UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1994

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 2A

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

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0803

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TY
	December, 1994	Final

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
USAF Summer Research Program - 1994 Summer Faculty Research Program Final Reports, Volume 2A, Armstrong Laboratory	
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 Summer Faculty Research Program (USAF- SFRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members 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	15. NUMBER OF PAGES
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES	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

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.

Due to its length, Volume 2 is bound in two parts, 2A and 2B. Volume 2A contains #1-22. Volume 2B contains reports #23-45. The Table of Contents for Volume 2 is included in both parts.

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

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%

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

DETERMINATION OF THE OXIDATIVE REDOX CAPACITY OF AQUIFER SEDIMENT MATERIAL BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson
Professor
and
Mark C. Delgado
Graduate Student
Department of Chemistry

University of Georgia
Athens, GA 30602-2556

Final Report for:
Summer Faculty Research Program
and
Graduate Student Research Program
Armstrong Laboratory, Environics Division
AL/EQC
Tyndall Air Force Base, Panama City, FL

Sponsored by:

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

September, 1994

DETERMINATION OF THE OXIDATIVE REDOX CAPACITY OF AQUIFER SEDIMENT MATERIAL BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson
Professor
and
Mark C. Delgado
Graduate Student
Department of Chemistry
University of Georgia

Abstract

Methodology was developed for determination of the oxidative redox capacity of aquifer sediment material by the method of spectroelectrochemical coulometric titration. This method involves the measurement of absorbance of sediment particle slurries at the maximum absorption wavelengths of the optically detectable mediator-titrant (reporter) molecules resorufin and methyl viologen as a function of the charge passed in a constant-potential coulometric titration. An approach which was successful for determination of the oxidative redox capacity of a pond sediment rich in organic matter and iron species was extended to an oxidized aquifer sediment material of low organic carbon and iron species content sampled from Columbus Air Force Base, Mississippi. Titration was carried out on diluted, dry-sieved material of particle size smaller than 75 μm diameter, suspended in aqueous, pH 7, 0.1 ionic strength phosphate buffer at 0.0426 % sediment by weight. Blank titration was carried out on a sample of identical composition but in absence of the aquifer material. In both cases, resorufin was reduced first, followed by methyl viologen. There was no perceptible delay between completion of titration of resorufin and the initiation of titration of methyl viologen. This behavior contrasted significantly with the titration of pond sediment of high organic and iron species content, which showed a very significant break between completion of titration of resorufin and initiation of titration of methyl viologen. Based on the uncertainties of measurement, it could be estimated that the upper limit of oxidative redox capacity of the Columbus aquifer material was ca. 3 microequivalents per gram of solid material. This estimate is in the vicinity of the values of redox capacity of aquifer material obtained from other sites by one other research group, but not consistent with the values reported by another group. More precise determination of oxidative redox capacity will require use of methods such as fluorescence which are more immune to the effects of scattered light than absorption spectrophotometry, and will allow higher loading of suspended solids than the current absorbance-based method. Additional studies identified the importance of thermal expansion of aqueous solutions as a cause of oxygen leakage into closed vessels when temperature is not regulated, and demonstrated that huge pressure changes (900 psi over a range of 22 °C) can occur when the temperature of an aqueous sample is allowed to vary by small amounts. Methods were devised to overcome this problem by the combination of a thermoisolation chamber to control the temperature of the sample and exclude oxygen from the titration zone.

DETERMINATION OF THE OXIDATIVE REDOX CAPACITY OF AQUIFER SEDIMENT MATERIAL BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson

Mark C. Delgado

Introduction

The remediation of many polluted sites, whether by in situ processes such as chemical or biological natural attenuation or by processes involving the addition of externally supplied chemical or biological agents, frequently involves redox processes. The feasibility and total transformation capacity of such processes is ultimately limited by the redox capacity of the environmental system to be transformed. In this context, redox capacity is a measure of the number of molar equivalents of electrons which can be donated (reductive redox capacity) or taken up (oxidative redox capacity) by the system in the course of a redox transformation of an externally added agent such as a pollutant. The redox capacity limits the ultimate quantity of pollutant which can be oxidatively or reductively transformed by the system, since for each mole of pollutant oxidatively or reductively transformed, a stoichiometrically related number of moles of oxidant or reductant in the environment must also be transformed. Once all electron donors or acceptors initially available in the system to drive the transformation of interest have been exhausted, the transformation must stop. In addition, the redox capacity is an important quantity in assessing the feasibility of driving a transformation process by the addition of an external chemical or biological agent, by providing an estimate of the quantity of external agent required either to supplement or to oppose the natural capacity of the system in driving a desired transformation.

Unfortunately, the present knowledge of redox capacity of natural systems is limited. Only a relatively small number of studies has been carried out to assess this important information (1-3) in sediments. Data from two of these studies on apparently very similar aquifer sedimentary materials are in significant conflict regarding the magnitude of the oxidative redox capacity of the sediments (1,2). The methodologies of these two studies are based on classical chemical titrations with very reactive chemical reductants, requiring rather long

equilibration times, rigorous oxygen exclusion over long periods, and centrifugation and filtration to overcome interference with spectrophotometric measurements due to light scattering by sediment particles.

We have developed an alternative indirect spectroelectrochemical coulometric titration approach in which the optical signal of an added indicator reagent is used together with coulometry to assess the number of oxidative equivalents of sediment components present. Reagents are generated in situ in the titration vessel as needed, thereby avoiding some of the problems in earlier work of storing very reactive, potentially unstable reagents for long periods of time. The approach also enables measurements in the presence of suspensions of sediment particles, without need for particle removal prior to spectrophotometric measurements. The indicator reagent provides the spectral signal, so that the approach is applicable even to samples that contain no chromophore. Our earlier work on redox capacity of a pond sediment (3,4) yielded results which were compatible with the results of Barcelona and Holm for aquifer material (1), but significantly higher than the results of Heron et al. for aquifer material (2).

In this investigation, we examine the use and limitations of the spectroelectrochemical methodology to probe the redox capacity of aquifer material more directly comparable to the material investigated by Barcelona and Holm (1) and Heron et al. (2). In addition, in the process of investigating the possible sources of an oxygen leak during the development of this methodology, we have discovered the enormous pressure changes (as great as 900 psi for a temperature change of 22 °C) that can be generated by thermal expansion of aqueous solutions in a closed vessel when temperature is allowed to vary. Methods were developed to overcome this practical problem in titration cells. A model for the dependence of pressure change on temperature was developed and compared with experiment. The enormous pressure change observed can be quite well explained based on the known thermal expansion coefficients and compressibility of water and their temperature dependence.

Methodology

Spectroelectrochemical methods have been previously used to coulometrically titrate biological components(5,6). A tin oxide working electrode is used to transfer electrons to or accept electrons from electrogenerated titrants which in turn transfer charge to or from the

biological species of interest. The working electrode may be used to drive either oxidative or reductive processes, by controlling the applied potential in an appropriate region. This approach is well-suited for reliable quantitation of low micromolar concentrations of spectrally visible and invisible species (7). Major advantages of the approach include the ability to work in a closed system of small volume, with oxygen removal before and exclusion during the titration; with accurate and quantitative addition of titrant at a controlled rate; and the feasibility of carrying out both forward and back titration of the component of interest to assess the reversibility of the process in a single titration experiment if desired (5).

In a conventional chemical titration, reagents must be added to the system from external reservoirs, thus continually changing the total mass, composition, and volume of the system. The spectroelectrochemical titration method described here does not suffer this problem, since the reagents are electrons generated or consumed by reactions at the working electrode in a closed system. Thus, the titrant species can be introduced into the system in a stable form which will not react with the species under study until the titration is initiated by applying an appropriate potential to the working electrode. Quantitation can be achieved electrically, by counting electrons (measured by the charge passed during the titration). In the experiments described here, the sample of interest was mixed with the reagents in inactive form, deoxygenated, and transferred into an electrochemical cell designed to isolate the solution from the ambient atmosphere. A detailed description of the cell and the degassing procedure follows.

Cell Design and Degassing Procedure

A diagram of the electrochemical cell and the electrodes appears in Figure 1. The cell consists of a main chamber constructed of 1 cm i.d. square Pyrex stock to which two sidearms and a valve with an inlet ground glass fitting have been attached. The valve allows introduction and isolation of samples from the ambient atmosphere. Samples are degassed in a degassing bulb connected to the valve via the ground glass fitting.

An inner Luer ground glass fitting was used for initial experiments, but was replaced with a 10/30 standard taper inner fitting due to significant problems in assuring that the joint did not leak and inadvertently admit oxygen to the solution. The two sidearms are fitted with an outer ground glass fitting (7/15). Each sidearm is joined to the main chamber through a medium

porosity frit. The reference and auxiliary electrode chambers are both made of Pyrex glass which terminate in 7/15 standard taper inner ground-glass fittings. The reference and auxiliary chambers are filled with 1.0 M KCl solution, and a Ag wire anodized in 6 M HCl to form a AgCl coating is inserted through a septum cap at the top of each electrode compartment to make the Ag/AgCl electrode.

Two chamber designs were used. The initial design had a single piece body, terminated with a porous Vycor frit epoxied into a 5 mm diameter glass tube extending from the lower end of the 7/15 fitting for contact with the cell solution. A serious problem with this design was the inability to maintain reproducibly any gas expansion volume when the cell was filled by vacuum degassing. This design was susceptible to leakage due to temperature elevation in the spectrometer sample compartment during an experiment, causing significant pressure increases due to water expansion. One or more 7/15 joints would open to relieve the otherwise disastrous pressure rise in the cell, giving rise to concomitant oxygen leakage into the cell. The revised electrode chamber was designed to allow retention of a gas space for liquid expansion, to prevent large pressure buildups due to thermal expansion of water in a system without such an expansion space.

In place of the septum cap, a 5 mm diameter Kontes Bevel-Seal threaded O-ring connector was sealed to the top of the 7/15 joint. A 3 mm diameter glass tube served as the reference or auxiliary electrode compartment. A piece of 3 mm porous Vycor rod was epoxied into a short length of 5 mm diameter tubing fused to the lower end of this 3 mm diameter tube. The tube passed through an O-ring placed above the 5 mm diameter section, through the 7/15 joint, and finally through the Bevel-Seal O-ring joint. The reference or auxiliary Ag/AgCl electrode protruded from the end of the tube, which was capped with an inverted septum cap. Because the tube passed through two O-rings, the electrode compartment could be slid up and down in the chamber. While the cell was being degassed prior to filling, the electrode compartment tube was slid down into the sidearm to attempt to make the gas volume above the lower O-ring accessible for oxygen removal and replacement by helium. When the cell was to be filled, the electrode compartment tube was pulled up so that the expanded lower end held the O-ring against the bottom of the inner 7/15 joint to preserve a gas space above the O-ring into which thermally expanding liquid could flow in the event of inadequate thermal control. This

approach should thus cut down on oxygen leakage. In fact the rate of oxygen leakage during titrations decreased more than an order of magnitude when the above sidearm design was used in conjunction with a thermostated chamber filled with helium to exclude oxygen from the atmosphere surrounding the cell during a titration. However, exclusion of oxygen was not consistently successful during the sample degassing process.

A section of 26 gauge platinum wire was flame-sealed into the cell to serve as a potentiometric electrode. The working electrode for the reduction steps was a 2.5 cm square piece of SnO_2 glass epoxied to the bottom of the main chamber of the cell with Devcon 2-Ton clear epoxy. The square edges of the working electrode were also used to align the cell in a square positioning recess in the optical train of the spectrophotometer. The main chamber of the cell held 1.85 mL of solution, which could be circulated by a magnetic stirrer. The stir bar was constructed by flame-sealing a ca. 7 mm long piece of steel paper clip inside Pyrex glass under vacuum. The stirrer was driven by a water-propelled magnetic impeller supplied with thermostated water from a circulating temperature regulator bath. The water was also circulated in the walls of an isolation chamber whose function was both to control the cell temperature and to bathe the cell in a nitrogen atmosphere to exclude oxygen access during the titration. Light was passed through the main chamber of the cell to determine the absorbance. Because the cell was made from Pyrex glass, light was detected primarily in the visible region of the spectrum.

Solutions to be studied were degassed and introduced into the cell under an inert atmosphere. The degassing assembly consisted of a Ridox catalyst, a helium inert gas line, a water-filled bubbler to saturate the inert gas with water, and a vacuum line with a liquid nitrogen trap and a Drierite drying column to prevent water from entering the mechanical vacuum pump. The helium was passed through a dryer and a Restek oxygen scrubber catalyst before it entered the Ridox catalyst chamber. The electrochemical cell could be either pressurized with the inert gas or evacuated via a two-way valve on the degassing assembly. All pieces of the degassing assembly were joined by ground glass fittings greased with Apiezon L or Apiezon N. The vacuum pump and the Drierite column were connected to the all-glass degassing apparatus by butyl rubber tubing.

The two-way valve was set to evacuate the solution degassing bulb/cell adapter and the cell for at least 15 min. prior to filling. The cell was then alternately evacuated for at least 1

min., followed by ca. 30 seconds of helium sparging, for a minimum of four cycles. The cell stopcock was closed off and the cell was removed from the degassing assembly. Both the reference and auxiliary electrodes chambers were degassed via a needle inserted through the septum cap while attached to the cell. They were evacuated and filled with an inert gas alternately for 2 min. for 5 cycles. After the electrodes had been degassed, they were evacuated and immediately filled by means of a 1 mL syringe with a previously degassed 1.0M KCl solution. If a bubble which might interfere with the solution conductivity appeared in a sidearm electrode compartment, the compartment was reevacuated and the degassing process was repeated for that electrode.

Once the reference and auxiliary electrode chambers were filled, the solution to be studied was introduced into the degassing bulb. At this point, both the bulb and the cell were reattached to the degassing assembly. The solution was then degassed by alternately exposing it to the inert gas and vacuum while it was being stirred. The solution was then introduced into the cell by opening the valves between the cell and the degassing bulb and simultaneously pulling vacuum in the cell and the degassing assembly. The cell and degassing bulb were tilted downward and the cell was filled by pressurizing the degassing bulb with the inert gas. To minimize bubble formation, vacuum was applied briefly and inert gas was again introduced into the degassing bulb. This procedure was repeated until no large bubbles were visible inside the cell.

Apparatus

The potentiostat which was used to control the experiment was custom-built for the purpose. Current output was converted to voltage by a current-to-voltage converter, and fed to a custom-built absolute value circuit which converted all signals to a positive value for input into a voltage-to-frequency converter (Datel) with a calibration factor of 10 kHz per volt at the input. The train of pulses from the voltage-to-frequency converter was fed to a counter-timer (Data Precision 5740) in the count mode. With a current output gain of 100 microamperes per volt, the counter had an output of 100,000 counts per millicoulomb of charge passed, with an accuracy better than 0.1 %.

The spectrophotometer was a Perkin-Elmer Model 3840 diode array spectrometer,

controlled by a Perkin-Elmer 7500 computer. All spectral scans were obtained in the survey mode, which had an excessive level of stray light (specified as 3 %, and measured to be ca. 2.5 %), due to mechanical problems which prevented operation in the high resolution mode, which had dramatically better stray light specifications. Spectra were recorded and stored on floppy diskettes. In some cases, data were converted to ASCII files and copied to a Sun Sparcstation computer for further processing.

A thermostated housing and platform were custom fabricated to adapt the nonstandard dimensions of the electrochemical cell to the spectrophotometer sample compartment. The thermostated housing was constructed from two 1/2 inch thick aluminum plates, through which water from a temperature bath was circulated, and poly(methylmethacrylate) to enable visual observation of the cell assembly. The housing was lowered over the cell and clamped to the platform. Nitrogen was circulated through the interior of the thermostated chamber to exclude oxygen from access to the exterior of the cell. Oil bubblers placed in the supply line to the thermostated chamber and in the outlet line from the chamber were used to visualize the flow of nitrogen into the system.

Pressure measurement equipment

Experiments were carried out to determine the dependence of pressure in a closed vessel on the temperature of the vessel. The vessel consisted of 1/4 inch diameter stainless steel tubing of ca. 1.25 mL internal volume, with a Helicoid 0-5000 pounds per square inch (psi) liquid chromatographic pressure gauge in a tee configuration, and two high pressure valves at either end. With the inlet valve closed and the outlet valve open, the system was evacuated. After closing the outlet valve, the inlet valve was opened and the system was pressurized to 1000 psi at 25 °C by pumping water in from an Altex Model 110A HPLC pump, and the inlet valve was closed. The vessel was equilibrated in a controlled temperature water bath overnight, to insure that there were no leaks. The temperature was then varied in random order over the range between 13 °C and 35 °C, in approximately 2 °C increments according to a sequence selected by use of a random number generator. At least two readings were obtained after the temperature had reached a stable, constant value. The final measurement was made at the initial temperature to check for any long-term leakage, which was found to be negligible. Temperatures were

measured with the temperature probe of a Yellow Springs Instrument YSI Model 33 conductivity meter and with a Radio Shack LCD readout temperature sensor (Archer catalog number 277-0123). Temperature readings with both probes were in good agreement.

Results and Discussion

The titration approach was applied to an aquifer material sample (sample tube number 13, Sample K-61, collected 4-6-90, depth 11 feet 5 inches to 11 feet 10 inches below the surface) taken from the Columbus Air Force Base site. The sample had been stored in the dry state, with no attempt to exclude oxygen. Thus the sample was well-oxidized. The sample was sieved dry through standard sieves to separate it into well-defined size fractions consisting of particle sizes > 850 μm, 425 - 850 μm, 250 - 425 μm, 106 - 250 μm, 75 - 106 μm, and < 75 μm. The fraction in the size range < 75 μm diameter was selected for titration to evaluate the spectroelectrochemical coulometric titration protocol for this aquifer material. This size fraction constituted 14.3 % by weight of the total sample weight, as determined by sieving. This percentage was nearly double the weight percent determined by sieving at the time of collection (8.63 %), although the percentages of other size fractions were in accordance with the as-sampled values. This behavior suggests that the particles aggregate to a significant extent, and that the percent of fine particles measured is significantly dependent on the degree of disturbance of the particles during sieving.

The pH of a suspension of < 75 μm diameter aquifer material in 5 mL of distilled water without added buffer was in the range of 5.1, depending on solids loading, stirring, and other factors. However, because the optical absorption properties and solubility of resorufin were not favorable at or below pH 6, samples were prepared for titration in a pH 7 phosphate buffer, with 0.1 ionic strength. The pH of a 1.2 % by weight suspension of aquifer material in pH 7.0 phosphate buffer was 7.03, indicating that the buffer capacity of the buffer was sufficient to control the pH of the suspension.

Spectroelectrochemical coulometric titrations were carried out in the controlled potential mode, at an applied potential of -0.65 V vs. Ag/AgCl/1 M KCl reference electrode. This potential was sufficiently negative that methyl viologen (MV^{2+}), the component with the most negative reduction potential, was reduced to the radical cation ($MV^{+\cdot}$), while any components

with more positive reduction potentials which could react directly with the electrode also were reduced. Methyl viologen could also react homogeneously with titratable components which do not react readily directly with the electrode surface. The reaction is thus catalytic, in the sense that reaction of $MV^{+\cdot}$ with species of more positive reduction potentials accelerates their reduction and regenerates MV^{2+} for further reaction. Upon addition of the desired quantity of charge, the applied potential was disconnected, and the potentiostat was placed in the potentiometric mode at zero current. While any oxidized species with reduction potentials more positive than that of methyl viologen remained in the system, any initial excess of $MV^{+\cdot}$ was consumed, and the system was allowed to come to equilibrium. Excess $MV^{+\cdot}$ remained in solution after equilibrium was reached, only after all species with more positive reduction potentials had been reduced. The appearance of excess $MV^{+\cdot}$ thus indicated the end of the titration.

Experiments were first carried out on a solution containing all of the reagents (10.1 micromolar resorufin, 0.408 millimolar methyl viologen) and the pH 7.0, ionic strength 0.1 phosphate buffer, but not the aquifer material. The resulting plot of absorbances at the wavelengths characteristic of resorufin (572 nm) and of methyl viologen radical cation (396 nm) is shown in Figure 2A. In all cases, absorbances are recorded in dual-wavelength difference mode with respect to the absorbance at 800 nm. Difference measurements are extremely valuable in compensating for the effects of scattered light, settling of suspension over time in the cell, and other nonidealities, since none of the species under study absorb appreciably at 800 nm, and the apparent absorbance at this wavelength is still affected by light scattering and other nonidealities, in a manner analogous to that at the analytically useful wavelengths of 572 and 396 nm. A plot of the potentiometric electrode potential vs. charge is shown in Figure 2B. The initial lag before resorufin is titrated is due to residual oxygen (ca. 3.5 μM, ca. 1 % of the ambient concentration prior to degassing and analysis) not completely removed during the degassing process. The quantity of residual oxygen varied, but the quantity of resorufin was consistent from run to run. It is clear in Figure 2A that there is essentially no break between completion of titration of resorufin and initiation of titration of methyl viologen. This behavior is to be expected for the system in absence of aquifer material, since the reduction potentials of resorufin and methyl viologen are in the order observed. The potentiometric potential seen in

Figure 2B also follows the trend expected, initially being governed by the redox equilibrium between oxidized and reduced components of resorufin, and then shifting to a value governed by the equilibrium between methyl viologen dication and methyl viologen monocation.

The weight percent of aquifer material in suspension during titrations was selected empirically based on the maximum quantity of suspended solids which afforded an acceptable level of light scattering. For the experiments reported here, the solids loading was 0.0426 % by weight. This loading afforded a slightly turbid solution with light scattering levels that were still acceptable. (Absorbance was elevated ca. 0.2 absorbance units at 396 nm relative to the same solution composition in absence of aquifer material.) A slightly higher loading level (perhaps threefold higher) would have been feasible if the spectrophotometer had had better stray light characteristics. The results of a spectroelectrochemical titration of an 0.0426 % by weight suspension of aquifer material under the same solution conditions as Figure 2A are shown in Figure 3A. The only difference evident between the response for the aquifer suspension and the blank control is a slightly lower residual oxygen content (ca. 0.7 % of the initial ambient concentration prior to degassing). In addition, the vertical shift of the absorbance of both curves in Figure 3A at ca. 3 mcoul charge reflects reoxidation of reduced resorufin during a waiting period of more than two hours before the next increment of charge was added, to test the susceptibility of the measurements to oxygen leakage over long periods. While some further improvements in preventing oxygen leaks would be beneficial, the observed leakage rate corresponds to the relatively low quantity of ca. 0.65 nmol O_2 per hour into the cell. After correction for the need to rereduce the resorufin that was reoxidized during this period, the charge consumed during titration of resorufin is in close agreement with the charge required for the blank, and the onset of excess $MV^{+\cdot}$ generation coincides closely with the completion of titration of resorufin. The plot of potentiometric potential vs. charge in the presence of sediment, seen in Fig. 3B, also corresponds closely to that for the blank in absence of sediment, except that the initial potential is more negative as a result of more effective degassing and oxygen removal.

The data support the conclusions reached previously that spectroelectrochemical titrations can be successfully applied to the investigation of redox capacity of soils. However, the aquifer sample represents a lower limit of the applicability of the technique with the solids loading used

here. The oxidative redox capacity of the aquifer material is so low that it cannot be resolved with respect to the relative uncertainty of the charge measurements. It can be estimated, however, based on the uncertainty of charge measurements, which is certainly less than 0.25 mcoul, that the oxidative redox capacity of this system is less than 3 microequivalents per gram of aquifer material. This is in the lower range reported by Heron et al. (2), but considerably lower than the values reported by Barcelona and Holm (1). Thus an upper limit can be established for the oxidative redox capacity of this aquifer material sample. This is considerably smaller than the results obtained from a very iron rich and organic carbon rich pond sediment investigated previously (3,4), which had an oxidative capacity of ca. 700 μequiv/g sediment for a particle size range smaller than 2 μm average diameter. The use of significantly larger particles in this sample is probably a factor, since there is evidence to suggest that the titratable oxidant in this aquifer material may be available iron (III) species, which are presumably distributed over the surface of sediment particles. Since surface area grows inversely in proportion to particle diameter, with a 1/r dependence if the particles can be treated as hard spheres, or a $1/r^{0.6}$ dependence if the particles are fractal (8), it is likely that these larger particles will inherently have a smaller redox capacity in mequiv/gram of sediment than the smaller particles used previously, but it is also clear that the oxidant loading per unit area of particle in the Columbus aquifer material is inherently lower than in the Beaver Dam sediment previously studied (3.4). Preliminary evidence from other workers at Tyndall Air Force Base suggests that the iron content of the Columbus aquifer material is also significantly lower than that of the Beaver Dam sediment. Clearly the relation between iron and redox capacity will need to be explored in greater depth.

Experiments to assess the temperature dependence of pressure in a closed vessel were carried out between the temperatures of 13 °C and 36 °C. The change dV in the volume of a closed vessel can be expressed as a function of changes in the pressure dp and temperature dT of the system, as given by the expression:

$$dV = \partial V/\partial p)_{T,c}\, dp + \partial V/\partial T)_{p,c}\, dT = (V/K_c)\, dp + \alpha_c\, V\, dT, \qquad (1)$$

where V is the initial volume at specified initial temperature and pressure, K_c is the bulk elastic modulus of the container material, and α_c is the volumetric coefficient of thermal expansion of the container material. The corresponding expression for the liquid sample in the container is:

$$dV = -\partial V/\partial p)_{T,l}\, dp + \partial V/\partial T)_{p,l}\, dT = -(V/K_l)\, dp + \alpha_l\, V\, dT, \qquad (2)$$

where K_l is the bulk elastic modulus of the liquid, and α_l is the volumetric coefficient of thermal expansion of the liquid. The pressure coefficient terms have opposite signs in equations (1) and (2), because an increase in internal pressure inside the container tends to increase the internal volume of the container, but tends to decrease the volume of the liquid contained therein. Equating the volume changes for container and liquid, equations (1) and (2) can be combined to yield the following expression for pressure change as a function of temperature:

$$dp = \frac{\alpha_l - \alpha_c}{\frac{1}{K_c} + \frac{1}{K_l}}\, dT \qquad (3)$$

The values of the parameters used were obtained from Perry and Chilton's *Chemical Engineering Handbook* (9): $\alpha_c = 4.80 \times 10^{-5}$ $(°C)^{-1}$ for stainless steel; $K_c = 2.8 \times 10^7$ pounds per square inch (psi) for stainless steel, and $K_l = 2.989 \times 10^5$ psi for water. The thermal expansion coefficient of the liquid was given by the expression

$$\alpha_l = a_1 + 2\, a_2\, T + 3\, a_3\, T^2, \qquad (4)$$

where $a_1 = -5.96 \times 10^{-5}$ $(°C)^{-1}$, $a_2 = 7.91 \times 10^{-6}$ $(°C)^{-2}$, and $a_3 = -4.09 \times 10^{-8}$ $(°C)^{-3}$, which were obtained from regression analysis of the temperature dependence of the thermal expansion coefficient of water over the range from 0 - 35 °C (9).

The predicted pressure change relative to an initial temperature of 25 °C was calculated by substituting equation (4) for the temperature dependence of α_l directly in equation (3) and by integrating equation (3) with respect to temperature from 25 °C to the temperature of measurement. Figure 4 illustrates the remarkable pressure change of 930 psi over the ca. 22 °C temperature range between 13.8 and 35.4 °C. Such a large pressure change implies that use of a closed cell in a thermally unregulated spectrophotometer will inevitably lead to leakage (or even destruction of the cell) unless a solution expansion volume is provided. Also shown in Figure 4 is a plot predicting the pressure change based on the known thermal expansion coefficients of water. The shape of the experimental plot of pressure vs. temperature is in good qualitative agreement with the shape predicted based on the thermal expansion coefficient of

water. The predicted pressure range is slightly greater than the observed change, but the predicted curve merges completely with the experimental curve if the predicted pressure range is normalized to the observed pressure range.

Conclusions

It was established that the Columbus aquifer material had a very low oxidation capacity, with an estimated upper limit of ca. 3 μequiv/g of sediment. This result is on the order of the lower redox capacity samples investigated by Heron (2), and considerably lower than the levels reported by Barcelona and Holm for similar aquifer materials (1), or reported by us for pond sediments (3,4). These results give support to the contention by Heron et al. that the results of Barcelona and Holm may be too high. We were able to establish that the use of absorbance measurements for indirect quantitation by monitoring of an optical reporter molecule is insufficiently sensitive to provide reliable quantitation at the redox capacity levels characteristic of Columbus aquifer material. The problem of insufficient resolution of the redox capacity of Columbus aquifer material must be resolved for further progress to be made. The most promising approach is to switch from absorbance as the optical probe to fluorescence. The detection limit of the method described here is limited by the quantity of sediment loading which can be tolerated while still obtaining acceptable absorbance signals. Fluorescence measurements are considerably more immune to light scatter, enabling considerably higher solids loading. We propose for future work to utilize fluorescence detection to improve the attainable resolution and thereby resolve the redox capacity of Columbus aquifer material.

It was also established that astonishingly high pressures can be generated in closed vessels completely filled with liquid, if temperature is not carefully regulated. These results make it clear that vessels used in experiments on liquids must have sufficient compressible volume, whether as a gas phase or a compressible insert (e. g. rubber or other elastic material), to allow liquid expansion without generation of ruinous internal pressures. Such volumes may be very modest, on the order of less than 10 μL, to eliminate this potential problem.

Acknowledgements

We wish to thank the Air Force Office of Scientific Research Summer Faculty Program

for financial support and the staff of the Environics Division, Armstrong Laboratory, Tyndall Air Force Base, for their hospitality and support of our research project. Particular thanks are extended to David Burris, Eila Burr, Tim Campbell, Chris Antworth, and William MacIntyre for their help and encouragement.

References

1. Barcelona, M.; and Holm, T., *Environ. Sci. Technol.*, **1991**, 25, 1565-1572.
2. Heron, G.; Crouzet, C.; Bourg, A. C. M.; and Christensen, T. H., *Environ. Sci. Technol.*, **1994**, 28, 1698-1705.
3. Sullins, T. V.; Anderson, J. L.; Burris, D. R.; and Wolfe, N. L., manuscript in preparation.
4. Sullins, T. V.; and Anderson, J. L., Final Completion Report, AFOSR Summer Faculty Research Program, Tyndall Air Force Base, 1993.
5. Hawkridge, F; and Kuwana, T., *Anal. Chem.*, **1973**, 45, 1021-1027.
6. Mackey, L; and Kuwana, T., *Bioelectrochemistry and Bioenergetics*, **1976**, 3, 596-613.
7. Anderson, J. L.; Kuwana, T.; and Hartzell, C. R., *Biochemistry*, **1976**, 15, 3847 - 3855.
8. Borkovec, M.; Wu, Q.; Degovics, G.; Laggner, P.; and Sticher, H., *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **1993**, 65-76.
9. Perry, R. H.; and Chilton, C. H., eds., *Chemical Engineering Handbook*, 5th Edition, McGraw-Hill, New York, NY, 1973.

Figure 1. Electrochemical Titration Cell Diagram

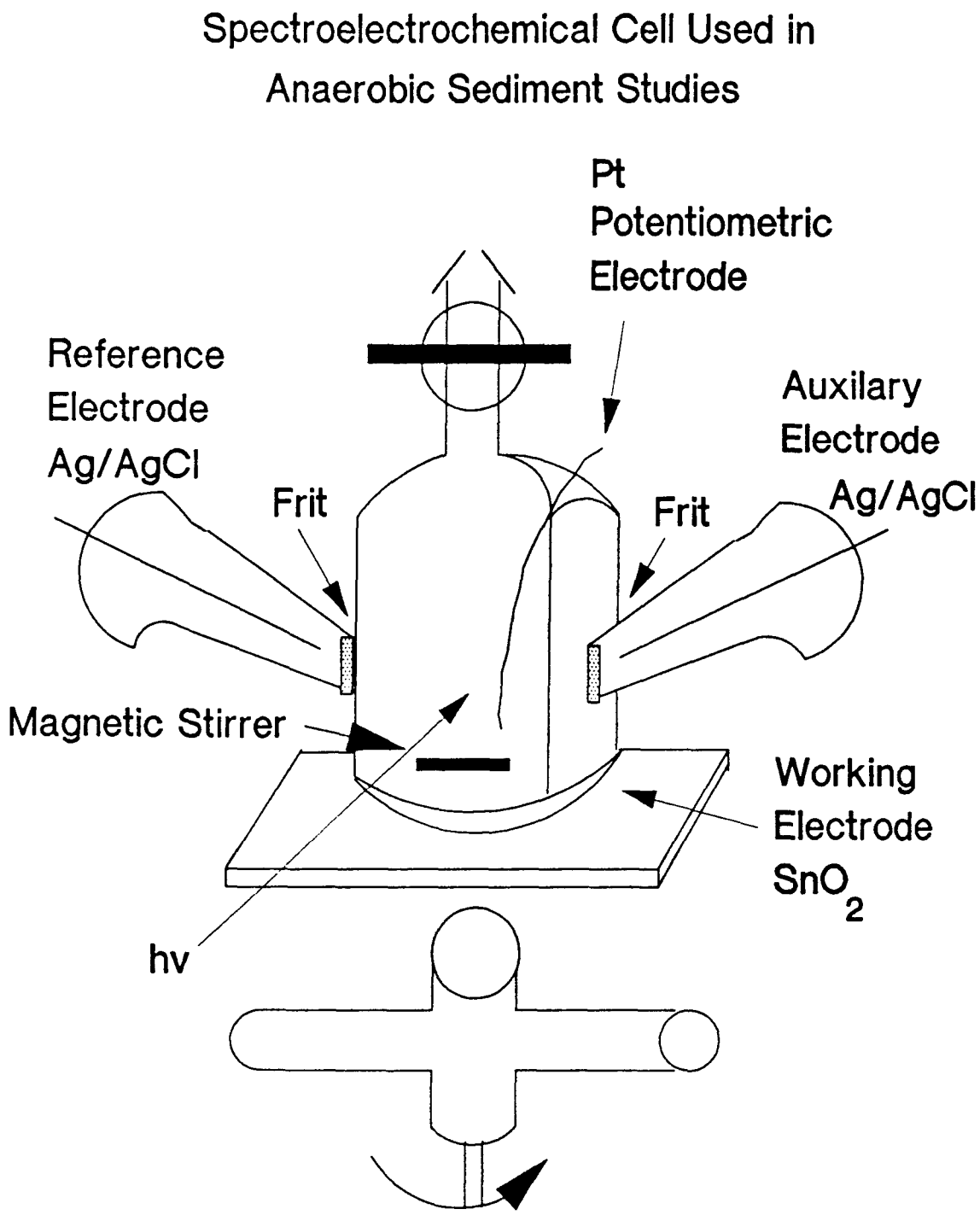
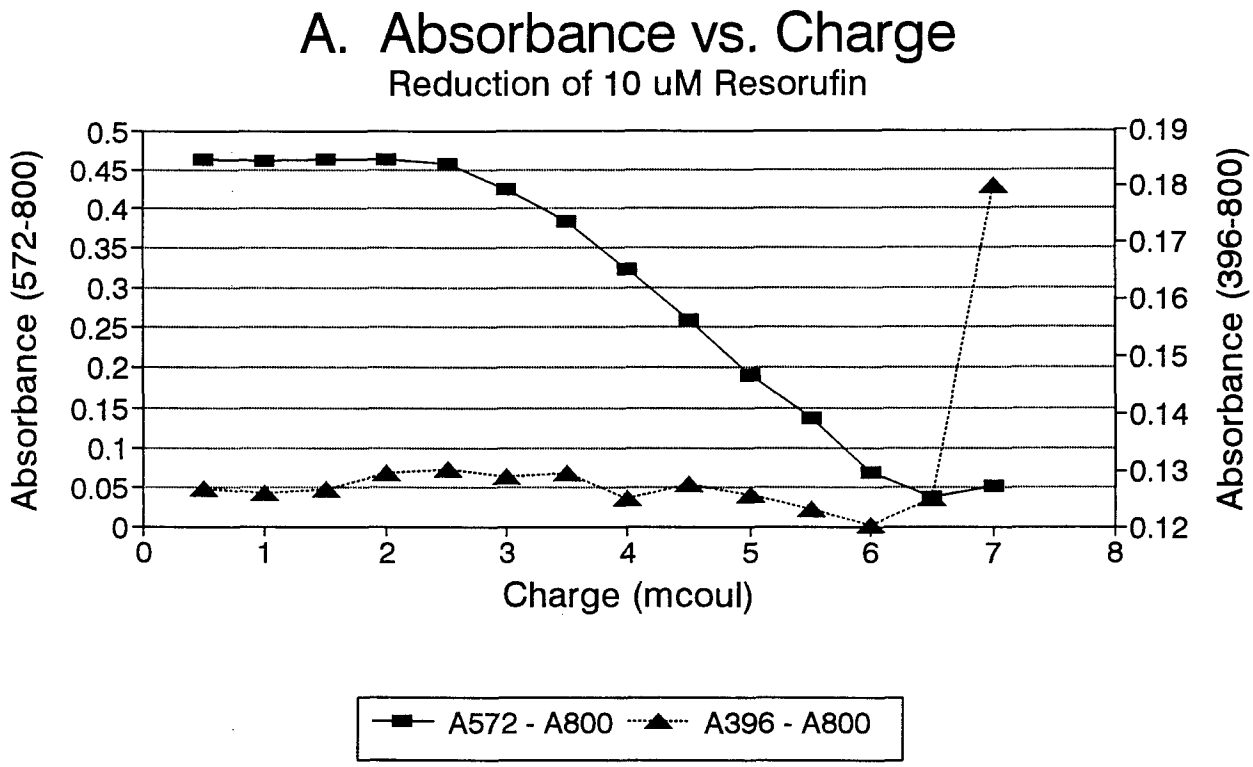


Figure 2. Spectroelectrochemical Titration Plots - Blank Run



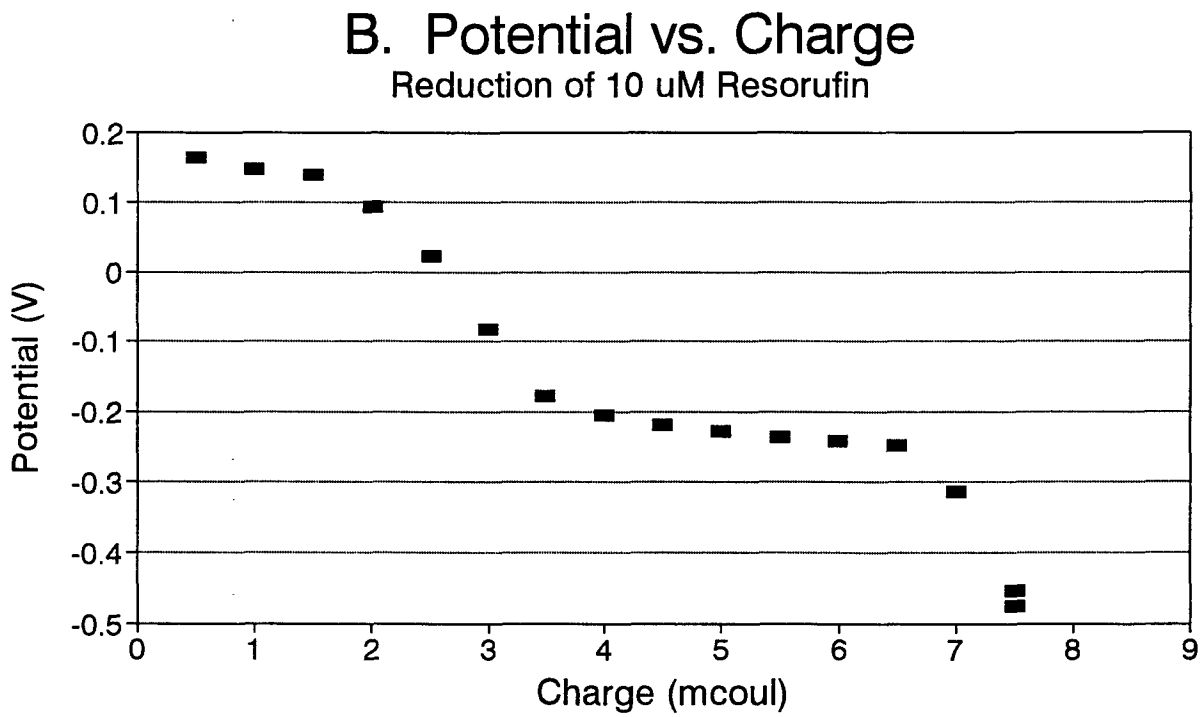
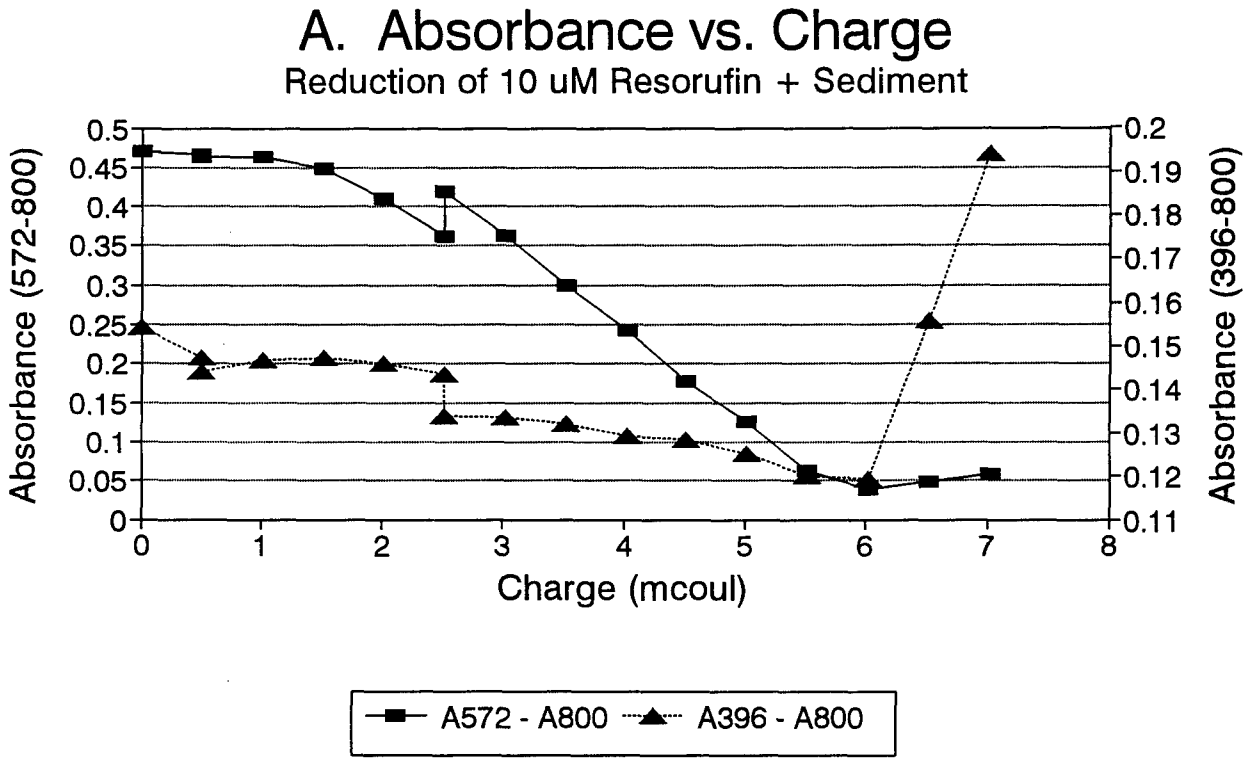


Figure 3. Spectroelectrochemical Titration Plots - Columbus Sediment Run



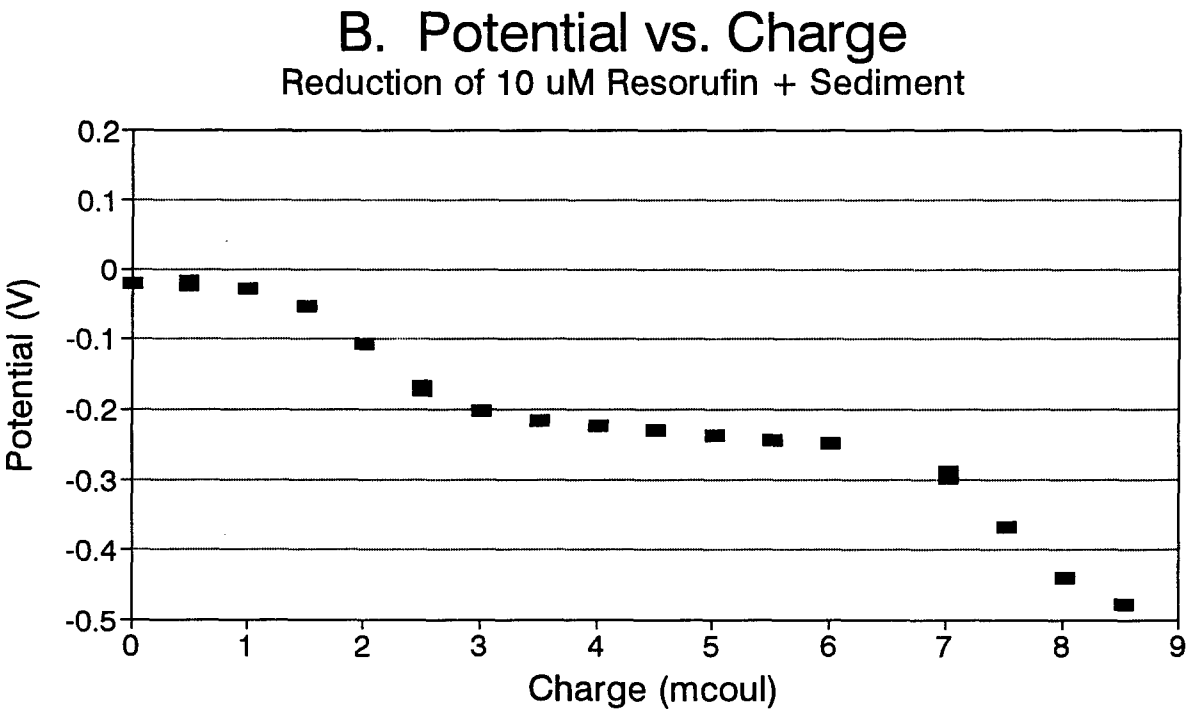
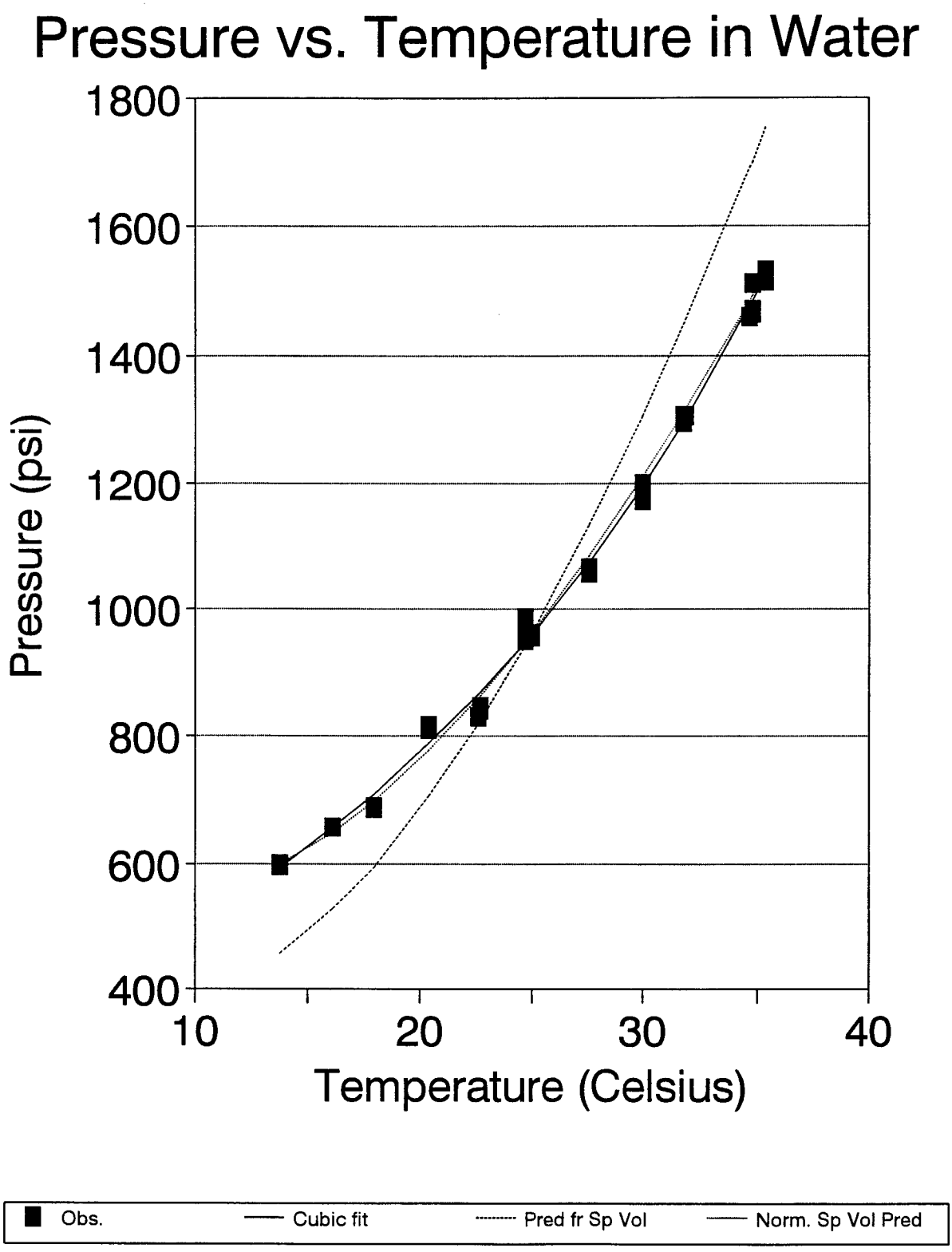


Figure 4. Temperature Dependence of Pressure in a Closed Vessel of Water



ATB SIMULATION OF DEFORMABLE MANIKIN NECK MODELS

Hashem Ashrafiuon
Assistant Professor
Department of Mechanical Engineering

Villanova University
Villanova, PA 19085

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

August 1994

ATB SIMULATION OF DEFORMABLE MANIKIN NECK MODELS

Hashem Ashrafiuon
Assistant Professor
Department of Mechanical Engineering
Villanova University

Abstract

The ATB (Articulated Total Body) is a body dynamic model of the human body used at the Armstrong Aerospace Medical Research Laboratory (AAMRL). The model is used to determine the mechanical response of the human body in different dynamic environments such as aircraft pilot ejection, sled test, etc. The new version of the ATB allows for segments to be treated as deformable bodies for more accurate prediction of dynamic response. However, accurate finite element models of the deformable segments are required for such analysis to be useful. In this study, finite element models of the Hybrid III and II dummy necks are incorporated into the revised version of the ATB model. Quasi-static Hybrid III and II neck simulations and several Hybrid III dynamic head-neck simulations are presented and compared with the experimental results where available. It is shown that the simulation results show good agreement with the available experimental results.

ATB SIMULATION OF DEFORMABLE MANIKIN NECK MODELS

Hashem Ashrafiuon

Introduction

The Articulated Total Body (ATB) is used at the Armstrong Aerospace Medical Research Laboratory (AAMRL) for predicting gross motion of the human body under various dynamic environments. The new version of the ATB model allows for treatment of the individual segments as deformable bodies (Ashrafiuon, 1993). The model assumes, however, that the displacement of a deformable body relative to its own reference frame is small and linear. Therefore, linear elastic displacement field of a segment may be defined by linear combinations of vibration normal (deformation) modes. The vibration normal modes are determined using finite element modeling and modal analysis (Shames, 1985) of the deformable bodies.

The new capability is particularly useful for modeling of human and dummy necks which are clearly deform in most dynamic situations. In this study, the relevant frequencies and mode shapes obtained from finite element models of the Hybrid III and Hybrid II dummy necks are used in the ATB model simulation of the Hybrid III and Hybrid II Static Neck Tester (Baughn, et al. 1993) and several Hybrid III Head/Neck Pendulum tests (Spittle, et al. 1992). The simulation results are compared with the experimental results where available.

This research is the continuation of the research performed by the Principal Investigator in the 1993 Summer Research Program. The mathematical theory and formulation of the problem is presented in the report associated with the previous research (Ashrafiuon, 1993). This research is also in conjunction with the research performed by R. Colbert under the 1994 Graduate Student Summer Research Program. Detailed description of the finite element models of the Hybrid II and Hybrid III manikin neck models are presented in the research performed by Colbert (1994). A short description of each model is presented in the following sections.

Hybrid II Model

The Hybrid II neck is a symmetric, cylindrical butyl rubber mold (E = 1200 psi) with steel end plates, as shown in Fig. 1. The cylinder has a 3" diameter and a length of almost 5". A 1/2" diameter hole runs through the length of the structure. Hybrid II is bolted to the manikin upper body at one end and pinned to the manikin head at the other. 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 set of bending modes at just over 41 Hz. The next two modes correspond to torsion and tension. The second set of bending modes are the fifth and sixth mode shapes.

Hybrid III Model

The Hybrid III neck only partially made with butyl rubber. The Hybrid III is segmented with aluminum plates in between the rubber sections to simulate the vertebral disks, as shown in Fig. 2. Length of the neck is 5.6" and disks have a 3.4" diameter while the rubber section has a 2.6" diameter. The rubber sections are offset 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 5/8" diameter hole in the neck. The cable is torqued to limit excessively large rotations in the neck.

The first two modes are similar to the first set of bending modes of a cantilever beam and have values of just over 36 Hz. These modes are here referred to as flexion/extension and lateral modes. The third mode is the first torsion mode occurring at a frequency of approximately 57 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 141 Hz and 144 Hz. Finally, the sixth mode shape is the second torsion mode at 184 Hz.

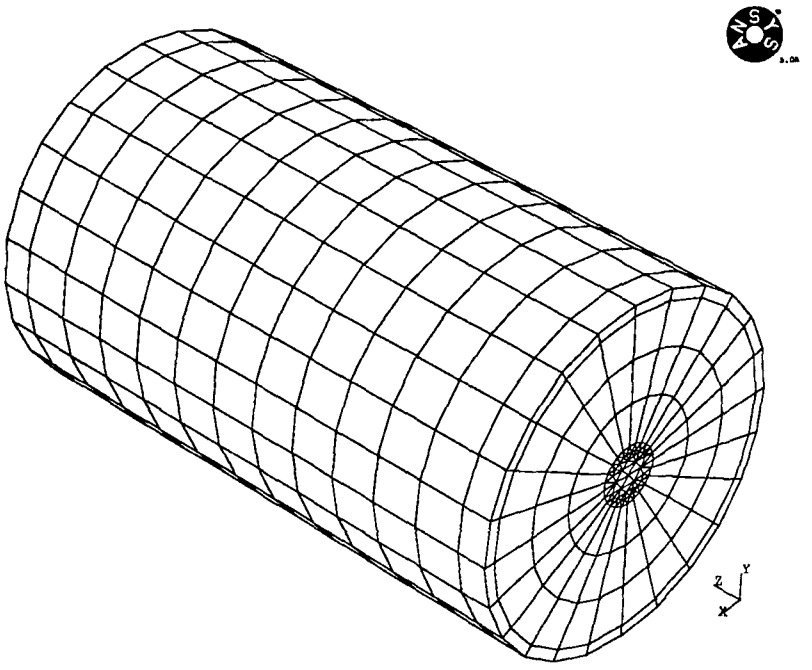


Figure 1. Hybrid II Finite Element Model

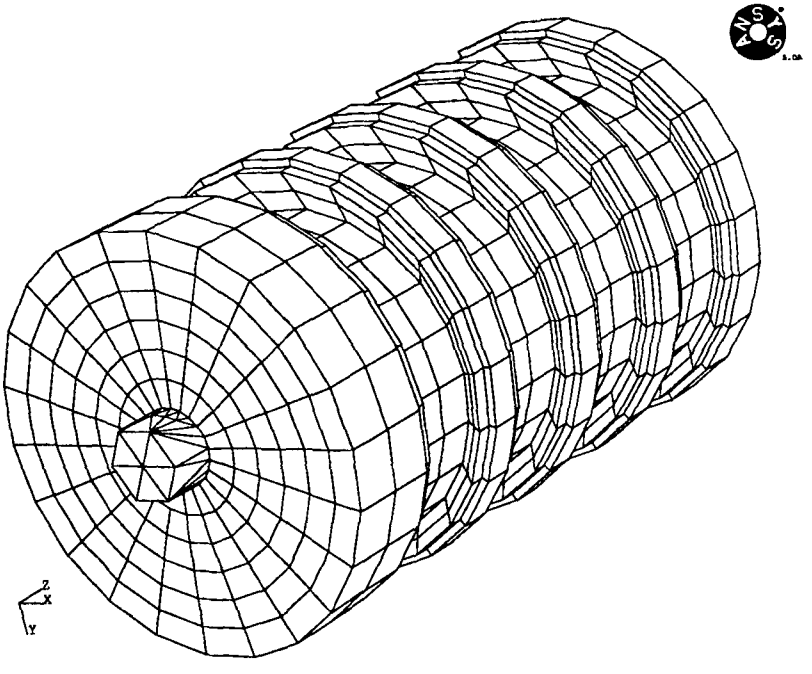


Figure 2. Hybrid III Finite Element Model

Static Test

To simulate the Static Neck Tester (SNT), Hybrid III (or Hybrid II) neck behaves similar to a cantilever beam with a load applied at its free end, as shown in Fig. 3. The load (F) is linearly increased from 0 to 400 lbs in 2 seconds. This simulation is "slow enough" such that dynamic effects are negligible. The bending modes obtained from finite element modeling (Colbert, 1994) are utilized to represent neck deformation. See Baughn, et al. (1993) for detailed description of the experiment.

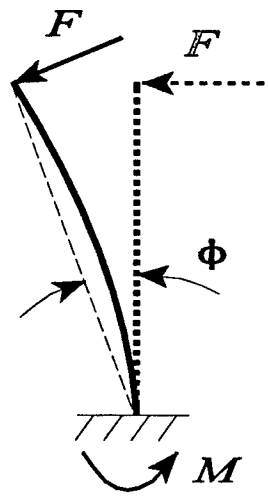


Figure 3. Static Test Neck Model

Figures 4 shows the plots of the static flexion test reaction moment (M) vs. the neck rotation angle (ϕ) for Hybrid II and both flexion and lateral tests for Hybrid III necks. Figure 5 shows a comparison of simulation and test results reported by Spittle, et al. (1992) for Hybrid III static flexion test. It can be seen that the bending stiffness is predicted to be much higher by the simulation for small deformation but slightly lower for large deformation. This is partly because dynamic elastic properties of rubber are used for the simulation which are normally higher than the static properties and also nonlinear (but still elastic) behavior of the rubber is ignored.

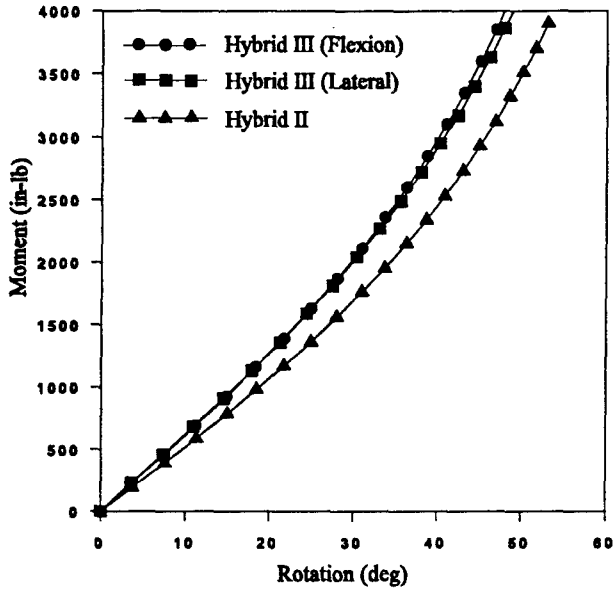


Figure 4. Moment Vs. Rotation in Static Neck Tests

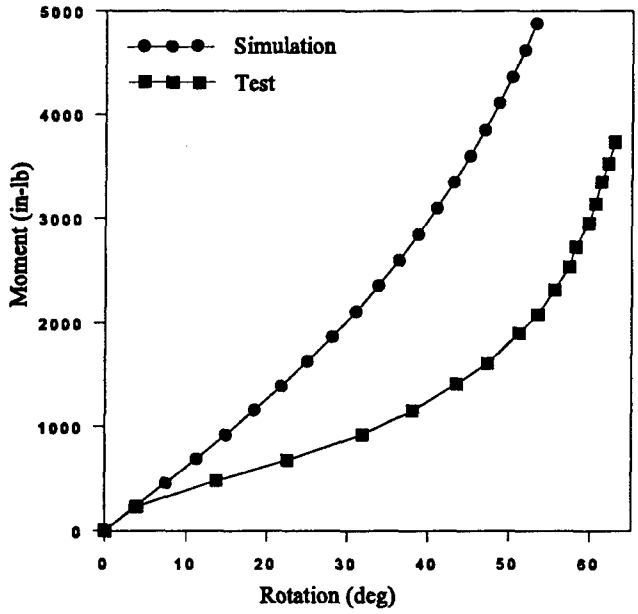


Figure 5. Moment Vs. Rotation in Hybrid III Static Test

Dynamic Test

An ATB model of the Head/Neck Pendulum (HNP) test is shown in Fig. 6. There are three segments (pendulum arm, neck, and head) and two fixed joints (j1 & j2) connecting them. Therefore, both neck rotation (ϕ') and head rotation (ϕ) are purely due to neck deformation. The mass and geometric properties reported by Kaleps, et al. (1988) are used. A damping ratio of $\zeta = 0.9$ is used for the neck which has an effective value of about 0.2 since the head introduces significant increase in system inertia. See Spittle, et al. (1993) for a detailed description of the experiment.

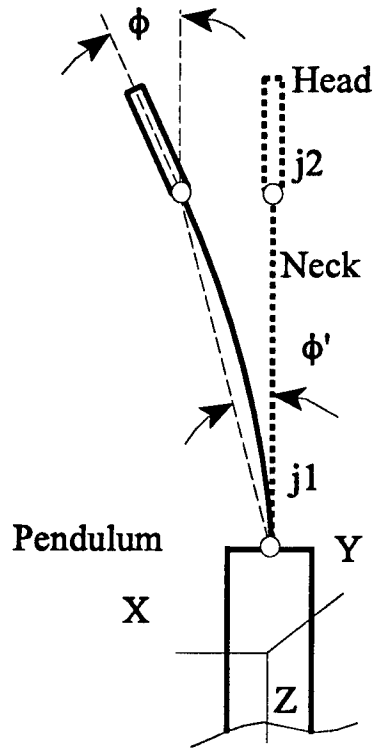


Figure 6. Head/Neck Pendulum Test Model

The excitation is provided through deceleration data obtained from the HNP tests. The deceleration data and the initial impact velocity are applied along X-axis of the pendulum arm for flexion/extension tests and Y-axis for lateral tests. Several flexion tests have been performed by dropping the pendulum arm from 20^0, 40^0, 60^0, 80^0, and 120^0 angles and one lateral test from 65^0 resulting in impact velocities ranging from 54.74 to 275.4 in/sec. The first four bending modes which represent two flexion/ extension and two lateral deformation modes are selected. These modes are

sufficient for accurate modeling of neck deformation due to physical characteristics of the system and range of frequencies obtained from the deceleration data.

Figures 7-12 show comparison of head and neck rotation (ϕ & ϕ') data obtained from the new ATB simulations and HNP tests. It can be seen that the simulation follows the test data closely in most cases. However, for the first three cases the test data seems to suggest that the neck is oscillating about a negative mean value. This could be due to a defect (slight permanent deformation) in the neck. Also note that the fundamental natural frequency of the system is about 5 Hz about 7 times smaller than neck's frequency. This, of course, is because of the addition of head mass to the neck which has a similar effect on the system damping as observed by the peak to peak ratio. Finally, the sharp peak observed in the 120^0 case (Fig. 11) may be due to the linear deformation assumption.

Figure 13-17 compares the head C.G. forward (x) accelerations. The simulation results in all cases follow the same pattern as the test data. However, the initial high acceleration peaks are not accurately predicted by the simulation. These accelerations may be due to activation of the nodding blocks at the head-neck joint (Spittle, et al. 1992). The nodding blocks may be modeled as a rotational spring at the joint about the pin direction (y axis). The interference from the pre-torqued cable which runs through the neck may have also contributed to the high accelerations. These effects are not included in the model at this point.

Figure 18 compares the resulting moments between the 60^0 flexion and 65^0 lateral simulations tests. This comparison clearly demonstrates that the high peak occurring at about 0.05 seconds into the simulation is due to the nodding blocks since the peak is not observed in the lateral test.

Conclusions

Finite element models of the Hybrid III and II have been incorporated into the new version of ATB for quasi-static (Static Neck Tester) and dynamic (Head/Neck Pendulum test) simulations and comparison with the experimental data. The models were shown to be stiffer than the actual neck since only elastic properties under dynamic loading were used. Dynamic simulations have been shown to have good agreement with the experimental results except for some high acceleration peaks which are not predicted accurately due to absence joint stiffness modeling.

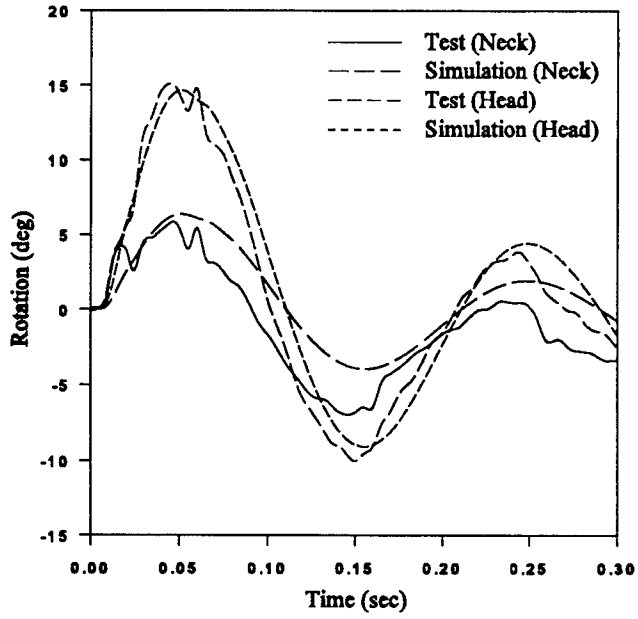


Figure 7. Hybrid III Rotations in HNP 20^0 Flexion Test

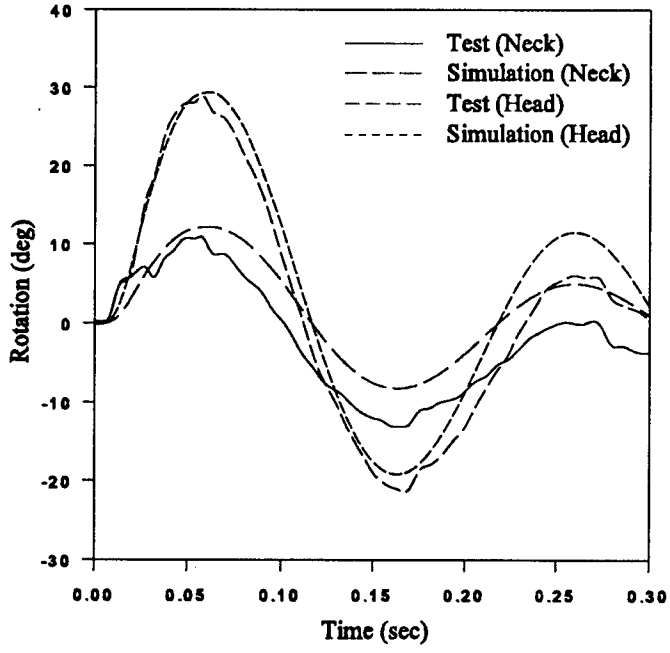


Figure 8. Hybrid III Rotations in HNP 40^0 Flexion Test

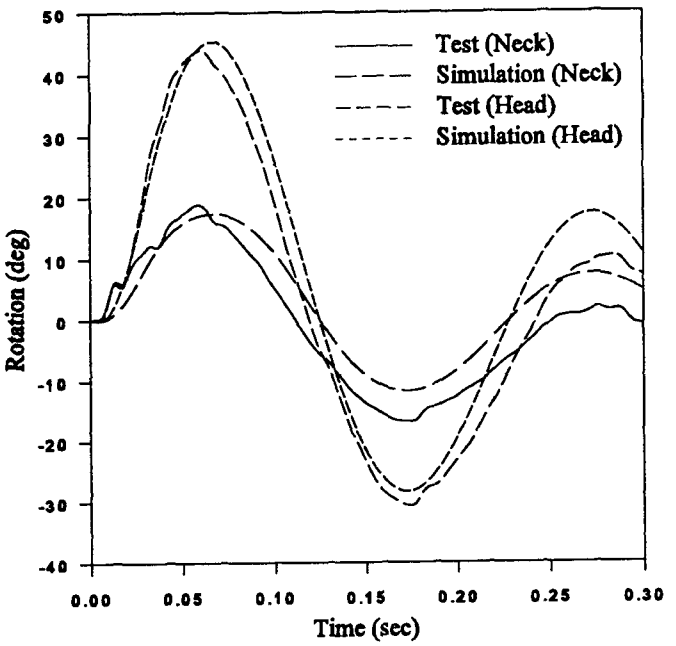


Figure 9. Hybrid III Rotations in HNP 60^0 Flexion Test

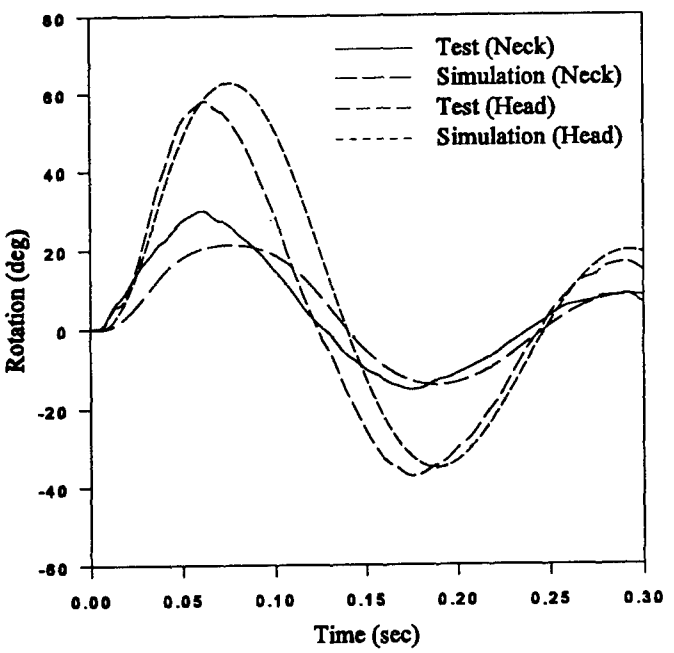


Figure 10. Hybrid III Rotations in HNP 80^0 Flexion Test

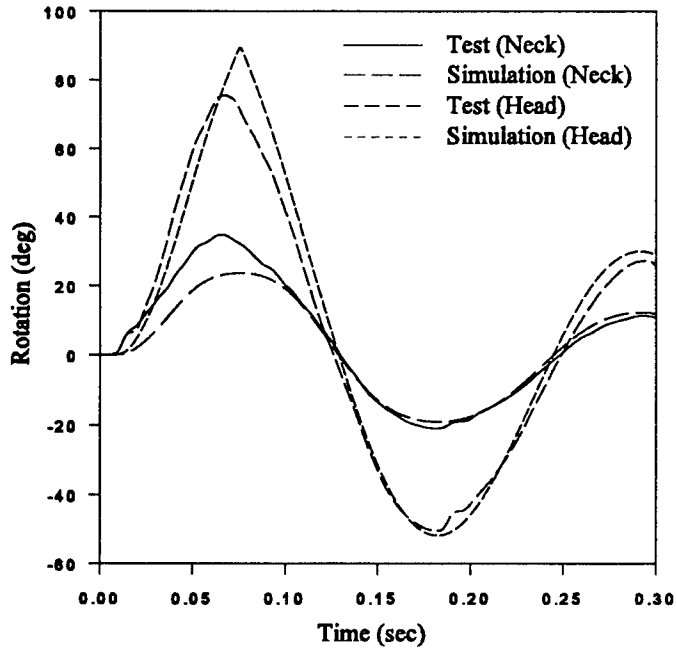


Figure 11. Hybrid III Rotations in HNP 120^0 Flexion Test

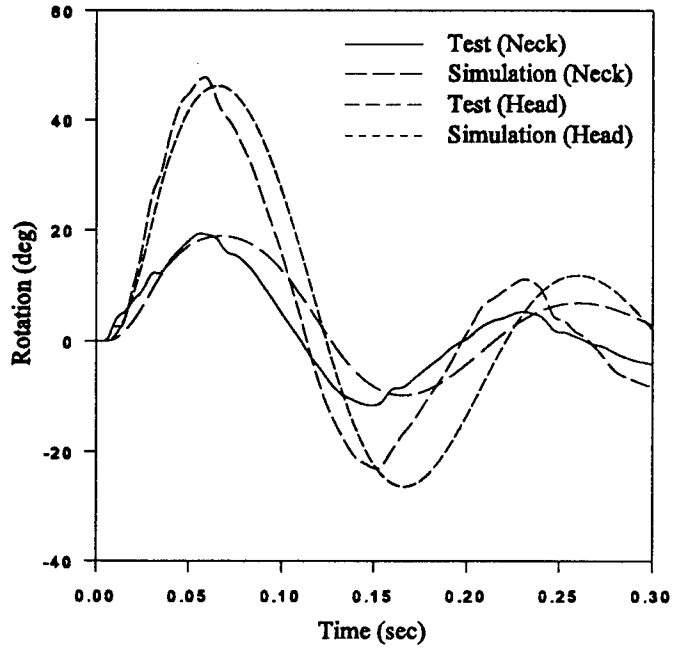


Figure 12. Hybrid III Rotations in HNP 65^0 Lateral Test

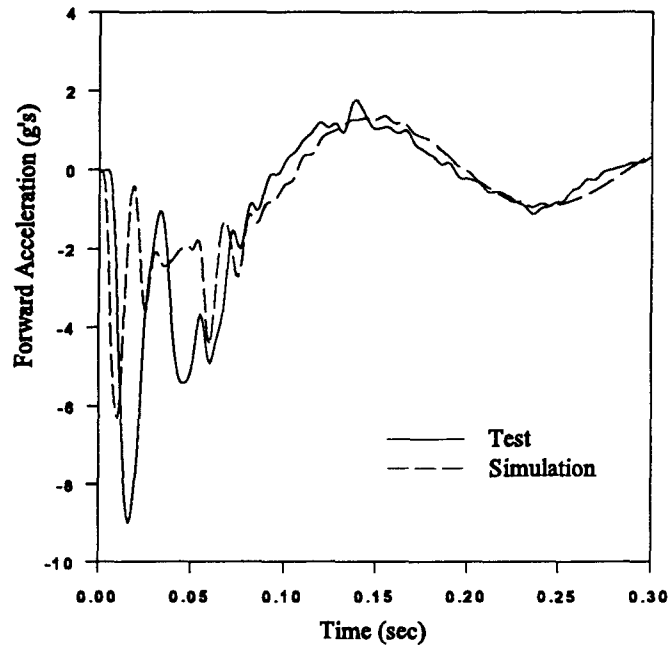


Figure 13. Head *x* Acceleration in HNP 20^0 Flexion Test

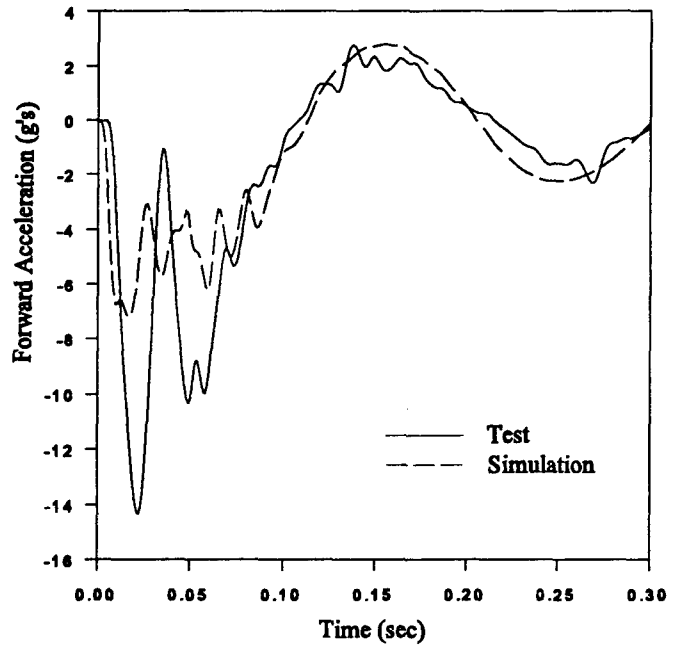


Figure 14. Head *x* Acceleration in HNP 40^0 Flexion Test

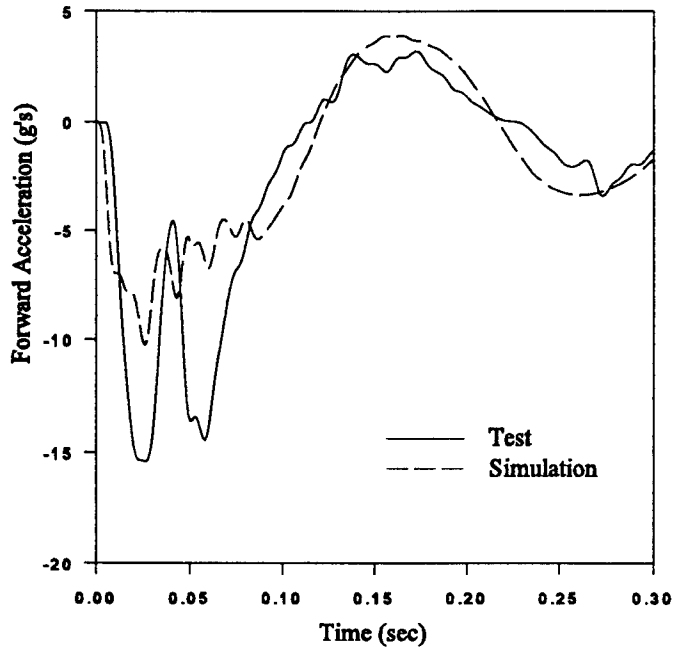


Figure 15. Head *x* Acceleration in HNP 60^0 Flexion Test

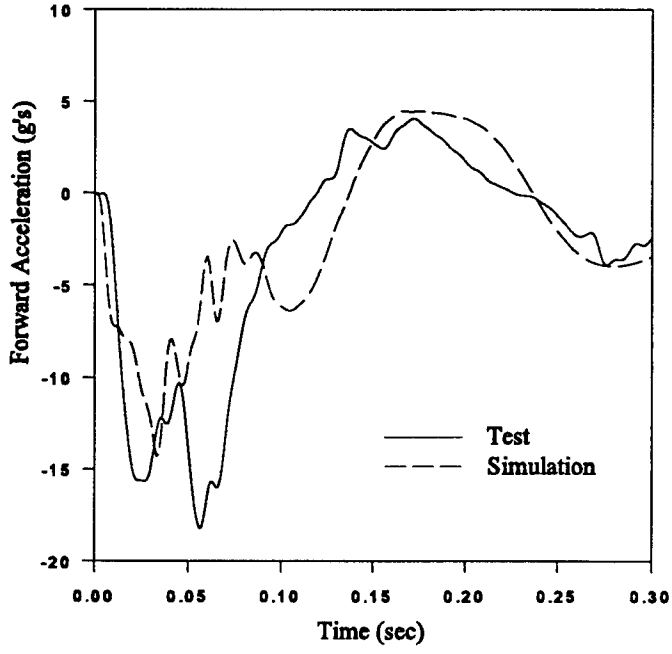


Figure 16. Head *x* Acceleration in HNP 80^0 Flexion Test

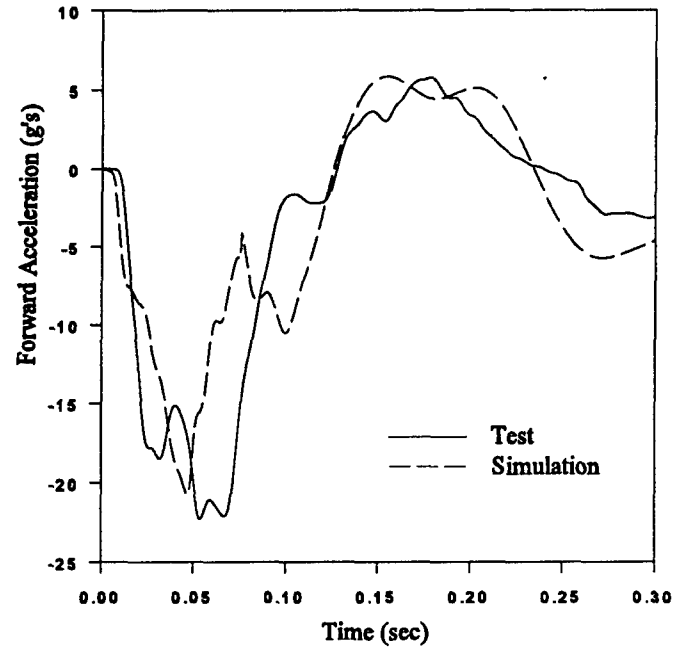


Figure 17. Head *x* Acceleration in HNP 120^0 Flexion Test

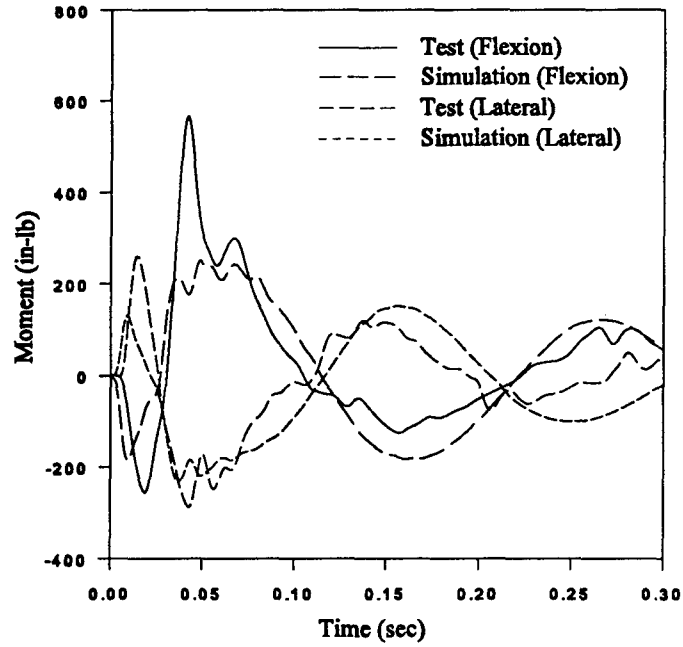


Figure 18. Moments for 60^0 Flexion and 65^0 Lateral Tests

REFERENCES

Baughn, D. J., Spittle, E. K., and Thompson, G., 1993, "A New Technique for Determining Bending Stiffness of Mechanical Necks," SAE Technical Paper No. 930099, SAE International Congress and Exposition, Detroit, MI.

Ashrafiuon, H., 1993, "Modeling of Flexible Bodies for The ATB Model," U. S. Air Force Summer Faculty Research Program, Armstrong Laboratory, Wright-Patterson Air Force Base, OH.

Colbert, R., 1994, "Finite Element Modeling of Manikin Necks for The ATB Model," U. S. Air Force Graduate Student Summer Research Program, Armstrong Laboratory, Wright-Patterson Air Force Base, OH.

Kaleps, I., White, R., Beecher, R., Whitestone, J., and Obergefell, L., 1988, "Measurement of Hybrid III Dummy Properties and Analytical Simulation Data Base Development," Technical Report, AAMRL-TR-88-005.

Spittle, E. K., and Baughn, D. J., 1992, "Measurement of Hybrid III Dummy Properties and Analytical Simulation Data Base Development," Technical Report, AL-TR-1992-0049.

Shames, I. H., and Dym C. L., 1985, Energy and Finite Element Methods in Structural Mechanics, Hemisphere Publishing Corporation, McGraw-Hill Book Company, New York.

PRE-SCREENING OF SOIL SAMPLES USING A SOLIDS INSERTION PROBE AND MASS SPECTROMETRY

Stephan B.H. Bach

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

Final Report for:
Summer Research Program
Armstrong Laboratory

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

September 1994

PRE-SCREENING OF SOIL SAMPLES USING A SOLIDS INSERTION PROBE AND MASS SPECTROMETRY

Stephan B.H. Bach
Assistant Professor
Division of Earth and Physical Sciences
University of Texas at San Antonio
San Antonio, Texas 78249-0663

Abstract

One of the primary difficulties with the analysis of environmental samples are the procedures used for the extraction of the target compounds from the sample matrix. It would be advantageous if samples that did not contain target compounds above the minimum detection limits stipulated by the United States Environmental Protection Agency could be identified before they were subjected to the entire extraction process. Our goal is to investigate and develop the use of a solids insertion probe coupled with a quadrupole mass spectrometer for the pre-screening of samples before they are subjected to extraction procedures.

Using sea sand to simulate the soil matrix, we have begun to examine the specific solids probe conditions and temperature profiles necessary for the pre-screening of samples. We have also examined minimum detection limits attainable using this technique.

PRE-SCREENING OF SOIL SAMPLES USING A SOLIDS INSERTION PROBE AND MASS SPECTROMETRY

Stephan B.H. Bach

INTRODUCTION

One of the primary difficulties with the analysis of environmental samples is the processes used for the extraction of the target compounds from the sample matrix. Target compounds that fall under the broad category of volatile organic compounds (VOC) are generally extracted from liquid samples via purge and trap techniques.[1,2] Extractions of the target compounds from solid matrices are more difficult and costly than the extraction of VOC's from liquid samples. In the analysis of semi-volatile target compounds contained in soil matrices the approved extraction methods involve the use of halogenated solvents. These solvents are then discarded thereby creating an additional waste stream. A majority of the soil samples that are processed will result in the identification of no contaminants. This results in the creation of a unnecessary halogenated waste stream. In addition, there is an investment in time and resources to arrive at this negative result. In order to prevent the creation of an unnecessary waste stream, we propose to pre-screen samples using a mass spectrometer coupled with a solids insertion probe. It would be advantageous if these negative samples could be identified before they were subjected to the entire extraction process, thereby saving time and eliminating the associated halogenated solvent waste stream.

There are several emerging methods for eliminating the halogenated solvents from the extraction process and achieving quantitative results. One method currently under development at United States Environmental Protection Agency Laboratories (USEPA) is a thermal vacuum desorption method. Hiatt, Youngman, and Donnelly[3] have reported a vacuum distillation procedure which is currently in the approval process for becoming an EPA test method. They investigated the use of vacuum distillation of water, soil, oil, and fish samples. The analyte recoveries were found to relate to the boiling points of the compounds unless solubilities of the compounds exceeded 5 g/L. Supercritical fluid extraction (SFE) methods using carbon dioxide are also being investigated and developed as alternatives to standard extraction methods.[4,5] The extraction for SFE is generally performed at low temperatures (between 40 °C and 100 °C) but at high pressures (between 170 atm and 400 atm), requiring specialized extraction vessels able to contain the high CO_2 pressures.

Even with these newer methods there is still considerable effort expended in preparing what may be a negative sample for analysis. Developing a routine using currently available and prevailing technology that would identify and eliminate these negative samples before they were processed for analysis would save considerable resources. Ascertaining the presence or absence of the target compounds would conceivably expedite the determination of the need for remediation at a particular site.

In order to address this issue, we investigated the use of mass spectrometry coupled with a direct insertion probe to determine the presence of target compounds in a soil matrix. The goal of this work is to determine the feasibility of analyzing soil

samples by loading a small portion into a capillary tube, mounting it into a direct insertion probe, inserting the probe into the ionizer of a quadrupole mass spectrometer, and analyzing the contents by running a temperature program on the solids probe. If no compounds of interest are detected using this process no further analysis of the sample would be necessary. If on the other hand any target compounds were detected, the remainder of the sample could be treated as usual and analyzed routinely.

EXPERIMENTAL

The mass spectrometric analyses for this investigation were accomplished on a modified Finnigan 5100 GC/MS. Modifications include plugging GC inlet into the ionizer, and replacing the 1/4 inch swagelok plug to the vacuum chamber with a vacuum manifold separated by a valve connecting it from the MS vacuum chamber. The front flange of the MS was replaced with a flange constructed from 304 stainless steel and the original vacuum interface for the direct insertion probe taken from an HP 5890A mass spectrometer. The dead space between the direct insertion probe and the valve opening to the 5100 vacuum chamber was evacuated via the vacuum manifold. The direct insertion probe and its power supply used were also taken from a Hewlett-Packard 5890A mass spectrometer.

Blank samples were prepared using sea sand (washed, Fisher Scientific). Initial experiments used unbaked sea sand, in later experiments the sea sand was baked at 200 °C for a minimum of 6 hours before use. The samples were loaded into a closed-end capillary tube which was then inserted into the solids insertion probe. (Passivated fiber

glass wool (Scientific Instrument Services) was used to keep the contents in the capillary tube.) Introduction of the solids probe into the vacuum chamber was through the above described vacuum lock, allowing the sample to be positioned next to the ionizer. The ion volume, which is designed to allow material from a solids probe into the ionizer, had been previously installed. The probe was heated using the HP power supply from the HP 5890A designed for the probe. Measurement of the temperature was accomplished with an iron-constantan thermocouple located at the tip of the solids probe. The voltages from the thermocouple were read using a Keithley digital multimeter (Model 191) and converted to degrees centigrade using tables found in the CRC.[6] Temperature control was effected by using the potentiometer located on the solids probe power supply.

The sea sand was doped using the Ultra Scientific GC/MS semi-volatiles tuning standard mixture containing decafluorotriphenylphosphine (1002.1 ug/mL), benzidine (1001.2 ug/mL), pentachlorophenol (1000.8 ug/mL), and 4,4'-DDT (1001.7 ug/mL) in methylene chloride. Only 3 uL of the calibration test mixture was introduced into the capillary tube containing the sand.

For each experiment, the data from the solids probe was treated as though a GC run were being performed. Instead of utilizing a temperature program on the GC column we were running a temperature program on the solids probe. The mass spectrometer scanned the contents of the ionizer every two seconds. Runs lasted approximately 20 to 30 minutes each. The resultant data appears as a chromatogram, and individual mass spectra could then be analyzed.

RESULTS

Initial experiments were done using the unbaked sea sand. These experiments were done to establish the presence of any contaminants in the sea sand before its use as the soil matrix for this project. Upon heating of the sea sand to 200 °C, several peaks were observed in the chromatogram (figure 1). Mass spectra for scan 141 (figure 2) and 322 (figure 3) are presented. Scan 141 has a base m/z of 44 indicating the presence of CO_2 in the sea sand. This most likely arises from the decomposition of carbonates as the sand was heated. There are no significant ions observed above m/z = 69 which is residual perfluorotributylamine (cal gas). Similar results are observed in scan 322. The carbon dioxide was most likely due to the presence of crushed sea shells in the sea sand. In order to eliminate the contamination, the sea sand was baked for several hours at 200 °C in an oven before being used as the soil matrix. The baked sea sand did not produce significant quantities of contaminants upon heating of the solids probe (figure 4). The relative ion counts for the baked sea sand were down to less than one-third of the counts observed for the unbaked sea sand. The mass spectrum in figure 5 taken from scan 436 in figure 4 shows a base peak at m/z = 40, but some m/z = 44 is still present. Again, there were no significant ions observed above m/z = 69.

The next step was to determine if organic chemicals contained with in the sea sand could be detected using the solids probe. For this part of the project we chose the semi-volatiles GC/MS tuning standard (Ultra Scientific). This mixture was chosen because it contains compounds with a range of volatilities and the compounds have unique and easily identifiable fragmentation patterns. Only 3 uL of the calibration mixture was

added to the sea sand contained in the test capillary. The results can be seen in the two mass spectra shown in figures 6 and 7. Even though the chromatographic resolution is poor, we were still able to observe the fragments of the some of target compounds placed into the capillary tube.

DISCUSSION

Even though this work is still in its preliminary stages, we have shown that detection limits below that required by EPA methods for soil analysis are achievable using a solids insertion probe. In our case, we were able to detect 3 ppm of organic contamination in the sea sand with out any kind of special tuning or modifications of the instrument. We strongly believe using a solids insertion probe is a promising tool for pre-screening soil samples and will provide substantial cost savings by eliminating further evaluation of uncontaminated samples.

The current difficulties with the method are obvious but not insurmountable. Improvement in the chromatography should be possible by better controlling the temperature on the solids probe. This will be accomplished by using a Finnigan 4610B already equipped with a solids insertion probe. The Finnigan 4610B controls the temperature of the solids probe using the same software that controls the GC temperatures. In this case it will be possible to get a direct correlation between the solids probe temperature and the mass spectrometer scan.

Up to this point only the calibration mixture has been used. Further investigations will involve using the target compounds. These compounds will at first be introduced in

a fashion similar to the calibration mixture. It will take time to determine the necessary temperature programs for the most efficient vaporization of the target compounds from the soil matrix and establish minimum detection limits for the target compounds using the solids probe.

Another point to be addressed in future investigations will be the homogeneity of the soil sample. This is a major source of concern to us since we examining only a small fraction of the sample. Our initial efforts will study methods of homogenizing the soil sample to ensure a representative sample for loading into the capillary tube.

The focus should remain on the fact that this is an efficient and cost effective method for pre-screening samples. This method will not generate a halogenated waste stream and will require less than 30 minutes to perform. Our goal in researching and developing this routine is not to quantify the amounts of target compounds in soil samples, but simply to determine whether any contamination is present above the prescribed minimum detection limits set by EPA.

REFERENCES

1. T.A. Bellar, J. Am. Water Works Assoc., 66, 739 (1974).
2. T.A. Bellar, J.J. Lichtenberg, R.C. Kroner, J. Am. Water Works Assoc., 66, 703 (1974).
3. M.H. Hiatt, D.R. Youngman, J.R. Donnelly, Anal. Chem. 66, 905 (1994).
4. A.P. Emery, S.N. Chesler, W.A. MacCrehan, J. Chrom., 606, 221 (1992).
5. D.R. Gere and E.M. Derrico, LC-GC, 12, 334 (1994).
6. "Handbook of Chemistry and Physics", 62nd Edition, R.C. Weast and M.J. Astle, ed., CRC Press, Inc., Boca Raton, FL, 1981, pg E 107-08.

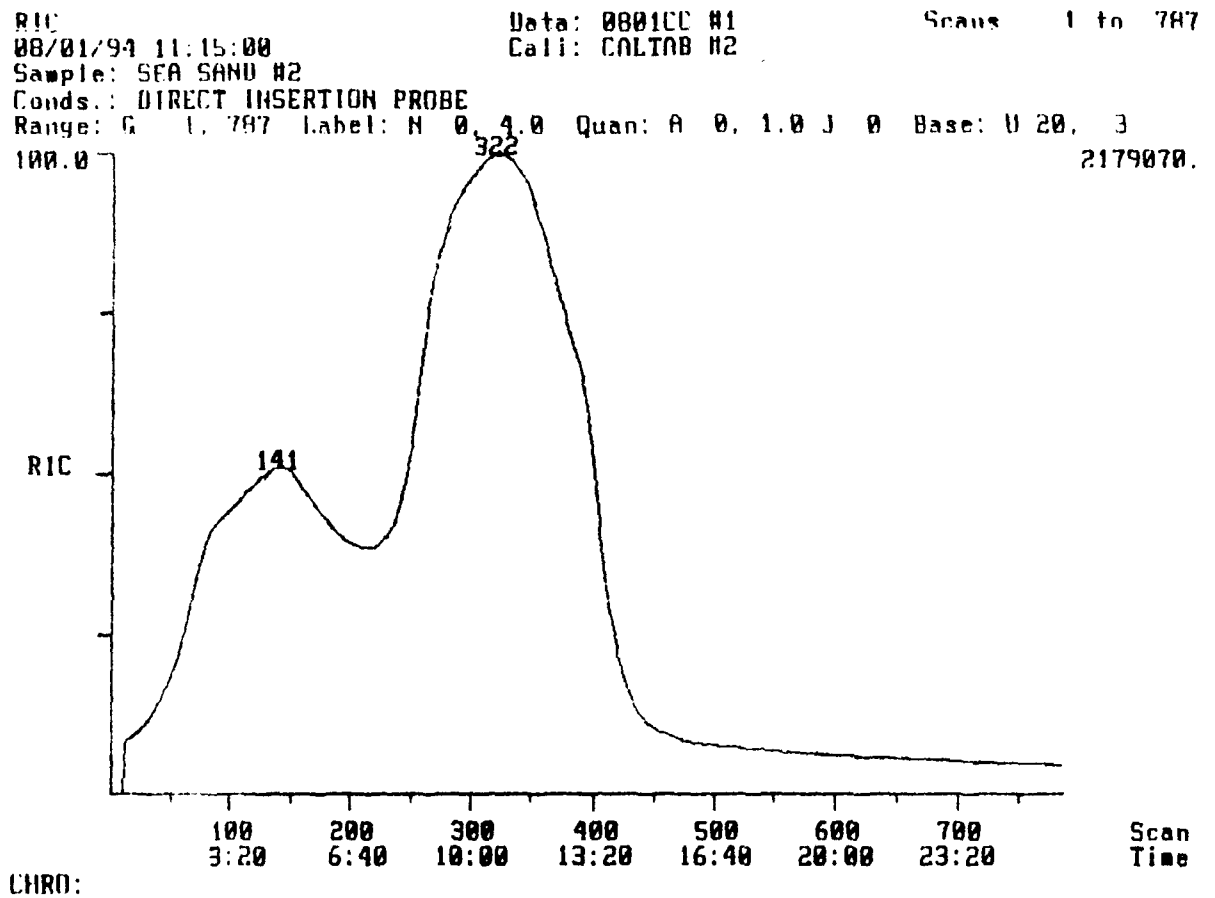


Figure 1: Sea sand chromatogram from heating solids probe, unbaked.

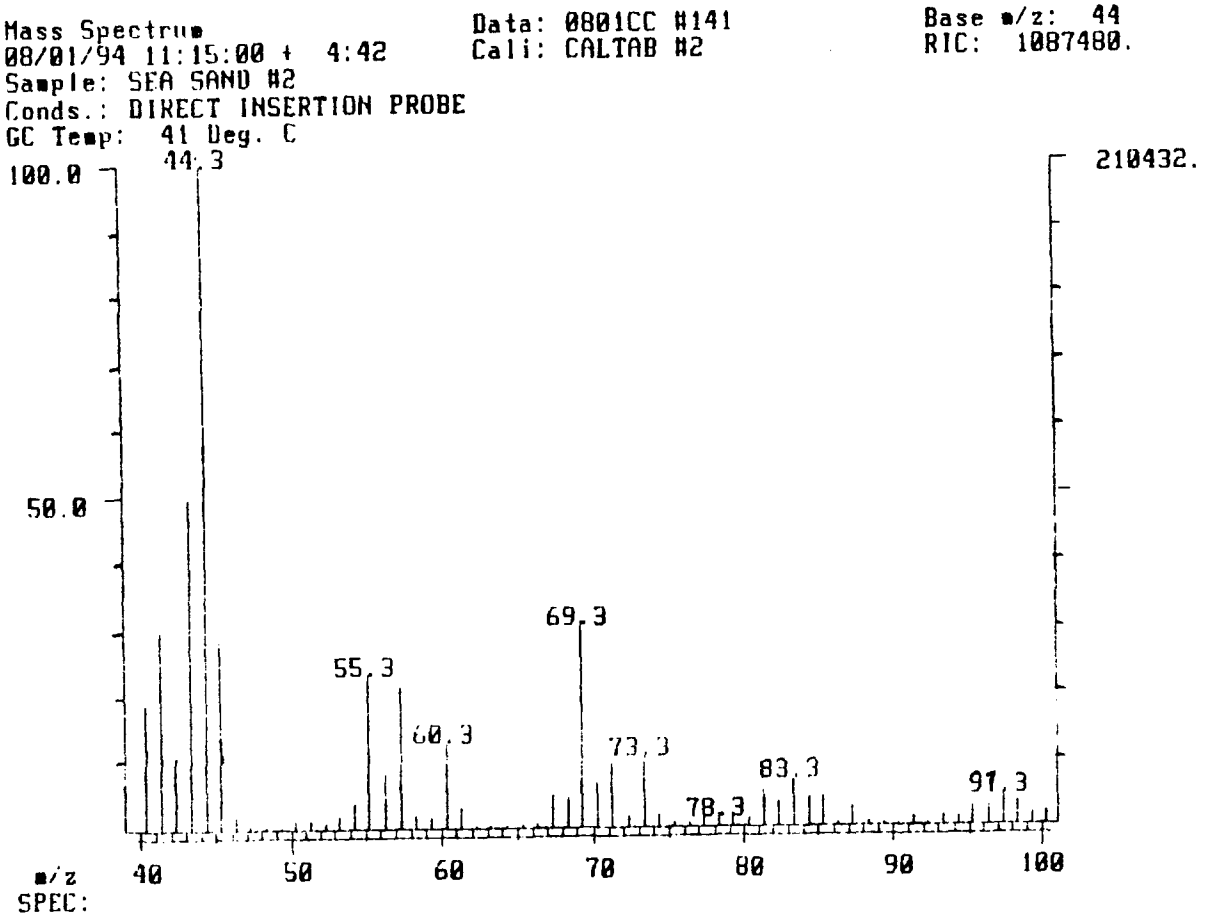


Figure 2: Mass spectrum scan from sea sand chromatogram, unbaked.

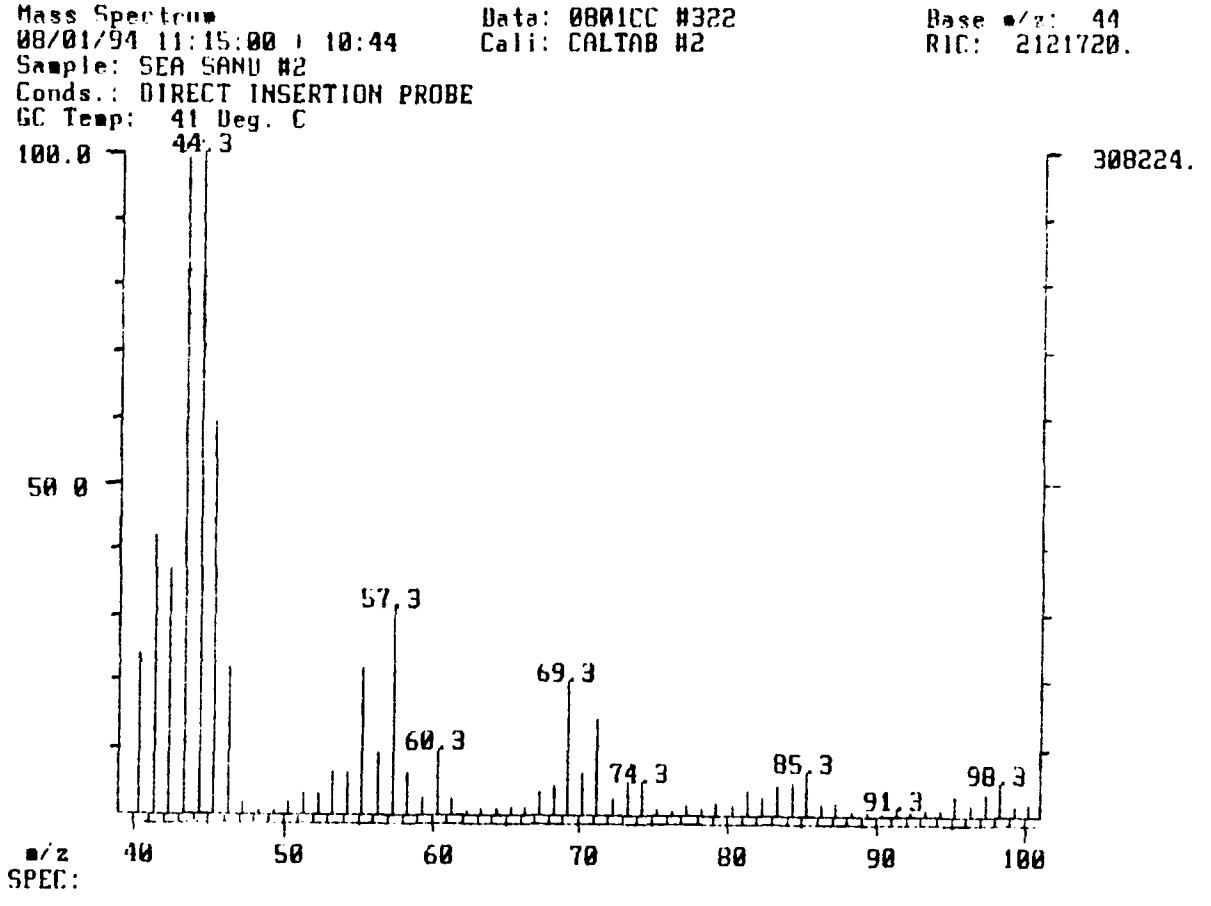


Figure 3: Mass spectrum scan from sea sand chromatogram, unbaked.

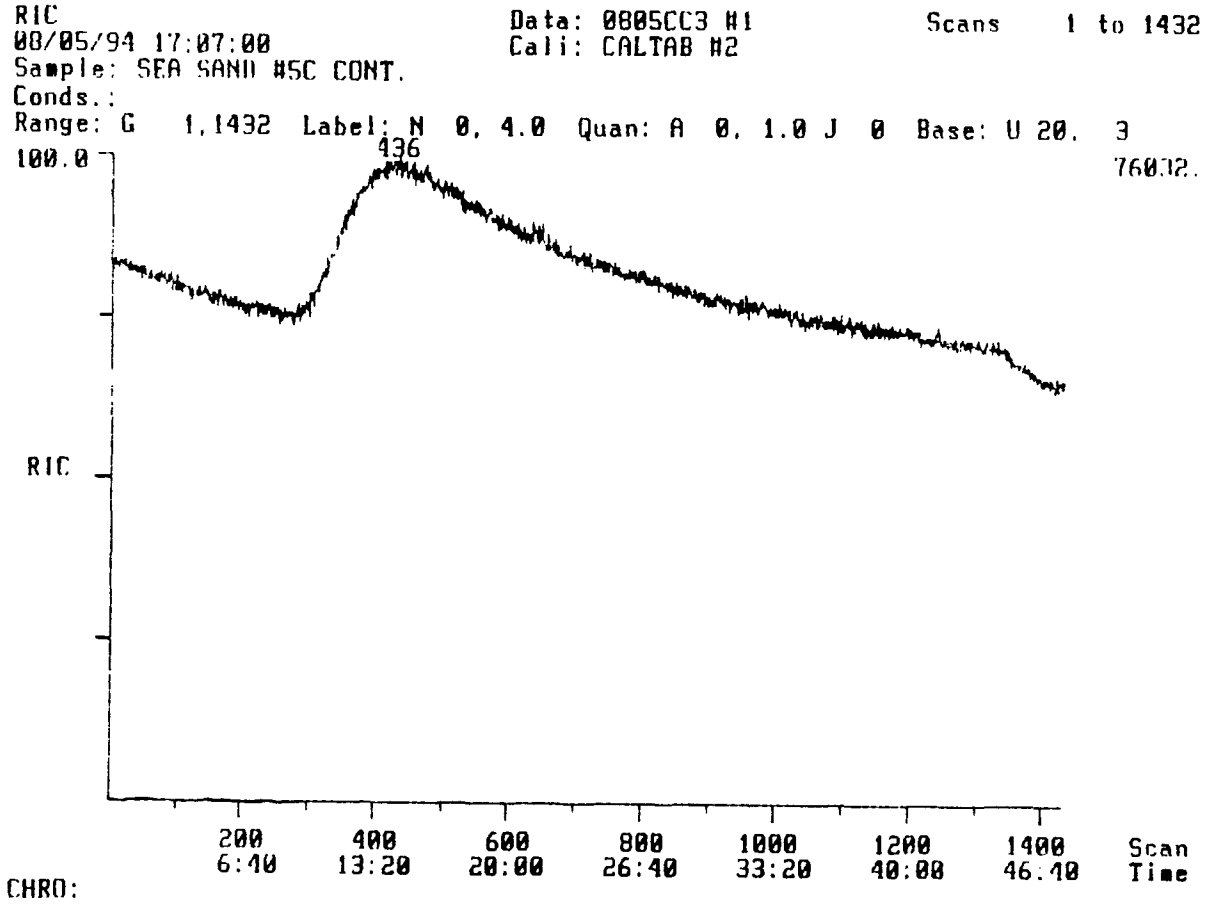


Figure 4: Sea sand chromatogram from heating solids probe, baked.

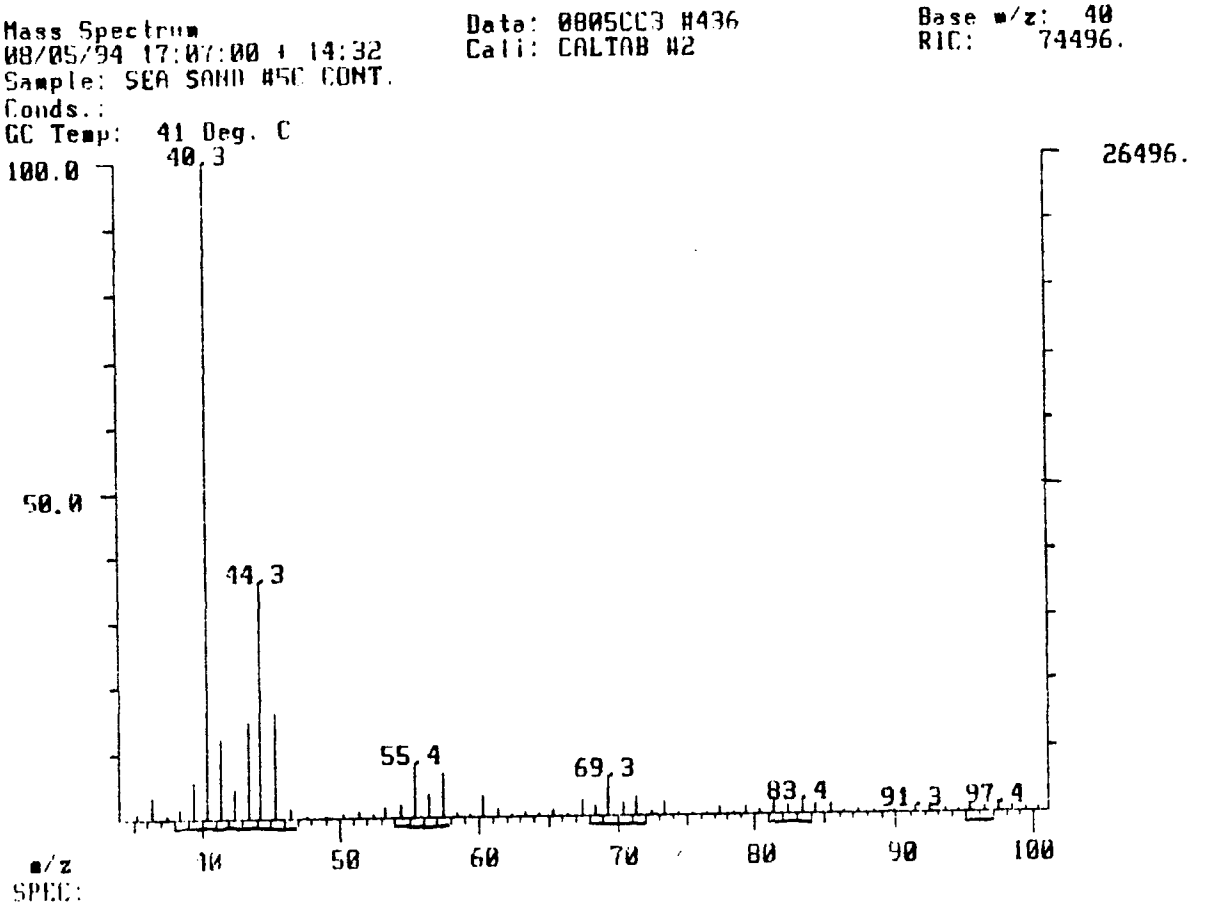


Figure 5: Mass spectrum scan from sea sand chromatogram, baked.

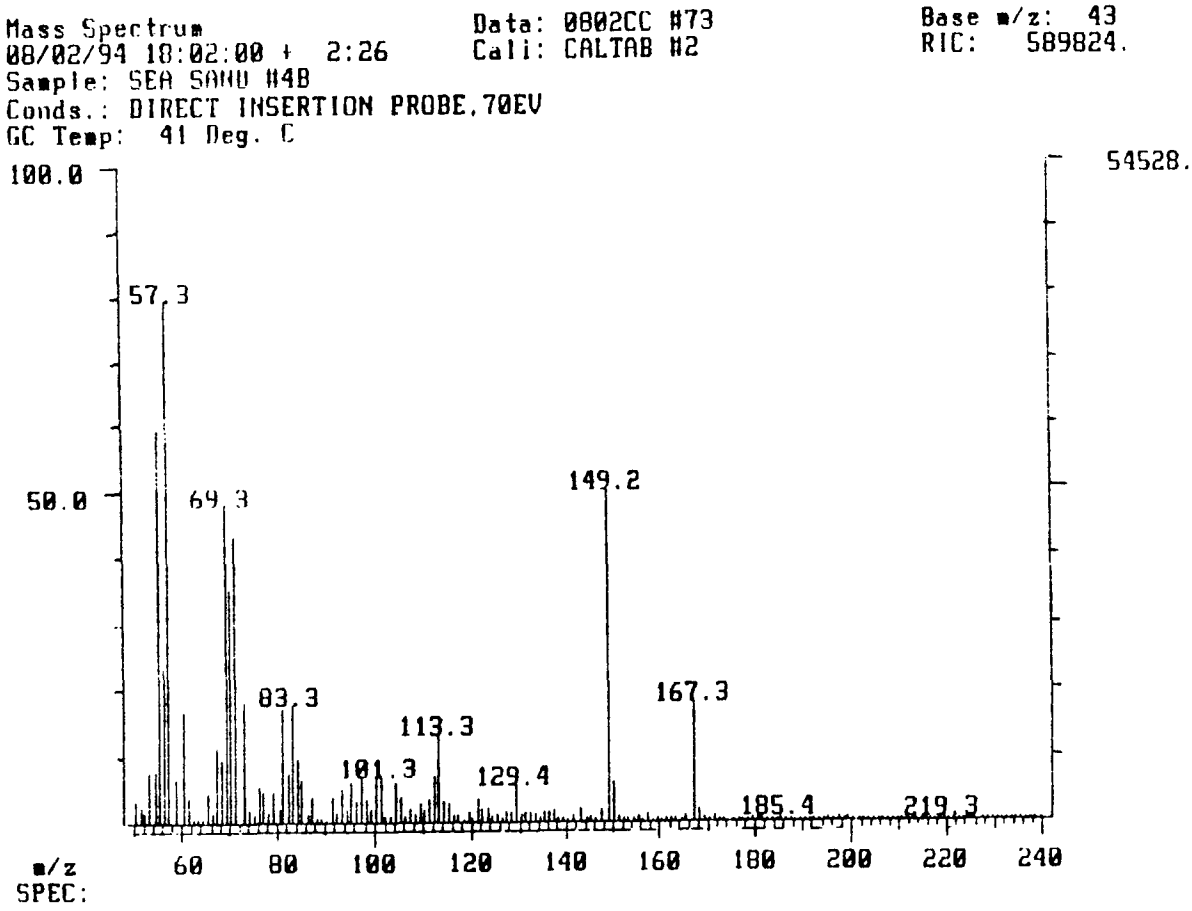


Figure 6: Mass spectrum scan from sea sand doped with 3 uL of the semi-volatile tuning mixture.

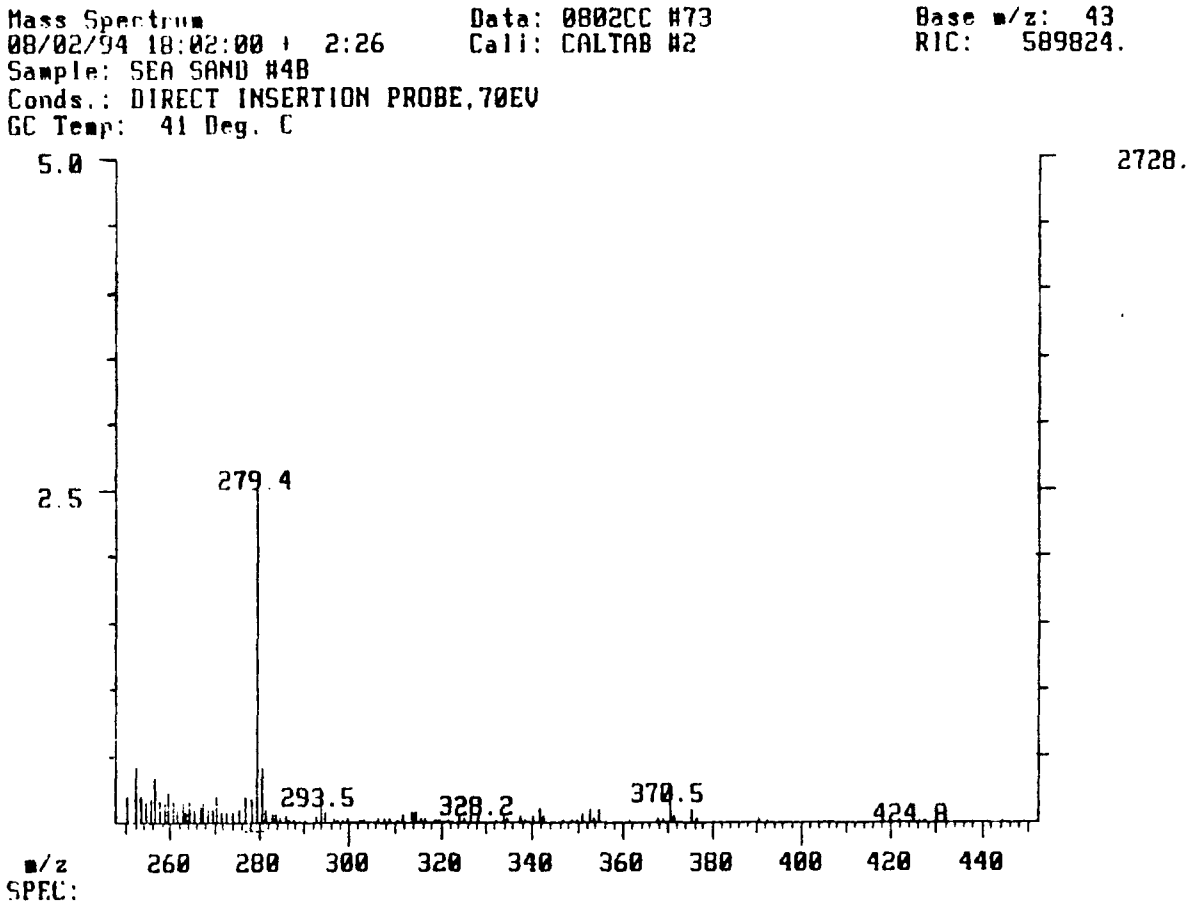


Figure 7: Mass spectrum scan from sea sand doped with 3 uL of the semi-volatile tuning mixture.

RAT PUP ULTRASONIC VOCALIZATIONS: A SENSITIVE INDICATOR OF TERATOGENIC EFFECTS

Suzanne C. Baker
Assistant Professor
Department of Psychology

James Madison University
Harrisonburg, VA 22801

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

July 1994

RAT PUP ULTRASONIC VOCALIZATIONS: A SENSITIVE INDICATOR OF TERATOGENIC EFFECTS

Suzanne C. Baker
Assistant Professor
Department of Psychology
James Madison University

Abstract

The ultrasonic vocalizations (UVs) normally emitted by rats in contexts which are assumed to be stress-inducing have been shown to be sensitive to the effects of various neuroactive substances. These vocalizations have been utilized by researchers as behavioral indicators of stress or emotionality, and they have provided a useful animal model of anxiety for the investigation of the effects of various anxiogenic and anxiolytic drugs. The research literature on ultrasonic vocalizations emitted by preweanling rats was reviewed in order to explore the potential usefulness of this behavior in testing teratogenic and toxicological effects of various substances using infant rats as subjects. Behavioral and methodological factors important in the use of these calls in research paradigms were identified.

RAT PUP ULTRASONIC VOCALIZATIONS: A SENSITIVE INDICATOR OF TERATOGENIC EFFECTS

Suzanne C. Baker

Rodents, particularly laboratory rats, are extremely useful subjects as animal models for the study of the effects of drugs and other substances on both physiology and behavior. The effects of various drugs, toxic agents, and teratologic agents on stress and anxiety states are of particular interest. Therefore, reliable, convenient, and efficient methods of assessing psychological states in rats (such as stress, emotionality, or anxiety) are very valuable. One behavioral pattern common to rats which has proved to be useful in this effort are the vocalizations typically emitted in contexts which are assumed to be stress-inducing. For example, vocalizations are emitted by adult rats in testing paradigms involving shock or acoustic startle stimulation (see Baker et al, 1991, for a review). While some vocalizations emitted by rats are audible, most of these vocalizations are "ultrasonic;" that is, they are above the range of human hearing.

The use of these ultrasonic vocalizations as an indicator of the emotional state of the animal has distinct advantages over other methods. Vocalization is a naturally-emitted behavior which occurs in response to naturally-occurring stressors in the rat's environment; the response thus has ethological validity. The vocalization response, because it is emitted naturally, does not require a training period (unlike, for example, avoidance conditioning). In addition, because the measurement of vocalization is a non-invasive procedure, it can be performed repeatedly on the same animal without interfering with the animal's ongoing behavior.

As part of a continuing program of research which examines the utility of rat ultrasonic vocalizations as indicators of the effects of various substances, several projects were undertaken during the Summer 1994 Research Program period.

1) <u>Rat pup ultrasonic vocalizations.</u> The existing research literature on the physical characteristics of the ultrasonic vocalizations emitted by rat pups, the natural contexts and laboratory testing conditions which elicit these calls, and the neuroactive substances which effect call emission was reviewed. This work was considered preparatory to the establishment of a testing program which utilizes rat pup ultrasonic vocalizations as an indicator of the effects of various drugs and teratologic agents.

2) <u>Adult rat ultrasonic vocalizations emitted under acoustic startle testing.</u>

Analyses of adult rat vocalizations emitted under the acoustic startle testing paradigm were continued. (This work is reported in a separate paper.) This work is important in determining detailed parameters of adult ultrasonic calls in order to examine the effects of various treatments on these parameters.

RAT PUP ULTRASONIC VOCALIZATIONS

A. Background

Immature rats (Rattus norvegicus) emit ultrasonic vocalizations under a variety of circumstances. These vocalizations are of interest as animal models of anxiety and distress because they are influenced by anxiolytic and anxiogenic drugs as well as other substances, and because they seem to be a sensitive measure of teratologic effects.

B. Contextual factors influencing vocalizations

In typical studies which utilize pup UVs as dependent measures, pups are removed from the nest and placed alone in a chamber, where calls are monitored and recorded or counted for a brief testing period (usually not more than 10 min). Calls which are emitted during this period are typically referrred to as "isolation" calls or, more rarely, "distress" calls. These are the calls most frequently examined in the literature.

Young rat pups are unable to regulate their own body temperature. When a pup is removed from the nest and isolated, as in the vocalization testing paradigm, it experiences a drop in ambient temperature, and researchers have noted the importance of temperature in eliciting these vocalizations. Rate of vocalization seems to be correlated with ambient temperature. For example, an early study (Allin & Banks, 1971) reported that pups tested at 35 deg C (which is within the pups' "thermoneutral" range) do not vocalize as much as pups tested at 2 or at 20 deg C. Testing is often done at "room temperature," approximately 22-24 deg C.

The effect of temperature on UVs depends on the age of the pups tested. For pups 1 week old or less, temperature seems to be the most important cue in eliciting UVs (although very young pups [newborns or day-old animals] vocalize very little under any circumstances). Most authors report that, for pups 2 weeks old or older, cues other than temperature come into play and influence the rate of vocalization. These cues include the presence of conspecifics (Hofer & Shair, 1978, 1980), handling or

tactile cues (Gardner, 1985; Okon, 1972; Elsner et al, 1990), olfactory cues (Conely & Bell, 1978; Oswalt & Meier, 1975; Lyons & Banks, 1982), nutritional cues (Blass & Fitzgerald, 1988; Shide & Blass, 1989), and other contextual factors (e.g., learning, Amsel et al, 1977).

Pups in the first week of life seem to respond primarily to temperature, while other cues become more effective in eliciting UVs during the second week. However, cues related to the presence of conspecifics may also have effects on vocalizations in younger pups as well (see Carden & Hofer, 1992 for 3-day-old pups).

It is important to note that the effectiveness of some of these cues (e.g., learning, handling, olfactory cues) in eliciting UVs has not been systematically investigated in much recent literature.

C. Functional ethological significance of pup UVs

Neonatal rat pups are dependent on the warmth provided by the mother and the nest environment. It has been suggested that vocalizing when separated from the mother or littermates, or when the ambient temperature drops, would be adaptive if the mother responded by retrieving the pup and placing it back in the nest. This appears to be the case (see Allin and Banks, 1972; Smotherman et al, 1974, 1978; Bell et al, 1974). It has been suggested that the function of UVs which are elicited by tactile stimulation is to cause the mother to break off contact with the pup in order to avoid damaging the pup (see Sales & Pye, 1974) or "to inhibit the aggression of the retrieving adult" (Sewell, 1968, p. 682). Hypotheses concerning the role of handling-induced UVs apparently have not been systematically examined in the literature.

In addition to being unable to thermoregulate on their own, neonatal pups are also unable to eliminate and rely on anogenital licking from the mother to accomplish this. Pup UVs apparently play a role in regulating the female's licking behavior (Brouette-Lahlou et al, 1992). There is also some evidence which suggests that exposure to rat pup UVs may result in a rise in a lactating female rat's prolactin levels (Terkel et al, 1979; but see also Voloschin & Tramezzani, 1984).

Taken cumulatively, these studies demonstrate that the vocalizations of the pups can significantly influence the mother's behavior toward them; however, in many cases, the details of how these calls modulate mother-pup interactions (e.g., do the physical parameters of calls emitted in these different contexts differ?) have not been elucidated.

D. Physical characteristics of rat pup ultrasonic vocaliztions

Few studies which utilize rat pup UVs have provided details on the physical characteristics of the calls, other than rate. The vast majority of these studies do not report detailed quantitative information on frequency parameters or duration of a large sample of calls; only generalized descriptive information is given and sometimes sonagrams of a single call or a small sample of calls are provided.

Noirot (1968) reported two types of calls. "Clicks" were very brief sounds. "Whistles" were sounds longer than 5 msec. These were usually between 40-50 kHz, and were of almost constant frequency. Okon (1972) reported changes with age in the spectral characteristics of rat pup UVs. UVs emitted by 1-5 day old pups were between 60-140 msec in duration. Frequency of the calls was generally between 45-65 kHz, and the calls had slow downward frequency drifts. By days 5-15 of life, frequency was generally between 40-50 kHz. The pulses became more variable, often involving "rapid frequency drifts, warble-like, step-like, and chirrup-like patterns." (More detailed data on these frequency patterns was not provided.) Sales & Pye (1974) reported that UVs of newborn rats were 4-65 msec in duration (up to 200+ msec); frequency was generally 40-75 kHz, but went as high as 112 kHz. For older pups, duration of the calls is reported as 5-65 msec (up to 150 msec), with frequency generally at 40-90 kHz (up to 100 kHz). The frequency pattern is described as a "single component," but no further information is given.

It is not clear whether these early studies were able to discriminate between actual vocalizations and sounds which could have been caused by the pup making contact with cage surfaces. The very brief duration signals reported (e.g., 5 msec) may be due to such incidental contact.

Naito & Tonoue (1987) also noted a change in the predominant frequencies of UVs during the first days of life. For pups 4-5 days old, most sound was in the range of 50 kHz. For 6-7 day-old pups, energy in the 40 kHz region predominated. For pups 10 days or older, most sound was in the 30 kHz region. Takahashi (1992a; also Takahashi, Baker, & Kalin, 1990) reports using a bat detector tuned to 40-50 kHz for detecting calls of 7- and 14-day old pups, but tuning the detector to 30-40 kHz to detect the UVs of 21-day-old pups. This indicates a developmental change in predominant frequency.

Naito & Tonoue (1987) also report sex differences in the frequency modulation patterns of the calls. They report that, in general, "male calls" are longer duration, and also have a longer duration central component which maintains a constant frequency. The frequency patterns of "female type" calls look more like an inverted V with steep

upsweeps and downsweeps. The frequency modulation patterns reported in this study appear to be more complex than those reported in other published studies, with calls showing a great deal of frequency modulation (steep upsweeps and downsweeps) over a very brief time period.

The results of these studies are somewhat inconsistent with one another. In general, however, the duration of these calls seems to be between 5 and 200 msec, with the most typical calls reported being around 50-100 msec. Frequency seems to be between 30-50 kHz for most calls, but may range up to over 100 kHz. There may be changes as the pup matures, with UVs becoming lower in frequency. Sex differences also may exist, although this has not been systematically examined in most research. The question of frequency modulation patterns has yet to be resolved, with some authors reporting very little modulation across the duration of the call, but others reporting a great deal (e.g., Naito & Tonoue, 1987; Okon, 1972).

Pups tested in the typical isolation paradigm typically emit calls at a high rate. Insel & Winslow (1991) report that, "In most laboratories, rat pups from 6-12 days of age will regularly emit at least 50 calls/min" (p. 19). Rates up to more than 150 calls per minute are reported in some studies.

As was noted above, it has been suggested that calls in response to handling may differ in their physical parameters from "isolation-induced" calls (e.g., Sales & Pye, 1974; Okon, 1971). In particular, early studies indicated that handling-induced UVs might be of greater intensity than isolation vocalizations (see Sewell, 1969; discussed in Sales & Pye, 1974). This early hypothesis has not been followed up in recent literature. In fact, there are apparently no systematic studies of the physical parameters of rat pup calls elicited in different contexts or by different stimuli. It is still not known whether the UVs emitted by pups when isolated or cooled differ in any of their physical parameters from UVs emitted during tactile stimulation.

E. Measurement/detection methods typically used

In most studies in which rat pup UV is a dependent variable, the vocalizations are detected using an ultrasonic detector ("bat detector") which heterodynes the ultrasonic signals into sounds within the range of human hearing. The ultrasound detector typically is tuned to 40-50 kHz (e.g., 42 kHz, Carden et al, 1993; 47+/-5 kHz, Hennessey et al, 1978; 40-44 kHz, Hofer & Shair, 1992; 40-45 kHz, Takahashi, 1992b). (It should be noted that, when used in this manner, the ultrasonic detector will not detect any signals occurring outside of the chosen frequency range; any others will be

missed.) The number of calls are then counted, either by simple listening or by using some type of automatic counting device. The most commonly used dependent measure is vocalization rate or number of UVs occurring within a specified time period.

Typically, testing lasts no more than 10 min. The likely reason for this is that the rate of isolation- or cold-induced UVs typically begins to drop off after a few minutes (e.g., Takahashi, 1992a; although a few studies have found UVs continuing for up to 30 min in isolated pups).

More detailed studies which involve recording and sonagraphic analyses of the calls are reported by Okon (1971), Sales & Pye (1974), and Naito and Tonoue (1987). This type of detailed analyses of these calls is rarely done, and there are few research reports which utilize measures of call duration, intensity, or spectral characteristics. Any and all of these parameters could be influenced by drugs or other agents, and any of them potentially could carry communicative information about the rat pup's current situation or anxiety state. Given that these calls have been shown to influence the behavior of the dam, including tactile contact, orientation, and anogenital licking (see above), it seems likely these largely unexamined parameters of the calls might carry such information.

F. Development of vocalizations

An early study by Noirot (1968) described the development of ultrasonic vocalization responses in infant rats. UVs were elicited by placing pups in isolation at cool temperatures. Few calls were emitted during the first 3 days of life. Number of calls rose to a maximum between 5 and 10 days of age, and calls began to decline rapidly at about 15-16 days of age. Very few calls (essentially none) were emitted after 20 days. This developmental pattern has since been reported by numerous researchers (e.g., Okon, 1971; Sales & Pye, 1974; Naito & Tonoue, 1987; Takahashi, 1992a) for isolation- or temperature-induced UVs, and also for UVs induced by handling (Okon, 1972). Several studies (Naito & Tonoue, 1987; Takahasi, 1992a; Takahashi, Baker, & Kalin, 1990) also report a change in frequency across age in isolation-induced UVs, with frequency becoming lower in older pups (see above).

Physical changes occur during this period of the pup's life may be correlated with changes in rate of UV emission. Noirot (1968) reported that the early increase in UV rate seemed to occur when the pups' ears unfolded, on about day 3. UV rate begins to decrease once the ability to thermoregulate begins to develop, at about 10 days of age. Rat pups are weaned at 21 days, at which time UVs in respose to isolation

have virtually completely disappeared (Okon, 1971).

As was noted above, there may be age-related changes in call frequency, there are also important age-related changes in stimuli which elicit UVs (see above). For pups younger than about 10 days, ambient temperature seems to be the most important stimulus. However, during the 2nd week of life, other factors (olfactory stimuli, etc) become influential, and for 14- or 15-day old pups, only extreme cold (2 deg C) affects vocalizations.

G. Applications

It has been recognized within the past 5-10 years that the ultrasonic vocalizations emitted by isolated, cold-exposed rat pups might provide a useful behavioral indicator of anxiety. Such an animal model, which utilizes a behavioral response which is ethologically relevant (i.e., naturally emitted by the animal under apparently stressful circumstances), and which has parameters which can be relatively easily quantified, has enormous utility in the study of many kinds of drug and treatment effects (see Insel & Winslow, 1991).

It has now been demonstrated that the rate of rat pup UV in the isolation-testing paradigm is sensitive to anxiolytic and anxiogenic drugs, indicating the potential usefulness of this behavior as an animal model for anxiety. For example, the calls are attenutated by both benzodiazepine anxiolytics and nonbenzodiazepine anxiolytics, such as the $5HT_{1A}$ agonist drugs buspirone and ipsapirone (e.g., Gardner, 1985; 1988; Insel et al, 1986; Klint & Andersson, 1994). Insel and Winslow (1991), in a review of this research literature, state that "the pharmacologic data demonstrate that the rat pup USV is exquisitely sensitive to manipulation of anxiolytic and anxiogenic drugs" (p. 23).

Additional substances which have been examined for their effects on these UVs include the serotonin reuptake inhibitors clomipramine, fluvoxamine, citalopram, paroxetine, and zimeldine, all of which selectively reduce the rate of UV (Mos & Olivier, 1989; Winslow & Insel, 1989); pentylenetetraxol (an anxiogenic drug which binds at the GABA-benzodiazepine receptor complex; Carden et al, 1993; Insel, Hill, & Mayor, 1983), and the kappa-opioid agonist U50,488, which increases the rate of UVs (Kehoe & Boylan, 1994; Carden, Barr, & Hofer, 1991; Carden et al, 1993).

Adams (1982) and Adams et al (1983) were among the first reports which examined the possibility that rat pup isolation- or cold-induced UVs might be a sensitive indicator of teratologic effects. Within the last 10 years, the effects of

prenatal exposure to haloperidol (Cagiano, Barfield, White, et al, 1988), diazepam (Cagiano, DeSalvia, Persichella, et al, 1990), methyl mercury (Cagiano, DeSalvia, Renna, et al, 1990; Cagiano, Cortese, DeSalvia et al, 1988; Elsner et al, 1988, 1990), flumazenil (Cagiano, DeSalvia, Giustino, et al, 1993), and ethanol (Kehoe & Shoemaker, 1991) have been examined using UVs as a dependent measure.

Other factors whose teratologic effects have been examined include exposure to a diabetic intrauterine environment (Johansson et al, 1991), maternal adrenalectomy (Hennessey et al, 1978), and maintaining the mother on a low protein diet during gestation (Hunt et al, 1976; Hennessey et al, 1978). Isolation-induced UVs have been shown to sensitive to these manipulations as well.

H. Considerations relevant to utilizing rat pup UVs as dependent measures

1. Eliciting conditions for UVs. It should be noted that, in the typical "isolation" testing paradigm, the pup simultaneously experiences numerous events, any one or all of which might elicit UVs. Pups are separated from their mother and from littermates. The pups must be handled by the experimenter in some way to transfer them to the testing chamber, so tactile stimulation is occurring. The olfactory environment of the testing chamber differs from that of the nest. Even when the pups are tested in a heated chamber, some temporary drop in temperature is likely to be experienced when the pup is moved to the testing chamber.

Therefore, despite the fact that in many studies the UVs emitted during testing are referred to as "isolation calls" or "distress calls," it should be remembered that it is seldom explicit which of the numerous sensory stimuli and changes are, in fact, eliciting the calls, or if, as seems likely in most cases, these environmental cues operate in some additive or interactive way to affect calling. In some of the work of Hofer and colleagues (e.g., Hofer & Shair, 1978; 1980), as well as some earlier research (e.g., Oswalt & Meier, 1975) there have been attempts to control various factors and test cues in isolation from one another. Hofer & Shair (1980) found that tactile, temperature, and odor cues were all important in influencing UV emission rate in isolated 2-week-old pups. No one cue was necessary or sufficient for a significant reduction in UVs (with the exception of tactile cues, which had a small effect when tested alone).

In contrast, most of the research utilizing rat pup UVs in toxicological or teratological studies does not attempt to separate out various factors which may be eliciting or affecting UVs during testing. This becomes an important consideration if

the goal of the research is to examine treatment effects on various sensory or physiological systems. If there are treatment effects on UVs, it often is not clear whether any substances used are affecting the pup's response to olfactory cues, to thermal cues, to tactile cues from handling, or to separation from the mother and littermates per se, since all of these things are changed simultaneously, at least in the typical testing paradigm in which the pup is removed from the nest and tested in isolation for a brief period.

2. Variability in pup UVs. It has been noted by several researchers (e.g., Insel & Winslow, 1991; Carden & Hofer, 1992) that, while it is possible to describe general characteristics of rat pup UVs (e.g., developmental time course, eliciting factors) there is a great deal of variability in physical characteristics of the call across individuals. Such a high degree of variability can, of course, influence results. The following factors may influence variability in call rate or characteristics.

a. Sex. As reported above, Naito and Tonoue (1987) reported sex differences in both frequency pattern and duration of rat pup UVs. While sex differences are reported by some investigators (e.g., Elsner et al, 1990), others do not find sex differences (e.g., Graham & Letz, 1979; Adams, 1982; Adams et al, 1983). These differences may sometimes be present only in some call parameters; for example, Elsner et al (1990) found no differences in the peak frequency of male and female calls, but females called significantly less than did males.

Collectively, the results of these and other studies indicate that sex of the pups may be an important variable affecting some call parameters. This factor should therefore be controlled. Numerous studies test only male pups; this approach obviously affects generalizability of the results.

b. Strain. All of the studies reported here utilized laboratory strains of Rattus norvegicus as subjects. It is important to note, however, that strain may be one source of variability in the calls, and this should be taken into account when interpreting results or when comparing results among studies which utilize different strains of rats. Most published studies have used Wistar rats as subjects, although a few have used Sprague-Dawley, Long-Evans (Cagiano, DeSalvia, Persichella, et al, 1990) or other strains.

Strain differences in UVs have rarely been systematically examined. Insel and Hill (1987) reported that 5-day-old pups of the Maudsley-reactive strain emitted approximately 5 times the number of UVs as Maudsley-nonreactive pups in an isolation testing paradigm. These two strains have been selected for extreme responses along a

continuum of "emotionality;" however, it is possible that differences of this type exist among other strains as well.

c. Litter. Another potential source of variability in rat pup UV data is inter-litter differences (see, e.g., Winslow and Insel, 1991). Graham & Letz (1979) performed what apparently is the only study to systmatically examine interlitter variability in rat pups UVs. They found a significant difference in UV rate due to litter, regardless of the rearing conditions of the pups. Although rate of vocalization was the only parameter of vocalization examined in this study, there may be many other significant interlitter differences as well, for example, in intensity patterns, spectral characteristics, or duration (Adams et al, 1983).

This factor can be controlled by testing pups from each litter under different treatment conditions. Cross-fostering of pups is another technique which may be useful.

3. Parameters of UVs to be measured. Another important consideration in interpreting the published research on both the communicative significance of rat pup UVs and the utility of these calls in the examination of anxiogenic and anxiolytic substances is the physical parameters of the calls which are utilized in these studies. As was noted above, with a few exceptions, the only physical parameter of the calls examined in most of these studies is number of calls emitted during a brief testing period, or rate of calling. Other characteristics of the calls, such as duration, intensity, and frequency, are rarely examined.

It seems likely that a primary reason most studies utilize rate of UV emission as the sole dependent variable is ease of measurement. However, there is no apparent reason (either an intuitive reason or a reason based on an empirical understanding of the ethological significance or clinical implications of rat pup UVs) why rate would be more likely than other measurements (e.g., intensity, duration) to reflect the emotional state of the animal, or why rate would be a more useful parameter for detection of anxiogenic, anxiolytic, or teratologic effects of various substances.

Therefore, for any study in which no effects on UVs are found, it is important to examine what call parameters were measured, and to question whether significant differences in other parameters (such as duration, intensity, or spectral characteristics) might be present even if there were no differences in rate, for example.

I. Factors in testing/recording

Stated very briefly, in a study which utilizes rat pup UVs as an indicator of teratologic effects, the following factors should be carefully considered:

Factors having to do with the testing situation or apparatus:

1. **Temperature at which animals are tested, including ambient temperature and temperature properties of any other objects or surfaces in the testing apparatus**
2. **Presence of conspecifics in the testing situation, including the mother, littermates, or any other "companions"**
3. **Handling - type of handling or manipulation used, e.g., to place the animals in the testing apparatus**
4. **Age of animals tested. There are clear age-related changes in rate of UV, as well as age-related changes in conditions which elicit UVs. There is also evidence that spectral characteristics of the calls may change with age.**
5. **Strain of rats used**
6. **Sex of subjects**
7. **Litter (genetic background) of subjects**
8. **Olfactory cues present in bedding or other material used in the testing apparatus**
9. **Tactile cues - cues having to do with the surface on which the subjects are placed for testing or whether they are allowed contact with an object during testing.**
10. **Duration of testing period**

Factors having to do with recording:

11. **Physical parameters of UVs - which physical parameters will be recorded? Possibilities include:**

    rate

    duration of individual calls

    percent of test time spent vocalizing

    intensity of calls

    frequency measures, including modulation patterns

Factors having to do with the substance being tested:

12. **Dosage of substance to be administered**

13. **Period during gestation at which substance is administered**

14. **Duration of exposure or number of exposures to substance during gestation**

REFERENCES

Adams, J. Ultrasonic vocalizations as diagnostic tools in studies of developmental toxicity: An investigation of the effects of hypervitaminosis A. Neurobehavioral Toxicology and Teratology, 4, 299-304 (1982).

Adams, J., Miller, D.R., & Nelson, C.J. Ultrasonic vocalizations as diagnostic tools in studies of developmental toxicity: An investigation of the effects of prenatal treatment with methylmercuric chloride. Neurobehavioral Toxicology and Teratology, 5, 29-34 (1983).

Allin, J.T., & Banks, E.M. Effects of temperature on ultrasound production by infant albino rats. Animal Behaviour, 20, 175-185 (1971).

Allin, J.T., & Banks, E.M. Functional aspects of ultrasound production by infant albino rats (Rattus norvegicus). Animal Behaviour, 20, 175-185 (1972).

Amsel, A., Radek, C.C., Graham, M., & Letz, R. Ultrasound emission in infant rats as an indicant of arousal during appetitive learning and extinction. Science, 197, 786-788 (1977).

Baker, S.C., Mulligan, B.E., & Murphy, M.R. Ultrasonic vocalizations by adult rats (Rattus norvegicus). USAF Technical Report AL-TR-1991-0135 (1991).

Bell, R.W., Nitschke, W., Bell, N., & Zachman, T. Early experience, ultrasonic vocalizations, and maternal responsiveness in rats. Developmental Psychobiology, 7, 235-242 (1974).

Blass, E.M., & Fitzgerald, E. Milk-induced analgesia and comforting in 10-day-old rats: Opioid mediation. Pharmacology Biochemistry & Behavior, 29, 9-13 (1988).

Brouette-Lahlou, I., Vernet-Maury, E., & Vigouroux, M. Role of pups' ultrasonic calls in a particular maternal behavior in Wistar rat: Pups' anogenital licking. Behavioural Brain Research, 50, 147-154 (1992).

Cagiano, R., Barfield, R.J., White, N.R., Pleim, E.T., Weinstein, M., and Cuomo, V. Subtle behavioural changes produced in rat pups by in utero exposure to haloperidol. European Journal of Pharmacology, 157, 45-50 (1988).

Cagiano, R., DeSalvia, M.A., Persichella, M., Renna, G., Tattoli, M., & Cuomo, V. Behavioural changes in the offspring of rats exposed to diazepam during gestation. European Journal of Pharmacology, 177, 67-74 (1990).

Cagiano, R., DeSalvia, M.A., Renna, G., Tortella, E., Braghiroli, D., Parenti, C., Zanoli, P., Baraldi, M., Annau, Z., & Cuomo, V. Evidence that exposure to methyl mercury during gestation induces behavioral and neurochemical changes in offspring of rats. Neurotoxicology and Teratology, 12, 23-28 (1990)

Cagiano, R., Cortese, I., DeSalvia, M.A., Renna, G., & Cuomo, V. Effects of prenatal exposure to methyl mercury on ultrasonic calling in rat pups. Pharmacological Research Communications, 20, 215-216 (1988).

Cagiano, R., DeSalvia, M.A., Giustino, A., LaComba, C., and Cuomo, V. Behavioral changes produced in rats by developmental exposure to flumazenil, a benzodiazepine receptor antagonist. Prog. Neuro-Psychopharmacol. & Biol.Psychiat., 17, 151-159 (1993).

Carden, S.E., Barr, G.A., & Hofer, M.A. Differential effects of specific opioid receptor agonists on rat pup isolation calls. Developmental Brain Research, 62, 17-22 (1991).

Carden, S.E., Bortot, A.T., & Hofer, M.A. Ultrasonic vocalizations are elicited from rat pups in the home cage by pentylenetetrazol and U50,488, but not naltrexone. Behavioral Neuroscience, 107, 851-859 (1993).

Carden, S.E., & Hofer, M.A. Effect of a social companion on the ultrasonic vocalizations and contact responses of 3-day-old rat pups. Behavioral Neuroscience, 106, 421-426 (1992).

Conely, L., & Bell, R.W. Neonatal ultrasounds elicited by odor cues. Developmental Psychobiology, 11, 193-197 (1978).

Elsner, J., Hodel, B., Suter, K.E., Oelke, D., Ulbrich, B., Schreiner, G., Cuomo, V., Cagiano, R., Rosengren, L.E., Karlsson, J.E., & Haglid, K.G. Detection limits of different apporaches in behavioral teratology, and correlation of effects with neurochemical parameters. Neurotoxicology and Teratology, 10, 155-167 (1988).

Elsner, J., Suter, D., & Alder, S. Microanalysis of ultrasound vocalizations of young rats: Assessment of the behavioral teratogenicity of methylmercury. Neurotoxicology and Teratology, 12, 7-14 (1990).

Gardner, C.R. Distress vocalizations in rat pups: A simple screening method for anxiolytic drugs. Journal of Pharmacological Methods, 14, 181-187 (1985).

Gardner, C.R. Potential use of drugs modulating 5-HT activity in the treatment of anxiety. Gen. Pharmacol., 19, 347-356 (1988).

Graham, M., & Letz, R. Within-species variation in the development of ultrasonic signaling of preweanling rats. Developmental Psychobiology, 12, 129-136 (1979).

Hennessy, M.B., Smotherman, W.P., Kolp, L.A., Hunt, L.E., & Levine, S. Stimuli from pups of adrenalectomized and malnourished female rats. Physiology &

Behavior, 20, 509-513 (1978).

Hofer, M.A., & Shair, H. Ultrasonic vocalization during social interaction and isolation in 2-week-old rats. Developmental Psychobiology, 11, 495-504 (1978).

Hofer, M.A., & Shair, H. Sensory processes in the control of isolation-induced ultrasonic vocalization by 2 week old rats. J. Comp. Physiol. Psychol., 94, 271-279 (1980).

Hofer, M.A., & Shair, H.N. Ultrasonic vocalization by rat pups during recovery from deep hypothermia. Developmental Psychobiology, 25, 511-528 (1992).

Hunt, L.E., Smotherman, W.P., Wiener, S.G., & Levine, S. Nutritional variables and their effect on the development of ultrasonic vocalizations in rat pups. Physiology & Behavior, 17, 1037-1039 (1976).

Insel, T.R., & Hill, J.L. Infant separation distress in genetically fearful rats. Biol. Psychiatry, 22, 786-789 (1987).

Insel, T.R., Hill, J.L., & Mayor, R.B. Rat pup isolation calls: Possible mediation by the benzodiazepine receptor complex. Pharmacology Biochemistry & Behavior, 24, 1263-1267 (1986).

Insel, T.R., & Winslow, J.T. Rat pup ultrasonic vocalizations: An ethologically relevant behaviour responsive to anxiolytics. In: Animal models in psychopharmacology, eds. B. Olivier, J. Mos, & J.L. Slangen. Basel: Birkhauser Verlag (1991).

Johansson, B., Meyerson, B., & Eriksson, U.J. Behavioral effects of an intrauterine or neonatal diabetic environment in the rat. Biol. Neonate, 59, 226-235 (1991).

Kehoe, P., & Boylan, C.B. Behavioral effects of kappa-opioid-receptor stimulation on neonatal rats. Behavioral Neuroscience, 108, 418-423 (1994).

Kehoe, P., & Shoemaker, W. Opioid-dependent behaviors in infant rats: Effects of prenatal exposure to ethanol. Pharmacology Biochemistry & Behavior, 39, 389-394 (1991).

Klint, T., & Andersson, G. Ultrasound vocalization is not related to corticosterone response in isolated rat pups. Pharmacology, Biochemistry, & Behavior, 47, 947- (1994).

Lyons, D.M., & Banks, E.M. Ultrasounds in neonatal rats: Novel, predator and conspecific odor cues. Developmental Psychobiology, 15, 455-460 (1982).

Mos, J., & Olivier, B. Ultrasonic vocalizations by rat pups as an animal model for anxiolytic activity: Effects of serotonergic drugs. In Bevan, P., Cools, A.R., Archer, T. (eds.), Behavioural pharmacology of 5-HT, pp. 361-366. Hillsdale, NJ: Lawrence Erlbaum Associates (1989).

Naito, H., & Tonoue, T. Sex difference in ultrasound distress call by rat pups. Behav. Brain Res., 25, 13-21 (1987).

Noirot, E. Ultrasounds in young rodents. II. Changes with age in albino rats. Animal Behaviour, 16, 129-134 (1968).

Okon, E.E. The temperature relations of vocalization in infant golden hamsters and Wistar rats. J.Zool., Lond., 164, 227-237 (1971).

Okon, E.E. Factors affecting ultrasound production in infant rodents. Journal of Zoology, 168, 139-148 (1972).

Oswalt, G.L., & Meier, G.W. Olfactory, thermal, and tactual influences on infantile ultrasonic vocalization in rats. Developmental Psychobiology, 8, 129-135 (1975).

Sales, G.D., & Pye, D. Ultrasonic communication by animals. London: Chapman & Hall (1974).

Sewell, G.D. Ultrasound in rodents. Nature, 217, 682-683 (1968).

Sewell, G.D. Ultrasonic signals from rodents. Ultrasonics, 8, 26-30 (1970).

Shide, D.J., & Blass, E.M. Opioidlike effects of intraoral infusions of corn oil and polycose on stress reactions in 10-day-old rats. Behavioral Neuroscience, 103, 1168-1175 (1989).

Smotherman, W.P., Bell, R.W., Starzec, J., Elias, J., & Zachman, T.A. Maternal responses to infant vocalizations and olfactory cues in rats and mice. Behavioral Biology, 12, 55-66 (1974).

Smotherman, W.P., Bell, R.W., Hershberger, W.A., & Coover, G.D. Orientation to rat pup cues: Effects of maternal experiential history. Animal Behaviour, 26, 265-273 (1978).

Takahashi, L.K. Developmental expression of defensive responses during exposure to conspecific adults in preweanling rats (Rattus norvegicus). Journal of Comparative Psychology, 106, 69-77 (1992a).

Takahashi, L.K. Ontogeny of behavioral inhibition induced by unfamiliar adult male conspecifics in preweanling rats. Physiology & Behavior, 52, 493-498 (1992b).

Takahashi, L.K., Baker, E.W., & Kalin, N.H. Ontogeny of behavioral and hormonal responses to stress in prenatally stressed male rat pups. Physiology & Behavior, 47, 357-364 (1990).

Terkel, J., Damassa, D., & Sawyer, C.H. Ultrasonic cries from infant rats stimulate prolactin release in lactating mothers. Homones and Behaviour, 12, 95-102 (1979).

Voloschin, L., & Tramezzani, J. Relationship of prolactin release in lactating rats to milk ejection, sleep state, and ultrasonic vocalization by the pups. Endocrinology, 114, 618-623 (1984).

White, N.R., Adox, R., Reddy, A., & Barfield, R.J. Regulation of rat maternal behavior by broadband pup vocalizations. Behavioral and Neural Biology, 58, 131-137 (1992).

Winslow, J.T., & Insel, T.R. Endogenous opioids: Do they modulate the rat pup's response to social isolation? Behavioral Neuroscience, 105, 253-263 (1991a).

Winslow, J.T., & Insel, T.R. Infant rat separation as a sensitive test for novel anxiolytics. Prog. Neuro-Psychopharmacol. & Biol. Psychiat., 15, 745-757 (1991b).

Winslow, J.T., & Insel, T.R. Serotonergic modulation of the rat pup ultrasonic isolation call: Studies with $5HT_1$ and $5HT_2$ subtype-selective agonists and antagonists. Psychopharmacology, 105, 513-520 (1991c).

KNOWLEDGE-BASED GROUPWARE FOR GEOGRAPHICALLY DISTRIBUTED COLLABORATIVE COMPUTING ENVIRONMENTS

Dr. Alexander B. Bordetsky
Department of Decision Sciences
School of Management

University of Texas at Dallas
P.O. Box 830688
Richardson, TX 75083-0688

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

KNOWLEDGE-BASED GROUPWARE FOR GEOGRAPHICALLY DISTRIBUTED COLLABORATIVE COMPUTING ENVIRONMENTS

Alexander Bordetsky
Decision Sciences
School of Management
University of Texas at Dallas

ABSTRACT

Feasibility analysis of the knowledge-based group decision support solutions based on groupware and collaborative computing technologies is the major subject of this study. Research on extending of current Group Research Laboratory for Logistics (GRLL) facilities in computer support of problem-solving groups into geographically distributed problem-solving environments is required for different group support projects at the Armstrong Lab Analysis of initial business requirements is based on the GRLL experience with different problem-solving sessions at the electronic meeting room. It is complemented by the results of information requirements analysis for the Quality Air Force Program at the Aeronautical Systems Center. Different business process engineering, process assessment, and quality management activities are considered subject to face-to-face, distributed asynchronous, and distributed synchronous forms of collaboration. Technology of collaborative computing is essentially capable to provide required distributed extension of an electronic meeting room environment by means of peer -to-peer and multiperson wide-area multimedia networking. It implies some specific communication constraints to be satisfied, but feasible solutions are already on the market, and their choice is a matter of test experiments. Another critical issue is an architecture of group decision support tools, subject to the changes in team communication. The proposed structure of application layer agents is based on evaluation of coordination support agents, learning features, and the specifics of distributed knowledge base management and models integration. Case-based reasoning technology is used to support the group coordinator to monitor consensus making on distributed network, as well as to incorporate an individual knowledge into the group memory representation. Results also include the sample of agent-facilitator for hypermedia-based asynchronous collaboration, and the plan of experiments with desktop videoconferencing environment.

1. INTRODUCTION

Historically the research and implementation of group support systems started at the Armstrong Laboratory on the basis of Group Research Laboratory for Logistics (Heminger et al, 94). Its physical setup represents an electronic meeting room, controlled by Group System V product (Nunumaker et al, 91) on Novell LAN network. Electronic meeting room is very effective for categorizing, idea generation , brainstorming, and group ranking, but is limited for the purpose of geographically distributed collaboration, essential to the Air Force problem-solving teams. It supports primarily the first of four well known forms of group communication (Jessup and Valacich, 1993):

A. Face-to-face synchronous communication, that occurs at the same time and place.

B. Distributed synchronous communication, that occurs at the same time at different places.

C. Asynchronous communication, that occurs at different time, but at the same place.

D. Distributed asynchronous communication, that occurs at different time at different places.

New technologies of videoconferencing, collaborative computing, and hypermedia Internet-based networking essentially extend an electronic meeting room environment by means of peer-to peer and multiperson asynchronous/synchronous communication via wide-area multimedia networking. In particular, collaborative computing (Hsu and Lockwood, 1993) makes immediately possible to incorporate time consuming asynchronous interactions with data bases, knowledge bases and decision support models (types B, C, and D) into the process of distributed group decision support.

2. GENERAL SPECIFICS OF DISTRIBUTED GROUP SUPPORT

From problem solving perspective distributed group decison support differes from face-to-face group decsision support environment by

- implementing mixed ussage pattern, that may be some times group-based and some times individual,
- in addition to brainstorming on structuring intially unstructured problem, it may as well incorporate some semi-structred model analysis, computer simulations and decison support calculations.

Therefore, incorporation of real-time synchronous and asynchronous communication types is critical factor for distributed group decision support.

There are two important features of distributed group decision support environment: shared data model and interpersonal communication space. In the electronic meeting room, supported by Group System V for example, (Nunumaker et al, 1991) data model is a meeting agenda, and text sharing features are fully supported at the degree required for brainstorming. Interpersonal communication space is given by the room environment: facilitor and group members communicate to each other on the basic of common "protocol" of face-to-face communication. Unlike the face-to-face environment, in geographically distributed case full scale interpersonal communication is not present It may only be reproduced at a certain degree.

Traditionally group support systems are implemented in the form of meeting room as oppose to desktop-based networking pattern typical for collaborative computing. Theoretically both cases are applIicable to support of geographically distributed problem solving teams, that is room-to-room or peer-to-peer. Room-to-room means that we have at least two remote parties, each equipped by electronic room

like PictureTel, collaborating through switched or satellite media. This form is very effective for short-term brainstorming, but doesn't actually allow to incorporate computerized data analysis into group meeting. Multiperson peer-to-peer communication means that we have multiple users of multimedia computers communicationg within wide-area switched networks.

Collaborative technology, that returns to distributed group of decison makers (DM) the following set of services:

- compressed full motion video,
- shared electronic workspace (shared data models, shared screen, shared graphics and animation),
- high speed file transfer,
- internetworking via switch-based services,

is potentially capable to substitute the lack of interpersonal communication. But DMs coordination becomes the major problem for such computerized group support environments (Bush, Hamalainen, Holsapple, Suh, and Whinston, 91). Software support of coordination may be also shaped as a set of so-called software agents (Shaw and Fox, 93). Authors, reviewing coordination issues from the different angles, such as:

- Coordination models for "nemawashi" distributed management technique in Japan (Watabe, Holsapple, and Whinston, 92),
- Issues on integration of collaborative technology and decision analysis techniques (Bhargava, Krishnan, and Whinston, 94),
- Learning mechanisms for intelligent distributed decision support (Sykara, 93), and
- Distributed AI for group decison support (Shaw and Fox, 93),

unanemously emphasise an integrated solution for coordination mechanism, based on multiple criteria solvers and knowledge-based agents.

Knowledge-based agents may be required for different levels of coordination in the distributed group decision support environment, namely for

- Making the trade-offs on conflicting criteria,
- Brainstorming support,
- Providing learning mechanisms, group memory support.

Multiple criteria models vary from one review work to another (although AHP technique is rather popular), but on the part of knowledge representation most authors refer to some type of case-based reasoning technique as most appropriate.

Case-based reasoning representations allow parties to intelligently navigate through the shared knowledge base, (Bush, Hamalainen, Holsapple, Suh, and Whinston, 91), and keep leaning the results of group problem solving (Sycara, 93). Typical steps of case-based reasoning process include:

- Retrieving approprite precedent cases from case memory;
- Selecting the most appropriate case(s) from those retrieved;
- Constructing a solution;
- Evaluating the solution for applicability to the current case.

"Retrieving" and "selecting" steps may typically be done at the stage of categorizing and commenting on the problem , while "constructing" and "evaluating" essentially incorporate brainstorming scenario.

Group decison making is appropriate for the new or unique system analysis and design problems, when no two projects are alike, and each one is custom designed. In such case simple data base search is not applicable . Retreaval based on similarity and adaptation is required, and case-based reasoning agent may well support it. Correspondently, case-based reasoning technique presents a perspective basis for designing of knowledge-based agents , that provide distributed collaboration.

Unlike the artificial neural nets, another known mechanism for keeping learning, case-based reasoning doesn't put on user the limitations of numerical representation for data input. All it takes, is natural language description of cases. While designing the agent we have to make a decision on structure (frame) of case, and don't really need to enter each case : users themselves of will do it along the way of collaboration on the network. Initaial case-base is certainly desirable. On of the most poweful features of constraint-based reasoning learning approach is that such system is aware of it's limitations (Stottler, 94). If no similar cases are retrieved system cannot make any advice. As the current problem moves outside the case range, fewer less similar cases are retrieved, but the change is not abrupt.

Given a distributed environment, the facilitator isn't in the same room with the participants, and may not be immediately available to every member of a group, that is, its communication may be dramatically delayed. Therefore, the role of agents, providing facilitator's critical knowledge transfer through the network is becoming vitaly important.

In order to identify the possible roles for the knowledge-based agents in typical group sessions, and specify their minimal combination for physical prototyping in distributed environment, let us consider some basic business problems requirements.

3. GRLL REQUIREMENTS TO DISTRIBUTED GROUP SUPPORT PROCESSES

GRLL (Group Research Laboratory for Logistics) is an electronic meeting room, equipped by the shared projection screen, 15 IBM/PC client computers linked by Novell LAN, and Group System V server providing the data sharing tools and interaction support software for categorizing, topic commenting, brainstorming, voiting, alternative evaluation, and policy formation. Air Force groups of different levels DMs are coming to the GRLL as scheduled (size of groups varies from 5 to 15), and groups are scheduled on a first-come, first served basis. Typical sessison take 6-8 hours. Obserevations on different groups made while the Summer Program, as well as the interviews with System Manager and Facilitator, enabled to retrieve the following communication models for basic phases of electronic meeting, namely preplanning session, and face-to-face session.

3.1 OBSERVATIONS ON PREPLANNING COMMUNICATION

Preplanning is used to design the meeting process that is later executed while face-to-face meeting. Preplanning goes through the technical phase of creating the meeting agenda, and a discussion phase of discovering the desired outcomes for the meeting. Preplannig is initiated from the side of the team (team leader, individual). Facilitator and System Manager respond by sharing their expertise on agenda design and the capabilities of the system. This step is rarely processed through voice conferencing, and more often is conducted as a short session in the same GRLL environment. Observations indicate , that at this stage facilitator would like two have two types of tools for computer-aided collaboration:

- interactive distributed communication with customer,
- knowledge-based agent, that may help customer to validate the applicability of his problem to the group environment. (Case-based reasoning representation is rather suitable for validation.)

For support of "discussion" phase of preplanning facilitator would like to transfer to remote DM his set of rules, containing general recommendations on how to implement the steps of the agenda : "what do you need to do now, what are you going to do next?".

3.2 OBSERVATIONS ON FACE-TO-FACE MEETING

Face-to-face meeting develops through the following stages:

1.Facilitator briefly introduces the features : number of questions, issues to address, ability to share the comments, an how the group may come up to consensus (10-15 min.).

2. Group tries to adjust the environment by starting verbally to discuss some critical items, at that time everebody is trying to make comments verbally, relying on System Manager quick input of the comments. At the background of verbal conversation (very intensive and fast) facilitator takes an initiative to move most preferable comment to the top of the list. This is the first look at the problem. Group organise the ideas. This could take from 15 minute to an hour , if it occurs.

3. On the basis of the group discussion, the participants start their individual work on commenting the agenda items. They do not communicate verbally much, but rather utilise shared computer communication between them. This is a brainstorming part of a session. This usually takes any where from twenty or thirty minutes to anhour or more.

4. The participants work with facilitator to understand and consolidate the items in brainstorming lists.

5. Participants vote on the items in the consolidated list. This is essentialy anonymous process of client-server interactions with Group Systems V tool. This may take from 5 minutes to a half an hour or more, depending on the length of the list and other factors.

This is the basic dynamic process of face-to-face meeting in the electronic meeting room. Sequence of steps may change subject to agenda, but the steps will be similar in many cases. From perspective of distributed communication steps 3 and 4 may be provided almost immediately, if shared workspace, and fast file transfer are supported by copmuter network . Interactivity is reduced, and minor delays, essential for wide-bandwidth Internet for example , may be accepatable. But steps 1, 2, and 4 are heavily based on voice and visual interpersonal communication, that can't be elimenated, and it's uncertain how much it can be reduced.

Therefore suitable distributed environment must incorporate the features of asynchronous information processing with fragments of real-time videoconferncing, that group coordinator and participants may flexibly exchange, following the meeting dynamics.

Personal comments of DMs provide some additional information on what the specific requirements to distributed group support system may be, and why the distributed environment is necessary. Here are some samples, provided by Aeronatical Systems Center 2-letter commanders, while structuring and brainstorming on the Key Processes for the Quailty Air Force Program:

-Put the model (map) of CCT structure for Key, Sustaining, and Enabling processes on the board.
-Let us compare the models.
-It'll be good to take a look to the list of processes we identified at previous meeting.
<At that moment I quickly developed a graphical representation of processes , and offered the idea on how to formally differentiate Enabling and Sustaining processes as providers:

- enabling process X: SP_i ====>SP_{i+1}, provides the transition from one subprocess SP_i to another, it's elimination may interrupt the whole Key Procees,

- sustaining process Y : SP_i ,simply identifies the subprocess, it's elimination doesn't lead to Key Process interrupt, process continues by passing the missed one.

Results of the modelling were immediately adopted by group, and brainstorming continued smoothly>.
-We need the presence of CCT Champion (upper level commander) for identifying this Key Process
-and so on...

For the next step of voting System Manager provided the following comment:
- In distributed setup I'd like to have a tool (knowledge-based essentially), that will help us to classify, what type of ordering is emerging for the current group, in order to apply an appropriate voting tool.

Based on above listed observations the following conclusions may be made:

1. Distributed communication is desirable improvement for both: preplanning and face-to-face meeting processes.

2. At both stages: preplanning and face-to-face meeting there is a demand to access the knowledge on similar cases. At preplanning it is important for validation of initial problem applicability to the group environment , and during the meeting it would be valuable for categorizing and brainstorming support.

3. Access to modelling and simulation tools, ability to integrate them into idea organizing and brainstorming processes is highly desirable feature. Such option isn't normally available at the electronic meeting room. Distributed environment makes it feasible to incorporate time consuming asynchronous interactions with models, data bases, and knowledge bases into the distributed group decision support process.

4. There are some preplanning and face-to-face meeting activities, that will immediately benefit from the rule-based type of knowledge representation . Namely, for distributed management of preplanning meeting, rule-base agent , transfering facilitator rules on structuring the agenda, is very important. Also, for distributed management of voting, rule-base, that advises the choice of voting tool for on-going meeting may be very valuable.

5. Suitable distributed environment must incorporate the features of asynchronous information processing with real-time video conferencing fragments available for scheduled and sponteneous interactions.

4. QUALITY AIR FORCE PROGRAM BUSINESS REQUIREMENTS TO DISTRIBUTED GROUP DECISION SUPPORT SYSTEM

Design and implementation of the Quality Air Force (QAF) Program at Aeronautical Systems Center (ASC) is continuing 2 year teamwork process. Currently ASC is working with two models for business process engineering: 16-Step "Blueprint" and 12-Step process from Texas Instruments.

16-Step model is conventional representation of the following processes (QAF, 94).
12-Step model is hierarchical representation, based on the life-cycle concept. The two models are compared below (Tab.1):

Tab.1

Step 1: Identify the Customer and its categories	Define/Review Core values
Step 2: Identify the Customer Life Cycle	Define /Review Missions
Step 3: Identify the Customer Care Abouts	Define/Review Key Business Factors
Step 4: Establish the Processes: Core, Sustaining, and Enabling	Revisit Mission Statement

Step 5: Process decomposition	Pyramid to Appropriate Levels
Step 6: Input/Output resolution: boundaries	Each level Do Steps 2&3
Step 7: Identify Process Measurements and Metrics	If "Conflict=yes" goto Step 2, else continue
Step 8: Assign Process Champions	Each level determine Key Processes & Goals
Step 9: Identify next steps and map validation approach	All levels link Mission , Key Business Factors, and Goals
Step 10: Identify and Prioritize Reengineering opportunities	Implement Improvement Cycle for Key Processes
Step 11: Identify field teams and "As is " models	Determine depth of QAF Assessment
Step 12: Perform Gap analysis and change Strategy	Establish/Maintain Baseline via QAF Assessment
Step 13:	Examine Score and document Areas for Improvement
Step 14:	Develop Improvement Plans
Step 15:	Incorporate into Strategic Plans
Step 16:	Implement Improvement Plans

Observations made on ASC initial discussions of both models, as well as the interviews with ASC Champions have been used to identify the specific requirements to distributed group decision support of QAF process. Questions and observations have been organized according to the following sample representation in which each process is considered as a case for distributed group support environment design.

4.1 GROUPWARE CASE-FRAME OF THE QAF PROCESS

Objective: to describe the processes of 16-Step and 12-Step models as objects or **cases** for distributed group decision support system design.

IDENTIFICATION FEATURES

Name :
Function (name of one or more of 16 Steps it provides):

Hierarchical order:
- **-critical process,**
- **-sub process,**
- **-enabling process**

input:
Output:
Life time:

COLLABORATIVE FEATURES

1. **Types of collaboration** (% of overall procees life time with number of people involved in brackets)

	person-to-person	**multiperson meeting**
face-to-face		
same time different places		
different time		

different places

different time
same place

2. **Transition diagram** (Gantt diagram for the different types of collaboration):

GEOGRAPHIC FEATURES

	Ave. Distance	**Number of lots & Allocation**
Local area net		
Metropolitan area net		
Wide area net		

SHARED WORKSPACE FEATURES
(what type of data needs to be shared)

- **shared text (text only)**
- **shared hypertext**
- **shared data base**
- **shared knowledge base**
- **shared computational model**

GROUP MEMORY FEATURES

1. **Long-term memory** (for one or more 16 -step cycles)
   - **documents to be saved,**
   - **knowledge to transfer.**

2. **Short-term memory** (for one or more meetings)

   - **types of shared documents**
   - **knowledge to exchange**

COMMUNICATION FEATURES
(available or required media capabilities)

1. **Wired media:**
   - **cable**
   - **private lines**
   - **switch lines:**
     ***14.4 Modem**
     ***ISDN**
     ***Switch 56**

2. **Wireless media**

- **microwave**
- **satellite**

WORKSTATION FEATURES:
(user interface requirements)

- **desktop,**
- **portable,**
- **windows type interface**
- **multimedia interface**
- **desktop videoconferencing features**

4.2 OBSERVATIONS

The analysis of QAF processes still continues, however initial observations make it already possible to derive some conclusions.

1. Most of the processes in 16-Step model are permanent or periodic group efforts:
   - Step 1 takes place within 1-2 months in group of 9-12 DMs,
   - Steps 2 & 3 take place within several months in groups of 50-60 members,
   - steps 11, 12, &13 (self-assessment part) represent design processes in groups of 50-60 members.
2. Some of the tasks for 50-60 members at Steps 11, 12, &13 are essentially asynchronous, and some elements of those steps have been already on trial support through Gopher server on Internet. On the other hand non of the Steps can be executed without periodical face-to-face meetings.
3. Groups are experiencing real difficulties in getting together and structuring the output for the large meetings in the electronic meeting room.

Different comments, made by the interviewed commanders on the information processing problems they face, revealed another two problems:
- Integrated assessment problem. It is classification problem that CCT Champion is experiencing. The return of 2-letter units on self-assessment is 3500 patterns (50 members x 70 questions). On the basis of such input CCT Champion needs to produce integrated evaluation of self-assessment. Knowledge-based agent, that may help to classify the answers, according to the small number of integrated categories, is extremely important for that purpose.
- Assignment problem. One of the big problems is an overhead of parallel assignments for the same processes. CCT Champion would like to have a tool enabling to locate the processes with overhead of parallel executives , and simulate reengineering of such processes.

Observations made while ASC teamwork with 12-Step model revealed similar communication specifics:

1. Identification of Customer Life Cycle, Customer Care Abouts, and structure of Core, Sustaining, and Enabling processes is done by group of 9-12 DMs.
2. Decomposition of the same processes to the level of 2-letter units is done by groups of 50-60 members.
3. More modeling activity, based on data flow and hierarchical diagramming (Steps 7, 9) as well as matrices analysis (Step 7) is involved in group decision making process.
4. Tracking the assignments and solving assignment problem for the process champion is required procedure.

Some conclusions :

A. Knowledge-based support for the assignment problem-solving and incorporating of modeling tools into the groupware for distributed environment become more critical for the 12-Step process vs. 16-Step model;

B. Communication requirements to distributed processing are practically the same for both models.

Let's take a look to some comments made by the group on 12-Step model processing:

-We need some users here to help to structure the life cycle.

-We need the tool check the multiple links of Care Abouts with sub processes.

-We need to develop 60 processes, how we can describe so many Core, Sustaining, and Enabling processes?

-We have to have the data base of samples for Sustaining processes.

From comments we may derive the following:

C. Availability of spontaneous communication with representatives of another level (another team) is important for 12-Step process, which is similar to what was indicated earlier about general process in GRLL.

D. Matrices mapping the relationships between processes and tools for their structure analysis should be included in the decision support groupware.

E. Knowledge of case-samples for different processes should be essential part of decision support environment.

Of interest is the fact, that at the beginning group was struggling the naming problem, at the middle of the meeting DMs started to look for the samples of Core processes, at the end for all possible samples of Enabling processes. It means that repository of prototypes, ability to retrieve and evaluate different similarities on design processes, is important part of generating group knowledge and making the consensus. Case-based memory may perfectly fit such requirement.

Another observation: while decomposition to Sustaining and Enabling processes group was struggling prioritization problem. Implementation of multiple criteria technique based on the input of parewise comparisons only (like AHP for example) might be helpful .

In general we may conclude that 12-Step model requirements to the potential distributed decision support groupware are rather similar to 16-Step requirements, they include :

-flexible incorporation of asynchronous information processing with real-time video conferencing and shared data model analysis.,

-possibility to interact spontaneously,

-availability of the case-memory of prototypes, that captures the knowledge on similarities and differences of designed or known processes,

-support in making classifications,

-support of assignment problem-solving,

-support in interrelationship control,

-availability of modeling facilities.

The difference is that 12-Step model requires more of a modeling to be incorporated into the group support environment vs. the 16-Step model, and is actually a tool for periodical 16-Step concept.

5. ANALYSIS OF COMMUNICATION TECHNOLOGY

Three new collaborative computing environments are essentially capable to extend collaborative opportunities of electronic meeting room into the geographically distributed environment by featuring some or all of the following services:

- compressed full motion video,
- shared electronic workspace (shared data models, shared screen, shared graphics and animation),
- high speed file transfer,

- internetworking with switched-based services.

They are: video conferencing, desktop videoconferencing , and multimedia Internet communication. We will consider them subject to media constraints and business requirements on distributed collaboration, that were identified earlier.

5.1 DATA FLOW CONSTRAINTS

In order to satisfy one of the basic requirements on flexible incorporation of shared data model processing, simulation, and real-time videoconferencing potential distributed groupware must transmit consolidated multimedia data flow, such as

$$F=(C1, C2, C3, C4, C5),$$

where

C1: still image and textual information,
C2: video records, animation,
C3: audio/voice
C4: video conference (uncompressed),
C5: video conference (compressed).

According to the results, presented by the researches for Norwegian Telemedicine Project (Kongsill, Slobak, and Vognild, 1993), such data flow with the minimal requirements on video quality may be described by the set of following constraints:

C1: about 4Mb/s,
C2: about 50 Mb/s,
C3: 64 Kb/s,
C4: 240 Mb/s

At such range of total transmission rate required there is no any reasonable solution within existing switch or satellite media to support the geographically distributed collaboration. An application of compression facilities at each node improves the conditions dramatically:

C5: [2-1O] Mb/s, C5 << C4.

Existing compression boards allow to reduce the image size for videoconferencing to the range of [16K, 64K]. With the speed as slow as 15 frames/sec it will draw us to much better conditions on transmission for videoconferencing:

- for two part compressed video conference total transmission rate averages as

$$C1+2*C5=8Mb/s,$$

$$C2+2*C5=12Mb/s$$

- with number of participants growing to n total maximum requirements are

$$(image+n*C5)=(4+n*2)Mb/s,$$

$$(video+n*C5)=(7.5+n*2)Mb/s$$

Such rates gets the case into the boundaries that modern switch media may already satisfy. For example, Northern Telecom Visit Video allows for geographically distributed group to collaborate via standard Switch 56 lines or ISDN lines. Minimal ISDN has two B-channels (64 Kbps) and one D-

channel (16Kbps). One B-channel is used for carrying voice, which has compatibility with telephone lines. The other B-channel is used for video images of participants and data transmission. For local transmission through LAN the data transfer rate is 1.5 Mbps.

Here are some results on availability of such service at Wright -Patterson AFB, collected while the interviews. Currently there is an Accunet Switched 56 service with FTS Switched Data service for 56 and 112 Kbps, already installed in 10 buildings for the trials with video conferencing. Cable connection to the Armstrong Lab building from the closest port of the Switched 56 service is possible. This Fall ASC Communications Computer System Group is planning to install communication SUPERNODE that will provide ISDN service for WPAFB customers. It will be Switch DMS-100 from Northern Telecom. ISDN service that is capable to fully support all types of interactive multimedia, including desktop videoconferencing. Switch 56 is an earlier type of similar facility. In addition ASC Computer System Group is particularly interested in the multiperson conferencing , that will incorporate users of MAC and IBM/PC computers. All this means, that media requirements for physical prototyping of geographically distributed system may be satisfied, and media facilities, that are necessary for proof of concept experiments are available.

5.2 COMPARATIVE EVALUATION OF COLLABORATIVE COMPUTING ENVIRONMENTS

Let us now compare three major alternatives : videoconferencing, desktop conferencing, and multimedia Internet communication on the basis of studies performed earlier by the other researches, and specific requirements derived from the QAF Program distributed group support problem.

5.2.1 MULTIMEDIA INTERNET

One of the reasons, that we start the comparison from asynchronous Internet communication, is that this form of communication may be available almost immediately at any working place with the standard PC or Workstation equipment. It's more of a cultural effort to start to apply it There are several well-known protocols and interfaces, that may be used for client-server group communication on the Internet: HTTP, FTP, Gopher, WAIS, MOSAIC, NNTP. Among them only MOSAIC is capable to transfer most of required multimedia data flow. It's script makes possible to incorporate text, computational model, sound, still image, and life video from the world wide locations into one shared document. In order to test the applicability of MOSAIC we have created a sample of MOSAIC-based agent-facilitator for the distributed group support environment. Results demonstrated good potential of the multimedia Internet to support shared workspace communication. Remote user was able to interact with agenda through hypermedia representation of it's content, and access the modelling tools incorporated in agenda. Due to the limitations on capturing the video, and limitations on it's real-time transmission through regular telephone lines, technology is certainly most appropriate for asynchronous fragments of group communication, specifically, when large number of people is involved. On the other hand, new wide-bandwidth facility on Internet may change the pattern to "low -level interactivity". Therefore, we consider as very important further experiments with multimedia on wide-bandwidth ISDN and Switch 56 networks.

5.2.2 VIDEO CONFERENCING

For the first glance video conferencing system looks like a perfect solution for reproducing group decision making processes in geographically distributed environment. There are several commercial products available. Two of them are most popular: Picture Tel (Danvers, MA), and VTEL (Austin, TX). They are installed as video conferencing rooms, operating at 112 Kbps bandwidth. Price is probably the first problem you meet, it ranges from $60,000 to $80,000. Usual assumption is, that video conferencing will cut down the travel bills. It doesn't happen in reality, moreover it increases travel expenses, because it leads to increased collaboration with distant parties (Gale, 92). A lot of communication related to room booking and personal scheduling take place before the meeting , some

times achieving meeting intention, and when the last-minute crises , the room is booked, so people had to fly (Marks, 94). Tang and Isaacs (1993) from Sun Microsystems Laboratories provided a detailed study on usage of distributed collaboration through 4 interconnected video conferencing rooms (located in California, Massachusets, Colorado, and North Carolina) at Sun.

From observations on working teams they found some problems, that didn't arise in face-to-face meetings:
-problematic audio collisions,
-difficulty in directing the attention of remote participants, and
-diminished interaction.
Here are some of their comments:
During the video conferences there were many instances of audio collisions. Although such collisions naturally occur in face-to-face and phone conversations, they were easily negotiated verbally (aided by gestures.) through precise timing (some times including overlapping talk), and systematic implicit organization. The 0.57 second one-way delay in transmitting audio between video conference rooms markedly disrupted these mechanisms for mediating turn-taking.
Missed glances is another problem that participants struggled with, and the meeting featured marked lack of humor (in part because humor relies on precise timing).

According to the authors, because of the communication difficulties, participants tried to avoid resolving the conflict and disagreement, and preferred to communicate in pairs. They also indicate that users prefer audio with minimal delay even at the expense of disrupting synchrony with video. As long as network constraints require trade-off to conserve bandwidth , experience of Sun Microsystems indicate that degrading video quality before degrading audio quality a more usable experience

Author of this report made several experiments with his students at the University of Texas at Dallas on incorporate decision support models into the group meeting . Because of communication difficulties, that group found the most appropriate way to negotiate the dialog with expert system installed at one location , that group found was to provide the oral answers to the group at the other location. In other words, two video conferencing rooms had a problem discussing and sharing the computer model while discussion.

5.2.3 DESKTOP VIDEO CONFERENCING

Desktop video conferencing is the central representation for collaborative computing environment. It makes available from the working desk all types of services, required for distributed multimedia collaboration:

- personal video conferencing (compressed full motion video),
- shared electronic workspace (shared data models, shared screen, shared graphics and animation),
- high speed file transfer,
- internetworking with switched-based services.

It is typical for desktop video conferencing, that members are located in their own offices where each person has access to his or her own resources and distractions: phone calls, e-mail arrivals, visitors (Tang & Isaacs, 93). Video conferencing rooms allowed only scheduled meetings in the rooms isolated from outside interrupts. Unlike videoconferencing rooms desktop conferencing allows spontaneous interactions between individuals or small groups.

One of desktop video conferencing distinctive features is that non of the participants of the same meeting need to be physically located in one room. They are distributed through the working desks and communicate via personal computers. One of the first experimental studies of multiparty desktop conferencing via ISDN lines was done with MERMAID system (Watabe et al., 90). The MERMAID system has been successfully tested in meetings connecting up to 4 locations, with number of participants

ranging from 2 to 8. The subjects of conferences include software specifications design, planning of research and development activities, and some other system analysis and design issues. Results are the following:

- Participants have noticed little delay in data transmission, even when number of them has increased up to more then four,
- Voice delay is too negligible to be noticed by participants. Though video signal delays from 0.2 to 0.5 second on ISDN (64 Kbps), participants rarely become impatient. It may be because facial expression of speaking member is transmitted in less than 0.3 second.
- Participants noticed delay in the transmission of handwritten data and window manipulations. Some became impatient, because it takes 10 to 20 seconds to send and display shared document.
- When more then four participants joined the conference it has been some difficult to determine who is speaking,
- Superiors and subordinates favored the designation mode to invoke shared windows,
- Persons of nearly equal rank preferred first-come-first-served mode,
- In brainstorming sessions free mode was used.

Researches from Sun Microsystems (Tang and Isaacs, 93) have studied 72 desktop video conferences between two remote locations in Massachusetts and California. Their observations revealed the following details:

- Group interaction that occurred in desktop video conferencing was more like in face-to-face meetings. Remote collaborators were able to interrupt each other, accomplish turn completions, and even occasionaly joke (features that were markedly absent in video conferencing room),
- During the study one-way audio delays were measured between 0.22-0.44 second, which is better then 0.57 delay in video conferencing room, but still noticeable,
- Desktop video conferencing reduced significantly number of e-mail messages per day,
- Desktop video conferencing almost eliminated the use of videoconferencing rooms.

Comparing the observations made on both video conferencing rooms and desktop video conferencing technologies, we may conclude in general, that some serious communication constraints exibited in video conferencing rooms are not observed while desktop video conferencing. Moreover, desktop video conferencing demonstrated less sensitivity to voice delays, immediate access to shared data model, and ability to communicate more like in face-to-face meeting.

For better comparison let us also take a look to the major business requirements, identified earlier for potential distributed group support system:
-flexible incorporation of asynchronous information processing with real-time video conferencing and shared data model analysis.,
-possibility to interact spontaneously,
-availability of the case-memory of prototypes, that captures the knowledge on similarities and differences of designed or known processes,
-support in making classifications,
-support of assignment problem-solving,
-support in interrelationship control,
-availability of modelling facilities.

It is clear that desktop video conferencing provides a better fit to business requirements. It satisfies essentially the requirement on flexible incorporation of information processing and real-time video conferencing , it makes possible spontaneous communication, provides a good potential to incorporate the knowledge base, and other problem-solving support tools.

Several commercial versions of desktop video conferencing products are already available on the market . Their number is rapidly growing. Most recommended are listed in table 2.

Table 2.

NAME	VENDOR	LAN	WAN	MULTIPERSON
Person-to-Person/	IBM	NetWare	TCP/IP	yes
NTV	Peregrine Systems	Video/Windows	-	-
Pro Share	Intel	-	ISDN	no info
Visit Video	Northern Telecom	-	ISDN/Switch 56	up to 20

Subject to the nature of research projects at the Armstrong Laboratory, product from Northern Telecom looks like most appropriate for prototyping of distributed system. It supports geographically distributed collaboration, admits multiperson capabilities, is capable to link MACs and IBM/PCs. allows to work through the old analog lines in addition to Switch 56/ ISDN. It was also tested through several years of implementation as a tool for global collaboration of NT research teams.

Desktop video conferencing is a new technology, that requires a lot of experimental justification as a part of distributed group decision support environment. Many functions and roles are changing: facilitator is no longer in the room, it's activity is subject to change, system manager controls the software via wide-area network , more interactions with software are involved in group collaboration, data sharing interface is new, etc. Therefore , proof of concept experiments are extremely important for prototype design,.

6. INTEGRATED SOLUTION

Desktop video conferencing is a practical means to extend basic group decision support facilities to geographically distributed environment. This form of collaborative computing may be used for preplanning, idea generation , categorizing, brainstorming, and analytical evaluation of alternatives. But it can not substitute all types of collaboration, that such complex design processes, as two year Quality Air Force Program, may need. Besides it's initially unclear how team will adjust to the new environment. Team members are used to face-to-face meetings for problem solving, and asynchronous communication via e-mail. Therefore, we suggest to develop experimental setup on the basis of integrated solution for distributed group support system (fig. 3). Solution is based on GRLL, LAN available to remote users through ISDN lines. One of computers on GRLL LAN is desktop video conferencing hub. At the beginning it will be just the node of desktop video conferencing branch . Remote nodes are connected to hub by ISDN or Switch 56 lines initially through the star topology. Besides the switched lines remote nodes and the hub are Internet nodes at the same time. In this integrated environment remote user and coordinator (facilitator) are capable to flexibly exchange asynchronous multimedia Internet communication, desktop video conferencing, face-to-face meeting at GRLL, and remote presence of some participants at face-to-face meeting of the rest of the group.

The main idea is to allow the group to evolve in such integrated environment in order to find the best usage pattern for long term collaborative problem solving process. It will help us to find experimental background for developing the decision support environment, and help the team to adjust to the new communication technology.

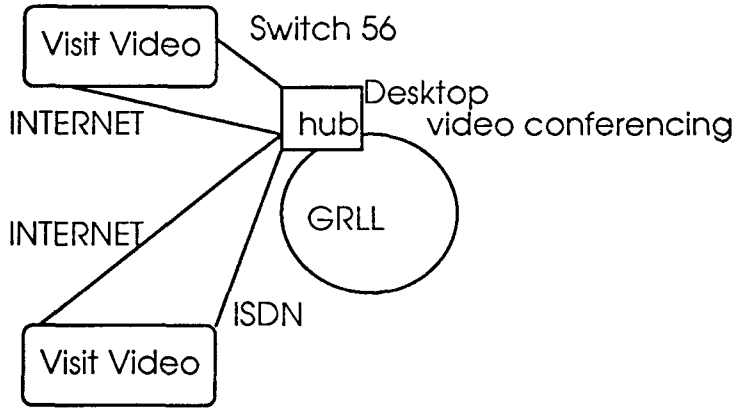


Fig. 3

7. KNOWLEDGE-BASED SUPPORT OF DISTRIBUTED GROUP COLLABORATION

One of the main changes in transition of group collaboration from electronic meeting room to the desktop video conferencing is the change in communication with facilitator and system manager. Theoretically it may even lead to the changes in their roles (that 's why proof of concept experiments enabling to observe those changes are so important). In the electronic meeting room facilitator accumulates and transfers the experience in structuring and solving group problem through his or her personal assistance. In the new distributed environment group-facilitator communication is available only through the software. Therefore the role of software in keeping the tracks of group sessions, accumulating group knowledge , transferring and sharing such information with participants, becomes critically important. Such knowledge transfer and learning features may be performed by knowledge-based agents, that will assist facilitator to coordinate group sessions.

7.1 ARCHITECTURE OF KNOWLEDGE-BASED AGENTS

Looking at observations made on meeting dynamics (pp.5-7), and recommendations provided by facilitator and GRLL system manager, we may initially identify such agents:

-Structuring agent,
-Brainstorming support agent,
-Voting support agent,
-Modelling support agent.

Structuring agent is needed to assist to validate the problem applicability to group environment while preplanning, and to support initial structuring of the group problem while the main phase of meeting. Case-based reasoning may be appropriate type of knowledge representation model for structuring agent

Brainstorming agent is necessary to support shared word processing operations required for categorizing and commenting on ideas. Standard categorizing tools from systems like Group System V, Vision Quest, and others may be imported to provide such service.

Voting support agent is important for system manager to assist in applying the voting model or their combination, that is most appropriate for the current group session. Most probably it may be a rule-based representation with links to multiple criteria solvers.

Modelling agent is an interface unit that will enable to integrate quantitative modelling and other decision analysis techniques with the script of meeting agenda. It may be a kind of HTTM document, similar to WWW MOSAIC. The sample of MOSAIC-based agent-facilitator have been developed for support of group asynchronous collaboration. The demonstrations indicated good potential and feasibility of approach.

Defining specifications for the knowledge-based agents, their prototyping and adaptation to Quality Air Force Program is subject of next step of the research, primarily experimental. According to proposed architecture of collaborative computing environment, group support agents will be on the top desktop video conferencing system application programming interface.

7.2 CASED-BASED REASONING TOOL

As it was mentioned earlier, we consider the case-based reasoning technique as a basic knowledge representation model for initial design and prototyping of distributed group support environment. We have evaluated the potential of this artificial intelligence technique by trial experiments with CBR Express product from Inference Co. (CBR, 94)

CBR Express environment for knowledge representation, problem solving. and learning is represented by the following data models:

Case Panel. In Case Panel new cases of problems , processes, or objects are defined, and old ones are modified. Collection of Case Panels may be used as a case memory.

Question Panel. The Question Panel is used to define the questions, related to case. Various pieces of information about each question and possible answers are used for calculating matching scores during a search.

Action Panel. The Action Panel is where we define the actions for use in case. Actions may be program calls, file transfer operations, browsing of visual image, rolling out video, sound transmission, etc.

Search Panel. It is major representation for problem solving. Through the Search Panel we describe the problem and observe the results of system search: list of similar cases with access to their descriptions, and associated actions.

Tracking Panel. In Tracking Panel various types of personal information on customers associated with case are identified.

If we compare this list of features with basic business requirements to distributed group support system (p.11), we may indicate almost perfect match with many of them:

-availability of prototypes for designed process (Search Panel),
-support in making classifications (Search and Question Panels),
-support in inerrelationship control (Tracking Panel),
-availability of modelling tools during the meeting (Actions Panel).

User describes the models in CBR Express, (and this is essential to case-base reasoning concept in general) in natural English language. That makes the technology specifically suitable for distributed group communication. There is another feature, that may be very useful for distributed group support. In desktop video conferencing environment, actions (identified in CBR Express in Actions Panel) may also be used to initiate video conferencing call to any person, that group or individual may need to interact, in association with case. Correspondingly we may describe communication cases, that will feature different coordination scenarios by monitoring conference calls and interrupts.

8. CONCLUSIONS

Research addresses feasibility analysis of the knowledge-based groupware solutions for support of distributed teams on the basis of collaborative computing technology. Feasibility analysis is the initial stage of the Armstrong Laboratory on-going research on extending the current group decision support facilities of GRLL into geographically distributed problem-solving environments. Different approaches and tools have been used during this study. Analysis of theoretical and experimental studies of other authors enabled to identify key problems that researches face in distributed group support systems design. This knowledge was used for conducting the observations on GRLL problem-solving sessions at the electronic meeting room. Analysis the business problem requirements is based on the interviews and observations on group sessions for the Quality Air Force Program at the Aeronautical Systems Center. Different business process engineering, process assessment, and quality management activities are considered subject to face-to-face, distributed asynchronous, and distributed synchronous forms of collaboration. An integrated solution based on desktop video conferencing link to GRLL meeting room is selected as most suitable for proof of concept experiments. Specific communication requirements include ISDN /Switch 56 media installations. Feasible solutions on physical setup for media are already available from the Communications Computer Systems Groups. Another critical issue is an architecture of group decision support tools, subject to the changes in team communication. Results include initial identification of knowledge based agents: structuring agent, brainstorming agent, voting support agent, modelling support agent. Case-based reasoning technique is selected as a basic knowledge representation model for prototyping of distributed group support environment. Experimental evaluation of CBR Experess system and sample design of agent-facilitator for hypermedia-based asynchronous collaboration, constitute the experimental part of research. Design of knowledge-based agents specifications and their prototypes is one of the tasks for further proof of concept experiments. Results of this research make possible to identify the plan and specifications on proof of concept experiments. Proof of concept experiments is a subject for next step of research .

Author would like to acknowledge Capt. Kennon Moen, Dr. Alan Heminger, Capt. Robert Goerke, and Janet Peasant for their great help and consistent support of this research.

REFERENCES

Ba, S., Lang,R., and Whinston A.(1994).Enterprise Modeling and Decision Support. Working paper, The University of Texas at Austin.

Balakrishnan,A. and Whinston, A.B.(1991). Information Issues on Model Specification. Information Systems Research, Vol.2, No.4, pp.263-286.

Bhargava, H., Krishnan, R., and Whinston,A.B.(1994). On Integarting Collaboration and Decision Analysis Techniques. Working paper, The University of Texas at Austin.

Bordetsky, A., Valtorta, M.(1993). Learning Empirical Constraints to Complement Diagnostic Models. Intelligent Engineering Systems Through Artificial Neural Networks, ASME Press, 1993, vol.3, pp. 97-102.

Bordetsky, A., Maslov, V. and Khavronina, M. (1992). Neural Network for Group Diagnosis in Cases of Compound Pathologies. Intelligent Engineering Systems Through Artificial Neural Networks, Vol.2, p. 993-998, ASME Press, New York.

Bordetsky, A., Maslov, V. and Khavronina, M. (1990). Search Space Decomposition for Distributed Expert Systems With Quantitative Knowledge Bases. News of USSR Academy

of Science. Journal of Computer and System Science, No. 6, p.219-224 (English translation by Script Technica, N.Y.).

CBR Express, User Guide, Inference Co., 1994.

Chang, A.M., Holsapple C.W., and Whinston, A.B.(1993). Model Management Issues and Directions. Decision Support Systems, No.9, pp.19-37.

Gale, S. (1992). Desktop Video Conferencing : Technical Aspects and Evaluation Issues. Computer Communications, Vol. 15, No.2, pp.517-526.

Greenberg,H.J.(1993). Rule-Based Intelligence to Support Linear Programming Analysis.Decision Support Systems, No.9, pp. 425-447.

Heminger, A., Moen, K., Peasant, J., and Wolf, D. (1994). The Group Research Laboratory for Logistics (GRLL): Development and First Use of a GSS Facility Within the Air Force Environment. Technical Report, Armstrong Laboratory, WPAFB, 19p.

Liang, T.-P. and Konsynski, B.R.(1993). Modeling by Analogy. Decision Support Systems, No.9, pp.113-125.

Marks, K. (1994). Candid Camera. LAN Magazine, July, pp.125-131.

Ozernoy, V. and Bordetsky, A.(1992). Using Expert Systems Technology for Multiple Criteria Decision Making: A Survey. Proceedings of Tenth International Conference on Multiple Criteria Decision Making, Taipei, Republic of China, 11p.

QAF, Quality Air Force Unit Self Assessment, ASC Quality Air Force Offfice, 1994, 48p.

Reeves, G., Bordetsky, A. (1994). A Framework for Interactive Multiple Criteria Group Decision Support. Journal on Group Decision and Negotiation, Vol.5, No.4

Shaw, M. and Fox, M.(1993). Distributed Artificial Intelligence for Group Decision Support. Decision Support Systems, No.9 pp.349-367.

Stottler, R. (1994). CBR for Cost and Sales Prediction. AI Expert, August, pp. 25-33.

Sycara,K.P.(1993. Machine Learning for Intelligent Support of Conflict Resolution. Decision Support Systems, No.10,pp.121-136.

Tang, J. and Isaacs, E. (1993). Why Do Users Like Video? Studies of Multimedia -Supported Collaboration. Computer Supported Cooperative Work, International Journal, Vol.1, No.3, pp. 162-196.

Watabe,K., Holsapple, C.W., and Whinston A.B.(1992). Coordinator Support in Nemawashi Decision Process. Decision Support Systems, No.8, pp.85-98.

Watabe, K. et al. (1990). Distributed Multiparty Desktop Conferencing System: MERMAID. CSCW Proceedings, Oct., pp. 27-38.

AN EMPIRICAL EXAMINATION OF THE EFFECT OF SECOND-ORDER SAMPLING ERROR ON ASVAB-TRAINING PROFICIENCY VALIDITY ESTIMATES

Michael J. Burke
Professor
Department of Psychology and Freeman School of Business

Tulane University
2007 Stern Hall
New Orleans, LA 70118

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
Brooks Air Force Base
San Antonio, TX

August 1994

AN EMPIRICAL EXAMINATION OF THE EFFECT OF SECOND-ORDER SAMPLING ERROR ON ASVAB-TRAINING PROFICIENCY VALIDITY ESTIMATES

Michael J. Burke
Professor
Department of Psychology and Freeman School of Business
Tulane University

Abstract

Within the context of estimating the criterion-related validity of the ASVAB Arithmetic Reasoning subtest and Mechanical Composite for predicting final grades in Air Force technical training schools, this study examined the influence of small sets of studies from the research domain of 191 apprentice (level 3) technical training schools on estimates of the mean and variance of validity coefficients in the research domain. More specifically, the effect of randomly sampling three different numbers of studies per meta-analysis (i.e., 5, 10, and 15) from the research domain on the estimates of the mean and variance of validity coefficients in the research domain was examined for two types of meta-analyses: (a) bare-bones meat-analyses where first-order sampling error was the only statistical artifact considered, and (b) meta-analyses involving corrections for sample-based artifacts (i.e., sample size, range restriction, and predictor reliability). In general, there are three primary conclusions from this study: (a) for both types of meta-analyses, when one desires to generalize to all studies in this research domain, small numbers of studies (in particular, samples of 15 studies) may provide adequate estimates of the mean and variance of validity coefficients, (b) when subjects and studies from relevant subpopulations (e.g., career fields) in the research domain are not sampled in proportion to their representation in the domain, the estimates of the mean of the validity coefficients may not closely approximate the mean of the validities in the research domain, and (c) given the possibility of overstating the minimum level of validity from a small number of studies in a meta-analysis, a suggestion would be to employ Ashworth, Osburn, Callender, and Boyle's (1992) for evaluating the robustness of meta-analytic findings.

AN EMPIRICAL EXAMINATION OF THE EFFECT OF SECOND-ORDER SAMPLING ERROR ON ASVAB-TRAINING PROFICIENCY VALIDITY ESTIMATES

Michael J. Burke

If the number of studies in a research domain is relatively large, then the **estimated** average population effect size and estimated variance of population effects from a meta-analysis of these studies will be approximately equal to the **actual** population values. However, for a number of reasons, a researcher often only has access to a sample of studies from the research domain. This sample of studies may even represent all studies that have been conducted at that point in time. When the meta-analysis is based on a small number of studies from the research domain, there will be sampling error in the meta-analytic estimates. This type of sampling is called second-order sampling error (cf. Schmidt, Hunter, Pearlman, & Hirsh, 1985).

As an example of second-order sampling error, consider the prediction of grades in Air Force technical training schools. The Air Force conducts technical training in over 200 schools for each of the Air Force Specialty Codes (AFSCs, Air Force jobs). In each of these schools, the Air Force is interested in the ability of the Armed Services Vocational Aptitude (ASVAB) to predict training school grades. If two researchers each randomly select 15 schools from the 200 schools, compute ASVAB-training grade validity coefficients in their respective sets of 15 schools, then conduct separate meta-analyses and find that their meta-analytic parameter estimates differ, second-order sampling error has influenced the findings in the two meta-analyses. An important unanswered question in the meta-analytic literature is to what extent do the estimated mean and variance of population validity coefficients from the meta-analysis (e.g., from a meta-analysis involving 15 technical training schools) differ from the population parameter values for the research domain (e.g., the estimated mean and variance of the population effects in the research domain of over 200 technical training schools).

It should be noted that even if the number of studies to be included in a meta-analysis is relatively large, these studies may not be representative of all of the studies in the research domain. As discussed by Raju and Dowhower (1991) and Wanous, Sullivan, and Malinak (1989), researchers make judgments about including specific studies in a meta-analysis which may systematically alter the number of subpopulations represented in the set of studies to be meta-analyzed. That is, the number of subpopulations included in the meta-analysis may not be representative of the research domain of interest. Although this latter issue concerns systematic deficiencies in the sampling of studies from the research domain, it is plausible that nonrepresentativeness of studies (i.e., subpopulations) may result from a random sampling process when the number of studies sampled from the research domain is relatively small. For example, nonrepresentativeness could occur when the number of studies randomly sampled from the research domain for inclusion in the meta-analysis is less than the number of subpopulations.

Raju and Dowhower (1991) examined the effect of the number of studies per meta-analysis and the representativeness of studies on "bare-bones" (i.e., only first-order sampling error is considered) meta-analytic estimates of the mean and variance of validity coefficients for simulated populations (i.e., populations created from empirical data). Their study indicated that the validity of a composite cognitive ability test score (i.e., Math and Word Knowledge) for predicting Air Force technical school grades generalized across three different sample sizes (i.e., 30, 68, 100) and 10 different numbers of studies per meta-analysis. Their results also demonstrated that when all relevant populations are not represented in a meta-analysis, the resulting meta-analytic parameter estimates are generalizable only to those populations that are included in the analysis.

In addition, in a study by Raju, Pappas, and Williams (1989), which used the same data base and simulated populations as Raju and Dowhower, the accuracy of three meta-analytic models (i.e., correlation, regression, and covariance) for three samples sizes (i.e., 30, 68, and 100) and 10 different numbers of studies per meta-analysis (ranging from 10 to 100) was examined. Their results indicated that when sampling error is the only artifact considered, all three models do well in estimating the relevant parameters (i.e., means and variances of correlations, regressions slopes, and covariance's) across the three samples sizes and 10 different numbers of studies per meta-analysis. Their research was not directly concerned with the issue of second-order sampling.

Two other studies (Schmidt & Hunter, 1984; Schmidt, Ocasio, Hillery, & Hunter, 1985) provide an indication of the potential impact of second-order sampling of studies even within the same setting. For instance, Schmidt and Hunter (1984) reanalyzed data reported in Bender and Loveless (1958) on a series of validity studies conducted annually on different cohorts of stenographers for four successive years in the same organization. The results of their reanalyzes indicated that the ratio of the standard deviation of validities predicted from sampling error to the standard deviation of the observed validities departed considerably from one when the number of studies per meta-analysis was four. However, this ratio was approximately equal to one when the number of studies included in the meta-analysis was 20.

An empirical assessment of the effect of second-order sampling error has not appeared in the literature. In particular, it would be informative to have an empirical assessment of the impact of second-order sampling on bare-bones meta-analytic parameter estimates (where first-order sampling error is the only artifact considered) as well as on corrected meta-analytic parameter estimates (where statistical artifacts such as criterion reliability and range restriction). Within the context of estimating the criterion-related validity of cognitive ability tests for predicting training proficiency in Air Force technical training schools, the present study will examine the influence of sampling small sets of studies (i.e., 5, 10, and 15) from the research domain on "bare-bones" and corrected meta-analytic estimates of the mean and variance of validity coefficients in the domain of all possible studies. These numbers of studies per meta-analysis are typical of distributions of effects in non-selection test validation meta-analyses (e.g., see Driskell, Willis, & Cooper, 1992; Fried, 1991; Mitra, Jenkins, & Gupta, 1992; Narby, Cutler, & Moran, 1993). Furthermore, numbers of studies in the range of 10 to 15 are representative of realistic numbers of studies for future Air Force criterion-related validation studies involving cognitive tests.

Although the degree of second-order sampling error is expected to decrease as the number of studies per meta-analysis increases, the magnitude of this decrease within a defined research domain is of primary interest in the present study. Furthermore, for Raju et al.'s (1991) meta-analytic procedure (hereafter referred to as the RBNL procedure), this study will provide a limited empirical examination of the efficacy of their procedure for estimating the sampling variance of the mean of corrected correlations (i.e., when M_ρ is based on three different numbers of studies: 5, 10, and 15).

The decision to use the RBNL meta-analytic procedure as opposed to employing an alternative correlational meta-analytic procedure (cf. Burke, 1984) is based on three considerations. The primary consideration is that the RBNL procedure has been shown to be more accurate than other meta-analytic procedures (Raju et al., 1991). Relatedly, since complete sample-based range restriction and predictor reliability data will be available, the RBNL procedure is more appropriate for estimating the mean and variance of corrected correlations. Third, as noted above, the RBNL procedure provides an estimate of the sampling variance of the mean of corrected correlations (i.e., $V_{M\rho}$) which can be used in constructing confidence intervals for M_ρ.

Thomas's (1990, 1991) mixture model could be considered in the present context for estimating the number of population correlations (i.e., ρ_i's), point values of these ρ_i's, and the proportion of observed correlations associated with each ρ_i. Mixture model-based estimates of the mean and variance of the ρ_i's, however, may not provide good estimates of these population parameters for disattenuated correlations to be studied here (cf. Thomas, 1990). Therefore, Thomas's procedure is not considered further in this study. Below, an overview of the RBNL meta-analytic procedure is presented.

Raju, Burke, Normand, and Langlois' (1991) Meta-Analytic Procedure

The equation that forms the basis of the Raju et al. (1991) procedure is

$$\hat{\rho}_i = \rho_i + e_i \qquad (1)$$

This equation indicates that the estimated true correlation, or effect, in an individual study is equal to the true correlation plus the sampling error associated with that corrected correlation in that study, where the estimated population correlation ($\hat{\rho}_i$), according to classical test theory, is defined as

$$\hat{\rho}_i = \frac{k_i r_i}{(r_{xixi}\, r_{yiyi} - r_i^2 + k_i^2 r_i^2)^{1/2}} \qquad (2)$$

where $k_i = 1/u_i$, u_i is the ratio of the restricted standard deviation in to the unrestricted standard deviation of the predictor, r_{xixi} is the sample-based predictor reliability, r_{yiyi} is the sample-based criterion reliability, and r_i is the correlation between x and y in sample i. Furthermore, Raju et al. (1991) presented the general formula for calculating an asymptotic estimate of the sampling variance of an individually corrected correlation ($\hat{V}_e$):

$$\hat{V}_{ei} = \frac{k_i^2 \, r_{xixi} \, r_{yiyi} \, (r_{yiyi} - r_i^2) \, (r_{xixi} - r_i^2)}{N_i W_i^3} \qquad (3)$$

where

$$W_i = r_{xixi} \, r_{yiyi} - r_i^2 + k^2 r_i^2 \qquad (4)$$

and where the other elements of Equation 3 and 4 are defined as above. Special cases of V_{ei} based on the availability of artifactual data were presented in Raju et al. (1991).

The sampling variance formula given in Equation 3 above (i.e., for the asymptotic variance of V_{ei}) is new to the literature. Importantly, the square root of the estimated sampling error variance in Equation 3 is the estimated standard error for a corrected correlation coefficient. This standard error can be employed in the construction of a confidence interval around an individually corrected correlation (ρ_i).

Once the two estimates of each ρ_i and V_{ei} are known, the sample-size weighted mean of $\hat{\rho}$'s and its sampling variance can be obtained as follows:

$$M_{\hat{\rho}} = w_1 \hat{\rho}_1 + w_2 \hat{\rho}_2 + \ldots + w_n \hat{\rho}_n \qquad (5)$$

where

$$w_i = \frac{N_i}{N_1 + N_2 + \ldots + N_n} \qquad (6)$$

and N_i represents the number of subjects in study i and n is the total number of studies.

Now, the sampling variance of M_ρ can be written as

$$V_{M\rho} = w_1^2 V_{e1} + w_2^2 V_{e2} + \ldots + w_k^2 V_{en} \qquad (7)$$

This estimate of the sampling variance of M_ρ may be used to set up confidence intervals around $M_{\hat{\rho}}$ to test the hypothesis that M_ρ is different from zero.

To test the hypothesis that the mean population effect is different from zero, an approximate 95-percent confidence interval for M_ρ in a meta-analysis is defined as

$$M_{\hat{\rho}} - 1.96(\sqrt{V_{M\hat{\rho}}}) \leq M_\rho \leq M_{\hat{\rho}} + 1.96(\sqrt{V_{M\hat{\rho}}}) \qquad (8)$$

where $V_{M\rho}$ is estimated from Equation 7. In the present study, 95% confidence intervals for $M_{\hat{\rho}}$ based on sample-size weighted estimates of each V_{ei} will be constructed. It is expected that the confidence interval based on sample-size weights will include the value of $M_{\hat{\rho}}$ from the research domain in a majority of the cases. In this sense, the present study will provide a limited empirical assessment of the efficacy of constructing a confidence interval for M_ρ based on sample-size weights for each V_{ei}.

Within the RBNL meta-analytic procedure, the basic equation for estimating V_ρ is

$$V_\rho = V_{\hat{\rho}} - V_e \qquad (9)$$

where V_e is the sample-size weighted variance of the specific V_{ei} estimates, which can be written as

$$V_e = \frac{N_1 V_{e1} + \ldots + N_n V_{en}}{\Sigma N_i} \qquad (10)$$

where each V_{ei} is computed in Equation 3 above. The estimate of V_ρ on the right-hand side of Equation 9 is computed as the sample-size weighted variance of the $\hat{\rho}_i$'s from Equation 2. The square root of the final estimate of V_ρ is typically employed in meta-analyses to construct a credibility interval for a distribution of effects.

Method

Subjects

The subjects were 88,188 non-prior-service Air Force recruits who were tested with parallel ASVAB Forms 11, 12, and 13 during 1984 to 1988. Only individuals who completed technical training school and who had final school grades were included in the study. The research domain was defined as all apprentice (AF level 3) technical training schools with a minimum of 30 individuals who had completed training. Only schools with 30 or more subjects were considered as part of the research domain since individual studies would likely not be conducted in smaller schools. This decision resulted in the elimination of 635 students in 61 schools, less than 1% of the subjects in the research domain. The demographic characteristics of the subjects in the research domain are presented in Table 1.

Insert Table 1 about here

Since the samples were to be drawn from military technical training schools, it was not expected that situational variables (e.g., psychological climate) would act as substantive causes of validity coefficients or statistical artifacts (cf. James, Demaree, Mulaik, & Ladd, 1991). In addition, recent research indicates that situational variables may not act as substantive causes of criterion-related validities and reliability coefficients (cf. Rupinski & Burke, 1994). Therefore, situational variables were not considered in the subsequent analyses.

Predictor and Criterion Measures

The predictor measures considered for this study were the Arithmetic Reasoning subtest and the Mechanical Composite of the ASVAB. The Arithmetic Reasoning subtest was chosen to represent a general cognitive ability test that has been considered a good indicator of the higher-order factor, psychometric g (cf. Ree & Carretta, 1994). Also, Earles and Ree (1992) have shown Arithmetic Reasoning to be one of the most generally valid tests in the ASVAB. Raw scores on the 30-item Arithmetic Reasoning test were employed in this study. The Mechanical Composite was randomly chosen to represent one of the four ASVAB composites. The Mechanical Composite is composed of three 25-item general knowledge subtests: Mechanical Comprehension, General Science, and Auto and Shop Information. Scores on the Mechanical Composite are in the metric of the normative reference standard scores which are based on a nationally representative sample of youth collected in 1980 (Department of Defense, 1984). Mechanical Composite scores are computed based on applying unit weights to Mechanical Comprehension and General Science and a weight of two to the Auto and Shop subtest.

The criterion measure was the final school grade (FSG) earned by each student in the 191 technical schools. The grades are based on an average of a series of multiple-choice tests administered during the course

(cf. Earles & Ree, 1992; Ree & Earles, 1991). The grades for the present group ranged from 60 to 99, with an average of 86.7 and a standard deviation of 6.4.

Procedure

Study 1. This study was designed to examine the accuracy of meta-analytic parameter estimates when the number of studies per meta-analysis varied and when sample-based predictor reliability estimates and range restriction effects were considered. Study 1 was conducted separately for three different numbers of studies per meta-analysis (i.e., 5, 10, and 15) and two types of tests (i.e., the ASVAB Arithmetic Reasoning subtest and Mechanical Composite). The procedure is described below for the condition of 5 studies per meta-analysis.

For the first meta-analysis, five Air Force Specialty Codes (AFSCs) were randomly sampled from the 191 AFSCs in the research domain. Each AFSC sample (study) was associated with a specific technical training school. For each of the five AFSC samples and for the predictor measure under consideration, a criterion-related validity coefficient, predictor reliability estimate, and range restriction value were computed. Sample-based criterion reliabilities could not be computed. Therefore, corrections for criterion unreliability were not made. In addition, corrections for criterion unreliability as well as multivariate range restriction (cf. Lawley, 1943) were not made since this study was not concerned with estimating fully corrected true validities or operational true validities. The decision to fix criterion reliability at 1.0 is consistent with previous research examining the criterion-related validity of the ASVAB (cf. Earles & Ree, 1992). However, for the theoretical purposes of evaluating the effect of second-order sampling with respect to sample-based artifact data, corrections for predictor reliability were made.

Since item-level data were unavailable, sample-based predictor (test) reliabilities were estimated with KR-21 (Kuder & Richardson, 1937). KR-21 is calculated from the mean, variance, and number of items on a test. If the item difficulties are not equal, KR-21 will underestimate KR-20 and a test's reliability. Sample-based reliabilities for the Mechanical Composite were estimated using stratified alpha (Rajaratnam, Cronbach, & Gleser, 1965). In order to apply stratified alpha, internal consistency reliability estimates (KR-21) for the subtests that comprise the Mechanical Composite were needed.

The range restriction value on the predictor for each of the five studies was initially computed as follows. First, for the five randomly sample studies, an estimate of the unrestricted standard deviation on the predictor was made. As noted above, only subtest or composite scores on a respective predictor were available. By pooling the data across the five studies, an estimate of the unrestricted standard deviation on the predictor could be made (cf. Glass & Stanley, 1970). That is, the predictor means, standard deviations, and number of subjects in each of the five studies were employed for estimating the unrestricted predictor standard deviation (in the research domain). Then, for each of the five studies, the ratio (i.e., u-ratio) of the predictor standard deviation in the study to the estimated unrestricted standard deviation on the predictor (based on the five studies) was computed. This u-ratio was initially considered as the estimated sample-based range restriction value. This

procedure for computing u-ratios was considered since it reflected how a researcher might estimate range restriction values with all available predictor data at the time of a meta-analysis.

Initial comparisons of u-ratios based on the above procedure to u-ratios computed with the unrestricted predictor standard deviation in the research domain indicated that the two u-ratios were approximately equal. Given these results and the fact that the RBNL procedures for estimating the sampling variances of individually corrected correlations do not incorporate the sampling error in sample-based range restriction values, a decision was made to compute each sample-based u-ratio with the sample predictor standard deviation and the predictor standard deviation in the research domain.

Then, for the five studies, the RBNL meta-analysis procedure was applied. The magnitude of second-order sampling effects was determined by comparing the estimates of the mean ($M_{\hat{\rho}}$) and variance of corrected correlations ($V_{\hat{\rho}}$) based on the five studies with the respective $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ values for the research domain (i.e., 191 AFSCs). Technically, the $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ values for the research domain are themselves subject to sampling error. However, for the purposes of the present study, the M_ρ and V_ρ estimates for the research domain served as the comparison standards for judging the magnitude of second-order sampling error.

The above process was repeated 5 times resulting in five meta-analyses (each with five studies) for the predictor measure. For the separate meta-analyses, studies were sampled with replacement. The rationale for sampling with replacement was that if a study were conducted in one Air Force technical training school, this school would not necessarily be excluded from future studies. Different random number seeds were employed for the analyses involving Arithmetic Reasoning and the Mechanical Composite which resulted, for the most part, in different samples (AFSCs) being drawn for the comparable meta-analyses with five studies per meta-analysis.

The entire process was repeated for 10 studies per meta-analysis and 15 studies per meta-analysis for both the Arithmetic Reasoning subtest and Mechanical Composite. It should be noted that two individual studies had negative validity coefficients for the Arithmetic Reasoning subtest and one study had a negative validity coefficient for the Mechanical Composite. Therefore, these studies were eliminated from the respective analyses involving the Mechanical Composite and Arithmetic Reasoning subtest.

Study 2. This study was similar to Study 1 with the exception that sampling error was the only artifact considered. The meta-analyses were carried out with respect to the Hunter and Schmidt (1990) "bare-bones" procedure.

Results

The results of the meta-analyses conducted in Studies 1 and 2 are summarized in Tables 2 through 5. Below, these results will be presented separately for meta-analyses involving corrections for sample-based artifact data (Study 1) and for meta-analyses where the only artifact considered was first-order sampling error (Study 2).

Study 1

The summary results for meta-analyses examining Arithmetic Reasoning- Final School Grade relationships with sample-based artifact data are given in Table 2. As shown, across the three conditions of 5, 10 and 15 studies per meta-analysis, the estimates of M_ρ were generally close to the value of $M_{\hat{\rho}}$ in the research domain (with corrections for sample size, predictor unreliability, and range restriction). The largest differences between the estimates of M_ρ and the estimated M_ρ in the research domain were for five studies per meta-analysis. The differences ranged from a low of -.005 to a high of .064. On average, these differences were small for all three conditions. In addition, the 95% confidence intervals for $M_{\hat{\rho}}$ contained the value of $M_{\hat{\rho}}$ from the research domain in 12 of the 15 meta-analyses.

Insert Table 2 about here

Table 2 also shows that, across the three conditions, the estimates of V_ρ in the respective meta-analyses were generally close to the value of $V_{\hat{\rho}}$ in the research domain. In each of the conditions, there was a tendency for the meta-analytic estimate of V_ρ to underestimate the value of $V_{\hat{\rho}}$ in the research domain. The largest range in V_ρ estimates across meta-analyses (of .016) occurred in the condition with five studies per meta-analysis. The meta-analytic estimate of V_ρ underestimated $V_{\hat{\rho}}$ in the research domain in 10 of the 15 meta-analyses.

The summary results for meta-analyses examining Mechanical Composite-Final School Grade relationships with sample-based artifact data are given in Table 3. In a majority of the cases and across the three conditions of 5, 10 and 15 studies per meta-analysis, the estimates of M_ρ were close to the value of $M_{\hat{\rho}}$ in the research domain. However, in five cases, the estimated difference between the $M_{\hat{\rho}}$ and the value of $M_{\hat{\rho}}$ in the research domain was .09 or greater in absolute value. Importantly, these over and under estimates were observed at each of the three conditions. The largest differences between the estimates of M_ρ and $M_{\hat{\rho}}$ from the research domain were with respect to five studies per meta-analysis. The range in these differences for five studies per meta-analysis was .23. In addition, the 95% confidence intervals for $M_{\hat{\rho}}$ contained the value of $M_{\hat{\rho}}$ from the research domain in 7 of the 15 meta-analyses.

Insert Table 3 about here

Table 3 also shows that, across the three conditions, the estimates of V_ρ in the respective meta-analyses were generally close to the value of $V_{\hat{\rho}}$ in the research domain. For all three conditions, there was a tendency for the meta-analytic estimate of V_ρ to underestimate the value of $V_{\hat{\rho}}$ in the research domain. The meta-analytic estimate of V_ρ underestimated $V_{\hat{\rho}}$ in the research domain in 11 of the 15 meta-analyses. The largest range in V_ρ estimates across meta-analyses (of .029) occurred in the condition with five studies per meta-analysis.

Study 2

The summary results for the bare-bones meta-analyses examining Arithmetic Reasoning- Final School Grade relationships are given in Table 4. As shown, across the three conditions of 5, 10 and 15 studies per meta-analysis, the estimates of M_ρ were generally close to the value of $M_{\hat{\rho}}$ in the research domain. The largest differences between the estimates of M_ρ and the value of $M_{\hat{\rho}}$ in the research domain were for ten studies per meta-analysis. The differences ranged from a low of -.047 to a high of .036. On average, these differences were small for all three conditions. In addition, the 95% confidence intervals for $M_{\hat{\rho}}$ contained the $M_{\hat{\rho}}$ value from the research domain in 12 of the 15 meta-analyses.

Insert Table 4 about here

Table 4 also shows that, across the three conditions, the estimates of V_ρ in the respective meta-analyses were generally close to the $V_{\hat{\rho}}$ in the research domain. In each of the conditions, there was a tendency for the meta-analytic estimate of V_ρ to be lower than the value of $V_{\hat{\rho}}$ in the research domain. The meta-analytic estimate of V_ρ was less than $V_{\hat{\rho}}$ in the research domain in 10 of the 15 meta-analyses.

The summary results for the bare-bones meta-analyses examining Mechanical Composite-Final School Grade relationships are given in Table 5. Across the three conditions of 5, 10 and 15 studies per meta-analysis, the estimates of M_ρ were generally close to the value of $M_{\hat{\rho}}$ in the research domain. In three of the 15 meta-analyses, the difference between the $M_{\hat{\rho}}$ and the value of $M_{\hat{\rho}}$ in the research domain was -.09 or greater. The largest range in these differences (.17) was for the condition of five studies per meta-analysis. In addition, the 95% confidence intervals for $M_{\hat{\rho}}$ contained the value of $M_{\hat{\rho}}$ from the research domain in 8 of the 15 meta-analyses.

Insert Table 5 about here

Table 5 also shows that, across the three conditions, the estimates of V_ρ in the respective meta-analyses were close to the $V_{\hat{\rho}}$ in the research domain. The meta-analytic estimates of V_ρ were less than the $V_{\hat{\rho}}$ value in the research domain in 10 of the 15 meta-analyses. The largest underestimates were in the condition of five studies per meta-analysis.

Discussion

In this study, the effect of randomly sampling small sets of studies (i.e., 5, 10, and 15) from a research domain on the estimates of the mean and variance of population parameters in the research domain was examined. Overall, the results indicate that when one desires to generalize to studies in this research domain, small samples of studies may provide adequate estimates of the mean and variance of validity coefficients in the

research domain. This conclusion is the same for meta-analyses involving only corrections for sampling error (i.e., bare-bones meta-analyses) and meta-analyses incorporating sample-based statistical artifact data. On average, the M_ρ and V_ρ estimates more closely approximated the values of $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ in the research domain as one moved from 5 to 15 studies per meta-analysis. These results would be expected from sampling theory. Furthermore, in a majority of the meta-analyses, the 95% confidence interval for $M_{\hat{\rho}}$ contained the $M_{\hat{\rho}}$ from the research domain. Although these latter findings do not directly address the accuracy of the present formula for estimating the standard error of the mean of corrected correlations, these findings provide tentative empirical support for the estimate of $SE_{M\rho}$ (the square root of Equation 7 above).

For the estimates of V_ρ, the present findings reflect the expected result that error due to second-order sampling can be reduced by including more studies in the meta-analysis or by conducting a second-order meta-analysis (i.e., a meta-analysis of each of the present sets of meta-analyses). It should be noted that the estimates of V_ρ were often less than the $V_{\hat{\rho}}$ in the research domain. The majority of these latter V_ρ estimates are not likely to influence decisions concerning the practical utility of the cognitive ability tests studied here. The reason being that, in all of these cases where the estimated V_ρ was less than $V_{\hat{\rho}}$ in the research domain, the estimated M_ρ's and lower bound 90% credibility values (not reported here) were positive. However, the V_ρ estimates that underestimate the value of $V_{\hat{\rho}}$ in the research domain could possibly influence conclusions concerning the generalizability of a particular test across jobs in the research domain. That is, these latter V_ρ estimates from the separate meta-analyses indicate that validity generalizes (at a minimum level) to a greater degree than is implied by the values of $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ for the research domain. Given the possibility that one might overstate conclusions from meta-analyses involving relatively small numbers of studies from a research domain, a suggestion would to employ the methodology proposed by Ashworth, Osburn, Callender, and Boyle (1992) for gaining insight into what the mean and standard deviation of unrepresented validity coefficients (i.e., in this case, in training schools not represented in the meta-analysis) would need to be in order to threaten a positive meta-analytic finding.

Although most of the estimates of M_ρ and V_ρ closely approximated the values for $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ in the research domain, there were several notable exceptions. In each case where the meta-analytic estimate of M_ρ substantially (i.e. by .10) underestimated or overestimated the value of $M_{\hat{\rho}}$ in the research domain, the difference was due to the influence of a large sample from a career field that was not representative of the research domain. For example, one study with a sample size of 3,429 caused the meta-analytic estimate of M_ρ in two meta-analyses for the Mechanical Composite (i.e., for 10 and 15 studies per meta-analysis with sample-based artifact data) to be substantially underestimated. This job is from a career field that has an estimated M_ρ of .21 in this research domain. This latter value for $M_{\hat{\rho}}$ is considerably less than the estimated M_ρ of .429 for the Mechanical Composite in the research domain. Furthermore, the jobs in this career field represent only 4% of the subjects in the research domain. In each of the respective meta-analyses where the $M_{\hat{\rho}}$ in the research

domain was substantially underestimated, the job comprised 57% and 47% of the subjects. When this latter study (job) was removed from each meta-analysis and the meta-analyses were rerun, the estimates of M_ρ more closely approximated $M_{\hat{\rho}}$ in the research domain. For instance, when this study was removed from the Mechanical Composite meta-analysis number 4 with 10 studies per meta-analysis, the revised $M_{\hat{\rho}}$ was .469. Although this latter value overestimated M_ρ in the research domain by .04, the magnitude of the difference was reduced (from -.11) and the 95% confidence interval for $M_{\hat{\rho}}$ contained the estimated M_ρ in the research domain. Since the $M_{\hat{\rho}}$ and the 95% confidence for $M_{\hat{\rho}}$ are based on sample-size weights, these results caution one to ensure that not only should efforts be made to include all relevant or primary subpopulations in a meta-analysis, but also to attempt to include sample sizes from such subpopulations that are proportional to their representation in the research domain.

Overall, these results are consistent with Raju and Dowhower's (1991) findings that when the validity studies included in a meta-analysis are not representative of all relevant subpopulations or when all relevant subpopulations are not adequately sampled, then the generalizability of meta-analytic results to the research domain is affected. Further empirical research on second-order sampling involving alternative predictor-criterion relationships, samples sizes that are more typical of those in the extant literature, and studies where sample-based criterion reliabilities can be estimated would be informative. Also, given the comparability of predictor and criterion metrics across studies (such as those in the present research domain), research on the effect of second-order sampling on regression slopes would add to our knowledge base. Finally, computer simulation studies examining the accuracy of the present formula for estimating the sampling variance of the mean of corrected correlations (Equation 7 and its square root, $SE_{M\rho}$) that incorporate alternative weighting procedures for the estimated sampling variances of individually corrected corrections in this formula would contribute to an evaluation of the efficacy of this estimate of $SE_{M\rho}$.

References

Ashworth, S.D., Osburn, H.G., Callender, J.C., & Boyle, K.A. (1992). The effects of unrepresented studies on the robustness of validity generalization results. Personnel Psychology, 45, 341-361.

Burke, M.J. (1984). A review and critique of the correlation model. Personnel Psychology, 37, 93-115.

Department of Defense. (1984). Test Manual for the Armed Services Vocational Aptitude Battery. North Chicago, IL: United States Military Entrance Processing Command.

Driskell, J., Willis, R., & Cooper, C. (1992). Effect of overlearning on retention. Journal of Applied Psychology, 77, 615-622.

Earles, J.A., & Ree, M.J. (1992). The predictive validity of the ASVAB for training grades. Educational and Psychological Measurement, , 721-725.

Fried, Y. (1991). Meta-analytic comparison of the job diagnostic survey and job characteristics inventory as correlates of work satisfaction and performance. Journal of Applied Psychology, 76, 690-697.

Glass, G.V., & Stanley, J.C. (1970). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall.

Hunter, J.E., & Schmidt, F.L. (1990). Methods of meta-analysis. Newbury Park,CA: Sage.

James, L. R., Demaree, R.G., Mulaik, S.A., & Ladd, R.T. (1992). Validity generalization in the context of situational models. Journal of Applied Psychology, 77, 3-14.

Kuder, G.F., & Richardson, M.W. (1937). The theory of the estimation of test reliability. Psychometrika, 2, 151-160.

Lawley, D.N. (1943). A note on Karl Pearson's selection formulae. Proceedings of the Royal Society of Edinburgh, Section A, 62, Part 1, 28-30.

Mitra, A., Jenkins, G., & Gupta, N. (1992). A meta-analytic review of the relationship between absence and turnover. Journal of Applied Psychology, 77, 879-889.

Narby, D., Cutler, B., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability. Journal of Applied Psychology, 78, 34-42.

Palmer, P., Hartke, D.D., Ree, M.J., Welsh, J.R., Jr., & Valentine, L.D., Jr. (1988). Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms, 8, 9, 10, and 11). (AFHRL-TP-87-48, AD A191 658). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. Psychometrika, 30, 39-56.

Raju, N.S., Burke, M.J., Normand, J., & Langlois, G.M. (1991). A new meta-analytic approach. Journal of Applied Psychology, 76, 432-446.

Raju, N.S., & Dowhower, D.P. (1991). The effect of second-order sampling on the accuracy of validity generalization results. Paper presented at the Sixth Annual Meeting of the Society for Industrial and Organizational Psychology, April, St. Louis.

Raju, N.S., Pappas, S., & Williams, C.P. (1989). An empirical monte carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. Journal of Applied Psychology, 74, 901-911.

Ree, M.J., & Carretta, T.R. (1994). Factor Analysis of the ASVAB: Confirming a Vernon-Like structure. Educational and Psychological Measurement, 54, 459-463.

Ree, M.J., & Earles, J. (1991). Predicting training success: Not much more than g. Personnel Psychology, 44, 321-332.

Rupinski, M.T., & Burke, M.J. (1994). Meta-analysis in the context of situational constraints. Paper presented at the Ninth Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, TN, April.

Schmidt, F.L., & Hunter, J.E. (1984). A within setting empirical test of the situational specificity hypothesis in personnel selection. Personnel Psychology, 37, 317-326.

Schmidt, F.L., Hunter, J.E., Pearlman, K., & Hirsh, H.R. (1985). Forty questions about validity generalization and meta-analysis. Personnel Psychology, 38, 697-798.

Schmidt, F.L., Ocasio, B.P., Hillery, J.M., & Hunter, J.E. (1985). Further within-setting empirical tests of the situational specificity hypothesis in personnel selection. Personnel Psychology, 38, 509-524.

Thomas, H. (1989). A mixture model for distributions of correlation coefficients. Psychometrika, 54, 523-530.

Thomas, H. (1990). A likelihood-based model for validity generalization. Journal of Applied Psychology, 75, 13-20.

Table 1

Demographic Characteristics of Subjects in Research Domain

Category	N	Percent
Gender		
Male	72,422	82.2
Female	15,696	17.8
Race/Ethnic Group		
American Indian	260	.3
Asian	1,572	1.8
Black	12,927	14.7
Hispanic	2,515	2.9
White	70,844	80.4
Educational Level		
Less than High School	745	.9
High School Graduate	69,919	80.2
Some College Experience	14,700	16.7
Associates Degree	1,208	1.4
College Graduate	1,546	1.8
Age At Entry		
17 - 18	25,521	29.0
19 - 20	33,128	37.6
21 - 22	16,759	19.0
23 +	12,710	14.4

Table 2

Summary of Population Parameter Estimates and Measures of Accuracy for Meta-Analyses with Sample-Based Artifact Data: Arithemtic Reasoning-Final School Grade Relationships

Meta-Analysis	Avg. N Per Study	$M_{\hat{\rho}}$	Difference	95% c.i. for $M_{\hat{\rho}}$	$V_{\hat{\rho}}$	Difference
			Five Studies Per Meta-Analysis			
1	125	.439	.022	.353, .526	.009	.001
2	599	.375	-.042	.336, .414	.000	-.008
3	234	.480	.064	.393, .568	.016	.006
4	566	.374	-.043	.334, .413	.001	-.007
5	929	.412	-.005	.380, .444	.000	-.008
Averages	491	.416	-.001		.005	-.003
			Ten Studies Per Meta-Analysis			
1	336	.378	-.039	.335, .422	.004	-.004
2	1124	.408	-.009	.389, .428	.004	-.004
3	658	.434	.017	.404, .464	.012	.004
4	384	.419	.002	.382, .455	.002	-.006
5	1282	.356	-.061	.336, .376	.006	-.002
Averages	757	.399	-.018		.006	-.002
			Fifteen Studies Per Meta-Analysis			
1	406	.435	.018	.406, .465	.013	.005
2	503	.423	.006	.398, .448	.009	.001
3	369	.441	.024	.410, .472	.004	-.004
4	495	.417	.000	.392, .443	.001	-.007
5	585	.407	-.010	.384, .429	.003	-.005
Averages	472	.425	.008		.006	-.002

Note. The $M_{\hat{\rho}}$ and $V_{\hat{\rho}}$ values for the 189 technical training schools were **.417** and **.008**, respectively.

Table 3

Summary of Population Parameter Estimates and Measures of Accuracy for Meta-Analyses with Sample-Based Artifact Data: Mechanical Composite-Final School Grade Relationships

Meta-Analysis	Avg. N Per Study	$M_{\hat{\rho}}$	Difference	95% c.i. for $M_{\hat{\rho}}$	$V_{\hat{\rho}}$	Difference
			Five Studies Per Meta-Analysis			
1	457	.451	.022	.408, .494	.001	-.002
2	295	.407	-.022	.349, .466	.000	-.018
3	152	.519	.090	.456, .583	.029	.011
4	349	.287	-.142	.236, .337	.010	-.008
5	148	.467	.038	.392, .542	.013	-.005
Averages	280	.426	-.014		.011	-.007
			Ten Studies Per Meta-Analysis			
1	1201	.411	-.018	.393, .429	.016	-.002
2	583	.423	-.006	.397, .448	.010	-.008
3	556	.300	-.129	.272, .328	.009	-.004
4	605	.317	-.112	.289, .344	.022	.004
5	550	.375	-.054	.348, .403	.008	-.010
Averages	699	.365	-.064		.013	-.005
			Fifteen Studies Per Meta-Analysis			
1	332	.493	.064	.465, .520	.006	-.012
2	208	.442	.013	.405, .479	.022	.004
3	491	.428	-.001	.404, .453	.002	-.016
4	817	.393	-.036	.375, .411	.021	.003
5	482	.342	-.087	.318, .366	.024	.006
Averages	466	.420	-.009		.015	-.003

Note. The estimates of M_ρ and V_ρ for the 190 technical training schools were **.429** and **.018**, respectively.

Table 4

Summary of Population Parameter Estimates and Measures of Accuracy for Bare-Bones Meta-Analyses: Arithemtic Reasoning-Final School Grade Relationships

Meta-Analysis	Avg. N Per Study	M_ρ	Difference	95% c.i. for M_ρ	V_ρ	Difference
			Five Studies Per Meta-Analysis			
1	125	.336	.017	.267, .405	.007	.004
2	599	.302	-.017	.270, .335	.000	-.003
3	234	.295	-.024	.243, .347	.009	.006
4	566	.308	-.011	.275, .341	.001	-.002
5	929	.318	-.001	.292, .344	.000	-.003
Averages	491	.312	-.007		.003	.002
			Ten Studies Per Meta-Analysis			
1	336	.283	.036	.252, .314	.002	-.001
2	1124	.337	.018	.320, .353	.003	.000
3	658	.285	-.034	.263, .307	.001	-.002
4	384	.346	.027	.318, .373	.004	.001
5	1282	.272	-.047	.256, .288	.002	-.001
Averages	757	.250	.000		.002	-.001
			Fifteen Studies Per Meta-Analysis			
1	406	.313	-.006	.290, .336	.002	-.001
2	503	.315	-.004	.295, .336	.001	-.002
3	369	.310	-.009	.286, .334	.001	-.002
4	495	.344	.025	.324, .364	.003	.000
5	585	.322	.003	.303, .341	.001	-.002
Averages	472	.321	.002		.002	-.001

Note. The estimates of M_ρ and V_ρ for the 189 technical training schools were **.319** and **.003**, respectively.

Table 5

Summary of Population Parameter Estimates and Measures of Accuracy for Bare-Bones Meta-Analyses: Mechanical Composite-Final School Grade Relationships

Meta-Analysis	Avg. N Per Study	$M_{\hat{\rho}}$	Difference	95% c.i. for $M_{\hat{\rho}}$	$V_{\hat{\rho}}$	Difference
			Five Studies Per Meta-Analysis			
1	457	.346	.010	.310, .382	.000	-.007
2	295	.301	-.035	.255, .348	.000	-.007
3	152	.406	.070	.347, .465	.009	.002
4	349	.233	-.103	.189, .277	.000	-.007
5	148	.351	.015	.288, .414	.005	-.002
Averages	280	.327	-.009		.003	-.004
			Ten Studies Per Meta-Analysis			
1	1201	.326	-.010	.310, .342	.004	-.003
2	583	.342	-.006	.319, .365	.002	-.005
3	556	.250	-.086	.225, .275	.003	-.004
4	605	.251	-.085	.227, .274	.010	.003
5	550	.316	-.020	.292, .339	.003	-.004
Averages	699	.297	-.041		.004	-.003
			Fifteen Studies Per Meta-Analysis			
1	332	.380	.044	.357, .404	.002	-.005
2	208	.326	-.010	.295, .357	.008	.001
3	491	.342	.006	.322, .362	.003	-.004
4	817	.315	-.021	.299, .331	.008	.001
5	482	.275	-.061	.254, .296	.011	.004
Averages	466	.328	-.008		.006	-.001

Note. The estimates of M_ρ and V_ρ for the 190 technical training schools were **.336** and **.007**, respectively.

A STUDY OF THE KINEMATICS, DYNAMICS AND CONTROL ALGORITHMS FOR A CENTRIFUGE SIMULATOR

Yu-Che (Jack) Chen
Assistant Professor
Department of Mechanical Engineering

The University of Tulsa
600 South College Avenue
Tulsa, OK 74104-3189

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 THE KINEMATICS, DYNAMICS AND CONTROL ALGORITHMS FOR A CENTRIFUGE SIMULATOR

Yu-Che (Jack) Chen
Assistant Professor
Department of Mechanical Engineering
The University of Tulsa

Abstract

A preliminary study of the kinematics, dynamics and control algorithms for a centrifuge simulator is conducted in this research. The centrifuge is modeled as a three joint manipulator. It is shown that the centrifuge simulator is an underactuated mechanism where the number of joints is less than the number of degrees of freedom needed to be controlled at the end effector (or the seat in the case of a centrifuge). Algorithms for solving the joint velocity and joint acceleration are quite different from those used in conventional manipulators. Here, we study the feasibility of various approaches for solving the joint velocity and joint acceleration with the prescribed trajectory of the end effector (of the pilot). In order to command the end effector (or the seat in the centrifuge) to follow the prescribed trajectory, various optimal control algorithms are proposed for the motion control of the centrifuge.

1. Introduction

The study of the motion and trajectory of fighter aircraft via motion simulators has enabled pilots to quickly visualize the unusual flight environments and develop solutions for such environments. At the Armstrong Laboratory, a centrifuge has been used to emulate different flight scenarios and to study the effect on pilots and equipment of exposure of these unusual motion fields. The purpose of this research is to investigate aspects in the kinematics, dynamics and the control of the centrifuge.

We model the centrifuge as a three-joint robotic manipulator. Its characteristics in kinematics are identified and problems encountered are addressed. Feasible solutions for these problems are proposed. A study of motion control of the centrifuge is also conducted in this research. Various approaches to the controller design of the centrifuge are proposed.

2. The Kinematics of the Centrifuge

Generally, the joint configuration of a manipulator θ is related to its end effector's position and orientation x by a nonlinear mapping described as follows :

$$x=f(\theta), \qquad x\in R^m \text{ and } \theta\in R^n \tag{1}$$

In equation (1), we have $m=n$ for non-redundant manipulators and $m<n$ for redundant manipulators. A popular technique for maneuvering a manipulator is resolved motion rate control which relates the end effector velocity and joint velocities by differentiating equation (1) with respect to time

$$[\dot{x}]_{m\times 1}=[J]_{m\times n}[\dot{\theta}]_{n\times 1} \tag{2}$$

where $[J]=\partial f(\theta)/\partial\theta$.

For $m=n$, the mapping $\delta\theta \rightarrow \delta x$ is unique and the solution of $[\dot{\theta}]$ for a prescribed $[\dot{x}]$ can be uniquely found. For $m<n$, due to the extra degree of freedom, there are infinite number of control strategies for redundant manipulators and a general approach for determining $[\dot{\theta}]$ for a prescribed $\dot{x}$ is given as:

$$[\dot{\theta}]= [J]^{+}[\dot{x}] +([I]- [J]^{+}[J])\ \xi \tag{3}$$

where $[J]^{+}=[J]^{T}([J][J]^{T})^{-1}$ and ξ is a $n\times 1$ column vector usually used for optimizing the joint motion. Physically, $([I]- [J]^{+}[J])\xi$ corresponds to the self-motion where the combined joint motions generate no motion at the end effector.

A centrifuge is a typical underactuated manipulator where the number of joints is less than the number of degree of freedom at the end effctor ($m>n$ with $m=6$ and $n=3$). For such manipulators, equation (3) is not valid since

the Jacobian has a dimension 6×3 and $[J][J]^T$, a 6×6 matrix, is always singular. For a prescribed end effector velocity $[\dot{x}]$, there does not exist any solution for $[\dot{\theta}]$. The following discussion on the kinematics for such manipulators is separated in two parts - the levels of joint velocity and joint acceleration.

2.1. The minimum norm solution for the joint velocity

Define the following norm for the joint and end effector velocities:

$$L = ||\dot{x}_d - J\dot{\theta}||_w = (\dot{x}_d - J\dot{\theta})^T[\bar{W}](\dot{x}_d - J\dot{\theta}) \qquad (4)$$

where $\dot{x}_d$ is the desired end effector velocity and $[\bar{W}]$ is a weighing matrix of dimension 6×6. Since there is no exact joint velocity $\dot{\theta}$ that will generate $\dot{x}_d$, the joint velocity that minimizes L can be found by using $\partial L/\partial\dot{\theta}=0$. This process is shown in the follows:

$$0=\frac{\partial L}{\partial\dot{\theta}}=-2[J]^T[\bar{W}]([\dot{x}_d]-[J][\dot{\theta}]) \quad , \; or \; [J]^T[\bar{W}][\dot{x}_d]=[J]^T[\bar{W}][J][\dot{\theta}]$$

and thus the joint velocity minimizing L can be found as:

$$[\dot{\theta}]=([J]^T[\bar{W}][J])^{-1}[J]^T[\bar{W}][\dot{x}_d] \qquad (5)$$

Depending the task, different sets of joint velocities can be achieved by adjusting the weighing matrix $[\bar{W}]$. For example, the least square norm joint velocity can be found by using $[\bar{W}]=[I]$ in equation (5). On the other hand, the following expression of $[\bar{W}]$ leads to the joint velocity that matches the linear part of the end effector velocity:

$$[\bar{W}]=\begin{bmatrix} I_{3\times3} & 0 \\ 0 & 0 \end{bmatrix}$$

Note that equation (5) will fail to give a joint velocity if $[J]^T[J]$ becomes singular and the Jacobian will have a rank no greater than two in this case. Similar to the case for redundant manipulators where the scalar $\det([J][J]^T)$ is used to quantify the quality of the arm posture, we define the following expression for the manipulability measure:

$$\text{Manipulability measure} = \det([J]^T[J]) \qquad (6)$$

We now extend our analysis to the level of the joint acceleration.

2.2. The joint acceleration

By differentiating equation (2) with respect to time, we reach the following expression which relates the joint and end effector velocities

$$\ddot{x}=\begin{bmatrix}\ddot{x}_L\\ \ddot{x}_R\end{bmatrix}=[J][\ddot{\theta}]+[\dot{J}][\dot{\theta}]=\begin{bmatrix}J_L\\ J_R\end{bmatrix}[\ddot{\theta}]+\begin{bmatrix}\dot{J}_L\\ \dot{J}_R\end{bmatrix}[\dot{\theta}] \qquad (7)$$

Equation (7) can be used in various scenarios. In the first case where the angular trajectory $[x_R]$ is estimated (for example the Herbst Maneuvering described in [Repperger 1992]) and a closed form expression of $\dot{x}_R$ and $\ddot{x}_R$ can be reached, the joint velocity and acceleration can be found by using part of equations (5) and (7) as follows:

$$[\dot{\theta}]=([J]^T[\bar{W}][J])^{-1}[J]^T[\bar{W}][\dot{x}_d],\ [\bar{W}]=\begin{bmatrix}0 & 0\\ 0 & I_{3\times 3}\end{bmatrix}$$
$$[\ddot{\theta}]=[J_R]^{-1}[\ddot{x}_R]-[\dot{J}_R][\dot{\theta}] \qquad (8)$$

By substituting the joint velocity and acceleration in equation (8) into the following expression, the linear acceleration of the end effector can be found:

$$[\ddot{x}_L]=[J_L][\ddot{\theta}]+[\dot{J}_L][\dot{\theta}] \qquad (9)$$

On the other hand, if the joint velocity has been determined using equation (5) and the goal is to find the joint acceleration which minimizes the linear acceleration experienced by the pilot, then we can define the following quantity:

$$L_2=G^TG=\ddot{x}_L^TR_0^5R_5^0\ddot{x}_L=\ddot{x}_L^T\ddot{x}_L=\ddot{\theta}^T[J_L]^T[J_L]\ddot{\theta}+2\dot{\theta}^T[\dot{J}_L]^T[J_L]\ddot{\theta}+\dot{\theta}^T[\dot{J}_L]^T[\dot{J}_L]\dot{\theta}$$

where G is the linear acceleration experienced by the pilot. To minimize L_2, we can set $\partial L_2/\partial\ddot{\theta}$ to zero and solve $[\ddot{\theta}]$ in terms of $[\dot{\theta}]$ as described in the follows:

$$0=\frac{\partial L}{\partial\ddot{\theta}}=2[J_L]^T[J_L][\ddot{\theta}]+2([\dot{J}_L]^T[J_L])^T[\dot{\theta}]=2[J_L]^T[J_L][\ddot{\theta}]+2[J_L]^T[\dot{J}_L][\dot{\theta}]$$

and

$$[\ddot{\theta}]=-([J_L]^T[J_L])^{-1}([J_L]^T[\dot{J}_L])[\dot{\theta}]=-[J_L]^{-1}[\dot{J}_L][\dot{\theta}] \qquad (10)$$

With the numerical values of $[\dot{\theta}]$ calculated using equation (5), the numerical value of $[\ddot{\theta}]$ can be found using equation (10). Equation (10) can also be verified by setting $[\ddot{x}_L]=0$ in equation (7) and solve for the first half of equation (7). In other words, the joint acceleration calculated using equation (10) gives zero linear acceleration at

the pilot's position.

Similarly, if the goal is to minimize $[\ddot{x}]$ instead of $[\ddot{x}_L]$, we can define $L_3=[\ddot{x}]^T[\ddot{x}]$ and go through the same process to find the joint acceleration in terms of the joint velocity calculated using equation (5) as:

$$[\ddot{\theta}]_{min}=-([J]^T[J])^{-1}([J]^T[\dot{J}])[\dot{\theta}] \qquad (11)$$

Finally, for a desired end effector acceleration $[\ddot{x}_d]_{6\times 1}$ with the joint velocity calculated from equation (5), we can obtain from equation (7) the following expression:

$$[\ddot{x}_d]-[\dot{J}][\dot{\theta}] =[J][\ddot{\theta}] \qquad (12)$$

Again, there are three equations and six unknown variables in equation (12) and there will be no exact solution for $[\ddot{\theta}]$. Defining $L_4=||[\ddot{x}_d]-[\dot{J}][\dot{\theta}]-[J][\ddot{\theta}]||=([\ddot{x}_d]-[\dot{J}][\dot{\theta}]-[J][\ddot{\theta}])^T([\ddot{x}_d]-[\dot{J}][\dot{\theta}]-[J][\ddot{\theta}])$ and going through the same process for minimizing L_4, we can reach an expression for $[\ddot{\theta}]$:

$$[\ddot{\theta}]=([J]^T[J])^{-1}[J]^T([\ddot{x}_d]-[\dot{J}][\dot{\theta}])=-([J]^T[J])^{-1}([J]^T[\dot{J}])\dot{\theta}-([J]^T[J])^{-1}[J]^T[\ddot{x}_d]$$

Substituting equation (11) into the first term of the above expression, we have the following expression for $[\ddot{\theta}]$:

$$[\ddot{\theta}]=[\ddot{\theta}]_{min}-([J]^T[J])^{-1}[J]^T[\ddot{x}_d] \qquad (13)$$

2.3. Solving the joint velocity and acceleration to match a desired G acceleration

For flight simulation, it is usually desirable to find the joint velocity and acceleration that generate the data of $[G]$ acceleration collected by the sensor. The $[G]$ acceleration is related to the joint velocity and acceleration using the first half of equation (7) as:

$$[G]=R_0^{pilot} \quad [\ddot{x}]=R_0^{pilot}\,([J_L][\ddot{\theta}]+[\dot{J}_L][\dot{\theta}]) \qquad (14)$$

If the joint velocity is already calculated using equation (5), then the joint acceleration can be immediately found using equation (14) from the given data of G. Now, suppose that both the joint velocity and acceleration are to be determined and we define the following quantity:

$$L_5=([G']-([J_L][\ddot{\theta}]+[\dot{J}_L][\dot{\theta}]))^T([G']-([J_L][\ddot{\theta}]+[\dot{J}_L][\dot{\theta}])), \quad [G']=R_{pilot}^{\ 0}\,[G] \qquad (15)$$

Note that the expression for $[\dot{J}_L][\dot{\theta}]$ can be written in the following forms:

$$[\dot{J}_L][\dot{\theta}]=\begin{bmatrix}[\dot{\theta}]^T[H_x][\dot{\theta}]\\ [\dot{\theta}]^T[H_y][\dot{\theta}]\\ [\dot{\theta}]^T[H_z][\dot{\theta}]\end{bmatrix}$$

where $[H_x]$, $[H_y]$, $[H_z]$ are the Hessians of the positional forward kinematic functions. Differentiate L_5 in equation (15) with respect to $\ddot{\theta},\dot{\theta}$, set the results to zero, and use the above expression to reach the following equation:

$$0=\frac{\partial L_5}{\partial \ddot{\theta}}=-2[J_L]^T[G']+2[J_L]^T[J_L][\ddot{\theta}]+2[J_L]^T\begin{bmatrix}[\dot{\theta}]^T[H_x][\dot{\theta}]\\ [\dot{\theta}]^T[H_y][\dot{\theta}]\\ [\dot{\theta}]^T[H_z][\dot{\theta}]\end{bmatrix}$$

$$0=\frac{\partial L_5}{\partial \dot{\theta}}=-2\begin{bmatrix}[\dot{\theta}]^T[H_x]\\ [\dot{\theta}]^T[H_y]\\ [\dot{\theta}]^T[H_z]\end{bmatrix}[G']+2\begin{bmatrix}[\dot{\theta}]^T[H_x]\\ [\dot{\theta}]^T[H_y]\\ [\dot{\theta}]^T[H_z]\end{bmatrix}[J_L][\ddot{\theta}]+2\begin{bmatrix}[\dot{\theta}]^T[H_x]\\ [\dot{\theta}]^T[H_y]\\ [\dot{\theta}]^T[H_z]\end{bmatrix}\begin{bmatrix}[\dot{\theta}]^T[H_x][\dot{\theta}]\\ [\dot{\theta}]^T[H_y][\dot{\theta}]\\ [\dot{\theta}]^T[H_z][\dot{\theta}]\end{bmatrix} \qquad (16)$$

Equation (16) consists of six unknown variables of $[\ddot{\theta}]$ and $[\dot{\theta}]$ in six nonlinear equations. Solution of equation (16) requires iterative programming. However, it gives a general relationship between those $[\ddot{\theta}]$ and $[\dot{\theta}]$ that generate the desired linear acceleration at the pilot's position.

3. Dynamics and Control of the centrifuge

This section deals with the dynamics formulation and control algorithms of the centrifuge. The equation of motion for the centrifuge was derived in a previous research [Repperger 1994]. Basically, it can be expressed as follows:

$$M(\theta)\,\ddot{\theta} + C(\theta,\dot{\theta})\,\dot{\theta}=\tau \qquad (17)$$

The goal of control is to regulate the torque τ such that the end effector follows the desired trajectory in the joint space. Recently, "exact linearization" or "computed-torque" control of nonlinear system as a method for control design has attracted considerable interest, both in theory and in such practical field as flight control and robotics. The scheme of computed-torque controlled is shown in Figure 1.

From Figure 1, it can be realized that the idea of computed-torque control is to use state feedback to enable exact cancellation of nonlinear terms and factors followed by optimal control design for the simplified system. The

control law is generally given as:

$$u=\ddot{\theta}_r+K_v(\dot{\theta}_r-\dot{\theta})+K_p(\theta_r-\theta) \qquad (18)$$

The torque outputted to the manipulator can then be expressed as:

$$M(\ddot{\theta}_r+K_v(\dot{\theta}_r-\dot{\theta})+K_p(\theta_r-\theta))+C(\theta,\dot{\theta})\,\dot{\theta}=\tau \qquad (19)$$

Comparing equations (17) and (19), we obtained the error dynamics for the system:

$$\dot{e}=Ae,\ [A]=\begin{bmatrix} 0_{n\times n} & I_{n\times n} \\ -K_p & -K_v \end{bmatrix},\ e=\begin{bmatrix} \theta_r-\theta \\ \dot{\theta}_r-\dot{\theta} \end{bmatrix} \qquad (20)$$

Two main issues in the computed-torque control are addressed here. First, all dynamic parameters such as the moment of inertia and the mass of each link are just estimated versions of the real parameters. Mismatch between the real the estimated versions of the parameters generally leads to non-exact cancellation of the nonlinear terms and large tracking error. To remedy this drawback, adaptive computed-torque is generally used. Figure 2 shows the scheme of adaptive computed-torque control.

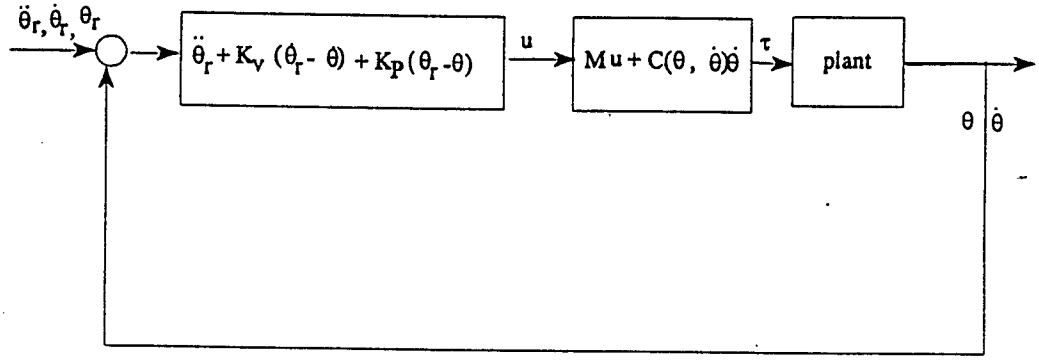


Figure 1. Block diagram for computed-torque control

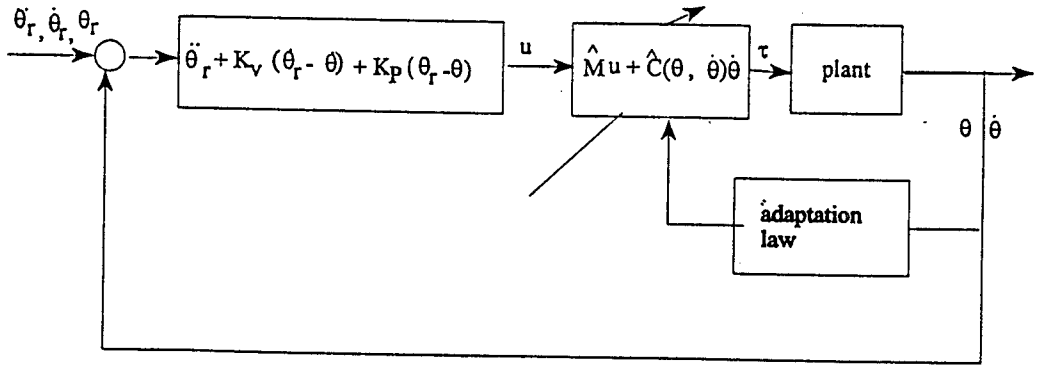


Figure 2. Block diagram for adaptive computed-torque control

The main difference in adaptive control from conventional control lies in the existence of the adaptation mechanism. It guarantees that the control system remains stable and the tracking error converges to zero as the parameters are varied. In adaptive control, the equation of motion is generally written in terms of the uncertain parameters. For example, the equation of motion of the centrifuge can be written in the following format:

$$M(\theta)\ddot{\theta}+C(\theta,\dot{\theta})\dot{\theta}=[W][\phi]=\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} & w_{36} & w_{37} \end{bmatrix} \begin{bmatrix} m_3 \\ I^*_{y1} \\ I^*_{x2} \\ I_{y2} \\ I_{x3} \\ I_{y3} \\ I_{z3} \end{bmatrix}$$

where

$w_{11}=L_2^2\ddot{\theta}_1,\ w_{12}=\ddot{\theta}_1,\ w_{13}=\ddot{\theta}_1,\ w_{14}=0,$

$w_{15}=s_2^2c_3^2\ \ddot{\theta}_1 - s_2s_3c_3\ \ddot{\theta}_2 + 2c_3^2\ s_2c_2\ \dot{\theta}_1\dot{\theta}_2 - 2s_2^2s_3c_3\ \dot{\theta}_1\dot{\theta}_3 - \cos(2\theta_3)\ s_2\ \dot{\theta}_2\dot{\theta}_3 - c_2s_3c_3\ \dot{\theta}_2^2$

$w_{16}=s_2^2s_3^2\ \ddot{\theta}_1 + s_2s_3c_3\ \ddot{\theta}_2 + 2s_3^2\ s_2c_2\ \dot{\theta}_1\dot{\theta}_2 + 2s_2^2s_3c_3\ \dot{\theta}_1\dot{\theta}_3 + \cos(2\theta_3)\ s_2\ \dot{\theta}_2\dot{\theta}_3 + c_2s_3c_3\ \dot{\theta}_2^2$

$w_{17}=c_2^2\ \ddot{\theta}_1 + c_2\ \ddot{\theta}_3 - 2s_2c_2\ \dot{\theta}_1\dot{\theta}_2 - s_2\ \dot{\theta}_2\dot{\theta}_3$

$w_{21}=w_{22}=w_{23}=0,\ w_{24}=\ddot{\theta}_2,$

$w_{25}=-s_2s_3c_3\ \ddot{\theta}_1 + s_3^2\ \ddot{\theta}_2 - s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_3 + 2s_2c_3\ \dot{\theta}_2\dot{\theta}_3 - c_3^2s_2c_2\ \dot{\theta}_1^2$

$w_{26}=s_2s_3c_3\ \ddot{\theta}_1 + c_3^2\ \ddot{\theta}_2 - 2s_2c_3\ \dot{\theta}_2\dot{\theta}_3 - s_3^2s_2c_2\ \dot{\theta}_1^2 + s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_3$

$w_{27}=s_2\ \dot{\theta}_1\dot{\theta}_3 + s_2c_2\ \dot{\theta}_1^2$

$w_{31}=w_{32}=w_{33}=w_{34}=0,$

$w_{35}=s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_2 + s_2^2s_3c_3\ \dot{\theta}_1^2 - c_3s_3\ \dot{\theta}_2^2$

$w_{36}=-s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_2 - s_2^2s_3c_3\ \dot{\theta}_1^2 + c_3s_3\ \dot{\theta}_2^2$

$w_{37}=c_2\ \ddot{\theta}_1 + \ddot{\theta}_3 + s_2\ \dot{\theta}_1\dot{\theta}_2$

M, C and ϕ are the actual values of the parameters and $\hat{M}$, $\hat{C}$, and $[\hat{\phi}]$ are the computed (or estimated) versions of the parameters. If the parameters are exact, then we have

$$[W(\ddot{\theta},\dot{\theta},\theta)][\phi]=\tau \qquad (21)$$

It should be noted that the formulation of equation (21) is not unique and it can be formulated in many different ways. Due to the mismatch between the parameters, the torque outputted to the manipulator becomes:

$$\hat{M}(\ddot{\theta}_r+K_v(\dot{\theta}_r-\dot{\theta})+K_p(\theta_r-\theta))+\hat{C}(\theta,\dot{\theta})\ \dot{\theta}=\tau$$

or

$$\hat{M}((\ddot{\theta}_r-\ddot{\theta})+K_v(\dot{\theta}_r-\dot{\theta})+K_p(\theta_r-\theta))+\hat{M}\ddot{\theta}+\hat{C}(\theta,\dot{\theta})\dot{\theta}=\hat{M}((\ddot{\theta}_r-\ddot{\theta})+K_v(\dot{\theta}_r-\dot{\theta})+K_p(\theta_r-\theta))+[W(\ddot{\theta},\dot{\theta},\theta)][\hat{\phi}] \quad (22)$$

Comparing equations (20) and (22), we can form the error dynamics as:

$$\ddot{e}+K_v\dot{e}+K_pe=\hat{M}^{-1}[W(\theta,\dot{\theta},\ddot{\theta})][\tilde{\phi}]\ ,\ [\tilde{\phi}]=[\phi]-[\hat{\phi}]$$

or

$$\dot{e}=Ae+B\hat{M}^{-1}(\theta)W(\theta,\dot{\theta},\ddot{\theta})[\tilde{\phi}],\ e=\begin{bmatrix}\theta_r-\theta\\ \dot{\theta}_r-\dot{\theta}\end{bmatrix},\ A=\begin{bmatrix}O_{n\times n} & I_{n\times n}\\ -K_p & -K_v\end{bmatrix},\ B=\begin{bmatrix}O_{n\times n}\\ I_{n\times n}\end{bmatrix} \quad (23)$$

In adaptive control, we usually update the parameters $[\phi]$ by a law given below:

$$[\hat{\phi}]=\Gamma[W]^T\hat{M}^{-1}B^TPe,\ \textit{where}\ \Gamma=diag(\gamma_1,\gamma_2,\cdots,\gamma_s) \quad (24)$$

where the subscript s is the number of uncertain parameters. The matrix P is determined by the Lyapunov equation shown below:

$$A^TP+PA=-Q \quad (25)$$

where Q is a positive-definite, symmetric matrix. The desingers' main task is to determine the matrix Q to obtain the matrix P and to decide the scalars γ_i, in the diagonal matrix Γ. With these parameters obtained, the adaptation law in equation (24) can be determined.

3.1. L-Q optimization for the linear feedback gains

A second issue in computed-torque control is to determine the gains for position and velocity feedback in matrix $[A]$ of equation (24). Usually these gains are determined by incooperating a performance index into the systems. A frequently used performance index which weighs the derivative of the error is given as:

$$I=\frac{1}{2}e^T(t_f)[K]e(t_f)+\frac{1}{2}\int_{t_0}^{t_f}[e^T(t)Q(t)e(t)+\dot{e}^T(t)R(t)\dot{e}(t)]dt\ ,[K]=\begin{bmatrix}K_p & 0\\ 0 & K_v\end{bmatrix}\ ,\ R(t)=\begin{bmatrix}0 & 0\\ 0 & R_0\end{bmatrix} \quad (26)$$

where Q is a real, nonnegative definite matrix and [K] and $[R_0]$ are real symmetric, positive definite matrices. It can be seen that the term $\dot{e}^T(t)R(t)\dot{e}(t)$ is simply the weighing of the joint acceleration, i.e., $(\ddot{\theta}_r-\ddot{\theta})^T\ [R_0](\ddot{\theta}_r-\ddot{\theta})$. Comparing equation (26) to the standard quadratic optimization where the Lagrangian is of the format $e^T(t)Q(t)e(t)+u^T(t)R(t)u(t)$, we learn that $u(t)=\dot{e}(t)$ and the control is in the sense of joint acceleration.

Sufficient conditions for the minimization of the performance index of equation (26) can be achieved by means of the Hamilton-Jacobi equation

$$\frac{\partial V}{\partial t} + \min_u H(e,u,\frac{\partial V}{\partial t},t)=0 \quad (27)$$

where the Hamiltonian of the system is defined as:

$$H(e,u,\frac{\partial V}{\partial t},t)=\frac{1}{2}[e^T(t)Q(t)e(t)+\dot{e}^T(t)R(t)\dot{e}(t)]+\frac{\partial V}{\partial e}\dot{e} \quad (28)$$

First differentiate equation (28) with respect to u (which is $\dot{e}$ in this case) and set the result to zero to find an expression of the optimal control law u^*. Then substitute the expression of u^* into equation (28) to get the expression for the optimal Hamiltonian as:

$$H^*(e,u^*,\frac{\partial V}{\partial t},t)=\frac{1}{2}e^T(t)Q(t)e(t)-\frac{1}{2}\frac{\partial V^T}{\partial e}S\frac{\partial V}{\partial e}+\frac{\partial V}{\partial e}F\ e\ ,\ S=\begin{bmatrix}0 & 0\\ 0 & R_0^{-1}\end{bmatrix}\ ,F=\begin{bmatrix}0 & I\\ 0 & 0\end{bmatrix} \quad (29)$$

Finally, choose the function $V(e(t),t)$ as shown below:

$$V(e(t),t)=1/2\ e^T(t)P(t)e(t) \quad (30)$$

Substitute equation (30) into equation (29) to reach the following Riccati equation

$$\dot{P}(t)+Q(t)-P(t)S(t)P(t)+P(t)F+F^TP(t)=0,\ P(t_f)=[K],\quad P(t)=\begin{bmatrix}P_{11} & P_{12}\\ P_{12} & P_{22}\end{bmatrix} \quad (31)$$

and the optimal control law becomes

$$u^*(t)=M(\theta)\{\ddot{\theta}_r+M^{-1}(\theta)C(\theta,\dot{\theta})\dot{\theta}+R_o^{-1}[P_{12}(t)\ (\theta_r-\theta)+P_{22}(t)\ (\dot{\theta}_r-\dot{\theta})]\} \quad (32)$$

When the matrices Q,S are constant and $t_f \to \infty$, the Riccati equation and the control law becomes

$$PF+F^TP-PSP+Q=0 \quad (33)$$

and

$$u^*(t)=M(\theta)[\ddot{\theta}_r+K_p\ (\theta_r-\theta)+K_v(\dot{\theta}_r-\dot{\theta})]+C(\theta,\dot{\theta})\dot{\theta} \quad (34)$$

where

$$K_p=R_o^{-1}P_{12}, \quad \text{and} \quad K_v=R_o^{-1}P_{22}. \tag{35}$$

The procedure here is to determine the matrix Q and thus find the matrix P from equation (33). Then, determine the matrix R_0 to obtain the feedback gains K_p and K_v. A practical choice of Q is to choose Q in the following format [Lewis 1993]:

$$Q=diag\{Q_p,Q_v\}, \quad \text{with } Q_p,Q_v \in R^{n\times n} \tag{36}$$

The formula for the optimal stabilizing gains becomes:

$$K_p=\sqrt{Q_pR_0^{-1}}, \; K_v=\sqrt{2K_p+Q_vR^{-1}} \tag{37}$$

3.2 Adaptive computed-torque control with torque optimization

In the approach discussed above, the energy associated with the torque τ(t) is not formally minimized and thus it is considered as a *suboptimal approach* with respect to the actual dynamics, although with respect to the error system e(t) and the control u(t), it is optimal. An optimal control approach that weights e(t) and τ(t) is discussed in the follows.

Let the state variables be the position and velocity error at the seat of the centrifuge:

$$x=\begin{bmatrix}\dot{\theta}-\dot{\theta}_r\\ \theta-\theta_r\end{bmatrix}=\begin{bmatrix}0 & -I_{n\times n}\\ -I_{n\times n} & 0\end{bmatrix} e \tag{38}$$

where the definition of the error e(t) is given in equation (23). The error dynamics of the manipulator can be written as:

$$\dot{x}=\begin{bmatrix}-M^{-1}(\theta)\;C(\theta,\dot{\theta}) & 0_{n\times n}\\ I_{n\times n} & 0_{n\times n}\end{bmatrix}x+\begin{bmatrix}-\ddot{\theta}_r-M^{-1}(\theta)\;C(\theta,\dot{\theta})\dot{\theta}_r\\ 0_{n\times n}\end{bmatrix}+\begin{bmatrix}I_{n\times n}\\ O_{n\times n}\end{bmatrix}M^{-1}(\theta)\tau$$
$$=\bar{A}(\theta,\dot{\theta})x+B_0(\ddot{\theta}_r,\dot{\theta}_r,\dot{\theta},\theta)+\bar{B}M^{-1}(\theta)\tau \tag{39}$$

It was shown in a previous approach that the applied torque affecting the kinematic energy is [Johansson 1990]:

$$\tau_k=M(\theta)\ddot{\theta}+\dot{M}(\theta,\dot{\theta})\dot{\theta}-\frac{1}{2}\frac{\partial(\dot{\theta}^TM(\theta)\dot{\theta})}{\partial\theta}=M(\theta)\ddot{\theta}+\frac{1}{2}\dot{M}(\theta,\dot{\theta})\dot{\theta}+N(\theta,\dot{\theta})\dot{\theta} \tag{40}$$

where $N(\dot{\theta},\theta)$ is a skew-symmetric matrix defined in equation (41) shown below and $N(\dot{\theta},\theta)\dot{\theta}$ represents the workless forces of the system.

$$N(\theta,\dot{\theta})=\frac{1}{2}\dot{M}(\theta,\dot{\theta})\dot{\theta}-\frac{1}{2}\frac{\partial(\dot{\theta}^T M(\theta)\dot{\theta})}{\partial\theta} \quad (41)$$

The work done on the system by the torque τ_k is given in the follows:

$$\int \tau_k^T\dot{\theta}\ dt=\int \dot{\theta}^T\ (M(\theta)\ddot{\theta}+\frac{1}{2}\dot{M}(\theta,\dot{\theta}))\ \dot{\theta}\ dt \quad (42)$$

To minimize the necessary torques, the control variable is defined in accordance with the integrand in equation (42):

$$u=[M(\theta),\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\theta)]\begin{bmatrix}\dot{z}_1\\ z_1\end{bmatrix}=[M(\theta),\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\theta)]\ T_1\ x \quad (43)$$

In equation (43), a new set of state variable $[z_1, z_2]^T$ is obtained by a transformation of the original state variables:

$$\begin{bmatrix}z_1\\ z_2\end{bmatrix}=T_0\ x=T_0\begin{bmatrix}\dot{\theta}-\dot{\theta}_r\\ \theta-\theta_r\end{bmatrix},\ T_0=\begin{bmatrix}T_1\\ T_2\end{bmatrix}=\begin{bmatrix}T_{11} & T_{12}\\ 0 & I_{n\times n}\end{bmatrix} \quad (44)$$

By setting $T_{11}=I_{n\times n}$ and $T_{12}=0$, the control algorithm becomes the traditional computed-torque control algorithm. The equation of motion in equation (39) can now be expressed in term of the control variable u:

$$\dot{x}=\bar{A}_1(\theta,\dot{\theta})x+\bar{B}_1 u \quad (45)$$

where

$$\bar{A}_1=T_0^{-1}\begin{bmatrix}-M^{-1}(\theta)(\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\dot{\theta})) & 0_{n\times n}\\ T_{11}^{-1} & -T_{11}^{-1}T_{12}\end{bmatrix}T_0\ ,\ \bar{B}_1=T_0^{-1}\bar{B}\ M^{-1}(\theta) \quad (46)$$

Now, defining a performance index PI and the lagrangian $L(x,u)$ in the follows:

$$PI=\int_{t_0}^{\infty}L(x,u)dt,\quad L(x,u)=\frac{1}{2}x^TQx+\frac{1}{2}u^TRu,\ Q=\begin{bmatrix}Q_{11} & Q_{12}\\ Q_{12}^T & Q_{22}\end{bmatrix}_{2n\times 2n} \quad (47)$$

Similar to the approach described in equations (27)-(30), sufficient conditions for the minimization of the performance index can be achieved by means of the Hamilton-Jacobi equation

$$\frac{\partial V}{\partial t} + \min_u H(x,u,\frac{\partial V}{\partial t},t)=0 \quad (48)$$

where the Hamiltonian of the system is defined as:

$$H(x,u,\frac{\partial V}{\partial x})=(\frac{\partial V}{\partial x})^T \dot{x}+L(x,u) \quad (49)$$

Choose the Hamilton principal function of optimization $V(x,t)$ as:

$$V=\frac{1}{2}x\ T_0 \begin{bmatrix} M & 0 \\ 0 & K \end{bmatrix} T_0\ x \quad (50)$$

Substitute equation (50) for $(\partial V/\partial t)$ and use equation (45) for $\dot{x}$ in the first term of equation (49) to obtain an expression for the Hamiltonian H. By setting $\partial H/\partial u=0$, we obtain the optimal control law u^* that minimizes the performance index PI [Johanson 1990]:

$$u^*= -R^{-1}\ \bar{B}^T\ T_0\ x= -R^{-1}\ T_1\ x= -R^{-1}\ [T_{11}\ (\dot{\theta}-\dot{\theta}_r)+T_{12}\ (\theta-\theta_r)] \quad (51)$$

where T_0 and the matrix K satisfy the following algebraic matrix equation

$$x^T \left(\begin{bmatrix} 0 & K \\ K & 0 \end{bmatrix} + Q - T_0^T \bar{B} R^{-1} \bar{B}^T T_0 \right) x=0 \quad (52)$$

Equation (52) and the algebraic Riccati equation are similar and the function $V(x,t)$ can be viewed as the aggregate of kinetic and potential energy inherent in a set of springs with a stiffness matrix K. The weighing matrices Q and R are usually chosen with Cholesky factors Q_1, Q_2, R_1 shown below:

$$Q=Q^T=\begin{bmatrix} Q_{11} & Q_{12} \\ Q^T_{12} & Q_{22} \end{bmatrix}=\begin{bmatrix} Q^T_1 Q_1 & Q_{12} \\ Q^T_{12} & Q^T_2 Q_2 \end{bmatrix},\ Q^T_1 Q_2+Q^T_2 Q_1-(Q^T_{12}+Q_{12})>0\ ,R=R^T=R^T_1 R_1>0 \quad (53)$$

The transformation matrix T_0 and the spring constant K can then be expressed in terms of Q and R as:

$$T_0=\begin{bmatrix} T_{11} & T_{12} \\ 0 & I_{n\times n} \end{bmatrix}=\begin{bmatrix} R^T_1 Q_1 & R^T_1 Q_2 \\ 0 & I_{n\times n} \end{bmatrix},\ K=K^T=\frac{1}{2}(Q^T_{12}+Q_{12})>0 \quad (54)$$

The function $V(x,t)$ of equation (50) is actually a Lyapunov function for the system. Under the optimal control law described by equation (51), it can be proved using equations (48) and (49) that the system is asymtopically stable

for $Q>0$ and $R>0$:

$$\frac{dV(x,t)}{dt}=\frac{\partial V}{\partial t}+\frac{\partial V}{\partial x}\dot{x}=-L(x,u^*)=-\frac{1}{2}x^T(T_0{}^T\bar{B}R^{-1}\bar{B}^TT_0+Q)x<0$$

The appropriate external torques τ^* to apply can be calculated in accordance with equations (45), (46) and the equation of motion (17).

$$\tau^*=M(\theta)(\ddot{\theta}_r-T_{11}^{-1}T_{12}(\dot{\theta}-\dot{\theta}_r)-T_{11}^{-1}M^{-1}(\theta)\ (-(\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\dot{\theta}))T_1x+u^*))+C(\theta,\dot{\theta})\dot{\theta} \quad (55)$$
$$u^*=-R^{-1}\bar{B}^TT_0x=-R^{-1}T_1x=-R^{-1}[T_{11}(\dot{\theta}-\dot{\theta}_r)+T_{12}(\theta-\theta_r)]$$

The control law is considerably simplified for a diagonal $T_{11}=t_{11}\ I_{n\times n}$ which is obtained for a special choice of Q and R. The control law becomes in the following format:

$$\tau^*=\frac{1}{t_{11}}(\ M(\theta)\ (t_{11}\ddot{\theta}_r-T_{12}(\dot{\theta}-\dot{\theta}_r)\ -\ (\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\dot{\theta}))T_1\ x+u*)+C(\theta,\dot{\theta})\dot{\theta})$$
$$=M(\theta)(\ddot{\theta}_r-\frac{1}{t_{11}}T_{12}(\dot{\theta}-\dot{\theta}_r))-\frac{1}{t_{11}}(\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\dot{\theta}))(\dot{\theta}-\dot{\theta}_r)-\frac{1}{t_{11}}(\frac{1}{2}\dot{M}(\theta,\dot{\theta})+N(\theta,\dot{\theta}))T_{12}(\theta-\theta_r)+u*)+C(\theta,\dot{\theta})\dot{\theta} \quad (56)$$

Comparing equation (56) with the traditional computed-torque with P.D. control given in equation (19), we found that equation (56) is similar to the computed-torque algorithms except that the position and velocity error is amplified by time-dependent gains.

Finally, with the simplified state transformation matrix $T_{11}=t_{11}\ I_{n\times n}$, the expression for the optimal torque can actually be written in the following:

$$\tau^*=\frac{1}{t_{11}}(W\hat{\phi}+W_0+u*) \quad (57)$$

where W_0 is the certain parameters in the equation of motion and $\hat{\phi}$ consists of those uncertain parameters in the equation of motion. For the centrifuge, we have the following formulation:

$$\mathrm{M}(\theta)\ddot{\theta}+\mathrm{C}(\theta,\dot{\theta})=[\mathrm{W}]\ [\phi]+[\mathrm{W}_0] \quad (58)$$

where

$$[W]=\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix},\ [\phi]=\begin{bmatrix} I_{y3}-I_{x3} \\ I_{z3} \end{bmatrix},\ [W_0]=\begin{bmatrix} (I^*_{y1}+I^*_{x2}+m_3L_2^2+I_{x3}c_3^2+I_{y3}s_3^2)s_2^2\ddot{\theta}_1+2(I_{x3}c_3^2+I_{y3}s_3^2)s_2c_2\dot{\theta}_1\dot{\theta}_2 \\ I_{y2}+I_{x3}s_3^2+I_{y3}c_3^2\ddot{\theta}_2-(I_{x3}c_3^2+I_{y3}s_2^2)s_2c_2\dot{\theta}_1^2 \\ 0 \end{bmatrix}$$

and

$w_{11}=s_2 s_3 c_3\ \ddot{\theta}_2 + 2s_2^2 s_3 c_3\ \dot{\theta}_1\dot{\theta}_3 + \cos(2\theta_3)\ s_2\ \dot{\theta}_2\dot{\theta}_3 + c_2 s_3 c_3\ \dot{\theta}_2^2$

$w_{12}=c_2^2\ \ddot{\theta}_1 + c_2\ \ddot{\theta}_3 - 2s_2 c_2\ \dot{\theta}_1\dot{\theta}_2 - s_2\ \dot{\theta}_2\dot{\theta}_3$

$w_{21}=s_2 s_3 c_3\ \ddot{\theta}_1 + s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_3 - 2s_2 c_3\ \dot{\theta}_2\dot{\theta}_3$

$w_{22}=s_2\ \dot{\theta}_1\dot{\theta}_3 + s_2 c_2\ \dot{\theta}_1^2$

$w_{31}=s_2\cos(2\theta_3)\ \dot{\theta}_1\dot{\theta}_2 - s_2^2 s_3 c_3\ \dot{\theta}_1^2 + c_3 s_3\ \dot{\theta}_2^2$

$w_{32}=c_2\ \ddot{\theta}_1 + \ddot{\theta}_3 + s_2\ \dot{\theta}_1\dot{\theta}_2$

An adaptation rule for updating the uncertain parameters was described in a previous work [Johansson 1990]:

$$[\dot{\hat{\phi}}]=-K_\phi W^T \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} \dot{q}-\dot{q}_r \\ q-q_r \end{bmatrix} = -K_\phi W^T [T_{11}(\dot{q}-\dot{q}_r)+T_{12}(q-q_r)] \quad (59)$$

Stability of the adaptation rule described in equation (59) has been proved [Johansson 1990]. The estimated version of the parameters were shown to converge to the real parameters.

4. Summary and discussions

The motion of a centrifuge is investigated in this report. On the aspect of kinematics, several characteristics of the centrifuge are identified and problems encountered are addressed. Feasible solutions for these problems are proposed. Results of these solutions are expected to help us to gain insight for the acceleration on the pilot induced by the specified trajectory of centrifuge. Future work on this aspect will be centered on numerical simulations for the proposed schemes to solve the joint velocity and acceleration.

A second set of the problem attached in this report is the design of the controller for the centrifuge. All algorithms proposed in this report are based on the concept of "exact-linearization" or "computed-torque" [Lewis] which cancels the nonlinear terms in the equation of motion by state feedback. The techniques of optimization with respect to the joint accelerations and joint torques are applied to determine the gains for the feedback position and velocity. Adaptation laws for updating the uncertain parameters in the equation of motion control are also proposed. Again, our future work will be focused on the simulation and implementation of these proposed algorithms.

5. References

1. Repperger, D.W., "A Study of Supermanuneuverable Flight Trajectories Through Motion Field Simulation of a Centrifuge Simulator", ASME Transaction on Dynamic Systems, Measurement and Control, Vol. 114, pp. 270-277, 1992.

2. Repperger, D.W., "Inverse Kinematic Control Algorithms with a Reduced Coriolis Component for Use in Motion Simulators",in Press, ASME Transaction on Dynamic Systems, Measurement and Control, 1992.

3. Repperger, D.W., Huang, M.Z., and Roberts, R.G., "Dynamic Controller Designer for A Fixed Base Motion Simulator", ACC, 1994.

4. Luo, G. L. and Saridis, G.N., "L-Q Design of PID Controllers for Robot Arms", IEEE Journal of Robotics and Automation, Vol. RA-1, No. 3, pp. 152-159, 1985.

5. Johansson, R., "Quadratic Optimization of Motion Coordination of Motion Coordination and Control", IEEE Tran. on Automatic Control, Vol. 35, No. 11, pp. 1197-1208,1990.

6. Lewis, F.L., Abdallah, C.T., and Dawson, D.M., Control of Robot Manipulators, Macmillian Publishing Company, New York, 1993.

7. Fu, K.S., Gonzalez, R.C., and Lee, C.S., Robotics-Control, Sensing, Vision, and Intelligence, McGraw-Hill Book Company, 1987.

THE BENCHMARK DOSE APPROACH FOR HEALTH RISK ASSESSMENT: TCE

DR. SHASHIKALA DAS
ADJUNCT FACULTY
DEPARTMENT OF PHYSICS

WRIGHT STATE UNIVERSITY
DAYTON OH 45431

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
WRIGHT PATTERSON AIR FORCE BASE, DAYTON , OH 45433

AUGUST 1994

THE BENCHMARK DOSE APPROACH FOR HEALTH RISK ASSESSMENT: TCE

Dr. Shashikala Das

Adjunct Faculty

Department of Physics

Wright State University

Dayton OH 45431

ABSTRACT

The Benchmark dose approach in health risk assessment is critically reviewed for noncancer endpoints. The algorithm for obtaining the benchmark dose for both quantal and continuous data is developed using SIMUSOLV software package prepared by the DOW Chemical company. The mathematical models used in modeling the dose-response relationship, the statistical methods used in testing the trends, goodness-of-fit and obtaining the upper confidence limits are presented. The benchmark dose for trichloroethylene (TCE) for reproductive endpoints data calculated using the developed algorithm is found to be 246 mg/kg/day for oral route and 57 ppm for inhalation route. Benchmark doses can be used in getting the reference doses by applying the uncertainty factors for setting up the regulatory standards for workers exposure to TCE or for environmental management purposes. Additional research is recommended using an improved algorithm that could substantially change the benchmark dose values calculated here.

I INTRODUCTION

In order to set up regulatory standards for human exposure to environmental noncarcinogenic toxicants, the EPA establishes a reference dose (RfD) or reference concentration (RfC), which is calculated by applying several uncertainty factors to the no-observed adverse-effect level (NOAEL). The use of the NOAEL for determining an RfD or RfC does not make full use of dose response data. NOAEL values depend on the dose levels tested and are not uniquely defined. The use of the NOAEL does not encourage improved experimental design. Since poorly designed experiments have less power to detect biological effects, they may yield higher NOAELs and hence higher RfD and RfC. Another important drawback of the NOAEL approach is that the risk of a NOAEL dose is not known and varies from case to case.

Because of these shortcomings in the NOAEL approach, an alternative method has been proposed by several authors (e.g. Crump (1984), Kimmel & Gaylor (1990)) in which the NOAEL is replaced by a benchmark dose (BMD), corresponding to a low level of risk of the order of 1% to 10%. These levels of risk can be estimated with adequate precision by mathematical modeling of dose-response data. A BMD is defined as a statistical lower confidence limit on the dose producing a predetermined level of adverse change in response compared to the response in untreated animals (control group). More specifically, the BMD is often defined as a lower 95% confidence limit estimate of dose corresponding to a 5% level of extra risk (ED05) over background level.

The BMD dose value for a toxin used for calculating RfD or RfC thus depends on these factors:

- The level of risk chosen.
- The chosen confidence limit.
- The quality of the experiments performed.

The format of the data recorded is important in determining the models chosen for determining the BMD. The noncancer health effects, which are of primary interest here, can be recorded in quantal (categorical) or continuous format. Examples of quantal response are the presence or absence of organ degeneration or the birth defects. Organ weight variation and serum enzyme levels are examples of

quantitative or continuous data . It is sometimes preferable to convert continuous data into quantal format for mathematical modeling. BMD should not depend on the mathematical models chosen to model the bioassay data, as the extrapolation to low doses far beyond experimental limits is not done in determination of BMD.

The risk estimation using the BMD approach for quantal data, where animals can be classified as with or without biological adverse effects (e.g. tumor or birth defects), has been discussed extensively by Crump (1984) using several mathematical models and dose response data for several toxicants. Crump, Allen and Faustman (1992) have also discussed in detail different models used for quantal and nonquantal data analysis along with the statistical techniques necessary to analyze data to obtain BMD. However, they did not attempt to estimate risk levels for continuous data. So the BMD values for these cases are not on a similar footing as the BMD values for the quantal data. Since for many chemicals one finds quantal as well as continuous responses when different endpoints are assessed for NOAEL or BMD, it is essential to treat all the data on equal footing to get meaningful estimate of BMD from different endpoints. Recently Gaylor and Slikker (1990,1994) Stiteler, Anatra-Cordone and Hertzberg (1993) and Kodell and West (1993) have proposed a technique for estimating the risk of an adverse effect for continuous data which will enable one to estimate BMD without the above mentioned problem.

In this report we present a brief critcal review of the benchmark dose methodology as applied specifically to noncancer endpoints, present the approach we have followed and then present the analysis of reproductive data for TCE obtained by the computer programs developed by us using the SIMUSOLV software package, introduced by the DOW Chemical company.

2 THE BENCHMARK DOSE METHODOLOGY

The determination of a reference dose (RfD), an estimate of daily oral exposure or an RfC for continuous inhalation exposure to a toxic chemical that is likely to be without an appreciable risk of an adverse effect during a human lifetime, involves three main steps:

- Selection of experiments and responses.
- Calculation of BMD instead of NOAEL from data.
- Application of appropriate uncertainty factors to BMD in order to obtain RfD or RfC.

We will discuss here only the first two steps. The calculation of RfD or RfC from BMD is discussed by Crump, Allen and Faustman (1992).

2.1 Selection of experiments and responses: The selection of experiments and responses for obtaining BMD is very important step. Several biological effect data sets for a toxic chemical may be available, but only the response data which satisfy the conditions given below need generally be chosen for dose response modeling.

- Overall good quality of experimental study
- Exposure route chosen to be one for which the RfD is required.
- Relevant adverse health effects which RfD is intended to cover.
- Statistical analysis shows significant trends in dose response.
- Response data at three or more dose levels including NOAEL and LOAEL.
- Critical studies which show toxic adverse effect at lowest dose level i.e. give lowest LOAEL.

The format of the data recorded is also important in determining the models chosen for determining the BMD. It is sometimes preferable to convert continuous data into quantal format for mathematical modeling. In order to do that, one first assigns a given level of response as adverse and then categorizes the animals according to whether the adverse biological effect is present or absent. The choice of adverse effect level response is a critical step in BMD determination. No unique criteria is established to choose this at present.

2.2 Calculations of BMD:

2.2.1 Statistical analysis of data

The trend tests and goodness-of-fit tests are performed before one calculates the BMD from selected dose -response data. If the dose response data fail to show any trend then one need not consider them for further analysis. Similarly if the goodness of fit test shows a poor fit of model to dose response data then one can choose another model which fits best to the data for BMD calculations. Some of these

tests are based on the maximum likelihood test or likelihood function ratio test (Crump et al 1992). For quantal data one can perform Mentel- Haenszel test (Haseman 1985) or Fisher's exact test as pairwise test (Bickel and Poksum 1977) can be used. For continuous data likelihood function ratio test or t-test may be applied to test the pairwise differences, when only group means and standard deviations are available.

2.2.2Statistical Modeling: The next step is to model statistically the selected dose-response data to calculate BMD. The various mathematical models commonly employed to fit the dose-response data are tabulated in Table 1. The quantal data should be modeled by only the quantal models. Quantal data gives the experimental doses, the total number of animals in each dose group, and the number of animals in each group with a response of interest. In the case of continuous data, one gets the experimental doses, the number of animals in each dose group, and the individual response of each animal in the experiment or the mean response in each group and the sample variance of the response in each group. The values of parameters of the model can be determined by various optimization techniques. The maximum likelihood method seems to give the best results and is described briefly below in other section. The goodness of the fit test and the χ^2 test are used to check how well the model is fitting the bioassay data. Next the upper 95% or 90 % confidence limit of response curve is calculated. The BMD for a given benchmark response (BMR), which is an extra response or risk above the control group with zero dose, is calculated from the 95% upper confidence limit (ED05) curve.

Statistical Assumptions in model selection

Statistical assumptions are very valuable in choosing the appropriate statistical model for analyzing the dose response data. In quantal models it is assumed that each subject is responding independently of all other subjects and all animals in a given dose group have equal probability of responding. In other words, it is assumed that quantal response results from binomial distribution of a dose-dependent number of responders. The continuous response data vary according to dose dependent normal probability of distribution. In other words here each subject assumed to respond independently of other subjects and does not have equal probability of response. These assumptions may not be valid in some cases such as the

developmental toxicity studies if responses within individual litters are considered or in species in which genetic polymorphism is observed, where the response cannot be approximated by any of the distributions mentioned above.

Biological Considerations in statistical modeling:

Sometimes the biological mechanism of the adverse effect caused by a toxicant is known. This knowledge may help in selecting the model for fitting the dose-response data. As an example, threshold versus nonthreshold models may be selected on the basis of biological mechanism. If the duration of exposure determines the response then the dose level as well as duration of response may need to be modeled for getting BMD.

Another important problem in fitting the mathematical model is the lack of fit observed at higher doses. Lack of fit occurs for several reasons such as :

- All the animals responding may not have been categorized properly.
- Interference in the response of interest by other forms of responses.
- The saturation of metabolic or delivery systems for the ultimate toxic substance.

The saturation of metabolic or delivery systems for the ultimate toxic substance may yield a plateau in responses at higher doses. In such cases one may replace the exposure doses by a tissue dose in the dose-response model. To achieve this, pharmacokinetic data on animals are used to estimate the internal dose delivered to the target tissue, for instance, using a physiologically based pharmacokinetic model. The BMD method then gives the internal measure of the dose for experimental animals. Human pharmacokinetic information can then be used to estimate RfDs for external doses. This method has the advantage of reducing the size of the uncertainty factors used in converting BMD to RfD. Another approach used to handle lack of fit when other types of response interfere with the endpoint under consideration at high dose levels is to drop the high level dose response data. This can be justified since other effects may be taking place at that dose level and the BMD is mainly determined from the shape of the dose response curve at low dose levels. In some cases the dose-response curve shows abnormal data at low doses due to excitation of some unknown mode which can not be fitted by a mathematical model. The

lack of fit here may be taken care of by disregarding the data at low levels. However, biological and toxicological justification is required for making decisions about dropping doses in modeling.

Goodness-of-fit-test:

To determine how good the model fits to the data one of the following tests can be performed. For quantal data χ^2 test is performed.

One calculates the quantity C defined by

$$C = \sum_{i=1,g} \frac{(Xi - Ni*P(di))^2}{Ni*P(di)(1-P(di))}$$

Where g = number of dose groups, and other quantities have been previously defined. The degree of freedom of this test = (g-No. of parameters of model). If, however, some of the parameters fall on the boundary of parameter space then the degree of freedom for that parameter is not lost. The value of C is then compared with quantile of a χ^2 distribution for the calculated degrees of freedom and if C equals or exceeds the quantile for (1-0.01) then it is concluded that the model did not fit the observed data.

For continuous data the F test is used. Defining $\underline{X}i$ and Si as

$\underline{X}_i$ = mean response in group i = $\Sigma_j Xij/Ni$

S_i^2 = sample variance of group i = $\Sigma_j (Xij-\underline{X}i)^2/(Ni-1)$

Ni being the number of animals in the group i. Let SSe, SSf be the sums of squares of errors due to the lack of fit and the experiment and dfe, dff ,be the number of degrees of freedom with experiment and fit.

dff = g -1 - Np, dfe = $\Sigma(ni-1)$, SSe = $\Sigma(Ni-1)S_i^2$, SSf = $\Sigma_i Ni(Xi-M(di))^2$,

where Np is the number parameters which do not lie on the boundary value of the parameter and the background parameter is estimated. Then the F test statistic is given by,

$$F = (SSf/dff)/(SSe/dfe)$$

The value of F is distributed according to F distribution with degrees of freedom dff and dfe. Again if the quantity F equals or exceeds the quantile corresponding to (1-0.01), then it is concluded that the the model did not fit the observed data.

2.2.3 Benchmark dose and measure of increased risk

For quantitative risk assessment using BMD approach, it is necessary to select a quantitative measure of an adverse response to determine the risk at a given exposure dose. The adverse response can be based on severity of response and (or) on the increased frequency of response. Since there is no clearcut definition of adverse response, one has to define the adverse response carefully based on the biological and social considerations because of its environmental , economical impact. In the case of quantal toxic responses such as cancer or birth defects, the definition of adverse effect may be self-evident and toxic events may be easily observed in individual subjects. However, one still needs to evaluate the the severity of the adverse effect in determining the BMD from these dose-response data. As an example, one may see different reproductive effects of a toxicant, but careful attention should be given to judge whether those effects are really unwanted before analyzing the data for BMD calculations. For a continuous or quantitative response, such as altered body weight, the adverse effect can not be as easily defined. One can not characterize risk in continuous response data directly. One could choose an adverse response level and determine the number of responders for each dose level. Because of these differences in quantal and continuous data, we will discuss next the determination of risk assessment and BMD in these two cases separately.

Quantal response Case: The calculation of the probability of response is rather straightforward in this case. If Xi is the number of subjects responding from a total number of subjects Ni in the i th dose group subjected to a dose di, then the probability of response of the group at dose di is given by

$$P(di) = Xi/ Ni$$

In order to measure increased response from the control group two methods have been proposed by Crump(1984). The additional risk (AR) over control group at dose d is given by

$$AR(d) = P(d) - P(0), \qquad (1)$$

where P(0) is the probability of response at zero or the background dose. The second measure called Extra Risk is defined by

$$ER(d)=\{P(d)- P(0)\}/\{1-P(0)\} \qquad (2)$$

Additional risk measures the additional proportion of total animals that would respond in the presence of dose. Whereas, the extra risk measures the fraction of animals that would respond to a dose d, who otherwise would not respond. Since the EPA uses extra risk measure in cancer risk assessment, we have also opted to use the extra risk measure to calculate the BMD.

Continuous or Quantitative Response

Analogous to quantal data , Crump(1984) has suggested two measures of increased response for quantitative data. If M(d) and M(0) are the mean responses at dose d and control group then the absolute difference, M(d)-M(0), is used to indicate the additional response. The second measure of increased response proposed by Crump (1984) is the extra response given by the absolute differences in the mean response (M(d)-M(0)) normalized by the background response M(0).These measures of response involve the fractional change in response rather than the absolute amount of change. These methods neglects the variability of response at control, and most important of all, does not give us a measure of risk or the probability of adverse response, at a given dose. If one calculates the BMD for continuous data using the dose-response data one will get an order of magnitude difference in BMD value, depending on whether one uses the actual response or the change in response (for example body weight), because the normalising response factor M(0) determines the scale of severity in BMD determination. The normalizing factor M(0) can be replaced by σ (0), the statistical error of mean response or the standard deviation of the control group, to measure extra adverse response as suggested by Crump (1984). This method has then the advantage that it measures the severity of response and the BMD is calculated on the basis of measured severity. Since the range of $2*\sigma$ (0) covers 68% of responders it is a better candidate for normalizing factor instead of M(o) and will not give an order of magnitude difference in BMD values in above mentioned cases . The BMD value determined thus will give us information about the severity of effect although it will not characterize the risk level .

Another method, suggested by Gaylor and Slikker (1990) to measure risk for quantitative response, involves the following steps:

- Fit a dose response model to observed continuous endpoints and obtain an estimate of M(d), the mean value of response at dose d and $\sigma(d)$, the standard deviation of the observations at that dose.
- Next define an abnormal response that shows an adverse biological effect.

Their suggested guideline to choose an abnormal response level is to consider the response to be normally distributed among the subjects and then find a level of response to which only 0.1% of the subjects may respond. For example, under the assumption of normal distribution, $\{M(0)-3.09*\sigma(0)\}$ response will be shown only by 1 in 1000 in the control group and could be defined to be adverse response level 'A'. One may also choose $\{M(0)-1.646*\sigma(0)\}$as an adverse response level 'A' since only 5% of the control group animals will show response less than the adverse level defined.

- The probability of adverse response at a dose d then can be calculated by calculating P(d) as

  P(d) = Probability (Response > A)

This is calculated as follows: Define $z = \{M(d)- A)/\sigma(d)$ and let 'α' be the area under standard normal distribution for a quantile $z(\alpha)=z$, then the probability of observing response greater than the adverse response A, P(d) is given by '$1-\alpha$'.

Thus using these calculated probabilities at different doses one can employ the mathematical models valid for quantal data to calculate the BMD using extra risk measures exactly in the same way as in a quantal response case. The advantage of the Gaylor and Slikker method is that it allows us to calculate BMD in both quantal and continuous data on common footing. This approach is illustrated with an example of neurotoxic effect analysis in rats and monkeys by Gaylor and Slikker (1994). The method of choosing the adverse effect level followed by Kodell and West (1993) is similar to that of Gaylor and Sliker (1994). In addition Kodell and West also discuss the estimation of upper confidence limits on the additional risk at a given dose over the background using two distinct algorithms and then use a Monte Carlo simulation technique to validate their methods for upper confidence interval. However, the problem with this approach is that one uses an arbitrary choice of level of adverse effect and the estimated risk is not accurate if the variance of the different dose group data is widely different than the control group. Moreover, if one has only the mean value of response and the variance instead of having the dose-response data for individual animals then in this approach one fits the mathematical model to only the mean

response to get m(d) and s. Thus one loses the details of the experimental information in this approach. Although for unequal variances one can apply a variance stabilizing transformation prior to the analysis, the definition of additional risk is not preserved under arbitrary transformation.

The alternative approach proposed by Stiteler et. al. (1993) for calculating the risk for continuous data involves transforming the continuous data to dichotomous or quantal form. In this method one does not have to select an arbitrary adverse response level. One makes use of both the mean value and the standard deviation of the response at all dose levels to change continuous response data to dichotomous form giving risk as a function of dose like quantal data. Then the quantal mathematical models can be fit to these data to give the BMD at added or extra risk of 5% above the background level. Thus, one can use common mathematical models and the same formalism to analyze the data for different endpoints in a consistent way. However their method of determining the responders to calculate risk does not sound reasonable in the case where the mean value of response in the control and treated group is same but the varience in treated group is widely different, since their method gives a high level of risk for this case. One can improve this approach by changing the definition of responders in their approach.

We have presented in this preliminary report the results of BMD calculations on reproductive effects of TCE only using the Crump (1984) approach and used the definition of extra risk in continuous data as

$$ER(d) = |\{M(d)-M(0)\}|/M(0) \qquad (3)$$

Complete BMD analysis based on the different methods discussed here for different noncancer endpoints for TCE will be published elsewhere. The work in that direction is under progress.

2.2.4 Calculations of the upper confidence limits on excess risk and the Benchmark dose:

After the parameters of the model are determined by maximum likelihood estimate, the upper confidence limit on the risk for a given dose and the lower confidence limit on dose for a given risk are calculated. Some of several approaches which can be taken for determining the confidence limits are discussed below.

(1)Asymptotic distribution of parameters: If the parameters of the model were estimated using the ordinary least squares method and the standard deviations of the parameters are calculated following the method discussed by Steiner, Rey and McCroskey (1990) (pages 5-56 to 5-61), then following Gallant (1987) (page 105), the upper and lower confidence interval for nonlinear parameters of univariate (One response variable) case is given by

$$q_u = q_i + t_{\alpha}/2s_i \text{ and } q_l = q_i - t_{\alpha}/2s_i, \quad (4)$$

where $t_{\alpha/2} = t^{-1}(1-\alpha/2; g-p)$ i.e. $t_{\alpha/2}$ denotes the upper $\alpha/2$ critical point of the t-distribution with g-Np degrees of freedom; g being the number of dose groups and Np is the number of parameters of the model, s_i is the standard deviation of the parameter . Using the upper confidence value of the parameter in the model the upper confidence limit of the risk Pd(di) can be calculated. The only draw back of the procedure is that the correlation of the parameters is not taken into account and hence the validity of the approach where the correlation of the parameters is large may be questionable. However, we tested this procedure for a nonlinear 4 parameter function and it gave almost identical results for the upper and lower 95% confidence limit value of the function as produced by SAS package.

(2) Asymptotic distribution of likelihood ratio Statistic: Following Cox and Hilkley (1974) Crump and Howel (1985) have found this approach convenient in dose-response analysis. In this method the log of the likelihood function LL(q) for the model is maximized by varying the parameters of the model to obtain the maximum likelihood estimate (MLE) of the parameters. Let Lmax be the maximum value of log likelihood function. Then the parameters are determined for which the loglikelihood function $LL(\theta)$ satisfies the relation

$$2 *(Lmax - LL(\theta)) = Z_{\alpha}^2 \quad (5)$$

Where $Z\alpha$ is the quantile for $100(1-\alpha)\%$ confidence interval of a standard normal distribution. Thus for getting 95% confidence interval $\alpha = 0.05$ and $Z\alpha = 1.645$ is chosen. Similarly for 90% confidence interval $a = 0.1$ and $z_a = 1.28$ are taken. These new parameters obtained from the constrained relationship then give the upper confidence limit of the risk when substituted in the mathematical model of the data.

The first step in calculations of BMD is to choose the benchmark respons level or the level of extra risk. Generally 10%, 5% or 1% values of extra risk are acceptable. If Pu(d) is the upper 95% confidence limit value of the risk at dose d, then the solution of the equation

$$(Pu(d)-Pu(0)) / (1-Pu(0))=0.05 \qquad (6)$$

gives the lower confidence value of dose d that gives extra risk of 5% and the BMD is then given by adding the threshold dose D0 to d. Similarly if m*(d) is upper 95% confidence limit value of response at dose d, then the solution of the equation

$$(m^*(d)- m^*(0)) / m^*(0) = 0.05 \qquad (7)$$

gives the lower confidence value of dose that gives extra response of 5% and the BMD then is given by adding the threshold dose value D0 to d.

3 BMD Determination : Reproductive and Developmental endpoints for TCE

A summary of the TCE experimental reproductive and developmental study information Tables 2-1 and 2-2 in ATSDTR (1992) , shows that the relevent data for BMD determination can be found in the following studies:

(1) NTP (1985) Which gives LOAEL of 750 mg/kg/day for decreased fetal and dam weight in mice fed microencapsulated TCE in food .

(2) Manson et al. (1984) Which gives 1000 mg/kg/day LOAEL for decreased dam weight and fetal mortality in rats fed TCE in corn oil by gavage.

(3) Zenick et al. (1984) Which gives LOAEL of 1000 mg/kg/day for impaired copulatory behavior of rats for TCE fed in corn oil by gavage.

(4) Land et al (1981) Which gave LOAEL of 2000 ppm for sperm morphology changes from TCE inhalation studies in mice.

The preliminary results of BMD calculation in some of these cases are summarized in table II. Here we have used the extra risk measure as defined in equations (2) and (3) for calculating the BMD. The Asymptotic distribution of parameters method was used for calculating the lower confidence limit on dose and equations (6) and (7) were solved for getting the BMD. Initially the computer programs written in

SIMUSOLV were validated by comparing results with those of Crump (1984) for data on carbon tetra chloride (continuous data) and ethylenethiourea (quantal data). Figure 1 shows the CPand CPR model fit curve to data of Land et al (1981) and the upper confidence limit curve for CPR model with the benchmark dose at 5% risk as an illustration of BMD calculations. The BMD value calculated from the inhalation data of Land (1981) at 5% risk level is smaller than the NOAEL reported for this study. On the other hand the values of BMD obtained from the study of Zenick et al are much higher than the reported NOAEL. The values of BMD based on male body weight differ by an order of magnitude depending on whether one considers the increase in body weight or the actual body weight. This difference is due to the m*(0) normalizing factor in the extra response equation (7) of continuous data.

We conclude the report with the statement that in this preliminary calculations we have not calculated the BMD on similar footing for the available continuous and quantal data and we do not know the risk involved BMD dose for the continuous data sets that are analyzed here. Thus before we could choose a BMD for calculating reference dose, we should choose an adverse effect level and convert the continuous data to quantal form or calculate risk using Kodell and West (1993) or Gaylor & Slikker (1994) approach mentioned above.

The author would like to thank Dr. H. Barton of Man Tech Environmental Technology Inc., for many helpful discussions, guidance and critical examination of this manuscript. The help of Dr. T. D. Rey of DOW Chemical Co. is specially acknowledged for suggesting the algorithm for calculating the confidence limit for BMD calculations. The author is also thankful to Dr. J. Byczkowski and C. Flemming of Man Tech Environmental Inc., for many helpful discussions and interest in the problem, Dr. J. Fisher of Armstrong Laboratory, Wright Patterson Air Force Base for encouragement and guidance and the US Air Force office of scientific research for the financial support through the summer faculty research program (SFRP) contract no F49620-93-C-0063.

REFERENCES

ATSDR (1992) , " Toxicological profile for TCE " , published by Agency for Toxic Substances and diseases Registry, ATSDR TP-92/19

Bickel P. and Doksum K.,(1977) Mathematical Statistics: Basic Ideas and Selected Topics., Holden-Day , Inc., Sanfransisco.

Cox and Hinkley, (1974) "Theoretical Statistics" , London Chapman and Hall

Crump and Howe, (1985) " A review of Methods for calculating confidence limits in low dose extrapolation" Toxicological Risk Assessment, Vol. 1.

Crump K. S., (1984) " A new method for determining allowable daily intakes" Fund. Appl. Toxicol. 4,854-871)

Crump K.S., Allen B.C. and Faustman E. M., (1992) " The use of the Benchmark dose approach in health risk Assessment", USEPA, ERG Technical Report, EPA Contract no. 68,C8-0036.

Gaylor D.W. and Slikker W. Jr., (1990) " Risk Assessment for Neurotoxic Effects", Neuro Toxicology 11,211-218

Gallant A. R., (1994) "Nonlinear Statistical Models" published by John Wiely and Sons Inc.,1987

Gaylor D.W. & Slikker W. Jr. " Modelling for Risk Assessment of Neurotoxic Effects", Risk Analysis 14, 337-338

Haseman J.,1984. Statistical issues in design ,analysis and interpretation of animal carcinogenicity studies. Environmental Health Perspect 58,385-392

Kimmel C.A. and Gaylor D. W., (1988) " Issues in qualitative and quantitative risk analysis for developmental toxicology", Risk Analysis 8,15-20(1988)8,15-20

Kodell R. L. and West R.W., (1993) " Upper Confidence limits on excess risk for quantitative Responses", Risk Analysis, 13, 177-182

Land P.C., Owen E.L., and Linde H. W., (1981) " Morphologic changes in mouse spermatozoa after exposure to inhalation anesthetics during early spermatogenesis", Anesthesiology 54,53-56

Manson J.M., Murphy M., Richardale N. and Smith M.K., (1984) " Effects of oral exposure to Trichloroethylene on Female Reproductive Function", Toxicology 32,229-242

NTP (1985) ,by George J.D. , Myers C. B., Lawton A. D. and Lamb J. C. 4 th, " Trichloroethylene : Reproduction and Fertility assessment in CD-1 Mice when administered in the feed", NTP 86-068

Steiner E. C., Rey T. D., and McCroskey P. S., SIMUSOLV: Modeling and simulation software. Vol. 2, Dow Chemical Company, Midland, MI 48674, pages 5-56 to 5-61.

Stiteler W. M., Anatra-Cordone M.and Hertzberg R.C., (1993) " Methods for converting continuous response data for dose-response modeling" Poster at DOD conference, Dayton, OH, 1993 Spring. See Also Stiteler W. M. , " Some statistical issues relating to the evaluations of risk levels at hazardous waste sites " , Proceedings of the 1990 Joint Statistical Meeting, Anaheim, CA, August 6, 1990

Zenick H. , Blackburn E., Richardale N. and Smith M.K. , (1984) " Effects of Trichloroethylene exposure on male reproductive function in rats", Toxicology, 31,237-250

TABLE I: List of Dose Response models used for estimating BMD

	Models for quantal data	
Quantal Polinomial Regression	(QPR)	$Pd = a_0 + (1-a_0)\{1- \exp(-a_1 d - a_2 d^{2} - \ldots\ldots - a_k d^{k})\}$
Quantal Wibul	(QW)	$Pd = a_0 + (1-a_0)\{1- \exp(-a_1 d^{a2})\}$
Log -Normal	(LN)	$Pd = a_0 + (1-a_0)\{N(a_1 + a_2 \log d)\}$
Probit	(PB)	$Pd = a_0 + (1-a_0)\{N(a_1 + \log d)\}$
Quantal quadratic	(QQR)	$Pd = a_0 + (1-a_0)\{1- \exp(-a_1 d^{*2})\}$
	Models for Continuous Data	
Continuous Polinomial Regression	(CPR)	$m(d) = a_0 + a_1 d^{*} + \ldots\ldots + a_k d^{*k}$
Continuous Quadratic Regression	(CQR)	$m(d) = a_0 + a_1 d^{*2}$
Continuous Power	(CP)	$m(d) = a_0 + a_1 d^{a2}$

N(x) = Normal cumulative distribution

$d^* = (d-d0)$, d0 being the threshold dose

$0 < ao < 1$

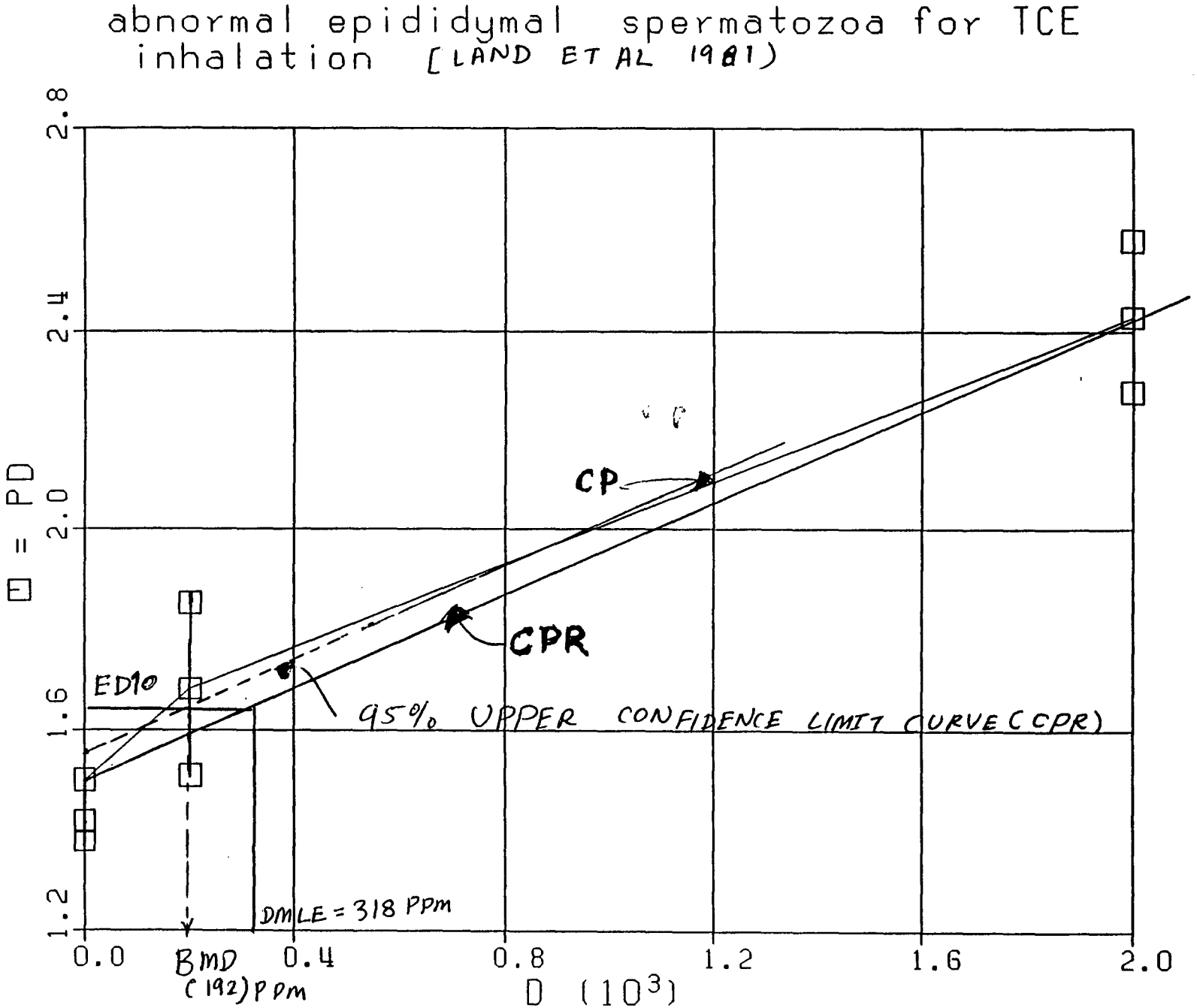

abnormal epididymal spermatozoa for TCE inhalation [LAND ET AL 1981)
E = PD
2.8
2.4
2.0
1.6
1.2
CP
CPR
ED10
95% UPPER CONFIDENCE LIMIT CURVE (CPR)
DMLE = 318 ppm
0.0
BMD (192) ppm
0.4
0.8
1.2
1.6
2.0
D (10³)

Table II

The summary of TCE reproductive and developmental Study along with calculated BMD values

	Species	route	Exposure	NOAEL (mg/kg/day)	LOAEL (mg/kg/day)	Endpoint	model	MLE(mg/kg/day) ED10	MLE(mg/kg/day) ED05	BMD(mg/kg/day) ED10	BMD(mg/kg/day) ED05
NTP (1985)	Mice	food	12 wk	375	750	*		*	*	*	*
Manson et al	Rat	OG	21 wk	100	1000	Fetal Survival	QPR	747	519	747	519
(1984)							QQR	751	524	747	524
						Dam Bodyweight	CQR	630	446	495	297
Zenick et al	Rat	OG	6 wk	100	1000	Male Bodyweight	CQR	987	698	894	608
(1984)						male (BW) increase	CQR	488	345	409	246
Land et al	Mice	IN	5 days	200 (ppm)	2000(ppm)	Sperm Morphology	CPR	318.6(ppm)	159(ppm)	192(ppm)	47(ppm)
(1981)							CP	152.6(ppm)	57.5(ppm)	152.6(ppm)	57(ppm)

OG corn oil gavage ; IN inhalation

* Analysis not completed awaiting reference data

NOISE AS A STRESSOR:

AN ASSESSMENT OF PHYSIOLOGICAL PARAMETERS

AND RADIOTELEMETRY EQUIPMENT AVAILABLE TO

STUDY ITS EFFECTS ON ANIMALS

Donald W. DeYoung
Chief, Biotechnology Support Section
University Animal Care
University of Arizona
Tucson AZ 85724

Final Report for:
Summer Faculty Research Program
Noise Effects Branch
Armstrong Laboratory

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

and

Noise Effects Branch
Armstrong Laboratory

September, 1994

NOISE AS A STRESSOR: AN ASSESSMENT OF PHYSIOLOGICAL PARAMETERS AND RADIOTELEMETRY EQUIPMENT AVAILABLE TO STUDY ITS EFFECTS ON ANIMALS

Donald W. DeYoung
University of Arizona
Tucson AZ 85724

Abstract

Noise as a stressor to animals is discussed with respect to the physiological parameters that can indicate that an animal is stressed. Known currently available radiotelemetry systems that can monitor some of these parameters are given and their sources are indicated. Data that can be reliably derived from parameters obtained by currently available radiotelemetry systems are presented. Possible pathology due to long-term stress as well as information for conducting animal studies, performing radiotelemetry implant surgery and anesthesia, and the potential complications of implant surgery are presented. Finally, recommendations for future study of noise as a stressor to animals are made.

Acknowledgments

Appreciation is extended to Major Robert C. Kull, Jr., Chief, AL/OEBN, my Effort Focal Point, and to all of the personnel of the Branch for their assistance and support; and for assimilating me into their group for the duration of my Summer Research Program opportunity. Appreciation is also expressed to AFOSR for sponsoring the worthwhile program. The cooperation of my colleagues at the University of Arizona is also acknowledged. Without them this effort would not have been possible.

Special thanks go to my wife, Karen, and sons, Rob and Michael, for their tolerance, understanding and encouragement during my absence to participate in this program.

NOISE AS A STRESSOR: AN ASSESSMENT OF PHYSIOLOGICAL PARAMETERS AND RADIOTELEMETRY EQUIPMENT AVAILABLE TO STUDY ITS EFFECTS ON ANIMALS

Donald W. DeYoung

Introduction

Sound is the propagation of pressure waves radiating from a vibrating body through an elastic medium (Sataloff and Sataloff, 1993). The physical attributes and corresponding psychological counterparts of sound are: amplitude--loudness, frequency--pitch, and complexity--timbre (Lipscomb, 1988). The intensity of sound waves decreases in inverse proportion to the square of the distance from the sound source (inverse-square law). The amplitude (intensity) of sound is measured in decibels, the frequency in octaves, while timbre permits the distinguishing between sounds (such as one musical instrument, airplane or voice from another).

Decibels afford a means of comparison or a ratio between two sound pressures. Decibels are generally described as being A, B, or C weighted; a fourth weighting system, D, has been described for aircraft noise (IEC, 1976). The A scale is most useful for low level sound; while the C scale is almost linear and the D scale may be more applicable for high levels (Lipscomb, 1988).

Noise is excessive or unwanted sound (Langdon, 1976). Noise can produce psychological and physiological effects. The psychological effects in people include: annoyance, disturbance, nuisance, etc. Some of these effects are attitudinal while others result in activities being interfered with. The physiological effects include temporary or permanent threshold shifts, hearing

loss and stress. Stress can result in alteration of body function and, if prolonged, pathology.

The effects of noise take two forms: those relating to the noise itself (external) and those relating to individual processes (internal). External factors include: loudness, sound character, number and duration of events, and time of day. Among the internal factors are individual sensitivity, and disturbance to sleep and activity patterns.

In man there is a tendency toward hearing loss at 90 dBA and above; the Occupational Safety and Health Administration (OSHA) requires employers to identify people who are exposed at or above an 85 dB 8 hour time weighted average by performing exposure measurements that include the 80 dB to 130 dB range. The effects of continuous noise ona number ofbody systems and a variety of species has been reviewed (Algers et al, 1978). While the effects ofsimulated aircraft noise on desert ungulates exposed to brief noise levels ranging from 92-112 dBA have also been studied (Krausman et al,1993). Factors that influence the noise level received from aircraft include: aircraft type, aircraft speed, the onset rate of the sound (which is related to aircraft speed), and the lateral offset between the subject and the aircraft's flight path, and the frequency of and length of time of exposure. These factors also play a role in the degree of startle, (surprise or alarm) that is experienced by the subject.

Noise and its effects are of interest to the U.S. Air Force because of the necessity for low-level flights for pilot training and proficiency maintenance. These flights occur over areas sparsely populated by humans but which may contain wildlife and especially endangered species. New flight paths or changes to existing flight paths require environmental impact

assessment and compliance with the National Environmental Policy Act. The effect of noise as a stressor to animals is important because this stress may acutely result in panic and chronically result in reduced marketibility or affect population dynamics by causing the animals to be less likely to survive and reproduce or by producing less healthy young and giving the young poorer care.

Stress and the Physiological and Pathological Responses to Stress

In the 19th century Claude Bernard indicated that health was dependent on a constant "milieu interieur." Later Walter Cannon replaced that term with "homeostasis" and Hans Selye indicated that stress was the body's non-specific response to requirements for change. A current definition of stress is: the effect of physical (environmental and external), physiologic or emotional factors (both internal) that induce alterations in the animal's homeostasis or adaptive state (Kitchen, et al, 1987). Thus, excessive or unwanted sounds (noise) can be a form of stress. The response may vary in accordance with the prior experience, sex, age, genetic profile and physiologic and psychological state (Kitchen, et al, 1987, Moberg, 1987). All animals have the same biologic responses with which to react to a stressor; but each may use a different type of biologic response; thus interanimal variability is also a component.

There are three categories of stress: neutral stress where the stimuli are not harmful and the responses neither benefit nor threaten the animal's well being; eustress where the stimuli are not harmful, but initiate responses that may be potentially beneficial; and distress where the stimuli may or may not be harmful but the animal is not able to adapt to the stimuli and may suffer

negative consequences as a result (Kitchen et al, 1987, and Breazile, 1987). Thus, stress by itself is not necessarily bad; but when carried to the extreme results in distress that should be avoided.

To cope with stress three major systems are employed: behavior, the autonomic nervous system and the neuroendocrine system. The behavioral response is often the simplest and most biologically economical (Moberg, 1987). It can involve increased alertness or movement to a different location. Different behavioral responses have been described for various common species in response to pain (Sanford et al, 1986). Although pain is usually considered much more severe than stress it is likely that there are subtle differences between species in their behavioral response to stress. These behavioral responses may indicate stress in an animal, but it has not been clearly demonstrated that these changes are necessarily harmful to an animal (Moberg, 1987). If the behavioral response does not alter the stress or if the stressor is of great intensity it is necessary that the animal advance to activating the remaining two systems (Moberg, 1987).

Both the autonomic and neuroendocrine systems are controlled by the hypothalamus (Moberg, 1985). The autonomic responses are fast and specific. They include alterations of function in many biological systems, such as: catecholamine release, changes in the cardiovascular and gastrointestinal systems, and exocrine gland secretions. Activation of the autonomic nervous system results in increases in heart and respiratory rate, vascular resistance, blood pressure and metabolism; as well as changes in gastrointestinal function and smooth muscle contraction. This is the classical fight or flight response as described by Cannon (Cannon, 1929). The neuroendocrine responses are largely mediated by hormones secreted by the

pituitary gland. These effects are longer term and influence reproduction, growth, metabolism, and immunity. These effects can be assayed by measuring the catecholamines epinephrine, norepinephrine, and dopamine. Other substances that are released and can be monitored include cortisol, ACTH, corticotropin releasing factor, atrial peptides, glucose, insulin, beta endorphins, growth hormone, prolactin, thyroid stimulating hormone, gonadotropins, antidiuretic hormone, cyclic AMP, vasopressin, renin, substance P, vasoactive intestinal peptide, neurotensin and neuropeptide Y (Muir, 1990, Moberg, 1987). Two concepts of stress were proposed: one that response to stress exhibits a generalized non-specific response to all stressors; and the other that there is a unique response to each type of stressor. A third view of stress was proposed (Engel, 1967): there is both a fight-or-flight and a conservation-withdrawal response to stress. Fight-or-flight involves an adrenal medulla response (increase in heart rate, arterial blood pressure cardiac output and changes in blood levels of glucose and lipids). The conservation-withdrawal response involves an adrenal cortex response, with chronic elevation of blood pressure, increased vagal activity and decreased gonadal steroids. The mode of response (fight-or-flight or conservation-withdrawal) that an individual chooses is primarily dependent on how the stressor is perceived by the individual. On the basis of this a model has been proposed (Kagan-Levi, 1974) that divides the response to stress as follows: 1) recognition of a threat to well-being, 2) the stress response, and 3) the biological consequences of stress. Using this as a pattern there occurs: a) a stimulus that is perceived as a threat, causing the b) biological defenses to organize, at which point a c) biological response occurs which results in either d) amelioration of the problem or a change in biological

function that is followed by a f) prepathological state and finally g) development of pathology (Moberg, 1985).

Examples of pathology as a result of stress include susceptibility to infectious disease, reduced growth, decreased reproduction, self-mutilation or other abnormal behavior, gastrointestinal ulceration and weight loss.

Stress can be acute or chronic depending on the length of time, magnitude and acclimation or habituation. The prepathologic and pathologic states are reached only if the response to stress is inadequate.

Moberg (1985) proposes that stress should not be examined only by its physiological responses but, also, that its effects on behavior, immunity, metabolism and reproduction should be considered.

Radiotelemetry and Stress Monitoring

Radiotelemetry is the measurement of data from a distance using radio waves to carry the parameters' signals. Biotelemetry is the transmission of physiological data from a transmitter, implanted in or located on the surface of a living subject, to the receiver.

Biotelemetry can be used for either physiological or ecological objectives. The physiological objectives are aimed at observing or explaining data from physiological parameters obtained under different situations. The physiological parameters can be used to study the behavior and health status of the animals. The data can also be used for ecological purposes (to monitor patterns of behavior and physiological and emotional status) (Lund, 1988).

Physiological parameters that do not require transduction (change from one form of energy to another, e.g., mechanical too electrical) are best suited for telemetry, and are the parameters that have been most commonly measured.

These parameters are biopotentials (events that produce measurable voltage) that require low power for filtration and amplification. These include: events of the cardiac cycle (ECG), eye movement (EOG), muscle activity (EMG) and brain waves (EEG), and temperature. Other useful parameters such as blood flow (by plethysmography) and blood pressure and any measurement requiring impedance (some respiratory measurements) are not as useful as they require transducers to make the signal useful and use more power. Catheter tip blood pressure sensors (with short life spans, ca 3.5 months) are currently available (See Appendix). Other data, such as respiratory rate, metabolic rate, etc., can sometimes be extrapolated from the signals of other parameters.

Power conservation is important for implanted units or those used on free ranging-animals as it is impractical or impossible to change batteries. The power source is an important factor in how far the transmitted signal will be received and for what time period the signal will be transmitted.

Telemetry has been used to study animal movements, position, group-behavior, mortality, and activity. Animal location has been accomplished by the use of satellites (GPS, Global Positioning System; and ARGOS) triangulation and by signal receipt by aircraft or personnel on the ground.

Historically, biotelemetry transmitter units have been custom made by "in-house" laboratories. Recently they have become more readily available commercially (see Appendix). Implantable systems have most frequently been single channel units. However, work is progressing on multiple channel, and ultrasonic and digitally encoded units (Cupal et al, 1989, Schild et al, 1989).

Transmitters that are placed on free-ranging animals or animals' that are in the environment need to be waterproof and shock resistant; additionally implanted transmitters must be biocompatible, so that minimal tissue reaction and no toxic reactions occur.

Transmitter size and configuration are important considerations. For smaller animals there is a 5% "rule", i.e., a transmitter should weigh no more than 5% of the animal's body weight or it will detrimentally affect energy, movement and foraging ability. Size has been demonstrated to result in ostracism and loss of position within an animal group (Bamberg et al, 1987). Reduction in size, streamlining and camouflaging may help reduce or solve this problem. Miniaturization will be of great value for both externally located and implanted telemetry units.

The Appendix contains a list of known sources of telemetry equipment.

Requirements for Animal Studies

Animal studies are subject to local, state, national and federal oversight, policies, guidelines, and regulations. Local and state issues may include those that are unique to an institution or geographical area, and may deal with subjects such as animal acquisition. Numerous national and federal groups impact animal research. These groups include the United States Department of Agriculture (USDA), which administers the Animal Welfare Act, the National Institutes of Health (NIH) which administers the Public Health Service (PHS) Policy, the National Science Foundation (which requires compliance with PHS Policy and extends that Policy to field research as well as laboratory research). Additionally there are voluntary organizations and societies that have guidelines or accreditation procedures for the use of

animals or animal research. Among these is the American Association for Accreditation of Laboratory Animal Care (AAALAC) a voluntary non-profit accreditation group. Accreditation by this group is a means of assuring the public and granting agencies that high quality animal care and use exist at the accredited facility. Facilities desiring accreditation submit an application and then receive a site visit for inspection, and review of the animal care and use program based on the NIH Guidelines for the Care and Use of Animals. The visiting team then submits a report to the AALAC Council recommending full accreditation, provisional accreditation with required correction items or withholding of accreditation. Following accreditation annual reports are submitted by the facility and the facility is re inspected at least every three years.

Additionally, several professional organizations have published field research or animal care guidelines. These include the American Society of Mammologists (Choate et al, 1987), the American Ornithologists' Union (Guidelines, 1988), the American Society of Ichthyologists and Herpetologists (Guidelines, 1988a) the American Fisheries Society (Guidelines, 1987a) and a Guideline for the Care and Use of Agricultural Animal in Agricultural Research and Teaching (Consortium, 1988).

The Animal Welfare Act and Public Health Policy both require the establishment of an Institutional Animal Care and Use Committee to review research protocols that include animal use and oversee the institutional animal care and use program. The NIH Guidelines and the AWA prescribe animal acquisition, care, and housing standards and require that a literature search using appropriate data bases (e.g., AGRICOLA) be performed. The AWA requires periodic facilities and records inspections by USDA personnel. The other guidelines

provide recommendations for the humane care and use of fish, birds or reptiles and amphibians, and farm animals. The farm animal guidelines are becoming accepted as standards in the same manner as the NIH Guidelines.

Implant Surgery and Complications of Implants

Animals should be stressed as minimally as possible before anesthetic induction. Restraint, immobilization and anesthetic agents appropriate for the species should be used. Special care should be taken when anesthetizing ruminants to prevent reflux and aspiration of ruminal contents. Acutely, this can cause reflux laryngospasm and bronchial constriction. Chronically, the result can be aspiration pneumonia. These life threatening complications can be avoided if the animal's head is kept at a level above the rumen and by passing a flexible tube down the esophagus and into the rumen to relieve any gas accumulation and to allow passage of liquid into a container.

Aseptic surgical techniques should be followed in appropriate surgical facilities. These include a wide surgical area clip and a three scrub surgical preparation technique (alternating disinfecting solutions with alcohol), appropriate surgical field draping to prevent contamination and resultant infection. Also included is the use of a surgical cap, mask, gown and gloves after a scrub of the surgeons' hands and arms. Instruments should be steam sterilized initially and between surgeries. If steam sterilization is not possible instruments should be cleaned, to remove debris, and chemically sterilized using freshly prepared sterlilization solution (e.g., chlorine dioxide) and not just a disinfecting solution. Telemetry transmitters should be sterilized by either gas sterilization (ethylene oxide) or chemical sterilization with freshly prepared chlorine dioxide solutions.

Disinfectant solutions (e.g., chlorhexidine or povidone iodine) should not be used for sterilization.

When surgery must be performed in conditions other than a dedicated surgery it is best to locate an enclosed structure in which to perform the surgery and to avoid performing the surgery under field conditions (in the open). Such a structure can be cleaned and will provide adequate protection from the elements and environmental contamination. The animals can be transported to the shelter by trailer or other vehicle. However, in some instances it may be necessary to perform the surgery in an outdoors setting.

The placement of a surgically implanted transmitter and surgical incision depend on the requirements of the project. Generally, it is preferable to not place the device subcutaneously. In this position it is subject to rubbing and placing pressure on the skin resulting in necrosis and extrusion of the transmitter through the skin. If the transmitter must be placed subcutaneously it is preferable to avoid a ventral or lateral location, as the transmitter is more susceptible to rubbing, and to pressure from being laid on (resulting in necrosis); also the skin is generally tighter in these areas. Suitable subcutaneous locations may include the base of the neck (in the area between the shoulder blades and adjacent) as there is more loose skin in these locations.

Abdominal implants are generally satisfactory. The incisions can be either ventral mid-line or lateral (flank). The implant should be placed more laterally than ventrally where the weight of the abdominal organs will rest on it. The transmitter can be enclosed in nylon or polyester fabric and sutured to the body wall with non-absorbable suture material to anchor it and reduce the chance of transmitter movement.

It is advisable to consider the administration of antibiotics to the surgical patient. If they are given, they should be administered prior to surgery so a therapeutic level is attained as soon as possible. If the surgical subject is wild and is to be turned loose after surgery the antibiotics should be long-acting as repeat injections will be impossible to give.

Surgery should be performed by well-trained, qualified individuals and should follow all applicable local, state, and federal guidelines and regulations.

There are several complications of implant surgery. The primary complications of subcutaneously implanted transmitters are infection and pressure necrosis of the skin. Lead wires that exit the skin are subject to trauma (breakage) and descending infection tracts. Intra-abdominal transmitter implants are susceptible to several complications. In armadillos and beavers these included small intestine incarceration and necrosis, adhesions to the falciform ligament, greater omentum and small intestine (Herbst, 1991; Guynn et al, 1987). Complications following heart-rate transmitter implantation were described by Wallace et at (1992). These included mortalities due to peritonitis, long term kidney failure, aspiration of rumen contents post-operatively, and incorporation (engulfment) of transmitters into the rumen or abomasum (presumably from being placed too ventrally).

Complications have also been reported with externally located (harness or collar) transmitters. Wild mallard ducks with transmitters attached by harnessing, gluing, or suturing have been reported to feed less while resting and preen more and lay smaller eggs and clutches than birds without transmitters (Pietz et al, 1993). Fallow deer fitted with collar transmitters were not accepted by other animals for one day in small enclosures (9 ha), for about 5 days in a 355 ha game park and had no social contact with uncollared

animals of the same species in the wild. During rutting the collars became too tight and the deer could not roar and consequently were unsuccessful during that breeding period. A previously territory holding buck that was fitted with a collar relinquished his position for that rutting season (Bamberg et al, 1987). In addition to changes in physical appearance transmitters that are externally located may become lodged or caught on physical structures, such as branches, while transmitters held in place by harnesses, sutures or adhesives may be shed or lost by the host.

References

Algers, B., I. Ekesbo, and S. Stromberg. 1978. The impact of continuous noise on animal health. Acta Vet. Scand. Suppl. 67:1-26.

Bamberg, F. B.,and Schwartau, B. 1987. The use of collar transmitters in biotelemetry in fallow deer. Z. Jagdwiss. 32:71-77.

Breazile, J. E. 1987. Physiologic basis and consequences of distress in animals. J.A.V.M.A. 191:1212-1215.

Cannon, W. B. 1929. Bodily changes in pain, hunger, fear and rage: an account of recent researches into the function of emotional excitement. Appleton, New York, New York.

Cupal,J..J., and R. W. Weeks. 1988. Digital encoding techniques for the telemetering of biologiical data. Pages 39-50 in C. J. Amlaner, Jr., ed. Biotelemetry, Proceedings of the tenth International Symposium on Biotelemetry. The University of Arkansas Press, Fayetteville.

Engel, G. L. 1967. A psychological setting of somatic disease: the "giving up-given up" complex. Proc.R. Soc. Med. 60:553-555.

Guynn, Jr., D. C., J. R. Davis, and A. F. Von Recum. 1987. Pathological potential of intraperitoneal transmitter implants in beaver. J. Wildl. Manage. 51: 605-606.

Henderson, D., M. Subramaniam, and F. A. Boettcher. 1993. Individual susceptibility to noise-induced hedaring loss: an old topic revisited. Eye and Hearing. 14:152-168.

Herbst, L. 1991. Pathological and reproductive effects of intraperitoneal telemetry devices on female armadillos. J. Wildl. Manage. 55:628-631.

International Electrotechnical Commission. 1976. Frequency weighting for the measurement of aircraft noise (D-weighting). I.E.C. Publication 537, Geneva.

Kitchen, H. 1987. Panel Report on the Colloquim on Recognition and Alleviation of Animal Pain and Distress. J.A.V.M.A. 191:1186-1191.

Krausman, P. R., M. Wallace, M. E. Weisenberger, D.W. DeYoung, O. E. Maughn. 1993, Effects of Simulated aircraft noise on heart-rate and behavior of desert ungulates. USAF, AL/OE-TR 1993-0185. 78 pp.

Langdon, F. J., 1985. Noise annoyance. Pages 143-176 in W. Tempest, ed. The Noise Handbook. Academic Press, London.

Lipscomb, D. M. 1988. What is this thing called "noise"? Pages 7-33 in D. Lipscomb, ed. Hearing Conservation in Industry, Schools, and the Military. Little, Brown and Co. Boston.

Moberg, G. P. 1987. Problems in defining stress and distress in animals. J.A.V.M.A. 191:1207-1211.

Moberg, G. P. 1985. Biological response to stress: Key to assessment of animal well-being? Pages 27-49 in G. P. Moberg, ed. Animal Stress. American Physiological Society, Bethesda.

Muir, W. W.. 1990. Editorial: The equine stress response to anaesthesia. Equine Vet. J. 22:302-305.

Pietz, P. J., G. L. Krapu, R. J. Greenwood, and J. T. Lokemoen. 1993. Effects of harness transmitters on behavior and reproduction of wild mallards. J. Wildl. Manage. 57: 696-703.

Sanford, J. 1986. Guidelines for the recognition and assessment of pain in animals. Vet. Rec. 118:334-338.

Sataloff, R. T., and J. Sataloff, 1993. The physics of sound. Pages 7-21 in R. T. Sataloff, and J. Sataloff, eds. Occupational Hearing Loss, 2nd ed. Marcel Dekker, Inc., New York.

Schild, J. H., P. H. Peckham and M. R. Neuman, 1989. A multichannel implantable transmitter for acquisition of physioogical data. Pages 430-443 in C. J. Amlaner, Jr., ed. Biotelemetry X, Proceedings of the Tenth International Symposium on Biotelemetry. The University of Arkansas Press, Fayetteville.

Wallace, M. C., P. R. Krausman, D. W. DeYoung, and M. E. Weisenberger. 1992. Problems associated with heart rate telemetry implants. Desert Bighorn Counc. Trans. 36:51-53.

The balance and final version

of this final report

can be obtained from:

AL/OEBN

2610 7th St.

Area B

Wright-Patterson Air Force Base,

Ohio 45433-7901

MEMORY FOR SPATIAL POSITION AND TEMPORAL OCCURRENCE OF DISPLAYED OBJECTS

Addie Dutta
Assistant Professor
Department of Psychology

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

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

July 1994

MEMORY FOR SPATIAL POSITION AND TEMPORAL OCCURRENCE OF DISPLAYED OBJECTS

Addie Dutta
Assistant Professor
Department of Psychology
Rice University

Abstract

Two experiments were conducted to assess memory for spatial and temporal occurrence attributes of visually displayed stimuli. A cueing procedure was used in which the subject was told whether to respond to spatial position or temporal occurrence on a trial by trial basis. The experiments examined whether the ability to recall the time of occurrence or spatial position of previously seen objects was a function of the relation between time and place of occurrence of the to-be-remembered items and whether it was a function of the time at which a cue was presented to indicate whether temporal or spatial occurrence was to be recalled. The results support the hypothesis that neither spatial location nor temporal occurrence can be completely ignored even when task demands are such that performance suffers from attendance to both attributes. Surprisingly, subjects were unable to ignore the occurrence attribute that was irrelevant on a given trial even when informed about which attribute to attend to before the presentation of the stimuli. Implications for the design of displays and for theories of memory representation are discussed.

MEMORY FOR SPATIAL POSITION AND TEMPORAL OCCURRENCE OF DISPLAYED OBJECTS

Addie Dutta

Introduction

Successful crew system performance often depends on the information processing capability of human operators. Some tasks, such as that of the Airborne Warning and Control Systems (AWACS) weapons director, depend heavily on the ability of the operator to remember and update large amounts of displayed information (Klinger, Andriole, Militello, Adelman, Klein, & Gomez, 1993). In dynamic tasks such as this one, the operator must process and remember not only the objects that are presented on the display screen, but also their spatial positions and times of occurrence. For example, it may be necessary to note where two planes are in relation to each other, when each appeared in the monitored area, and how quickly the positions of each of the planes are changing. Because adequate performance depends on the rapid processing and recall of such information, it is vital that the nature of spatial and temporal location processing be understood. It is also important to examine further attributes of the displayed information that facilitate the recall of both item and position information. However, previous work has dealt primarily with global aspects of performance in complex tasks such as that of the weapons director, with relatively little emphasis on the component processes that contribute to successful performance (Klinger et al., 1993).

In the task described above, many attributes of stimuli must be coded and remembered. Because the displays are large and complex, it is impossible for the weapons director to apprehend and follow all of the relevant attributes simultaneously. Of critical importance, then, is an understanding of how such attributes as item identity (e.g., plane or helicopter), color (e.g., as used to code "friend" or "foe" status), position (e.g., altitude and flight path), and time at which the object enters and leaves the monitored airspace, as well as the rate of movement, are coded and represented in memory. Much

controversy exists in the memory literature on how different attributes of stimuli are represented and processed (Jones, 1987). Examples of both independence (e.g., Stefurak & Boynton, 1986) and dependence (e.g., Well & Sonnerschein, 1973) in recall of multiple stimulus attributes have been found. Moreover, theoretical accounts of how multiple attributes of remembered stimuli are represented differ according to whether integration (e.g., Massaro, Weldon, & Kitzis, 1991) or independence (e.g., Jones, 1976) is assumed. In short then, additional work assessing the representation of multiple attributes is warranted. Because spatial position and temporal occurrence seem to have a special role in attentional orienting and selection (e.g., Keele, Cohen, Ivry, Liotti, & Yee, 1988) and are attributes that play a key role in situational awareness, the present experiments concentrate on recall of where and when stimuli are displayed.

Experiment 1

Dutta and Nairne (1993) showed that when subjects are required to respond to both the spatial and temporal occurrence of two stimuli, responses are faster and more accurate when the two dimensions correspond such that the stimulus presented first in time is also presented in the top position of the display and the second stimulus is presented in the bottom position of the display than when the spatial positions of the first and second stimuli are reversed. This so-called congruity effect was found to hold in experiments which required a speeded response to temporal position coupled with an unspeeded response to spatial position (Experiment 3) or a speeded response to spatial position coupled with an unspeeded response to temporal position (Experiment 4). The present experiments investigate whether such a finding will be obtained when only one dimension must be responded to on a given trial. If subjects are able to attend selectively to either spatial or temporal location when they know in advance which dimension is relevant, reaction time to recall the relevant (i.e., cued) dimension should be uninfluenced by the value of the irrelevant dimension. This hypothesis was tested by comparing three cueing conditions. In the prior cue condition, the cue "T" or "S" was presented 2 s prior to the stimulus display to inform subjects whether to respond to the temporal or spatial occurrence of the probed stimulus, respectively. In the immediate condition, the "T" or "S" cue was presented

immediately after the stimulus display but 1 second before the probe stimulus. Finally, in the delayed condition, the cue was presented 2 seconds after the stimuli but before the probe stimulus.

Several effects are of interest. One such effect is that of cue type. Earlier studies show mixed results regarding the relative dominance of spatial over temporal processing (e.g., Dutta & Nairne, 1993). By comparing responses to temporal and spatial position within a single experiment, it is possible to compare the relative ease of recall of the two dimensions as well as the extent to which recall of one dimension is influenced by the other. The effect of cue presentation time (prior to, immediately after, or 2 s after the stimulus presentation) is of interest for several reasons. As discussed above, comparison of the prior cue condition with the other conditions addresses how effectively subjects can attend to just one of the occurrence dimensions of the stimuli. By comparing performance in the immediate and delayed cue conditions, we can address additional questions regarding the nature of the rehearsal and maintenance of the stored information. Moreover, an examination of the influence of the nature of the variation on the irrelevant dimension (i.e., the correlation between time and place of occurrence) may reveal whether the occurrence dimensions are represented in an integrated or separate manner.

Method

Subjects. Air Force Basic Trainees volunteered to participate in the experiment as part of a three-hour testing session. Although 75 trainees were tested, data from 15 subjects had to be dropped due to a failure to comprehend instructions as evidenced by chance accuracy in at least one condition. Although this failure rate is high, it is not too surprising since approximately 40 subjects were tested at one time, making it difficult to ensure that instructions were followed. The chance of failure did not depend on the experimental condition received first.

Apparatus. The experiment was conducted on 286 IBM-compatible computers equipped with EGA monitors. The testing room contained approximately 60 computers, of which at most 44 were used for this experiment. Each testing station was isolated by solid partitions to each side and in back of the computer, making it unlikely that a given subject would view the performance of others. The room was

moderately lit and each subject was seated so that the viewing distance to the computer screen was approximately 50 cm.

Stimuli. Stimulus presentation and response collection were controlled by a program written using the Micro Experimental Laboratory (Schneider, 1988). The stimuli were the standard text characters "#" and "@"; each character measured 5 mm in height and 3 mm in width. The "+" was used as a fixation point and was 3 mm in height and width. The fixation point was centered on the screen, and the two stimuli were separated from the fixation point by 11 mm. The display was vertical, with the stimuli shown above and below fixation. Based on the average viewing distance of 50 cm, the visual angle subtended by the complete display was 4.0 deg. On each trial, a "T" (for "time") or "S" (for "space") was presented to cue the subject to respond either to the time of presentation or spatial position of one of the stimuli. This response cue was presented for 1 s and replaced the fixation point for that duration. The time at which the response cue was presented depended on the experimental condition: In the prior cue condition, the cue was shown 1 s before the stimuli for the trial, in the immediate cue condition the cue was shown immediately after the offset of the second stimulus, and in the delayed cue condition the cue was presented 2 s after the offset of the second stimulus. Each stimulus was shown for 1 s. After the stimuli and response cue had been presented, the "#" or "@" was presented at fixation as the signal to respond. This probe stimulus remained in view until the response was executed. For example, the series of events in the immediate condition was: fixation alone for 1 s, fixation plus first stimulus for 1 s, fixation plus second stimulus for 1 s, response cue alone for 1 s, and probe stimulus alone until a response was made. Trials were separated by a 1 s inter-trial interval.

Procedure. Sixteen trial types were constructed by forming all possible combinations of stimulus time and position of occurrence, response cue, and probe. The trial was called positively correlated if the stimulus that was shown first was also above fixation; otherwise it was called negatively correlated (see Dutta & Nairne, 1993, for a further discussion of this distinction). Cue type ("T" or "S") and correlation were varied within blocks of trials for which only one cue presentation time was used. Each subject received 4 instruction, 16 practice, and 64 test trials for each of the cue conditions. The

order in which the conditions were presented was completely counterbalanced across subjects. Trials were presented in random order within the test blocks with the constraint that each trial type occurred four times. Subjects were allowed to take breaks between the practice and test trials and in the middle and at the end of each block.

The computer keyboard was used to collect responses. Two response sets were used, with equal numbers of subjects in each of the presentation orders using each set. For Response Set 1, the "E" and "D" keys were used to respond "above" and "below," respectively, and the "O" and "P" keys were used to respond "first" and "second," respectively using the left middle, left index, right index, and right middle fingers. In Response Set 2, "Q" and "W" were assigned to "first" and "second, "respectively, and "L" and "P" were used to respond "below" and "above," respectively using the left middle, left index, right index, and right middle fingers. Subjects were instructed to keep their hands on the keyboard for the duration of the experiment. If an error was made, the computer screen displayed the message, "Error! Incorrect response" and a diagram of the response assignment. The error screen remained in view until the subject pressed a key to resume the experiment.

Subjects were tested in groups, as described above. All instructions were presented on the computer screen. Three experimenters patrolled the room to answer questions as required.

Data analysis. Both mean reaction times (RTs) and proportion correct were analyzed. Only nonerror trials were included in the RT analyses. Trials on which the RT was faster than 200 ms or slower than 2500 ms (less than 1% of trials) were not analyzed.

Results

Mean RTs and proportion correct as a function of condition, cue type, and correlation are shown in Figure 1. Generally, responses were faster and more accurate when spatial position, rather than time of occurrence, was recalled (816 vs. 845 ms, respectively); when the relevant dimension was cued before the stimuli for the trial were shown (731 vs. 886 and 875 ms for the prior, immediate, and delayed cue conditions, respectively), and when the stimuli were positively correlated (817 vs. 844 ms for positively and negatively correlated trials, respectively). These observations were confirmed by two

separate analyses of variance (ANOVAs) conducted on mean RT and proportion correct with condition, cue, and correlation as within subject factors and order and response set as between subject factors. To conserve space, only significant results are reported.

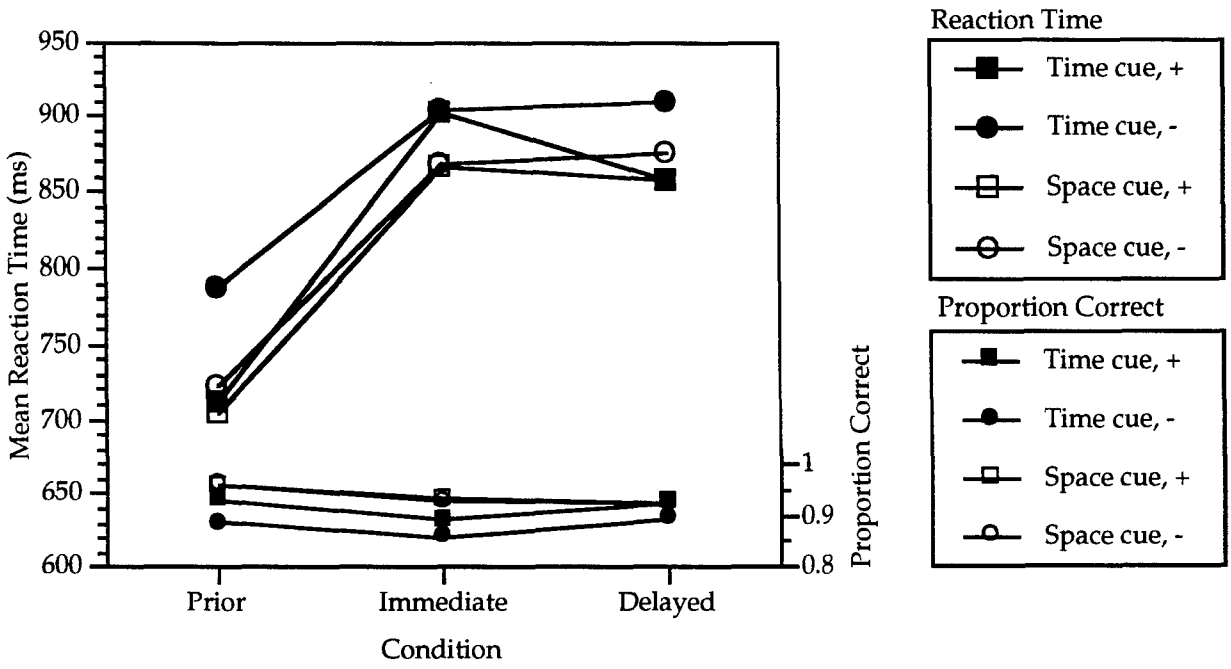


Figure 1. Mean reaction time and proportion correct as a function of condition, cue, and correlation in Experiment 1 (positive correlation = "+"; negative correlation = "-").

The ANOVA on mean RT revealed significant main effects of condition [$F(2, 96) = 27.96$, $p < .001$, $MS_e = 63{,}719$], cue [$F(1, 48) = 5.88$, $p < .02$, $MS_e = 26{,}076$], and correlation [$F(1, 48) = 10.83$, $p < .002$, $MS_e = 12{,}821$]. The Cue x Correlation interaction also was significant, [$F(1, 48) = 4.40$, $p < .05$, $MS_e = 9{,}472$], reflecting a greater effect of correlation when time of occurrence was responded to than when spatial position was recalled (see Figure 2).

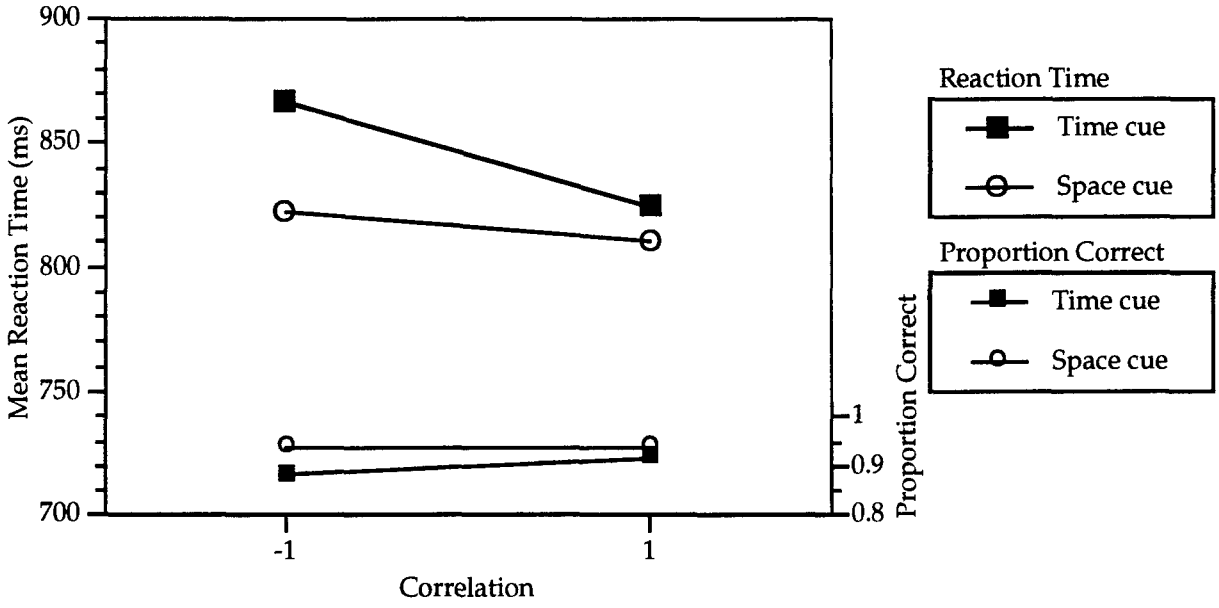


Figure 2. Mean reaction time and proportion correct as a function of correlation (positive correlation = "+ 1"; negative correlation = "- 1") and cue type in Experiment 1.

The results of the ANOVA conducted on proportion correct were consistent with the RT analysis and showed no evidence of speed-accuracy trade-off. The only effect to reach significance in the accuracy analysis that was not significant in the RT analysis was the Cue x Condition interaction [$F(2, 96) = 7.32$, $p < .002$, $MS_e = 0.004$]. Pairwise tests comparing the effect of condition within cue type showed that whereas the responses in the delayed condition were more accurate than immediate responses with the time cue [$F(1, 48) = 9.38$, $p < .004$], there was no difference between these conditions with the space cue.

Although there were no main effects of response set or order, several interactions involving these nuisance factors reached significance. The Condition x Order interaction [$F(10, 96) = 3.67$, $p < .001$] reflects a practice effect such that the magnitude of the differences between conditions depended on which condition was experienced first. In all cases, mean RT was fastest in the immediate condition. The Condition x Order x Response Set interaction [$F(10, 96) = 2.01$, $p < .05$] appeared to be attributable to especially slow responses by one subject, so will not be discussed. The Response Set x Correlation

interaction [$F(1, 48) = 4.09$, $p < .05$] is attributable to a greater effect of correlation with Response Set 1 than with Response Set 2. Inspection of the data revealed this to be true only when the cue was for spatial position; it is not clear why this should be the case [$F(10, 96) = 3.67$, $p < .001$, for the Cue x Correlation x Response Set interaction].

Discussion

The results of the experiment demonstrate that subjects are sensitive to space-time correlation even when cued to attend to only spatial position or time of occurrence. In fact, the effect of correlation was as great for the prior cue condition as for the delayed cue condition. However, responses were much faster overall in the prior cue condition. One possible explanation for the combination of faster responding but continued sensitivity to time of cue presentation is that subjects were able to use the cue to prepare a subset of the responses; according to this view, although selective attention could operate on the response selection processes, space-time correlation still affected memory retrieval processes.

Experiment 2

Experiment 1 demonstrated strong effects of the correlation between spatial location and temporal occurrence. Experiment 2 extends the generality of the finding in two ways. First, three rather than two stimuli are used. This makes the task significantly more difficult and obviates any strategy specific to the case where only two stimuli are involved. Second, the stimuli are arranged in a horizontal row rather than above and below a fixation point. If the effects of correlation are robust, they should still be obtained in this case. Finally, because three stimuli are used, it is possible to examine a greater range of correlations between space and time and thus to evaluate further the effect of correspondence of the two attributes on the recall of just one of them.

Method

Subjects and apparatus. The same 75 subjects who participated in Experiment 1 took place in this experiment. Three subjects were dropped from the analysis because of missing data. Experiments 1 and 2 were separated by a third, unrelated experiment that lasted approximately 40 min. Equal numbers of subjects in each of the experimental orders of Experiment 1 were in each condition of

Experiment 2. The testing conditions and equipment were identical to those of Experiment 1.

Stimuli and procedure. Individual stimuli were the same as in Experiment 1, with the addition of a third stimulus, "&," which had the same dimensions as the others. Rather than displaying the stimuli above and below a fixation point, the three stimuli were shown in left-to-right order. A row of three underscores served as a fixation symbol ("_ _ _"). The display, including both the fixation and the stimuli, subtended a visual angle of 3.78 deg in width and 1.49 deg in height. The "T" or "S" cue was shown directly below the center "_," as was the probe stimulus. All aspects of the trial procedure were the same as in Experiment 1, except that three rather two stimuli were shown. Because of the addition of the third stimulus, 216 unique trial types were possible, and space-time correlations of -1, -.5, .5, and 1 were represented. All possible trial types were used for each subject. Because of the large number of trials, condition (prior, immediate, or delayed cue) was manipulated between subjects. Subjects were allowed to take a rest break after the practice trials and after each set of 54 test trials.

As in Experiment 1, two response sets were used. If a subject used Response Set 1 in Experiment 1, they also used Response Set 1 in Experiment 2 (i.e., the same hand was used to respond to "space" in both experiments). For Response Set 1, the keys "A," "S," "D," "O," "K," and "M" corresponded to the responses "left," "middle," "right," "first," "second," and "third" and were operated with the left ring, left middle, left index, right ring, right middle, and right index fingers, respectively. For Response Set 2, the keys "W," "S," "X," "L," "K," and "J" corresponded to the responses "first," "second," "third," "right," "middle," and "left" and were operated with the left ring, left middle, left index, right ring, right middle, and right index fingers, respectively.

Data analysis. Trials on which the RT was less than 200 ms or over 3500 ms (less than 1 % of trials) were excluded. Error trials were excluded from the RT analyses.

Results and Discussion

Reaction time analysis. Response times were slower than in the first experiment, and the error rates were higher. However, in most other respects the data are similar. Figure 3 shows the mean RT and proportion correct as a function of condition and cue.

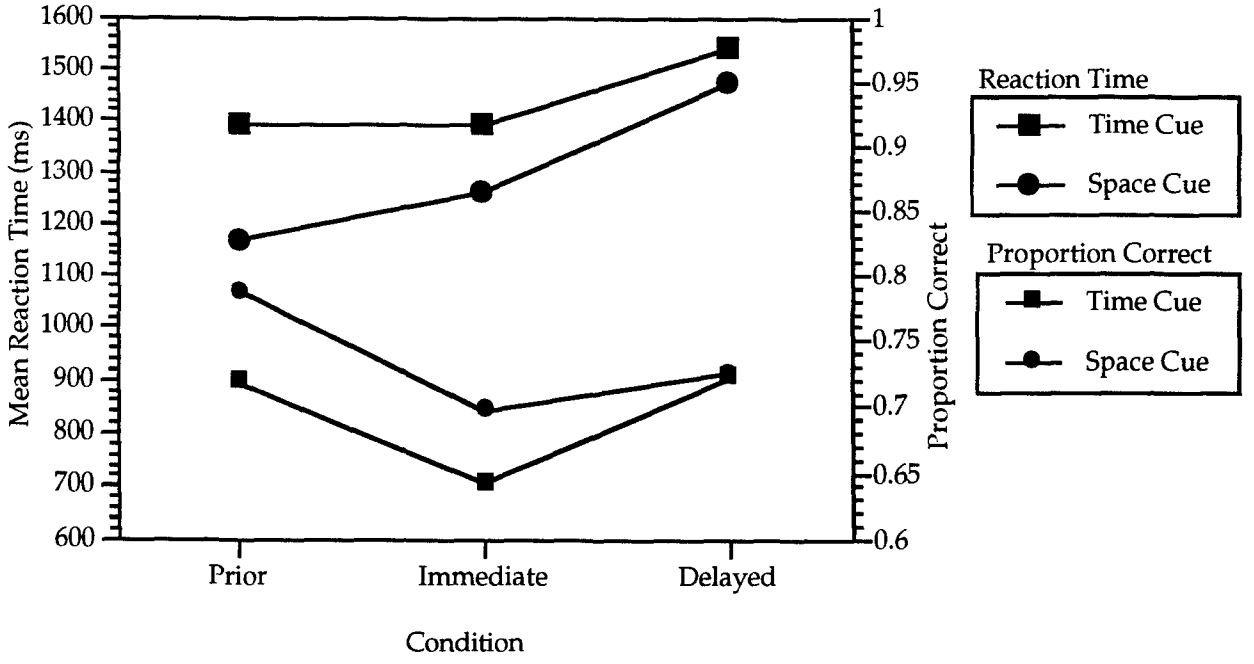


Figure 3. Mean reaction times and proportion correct as a function of condition and cue in Experiment 2.

Responses were faster when spatial position rather than temporal occurrence was cued [1300 vs. 1426 ms, respectively, $F(1, 66) = 37.26$, $p < .001$, $MS_e = 61{,}144$]. Responses were fastest and most accurate in the prior cue condition, intermediate in the immediate cue condition, and slowest in the delayed cue condition [1,269, 1,318, and 1,502 ms, respectively, $F(2, 66) = 3.64$, $p < .04$, $MS_e = 794{,}567$; Tukey's HSD test showed that all means were different from each other]. The effect of positive vs. negative correlation was similar to that of Experiment 1: Responses were 46 ms faster overall on + 1 correlation trials than on -1 correlation trials [$t(71) = 2.37$, $p < .025$] and 38 ms faster on + .5 than on - .5 correlation trials [$t(71) = 2.61$, $p < .02$]. However, both of the perfect correlation trial types produced faster and more accurate responding than did the +.5 and -.5 correlation trials (mean RT = 1,304, 1,350, 1380, and 1,418 ms for the +1, -1, +.5, and -.5 trials, respectively). This advantage for perfect correlation, even when negative, has also been found when stimulus-response assignments are varied to change the correlation of positions in the stimulus set to positions in the response set (e.g., Fitts & Deininger, 1954). This suggests that subjects are sensitive to both the correlation of spatial and temporal occurrence, as

well as the regularity of that correlation. That is, they appear to be able to process a "reverse" order of spatial or temporal positions better than a more random order, even though more individual spatial and temporal positions correspond in the latter case.

The only interactions to reach significance were that of condition and cue [$F(2, 66) = 4.31$ $p < .02$, $MS_e = 61,144$] and response set, cue, and correlation [$F(3, 198) = 3.94$, $p < .01$, $MS_e = 22,260$]. As shown in Figure 3, the difference in response times as a function of condition depended on whether time or space was cued. However, an inspection of the proportion correct in each condition suggests that the lack of difference between the prior and immediate cue conditions for the time cue may be attributable to speed-accuracy trade-off. The Response Set x Cue x Correlation interaction appears to be due primarily to relatively fast responding using Response Set 1 on -1 correlation trials when the time cue was presented, whereas for the space cue, these responses were relatively slow.

Error Analysis. The results of the ANOVA on proportion correct are similar to those of the RT analysis, except that there was no main effect of condition. There were main effects of cue [$F(1,66) = 18.68$, $p < .001$, $MS_e = 0.014$] and correlation [$F(3,198) = 5.16$, $p < .002$, $MS_e = 0.009$]. The Condition x Cue interaction [$F(2,66) = 4.72$, $p < .02$, $MS_e = 0.014$] appears to be mainly due to relatively low accuracy with the space cue in the delayed condition.

Because the error rates in this experiment were relatively high, additional analyses looking at the types of errors made were carried out. The data were coded by the correct response, the response actually selected, and what the correct response would have been had the other cue been presented. For example, on a trial for which the stimuli appeared in the positions "middle," "left," and "right," in that temporal order, if the space cue was presented and the probe was the stimulus that occupied the middle position, the corresponding time of occurrence of the probed stimulus would be "first." The average number of responses for each of the correct answers are shown in Figure 4. It should be noted that similar graphs were drawn for all possible combinations of correlation, cue, and condition and the same general pattern was observed. It can be seen that for both the space and time cues errors tended to be intrusions of responses that were closest in space or time to the correct response. Thus, it appears that

subjects treat both space and time as ordered dimensions (see Nairne & Dutta, 1992, for further evidence of this).

An additional analysis was carried out on the number of errors made by responding to the uncued rather than the cued dimension. For this analysis, errors were counted as "correct" if the response would have been correct for the corresponding cue type and "incorrect" otherwise. Four subjects made no errors of either type and so were not included in the analysis. There were main effects of cue [$F(1, 67) = 28.79$ $p < .001$, $MS_e = 5.6$] and type of answer [$F(1, 67) = 35.06$ $p < .001$, $MS_e = 4.34$], and the Cue x Answer Type interaction [$F(1, 67) = 26.52$ $p < .001$, $MS_e = 2.74$] was significant. More errors were made when time was cued (average = 3.43 errors) than when space was cued (average = 1.89) and more "correct" than "incorrect" responses were made (3.41 vs. 1.91, respectively). Post hoc tests showed that , when the time cue was presented and the subject erred by responding to spatial position, they tended to make the response that would have been correct if space were cued rather than any other spatial response ($p < .0001$; average number of "correct" responses = 4.69, average number of "incorrect" responses = 2.16). When the space cue was presented but time of occurrence was responded to, there was no difference in the mean frequency of "correct" and "incorrect" responses ($p > .10$; average number of "correct" responses = 2.12, average number of "incorrect" responses = 1.65). This finding suggests that recall of temporal order is more influenced by prior spatial position than vice versa.

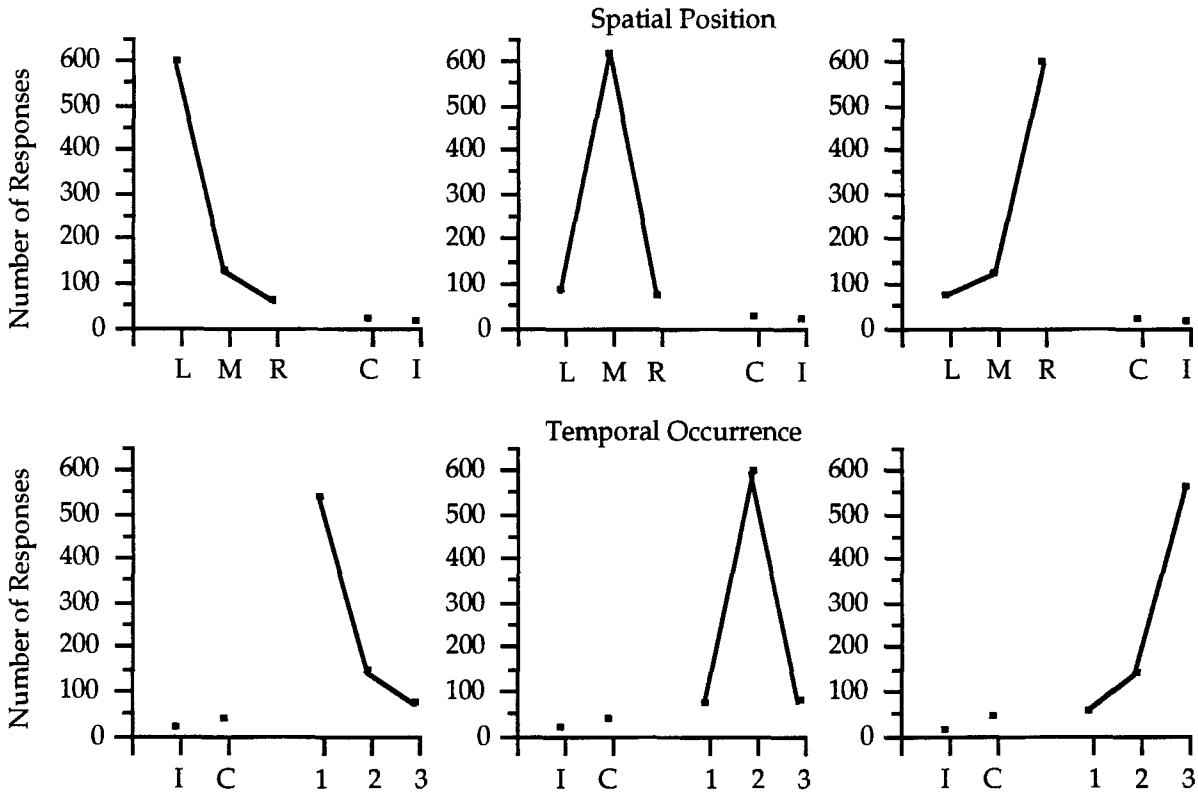


Figure 4. Correct and incorrect responses as a function of cue type. "I" means incorrect intrusion from the noncued dimension and "C" indicates the response that would have been correct had the other dimension been cued. "L," "M," "R," "1," "2," and "3" denote the responses "left," "middle," "right," "first," "second," and "third," respectively.

General Discussion

In Experiment 1, two stimuli were presented one at a time, above or below fixation, and their spatial position or temporal occurrence were to be remembered. Experiment 2 was similar, but used three stimuli that appeared in three horizontal positions. In both experiments, evidence of the interaction of spatial and temporal location was obtained. The interference of spatial position on recall of temporal information was greater than that of temporal occurrence on spatial position. This was evident both in reaction time and error analyses and in an analysis of the types of intrusions made

by subjects on error trials. This suggests that special care must be taken in presenting temporal information. Either the information should be presented in a nonspatial format (e.g., in a tabular format) or it should be presented in such a way that the spatial information does not conflict with the temporal information. Another way to describe the pattern of results would be to say that spatial information is more salient than temporal information. This also has implications for design. For example, if a graphic display is used to present a time history of a plane's progress (e.g., by leaving a trail of dots behind the planes display symbol; Klinger et al., 1993) more information would be conveyed if the relative spacing of the dots corresponded to the distance traveled in a fixed amount of time than if the spacing corresponded to time elapsed over fixed distances. Since it appears to be easier for subjects to process spatial information, the former scenario, which minimizes the processing of temporal information, should lead to better performance.

It is important to note that the results were qualitatively similar in both experiments, despite the change in processing demand. This suggests that the interactions observed here will also be observed in more complex, real-world tasks, although further research is required to determine more fully how spatial and temporal information interact in memory.

References

Dutta, A., & Nairne, J. S. (1993). The separability of space and time: Dimensional interaction in the memory trace. Memory & Cognition, 21, 440-448.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. Journal of Experimental Psychology, 48, 483-492.

Jones, G. V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. Journal of Experimental Psychology: General, 105, 277-293.

Jones, G. V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. Psychological Review, 94, 229-235.

Keele, S. W., Cohen, A., Ivry, R., Liotti, M., & Yee, P. (1988). Tests of a temporal theory of attentional binding. Journal of Experimental Psychology: Human Perception and Performance, 14, 444-452.

Klinger, D. W., Andriole, S. J., Militello, L. G., Adelman, L., Klein, G., & Gomez, M. E. (1993). Designing for performance: A cognitive systems engineering approach to modifying an AWACS human-computer interface. Final Technical Report #KATR9014-93-01Z, Fairborn, OH: Klein Associates, Inc.

Massaro, D. W., Weldon, M. S., & Kitzis, S. N. (1991). Integration of orthographic and semantic information in memory retrieval. Journal of Experimental Psychology: Learning, Memory, & Cognition, 17, 277-287.

Nairne, J. S., & Dutta, A. (1992). Spatial and temporal uncertainty in long-term memory. Journal of Memory and Language, 31, 396-407.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM-PC compatibles. Behavior Research Methods, Instruments, and Computers, 20, 206-217.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. Perception & Psychophysics, 39, 164-174.

Well, A. D., & Sonnenschein, B. (1973). Effects of irrelevant stimulus dimensions on selection in immediate memory. Journal of Experimental Psychology, 99, 283-285.

Acknowledgments

I am grateful for the collaboration of Kelly Neville, who assisted in the development and execution of the experiments. Cynthia Guerrero provided valuable assistance in creating the many, many figures that aided in the interpretation of the data.

FUEL IDENTIFICATION BY NEURAL NETWORK ANALYSIS OF GAS CHROMATOGRAPHY DATA

Paul A. Edwards, Ph. D.
Associate Professor
Department of Chemistry

Cooper Hall
Edinboro University of Pennsylvania
Edinboro, Pennsylvania 16444

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

FUEL IDENTIFICATION BY NEURAL NETWORK ANALYSIS OF GAS CHROMATOGRAPHY DATA

Paul A. Edwards, Ph. D.
Associate Professor
Department of Chemistry
Edinboro University of Pennsylvania

Abstract

Neural network methods including back-propagation have been successfully applied to the analysis of jet fuel gas chromatography data. The gas chromatographic profiles of aviation fuels have been used to train artificial neural networks to correctly classify fuels in two different data sets. Attention was paid to minimizing the number of features and optimizing the network architecture required to accomplish classification.

FUEL IDENTIFICATION BY NEURAL NETWORK ANALYSIS OF GAS CHROMATOGRAPHY DATA

Paul A. Edwards

Introduction

The rapid and reliable identification of aviation fuel in soil and water samples is important in the detection and monitoring of fuels in the environment. As the United States military *downsizes* and bases are converted to civilian use, it is imperative that techniques be developed to identify and monitor fuels in the soil and water at former military sites. Significant improvement in the classification of such samples has been obtained by applying neural network methods to the analysis of gas chromatography data of aviation fuels.

Artificial neural networks (ANN) are computer simulations of biological nervous systems. In general terms, numerical information is entered into a network through a layer of input neurons or nodes and exits through a layer of output nodes. Information is passed from the input to the output layer through a hidden layer or layers that also have nodes. As information is passed through the layers, weights, biases and transfer functions are applied which adjust the transfer of information between the nodes and the output of the network. A network is trained with a set of input and corresponding output patterns or vectors. The weights and biases are adjusted until the output patterns generated by the network match those in the training data set.

Dr. Howard Mayfield and co-workers have been exploring the application of artificial neural networks (ANNs) to the analysis of jet fuel chromatographic data (1). Recent work by Faruque and Mayfield (2) has resulted in the development of FIP, a "Fuel Identification Program". FIP is a marvelous suite of MATLAB^{tm} (3a.) and Neural Network TOOLBOX^{tm} (3b.) functions that provides a *user-friendly* environment for classifying jet fuels based on gas chromatography data. The work described here shifts the focus from development of FIP to its application.

A recent review of pattern recognition techniques by Brown *et al* (4) noted "The most novel research in pattern recognition involved work with artificial neural networks." Indeed, not only have ANNs been used in this laboratory (1) along with more classical methods (5-8) to classify fuels, ANNs have also been used to classify fuels based on laser induced fluorescence spectra (9). However, questions of feature selection (10), and architecture optimization (11,12) continue, and are addressed in this report.

In this context, feature selection is the process of determining how many and which peaks from the chromatograms are to be included in the data matrix input to the ANN. The process starts in the data preprocessing phase when the gas chromatograms are searched for peaks that appear in a significant fraction of the profiles in the data set. However, this preprocessing generally yields a data matrix that has too many features. Classical pattern recognition methods require many more samples (chromatograms) than variables for training to be statistically valid. Including too many variables in an ANN can lead to over-fitting of the training set and the concomitant loss of robustness. The dilemma in applying neural networks is no clear guidelines as yet exist for determining a safe ratio of samples to parameters or variables, in part because the parameters are coupled to varying degrees (13). In this paper, the terms parameters or variables will refer to the weights and biases adjusted during the training process. Thus feature selection becomes the process of reducing the number of features in the data matrix such that the ratio of the numbers of samples to variables is a maximum.

Architecture optimization, in this context, is the process of determining the number of input, hidden and output nodes required in the ANN to yield the best possible classifications. The number of output nodes in the ANN is fixed by the number of classes represented in the data matrix. Niebling (11) claims there is also a minimum number of hidden nodes required for the ANN. He proposes the minimum number of what he calls inner units (n) is related to the number of identification areas (k): $k \leq [n(n + 1)/2] + 1$. Finally, architecture optimization and feature selection are interconnected because the number of input nodes is determined by the number of features in the data matrix.

Ideally, there should be several possible architectures that fulfill the following requirements. First, and most obviously, the number of output nodes should equal the number of classes in the data matrix. Second, there should be at least the number of hidden nodes as claimed by Niebling (8) to accomplish classification. Third, the number of input nodes should be small enough that the ratio of samples to variables is equal to or preferably greater than one. Fourth, classification must be successfully accomplished. Thus, architecture optimization becomes the process of finding the architecture that best fulfills these requirements.

Methodology

Collection of the kfr Data Set: The kfr data set represents 264 chromatograms of neat fuels analyzed using a high-speed gas chromatography proceedure. The fuels were diluted with methylene chloride and analyzed by gas chromatography, using a flame ionization detector and a fused silica capillary column, 10 m long, with an internal diameter of 0.10 mm, and coated with 0.34 μm of 5% phenyl substituted polymethylsiloxane (HP-5, Hewlett-Packard Co.). A high speed temperature program was used to elute the fuel components through the column, which yielded increased throughput and enabled a large number of chromatographic analyses to be performed in a short

period. The data were originally collected from the GC/FID signal using an HP-3357 Laboratory Automation System (LAS) (1). For this data analysis, the raw data files were transferred to an HP-3350 Laboratory Automation System, translated into the new system's file format, and re-integrated with a new integration method designed to ignore the methylene chloride solvent peak. An internal standard, D_{10}-anthracene, was also spiked into each diluted fuel prior to analysis, but after some trial and error it was decided to transduce the data based on normalized peak areas, i.e. percentage areas, rather than to apply an internal standard correction. In transducing the data set, retention time variations were corrected through the use of Kovat's retention indices, calculated using a linear temperature programming formula. The percentage areas of 85 peaks, found to occur in a satisfactory portion of the chromatograms, were used as features to transduce the chromatographic integration reports into data vectors. Integration reports from 25 chromatograms of recovered fuels and environmental extracts analyzed under the same conditions as the initial 264 fuel samples were also transduced to produce a prediction set. The resulting data set consists of a training set of 264 objects X 85 dimensions, and a prediction set of 25 objects X 85 dimensions.

Collection of the wfd Data Set: The wfd data set is derived from 134 chromatograms of the water-soluble fractions of jet fuels. The fuels were equilibrated with water in vessels which kept the fuel and water layers carefully separated. Aliquots of the resulting aqueous phase were extracted using a positive pressure type solid phase extraction cartridge (Sep-Pak^{tm}, Millipore, Inc) and the organic components were eluted from the extraction cartridge with CS_2. The resulting extracts in CS_2 were analyzed by gas chromatography/mass spectrometry using a fused silica capillary column, 60 m long, with an internal diameter of 0.25mm, and coated with 0.25 μm of a bonded polyethylene glycol stationary phase (DBWAX, J&W, Inc). Total ion chromatograms were integrated, and the peak areas were corrected for response using an internal standard, D_{10}-ethylbenzene. Corrected peak areas from 48 peaks occuring in most of the the integration reports were used as features to transduce the chromatograms into data vectors. The resulting 134 object X 48 dimensional data set is designated wfd. This data set has been described in previous reports as water-soluble fraction data (1, 5).

The Data Sets: The composition of the two data sets is displayed in Table I. The necessity of the modified kfr data set will be discussed in the Results and Discussion section. The training and prediction subset (tset/pset) pairs of these data sets also will be discussed later. There were two differences between the kfr and wfd data sets that are important to this discussion. First, one of the 48 features in the wfd data set was a constant spike of D_{10}-ethylbenzene. That feature was removed from the data matrix. Second, 22 other features in the chromatograms in the wfd data set had been identified. As features were removed from consideration, an attempt was made to maintain the maximum number of identified features.

Neural Network Analysis: The main tool in the neural network analysis of the data sets was FIP, a "Fuel Identification Program," by Faruque and Mayfield (2). FIP is a suite of MATLAB^{tm} (3a.) and Neural Network TOOLBOX^{tm} (3b.) functions developed for the classification of fuels from gas chromatography data. FIP has been implemented on two platforms. It has been implemented on Sun SPARCstations operating under SunOS 4.1.3 (Solaris 1.1.1) and OpenWindows. It has also been implemented on IBM-compatible 486 computers operating under Windows.

Feature Selection: Initially, the data matrix was searched for features with high pairwise correlations (i.e., > 0.90). Selecting which of a pair of correlated features to delete is a somewhat arbitrary process. Features correlated to more than one other were removed first to minimize the number removed at each step in the process. In the analysis of the wfd data set, identified peaks in the chromatograms were maintained as much as possible. The network was then trained with the reduced data matrix. If the network trained to essentially the same or fewer misclassifications, the data matrix was searched again. Features with the highest remaining correlations were removed and the network retrained. The ratio of samples to variables was also calculated for each data matrix. The sum of the square of the weights connecting each input node to the hidden layer was also calculated. Features with low values for this quantity were also considered for deletion. This cyclic process of reducing the data matrix and training was repeated until the network could not be trained to an acceptable level and the ratio of samples to variables was greater than one.

Training and Architecture: Initially the architecture of the ANN was the minimum. The number of input nodes was the number of features in the data matrix. The number of output nodes was eight. The number of nodes in the hidden layer was four as calculated using Niebling's relationship (8). Training was done using back propagation (BP) methods unless otherwise noted. Generally the minimum ANN could be trained to give zero or one misclassifications. Several radial basis function (RBF) neural networks (13, 14) were also trained.

Results and Discussion

The kfr data set: The original data matrix for the kfr data set contained 85 features as shown in Table I. The ratio of samples in the complete data set to variables in an otherwise minimum network architecture would be 0.69. For the ratio of samples to variables to be greater than one, the number of features in the data matrix would have to be reduced to at least 55.

There was concern early in the feature selection process that there might be a problem within the data set. Samples of JP-8 were regularly misclassified during training. It is possible for a fuel to meet the specifications for both JP-5 and JP-8, and batches of JP-5 have been recertified as JP-8 after initial processing (2). Thus the true identity of some JP-8 samples may be unknown unless the initial

processing and certification is known. In light of this problem, the kfr data set was modified with all JP-5 and JP-8 samples being put into the JP-5 class. In practice this would mean that two ANNs would be required to completely classify a fuel. The first network would assign fuels to one of seven classes, including the combined JP-5/JP-8 class. The second would be required to separate the combined class into separate JP-5 and JP-8 subclasses.

It was possible to reduce the data matrix to 29 features using the modified kfr data set and the process described above. There were no pairwise correlations above 0.60 in the reduced data matrix for the complete data set. A minimum network of 29 input, 4 hidden and 8 output nodes was trained to zero misclassifications. The ratio of samples in the full data set to variables in this network was 1.65.

In order to validate the network and optimize the architecture, the kfr data set was divided into two training and prediction (tset/pset) subset pairs. As shown in Table II, there were 251 and 13 samples, respectively, in both tset/pset pairs. The difference was that the JP-5 and JP-8 samples were combined into a single class in one of the pairs, while they were in separate classes in the other. Because there were two similar tset/pset pairs, it first had to be determined which would lead to better predictions. Then it could be determined if there was an optimum architecture. All training and predictions were done with reduced data matrices. Several architectures were explored ranging from the minimum to larger ones up to those with sample to variable ratios of one for both tset/pset pairs.

Neither training set gave satisfactory predictions using 29 features in the data matrix. Indepedent of architecture, either the network could not be training to zero misclassifications or the number of prediction samples misclassified was unacceptably high when 29 features were used in the data matrix to train the network. It was necessary to use a larger data matrix with 33 features. There were no pairwise correlations above 0.65 in the 33 feature data matrix for the complete data set.

Three architectures with one hidden layer were considered ranging from the minimum of 33 input, 4 hidden and 8 output nodes to 33 input, 6 hidden and 8 output nodes. The values for the ratio of samples in the training sets to variables in each network ranged from 1.43 to 0.97. Two architectures with two hidden layers were also studied. Both of these networks had 33 input and 8 output nodes. One network had 4 nodes in both the first and second hidden layers, while the other had 5 nodes in the first and 4 nodes in the second hidden layers. The values for the ratio of samples in the training set to variables in these two networks were 1.28 and 1.07, respectively.

The best results for the data set with JP-5 and JP-8 samples in the same class were obtained using a minimum network of 33 input, 4 hidden and 8 output nodes. The training set was used to train this

network to zero misclassifications. There were no misclassifications when the network was applied to the prediction set. The value for the ratio of samples in the training set to variables in this network was 1.43.

The best results for the data set with JP-5 and JP-8 samples in the separate classes were obtained using a network of 33 input, 8 output and 4 nodes in each of two hidden layers. The training set was used to train this network to zero misclassifications. There were no misclassifications when the network was applied to the prediction set. The value for the ratio of samples in the training set to variables in this network was 1.28.

Slightly different results were obtained when the RBF method was used to train the network instead of BP. The need for 33 features in the data matrix instead of 29 was again clear. The number of prediction samples misclassified was unacceptably high (10 of 13) when 29 features were used in the data matrix to train the network. There was only one misclassified prediction sample when 33 features were used to train the network. Also similar to what was observed using BP methods, the results were essentially the same whether JP-5 and JP-8 samples were in the same or separate classes. In both cases it was possible to train a network that gave one misclassification when applied to the prediction set. Contrary to what was observed using BP, the network trained with 33 features in the data matrix misclassified one sample, instead of zero, when applied to the prediction set.

The wfd data set: The original data matrix for the wfd data set contained 48 features as shown in Table I. Assuming an otherwise minimum architecture for the network, the ratio of samples to variables would be 0.57. For the ratio of samples to variables to be greater than one, the number of features in the data matrix would have to be reduced to at least 22. In fact, it was possible to reduce the data matrix to 10 features using the process described above. There were no pairwise correlations above 0.50 in the reduced data matrix for the complete data set. A minimum network of 10 input, 4 hidden and 8 output nodes was trained to zero misclassifications. The ratio of samples in the full data set to variables in this network was 1.60.

Mayfield and Henley (5) had previously studied the wfd data set using k-nearest neighbor techniques. They identified four features, benzene, 1,2-diethylbenzene, 1,2,3,4-tetramethylbenzene and 1,2,3,4-tetrahydronaphthalene, which could be used to classify the data set into JP-4, AVGAS and "other" categories. They also identified four more features, isopropylbenzene, 1-methylnaphthalene, an unidentified peak, and 1,2,3,4-tetrahydronaphthalene, which could be used to classify the "other" profiles into JETA, JP-5, JP-7, JPTS and diesel categories.

Of the ten features in the data matrix deduced in this study, six are identified and four are unidentified. The six identified features are benzene, toluene, ethylbenzene, isopropylbenzene,

1,2,3,4-tetrahydronaphthalene and naphthalene. Three of these features are common to those deduced by Mayfield and Henley (5). One feature, naphthalene, deduced in this study was highly correlated with a feature, 1-methylnaphthalene, found to be important by Mayfield and Henley. The naphthalene feature was replaced with 1-methylnaphthalene in the reduced data matrix and the minimum ANN trained to zero misclassifications. Interestingly, there was a pairwise correlation greater than 0.50 in this 10 feature data matrix; the 1-methylnaphthalene was correlated to an unidentified peak. The unidentified peak was removed and the resultant nine feature data matrix used to train a minimum ANN to zero misclassifications! The ratio of samples in the full data set to variables in the minimum ANN was 1.68.

The final nine feature data matrix included six identified and three unidentified features. The six identified features are benzene, toluene, ethylbenzene, isopropylbenzene, 1,2,3,4-tetrahydro-naphthalene and 1-methylnaphthalene. It is interesting that four features deduced in this study, benzene, isopropylbenzene, 1,2,3,4-tetrahydronaphthalene and 1-methylnaphthalene, are common to those found by Mayfield and Henley (5). The six identified compounds are also interesting in that they can be taken to represent different series of compounds found in fuels. The list includes benzene and three related compounds (i.e. benzene plus C_1, C_2 and C_3 alkyl derivatives). The list also includes a C_1 alkyl derivative of naphthalene in the data matrix with 9 features, or naphthalene in the data matrix with 10 features.

In order to validate the network and optimize the architecture, the wfd data set was divided into two training and prediction (tset/pset) subset pairs. As shown in Table II, there were 121 and 13 samples, respectively, in the first tset/pset pair. Each subset contained one of the two JP-8 samples in the full data set. There were 118 and 16 samples, respectively, in the second tset/pset pair. Again, each subset contained one of the two JP-8 samples, but they had been switched. That is, the JP-8 sample in the training set in the first pair, was in the prediction set of the second, and vice versa. Other samples in the prediction subsets were randomly chosen from the complete data set. All training and predictions were done with reduced data matrices containing only the nine features previously discussed. Several architectures were explored ranging from the minimum to larger ones up to those with sample to variable ratios of one.

The best results for the first tset/pset pair were obtained with two different architectures. Networks with 4 and 5 nodes in a single hidden layer were trained to zero misclassifications with the first training set. The ratio of samples in the training set to variables in these networks was 1.51 and 1.23, respectively. One fuel, a JP-4, was misclassified as a JPTS, when those networks were applied to the prediction set. A network with 6 nodes in the hidden layer (sample to variable ratio 1.04) was also trained to zero misclassifications. There were no misclassifications when that network was applied to the prediction set. A network with 5 nodes in the first hidden layer and 4 in the second

was also trained to zero misclassifications for both the training or prediction data sets. The ratio of samples in the training set to variables in that network was 1.06. However, networks with 4 nodes in both first and second hidden layers would not train to zero misclassifications for the training set. Networks with more than 6 nodes in a single layer or 4 and 5 nodes in two layers were not considered because they contained more variables than samples in the training set.

The results for the second tset/pset pair were better than those for the first tset/pset pair. A network with four nodes in a single hidden layer was trained to zero misclassifications. There was one misclassification when this network was applied to the prediction set; a sample of JP-7 was classified as a JP-5. Similarly, a network with four nodes in both a first and second hidden layer, results in zero misclassifications in training and one misclassification in prediction. Three different architectures were all trained to zero misclassifications and gave zero misclassifications when applied to the prediction set. Two of three networks had single hidden layers with five or six nodes. The third network had five nodes in the first hidden layer and four in the second. The ratio of samples in the training set to variables in these three network were 1.20, 1.02 and 1.04, respectively.

Two results from both training and prediction sets are important. First, the architecture with six nodes in a single hidden layer gave 100% correct classifications for training and prediction for both tset/pset pairs. Second, it is striking that the JP-8 samples in both training and prediction set pairs can be properly classified with only one JP-8 sample in the training data set!

The results obtained when the RBF method was used to train the network were not as good as those obtained using BP. The best results for the first tset/pset pair showed one sample, a JP-4, misclassified as a JPTS. The best result obtained for the second tset/pset pair had zero misclassifications for prediction set.

Conclusion: The data matrices for both data sets were reduced in size significantly by removing features with high pairwise correlations. The kfr data matrix was reduced from 85 to 33 features and the wfd data matrix was reduced from 48 to 9 features. In the process the ratio of samples in the full data set to variables in a minimum ANN increased from 0.69 to 1.43 for the kfr data set, and from 0.57 to 1.68 for the wfd data set. This increase in the ratio of samples to variables should increase confidence in the statistical validity of the network.

The reduced data matrices were used to train neural networks with as good if not better predictive power. Architectures were found that gave zero misclassifications in both training and prediction. It was also easier to make chemically meaningful inferences from the reduced data matrix. As demonstrated with the wfd data set, it was possible to recognize at least some of the compounds that led to classification.

Generally, the results obtained using the RBF method to train the network were not as good as those obtained using BP. It was possible to train networks using BP for tset/pset pairs of both data sets to give zero misclassifications for prediction. However, when RBF was used, the best results with the same data matrices for the kfr data set was one misclassified prediction sample. When RBF was used with the same wfd data matrices, one tset/pset pair had one misclassified sample. The second tset/pset pair had zero misclassifications.

Further Comments: Dr. Abdullah Faruque is developing algorithms for reducing the size of the data matrix and modifying the FIP program to include these options. The new algoritms not only explore the correlations within the data matrix but the covariance. It is possible to reduce the size of the kfr data matrix farther than indicated in this report with good results for training and prediction.

There does however appear to be one disadvantage to this automated procedure. It will be difficult to preferentially keep assigned features in the chromatograms while removing unassigned features as was done with the wfd data set.

References

1. Long, J. R.; Mayfield, H. T.; Henley, M. V.; Kromann, P. R. *Anal. Chem.* **1991**, *63*, 1261-1264.

2. Faruque, A. and Mayfield, H. T., abstract submitted to Pittsburgh Conference, New Orleans, LA, March 5-10, 1995.

3. (a.) MATLAB[tm], High-Performance Numerical Computation and Visualization Software; (b.) Neural Network TOOLBOX: For Use with MATLAB[tm]; The Mathworks, Inc., Natick, Massachusetts, 1992.

4. Brown, S. D.; Blank, T. B.; Sum, S. T.; Weyer, L. G. *Anal. Chem.* **1994**, *66*, 315R-359R.

5. Mayfield, H. T.; Henley, M. V. In: *Monitoring Water in the 90's: Meeting New Challenges, ASTM STP 1102*; Hall, J. R.; Glysson, D. G., Eds., American Society for Testing and Materials, Philadelphia, 1991.

6. Lavine, B. K.; Qin, X; Stine, A.; Mayfield, H. T. *Process Control Quality*, **1992**, *2*, 347-355.

7. Lavine, B. K.; Stine, A.; Mayfield, H. T. *Anal. Chimica Acta.*, **1993**, *277*, 357-367.

8. Lavine, B. K.; Stine, A. B.; Mayfield, H.; Gunderson, R. J. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 826 - 834.

9. Andrews, J. M.; Liebermann, S. H. *Anal. Chim. Acta*, **1994**, *285*, 237-246.

10. Duin, R. P. W. *Patt. Recog. Lett.* **1994**, *15*, 215-217.

11. Niebling, G. *Sens. Actuators B.* **1994**, *18-19*, 259-263.

12. Wythoff, B. J. *Chemom. Intell. Lab Syst.* **1993**, *18*, 115-155.

13. Carlin, M.; Kavli, T.; Lillekjendlie, B. *Chem. Intell. Lab. Syst.*, **1994**, *23*, 163-177.

14. Lohninger, H. *J. Chem. Inf. Comput. Sci.*, **1993**, *33*, 736-744.

Table I: Composition of Data Sets

	Date Set		
Fuel	kfr	kfr(modified)	wfd
JP-4	54	54	31
JETA	70	70	27
JP-7	32	32	8
JPTS	29	29	20
JP-5	43	57	18
JP-8	14	0	2
Diesel	0	0	12
AVGAS	22	22	16
Total:	264	264	134
No. Features	85	85	48

Table II: Composition of Prediction/Training Set Pairs

	kfr Data Set		wfd Data Set	
Fuel	TSET1/PSET1	TSET2/PSET2	TSET1/PSET1	TSET2/PSET2
JP-4	52/2	52/2	29/2	28/3
JETA	68/2	68/2	25/2	24/3
JP-7	30/2	30/2	7/1	7/1
JPTS	27/2	27/2	18/2	18/2
JP-5	41/2	54/3	16/2	16/2
JP-8	13/1	0/0	1/1	1/1
Diesel	0/0	0/0	11/1	10/2
AVGAS	20/2	20/2	14/2	14/2
Total:	251/13	251/13	121/13	118/16

REGIONAL ARTERIAL COMPLIANCE AND RESISTANCE CHANGES FOR TRANSIENT +Gz PROFILES

Richard D. Swope
Professor
Department of Engineering Science
Trinity University
715 Stadium Drive
San Antonio, TX 78212-7200
and
Daniel L. Ewert
Assistant Professor
Department of Electrical Engineering
North Dakota State University
Fargo, ND 58105

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

ABSTRACT

The primary aim of this research is to determine regional variations in peripheral resistance, blood volume and arterial compliance caused by transient +Gz loads. A model previously used to analyze systemic arterial compliance and total peripheral resistance is extended to allow similar calculations for the head, lungs and body as well as shifts in blood volume between these regions. Gravitational loss of consciousness (G-LOC) is a direct result of a prolonged blood volume shift from the head to the body and the new model allows a study of the relationship between this shift and regional changes in resistance and compliance on a beat to beat basis. Practical surgical limitations require the development of a new transducer for measuring pressure and flow in the pulmonary artery and the aorta before the method can be implemented. Preliminary work with a modified transit time ultrasonic transducer shows promise as a solution to this problem.

Background

The vulnerability of cardiovascular function to the increased gravitational (+Gz) loads seen in military aircraft has spawned a large body of applied research over the last 50 years. In recent years the need to wear chemical and biological protective gear has increased this vulnerability because the aircrew often become dehydrated as a result of the added thermal load. Under normal (hydrated) circumstances +Gz loads lead to a pooling of the blood volume in the lower body and legs and a consequent decreased stroke volume of the heart. This limits the ability of the baroreflexive mechanisms to restore arterial cerebral perfusion pressure and if the blood supply to the brain is reduced for a long enough time loss of consciousness (G-LOC) results. With dehydrated crew members plasma volume is reduced and it is expected that the drop in stroke volume will occur at lower +Gz loads. In fact, given enough dehydration, one would expect problems even at the +Gz loads found in some helicopter maneuvers, (3 to 4 G).

Total peripheral resistance (TPR) and systemic arterial compliance (SAC) are two parameters which may be used to help understand and ultimately provide solutions to, the problems of +Gz loads. TPR, a measure of the hydraulic resistance or vascular load of the left ventricle during ejection, is a significant player in determining aortic blood pressure. SAC, a measure of the elastance of the arterial system is also significant; together they determine the

systemic impedance load for the left heart and the pressure/flow characteristics in the aorta.

Researchers historically have had difficulty translating physiological signals into useful descriptions of short-term systemic pressure regulation, especially under non-stationary conditions. Underlying this has been an inability to determine TPR and SAC during non-steady-state periods of +Gz exposure. Standard methods that use mean pressure and flow (1) are inappropriate under transient conditions because of the varying amounts of blood stored in the arterial compliance (2). As a result non-steady-state TPR can only be correctly derived when SAC is taken into account.

We have developed a method (3) which provides beat to beat values of TPR and SAC under transient conditions similar to those found in military aircraft. Modifying the model to include separate vascular beds for the head, lungs, and remainder of the CV system will allow a determination of regional blood volume shifts during transient +Gz loads and a better understanding of the effects of hydration state.

Proposed Model for Dehydration Studies

Our previously developed model for computing TPR and SAC is a two element Windkessel consisting of a lumped systemic arterial compliance (C_{ao}) and a lumped resistance ($R_{arterial}$) as shown in Figure 1.

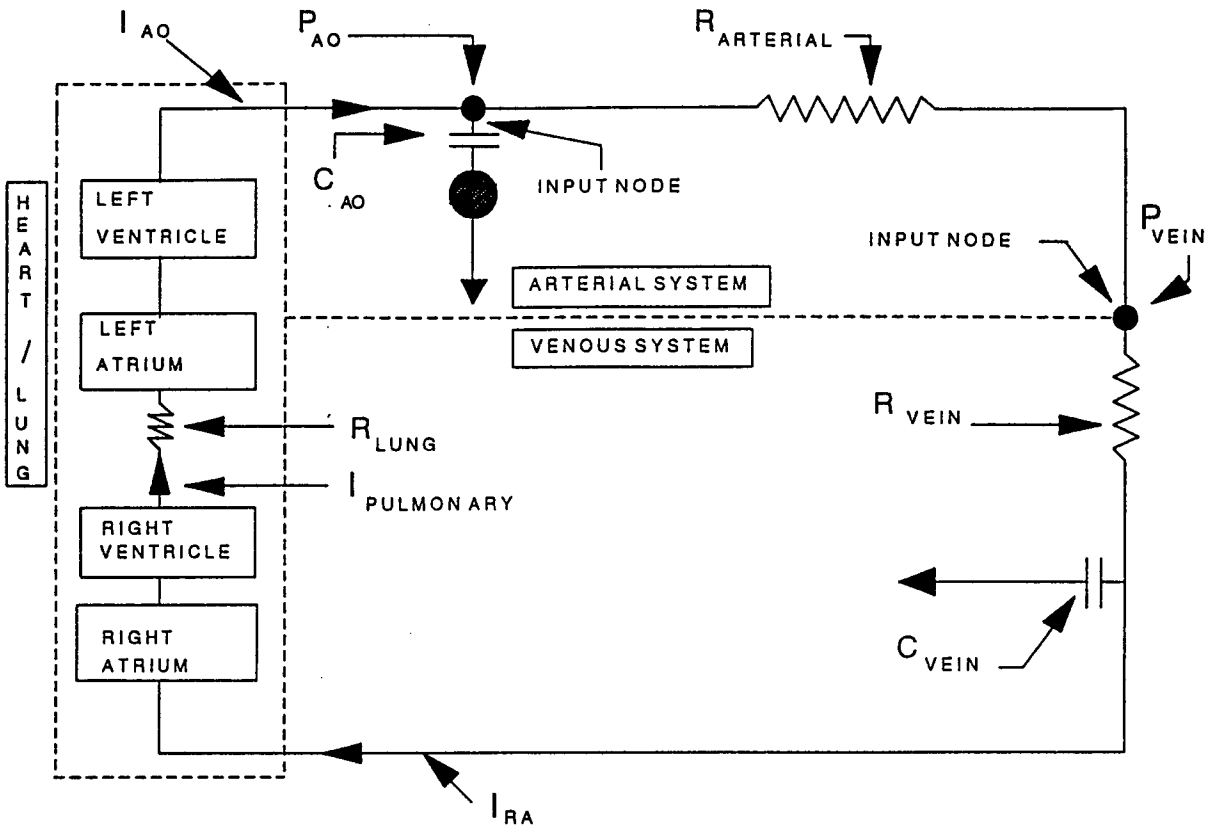


Figure 1

By applying Kirchhoff's current law to the input node and translating the result to cardiovascular terms we get

$$I_{ao} = C_{ao}\frac{d(P_{ao} - P_{pleural})}{dt} + (P_{ao} - P_{ra})/R_{arterial} \qquad (1)$$

where,

$$C_{ao}\frac{d(P_{ao} - P_{pleural})}{dt}$$

represents that portion of the aortic flow which goes into charging the aortic capacitance (P_{ao} = aortic root pressure, $P_{pleural}$ = extravascular thoracic pressure) and $(P_{ao} - P_{ra})/R_{aterial}$ is the flow

through the resistor. (P_{ra} = right atrial pressure, is taken to be equal to venous pressure after correcting for hydrostatic offset). The computational details of the procedure for calculating TPR and SAC are given in reference 3. It is noted here that the evaluation of these parameters requires continuous recordings of aortic root pressure, aortic root flow and right atrial pressure.

The proposed model uses the same basic elements as above for studying the effects of hydration states but distributes them into three regions as shown in Figure 2.

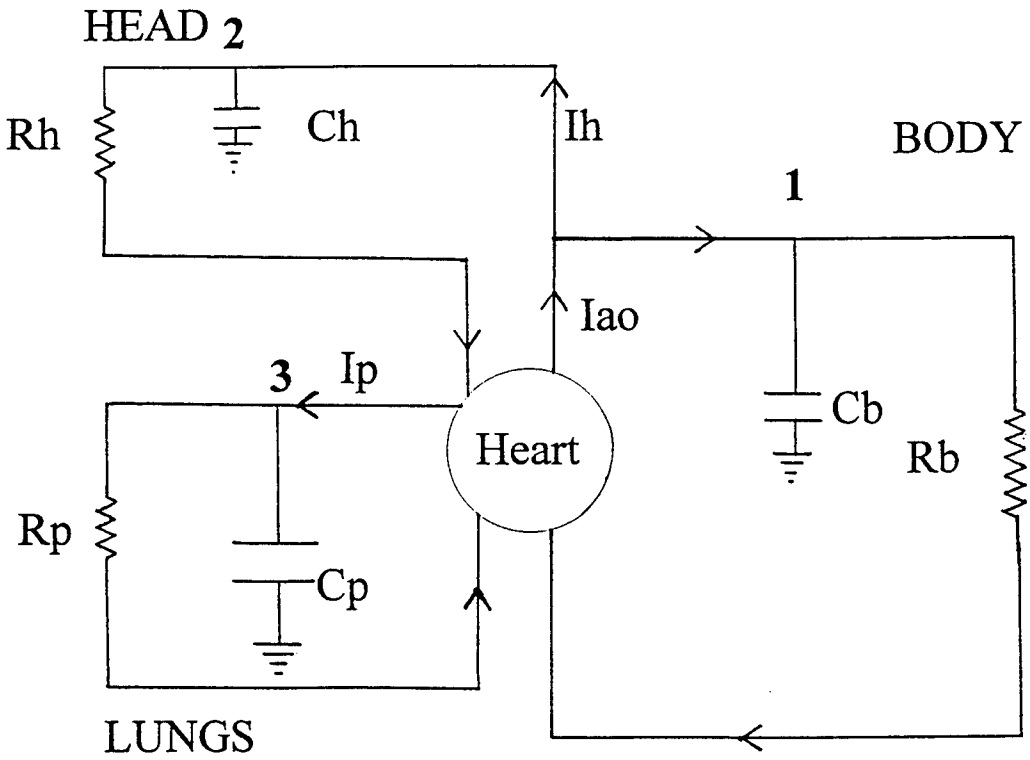


Figure 2

Thus, the head, lungs and body are each represented by a two element Windkessel. This is the simplest model suitable for studying fluid volume shifts and changes in vascular bed properties. It requires the continuous measurement of aortic root flow (I_{ao}) as well as flow to the head (I_h) and Pulmonary artery flow (I_p). Required pressure measurements include: aortic root pressure, Pulmonary Artery pressure, plural pressure and right and left atrial pressures.

Applying Kirchhoffs current law to nodes 1, 2 and 3 of Figure 2 and translating the results into cardiovascular terms we get.

for node 1

$$I_{ao} - I_h - C_b \frac{d(P_{ao} - P_{pleural})}{dt} + (P_{ao} - P_{ra})/R_b \qquad (2)$$

for node 2

$$I_h = C_h \frac{d(P_{ao} - P_{cran})}{dt} + (P_{ao} - P_{ra})/R_h \qquad (3)$$

for node 3

$$I_p = C_p \frac{d(P_p - P_{pleural})}{dt} + (P_p - P_{la})/R_p \qquad (4)$$

Where P_{cran} = cranial pressure and P_{la} = left atrial pressure and C_b C_n, and C_p are compliances for the arterial components of the body, head and lungs respectively. R_b, R_h and R_p are the corresponding resistive elements. With corrections for hydrostatic offset and the

measurements mentioned above, the procedure used in our earlier two element model may now be applied to each model equation to generate beat to beat values for C_b, C_n, C_p, R_b, R_n and R_p. Furthermore, by integrating the flow rates I_{ao}, I_h and I_p and subtracting the results we can quantify blood volume shifts (between the three regions of the cardiovascular system) over the course of transient +Gz load variations.

The added information available with this model comes with a price. In our first model only two pressures and one flow rate needs to be measured. In our proposed model it is necessary to continuously measure 6 pressures and 3 flows. It might be possible to reduce the number of measured pressures by making some assumptions about changes in pleural and cranial pressures which are relatively small compared with aortic and pulmonary artery pressures. Even if these assumptions can be made, four pressures and three flows will be needed and thus surgical limitations must be considered. Personal communications with Dr. John Fanton of the Armstrong Laboratory indicate that the only practical problems to be overcome are associated the measurement of pulmonary artery flow and pressure and flow and pressure in the aorta at the same time. The problem is basically one of "real estate". There is not enough room to place all of the needed transducers. Thus, there is a need for a new transducer which is smaller than the ones we have used in the past. The development of such a transducer is the goal of an AFOHSR proposal from Dr. Dan Ewert (4). Other goals of this proposal include modified surgical techniques intended to dramatically

reduce recovery time and costs.

We have completed in-vito testing of a relatively new Triton active-redirectional transit time (ART^2) flow transducer to determine if it is suitable as a basic foundation for further development. Our intention is to incorporate a pressure transducer in the housing of the flow transducer to allow simultaneous extravascular measurement of pressure and flow.

Two adult rhesus monkeys were instrumented with Triton ART^2 probes and flow readings were recorded (and calibrated against Thermal Dilution (TDL) measurements) for 5 days over a 3 month period. On each day a total of fifteen recordings took place: 5 baseline TDLs, 5 TDLs after administering nitropruside and 5 TDLs after administering phenelephrine. The drugs were used to produce a wide cardiac output range.

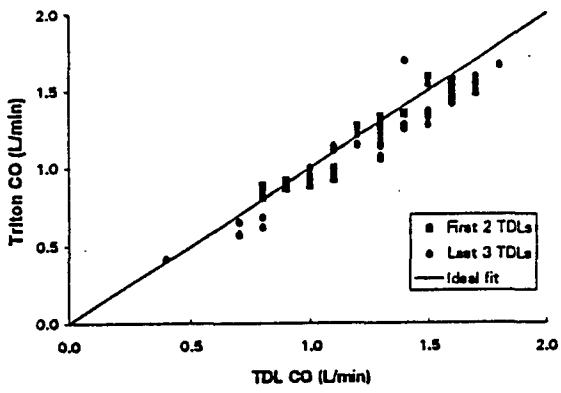


Figure 3

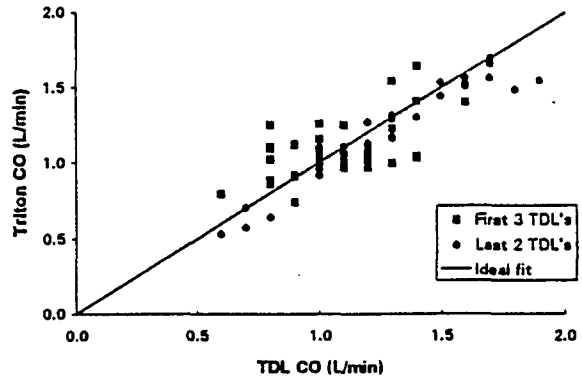


Figure 4

The results of the study are shown in Figures 3 and 4. Figure 3 is for one test animal and 4 is for the other. In each figure a straight line with a slope of 1 is drawn through the origin. The data generally fit the line fairly well. It should be noted that the accuracy of the TDL method is about ±15% which could account for some of the scatter shown. Bench tests of the same flow probes using a bucket and stop watch method for calibration showed them to be linear and accurate to ±10% over the flow range shown in the figures.

Wet Lab Verification of Model

TPR and SAC values computed with our first model are comparable to those found for the steady state condition computed by other methods. This agreement increases our confidence in the validity of our approach but since there is no way to independently measure these parameters in-vivo we cannot be 100% sure of our results. One way of approaching a 100% confidence level is to compare predicted values of TPR and SAC with known (measured) values in an in-vitro setting.

During this research period we have continued the development of a "Wet Lab" (5) which can be used as a test bed for evaluating transducers under controllable and realistic conditions. The facility is now developed to the point where it can be used to validate our transient 2-element windkessel model method. We can program a servo controlled pump system to produce realistic aortic

pressure waveforms. Figure 5 shows measured left ventricular and aortic pressures as well as the pump piston displacement (LVDT) as functions of time.

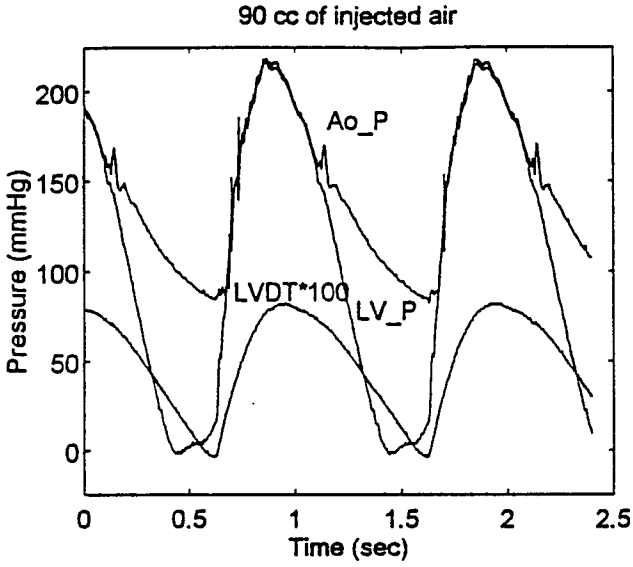


Figure 5

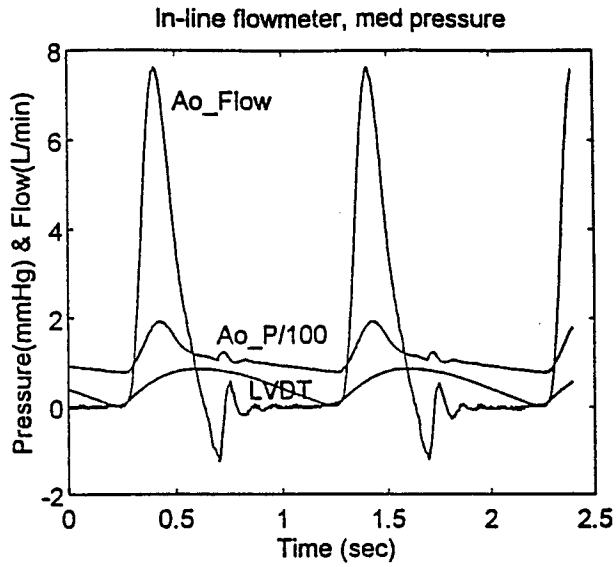


Figure 6

While the pressure peaks are a bit high, the waveforms are realistic and we have learned that the peaks are influenced by the amount of air trapped in the pump system "left ventricle". Future efforts include introducing a larger air cushion to reduce the peak pressure to the 130 to 140 mm Hg range. Figure 6 shows aortic flow superimposed on the pressure waveform and the LVDT trace. The flow is very realistic. Our pump facility has many adjustable parameters and we have not yet found a combination which simultaneously produces realistic pressure and flow waveforms simultaneously. We do however, understand the influences and should be able to do so

with little additional work.

To complete the in-vitro validation we need to construct an experimental model of the two element Windkessel using elastic tubing for the capacitance and an adjustable needle/plug valve for the resistance. For realistic simulation the tubing should have a compliance in the range of 0.5 to 2.0 cc/mmHg and the resistance range should be adjustable from about 1000 to 5000 dyne sec/cm^5. Figures 7 and 8 show the calibration of candidate tubing and resistance valves. Three lengths of latex rubber tubing were tested (1,2 and 3 ft). For pressures in the range from 50 to 150 mmHg the compliance is a linear function of pressure and length. One would expect compliance to be a linear function of length for any pressure range and our data seem to bear this out for the limited range of the tests. A 1 to 2 foot length of this tubing should be appropriate as the human capacitance equivalent for our two element Windkessel model of SAC.

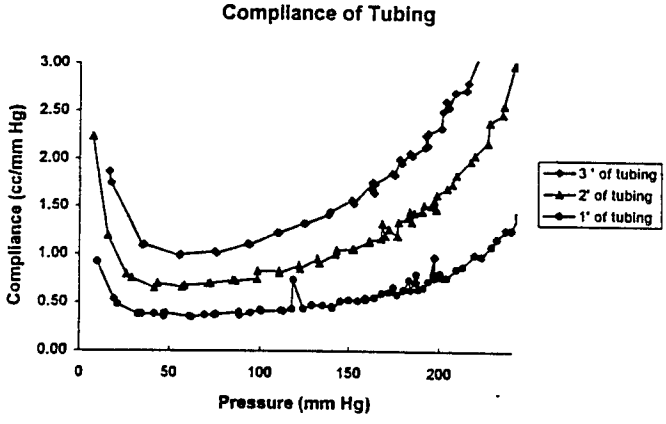


Figure 7

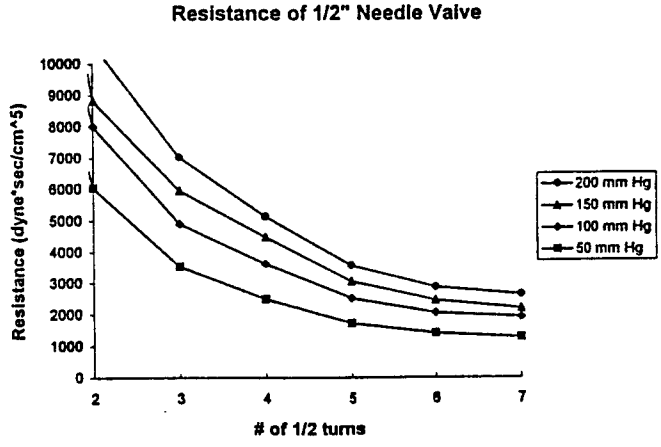


Figure 8

Figure 8 shows that a 1/2" needle valve opened two to three turns gives an appropriate resistance for a range of pressure drops from 50 to 200 mm Hg. Even though the resistance is a function of pressure drop it is expected to be fairly constant over the course of a beat because of the windkessel effect. It is expected that the valve will see an input pressure approximately equal to the mean "aortic pressure".

A student (Jeremy Schaub) from Trinity University will continue to work on this project part time during the 94/95 academic year. He will construct a flow loop model of the arterial system shown in Figure 1 (i.e. the 2 element windkessel) and measure the input node pressure and flow variations for pump produced aortic pressures and flows. These data will then be used in conjunction with our beat to beat method (2) for computing SAC and TPR. A comparison of the computed and measured values will determine the "goodness" of our method.

If our method is proven valid (we think it will be) a flow loop equivalent of Figure 2 will be constructed on a tilt table and used to study regional fluid volume shifts and regional changes in SAC and TPR.

Conclusions

The two element windkessel is an adequate but simple model for describing resistance and compliance changes associated with

transient +Gz loads. By representing the head, lungs and remaining body as a combination of three two-element windkessels regional fluid volume shifts under transient +Gz loads may be determined.

The Triton ART^2 flow probe is suitable for our needs to measure arterial blood flow. It also is suitable for adaptation to measure pressure but additional development is needed.

References

1. Nichols, W.W., O'Rourke, M.F., McDonald's Blood Flow in Arteries, Lea and Febiger, 3rd Edition, 1990.

2. Toorop, G.P., Westerhof, N., and Elzinga, G., Beat-to-Beat Estimation of Peripheral Resistance and Arterial Compliance During Pressure Transients, Am. J. Physiol., 1987, 252: H1275-H1283.

3. Self, D.A., Ewert, D.L., Swope, R.D., Crisman, R.P. and Latham, R.D., Beat-to-Beat Determination of Peripheral Resistance and Arterial Compliance During +Gz Centrifugation, Aviation, Space and Environmental Medicine, May 1994.

4. Development of an Extravascular Pressure/Flow Transducer, AFOSR Summer Research Extension Program Proposal; submitted by Dr. Dan Ewert, August, 1994.

5. Development of an Enhanced Hydraulic Cardiovascular Model/Test Apparatus for In-Vitro Simulations in Altered-G Environments. R.D. Swope, Final Report, USAF/RDL Summer Faculty Research Program, September, 1991.

LASER INDUCED BUBBLE FORMATION IN THE RETINA

Bernard S. Gerstman
Associate Professor
Department of Physics

Florida International University
University Park
Miami, FL 33199

Final Report for:
Summer Faculty Research Program
Laser Laboratory, Armstrong Laboratory
Brooks Air Force Base, TX

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

and

Armstrong Laboratory

September 1994

LASER INDUCED BUBBLE FORMATION IN THE RETINA

Bernard S. Gerstman
Associate Professor
Department of Physics
Florida International University
University Park
Miami, FL 33199

Abstract

The immediate thermodynamic effects of absorption of a laser pulse in the retina were theoretically investigated. The absorption occurs in a retinal pigment epithelium modeled as an aqueous environment with absorption occurring at small spherical sites with absorption coefficients representative of melanosomes. For laser pulse durations of less than 10^{-6} seconds, heat conduction is negligible during energy deposition and the resulting large energy density in the melanosome will cause vaporization of the surrounding medium. We develop expressions for calculating the size of the bubbles produced as a function of laser characteristics and melanosome properties. We also show that for pulse durations between 10^{-6} to 10^{-9} seconds, bubble formation will occur for laser fluences that are smaller than those required to cause Arrhenius type thermal damage. Therefore bubble formation is likely to be the source of threshold damage to the retina for laser pulse durations in this regime.

LASER INDUCED BUBBLE FORMATION IN THE RETINA

Bernard S. Gerstman

INTRODUCTION

The theoretical research described here was undertaken in order to understand the primary effects of the laser energy immediately after absorption in the retina. It is well known that damage can occur in cells due to temperature rises. This thermal damage has been modeled [1,2,3] in terms of an Arrhenius type activation process. However, it is also known that damage can occur on the cellular level due to bubble formation [4]. Significant heat conduction away from a melanin granule requires timescales of the order of microseconds. For pulses of shorter duration than this, at the end of the pulse most of the pulse energy is still localized at the absorbing melanosome and temperature rises are expected to be high enough to cause vaporization of the immediate surrounding medium. This will create a bubble that then expands outward from the melanosome. In this paper, expressions are developed to calculate the maximum size expected for the expanding bubble as a function of the laser pulse parameters and properties of the melanosomes and surrounding cellular medium. Tables are included with a range of representative cases. We also show that for pulse durations of 10^{-6}-10^{-9} seconds, damage due to bubble formation and growth will occur at fluences (J/cm^2) lower than those needed for Arrhenius thermal damage. This implies that the mechanism for threshold damage in this pulse duration regime is bubble formation.

Functional impairment to the visual system will occur if there is damage to photoreceptors or their associated nerve transmission pathways. This can occur as a secondary effect of damage that occurs initially in other cellular layers in the retinal region. In this paper we investigate damage due to bubble formation in the RPE. We refer the reader to Ref. [5] for discussions of other damage mechanisms which are not relevant to threshold damage for submicrosecond pulses.

The most important site for retinal damage is the retinal pigment epithelium which absorbs approximately 50% of incident visible radiation [6], an order of magnitude more energy than absorbed by the photoreceptors. This strong absorption by the melanosomes makes the RPE the likely location for the source of temperature rises and bubble formation that can lead to damage to the retina at threshold levels of irradiance [7,8]. Evidence that near threshold damage is centered in the RPE was observed by Gueneau, et. al. [9].

ANALYSIS OF BUBBLE FORMATION AND GROWTH

In our model the absorption of light occurs in melanosomes which are represented as spheres described by two parameters that can be varied: the radius R_a, and the absorption coefficient α.

These absorbing spheres are also given the following thermal characteristics: the specific heat of melanin of c_m=2.51 J/g·°C, and the density of melanin of ρ_m=1.35 g/cm^3 [10]. These melanosomes are embedded in a surrounding cellular medium that has the thermal characteristics of water; heat capacity c=4.19 x 10^3 J/kg·°C, density $\rho=10^3$ kg/m^3, thermal conductivity κ=0.57 J/m·s·°C.

Thermodynamic Conditions of Bubble Growth

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

$$v(\text{thermal}) \approx \kappa/\rho cL \approx 0.1 \text{ m/s} \qquad (1)$$

During a laser pulse of $\tau_p < 10^{-6}$ seconds, heat conduction occurs over a negligible distance from the melanosome and thus negligible heat loss occurs during absorption and subsequent bubble growth.

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

$$\frac{V}{V_o} = \left(\frac{P_o}{P}\right)^{1/\gamma} \qquad (2)$$

where V_o and P_o are the initial volume and pressure of the bubble at the end of the laser pulse when it starts its adiabatic expansion, and γ is the ratio of the specific heat of the vapor at constant pressure to the specific heat at constant volume. Equation (2) shows that the maximum radius attained by the bubble after expansion can be calculated if the initial radius of the bubble surrounding the melanosome at the end of the laser pulse can be determined. The radius of the melanosome itself will be denoted by R_a (or 'a'). The radius of the bubble and the pressure within will be denoted by r and P, and their values immediately after laser absorption will be r_o and P_o. The volume of the vapor at any time is therefore

$$V(r) = \frac{4\pi}{3}\,(r^3 - R_a^3) \qquad (3)$$

Combined with Eq. (2), this gives

$$r^3 = R_a^3 + (r_o^3 - R_a^3)(P_o/P)^{1/\gamma} \qquad (4)$$

which can be rewritten in terms of dimensionless quantities as

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

Equation (5) is an expression for the radius of a bubble as a function of the pressure of the vapor within the bubble.

In order to determine r_o and P_o, we follow the work of Cleary [11] in which the initial pressure P_o at the end of the laser pulse will be taken to be the critical pressure of water of 218 atmospheres. The reasoning behind this is as follows. The kinetics of vaporization is a non-equilibrium process, but eventually equilibrium will be reached between the vapor phase within the bubble and the liquid phase surrounding it. Since there are two distinct phases, the temperature and pressure cannot go above the critical values, which for water are T_c=374 C and P_c=218 atmospheres (221 bars) with a critical density of ρ_c=0.315 g/cm^3 [12]. Since the rate of energy input is faster than either the expansion of the bubble or the heat conduction rate, the critical conditions will at some time be reached if enough energy is absorbed by the melanosome for it to reach 374 C. Therefore, bubble formation can be treated as a process in which the laser energy absorbed by the melanosome creates a saturated vapor with T_o=374°C and P_o=218 atmospheres and whose initial radius r_o is determined by the total energy absorbed by the melanosome. Bubble growth then occurs in an adiabatic expansion that is rapid compared to heat loss. (Note that the maximum volume of the bubble depends on the product of P_oV_o and that the energy of the laser pulse is used to both vaporize cellular fluid in creating V_o, and increase the initial pressure P_o. Thus, for a given amount of energy, if the initial non-equilibrium vaporization process does not raise the pressure to a P_o of the full 218 atmospheres, there will be additional energy available to vaporize more cellular fluid leading to a larger V_o. This tends to limit the variation in the product P_oV_o and thus the final volume (and radius) are not especially sensitive to the actual value used for P_o.)

In order to justify the adiabatic treatment of the bubble growth, the velocity of expansion must be much greater than the rate of heat loss. The characteristics of the expansion can be determined by following the treatment of Lamb [13] for the rate of expansion of the bubble radius

$$\dot{r}^2 = \frac{2c_o^2}{3(\gamma-1)}\left[\left(\frac{r_o-R_a}{r-R_a}\right)^3 - \left(\frac{r_o-R_a}{r-R_a}\right)^{3\gamma}\right] \qquad (6)$$

where $c_o=\sqrt{P_o/\rho}$ and ρ is the density of the liquid. The value of r at which the speed of expansion is a maximum is obtained

from Eq. (6) by taking the derivative of $\dot{r}$ with respect to r, and gives

$$(r-R_a) = \gamma^{1/(3\gamma-3)} (r_o-R_a) \qquad (7)$$

The maximum rate of expansion that occurs at this value of r is (the following equation is the corrected version of Eq. (12), p. 123 of Lamb, and Eq. (8) in Cleary [11])

$$\dot{r}_{max}^2 = \frac{2}{3} c_o^2 \gamma^{\gamma/(1-\gamma)} \qquad (8)$$

Finally, the time at which a bubble reaches a radius r during its growth phase can be gotten from Eq. (6)

$$t = \frac{r_o-R_a}{c_o} (2Z)^{1/2} \left(1 + \frac{2}{3}Z + \frac{Z^2}{5}\right) \qquad (9)$$

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

If we insert $\gamma=4/3$ [11,13] into Eq. (8), the maximum speed of bubble growth is $\dot{r}_{max}=0.46c_o$. Using $P_o=218$ atmospheres and $\rho=1 g/cm^3$ gives $c_o \approx 150$ m/s and $\dot{r}_{max} \approx 70$ m/s, which is more than two orders of magnitude larger than the characteristic thermal conduction rate of Eq. (1). Thus, expansion occurs on a timescale that is much shorter than heat loss and this justifies the use of an adiabatic treatment during expansion. This can also be seen from Eq. (9). If we use representative values of $R_a=10^{-6}$m, $r_o=2R_a$, the time it takes a bubble to grow to $2r_o$ is on the order of 10^{-7} seconds.

BUBBLE SIZE AS A FUNCTION OF LASER FLUENCE

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

Using Eq. (5) to get the maximum bubble size we obtain

$$(r_m/R_a)^3 = 1+[(r_o/R_a)^3-1](P_o/P_{min})^{1/\gamma} \qquad (10)$$

The minimum pressure, at which the bubble stops expanding, is taken to be the ambient pressure of one atmosphere and this will be the same for all bubbles. In actuality, the outward momentum of the liquid cellular medium may cause an overshoot in which the bubble's vapor has a pressure of less than one atmosphere. However, this will have a small effect on r_m since the bubble radius has a dependence on pressure of $r_m \propto P_{min}^{-1/3\gamma} = P_{min}^{-1/4}$. For example, an overshoot in which the pressure drops to ½ atmosphere will result in a final bubble radius that is only 19%

larger than if the pressure goes no lower than one atmosphere. Furthermore, this inertial tendency for overshoot tends to be counterbalanced by energy loss during expansion from viscous forces. We therefore use Eq. (10) with P_{min}=1 atmosphere=$1.013 \times 10^5 N/m^2$.

Using P_o=218 atmospheres, the ratio P_o/P_{min} is equal to 218 for all laser pulses that have sufficient fluence to raise the melanosome to T_c=374°C. Therefore, Eq. (2) tells us that the maximum volume of the vapor in the bubble is $218^{3/4}$=56.7 times larger than V_o, with the melanosome occupying constant volume inside the bubble. This will be true even if the melanosome breaks apart during the process, as long as the pieces remain inside the bubble. In order to calculate r_m from Eq. (10) therefore requires only an expression for r_o/R_a, which depends on the energy absorbed by the melanosome.

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

$$V_o = \frac{4\pi}{3}(r_o^3 - R_a^3) = \frac{E - E_m}{q} \times \frac{1}{\rho_c} \qquad (11)$$

where E_m is the energy required to raise a melanosome from 37°C to 374°C and is equal to $E_m = c_m \rho_m (4\pi/3\ R_a^3)\Delta T$. For $R_a = 10^{-6}$m, this gives a value of $E_m = 4.8 \times 10^{-9}$ J.

The calculation for r_o continues by evaluating the energy absorbed by the melanosome. For a path length of d, through a material with an absorption coefficient α, the fraction of light absorbed is $(1-e^{-\alpha d})$. If H_o is the fluence of the laser in J/cm^2, then the energy incident on a spherical melanosome is $E_o = \pi a^2 H_o$, where $a \equiv R_a$ is the radius of the melanosome. The rigorous expression for the total energy absorbed by a spherical absorber of radius a and absorption coefficient α is derived in Appendix II and is given by

$$E = E_o - E_T = \pi a^2 H_o \left[1 - \frac{1}{2\alpha^2 a^2}\left(1 - e^{-2\alpha a}(1+2\alpha a)\right)\right] = C(\alpha, a)\pi a^2 H_o \qquad (12)$$

Using representative values for α of melanin [14,15] gives $C(\alpha,a)$ for a melanosome of

$$C(1000\ cm^{-1}, 10^{-6}m) = 0.124 \qquad C(1800\ cm^{-1}, 10^{-6}m) = 0.210$$

$C(\alpha,a)$ can be approximated to second order by $C(\alpha,a) \approx \frac{4}{3}\alpha a - (\alpha a)^2$ giving for the total energy absorbed

$$E \approx [\frac{4}{3}\alpha a - (\alpha a)^2]\, \pi a^2 H_o \qquad (13)$$

This approximate expression for $C(\alpha,a)$ is accurate to within 1.5% for $\alpha a = .18$ and accurate to 0.5% for $\alpha a = .10$.

We can use these numbers to get an approximate value for the minimum fluence necessary to produce a bubble in the manner described above in which the rapid and intense heating of the melanosome leads initially to a thin shell of vapor close to the critical point. For a 1μm radius melanosome, we find from Eq. (12) or Eq. (13) that for $\alpha = 1000\ cm^{-1}$, the energy absorbed by a melanosome is $E = 3.9\times10^{-9} cm^2 H_o$. Using this in Eq. (11) leads to an initial volume that is vaporized of

$$V_o = \frac{4\pi}{3}(r_o^3 - 10^{-12} cm^3) = \frac{3.9\times10^{-9} cm^2 H_o - 4.9\times10^{-9} J}{q} \times \frac{1}{.315 g/cm^3} \qquad (14)$$

Equation (14) represents the scenario in which any energy above $E_m = 4.8\times10^{-9}$ J absorbed by a 1μm melanosome is used to produce an initial bubble with the vapor at the critical point. In order for at least 4.8×10^{-9}J to be absorbed, Eq. (14) shows that the fluence hitting the melanosome must be at least

$$H_o^{min} = 4.8\times10^{-9} J / 3.9\times10^{-9} cm^2 = 1.23\ J/cm^2 \quad (\alpha = 1000\ cm^{-1}) \qquad (15a)$$

If the same calculation is done with $\alpha = 1800\ cm^{-1}$, the energy absorbed becomes $E = 6.6\times10^{-9} cm^2 H_o$ and there is a decrease to

$$H_o^{min} = 0.73\ J/cm^2 \quad (\alpha = 1800\ cm^{-1}) \qquad (15b).$$

Two comments must be made concerning these H_o^{min}:

1) The value of H_o^{min} will vary as a function of several factors; wavelength (α), shape, and size (volume of melanosome to be heated depends on a^3 but absorbed energy depends on $C(\alpha,a)$ of Eq. (12) which has a complicated dependence on αa).

2) A fluence of less than H_o^{min} does not mean that no bubble is formed but instead that the initial conditions of the bubble are less extreme than those of a vapor with the critical values of $T_o = 374°C$ and $P_o = 218$ atmospheres, and initial bubble formation cannot be treated in the manner leading to Eqs. (14) and (15). However, the values of H_o^{min} just calculated are very close to the experimental values for ED_{50} measurements of a retinal fluence of approximately 1 J/cm^2 for pulse durations less than 10^{-6} seconds. This shows that this treatment of bubble formation

and growth is appropriate for analyzing threshold damage leading to minimum visible lesion (MVL).

In order to get an expression for the maximum radius attained by the bubble, r_m, we continue with our analysis. Equation (10) shows that r_m/R_a depends on $[(r_o/R_a)^3-1]$. Using Eq. (12) for the energy absorbed, and the expression for E_m following Eq. (11) (with R_a given in cm) for the energy needed to raise the temperature of a melanosome from 37°C to 374°C, Eq. (11) leads to

$$\left(\frac{r_o}{R_a}\right)^3 - 1 = \frac{1}{q\rho_c} \times \left[\frac{3}{4}\frac{C(\alpha a)H_o}{R_a} - c_m\rho_m\Delta T\right] \qquad (16)$$

Using this expression in Eq. (10) gives

$$\left(\frac{r_m}{R_a}\right)^3 = 1 + \frac{1}{q\rho_c} \times \left[\frac{3}{4}\frac{C(\alpha a)H_o}{R_a} - c_m\rho_m\Delta T\right]\left(\frac{P_o}{P_{min}}\right)^{1/\gamma} \qquad (17)$$

with R_a in cm, α in cm^{-1}, and H_o the retinal fluence in J/cm^2. Using the rigorous expression for $C(\alpha,a)$ from Eq. (12) gives

$$\left(\frac{r_m}{R_a}\right)^3 = 1 + \frac{1}{q\rho_c} \times \left[\frac{3}{4R_a}\left[1 - \frac{1}{2\alpha^2a^2}\left(1 - e^{-2\alpha a}(1+2\alpha a)\right)\right]H_o - c_m\rho_m\Delta T\right]\left(\frac{P_o}{P_{min}}\right)^{1/\gamma} \qquad (17a)$$

A quicker estimate can be made using the simple expansion for $C(\alpha,a)$ given in Eq. (13), which results in

$$\left(\frac{r_m}{R_a}\right)^3 = 1 + \frac{1}{q\rho_c} \times \left[\left(\alpha - \frac{3}{4}\alpha^2R_a\right)H_o - c_m\rho_m\Delta T\right]\left(\frac{P_o}{P_{min}}\right)^{1/\gamma} \qquad (17b)$$

COMPARISON OF THRESHOLD FLUENCES (H_o) FOR BUBBLE DAMAGE VERSUS THERMAL DAMAGE

By comparing the fluence needed for bubble damage to the fluence needed for thermal damage we now show that for laser pulse durations in the nanosecond to microsecond range, bubble formation is the mechanism that determines the damage threshold fluence. The values graphed in Fig. 1 are retinal fluences. With the focusing power of the eye of approximately 10^5, the equivalent corneal fluences can be obtained by multiplying the retinal fluence by 10^{-5}.

We used the following criteria for calculating and comparing damage fluences. First, we assume that a bubble causes damage only out to distances from the surface of a melanosome that the

bubble has actually expanded into. Thus we are under estimating bubble damage by ignoring any damaging effects due to compression and movement of the cellular material outside the actual volume cleared out by the bubble. Next, we assume that cellular damage leading to MVL will occur if primary damage (bubble or thermal) occurs at a distance of 1 μm from the surface of 1 μm melanosome (i.e. r_m=2 μm from the center of a melanosome). The choice of this distance is not based on any specific experimental observations but is based on the following conservative reasoning:

1)Assume an RPE cell is 6000 μm^3 in volume (15μmx20μmx20μm).
2)An RPE cell contains approximately 100 melanosomes.
3)The average volume of a melanosome is approximately 4.2μm (sphere of radius 1μm) so melanosomes occupy 420 μm^3 which is 0.07 of the total volume of a cell.
4)If damage (bubble or thermal) occurs out to a radius of 2 μm than each spherical damage zone now occupies a volume of 113 μm^3.
5)100 of these damage spheres, if they did not overlap, would occupy 11300 μm^3 which is greater than the cell's total volume.
6)Assume that this damage is more than enough to kill a cell, and thus overestimates the fluence needed for threshold damage.

In Figure 1 we plot three curves. The curve labeled bubble damage is the retinal fluence H_o needed to produce a bubble that expands out 1μm from the surface of a 1μm radius melanosome with an absorption coefficient α=1000 cm^{-1}. The fluence is calculated using the near critical point bubble formation described above which leads to Eq. (16a) and gives a fluence of 1.4 J/cm². This curve is only plotted in the regime in which it is valid; laser pulses of durations between 10^{-6} and 10^{-9} seconds, and in this regime the required fluence is constant. If α=1800 cm^{-1}, this curve drops to H_o=0.79 J/cm².

The curve labelled Arrhenius Thermal Damage is the retinal fluence needed to produce thermal damage at a distance of 1μm from the surface of a 1μm melanosome, again with an absorption coefficient α=1000 cm^{-1}. The fluence is calculated by using an Arrhenius expression to model the thermal damage [1,2,3] in which damage occurs when

$$\sum C_1 e^{\frac{-C_2}{310K+\Delta T(t)}} \Delta t = 1 \qquad (18)$$

The values for C_1 and C_2 were taken from Takata [2]:

$C_1 = 4.322 \times 10^{64}$/sec $T \leq 323$ K
$C_2 = 50{,}000$ K

$C_1 = 9.389 \times 10^{104}$/sec $T > 323$ K
$C_2 = 80{,}000$ K

The fluence necessary to cause thermal damage is determined by Eq. (18) the time dependence of ΔT on the fluence. The details of the computations will be published in a separate paper.[16] It is important to note that for thermal damage, we looked at a point that is 1μm from the surface of a specific melanosome, but

that 99 other melanosomes were randomly placed by the computer code in a cellular volume of 6000 μm^3. Thus, as expected in actual experiments, the location received heat from other melanosomes that were further away than 1μm, but could still be relatively nearby. The result for this computation was that for pulses of duration less than 10^{-6} seconds, 2.1 J/cm^2 were needed to cause thermal damage, noticeably higher than the 1.4 J/cm^2 calculated for bubble damage.

Finally, the third curve in Fig. 1, labelled Vaporization Threshold, shows the retinal fluence needed to raise the temperature of a melanosome to 100 C, at which point bubble formation will begin, though not with the near critical point conditions discussed in this paper. This fluence of approximately 0.29 J/cm^2 is well below measured threshold ED_{50}'s and therefore is evidently insufficient for causing MVL. This curve does have the value however, of showing how heat conduction away from a melanosome plays an important role for time scales greater than 10^{-6} seconds. Thus for laser pulses of duration greater than a microsecond, significant energy conducts away from the melanosome into the cellular media during the time in which energy is being deposited. This is why thermal damage, which requires smaller temperature rises but that last for extended times, becomes easier to cause (smaller H_o) than bubble formation for pulse durations greater than 10^{-5} seconds.

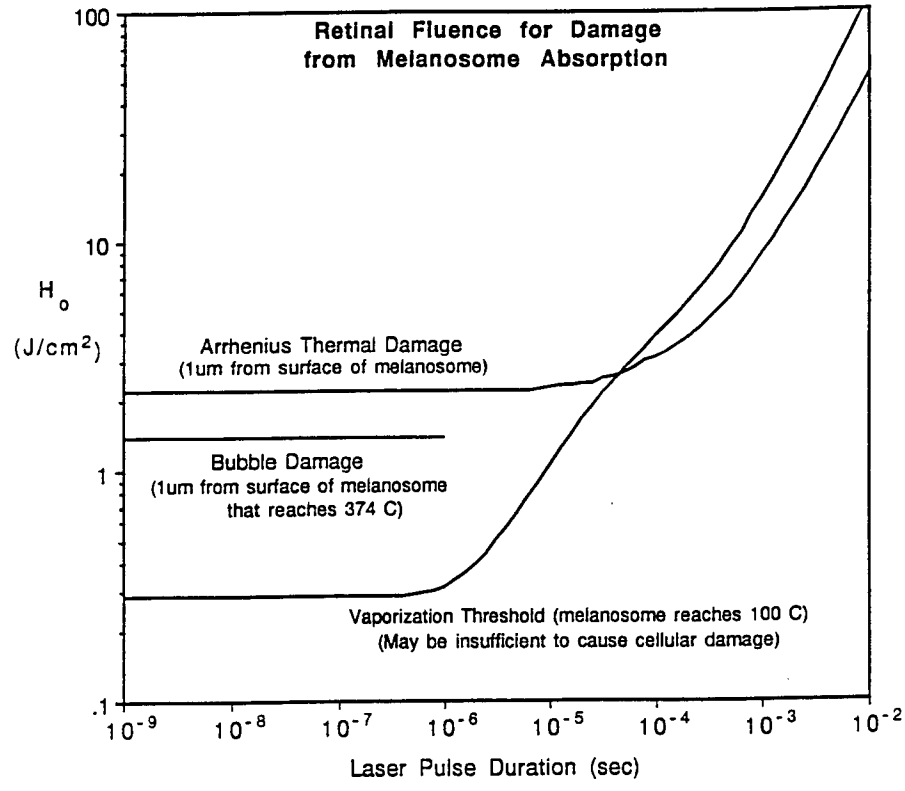


Figure 1. Retinal fluence necessary for damage caused by different mechanisms. Arrhenius thermal damage or bubble damage is at a location that is 1um from surface of melanosome of 1um radius. Absorption coefficient for melanosome is taken to be 1000 cm^{-1} Equivalent corneal fluence is approximately 10^{-5} x Ho.

RESULTS

Equations (17) show explicitly how r_m depends on the pulse fluence H_o, and on melanosome properties; α, R_a, and q. We first get a representative value for r_m by using representative values for all the parameters in Eq. (17a) as listed earlier in this report: $a \equiv R_a = 10^{-6}m = 10^{-4}cm$, $c_m = 2.51$ J/g·°C, $\rho_m = 1.35$ g/cm^3, $\Delta T = 374°C - 37°C = 337°C$, $\rho_c = 0.315$ g/cm^3, $P_o = 218$ atmospheres, $P_{min} = 1$ atmosphere, $\gamma = 4/3$, and $q = 2770$ J/g. As an example, for $H_o = 1.5$ J/cm^2, Eq. (17a) gives the following results

$$\alpha = 1000\ cm^{-1} \rightarrow r_m = 2.6\ \mu m\ , \qquad \alpha = 1800\ cm^{-1} \rightarrow r_m = 4.3\ \mu m$$

Tables I show how the maximum bubble radius, r_m, varies as a function of different parameters. Each Table shows the variation of r_m with the melanosome properties a and absorptivity α, for a specific laser fluences H_o. Entries of 0.0 do not mean that no bubble is formed but instead signify that the present model is not applicable because not enough energy was absorbed by the melanosome to raise the temperature of the melanosome to T_c.

CONCLUSIONS

The calculation leading to the results of Eqs. (15) is strong support for both the validity of the model used in this report and of the importance of bubble formation in causing minimal visible lesions (MVL) for short laser pulses (10^{-6}-10^{-9} seconds) incident on the eye. Our calculations show that bubble damage will occur for a retinal fluence of 1.4 J/cm^2 for an "average" melanosome with a radius of 1 μm and $\alpha = 1000\ cm^{-1}$, or only 0.79 J/cm^2 if $\alpha = 1800\ cm^{-1}$. These calculated H_o^{min} agree well with the experimental ED_{50}'s measured for short pulses, which are found to be approximately 1 J/cm^2. This close agreement supports the idea that bubble formation is a cause of MVL in short pulses and that this model is a reasonable theoretical treatment for calculating the size of damage causing bubbles as a function of the relevant parameters.

The importance of various parameters for bubble growth can be ascertained from Eqs. (17) and the information in Tables I:

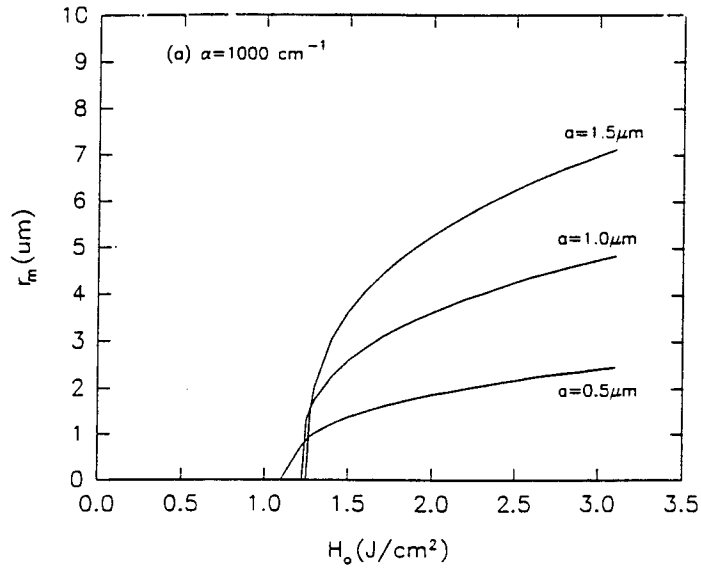
1) The dependence of r_m on the ratio of the pressures at the beginning and end of the bubble expansion is $(P_o/P_{min})^{-1/3\gamma}$. With $\gamma = 4/3$, this gives a weak dependence on $(P_o/P_{min})^{-1/4}$. Thus a major change in the ratio of P_o/P_{min}, such as by a factor of two, leads to a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

2) The dependence of r_m on the melanosome radius a, absorptivity α, and the fluence H_o is complex due to the non-linear way in which these parameters influence r_m. Thus, how strong a dependence r_m has on any one of these depends on the specific values of the other two. A few general trends are discernible:

   a) r_m increases monotonically with α and H_o, but not linearly, as seen in Tables I, and plotted in Figures (2a) and (2b). The dependence is better categorized as a threshold dependence, as expected from Eqs. (17).

   b) Because of its appearance in several terms in Eq. (17a), the

dependence on $a(R_a)$ is much more complicated, and not always monotonic. Even the ratio r_m/R_a is not monotonic and for certain values of α and H_o, increasing R_a leads to *decreasing* r_m. This occurs around threshold values for H_o; see the columns with α=1800 cm^{-1} in Table I.2 and 1400 cm^{-1} in Table I.3.



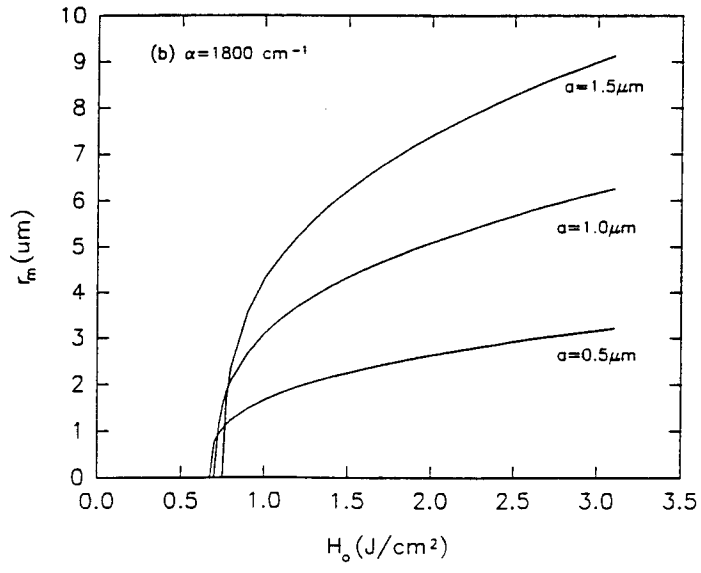


Fig. 2. Maximum bubble radius as a function of retinal fluence, showing a threshold dependence. Figs. (2a) and (2b) differ in the absorption coefficient used for the melanosome.

Table I. Maximum bubble radius r_m, calculated from Eq. (17a), as a function of melanosome radius 'a', melanosome absorptivity 'α', and laser retinal fluence 'H_o'. Other parameters: c_m=2.51 J/g·°C, ρ_m=1.35 g/cm^3, ΔT=337°C, ρ_c=0.315 g/cm$_3$, P_o/P_{min}=218, γ=4/3, q=2770 J/g.

Table I.1) H_o(J/cm^2): 0.50

		α(cm^{-1})			
	rm	600.	1000.	1400.	1800.
	0.50	0.00	0.00	0.00	0.00
	0.75	0.00	0.00	0.00	0.00
	1.00	0.00	0.00	0.00	0.00
a(μm)	1.25	0.00	0.00	0.00	0.00
	1.50	0.00	0.00	0.00	0.00
	1.75	0.00	0.00	0.00	0.00
	2.00	0.00	0.00	0.00	0.00

Table I.2) H_o(J/cm^2): 0.75

		α(cm^{-1})			
	rm	600.	1000.	1400.	1800.
	0.50	0.00	0.00	0.00	1.04
	0.75	0.00	0.00	0.00	1.38
	1.00	0.00	0.00	0.00	1.55
a(μm)	1.25	0.00	0.00	0.00	1.38
	1.50	0.00	0.00	0.00	0.00
	1.75	0.00	0.00	0.00	0.00
	2.00	0.00	0.00	0.00	0.00

Table I.3) H_o(J/cm^2): 1.00

		α(cm^{-1})			
	rm	600.	1000.	1400.	1800.
	0.50	0.00	0.00	1.18	1.65
	0.75	0.00	0.00	1.67	2.40
	1.00	0.00	0.00	2.07	3.09
a(μm)	1.25	0.00	0.00	2.38	3.71
	1.50	0.00	0.00	2.55	4.28
	1.75	0.00	0.00	2.53	4.77
	2.00	0.00	0.00	2.13	5.19

Table I.4) $H_o(J/cm^2)$: 1.25

		$\alpha(cm^{-1})$			
	rm	600.	1000.	1400.	1800.
	0.50	0.00	0.86	1.63	2.00
	0.75	0.00	1.15	2.38	2.92
	1.00	0.00	1.31	3.09	3.80
a(μm)	1.25	0.00	0.00	3.75	4.64
	1.50	0.00	0.00	4.35	5.42
	1.75	0.00	0.00	4.91	6.16
	2.00	0.00	0.00	5.41	6.85

Table I.5) $H_o(J/cm^2)$: 1.50

		$\alpha(cm^{-1})$			
	rm	600.	1000.	1400.	1800.
	0.50	0.00	1.37	1.92	2.25
	0.75	0.00	2.00	2.82	3.31
	1.00	0.00	2.59	3.68	4.32
a(μm)	1.25	0.00	3.13	4.51	5.29
	1.50	0.00	3.63	5.30	6.22
	1.75	0.00	4.07	6.04	7.11
	2.00	0.00	4.46	6.75	7.95

Table I.6) $H_o(J/cm^2)$: 2.00

		$\alpha(cm^{-1})$			
	rm	600.	1000.	1400.	1800.
	0.50	0.71	1.87	2.32	2.63
	0.75	0.97	2.76	3.43	3.88
	1.00	1.11	3.62	4.50	5.09
a(μm)	1.25	0.00	4.46	5.54	6.26
	1.50	0.00	5.27	6.55	7.39
	1.75	0.00	6.04	7.52	8.48
	2.00	0.00	6.79	8.46	9.53

Table I.7) $H_o(J/cm^2)$: 3.00

		$\alpha(cm^{-1})$			
	rm	600.	1000.	1400.	1800.
	0.50	1.72	2.43	2.85	3.17
	0.75	2.56	3.61	4.23	4.69
	1.00	3.38	4.76	5.57	6.17
a(μm)	1.25	4.18	5.89	6.89	7.60
	1.50	4.96	7.00	8.16	9.00
	1.75	5.73	8.08	9.41	10.36
	2.00	6.47	9.14	10.63	11.67

This paper presents a theoretical approach for calculating maximum bubble size formed in retinal pigment epithelium cells due to short laser pulses with pulse durations in the range of 10^{-6}-10^{-9} seconds. The agreement between the threshold energy for the formation of bubbles calculated by the model with the experimental ED_{50} shows the relevance of the model for damage assessment. A full understanding of the damage process requires additional work directed towards understanding the mechanisms by which bubbles actually cause cellular damage. These mechanisms obviously derive from the manner in which the physical expansion destructively couples to the functioning of a cell. Does the expansion destroy enough cellular proteins to cause immediate MVL or is the damage initially minor, but enough to prevent the cell's biochemical pathways from repairing the initially damaged areas, as well as disrupting transport channels. In order to distinguish between mechanisms such as these, or others, experiments must be done that look at immediate effects of the laser pulse. Immediate effects here mean as soon as the bubble has finished expansion which is on the sub millisecond time scale. This requires use of experimental methods such as pump-probe optical techniques that look for changes in protein absorption characteristics on these time scales.

Finally, we note that another paper [16] reports on a more accurate computational method for predicting temperature rises produced by relatively long laser pulses ($\tau > 10^{-5}$ seconds) for which heat is conducted away fast enough so that ED_{50}'s imply no vaporization. For laser pulses of duration in the range of 10^{-5}-10^{-6} seconds, the theoretical treatment will be complicated by the presence of both vaporization and conduction. This middle regime remains to be investigated in detail. Also, for pulse durations less than 10^{-9} the damage analysis of this paper may not be valid due to shock wave formation. In this sub-nanosecond pulse length regime of "stress confinement", mechanical waves created by the laser do not have time to leave the melanosome during the duration of the laser pulse and a significant fraction of the absorbed energy can be used in generating mechanical stress and shock waves, rather than bubble formation. This may be responsible for the apparent lowering of the ED_{50} threshold for damage for laser pulses shorter than 10^{-10} seconds.

APPENDIX I: Calculation of q

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

$$\Delta E = \Delta H - \Delta(PV) \qquad \text{(A1)}$$

where the energy change q is represented in Eq. (A1) by E, and H is the enthalpy of the process. Since the change in energy is a state function, we can use any path in P-T space to evaluate ΔE. We use a path in which first, at constant pressure of 1

atmosphere, water is raised from body temperature to the critical temperature. Since dH=dQ+VdP, during this constant pressure part of the process $\Delta H=Dq$ even though volume changes do occur. The heat required to raise water from body temperature of 37°C to 100°C is 263 J/g. The water is then transformed to vapor at 1 atmosphere pressure which requires heat of 2262 J/g. Heat is then used to raise the temperature of the steam from 100°C (373°K) to the critical temperature of 374°C (647°K). The heat capacity for steam at a constant pressure of 1 atmosphere increases over this temperature range and we use the following expression for c_p [17]:

$$c_p = a + (b \times 10^{-3})T + (c \times 10^{-6})T^2 \qquad (A2)$$

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

The remaining changes in energy needed for evaluating Eq. (A1) are: ΔH due to the step in which the temperature remains constant at 647°K and the pressure is increased isothermally from 1 atmosphere to 218 atmospheres, as well as the evaluation of $\Delta(PV)$ in Eq. (A1) for the entire process. Since the product PV is also a state function, we can ignore intermediate steps and evaluate $\Delta(PV)=P_cV_c-P_oV_o$ with P_c=218 atmospheres = 221 x 10^5N/m², V_c=3.17 cm³/g, P_o= 1 atmosphere = 1.013 x 10^5N/m², V_o=1 cm³/g. This gives $\Delta(PV)$=70 J/g. The ΔH for the isothermal compression of the steam at 647°C from P_o=1 atmosphere to P_c=218 atmospheres can be evaluated by rewriting Eq. (A1) as $\Delta H=\Delta E+\Delta(PV)$. If the steam behaved as an ideal gas during the isothermal compression then we would have $\Delta H=0$ since E=E(T) for an ideal gas and PV=RT, so $\Delta E=0$ and $\Delta(PV)=0$. In actuality the steam may behave as an ideal gas at P_i=1 atmosphere but does not behave as an ideal gas as the critical pressure is reached. Treating steam as an ideal gas at 1 atmosphere and using ρ_c=.315 g/cm³ allows the evaluation of $\Delta(PV)=P_cV_c-P_iV_i$ for this step: P_cV_c=70 J/g, and P_iV_i(ideal)=RT= 299 J/g. This gives $\Delta(PV)$=-230 J/g. The evaluation of ΔE is not as straightforward. During this isothermal compression process, work is done on the steam which tends to increase its energy. However, in order for the temperature to remain constant, this energy must be either lost to heat ΔH, or used in bond modifications. If the steam behaved as an ideal gas then its internal energy, which is purely kinetic for an ideal gas, would not change and this $\Delta E=0$ would imply $\Delta H=\Delta(PV)$. In a non-ideal gas however, there are bonds between molecules and it is possible to increase the energy of the system without a corresponding temperature increase. Thus, in compressing the steam isothermally, some of the energy put in as work can stay in the system and need not be lost as heat. Nevertheless, since we are compressing a vapor in which the interactions between molecules are much weaker than in liquids, and since a phase transformation does not occur during the compression, we will

assume that the change in internal energy is small since T remains constant and set $\Delta E=0$ for this isothermal compression. The actual value could be determined from the virial coefficients of steam under these conditions of temperature and pressure. With $\Delta E=0$, we have $\Delta H=\Delta(PV)=-230$ J/g for the isothermal compression.

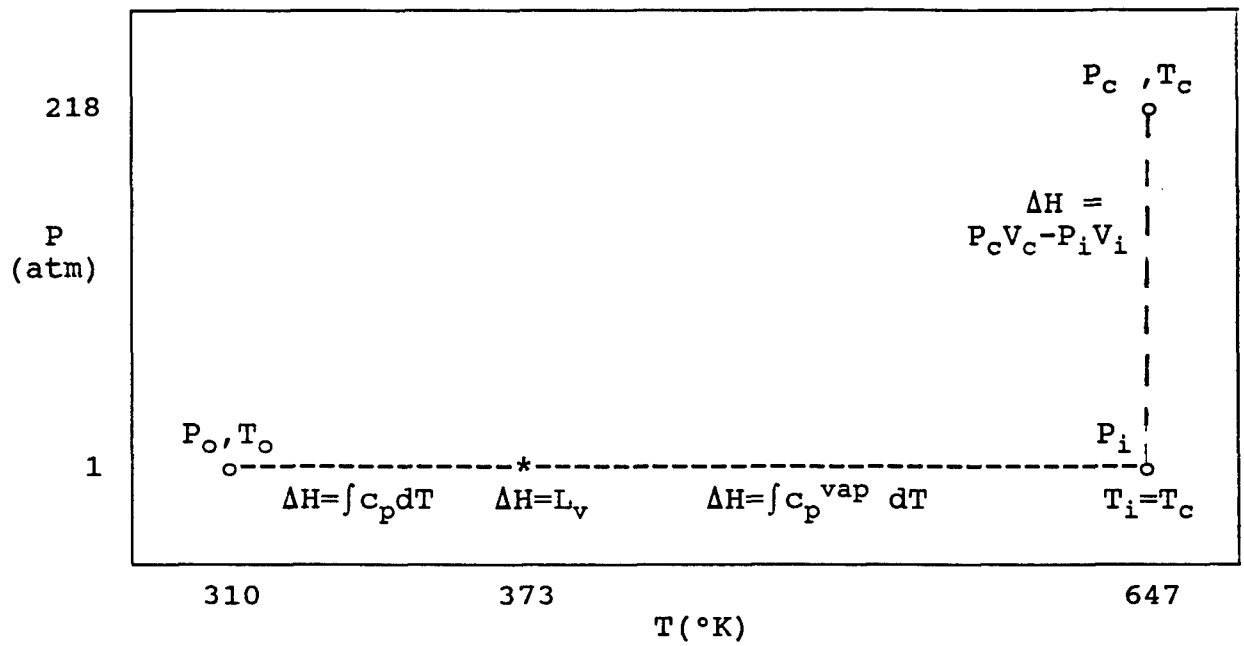


Fig. A1. Evaluation of enthalpy changes, ΔH, in determination of $q\equiv\Delta E=\Delta H-\Delta(PV)$. In addition to ΔH, there is the overall $\Delta(PV)=P_cV_c-P_oV_o$. An expression for c_p^{vap} is given in Eq. (A2).

Adding up all the contributions to ΔH we get $\Delta H=3066$ J/g - 230 J/g = 2836 J/g. The overall $\Delta(PV)=70$ J/g. Inserting these numbers into Eq. (A1) gives $q=\Delta E\approx 2770$ J/g for use in Eqs. (17), with an uncertainty due to setting $\Delta E=0$ for the isothermal compression as explained in the previous paragraph.

APPENDIX II: Energy Absorption by a Spherical Absorber

For light of uniform fluence H_o(J/cm²) hitting a sphere of radius 'a' with uniform absorption coefficient α, the fraction of light absorbed is calculated as follows. The total energy hitting the sphere is $E_o=\pi a^2 H_o$. For a path length of d, the fraction of light that passes through is $e^{-\alpha d}$. The average of this fraction over a sphere gives the energy E_T that is transmitted

$$\frac{E_T}{E_o}=\frac{1}{\pi a^2}\int_o^a e^{-2\alpha\sqrt{a^2-r^2}}2\pi r dr \qquad (A.3)$$

where r for a light ray is the distance of closest approach to the center, and $2\sqrt{a^2-r^2}$ is the path length of the light ray

through the melanosome. Figure A2 is a diagram of the process. The integration can be performed with a change of variable of $u=2\alpha\sqrt{a^2-r^2}$ and gives

$$\frac{E_T}{E_o}=\frac{1}{2\alpha^2a^2}[1-e^{-2\alpha a}(1+2\alpha a)] \qquad (A.4)$$

The energy absorbed is

$$E = E_o-E_T = \pi a^2H_o[1 - \frac{1}{2\alpha^2a^2}(1-e^{-2\alpha a}(1+2\alpha a))] = C(\alpha,a)\pi a^2H_o \qquad (A.5)$$

For a sphere that has the properties of a melanosome with a=1μm and a visible light α=1000 cm^{-1}, the fraction of light absorbed C(α,a), is 0.124 and $E=0.124E_o=0.124(\pi a^2H_o)$. For α=1800 cm^{-1}, C(α,a)=0.210.

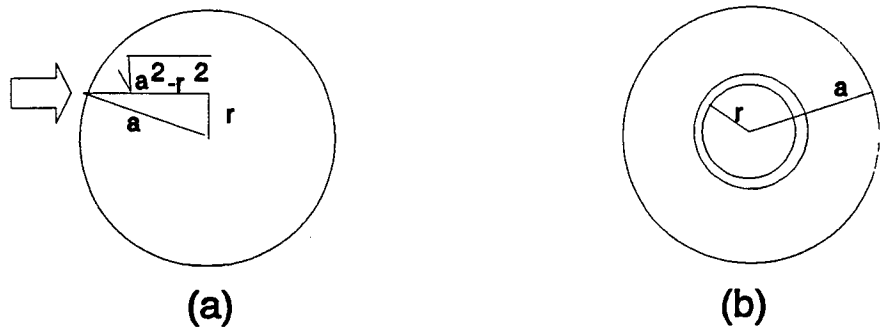


Figure A.1: Absorption by spherical melansome. a)Side view showing distance r of light from center of melanosome. b)Side view of same.

References

1. Henriques FC. Studies of thermal injury. Archives of Pathology 1947; 43:489-502.

2. Takata AN. Development of criterion for skin burns. Aerospace Medicine 1974; 45:634-637.

3. Birngruber R, Hillenkamp F, Gabel VP. Theoretical investigations of laser thermal retinal injury. Health Physics 1985; 48:781-796.

4. Flotte TJ, Frisoli JK, Goetschkes M, Doukas AG. Laser induced shock wave effects on red blood cells. In: "Laser-Tissue Interaction II," SPIE 1991; 1427:36.

5. Sliney DH. Interaction mechanisms of laser radiation with

ocular tissues. In: Court L, Duchene A, Courant D, eds. "First International Symposium on Laser Biological Effects and Exposure Limits." Paris; 1986, 64-83.

6. Ham WT, Mueller HA, Goldman AI, Newnam BE, Holland LM, Kuwabara T. Ocular hazard from picosecond pulses of Nd:YAG laser radiation. Science 1974; 185:362-363.

7. Wolbarsht ML, Walsh AL, George G. Melanin, a unique absorber. Applied Optics 1981; 20:2184-2186.

8. Cubeddu R, Docchio F, Ramponi R, Boulton M. Time-resolved fluorescence spectroscopy of the retinal pigment epithelium: age related studies. IEEE Journal of Quantum Electronics 1990; 26:2218-2225.

9. Gueneau G, Baille V, Courant D, Dubos M, Court L. Histopathology and ultrastructure of retinal injuries produced in rabbits by low energy laser irradiations. In: Court L, Duchene A, Courant D, eds. "First International Symposium on Laser Biological Effects and Exposure Limits." Paris; 1986, 176-191.

10. Hayes JR, Wolbarsht ML. Thermal model for retinal damage induced by pulsed lasers. Aerospace Medicine 1968; 39:474-480.

11. Cleary SF. Laser pulses and the generation of acoustic transients in biological material. In: Wolbarsht ML, ed. "Laser Applications in Medicine and Biology." New York: Plenum Press 1977, pp 175-219.

12. Ražnjević K, ed. "Handbook of Thermodynamic Tables and Charts." New York: McGraw Hill Book Company, 1976; 61.

13. Lamb H. "Hydrodynamics," New York: Dover, 1945;123.

14. L'Esperance F. "Ophthalmic Lasers," St. Louis: C. V. Mosby Company, 1989.

15. Vogel A, Birngruber R. Temperature profiles in human retina and choroid during laser coagulation with different wavelengths ranging from 514 to 810 nm. Lasers and Light in Ophthalmology 1992; 5(1): 9-16.

16. Thompson CR, Gerstman BS, Jacques SL, Rogers ME. Melanin granule model for laser-induced thermal damage in the retina. In preparation.

17. Weast RC, ed. "Chemical Rubber Company Handbook of Chemistry and Physics," Boca Raton, FL; CRC Press, 70th Edition, 1989-1990, pp. D42, D47.

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.

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.

USING ELECTRONIC BRAINSTORMING TOOLS TO VISUALLY REPRESENT THE IDEAS OF OTHERS: A PROPOSAL FOR RESEARCH

Kenneth A. Graetz

and

Scott MacBeth

Interaction, Creativity, and Electronic Brainstorming Systems

Commonsense dictates that interactive groups, where individual group members communicate with one another while working cooperatively toward a common goal, are often more productive than the same number of individuals working in isolation (i.e., *nominal groups*). Obviously, this is true for a wide variety of tasks such as competitive contests, large scale conflict resolution, and the execution of certain performance or psychomotor activities (McGrath, 1984). The presumed benefits of social interaction also provide the rationale for assigning creative tasks (e.g., idea or plan-generation tasks) to interacting groups. Osborn (1957), who popularized the *brainstorming* technique, claimed that free-wheeling, nonevaluative communication during idea-generation sessions could double the number of novel, creative ideas produced by any member of the group. This prediction derives from the popular belief that exposure to the creativity of others stimulates individual creativity. Osborne (1957) viewed the creative process as one positively influenced by social interaction.

As research progressed, it became apparent that brainstorming proponents had seriously underestimated the detrimental effects of group interaction. A large number of studies comparing nominal with interactive brainstorming groups highlighted the costs of face-to-face group discussion (Diehl & Strobe, 1987, 1991; Mullen, Johnson, & Salas, 1991). Of the 22 brainstorming studies reviewed by Diehl and Strobe (1987), 18 revealed nominal brainstorming groups, in which group members worked in isolation, to be significantly more creative than

interactive groups. Four studies obtained no significant difference in productivity. The strong evidence against group brainstorming prompted McGrath (1984) to state:

> "...the evidence speaks loud and clear: *Individuals working separately generate many more*, and *more creative ideas* (as rated by judges) *than do groups*, even when the redundancies among member ideas are deleted, and, of course, without the stimulation of hearing and piggybacking on the ideas of others. The difference is large, robust, and general." (p. 131)

In hindsight, it is not difficult to understand why being expected to contribute exciting, new ideas during face-to-face discussions with other, like-minded individuals might curb individual creativity. Most explanations focus on the dynamics of the brainstorming group itself. First, group discussions require a certain level of communicative coordination (e.g., only one person may talk at a time) that may block individuals from expressing their ideas at the moment of inspiration. Second, the presence of others may arouse evaluation apprehension. Even when instructed to be uninhibited and uncritical, it is the rare individual who can publicly disgorge every idea that comes to mind, no matter how fantastic or tangential. Finally, the presence of coworkers distributes the responsibility for accomplishing the task across the entire group. This may have serious ramifications for individual effort, with some group members slacking or loafing in the hopes that others will move the group toward its goal. Thus, *production blocking*, *evaluation apprehension*, and *social loafing* have all been offered as possible explanations for the apparent lack of creativity of interacting groups. Of the three potential causes, current research indicates that production blocking accounts for most of this variability in productivity (Diehl & Strobe, 1991).

One recent development that promises to unlock the creative potential of interacting groups is the emergence of computerized group decision support systems (GDSS). While each existing GDSS is composed of a wide variety of unique decision-making tools, most include an electronic brainstorming system (EBS). GroupSystems® (Dennis, George, Jessup, Nunamaker, & Vogel, 1987), for example, contains the text-based EBS represented in Figure 1. Each member of the group can enter ideas, send ideas to a common location (a group database and/or a public screen), and view the ideas of other group members. Sage®, a Macintosh-based GDSS (Wagner, Wynne,

& Mennecke, 1993), includes an EBS with a graphical interface designed to represent a deck of index cards. The individual types each idea on a separate card along with a subject heading. Users can then group and search the ideas by subject heading. Again, individuals can access a common database of ideas generated by other members of the group. Finally, CM/1® (Yakemovic & Conklin, 1990) provides users with a set of electronic drawing tools and a large, blank drawing area. Brainstormers can choose from among a wide variety of icons in displaying the categorical structure of the idea-generation task. The final product, an example of which is illustrated in Figure 2, is similar to a graphical flow chart in which each icon contains a number of ideas. CM/1® can be used by a group in either a face-to-face setting, with the diagram displayed on a public screen, or in a distributed meeting environment, with each group member independently accessing and editing the group's diagram.

These EBS products, while still in early stages of development, differ along various dimensions. These differences and similarities are summarized in Table 1. First, while all existing software allows users access to the specific ideas of others, products vary 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. GroupSystems®, for example, displays a random set of ideas on the user's screen whenever the user enters an idea into the common database. Thus the user has very little control over the level of specificity and the domain of information displayed. With CM/1®, individual users must consider the group's categorizational structure, as represented by the flow chart, but they are not exposed to the specific ideas of others unless they actively select a particular icon. Thus, CM/1® provides some control over the level of specificity and complete control over the domain of information accessed. Sage®, on the other hand, provides users with complete control over both level of specificity and information domain. Sage® users are not automatically exposed to either the specific ideas or the categorical structures of other group members. If users choose to access information, they can select from a list of idea categories.

Products also differ with respect to the display of information on the screen. Both Sage® and GroupSystems® rely on the text-based display of specific ideas. Sage® highlights the categorical structure of ideas by displaying heading text above each idea. CM/1®, on the other hand, represents the general trend in software development toward a graphical, point-and-click interface. Here, categories are displayed as icons which can be moved and dropped anywhere on the screen. Also, the user can connect categories together with lines and arrows in order to represent subcategorical structures or tangential search paths. Different icons can be used to represent different types of categories or other information. For example, users might use a question mark icon to store a key question that is guiding their search along a particular path. By selecting appropriate icons, CM/1® users can visually code ideas along a variety of different dimensions, leaving traces of their cognitive processes as well as their specific ideas. This opportunity is not afforded by either GroupSystems® or Sage®.

Finally, these products vary with respect to whether group versus individual productivity is emphasized. Both GroupSystems® and CM/1® place the group product at the forefront. With GroupSystems®, for example, each idea is placed immediately into the common database and the individual user is left without a record of their own productivity. Sage®, on the other hand, is designed to help users build their own list of ideas. While ideas are stored in a common database, their accumulation in the individual's own database serves as an indicator of individual ideational performance. While these three products are not the only EBS available (for a recent review of individual brainstorming tools, see Schorr, 1994), they represent well the various approaches to electronic brainstorming assistance.

In most cases, EBS developers argue that electronically assisted brainstorming allows individuals to enter their ideas anonymously, thereby reducing evaluation apprehension (Connolly, Jessup, & Valacich, 1990). More importantly, EBS tools presumably allow for the simultaneous entry of ideas by all members of the group, thereby reducing the level of production blocking (Dennis & Valacich, 1993). While limited in scope to the GroupSystems® EBS, recent research comparing electronically assisted, interacting groups with nominal groups provides little evidence

that the EBS tools unleash a group member's creative power. Two recent investigations (Connolly, Jessup, & Valacich; 1990; Valacich, Dennis, & Nunamaker) of anonymity versus indentifiability in electronic brainstorming groups found only minimal effects on ideational performance. Gallupe, Bastianutti, and Cooper (1991) found that electronically assisted, nominal groups actually outperformed electronically assisted, interacting groups. The only evidence that EBS tools facilitate the creativity of interacting groups comes from studies of relatively large (e.g., 12 member) groups (Dennis & Valacich, 1993). This is problematic considering that the brainstorming technique was not designed for groups of this size. Osborne (1957) warns against brainstorming in groups with more than eight members.

In many ways, the findings from EBS evaluation research are much more surprising than the observation of poor performance in unassisted, interacting groups. How is it possible that consideration of others' ideas in an environment that eliminates the detrimental social effects of face-to-face interaction does not yield any appreciable benefit for individual creativity? Is it that, as McGrath (1984) suggests, "the set of facilitative forces posited for groups is really not very powerful" (p. 132)? While explanations for process losses in interacting brainstorming groups are many, arguments as to why exposure to the ideas of others does not lead to significant process gains are virtually nonexistent.

Blind Searches and Category Activation

In developing such an explanation, it may be beneficial to represent the isolated brainstormer as engaged in the *blind search* of a large problem space. Blind searches use only the structure of the space of alternatives in selecting the next alternative (Nilsson, 1971). The defining characteristic of a blind search is its intention to explore the entire space. While various techniques might bias the search in one direction or another, no domain is given preferential treatment; the ultimate goal is to fully exhaust the problem space (Nilsson, 1971). Examples of blind search techniques include *depth-first* and *breadth-first* searches. An individual brainstormer might attempt to delineate all of the possible categories of ideas within the problem space before exploring any

one category in detail (i.e., breadth-first). Alternatively, a brainstormer might attempt to develop an exhaustive list of all the ideas in a particular category before turning to another category (i.e., depth-first). In either case, the only goal is the complete exploration of the ideational landscape.

Exposure to the ideas of others automatically eliminates blind search as an option available to the brainstormer. Instead, the informed individual must now employ a *heuristic search*, in which an algorithm is used to limit the search process to only the most promising or viable alternatives (Stillings, Feinstein, Garfield, Rissland, Rosenbaum, Weisler, & Baker-Ward, 1987). At the very least, all informed brainstormers now adhere to the *redundancy rule*: redundant ideas have no creative value. This is a trivial, universal rule governing face-to-face brainstorming groups; all ideas are viable except those that have already been generated by other brainstormers. While tagging redundant ideas complicates the search process to some extent, there is no reason to expect that the redundancy rule reduces the degree to which individuals are stimulated by the ideas of others. One possible exception involves exposure to the results of depth-first searches in which other brainstormers have generated a large number of ideas within a certain ideational category. In this case, an individual may tag an entire category as redundant. If numerous categories are excluded in this way, individual idea production may suffer. Given a large problem space however, extensive redundancy tagging could be expected to occur only when the problem space is approaching exhaustion. The use of the redundancy rule early in the interactive brainstorming session should not lead to the wholesale exclusion of entire idea categories. On the contrary, specific ideas should prime ideational categories, leading individual brainstormers to generate additional ideas within those categories. While the term has never been defined by brainstorming researchers, this is presumably what is meant by *creative stimulation* in interacting groups. Here, the term *category activation* will be used.

The Trailblazing Heuristic

Following this logic, exposure to the specific ideas of others during interactive brainstorming should activate ideational categories. Given that these categories have not been previously activated

by the individual brainstormers themselves, idea production should be enhanced. It is conceivable, however, that the interactive brainstorming task primes another heuristic that seriously impedes an individual's search process: the *trailblazing heuristic*.

The trailblazing heuristic is defined here as the belief that the value of an idea is inversely proportional to its similarity to the ideas of others. The trailblazing heuristic leads brainstormers to redefine their implicit definition of creativity. Ideas are no longer creative in and of themselves. Much of their creative value now derives from the extent to which they differ from the ideas of others. The trailblazing heuristic also affects the ideational search. Brainstormers now allocate more attention to formerly unexplored domains within the problem space. At the very least, trailblazing requires that certain ideas be designated or tagged as "discovered" and that trailblazers divert their search away from these ideas. In some cases, as with the redundancy rule, an entire area of the ideational landscape is cordoned off. This subdivision tends to occur much earlier in the brainstorming process, however, and has predictable, detrimental effects on creative performance. When the individual is exposed to a wide variety of ideas encompassing a large number of categories, unexplored territory becomes more difficult to locate; similarity judgments become more complex and more time consuming. Ultimately, the unfortunate consequence of trailblazing during interactive brainstorming tasks is that, by purposely diverting their attention away from specific ideas and entire ideational categories, trailblazers miss the opportunity to contribute ideas that may be far superior to those already residing within that domain. Thus, trailblazing limits creative production, reducing the probability that individuals will generate additional ideas within a previously discovered category.

It should be pointed out that a trailblazer's implicit definition of creativity is not necessarily based upon a competitive social motive. The trailblazing heuristic assigns a higher value to ideas that are different from, not necessarily better than, the ideas of others. Trailblazers simply view unexplored areas of the problem space as potentially more fruitful than well traveled portions. This is not to say that a competitive environment will not increase the prevalence of trailblazing. It would be reasonable to assume that the trailblazing heuristic is more common in competitive versus

cooperative environments. When individual brainstormers are competing to generate the most creative ideas, this only serves to increase the attractiveness of previously unexplored territory. No research has ever investigated the effect of a competitive motive on ideational performance in interacting groups, yet it is a natural extension of the evaluation apprehension notion. Presumably, individual brainstormers are apprehensive because they desire a positive evaluation from the group. This requires generating ideas that are at least as creative, and possibly more creative, than the ideas of other members. It could be argued that the mere act of brainstorming in a face-to-face group promotes a *creative tournament* of sorts. This competitive environment may increase the likelihood that any individual brainstormer will adopt a trailblazing algorithm when searching the ideational problem space.

Research Proposal

These ideas will be tested in a series of studies involving manipulations of information access and information display.

Experiment 1. The first study will employ a 2x4 experimental design with display of information and motivational set as between-subjects independent variables. Inspiration®, an EBS similar to CM/1®, will be used to manipulate information display. Participants will arrive in groups of three and will be introduced to the study by either a male or female experimenter. During the introduction, participants will receive training on Inspiration® using a practice brainstorming task (e.g., generate creative uses for a brick). Following the introduction and training phase, participants will be escorted to private rooms where they will engage in the experimental brainstorming task (e.g., creative ideas for new television shows) using Inspiration®. Traditional brainstorming instructions will be issued to each participant. After 10 minutes of idea generation, most individuals will be presented with the Inspiration® diagram ostensibly generated by another group member. Participants will be told that the diagram was randomly selected from one of the other two brainstormers. In fact, these stimulus materials will have been generated prior to the experiment in order to introduce three levels of the information display variable. Some of the

participants will view a diagram depicting the main category icon (i.e., "TV Shows") and six specific ideas from three separate categories. This is the *ideas only condition*. Some of the participants will view a diagram depicting the main category icon and three subcategory icons. This is designated as the *categories only condition*. Finally, some participants will view a diagram consisting of the main category icon, the subcategory icons, and the specific ideas (i.e., the *ideas plus categories condition*). Examples of these diagrams are included in Figure 3. The fourth level of the display variable will be a control condition in which participants will not be given access to the ideas of coworkers but will view a control diagram from a different brainstorming task (e.g., solutions to the parking problem on campus). Participants will then be given 20 additional minutes to generate ideas. During this time, individuals will have constant access to the stimulus diagram.

The participant's motivational orientation, cooperative versus competitive, will also be manipulated. In order to improve overall motivational level, all participants will be told that the study is being funded by a national broadcasting company in an attempt to improve their programming and to attract college-aged viewers. Participants will be told that the company has provided $100 cash prizes to each participating university. Brainstormers in the cooperative condition will be told that the most creative groups will be registered for a $300 (i.e., $100 for each group member) random drawing. Participants in the competitive condition will be told that the name of the most creative individual from each session will be registered for a $100 drawing.

Dependent variables will include measures of overall ideational performance and trailblazing. The ideas generated by participants following exposure to the stimulus diagram will be coded by trained judges for quantity and creative quality. Past research has shown this to be an effective and reliable measure of ideational performance (Diehl & Strobe, 1987). In addition, judges will group ideas into categories. The degree to which individual brainstormers add ideas to categories already represented on the stimulus diagrams will be used as an indicator of the trailblazing heuristic. In Experiment 1, the following hypotheses are advanced:

Hypothesis #1: A main effect for information display is predicted regarding both overall ideational performance and trailblazing. Participants in the categories only condition are

expected to generate significantly more creative ideas and to show less evidence of a trailblazing heuristic than participants in other display conditions.

Hypothesis #2: A similar main effect is expected for motivational set. Participants in the competitive condition are expected to generate significantly fewer creative ideas and to show more evidence of a trailblazing heuristic than participants in other display conditions.

Hypothesis #3: An interaction between motivational set and information display is predicted with respect to overall ideational performance. It is hypothesized that the creativity of competitive brainstormers will be reduced through exposure to the specific ideas of others. Thus, competitive brainstormers are expected to generate more creative ideas in the no information condition than in any other display condition. Cooperative brainstomers, on the other hand, will benefit from category activation, showing higher ideational performance when exposed to the ideas of others.

Experiment 2. The second study will utilize a 2x3 experimental design to investigate the effects of the iconic representation of categories on brainstorming performance. The procedure and brainstorming task (e.g., creative ideas for new television shows) will be identical to those employed in Experiment 1. Again, a cooperative versus competitive orientation will be induced. Individual brainstormers will be exposed to the ideas of others in the form of an Inspiration® diagram. In this study, however, only the idea categories will be displayed, in the form of icons, on the screen. The breadth of the iconic representation will be manipulated as wide versus narrow. As in Experiment 1, no exposure control groups will be included for comparison purposes. In one control group, brainstormers will be exposed to a wide breadth stimulus diagram from another brainstorming task (e.g., solutions to the parking problem on campus); in the other control group, the stimulus diagram will reflect a narrow representation. The major hypothesis is that, due to the increased likelihood of trailblazing, the creativity of competitive individuals should be reduced as the breadth of the stimulus diagram increases. Cooperative individuals, on the other hand, should benefit most from a wider breadth of category icons.

Experiment 3. In the third proposed study, brainstormers will have genuine access to the ideas of others via a group editing task. Groups of two members each will use Inspiration® to privately generate a graphical representation of the television brainstorming problem. In this experiment, however, a group member will be randomly selected to begin the diagram. This individual will have 10 minutes to brainstorm after which the diagram will be passed to the other group member. This member will then have 10 minutes of brainstorming time. Each brainstormer will participate in two, 10 minute rounds. The input of each participant will be identified based upon the color of the icons used. Again, a cooperative versus competitive orientation will be induced and a no information control group will be used. The major hypothesis in Experiment 3 is that competitive individuals will trailblaze by avoiding the categories previously generated by the opponent. Cooperative individuals will be more likely to extend, depth-wise, the categorical structure provided by the partner.

The Practical Relevance of the Proposal

While this research is designed to answer a number of basic theoretical questions concerning creativity in groups, it also addresses several important EBS design issues. As user interfaces make the transition from textual to graphical, electronically assisted brainstormers are afforded the opportunity to portray visually not only a list of ideas but their entire cognitive map of the problem space. The effects of this type of information display on an individual's cognitive processing are as yet unspecified. More importantly for EBS developers, the cognitive effects of the exposure to another individual's visual representation are also unknown. It is presumed, for example, that access to such information allows individuals to quickly distinguish explored from untouched domains within that space. This may or may not facilitate the individual's ideational performance. If, as in a competitive environment, an individual is using a trailblazing heuristic, such knowledge may complicate and confuse the search process. In a cooperative climate, such information may be quite helpful.

The answers to these questions are relevant to developers who are trying to determine the extent of automatic versus controlled information display and the level of automaticity and specificity of displayed information that should be programmed into new computer tools. Should specific ideas be automatically displayed on the screen? Should subcategories be displayed and, if so, what level of depth should be employed? Is there a benefit to visually identifying individual user input on a communal diagram? Are there other dimensions, such as the level of specific user input, that might benefit brainstormers if represented visually? Hopefully, the proposed research will begin to shed some light on these important issues.

References

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. Management Science, 36, 698-703.

Dennis, A. R., George, J. F., Jessup, L. M., Nunamaker, J. F., Jr., & Vogel, D. R. (1988). Information technology to support electronic meetings, MIS Quarterly, 12, 591-616.

Dennis, A. R., & Valacich, J. S. (1993). Computer brainstorms: More heads are better than one. Journal of Applied Psychology, 78, 531-537.

Diehl, M., & Strobe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. Journal of Personality and Social Psychology. 61, 392-403.

Diehl, M., & Strobe, W. (1987). Productivity loss in idea-generating groups: Toward the solution of a riddle. Journal of Personality and Social Psychology, 53, 497-509.

Gallupe, R.B., Bastianutti, L. M., and Cooper, W.H. (1991). Unblocking brainstorms. Journal of Applied Psychology, 76, 137-142.

McGrath, J. E. (1984). Groups: Interaction and performance. Englewood Cliffs, NJ.:Prentice-Hall

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. Basic and Paaplied Social Psychology, 12, 3-23.

Nilsson, N. J. (1971). Problem-solving methods in artificial intelligence. New York: McGraw-Hill.

Osborne, A. F. (1957). Applied imagination (Rev. ed.) New York: Scribner.

Schorr, J. (1994). Smartthink: Eight programs that help you think creatively and plan. Macworld, May, 138-142.

Stillings, N.A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. E., & Baker-Ward, L. (1987) Cognitive science: An introduction. Cambridge, MA: MIT Press.

Valacich, J. S., Dennis, A. R., & Nunamaker, J. F., Jr., (1991). Electronic meeting support: The GroupSystems concept. International Journal of Man-Machine Studies, 34, 261-282.

Yakemovic, K. C. B., & Conklin, E. J. (1990). Report on a development project use of an issue-based information system. Proceedings: Computer Supported Cooperative Work, October, 1990.

Table 1

A Comparison of Electronic Brainstorming Products Regarding Access to Group Information, Visual Display of Information, and Focus of Productivity

	Electronic Brainstorming System		
Criteria	GroupSystems®	Sage®	CM/1®
Access to group information			
Specific ideas			
Automatic	√	—	—
Controlled	—	√	√
Categories			
Automatic	—	—	√
Controlled	—	√	—
Visual display of information			
Text	√	√	—
Graphics	—	—	√
Focus of productivity			
Group	√	—	√
Individual	—	√	—

Figure 1. Using GroupSystems' (1987) EBS to generate ideas for a television show.

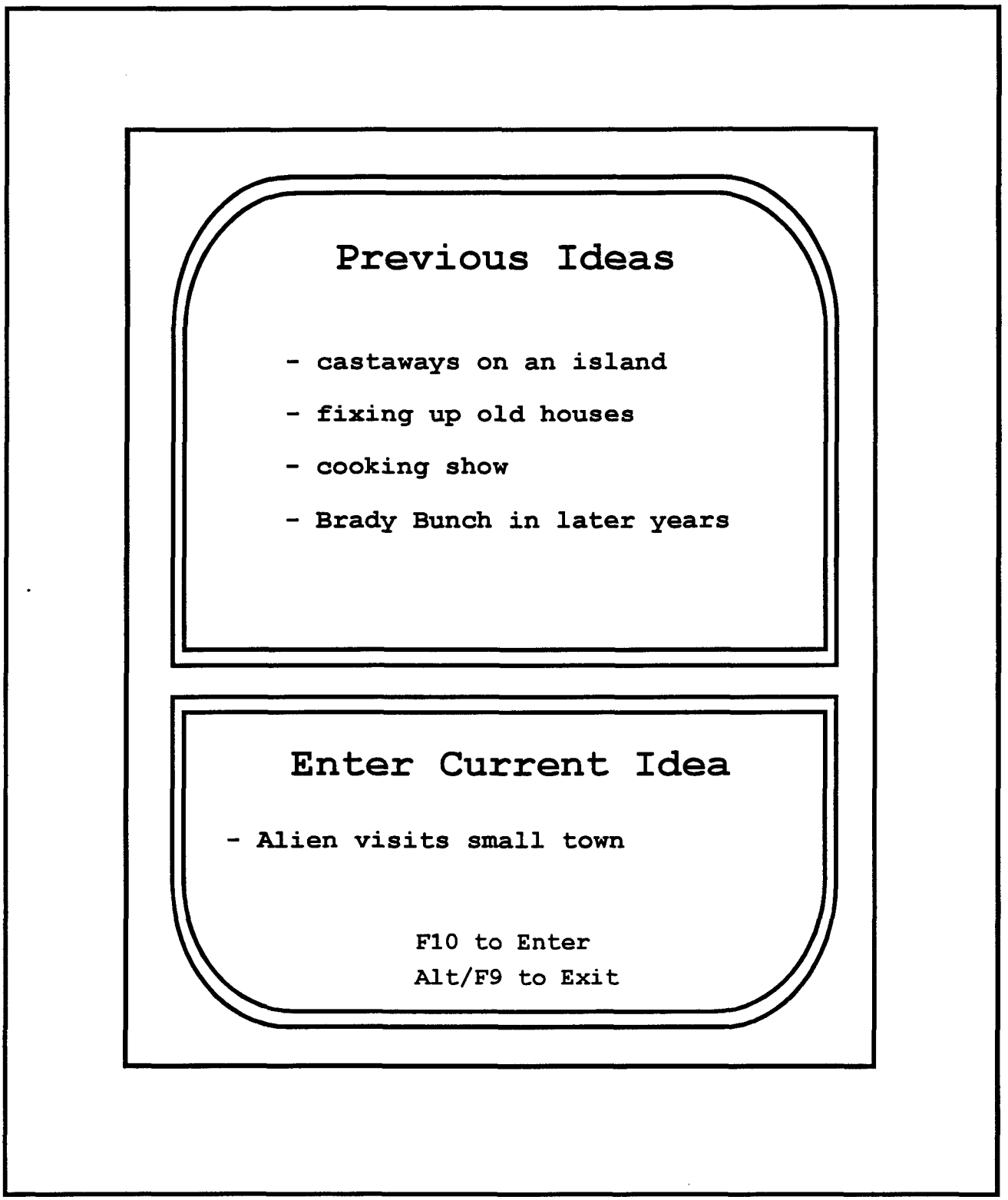


Figure 2. A sample CM/1® flow chart.

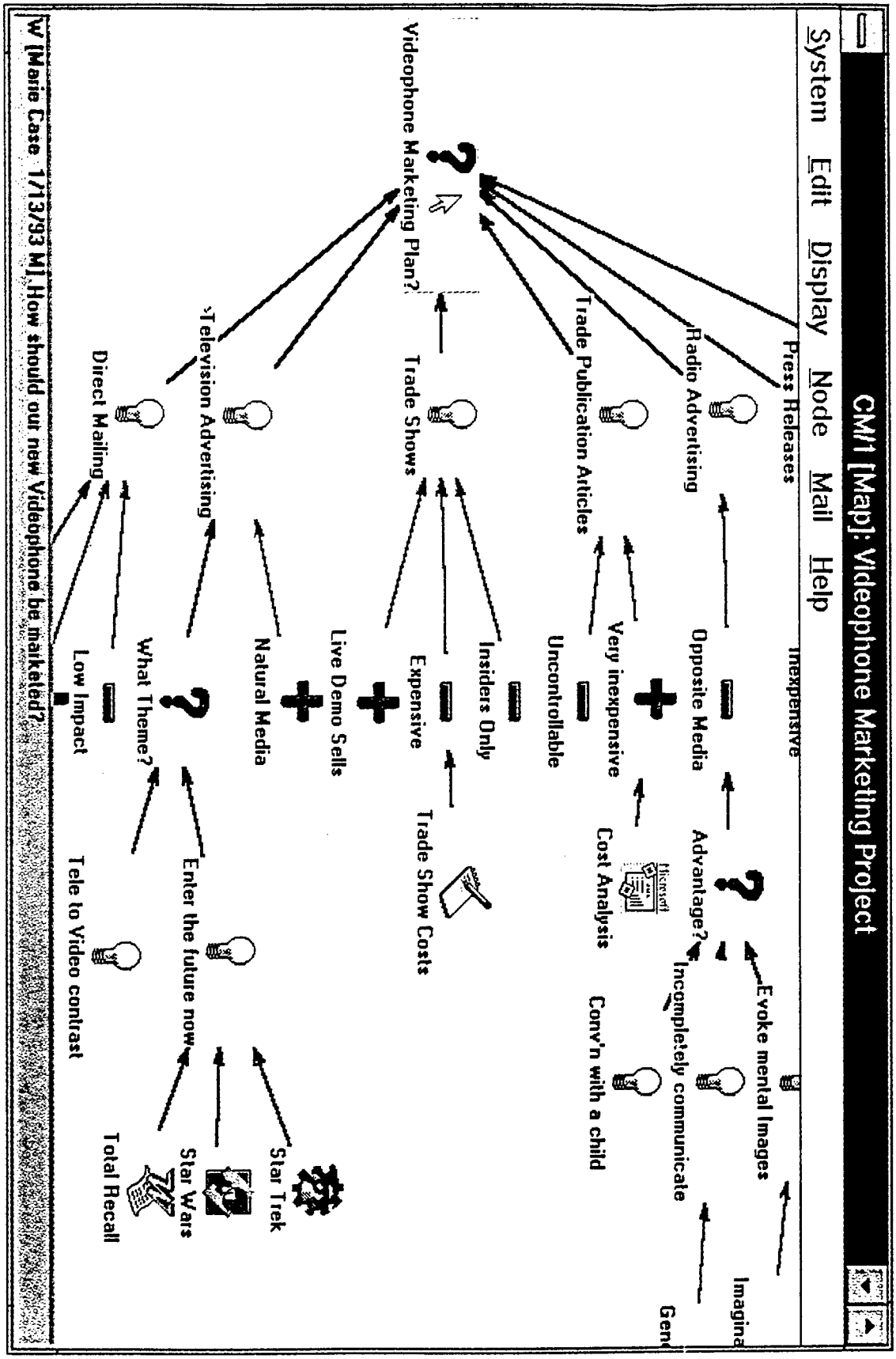
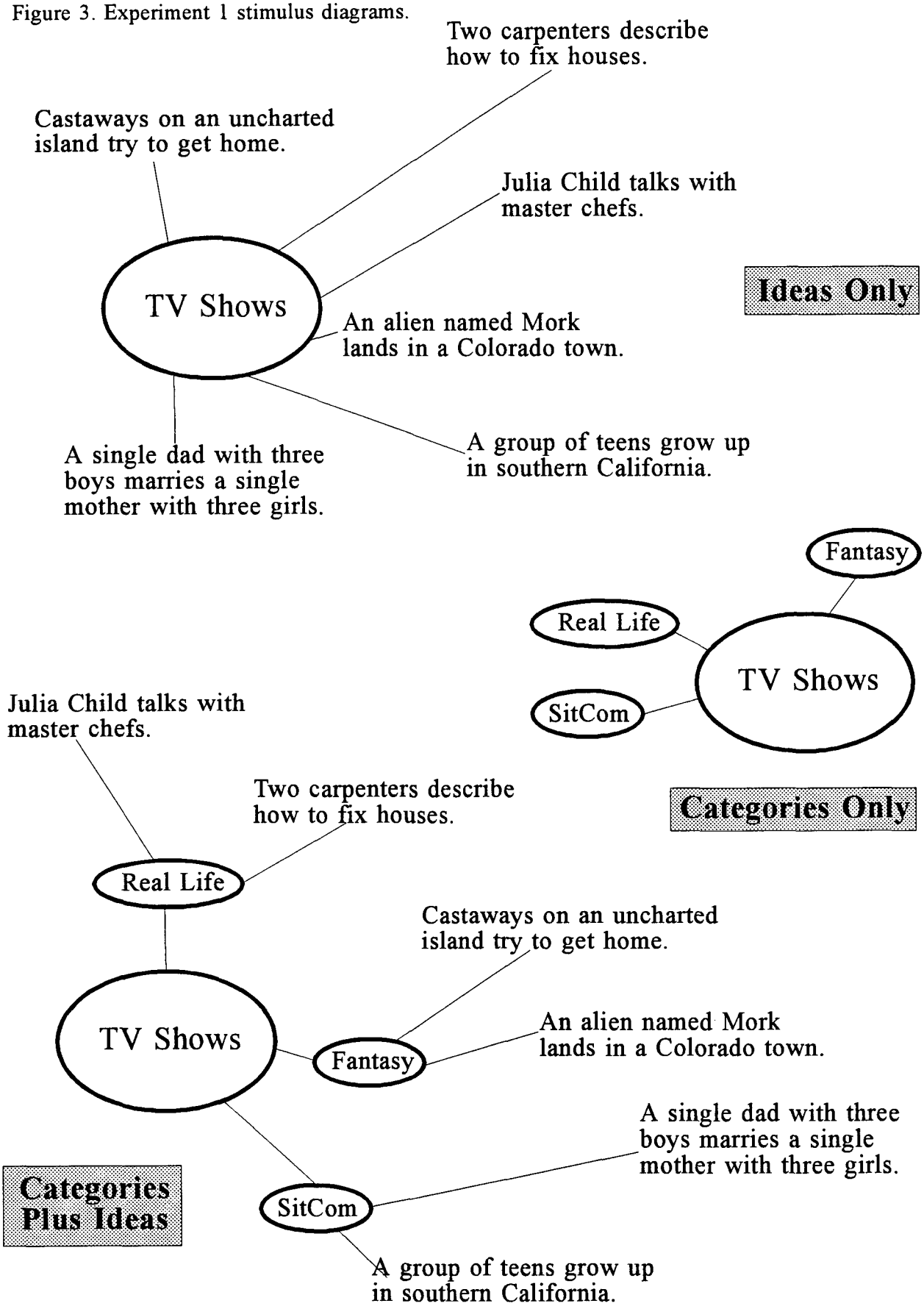


Figure 3. Experiment 1 stimulus diagrams.



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.

PIEZOMETER LOCATIONS
Made-2 Coordinates

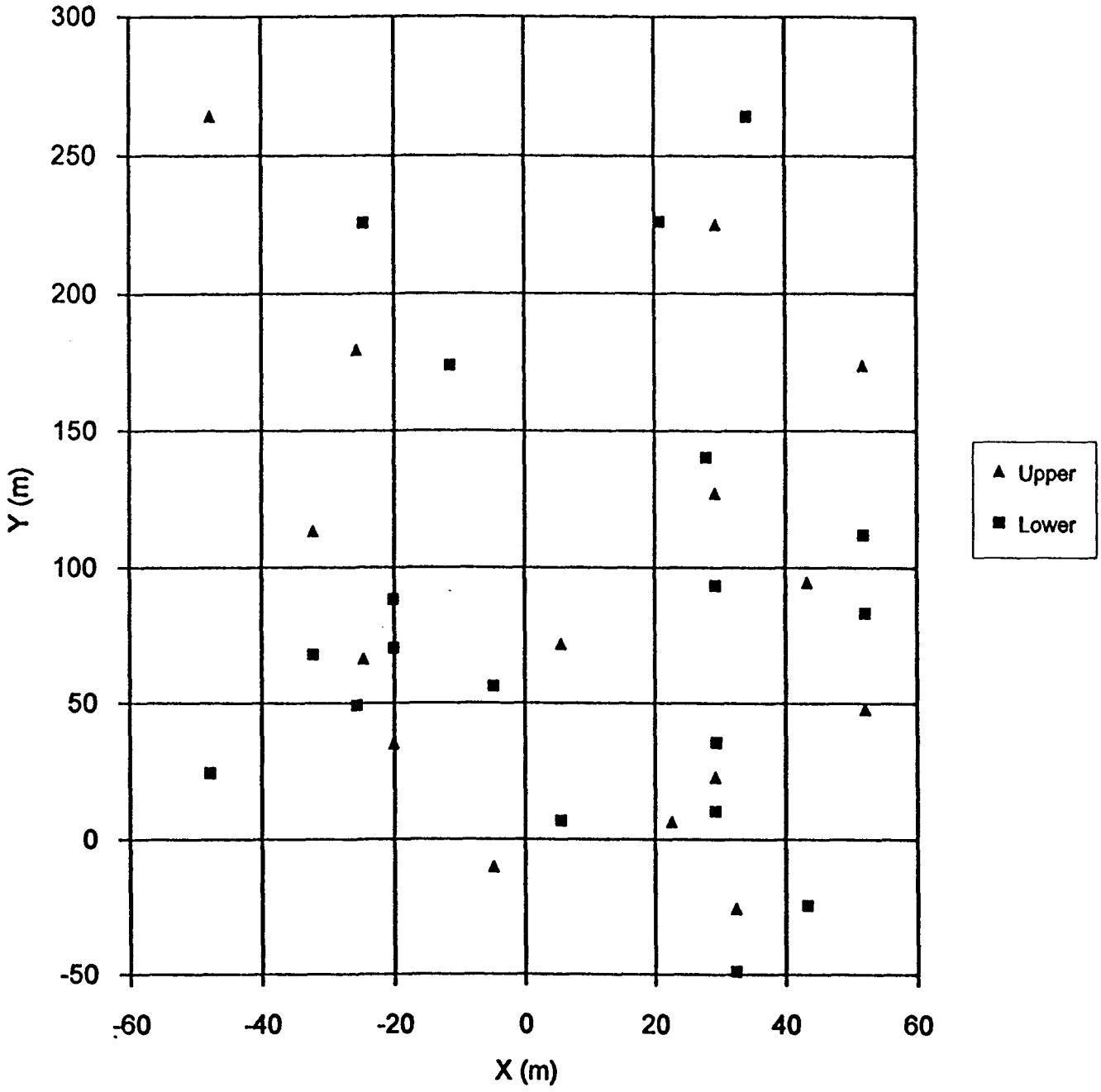


Figure 1. Locations of upper (squares) and lower (triangles) piezometers used to establish initial and boundary conditions.

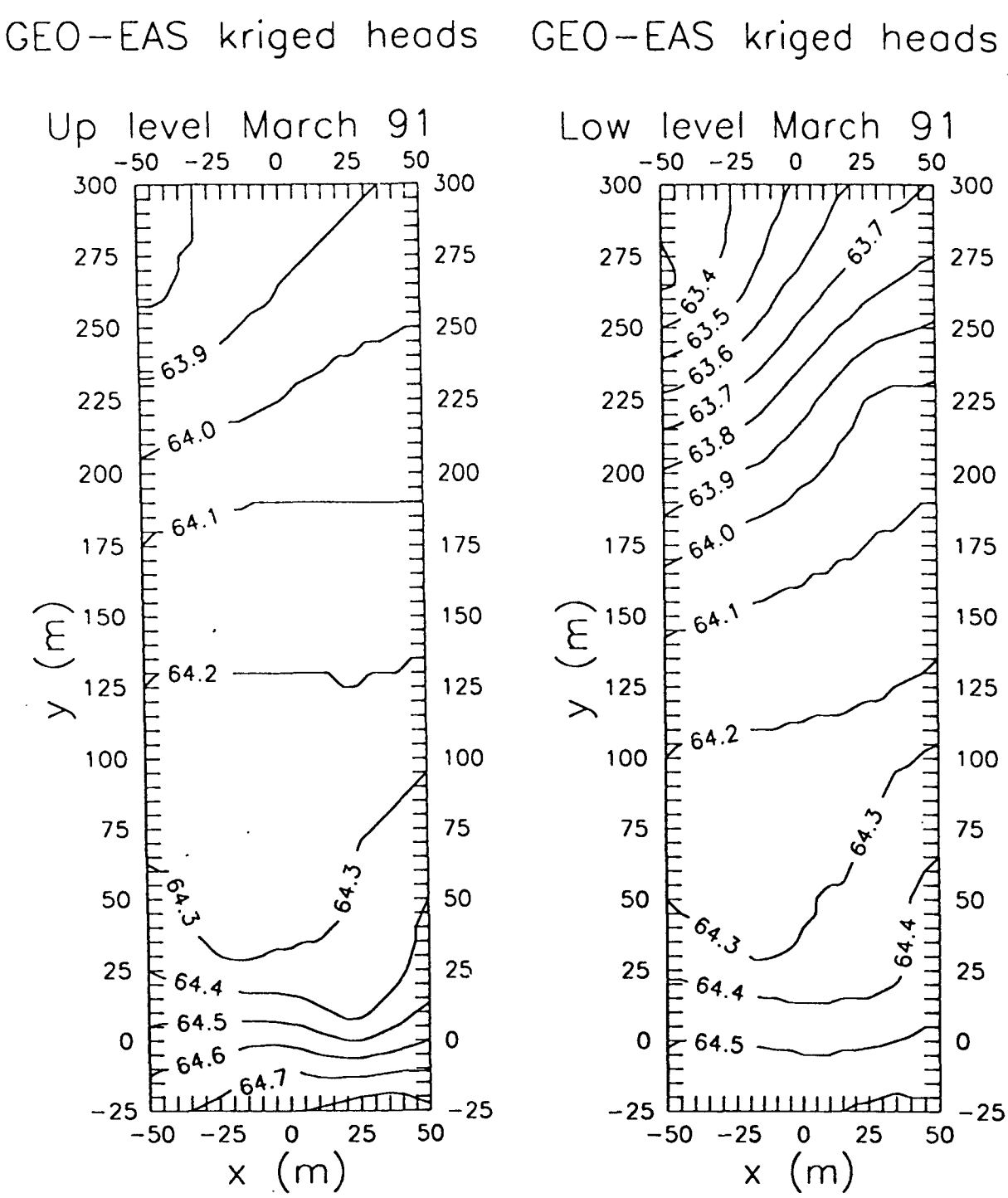


Figure 2. Upper (left) and lower (right) kriged head distributions for simulation day 270 (March 8, 1991). Heads are in meters.

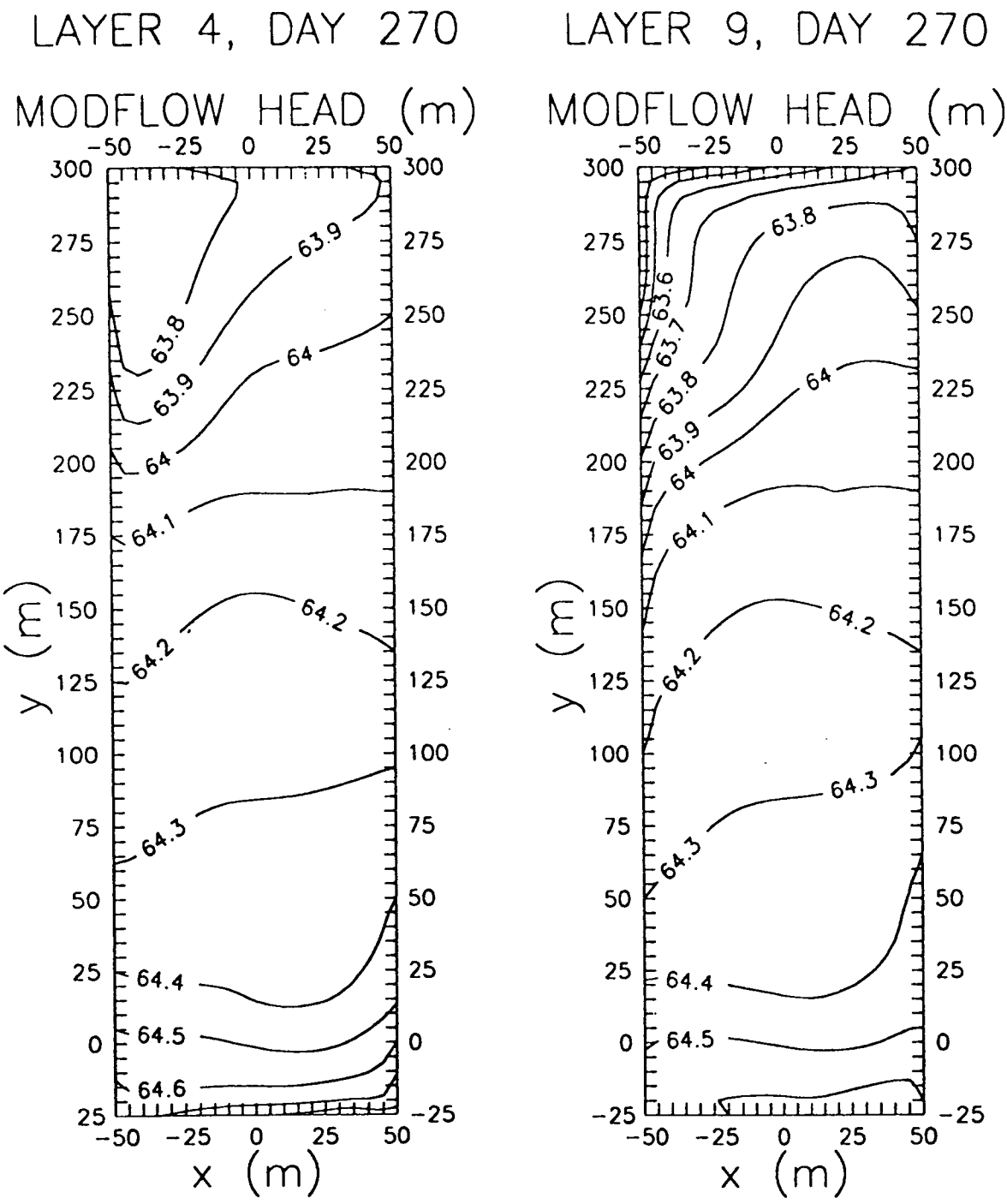


Figure 3. MODFLOW Case 1 simulated heads for layers 4 (left) and 9 (right) for simulation day 270 (March 8, 1991).

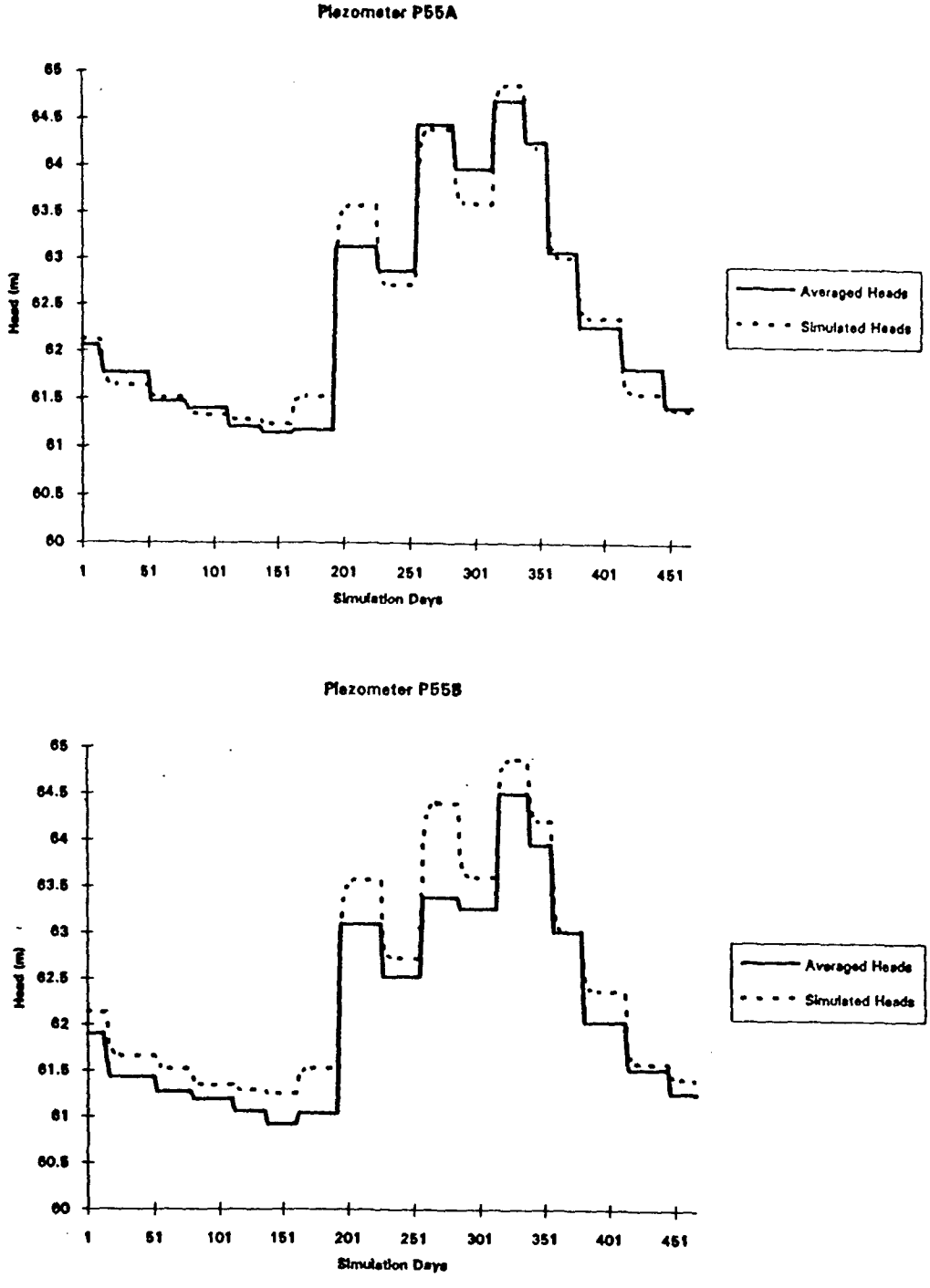


Figure 4. Comparison of MODFLOW Case 1 simulated heads with observed heads averaged over stress periods at piezometers P55a (top) and P55b (bottom). Heads are in meters.

REGRESSION TO THE MEAN IN HALF-LIFE STUDIES

Pushpa L. Gupta
Professor
Department of Mathematics & Statistics

University of Maine
321 Neville Hall
Orono, ME 04469

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

REGRESSION TO THE MEAN IN HALF-LIFE STUDIES

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

ABSTRACT

Half-life studies of biomarkers for environmental toxins in humans are generally restricted to a few measurements per subject taken at least one half-life after exposure. The initial dose is usually unknown because the exposure occurred before the substance was known to be toxic. In this setting, subjects are selected for inclusion in the study if their measured body burden is above a threshold (C), determined by the distribution of the biomarker in a control population. We assume a simple one-compartment first order decay model and a log-normal biomarker distribution, which together imply a repeated measures linear model relating the logarithm of the biomarker and time, with the slope being the negative of the decay rate (λ). Unless the data set is properly conditioned, we show that ordinary weighted least squares estimates of λ are biased due to regression toward the mean. In practice, the last measurement is taken to be greater than the threshold value C. Formulae are presented in the special case that 3 measurements per subject are available. Generalizations to k measurements per subject are straightforward.

REGRESSION TO THE MEAN IN HALF LIFE STUDIES

Pushpa L. Gupta

1. INTRODUCTION

Half-life studies of biomarkers for environmental toxins in humans are generally restricted to a few measurements per subject taken at least one half-life after exposure. The initial dose is usually unknown because the exposure occurred before the substance was known to be toxic. In this setting, subjects are selected for inclusion in the study if their measured body burden is above a threshold (C), determined by the distribution of the biomarker in a control population. We assume a simple one-compartment first order decay model and a log-normal biomarker distribution, which together imply a repeated measures linear model relating the logarithm of the biomarker and time, with the slope being the negative of the decay rate (λ). Unless the data set is properly conditioned, we show in section 2 that ordinary weighted least squares estimates (WLSE) of λ are biased due to regression toward the mean (James (1973); Senn & Brown (1985)). If the within-subject correlation matrix is banded, we show that the weighted least squares estimate of λ is unbiased when the data set is conditioned on all repeated measures being above a line with slope $-\lambda$. In particular if the within-subject correlation matrix is auto regressive of order 1 or has compound

symmetry, then the unbiased-ness of WLSE of λ is automatically satisfied. In section 3 the Air Force data on 240 Ranch Hands were analyzed for the bias of β_1 (WLSE of $-\lambda$). The results are displayed in Figures 1-5. Finally, in section 4 some conclusions and recommendations are presented for future work.

2. BIAS IN THE ESTIMATION OF DECAY RATE

We have assumed that a single exposure produced an elevation of the TCDD body burden above background level and that the first-order kinetics model

$$C_t = C_o e^{-\lambda t}. \qquad (2.1)$$

holds, where C_t is the TCDD concentration t years after exposure measured in parts per trillion, C_o is the (unknown) initial exposure, and λ is a constant but unknown decay rate. Based on (2.1), the true population half-life is $t_{1/2} = \frac{\ln 2}{\lambda}$. If we take the natural logarithm of (2.1) we obtain

$$\ln C_t = \ln C_o - \lambda t. \qquad (2.2)$$

Thus, (2.2) can be regarded as a motivating equation for a model which can accommodate multiple measurements per subject as well as covariates. Such a model is known as a fixed subject-effects model with repeated measures and is described below:

$$Y_{ij} = \mu + \tau_i + \beta_1 t_{ij} + \epsilon_{ij} \quad j = 1, 2, 3; \; i = 1, 2, \ldots, n, \qquad (2.3)$$

where Y_{ij} represents the natural logarithm of the jth background

corrected TCDD measurement on the ith subject t_{ij} years after exposure, $-\beta_1$ represents the common decay rate λ, β_0 represents the average intercept for all subjects, τ_i represents the fixed subject effect for the ith subject, and ϵ_{ij} is the residual error term for Y_{ij}. Let $Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix}$ be the observation vector for the ith individual such that

$$Y_i \sim N_3(X_i\beta, \Sigma), \quad i = 1, 2, \ldots, n.$$

Here X_i is a 3x(n+2) matrix of known design indicators given by

$$X_i = \left[\begin{array}{ccc|c} 1 & t_{i1} & 1 & \\ 1 & t_{i2} & 1 & O_{3x(n-1)} \\ 1 & t_{i3} & 1 & \end{array}\right] E_i,$$

E_i is an elementary matrix obtained from the (n+2) x (n+2) identity matrix by the interchange of 3rd and (i+2)th columns, $\beta = (\mu\ \beta_1\ \tau_1\ \tau_2\ \ldots\ \tau_n)'$ is a (n+2) x 1 vector of unknown regression coefficients, and $\Sigma = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 \\ \theta_1 & \theta_0 & \theta_3 \\ \theta_2 & \theta_3 & \theta_0 \end{bmatrix}$ is the unknown variance-covariance matrix with equal variances. We can write the combined model for all of the data in a matrix form by letting

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \qquad \text{and } V = diag(\Sigma, \Sigma, \ldots, \Sigma)_{(3n)x(3n)}.$$

Then the model for the entire observation vector is

$$Y \sim N(X\beta, V). \qquad (2.4)$$

When Σ is known, the generalized least-squares estimator β is found by minimizing the quadratic form $\sum_{i=1}^{n} Q_i(\beta,\Sigma)$, where

$$Q_i(\beta,\Sigma) = (Y_i - X_i\beta)^T \Sigma^{-1} (Y_i - X_i\beta).$$

The solution is $\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} Y$.

When Σ is unknown and therefore V is unknown, which is usually the case in practice, V is replaced by its estimator S, so then $\hat{\beta} = (X' S^{-1} X)^{-1} X' S^{-1} Y$,. Under the normal distribution assumption $\hat{\beta}$ is also the maximum likelihood estimator of β.

In this section we would first like to find the estimator of β and in turn the estimator of β_1 and then find the expected value of $\hat{\beta}_1$ under truncation, i.e. when $(Y_{i1}\ Y_{i2}\ Y_{13}) > (C_1\ C_2\ C_3)$, where C_1, C_2, C_3, are known constants. $\hat{\beta}$ for model (2.4) is given by

$$\hat{\beta} = (X' V^{-1}X)^{-1} X'V^{-1}Y,$$

where

$$X' V^{-1} X = \frac{1}{|\Sigma|} \left[\begin{array}{cc|ccc} n a_{11} & \sum_{i=1}^{n} a_{21}^{(i)} & a_{11} & a_{11} & \cdots\ a_{11} \\ \sum_{i=1}^{n} a_{21}^{(i)} & a_{22} & a_{21}^{(1)} & a_{21}^{(2)} & \cdots\ a_{21}^{(n)} \\ \hline a_{11} & a_{21}^{(1)} & & & \\ a_{11} & a_{21}^{(2)} & & a_{11} I_n & \\ \vdots & \vdots & & & \\ a_{11} & a_{21}^{(n)} & & & \end{array}\right] \tag{2.5}$$

$$X'V^{-1}Y = \frac{1}{|\Sigma|} \begin{bmatrix} \sum_{j=1}^{3} b_{1j} S_{yj} \\ \sum_{i=1}^{n} \sum_{j=1}^{3} b_{2j}^{(i)}\ Y_{ij} \\ b_{11}Y_{11} + b_{12}Y_{12} + b_{13}Y_{13} \\ \vdots \\ b_{11}Y_{i1} + b_{12}Y_{i2} + b_{13}Y_{i3} \\ \vdots \\ b_{11}\ Y_{n1} + b_{12}\ Y_{n2} + b_{13}Y_{n3} \end{bmatrix} \tag{2.6}$$

The following notations have been used above:

$a_{11} = b_{11} + b_{12} + b_{13}$

$a_{21}^{(i)} = t_{i1}\, b_{11} + t_{i2}\, b_{12} + t_{i3}\, b_{13}$, *where*

$b_{11} = (\theta_0 - \theta_3)\ (\theta_0 + \theta_3 - \theta_1 - \theta_2)$,

$b_{12} = (\theta_0 - \theta_2)\ (\theta_0 + \theta_2 - \theta_1 - \theta_3)$,

$b_{13} = (\theta_0 - \theta_1)\ (\theta_0 + \theta_1 - \theta_2 - \theta_3)$.

$a_{22} = \sum_{i=1}^{n} t_{i1}\, b_{21}^{(i)} + t_{i2}\, b_{22}^{(i)} + t_{i3}\, b_{23}^{(i)}$, *where*

$b_{21}^{(i)} = t_{i1}\ (\theta_0^2 - \theta_3^2) + t_{12}\ (\theta_2\, \theta_3 - \theta_0\, \theta_1) + (\theta_1\, \theta_3 - \theta_0\, \theta_2)$,

$$b_{22}^{(i)} = t_{i1}\,(\theta_2\,\theta_3 - \theta_0\,\theta_1) + t_{i2}\,(\theta_0^2 - \theta_2^2) + t_{i3}(\theta_1\,\theta_2 - \theta_0\,\theta_3),$$

$$b_{23}^{(i)} = t_{i1}\,(\theta_1\,\theta_3 - \theta_0\,\theta_2) + t_{i2}\,(\theta_1\,\theta_2 - \theta_0\,\theta_3) + t_{i3}\,(\theta_0^2 - \theta_1^2).$$

$$|\Sigma| = \theta_0\,(\theta_0^2 - \theta_1^2 - \theta_2^2 - \theta_3^2) + 2\theta_1\,\theta_2\,\theta_3, \text{ and}$$

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} \theta_0^2 - \theta_3^2 & \theta_2\,\theta_3 - \theta_0\,\theta_1 & \theta_1\,\theta_3 - \theta_0\,\theta_2 \\ \theta_2\,\theta_3 - \theta_0\,\theta_1 & \theta_0^2 - \theta_2^2 & \theta_1\,\theta_2 - \theta_0\,\theta_3 \\ \theta_1\,\theta_3 - \theta_0\,\theta_2 & \theta_1\,\theta_2 - \theta_0\,\theta_3 & \theta_0^2 - \theta_1^2 \end{bmatrix}.$$

$$s_{yj} = \sum_{i=1}^{n} Y_{ij}, \quad j = 1,2,3. \tag{2.7}$$

In order to find $\hat{\beta}_1 = [2nd \text{ row of } (X'V^{-1}X)^{-1}]X'V^{-1}Y$, first we need to find the 2nd row of $(X'V^{-1}X)^{-1}$. Second row of $(X'V^{-1}X)^{-1}$ can be obtained by considering a generalized inverse of $X'V^{-1}X$ since it is singular. For obtaining the generalized inverse we first drop the last column and last row of $X'V^{-1}X$ and write the remaining matrix as $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, where

$$A_{11} = \frac{1}{|\Sigma|}\begin{bmatrix} n\,a_{11} & \sum_{i=1}^{n} a_{21}^{(i)} \\ \sum_{i=1}^{n} a_{21}^{(i)} & a_{22} \end{bmatrix}, \qquad A_{12} = \frac{1}{|\Sigma|}\begin{bmatrix} a_{11} & a_{11} & \cdots & a_{11} \\ a_{21}^{(1)} & a_{21}^{(2)} & \cdots & a_{21}^{(n-1)} \end{bmatrix}$$

$$A_{22} = \frac{a_{11}}{|\Sigma|}\,I_{n-1}, \text{and } A_{21} = A_{12}'.$$

Consider the following:

$$\text{(i)} \quad (A_{11} - A_{12}\, A_{22}^{-1}\, A_{21})^{-1} = \frac{|\Sigma|}{D} \begin{bmatrix} a_{22} - \frac{1}{a_{11}} \sum_{i=1}^{n} (a_{21}^{(i)})^2 & -a_{21}^{(n)} \\ -a_{21}^{(n)} & a_{11} \end{bmatrix}$$

$$\text{(ii)} \quad -(A_{11} - A_{12}\, A_{22}^{-1}\, A_{21})^{-1} A_{12}\, A_{22}^{-1} =$$

$$\frac{-|\Sigma|}{a_{11} D} \begin{bmatrix} \text{<----------first row----------->} \\ a_{11}(a_{21}^{(1)} - a_{21}^{(n)}) \;\; a_{11}(a_{21}^{(2)} - a_{21}^{(n)}) \ldots a_{11}(a_{21}^{(n-1)} - a_{21}^{(n)}) \end{bmatrix}$$

where $D = a_{11}\, a_{22} - \sum_{i=1}^{n} (a_{21}^{(i)})^2$

The second row of $(X'\, V^{-1}\, X)^{-1}$ amounts to finding the second row of the matrix

$$[(A_{11} - A_{12}\;\; A_{22}^{-1}\, A_{21})^{-1} \quad -(A_{11} - A_{12}\, A_{22}^{-1}\, A_{21})^{-1} A_{12}\, A_{22}^{-1} \;\; 0]$$

$$= \frac{|\Sigma|}{D} \begin{bmatrix} \text{<----------first row----------->} \\ -a_{21}^{(n)} \;\; a_{11} \;\; -(a_{21}^{(1)} - a_{21}^{(n)}) \;\; -(a_{21}^{(2)} - a_{21}^{(n)}) \ldots \;\; -(a_{21}^{(n-1)} - a_{21}^{(n)}) \, 0 \end{bmatrix}$$

Thus $\hat{\beta}_1 = [2nd \text{ row of } (X'\, V^{-1}\, X)^{-1}]\; X' V^{-1}\, Y$

$$= \frac{1}{D} \sum_{i=1}^{n} \left\{ \begin{matrix} [a_{11}\, b_{21}^{(i)} - b_{11}\, a_{21}^{(i)}]\, Y_{i1} + [a_{11}\, b_{22}^{(i)} - b_{12}\, a_{21}^{(i)}]\, Y_{i2} \\ + [a_{11}\, b_{23}^{(i)} - b_{13}\, a_{21}^{(i)}]\, Y_{i3} \end{matrix} \right\} \qquad (2.8)$$

As a special case let

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu+\tau \\ \mu \\ \mu-\tau \end{pmatrix}, \Sigma \right)$$

where Σ is an equivariance matrix specified earlier, $i = 1, 2, \ldots, n$. We have introduced this model to study the effect of regression to the mean (James (1973); Senn & Brown (1985)). It may be mentioned that $\lambda = \tau/\Delta (= -\beta_1$ in the repeated measures model (2.4)).

Then

$$E(\hat{\beta}_1 | Y_1 > C_1,\ Y_2 > C_2,\ Y_3 > C_3)$$

$$= \frac{1}{D} \left\{ \begin{array}{l} \sum_{i=1}^{n} (a_{11}\, b_{21}^{(i)} - b_{11}\, a_{21}^{(i)}) \quad (\mu + \tau + \frac{1}{\alpha \sqrt{\theta_0}} (A\theta_0 + B\theta_1 + C\theta_2)) \\ + \sum_{i=1}^{n} (a_{11}\, b_{22}^{(i)} - b_{12}\, a_{21}^{(i)}) \quad (\mu + \frac{1}{\alpha \sqrt{\theta_0}} (A\theta_1 + B\theta_0 + C\theta_3)) \\ \sum_{i=1}^{n} (a_{11}\, b_{23}^{(i)} - b_{13}\, a_{21}^{(i)}) \quad (\mu - \tau + \frac{1}{\alpha \sqrt{\theta_0}} (A\theta_2 + B\theta_3 + C\theta_0)) \end{array} \right\}$$

(see Tallis (1961)),

$$= \left\{ \begin{array}{l} \tau \sum_{i=1}^{n} a_{11}\, (b_{21}^{(i)} - b_{23}^{(i)}) \quad - (b_{11} - b_{13})\, a_{21}^{(i)} \\ + \frac{1}{\alpha \sqrt{\theta_0}} A \left[\begin{array}{l} \sum_{i=1}^{n} (a_{11}\, b_{21}^{(i)} - b_{11}\, a_{21}^{(i)}) \theta_0 + (a_{11}\, b_{22}^{(i)} - b_{12}\, a_{21}^{(i)})\ \theta_1 \\ \qquad + (a_{11}\, b_{23}^{(i)} - b_{13}\, a_{21}^{(i)})\ \theta_2 \end{array} \right] \\ + \frac{1}{\alpha \sqrt{\theta_0}} B \left[\begin{array}{l} \sum_{i=1}^{n} (a_{11}\, b_{21}^{(i)} - b_{11}\, a_{21}^{(i)}) \theta_1 + (a_{11}\, b_{22}^{(i)} - b_{12}\, a_{21}^{(i)})\ \theta_0 \\ \qquad + (a_{11}\, b_{23}^{(i)} - b_{13}\, a_{21}^{(i)})\ \theta_3 \end{array} \right] \\ + \frac{1}{\alpha \sqrt{\theta_0}} C \left[\begin{array}{l} \sum_{i=1}^{n} (a_{11}\, b_{21}^{(i)} - b_{11}\, a_{21}^{(i)}) \theta_2 + (a_{11}\, b_{22}^{(i)} - b_{12}\, a_{21}^{(i)})\ \theta_3 \\ \qquad + (a_{11}\, b_{23}^{(i)} - b_{13}\, a_{21}^{(i)})\ \theta_0 \end{array} \right] \end{array} \right\}.$$

For equal Δ,

$$E\,(\hat{\beta}_1 \mid Y_1 > C_1\ ,\ Y_2 > C_2\ ,\ Y_3 > C_3)$$

$$= -\frac{\tau}{\Delta} + \frac{n\Delta\,|\Sigma|}{\alpha\sqrt{\theta_0 D}}\ \left(-(b_{12} + 2b_{13})\ A - (b_{13} - b_{11})\ B + (2b_{11} + b_{12})\ C\right), \qquad (2.9)$$

$$= \hat{\beta}_1 + bias,$$

where $|\Sigma| = \theta_0^3 - \theta_0(\theta_1^2 + \theta_2^2 + \theta_3^2) + 2\,\theta_1\,\theta_2\,\theta_3$

$$D = n\Delta^2\,[a_{11}(2\,\theta_0^2 - \theta_1^2 - \theta_3^2 - 2\,(\theta_1\,\theta_3 - \theta_0\,\theta_2)) - (b_{13} - b_{11})^2]$$

$$a_1 = \frac{c_1 - (\mu + \tau)}{\sqrt{\theta_0}},\ a_2 = \frac{c_2 - \mu}{\sqrt{\theta_0}},\ a_3 = \frac{c_3 - (\mu - \tau)}{\sqrt{\theta_0}}$$

$$A = \phi(a_1)\ \Phi_2(A_{12},\ A_{13},\ \rho_{23.1}),\ \rho_{23.1} = \frac{\theta_0\,\theta_3 - \theta_1\,\theta_2}{\sqrt{(\theta_0^2 - \theta_1^2)\ (\theta_0^2 - \theta_2^2)}}$$

$$B = \phi(a_2)\ \Phi_2(A_{21},\ A_{23},\ \rho_{13.2}),\ \rho_{13.2} = \frac{\theta_0\,\theta_2 - \theta_1\,\theta_3}{\sqrt{(\theta_0^2 - \theta_1^2)\ (\theta_0^2 - \theta_3^2)}}$$

$$C = \phi(a_3)\ \Phi_2(A_{31},\ A_{32},\ \rho_{12.3}),\ \rho_{12.3} = \frac{\theta_1\,\theta_0 - \theta_2\,\theta_3}{\sqrt{(\theta_0^2 - \theta_2^2)\ (\theta_0^2 - \theta_3^2)}}$$

$$A_{12} = \frac{\theta_0\,a_2 - \theta_1\,a_1}{\sqrt{\theta_0^2 - \theta_1^2}},\quad A_{21} = \frac{\theta_0\,a_1 - \theta_1\,a_2}{\sqrt{\theta_0^2 - \theta_1^2}}$$

$$A_{13} = \frac{\theta_0\,a_3 - \theta_2\,a_1}{\sqrt{\theta_0^2 - \theta_2^2}},\quad A_{31} = \frac{\theta_0\,a_1 - \theta_2\,a_3}{\sqrt{\theta_0^2 - \theta_2^2}}$$

$$A_{23} = \frac{\theta_0 a_3 - \theta_3 a_2}{\sqrt{\theta_0^2 - \theta_3^2}}, \quad A_{32} = \frac{\theta_0 a_2 - \theta_3 a_3}{\sqrt{\theta_0^2 - \theta_3^2}}$$

$$b_{11} = \theta_0^2 - \theta_3^2 + \theta_2 \theta_3 - \theta_0 \theta_1 + \theta_1 \theta_3 - \theta_0 \theta_2$$

$$b_{12} = \theta_2 \theta_3 - \theta_0 \theta_1 + \theta_0^2 - \theta_2^2 + \theta_1 \theta_2 - \theta_0 \theta_3$$

$$b_{13} = \theta_1 \theta_3 - \theta_0 \theta_2 + \theta_1 \theta_2 - \theta_0 \theta_3 + \theta_0^2 - \theta_1^2$$

$\alpha = P(X_1 > a_1, X_2 > a_2, X_3 > a_3)$, where $[X_1 X_2 X_3]' \sim N_3(0, \Sigma)$,

$$\Sigma = \begin{bmatrix} 1 & \rho_{12.3} & \rho_{13.2} \\ \rho_{12.3} & 1 & \rho_{23.1} \\ \rho_{13.2} & \rho_{23.1} & 1 \end{bmatrix}.$$

For equal Δ, when Σ on page 5 is a banded matrix, i.e. when $\theta_1 = \theta_3$, $b_{11} = b_{13}$ and therefore $b_{12} + 2\, b_{13} = 2\, b_{11} + b_{12}$. Then (2.9) becomes

$E(\hat{\beta}_1 | Y_1 > C_1, Y_2 > C_2, Y_3 > C_3)$

$$= -\frac{\tau}{\Delta} + \frac{n\Delta |\Sigma|}{\alpha \sqrt{\theta_0} D} (b_{12} + 2\, b_{13}) (-A + C) = \beta_1 + \text{ bias.}$$

In addition, if we take the cut points $a_1 = a_3$, then $A_{12} = A_{32}$, $A_{13} = A_{31}$. Since, in the banded case, $\theta_1 = \theta_3$ we have $\rho_{23.1} = \rho_{12.3}$. Therefore A = C and

$$E(\hat{\beta}_1 | Y_1 > C_1, Y_2 > C_2, Y_3 > C_3) = \frac{-\tau}{\Delta} = \beta_1.$$

Hence $\hat{\beta}_1$ is an unbiased estimator of β_1.

Remark 1 If we take $a_1 = a_2 = a_3 = a$ in the general case, then in

(2.9) $A = \phi(a)\,\phi_2(A_{12}, A_{13}, \rho_{23.1})$,

$B = \phi(a)\,\phi_2(A_{12}, A_{23}, \rho_{13.2})$,

$C = \phi(a)\,\phi_2(A_{13}, A_{23}, \rho_{23.1})$,

where

$$A_{12} = \sqrt{\frac{\theta_0 - \theta_1}{\theta_0 + \theta_1}} = A_{21},$$

$$A_{13} = \sqrt{\frac{\theta_0 - \theta_2}{\theta_0 + \theta_2}} = A_{31},$$

$$A_{23} = \sqrt{\frac{\theta_0 - \theta_3}{\theta_0 + \theta_3}} = A_{32}.$$

Remark 2 It may be noted that $\hat{\beta}_1$ is an unbiased estimator of $\beta_1 = \frac{-\tau}{\Delta}$ when Σ has compound symmetry or is autoregressive, since both are special cases of when Σ is a banded matrix.

3. BIAS VERSUS SLOPE

In this section we present the analyses of the bias using the Ranch Hand data set of 240 subjects with three measurements. Bias is graphed as a function of γ (the slope of the line passing through $((t_1, c_1), (t_2, c_2), (t_3, c_3))$ for different values of the other parameters. Figures 1 and 2 show that bias is an increasing function of γ as well as σ. Bias is negative for $\gamma < -.08$ and positive for $\gamma > -.08$ and is zero for $\gamma = -.08$. The effect of σ on the bias is negligible for values of $\gamma < -.05$ and

is more pronounced for larger values of γ. It can also be seen from these two figures that the bias is decreasing function of ρ. The same fact is demonstrated in Figure 3. Figure 4 shows that the bias is an increasing function of γ as well as that of the cut points C_3. Finally Figure 5 shows that the bias is an increasing function of γ as well as θ_3. It can be seen that the bias can be zero even for $\theta_1 = \theta_3$ for some particular values of the parameters.

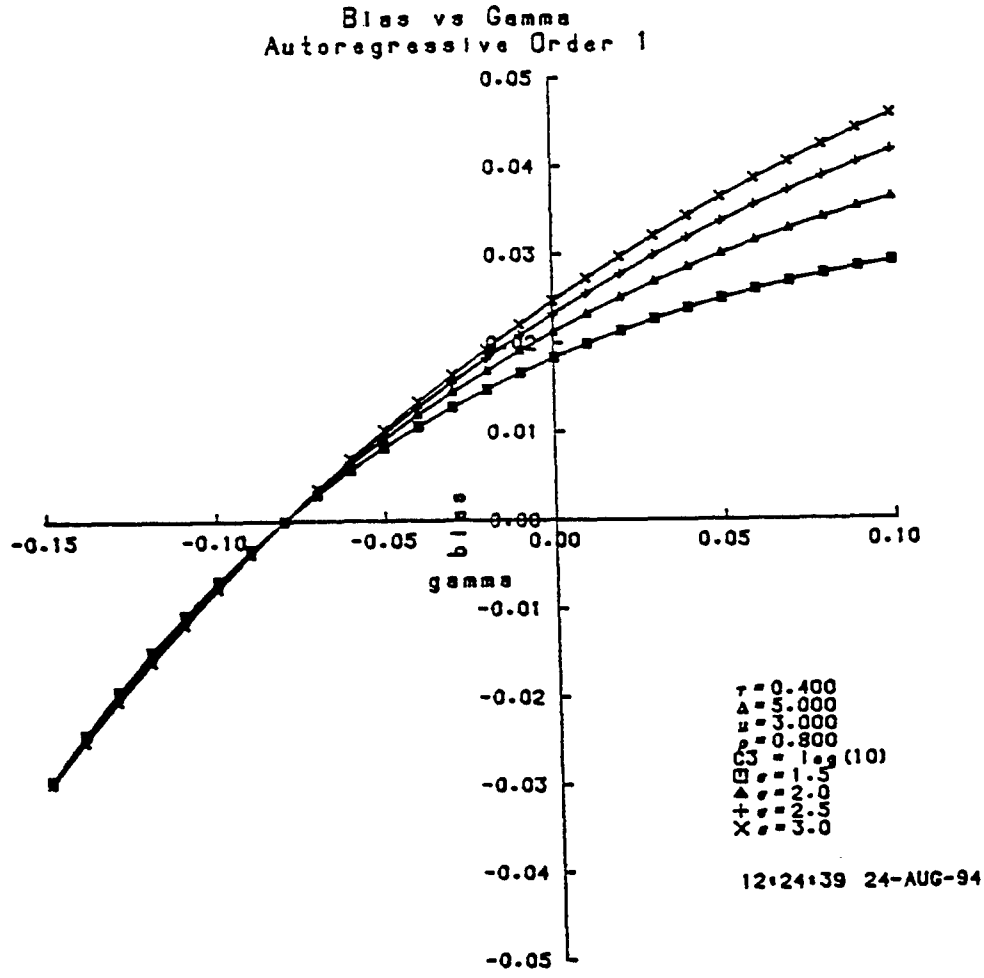


Figure 1

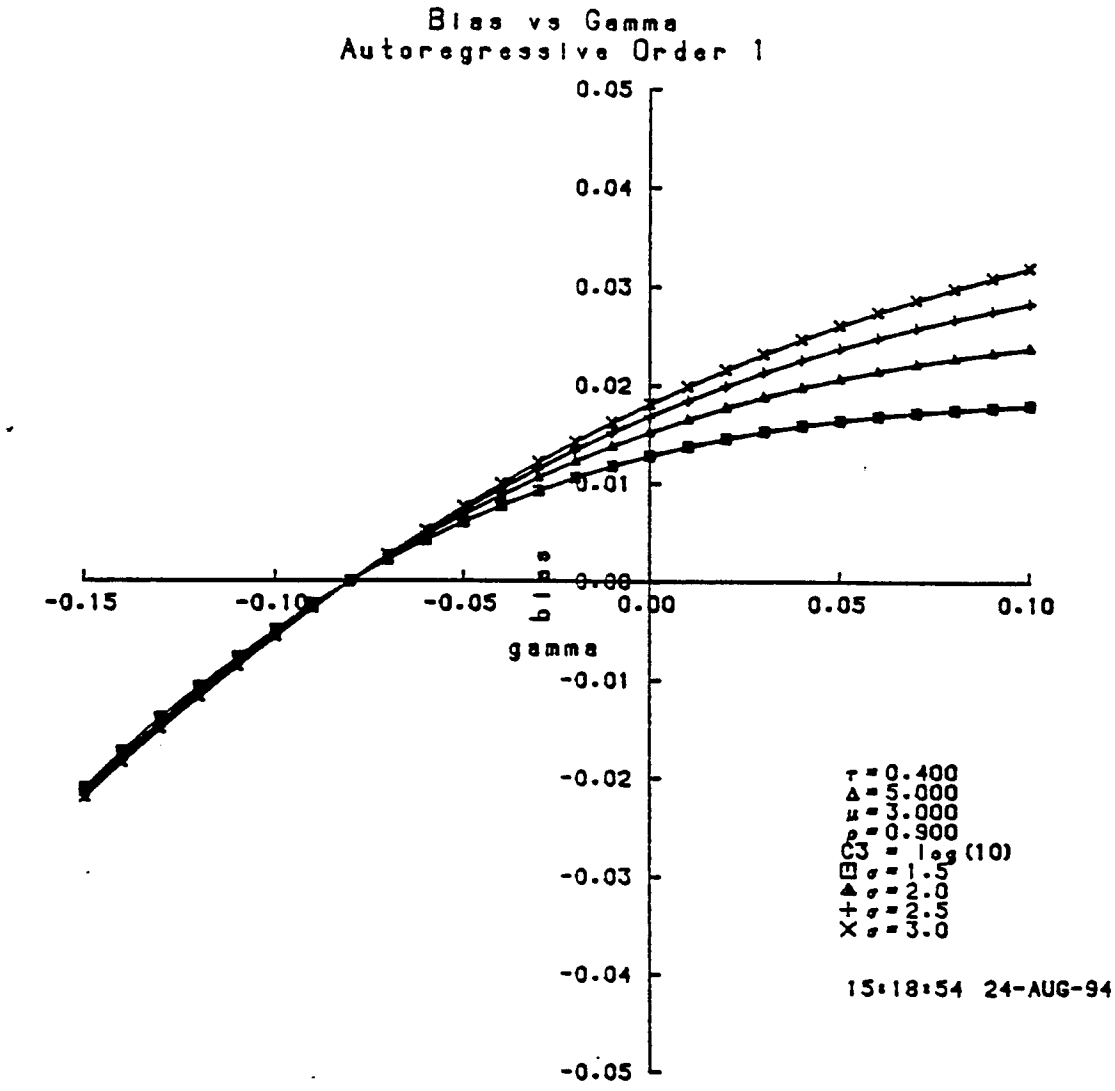

Bias vs Gamma
Autoregressive Order 1
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.15
-0.10
-0.05
0.00
0.05
0.10
gamma
τ = 0.400
Δ = 5.000
μ = 3.000
ρ = 0.900
σ3 = log (10)
σ = 1.5
σ = 2.0
σ = 2.5
σ = 3.0
15:18:54 24-AUG-94

Figure 2

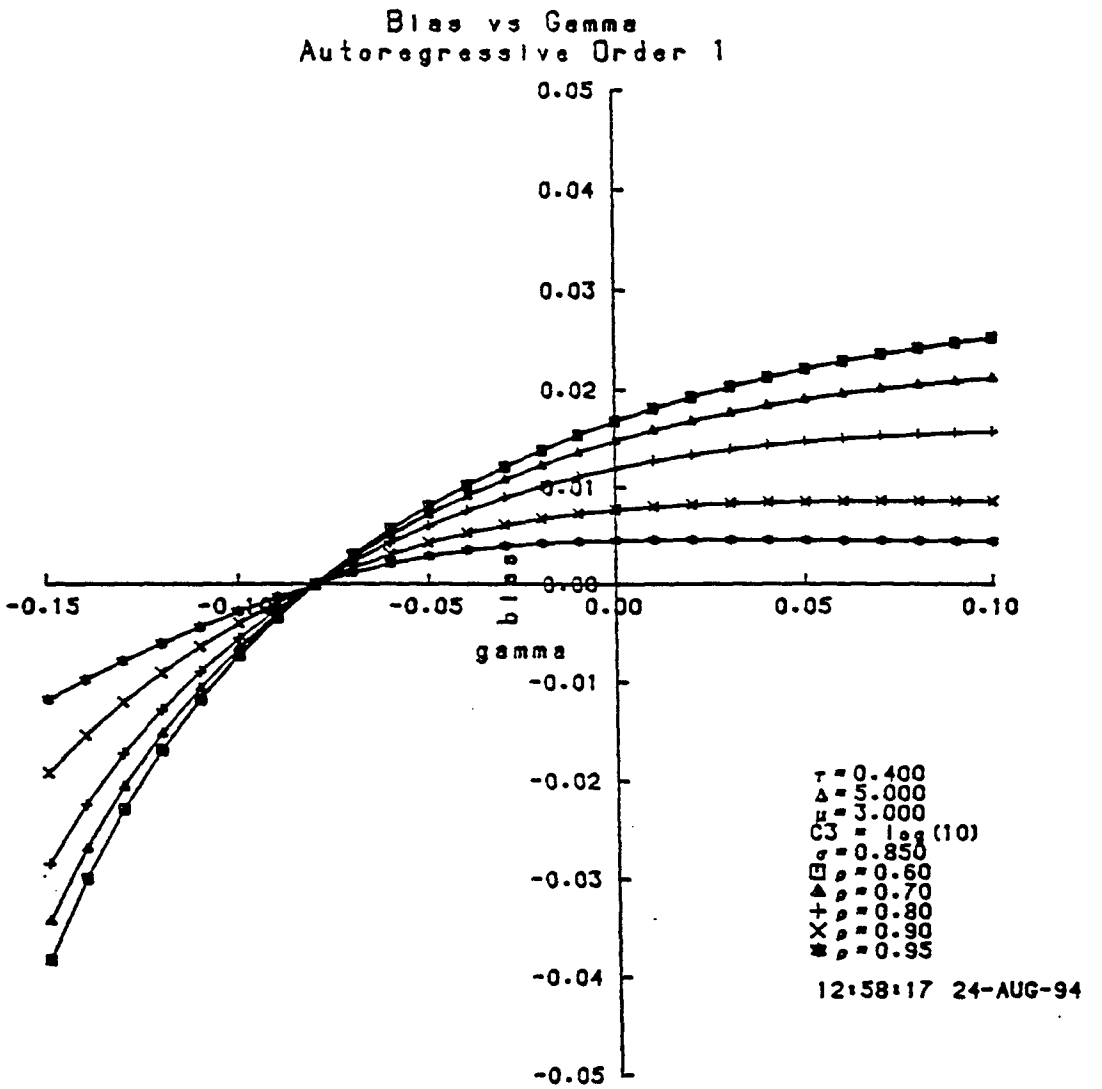

Bias vs Gamma
Autoregressive Order 1
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.15
-0.05
0.00
0.05
0.10
bias
gamma
τ = 0.400
Δ = 5.000
H = 3.000
CJ = log (10)
q = 0.850
ρ = 0.60
ρ = 0.70
ρ = 0.80
ρ = 0.90
ρ = 0.95
12:58:17 24-AUG-94

Figure 3

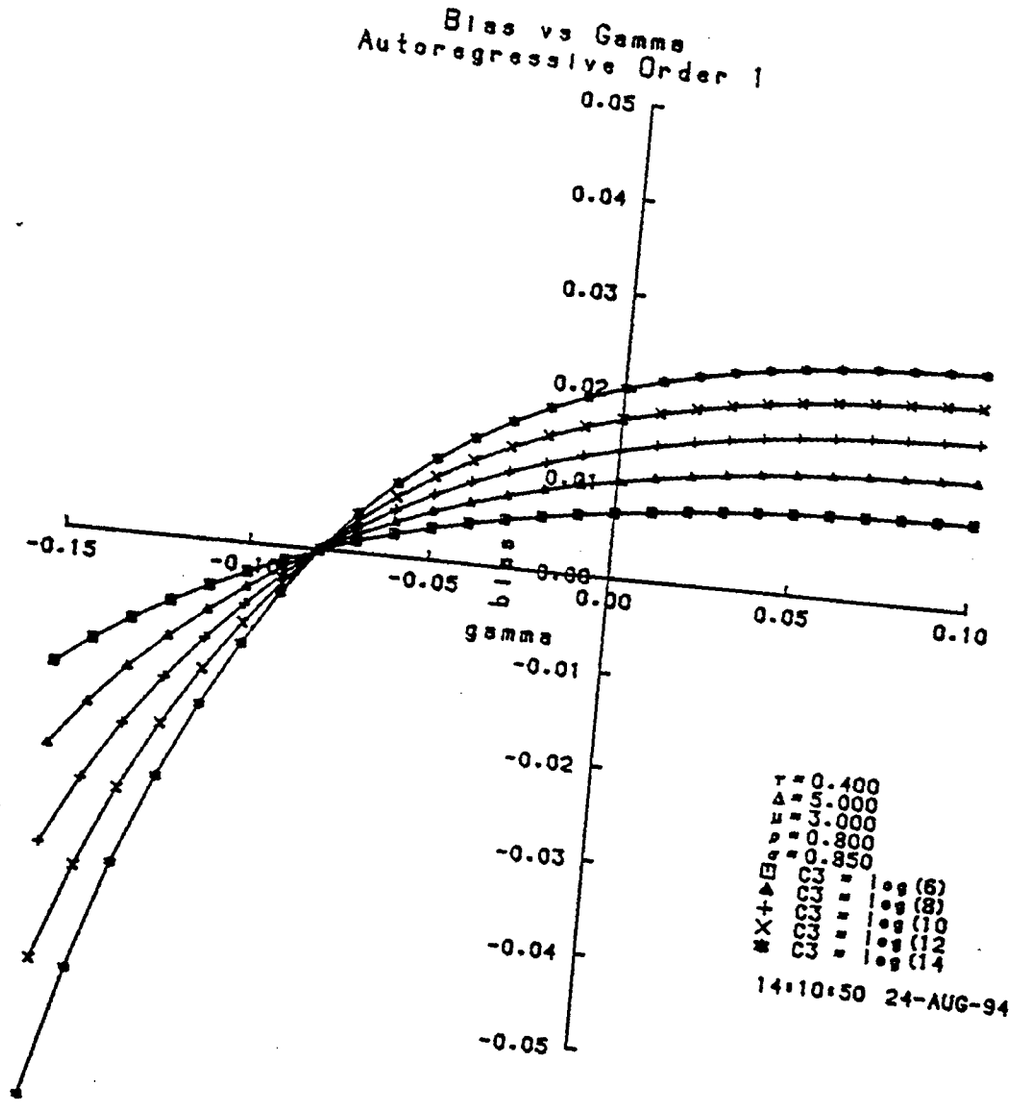

Bias vs Gamma
Autoregressive Order 1
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.15
-0.10
-0.05
0.00
0.05
0.10
gamma
T = 0.400
Δ = 5.000
m = 3.000
p = 0.800
q = 0.850
log (6)
log (8)
log (10
log (12
log (14
14:10:50 24-AUG-94

Figure 4

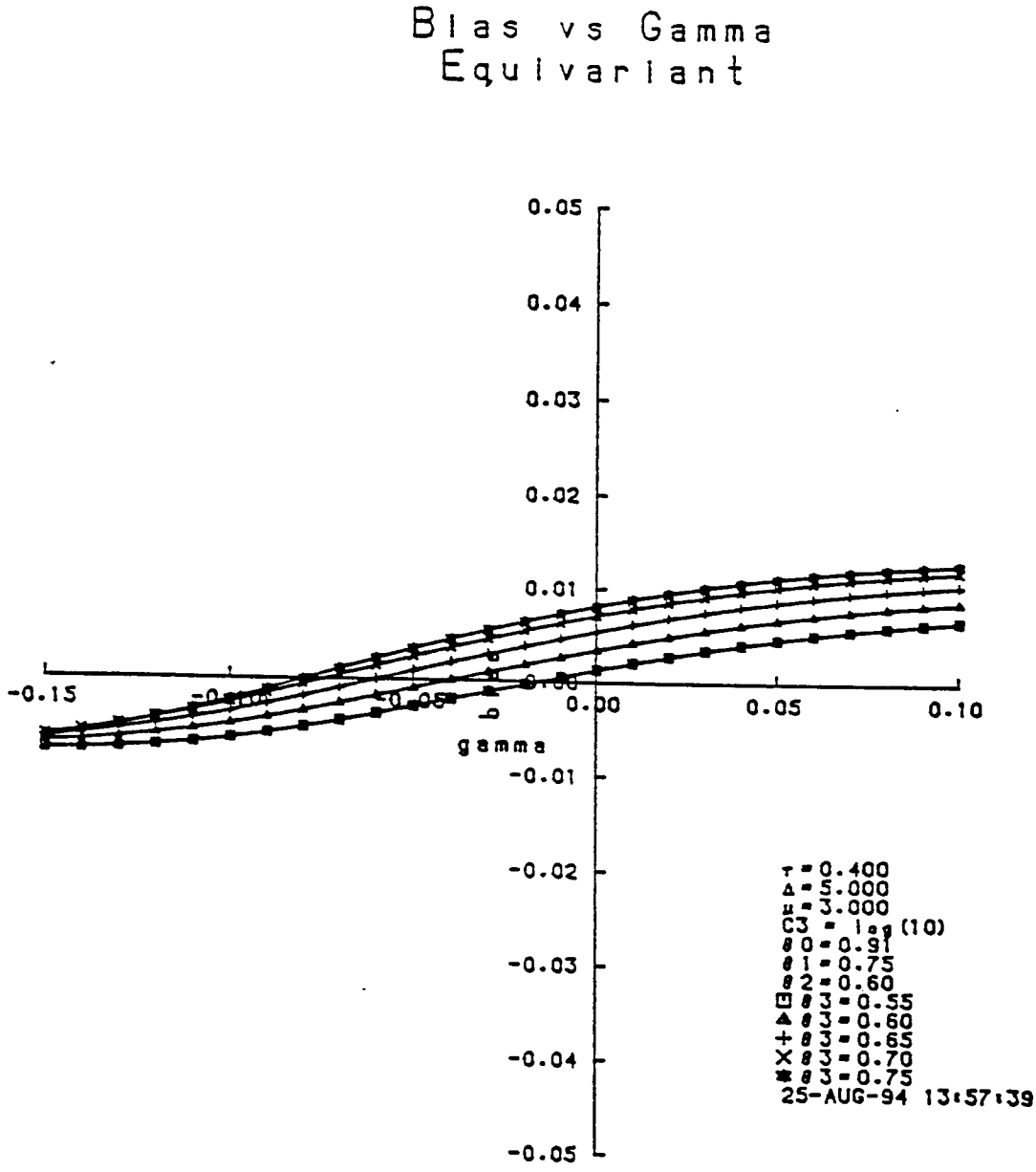

Bias vs Gamma
Equivariant
0.05
0.04
0.03
0.02
0.01
-0.01
-0.02
-0.03
-0.04
-0.05
-0.15
0.00
0.05
0.10
gamma
τ = 0.400
Δ = 5.000
μ = 3.000
C3 = log(10)
θ0 = 0.91
θ1 = 0.75
θ2 = 0.60
θ3 = 0.55
θ3 = 0.60
θ3 = 0.65
θ3 = 0.70
θ3 = 0.75
25-AUG-94 13:57:39

Figure 5

4. CONCLUSIONS AND RECOMMENDATIONS

In this project we show that unless the data set is properly conditioned ordinary weighted least squares estimates of λ are biased due to regression toward the mean. If the within-subject correlation matrix is banded, we show that the weighted least squares estimate of λ is unbiased when the data set is conditioned on all repeated measures being above a line with slope $-\lambda$.

In the end the analyses of the bias, using the Ranch Hand data set of 240 subjects with three measurements, is given. Bias is graphed as a function of λ (the slope of the line through (t_1, C_1), (t_2, C_2) *and* (t_3, C_3)).

Considering the importance of this work, it is hoped that this work will be a significant contribution to the Air Force Health Study. Due to the shortage of time, this work could not be generalized to K measurements for equivariant Σ. This and other issues will be considered in a future study if funding is available.

ACKNOWLEDGEMENTS

I would like to thank the Armstrong Laboratory and the Air Force Office of Scientific Research for sponsoring this research. I would also like to thank the RDL for their efficiency in administering the Summer Faculty Research Program.

My special thanks are due to Dr. Joel E. Michalek for providing me some insight in to this problem. I would also like to thank Mr. Tom White for providing me computational help.

REFERENCES

James, K.E. (1973). Regression toward the mean in uncontrolled clinical studies. Biometrics, 29, 121-130.

Senn, S.S. and Brown, R.A. (1985). Estimating treatment effects in clinical trials subject to regression to the mean. Biometrics, 41, 555-560.

Tallis, G.M. (1961). The moment-generating function of the truncated multinormal distribution. J. R. Statist. Soc. B, 23, 223-229.

SUMMARY OF THREE PAPERS:

I. AN EXPANDED VERSION OF THE KULHAVY/STOCK MODEL OF INSTRUCTIONAL FEEDBACK

II. LOGISTIC MODELS USING MULTIPLE INDICATORS OF RESPONSE STATE TO PREDICT SUBSEQUENT RESPONSE CORRECTNESS

III. GENDER AND DEVELOPMENTAL DIFFERENCES IN ACADEMIC STUDY BEHAVIORS

Thomas E. Hancock
Assistant Professor
Educational Psychology

Grand Canyon University
3300 West Camelback University
Phoenix, Arizona 85017

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

SUMMARY OF THREE PAPERS:
I. AN EXPANDED VERSION OF THE KULHAVY/STOCK MODEL OF INSTRUCTIONAL FEEDBACK
II. LOGISTIC MODELS USING MULTIPLE INDICATORS OF RESPONSE STATE TO PREDICT SUBSEQUENT RESPONSE CORRECTNESS
III. GENDER AND DEVELOPMENTAL DIFFERENCES IN ACADEMIC STUDY BEHAVIORS

Thomas E. Hancock
Assistant Professor
Educational Psychology
Grand Canyon University

Abstract

This report is a summary of three journal articles completed this summer. The results of previous papers and reports, including additional analyses, have been drawn together and reinterpreted. The first draws together literature regarding the Kulhavy/Stock model of instructional feedback, then summarizes data that support an expanded version of their model. It is suggested that the model include a higher order control system which accounts for learners' goals and which system governs the behavior in the lower order systems focused on responding correctly to instructional demands. It is further proposed that analyses and modeling be conducted on a subject by subject basis and that the use of feedback be understood more in terms of each learner's control systems and less in terms of an S-R or cause-effect orientation. The second summary reports measures of the response state of a learner as a means of predicting posttest correctness. It has been suggested that student modeling could incorporate more measures of the cognitive state at the time of responding. Comments are made regarding the future of such work.And finally, a paper is summarized which reports on what we believe is the initial source of study behaviors in adults: the development of such behaviors in elementary school children. Results provide explanations for persisting gender differences in the performance of complex academic tasks and demonstrate the importance of interpreting students' behaviors in terms of environmental and personalogical factors.

SUMMARY OF THREE PAPERS

Thomas E. Hancock

I. AN EXPANDED VERSION OF THE KULHAVY/STOCK MODEL OF INSTRUCTIONAL FEEDBACK

Kulhavy and Stock (1989) have recently articulated a control model which explains the use of instructional feedback. This model has been cited numerous times and is taken by some, including the present author, as feedback's most definitive model (e.g Dempsey, Driscoll, & Swindell, 1993; Hancock, Hubbard, & Thurman, 1992a; Mory, 1992; Shutz, 1993). The following are some of its strengths. It clearly defines the instructional feedback paradigm as one with three cycles. It breaks down the feedback message into specific and testable components. It emphasizes the internal processing of the feedback message, arguing that this has been too long ignored in research. It calls attention to the cognitive correlates that are associated with each response in each cycle. It is expressed in terms that are testable. And it builds on a long tradition of researchers appealing for a closed-loop explanation for the function of feedback (Adams, 1968; Guthrie, 1971; Talyzina, 1981; Smith & Smith, 1966).

The literature has yielded several suggestions for the extension of the Kulhavy/Stock model. Some of those are mentioned here. The most common is that there should be a place for the learner's goals in any model explaining the use of feedback (Bangert-Drowns, Kulik, Kulik, & Morgan,1992; Butler, Winne, McGinn, 1993; Dempsey,et. al., 1993; Hancock, Thurman, & Hubbard, 1993a; Mory, 1992; Shutz, 1992). A second comment from the literature is that feedback can be processed mindfully or mindlessly (Bangert-Drowns, et. al.,1992) but that the Kulhavy/Stock model assumes that processing is automatic (Hancock, et. al., 1992a; Swindell & Walls, 1993). Also it is suggested that a model for the use of feedback must be expressed in terms that incorporate each learner's unique characteristics (Butler, et. al., 1993; Dempsey,et. al., 1993; Hancock, Hubbard, Thurman, 1992b). And related to that, it is suggested that for the most practical relevance in applied settings, such as with intelligent computer-aided instruction, a model must be applied not solely according to group-based conclusions (Sales, 1992), but with statistical fits on a subject by subject basis (Hancock, Thurman, & Hubbard, 1993b). And finally, it is pointed out (Hancock, Thurman, & Hubbard, in submission) that though the Kulhavy/Stock model does recognize the importance of the closed-loop understanding of human

functioning and the importance of the learner's perception, it still is overly dependent on an inanimate physical world model, where feedback is treated as a stimulus or force, that somehow is to cause a learner to perform in a certain manner, and that the learner should respond according to some natural laws of the feedback message.

The purpose of this paper is to build on the pioneering work of Kulhavy and Stock. (Hereafter in this paper the Kulhavy/Stock model will be referred to as The Model.) A pivotal part of The Model is the certitude measure and the construct of discrepancy in a control system framework. Typically the certitude measure is a metacognitive rating by the subject immediately following responding to a criterion task. The subject is asked, "How certain are you?" Kulhavy and others have found that certitude or confidence measures are predictive of future performance (Hancock, Stock & Kulhavy, 1992; Kulhavy, Stock, Hancock, Swindell, & Hammrich, 1990; Shutz, 1993; Stock, Kulhavy, Pridemore & Krug, 1992; Swindell & Walls, 1993; Webb, Stock & McCarthy, 1994). In particular, feedback frame latencies increase as certitude decreases following correct responding but latencies decrease as certitude decreases following incorrect responding (Hancock, Stock, & Kulhavy, 1992; Kulhavy & Stock, 1989).

The Model proposes that humans function as control systems when they process feedback. The perception of the task demand is compared to the related cognitive referents and an error signal is output. That error signal appears to be measured with a discrepancy scale based on certitude and initial response correctness. If The Model's discrepancy scale is valid then as discrepancy increases so should the opposition to it. In general this is what happens. That is, as discrepancy increases so does feedback processing time, and thus the plausibility of a control explanation is supported — control systems oppose error signals. Subjects perceive an instructional demand and they have some error signals related to that demand, which error signals can be opposed or eliminated by studying feedback.

In the present paper it is proposed that the closed loop understanding of humans is exactly what is needed (cf. Shutz,1993), not only to account for the basic level of perceiving task demands and responding correctly, but also to account for the place of goals, mindful or mindless processing, and the uniqueness of each individual.

As a first step in an expansion to The Model, a higher order control system above the cognitive referents in each cycle is needed. Thus the application of the cognitive referents to the perceived task demand are controlled by another control system. This control system's reference standards would be the subject's goals at the time of

responding. Based upon the discrepancy between goals and present time attainment of those goals, there would be error signals output. In instructional feedback sequences these varying strength error signals would go to the "responding correctly" systems, which are Kulhavy and Stock's basic concern with the subject's perception of the task demand. However, whether a subject will have discrepancy for responding correctly depends, not simply on the perception of the task demand, but on the signal from the higher level goal system. Thus the subject will not automatically process feedback in order to respond correctly.

In this Expanded Model it is pivotal that every human student has certain reference standards (or goals or what's deemed important to that individual) while engaged in an instructional sequence. The subjects with goals for correct responses have reference standards requiring that they see themselves getting correct response messages. They are most likely subjects who hold what has been called "performance goals" (Dweck, 1986; Nichols, 1984). If they do not see correct answers on the screen there will be error signals output from the learners' "correct response" control system; they will activate some programs, such as studying feedback, to do what they can to see correct responses. Therefore, upon the display of the "wrong" feedback message these subjects should exhibit longer latencies than upon the display of a correct message.

Other subjects have goals not only for correct responding, but also for learning and understanding. They hold "learning goals" (Dweck, 1986; Nichols, 1984). They have reference standards requiring that they see themselves learning and understanding. If, for example, they are not certain of a response even though it is correct, there will be error signals output from the subjects' "learning or certainty" control systems; they will activate some programs, such as studying feedback, to do what they can to see themselves being certain. Therefore, at the feedback frame with the redisplay of the stimulus array and the correct name, these subjects should exhibit longer latencies as metacognitive ratings of certainty decrease, even when the response has been outwardly correct. They should use feedback the most and they should exhibit the most learning. These are the subjects that The Model assumes.

And finally any comprehensive explanation of the function of feedback must additionally treat those who do little or nothing with the feedback message — those who do not have strong goals for learning or for responding correctly. For example, a subject (or student) may be controlling for getting to the end of the task —

participation in an experiment or in a learning task is not important. Other goals, from higher level systems, are sending out stronger signals than ones from the control systems for learning or responding correctly. So when feedback is administered, the only way it can help them to reduce error signals from the "get to the end of the task" control system, for example, is to quickly go on to the next frame regardless of certainty or outward correctness. These subjects consequently should not achieve high rates of correct responding. Thus, the use of the feedback message by lower achieving students (ones with the lowest mean correct responses) should show little systematic relationship between feedback latencies and correctness or certitude.

Two experiments were arranged whose results fit well with the above discussion on the varied use of the feedback message.

Experiment 1

This methodology is reported by the author in the final report of the Summer Faculty Research Program of 1992, but is summarized here for ease of reading.

The basic experimental design was a mixed factorial with one between subjects factor — 5 levels of feedback, and two within subjects factors — 5 levels of certitude estimate, and 2 levels of response correctness The predicted value was the feedback frame latencies (feedback study time). Fifty-four university undergraduates participated for partial course credit. MacIntosh Plus computers were programmed to present the task and collect data.

The stimuli consisted of 27 separate items each presented as a screen of information. Each item consisted of three separate but simultaneous displays of graphic, aural and iconic information. In addition, the graphic, aural and iconic information could be presented at one of three levels, yielding a total of 27 combinations. The right side of the screen also displayed 27 names. Each display item was associated with one and only one of these names. The subject's task was to identify the item by clicking on the appropriate name with the computer's mouse.

The certainty rating screens included a certainty rating scale: "How certain are you that your response is correct?" 100% certain, 75% certain, 50% certain, 25% certain, 0% certain. There was a radio button to the left of each level of certainty.

Each feedback screen displayed one of five levels of response sensitive feedback information which varied between subjects. In every case the feedback displayed basic response sensitive verification information : "No, (or yes) the correct response was (name button)" (Kulhavy & Stock, 1989), plus the redisplay of the waveform in

its correct position. The procedure during each session was as follows: 1. view stimulus item and "click on" a name button; 2. view a certainty rating scale and select a rating; 3. view the feedback screen, and press a "continue" button; 4. view the next stimulus item, etc. Each session was about 40 minutes.

The initial results are reported in the 1992 Summer Report. The predictions were confirmed. Subjects were grouped by ability (three groups of mean correctness), then the Proc GLM procedure (SAS Institute) was applied to each subject individually. Significant predictors of posttest response correctness were identified. The following was the breakdown of significant predictors: top ability — correctness, 100%; certitude, 61%; correctness by certitude, 30%; middle ability: correctness, 68%; certitude, 36%; correctness by certitude 11%; lowest ability: correctness, 20%; certitude, 20%; correctness by certitude, 0%.

Then analyses were conducted by ability group, response correctness, certitude and the interactions. Main effects were significant (alpha = .01), but more importantly the two-way interactions with ability group were all significant. Inspection of means confirms predictions. Subjects in the top ability group spend more time with feedback after incorrects and following lower certitude corrects. Subjects in the middle group display an effect for correctness but not so much for certitude. And subjects in the lower group display little difference in feedback frame latencies following incorrects or corrects and after low or high certainty.

In addition, more direct evidence for the influence of the subject's goals was obtained from a post-experiment questionnaire which was first analyzed this summer. A subset of the original sample (n = 23) had been surveyed. These subjects were from one of the classes of the senior author. In this class, goals and priorities had been discussed as part of the regular instruction. After the experiment, these students were asked to rate their two top priorities or goals while they were participating in the experiment. The subjects were grouped fairly evenly according to goals and their mean feedback time was calculated. The response patterns discussed above are evident: Getting finished and being correct = 5.0 seconds; Being correct (both goals) = 5.1 seconds; Learning and getting finished = 5.4 seconds; Learning and being correct = 5.7 seconds; Using efficient strategies = 7.8 seconds; Just learning (both goals) =10.8 seconds.

Thus, it is confirmed that those subjects who have a goal to be finished tend to spend less time with feedback. Those who have a goal to be correct spend more time. And those whose goal is related to learning spend the most time. We see that subjects'

effort is logically related to their goals.

Experiment 2

The methodology was reported in the final report of Summer Faculty Research Program of 1993. The analyses are new.

In Experiment 2 we used the basic instructional stimuli from Experiment 1, but we changed the program slightly. We moved the certitude estimate — in this case to the frame following the feedback. And also we chose to have only one type or load of feedback, one not included in Experiment 1, but one which results indicated might be the most effective. Twenty-six university undergraduates participated.

The feedback was a basic response verification, an indication of the correct response, in addition to name buttons for both the correct and incorrect which were enabled to display visual and echoic elaborative information upon learner selection — a redisplay of the task stimuli.

Thus we sought to obtain another type of evidence for the use of the feedback message. Instead of a dependent measure of feedback frame latencies we used choice of feedback: that is, the selection of elaborative feedback information was learner controlled and the number of selections of that feedback was recorded automatically.

A simple chi-square of the probability of selecting elaborative feedback when correct was significantly greater for the high ability group than for the lower ability groups, chi-square (2) = 243.85. The mean percentages were as follows: low ability, 3.9%; middle ability, 1.2%; high ability, 12.5%.

And following incorrects, the three way interaction for the probability of selecting each type of elaborative feedback was significant. The mean percentages by ability group and type of feedback were as follows: low ability — wrong, 7%; correct, 10%; both, 9%; none, 75%; middle ability — wrong, 6%; correct, 6%; both, 29%; none, 58%; high ability — wrong, 8%; correct, 11%; both, 43%; none, 38%.

The results of Experiment 2 provide further evidence that subjects of various ability levels treat feedback differently. Those who are the better students are those who choose to study feedback more often both following correct responses and following incorrect responses. And following incorrect responses, they choose to study the elaborative information regarding the incorrect and the correct more than the lower achieving students.

General Discussion

On the basic level the results demonstrate that subjects process feedback differently. And these differences are systematically related to performance on the

learning task. These empirical trends are noteworthy. However, we believe that the primary importance of the present work is the strengthening of the Kulhavy/Stock model. Specifically, we have demonstrated that positing a higher level system related to goals helps explain varying patterns for the use of the feedback message. In addition, this paper has demonstrated the benefit of performing analyses on a subject by subject basis.

Future research could investigate aspects of this expanded Kulhavy/Stock model. And it may be that some would want to further increase the power of the model by including at least one other of the 9 levels of perceptual control theory: the program level (see Powers, 1973). That is, subjects not only should have goals which are superordinate to and control the responding correctly systems, but subjects also should have programs, which are subordinate control systems, for studying the feedback message. If these programs could be precisely measured on a subject by subject level, even at a a nominal level, and incorporated into the basic control model, then our predictive and explanatory power may be maximized.

II. Logistic Models Using Multiple Indicators of Response State to Predict Subsequent Response Correctness

A continuing need in instructional research is the identification of more reliable methods of calculating the probability that a particular response from a particular learner will be correct at future trials. With such information about the student, the instructional system, machine or human, may be more able to provide feedback or some other instructional manipulation that is appropriate for that learner. For example, in the domain of computer mediated learning, when the system can identify the probability that a particular item is learned, that probability can be used in the construction of the student model (Atkinson, 1976; Park & Tennyson, 1983). Then feedback and instructional sequences can be generated which are most appropriate for that student model.

In both research and practice, the administration of instructional feedback, has traditionally been based on the correctness of a response (Alessi & Trollip, 1991; Kulhavy & Stock, 1989). Even in the domain of intelligent tutor systems (ITS), though the cognitive state is a concern, the major source of inferences is the outward response history (e.g. Chin, 1993; Johnson & Norton, 1992).

The problem is that such reliance on response alone tends to ignore that each is generated from a particular cognitive state, which we will refer to as the response

state. Hence, instructional interventions tend to be directed toward responses and inferences are made from those responses rather than the more direct measurement of the cognitive state underlying the responses. This may be one reason there is still much uncertainty in constructing student models which adequately predict and match student performance (Elsom-Cook, 1993; Regian & Shute, 1992), and why it is recommended that more and varied measures of the learner's cognitive state be tested (Winne, 1989).

The use of the certitude measure as one means of measuring the response state of the learner has been proposed (cf. Kulhavy & Stock, 1989). And much evidence has accrued that this measure provides predictive power regarding the learner's subsequent responding (Hancock, Stock, & Kulhavy, 1992; Kulhavy, et. al., 1990; Shutz, 1993; Stock, et. al.,, 1992; Swindell & Walls, 1993; Webb, et. al., 1994). In addition, the use of such metacognitive measures has been proposed in improving student modeling in intelligent computer aided instruction (Noble, 1991; Steinberg, 1991; Winne, 1989).

A primary concern of this present paper is that, in addition to certitude estimates, other potentially powerful measures of response state may be available for student model construction. For example, the basic psychological literature indicates that response latency is a measure of the cognitive state and is predictive of subsequent performance (Anderson, 1983; Kyllonen, Tirre & Christal, 1991; Logan & Stadler, 1991; Meyer, Osman, Irwin & Yantis, 1988; Simon & Croft, 1989). Also study time, such as for instructional feedback, has been related to the cognitive state and to subsequent performance (Hancock, Stock, & Kulhavy, 1992; Kulhavy & Stock, 1989; Mazzoni & Cornoldi, 1993; Nelson & Leonesio, 1988; Shafir & Pascual-Leone, 1990; Webb, et al., 1994).

Data are used from three experiments where the response state of the learner is monitored with several measures: certitude estimates and three latency measures (response or criterion task latency, certitude rating latency, and feedback frame latency). It was reasoned that these variables should interact with the load of feedback information and also with the inter-stimulus delay. Both feedback and load have been shown to relate, at least in some cases, to subsequent performance (Dempster & Farris, 1990; Kulhavy & Stock, 1989).

Our primary objective was to determine whether multiple measures of response state could predict posttest performance. The basic method of analysis was to attempt to fit logistic regression models for each group and then separately for each subject.

We have not found any evidence of similar model construction in the feedback or the computer-mediated learning literature.

The methodology of two of these experiments is summarized above in Part I. The third experiment is reported in the 1993 Summer Report.

Data are from a total of 97 subjects. The group-based models did not tend to be acceptable fits, but logistic regression models were fit for about 80% of the individual subjects in each of the three experiments. In addition, in all experiments all of the response state measures were significant for some of the subjects, according to the chi-square goodness of fit test. But the significant predictors varied from one subject to the next.

With the Experiment 3 data, a new analysis was conducted to test the efficacy of a group-based model which accounts for the individual differences interacting with each variable. In this case the total model could not be rejected, chi-square (399) = 401.18, p = .4600. The Wald statistics were as follows: response correctness, chi-square (1) = 32.99, p = 0.000; delay, chi-square (3) = 297.43, p = 0.000; certitude, chi-square (4) = 31.40, p = 0.000; response latency, chi-square (1) = 0.46, p = 0.4992; certitude latency, chi-square (1) = 1.17, p = 0.2793; feedback latency, chi-square (1) = 0.00, p = 0.9468; response latency X subject, chi-square (22) = 30.98, p = 0.0965; certitude latency X subject, chi-square (22) = 29.15, p = 0.1405; feedback latency X subject, chi-square (22) = 45.88, p = 0.0020; subject X delay, chi-square (66) = 115.15, p = 0.0002; subject X response correctness, chi-square (1) = 29.19, p = 0.1394; subject, contained restricted parameters; subject, X certitude, contained restricted parameters.

It is clear that in order for the total model to fit it must account for the subject interaction. This just underlines again the importance of increasing precision by creating a separate model for each subject.

General Discussion

It was demonstrated that posttest performance can be predicted by various measures which are easily gathered in an interactive computer environment.

It is noteworthy that the group-based models did not generally fit, unless the subject interactions were included. In addition, it is even more noteworthy that logistic models were fit for most of the subjects individually with every one of the factors being significant for at least some of the individual subjects.

There are at least two sensible conclusions to be made. First, each of these response state measures may have predictive potential for use in various instructional systems. Second, using a single group-based model in order to predict

the probability of a subsequent correct response may not be as effective as providing some means for accounting for the uniqueness of each subject, such as by using a separate model which is fit with each subject's significant predictors. Other researchers corroborate this finding (e.g. Runkel, 1990; Walker & Catrambone, 1993).

Future Research

The focus of the present paper has been primarily empirical. However, there remains the question of how to explain the cognition underlying these models. This has been a primary concern in the work performed this summer. Comments regarding the progress along these lines follow.

First of all, we do know that an explanation of the cognition underlying our response state measures should take account of the significant predictors which vary from subject to subject. Our present thought is that such variation is due to the variation in reference standards or goals (see Powers, 1973;1992). Though all subjects are performing the same outward tasks, what is driving their systems — why they are behaving — may differ between subjects (see Cziko, 1991; Runkel, 1990). Thus different predictors may be indicative of different goals and individual parameters.

However, the conclusion is that explanations of human functioning, in instructional feedback episodes for example, should ultimately be tested with mathematic models which specify the functions relating each of the variables. Then the modeled behavior would be compared with the actual. Though the logistic models provide more predictive power than previously available, they do not define a model of an actual human system such as used with the physical sciences or with the newly emerging perceptual control theory (Cziko, 1991; Marken, 1986, 1991; Powers, 1978).

In the attempts at such modeling, it has been realized that various response latencies, though helpful in logistic models and though interesting in their own right, cannot be included as meaningful components of an actual model of a human. Such a model needs precise articulation of the cognitive correlates of the response state — what are the perceptions, what are the specific goals that are operative, and what are the precise strategies or programs being used; and all of this occurs within real time not as a result of time. In other words latency variation is a product of the human processing, but not a cause.

III. Gender and Developmental Differences in Academic Study Behaviors

The third paper completed this summer (Hancock, Stock, Kulhavy, & Swindell, in

submission) was partially motivated with a concern about studying. (The data were gathered several years ago — but never published, and the relationship of the results to the research literature had never been investigated.) A student model in instructional sequences involving feedback should include the studying component. Thus the process of studying needs to be understood more precisely so that individual variation can be more precisely specified.

Examination of reviews of research on academic study behaviors (e.g., Anderson and Armbruster, 1984; Kulhavy and Kardash, 1988; Rohwer, 1984; Snowman, 1986; Thomas and Rohwer, 1986; Weinstein & Mayer, 1986) reveals that researchers have most often focused on specific study skills and their relation to academic performance.

The majority of this study behavior data has been collected from high school and college students, where study behaviors are already well-formed. Some projects reach down to junior high populations (e.g. Nolen, 1988; Thomas & Rohwer, 1987) and earlier school ages (e.g., Brown, 1981). Much of this work has involved manipulation of study behaviors in laboratory settings.

In the present research we were interested in how elementary age children develop and use study procedures in everyday classroom situations. We chose to use a variation of the broad-based approach (see Kulhavy & Kardash, 1988) which is characteristic of European/English researchers (e.g. Biggs, 1976, 1993; Entwistle & Tait, 1990; Marton & Saljo, 1976). Our intent was to identify how naturally occurring study activities group into broad classes of study behavior. Naturally occurring study activities were obtained from children's self-reports of specific study events using the critical incident technique developed by Flanagan (1954). We expected that an analysis of a large number of such incidents would provide markers for a detailed description of the development and use of study procedures across elementary grade levels.

We identified clusters of study behaviors common to elementary school students, derived critical incidents of study behavior from student interviews, and developed a 40-item study behavior questionnaire from the resulting data. Factor analysis of questionnaire data from 803 elementary students yielded significant grade and gender differences. These factors are interpreted in light of the research literature.

The general consensus among the investigators of broad classes of study behavior is that there are certain universal characteristics in academic studying (e.g. Andrews, Violato, Rabb, & Hollingsworth, 1994; Biggs, 1993; Entwistle, 1991; Harper &

Kember, 1989; Marton & Saljo, 1976; Schmeck, Geisler-Bernstein, & Cercy, 1991; Trigwell & Prosser, 1991; Volet & Chalmers, 1992): 1. a deep versus a surface approach to learning is the most basic distinction.2. an achievement or goal orientation which interacts with the deep or surface approach.

Though most study skills' factor analyses have been performed on data gathered at the college level, our sample was from students who are just beginning to develop the kinds of study behavior defined in those inventories. And though most studies are constrained by theory or previous inventories, our approach was initially wide-open with a complete array of naturally occurring study incidents. It is interesting to note that the factors identified above from other research are evident in our elementary student sample. But the unique way in which they manifest is what particularly adds to our understanding of study behaviors.

It appears that boys and girls at fourth grade have certain study behavior similarities: they primarily emphasize overt study activities, and they secondarily show both a concern with tests and for the use of deeper thinking. However, the girls are more occupied with script, their thinking appears to be somewhat deeper, and their study behavior appears to be more deliberate. (For a more detailed discussion of the fourth grade factors see the complete paper.)

By sixth grade the divergence in broad classes of study behavior is greater. It appears that each gender's factors at fourth grade can be traced to sixth grade.

The girls' minor factor at fourth grade which indicates deliberate and planful encoding for tests or criterion success has apparently developed into the girls' dominant factor at sixth grade, where their main concern is with review and rehearsal. Research indicates that there is increasing use of planful activity beyond sixth grade (Christopolous, Rohwer, & Thomas, 1987). Therefore, it appears that girls develop before boys in terms of dominant concern for planful behaviors focused on preparing for tests. We see their strategy indicated in items such as "write over and over," "repeat things," "reread," "read over," "go over things again."

The girls' elaboration and deeper thinking at fourth grade appears only as a minor factor in sixth grade, visual retrieval. Thus, it appears that the overriding concern with tests may have caused girls to be more concerned with review and rehearsal than with deep thinking.

Indeed, other research does indicate that there is a tendency for girls to develop less efficient study behaviors (Boggiano, Main, & Katz, 1991; Licht, Linden, Brown, & Sexton,1984) and also to excel in rote recall (Anastasi, 1958; Maccoby & Jacklin, 1974).

The reason for this developmental change, may be seen in the following research findings: 1) girls are more extrinsically oriented (Boggiano, Main, & Katz, 1991) and less efficient strategies, such as rote recall, are employed when motivation is extrinsic (Kimball, 1989); 2) girls are more influenced by the presence of an adult and are more compliant (Boggiano, Main, & Katz, 1991; Harter, 1977), thus, it is sensible that they would focus on pleasing the teacher or working hard to perform well in the way the teacher has emphasized. Indeed very few teachers suggest cognitive strategies to students (Moeley, Hart, Leal, Santulli, Rao, Johnson, & Hamilton, 1992) so girls appear to revert to review or rehearsal; 3) girls have lower self-concepts than boys (Stipek & Gralinski,1991) even when achievement is actually the same (Fennema & Sherman, 1978), and this gap increases with age (Block & Robins, 1993) so it may be that they are compensating by attempting for superior performance on tests; 3a) when self-concept is low development is impeded (Block & Robins, 1993); 3b) control-deprived students tend to encode new information more deliberately (Pittman & D'Agostino, 1989); 4) girls receive more low-level questions (Barba & Cardinale, 1991) and thus are not stimulated with higher level thinking as boys may be.

In contrast, the boys Factor One at sixth grade indicates that elaboration has developed for them, and particularly with covert activities of retrieval of information that is non-text. The boys have not yet become centered around script or text, so it may be that the research finding of young children being dependent on oral behaviors and not understanding text elements (Rybszynski, 1992) applies particularly to boys who may be slower at developing the traditional school-type behaviors. The boys' previous minor concern with tests does not appear as any consistent concern at sixth grade, in marked contrast to the girls. And the boys' fourth grade concern with paying attention (for performance) appears to have developed into aural comprehension.

The sixth grade boys' "non-text" and their third marker, aural comprehension, fit with the research: 1) boys receive more interaction from the teacher (Jones & Wheatley, 1990; Parsons,Kaezala, & Meece,1982; Simpson & Erikson, 1983;) rather than the text; 2) boys feel more independent (Grieb & Easley, 1984; Kimball, 1989), and hence may not be as concerned with the text and the teachers' structuring instruction around the text; 3) boys are not as good readers (Feingold, 1993; Leinhardt, Seewald, & Engel, 1979) and thus may prefer to listen; 4) boys tend to be better listeners (Brimer, 1969).

The sixth grade boys' major emphasis on "retrieval" can be understood by the following: 1) boys tend to be given higher level thinking demands (Barba & Cardinale, 1991); 2) they are less affected by the teacher's demands and adult presence (Boggiano, Main, & Katz, 1991; Dweck, Davidson, Nelson, & Enna,1978), which would tend to make them conform to outward school-type behaviors; 3) boys are more intrinsically oriented (Boggiano, Main, & Katz, 1991), especially as they get older and thus might tend to emphasize retrieval for present classroom needs rather than deliberate encoding for future demands.

These developmental trends may explain why adolescent girls do not do as well as boys on achievement tests, especially in general knowledge, math, and science (Anastasi, 1958: Feingold, 1993; Hyde, 1981; Maccoby, 1966; Maccoby & Jacklin, 1974; Tyler, 1965), and why these differences get larger after elementary school (Born & Lynn, 1994; Feingold, 1993; Fennema & Sherman, 1977; Maccoby & Jacklin, 1974; Meece & Eccles, 1993). That is, though girls work harder at deliberate encoding and emphasizing text material, they do so with what appears to be more of a performance orientation (Dweck, 1986; Nichols, 1984) with more shallow processing. The boys appear to have more of a learning orientation (Dweck, 1986; Nichols, 1984) using study behaviors which are concerned with elaboration, retrieval, and comprehension.

Thanks to Dr. Richard Thurman, AL/HRA and Dr. David C. Hubbard, UDRI, for their help with the text and analyses. Also, thanks to Debra Bolin for her help with figures.

References

ADAMS, J. A. (1968). Response feedback and learning. *Psychological Bulletin, 70*, 486-504.

ALESSI, S. M., & TROLLIP, S. R. (1991). *Computer-based instruction: Methods and development*, (pp 70-77). Englewood Cliffs: Prentice Hall.

ANASTASI, A. (1958). *Differential psychology* (3rd ed.) New York: MacMillan.

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

ANDERSON, T. H., & ARMBRUSTER, B. B. (1984). Studying. In P. D. Pearson (Ed), *Handbook of Reading Research*. New York: Longman.

ANDREWS, J., VIOLATO, C., RABB, K., & HOLLINGSWORTH, M. (1994). A validity study of Bigg's three-factor model of learning approaches: A confirmatory factor analysis employing a Canadian sample. *British Journal of Educational Psychology, 64*, 179-185.

ATKINSON, R. C.(1976). Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and Instruction* (pp. 81-115). Hillsdale, NJ: Lawrence Erlbaum.

BANGERT-DROWNS, R. L., KULIK, C. C., KULIK, J. A., & MORGAN, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

BARBA, R., & CARDINALE, L. (1991). Are females invisible students? An investigation of teacher student questioning interactions. *School Science and Mathematics, 91(7),* 306-?.

BIGGS, J. B. (1976). Dimensions of study behavior: Another look at it. *British Journal of Educational Psychology, 46,* 68-80.

BIGGS, J. B., (1993). What do inventories of student processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63,* 3-19.

BLOCK, J., & ROBINS, R W. (1993). A longitudinal study of consistency and change in self-esteem from early adolescence to early adulthood. *Child Development, 64,* 909-923.

BOGGIANO, A. K., MAIN, D. S., & KATZ, P. (1991). Mastery motivation in boys and girls: The role of intrinsic versus extrinsic motivation. *Sex Roles, 25(9/10),* 511-520.

BORN, M. P., & LYNN R. (1994). Sex Differences on the Dutch WISC-R: A comparison with USA and Scotland, *Educational Psychology, 14(2),* 249-254.

BRIMER, M. A. (1969). Sex differences in listening comprehension. *Journal of Research and Development in Education, 9,* 171-179.

BROWN, A. L. (1981). Metacognition: The development of selective attention strategies for learning from texts. In M. L. Kamil (Ed.), *Directions in reading: Research and instruction,* (Pp. 21-43). Washington, D. C.: National Reading Conference.

BUTLER , D. L., WINNE, P. H., & MCGINN, M. K. (April,1993). *A model of feedback in self-regulated learning.* Paper presented at the Annual Meeting of the American Educational Research Association, in Atlanta, Georgia.

CHIN, D. N. (1993). Acquiring user models. *Artificial Intelligence Review, 7,* 185-197.

CHRISTOPOLOUS, J. P., ROHWER, W. D., & THOMAS, J. W. (1987). Grade level differences in students' study activities as a function of course characteristics. *Contemporary Educational Psychology, 12,* 303-323.

CZIKO G. A. (1992). Purposeful behavior as the control of perception: Implications for educational research. *Educational Researcher, 21(9),* 10-18.

DEMPSEY, J. V., DRISCOLL, M. P., & SWINDELL, L. K. (1993). Text -based feedback. In J. V. Dempsey & G. C. Sales (Eds.) *Interactive instruction and feedback.* Englewood Cliffs, NJ: Educational Technology Publications.

DEMPSTER, F. N., & FARRIS, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23,* 97-114.

DWECK, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41(10),* 1040-1048.

DWECK, C. S., DAVIDSON, W., NELSON, S., & ENNA, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom. III. An experimental analysis. *Developmental Psychology, 14,* 268-276.

ELSOM-COOK, M. (1993). Student modeling in intelligent tutoring systems. *Artificial Intelligence Review, 7,* 227-240.

ENTWISTLE, N. J. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education, 22,* 201-204.

ENTWISTLE & TAIT, (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education, 19,* 169-194.

FEINGOLD, A. (1993). Cognitive gender differences: A developmental perspective. *Sex Roles, (29),* 91-112.

FENNEMA, E. L., & SHERMAN, J. (1978). Sex related differences in mathematic achievement and related factors: A further study. *Journal of Research in Mathematics Education, 9,* 189-203.

FLANAGAN, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

GRIEB, A. & EASLEY, J. (1984). A primary school impediment to mathematical equity: Case studies in rule-dependent socialization. In M. W. Steinkamp & M. L. Maehr (Eds.) *Advances in motivation and achievement: Women in science* (Vol. 2). Greenwich, CT: JAI Press.

GUTHRIE, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior, 76,* 1-15.

HANCOCK, T. E., HUBBARD, D. C., & THURMAN, R. T (1992, June). *The Discrepancy Construct in the Kulhavy/Stock Model of Instructional Feedback.* A paper presented at the Annual Convention of the American Psychological Society, San Diego.

HANCOCK, T. E., HUBBARD, D. C., & THURMAN, R. T (1992, August). *Modeling student performance and cognition using multiple CBT sources (for more intelligent feedback).* A paper presented at the Annual Convention of the American Psychological Association, Washington, D. C.

HANCOCK, T. E., STOCK, W. A., & KULHAVY, R. W. (1992). Predicting feedback effects from response-certitude estimates. *Bulletin of the Psychonomic Society, 30(2),* 173-176

HANCOCK, T. E., STOCK, W. A., KULHAVY, R. W., & SWINDELL, L. K. (in revision). Gender and developmental differences in the academic study behaviors of elementary school children.*Journal of Experimental Education.*

HANCOCK, T. E., THURMAN, R. A., & HUBBARD, D. C., (June, 1993). *Computer-based drill performance predicted by feedback processing: Explained by perceptual control theory.* A paper presented at the Annual Convention of the American Psychological Society, Chicago.

HANCOCK, T. E., THURMAN, R. A., & HUBBARD, D. C., (August, 1993). *Performance predicted by student processing models..* A paper presented at the Annual Convention of the American Psychological Association, Toronto.

HANCOCK, T. E., THURMAN, R. A., & HUBBARD, D. C. (in submission). An expanded model for the use of instructional feedback. *Contemporary Educational Psychology.*

HARPER, G., & KEMBER, D. (1989). Interpretation of factor analysis from the approaches to studying inventory. *British Journal of Educational Psychology, 59,* 66-74.

HARTER, S. (1977). The effects of social reinforcement and task difficulty on the pleasure derived by normal and retarded children from cognitive challenge and mastery. *Journal of Experimental Child Psychology, 17,* 476-494.

HYDE, J. S. (1981) How large are cognitive gender differences? A meta-analysis using w^2 . *American Psychologist, 36,* 892-901.

JOHNSON, W. B. & NORTON, W. A. (1992). Modeling student performance in diagnostic tasks: A decade of evolution. In J. W. Regian & V. J. Shute (Eds.), *Cognitive Approaches to Automated Instruction,* (pp 195-216). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

JONES, M. G., & WHEATLEY, J. (1990). Gender differences in teacher-student interactions in science classrooms. *Journal Differences in Science Teaching, 27(9), 861-874.*

KIMBALL, M. M. (1989). A new perspective on women's math achievement motives. *Psychological Bulletin, 105,* 198-214.

KULHAVY, R. W., & KARDASH, C. A. (1988). Studying study: An analysis of instructional encoding behavior. *Educational and Psychological Research, 8,* 25-37.

KULHAVY, R. W., & STOCK, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1,* 279-308.

KULHAVY, R. W., STOCK, W. A., HANCOCK, T. E., HAMMRICH, P., & SWINDELL, L. K. (1989). Written feedback: Response certitude and durability. *Contemporary Educational Psychology, 17,* 319-332.

KULHAVY, R. W. & WAGER, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. V. Dempsey & G. C. Sales (Eds.) *Interactive Instruction and Feedback.* Englewood Cliffs, New Jersey: Educational Technology Publications.

KYLLONEN, P.C., TIRRE , W. C., & CHRISTAL, R. E. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General, 120,* 57-79.

LEWIS, C. H., & ANDERSON, J. R. Interference with real world knowledge. *Cognitive psychology, 7,* 311-335.

LEINHARDT, G., SEEWALD, A. M. & ENGEL, M. (1979). Learning what's taught: Sex differences in instruction. *Journal of Educational Psychology, 71(4),* 432-439.

LICHT, B. G., LINDEN, T. A., BROWN, D. A. & SEXTON, M. A. (1984). Determinants of academic achievement: The interaction of children's achievement orientations with skill area. *Developmental Psychology, 20,* 628-636.

LOGAN, G. D. ,& STADLER, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 478-496.

MACCOBY, E. E. (1966). Sex differences in intellectual functioning. In E. E. Maccoby (Ed.), *The development of sex differences*. Stanford, CA: Stanford University Press.

MACCOBY E., & JACKLIN, C. (1974). Sex differences in intellectual functioning. *The Psychology of Sex Differences*. Stanford: Stanford University Press.

MARKEN, R. S. (1986). Perceptual organization of behavior: A hierarchical control model of coordinated action. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 267-276.

MARKEN, R. S. (1991). Degrees of freedom in behavior. *Psychological Science, 2*, 92-100.

MARTON, F., & SALJO, R. (1976). On qualitative differences in learning: I-outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

MAZZONI, G., & CORNOLDI , C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122*, 47-60.

MEECE, J. L., & ECCLES, J. S. (1993). Introduction: Recent trends in research on gender and education. *Educational Psychologist, 28*, 313-319.

MEYER, OSMAN, IRWIN & YANTIS, Meyer, D. E., Osman, A. M., Irwin, D. E., Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26*, 3-67.

MOELEY, B. E., HART, S. S., LEAL, L., SANTULLI, K. A., RAO, N., JOHNSON, T., & HAMILTON, L. B. (1992). The teacher's role in facilitating memory and study strategy development in the elementary school children. *Child Development, 62*, 653-672.

MORY, E. H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research & Development, 40(3)*, 5-20.

NELSON, T. O., & LEONESIO, R. J. (1988). allocation of self-paced study time and the "labor -in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676-686.

NICHOLS, J. G. (1984) Conceptions of ability and achievement motivation. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 1). New York: Academic Press.

NOBLE, D. D. (1991). *The classroom arsenal: Military research, information technology and public education*, (pp. 155-157). London: The Falmer Press.

NOLEN S. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction, 5*, 269-287.

PARK, O. C., & TENNYSON, R. D. (1983). Computer-based instructional systems for adaptive education: A review. *Contemporary Educational Review, 2*, 121-135.

PARSONS, J. E., KAEZALA, C. M., & MEECE, J. L. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development, 53*, 322-339.

PITTMAN, T., & D'AGOSTINO, P. (1989). Motivation and cognition: Control deprivation and the nature of subsequent information processing. *Journal of Experimental Social Psychology, 81*, 384-391.

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

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

POWERS, W. T. (1992). *Living control systems II*. Gravel Switch, Kentucky: The Control Systems Group.

REGIAN, V. J. & SHUTE, J.W. Automated instruction as an approach to individualization. In J. W. Regian & V. J. Shute (Eds.), *Cognitive Approaches to Automated Instruction*, (pp 1-13). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

ROHWER, W. D. (1984). An invitation to an educational psychology of studying. Educational *Psychologist, 19*, 1-14.

RUNKEL (1990) P. J. (1990). *Casting Nets and Testing Specimens: Two Grand Methods of Psychology*. New York: Praeger.

RYBCZYNSKI, M. (1992). Audience adaptation and persuasive strategies: A study of letters by sixth-grade students. *Journal of Research and Development in Education, 26*, 15-23.

SALES, G. C. (1993). Adapted and adaptive feedback in technology-based instruction. In J. V. Dempsey & G. C. Sales (Eds.) *Interactive instruction and feedback*. Englewood Cliffs, NJ: Educational Technology Publications.

SAS INSTITUTE, (1985). *SAS User's Guide: Statistics*. Cary, NC: SAS Institute Inc.

SCHMECK, R. R., GEISLER-BERNSTEIN, E., & CERCY, S. P. (1991). Self-concept and learning: the revised inventory of learning processes. *Educational Psychology*, 11(3/4), 343-362.

SHAFIR, U., & PASCUAL-LEONE, J. (1990). Postfailure reflectivity and spontaneous attention to errors. *Journal of Educational Psychology, 82*, 378-387.

SHUTZ, P. A. (1993). Additional influences on response certitude and feedback requests. *Contemporary Educational Psychology, 18*, 427-441.

SIMON, J. R., & CROFT, J. L. (1989). The effect of prediction accuracy on choice reaction time. *Memory & Cognition, 17*, 503-508.

SIMPSON, A. W. & ERIKSON, M. T. (1983). Teachers' verbal and nonverbal communication patterns as a function of teacher race, student gender, and student race. *American Educational Research Journal, 20*, 183-198.

SMITH, K. U., & SMITH, M. F. (1966). *Cybernetic principles of learning and educational design*. New York: Holt, Rinehart, & Winston.

SNOWMAN, J. (1986). Learning tactics and strategies. In G. D. Phye & T. Andre (Eds.), *Cognitive classroom learning: Understanding, thinking, and problem solving* (pp. 243-276). New York: Academic Press.

STEINBERG, E. R. (1991). *Computer-assisted instruction: A synthesis of theory, practice, and technology*, (page 187). Hillsdale, NJ: Lawrence Erlbaum.

STIPEK, D. J., & GRALINSKI, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology, 83(3)*, 361-371.

STOCK, W. A., KULHAVY, R. W., PRIDEMORE, D. R., & KRUG, D. (1992). Responding to feedback after multiple-choice answers: The influence of response confidence. *Quarterly Journal of Experimental Psychology, 45(a)*, 649-667.

SWINDELL, L. K. & WALLS, (1993). Response confidence and the delay-retention effect. *Contemporary Educational Psychology, 18*, 363-375.

TALYZINA, N. (1981). *The psychology of learning*. Moscow, Russia: Progress Publishers.

THOMAS, J. W., & ROHWER, W. D. (1986). Academic studying: The role of learning strategies. *Educational Psychologist, 21*, 1941.

THOMAS, J. W., & ROHWER, W. D. (1987). Grade-level and course-specific differences in academic studying: Summary. *Contemporary Educational Psychology, 12*, 381-385.

TRIGWELL, K., & PROSSER, M. (1991). Relating approaches to study and quality of learning outcomes at the course level. *British Journal of Educational Psychology, 61*, 265-275.

TYLER, L. E. (1965). *The psychology of human differences* (3rd edition). New York: Appleton.

VOLET, S. E., & CHALMERS, D. (1992). Investigation of qualitative differences in university students' learning goals, based on an unfolding model of stage development. *British Journal of Educational Psychology, 62*, 17-34.

WALKER, N. & CATRAMBONE, R. (1993). Aggregation bias and the use of regression in evaluating models of human performance. *Human Factors, 35(3)*, 397-411.

WEBB, J. M., STOCK, W. A., & MCCARTHY, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology. 19*, 251-265.

WEINSTEIN, C. E., & MAYER, R. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.) *Handbook of Research on Teaching*. New York MacMillan.

WINNE, P. H. (1989). Theories of instruction and of intelligence for designing artificially intelligent tutoring systems. *Educational Psychologist, 24(3)*, 229-259.

PRELIMINARY RESULTS OF THE NEUROPSYCHIATRICALLY ENHANCED FLIGHT SCREENING PROJECT

Alexis G. Hernandez
Associate Dean
Department of the Dean of Students

The University of Arizona
Old Main, Room 211
Tucson, Arizona 85721

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

and

Armstrong Laboratory

August 1994

PRELIMINARY RESULTS OF THE NEUROPSYCHIATRICALLY ENHANCED FLIGHT SCREENING PROJECT

Alexis G. Hernandez
Associate Dean
Department of the Dean of Students
The University of Arizona

Abstract

The United States Air Force has yet to find useful predictive measures of success as an aviator. Within the spectrum of human factors, personality presents one variable, and if properly measured, may have predictive validity in determining who becomes a successful military aviator. Personality factors may also be useful in the cockpit assignment process. This study represents the first phase of the neuropsychiatrically enhanced flight screening project, a longitudinal study searching for valid predictive measures of success as a military aviator; and providing baseline psychological information about these candidates should it be needed. Candidates entering the Air Force enhanced flight screening program were tested using the following psychological instruments: Revised NEO Personality Inventory, Multidimensional Aptitude Battery, and Personal Characteristics Inventory. Results are evaluated for their usefulness in describing current flight training candidates. Discussion focuses on how these tests may help in selecting the best qualified pilot candidates (select-in measures) or ensuring candidates meet minimum standards (select-out measures).

PRELIMINARY RESULTS OF THE NEUROPSYCHIATRICALLY ENHANCED FLIGHT SCREENING PROJECT

Alexis G. Hernandez

Introduction

If some predictive measure of success in becoming a mission-qualified pilot is possible, the United States Air Force (USAF) has not found it or has kept it hidden. The USAF has been training pilots for almost 50 years and has employed various screening criteria to determine who enters pilot training. Nevertheless, data on who becomes a successful military aviator has not yet been published. The screening methods have consisted of various combinations of academic performance, physical fitness, medical fitness, motor coordination tests, psychological tests, and a psychiatric interview (Baker, 1989; Carretta, 1989; Carretta & Ree, 1993a; Long & Varney, 1975). Although these techniques contributed to reduced attrition rates, from 75% to 30% (Long & Varney, 1975), further research may improve these figures. With decreasing budgets and increasing mission demands, the Department of Defense would welcome improved methods of selecting the best candidates for aviation training programs.

In aviation, human factors account for training attrition and many lost lives and flying resources. Foushee and Helmreich (1988) and Nance (1986), among others in civilian and military aviation, have recognized human factors as a significant contributor to, and cause of, aircraft mishaps. Pilot personality, as one such factor, has been studied from various angles and for various purposes, including determining what is the "right stuff," who has it, and who functions best within certain cockpits (Ashman & Telfer, 1983; Chidester, Helmreich, Gregorich, & Geis, 1991; Lardent, 1991; Novello & Youssef, 1974a, 1974b; Picano, 1991; Siem & Murray, 1994). Nevertheless, pilot personality has not been fully explored as an effective predictor of military aviation success.

Psychological factors have been given little consideration in the selection process. The adaptability interview for Air Force pilot candidates, formally known as the Adaptability Rating for Military Aviation (ARMA), has remained unchanged for over 30 years (Verdone, Sipes, & Miles, 1993). Coincidentally, the attrition rate in pilot training, ranging from 11% to 37% (Eisen, 1988; Knutsen, 1988), has also remained unchanged for many years (Long & Varney, 1975; Olea & Ree, 1993; Retzlaff & Gibertini, 1987; Walters, Miller, & Ree, 1993).

A variety of reasons account for the limited aviator psychometric norms. Researchers may report similar findings, as did Chidester et al. (1991), Picano (1991), and Retzlaff and Gibertini (1987), yet direct comparisons across studies are difficult and, at best, represent guesses. Samples often come from specialized populations such as astronaut candidates (Fine & Hartman, 1968; Santy, Holland, & Faulk, 1991), fighter pilots (Flynn, Sipes, Grosenbach, & Ellsworth, 1993), tanker, transport, and bomber pilots (Chidester et al., 1991), undergraduate pilot training candidates (Retzlaff & Gibertini, 1988; Siem, 1992), specialized or psychiatric evaluation referrals (King, 1994; Levy, Tolson, & Carlson, 1979), and mishap pilots (Lardent, 1991). Additionally, a variety of instruments have been utilized across studies: Millon Clinical Multiaxial Inventory (MCMI; King, 1994; Retzlaff & Gibertini, 1987), Occupational Personality Questionnaire (OPQ; Picano, 1991), Edwards Personal Preference Schedule (EPPS; Ashman & Telfer, 1983; Novello & Youssef, 1974a, 1974b), Personality Research Form (PRF; Goeters, Timmermann, & Maschke, 1993), 16 Personality Factors (16PF; Galloway, Ogle, & Malmstrom, 1991; Lardent, 1991), Basic Attributes Test (BAT; Carretta, 1989), Eysenck Personality Inventory (EPI; Jessup & Jessup, 1971), and Personal Characteristics Inventory (PCI; Chidester et al., 1991), among others. Some studies used no psychological tests and instead utilized semi-structured clinical interviews (Santy et al., 1991). Data on female aviators present the same mixed picture for the same reasons (Galloway et al., 1991; Jones, 1983; Novello & Youssef, 1974b), although fewer studies exist.

Regardless of the instruments used, however, researchers have concluded that distinct "pilot personality types" exist (Ashman & Telfer, 1983; Galloway et al., 1991; Novello & Youssef, 1974b; Picano, 1991; Retzlaff & Gibertini, 1987). Female aviators also present distinct personality types (Galloway et al., 1991; Novello & Youssef, 1974b). With such differences between pilots and the general population, Air Force psychologists use norms specifically established for pilots when evaluating aviators (Fine & Hartman, 1968; King, 1994; Retzlaff & Gibertini, 1988; Wheatley, 1979). Pilots often look pathological when compared with the general population, yet have "normal" profiles when compared to other pilots (King, 1994).

Neuropsychiatrically Enhanced Flight Screening Project

Previous studies of fledgling aviators used the completion of training as the criterion of success rather than actual mission readiness or performance in the aircraft cockpit (Carretta, 1989; Carretta & Ree, 1993b; Siem, 1992; Walters et al., 1993). These studies provide limited usefulness because not every pilot successfully transitions from undergraduate pilot training to advanced or upgrade training. Helmreich, Sawin, and Carsrud (1986) postulate such short-term research reflects a "honeymoon effect." Subjects purposely look their best and maintain high performance levels during short research periods. After this initial period, other factors such as motivation, personality characteristics, and perhaps luck probably become more important predictors of performance (Chidester et al., 1991). Long term studies can help determine what factors affect later performance and achievement. Although some studies present post-hoc data on mishap aviators, this information does not help in predicting success (e.g., Lardent, 1991).

The Neuropsychiatrically Enhanced Flight Screening (N-EFS) program hopes to address the above mentioned issues. This longitudinal project explores the use of the Revised NEO Personality Inventory (NEO-PI-R), the Multidimensional Aptitude Battery (MAB), the Personal Characteristics Inventory (PCI), and the CogScreen as tools to determine if personality traits and cognitive functioning can predict success in military aviation. The N-EFS program

defines success as becoming a mission-qualified pilot in the assigned cockpit within the standard time for completing such advanced training.

N-EFS hopes to help reduce attrition rates by providing information about who most likely completes military flight training and becomes a mission-qualified pilot. This information may then be used to determine who enters flight training. This methodology, known as a select-in process, promises greater usefulness because aviators and student aviators, as a group, appear to have little psychopathology (King, 1994). Hence, searching for and excluding those with psychopathology, known as a select-out methodology, offers limited assistance.

Method

Subjects

Students entering the Air Force enhanced flight screening program were offered an opportunity to take a battery of psychological tests. Of those 87 students, seven (7) females and 79 males (N=86) agreed to participate. ROTC cadets between their junior and senior year in college comprised the majority of participants. Others were commissioned second lieutenants and enlisted national guard personnel.

Instruments

Revised NEO Personality Inventory. The NEO-PI-R measures the five major domains of personality, with each domain further subdivided into six defining facets, for a total of 30 facet scales (see Table 1 for domain and facet labels). The 5 domain scales and the 30 facet scales of the NEO-PI-R allow for a comprehensive assessment of adult personality (Costa & McCrae, 1992). The materials consisted of the Professional Manual, Form S reusable item booklet, hand-scoring answer sheets, and profile forms. We administered this inventory in groups.

Multidimensional Aptitude Battery. The MAB, a wide-range assessment tool for adolescents and adults, provides a convenient, objectively-scored measure of general aptitude or intelligence (Jackson, 1984), designed for either group or individual administration. The battery consists of 5 verbal

subtests and 5 performance subtests, yielding ten subscale scores, a Verbal IQ, a Performance IQ, and a Full Scale IQ. Jackson (1984) reports a correlation coefficient of .91 between the MAB and the Wechsler Adult Intelligence Scale-Revised (WAIS-R) Full Scale score. We administered the MAB in groups.

Personal Characteristics Inventory. Gregorich et al. (1989) describe the PCI as useful in identifying subpopulations among pilots. The PCI measures the positive and negative components of two traits, instrumentality (goal orientation) and expressivity (interpersonal orientation). We administered the PCI in groups.

CogScreen. Horst and Kay (1991) developed this computerized, self-administered cognitive function screening test for the Federal Aviation Administration's pilot medical recertification procedure. The tests invoke a wide range of perceptual/cognitive processes thought necessary for skilled aviation performance, and reveal the presence of both specific cognitive deficits and generalized deficits. At the start of N-EFS, however, this battery was not available to the researchers. Therefore, the project directors decided to proceed without the CogScreen.

Design and Procedure

At the beginning of the U.S. Air Force initial flight screening and training program, students were offered an opportunity to complete a battery of psychological tests. A licensed psychologist thoroughly explained the informed consent forms and the reasons for the research, and answered participants' questions. All students were required to take the MAB as part of the medical baseline procedure. Volunteer subjects additionally completed the NEO-PI-R and PCI, and allowed their MAB to be used for research purposes. The testing was conducted during a 6 hour block of time, with rest breaks and lunch between tests to avoid problems with mental and physical fatigue. The licensed psychologist debriefed all the participants.

The results of these tests will be correlated with the criterion measure--becoming a mission-qualified pilot in the assigned cockpit within the

standard time for completing such upgrade training. The earliest this information will be available is mid- to late-1997. Results will eventually be analyzed to determine the predictive validity of these psychological tests. For the purposes of this study, the initial NEO-PI-R results are analyzed to determine personality traits and personality profiles for this group of candidates.

Results

The NEO-PI-R was scored for all participants to date (N=86). The means and standard deviations were calculated for the five domain and 30 facet T scores, and are presented in Table 1 for the total sample and divided by gender. Figure 1 profiles the average domain T scores and figure 2 presents the same information for the facet scores. Males and females present similar

---

Insert Table 1, Figure 1, and Figure 2 about here

---

but not identical profile patterns. As anticipated with the males, the Extraversion (60.62) and Conscientiousness (54.70) domains had relatively high scores, with Agreeableness (45.48) and Neuroticism (46.35) at the low end. Openness was in the middle with a T score of 49.23.

Similarly, females averaged highest on Extraversion (70.17), but their second highest score was Openness (56.17). Like the males, Neuroticism and Agreeableness were the low scores with averages of 46.14 and 44.50, respectively. The female profile had a significantly more extreme Extraversion score, although this may be an artifact of the small sample size.

Within the Extraversion domain, females averaged high scores in five of the six facets: excitement-seeking (68.14), assertiveness (63.00), gregariousness (62.14), positive emotions (62.14), and activity (61.57). Costa and McCrae (1992) describe those scoring high on these facets as dominant, forceful, socially ascendant, and group leaders. They lead fast-paced lives, enjoy the company of others, like bright colors and noisy environments, crave excitement and stimulation, are cheerful and optimistic,

and laugh easily and often. These characteristics would be expected, and perhaps required, of anyone entering military flight training. Nevertheless, on all facets of Extraversion, this sample of female student pilots averaged much higher than the standard population of females or their male student pilot counterparts.

The females also averaged higher scores on four of the six Openness facets: feelings (58.00), actions (61.00), ideas (55.71), and values (52.43). They scored lower on fantasy (48.29), and about the same on aesthetics (49.86). Costa and McCrae (1992) describe those with high scores in these facets as individuals who experience their emotions deeper and more intensely than others. They prefer novelty and variety, and are open-minded and willing to consider new and unconventional ideas, and reexamine social, political, and religious values. One might argue this result is expected as females are traditionally considered to be more aware of their feelings than men, especially the stereotypical male pilot populations.

The males scored only slightly higher than the females on all facets of Conscientiousness, although the slopes of the profiles parallel each other. Characteristics of high scores on these facets, which apply to both the males and females, include feeling well prepared to deal with life, keeping things in their proper places, and strictly adhering to one's ethical principles. These individuals have high aspiration levels, work hard to achieve their goals, and motivate themselves to get the job done. Lastly, they are cautious and deliberate.

General Discussion

In an overall sense, these results are consistent with earlier findings and the descriptions of other student pilots (Retzlaff & Gibertini, 1988) and female pilots (Galloway et al., 1991; Novello & Youssef, 1974b). These student pilots seek excitement and are conscientious, yet demonstrate little psychopathology. As the N-EFS project continues, we shall learn more about what role these personality factors play as a subset of these students becomes mission-qualified pilots.

Female students presented themselves in a manner similar to the male students, with some interesting differences. Surprisingly, the females scored much higher on the Extraversion domain, and somewhat higher on the Openness domain. These findings may have resulted from the small sample size and may not accurately reflect female student pilots. Hence, conjecture on the high Extraversion scores are premature. Will the differences hold up as sample size increases? The answer is possible within the next year as data collection continues.

By changing its policy and allowing females to fly combat aircraft, the military may have inadvertantly opened the door for reconsideration of the "right stuff." As currently accepted by the Air Force, flying with the "right stuff" means losing, on average, one aircraft per week, worldwide, outside of combat. Consequently, reconsidering this concept may not be a bad idea. Scoring the female participants' NEO-PI-R's using male norms would allow more direct comparisons of gender-based differences. This research, while enhancing our appreciation for what females bring to the aviation environment, may expand, or modify our understanding of the "right stuff." Scoring the males using female norms may help us understand how males compare to the female version of the "right stuff." Regardless of the findings, this line of investigation would further our understanding of the complexities and nuances of personality and the motivation of pilots in military aviation, and their relation to the "right stuff."

Data on NEO-PI-R profiles of student pilots has not yet been published. As an excellent parallel project, NEO-PI-R profiles of currently mission-qualified pilots could be collected. Comparisons between the profiles of pilots and the N-EFS students may help in resolving questions about changes, if any, facilitated by military flight training. Furthermore, a factor analysis is strongly recommended, specifically to make comparisons with other published studies; in particular Picano (1991) and Retzlaff and Gibertini (1987). A factor analysis would allow for further exploration of any differences between pilots of various types of aircraft.

As N-EFS progresses, comparisons between the academy, ROTC, and other students will be of interest. Should differences in rates of continued progress toward graduation (persistence), and graduation rates exist, the data may yield compelling support for the academy and/or provide evidence of the strength and efficacy of the ROTC programs. Ultimately, the data may help direct future commissioning and pilot selection procedures. This sub-project is recommended as the N-EFS project collects objective data on who has become a mission-qualified pilot.

N-EFS represents a long overdue and monumental undertaking with limitless possibilities for gaining psychological information on pilot profiles. Such projects require critical support from all levels of the Air Force command structure. Short-sighted decisions to cancel or reduce N-EFS support hurts military aviation, especially in light of the role human factors inevitably play in aircraft mishaps. The Air Force has an excellent opportunity to improve the selection, training process, and flying mission by improving rates of candidates who ultimately become mission-qualified pilots.

References

Ashman, A. & Telfer, R. (1983). Personality profiles of pilots. Aviation, Space, and Environmental Medicine, 54(10), 940-943.

Baker, L. E. (1989). The effect of higher education variables on cadet performance during 1987 light aircraft training (AD-A210 199, AU-ARI-88-9). Maxwell AFB, AL: Airpower Research Institute, Air University.

Fine, P. M., & Hartman, B. O. (1968). Psychiatric strengths and weaknesses of typical air force pilots (SAM-TR-68-121). Brooks AFB, TX: USAF School of Aerospace Medicine, Aerospace Medical Division.

Carretta, T. R. (1989). USAF pilot selection and classification systems. Aviation, Space, and Environmental Medicine, 60, 46-49.

Carretta. T. R., & Ree, M. J. (1993a). Basic attributes test: Psychometric equating of a computer-based test. International Journal of Aviation Psychology, 3(3), 189-201.

Carretta, T. R., & Ree, M. J. (1993b). Pilot candidate selection method (PCSM): What makes it work? (AL-TP-1992-0063). Brooks AFB, TX: Manpower and Personnel Research Division, Human Resources Directorate, Armstrong Laboratory.

Chidester, T. R., Helmreich, R. L., Gregorich, S. E., & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. The International Journal of Aviation Psychology, 1(1), 25-44.

Costa, P. T., Jr., & McCrae, R. R. (1992). NEO-PI-R professional manual: Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources, Inc.

Eisen, S., Jr. (1988). USAF flying screening: First step on the road to wings (AD-A192 613, 88-0850). Maxwell AFB, AL: Air Command and Staff College, Air University.

Flynn, C. F., Sipes, W. E., Grosenbach, M. J., & Ellsworth, J. (1993). Field test of a computer-driven tool to measure psychological characteristics of aircrew (AL-TR-1992-0171). Brooks AFB, TX: Aerospace Medicine Directorate, Clinical Sciences Division, Armstrong Laboratory.

Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. L. Weiner & D. C. Nagel (Eds.), Human factors in aviation (pp. 189-227). San Diego: Academic Press.

Galloway, M. L., Ogle, C. D., & Malmstrom, F. V. (1991). Comparing the Cattell 16PF profiles of male and female commercial airline pilots. Proceedings of the Human Factors Society 35th Annual Meeting.

Goeters, K., Timmermann, B., & Maschke, P. (1993). The construction of personality questionnaires for selection of aviation personnel. International Journal of Aviation Psychology, 3(2), 123-141.

Gregorich, S. E., Helmreich, R. L., Wilhelm, J. A., & Chidester, T. (1989). Personality based clusters as predictors of aviator attitudes and performance. In R. S. Jensen (Ed.), Proceedings of the symposium on aviation psychology (pp. 686-691). Columbus: Ohio State University.

Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986). The honeymoon effect in job performance: Delayed predictive power of achievement motivation. Journal of Applied Psychology, 71, 1085-1088.

Horst, R. L., & Kay, G. G. (1991). CogScreen user's manual (FAA/933-015-90). Washington, D.C.: Georgetown School of Medicine.

Jackson, D. N. (1984). Multidimensional aptitude battery manual. London, Ontario, Canada: Research Psychologists Press, Inc.

Jessup G., & Jessup, H. (1971). Validity of the Eysenck Personality Inventory in pilot selection. Occupational Psychology, 45, 111-123.

Jones, D. R. (1983). Psychiatric assessment of female fliers at the U.S. Air Force School of Aerospace Medicine (USAFSAM). Aviation, Space, and Environmental Medicine, 54, 929-931.

King, R. E. (1994). Assessing aviators for personality pathology with the Millon Clinical Multiaxial Inventory (MCMI). Aviation, Space, and Environmental Medicine, 65, 227-231.

Knutsen, D. W. (1988). Proposed improvements to the USAF flight screening program (AD-A194 506, 88-1485). Maxwell AFB, AL: Air Command and Staff College, Air University.

Lardent, C. L. (1991). Pilots who crash: Personality constructs underlying accident prone behavior of fighter pilots. Multivariate Experimental Clinical Research, 10(1), 1-25.

Levy, R. A., Tolson, D. B., & Carlson, E. H. (1979). Student pilots referred to the neuropsychiatry branch, USAFSAM 1968-78: Implications for selection. Aviation, Space, and Environmental Medicine, 50, 1173-1175.

Long, G. E., & Varney, N. C. (1975). Automated pilot aptitude measurement system (AFHRL-TR-75-58, AD-A018 151). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Nance, J. J. (1986). Blind Trust. New York: William Morrow & Company.

Novello, J. R., & Youssef, Z. I. (1974a). Psycho-social studies in general aviation: I. Personality profile of male pilots. Aerospace Medicine, 45(2), 185-188.

Novello, J. R., & Youssef, Z. I. (1974b). Psycho-social studies in general aviation: II. Personality profile of female pilots. Aerospace Medicine, 45(6), 630-633.

Olea, M. M., & Ree, M. J. (1993). Predicting aircrew training performance with psychometric g (AL-TP-1993-0011). Brooks AFB, TX: Human Resources Directorate, Technical Training Research Division, Armstrong Laboratory.

Picano, J. J. (1991). Personality types among experienced military pilots. Aviation, Space, and Environmental Medicine, 62, 517-520.

Retzlaff, P. D., & Gibertini, M. (1987). Air Force pilot personality: Hard data on the "right stuff." Multivariate Behavioral Research, 22, 383-399.

Retzlaff, P. D., & Gibertini, M. (1988). Objective psychological testing of the U.S. Air Force officers in pilot training. Aviation, Space, and Environmental Medicine, 59, 661-663.

Santy, P. A. (1994). Choosing the right stuff: The psychological selection astronauts and cosmonauts. Westport, CT: Praeger Publishers.

Santy, P. A., Holland, A. W., & Faulk, D. M. (1991). Psychiatric diagnoses in a group of astronaut applicants. Aviation, Space, and Environmental Medicine, 62, 969-973.

Siem, F. M. (1992). Predictive validity of an automated personality inventory for Air Force pilot selection. International Journal of Aviation Psychology, 2(4), 261-270.

Siem, F. M., & Murray, M. W. (1994). Personality factors affecting pilot combat performance: A preliminary investigation. Aviation, Space, and Environmental Medicine, 65(5, Suppl.), A45-48.

Verdone, R. D., Sipes, W., & Miles, R. (1993). Current trends in the usage of the adaptability rating for military aviation (ARMA) among USAF flight surgeons. Aviation, Space, and Environmental Medicine, 64, 1086-1093.

Walters, L. C., Miller, M. R., & Ree, J. M. (1993). Structured interviews for pilot selection: No incremental validity. The International Journal of Aviation Psychology, 3(1), 25-28.

Wheatley, R. (1979). Descriptive data for three selected USAF groups psychometrically evaluated at USAF School of Aerospace Medicine. Unpublished manuscript available through Brooks AFB, TX: Aerospace Medicine Directorate, Clinical Sciences Division, Neuropsychiatry Branch.

Table 1

Average Domain and Facet T-scores Males, Females, and Both

Student Aviator		Males	n = 79		Both	n = 86		Females	n = 7
Domain		Mean	SD		Mean	SD		Mean	SD
Neuroticism	N	46.35	8.66		46.34	8.68		46.14	9.62
Extraversion	E	60.62	9.95		61.33	9.97		70.17	6.31
Openness	O	50.82	12.42		51.20	12.24		56.17	10.38
Agreeableness	A	45.48	10.36		45.33	10.63		44.50	15.35
Conscientiousness	T	54.70	8.95		54.29	8.88		49.33	7.69
Facets		Mean	SD		Mean	SD		Mean	SD
Anxiety	N1	49.25	9.09		49.28	9.17		49.57	10.88
Angry Hostility	N2	49.04	10.09		49.08	10.10		49.57	11.01
Depression	N3	46.82	8.23		46.57	8.17		43.71	7.34
Self Consciousness	N4	47.82	9.67		47.49	9.58		43.71	8.28
Impulsiveness	N5	47.73	9.32		47.95	9.08		50.43	5.62
Vulnerability	N6	42.29	8.21		42.43	7.99		44.00	5.03
Warmth	E1	51.82	10.58		52.19	10.51		56.29	9.52
Gregariousness	E2	55.56	10.03		56.09	10.13		62.14	9.99
Assertiveness	E3	59.18	9.47		59.49	9.26		63.00	5.63
Activity	E4	60.01	8.89		60.14	8.64		61.57	5.19
Excitement-seeking	E5	62.19	7.08		62.67	7.41		68.14	9.42
Positive Emotions	E6	54.61	10.99		55.22	11.06		62.14	10.09
Fantasy	O1	54.23	10.28		53.74	10.34		48.29	10.23
Aesthetics	O2	49.35	11.00		49.40	10.90		49.86	10.45
Feelings	O3	52.06	12.90		52.55	12.56		58.00	5.77
Actions	O4	52.63	10.81		53.31	10.91		61.00	9.56
Ideas	O5	52.99	10.90		53.21	10.63		55.71	7.06
Values	O6	46.37	9.76		46.86	10.04		52.43	12.31
Trust	A1	48.53	11.17		48.48	11.16		47.86	11.95
Straightforwardness	A2	47.89	9.86		47.73	9.71		46.00	8.19
Altruism	A3	51.04	9.40		51.06	9.89		51.29	15.42
Compliance	A4	44.20	10.18		43.78	10.27		39.00	10.83
Modesty	A5	46.81	10.37		46.80	10.43		46.71	11.94
Tender-Mindedness	A6	45.37	10.91		45.52	10.99		47.29	12.59
Competence	C1	56.19	7.98		55.72	8.18		50.43	9.13
Order	C2	51.72	10.57		51.44	10.36		48.29	7.39
Dutifulness	C3	52.49	9.01		52.27	8.89		49.71	7.48
Achievement-Striving	C4	57.52	13.22		57.33	12.80		55.14	6.39
Self-Discipline	C5	52.63	8.59		52.36	8.40		49.29	5.22
Deliberation	C6	48.67	9.95		48.56	9.83		47.29	8.96

Figure 1

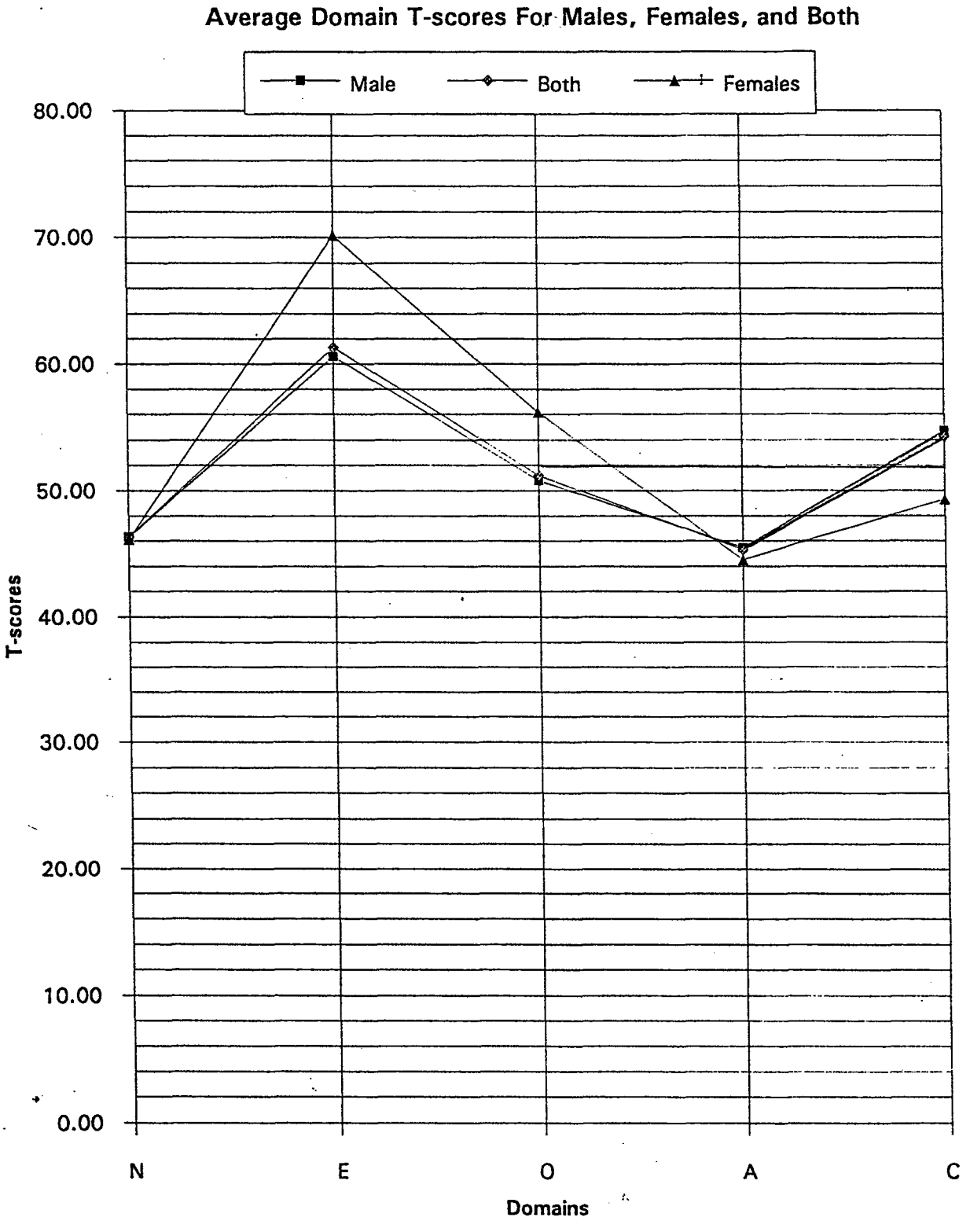
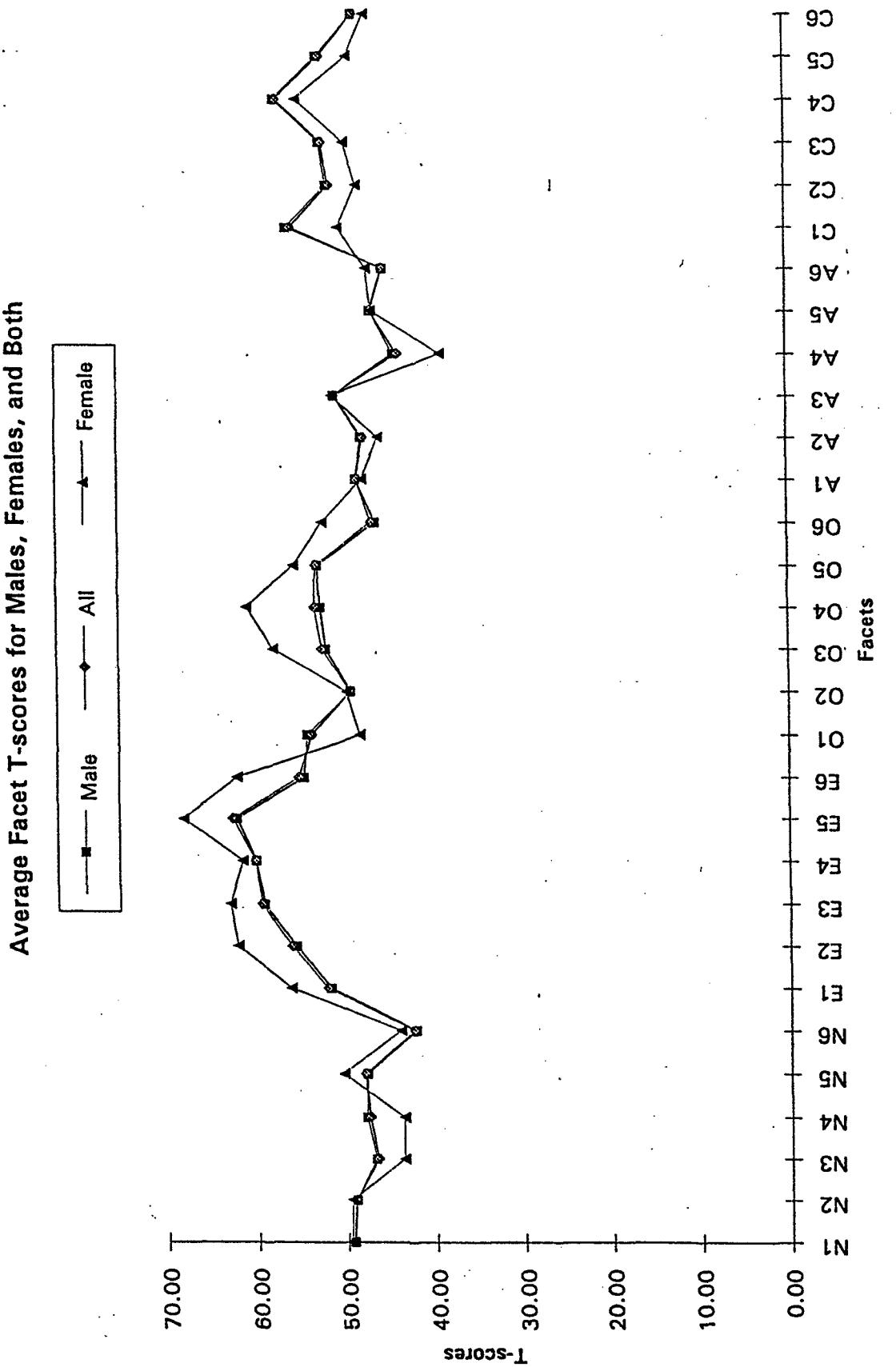

Average Domain T-scores For Males, Females, and Both
Male
Both
Females
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
T-scores
N
E
O
A
C
Domains

Figure 2


Average Facet T-scores for Males, Females, and Both
Male
All
Female
T-scores
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Facets
N1
N2
N3
N4
N5
N6
E1
E2
E3
E4
E5
E6
O1
O2
O3
O4
O5
O6
A1
A2
A3
A4
A5
A6
C1
C2
C3
C4
C5
C6

DISTRIBUTED SENSORY PROCESSING DURING GRADED HEMODYNAMIC LOADS

Arthur Koblasz
Associate Professor
School of Electrical & Computer Engineering

Georgia Institute of Technology
777 Atlantic Drive
Atlanta, GA 30332-0250

Final Report for:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

October 1994

DISTRIBUTED SENSORY PROCESSING DURING GRADED HEMODYNAMIC LOADS

Arthur Koblasz
Associate Professor
School of Electrical & Computer Engineering
Georgia Institute of Technology

Abstract

A new protocol was studied which will use cortical evoked responses to characterize how visual and auditory sensitivities are altered by physical workloads which produce hemodynamic stress. The cortical evoked responses will be measured during periods of fatigue. The subjects will be given simultaneous vestibular, visual and auditory stimuli. The stimuli will be random binary modulations which will allow the simultaneous responses to be identified for each stimulus.

DISTRIBUTED SENSORY PROCESSING DURING GRADED HEMODYNAMIC LOADS

Arthur Koblasz

Background

Tactical missions often impose high physical workloads on pilots, which produce measurable levels of emotional and hemodynamic stress. Previous research has demonstrated that physical exertion can affect a subject's ability to respond to simultaneous sensory ques. However, these studies have not normalized the physical workload to be relative to each subject's work capacity--percent physical work capacity (%PWC). Furthermore, none of the previous studies were able to simultaneously characterize vestibular, visual and auditory responses, which is possible using random binary stimuli.

Methodology

Subjects with different PWC's will be asked to perform at specified levels of their individual PWC. At each PWC, we will measure the subject's cortical evoked responses to randomly modulated vestibular, visual and auditory stimuli. The vestibular stimulus will be created using random binary inputs to a servo controlled turntable. The angular accelerations will shift between two levels with random duration of each on/off period. The visual stimulus will be generated using a standard checkerboard pattern with random intervals between the pattern switches. The auditory stimulus will be created using a pure tone at 1000 Hz with random duration of each on/off period. The 3 random binary stimuli will be presented simultaneously for a duration of 5 minutes at each of the %PWC's described below.

Each subject will be asked to reach a constant %PWC by hand pedaling a stationary

ergometer while seated on the rotating turntable. The %PWC will be measured from the subject's expired gases using a metabolic cart.

During each experiment, the cortical potentials will be measured continuously at 8 different locations on the scalp. Each of the cortical response signals will be crosscorrelated with the 3 separate binary stimuli to identify the average step-on response for each stimulus. The same 3-input protocol and analytical methods will be repeated at 50%, 70% and 90% PWC for a sustained period of 5 minutes at each level. Changes in the average cortical responses will be identified for each %PWC.

During each experiment, horizontal eye movements also will be measured continuously using EOG electrodes. The continuous EOG signal will be crosscorrelated with the vestibular (random binary) stimulus to identify the average VOR response to a step-on of angular acceleration. Changes in the average VOR will be identified for each %PWC.

Engelken, Edward J. and Stevens, Kennith W., "Saccadic Eye Movements in Response to Visual, Auditory, and Bisensory Stimuli", *Aviation, Space, and Environmental Medicine*, August 1989, pp. 762.

Engleken, Edward J., Stevens, Kennith W. and Enderle, John D., "Relationships between Manual Reaction Time and Saccade Latency in Response to Visual and Auditory Stimuli", *Aviation, Space and Environmental Medicine*, April 1991, pp. 315.

Basil-Neto, Joaquim P., Pascual-Leone, Alvaro, Valls-Sole, Josep, Cammarota, Angel, Cohen, Leonardo G. and Hallett, Mark, "Postexercise Depression of Motor Evoked Potentials: A Measure of Central Nervous System Fatigue", *Experimental Brain Research*, 1993.

Watten, Reidulf G. and Lie, Ivar, "Time Factors in VDT-Induced Myopia and Visual Fatigue: An Experimental Study", *J. Human Ergol.*, 1992, pp. 13-20.

Saito, Susumu, "Does Fatigue Exist in a Quantitative Measurement of Eye Movements?", *Ergonomics*, 1992, Vol. 35, Nos. 5/6, pp. 607-615.

Takeda, Masaharu, "Basic Study on the Measuring of Fatigue by the Minimum Audible Pressure", *Ergonomics*, 1992, Vol. 35, Nos. 5/6, pp. 637-646.

Rosenberg, Martin E., Barson, John V. and Pollard, Andrew J., "Auditory Localizing Acuity During Acute Exposure to Altitude", *Aviation, Space and Environmental Medicine*, July 1994, pp. 649.

Morris, Alison M., So, Yuen, Lee, Kathryn A., Lash, Andrea A. and Becker, Charles E., "The P300 Event-Related Potential-The Effects of Sleep Deprivation", *JOM*, Vol. 34, No. 12, December 1992, pp. 1143.

Prasher, Deepak, Smith, Alison and Findley, Leslie, "Sensory and Cognitive Event-Related Potentials in Myalgic Encephalomyelitis", *Journal of Neurology, Neurosurgery, and Psychiatry*, 1990, Vol. 53, pp. 247-253.

Murata, Katsuyuki, Araki, Shunichi, Kawakami, Norito, Saito, Yuko and Hino, Eiko, "Central Nervous System Effects and Visual Fatigue in VDT Workers", *Occupational-Environmental Health*, 1991, Vol. 63.

APPLICATION OF THE MT3D SOLUTE TRANSPORT MODEL TO THE MADE-2 SITE

Dr. Manfred Koch
Research Scientist

Geophysical Fluid Dynamics Institute
Florida State University
Tallahassee, FL 32306

Final report for:
Summer Faculty Research Program
Armstrong Laboratory

October, 1994

APPLICATION OF THE MT3D SOLUTE TRANSPORT MODEL TO THE MADE-2 SITE

Dr. Manfred Koch
Research Scientist

Geophysical Fluid Dynamics Institute
Florida State University

ABSTRACT

The applicability of the public domain 3D solute transport code MT3D to model the migration of a conservative tracer and of reactive hydrocarbons at the MADE-2 site is investigated. To the author's knowledge this is the first time that MT3D has been applied to such a complex and heterogeneous groundwater aquifer system as is the MADE-2 site. The results of the study are very encouraging. In spite of the extreme numerical difficulties encountered initially with the code, which required several modifications and adaptations, MT3D has been able to mimic qualitatively the essential features of the tritium tracer plume and of the possibly biodegrading hydrocarbon p-xylene. Visual differences between the modelled and the observed tritium plume are noticeable after about 1 year after the start of the experiment and are to be attributed mainly to insufficient calibration of the head and flow fields computed by means of the MODFLOW model. Although more quantitative calibrations, using moment analyses of the observed and modelled plumes, and better numerical verification of the MT3D code must be left to future studies, it is the author's belief that this model has enormous potentials for unravelling some of the most important physico-chemical transport and fate mechanisms that have been proposed for the MADE-2 site by scientists at the Armstrong Laboratory. However, it also became clear during the course of the study that 3D flow and transport modelling requires extremely powerful computational platforms and, even more importantly, efficient 3D Graphics visualization software. The positive experience obtained from the latter serves also as a reminder to modellers who are still ingrained in a 2D thinking that their attempts to use classical 2D-contouring software will be inefficient, excruciatingly time-consuming, and still incapable of providing a full perspective of the complex 3D spatial model.

APPLICATION OF THE MT3D SOLUTE TRANSPORT MODEL TO THE MADE-2 SITE

Dr. Manfred Koch
Research Scientist

Geophysical Fluid Dynamics Institute
Florida State University

INTRODUCTION

Objective of the MADE-2 (Macrodispersion Experiment 2) experiment at the Columbus, Mississippi, Air Force Base was to investigate the the feasibility of remediation by natural attenuation of an alluvial aquifer contaminated by aromatic hydrocarbons The results of this experiment, which went for about 450 days, do in fact show strong experimental evidence for the above conjecture (Stauffer et al., 1993). In contrast to the observed concentrations of the 'conservative' tracer tritium, those of the four aromatic compounds used in the experiment, benzene, naphthalene, p-xylene and o-dichlorobenzene, decreased significantly over the time-period of the experiment. These reductions were explained by Stauffer et al. (1993) by means of first-order-kinetics degradation processes. The important implications of the MADE-2 results are that if such natural attenuation is a regularly occurring phenomenon in contaminated aquifers, it would alleviate the problem of active remediation of the latter. The cost-savings to the society in the whole would be huge.

In order to get a more detailed understanding of the most important physico-chemical transport and fate mechanisms acting in the MADE-2 experiment, the use of a numerical solute transport model is required. This has been the objective of the present author's 2-months stay at the Armstrong Lab. After some initial surveying it was agreed to employ the MT3D solute transport model (Zheng, 1990), since is one of the rare 3D models available in the public domain. This was done in spite of the negative results of Gray and Rucker (1994) who could not report any positive MT3D model run during their 3-months stay at the Armstrong Lab. The reasons for their failure are numerous, none of the least might be that several programming bugs were found in the code by this author, after a relatively lengthy test-period. Others are to be attributed to a non-appropriate setting of relevant input parameters for the code and finally also in systematic insufficiencies of the code itself, that had to corrected before successful runs were obtained. Although the final results of the present modelling efforts are overall positive, it is the conclusion of this author, that practical use of MT3D in its present form requires a deep understanding of numeric modelling as a whole.

The MTD3 solute transport model (Zheng, 1990, 1993) belongs to the so-called class of mixed Eulerian-Lagrangian methods that have become popular in the last decade for the numerical solution of advection-dominated transport problems in porous media. As reviewed in detail in Koch (1994), contrary to conventional Eulerian methods which solve the governing parabolic/hyperbolic transport equation on a fixed mesh (either by finite differences (FD) as in the 3D HT3D model, or by finite elements (FE) as in the 2D FEMWASTE model) and which are often plagued with numerical problems of numerical dispersion and/or oscillations, mixed Eulerian-Lagrangian are often devoid of these numerical hassles. These methods are essentially based on operator splitting techniques that split the transport equation into the hyperbolic (advection) portion that is solved by a Method of Characteristics (MoC) or by particle tracking and into the parabolic (diffusion) portion that is solved by a classical FD of FE method. One of the first techniques of this kind is the classical 2D MoC method of Konikow and Bredehoeft (1977). Here a set of particles is injected in a cell and tracked forward along the characteristical line. Because the advective part of the problem is solved very accurately, the method is essentially oscillation- and dispersion-free, i.e. it is very suitable for sharp front problems. Applications of a MoC to the simulations of density-dependent solute transport and of finger instabilities that may arise in such situations have been presented in Koch and Zhang (1992) and Koch (1992; 1993).

The drawback of the MoC-method is that, in order to reduce mass-balance errors, many particles have to be tracked which makes the method computationally very expensive. In addition, since the time integration of the transport equation is usually done by an explicit method (which eliminates the need for the solution of large matrix systems), stability considerations put upper limits on the allowable timesteps. In fact, the so-called Fourier-number stability criterium states that the permissible timestep Δt is inversely proportional to the hydrodynamical dispersion coefficient D. As it will be demonstrated, this behavior can have disastrous implications on the practical application of the MTD3 model, which is mainly only a 3D-extension of the classical MoC code.

A computationally more efficient MoC technique has been proposed recently in form of the Modified Method of Characteristics (MMoC) (cf. Douglas and Russel, 1982; Russel and Wheeler, 1983) where only one particle is placed at a nodal point and traced backwards for one timestep to the foot of the characteristic. This is the second model option included in the MT3D code. The disadvantage with the MMoC approach is that it appears to be too dispersive in strongly advection-dominated transport problems, if proper care with the interpolation to the fixed Eulerian nodes is not taken. Moreover, MMoC is known to be plagued with large mass errors, though the recent development of a flux-based MMoC (Roache, 1992) appears to be more mass-conservative. In fact, as will be demonstrated, such a behavior of the MMoC version of the MT3D code is also observed during the numerical simulations.

With the MoC and the MMoC model options of MT3D having each complementary advantages and limitations, a third option available in the model is the Hybrid Method of Characteristics (HMoC) which

employs the MoC in the vicinity of sharp concentration gradients and the MMoC elsewhere. This HMoC option has mostly been used in the numerical simulations. On the other hand, the final option available in the code, the classical Upstream Finite Difference (UFD) has not been investigated at all in the present study.

INITIAL TESTING AND DEBUGGING OF THE MT3D-CODE

First impressions with the MT3D code, as delivered on the PC-diskettes by the International Groundwater Center, were negative. The code would run for some of the test examples, but crashed for others, as well, as for the MADE-2 example that had been set up earlier by Rucker and Gray (1994). Numerous debugging tests were then performed with the source code to find the location where the executable crashed. Eventually, after several days of debugging and analysis of the code, a programming bug in the form of a 'dimension error' was found in the advection package of the code which essentially acted like a virus, i.e. showed up in some runs and not in others (a typical feature for this kind of programming error).

The MT3D code then ran successfully variations of the model set-up, as had been prepared by Rucker and Gray (1994) on the SUN Sparc-10 workstation. However, the simulation would never go beyond 140 days or so of real time, after it had run for six to ten hours hours on the Sparc-10. This was exactly what had been reported by Gray and Rucker (1994).

From there on the tedious process of analyzing and dissecting the MT3D was undertaken. It was clear that this task could not been carried out by running jobs on the Sparc-10 for several hours and watching idly what would happen eventually. First of all, the longitudinal dispersivity α_L was reduced from 10 to 1 which, because of the Fourier-stability criterium, increased the time-steps taken by a factor of ten, with a corresponding reduction in total run-time. The number of particles injected was reduced, to further speed up the code (it turns out that this does not result in too much of a savings, but may deteriorate the solution significantly). The code would still crash at around day 140. It turned out eventually that the job was stalling during the last days of the simulation, by using conspicuously small transport time steps, running 'dead', until it eventually crashed after umpteen of hours on the Sparc-10.

From there on it was decided to use the powerful DEC-Alpha workstations at the Geophysical Fluid Dynamics Institute, which are about 6-10 times faster than a SUN Sparc-10 workstation, as the major platform for the testing of the MT3D model. With the Alpha's it was possible to attain the 'disastrous' day 140 in about 10 minutes of run time-time, giving the possibility of interactive and efficient debugging tests.

What was found after numerous tests was that the artificially small time steps were created from the stability criterium that is set up for the sink sink/source term in the code (see. Eqs. 4.29-.4.31 of Zheng, 1990), namely, at the locations of the surficial recharge source, i.e. at the upper (the water-table) boundary of the model. The way the MODFLOW model has prepared by Gray and Rucker (1994) to simulate aquifer recharge through precipitation, is to use negative recharge (mimicking evaporation) up to about day 130 and positive recharge (for the wet season) thereafter. It is shortly after that time that this positive recharge leads

to rewetting of previously dry cells at the water-table surface. It is then when the recharge source (which is passed from the MODFLOW program to the MT3D program) acts in the disastrous ways described. The interesting thing was that the cell responsible for the problem did not even contain any contaminant concentration for that time, and computing a (wrong) upper timestep (s. Eq. 4.32) does not even make sense. Through some minor changes in the program the problem was fixed and small timesteps as a result of the sink/source term never occurred henceforward.

NUMERICAL MODELLING OF THE MADE-2 EXPERIMENT

As listed in the following table, most of the MT3D models that were run simulate the transport and fate of the tritium plume of the MADE-2 experiment. Only one model experiment has been carried out up-to-date to model the (possibly!) degradable p-xylene plume. In the later case a degradation-rate constant, as reported by Stauffer et al. (1993) has been used in reaction package of the MT3D program.

Models:

M1:	Tritium (no decay; i.e. conservative), α=1: OK
M2:	Tritium (with decay), α_L=1: OK, CPU-time= 2:20 hours
M3:	Tritium α_L=5: Runs until 350 days
M4:	Tritium, less particles: α_L=5: Runs until 170 days
M5:	Tritium, less particles, α_L=0.5, α_V/α_L =0.1: OK, CPU-time= 1 hour
M6:	Like M5, use MMOC alone: OK, CPU-time= 50 minutes
M7:	Like M5, with Euler particle tracker, MMOC+MOC: OK
M8:	p-Xylene with degradation, α_L=1: OK, CPU-time= 2 hours

All of the above models were run with some standard hydraulic and transport parameters, as they have been set up by Gray and Rucker (1994). Moreover, most of the MODFLOW-flow calibration of these authors were not changed which results in the similar differences between observed and modelled flow field (and ergo for the plume migrations) in the northern section of the model, as also reported by these authors. In the above table α_L and α_V denote the longitudinal and vertical dispersivities, respectively.

As is noticeable from the table, not all models ran successfully. So far, model-runs with longitudinal dispersivities larger than 5 (as induced from the MADE-1 experiment (Rehfeldt et al., 1990)) could not be

Figure 1. 3D-visualization of modelled tritium plume

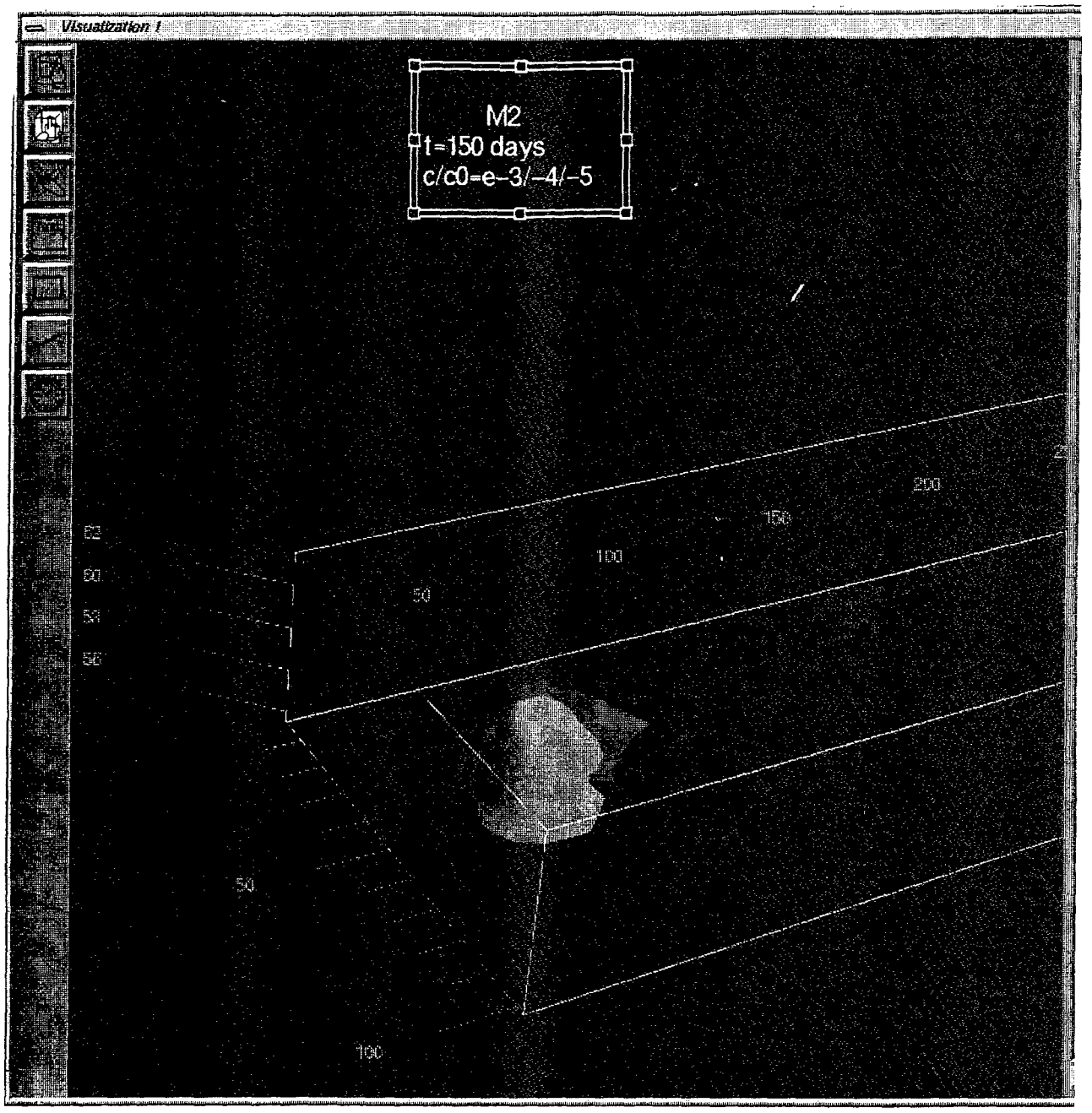


Figure 2. 3D-visualization of modelled tritium plume

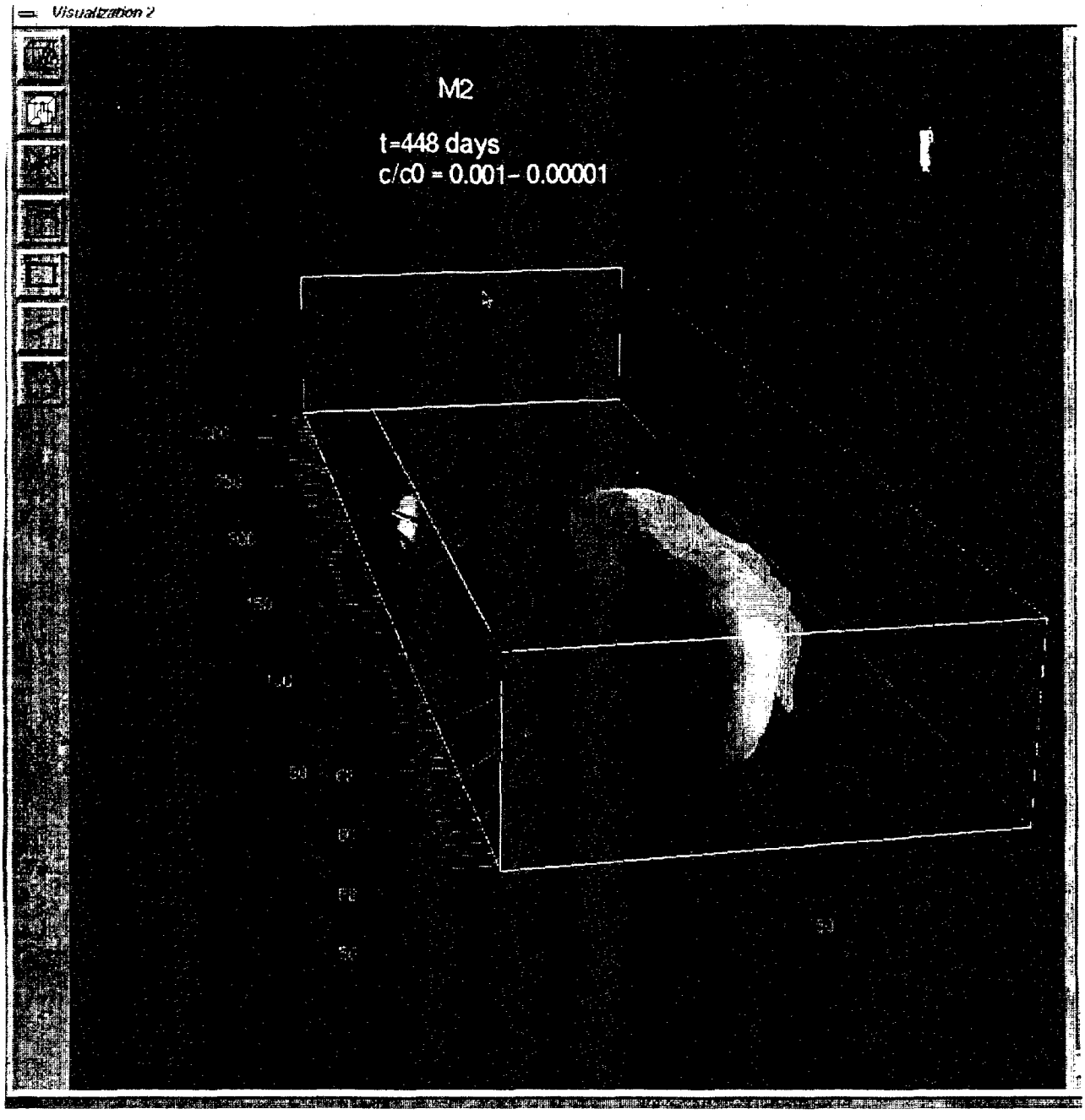


Figure 3. 3D-visualization of modelled tritium plume

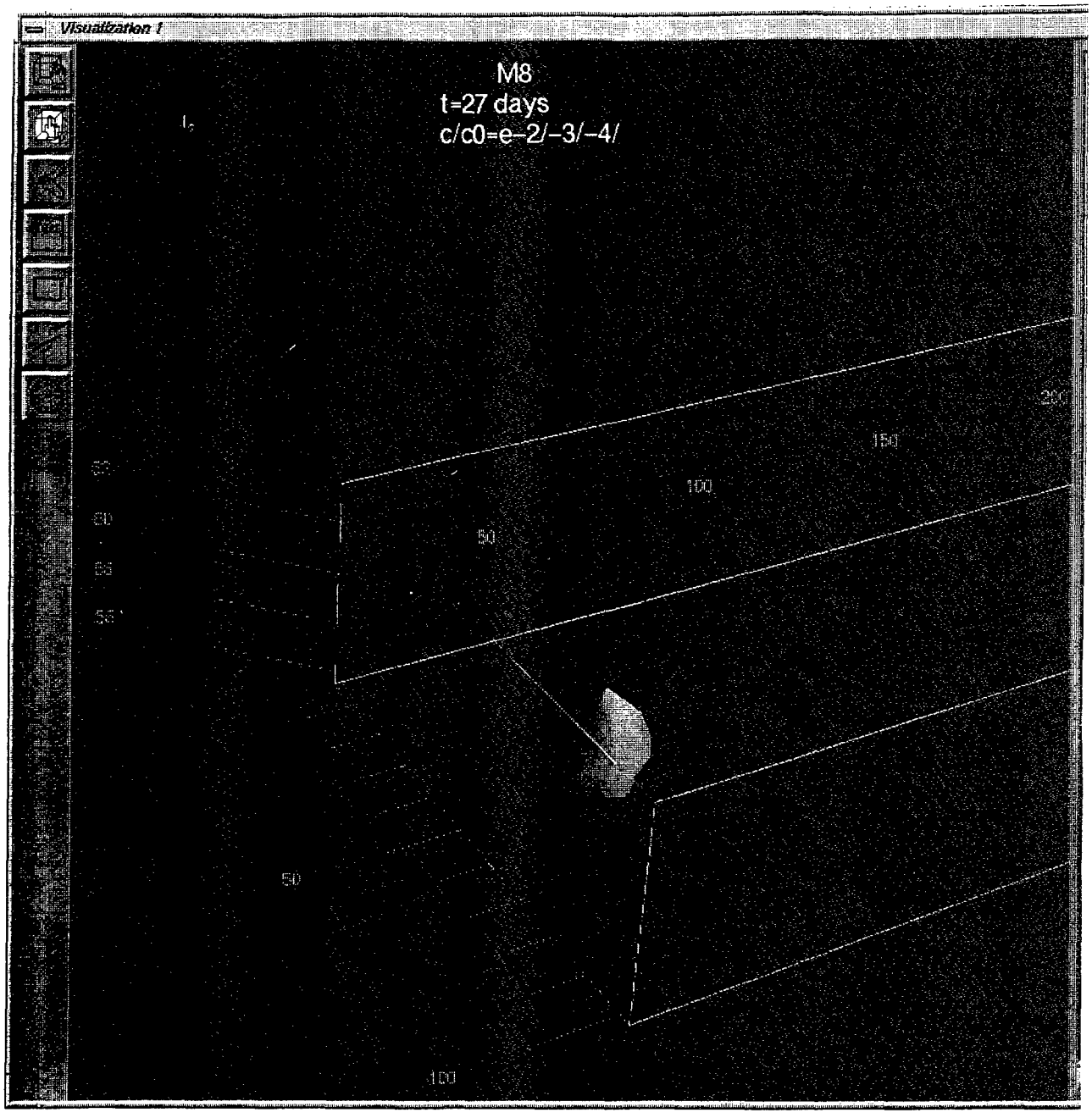


Figure 4. 3D-visualization of modelled p-xylene plume

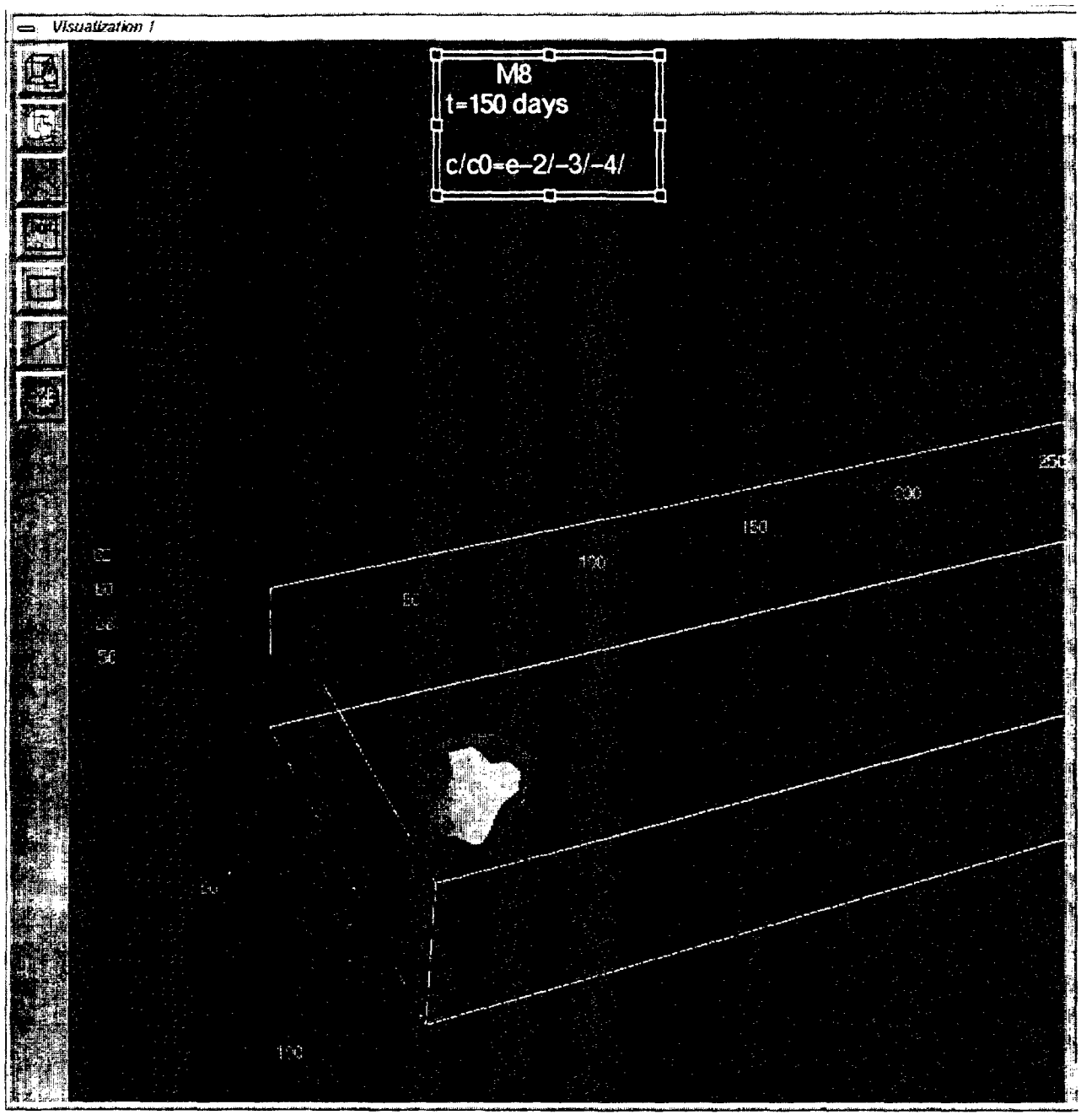


Figure 5. 3D-visualization of modelled p-xylene plume

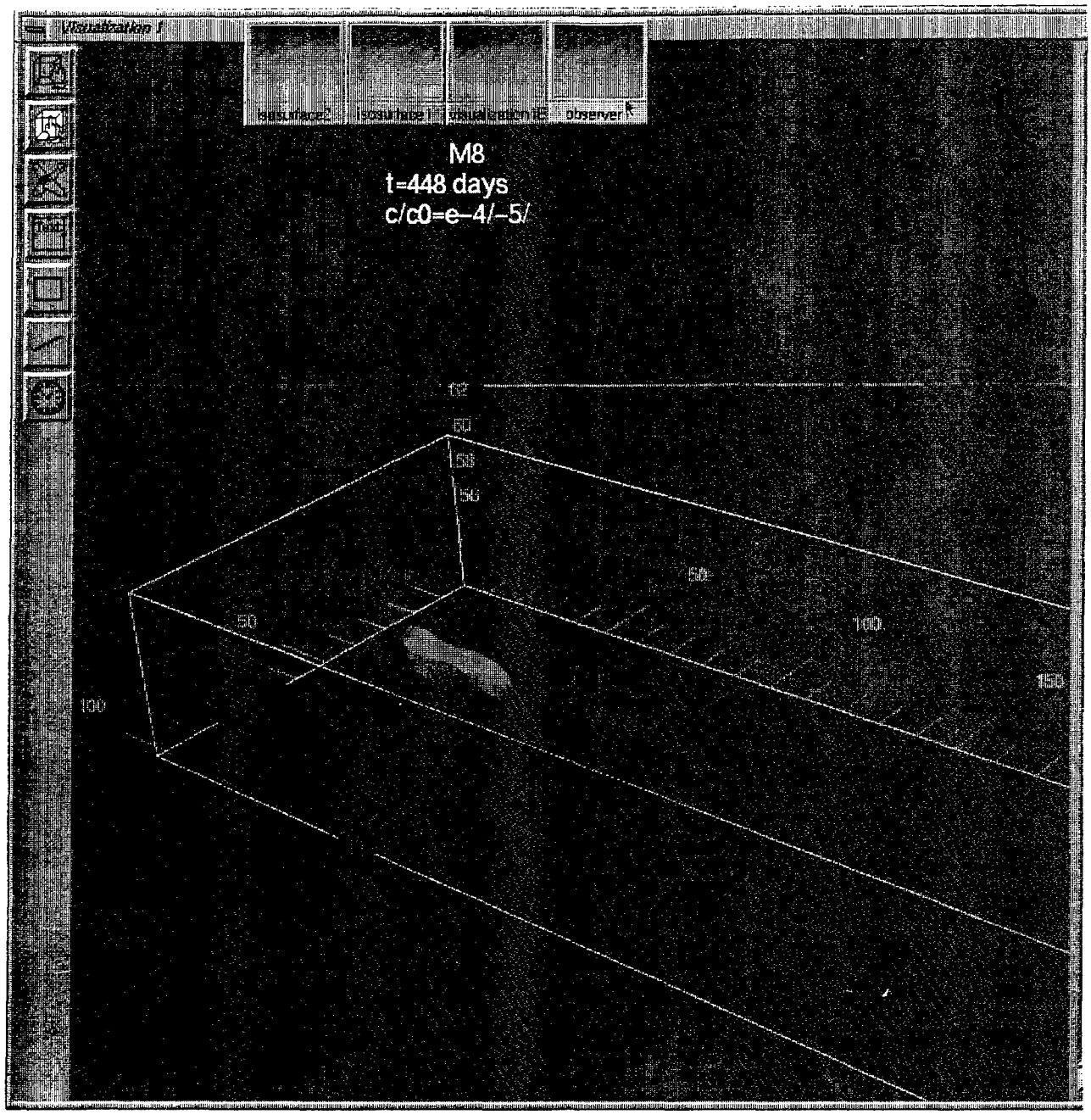


Figure 6. 3D-visualization of modelled p-xylene plume

completed successfully and eventually crashed. This appears to be related to the difficulty of the MoC-option mentioned earlier to model very dispersive transport processes. More research is needed.

The cumulative mass-balance errors for the successful models range between 3 and to 5%, when using the HMoC option of the code. However, they are of O(20%) when the somewhat faster Modified Method of Characteristics (MMoC) is used. The plumes appear also to be more dispersed which is a typical characteristic of this technique.

Only the two important models M2 for the tritium plume and model M8 for the p-Xylene plume have been visualized up-to-date by means of the *SCIAN* graphics package. Three colour snapshots for the tritium model M2 are shown in Figs. 1, 2 and 3 and for the p-Xylene plume in Figs. 4 and 5 and 6 after different times since the beginning of the experiment. Each of the plume pictures shows three isosurfaces with the relative concentrations c/c_0, as indicated. The outer two isosurfaces (with the lower c/c_0) are transparent and so allow to look into the interior of the plume. Unfortunately the hardcopies of these screen-pictures are no match to the original ones on the Silicon Graphics terminal and details of the plumes pattern are harder to detect.

The model plumes appear to mimic the observed plumes reasonably well, particularly for the tritium plume. Similar to the experimental plume, the model tritium plume does not move much in the first 150 days of the experiment, since is is stuck in the zone of low conductivity. After 448 days of simulation time it extends to about 250 m from the source. However, it begins to veer into the western direction, in contrast to the to the real plume which tends eastwards. This is in agreement with the modelled flow field of Gray and Rucker (1994) and shows that the MODFLOW calibration needs to be redone.

The degradation of the p-xylene plume is visible from Figs. 4-6. However noticeable decay starts only after about 100 days. After the end of the experiment at 448 days, only a small core of p-xylene is left in the vicinity of the source. This is pretty much in agreement with the observations.

CONCLUSIONS AND RECOMMENDATIONS

The applicability of the public domain 3D solute transport code MT3D to model the migration of a conservative tracer and of reactive hydrocarbons at the MADE-2 site has been shown. However, the success of the simulations came about only after several days of a detailed 'numerical dissection' of the code, fixing a bug to get it running in the first place, understand its numerical limitations under particular transport situations, and make minor modifications in the code to prevent it from failing under various circumstances. In spite of these changes, the code is by no means 'fool-proof' at the present time and will need numerous improvements to make it palatable to the ordinary groundwater modeller.

The success of the modelling efforts has also to be attributed to the outstanding computational and graphics facilities available at Florida State University which were used intermittently during the official contract period and more so afterwards. DEC-Alpha workstations which are many times faster than the

completed successfully and eventually crashed. This appears to be related to the difficulty of the MoC-option mentioned earlier to model very dispersive transport processes. More research is needed.

The cumulative mass-balance errors for the successful models range between 3 and to 5%, when using the HMoC option of the code. However, they are of O(20%) when the somewhat faster Modified Method of Characteristics (MMoC) is used. The plumes appear also to be more dispersed which is a typical characteristic of this technique.

Only the two important models M2 for the tritium plume and model M8 for the p-Xylene plume have been visualized up-to-date by means of the *SCIAN* graphics package. Three colour snapshots for the tritium model M2 are shown in Figs. 1, 2 and 3 and for the p-Xylene plume in Figs. 4 and 5 and 6 after different times since the beginning of the experiment. Each of the plume pictures shows three isosurfaces with the relative concentrations c/c_0, as indicated. The outer two isosurfaces (with the lower c/c_0) are transparent and so allow to look into the interior of the plume. Unfortunately the hardcopies of these screen-pictures are no match to the original ones on the Silicon Graphics terminal and details of the plumes pattern are harder to detect.

The model plumes appear to mimic the observed plumes reasonably well, particularly for the tritium plume. Similar to the experimental plume, the model tritium plume does not move much in the first 150 days of the experiment, since is is stuck in the zone of low conductivity. After 448 days of simulation time it extends to about 250 m from the source. However, it begins to veer into the western direction, in contrast to the to the real plume which tends eastwards. This is in agreement with the modelled flow field of Gray and Rucker (1994) and shows that the MODFLOW calibration needs to be redone.

The degradation of the p-xylene plume is visible from Figs. 4-6. However noticeable decay starts only after about 100 days. After the end of the experiment at 448 days, only a small core of p-xylene is left in the vicinity of the source. This is pretty much in agreement with the observations.

CONCLUSIONS AND RECOMMENDATIONS

The applicability of the public domain 3D solute transport code MT3D to model the migration of a conservative tracer and of reactive hydrocarbons at the MADE-2 site has been shown. However, the success of the simulations came about only after several days of a detailed 'numerical dissection' of the code, fixing a bug to get it running in the first place, understand its numerical limitations under particular transport situations, and make minor modifications in the code to prevent it from failing under various circumstances. In spite of these changes, the code is by no means 'fool-proof' at the present time and will need numerous improvements to make it palatable to the ordinary groundwater modeller.

The success of the modelling efforts has also to be attributed to the outstanding computational and graphics facilities available at Florida State University which were used intermittently during the official contract period and more so afterwards. DEC-Alpha workstations which are many times faster than the

5. An experimental (the chemical data) and numerical (the code) investigation of the minimal reliable threshhold background concentration for the various MADE-2 plumes. This is a major general issue in any practical site-modelling effort. Because of physical and numerical dispersion (the latter should be theoretically small in the MT3D code, since the MOC technique is used) and of rounding errors in the code itself, small numbers for the concentrations are still produced in regions beyond the plume edges. The question then is whether such numbers are numerically significant and whether they are supported by measured concentrations. As for the tritium plume, there is a natural background activity of about 10^{-4} which sets a lower threshhold concentration that should not be undercut by the model concentrations.

With regard to a more detailed numerical analysis of the MT3D code, this should include:

6. Investigation of the numerical problems that occur with the MT3D code when large values for the hydrodynamic dispersivity are used. This is contrary to ordinary finite difference methods (such as HTD3) which performs particularly well in the presence of large physical dispersion. So far it appears that this problem is related to the Fourier stability criteria which imposes timesteps too small for large dispersivities, similar to the USGS MOC code (Konikow and Bredehoeft, 1977).

7. Set-up of criteria for proper selection of the number of particles, of the time-integration method, and of acceptable minimal timesteps in order to prevent the code from stalling in time and eventually failing, as has happened occasionally.

8. Modification of the MT3D code to account for effective rewetting of the surficial water table cells and to account for the proper mass-balance of the contaminants in such cells. Although the BCF2-rewetting option is used in the present MODFLOW version, dried-out cells are not properly reactivated in the MT3D code if the cell is rewetted during a later stage of the simulations. Such a situation occurs exactly in the MADE-2 models, when after about 130 days of dry season and negative recharge (mimicking evaporation), increased precipitation induces positive recharge.

Once the issues above are satisfactorily addressed, a more quantitative calibration of the MADE-2 transport models that goes beyond the 3D graphical inspection can be endeavoured by comparison of the various theoretical moments of the observed and modelled plumes. It is the author's belief that the MT3D code has eventually the potentials for unravelling some of the most important physico-chemical transport and fate mechanisms that have been proposed for the MADE-2 site so far.

Acknowledgements

I'm particularly indebted to Dr. Tom Stauffer for making my stay at the Armstrong Lab. possible and the numerous suggestions concerning the subject. Many thanks are due to Dr. Howard Mayfield with his patient support in my dealing with the new computer hardware and software at the Lab. Thanks are

also due to Mr. Chris Antworth for help with the PC. Finally I thank Prof. Dr. 'Mano' for interesting discussions concerning the numerical analysis of the model.

REFERENCES

Douglas, J., Jr. and T.F. Russel, Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element of finite difference procedures, SIAM J. Numer. Anal., 19, 871–885, 1982.

Gray, D.D. and D.F. Rucker, Improved numerical modeling of groundwater flow and transport at the MADE-2 site, Final report for the Summer Faculty Research Program, Armstrong Laboratory, October, 1994

Koch, M., Numerical simulation of finger instabilities in density and viscosity dependent miscible solute transport, in: Proceedings of the 'IX International Conference on Computational Methods in Water Resources', Denver, CO, June 9–12, 1992, Vol. 2, Mathematical Modeling in Water Resources, Russel, T.F., R.E. Ewing, C.A. Brebia, W.A. Gray and G.F. Pinder (eds.), pp. 155–162, Computational Mechanics Publications, Southampton, UK, 1992.

Koch, M., Modeling the dynamics of finger instabilities in porous media: Evidence for fractal and nonlinear system behavior, in: Advances in Hydroscience and -Engineering, Volume I, (Wang, Sam S.Y., ed.), pp. 1763–1774, Center for Computational Hydrosciene and Engineering, The University of Mississippi, University, MS, 1993.

Koch, M., A mesh-adaptive collocation technique for the simulation of advection-dominated single-and multi-phase transport phenomena in porous media, In: Modeling, Mesh Generation and Adaptive Numerical Methods for Partial Differential Equations, IMA volume series in Mathematics and its Applications, J.E. Flaherty, I. Babuska and J.E. Oliger, (eds.) Springer, New York, 1994 (In Press).

Koch, M. and G. Zhang, Numerical simulation of the migration of density dependent contaminant plumes, Ground Water, 5, 731–742, 1992.

Konikow, L.F. and J.D. Bredehoeft, Computer model for two-dimensional solute transport and dispersion in groundwater, US. Geol. Surv. Tech. Water Resour. Invest. Rep. 77-19, 1977.

McDonald, M.G. and Harbaugh, A.W., "A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model ", Open-File Report 83-875, U.S. Geological Survey, 1984.

Rehfeldt, K.R., J.M. Boggs and L.W. Gelhar, Field study of dispersion in a heterogeneous aquifers. 3. Geostatistical analysis of hydraulic conductivity, Water Resour. Res., 28, 3309–3324, 1992.

Roache, P.J., Validation exercises of one-dimensional flux-based modified methods of characteristics, in: Proceedings of the 'IX International Conference on Computational Methods in Water Resources', Denver, CO, June 9–12, 1992, Vol. 1, Numerical Modeling in Water Resources, Russel, T.F., R.E. Ewing,

C.A. Brebia, W.A. Gray and G.F. Pinder (eds.), pp. 69–76, Computational Mechanics Publications, Southampton, UK, 1992.

Russel, T.F. and M.F. Wheeler, Finite element and finite difference methods for continuous flows in porous media, in Ewing, R. E. (ed.), The mathematics of reservoir simulation, Society for Industrial and Applied Mathematics, Philadelphia, PY, pp. 35–106, 1983.

Stauffer, T.B., C.P. Antworth, R.G. Young, W.G. MacIntyre, J.M. Boggs, and L.M. Beard, 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-1933-0007, 1993

Zheng, C., MT3D, A modular three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems, prepared for the U.S. EPA Roberts S. Kerr Environmental Research Laboratory, Ada, Oklahoma, by S.S. Papadopoulus & Associates Inc., Rockville, Maryland, 1990.

Zheng, C., Extension of the method of characteristics for simulation of solute transport in three dimensions, Ground Water, 31, 456–465, 1993.