di}{dt} + Rs \cdot i \qquad \text{(Eq. 8)}$$

Model Limitations

Although the model reproduces physiological beats well, there are limitations. For example, wave reflections are ignored in the optimization routine and coronary flow has not been included due to difficulty of implementation. The method of estimating the left ventricular source resistance constant of proportionality, k, should either be improved or an uncertainty analysis should be performed to test the sensitivity of the circuit to using a constant k. The effective volume estimation depends on the estimation of Pmax. The uncertainty in predicting Pmax and Ees also produces uncertainty in predicting the effective volume. However, since Elv is actually calculated using the estimated values for k and effective volume, the resultant left ventricle pressure produced in the simulation is close to the actual data as long as the rest of the model parameters are estimated accurately.

Acknowledgments

The authors wish to thank the Physiological Research team at Brooks A.F.B. for the use of their time, knowledge and resources. The views expressed herein are the private views of the authors and not to be construed as representing those of the Department of Defense or Departments of the Army or Air Force. This work was supported in part by the Air Force Office of Scientific Research Summer Research Program.

References:

1. Berger, D. S., and J. Li. Temporal relationship between left ventricular and arterial system elastances. *IEEE Trans. Biomed. Eng.* 39(4): 404-410, 1992.

2. Fanton, J. W., R. D. Latham, C. W. White, D. A. Self, C. P. Kingsley, and C. D. DiCarlo. A conscious baboon model for evaluation of hemodynamics in altered gravity. *J. Invest. Surg.* 6:451-9, 1993.

3. Guyton, A.C. Vascular distensibility and functions of the arterial and venous systems. In: *Textbook of Medical Physiology*, 8th Ed., Philadelphia: W.B. Saunders Company, 1991, p. 160.

4. Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel. *Robust statistics: the approach based on influence functions.* New York: John Wiley, 1986.

5. O'Rourke, F., A. P. Avolio, and W. W. Nichols. Left ventricular-systemic arterial coupling in humans and strategies to improve coupling in disease states. In: *Ventricular/Vascular Coupling,* edited by F. C. P. Yin. New York: Springer-Verlag Company, 1987, p. 1-19.

6. Self, D. A., D. L. Ewert, R. D. Swope, R. P. Crisman, and R. D. Latham. Beat-to-beat determination of peripheral resistance and arterial compliance during +Gz centrifugation. *Aviat. Space Environ. Med.* 65: 396-403, 1994.

7. Shroff, S. G., J. Janicki, and K. Weber. Evidence and quantitation of left ventricular systolic resistance. *Am. J. Physiol.* 249: H358-H370, 1985.

8. Sunagawa, K., W. L. Maughan, D. Burkoff, K. Sagawa. Left ventricular interaction with arterial load studied in isolated canine ventricle. *Am. J. Physiol.,* 245: H733-H780, 1983.

9. Sunagawa, K., W. L. Maughan, and K. Sagawa. Optimal arterial resistance for the maximal stroke work studied in isolated canine left ventricle. *Circ. Res.* 56:586-595, 1985.

10. Sunagawa, K., A. Yamada, Y. Senda, Y. Kikuchi, M. Nakamura, T. Shibahara, and Y. Nose. Estimation of the hydromotive source pressure from ejecting beats of the left ventricle. *IEEE Trans. on Biomed. Eng.* 27(6), 299-305, 1980.

11. Takeuchi, M., Y. Igarashi, S. Tomimoto, M. Odake, T. Hayashi, T. Tsukamoto, K. Hata, H. Takaoka, and H. Fukuzaki. Single-beat estimation of the slope of the end-systolic pressure-volume relation in the human left ventricle. *Circulation* 83:202-212, 1991.

Tactile Perception in a Virtual Environment

Katherine M. Specht
Graduate Student
Department of Speech & Hearing Science

The Ohio State University
1070 Carmack Road
Columbus, Ohio 43220

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

Tactile Perception
in a Virtual Environment

Katherine M. Specht

Abstract

Previous tactile pattern perception research suggests that moving the fingertip relative to a fixed pattern is superior to other presentation modes, such as moving the pattern across a stationary fingertip. In the present study, several experiments were conducted to determine if performance in other presentation modes could be facilitated or degraded through manipulation of factors such as display size, repeated looks, and scan directions. Experiments 1 and 2 of the present study were designed to determine how far a field of view can be reduced before performance was degraded.

Experiment 3 of the present study allowed subjects to have repeated "looks" at the stimulus during static and passive scan presentation modes. This experiment was conducted in order to determine whether the ability to repeat the stimulus was the aspect of the haptic mode that had led to the performance levels observed by Weisenberger and Hasser (1994). Experiment 4 manipulated the scan direction during the passive scan presentation to determine whether advantages found in the haptic presentation could be attributed to the capability of scanning in any direction (right, left, up, down, and any oblique). Eight different scan directions were permitted.

Data from these experiments suggest that display size can be reduced to as few as 4-pins before performance is degraded. In addition, the ability of the subject to choose to repeat a stimulus and to scan in multiple directions actually facilitates performance in the static and passive scan presentation modes. Utilization of these strategies in the static and passive scan modes appears to aid processing of even complex patterns to a point which approximates haptic exploration.

The results of the present study suggest that future scan displays should continue to utilize horizontal scan directions to encourage optimal pattern identification performance. These data also imply that future displays can be constructed for practical applications in telerobotics and virtual reality research, particularly in the development of wearable displays.

Tactile Perception in a Virtual Environment

Katherine M. Specht

Introduction

Few studies in tactile perception have addressed the question of how pattern identification is affected by reducing the "field of view" of the fingertip (i.e. the area of skin surface to which a pattern is presented). One goal of the present research was to determine how far the field of view can be restricted before pattern identification is degraded. A second goal of this research was to examine whether permitting multiple opportunities to view a pattern and/or multiple scan directions from which to view a pattern in the static and passive scan presentation modes facilitate pattern identification.

Past research has suggested that tactile information distributed to spatially separate sites on the fingertip may interfere with the identification of a target pattern, a phenomenon known as masking (Weisenberger 1981; Weisenberger & Craig, 1982). Similarly, in a detection task, Sherrick (1964) investigated the influence of masking on target detection by presenting a target on one fingertip and a masker on another fingertip. Sherrick found that the amount of masking is a function of the spatial distance between the masker and target stimuli. These data suggest that information from different fingertips can be integrated into a single field of view.

Expanding the field of view to spatially distributed areas of the body has also been investigated by researchers such as Lappin & Foulke (1973) and Hill (1974). Lappin and Foulke (1973) studied tactile perception with braille letters simultaneously presented to one, two, or four fingers, on the same hand or on opposite hands. Subjects were most adept at the task when two fingers of different hands were utilized, but error rates were lowest when only one finger was employed. Hill's study was similar, although subjects were asked to identify alphabetic characters instead of braille letters. The results from Hill's study support Lappin and Foulke's findings that the fewest errors were found in the single finger condition. These findings suggest that tactile perception from spatially distributed sites can be integrated to a limited

degree, but may interfere with feature information that is important for correct identification.

Research by Loomis (1974) indicated that performance with a visual aid for the blind, called the TVSS, was better for slit-scan modes than for static and passive modes. Loomis argued that the reason for the superior performance was because the slit-scan modes represented phase information in addition to spatial information. In 1980, Loomis again investigated differences in presentation modes that were analogous to reductions in the tactile field of view. Stimulus patterns were presented to a single fingertip in full-field of view mode as well as in slit-scan mode, where a moving slit was moved across a stationary pattern. Results from this study indicated that the slit-scan condition was better for identification than the full-field condition when the pattern was small and dense. This superiority in performance was attributed to limitations in the tactile system's spatial resolution abilities when its low-pass sensing properties are challenged.

However, Craig (1980) also asked the question of how manipulating the presentation mode would affect tactile perception. Several presentation modes were targeted, including static mode (no pattern movement), scan mode (pattern moving right to left across a stationary site), slit-scan mode (moving slit across the pattern allowing limited view at any given time), continuous sequential mode (successive elements of the pattern being activated, and staying activated once on), and a discontinuous mode (pattern elements were activated, but turned off as the next element became activated). All presentations were made to a single finger. The results of Craig's study did not support Loomis' results. Craig found that the full-field mode was the best, and that reduced view presentation modes resulted in poorer identification. Loomis argued that Craig's stimuli must not have taxed the spatial resolution properties of the tactile system, whereas Loomis's 1974 experiment with smaller patterns had taxed these properties of the tactile system. Loomis conducted a follow-up investigation and found that when larger patterns were utilized, his results approximated Craig's findings.

Craig (1981) explored the pattern recognition of individuals using the Optacon display, a reading aid for blind persons comprised of stimulators that vibrate at 230 Hz when contacted by the finger. The Optacon display was designed to present a moving pattern to a stationary finger. Results from this study indicated that pattern recognition was better when there was no movement of the finger or the pattern

than for any of the conditions involving movement (slit-scan, scan, continuous, discontinuous modes).

In the present study, we were interested in investigating the spatial and temporal information obtained through a tactile display. The question arises whether decreasing the spatial size of the field of view can be compensated via temporal integration of pattern elements encountered sequentially. Likewise, it is interesting to question whether maximizing the temporal properties of the display will affect the spatial sensing properties of the tactile system. These question are important from a theorectical standpoint because results may provide more information on how patterns are processed by the tactile system. The questions are also important for practical applications in telerobotics and virtual reality research, particularly in the development of wearable displays.

These questions can be addressed through the manipulation of factors such as presentation mode, duration of view, and relative field of view. For instance, reducing the field of view is anticipated to result in increased importance of kinesthetic movement to determine spatial aspects of the stimulus. Also, reducing the field of view is likely to maximize the observer's reliance on temporal properties of the stimulus. Experiments 1 and 2 of the present study were designed to determine how far a field of view of a given display can be reduced while still allowing pattern identification.

Other research has been conducted with mobile tactile displays instead of the former stationary display with moving patterns. Weisenberger and Hasser (1994) noted that in normal touch, maximal information is gained when the finger is moved relative to the surface being sensed, creating a shearing motion. They suggested that Craig's results may have been attributable to the fact that the finger was held stationary, and patterns moved across it, thus eliminating the kinesthetic cues provided in active haptic movement of the fingertip. Weisenberger and Hasser investigated pattern identification in the haptic presentation mode (active movement of a display in relation to a surface) versus that in a passive scan mode and a static mode (similar to those studied by Craig, 1980). Under the haptic condition, kinesthetic movement of the finger in relation to the pattern was anticipated to have a facilitating effect on pattern identification. Using a 5 x 6 array of shape memory alloy actuators, Weisenberger and Hasser found no difference in performance for different presentation modes when the patterns consisted of simple geometric shapes. However, the results showed that the haptic presentation mode led to slightly better

identification for a set of letters from the alphabet [A, X, B, R, G, C, D, O, Q]. These data suggest that the movement of the fingertip relative to the pattern does facilitate identification when patterns are complex.

Confusion matrices for these data revealed similar confusions across the three presentation modes. Confusions are most often made between the letters "X" and "K", "R" and "K", and "B" and "R". There are several possible explanations for the superior performance of subjects in the haptic mode. First, subjects were not constrained by duration or rate of scan in the haptic presentation. Second, haptic presentation allowed the subjects to "look" at the stimulus as often as desired before responding, whereas only one "look" was presented in passive scan and static modes. Third, subjects were able to scan the stimulus from any direction during haptic scan, whereas only a right-to-left scan was permitted in the passive scan mode. Finally, it may be that the kinesthetic feedback provided in the haptic mode may have improved performance. Further experiments are needed to determine the influence of such variables as duration, number of looks, and scan direction on subject performance. A second goal of the present study was to address this question.

Experiment 3 of the present study allowed subjects to have repeated "looks" at the stimulus during static and passive scan presentation modes, to determine whether the ability to repeat the stimulus was the aspect of the haptic mode that had led to the performance levels observed by Weisenberger and Hasser (1994). Experiment 4 addressed subjects' control of the stimulus scan direction during the passive scan presentation to determine whether advantages found in the haptic presentation could be attributed to the capability of scanning in any direction (right, left, up, down, and any oblique). Patterns identical to those used in the Weisenberger and Hasser study were chosen for all experiments in the present study (Figure 1a and 1b).

Experiment 1

Methodology

The goal of Experiments 1 and 2 was to determine the reductions in the tactile field of view had deleterious effects on performance. Experiment 1 employed two sets of patterns, shapes and letters.

The stimuli in these blocks were all larger than the field of view provided by the stimulator, so active haptic manipulation was necessary for accurate identification of the patterns.

Three subjects (2 female, 1 male) participated in this study. Two subjects (CJH and JMW) had previous experience with tactile experiments similar to the present one. One subject (KMS) had no previous experience with tactile experiment, but received several weeks of practice in tactile pattern perception tasks prior to data collection.

The display used in this experiment was a 3 x 3 array of shape memory alloy (SMA) stimulators housed in a plastic casing. This display, manufactured by TiNi Alloy Corporation (A. Johnson, 1991), is composed of 9 metal stimulator elements, connected to a SMA wire. The SMA wires when heated by input current to the display, contract, forcing the stimulators directly upward. Activation of the stimulators causes them to protrude through holes in the display. As the input current stops and the SMA wire cools, each element returns to its resting state below the plastic casing of the display. The nine elements are spaced approximately 3.0 mm apart. All nine elements could be activated during the first experiment.

A CalComp Drawing Board II digitizing pad was utilized for the haptic presentation mode. The 3 x 3 display was mounted onto a mouse pointer to permit its mobility while allowing the computer to register the display's coordinates along the digitizing pad. Movement of the SMA display, then, results in the perception of a stimulus pattern on the pad. All stimuli for the present experiment were presented in the haptic scan mode.

Both the 3 x 3 display and digitizing pad were interfaced to an 80386-based PC computer via its serial port. C Software controlled stimulus presentation, response collection, and data storage. Duty cycle was fixed at 50% for the present experiment, and stimulus duration was controlled by each individual subject based on duration of active scanning.

The stimuli used in this experiment were chosen for comparability to data from a similar experiment by Weisenberger and Hasser (1994) using a 5 x 6 SMA array. The first set of patterns were 8 geometric shapes from a combination of horizontal, vertical, and diagonal lines. The second set of patterns were 10 "letters" [A, X, B, R, G, C, D, O, Q]. Previous data suggested that items from the set of "letters" were significantly more difficult for subjects to identify than those from the set of shapes when utilizing the 5 x 6

array.

Subjects were instructed to rest their index finger on the SMA display and then to move the display across digitizing pad to identify a fixed pattern. Scanning had to be done slowly for the percept actually to be felt by the subject, due to relatively low system bandwidth. An icon of each pattern was displayed on the PC computer screen during all trials. Response was made on the PC keyboard and trial by trial feedback was delivered to the subject via the PC monitor.

Sixteen blocks of trials were completed by each subject (40-trial blocks for shapes, 50-trial blocks for letters). All blocks of shape patterns were completed first; followed by the blocks of letter patterns. Stimulus frequency was varied across blocks and included 10, 20, 50, 200 Hz . It is known that the slowly adapting mechoreceptors of the tactile system which have the best spatial sensitivity, are most responsive to low frequencies of vibration, whereas the Pacinian corpuscles are least spatially sensitive of the mechoreceptors and respond best to high frequencies. In light of this knowledge, the use of different frequencies in the present study was examined to assess how the mechoreceptors interact under different stimulus frequencies. Stimulus frequency was fixed within each block. Testing was conducted in 1-2 hour sessions, and subjects were given frequent rest periods to minimize fatigue.

Results and Discussion

Performance for each subject was averaged as a function of frequency. Pattern identification for the set of shapes was high for all subjects, averaging above 90% across frequencies. Average identification for the set of letters, however, was 79.5%. This difference in performance across pattern sets indicates that the set of letters contained more complex patterns. These results are very similar to those of Weisenberger and Hasser (1994) for a 30 pin full-finger display, and suggest that changing the field of view from 30 pins to 9 pins, while keeping the size of the pattern constant, is not detrimental to performance. A slight frequency effect was noted and indicated better performance obtained at 10 Hz and the lowest performance at 200 Hz, suggesting that the slowly adapting receptors contribute most to pattern identification in haptic exploration and the Pacinian corpuscles contribute less information. This frequency effect was not significant.

Confusion matrices for these data averaged across subjects and frequency are shown in Figures

2 and 3 for the two pattern sets. Figure 2 indicates that few confusions were made on the shapes set, which is expected given the high performance levels. The confusions that were made were often composed of shapes with similar features. For example, "X" was confused most frequently with patterns 1 and 6, each of which contain one diagonal line. Figure 3 indicates confusion items for the set of letters. Confusions made in this set were also made between patterns with similar characteristics, such as "G" for "Q", "R" for "K", and "D" for "O".

These confusions are similar to ones reported by Weisenberger & Hasser's (1994), using the 30 pin array. The similarities between confusions made with the two different arrays suggests that limiting the field of view of the stimulators does not constrain the transmission of feature information, and in fact, that the information provided by the two displays was similar. These findings give rise to the another question of how far the field of view can be reduced before there are detrimental effects of pattern identification. From a practical standpoint, reducing the field of view may simplify construction of a wearable tactile display for use in telerobotics and virtual environment applications.

Experiment 2

Methodology

The second experiment addressed the perception of different pattern sets with further reductions in field of view. Again, the stimuli in these blocks were larger than the field of view provided by the display, so that active haptic manipulation was necessary for accurate identification of the patterns. The 3 x 3 array used in this experiment was the same as that used in Experiment 1. However, software was enhanced to allow the experimenter to mask stimulator elements. This enabled the array to be activated with different fields of view (9-pins, 4-pins, 1-pin).

Three subjects (2 female, 1 male) participated in this study. Two subjects (KMS and JMW) were involved in Experiment 1. A third subject (KJS) had participated in Weisenberger and Hasser's (1994) study.

Eighteen blocks were completed by all subjects (50-trial blocks for shapes, 40-trial blocks for letters). Patterns were randomly presented in each block of trials. The three different fields of view were presented in a random order across blocks, but were held constant with a block. Stimulus frequency was held constant at 20 Hz. Duty cycle was fixed at 35% for all blocks. When subjects moved the display into contact with the pattern on the digitizing pad, a visual display of random flashing lights was activated. Subjects were instructed that they could use this visual display to locate the pattern on the digitizing pad, but were not to fixate on the visual display. Testing was conducted in 1-2 hour sessions, and subjects were given frequent rest periods to minimize fatigue.

Results

Figure 4 show average pattern identification across subjects for each presentation mode (9-pin, 4-pin, 1-pin) for the sets of shapes and letters, respectively. It is clear that performance with both 9-pin and 4-pin displays were similar. However, performance with 1-pin was substantially reduced for both pattern sets, and fell to only slightly above chance levels for the set of letters, indicating that the 1-pin field of view does not offer enough information for consistent pattern identification when patterns are complex.

A two-way, within subjects analysis of variance on arcsine-transformed data is planned to determine significance of differences in performance as a function of pattern set and display size. Further data collection is required for one subject before statistical analysis can be completed. However, results suggest substantial effect on performance for both number of elements in a display and for pattern set.

Discussion of Experiments 1 and 2.

Subjects reported that items in the set of shapes were easier to recognize than those in the set of letters in both Experiments 1 and 2. No differences in performance across frequencies were found for any of the subjects, suggesting that the tactile mechoreceptors interact to offer appropriate spatial sensitivity regardless of stimulus frequency.

Subjects also indicated that the field of view with only one pin activated was too limiting to identify patterns accurately. They expressed considerable difficulty in spatially locating and maintaining the pattern on the digitizing pad when using the 1-pin display. The data support this observation.

Differences in scanning behavior were measured for one subject (KMS) across the 9-pin, 4-pin,

and 1-pin display modes. A total of 10 stimulus presentations were timed for each display mode, and averages were calculated. Average scanning time for the 9-pin display was 9.1 s. For the 4-pin , the scanning time lengthened to 15.8 s. Interestingly , the scanning time for the 1-pin display was appreciably longer at 24.2 s. These differences in scanning time across the three display modes suggest that they are not equally easy to use, with the 1-pin display mode being the most difficult. These data parallel the percent correct identification results. This is true because the 1-pin display results in a task which relies heavily on temporal integration and severely limits the spatial cues, except those acquired via kinesthetic motion. Results such as these demonstrate the problems in reducing the field of view to a point where spatial mapping and temporal integration becomes difficult for the tactile system. It would appear, however, that development of a stimulator with 4-pins would still allow acceptable performance levels, while reducing the actual size of the display.

Experiment 3

Methodology

Experiments 3 and 4 were designed to address possible explanations for Weisenberger and Hasser's (1994) finding that haptic scanning of stimulus patterns was superior to the passive scanning mode.with a 5 x 6 array. A 5 x 6 array was utilized in this experiment, identical to the one used during the previous Weisenberger and Hasser (1994) study, and similar in design to the 3 x 3 array used in the other set of experiments. Experiment 3 examined subject performance when multiple "looks" are allowed in the passive scan and static presentation modes, as well as in the haptic mode.

The display used in this experiment was a 5 x 6 array of shape memory alloy (SMA) stimulators housed in a plastic casing. This display, manufactured by TiNi Alloy Corporation (A. Johnson, 1991), is composed of 30 stimulator elements made of beryllium-copper rod, connected to a SMA wire. The SMA wires when heated by input current to the display, contract, forcing the stimulators upward. Activation of the stimulators causes them to protrude through holes in the display. As the input current stops and the SMA wire cools, each element returns to its resting state below the plastic casing of the display. The 30

elements are spaced approximately 3.0 mm apart. All thirty elements could be activated during the first experiment.

Four subjects (1 female, 3 male) participated in this study. Two subjects (KMS and KJS) had been involved in previous tactile experiments (Experiments 1 and 2 of the present study). Two subjects (BTP and RGH) had limited exposure to tactile pattern perception in the past, but received practice in tactile perception tasks prior to data collection.

Only the set of letters were utilized in this experiment because when the set of shapes was used in earlier experiments, no significant differences among presentation modes could be identified due to ceiling effects (Weisenberger and Hasser, 1994). The set of letters, however, were more difficult for subjects to identify, allowing significant differences to be examined.

Eighteen blocks were completed by each of the subjects (40-trial blocks). Stimulus patterns were randomly presented in each block of trials. Stimulus frequency was varied across blocks, including 20, 50, and 200 Hz. Stimulus frequency was fixed within each block. Duty cycle was fixed at 75% for all blocks. Testing was conducted in 1-2 hour sessions, and subjects were given frequent rest periods to minimize fatigue.

During static and passive scan trials, subjects were instructed to rest their index finger on the display and wait for the stimulus presentation. The C software initiated the first presentation of a pattern. In the passive scan presentation, stimulus patterns moved from right-to-left across the display . An icon of each pattern was displayed on the PC computer screen. Following the initial presentation, the subject was prompted to chose either to repeat the stimulus presentation by selecting an "R" on the PC keyboard or to select a response between 0 and 7. Subjects were permitted a maximum of three repeats in addition to the initial presentation. Trial by trial feedback was delivered to the subject via the PC monitor.

Results and Discussion

Figure 5 shows average pattern identification across subjects for each presentation mode. Pattern identification was similar across presentation modes and was between 42-82%. Figure 5 also suggests that there is large intersubject variability in the data.

A two-way, within subjects analysis of variance was performed on arcsine-transformed data to determine the significance of differences in performance as a function of presentation mode and presentation frequency. Results indicated no significant differences (ns) across presentation modes ($F(2,6) = 1.72$, ns) or presentation frequencies ($F(2, 6) = 1.02$, ns), indicating that performance was equivalent regardless of mode or frequency. No interaction was observed between mode and frequency. These data suggest that the ability of repeating the stimulus is that factor in the Weisenberger and Hasser (1994) study that accounts for the superior performance of the haptic presentation mode.

Data were also collected on the number of repeats used by each subject under different presentation modes to determine if there are differences in difficulty across modes. The average number of repeats in the static presentation mode was 2.045, and in the scan mode was 0.72. A t-test revealed a significant effect ($t(3) = 10.5$, $p < .01$), indicating that subjects needed significantly more looks in the static presentation mode than in the passive scan mode. The fact that subjects requested more looks in the static mode suggests that even though differences in performance were not significant, subjects processed the stimulus patterns differently in the static mode than in the passive scan mode.

Experiment 4

Methodology

The 5 x 6 array used in Experiment 3 was also utilized in Experiment 4. This fourth experiment examined subject performance when multiple scan directions were permitted in the passive scan presentation mode. Experiment 3 suggested that repeated looks at the stimulus patterns does reduce the differences in performance across presentation modes that was found in Weisenberger and Hasser (1994). Experiment 4 was designed to determine whether allowing subjects to choose the direction of scan in the passive scan presentation mode facilitates performance.

Four subjects (1 female, 3 male) participated in this experiment. Two subjects (KMS and KJS) were involved in Experiments 1, 2, and 3 of the present study. Two subjects (BTP and RGH) had participated in Experiment 3.

The set of letters delivered in this experiment were identical to those in Experiment 3. Software created for this experiment permitted the subjects to choose among 8 different directions from which to scan the set of letters. Three stimulus presentation conditions were explored. The first condition, termed "single," allowed only a single look at the pattern from a specified direction. The second condition, "multi-same," permitted multiple looks at the pattern, but constrained the subject always to look at the pattern from the same specified direction. The third condition, "multi-different," allowed multiple looks at the pattern, but constrained the subject's initial look to a specific direction and subsequent looks to be from scan directions determined by the subject. In this third condition, subjects were instructed that no scan direction could be chosen twice. Figure 6 shows the eight scan directions tested in this experiment.

Forty-eight blocks were completed by all subjects (40-trial blocks). Stimulus patterns were randomly presented in each block of trials. Patterns were presented at 20 Hz. Experimental condition (single, multi-same, multi-different) were randomized across blocks, and were fixed within each block. Duty cycle was fixed at 75% for all blocks. Testing was conducted in 1 hour sessions, and subjects were given frequent rest periods to minimize fatigue.

Results and Discussion

Data were collected and averaged across subjects for the single, multi-same, and multi-different conditions. Figure7 shows the data for three of the four subjects (the fourth subject had not completed data collection at the time of this report). Large intersubject differences were observed among the data. Across experimental conditions, averages for scan direction 2 were the highest, suggesting that the subjects preferred to view the patterns from this direction.

There are no substantial differences between the overall averages from the single and multi-same conditions. However, comparison of the single and multi-same averages as a function of scan direction yields an interesting trend. Here it is apparent that performance does improve slightly in the multi-same condition for all directions except 5 and 7. This increase in performance demonstrates that subjects were benefitting from repeated looks at the patterns. Averages for the multi-different condition appear to be somewhat better than those for the other two conditions. In the multi-different condition, there is only small intersubject variability in performance. This may be attributed to the fact that all of the subjects used

a similar scan strategy. Analysis of the optional scan directions chosen by the subjects indicates that the subjects almost always chose scan direction 2 on the second look.

Because subjects were trained for Experiments 3 and 4 using scan direction 2, it was anticipated that subjects would prefer scanning from this direction over the others. Further, it was originally postulated that either horizontal scan direction, 2 or 3, would offer the most information to subjects for pattern identification and that the oblique scan directions, 4 - 7, would offer only minimal information. These assumptions were only partially correct. Performance was higher across the single and multi-same conditions for scan directions 2 and 3 than for the oblique directions. In fact, performance across presentation modes was better for scan direction 2, suggesting that subjects could process information from this direction more easily than from the other directions.

Overall, it appears that scan directions 5 and 7 offer the least information for pattern identification across presentation modes. However, some of the scan directions that were not expected to facilitate performance do appear to offer a substantial amount of information. For example, averages for scan direction 6 were high across all conditions.

Average performance for the multi-different experimental condition is better than for either the single or multi-same conditions. In fact, the multi-different condition results in superior performance for all scan direction except direction 2 in which case the multi-same condition showed the best performance. This is likely to be due to the subjects preference for scan direction 2. Examination of the data indicates that subjects tended to chose scan direction 2 as one of their repeats in all of the multi-different blocks.

General Conclusions

In summary, the four experiments completed on this project offer important information toward the optimal design of a tactile display. Experiment 1 suggests that the vibrating frequency of the stimulator will not be a critical factor in the design, because no significant differences in performance across frequencies were found for any of the subjects. Both Experiments 1 and 2 suggest that subjects are able to discriminate both simple and complex stimuli. These experiments also reveal that similar performance was observed with the 3 x 3 array as with the 5 x 6 array in earlier research by Weisenberger and Hasser (1994). Experiment 2 further suggests that stimulator designs can be reduced to as few as 4-pins before

performance is degraded. Differences among scanning behaviors across the 9-pin, 4-pin, and 1-pin displays demonstrate the problems in reducing the field of view further than 4-pins. These findings suggest that wearable tactile displays for sensing need not cover the entire fingertip, facilitating incorporation of such displays into a data glove.

Experiment 3 suggests that subjects needed more looks when patterns were presented in a static mode than in a passive scan mode. Passive scan and static mode performances approximate performance in the haptic mode when multiple looks at the patterns are permitted. These findings suggest that repeated looks at the stimulus do facilitate pattern identification.

Finally, Experiment 4 demonstrated the subjects' ability to integrate information from most of the eight scan directions. From a practical standpoint, this is interesting because it suggests that performance in the passive scan mode is only somewhat dependent on scan direction. Nonetheless, future scan displays should continue to utilize horizontal scan directions to encourage optimal pattern identification performance.

REFERENCES

Craig, J.C. (1980). Modes of vibrotactile pattern generation. Journal of Experimental Psychology: Human Perception and Performance, 6, 151-166.

Craig, J.C. (1981). Tactile letter recognition: Pattern duration and modes of pattern generation. Perception & Psychophysics, 31, 540-546.

Hill, J.W. (1974). Limited field of view in reading letter shapes with the finger. In F.A. Geldard (Ed.), Cutaneous communication systems and devices. Austin, TX: The Psychonomic Society.

Johnson, A.D. (1990). Shape-memory alloy tactile feedback actuator. Technical Report AAMRL-TR-90-039. TiNi Alloy Company, 1144 65th Street, Oakland CA, August.

Lappin, J.S. and Foulke, E. (1973). Expanding the tactual field of view. Perception & Psychophysics, 14, 237-241.

Loomis, J.M. (1974). Tactile letter recognition under different modes of stimulus presentation. Perception and Psychophysics, 16, 401-408.

Loomis, J.M. (1980). Interaction of display mode and character size in vibrotactile letter recognition. Bulletin of the Psychonomic Society, 16, 385-387.

Sherrick, C.E. (1964). Effects of double simultaneous stimulation of the skin. American Journal of Psychology, 77, 42-53.

Weisenberger, J.M., & Craig, J.C. (1982). A tactile metacontrast effect. Perception & Psychophysics, 31, 530-536.

Weisenberger, J.M., & Hasser, C.J. (1994). Role of active and passive movement in vibrotactile pattern perception. Submitted for publication.

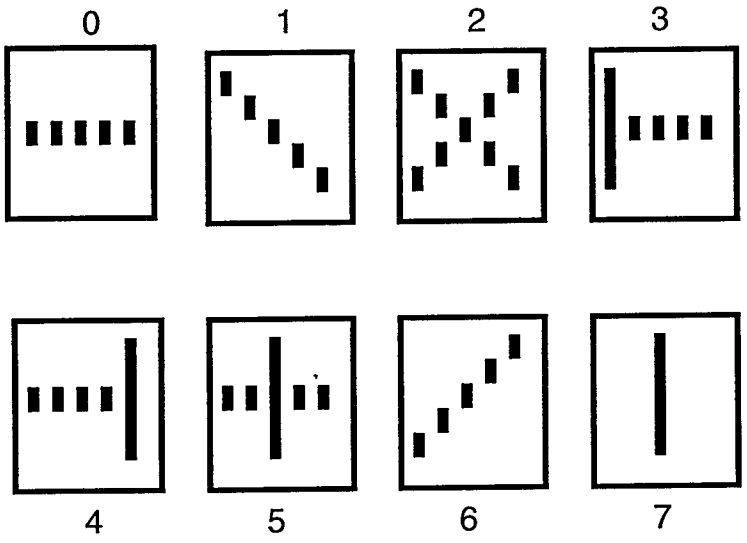


Figure 1a. Icons for stimuli in the set of shapes.

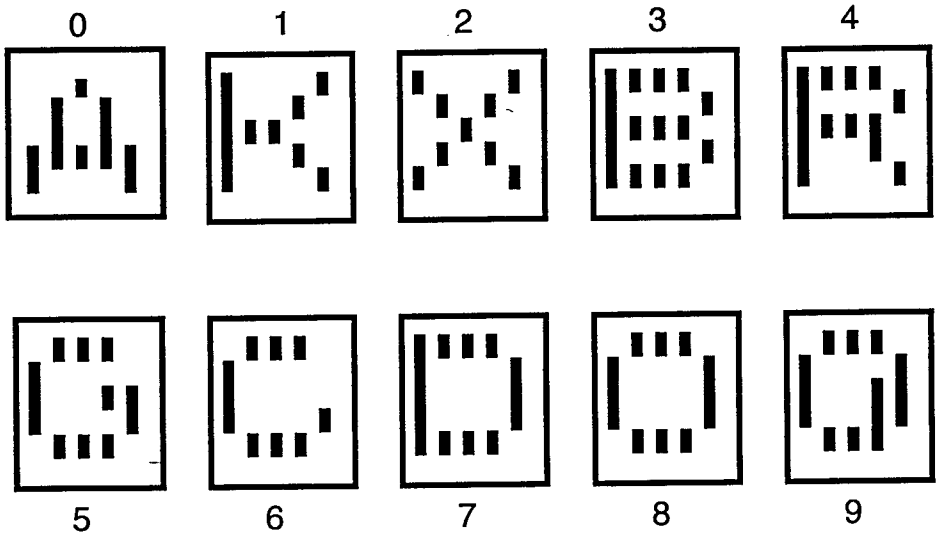


Figure 1b. Icons for the stimuli in the set of letters.

		RESPONSE (in percent)									
		0	1	2	3	4	5	6	7	8	9
	0	97.5	1.7	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0
S	1	0.0	99.2	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0
T	2	0.0	0.0	90.8	0.0	0.0	7.5	1.7	0.0	0.0	0.0
I	3	0.8	0.0	1.7	93.3	0.0	0.8	0.0	3.3	0.0	0.0
M	4	2.5	0.0	2.5	0.0	90.8	3.3	0.8	0.0	0.0	0.0
U	5	0.8	0.0	3.3	7.5	2.5	85.8	0.0	0.0	0.0	0.0
L	6	2.5	0.0	0.0	0.0	0.0	0.0	97.5	0.0	0.0	0.0
U	7	0.8	0.0	0.0	0.0	0.0	0.0	0.0	99.2	0.0	0.0
S	8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 2. Confusion matrix showing percent correct as a function of stimulus pattern for the set of shapes. Data are averaged across subjects and frequency.

		RESPONSE (in percent)									
		0	1	2	3	4	5	6	7	8	9
	0	86.7	1.7	1.7	1.7	5.0	1.7	0.0	0.0	0.0	1.7
S	1	8.3	79.2	8.3	1.7	2.5	0.0	0.0	0.0	0.0	0.0
T	2	5.0	3.3	91.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	3	1.7	0.0	0.0	81.7	0.8	5.0	0.0	1.7	0.8	8.3
M	4	5.0	3.3	0.0	0.8	89.2	0.8	0.0	0.0	0.0	0.8
U	5	0.8	0.0	0.0	2.5	0.0	74.2	3.3	4.2	5.8	9.2
L	6	0.0	0.0	0.0	0.8	0.0	0.8	97.5	0.0	0.8	0.0
U	7	0.0	0.0	0.8	0.8	0.8	5.0	0.0	71.7	17.5	3.3
S	8	0.0	0.0	0.0	1.7	0.0	5.0	0.0	20.8	66.7	5.8
	9	0.0	0.0	0.8	11.7	0.8	21.7	0.0	0.8	7.5	56.7

Figure 3. Confusion matrix showing percent correct as a function of stimulus pattern for the set of letters. Data are averaged across subjects and frequency.

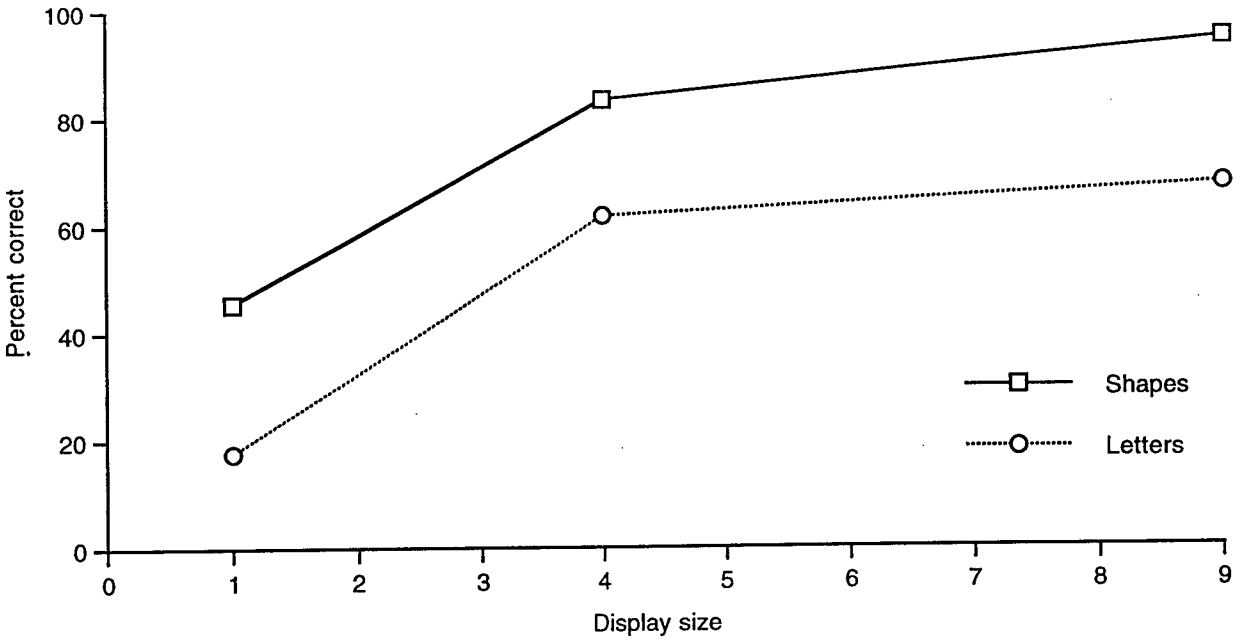


Figure 4. Percent correct identification performance for the 1-element, 4-element, and 9-element displays, for the sets of shapes and letters. Data are averaged for two subjects.

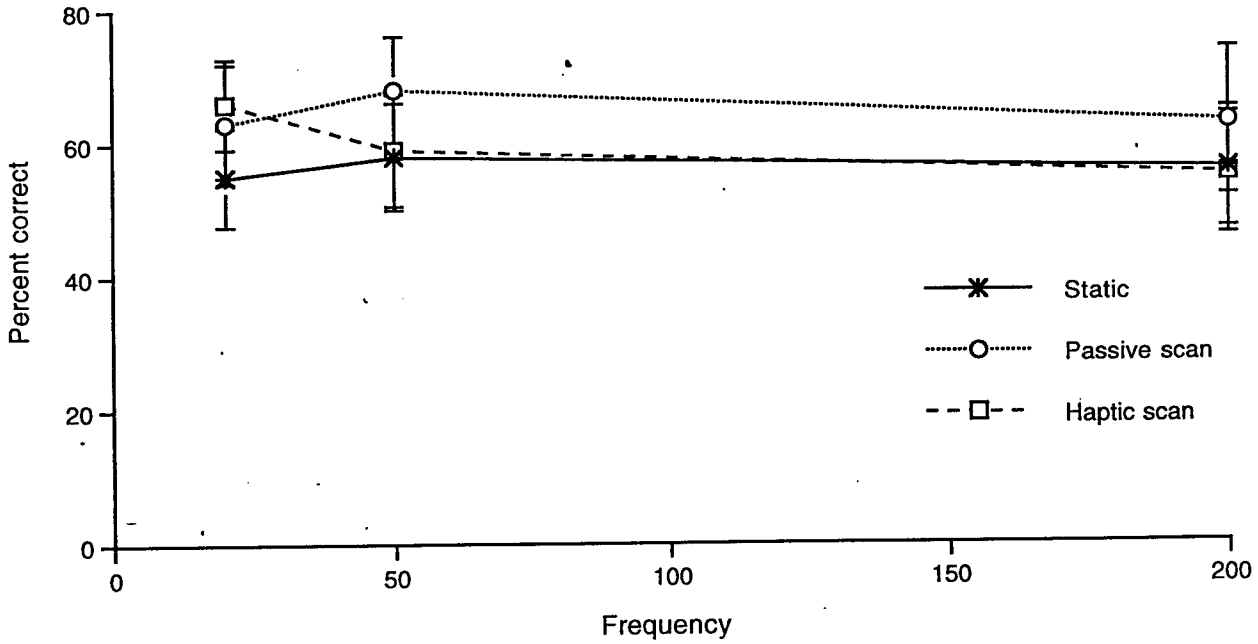


Figure 5. Percent correct identification performance as a function of frequency for static, passive scan, and haptic scan presentation modes. Data are averaged for three subjects. Note that large individual variability indicates no significant differences across presentation modes.

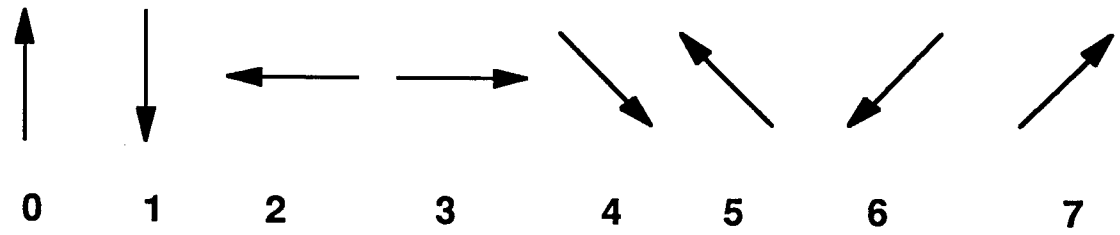


Figure 6. Scan directions employed in the passive scan presentations in Experiment 4.

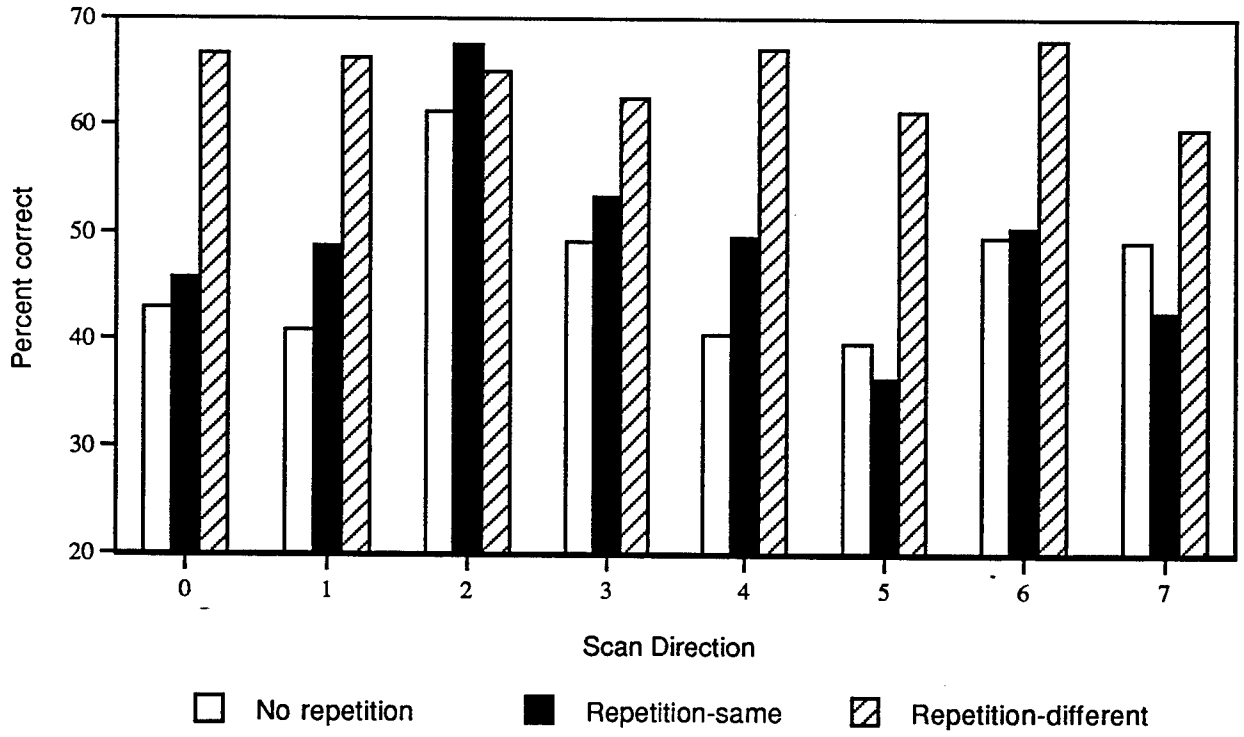


Figure 7. Percent correct letter identification as a function of scan direction, for no repetition, repetition-same, and repetition-different presentation conditions. Data are averaged for three subjects.

PREDICTING PILOT TRAINING SUCCESS WITH LOGISTIC OR LINEAR REGRESSION: AN EXAMPLE WHERE IT DOESN'T MATTER AND WHY

Joseph M. Stauffer

Department of Management and Organizations

University of Iowa

108 PBAB

Iowa City, IA 52242-1000

Final Report for:

AFOSR Summer Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

July 1994

Abstract

Many studies of human attributes related to success in pilot training or job performance use linear statistical methods. For statistical prediction, linear models make few assumptions, have well known statistical characteristics, and are robust to violation of assumptions. Alternatives to these linear statistical models are classes of nonlinear statistics. A nonlinear analogue of linear prediction is logistic regression. Existence of a dichotomous criterion is frequently seen as sufficient and compelling reason for the use of logistic regression. Using the dichotomous criterion of passing-failing pilot training, we demonstrate that linear and logistic regression can yield corresponding results and would rank applicants virtually identically. Certain practical, psychometric, and interpretive advantages accrue to linear regression. A comparative discussion of linear and logistic regression is included.

Predicting Pilot Training Success with Logistic or Linear Regression:

An example where it doesn't matter and why

For most measures of training and job performance, a simple linear regression (LR) model relating various predictors to the performance criterion works very well. Typically, LR uses least-squares estimation and, therefore, has convenient and well-known statistical properties and is widely understood. LR is even robust in the sense that it works well despite moderate violations of its assumptions. These assumptions are sometimes known as the Gauss-Markov assumptions and state that: (1) the predictor variables are fixed, or nonrandom, with no linear dependencies among them, and (2) the disturbance, or error, terms have identical distributions with a mean of 0, have equal variance (homoscedasticity), and are not intercorrelated (Huang, 1970). Another assumption is linearity of form of the regression.

Dichotomous Criteria

When modeling these criterion-predictor relationships, continuous criteria should be used. Cohen (1983) points out that predictive efficiency and statistical power are reduced substantially when a criterion is artificially dichotomized. Artificial dichotomization occurs when trainees are classified as having passed or failed on the basis of a cut score applied to a continuous distribution of scores. Typically one-fifth to two-thirds of the predictive efficiency (Brogden, 1946) is lost when a continuous performance measure is reduced to a dichotomous measure of success/failure. Often, however, in assessing pilot selection systems, continuous measures may not be available for all trainees, particularly those who fail. Often trainees leave training prior to completing all phases of instruction that contribute to the continuous criterion. Consequently, a dichotomous measure, 1 for successful completion and 0 for failure, would be used. Additionally, the probabilities of success given scores on a set of predictors may be of interest. In that case, a dichotomous criterion could be used to obtain estimates of these probabilities of success (Raju, Steinhaus, Edwards, & DeLessio, 1991). However, exact probabilities (i.e. .83, .67, .94, etc.) of success are only infrequently desired. More frequently ranking of the probability of success or membership in the success or failure group is used in practice.

We are warned, however, that LR is inappropriate when the criterion measure is dichotomous, because such a model violates certain Gauss-Markov assumptions as well as the implicit LR assumption that the criterion

is continuous (see, e.g., Aldrich & Nelson, 1984).

The equation for the LR model is stated as

$$Y_i = \alpha + \Sigma\beta_j X_{ij} + u_i, \quad (1)$$

where Y_i is the dichotomous criterion for the ith trainee, 1 for successful completion (pass) and 0 for failure, α is the intercept, $\Sigma\beta_j X_{ij}$ is the sum of the product of the predictor scores and their respective regression slope coefficients, and u_i is the model error term. When the criterion is dichotomous and is scored as 0 or 1, the systematic portion of Equation (1), $\alpha + \Sigma\beta_j X_{ij}$, can be interpreted as the probability that person i will successfully complete the training given his or her scores on the predictors, X_{ij}. Such an LR model is known as the Linear Probability Model (LPM) (Aldrich & Nelson, 1984; Maddala, 1983).

Criticisms of Linear Regression

Heteroscedasticity

One criticism of LR based models, such as the LPM, is that they violate the assumption of homoscedasticity. With a dichotomous criterion, the error term, u_i, can take on only one of two values. If $Y_i = 1$, u_i is equal to $1 - \alpha - \Sigma\beta_j X_{ij}$. If $Y_i = 0$, then $u_i = -(\alpha + \Sigma\beta_j X_{ij})$. As a result, the variance of u_i is given by the following equation:

$$\text{Var}(u_i) = (\alpha + \Sigma\beta_j X_{ij})(1 - \alpha - \Sigma\beta_j X_{ij}). \ ^{1} \quad (2)$$

It can be seen from Equation (2) that $\text{Var}(u_i)$ varies directly, i.e., systematically, as a function of the values of the X_{ij}. It is interesting and important to note that the mean value of u_i is still expected to be 0, ***despite*** this violation of homoscedasticity. This means that LR would still yield unbiased estimates of the β_j. These estimates would not be the most ***efficient***, however, meaning that the sampling error of the estimates would not be optimally minimum (Hogg & Craig, 1978).

The practical impact of such a violation is, for the most part, an empirical question. Cox (1970), for example, pointed out that the efficiency lost is rather minor except in the case where the probabilities, $\alpha + \Sigma\beta_j X_{ij}$, are relatively extreme (less than .2 or higher than .8), and McGillivray (1970) showed that the estimated

---

[1] This equation is directly analogous to the binomial variance: $\sigma^2 = pq$, where p is the proportion of trainees who pass and $q = (1 - p)$, the proportion of trainees who fail.

variance is a *consistent* estimator of the true binomial variance. Additionally, it can be shown that with sample sizes greater than 100, LR yields results similar to generalized least squares estimates that do not rely on the assumption of homoscedasticity (Kuechler, 1980; Smith, 1973). Therefore, despite heteroscedasticity, LR can be expected to yield unbiased, consistent, and fairly efficient estimates needed to predict probabilities.

A related concern stemming from heteroscedasticity is that the multiple correlation coefficient would not be meaningful. Theil (1971), however, explains that the multiple correlation coefficient in such a case is actually a canonical correlation coefficient and can be meaningfully interpreted as an index of predictive efficiency (Brogden, 1946).

Linearity

A second criticism of LR is that the predictor-criterion relationship may be nonlinear. When predicting performance from measures of human attributes, an assumption of linearity seems reasonable. Research shows that human abilities are consistently linearly related to performance measures of all kinds. Specifically, tests of over 150 job predictor-criterion relationships in utility analysis research show that such relationships are linear. Even if one were to assume that the relationship were logistic, a linear model would fit the data very well over most values of the predictor. For probabilities between .25 and .75, the logistic curve is approximately linear (Goodman, 1976; Knoke, 1975).

Range of Predicted Values

Another major criticism, what Maddala (1983) calls "[t]he most important criticism" (p. 16), of LR is that the predicted probabilities can fall outside the 0-1 range. That is, it is possible that some predicted probabilities can fall below 0, and others above 1. This is a trivial concern in practical personnel selection where the goal is to rank order applicants and select from the top down. An applicant receiving a predicted probability of 1.10 would be expected to do better on the criterion than one with a predicted probability of .50, regardless of whether or not his or her probability was outside the 0-1 range. Conversely, an applicant with a score of -.10 would be expected to do worse than an applicant with a probability of .50 and would be ranked accordingly. Applicants are rarely selected on the basis of exact probabilities. The criterion could be rescaled if the goal were, in fact, to produce a probability and, additionally, other methods exist for computing probabilities

from linear models (Hogg & Craig, 1978; Raju et al., 1991).

Logistic Regression

An alternative that seeks to overcome these criticisms is logistic regression (LOGR). In its most basic form, LOGR defines the expected probability of success through a nonlinear transformation of $\alpha + \Sigma\beta_j X_{ij}$:

$$\text{predicted probability}_i = \frac{e^{(\alpha+\beta_j X_{ij})}}{1 + e^{(\alpha+\beta_j X_{ij})}} \qquad (3)$$

Where e is the base of the natural log and has the constant value 2.718. This logistic probability function is continuous and takes on a value from 0 to 1 (0 when $\alpha + \Sigma\beta_j X_{ij} = -\infty$ and 1 when $\alpha + \Sigma\beta_j X_{ij} = +\infty$). Because its predicted values fall within the theoretically permissible range of values, LOGR is seen as the more appropriate modeling technique for relating a set of predictors to a dichotomous criterion (Aldrich & Nelson, 1984).

LR versus LOGR

Certainly, there are violations of the assumptions of LR when the criterion is dichotomous and the resulting predicted values can lie outside the 0-1 range. The question is whether these violations make LR based models, like the LPM, less desirable to alternative models like LOGR. In can be shown, for example, that when the predictors can be assumed to be distributed as multivariate normal, LR, as a special case of discriminant analysis (Tatsuoka, 1988, p. 228), yields estimates that are the true maximum likelihood estimates and is more efficient than LOGR estimation (Maddala, 1983, p. 27).

Nearly thirty years of research pitting LR against LOGR in econometrics, sociology, and biometrics shows that, despite theoretical concerns, LR and LOGR produce very similar results except in the unlikely case where a large proportion of the population has probabilities near 0 or 1 (see, e.g., Cleary & Angel, 1984; Hanushek & Jackson, 1977; Press & Wilson, 1978). Even where LR is not expected to hold up well (e.g., all predictors are dichotomous and, therefore, clearly not multivariate normal, Gilbert, 1968), it can perform as well as LOGR (see also Moore, 1978). This is an important finding for personnel selection.

Clearly, violations of LR assumptions do not necessarily argue in favor of LOGR models. There are a

number of practical, psychometric, and interpretive advantages to using LR models. This paper directly compares the estimated probabilities of success based on an artificially dichotomized criterion (pass versus fail) in a military pilot training program derived from both LR and LOGR to answer the research question of does the form of the regression make a difference.

Method

Subjects

The subjects were 1,228 officers enrolled in U.S. Air Force Undergraduate Pilot Training (UPT). All subjects had complete data with 1,060 (86.3%) passing and 168 (13.7%) failing training. They were all college graduates at time of training and had been selected on the basis of academic achievement, desire to fly and, at least in part, on the basis of scores on the Air Force Officer Qualifying Test (AFOQT).

Measures

Predictors. The predictors were scores from the eight components of the Pilot Candidate Selection Method (PCSM) model. These scores are currently used by the Air Force to select candidates for pilot training (Carretta, 1992a). These included the Pilot Composite of the AFOQT (Skinner & Ree, 1987), a psychomotor composite, response time on a set of mental rotation tasks, response time from short-term memory/item recognition tasks, a measure of tracking difficulty/time-sharing (North & Gopher, 1976), response time and choice on an activities interest inventory designed to measure attitudes toward risk-taking, and a measure of previous flying experience. See Carretta (1992b) for a more complete description.

Criterion. UPT is a 53-week flight training course with an academic phase running concurrently with initial and advanced jet training phases. UPT graduates typically complete about 190 hours of flying. The criterion was the dichotomy of pass/fail in UPT based on a series of continuous ratings of flying performance and academic grades during training. This dichotomous measure was coded 1 if the trainee passed, 0 if the trainee failed.

Procedure

Two sets of analyses were conducted. First, the relationship between the dichotomous criterion and all eight

predictor variables was modeled using both LR (Equation 1) and LOGR (Equation 3). Both models were applied to produce predicted probabilities of success for each pilot trainee. These predicted probabilities were then correlated with each other and the criterion using both Pearson product-moment and Spearman rank-order correlation coefficients. The Spearman correlation was used in addition to the Pearson correlation, because it provides a measure of equality of ranking and does not require the assumption of linearity in the form of the relationship. Nonlinearity was introduced by the computation of LOGR. It is advantageous to use the Spearman coefficient in situations where ordering of the observations is the operational application. For example, we typically order applicants for jobs or training, and the absolute magnitude of the predicted score is less important than the rank of the applicant.

The second set of analyses used various multiple regression forward selection procedures to select the "best" subset of predictors from the eight variables.[2] For LR, regular forward selection and stepwise selection were used. For LOGR, two forward selection methods were used: one was based on a Wald statistic, the other, a general likelihood ratio statistic.[3] No commonly accepted true stepwise LOGR is generally available. The predicted probabilities for all models were then correlated with each other and the criterion.

Results

Table 1 shows the intercorrelations of the predictor and criterion variables and their means and standard deviations. The raw score (unstandardized) regression weights for both LR and LOGR using all eight predictors were examined. Both models yielded weights with the proper signs. No variables were weighted with signs that differed from the sign of their zero-order correlation with the criterion. The weights were positive for variables where higher scores indicated better performance and negative where lower scores indicated better performance (e.g., response time). Figure 1 is a bivariate scatterplot showing the relationship between the

---

[2] SPSS for Windows, version 5.0.2, was used. We retained the default values for the analyses. For example, .05 was the significance level for including a variable, .10 was the level for excluding one, and 20 was the maximum number of iterations for the LOGR analyses.

[3] Two different likelihood ratio decision criteria (i.e., indices of the change in log likelihood) were used, but they resulted in identical solutions. Rather than go into the technical differences between the two, we opted to discuss them as one general likelihood ratio method.

probabilities obtained by LR and by LOGR. The correlations describing this relationship were .9587 using a Pearson coefficient and .9967 using a Spearman rank-order coefficient. The difference in the two correlation coefficients is clearly due to a slight nonlinear relationship between the two sets of probabilities introduced by LOGR. The ranking of trainees on the basis of the eight predictors, however, would be essentially identical using LR or LOGR. When the predicted probabilities were correlated with the original dichotomous criterion, LR proved slightly better with R_{LR}=.2306, while R_{LOGR}=.2284 for correlation of the predicted scores with the criterion.

Table 2 show the intercorrelations of the predicted probabilities resulting from regressions on the predictors selected by the various forward selection procedures. All selection procedures, the two LR and the two LOGR, selected the same three predictors for this sample: the AFOQT Pilot Composite, response time on the short-term memory/item recognition tasks, and response time on the activities interest inventory. The predicted probabilities were perfectly correlated within regression method (i.e., among LR methods and among LOGR methods) and nearly so between method (Pearson correlation = .9672, Spearman correlation = .9985). Here again the trainees would have been ranked almost identically regardless of the model, and LR holds the slight advantage with R_{LR}=.2113 and R_{LOGR}=.2015.

Discussion

Despite the high pass rate (86.3%) in this sample, the resulting expected loss of efficiency, and the other criticisms of LR, it is difficult to argue that LOGR should be preferred. In fact, based on the correlations between the predicted probabilities and the criterion measure, LR actually produced slightly better results than LOGR. More importantly, LR and LOGR would essentially rank all trainees the same, making neither method preferable in that respect. If both regression methods were used and applicants were ranked and selected from the top of the rankings, the same applicants would be selected under both methods. Additionally, the same predictors would be selected for inclusion in the models.

The lack of any practical difference between the two methods is because they are both misspecifications of the predictor-criterion relationship. The criterion is dichotomous, making the relationship between it and any predictor discontinuous. There are no criterion values *between* 0 and 1. Both LR and LOGR models yield

continuous predicted values such as .70, .39, and so forth. Any model that posits a continuous relationship is merely arbitrary and will suffer specification error except under conditions of perfect prediction. Perfect prediction occurs when all failures score below some value on the predictor, say X_c, and all successes score above X_c and in this case the Pearson correlation would be 1.0.

Despite the fact that both methods would rank order applicants similarly, there are a number of practical reasons for preferring LR. Anyone who has tried explaining correlation or simple regression to managers will appreciate the difficulty in explaining a LOGR selection model. In LR, for example, the β_j can be interpreted as the effect of a unit change in X_{ij} on trainee *i*'s probability of passing. Interpreting LOGR weights is not as straightforward. The influence of X_{ij} on the probability of passing varies with the value of X_{ij}. In other words, the effect of the variable depends on its value. Even though partial derivatives can be employed to aid in interpreting the LOGR model, they would be of little value in convincing managers. The ability to clearly interpret the LOGR model grows more complicated as the number of predictors grows. Practitioners would no doubt find the LOGR model difficult to apply.

In contrast, there are a number of reasons for preferring LR. There are, for example, well-established methods for detecting bias in predictors using LR (Lautenschlager & Mendoza, 1986). There are no such universally accepted procedures for LOGR.

Although Raju et al. (1991) point out that the effect of restriction in range on LOGR coefficients is small in their study, Lawley (1943) demonstrates that there is no effect of range restriction on LR coefficients. Further, Lawley (1943) presents a proof of a method to correct the LR derived multiple correlation for range restriction. No such proof exists for LOGR. In fact, the results of Raju et al. (1991) suggest that LOGR coefficients developed in a trainee pool would not be appropriate for an applicant pool. This is most likely due to the LOGR model's reliance on the proportions passing and failing which will differ dramatically between applicant and trainee samples. This could be a serious problem that renders LOGR untenable in practice. If we develop models using trainee samples that do not hold for applicant samples, we cannot apply them with any confidence. In LR, however, the coefficients estimated in one sample can be expected to hold in the other sample.

LR derived multiple correlations can also be corrected to account for dichotomization. Dichotomous criteria in selection settings such as pass/fail are dichotomizations of continuous variables. In this study the continuous variable was a composite measure of ratings of flying and academic grades. As pointed out above, much of the predictive validity is lost when the criterion is artificially dichotomized. Corrections for LR correlations are readily available (see, e.g., Hunter & Schmidt, 1990), but LOGR assumes that the criterion is truly dichotomous; therefore, no such correction exists for LOGR.

Summary

When assessing selection models, continuous criteria should be used. When a criterion is dichotomized, estimates of LR coefficients may not be as efficient as they would be with a continuous criterion. If the predicted values are to be interpreted as probabilities of success, some of these values could fall outside the 0-1 range. Predicted probabilities from a LOGR model would fall within the 0-1 range. This study demonstrated that this may be LOGR's only advantage. Both methods ranked trainees the same. In practice this means that the same applicants would be selected and the aggregate training outcome (i.e., proportion passing) would be the same.

However, there are many practical reasons for preferring LR: the ability to detect bias, correct for range restriction, and ease of interpretation and use by practitioners and decision makers. A critical question for future research is the extent to which LOGR models developed with a trainee are appropriate for the applicant sample. A large discrepancy in the proportions of passing and failing between trainee and applicant samples would strongly suggest that LOGR models developed in each sample would also differ substantially. In sum, there are no quantitative advantages in practice to using LOGR models in pilot selection, while there are many practical advantages to using LR selection models.

References

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models.* Sage University Paper series on Quantitative Applications in the Social Sciences, 07-045. Newbury Park CA: Sage.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37,* 65-76.

Carretta, T. R. (1992a). Recent developments in U. S. Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine,* 63, 1112-1114.

Carretta, T. R. (1992b). Short-term retest reliability of an experimental U. S. Air Force pilot candidate selection test battery. *The International Journal of Aviation Psychology,* 2, 161-173.

Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. *Journal of Health and Social Behavior,* 25, 334-348.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement,* 7, 249-253.

Cox, D. R. (1970). *The analysis of binary data.* London: Chapman & Hall.

Gilbert, E. S. (1968). On discrimination using qualitative variables. *Journal of the American Statistical Association,* 63, 1399-1412.

Goodman, L. A. (1976). The relationship between modified and usual multiple-regression approaches to the analysis of dichotomous variables. In D. R. Heise (ed.), *Sociological Methodology.* San Francisco: Jossey-Bass.

Hanushek, E. A., & Jackson, J. E. (1977). *Statistical methods for social scientists.* New York: Academic Press.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics.* New York: Macmillan.

Huang, D. S. (1970). *Regression and econometric methods.* New York: Wiley.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park CA: Sage.

Knoke, P. (1975). A comparison of log-linear and regression models for systems of dichotomous variables. *Socilogical Methods and Research,* 3, 416-434.

Kuechler, M. (1980). The analysis of nonmetric data: The relationship of dummy dependent variable regression using an additive-saturated Grizzle-Starmer-Koch Model. *Sociological Methods and Research,* 8, 369-388.

Lautenschlager, G. J., & Mendoza, J. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133-139.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburg, Sect. A. 62,* Part I, 28-30. Lawley (1943).

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge: Cambridge University Press.

McGillivray, R. G. (1970). Estimating the linear probability function. *Econometrica*, 38, 775-776.

Moore, D. M. (1973). Evaluation of five discriminant procedures for binary variables. *Journal of the American Statistical Association*, 68, 399-404.

North, R. A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors*, 18, 1-14.

Press, S., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73, 699-705.

Raju, N. S., Steinhaus, S. D., Edwards, J. E., & DeLessio, J. (1991). A logistic regression model for personnel selection. *Applied Psychological Measurement*, 15, 139-152.

Smith, V. K. (1973). *Monte Carlo methods: Their role for econometrics.* Lexington MA: Heath.

Skinner, J., & Ree, M. J. (1987). *Air Force Officer Qualifying Test (AFOQT): Item and Factor analyses of Form O* AFHRL Tech. Rep. No. 86-68). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research.* New York: Macmillan.

Theil, H. (1971). *Principles of econometrics.* New York: Wiley.

Figure 1. Predicted Probabilities from Linear and Logistic Regression on all 8 PCSM Predictors

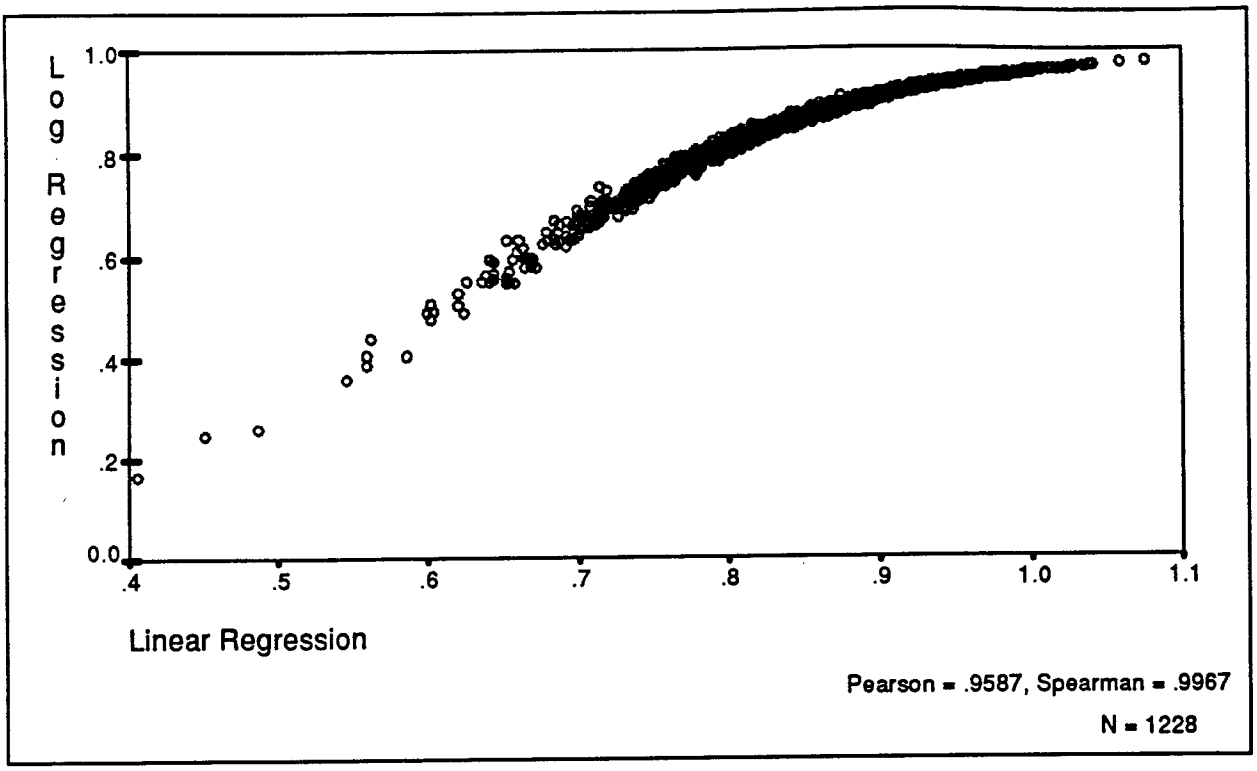

Log Regression
1.0
.8
.6
.4
.2
0.0
.4
.5
.6
.7
.8
.9
1.0
1.1
Linear Regression
Pearson = .9587, Spearman = .9967
N = 1228

Table 1.

Means, Standard Deviations, and Intercorrelations Among the Variables

	1	2	3	4	5	6	7	8	9
	1.00								
	-.15	1.00							
	-.16	.09	1.00						
	-.07	.06	.35	1.00					
	-.05	.13	.10	.08	1.00				
	.15	-.02	-.23	-.20	-.05	1.00			
	.06	.06	-.21	-.14	.04	.24	1.00		
	.10	-.07	-.14	-.26	.06	.24	.17	1.00	
	.00	-.10	-.13	-.10	.03	.11	.08	.16	1.00
Means									
	65.8	4453.9	825.7	1027.6	3.9	0.018	245.4	76.9	0.86
Standard Deviations									
	12.9	1014.5	253.6	508.9	3.2	0.85	36.8	15.5	0.34

Note: The variables are: 1 = activities interest score, 2 = activities interest response speed, 3 = item recognition score, 4 = mental rotation response speed, 5 = rating of previous flying experience, 6 = psychomotor composite score, 7 = tracking difficulty task score, 8 = AFOQT Pilot Composite score, 9 = pass or fail flight training. See Carretta (1992b) for a detailed description.

Table 2.

Intercorrelations of the Predicted Scores from the Various Regression Methods.

	LINF	LINS	LOGL	LOGW	
LINF	1.000	1.000	.998	.998	Forward Selection
LINS	1.000	1.000	.998	.998	Stepwise Selection
LOGL	.965	.965	1.000	1.000	Likelihood Function
LOGW	.965	.965	1.000	1.000	Wald Statistic

Note: Values below diagonals are Pearson correlation coefficients; values above are Spearman rank-order coefficients. LINF is LR forward selection, LINS is LR stepwise selection, LOGL is LOGR selection by likelihood function and LOGW is LOGR with selection based on the Wald statistic.

ANALYSIS OF THE ABSORPTION AND METABOLISM OF TRICHLOROETHYLENE AND ITS METABOLITES BY THE RAT SMALL INTESTINE AND MICROFLORA

Scott G. Stavrou
Former Graduate Student
Department of Biology

University of Central Arkansas
201 S. Donaghey
Conway, AR 72305

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September 1994

ANALYSIS OF THE ABSORPTION AND METABOLISM OF TRICHLOROETHYLENE AND ITS METABOLITES BY THE RAT SMALL INTESTINE AND MICROFLORA

Scott G. Stavrou
Former Graduate Student
Department of Biology
University of Central Arkansas

Abstract

An isolated vascular perfused intestinal system was used to analysis the absorption and metabolism of trichloroethylene (TRI) by the small intestine of F-344 rats. The uptake of TRI was studied at doses of 50, 25, and 5 mg/kg body wt. The maximum cumulative uptake of TRI was found not to exceed 0.01% of any of the administered doses. The formation of metabolites of TRI by enzymes of intestinal mucosa was not observed. The low absorption and lack of metabolite formation may contribute to TRI's differential carcinogenic potential for different species. Additionally, the microflora of the small intestine, cecum, and large intestine were analyzed for the ability to metabolize TRI and its metabolites under aerobic and anaerobic conditions. Under anaerobic conditions the formation of large amounts of dichloroacetic acid (DCA) from spikes of trichloroacetic acid (TCA) was observed. Formation of DCA was often associated, but not limited to, gut contents obtained from the cecum. Also, trichloroethanol (TCOH) was formed from chloral hydrate (CH) under anaerobic conditions from cecum and large intestine samples. The degradation of TRI did not occur under aerobic conditions, but under anaerobic conditions the formation of low levels of DCA was observed. These findings show that the microflora can clearly metabolize TRI metabolites and suggest that the microflora should potentially be considered as a separate compartment within physiologically based pharmacokinetic models.

ANALYSIS OF THE ABSORPTION AND METABOLISM OF TRICHLOROETHYLENE AND ITS METABOLITES BY THE RAT SMALL INTESTINE AND MICROFLORA

Scott G. Stavrou

Introduction

Trichloroethylene (TRI) is a man-made chlorinated hydrocarbon mainly used as a solvent for the vapor degreasing of metals {1}. This application alone accounts for 80% of the TRI produced in the United States {2}. Currently, TRI is also used as a solvent during textile manufacturing, extraction processes, and as an ingredient of solvent blends {1}. Historically, this volatile organic compound has also served as a fumigant, obstetrical anesthetic, analgesic, disinfectant, and as a extractant in the decaffeination process of coffee {3,4}.

TRI, Chemical Abstracts Service (CAS) registry # 79-01-6, has become a chemical of interest to the United States Air Force due to inhalation and dermal exposure of Air Force personnel during vapor degreasing procedures. Also, additional exposure may occur due to subsequent water and soil contamination in and around Air Force installations that employ this chemical agent. TRI has been detected in up to 34% of the drinking water supplies tested in a nationwide survey, and in 1986 became listed as a chemical contaminant for national regulation under the amended Safe Drinking Water Act of 1974 {1,5}. Besides being a common environmental contaminant, TRI has also fallen under scrutiny due to its structural chemical similarity to vinyl chloride, which is carcinogenic in humans and animals {5,6}.

TRI and its metabolites have been found to be cancer producing in rodents . However, the degree to which these results can be extrapolated in determining the carcinogenic risk to humans is equivocal. Therefore, TRI has been classified as a B2 carcinogen by the EPA {1,7,8}. Part of this ambiguity can be linked to inconclusive epidemiological studies {1}. Physiologically based pharmacokinetic models (PBPK) provide predictive risk assessment for humans exposed to chemical agents. This assessment is based upon the extrapolation of data

obtained from carcinogenic studies on test animals. and takes into account interspecies differences. PBPK models have been established for TRI. These models attempt to address the fate of TRI and its most notable metabolites trichloroacetic acid (TCA) and dichloroacetic acid (DCA). These metabolites have been found to be cancer forming in rodents {1,6}.

The toxicity of chemicals can rely upon the intestinal tracts ability to absorb and metabolize chemical agents. However. current PBPK models for TRI do not address the intestines as a separate compartment of absorption and metabolism within their design. In many reports. observations are made concerning the hepatic effects of TRI. Therefore it is important to quantify and qualify the intestinal role in the delivery of TRI and its metabolites to the liver. Previous work {Paula Adams. unpublished} with an isolated vascular perfused intestinal system. has shown the absorption of 74-83% of parent 1,3-Dintrobenzene. Additionally. the biotransformation of 1,3-dinitrobenzene into the metabolite 3-nitroaniline by drug metabolizing enzymes of the intestinal mucosa was also found {9} . Studies by Hirayama and Pang (1990) using a vascular perfused rat intestine-liver system reported the intestinal formation of gentisamide-5-glucoronide and gentisamide-2-sulfate from gentisamide {10}.

The metabolic contributions of the microbial complement of the gut is also not taken into consideration as a compartment within PBPK models for TRI. Walton and Anderson (1990) found an enhancement of TRI degradation within rhizosphere soils versus edaphosphere soils {11}. This was most likely brought about by microorganisms within the rhizospere. Additionally. cytochrome P-450 systems. which are responsible for a diverse range of biotransformations within humans. are also found in microorganisms {12}. Studies with C3H/HE mice have shown that mice with conventional microflora had a higher incidence of hepatic tumors than their germfree counterparts {13}. While some microorganisms of the intestinal tract will metabolize parent compounds into carcinogenic metabolites. others may detoxify hazardous chemicals into relatively harmless metabolites. This is the case with methylmercury. a potent neurotoxin. which is demethylated by a bacterial enzyme into mercuric mercury. a metabolite that can be readily excreted through the feces {13}.

Research is being done to clearly elicit TRI's carcinogenic potential in humans. The establishment of an isolated vascular perfused intestinal model for rats will enhance future PBPK models and add to the understanding of the parameters which control TRI uptake and metabolism. In addition, analysis of the metabolic roles of the microflora in degrading TRI and its metabolites will further contribute to the refinement of future PBPK models.

Material and Methods

Animals. Adult male F-344 rats (Charles River Breeding Laboratories, Raleigh, NC) weighing between 200-300 grams were kept two per polycarbonate cages with hardwood chips as bedding. Cages were suspended within partitioned enclosures with High-Efficiency Particulate Air (HEPA) filter units controlling air quality and flow. Animals were housed in a temperature and humidity controlled room, with a 12 hour light:dark cycle (0600-1800 and 1800-0600, respectively). Rats were maintained on a commercial diet (Purina rat chow, Ralston Purina, St. Louis, MO) and *Pseudomonas*-free water available ad libitum. Rats were not fasted prior to microflora or intestinal perfusion experiments.

Perfusion Solutions. A Kreb's-Ringer bicarbonate buffer, pH 7.4, was prepared with the following chemicals: 0.5 mM $MgCl_2 \cdot 6H_20$, 4.6 mM KCl, 120 mM NaCl, 0.7 mM Na_2HPO_4, 1.5 mM NaH_2PO_4, 15 mM $NaHCO_3$, and 10 mM D-Glucose. All chemicals were reagent grade. The perfusate for the flushing of the intestinal vasculature (flushing perfusate) consisted of the aforementioned bicarbonate buffer with the addition of 7mg/ml of Bovine Serum Albumin (BSA), fraction V, and 5 USP/ml of Heparin. The flushing perfusate was oxygenated within an I.V. drip bag connected to a bench-top air line through a bubble trap. Air flow was controlled by a pinch clamp. The perfusate for data collection (collection perfusate) was a mixture of the flushing perfusate and human red blood cells (RBC) suspended at 20% (vol./vol.). Packed human RBC that had recently expired were obtained from Wright-Patterson Medical Center's Blood Blank. On the days of the perfusion experiments packed RBC were washed twice in bicarbonate buffer at 800 X g for 10 minutes at 4^0C prior to resuspension in flushing perfusate. Oxygenation of RBC was accomplished through a combination of gentle hand shaking and placement within the perfusion box upon a magnetic stirrer set at medium-low speed. Additionally, a solution consisting of flushing

perfusate with 20% (vol./vol.) India Ink was prepared to provide photographic evidence of the tissues perfused during experiments.

Apparatus for Isolated Vascular Perfused Rat Intestine. A plexiglass box (18" h X 18" w X 23" deep) constructed by engineers at Wright-Patterson Air Force Base contained most of the equipment needed for oxygenating and warming the collection perfusate. Heating of the box was accomplished by heating tape mounted along the interior sides of the box. A muffin fan (Radio Shack) was mounted on the side of the box and was used to circulate the warm air. In addition, a thermostat (Penn Automatic from Johnson Controls) controlled the temperature within the chamber. An Ismatec IPN model 7618-40 peristaltic pump (Cole-Parmer Instrument Comp., Chicago, IL) was engaged in circulating the collection perfusate from a 250ml beaker placed upon a magnetic stirrer. The magnetic stirrer was used to oxygenate the collection perfusate. Collection perfusate was pumped from the beaker through a course metal filter to a bubble trap (modified 20cc plastic syringe). A pressure regulator (20cc plastic syringe) was connected to the top of the bubble trap via extension tubing. Temperature of exposed organs was maintained by a visible light heating lamp (250 watt, Sunnex Corp., Needam, MA).

Surgery and Perfusion Technique. Rats were anesthetized by an intramuscular (i.m.) injection of ketamine (Ketamine-HCl, Fort Dodge, IA) and xylazine (Rompun, Mobay Corp., Shawnee, KS) mixture (90 mg/kg and 10 mg/kg body wt, respectively) into the groin area of one of the hind legs. This dose was approved by the Armstrong Laboratory Animal Care and Use Committee and the American Veterinary Medical Association (AVMA). Rats were allowed ~ 15 minutes to become sedated and then were checked for reflex response by gently pinching the tail and hind paws with broad nosed tweezers. A modified version of the isolated vascular perfused intestinal system employed by Pang et al (1985) was used for experiments {14}. Upon complete sedation rats were opened by midline and lateral abdominal incisions. Exposed organs and vasculature were exteriorized and surrounded with surgical gauze that had been soaked in warm 0.9% (wt./vol.) NaCl solution. The left renal artery and vein were ligated, and a loose tie was placed around the right renal artery and vein. The pyloric vein and bile duct

were then ligated. The superior mesenteric artery was quickly isolated by two loose ties (one tie proximal of the other). Additionally, the hepatic portal vein was isolated in the same manner. The proximal loose ties on the superior mesenteric artery and hepatic portal vein were quickly tightened. The superior mesenteric artery was cut between the distal and proximal ties, and then cannulated with PE-50 tubing (polyethylene. 0.022" i.d. X 0.042" o.d.). The cannulated superior mesenteric artery was then connected to an extension line running from the I.V. drip bag. Oxygenated flushing perfusate was circulated through the intestinal vasculature to remove blood and maintain oxygenation. A 22-gauge catheter was inserted into the hepatic portal vein for outflow of perfusion solutions. The cannula and catheter were initially secured by tightening the distal ties (later cyanoacrylate glue was used for additional securing of tubing and catheter). The chest was opened and blood flow was halted by injection of ketamine and xylazine directly into the heart, and the loose tie around the right renal artery and vein was tightened. Perfusion with the collection perfusate was begun upon completion of ligations and securing of cannula and catheter.

At this time a dose of TRI (Trichloroethylene[CAS# 79-01-6] spectrophotometric grade. 99+% pure. Aldrich Chemical Co., Inc., Milwaukee, WI) in a corn oil vehicle (Mazola corn oil) was delivered to lumen of the proximal 1/3 of the small intestine with a .25cc glass syringe. The injection site was sealed with a hand held cauterizing pen. Samples of outflowing collection perfusate from the hepatic portal vein were collected on ice to be later analyzed for TRI and metabolites.

Rats were dosed with TRI at 50, 25, or 5 mg/kg body wt.. Two rats were dosed at 50 mg/kg, one of which received 0.22 ml of a 50mg/ml TRI solution, delivering 11.0 mg of TRI. The second was dosed with 0.25 ml of a 50 mg/ml TRI solution, delivering 12.5 mg of TRI. At 25 mg/kg, a rat was dosed with 0.21 ml of TRI solution delivering 5.25 mg of TRI. At 5 mg/kg , two rats were each dosed with 0.24 ml of a 5 mg/ml TRI solution exposing the intestine to 1.2 mg of TRI.

Analytical Methods. Quantitative measurements of TRI in collected samples were analyzed following

modifications of the procedure of Chen et al (1993) {15}. A Hewllet Packard model 5890 Series II gas chromatograph fitted with an electron capture detector (GC-ECD) was employed for quantifying TRI found in collection perfusate. The GC-ECD contained a stainless steel packed column. 150ul samples of TRI exposed collection perfusate were added to 50ul of distilled H_20 within 2ml extraction vials. This was followed by 30-60 seconds of vortexing to lyse the RBC. Immediately following the lysing period. 1ml of 2,2,4-Trimethylpentane (isooctane. 99.98% purity) purchased from Aldrich (Milwaukee. WI) was added to samples. TRI was extracted into solvent phase by vortexing for 5 min. at room temperature. Samples were then centrifuged for 10 min. at 800 X g at 4°C. After centrifugation. vials were placed inside a -80°C freezer for 10 min. to solidify lipid material. Solvent phase was then transferred with glass pipettes to 2ml sample vials to be analyzed for TRI uptake. Direct injections of known quantities of TRI into isooctane were used to produce standard curves. Direct injections of known quantities of TRI into collection perfusate were used to determine extractions efficiencies for each perfusion experiment.

Additionally. samples quenched 1:2 with 20% (wt./vol.) lead acetate were analyzed to determine TCA. DCA and trichloroethanol (TCOH) formation by enzymes of the intestinal mucosa. These samples were acidified and extracted into hexane and derivitized with dimethylsulfate following modification of the procedure of Maiorino et al (1980). Analysis of TCA. DCA and TCOH was accomplished by GC-ECD. The DCA methyl ester was analyzed by GC liquid injections. Samples were injected by a Tekmar headspace analyzer. Area counts were integrated through a P.E. Nelson Turbochrome 3 data analysis system using external standards and dichloroproprionic acid as the internal standard for DCA. Disappearance of DCA was determined by subtracting the quantity of DCA remaining after the incubations from the concentration initially present.

Glove Bag. Rats were euthanized by CO_2 asphyxiation and carcasses were placed into glove bags containing a nitrogen atmosphere. The abdominal cavities of rats were opened within the confines of the glove bag by midline and lateral incisions. The intestines were removed and separated as entire small intestine. proximal small

intestine. distal small intestine. cecum. or large intestine. Separated intestinal segments were placed into individual beakers (30ml) containing enough oxygen purged .1 M KH_2PO_4, pH 7.4. buffer to keep tissue from drying out. Separated intestinal segments were milked of contents into 10 ml beakers. Gut contents were mixed prior to allocation. During milking process lightest amount of pressure was used so as not to strip out the intestinal lining. Intestinal contents were then aliquoted into Hewlett Packard 10 ml vials and sealed with metal crimp caps with Teflon septum. This was done within the glove bag to preserve the anaerobic conditions. TCA. DCA. and CH spiked into vials were prepared in .1 M KH_2PO_4 buffer. Spikes of TRI were prepared in polysorbate 20. Any volume adjustments were made with purged buffer. Incubations were carried out on a Haakebuchler Vortex-Evaporator for varying time periods at 37°C. Metabolism of spiked chemicals were halted by heat quenching vials immediately after prescribed time periods. Metabolite formation was measured using the same methods used in perfusion experiments. Studies of microflora degradation of TRI and TCA under aerobic conditions were conducting in the same manner as anaerobic studies in all respects except for elimination of glove bag and purged buffer.

Results

The oxygenation process used during perfusions was validated with the use of an Instrumentation Labs IL-282 Co-oximeter. Oxyhemoglobin was found to be at 96.5%. It was determined that no further oxygenation procedures were required. Additionally. the methods used for washing and preparing RBC suspension in flushing perfusate was shown to yield ~20% suspension of packed RBC in the flushing perfusate. This was verified by the use of Techtron Hematology System.

The photographs of figure 1 show the vasculature perfused during these experiments. The vasculature of the small intestine and the cecum were shown to be the organs of perfusion for our experimentation. Slides of small intestine cross sections before and after use of the flushing perfusate were stained using H and EM stain. Blood was visible in the vessels prior to circulation of the flushing perfusate. After use of flushing perfusate there. was little evidence of rat blood in the vasculature and the structural viability of the vasculature was maintained. In

future experiments, radiolabelled polyethylene glycol (PEG) injected into the lumen of the small intestine will be used to determine whether the tight junctions of the intestinal endothelium remain intact during perfusion procedures. PEG should not be detectable in effluent collected from the hepatic portal vein if endothelium junctions are maintained throughout perfusions. Also, the viability of the intestine will be evaluated by its ability to utilize glucose introduced to the system. This may also help in finding the the upper range of time limits that can be approached before the system begins to loose its viability.

Metabolite Formation. In two of the five perfusion studies, there was an indication of metabolite formation. TCOH was determined to be present in the effluent collected from the rat exposed to 11mg of TRI. The range of TCOH found was from 0.30 to 1.46ug/ml. Additionally, TCOH was found in the effluent collected from the intestine of a rat exposed to 1.2mg of TRI at ranges from n.d. to 0.31ug/ml. Complete analysis of this data is not given because it is believed that the detection of TCOH was due to the representative samples not being quenched with 20% lead acetate prior to analysis for metabolite formation. In the other three perfusion experiments, allocated samples that were to be tested for metabolite formation were treated with 20% lead acetate. There was no indication of TCOH formation. TCA and DCA levels if at all present in any samples for the perfusion experiments could be accounted for from controls.

Absorption of TRI. With a dose of 95.129 nmol (12.5 mg) of TRI injected into the lumen of the small intestine, the cumulative uptake of TRI was found to be 88.974 nmol after 20.17 min.. This is representative of 0.094% of the administered dose. As figure 2 depicts, the uptake has not leveled off. However, perfusions were halted due to insufficient amounts of collection perfusate to continue the experiments. Flow rate of the effluent was observed to be 5.5 ml/min, and remained constant throughout the perfusion. Dosing with 83.714 nmol (11 mg) of TRI with ~ the same effluent flow rate yielded a cumulative uptake of 10.335 nmol of TRI for 10 min (figure 3). This accounted for only 0.012% of the administered dose.

In two rats dosed at 5 mg/kg body wt., the cumulative uptake of TRI was 0.032% for 11 min. and 0.008% for 30

min. in two separate perfusion studies. However. the 9.132.4 nmol dose was subjected to effluent flow rates of 3.6 ml/min. and 0.67 ml/min. The greater percentage of uptake was under the influence of the higher flow rate. The cumulative uptake for the perfusion lasting 11 min. was 2.952 nmol while that for the perfusion lasting 30 min. was 0.737 nmol (figures 4 and 5).

TRI was also dosed at 25 mg/kg body wt.. The cumulative absorption of TRI was measured at 15.351nmol for samples that did not have the addition of distilled H_2O prior to lysing. Samples with the addition of 50ul of distilled H_2O before lysing had a cumulative uptake of 10.995 nmol of TRI (figure 6). This translates into 0.038% and 0.028% of the 39.954 nmol dose of TRI introduced into the small intestine. respectively. The flow rate as observed in the collected effluent was 2.35 ml/min. on average. During onset of collections. effluent flow rate was calculated at 3.7 ml/min.. but by the end of the 60 minutes it had fallen to 1 ml/min.

Aerobic Microflora. Within the aerobic studies to determine TRI degradation by the microflora. the gut contents of the small intestine were distinguished as being from the proximal. distal. or complete small intestine. Additionally. the gut contents of the cecum were also analyzed for TRI degradation. All gut content samples were adjusted to 300mg/ml. Incubations were run for 30. 60. and 90 minutes at TRI doses of either 10 mg or 0.5 mg. In experiments studying the metabolic capabilities of the microflora to degrade TRI under aerobic conditions. there was no reported formation of TCA. DCA. or TCOH.

Experiments were also undertaken to determine the breakdown of TCA by microflora under aerobic conditions. Gut contents were recognized as being milked from the complete small intestine or cecum. Contents were adjusted to 300mg/ml and spiked with 20ug/ml of TCA. There was no detectable formation of DCA or TCOH from samples representing the small intestine. However. in three samples harvested from the cecum and incubated for 30 min. there was a mean DCA formation of 0.60ug/ml with a range from 0.55 to 0.68 ug/ml. Additionally. in one cecum samples incubated for 60 min. there was 0.21 and 0.24 ug/ml of DCA and TCOH formed. respectively.

Anaerobic Microflora. Gut contents were incubated in the presence of corresponding tissue segments under anaerobic conditions. Tissue segments and gut contents were from the distal small intestine or cecum . Samples were exposed to either 10 or 0.5 mg of TRI for 15 or 30 minutes incubations. There was no detectable formation of TCA. DCA. or TCOH metabolites. In a companion study. gut contents exposed to 0.5 mg of TRI were incubated for 15. 30. 60. and 90 min. without intestinal tissue segments. Formation of low levels of DCA in the samples from the distal small intestine and cecum were observed. Although all samples had been adjusted to 300mg/ml replicates for the same incubation times did not have similar DCA levels. Notable amounts of DCA were formed in distal small intestine contents (0.49 mg/ml at 60 min.) and in cecum samples (0.29 ug/ml at 60 min. and 0.30 ug/ml at 90 min.).

Initial anaerobic experiments involved the spiking of 20 ug/ml of TCA. DCA. and CH into the gut contents milked from the proximal small intestine. distal small intestine. cecum. or large intestine. These samples were brought to 1ml and incubated for 30 or 60 min. before heat inactivation. Spikes of TCA into cecum and distal small intestine samples incubated for 60 min. resulted in the formation of DCA. The largest amount of DCA formed in the 60 min. incubation was 10.38 ug/ml in a cecum sample. The amount of DCA formed seemed to be dependent upon the amount of gut contents (figure 7). The relationship between the amount of gut contents and DCA formation was also observed in the 30 min. incubations (figure 8). Here the highest level of DCA formed was 8.33 ug/ml from a cecum sample. Additionally. at 60 min. incubations. spikes of CH yielded TCOH levels of 9.39 ug/ml in a cecum sample. and 7.64 ug/ml in a large intestine sample. Incubations of 30 min. with CH yielded TCOH at 3.5 and 0.77 ug/ml in cecum samples of 248 and 266 mg/ml. In large intestine samples. TCOH was observed at 2.06 and 0.73 ug/ml for 102 and 167 mg/ml.

In follow up studies looking exclusively at the breakdown of TCA and formation of DCA. gut contents of the proximal small intestine. distal small intestine. and cecum were adjusted to 300mg/ml. Some cecum samples were adjusted to 150 mg/ml. All samples were spiked with TCA at 20ug/ml. and then incubated for 30 or 60 min. at 37°C. The average amount of DCA formed for replicate samples was plotted against time (figure 9). As expected

the largest amounts of DCA formed came from cecum samples incubated for 60 min. Low amounts of DCA (< 0.165 ug/ml) were observed in samples from the proximal and distal small intestine. Additionally, the amount of TCA degraded by the gut contents was plotted against time. Figure 10 shows that cecum samples adjusted to 300 mg/ml accounted for the greatest amount of TCA degradation at respective time points.

Discussion

TRI has been found in drinking water supplies at a levels of 130 ppb and in contaminated wells at levels as high as 27,000 ppb {1}. This clearly presents a potential for repeated oral exposure for individuals who use these contaminated water supplies. In studies involving gavage dosing of test animals, it is assumed that 100% of the dose is absorbed as the administered parent compound without the formation of metabolites before reaching the systemic circulation and delivery to target organs. In many gavage studies, observations are made concerning the hepatic effect of the administered chemical without taking into consideration the intestinal formation of metabolites, some of which may be more toxic than the parent compound.

However, previously cited experiments with isolated perfused intestinal systems have shown that the intestine can play a role in the formation of metabolites (9,10). The use of an isolated vascular perfused intestinal system used in this study could not account for the formation of the TRI metabolites TCA, DCA, or TCOH by the enzymes of the intestinal mucosa. In addition, the amount of TRI uptake was observed not to excede more than 0.01% of the nmol dose that was administered . The low cumulative uptake of TRI and the lack of metabolite formation in rats may effect the intestine as a compartment within PBPK models. This low uptake and lack of metabolism in the rat small intestine may contribute to the species differences in susceptibility to the carcinogenic effects of TRI that exists between rats and mice {16}. Experimentation with glucose consumption and radiolabelled PEG will further validate the techniques used and results observed in these perfusion experiments. Additionally, research with S9 fractions prepared from intestinal tissues of mice, rats, and humans may support these findings, and detail interspecies difference that exist concerning metabolite formation. The use of a pressure meter and regulator coupled with a peristaltic pump of greater sensitivity and capacity would help in the maintenance of

normal arterial pressure and insure consistent effluent flow rates.

There is strong evidence from studies with the gut contents. that the microflora of the intestinal tract can degrade TCA into DCA and CH into TCOH. Microbes associated with the root structures of vegetation have been shown to be capable of degrading TRI {11}. Additionally, there was some indication of aerobic metabolism of TCA in the cecum. Because of the low levels encountered. and the anaerobic conditions normally present. further experimentation will need to be done to validate this occurrence. TCA arising from TRI in the liver and reintroduced into the gut via the bile duct could be acted upon by the microflora of the gut. The role of the microflora in the disposition of TRI and its metabolites is an area needing further research. Refinements of the techniques used in this study. such as the use of an anaerobic indicator. could further validate these studies by insuring that proper conditions were met throughout the experiments. The microflora of the cecum. a structure not present in humans. made the most noticeable metabolic contributions. This is in line with the abundance of microorganism in the cecum and the important caloric contributions they make to their host {13.17}. Future studies with rats treated with antibiotics to deplete the microflora. and preparation of S9 fractions from the cecum may further define the metabolic contributions of the microflora.

We have established and validated methods by which the role of the intestines in metabolism and absorption of TRI can be qualified and quantified. Further research with the microflora of the intestinal tract may account for different ways in which metabolites of TRI can be formed. These two lines of experiments will enable a greater understanding of the distribution of TRI and its metabolites within humans. This will lead to risk predictions of greater accuracy.

Acknowledgment

I would like to give thanks to Capt. J.C. Lipscomb for allowing me the opportunity to work with him and to learn from him. Also. his help in preparing this report was greatly appreciated. Deirdre Mahle is also due some kudos and thanks for as much as I bothered her. I would also like to give a nod to Carol Huffman who saved my life many times from monster rats bent on my destruction.

References

1. U.S. Department of Health and Human Services. 1991. Toxicology Profile for Trichloroethylene. Agency for Toxic Substances and Disease Registry. Atlanta. GA.

2. Chemical Marketing Reporter. January 27. 1986. Chemical Profile-Trichloroethylene.

3. International Agency for Research on Cancer. 1979.Trichloroethylene. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Lyon. France. 20: 545.

4. National Institute for Occupational Safety and Health. Special Occupational Hazard Review of Trichloroethylene. DHEW Publ. No. 78-130. Washington. DC

5. Bruckner. J.V.. Davis. B.D.. and Blancato. J.N. 1989. Metabolism. toxicity. and carcinogenicity of trichloroethylene. Critical Reviews in Toxicology. 20: 31-50.

6. Dekant. W.. Metzler. M.. and Henschler. D. 1984. Novel metabolites of trichloroethylene through dechlorination reactions in rats. mice. and humans. Biochem. Pharmacology. 33: 2021-2027.

7. Allen. B.C. and Fisher. J.W. 1993. Pharmacokinetic modeling of trichloroethylene and trichloroacetic acid in humans. Risk Analysis. 13: 71-86.

8. Environmental Protection Agency. Health Assessment Document for Trichloroethylene. EPA/600/8-82/006F. Final Report. RTP. NC. 1985.

9. Paula Adams. unpublished. Glaxco Pharmaceuticals. Inc.. RTP. NC and the University of North Carolina. Chapel Hill. NC. 1994.

10. Hirayama. H. and Pang. K.S. 1990. First pass metabolism of gentisamide: influence of intestinal metabolism on hepatic formation of conjugates. Drug Metab. Dispos. 18: 580-587.

11. Walton. B.T. and Anderson. T.A. 1990. Microbial degradation of trichloroethylene in the rhizosphere: potential application to biological remediation of waste sites. Applied and Environmental Microbiology. 56: 1012-1016.

12. Sariaslani. F.S. 1991. Microbial cytochrome p-450 and xenobiotic metabolism. Advances in Applied Microbiology. 36: 133-178.

13. Rowland. I.R. 1988. Interactions of the gut microflora and the host in toxicology. Toxicologic Pathology. 16: 147-153.

14. Pang. K.S.. Cherry. W.F.. and Ulm. E.H. 1985. Disposition of enalapril in the perfused rat intestine-liver preparation: absorption. metabolism. and first-pass effect. Drug Metab. Dispos. 233: 788-795.

15. Chen. X.M.. Dallas. C.E.. Muralidhara. S.. Srivatan. V.. and Bruckner. J.V. 1993. Analyses of volatile C_2 haloethanes and haloethenes in tissues: sample preparation and extraction. J. Chromatogr. 619: 199-208.

16. Larson. J.L. and Bull. R.J. 1992. Species differences in the metabolism of trichloroethylene to the carcinogenic metabolites trichloroacetate and dichloroacetate. Toxic. Appl. Pharmacol. 115: 278-285.

17. Gerrity. T.R. and Henry. C.J. ed. 1990. Principles of route-to-route extrapolation for risk assessment. Elsevier Science Publishing Co.. Inc. New York.

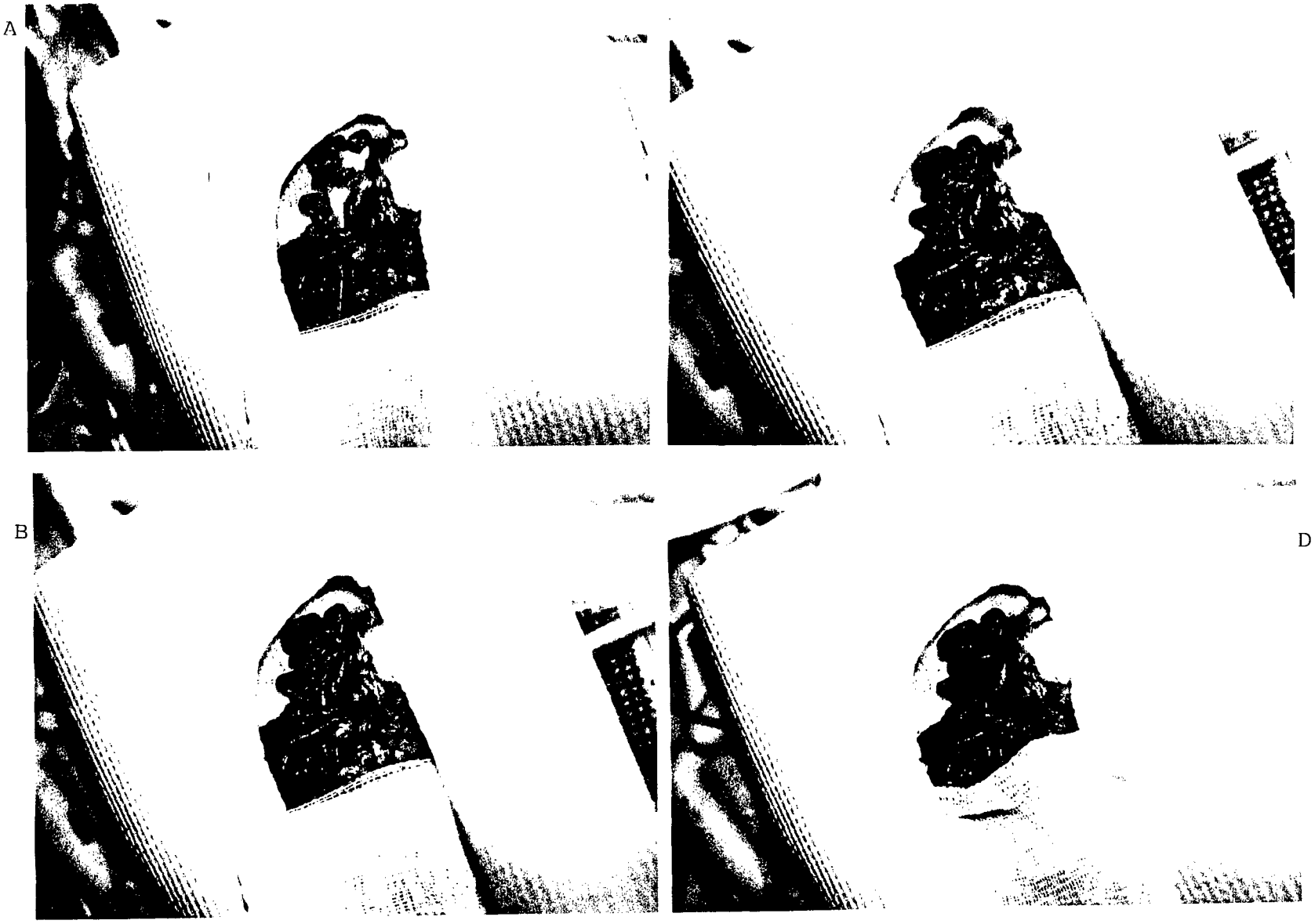


Figure 1. Photographs of perfusions with 20% India Ink. Sequence is A,B,C,D.

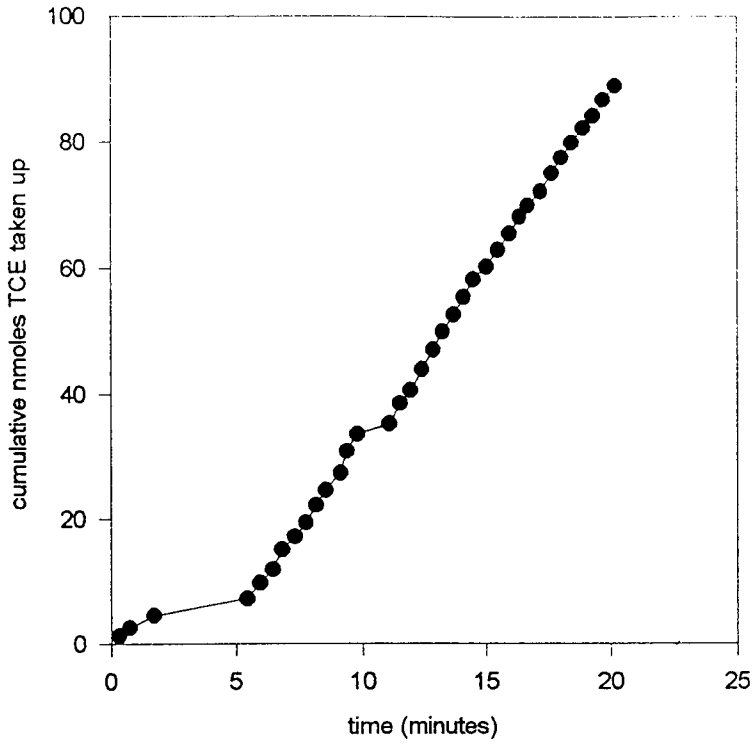


Figure 2. Cumulative uptake of TRI from a 95,129 nmol dose.

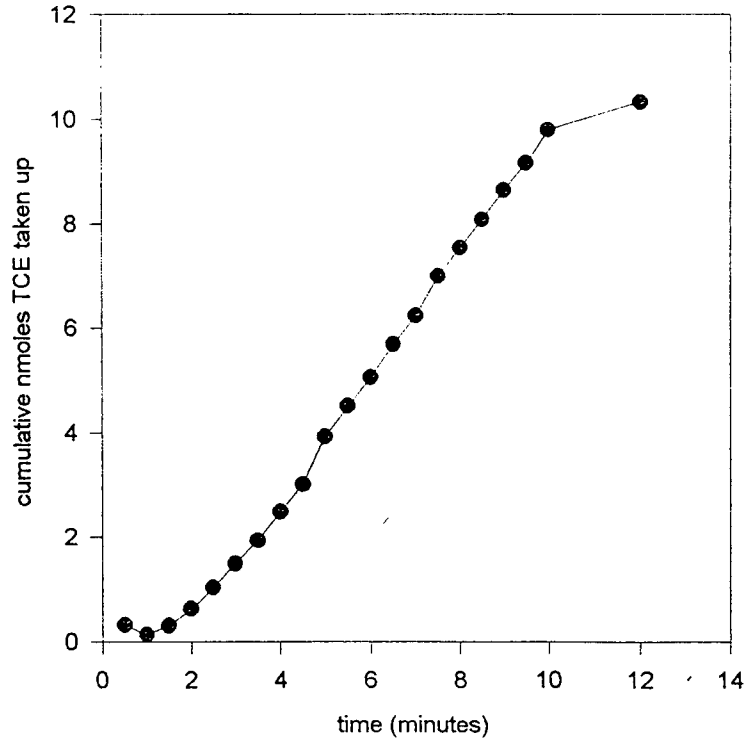


Figure 3. Cumulative uptake of TRI from a 83,714 nmol dose.

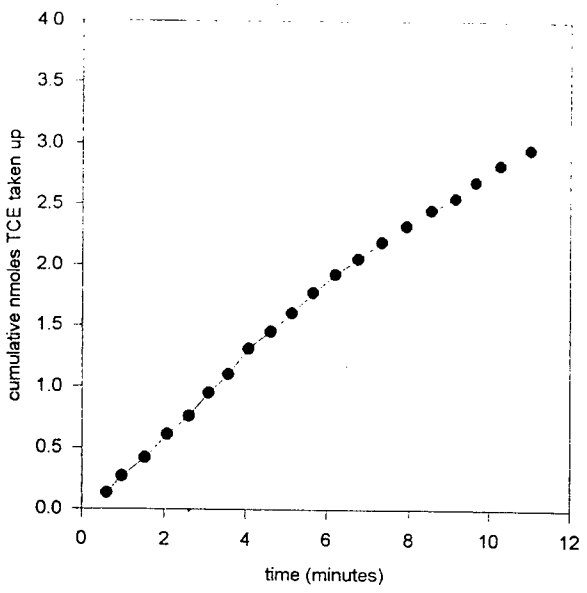


Figure 4. Cumulative uptake of TRIfrom a 9,132.4 nmol dose.

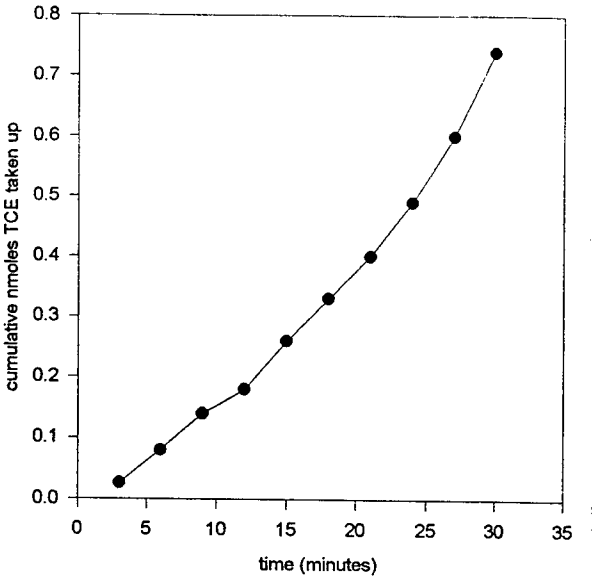


Figure 5. Cumulative uptake of TRI from a 9,132.4 nmol dose.

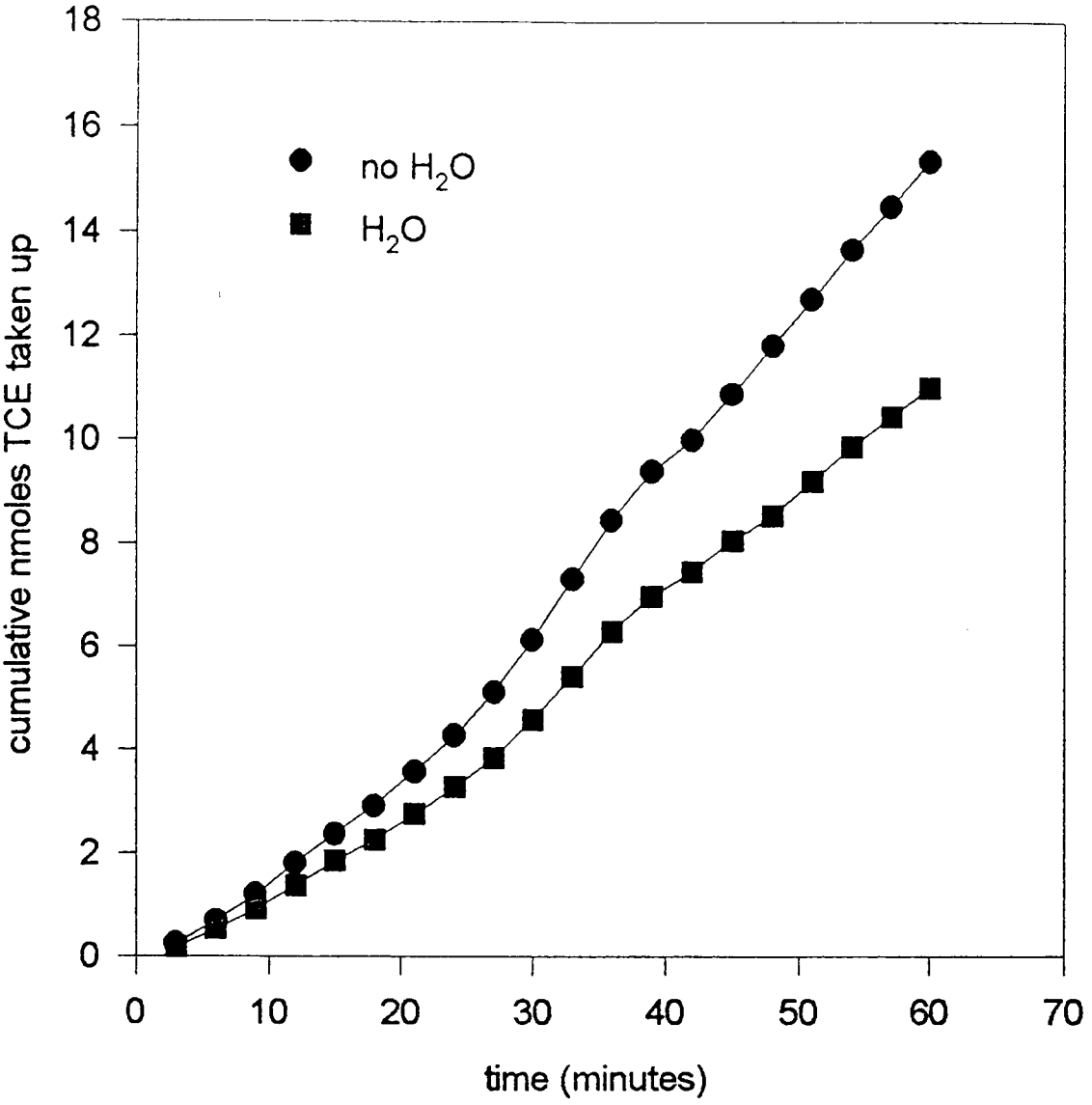


Figure 6. Cumulative uptake of TRI from two separate procedures to lyse red blood cells prior to extraction into solvent. Intestine was dosed with 39,954 nmol of TRI.

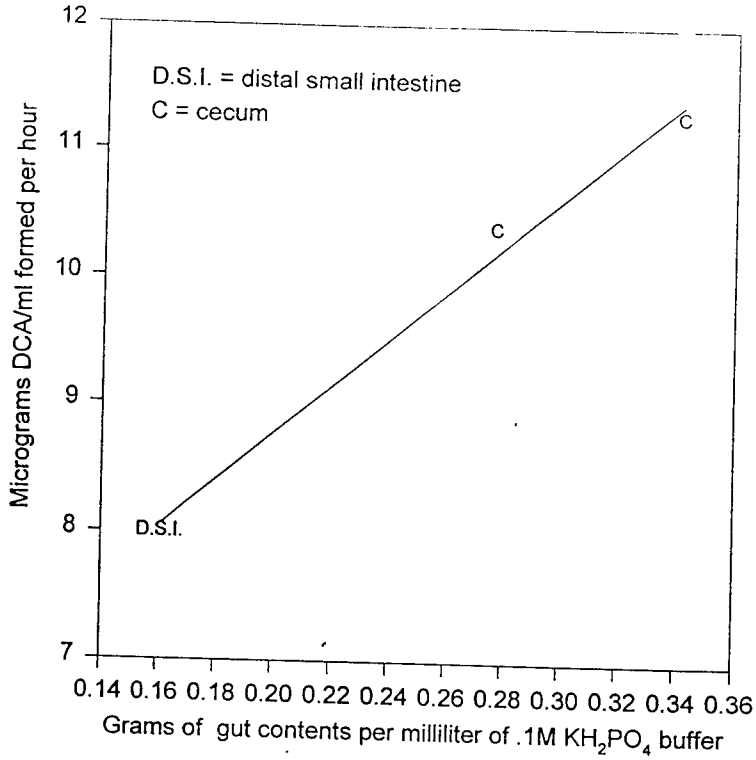


Figure 7. DCA formation from spikes of 20 ug/ml of TCA and then incubated for 60 min. at 37°C.

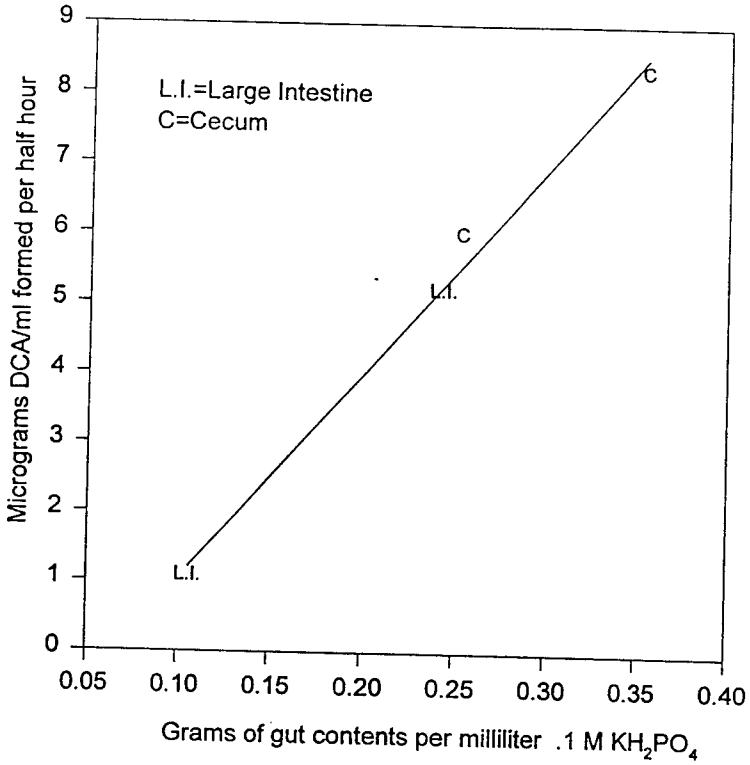


Figure 8. DCA formation from spikes of 20 ug/ml of TCA followed by 30 min. incubation at 37°C.

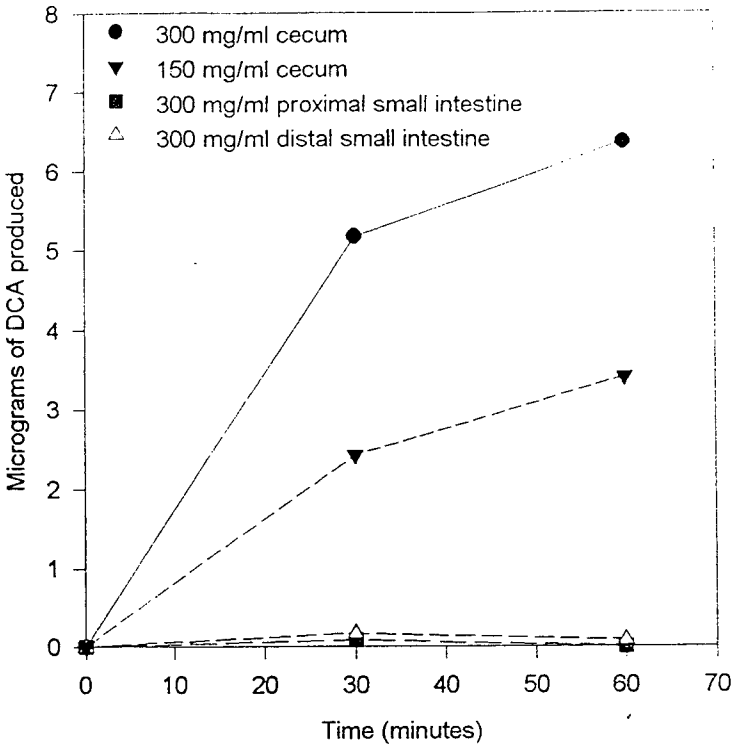


Figure 9. DCA formation from 20 ug/ml spikes of TCA

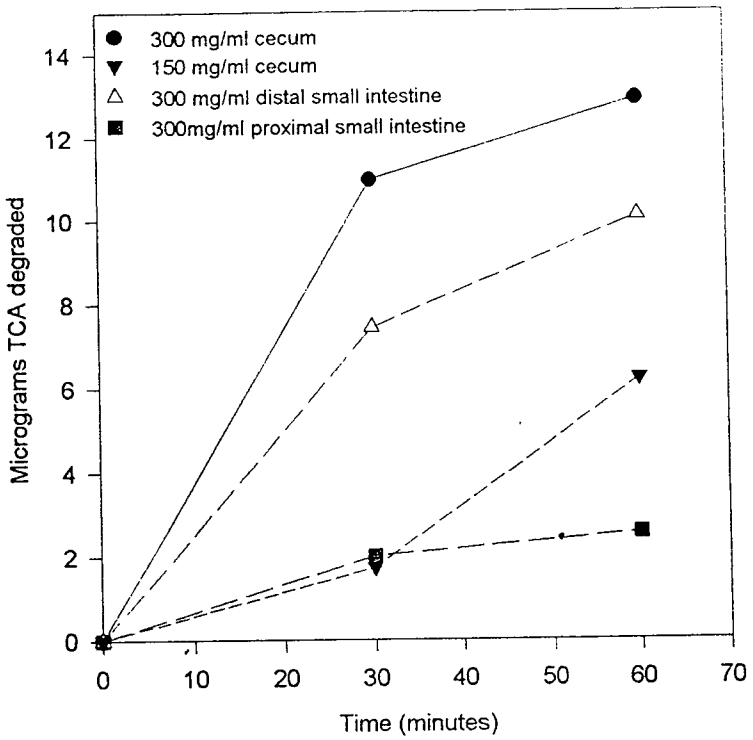


Figure 10. Amount of TCA degradation observed in spiked cecum and small intestine contents.

SOLID PHASE MICROEXTRACTION AS A METHOD FOR QUANTIFYING JET FUEL CONTAMINATION IN WATER

Virginia K. Stromquist
Graduate Student
Department of Civil and Environmental Engineering

Michigan Technological University
1400 Townsend Dr.
Houghton, MI 49931-1295

Final Report for:
Summer Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

SOLID PHASE MICROEXTRACTION AS A METHOD FOR QUANTIFYING JET FUEL CONTAMINATION IN WATER

Virginia K. Stromquist
Graduate Student
Department of Civil and Environmental Engineering
Michigan Technological University

Abstract

Solid-phase microextraction (SPME) with capillary gas chromatography (GC) was evaluated as a method for quantifying jet fuel contamination in groundwater. Solid-phase microextracts were analyzed by gas chromatography using a split/splitless injection port, a fused silica capillary column (10m length, 0.10mm internal diam., 0.34μm HP-5 stationary phase) and a flame ionization detector (FID). Several components of jet fuels thought to be soluble in water were evaluated in depth. Water soluble fractions of benzene, toluene, ethylbenzene, and m-xylene gave linear responses and were quantifiable over the range of 10-1000 ppb. Water soluble fractions of n-butylbenzene and n-propylbenzene demonstrated potential for quantification over a larger range and were detectable at a concentration of 1 ppt. The effects of increasing the salinity of the sample solution on analyte response were investigated. Introduction of internal standards to the water soluble fraction of jet fuel was shown to be possible and necessary for quantification. The variation of response with thickness of the extracting fiber was investigated using fibers with 7, 20, and 100 μm thicknesses of polydimethylsiloxane (PDMS). The water-soluble fraction of JP-8 jet fuel was obtained by equilibrating the fuel with water. SPME combined with GC was used to analyze the resulting aqueous samples.

SOLID PHASE MICROEXTRACTION AS A METHOD FOR QUANTIFYING JET FUEL CONTAMINATION IN WATER

Virginia K. Stromquist

INTRODUCTION

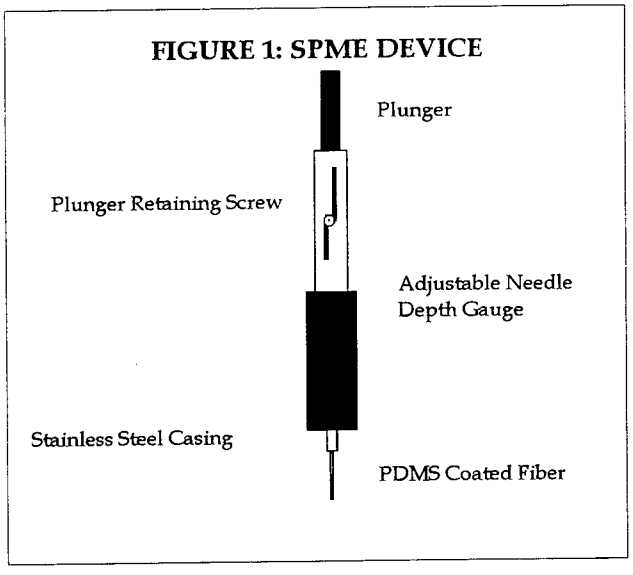
Groundwater contamination from leaking storage tanks and jet fuel spills is a major concern of the Air Force. In order to remediate groundwater contamination and to meet environmental regulations, it is necessary to know the type and extent of contamination in the water. To date, there are no methods that simply and efficiently provide this information for jet fuel contamination. Purge and trap is the current method used to assess dissolved hydrocarbons in water. However, it has drawbacks. Compounds in jet fuels that have low volatility are not well characterized by purge and trap methods[1]. Purge and trap also expends quite a bit of solvent and is relatively time consuming. For these reasons, it is necessary and beneficial to find new methods to replace or augment purge and trap.

Solid-phase microextraction (SPME) is a relatively new technique used to qualitatively identify organic compounds in water. It has been used to identify caffeine in beverages[2], organic contaminants in drinking water and wastewater[3], and to quantify BTEX compounds in solvent and water[4]. In this study, solid-phase microextraction was investigated as a method for quantifying water the soluble fraction of jet fuels in water.

THEORY

SPME is a very quick and simple technique. It involves the use of a silica fiber coated with polydimethylsiloxane (PDMS) attached to a syringe and encased in stainless steel tubing. (See Figure 1.) When the coated fiber is exposed to an aqueous sample, organic compounds are absorbed into the PDMS. (Head space analysis is also

FIGURE 1: SPME DEVICE



effective with SPME[3], although this was not explored here.) Diffusion governs the migration of the analytes to the stationary phase and distribution coefficients of organic analytes determine the extraction efficiency of the PDMS for each analyte[5]. After exposure to a sample, the fiber is placed into the gas chromatograph (GC) injection port. Desorption of the fiber occurs at a high temperature the organics move through the column of the GC, just as in a regular liquid injection.

EXPERIMENTAL

QUANTIFICATION, SALINITY, AND FUEL EQUILIBRATION EXPERIMENTS

Materials

SPME device and 7, 20, and 100 μm PDMS coated fibers from SUPELCO. Forty milliliter vials with screw tops and septa. Ten, 250, and 1000 μL Hamilton syringes. Benzene, toluene, ethylbenzene, and m-xylene (BTEX), D10 ethylbenzene, n-butylbenzene, n-propylbenzene, and methanol. JP-8 jet fuel. Two equilibration vessels (see Figure 6) fitted with extra bottom sampling arms. HP 5890 GC with flame ionization detector (FID) and subambient temperature control, equipped with a fused silica capillary column, 10 m long with internal diameter of 0.11 mm and coated with 0.34 μm HP-5.

Methods

Two standards (BTEX and n-propyl and n-butylbenzene) were prepared in methanol in 2 mL vials. For each experiment, 10 μL of a standard were injected with a syringe into 40 mL of distilled/deionized water. For the salinity experiment, 1.5%, 2.5%, 3.5%, and 4.5% solutions of NaCl in distilled/deionized water were prepared and the BTEX standard was injected into 40 mL of NaCl/water solution. D10 Ethylbenzene was added to the samples as an internal standard to produce a concentration of 500 ppb for the quantification experiment and 200 ppb for the salinity experiment. The samples were then stirred rapidly for 10 minutes. The SPME fiber was inserted through the septum of the 40 mL vial and exposed for 5 minutes. The water soluble fraction of JP-8 was obtained by using the approaches of Mayfield and Henley, 1991 with slight modifications. D10 Ethylbenzene was prepared in methanol and spiked into 250 mL of distilled/deionized water to produce a concentration of 1 ppm. The spiked water was added to the equilibration vessel. (Refer to Figure 6.) A delivery tube was placed into the vessel so that its end was submerged in the water. Two milliliters of JP-8 was added to the vessel with a syringe through the top arm. The fuel was stirred very gently with a minute stir bar for 18

hours. Sampling was done by inserting the SPME fiber through the septum in the bottom flask arm or by extracting from a vial containing an aliquot of the equilibrated water. Aliquots were taken from the equilibration vessel through the delivery tube by introducing air through the top arm of the vessel with a syringe. All sample extracts were immediately injected into the GC. For the salinity and quantification experiments, the GC program was as follows: initial oven temperature -10° C held for 3 min, oven ramped at 18° C/min to final temperature of 150° C and held for 1 min. The injection port temperature was held at 250° C. The injection port was purged with purge flow of 30 mL/min of helium, which was interrupted for 3 min during each injection to produce splitless injections. For the fuel equilibration experiment, the GC program was the same as above with the following modifications: oven ramped at 12° C/min to final temperature of 150° C and held for 5.67 min.

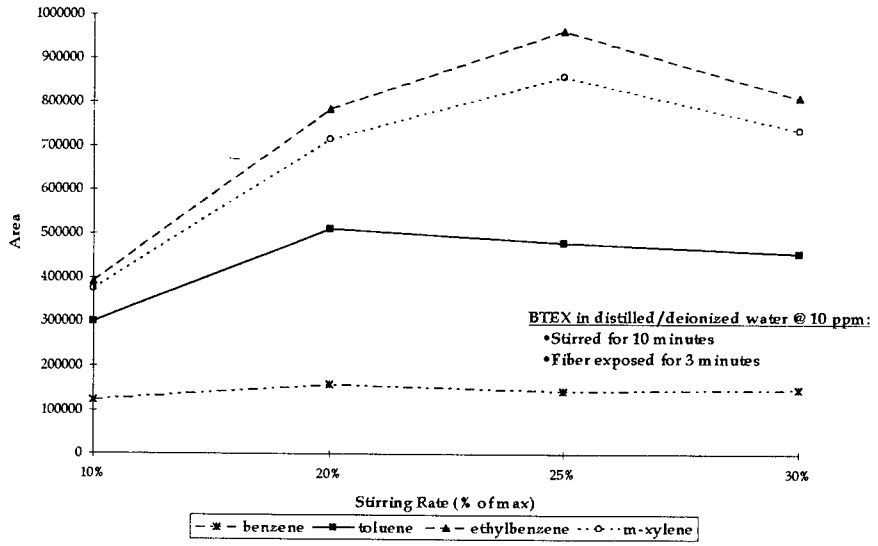
RESULTS

Quantification and Salt Addition

Sample Preparation

Since the extraction of organics with SPME is dependent on diffusion and analyte distribution coefficients, it is imperative that samples be at equilibrium before extraction. Thorough mixing is a necessity if reproducible results are to be obtained. Figure 2 shows the change in response for BTEX with increasing stirring rates. As seen in Figure 2, there is a mixing rate for each compound at which its response is at a maximum. For this study, the 20% mixing rate was used throughout. Figure 2 may be compared with results obtained by Pawliszyn, *et. al.* 1992, p. 1962, Figure 3.

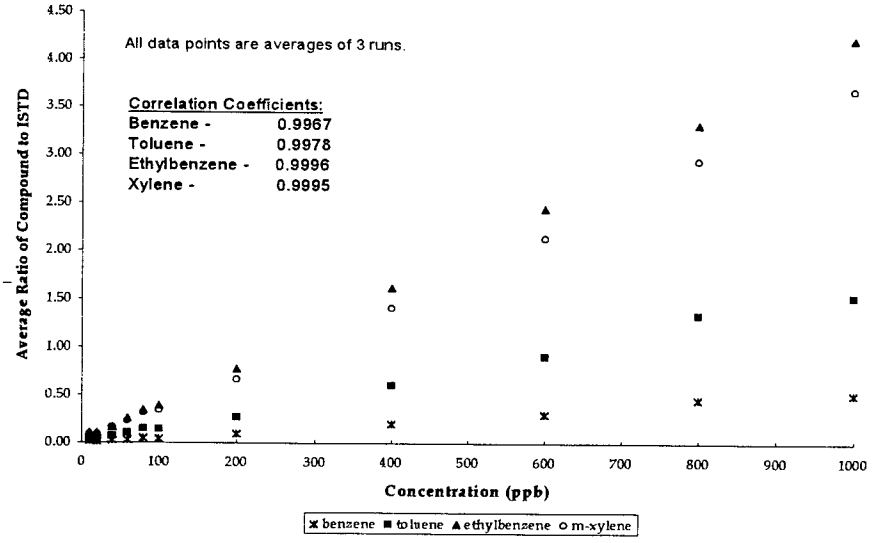
FIGURE 2:
Effect of Stirring Rate on Amount of Analyte Absorbed



Internal Standards

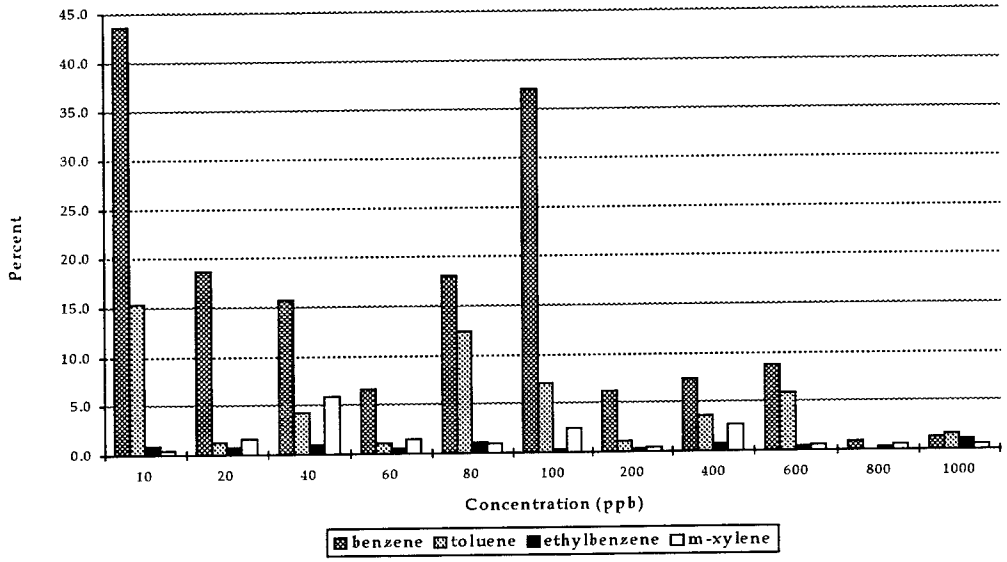
Reproducibility of peak areas is very difficult to achieve with SPME, due to the strong dependencies on time of extraction, stirring efficiency, and fiber placement in the sample (i.e., sample container geometry). Absolute peak areas of all compounds examined were not reproducible. However, ratios of analyte response to internal standard response were reproducible. Internal standards (ISTDs) are absolutely necessary in samples extracted with SPME. Benzene, toluene, ethylbenzene, and xylene were quantifiable in this experiment with SPME from 10-1000 ppb, when D10 ethylbenzene was used as an ISTD. (See Figure 3.)

FIGURE 3:
Average Ratios of Each Compound to D10 Ethylbenzene (ISTD)



Benzene and toluene are not as readily extracted with the fiber as ethylbenzene and xylene, due to their higher solubility in water. This is illustrated clearly in Figure 4. This figure shows normalized standard deviations for the data shown in Figure 3. The normalized standard deviations of m-xylene and ethylbenzene are, for the most part, below 5%. However, benzene and toluene data varies to a much greater extent, as high as 43% in one case. For higher concentrations, reproducibility increases for the toluene and benzene peaks.

FIGURE 4:
Normalized Standard Deviation of BTEX Ratios to D10 Ethylbenzene



In order to improve extraction of benzene and toluene, the effect of increasing the salinity of the distilled/deionized water to force the compounds onto the stationary phase was investigated. The absolute response of BTEX and D10 ethylbenzene increased with concentrations of NaCl up to 3.5%. (Figure 5) The addition of NaCl did not effect the ratios of BTEX to the ISTD. (Table 1) Therefore, it is still possible to quantify the compounds when NaCl has been added to the solution.

TABLE 1: Average Ratios of BTEX to D10 Ethylbenzene with NaCl in Solution

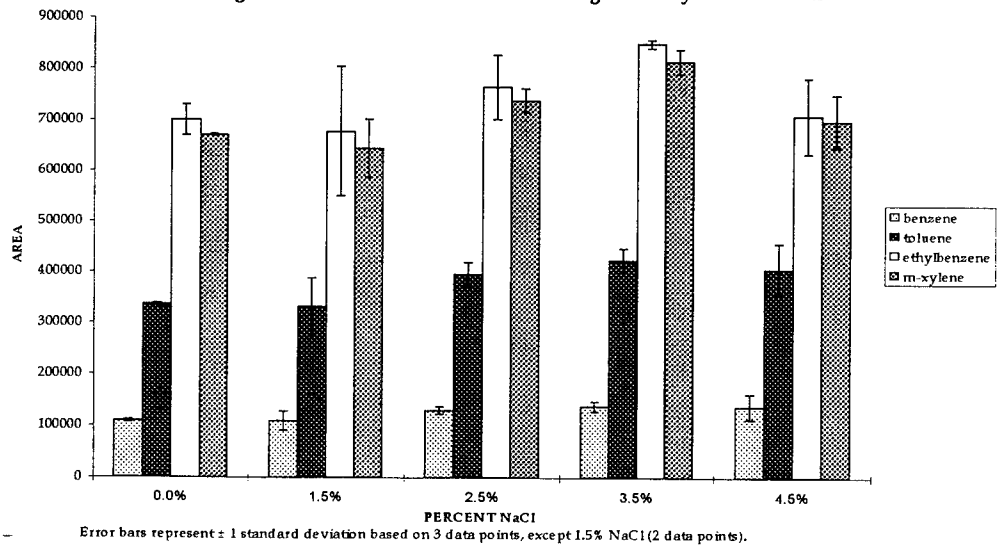
RATIOS	0.0%	1.5%	2.5%	3.5%	4.5%	STDEV
benzene	0.63	0.64	0.62	0.65	0.56	0.04
toluene	0.20	0.21	0.20	0.21	0.19	0.01
ethylbenzene	0.10	0.10	0.11	0.10	0.11	0.00
m-xylene	0.10	0.11	0.11	0.11	0.11	0.00

All averages represent 3 data points, except 1.5% NaCl (2 data points).

Responses for benzene and toluene dropped significantly at 10 ppb in the quantification experiment. Adding NaCl to the sample before extraction could result in better quantification of benzene and toluene at these lower concentrations. Another important thing to note about benzene and toluene peaks obtained with SPME: the peaks are very broad compared to those of ethylbenzene and toluene. It is necessary to use oven cryogenics to sharpened the peaks in order to get satisfactory integration results. However, if benzene and toluene are not being analyzed, cryogenics is not necessary, since the ethylbenzene and xylene peaks are sharp enough with the column held at ambient temperature during desorbtion of the fiber in order to get satisfactory integration results. However, if benzene and toluene are not being analyzed, cryogenics is not necessary, since the ethylbenzene and xylene peaks are sharp enough with the column held at ambient temperature during the desorbtion of the fiber.

N-butylbenzene and n-propylbenzene were also examined with SPME. Both compounds showed potential for quantification over the range of 1 ppt - 10 ppm. A chromatogram of the sample spiked at 1 ppt showed peaks for both n-butylbenzene and n-propylbenzene. However, impurities in the methanol (Fisher, HPLC Grade) were at a higher concentration than the analytes, making it difficult for proper peak integration.

FIGURE 5:
Average Areas of BTEX with Increasing Salinity in Solution



Error bars represent ± 1 standard deviation based on 3 data points, except 1.5% NaCl (2 data points).

Fuel Equilibration

Fiber Coating Thickness

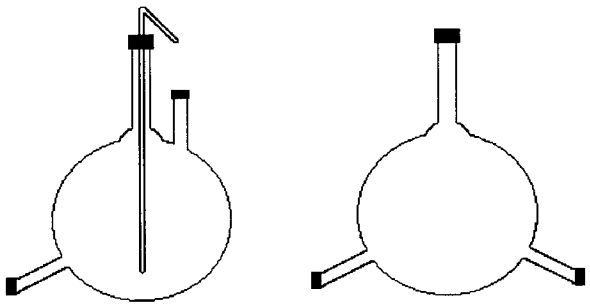
Three thicknesses (7, 20, and 100 μm) of the PDMS fiber coating were tested. The 100 μm coating was a non-bonded stationary phase. This fiber gave excellent analyte response on the

chromatograms, due to its ability to extract a larger mass of the analytes. However, in the equilibrated fuel experiments, the 100 μm coating became saturated with hydrocarbons, swelled, and slid off the silica fiber. The 7 μm coating of PDMS is a bonded phase. The response of this fiber was less than satisfactory. Responses with the 7 μm coating are lower than those of the 100 μm coating by approximately a factor of 10. Also, the 7 μm coating is not suitable for extracting hydrocarbons with 7 or less carbon atoms[6]. Chromatograms of the water soluble fraction of JP-8 created with the 7 μm fiber lacked the toluene peaks seen with the 100 μm coated fiber. The 20 μm bonded phase fiber gave analyte responses similar to that of the 7 μm, although it was expected that the responses would increase with the larger coating thickness. The 20 μm fiber did, however, extract toluene from the water soluble fraction of the fuel, where the 7 μm fiber did not.

Extraction Efficiency of Water Soluble Fractions

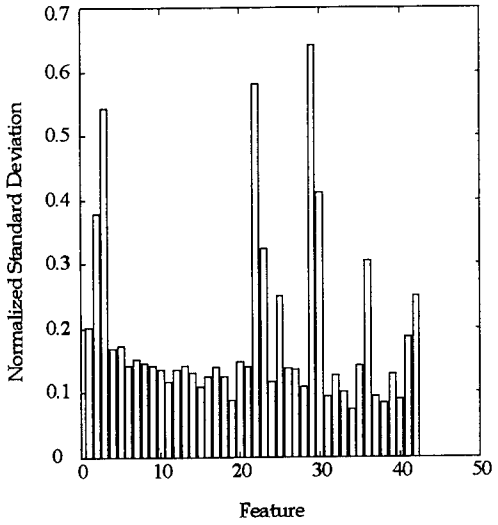
SPME proved to be effective for extracting the water soluble fraction of JP-8 from the equilibration vessel. Two flask types were used in the experiment. (Figure 6.) Peak areas of the water soluble compounds seemed to vary significantly from flask to flask. Peak areas from Flask A were significantly lower than those of Flask B. A *t-test* done on data from the two flasks showed that for 38 of the 42 features, the mean of the peak area of Flask A was less than that of Flask B. This is congruent with similar findings in the quantification experiment, where differences in peak areas were seen among 40 mL vials that were not of precisely the same volume. Differences in the fuel equilibration experiment are greater, though, presumably due to the difference in mixing efficiency between the two flasks. Flask B has an extra bottom arm that conceivably creates a "dead space" for mixing. Also, Flask A has an extra top arm that provides a larger volume of head space in the

FIGURE 6: Fuel Equilibration Flasks



Flask A Flask B

FIGURE 7: Normalized Standard Deviation of Flask B

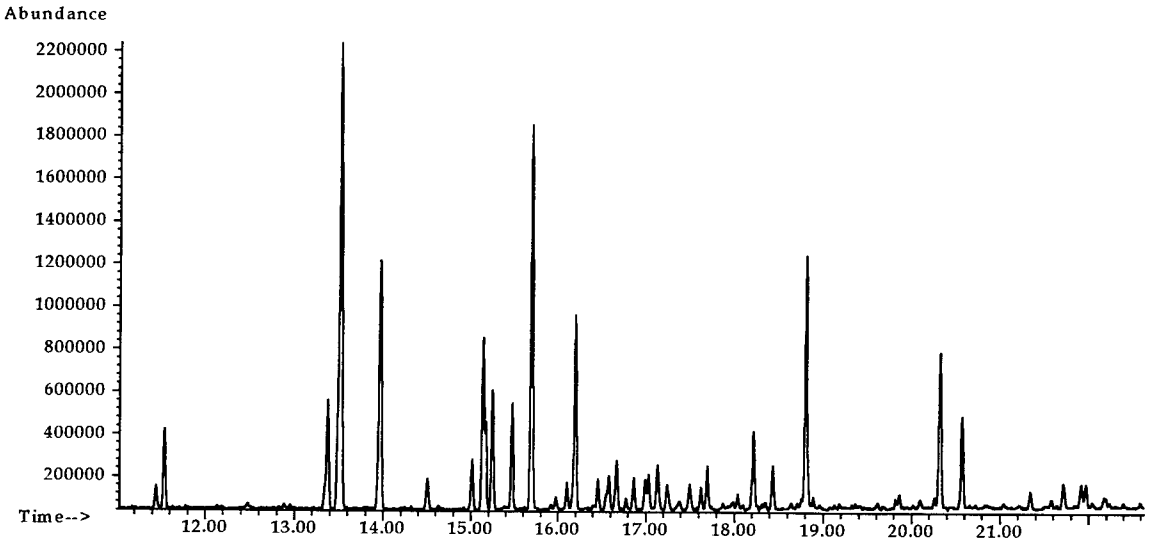


Values based on 5 data points

vessel. Figure 7 shows the nominalized standard deviation for 5 samples extracted from Flask B. Peak areas were reproducible without an ISTD within 20% error for most features in the chromatogram. (Note: Original number of features was 80, which was reduced to 42 features present in 80% of samples.)

Figure 8 shows a chromatogram of the water soluble fraction from JP-8 that was obtained with SPME using a 100 μm PDMS coated fiber. SPME proved to be very successful at extracting the compounds in the water soluble fraction. Previous extraction methods used by Mayfield and Henley did not extract any compounds past 21 minutes where the methylnaphthalenes come out. Most of these peaks have been tentatively identified as dimethylnaphthalenes. One of the early peaks in this group was tentatively identified as 1,1′ biphenyl.

FIGURE 8: JP-8 Water Soluble Fraction



CONCLUSIONS

Solid phase microextraction (SPME) is a relatively new technique used to extract organic compounds from aqueous samples. It is a quick and simple technique that could be of much benefit for the U.S. Air Force if it could be used to quantify jet fuel contamination in groundwater samples.

SPME is able to quantify BTEX compounds and other organics successfully. To properly quantify benzene and toluene, it is necessary to implement oven cryogenics and to increase the salinity of the solution. An internal standard is necessary for quantification of all organics with SPME, as absolute peak areas of identical samples vary significantly. With an internal standard, variation among ratios of analytes to the internal standard can be kept below 5% for ethylbenzene, xylene, and presumably all benzene derivatives that are more highly substituted.

Profiles of the water soluble fractions of jet fuels can be obtained with SPME. SPME is able to extract some compounds in the water soluble fraction of a jet fuel which have not been extracted by other methods. The compounds in the water soluble fractions can conceivably be quantified if an internal standard is used in the fuel sample.

SPME is an excellent screening technique. The results of this work suggest that in the future SPME can also be used successfully to quantify organic contamination in aqueous samples.

References

1. Personal communication with Howard T. Mayfield, July 1994

2. Hawthorne, S.B.; Miller, D.J.; Pawliszyn, J.; Arthur, C.L. *Journal of Chromatography* 1992, **603**, 185-191

3. Arthur, C.L.; Pratt, K.; Motlagh, S.; Pawliszyn, J. *Journal of High Resolution Chromatography* 1992, **15**, 741-744

4. Potter, D.W. and Pawliszyn, J. *Journal of Chromatography* 1992, **625**, 247-255

5. Louch, D.; Motlagh, S.; Pawliszyn, J. *Analytical Chemistry* 1992, **64**, 1187-1199

6. Personal communication with Jim Byron, SUPELCO, July 1994

7. Arthur, C.L.; Killam, L.M.; Buchholz, K.D.; Pawliszyn, J. *Analytical Chemistry* 1992, **64**, 1960-1966

8. Mayfield, H.T. and Henley, M.V., "Classification of Jet Fuels Using High Resolution Gas Chromatography and Pattern Recognition," Monitoring Water in the 1990's: Meeting New Challenges, American Society for Testing and Materials, 1991

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser
Research Associate
Department of Biology

Trinity University
715 Stadium Dr.
San Antonio, TX 78212-7200

Final Report for:
Summer Graduate Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser
Associate Researchers
Department of Biology
Trinity University

Abstract

In ketamine-anesthetized rats, sustained whole-body exposure to 35-GHz millimeter wave radiofrequency radiation (RFR) produces hyperthermia, visceral vasodilation, and subsequent hypotension resulting in death of the subject (Physiologist 34:246, 1991). This study sought to determine whether this phenomenon (i.e., eradication of compensatory splanchnic vasoconstriction precipitating hypotension) is caused by vasodilatory factor(s) present in the circulating blood during circulatory failure. In search of evidence for a humoral visceral vasodilator, we performed a blood transfusion experiment. Two groups of rats (n=10 for each group) were used for the protocol. In the experimental group, one rat (donor rat) was exposed to RFR until mean arterial pressure (MAP) fell to 75 mmHg (arbitrarily assigned point of shock induction from previous work). At this point, 5 ml of blood were withdrawn from the hypotensive rat via the left carotid artery. This blood was subsequently infused into the recipient rat via the right jugular vein while an equal volume of blood was withdrawn simultaneously from the right femoral artery. MAP was monitored on the recipient rat for a 5 minute control period prior to transfusion and during the entire transfusion. In the control group, the same procedure was employed without exposing the donor subject to RFR. Therefore, in the control paradigm, the donor subject was normotensive when the blood was withdrawn. Immediately following transfusion in both groups, we observed an initial decrease in MAP followed by a similar increase returning MAP to control period levels. The recipient rats in the experimental paradigm did demonstrate a more pronounced decline in MAP post-transfusion as compared to the recipient rats in the control group (20.4 mmHg to 9.3 mmHg, respectively); however, those differences in mean maximum decrease in MAP were not shown to be significant (p=0.051). Therefore, we conclude that the vasodilatory factor(s) is not a humoral agent.

MILLIMETER WAVE-INDUCED HYPOTENSION DOES NOT INVOLVE HUMORAL FACTOR(S)

Amber Luong and Eric Wieser

Introduction

In humans and other mammals, maintenance of homeostasis is vital to survival. Homeostasis involves the regulation of physiological variables within a very narrow range. One of the many regulated variables is internal body temperature. Although all mammals possess thermoregulatory mechanisms that maintain their respective optimal internal temperature, prolonged extreme temperature changes can result in failure of the thermoregulatory system.

A primary mechanism of heat loss during thermal stress is through dilation of the cutaneous vasculature. In mild to moderate heat stress, arterial blood pressure is maintained at normal levels despite the marked cutaneous vasodilation by both an increase in cardiac output and a redistribution of blood flow from the viscera to the skin. That is, cutaneous vasodilation is normally accompanied by a compensatory vasoconstriction in visceral vascular beds that is primarily mediated by increases in sympathetic nervous system activity (Rowell, 1986; Kregel and Gisolfi, 1989).

Severe hyperthermia, however, may result in heat stroke, a condition characterized by a precipitous fall in arterial blood pressure. Heat stroke may, in turn, lead to a state of circulatory shock, in which tissue hypoperfusion occurs. Although the mechanism(s) responsible for this circulatory dysfunction is still in question, it now appears that a significant loss of peripheral vascular tone occurs in vascular beds that were previously constricted. Adolph (1923/4) first suggested that circulatory failure contributes to heat-induced circulatory shock. Subsequently, Daily and Harrison (1948) demonstrated that the hypotension and decreased cardiac output attendant to severe hyperthermia in humans were the result of peripheral pooling of blood. Kielblock et al. (1982) later proposed that fatal heat-induced shock resulted from cardiac failure due to a marked decline in vascular resistance after the loss of compensatory

vasoconstriction.

Kregel et al. (1988) directly measured the sequence and nature of vascular responses to environmental heat stress in conscious and anesthetized rats. In these heat-stressed rats, mean arterial pressure (MAP) increased until core temperature reached 41.5°C, at which point MAP fell precipitously. Mesenteric vascular resistance increased during the early stages of heat but declined sharply before the sudden fall in MAP. Thus, a selective loss of compensatory splanchnic vasoconstriction appears to trigger the circulatory collapse associated with severe hyperthermia. The sudden splanchnic vasodilation, combined with continued cutaneous vasodilation, produces hypotension by decreasing both total peripheral vascular resistance and venous return; the latter ultimately results in decreased cardiac output.

Visceral vasodilation preceding shock induction has been demonstrated during millimeter wave (MMW) irradiation, as it does during environmental heat-induced shock. In our model of heat stress (i.e. MMW exposure), using ketamine-anesthetized rats, mesenteric blood flow decreased during the early stages of MMW irradiation but then dramatically increased immediately prior to the onset of hypotension (Frei, *et al*, in preparation). Therefore, our model of heat-induced shock induction is analogous to that produced by environmental heating because, in both cases, eradication of compensatory splanchnic vasoconstriction precipitates hypotension.

There are several known possible endogenous vasodilators including opiates, catecholamines, nitric oxide, cytokines, arachidonic acid metabolites, bradykinin, histamine and some other small humoral peptides. Kregel *et al.* (1990) ruled out opiates, splanchnic sympathetic neurotransmitters and catecholamines as possibilities, since blockade of each of these potential mediators failed to prevent visceral vasodilation. In the MMW-induced heat stress model, nitric oxide, a potent gaseous vasodilator implicated in several other forms of circulatory failure, does not appear to be responsible for the noted hypotension. Chronic nitric oxide synthetase blockade studies concluded that nitric oxide was not the vasodilator (Wieser *et al.*, 1994). Although

several of these vasodilator possibilities have been extensively studied, the primary factor(s) involved have yet to be identified.

In order to narrow down the remaining possible vasodilator candidates, the present study, employing the MMW-induced heat stress model, sets out to determine if the factor(s) is present in the circulating blood during circulatory failure.

MATERIALS AND METHODS

Animals and Surgical Preparation

Forty male Sprague-Dawley rats (Charles River Laboratories), weighing between 328 and 402 g (368 ± 5g) were used in this study. Animals were housed in polycarbonate cages and provided food and water *ad libitum*. The rats were maintained on a 12 h/12 h, light/dark cycle (lights on at 0600) in a climatically controlled environment (ambient temperature of 24.0 ± 0.5°C).

Immediately prior to experimentation, two rats were anesthetized with ketamine HCl (150 mg/kg, I.M.). Administration of ketamine at this dose level provides prolonged anesthesia in Sprague-Dawley rats (Smith *et al.*, 1980; Jauchem *et al.*, 1984). Supplemental ketamine injections were administered throughout the duration of the experiment to ensure proper anesthetized conditions for the subjects.

Donor Subject

The larger of the two rats was designated as the donor subject. A catheter (Teflon, 28 gauge i.d.) was placed into the aorta via the left carotid artery for measurement of mean arterial blood pressure and later used for blood withdrawal. After surgery, the rat was placed on a holder consisting of seven 0.5-cm (O.D.) Plexiglas rods mounted in a semicircular pattern on 4 X 6 cm Plexiglas plates (0.5 cm thick). The electrocardiogram (ECG), mean arterial pressure (MAP), respiration and temperatures at five locations were continuously monitored using a Gould TA 2000 recorder. A Lead II ECG was used to monitor the subject with subcutaneous nylon-covered flurocarbon leads in the right arm, right leg and left leg (ground). The arterial catheter was attached to a pre-calibrated blood

pressure transducer (P10EZ, Statham) which was interfaced with a pressure processor (Gould 13-4615-52). Respiratory rate was monitored by a pneumatic transduction method employing a piezoelectric pressure transducer (Model 320-0102-B, Narco Biosystems). Heart rate (HR) was determined from ECG readings. Temperature was recorded from five sites: (1) colonic (T_c) (5-6 cm post-sphincter), (2) left subcutaneous (T_{sl})(lateral, midthoracic, side facing the source of radiation), (3) right subcutaneous (T_{sr})(lateral, midthoracic, side away from radiation source), (4) right tympanic (T_t), and (5) tail (T_{ta}). Tail temperature was measured subcutaneously from the dorsal surface approximately 2 cm from the base of the tail. All of the above recorded variables were monitored by a Unisys computer system via a software program specifically developed for physiological measurements (Berger *et al.*, 1991).

Recipient Subject

The smaller rat was designated the recipient subject, and catheters were inserted into three different locations: aorta via the left carotid artery, right jugular vein, and the left femoral artery. The left carotid artery was used to measure mean arterial pressure; the right jugular vein was used for the infusion of blood while the femoral artery served as a means of blood withdrawal.

During the surgical procedures on both rats, T_c was measured using an electrothermia monitor (Vitek, model 101) and was maintained at a temperature of 37.5 $\pm$ 0.5°C.

Exposure Conditions and Equipment

Experimental donor rats were individually exposed to 35-GHz continuous wave radiofrequency radiation (RFR) at an incident power density resulting in a whole body average specific absorption rate of 13 W/kg. The animals were aligned in the E orientation (long axis parallel to the electric field) during the exposure time. Prior to exposure, physiological control readings were recorded for a five minute period. The control period was subsequently followed by 35-GHz RFR. Irradiation was continued until mean arterial pressure deceased to 75 mmHg

(arbitrarily defined from previous work as the point of shock induction), at which point the RFR was turned off and the animal was prepared for blood withdrawal.

RF fields were generated by an Applied Electromagnetics Millimeter Wave Exposure System and were transmitted by a model 3-28-725 standard-gain horn antenna (Macom Millimeter Products, Inc.). Irradiation was performed under far-field conditions (animals positioned 110 cm from the antenna). The incident power density (75mW/cm^2) of the RFR fields was determined with an electromagnetic radiation monitor (Model 8600, Narda Microwave Corporation), employing a Model 8623D probe. During exposures, generator power output was monitored continuously with a Model 432B Hewlett Packard power meter. Irradiation was conducted in an Eccosorb RF-shielded anechoic chamber (Rantec, Emerson Electric Co.) at Brooks Air Force Base, Texas. The chamber temperature and relative humidity were maintained at 27.0±0.5° and 20±5% RH, respectively.

Transfusion Procedures

Immediately following shock induction in the irradiated rat, 5 ml of blood were withdrawn via the left carotid artery. The withdrawal was performed using a Harvard Apparatus 44 pump (model 55-1144) at a rate of 1 ml/min. The blood was collected in a heparinized syringe.The collection of 5 ml of blood from the donor rat in conjunction with the shock induction resulted in the death of this subject shortly after the withdrawal was complete.

During the withdrawal of blood from the donor subject, control readings of MAP and respiratory rate were obtained on the recipient rat for five minutes via the same recording apparatus as described for the donor rat. Also, T_c was monitored via an electrothermia monitor (Vitek, model 101).

The syringe containing the blood withdrawn from the donor was subsequently placed on a Razel Syringe pump (model 4-99..M)and connected to the catheter in the right jugular vein of the recipient rat. An empty heparinized plastic syringe was mounted onto the Harvard Apparatus 44 pump (model 55-1144) and connected to the catheter in the left femoral artery. Withdrawal of blood from

the left femoral artery and infusion of the blood from the donor rat occurred concurrently at a rate of 1 ml/min in order to maintain a constant blood volume. During the transfusion of blood into the recipient rat, the MAP and T_c were continuously recorded. These parameters were monitored for thirty minutes after the completion of the transfusion procedure.

The recipient rat was euthanized with a overdose of ketamine HCl at the end of the experiment.

The rats were divided into two groups: (1) a control group (n=10) in which transfusion occurred between two non-irradiated rats and (2) an experimental group (n=10) where the transfusion occurred between an irradiated and a non-irradiated rat.

For the control group, the donor rat was monitored for the same physiological parameters as the donor subject in the experimental group; however, no radiation was applied. Similar to the experimental group, five minutes of control readings for the donor rat were attained with T_c between 37.0 ± 0.5°C. These parameters were recorded for an additional thirty minutes, approximately the amount of time required for shock induction in the irradiated rats from the experimental group. At the end of the thirty minutes, the transfusion procedure was performed as described above.

Data Analysis

Preliminary statistical comparisons of MAP in the recipient rat between control and experimental group were performed at twelve different time intervals: control (mean of MAP values 2 min prior to transfusion), pre-transfusion (MAP immediately prior to transfusion), 0 min (the last MAP value during the transfusion), 0.5, 1, 2, 3, 4, 5, 10, 20, and 30 minutes post-transfusion. Statistical comparisons of each time period were accomplished by a two-way analysis of variance (ANOVA) with repeated measures.

Statistical comparison of mean maximum decrease in MAP in the recipient rat following transfusion were performed comparing control and experimental groups. The mean maximum decrease in MAP was calculated by taking the difference between

a mean from 2.5 minutes of MAP values at the end of the transfusion and the lowest MAP reading post-transfusion.

RESULTS

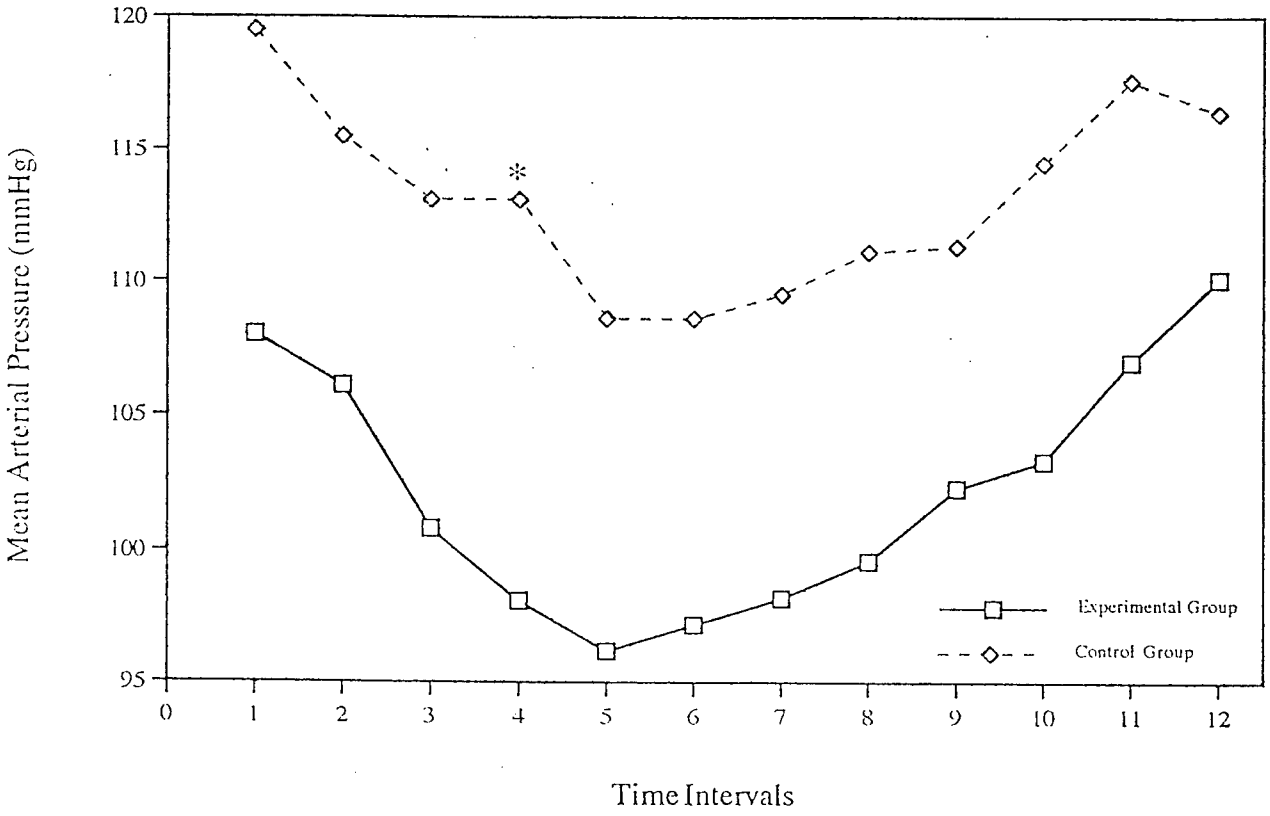
Table 1 shows that the time from the control period until the beginning of the transfusion used in the experimental paradigm is similar to the allotted 30 minute time prior to transfusion for the control paradigm. In the exposed animals, the mean T_c and T_{sl} reached 40.1°C and 45.0°C, respectively, prior to transfusion. The control group's mean T_c and T_{sl} remained constant during the 30 minutes prior to transfusion at 36.9°C and 35.3°C, respectively.

Table 1. Donor rat's parameters (mean values with n=10) prior to cross-circulation.

	Experimental	Control
Time to shock (min)	32.54	30.00
Tc (C)	40.13	36.86
Tls (C)	45.00	35.25

Figure 1 graphs the MAP over time for both the control and experimental group plotting 12 time intervals: control period, pretransfusion, post transfusion, .5 min, 1 min, 2 min, 3 min, 4 min, 5 min, 10 min, 20 min, and 30 min. The upper line represents data from the control group, while the lower line shows values for the experimental group. There were no significant differences of MAP values between control and experimental groups except at 0.5 min post-transfusion. Both groups show similar trends in MAP changes (i.e., initial decrease followed by a slow increase in MAP) with the lowest MAP value occurring at 1 min post-transfusion.

Figure 1. MAP graph versus time for both experimental and control groups.



* Significant difference between control and experimental values ($p<0.05$)

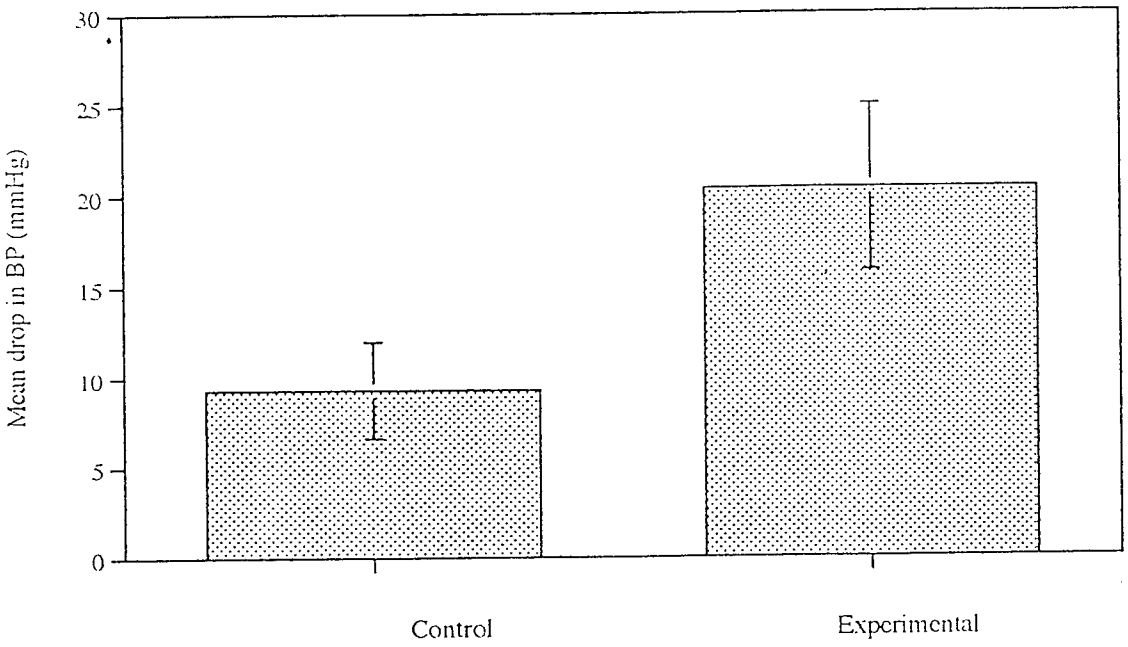
LEGEND FOR TIME INTERVALS

1- Control period	7- 3 min post-transfusion
2- Pre-transfusion	8- 4 min post-transfusion
3- 0 min	9- 5 min post-transfusion
4- 0.5 min post-transfusion	10- 10 min post-transfusion
5- 1 min post-transfusion	11- 20 min post-transfusion
6- 2 min post-transfusion	12- 30 min post-transfusion

Figure 2 is a bar graph showing the mean maximum changes in MAP for both groups. This change was calculated by obtaining a mean BP value from 2.5 min at the end of the transfusion and the minimum BP following transfusion. The difference between these values represents the maximum change in MAP. The mean of these maxima is represented in the bar graph.

Although there was no significant difference (p=0.051) in the maximum change in MAP between the control and experimental group, the experimental group shows a greater change in MAP than the control group. The mean maximum change in MAP of the control and experimental groups were 9.3 mmHg and 20.4 mmHg, respectively. As the p-value indicates, the values between the two groups were borderline to being significantly different.

Figure 2. Control and Experimental Groups Mean Maximum Change After Transfusion



DISCUSSION

Our results suggest that the vasodilator(s) responsible for the MMW heat-induced hypotension is either not humoral in nature or not detectable via our transfusion protocol. Figure 1, showing the MAP over time, depicts similar trends in both the control and experimental groups (i.e., initial decrease followed by a similar increase in MAP after the transfusion). This suggests that an artifact of the protocol may be partially responsible for the drop in MAP during and immediately following transfusion.

However, it appears that some of the drop in MAP may not be entirely due to protocol technique. There was a noted significant difference in MAP values at 0.5 minutes following transfusion between the control and experimental group, with the experimental group experiencing a greater drop in MAP immediately following transfusion. Also, Figure 2 shows that the experimental group had a greater mean maximum decrease in MAP than the control group. Although this difference was not significant, there was a trend in greater mean maximum decrease in MAP in the experimental group that just failed to reach significance (p-value=0.051). These findings of a greater drop in MAP following transfusion for the experimental group suggest the presence of some blood-borne vasodilator. Therefore, we do not completely discount the possible existence of a humoral vasodilatory factor(s).

REFERENCES

Adolph, E.F., and W.B. Fulton. The effects of exposure to high temperatures upon the circulation in man. *Am. J. Physiol.* 67:573-588, 1923/4.

Berger, R.E., M.R. Frei, and J.R. Jauchem. Computer-based monitoring of thermoregulatory responses of anesthetized rats exposed to radiofrequency radiation. *Physiologist* 34:246.

Daily, W. M. and T. R. Harrison. A study of the mechanism and treatment of experimental heat pyrexia. *Am. J. Med. Sci.* 215:42-55, 1948.

Frei, M.R., R.E. Berger, and J.R. Jauchem. Visceral vasodilation precedes circulatory shock during millimeter wave irradiation. In preparation.

Gisolfi, C.V., Matthes, R.D., Kregel, K.C., and Oppliger, R. Splanchnic sympathetic nerve activity and circulating catecholamines in the hyperthermic rat. *J. Appl. Physiol.* 70:1821-1826, 1991.

Hecker, M., M.L. Foegh, and P.W. Ramwell. The Eicosanoids: Prostaglandins, Thromboxanes, Leukotrienes, and Related Compounds. In: *Basic & Clinical Pharmacology*. Connecticut: Appleton& Lange, 1992, p. 263-277.

Kielblock, A.J., N.B. Strydom, F.J. Burger, P.J. Pretorius and M. Manjoo. Cardiovascular origins of heatstroke pathophysiology: an anesthetized rat model. *Aviat. Space Environ. Med.* 53:171-178, 1982.

Kregal, K.C. and Gisolfi, C.V. Circulatory responses to heat after celiac ganglionectomy or adrenal demedullation. *J. Appl. Physiol.* 66:1359-1363, 1989.

Kregal, K.C., P.T. Wall, and C.V. Gisolfi. Peripheral vascular responses to hyperthermia in the rat. *J. Appl. Physiol.* 64:2582-2588, 1988.

Kregal, K.C. and Gisolfi, C.V. Circulatory responses to vasoconstrictor agents during passive heating in the rat. *J. Appl. Physiol.* 68:1220-1227, 1990.

Parratt, J.R. Neurohumoral agents and their release in shock. In: *Handbook of Shock and Trauma*. edited by B.M. Altura, A.M. Lefer, and W. Schumer. New York: Raven Press, 1983, p. 311-336.

Rowell, L.B. *Human Circulation: Regulation during Physical Stress*. New York: Oxford University Press, 1986.

Schauer, A. Initial liberation of biogenic amines and effect of further mediators following application of endotoxins. In: *Gram-Negative Bacterial Infections*, edited by B. Urbaschek, R. Urbaschek, and E. Netter. Springer, Berlin, 1975, p. 315-322.

Wieser, E.S., Luong, A.U., Ryan, K.L., Frei, M.R., Berger, R.E., and Jauchem, J.R. (Abstract) Chronic nitric oxide (NO) inhibition: Effects on hypotension induced by 35-GHz radiofrequency radiation (RFR) exposure. *The FASEB J.* 2:A593, 1994.

A STUDY OF THE USE PREDICTIVE MODELING FOR DYNAMIC LOADING OF THE SPINE

Gregory S. Zinkel
Department of Engineering Science and Mechanics

Georgia Institute of Technology
Atlanta, GA 30332

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

September, 1994

A STUDY OF THE USE PREDICTIVE MODELING FOR DYNAMIC LOADING OF THE SPINE

Gregory S. Zinkel
Department of Engineering Science and Mechanics
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

The applicability of finite element modeling to predict the response of the cervical spine under high dynamic loading was studied. MRI images were to be used to build a 3-D geometry of individual pilots which would then be analyzed to determine the response of the spine to high G loading. However, it was found that a 3-D model could not be built from the images directly. In addition, the lack of data covering the responses of the spine under high dynamic loads prevented the building of a validatable model by hand. Although finite element modeling with a dynamic analysis would be able to predict the response of the spine, the following conditions must be met first: 1. further experiments must be conducted to ascertain the mechanical properties of the spine under high dynamic loads, 2. injury mechanisms from high dynamic loading must be better characterized so that a model can be validated acceptably, and 3. the medical images must be enhanced from what we had done.

Introduction

As technological improvements have increased the performance of military aircraft, injuries to pilots after ejection have increased in both severity and frequency. Of great importance then is to be able to determine what the specific injury mechanisms are. This information can then be used to develop improved ejection protocols and to identify individuals who are at risk of injury due to an existing pathology.

From a modeling perspective, the structure of the spine is a very complex three-dimensional structure, consisting of both hard and soft tissue interconnected with ligaments and muscles. The spine has been shown to exhibit time dependent, viscoelastic responses to loading. Thus an accurate model would be a three-dimensional model, derived from three-dimensional, time dependent data. Of interest would be a predictive model, from which an analysis could be performed on pilots with existing pathologies to determine if they can safely eject from their aircraft.

Developing proper techniques to minimize acceleration injuries require a better understanding of the factors involved in the dynamic loading of the spine. However there are many problems which hinder our understanding of the injury mechanisms. In particular, the lack of sufficient data on high G loading of the spine prevents proper validation of the proposed mechanisms of injury. Experiments have been performed on volunteers to investigate the effects of low G loading, but data on high G loads can only be obtained from animal studies or from human cadavers. This data may or may not accurately represent the response of a real pilot during ejection.

Anatomy of the Cervical Spine

The cervical column consists of 7 vertebra. Except for between C-1 and C-2, all have intervertebral disks between them. In addition, C-1 and C-2 are atypically shaped to allow the head to move vertically and laterally. The other disks, C-3 to C-7, consist of an oval body, two processes on each side, a triangular vertebral foramen, and a spinous process pointing posteriorly from the vertebral arch. The vertebra increase in size from top to bottom. The

interior of the vertebral body is composed of trabecullar (cancellous or spongy) bone and is encased by cortical (compact) bone. The trabecullar bone is composed of spongy appearing bone with open spaces filled with fluid. It is softer than the surrounding cortical bone (Tortora, 1983 and Belyschenko and Privitzer, 1978). It is questionable whether the fluid in the trabecullar space provides any mechanical response to the vertebrae. Some researchers have reported that the fluid greatly affects the response of the vertebrae under compression loading, while others assert that the fluid imparts nothing to the mechanical properties of the disk (Belyschko, 1985). The effects of the fluid in the trabecullar spaces will have to be investigated further in order to accurately predict the dynamic response of the vertebrae under large acceleration forces. The cortical bone is a shell of stiff bone surrounding the trabecular tissue. Its elastic modulus is roughly 200 times that of the trabecular bone. The upper and lower surfaces of the vertebral body are called end plates. They are of interest since a common injury mechanism is a fracture of the lower end plate. The processes extending from the sides of the vertebrae are in general for the protection of the nerves running through the spine, and do not contribute significantly to the mechanical properties of the vertebrae. Many researchers have ignored the processes when developing their model of the spine. The intervertebral disks lying between vertebrae C-2 through C-7 consist of a tough, fibrous shell surrounding a semifluid medium. The fibrous shell, the annulus fibrosus, consists of tough collagen fibers which lie in sheets which run in alternating fashion, one running about 30° from horizontal plane one way and the next layer running 30° from horizontal the other way. The annulus fibrosis provides active resistance to pressure placed on the disk. However, in torsion, damage can occur to the disk. The fluid medium within the disk, the nucleus pulposis, is a gel like substance early in life, however as one ages, the nucleus loses its gel like properties. This causes the loads on the disk to be distributed anisotropically (Williams and Belytschko, 1981). The vertebrae are also connected to muscle and ligaments via the spinal processes. An accurate model of the spine will have to take into account the actions of these muscles and ligaments.

Injury Data

The data which exists on injuries during pilot ejection is very difficult to analyze due to the number of variables involved and the scarcity of data. Variables include the aircraft's trajectory prior to ejection, the type of ejection mechanism, and the pilot's orientation prior to ejection, along with many others. Data from actual ejections generally consist of an estimate of the aircraft's speed and orientation prior to ejection. Although the data is sporadic, general trends among the types of injuries sustained by pilots and inferences about the injury mechanisms can be obtained. It has been observed in testing that in addition to high accelerations in the vertical direction (G_Z) of over 10G, pilots are also subject to high lateral (G_Y) accelerations, sometimes exceeding 15G. This is due to actions such as parachute deployment. The difficulty in identifying the actual loads which caused an injury are that these dynamic loads are transient, lasting only a fraction of a second and that these loads leave no indication of their presence, other than the actual injury. What is known is that among the type of injuries suffered by pilots during ejection, neck injuries are both the most common and the most serious. Moderate to severe neck injuries have occurred in about 11% of all ejections and the injury rate is increasing today. Severe injuries to the neck have found to be localized at the C-2, C-5, and C-6 cervical vertebrae (Guill and Herd, 1989 and Guill, 1989).

Data from Pilots with Pre-existing Conditions

The purpose of this study is to determine how a pilot with a pre-existing condition can be evaluated. Attempting to identify injuries and abnormalities which would prevent a pilot from safely ejecting is difficult. Of all pilots who have documented injuries during ejection, we anticipate very few will have had a known pre-existing condition. Thus we expect to be unable to draw significant conclusions from this data and will be forced to rely on predictive analysis to evaluate an individual. After developing an accurate model, one can apply anticipated dynamic loads on the spine and determine its response. We expect that this information can then be used to predict in a rough sense what would happen to an individual during an ejection.

Developing a Model

It was our intention to develop a finite element model of the head and cervical spine and then use this model to evaluate pilots with back pathologies. The individual's geometry would be introduced by building a three-dimensional representation of the spine from MRI data. This would be done in a similar fashion as had been previously described in the literature for CT scans (Breau, Shirazi-Adl, and de Guise, 1990). Unfortunately, the MRI data did not lend itself to building a three-dimensional representation as had been anticipated. An MRI image gives accurate information on soft tissues, but is not well suited for visualizing bony structures. In contrast, CT scans are excellent for visualizing bony tissue, but not soft tissue. MRI data was chosen since soft tissue information for this study is essential. Many of the pathologies in the back are due to abnormalities in the intervertebral disks and other soft tissues. The MRI data would be presented in a series of 10 to 20 slices, from which a three-dimensional model could be developed. We attempted to input the MRI data into our finite element software by processing the digital image with an image processing program (NIH Imaging). However, after much manipulation, the bony tissues were determined to be of not sufficient resolution to be able to transfer the data directly to the finite element program. Thus it would be necessary to build the geometry from the vertebrae from scratch within the finite element program. In this study, we did not have the time required to be able to build the geometry by hand. However the particular software we were using could build a 3-D model for finite element analysis from either a wire-frame outline generated manually or from a computer image.

Conclusions

Finite element methods have been proven over the last thirty years to be an accurate method for modeling the human spine. Researchers have shown the applicability of finite element methods for static and viscoelastic models of the spine once the proper elastic and viscoelastic parameters have been determined. We believe that finite element methods are equally applicable to modeling high G loading of the cervical spine. However, we have had several problems in trying to implement a finite element model. Our original intent was to build a finite

element model from MRI images. MRI imaging has very good resolution for soft tissue, but is not especially applicable for analyzing hard tissues, such as bone. We found that we could not adequately build a model from the MRI images since the vertebrae could not be resolved from the remaining vicera. A solution to this problem would be to further manipulate the MRI to enhance the contrast between the bone and vicera or to use CT scans in conjunction with MRI images so that one is able to obtain an accurate model of the vertebrae, the intervertebral disks, and the surrounding vicera. We could build a model by creating wire-frame outlines of the spine by hand from medical images and then generate a 3-D model from these frames, but we found that there is little mechanical data on the characteristics of the spine under high dynamic loads, thus our model would be of questionable validity. Thus it is our contention that it is feasible to use predictive analysis to analyze the response of an individual's spine to high dynamic loads subject to the following conditions: 1. further experiments must be conducted to ascertain the mechanical properties of the spine under high dynamic loads, 2. injury mechanisms from high dynamic loading must be better characterized so that a model can be validated acceptably, and 3. the medical images must be enhanced from what we had done.

References

Belytschko, T., Rencis, M., and Williams, J., (1985), "Head-Spine Structure Modeling: Enhancements to Secondary Loading Path Model and Validation of Head-Cervical Spine Model", Northwestern University, Report to Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AAMLR-TR-85-019.

Belytschko, T. and Privitzer, E., (1978), "Refinement and Validation of a Three-Dimensional Head-Spine Model", University of Illinois at Chicago, Report to Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AMLR-TR-78-7.

Breau, C., Shirazi-Adl, A., and de Guise, J., (1991), "Reconstruction of a Human Ligamentous Lumbar Spine Using CT Images - A Three Dimensional Finite Element Mesh Generation", *Annals of Biomedical Engineering 19*, 291-302.

Guill, F. and Herd, G., (1989), "An Evaluation of Proposed Causal Mechanisms for 'Ejection Associated' Neck Injuries", *Aviation, Space, and Environmental Medicine 60*:7, Supplement, A26-A47.

Guill, F., (1989), "Ascertaining the Causal Factors for 'Ejection Associated' Injuries", *Aviation, Space, and Environmental Medicine 60:7*, Supplement, B44-B71.

Panjabi, M., Oxland, T., Yamamoto, I., and Crisco, J., (1994), "Mechanical Behavior of the Human Lumbar and Lumbosacral Spine as Shown by Three-Dimensional Load-Displacement Curves", *Journal of Bone and Joint Surgery 76-A:3*, 413-424.

Tortora, G., (1983), *Principles of Human Anatomy*, New York, Harper & Row.

Williams, J. and Belytschko, T., (1981), "A Dynamic Model of the Cervical Spine and Head", Northwestern University, Report to Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AFAMRL-TR-81-5.