AD-A278 693

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1993

SUMMER RESEARCH PROGRAM FINAL REPORTS

VOLUME 2

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES
5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Col. Hal Rhoades

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 1993

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution /
Availability Codes
Dist Avail and/or Special
A-1

Best
Available
Copy

MEMORANDUM FOR DTIC (Acquisition)
(Attn: Pat Mauby)

SUBJECT: Distribution of USAF (AFOSR Summer Research Program (Air Force Laboratories) and Universal Energy Systems, Inc., and the Research Initiation Program

FROM: AFOSR/XPT
Joan M. Boggs
110 Duncan Avenue, Suite B115
Bolling AFB DC 20332-0001

1. All of the books forwarded to DTIC on the subjects above should be considered Approved for Public Release, distribution is unlimited (Distribution Statement A).

2. Thank you for processing the attached information.

Joan M. Boggs

JOAN M. BOGGS
Chief, Technical Information Division

1993 SRP Vol 1, 2, 3, 4, 5A, 5B, 6, 7, 8, 9
(10 Books)

Master Index for Faculty Members

Abbott, Ben
Research, MS
Box 1649 Station B
Vanderbilt University
Nashville, TN 37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 6- 1

Abrate, Serge
Assistant Professor, PhD
Mechanical & Aerospace En
University of Missouri - Rolla
Rolla, MO 65401-0249

Field: Aeronautical Engineering
Laboratory: WL/FI

Vol-Page No: 5-15

Almallahi, Hussein
Instructor, MS
P.O. Box 308
Prairie View A&M University
Prairie View, TX 77446-0000

Field: Electrical Engineering
Laboratory: AL/HR

Vol-Page No: 2-25

Anderson, James
Associate Professor, PhD
Chemistry
University of Georgia
Athens, GA 30602-2556

Field: Analytical Chemistry
Laboratory: AL/EQ

Vol-Page No: 2-18

Anderson, Richard
Professor, PhD
Physics
University of Missouri, Rolla
Rolla, MO 65401-0000

Field: Physics
Laboratory: PL/LI

Vol-Page No: 3- 7

Ashrafiuon, Hashem
Assistant Professor, PhD
Mechanical Engineering
Villanova University
Villanova, PA 19085-0000

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2- 6

Backs, Richard
Assistant Professor, PhD
Dept. of Psychology
Wright State University
Dayton, OH 45435-0001

Field: Experimental Psychology
Laboratory: AL/CF

Vol-Page No: 2- 7

Baginski, Thomas
Assoc Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-40

SFRP Participant Data

Baker, Suzanne
Assistant Professor, PhD
Dept. of Psychology
James Madison University
Harrisonburg, VA 22807-0000

Field:
Laboratory: AL/OE

Vol-Page No: 2-36

Baker, Albert
Assistant Professor, PhD

University of Cincinnati
, - 0

Field: Electrical Engineering
Laboratory: WL/MT

Vol-Page No: 5-53

Balakrishnan, Sivasubramanya
Associate Professor, PhD

University of Missouri, Rolla
, - 0

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-41

Bannister, William
Professor, PhD

Univ of Mass.-Lowell
Lowell, MA 1854-0000

Field: Organic Chemistry
Laboratory: WL/FI

Vol-Page No: 5-16

Barnard, Kenneth
Assistant Professor, PhD

Memphis State University
, - 0

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 1

Bayard, Jean-Pierre
Associate Professor, PhD
6000 J Street
California State Univ-Sacramen
Sacramento, CA 95819-6019

Field: Electrical/Electronic Eng
Laboratory: RL/ER

Vol-Page No: 4- 7

Beardsley, Larry
Research Professor, MS

Athens State College
, - 0

Field: Mathematics
Laboratory: WL/MN

Vol-Page No: 5-42

Beecken, Brian
Assoicate Professor, PhD
3900 Bethel Dr.
Bethel College
St. Paul, MN 55112-0000

Field: Dept. of Physics
Laboratory: PL/VT

Vol-Page No: 3-23

SFRP Participant Data

Bellem, Raymond
Dept, CHM. EE cs, PhD
3200 Willow Creek Road
Embry-Riddle Aeronautical Univ
Prescott, AZ 86301-0000

Field: Dept. of Computer Science
Laboratory: PL/VT

Vol-Page No: 3-24

Bellem, Raymond
Dept, CHM. EE cs, PhD
3200 Willow Creek Road
Embry-Riddle Aeronautical Univ
Prescott, AZ 86301-0000

Field: Dept. of Computer Science
Laboratory: /

Vol-Page No: 0- 0

Bhuyan, Jay
Assistant Professor, PhD
Dept. ofComputer Science
Tuskegee University
Tuskegee, AL 36088-0000

Field: Computer Science
Laboratory: PL/WS

Vol-Page No: 3-33

Biegl, Csaba
Assistant Professor, PhD
Box 1649 Station B
Vanderbilt University
Nashville, TN 37235-0000

Field: Electrical Engineering
Laboratory: AEDC/

Vol-Page No: 6- 2

Biggs, Albert
Professor, PhD
Electrical Engineering
Univ. of Alabama, Huntsville
Huntsville, AL 35899-0000

Field:
Laboratory: PL/WS

Vol-Page No: 3-34

Blystone, Robert
Professor, PhD
Trinity University
715 Stadium Drive
San Antonio, TX 78212-7200

Field: Dept of Biology
Laboratory: AL/OE

Vol-Page No: 2-37

Branting, Luther
Assistant Professor, PhD
PO Box 3682
University of Wyoming
Laramie, WY 82071-0000

Field: Dept of Computer Science
Laboratory: AL/HR

Vol-Page No: 2-26

Bryant, Barrett
Associate Professor, PhD
115A Campbell Hall
University of Alabama, Birming
Birmingham, AL 35294-1170

Field: Computer Science
Laboratory: RL/C3

Vol-Page No: 4- 1

Callens, Jr., Eugene
Association Professor, PhD
Industrial
Louisiana Technical University
Ruston, LA 71270-0000

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-43

Cannon, Scott
Associate Professor, PhD
Computer Science
Utah State University
Logan, UT 84322-0000

Field: Computer Science/Biophys.
Laboratory: PL/VT

Vol-Page No: 3-25

Carlisle, Gene
Professor, PhD
Dept. of Physics
West Texas State University
Canyon, TX 79016-0000

Field: Killgore Research Center
Laboratory: PL/LI

Vol-Page No: 3- 8

Catalano, George
Associate Professor, PhD
Mechanical Engineering
United States Military Academy
West Point, NY 10996-1792

Field: Department of Civil &
Laboratory: AEDC/

Vol-Page No: 6- 3

Chang, Ching
Associate Professor, PhD
Euclid Ave at E. 24th St
Cleveland State University
Cleveland, OH 44115-0000

Field: Dept. of Mathematics
Laboratory: WL/FI

Vol-Page No: 5-17

Chattopadhyay, Somnath
Assistant Professor, PhD

University of Vermont
Burlington, VT 5405-0156

Field: Mechanical Engineering
Laboratory: PL/RK

Vol-Page No: 3-14

Chen, C. L. Philip
Assistant Professor, PhD
Computer Science Engineer
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/ML

Vol-Page No: 5-26

Choate, David
Assoc Professor, PhD
Dept. of Mathematics
Transylvania University
Lexington, KY 40505-0000

Field: Mathematics
Laboratory: PL/LI

Vol-Page No: 3- 9

SFRP Participant Data

Chubb, Gerald
Assistant Professor, PhD
164 W. 19th Ave.
Ohio State University
Columbus, OH 43210-0000

Field: Dept. of Aviation
Laboratory: AL/HR

Vol-Page No: 2-27

Chuong, Cheng-Jen
Associtae Professor, PhD
501 W. 1st Street
University of Texas, Arlington
Arlington, TX 76019-0000

Field: Biomedical Engineering
Laboratory: AL/CF

Vol-Page No: 2- 8

Citera, Maryalice
Assistant Professor, PhD
Department of Psychology
Wright State University
Dayton, OH 4-5435

Field: Industrial Psychology
Laboratory: AL/CF

Vol-Page No: 2- 9

Collard, Jr., Sneed
Professor, PhD
Ecology & Evolutionary Bi
University of West Florida
Pensacola, FL 32514-0000

Field: Biology
Laboratory: AL/EQ

Vol-Page No: 2-19

Collier, Geoffrey
Assistant Professor, PhD
300 College St., NE
South Carolina State College
Orangeburg,, SC 29117-0000

Field: Dept of Psychology
Laboratory: AL/CF

Vol-Page No: 2-10

Cone, Milton
Assistat Professor, PhD
3200 Willow Creek Road
Embry-Riddel Aeronautical Univ
Prescott, AZ 86301-3720

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 2

Cundari, Thomas
Assistant Professor, PhD
Jim Smith Building
Memphis State University
Memphis, TN 38152-0000

Field: Department of Chemistry
Laboratory: PL/RK

Vol-Page No: 3-15

D'Agostino, Alfred
Assistant Professor, PhD
4202 E Fowler Ave/SCA-240
University of South Florida
Tampa, FL 33620-5250

Field: Dept of Chemistry
Laboratory: WL/ML

Vol-Page No: 5-27

SFRP Participant Data

Das, Asesh
Assistant Professor, PhD
Research Center
West Virginia University
Morgantown, WV 26505-0000

Field: Concurrent Engineering
Laboratory: AL/HR

Vol-Page No: 2-28

DeLyser, Ronald
Assistant Professor, PhD
2390 S. York Street
University of Denver
Denver, CO 80208-0177

Field: Electrical Engineering
Laboratory: PL/WS

Vol-Page No: 3-35

DelVecchio, Vito
Professor, PhD
Biology
University of Scranton
Scranton, PA 18510-4625

Field: Biochemical Genetics
Laboratory: AL/AO

Vol-Page No: 2- 1

Dey, Pradip
Associate Professor, PhD

Hampton University
, - 0

Field: Computer Science
Laboratory: RL/IR

Vol-Page No: 4-16

Ding, Zhi
Assistant Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-44

Doherty, John
Assistant Professor, PhD
201 Coover Hall
Iowa State University
Ames, IA 50011-1045

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-21

Dolson, David
Assistant Professor, PhD

Wright State University
, - 0

Field: Chemistry
Laboratory: WL/PO

Vol-Page No: 5-56

Dominic, Vincent
Assustant professor, MS
300 College Park
University of Dayton
Dayton, OH 45469-0227

Field: Electro Optics Program
Laboratory: WL/ML

Vol-Page No: 5-28

SFRP Participant Data

Donkor, Eric
Assistant Professor, PhD
Engineering
University of Connecticut
Stroes, CT 6269-1133

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-22

Driscoll, James
Associate Professor, PhD
3004 FXB Bldg 2118
University of Michigan
Ann Arbor, MI 48109-0000

Field: Aerospace Engineering
Laboratory: WL/PO

Vol-Page No: 5-57

Duncan, Bradley
Assistant Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-0226

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 3

Ehrhart, Lee
Instructor, MS
Communications & Intellig
George Mason University
Fairfax, VA 22015-1520

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 2

Ewert, Daniel
Assistant Professor, PhD
Electrical Engineering
North Dakota State University
Fargo, IN 58105-0000

Field: Physiology
Laboratory: AL/AO

Vol-Page No: 2- 2

Ewing, Mark
Associate Professor, PhD
2004 Learned Hall
University of Kansas
Lawrence, KS 66045-2969

Field: Engineering Mechanics
Laboratory: PL/SX

Vol-Page No: 3-22

Foo, Simon
Assistant Professor, PhD
College of Engineering
Florida State University
Tallahessee, FL 32306-0000

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-45

Frantziskonis, George
Assistant Professor, PhD
Dept of Civil Engrng/Mech
University of Arizona
Tuson, AZ 85721-1334

Field: College of Engrng/Mines
Laboratory: WL/ML

Vol-Page No: 5-29

SFRP Participant Data

Frenzel III, James
Assistant Professor, PhD
Dept of Electrical Engnr
University of Idaho
Moscow, ID 83844-1023

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 4

Fried, Joel
Professor, PhD
Chemical Engineering
University of Cincinnati
Cincinnati, OH 45221-0171

Field: Polymer Science
Laboratory: WL/PO

Vol-Page No: 5-58

Friedman, Jeffrey
Assistant Professor, PhD
Physics
University of Puerto Rico
Mayaguez, PR 681-0000

Field: Physics/Astrophysics
Laboratory: PL/GP

Vol-Page No: 3- 1

Fuller, Daniel
Dept. Chairman, PhD
Chemistry & Physics
Nicholls State University
Thibodaux, LA 70310-0000

Field: Chemistry
Laboratory: PL/RK

Vol-Page No: 3-16

Gao, Zhanjun
Assistant Professor, PhD
203 W. Old Main, Box 5725
Clarkson University
Potsdam, NY 13699-5725

Field: Mechanical/Aeronautical E
Laboratory: WL/ML

Vol-Page No: 5-30

Gavankar, Prasad
Asst Professor, PhD
Campus Box 191
Texas A&I University
Kingsville, TX 78363-0000

Field: Mech & Indust Engineering
Laboratory: WL/MT

Vol-Page No: 5-54

Gebert, Glenn
Assistant Professor, PhD
Mechanical
Utah State University
Logan, UT 84339-0000

Field: Aerospace Engineering
Laboratory: WL/MN

Vol-Page No: 5-46

Gerdom, Larry
Professor, PhD
Natural Science
Mobile College
Mobil, AL 36663-0220

Field: Chemistry
Laboratory: AL/EQ

Vol-Page No: 2-20

SFRP Participant Data

Ghajar, Afshin
Professor, PhD
Mech. & Aerospace Enginee
Oklahoma State University
Stillwater, OK 74078-0533

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-59

Gopalan, Kaliappan
Associate Professor, PhD
Dept of Engineering
Purdue University, Calumet
Hammond, IN 46323-0000

Field:
Laboratory: AL/CF

Vol-Page No: 2-11

Gould, Richard
Assistant Professor, PhD
Mechanical & Aerospace En
N.Carolina State University
Raleigh, NC 27695-7910

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-60

Gowda, Raghava
Assistant Professor, PhD
Dept of Computer Science
University of Dayton
Dayton, OH 45469-2160

Field: Computer Information Sys.
Laboratory: WL/AA

Vol-Page No: 5- 5

Graetz, Kenneth
Assistant Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-1430

Field: Department of Psychology
Laboratory: AL/HR

Vol-Page No: 2-29

Gray, Donald
Associate Professor, PhD
PO Box 6101
West Virginia Unicersity
Morgantown, WV 20506-6101

Field: Dept of Civil Engineering
Laboratory: AL/EQ

Vol-Page No: 2-21

Green, Bobby
Assistant Professor, MS
Box 43107
Texas Tech University
Lubbock, TX 79409-3107

Field: Electrical Engineering
Laboratory: WL/FI

Vol-Page No: 5-18

Grubbs, Elmer
Assistant Professor, MS
Engineering
New Mexico Highland University
Las Vegas, NM 87701-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 6

Guest, Joyce
Associate, PhD
Department of Chemistry
University of Cincinnati
Cincinnati, OH 45221-0172

Field: Physical Chemistry
Laboratory: WL/ML

Vol-Page No: 5-31

Gumbs, Godfrey
Professor, PhD
Physics & Astronomy
University New York Hunters Co
New York, NY 10021-0000

Field: Condensed Matter Physics
Laboratory: WL/EL

Vol-Page No: 5-12

Hakkinen, Raimo
Professor, PhD
207 Jolley Hall
Washington University
St. Louis, MO 63130-0000

Field: Mechanical Engineering
Laboratory: WL/FI

Vol-Page No: 5-19

Hall, Jr., Charles
Assistant Professor, PhD
Mech & Aerospace Engr.
North Carolina Univ.
Raleigh, NC 27695-7910

Field:
Laboratory: WL/FI

Vol-Page No: 5-20

Hancock, Thomas
Assistant Professor, PhD

Grand Canyon University
, - 0

Field: Educational Psychology
Laboratory: AL/HR

Vol-Page No: 2-30

Hannafin, Michael
Visiting Professor, PhD
305-D Stone Building,3030
Florida State University
Tallahassee, FL 3-2306

Field: Educational Technology
Laboratory: AL/HR

Vol-Page No: 2-31

Helbig, Herbert
Professor, PhD
Physics
Clarkson University
Potsdam, NY 13699-0000

Field: Physics
Laboratory: RL/ER

Vol-Page No: 4- 8

Henry, Robert
Professor, PhD
Electrical Engineering
University of Southwestern Lou
Lafayette, LA 70504-3890

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 3

SFRP Participant Data

Hong, Lang
Assistant Professor, PhD
Dept of Electrical Engin
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 7

Hsu, Lifang
Assistant Professor, PhD

Le Moyne College
, - 0

Field: Mathematical Statistics
Laboratory: RL/ER

Vol-Page No: 4- 9

Huang, Ming
Assistant Professor, PhD
500 NW 20th Street
Florida Atlantic University
Boca Raton, FL 33431-0991

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2-12

Humi, Mayer
Professor, PhD
Mathematics
Worchester Polytechnic Institu
Worchester, MA 1609-2280

Field: Applied Mathematics
Laboratory: PL/GP

Vol-Page No: 3- 2

Humi, Mayer
Professor, PhD
Mathematics
Worchester Polytechnic Institu
Worchester, MA 1609-2280

Field: Applied Mathematics
Laboratory: /

Vol-Page No: 0- 0

Jabbour, Kamal
Associate Professor, PhD
121 Link hall
Syracuse University
Syracuse, NY 13244-1240

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 4

Jaszczak, John
Assistant Professor, PhD
Dept. of Physics
Michigan Technological Univers
Houghton, MI 49931-1295

Field:
Laboratory: WL/ML

Vol-Page No: 5-32

Jeng, San-Mou
Associte, PhD
Mail Location #70
University of Cincinnati
Cincinnati, OH 45221-0070

Field: Aerospace Engineering
Laboratory: PL/RK

Vol-Page No: 3-17

Johnson, David
Associate Professor, PhD
Dept of Chemistry
University of Dayton
Dayton, OH 45469-2357

Field: Chemistry
Laboratory: WL/ML

Vol-Page No: 5-33

Karimi, Amir
Associate, PhD
Division Engineering
University of Texas, San Anton
San Antonio, TX 7824-9065

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-26

Kheyfets, Arkady
Assistant Professor, PhD
Dept. of Mathematics
North Carolina State Univ.
Raleigh, NC 27695-7003

Field:
Laboratory: PL/VT

Vol-Page No: 3-27

Koblasz, Arthur
Associate, PhD
Civil Engineeing
Georgia State University
Atlanta, GA 30332-0000

Field: Engineering Science
Laboratory: AL/AO

Vol-Page No: 2- 3

Kraft, Donald
Professor, PhD
Dept. of Computer Science
Louisiana State University
Baton Rouge, LA 70803-4020

Field:
Laboratory: AL/CF

Vol-Page No: 2-13

Kumar, Rajendra
Professor, PhD
1250 Bellflower Blvd
California State University
Long Beach, CA 90840-0000

Field: Electrical Engineering
Laboratory: RL/C3

Vol-Page No: 4- 5

Kumta, Prashant
Assistant Professor, PhD
Dept of Materials Science
Carnegie-Mellon University
Pittsburgh, PA 15213-3890

Field: Materiels Science
Laboratory: WL/ML

Vol-Page No: 5-34

Kuo, Spencer
Professor, PhD
Route 110
Polytechnic University
Farmingdale, NY 11735-0000

Field: Electrophysics
Laboratory: PL/GP

Vol-Page No: 3- 3

Lakeou, Samuel
Professor, PhD
Electrical Engineering
University of the District of
Washnington, DC 20008-0000

Field: Electrical Engineering
Laboratory: PL/VT

Vol-Page No: 3-28

Langhoff, Peter
Professor, PhD

Indiana University
Bloomington, IN 47405-4001

Field: Dept. of Chemistry
Laboratory: PL/RK

Vol-Page No: 3-18

Lawless, Brother
Assoc Professor, PhD
Dept. Science /Mathematic
Fordham University
New York, NY 10021-0000

Field: Box 280
Laboratory: AL/OE

Vol-Page No: 2-38

Lee, Tzesan
Associate Professor, PhD
Dept. of Mathematics
Western Illinois University
Macomb, IL 61455-0000

Field:
Laboratory: AL/OE

Vol-Page No: 2-39

Lee, Min-Chang
Professor, PhD
167 Albany Street
Massachusetts Institute
Cambridge, MA 2139-0000

Field: Plasma Fusion Center
Laboratory: PL/GP

Vol-Page No: 3- 4

Lee, Byung-Lip
Associate Professor, PhD
Engineering Sci. & Mechan
Pennsylvania State University
University Park, PA 16802-0000

Field: Materials Engineering
Laboratory: WL/ML

Vol-Page No: 5-35

Leigh, Wallace
Assistant Professor, PhD
26 N. Main St.
Alfred University
Alfred, NY 14802-0000

Field: Electrical Engineering
Laboratory: RL/ER

Vol-Page No: 4-10

Levin, Rick
Research Engineer II, MS
EM Effects Laboratory
Georgia Institute of Technolog
Atlanta, GA 30332-0800

Field: Electrical Engineering
Laboratory: RL/ER

Vol-Page No: 4-11

SFRP Participant Data

Li, Jian
Asst Professor, PhD
216 Larsen Hall
University of Florida
Gainesville, FL 32611-2044

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 8

Lilienfield, Lawrence
Professor, PhD
3900 Reservoir Rd., NW
Georgetown University
Washington, DC 20007-0000

Field: Physiology & Bio s
Laboratory: WHMC/

Vol-Page No: 6-14

Lim, Tae
Assistant Professor, PhD
2004 Learned Hall
University of Kansas
Lawrence, KA 66045-0000

Field: Mechanical/Aerospace Engr
Laboratory: FJSRL/

Vol-Page No: 6- 8

Lin, Paul
Associate Professor, PhD
Mechanical Engineering
Cleveland State University
Cleveland, OH 4-4115

Field: Associate Professor
Laboratory: WL/FI

Vol-Page No: 5-21

Liou, Juin
Associate Professor, PhD
Electrical & Computer Eng
University of Central Florida
Orlando, FL 32816-2450

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 5-13

Liu, David
Assistant Professor, PhD
100 Institute Rd.
Worcester Polytechnic Inst.
Worcester, MA 1609-0000

Field: Department of Physics
Laboratory: RL/ER

Vol-Page No: 4-12

Losiewicz, Beth
Assistant Professor, PhD
Experimental Psychology
Colorado State University
Fort Collins, CO 80523-0000

Field: Psycholinguistics
Laboratory: RL/IR

Vol-Page No: 4-17

Loth, Eric
Assistant Professor, PhD
104 S. Wright St, 321C
University of Illinois-Urbana
Urbana, IL 61801-0000

Field: Aeronaut/Astronaut Engr
Laboratory: AEDC/

Vol-Page No: 6- 4

Lu, Christopher
Associate Professor, PhD
300 College Park
University of Dayton
Dayton, OH 45469-0246

Field: Dept Chemical Engineering
Laboratory: WL/PO

Vol-Page No: 5-61

Manoranjan, Valipuram
Associate Professor, PhD
Neill Hall
Washington State University
Pullman, WA 99164-3113

Field: Pure & AppliedMathematics
Laboratory: AL/EQ

Vol-Page No: 2-22

Marsh, James
Professor, PhD
Physics
University of West Florida
Pensacola, FL 32514-0000

Field: Physics
Laboratory: WL/MN

Vol-Page No: 5-47

Massopust, Peter
Assistant Professor, PhD

Sam Houston State University
Huntsville, TX 77341-0000

Field: Dept. of Mathematics
Laboratory: AEDC/

Vol-Page No: 6- 5

Miller, Arnold
Senior Instructor, PhD
Chemistry & Geochemistry
Colorado School of Mines
Golden, CO 80401-0000

Field:
Laboratory: FJSRL/

Vol-Page No: 6- 9

Misra, Pradeep
Associate Professor, PhD

University of St. Thomas
, - 0

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5- 9

Monsay, Evelyn
Associate Professor, PhD
1419 Salt Springs Rd
Le Moyne College
Syracuse, NY 13214-1399

Field: Physics
Laboratory: RL/OC

Vol-Page No: 4-23

Morris, Augustus
Assistant Professor, PhD

Central State University
, - 0

Field: Biomedical Science
Laboratory: AL/CF

Vol-Page No: 2-14

Mueller, Charles
Professor, PhD
W140 Seashore Hall
University of Iowa
Iowa City, IA 52242-0000

Field: Dept of Sociology
Laboratory: AL/HR

Vol-Page No: 2-32

Murty, Vedula
Associate Professor, MS

Texas Southern University
, - 0

Field: Physics
Laboratory: PL/VT

Vol-Page No: 3-29

Musavi, Mohamad
Assoc Professor, PhD
5708 Barrows Hall
University of Maine
Orono, ME 4469-5708

Field: Elect/Comp. Engineering
Laboratory: RL/IR

Vol-Page No: 4-18

Naishadham, Krishna
Assistant Professor, PhD
Dept. of Electrical Eng.
Wright State University
Dayton, OH 45435-0000

Field: Electrical Engineering
Laboratory: WL/EL

Vol-Page No: 5-14

Noel, Charles
Associate Professor, PhD
151A Campbell Hall
Ohio State University
Columbus, OH 43210-1295

Field: Dept of Textiles & Cloth
Laboratory: PL/RK

Vol-Page No: 3-19

Norton, Grant
Asst Professor, PhD
Mechanical & Materials En
Washington State University
Pullman, WA 99164-2920

Field: Materials Science
Laboratory: WL/ML

Vol-Page No: 5-36

Noyes, James
Professor, PhD
Mathematics & Computer Sc
Wittenberg University
Springfield, OH 45501-0720

Field: Computer Science
Laboratory: WL/FI

Vol-Page No: 5-22

Nurre, Joseph
Assistant Professor, PhD
Elec. & Computer Engineer
Ohio University
Athens, OH 45701-0000

Field: Mechanical Engineering
Laboratory: AL/CF

Vol-Page No: 2-15

Nygren, Thomas
Associate Professor, PhD
1885 Neil Ave. Mail
Ohio State University
Columbus, OH 43210-1222

Field: Department of Psychology
Laboratory: AL/CF

Vol-Page No: 2-16

Osterberg, Ulf
Assistant Professor, PhD
Thayer School of Engrg.
Dartmouth College
Hanover, NH 3755-0000

Field:
Laboratory: FJSRL/

Vol-Page No: 6-10

Pan, Ching-Yan
Associate Professor, PhD
Physics
Utah State University
Logan, UT 84322-4415

Field: Condensed Matter Physics
Laboratory: PL/WS

Vol-Page No: 3-36

Pandey, Ravindra
Assistant Professor, PhD
1400 Townsend Dr
Michigan Technological Univers
Houghton, MI 49931-1295

Field: Physics
Laboratory: FJSRL/

Vol-Page No: 6-11

Patton, Richard
Assistant Professor, PhD
Mechanical&Nuclear Engine
Mississippi State University
Mississippi State, MS 39762-0000

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-30

Peretti, Steven
Assistant Professor, PhD
Chemical Engineering
North Carolina State Univ.
Raleigh, NC 27695-7905

Field:
Laboratory: AL/EQ

Vol-Page No: 2-23

Petschek, Rolfe
Associate Professor, PhD
Department of Physics
Case Western Reserve Universit
Cleveland, OH 44106-7970

Field: Physics
Laboratory: WL/ML

Vol-Page No: 5-37

Pezeshki, Charles
Assistant Professor, PhD

Washington State University
Pullman, WA 99164-2920

Field: Mechanical Engineering
Laboratory: FJSRL/

Vol-Page No: 6-12

Piepmeier, Edward
Assistant Professor, PhD
College of Pharmacy
University of South Carolina
Columbia, SC 29208-0000

Field:
Laboratory: AL/AO

Vol-Page No: 2- 4

Pittarelli, Michael
Associate Professor, PhD
PO Box 3050, Marcy Campus
SUNY, Institute of Technology
Utica, NY 13504-3050

Field: Information Sys & Engr.
Laboratory: RL/C3

Vol-Page No: 4- 6

Potasek, Mary
Research Professor, PhD

Columbia University
, - 0

Field: Physics
Laboratory: WL/ML

Vol-Page No: 5-38

Prasad, Vishwanath
Professor, PhD

SUNY, Stony Brook
Stony Brook, NY 11794-2300

Field: Mechanical Engineering
Laboratory: RL/ER

Vol-Page No: 4-13

Priestley, Keith
Research Scientist, PhD

University of Nevada, Reno
, - 0

Field: Geophysics
Laboratory: PL/GP

Vol-Page No: 3- 5

Purasinghe, Rupasiri
Professor, PhD
5151 State Univ. Dr.
California State Univ.-LA
Los Angeles, CA 90032-0000

Field: Dept of Civil Engineering
Laboratory: PL/RK

Vol-Page No: 3-20

Raghu, Surya
Assistant Professor, PhD
Mechanical Engineering
SUNY, Stony Brook
Stony Brook, NY 11794-2300

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-62

Ramesh, Ramaswamy
Associate Professor, PhD
School of Management
SUNY, Buffalo
Buffalo, NY 14260-0000

Field: Magement Science/Syatems
Laboratory: AL/HR

Vol-Page No: 2-33

Ramm, Alexander
Professor, PhD
Mathematics
Kansas State University
Manhattan, KS 66506-2602

Field:
Laboratory: AL/CF

Vol-Page No: 2-17

Ray, Paul
Assistant Professor, PhD
Box 870288
University of Alabama
Tuscaloosa, AL 35487-0288

Field: Industrial Engineering
Laboratory: AL/OE

Vol-Page No: 2-40

Reimann, Michael
Assistant Instructor, MS
Information Systems
The University of Texas-Arling
Arlington, TX 76019-0437

Field: Computer Science
Laboratory: WL/MT

Vol-Page No: 5-55

Rodriguez, Armando
Assistant Professor, PhD

Arizona State University
Tempe, AZ 85287-7606

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-48

Rohrbaugh, John
Research Engineer, PhD
347 Ferst St
Georgia Institute of Technolog
Atlanta, GA 30332-0800

Field: Sensors & Applied Electro
Laboratory: RL/ER

Vol-Page No: 4-14

Roppel, Thaddeus
Associate Professor, PhD
200 Broun Hall
Auburn University
Auburn, AL 36849-5201

Field: Electrical Engineering
Laboratory: WL/MN

Vol-Page No: 5-49

Rosenthal, Paul
Professor, PhD
Mathematics
Los Angeles City College
Los Angeles, CA 90027-0000

Field: Mathematics
Laboratory: PL/RK

Vol-Page No: 3-21

Rotz, Christopher
Associate Professor, PhD

Brigham Young University
Provo, UT 84602-0000

Field: Mechanical Engineering
Laboratory: PL/VT

Vol-Page No: 3-31

Rudolph, Wolfgang
Associate Professor, PhD
Dept of Physics and Astro
University of New Mexico
Albuquerque, NM 84131-0000

Field: Physics
Laboratory: PL/LI

Vol-Page No: 3- 0

Rudzinski, Walter
Professor, PhD
Dept. of Chemistry
Southwest Texas State Universi
San Marcos, TX 78610-0000

Field: Professor
Laboratory: AL/OE

Vol-Page No: 2-41

Rule, William
Asst Professor, PhD
Mechanical Engineering
University of Alabama
Tuscaloosa, AL 35487-0278

Field: Engineering Mechanics
Laboratory: WL/MN

Vol-Page No: 5-50

Ryan, Patricia
Research Associate, MS
Georgia Tech Research Ins
Georgia Institute of Tech
Atlanta, GA 30332-0000

Field: Electrical Engineering
Laboratory: WL/AA

Vol-Page No: 5-10

Saiduddin, Syed
Professor, PhD
1900 Coffey Rd
Ohio State University
Columbus, OH 43210-1092

Field: Physiology/Pharmacology
Laboratory: AL/OE

Vol-Page No: 2-42

Schonberg, William
Assoc Professor, PhD
Engineering Dept.
University of Alabama, Huntsvi
Huntsville, AL 35899-0000

Field: Civil & Environmental
Laboratory: WL/MN

Vol-Page No: 5-51

Schulz, Timothy
Assistant Professor, PhD
1400 Townsend Dr
Michigan Technological Univers
Houghton, MI 49931-1295

Field: Electrical Engineering
Laboratory: PL/LI

Vol-Page No: 3-11

Shen, Mo-How
Assistant Professor, PhD
2036 Neil Ave.
Ohio State University
Columbus., OH 43210-1276

Field: Aerospace Engineering
Laboratory: WL/FI

Vol-Page No: 5-23

Sherman, Larry
Professor, PhD
Dept. of Chemistry
University of Scranton
Scranton, PA 18510-4626

Field: Analytical Chemistry
Laboratory: AL/OE

Vol-Page No: 2-43

Shively, Jon
Professor, PhD
Civil & Industrial Eng.
California State University, N
Northridge, CA 91330-0000

Field: Metallurgy
Laboratory: PL/VT

Vol-Page No: 3-32

Snapp, Robert
Assistant Professor, PhD
Dept of Computer Science
University of Vermont
Burlington, VT 5405-0000

Field: Physics
Laboratory: RL/IR

Vol-Page No: 4-19

Soumekh, Mehrdad
Associate Professor, PhD
201 Bell Hall
SUNY, Buffalo
Amherst, NY 14260-0000

Field: Elec/Computer Engineering
Laboratory: PL/LI

Vol-Page No: 3-12

Spetka, Scott
Assistant Professor, PhD
PO Box 3050, Marcy Campus
SUNY, Institute of Technology
Utica, NY 13504-3050

Field: Information Sys & Engrg
Laboratory: RL/XP

Vol-Page No: 4-26

Springer, John
Associate Professor, PhD

Fisk University
, - 0

Field: Physics
Laboratory: AEDC/

Vol-Page No: 6- 6

Stevenson, Robert
Assistant Professor, PhD
Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556-0000

Field: Electrical Engineering
Laboratory: RL/IR

Vol-Page No: 4-20

Stone, Alexander
Professor, PhD
Mathematics & Statistics
University of New Mexico
Alburquerque, NM 87131-1141

Field:
Laboratory: PL/WS

Vol-Page No: 3-37

Sveum, Myron
Assistant Professor, MS
Electronic Engineering Te
Metropolitan State College
Denver, CO 80217-3362

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-24

Swanson, Paul
Research Associate, PhD
Electrical Engineering
Cornell University
Ithaca, NY 14853-0000

Field: Electrical Engineering
Laboratory: RL/OC

Vol-Page No: 4-25

Swope, Richard
Professor, PhD
Engineering Science
Trinity University
San Antonio, TX 78212-0000

Field: Mechanical Engineering
Laboratory: AL/AO

Vol-Page No: 2- 5

Tan, Arjun
Professor, PhD
Physics
Alabama A&M University
Normal, AL 35762-0000

Field: Physics
Laboratory: PL/WS

Vol-Page No: 3-38

Tarvin, John
Associate Professor, PhD
800 Lakeshore Drive
Samford University
Birmingham, AL 35229-0000

Field: Department of Physics
Laboratory: AEDC/

Vol-Page No: 6- 7

Taylor, Barney
Visiting Assist Professor, PhD
1601 Peck Rd.
Miami Univ. - Hamilton
Hamilton, OH 4-5011

Field: Dept. of Physics
Laboratory: WL/ML

Vol-Page No: 5-39

Thio, Y.
Associate Professor, PhD

University of Miami
Coral Gables, FL 33124-0530

Field: Physics Dept.
Laboratory: PL/WS

Vol-Page No: 3-39

Tong, Carol
Assistant Professor, PhD
Electrical Engineering
Colorado State University
Fort Collins, CO 80523-0000

Field:
Laboratory: WL/AA

Vol-Page No: 5-11

Truhon, Stephen
Associate Professor, PhD
Social Sciences
Winston-Salem State University
Winston-Salem, NC 27110-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 2-34

Tzou, Horn-Sen
Associate Professor, PhD
Mechanical Engineering
University of Kentucky
Lexington, KY 40506-0046

Field: Mechanical Engineering
Laboratory: WL/FI

Vol-Page No: 5-24

Vogt, Brian
Professor, PhD

Bob Jones University
, - 0

Field: Pharmaceutical Sciences
Laboratory: AL/EQ

Vol-Page No: 2-24

Wang, Xingwu
Asst Professor, PhD
Dept. of Electrical Eng.
Alfred University
Alfred, NY 14802-0000

Field: Physics
Laboratory: WL/FI

Vol-Page No: 5-25

Whitefield, Philip
Research Assoc Professor, PhD
Cloud & Aerosol Sciences
University of Missouri-Rolla
Rolla, MO 65401-0000

Field: Chemistry
Laboratory: PL/LI

Vol-Page No: 3-13

Willson, Robert
Research Assoc Professor, PhD
Robinson Hall
Tufts University
Medford, MA 2155-0000

Field: Physics and Astronomy
Laboratory: PL/GP

Vol-Page No: 3- 6

Witanachchi, Sarath
Assistant Professor, PhD
4202 East Fowler Avenue
University of South Florida
Tampa, FL 33620-7900

Field: Department of Physics
Laboratory: FJSRL/

Vol-Page No: 6-13

Woehr, David
Assistant Professor, PhD
Psycholoogy
Texas A&M University
College Station, TX 77845-0000

Field: Psychology
Laboratory: AL/HR

Vol-Page No: 2-35

SFRP Participant Data

Xu, Longya
Assistant Professor, PhD
Electrical Engineering
The Ohio State University
Columbus, OH 43210-0000

Field: Electrical Engineering
Laboratory: WL/PO

Vol-Page No: 5-63

Yavuzkurt, Savas
Associate Professor, PhD

Pennsylvania State University
University Park, PA 16802-0000

Field: Mechanical Engineering
Laboratory: WL/PO

Vol-Page No: 5-64

Zhang, Xi-Cheng
Associate Professor, PhD
Physics Department
Rensselaer Polytechnic Institu
Troy, NY 12180-3590

Field: Physics
Laboratory: RL/ER

Vol-Page No: 4-15

Zhou, Kemin
Assistant Professor, PhD
Dept. of Elec & Comp. Eng
Louisiana State University
Baton Rouge, LA 70803-0000

Field:
Laboratory: WL/MN

Vol-Page No: 5-52

Zimmermann, Wayne
Associate Professor, PhD
P.O. Box 22865
Texas Woman's University
Denton, TX 76205-0865

Field: Dept Mathematics/Computer
Laboratory: PL/WS

Vol-Page No: 3-40

Development and testing of DNA probes specific for *Escherichia coli* strain 0157:H7, *Ureaplasma urealyticum*, and *Mycoplasma hominis*

Vito G. DelVecchio
Professor
Department of Biology

University of Scranton
Scranton, PA 18510

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

September 1993

Development and testing of DNA probes specific for *Escherichia coli* strain 0157:H7, *Ureaplasma urealyticum*, and *Mycoplasma hominis*.

Vito G. DelVecchio
Professor of Biology
University of Scranton
Scranton, PA 18510

Abstract

Escherichia coli strain 0157:H7 is enterohemarrhagic causing severe bloody diarrhea in humans. It is often contracted by ingestion of contaminated hamburger meat. Present methods used for diagnosis of this pathogen are time-consuming, expensive, and not of great sensitivity. Attempts to develop a DNA probe for this plasmid-borne virulence were initiated to circumvent these problems and to provide better patient care.

Ureaplasma urealyticum and *Mycoplasma hominis* are among the smallest of all free-living organisms. They have been implicated as the etiologic agent for a variety of disease conditions. DNA probing systems, which are specific for these suspected pathogens, have been developed. These probes were applied using *in situ* hybridization and DNA amplification assays. Approximately 80 clinical specimens were used to test the accuracy of these systems. The results obtained were in complete agreement with results obtained from cultural diagnosis.

Development and testing of DNA probes specific for *Escherichia coli* strain 0157:H7, *Ureaplasma urealyticum*, and *Mycoplasma hominis*

Vito G. DelVecchio

Introduction

Enterohemarrhagic Escherichia coli (EHEC) 0157:H7 was first recognized in 1982 during an outbreak of severe bloody diarrhea caused by contaminated hamburger meat. Infection is commonly associated with the consumption of undercooked ground beef but can also occur from drinking raw milk or sewage contaminated water. Infection not only leads to bloody diarrhea but also abdominal cramps and sometimes hemolytic uremic syndrome (HUS). HUS is characterized by destruction of red blood cells and kidney failure. Currently, there are several clinical tests available for EHEC but they are relatively expensive, time-consuming, and not always accurate. The lack of standardization of these assays from one laboratory to another represents an obstacle in diagnosis of EHEC. The genes for EHEC virulence reside on a 60 megadalton (mD) plasmid. A DNA probe specific for EHEC would circumvent these problems and also result in more rapid therapy for patients. The present study is concerned with development of a DNA probe for EHEC strain 0157:H7. Such a diagnostic assay would include *in situ* hybridization and nucleic acid amplification of EHEC 0157:H7. DNA probes are economical and highly specific ways of diagnosing illness.

Methodology

Plasmid DNA from *E. coli* strain 0157:H7 was isolated using the Magic Miniprep Method kit (Promega). The 60 mD plasmid was separated from genomic

DNA and other plasmids by agarose gel electrophoresis using a 1% agarose gel. The resulting band of plasmid DNA was excised from the gel and filtered through a 0.45 um Ultrafree-MC Filter Unit (Millipore). The plasmid DNA was precipitated in 7.5M ammonium acetate and 95% ethanol followed by centrifugation at 14000xg and resuspension in distilled water. It was then hydrolyzed with the restriction endonuclease Pst I. The resulting fragments were Genecleaned (Bio 101) and ligated into the plasmid vector pUC18, and the recombinant DNA molecules were cloned into *E. coli* DH5α competent cells. These cells were transformed by the method of Hanahan (1983) and grown on Luria-Bertani agar containing ampicillin in the presence of X-gal (5-bromo-4-chloro-3-indoyl-B-D-galactoside) and IPTG (isopropyl-B-D-thiogalactopyranoside). Colonies containing plasmids with inserts appeared yellow in color and non-recombinants were blue. The recombinant plasmids were isolated from each clone by the miniprep boiling lysis method of Holmes (1981), hydrolyzed with Pst I, and electrophoresed on a 1% agarose gel containing ethidium bromide. DNA was visualized using an ultraviolet trans-illuminator and sized by comparison with a standard 1 Kb DNA ladder (Gibco-BRL).

Results

Several attempts were made at forming a clone library of the 60 mD plasmid of EHEC. The majority of clones contained inserts which were identical in size (3600 bp) some clones contained additional DNA segments which ranged in size from 100 to 3600 bp - however there was not a great diversity of insert sizes.

Discussion

The reason for the general lack of variation in insert size is not known at this point. The use of other restriction endonucleases to generate libraries or electroporation to cause transformation may result in more diverse libraries.

The resulting library for EHEC 0157:H7 specific genes will be analyzed as follows: Recombinant plasmids containing inserts will be blotted onto positively charged nylon membrane (Boehringer Mannheim) by the method of Southern (1975). Probing with DNA of other diarrheagenic strains (*E. coli* strains 188, 189, 194, 221) will detect inserts which are not EHEC-specific. Those inserts which do not hybridize with these strains will be considered possible EHEC probing candidates.

The Southern blot will be probed with digoxigenin-labeled DNA probes. The The hybrids will be detected by complexing digoxigenin DNA with anti-digoxigenin antibodies complexed with alkaline phosphatase. The presence of hybridized probes will be signaled by the action of alkaline phosphatase on Lumi-Phos 530 (Boehringer Mannheim). The enzyme causes the hydrolysis of Lumi-Phos 530 which results in the emission of photons which are analyzed on Fuji Medical X-Ray Film (Fuji Photo Film Company). Clones of interest will be sequenced to determine the nucleotide content of the insert DNA.

Sequencing will be accomplished using a non-isotopic DNA Sequenase Imaging System (United States Biochemical Corporation) in conjunction with the Auto Trans 530 direct transfer electrophoresis system (Betagen). After the sequence has been read, PCR primers will be synthesized and amplification studies will be done. At this point polymerase chain reaction (PCR) analysis of EHEC strain 0157:H7 can be accomplished and used as a diagnostic tool.

Since the 60 mD plasmid is found in multicopies in *E. coli*, *in situ* DNA probe analysis is also possible. Fluorescent-labeled oligonucleotides specific for 0157:H7-unique 20 mers on the plasmid can be used to diagnose the presence of this strain of *E. coli*. *In situ* detection can be used in conjunction with PCR or by itself. Thus the two different probing systems can easily be adapted to most clinical laboratory settings.

PCR Detection of *Mycoplasma hominis* and *Ureaplasma urealyticum* in Clinical Isolates.

Introduction

Mycoplasma hominis and *Ureaplasma urealyticum* have been associated with infertility, pregnancy wastage, pelvic inflammatory disease, pneumonia of neonates, and non-specific urethritis. There has been no satisfactory method for the direct identification of these organisms in clinical samples since culture assays involve complex media, lengthy incubations, frequent growth failure and contamination; therefore their total involvement in the disease process has never been elucidated. With this in mind we have established DNA libraries of these organisms in *E. coli*. These libraries have afforded a probes collection of varying specificity and sensitivity as defined by hybridization and visualization by the use of Lumi-Phos 530.

Since PCR has become available to researchers and clinicians, the application of this technique has revolutionized diagnostic medicine for it offers the most sensitive, rapid, and less labor-intensive means of detecting specific nucleic acid segments. Thus we applied this technique's sensitivity to detect the cloned *Ureaplasma* and *Mycoplasma* sequences present in clinical samples.

PCR is based upon the natural DNA replication process since the number of DNA molecules doubles after each cycle. Target DNA is first denatured, or converted from the double-to-single-standard state, at a high temperature (95^0 - 100^0 C). The denatured single-stranded DNA will then anneal to complementary primer once the temperature of the reaction is lowered. Primers are single-stranded oligonucleotides which limit and define a target DNA segment of gene or organism. Primers are added in molar excess so that they may easily anneal to disassociated target single-stranded DNA. Once annealing has occurred, the enzyme DNA polymerase catalyzes the synthesis of new strands of target DNA segments. The synthesis occurs by the DNA polymerase adding nucleotides which are complimentary to the unpaired DNA strands onto the annealed primers 3'OH. After 30 cycles of denaturation, linker annealing and primer extension a single target DNA can be amplified or duplicated up to 1,000,000 copies in the space of 3 hours. The amplified DNA can then be identified via agarose gel electrophoresis or hybridization techniques.

Methodology

Clinical samples used included: throat swabs, blood samples, tracheal aspirants, lung biopsies, urethral and vaginal swabs, and placental tissue. The specimens were shipped in various types of transport media, including: Mycotrans transport media (Irvine Scientific), Remel Arginine broth (Remel), and Remel 10B media. *Mycoplasma hominis* identification was done by microscopic observation of colony characteristics on Remel A8 media and Mycotrim GU media (Irvine Scientific).

Clinical samples were prepared for PCR analysis by aliquoting 1 ml of transport media into a microcentrifuge tube. This was centrifuged at 14,000 x g for 15 minutes in a microcentrifuge. The supernatant was discarded and the remaining pellet was resuspended in 100 ul of sterile distilled water, boiled for 10 minutes, and stored at -20^{0} C until use.

Clinical samples containing cotton swabs were placed in 1 ml of sterile distilled water and vortexed gently for 15 seconds. The swab was then discarded and the liquid portion was centrifuged at a low speed for 3 min. The supernatant was transferred to a microcentrifuge tube, boiled for 10 minutes, and stored at -20^{0} C until use.

PCR amplification was done using the Gene Amp Kit with AmpliTaq DNA Polymerase (Perkin Elmer-Cetus) in the Techne PHC-3 Dri Block Cycler (Techne Corporation). Twenty-five ul of isolated clinical sample was added to 75 ul of PCR reaction mix containing: 50 mM Tris (pH 8.3); 1.5 mM MgCl; 200 uM dNTP's; 0.25 ul of Taq Polymerase; and 0.15 mM 20-mer primers Mh-2 (5'-GGTGATTCACGTTGTATGC-3') and Mh-3 (5'-GGTCCTAGACAACTTATAAG-3'). The PCR reaction mix was overlaid with 50 ul of sterile mineral oil (Sigma Chemical Company). PCR times and temperatures were as follows: initial denaturation at 95^{0} C for 3 min; amplification using 35 cycles of: denaturation at 94^{0} C for 1 min, annealing at 60^{0} C for 1 min, and extension at 72^{0} C for 1 min. An additional 5 min at 72^{0} C was added at the end of the 35 cycles to allow for complete extension of the primers.

Twenty-five ul of amplified PCR product was electrophoresed for 60 min using a 2% Nu Sieve agarose gel (FMC Bioproducts). The gel was prepared with Tris-Acetate-EDTA (40 mM Tris, 20 mM Acetic Acid, 1 mM EDTA, pH 8.3), containing ethidium bromide. The PCR amplicons were visualized on an

ultraviolet transilluminator (Fotodyne) and sized by comparison with the standard 100 bp DNA ladder (Gibco BRL).

Results

Those amplicons which tested positive for *M. hominis* displayed the 152 bp band. A total of 82 clinical samples were tested for *M. hominis* and the PCR based results were in 100% accordance with the media based results recorded in the Epidemiologic Research Division/Bacteriology Section at Brooks Air Force Base, San Antonio, Texas, as is indicated in Table 1. PCR products resulting from amplification of *U. urealyticum* target segments resulted in a band consisting of 186 bp. The clinical samples tested were in total agreement with the cultural assay as can be seen in Table 2.

Discussion

The specificity and sensitivity characteristics of the *M. hominis* and *U. urealyticum* diagnostic PCR assay was confirmed. All results were in accordance with data recorded using the cultural diagnostic assays. PCR diagnosis for the presence of these bacteria is a considerable improvement over the cultural systems for diagnosis can take place in as few as 8 hours. PCR assay predicted the outcome of the cultural assay days before results were obtained. In summation, the DNA probes are specific, sensitive, rapid, and not labor-intensive. Routine application of the PCR assay is now beginning.

TABLE 1

Clinical Data
Mycoplasma hominis

No.	Sample No.	Media	PCR	Culture
1	924	10B	-	-
2	925	Arg. Broth	+	+
3	1231	Arg. Broth	-	-
4	2057	Arg. Broth	-	-
5	1239	Arg. Broth	+	+
6	2513	Arg. Broth	+	+
7	2672	10B	-	-
8	2463	Arg. Broth	+	+
9	2044	Arg. Broth	+	+
10	1641	Arg. Broth	-	-
11	2016	Arg. Broth	-	-
12	2017	Arg. Broth	-	-
13	2018	Arg. Broth	-	-
14	2019	Arg. Broth	-	-
15	1293	Arg. Broth	+	+
16	1498	Arg. Broth	+	+
17	2153	Arg. Broth	-	-
18	2116	Arg. Broth	+	+
19	523	Arg. Broth	-	-
20	2838	Arg. Broth	-	-
21	1048	Arg. Broth	+	+
22	2645	Arg. Broth	+	+
23	2447	Arg. Broth	-	-
24	788	Arg. Broth	+	+
25	2451	Arg. Broth	-	-

No.	Sample No.	Media	PCR	Culture
26	2603	Arg. Broth	+	+
27	2696	Mycotrans	-	-
28	2960	Mycotrans	-	-
29	2448	Mycotrans	-	-
30	1286	Mycotrans	+	+
31	3562	Mycotrans	-	-
32	3563	Mycotrans	-	-
33	2115	Mycotrans	+	+
34	3299	Mycotrans	-	-
35	1292	Mycotrans	-	-
36	1438	Mycotrans	-	-
37	2450	Mycotrans	-	-
38	1569	Mycotrans	-	-
39	1309	Mycotrans	-	-
40	3453	Mycotrans	-	-
41	1074	Mycotrans	-	-
42	1913	Mycotrans	-	-
43	1565	Mycotrans	-	-
44	1513	Mycotrans	-	-
45	710	Mycotrans	-	-
46	1672	Mycotrans	-	-
47	329	Mycotrans	-	-
48	2242	Mycotrans	-	-
49	1768	Mycotrans	-	-
50	1034	Mycotrans	-	-

No.	Sample No.	Media	PCR	Culture
51	787	Mycotrans	-	-
52	1838	Mycotrans	-	-
53	1713	Mycotrans	-	-
54	1839	Mycotrans	-	-
55	1847	Mycotrans	-	-
56	1916	Mycotrans	+	+
57	1782	Mycotrans	-	-
58	2802	Mycotrans	-	-
59	1751	Mycotrans	-	-
60	1794	Mycotrans	-	-
61	2807	Mycotrans	-	-
62	95	Mycotrans	-	-
63	1666	Mycotrans	-	-
64	2270	Mycotrans	-	-
65	2292	Mycotrans	-	-
66	2425	Mycotrans	-	-
67	2849	Mycotrans	-	-
68	655	Mycotrans	-	-
69	2700	Mycotrans	-	-
70	1679	Mycotrans	-	-
71	1882	Mycotrans	-	-
72	3444	Mycotrans	-	-
73	2695	Mycotrans	-	-
74	1844	Mycotrans	+	+
75	1945	Mycotrans	+	+
76	1948	Mycotrans	-	-
77	1939	Mycotrans	-	-
78	1968	Mycotrans	-	-
79	1965	Mycotrans	-	-
80	2003	Mycotrans	-	-
81	2006	Mycotrans	-	-
82	1592	Arg. Broth	-	-

TABLE 2

Clinical Data
Ureaplasma urealyticum

No.	Patient No.	Date	Media	PCR	Culture
1	2696	10/31/92	M	-	-
2	2643	10/08/92	M	+	+
3	2017	08/18/92	M	+	+
4		07/29/92	M	-	-
5	2594	10/05/92	M	-	-
6	2521	09/29/92	M	-	-
7	2016	08/19/92	M	+	+
8	2019	08/19/92	M	-	-
9	1239	04/18/91	Arg	+	+
10	1286	05/07/93	Arg	+	+
11	2513	10/13/92	Arg	-	-
12	1592	06/10/93	M	-	-
13	1604	06/11/93	M	+	+
14	1264	05/04/93	M	-	-
15	3300	12/04/92	M	-	-
16	3092	11/16/92	M	-	-
17	2018	08/19/92	M	-	-
18	2115	08/27/92	M	-	-
19	2116	08/27/93	M	-	-
20	788	03/19/93	MT	-	-
21	1744	05/09/92	M	-	-
22	2450	09/24/92	M	+	+
23	2849	12/22/66(DOB)	M	+	+
24	1293	03/10/93	M	+	+
25	1338	04/21/93	M	+	+

No.	Patient No.	Date	Media	PCR	Culture
26	1292	05/06/93	M	+	+
27	1565	06/08/93	M	-	-
28	1610	06/15/93	U. Media	+	+
29	1438	03/25/93	M	+	+
30	1074	04/14/93	M	-	-
31	1813	06/02/93	M	+	+
32	610	03/02/93	U. ?	+	+
33	1263	05/04/93	M	+	+
34	323	02/04/93	M	+	+
35	1112	04/21/93	M	+	+
36	1048	04/13/93	M	-	-
37	1513	06/02/93	M	+	+
38	sample contamination				
39	1672	06/21/93	M	-	-
40	1666	06/21/93	M	-	-
41	1783	06/29/93	M	-	-
42	1758	06/29/93	M	-	-
43	1772	06/30/93	U. Media	-	-
44	1776	06/30/93	M	+	+
45	1779	06/30/93	M	+	+
46	1679	06/23/93	M	+	+
47	1794	07/01/93	M	+	+
48	1794	07/01/93	10B	+	+
49	1838	07/07/93	M	+	contaminated
50	1894	07/07/93	M	+	+

No.	Patient No.	Date	Media	PCR	Culture
S1	1777	05/27/92	U. Media	-	-
S2	2047	08/20/92	U. Media	-	-
S3	1883	08/03/92	U. Media	-	-
S4	95	01/12/93	MT	+	+
S5	1592	06/10/93	Arg	-	-
S6	1590	06/10/93	10B	-	-
S7	655	03/05/93	MT	+	+
S8	1074	04/10/93	10B	+	+
S9	2002	10/06/92	Arg	+	+
60	2285	07/23/92	10B	+	+
61	710	03/12/93	MT	+	+
62	877	03/26/93	10B	+	+
63	3175	11/20/92	10B	+	+
64	....	06/04/92	M	-	-
65	1892	07/09/93	M	-	-
66	1916	07/12/93	MT	+	+
67	2006	07/19/93	M	+	+
68	1965	07/15/93	M	+	+
69	2021	07/21/93	M	-	-
70	2028	07/21/93	M	-	-
71	2019	07/21/93	M	-	-
72	1939	07/14/93	M	-	-
73	1948	07/14/93	M	-	-
74	1971	07/13/93	M	-	-
75	1782-84	06/30/93	M	+	+

References

Hanahan, D. 1983. Studies on transformation of *E. coli* with plasmids. Journal of Molecular Biology. 166:557-580.

Holmes, D.S. and Quigly, M. 1981. A rapid boiling method for the preparation of bacterial plasmids. Analytical Biochemistry. 114:193-197.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology. 98:503-517.

ENTROPY GENERATION OF THE CARDIOVASCULAR CYCLE

Richard D. Swope

Professor

Department of Engineering Science

Trinity University

715 Stadium Drive

San Antonio, Texas 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.,

Trinity University

and

North Dakota State University

September 1993

ABSTRACT

The primary aim of this research is to develop mathematical descriptions of the entropy generation of the cardiovascular system. To accomplish this aim, the cardiovascular cycle is modeled by four discrete thermodynamic components - left and right heart, systemic and pulmonic circulation beds. Using the first two laws of thermodynamics, mathematical expressions for the irreversibility of the components are obtained. Blood flow pressure drop and extraction of metabolic fuel largely contribute to the irreversibility of the cardiovascular cycle. For the left and right heart the irreversibility is found to be due to compression losses in the ventricles, metabolic losses, and fluid flow pressure drop. In the systemic circulation, irreversibility was primarily due to fluid flow pressure drop and oxygen and fuel exchange with the tissue. Pulmonary circulation irreversibility was comprised of fluid flow pressure drop and oxygen exhange.

INTRODUCTION

Purpose:

The objective of this research was to describe, in mathematical terms, the entropy generated by the cardiovascular cycle.

Background:

The primary question that motivated this research was: how is the heart hydraulically coupled to its vascular load? Hydraulic coupling theories based on maximum first law thermodynamic efficiency or maximum external work transfer have been proposed (Suga et al, 1985). However, it appears that these approaches are not sufficient to describe the hydraulic coupling in every situation (Hayashida et al,1992). Another approach seemed warranted.

In an excellent review, Gibbs and Chapman (1979) detailed early thermodynamic approaches to cardiac and skeletal muscle tissue. Much useful work has been done in the biothermodynamic analysis at the tissue level. However, the authors performed multiple automated and manual literature searches for previous work in which the cardiovascular cycle (CVC) was analyzed for entropy generation, lost available work, irreversibility, exergy or second law views, but found none.

In contrast, much second law analysis has been performed on engineering thermodynamic cycles such as Otto, Brayton, and vapor compression. It is a well-accepted approach for obtaining additional information about the performance of such systems. (Black and Hartley, 1991, Van Wylen and Sonntag, 1991) Therefore, it is puzzling that this approach has not been used to analyze the CVC. One good reason for studying the CVC from an entropy generation viewpoint is that to minimize irreversibility in a cycle, means that the fuel obtained by the cycle is utilized as effectively as possible. A good strategy for survival is to use what energy one has as effectively as possible - given certain life supporting constraints. Other reasons supporting this approach are that:

2) second law analysis relies on a relatively simple model where a few measurements of state properties can yield much information

3) viewing CV performance from this perspective may help obtain an increase in G-tolerance of tactical aircraft pilots.

4) viewing CV performance from this perspective may help improve clinical diagnosis and therapeutic approach of the general cardiac patient.

Scope of Report:

This report will discuss the basic thermodynamic laws which form the foundation for the mathematical development. Next, the development of the CVC model will be discussed. Then, the mathematical equations describing the entropy generation of each component in the CVC will be developed. Finally, the utility and future promise of this approach will be discussed.

BASIC THERMODYNAMIC LAWS

In this section, the general laws of thermodynamics will be shown along with certain simplifying assumptions which allow for a more convenient mathematical form.

1st law of Thermodynamics for a control volume:

$$\dot{Q}_{c.v} = \frac{d}{dt}\int_V e\rho\, dV + \int_A \left(h + \frac{\upsilon^2}{2} + gZ\right)\rho\upsilon_{rn}\, dA + \dot{W}_{c.v.}$$

Where:

$\dot{Q}_{c.v}$ ≡ rate of heat transfer into the control volume (c.v.) surface.
ρ ≡ density of fluid
e ≡ energy contained in volume, dV
h ≡ enthalpy of substance
υ ≡ velocity of substance
g ≡ gravitational constant
Z ≡ elevation of substance
$\upsilon_{r.n.}$ ≡ outward-directed normal velocity
dA ≡ area of flow, $\dot{m}$
$\dot{W}_{c.v}$ ≡ done by the control volume.

In words the first law states for a control volume ...

rate of heat transfer = rate of energy stored + rate of net energy entering + power produced

For uniform states of mass crossing the control surface, the first law becomes ...

$$\dot{Q}_{C.V.} + \sum \dot{m}_i\left(h_i + \frac{\upsilon_i^2}{2} + gZ_i\right) = \frac{dE_{C.V.}}{dt} + \sum \dot{m}_e\left(h_e + \frac{\upsilon_e^2}{2} + gZ_e\right) + \dot{W}_{C.V.}$$

Where:

$\dot{m}_i$ ≡ mass flow rate into control volume

$\dot{m}_e$ ≡ mass flow rate out of control volume

Furthermore, if one assumes steady energy content in the control volume, that the control volume does not move relative to the coordinate frame, and that the mass flux and its state do not change in time, and finally, that $\dot{Q}_{C.V.}$ and $\dot{W}_{C.V.}$ remain constant, one

obtains a steady state steady flow (SSSF) equation.

$$\dot{Q}_{C.V.} + \sum \dot{m}_i \left(h_i + \frac{\upsilon_i^2}{2} + gZ_i \right) = \sum \dot{m}_e \left(h_e + \frac{\upsilon_e^2}{2} + gZ_e \right) + \dot{W}_{C.V.}$$

If a chemical reaction takes place or if there is a transformation of material, the enthalpies take the form:

$$m_i h_i = H_i = n_1 \bar{h}_1 + n_2 \bar{h}_2 + \ldots + n_n \bar{h}_n$$

Where:

$n_n \bar{h}_n$ = product of moles and molar enthalpy of species n

and

$$\bar{h}_n = \left[h_f^o + \Delta \bar{h} \right]_n$$

Where:

h_f^o ≡ enthalpy of formation at reference pressure and temperature

$\Delta \bar{h}$ ≡ enthalpy change due to non reference pressure and temperature.

Thus, under SSSF conditions and neglecting kinetic energy and potential energy ...

$$\dot{Q}_{C.V.} + \sum \dot{n}_i \bar{h}_i = \sum \dot{n}_e \bar{h}_e + \dot{W}_{C.V.}$$

Certainly, the CV system does not operate under SSSF conditions over a heart beat. The ventricles produce flow intermittently and the condition of the system fluctuates. But over a "sufficiently" long time, the CV system approaches a SSSF condition. It is not necessary to limit the thermodynamic analysis to SSSF conditions, but it helps to focus on the important issues, by taking this simpler approach. Just as Poiselle's law (valid for steady flow conditions) has great utility in explaining nonsteady hemodynamics, the authors believe that a SSSF approach can be of great utility in explaining CV thermodynamics as well.

<u>2nd Law of Thermodynamics for a control volume (general form):</u>

$$\frac{d}{dt} \int_V s\rho dV + \int_A s\rho v_{r.n.} dA = \int_A \left(\frac{\dot{Q}/A}{T} \right) dA + \int_V \left(\frac{\dot{L}W}{T} \right) dV$$

Where:

s ≡ entropy
T ≡ Temperature
$L\dot{W}$ ≡ lost work due to irreversibility

In words, the second law states that for a control volume the ...

rate change in entropy + net entropy leaving = entropy due to heat transfer + entropy generation

For a SSSF process ...

$$\sum \dot{m}_e s_e - \sum \dot{m}_i s_i - \int_A \left(\frac{\dot{Q}_A}{T} \right) dA = \sigma$$

Where:

σ ≡ rate of entropy generated
s_e ≡ entropy exiting the control volume
s_i ≡ entropy leaving the control volume

Continuity Equation (general form):

The continuity equation for a control volume is:

$$\frac{d}{dt}\int_v \rho dV + \int_A \rho \upsilon_{r.n} dA = 0$$

For SSSF conditions:

$$\sum \dot{m}_i = \sum \dot{m}_e$$

Where:

$$\dot{m} = \rho A \upsilon$$

Now these SSSF equations can be applied to the CV system and in the next section the CVC model will be developed.

The CV system is thermodynamically represented by four compartments. These are 1) left heart, 2) systemic circulation, 3) right heart, and 4) pulmonic circulation. The CV system schematic is shown below. A control surface is drawn around the outer surface of the heart.

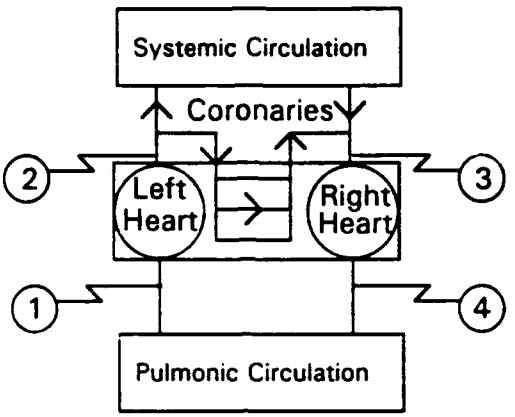


Figure 1 - CV system schematic

While the coronary flow is part of the systemic circulation, it is included with the heart as a separate flow from the "systemic circulation" compartment so that the overall thermodynamics of the heart can be determined. Note that the systemic circulation can be subdivided into individual organ systems so that the organ thermodynamics can be determined in a similar manner as the heart.

Now that the CVC has been thermodynamically represented in schematic form, it can also be represented on a fluid property diagram, such as a temperature-entropy (TS) diagram. However, unlike a common fluid such as water, blood is not truly a thermodynamically pure substance. A thermodynamically pure substance is one that has a homogeneous and invariable chemical composition and blood is actually a suspension. Also, since blood carries reacting and reacted components in it, blood cannot be considered a simple compressible substance. A simple compressible substance is one where knowing two independent variables, say pressure and specific volume, a third variable is automatically known, say temperature, through an equation of state or property chart. Unfortunately, to the authors' knowledge, an equation of state or property chart does not exist for blood. But later it will be shown, that to an engineering degree of accuracy, under certain conditions, blood may be considered a thermodynamically pure substance and a simple compressible substance. Once an equation of state or property chart for blood has been developed, the analysis suggested in this report will become much more straightforward.

Consider the following model of the heart as two pumps with the energy for the pumps derived from a fuel/oxygen reaction taking place in the cardiac muscle (supplied by the coronaries).

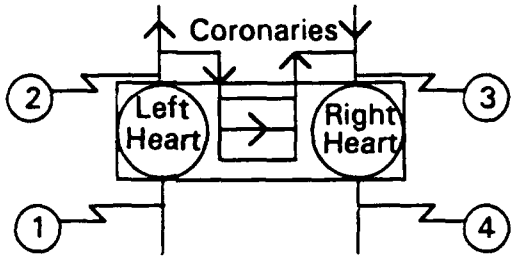


Figure 2 Left and Right Heart Schematic

Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity Equation (Heart):

Conservation of mass yields the following equations.

$$\dot{m}_1 = \dot{m}_2 + \dot{m}_C$$

$$\dot{m}_3 + \dot{m}_C = \dot{m}_4$$

$$\dot{m}_4 = \dot{m}_1, \quad \therefore \quad \dot{m}_2 = \dot{m}_3$$

1st Law of Thermodynamics:

Application of the 1st Law produces ...

$$\dot{Q} + \sum_R \dot{N}_i \left[h_f^\circ + \Delta h\right]_i = \dot{W} + \sum_P \dot{N}_i \left[h_f^\circ + \Delta h\right]_i$$

Where, R = reactants and P = products

Expanding the summation signs and assuming heat and work transfer rates are negligible across the outer surface of the heart ...

$$\dot{N}_{c_2}\left[h_f^\circ + \Delta h\right]_{c_2} + \dot{N}_1\left[h_f^\circ + \Delta h\right]_1 + \dot{N}_3\left[h_f^\circ + \Delta h\right]_3 = \dot{N}_{c_3}\left[h_f^\circ + \Delta h\right]_{c_3} + \dot{N}_2\left[h_f^\circ + \Delta h\right]_2 + \dot{N}_4\left[h_f^\circ + \Delta h\right]_4$$

where:

c_2 ≡ coronary inf low

c_3 ≡ coronary outflow

Substituting the following ...

$$\dot{m}_C HV = \dot{N}_{c_2}[h_f^\circ]_{c_2} - \dot{N}_{c_3}[h_f^\circ]_{c_3}$$

$$\dot{m}_C \Delta h = \dot{N}_{c_3}[\Delta h]_{c_3} - \dot{N}_{c_2}[\Delta h]_{c_2}$$

$$\dot{m}_{LH}\,\Delta h_{LH} = \dot{N}_2[\Delta h]_2 - \dot{N}_1[\Delta h]_1, \text{ where } \dot{N}_1[h_f^\circ]_1 = \dot{N}_2[h_f^\circ]_2$$

$$\dot{m}_{RH}\,\Delta h_{RH} = \dot{N}_3[\Delta h]_3 - \dot{N}_4[\Delta h]_4, \text{ where } \dot{N}_3[h_f^\circ]_3 = \dot{N}_4[h_f^\circ]_4$$

and where under SSSF conditions ...

$$\dot{m}_{LH} = \dot{m}_{RH} = \dot{m}_H.$$

Yields ...

$$\dot{m}_C\, HV = \dot{m}_C \Delta h_C + \dot{m}_H\,(\Delta h_{RH} + \Delta h_{LH})$$

At this point, it may be helpful to view the coronary flow and its thermodynamic equations by examining the schematic in Figure 3. Here Q_2 is the heat that is lost either across the outer surface of the heart or that which is lost to the blood in the atria or ventricles. Since the control surface was drawn at the outside surface of the heart and Qc.v. = 0, Q_2 is heat transferred to the blood in the atria and the ventricles. Practically, this is probably very small compared to the heat transferred to the coronary blood, Q. W is the work performed by the heart on the blood in the atria and ventricles.

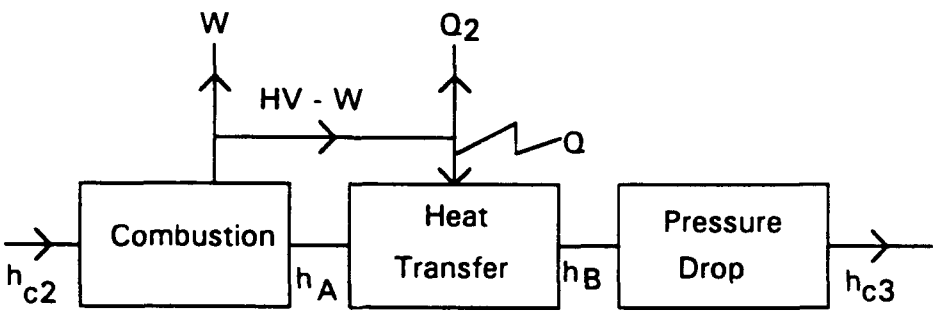


FIGURE 3 Coronary Flow Schematic

If one assumes that state point A is at the same temperature and pressure as state point c2 then $h_{c2} - h_A = HV$. It is also apparent that $HV = W + Q_2 + Q$. State point B may be at a higher temperature than A, so that $h_B - h_A = c\Delta T_{BA}$ where c is specific heat of coronary blood. Block 3 accounts for any pressure drop in the coronary flow.

Thus,

$$\Delta h_{C3,C2} = \Delta h_{A,C2} + \Delta h_{B,A} + \Delta h_{C3,B}$$

and

$\Delta h_{C3,B} = 0$, Throttling process

$\Delta u_{C3,B} = -\upsilon \Delta P_{C3,B}$

$$\Delta h_{C3,C2} = -HV + \Delta h_{B,A}$$

$$\Delta h_{C3,C2} = -HV + c\Delta T_{B,A}$$

$$\Delta h_{C3,C2} = -HV + c\Delta T_{C3,C2}$$

$$\Delta h_C = -HV + c\Delta T_C$$

$$\dot{m}_C \Delta h_C = -\dot{m}_C HV + \dot{m}_C C \Delta T_C$$

and with substitution and rearrangement

Representing the total enthalpy increase of both left and right hearts as ...

$$\Delta h_H = (h_2 - h_1) + (h_4 - h_3)$$

$$(h_2 - h_1) = c\Delta T_{LH} + v\Delta P_{LH}$$

$$(h_4 - h_3) = c\Delta T_{RH} + v\Delta P_{RH}$$

next applying the 1st law to all the flows of the heart one obtains...

$$\dot{m}_H v(\Delta P_{LH} + \Delta P_{RH}) = \dot{m}_C HV - \dot{m}_C c\Delta T_C - \dot{m}_H \left(c(\Delta T_{LH} + \Delta T_{RH}) \right)$$

Compare this equation with the equation below.

$$\dot{W} = \dot{m}_C HV - Q - Q_2.$$

This equation shows that the power output of the heart (left side of equation) is the total heating value of the fuel less some energy loss terms. These loss terms represent the net power change in the coronary flow due to heating from waste heat of combustion and pressure drop, and the net change in energy of the blood due to temperature increase of the blood as it is pressurized in the heart ventricles. If there were no loss terms, all the fuel could be converted to power output of the heart. But because of thermodynamic energy losses when energy is transformed from one form to another, such as metabolism, friction, and heat transfer, the power output is less than the energy content in the fuel.

Second Law of Thermodynamics:

$$\sigma = \sum_{out} \dot{m}s - \sum_{in} \dot{m}s - \sum_{i} \frac{Q_i}{T_i}$$

Assuming that ...

$$\sum_{i} \frac{Q_i}{T_i} = 0, \text{ because } Q_2 << Q$$

$$\sigma = \dot{m}_C(s_{C3} - s_{C2}) + \dot{m}_H(s_2 + s_4 - s_1 - s_3)$$

$$\sigma = \dot{m}_C(\Delta s_C) + \dot{m}_H(\Delta s_{LH} + \Delta s_{RH})$$

Recall, the TΔS formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_{i} \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_{i} \mu_i \Delta N_i$$

$$\sum_{i} \mu_i \Delta N_i = \Delta G)_{T,P}$$

where G is the Gibbs' free energy function...

Thus,

$$T\Delta s_{LH} = \Delta h_{LH} - v\Delta P_{LH} = c\Delta T_{LH}$$

$$T\Delta s_{RH} = \Delta h_{RH} - v\Delta P_{RH} = c\Delta T_{RH}$$

$$T\Delta s_C = \Delta h_{C3,C2} - v\Delta P_{C3,C2} - \sum_{i} \mu_i \Delta N_i = -HV + v\Delta P_{C2,C3} + \Delta G)_{T,P}$$

$$\sigma = -HV + \Delta G)_{T,P} + \upsilon\Delta P_{C2,C3} + c\Delta T_{LH} + c\Delta T_{RH} + c\Delta T_c$$

Since

$$\Delta G)_{T,P} \equiv W_{rev}$$

and

$$W_{act} = HV - C\Delta T_c, \text{ assuming } Q_2 \approx 0$$

$$\sigma_c = W_{rev} - W_{act} + \upsilon\Delta P_{C2,C3}$$

Lost Work (LW) = W_{rev}-W_{act}

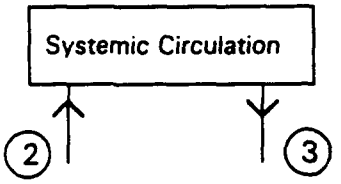
$$\sigma_c = LW + \upsilon\Delta P_{C2,C3}$$

σ_c = Lost work due to combustion inefficiency and coronary flow pressure drop.

$$\sigma = \sigma_c + c(\Delta T_{LH} + \Delta T_{RH})$$

SYSTEMIC CIRCULATION

The analysis for the systemic circulation follows closely the analysis for the coronary flow. First, a model will be defined.



Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity equation (systemic Circulation)

Conservation of mass yields the following equation.

$$m_3 = m_2$$

First Law Analysis (systemic circulation)

At this point, it may be helpful to view the systemic flow and its thermodynamic equations by examining the schematic in Figure 4. Here Q_{SKIN} is the heat that is lost to the outside tissue and Qs is the heat that is returned to the capillary across the control surface. Wc.v. = 0.

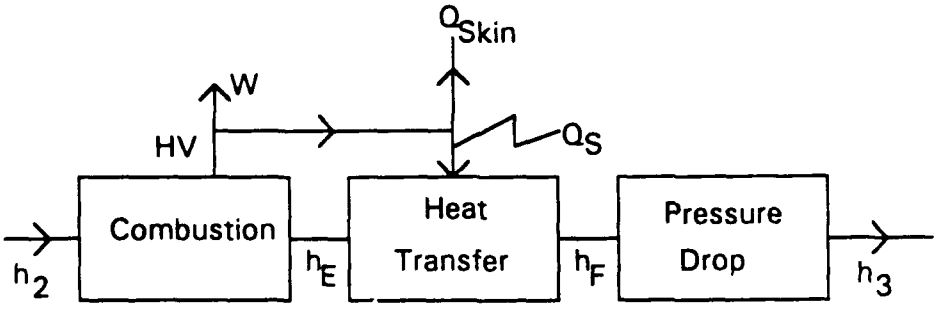


FIGURE 4 Systemic Block Diagram

If one assumes that state point E is at the same temperature and pressure as state point 2 then $h_2 - h_E = HV$. State point F may be at a higher temperature than E, so that $h_F - h_E = c\Delta T_{FE}$. Block 3 accounts for any pressure drop in the systemic bed.

Thus,

$$\Delta h_{32} = \Delta h_{E2} + \Delta h_{FE} + \Delta h_{3F}$$

$$\Delta h_{3F} = 0, \text{ Throttling Process}$$

$$\Delta u_{3F} = -v\Delta P_{3F}$$

$$\Delta h_{32} = -HV + \Delta h_{FE}$$

$$\Delta h_{32} = -HV + c\Delta T_{FE}$$

and with substitution and rearrangement

$$\dot{m}_H \Delta h_{32} = m_H(-HV + c\Delta T_{32})$$

This equation shows that the enthalpy change is the heating value of the fuel less the heat returned to the systemic circulation.

Second Law Analysis

$$\sigma = \sum_{out} ms - \sum_{in} ms - \sum_{i} \frac{Q_i}{T_i}$$

$$\sigma = m_H(\Delta s_{32}) - \frac{Q_s}{T}$$

where Q_S is the heat transferred across the circulatory system vessels.

Recall, the $T\Delta S$ formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_i \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_i \mu_i \Delta N_i$$

$$\sum_i \mu_i \Delta N_i = \Delta G)_{T,P}$$

Thus,

$$T\Delta s_{32} = \Delta h_{32} - v\Delta P_{32} - \sum_i \mu_i \Delta N_i = -HV + c\Delta T_{32} + v\Delta P_{23} + \Delta G)_{T,P}$$

$$\sigma = \dot{m}_s\left(-HV + c\Delta T_{32} + v\Delta P_{23} + \Delta G)_{T,P}\right) - \frac{Q_s}{T}$$

$$\sigma = \dot{m}_s\left(-HV + \Delta G + v\Delta P_{23})_{T,P}\right)$$

PULMONARY CIRCULATION:

Shown below is the schematic of the last component of the CVC system.

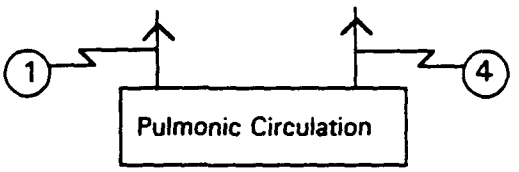
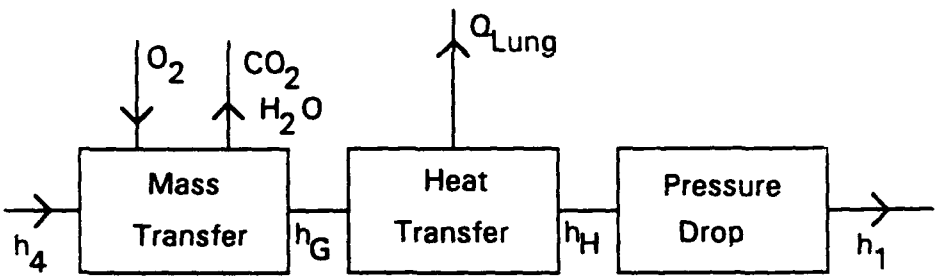


Figure 5 Pulmonary Circulation Schematic

The control surface for the control volume was drawn around the pulmonary circulation. It was assumed that the boundary work done on the pulmonary circulatory system is much smaller than the amount of heat transferred across the control surface. Given these constraints, a thermodynamic block diagram similar to the systemic circulation was used.



The first block represents products of metabolism entering with enthalpy h_4. The fluid gains oxygen (a reactant). For simplicity, it was assumed that the blood also acquires fuel at this point. While physiologically, the fuel is acquired by a number of methods and anatomical locations, thermodynamically it is a good approximation to have the fuel enter with the oxygen. Physiologically, this is termed the energy equivalent of oxygen and is about 4.8 Calories/L oxygen. Furthermore it was assumed that state 4 and state G are at the same temperature and pressure. State H is typically at a lower temperature than G and

state 1 is at the same temperature as H but a lower pressure.

Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity equation (Pulmonary Circulation)

Conservation of mass yields the following equation.

$$m_4 = m_1$$

First Law Analysis (Pulmonary circulation)

$$\Delta h_{1H} = 0, \text{ Throttling Process}$$
$$\Delta u_{1H} = -v\Delta P_{1H} = v\Delta P_{41}$$

Thus,

$$\Delta h_{14} = \Delta h_{G4} + \Delta h_{HG} + \Delta h_{1H}$$

$$\Delta h_{G4} = HV$$

$$\Delta h_{14} = HV + \Delta h_{HG}$$

$$\Delta h_{14} = HV + c\Delta T_{HG} = HV + c\Delta T_{14}$$

and with substitution and rearrangement

$$\dot{Q}_{LUNG} = m_H \, c\Delta T_{14}$$

This assumes that most of the temperature change is due to heat-transfer with the pulmonary gases and not from metabolic waste heat from the lung tissue.

Second Law Analysis (Pulmonary Circulation):

$$\sigma = \sum_{out} \dot{m}s - \sum_{in} \dot{m}s - \sum_{i} \frac{Q_i}{T_i}$$

$$\sigma = m_H(\Delta s_{14}) - \frac{Q_{LUNG}}{T}$$

Recall, the TΔS formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_i \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_i \mu_i \Delta N_i$$

$$\sum_i \mu_i \Delta N_i = \Delta G)_{T,P}$$

Thus,

$$T\Delta s_{14} = \Delta h_{14} - v\Delta P_{14} - \sum_i \mu_i \Delta N_i = HV + c\Delta T_{14} + v\Delta P_{14} + \Delta G)_{T,P}$$

$$\sigma = \dot{m}_H \left(HV + c\Delta T_{14} + v\Delta P_{14} + \Delta G)_{T,P} \right) - \frac{Q_{LUNG}}{T}$$

$$\sigma = \dot{m}_H \left(HV + v\Delta P_{14} + \Delta G)_{T,P} \right)$$

CONCLUSIONS

The CV system was modeled as a thermodynamic system to investigate the theoretical foundations of irreversibility minimization as a control strategy for the circulatory system. Simplifying assumptions, such as SSSF conditions, were imposed. As one might suspect, irreversibility was generated in each component from fluid flow pressure drops, metabolic processes and mass transfer. Although much more work needs to be done, the foundations of the concept have been developed.

REFERENCES

Suga H, Igraashi Y, Yamada O, Goto Y: Mechanical efficiency of the left ventricle as a function of preload, afterload and contractility. Heart Vessel 1:3-8, 1985

Hayashida K, Sunagawa K, Noma M, Sugimachi M, Ando H, Nakamura M: Mechanical Matching of the left ventricle with the arterial system in exercising dogs. Circulation Research 71:481-489, 1992

Gibbs C, Chapman J: Cardiac Energetics. In: Bern RM, Sperelakis N, Geiger S, eds.: Handbook of Physiology. Sec 2: The Cardiovascular system, Vol. 1: The Heart. Bethesda: American Physiology Society, 1979, pp 775-804.

Black W and Hartley J: Thermodynamics, 2nd ed, Hraper Collins, 1991

Van Wylen G and Sonntag R: Fundamentals of Classical Thermodynamics, 3rd ed, John Wiley and Sons, 1991

WHITE-NOISE ANALYSIS OF CAROTID BARORECEPTOR FUNCTION IN BABOONS

Arthur J. Koblasz, PhD
Associate Professor
School of Civil Engineering

Georgia Institute of Technology
Atlanta, GA 30332-0355

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 1993

WHITE-NOISE ANALYSIS OF CAROTID BARORECEPTOR FUNCTION IN BABOONS

Arthur J. Koblasz
Associate Professor
School of Civil Engineering
Georgia Institute of Technology

Abstract

A white-noise protocol was evaluated for characterizing carotid baroreceptor function in three adult male baboons. The white-noise (pseudo-random binary) stimulus was created by varying the pressure in the right carotid sinus. The pseudo-random stimulus was sustained for a period of 3 minutes at 3 different mean levels-- 40 mm Hg, 70 mm Hg and 100 mm Hg. The baroreceptor response at each mean level was indicated by continuously measuring the pressure in both the right atrium and the aorta. First-order Wiener Kernels were calculated from this stimulus-response data (see Appendix A for analytical details). The first-order Wiener Kernels were then used to predict the pressure changes in the right atrium and aorta which would result from a pulse of pressure at the right carotid sinus.

The white-noise protocol reveals that the carotid baroreceptors are still effecting right atrium pressure more than 50 seconds after the pressure in the right carotid has been pulse modulated. If our predicted responses are correct, then the time-constants for g-induced cardiovascular changes can be much longer than expected. Also, the delayed response is in the wrong direction-- a positive pulse at the carotid sinus causes vasoconstriction.

These unexpected results could be the result of reduced pO2 at the right carotid sinus, which is a common situation in prolonged air combat. Other investigators have reported that the carotid baroreceptor response is entirely eliminated under similar conditions. In future experiments, we will investigate the effect of pO2 on carotid baroreceptor transduction speed. We will also use aperiodic pulse stimuli, i.e. random intervals between the pulses, and will continue to measure the response to each pulse stimulus for 100 seconds after each stimulus.

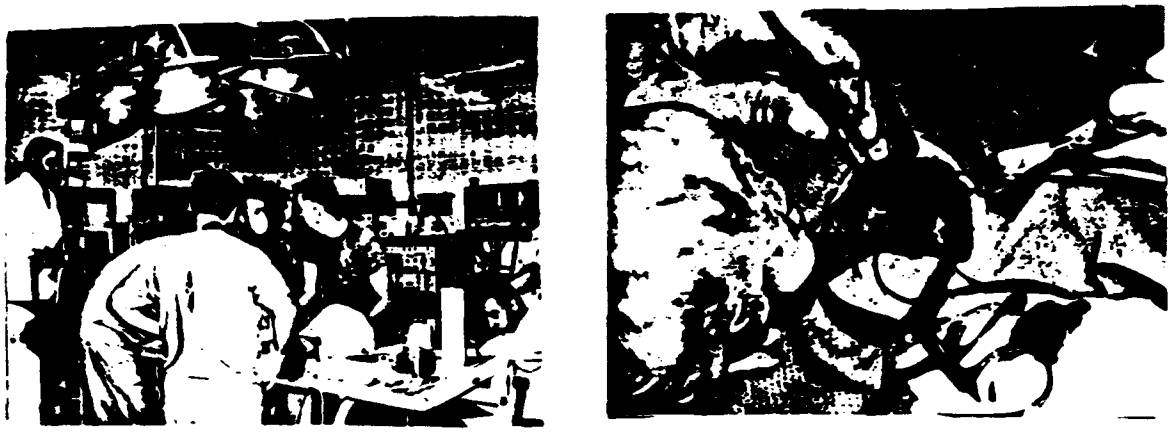


Figure 1.

Background

The white-noise method for characterizing nonlinear biological systems has been successfully demonstrated in catfish retina studies (Marmarelis and Naka, 1973), in human ERG studies (Koblasz, et al, 1980) and in a variety of other vertebrate studies (Marmarelis, 1976). Similar nonlinear characterization schemes have been demonstrated using random square wave stimuli (Yasui and Koblasz, 1984) and random pulse sequences (Fricker and Sanders, 1975).

Recently, a white-noise protocol has been demonstrated in a study of aortic baroreceptor function using in vivo rabbit preparations (Masaru, et. al., 1990). The aortic pressure was modulated by electrically stimulating the right ventricle using pacing electrodes triggered at a constant frequency of 400 beats per minute. The pacing was sustained for variable durations of greater than 1 second per burst, and the interval between bursts was varied to produce aortic pressure fluctuations with a fairly flat power spectrum over the frequency range of DC to 1/2 Hz. (In our protocol, we will block the normal pulsatile flow into the left carotid artery and will then modulate the left carotid pressure to produce a stimulus power spectrum which is very flat over the frequency range of DC to 5 Hz.)

In Masaru's protocol, the aortic pressure was measured using a high-fidelity micromanometer (Millar MPC-500), and the baroreceptor response was measured using Ag/AgCl bipolar electrodes positioned at the distal end of the (desheathed) left aortic depressor nerve. Linear transfer functions were calculated using these stimulus/response data. The resulting math model characterized the combination of aortic wall mechanics followed by neural transduction and encoding mechanisms.

EXPERIMENTAL PROTOCOL

Three adult male baboons were anesthetized with Ketamine (30mg IM) followed by periodic doses of a-Chloralose (50mg/kg for the first dose and 20 mg/hr thereafter). Pancuronium Bromide (.1mg/kg IV) was also given to each animal to reduce muscle activity. A 3-French Millar pressure transducer (single-tipped) was inserted into the right femoral artery to record systemic pressure changes resulting from the pressure stimuli applied to the right carotid sinus. EKG was also recorded to indicate changes in heartrate. Figure 1 presents a collection of photographs which were taken during the animal surgery.

In the first set of experiments, the variations in carotid pressure were produced by randomly varying the flow rate of IV fluid injected into the common carotid artery. The flow rate was modulated using a stepping motor connected to a linear hydraulic valve, as depicted in Figure 2. At periodic intervals, equivalent volumes of blood were removed from the right femoral vein. Unfortunately, the relationship between the volume of blood injected into the carotid artery and the carotid sinus pressure is much more complicated than we had expected. Futhermore, it is difficult to maintain a constant total blood volume during a prolonged experiment.

In the second set of experiments, the right carotid sinus was isolated by ligating all incoming and outgoing vessels approximately 1 cm above and below the carotid sinus. A small catheter was then inserted into the carotid sinus, along with a 2-French Millar (single-tipped) pressure transducer. The catheter was connected to a Skinner three-way hydraulic valve (Type B14), which allowed the catheter to be switched between two reservoirs of warmed physiological saline solution. The reservoirs were positioned at different heights above the carotid sinus; hence, switching between the two reservoirs produced step changes in pressure inside the carotid sinus. Since the carotid sinus was completely isolated, the pressure shifts occurred nearly instantaneously with no significant flow of saline into or out of the sinus.

The circuit shown in Figure 3 was used to create a pseudo-random binary (two-level) signal and was also used to generate periodic step and pulse signals-- the latter being more conventional stimuli. The pseudo-random binary signal was used as a control input to the three-way hydraulic valve, thus producing pseudo-random binary pressure changes inside the carotid sinus. The more conventional step and pulse stimuli were also generated to validate the pseudo-random data (see report for Summer 1992).

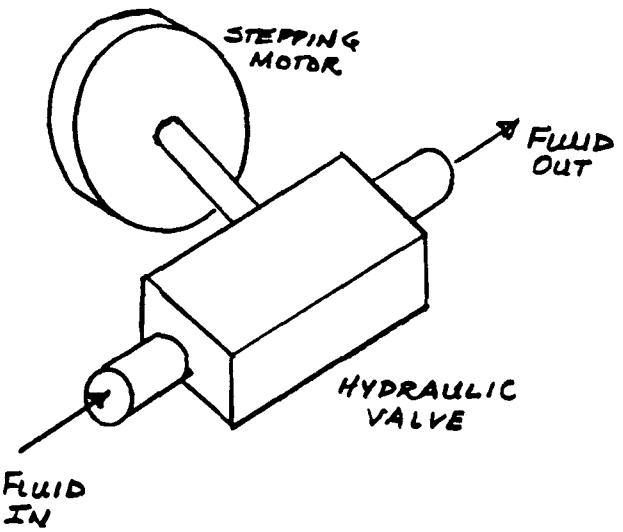


FIGURE 2
Linear hydraulic valve

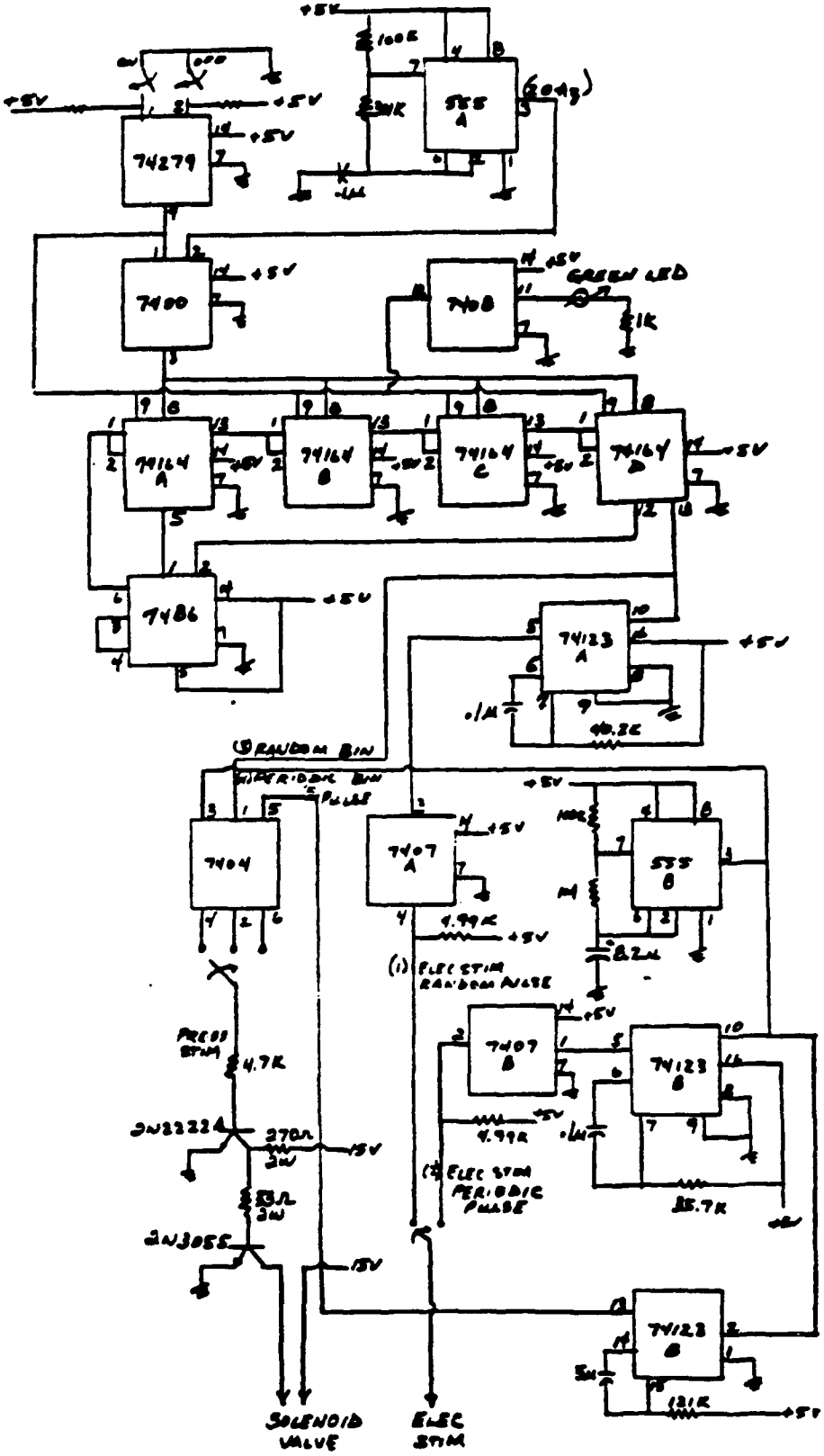


Figure 3
Electrical Circuit

During each period of data acquisition, the pseudo-random binary stimulus was sustained for three minutes . In the first run, the two reservoirs were positioned to provide a mean carotid pressure of 40mm Hg, and the random binary stimulus fluctuated about the mean by ± 20mm Hg. In the second run, the mean was raised to 70mm Hg, and the binary fluctuations were first set at ± 20mm Hg and then increased to ± 50mm Hg. In the last run, the mean was set at 100mm Hg and the binary fluctuations were ± 20mm Hg.

RESULTS

Figure 4 shows a plot of the carotid pressure (i.e. stimulus) during the first run on the third animal, when the mean was set at 40mm Hg, and the random binary fluctuations were + 20mm Hg. Figure 5 is a log-log plot of the power spectrum of this signal. The pressure stimulus appears to be flat over the bandwidth, DC to .75Hz.

Figure 6 presents a plot of the pressure measured in the right atrium (i.e. response) during the same period of time as Figure 4. The low frequency periodic component (approximately .2 Hz) is related to respiration, and the high frequency component (approximately 3 Hz) is the typical systolic/diastolic waveform. Figure 7 is a plot of the EKG interval during this same period.

Figure 8 displays the first-order Wiener kernels (see Appendix A for formula) for each of the animals at the same stimulus mean level of 70mm Hg. These transfer functions were calculated by defining the pressure in the carotid sinus as the stimulus and the pressure in the right atrium as the response. The response data was low-pass filtered (corner=1.5 Hz) to remove the systolic/diastolic fluctuations-- without attenuating the lower frequency, respiratory-dependent fluctuations. Figure 9 shows the comparable plots of the first-order kernels when the aortic pressure is considered to be the response.

Under certain conditions (see Appendix A for details), the above first-order Wiener kernels predict the response to a pulse stimulus occurring at the origin of each plot. In all of the kernels presented in Figures 8 and 9, the predicted responses occur more than 50 seconds after the hypothetical pulse stimulus. This unexpected long time-constant is undoubtedly the result of reduced pO2 in the carotid sinus. It is also surprising that the predicted pulse responses are in the wrong direction; a positive pulse at the carotid sinus causes vasoconstriction. Since the predicted negative pulse response is nearly the same, the carotid baroreceptors may be only responding to the negative-going edge of the stimulus.

Figure 10 shows how the first-order Wiener kernel changes when the mean pressure level in the carotid sinus is changed. At the higher mean levels, the latencies appear to be smaller. Hence, the longest time-constants occur at low pO2 and low mean carotid pressures, which is a common situation in prolonged high-G combat. Figure 11 presents these same Wiener kernels in a different perspective-- the peak values are shown for each respiratory period.

The obvious next step in the analytical progression is to calculate the average responses to the pulse and step stimuli which were used-- comparing the measured pulse/step responses with the predicted responses obtained from the white-noise data. Unfortunately, we were not expecting latencies as long as the above white-noise data indicates; therefore, the pulse and step stimuli were presented with short intervals-- an order of magnitude smaller than we needed. In future experiments it will be important to use aperiodic pulse and step stimuli with much longer intervals between stimuli.

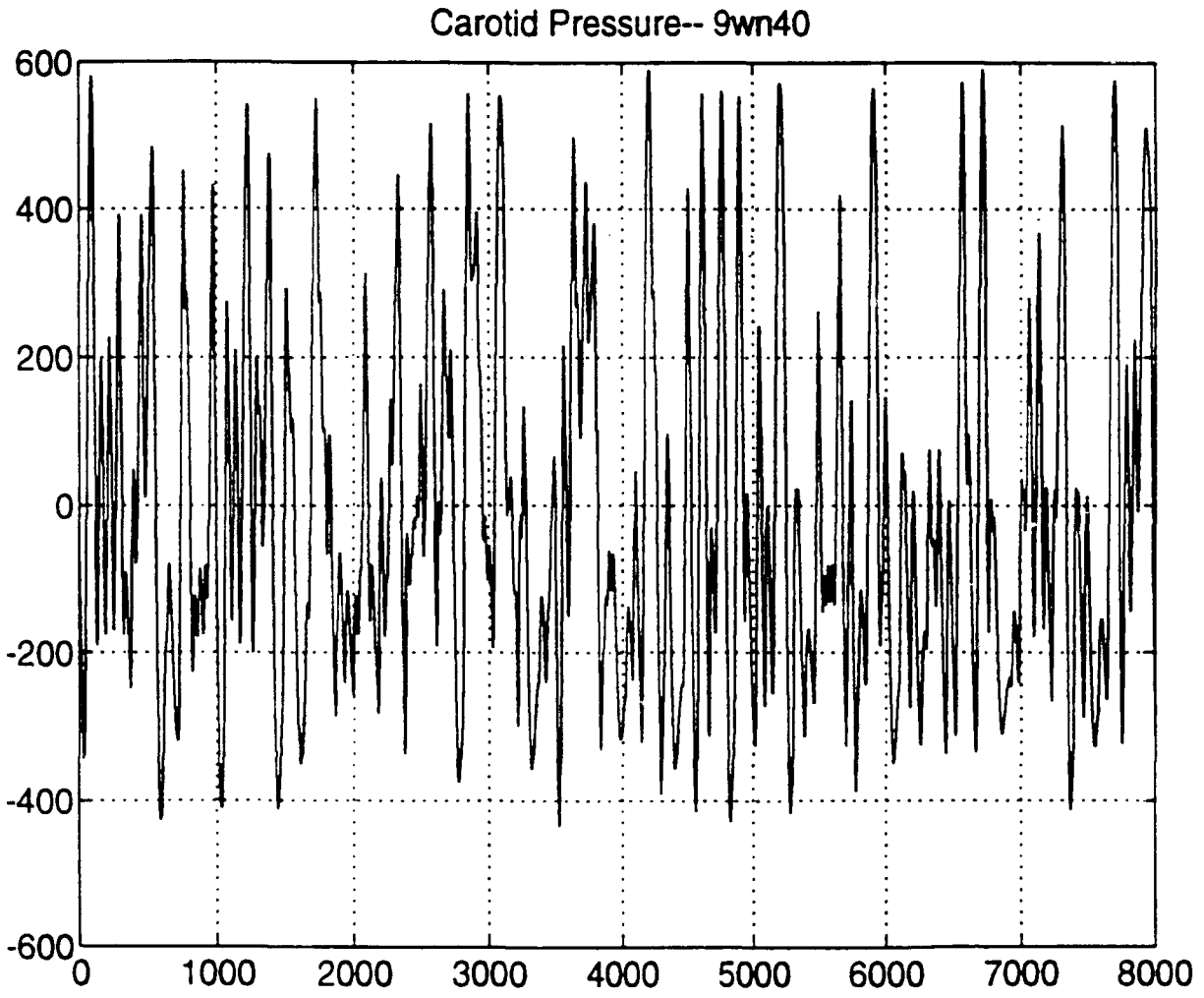


Figure 4
Carotid pressure measured during white-noise protocol

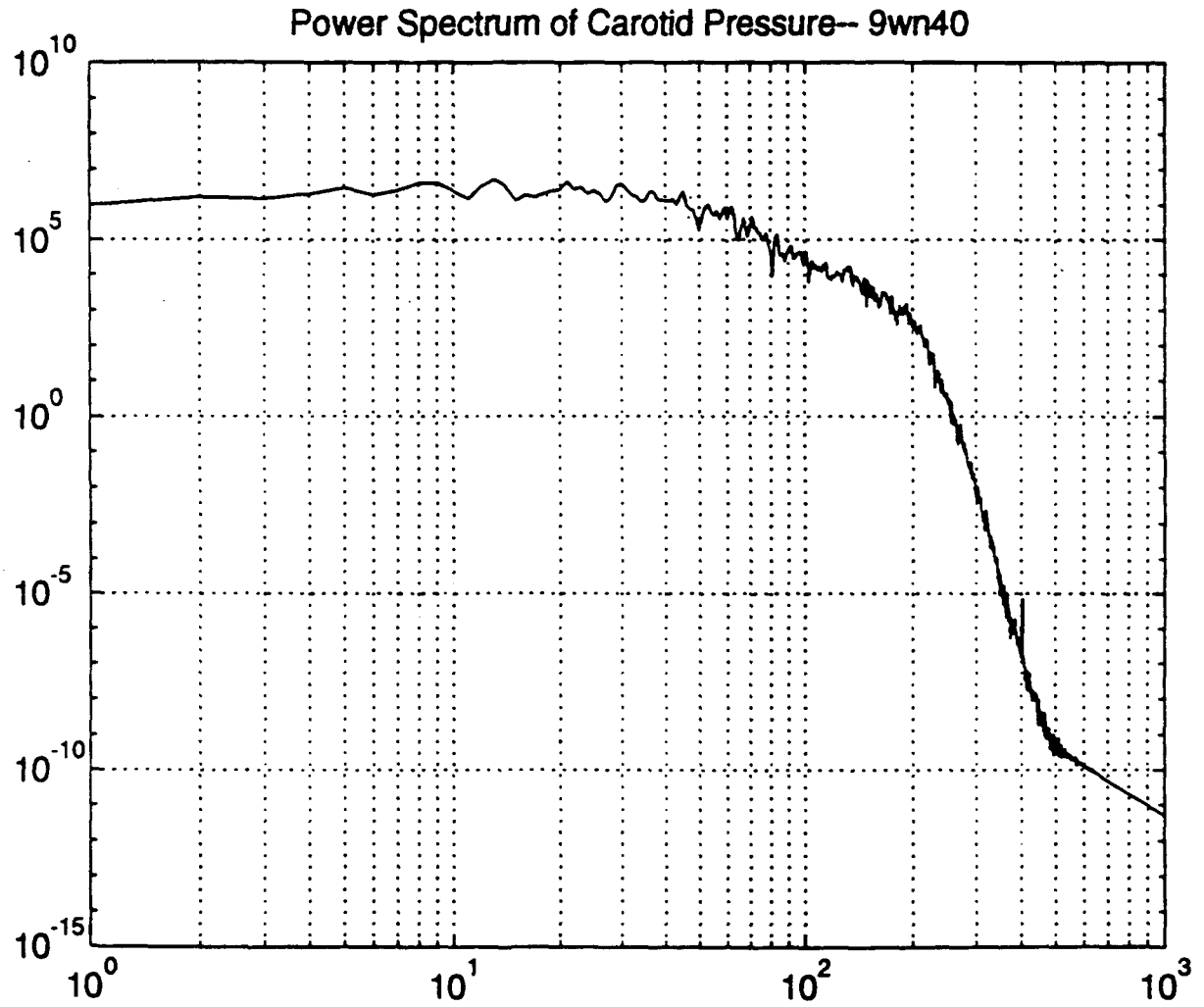


Figure 5
Power spectrum of white-noise stimulus

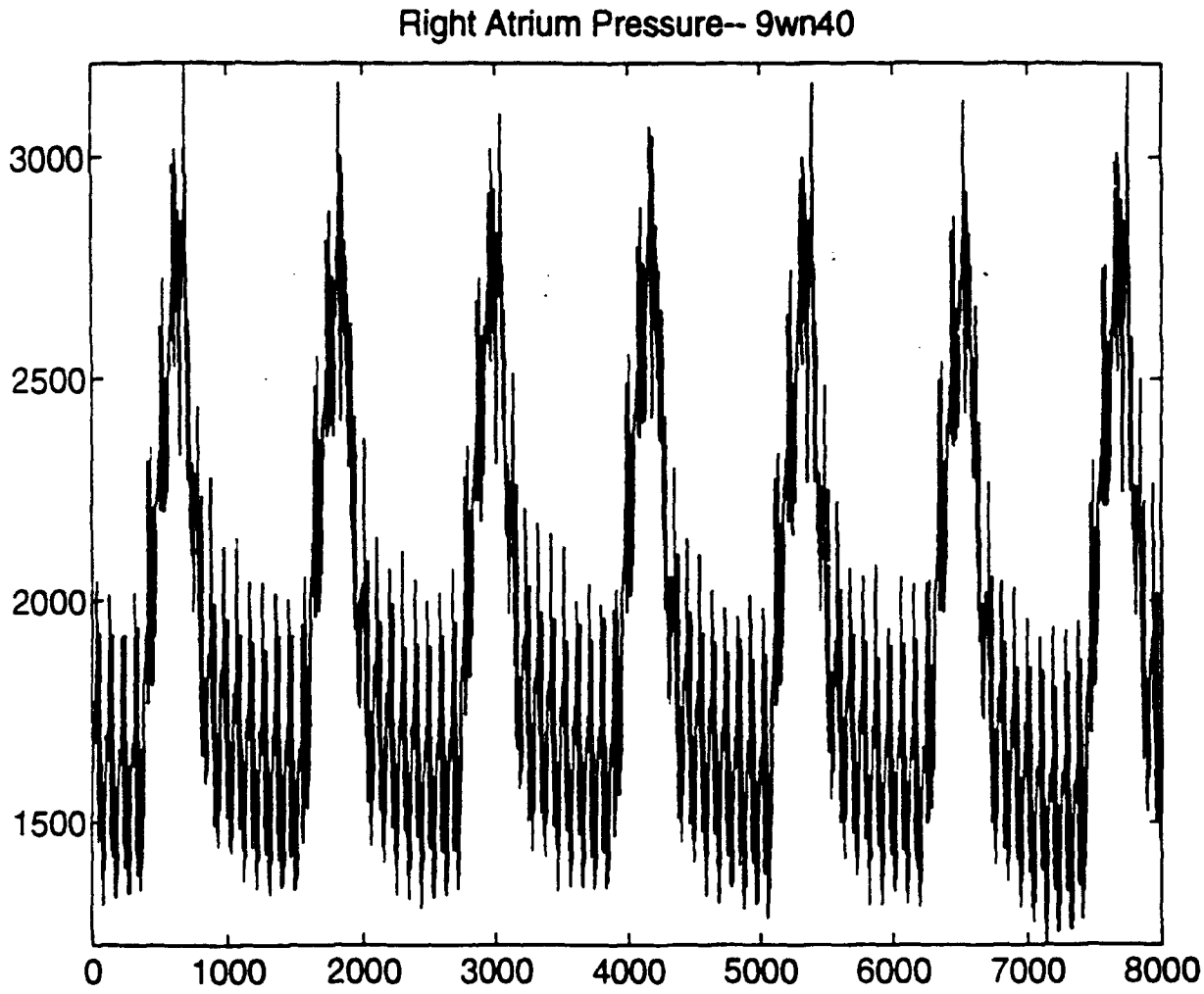


Figure 6
Pressure measured in Right Atrium during white-noise experiment

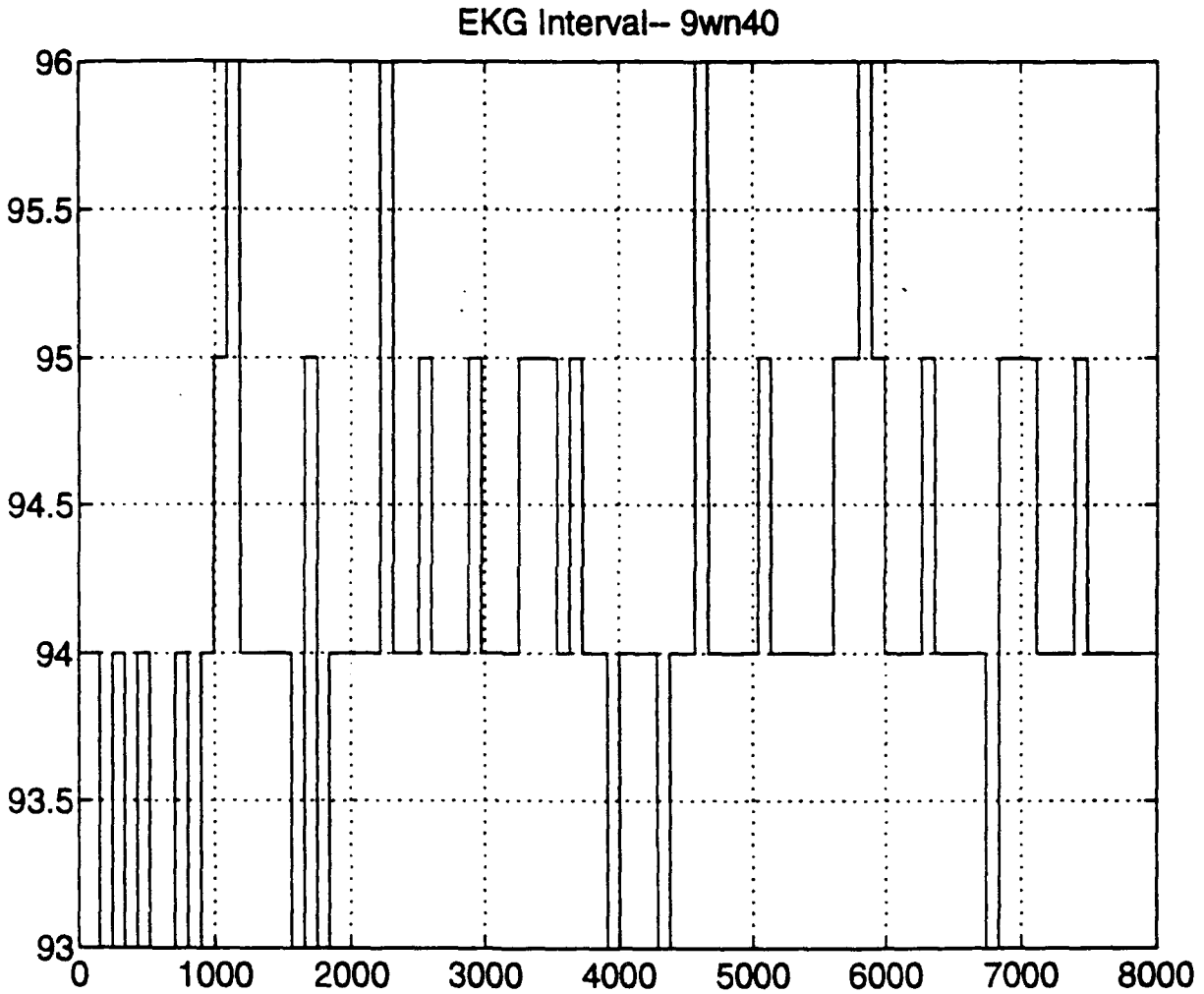


Figure 7
EKG interval measured during white-noise experiment

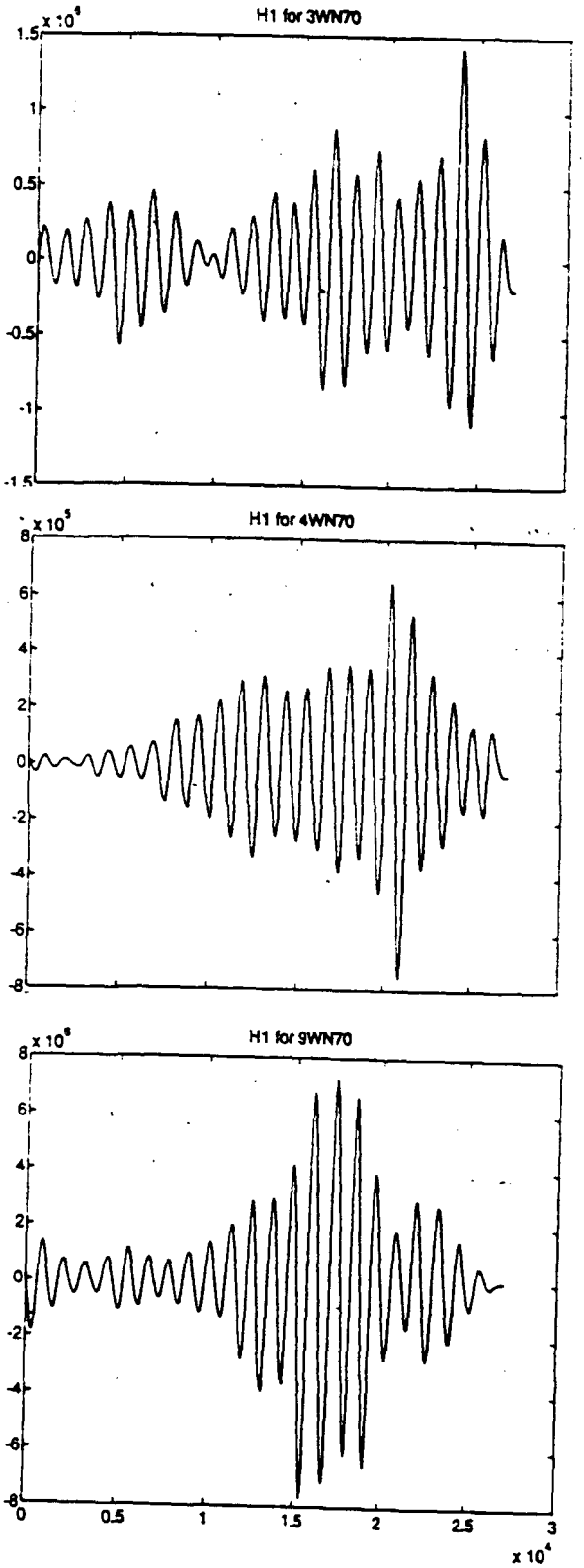


Figure 8
First-order Wiener kernels for 3 different baboons
at mean stimulus level of 70 mm Hg--
response measured at Right Atrium

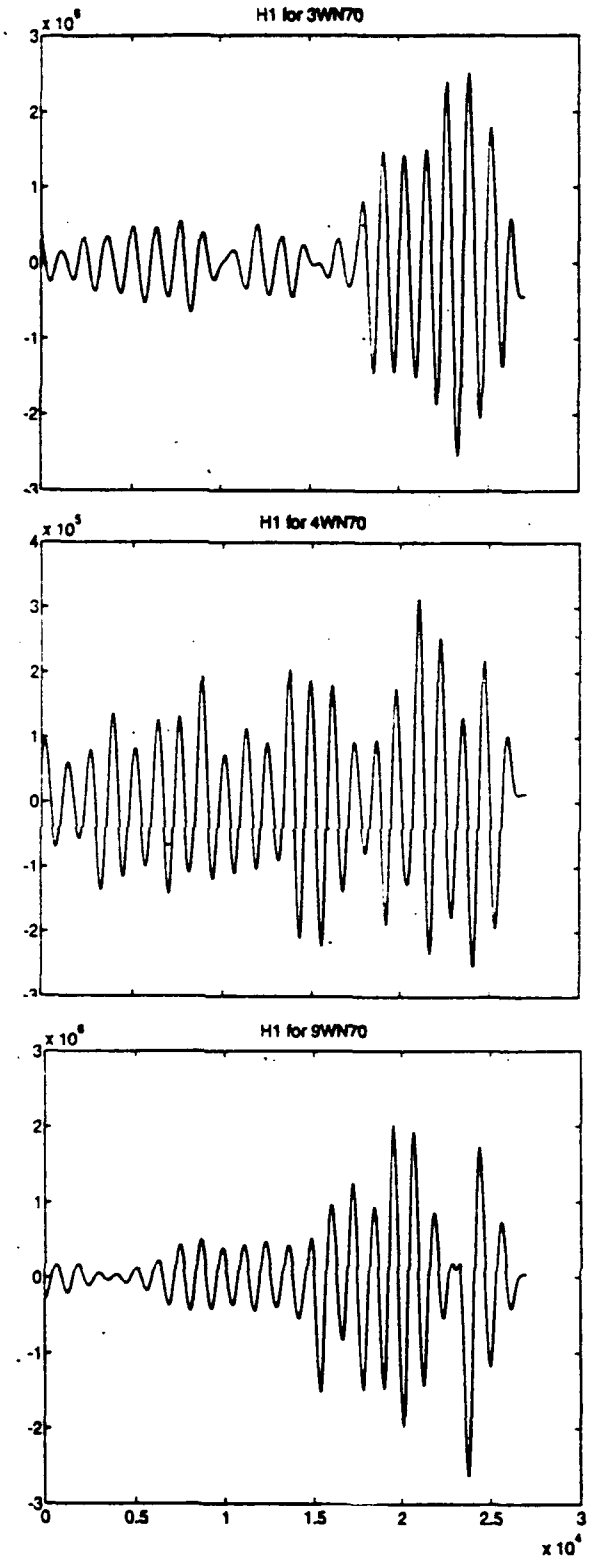


Figure 9
First-order Wiener kernels for 3 different baboons
at mean stimulus level of 70 mm Hg--
response measured at Aorta

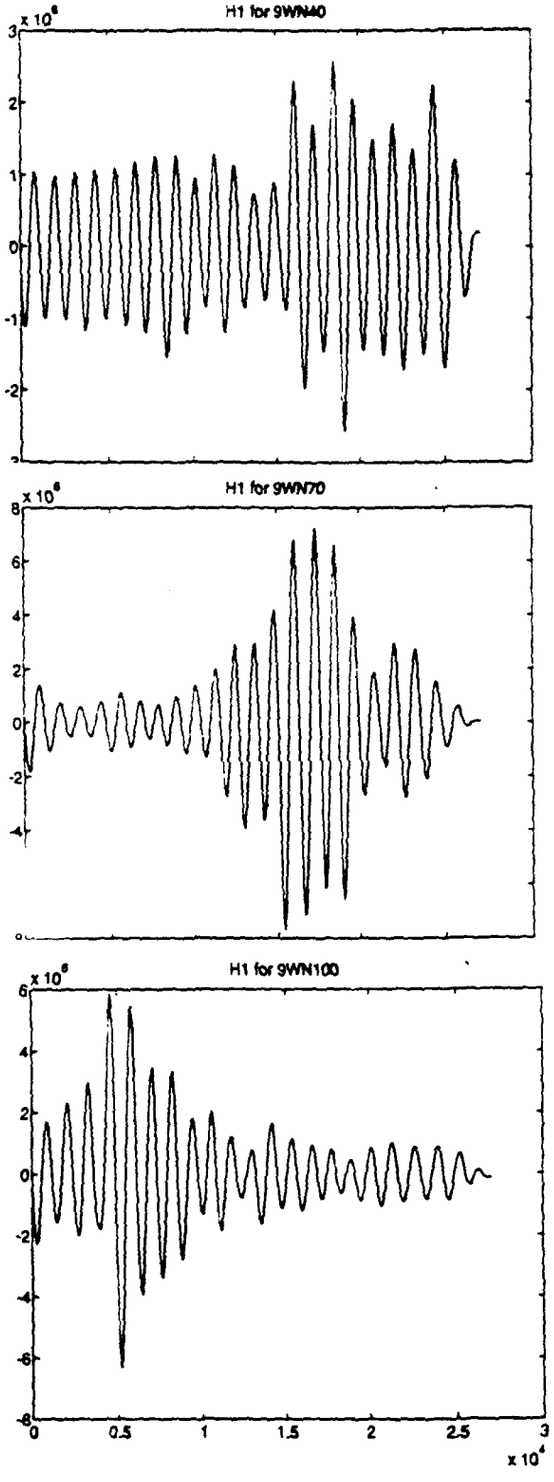


Figure 10
First-order Wiener kernels at 3 different mean levels--
40 mm Hg, 70 mm Hg and 100 mm Hg

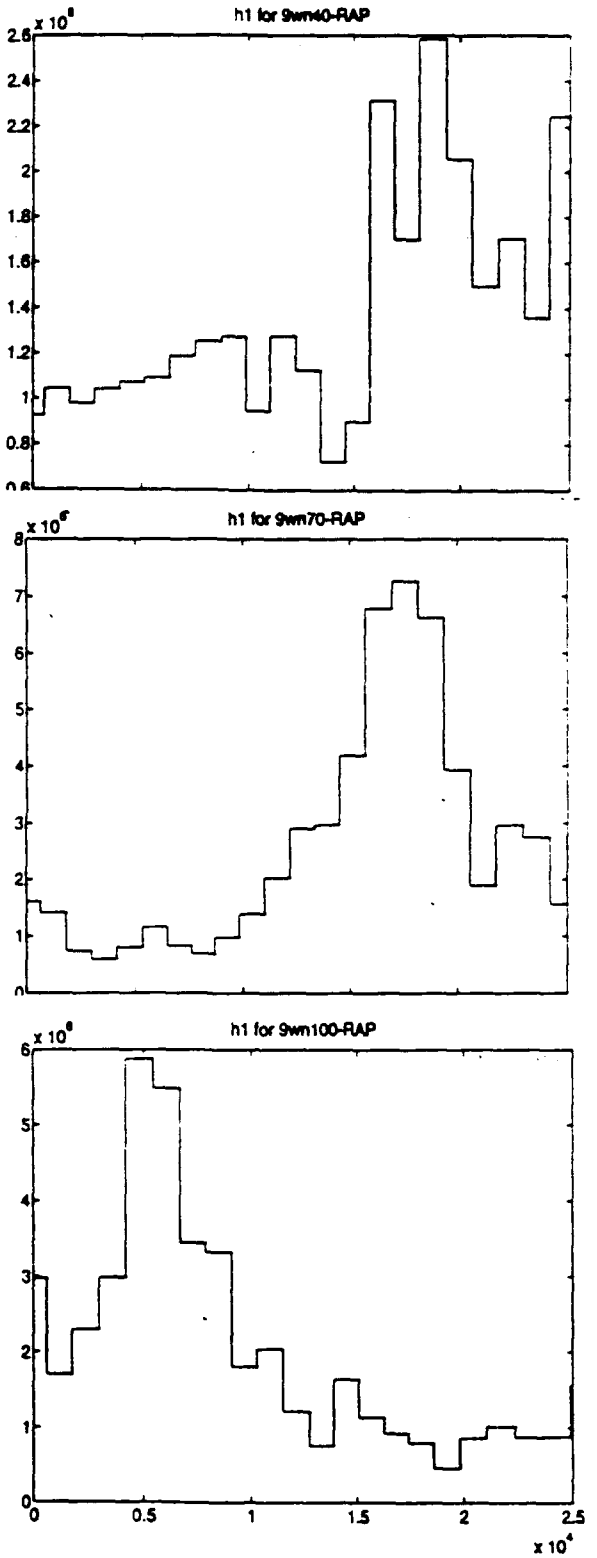


Figure 11
First-order Wiener kernels at 40 mm Hg, 70 mm Hg and 100 mm Hg--plotting peak pressures during each respiratory period

References

Andresen, M J Krauhs and A Brown. 1978. Relationship of aortic wall and baroreceptor properties during development in normotensive and spontaneous hypertensive rats. Circulation Research. 43:728-738.

Brown, A. 1980. Receptors under pressure. An update on baroreceptors. Circulation Research. 46:1-10.

Fricker, S and J Sanders. 1975. A new method of cone electroretinography: the rapid random flash response. Investigative Ophthalmology. 14:131-137.

Koblasz, A, J Rae, M Correia and M Ni. 1980. Wiener kernels and frequency response functions of the human retina. IEEE Biomedical Engineering. 27:68-75.

Marmarelis, P and K Naka. 1973. Nonlinear analysis and synthesis of receptive field responses in the catfish retina. J Neurophysiology. 36:619-653.

Marmarelis, V. 1976. Identification of nonlinear systems through quasi-white test signals. Caltech PhD Thesis.

Masaru, S, T Imaizumi, K Sunagawa, Y Hirooka, K Todaka, A Takeshita and M Kakamura. 1990. American Journal of Physiology. H887-H895.

Yasui, S and A Koblasz. 1984. Transmission of array data by multi-variate convolution and cross-correlation using white-noise reference signal. International Journal of Systems Science. 15:525-541.

INTERACTION BETWEEN NITRIC OXIDE AND OXYGEN DURING HYPERBARIC OXYGENATION

Edward H. Piepmeier, Jr.
Assistant Professor
College of Pharmacy

University of South Carolina
Columbia, SC 29208

Final Report for:
Summer Research Program
Armstrong Laboratory

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

August 1993

INTERACTION BETWEEN NITRIC OXIDE AND OXYGEN DURING HYPERBARIC OXYGENATION

Edward H. Piepmeier, Jr.
Assistant Professor
College of Pharmacy
University of South Carolina

Abstract

The interactions between oxygen, nitric oxide and some of the effects of these interactions on the immune system and inflammatory response were identified. Hyperbaric oxygenation treatment was used to alter oxygen levels both in chronic and acute studies in vivo in both a human and rats. Microwave treatment was used to induce a shock response in rats. In vitro studies were used to supplement information gained from in vivo studies.

The effects of hyperbaric oxygenation treatment and microwave exposure on mitogenic stimulation of lymphocytes, nitrate production of macrophages, and time to induction of shock and death were determined in rats. The effects of hyperbaric oxygenation treatment on CD4/CD8 surface protein ratios, mitogenic stimulation of lymphocytes, and nitrate production of macrophages was tested in a human. Finally, the mitogenic and toxic effects of nitric oxide, 3-amino-tyrosine and glucose on human lymphocytes were measured.

INTERACTION BETWEEN NITRIC OXIDE AND OXYGEN DURING HYPERBARIC OXYGENATION

Edward H. Piepmeier, Jr.

INTRODUCTION

Nitric oxide and oxygen both have vasopressive effects. Nitric oxide causes vasodilation and oxygen causes vasoconstriction. An imbalance in these molecules may be responsible for the lethal effects associated with shock. In shock, vasoconstriction is followed by vasodilation which may be followed by visceral blood pooling. The rapid reperfusion associated with this action can result in tissue damage and ultimately death.

The interaction between nitric oxide and oxygen is unclear. Superoxide complexes with nitric oxide to form the peroxynitrate ion. If superoxide dismutase levels are induced then the depletion of superoxide may result in increased concentrations of nitric oxide. Alternately nitric oxide may decrease due to an increase in superoxide resulting from depleted superoxide dismutase.

This study examined the effects of induction of the shock response with microwaves in conjunction with hyperbaric oxygenation therapy in rats. Rats were monitored for changes in time to shock and time to death following exposure to 350 GHz both with and without prior hyperbaric oxygenation treatment. Changes in response of the mitogenic response of rat lymphocytes was monitored. The production of nitrates by macrophages was also measured.

Human lymphocytes were stimulated with different amounts of nitric oxide or 3-amino tyrosine to identify a mitogenic response which may be associated with elevated levels of each molecule. The activity of lymphocytes and macrophages in a hyperbaric oxygenation patient were observed before and after therapy. Changes in the CD4/CD8 receptor ratio, mitogen induction, of lymphocytes and nitrate production by macrophages were monitored.

METHODS & RESULTS

All cells were grown in RPMI-1640 with 10% fetal bovine serum and without phenol red. Concanavalin A was made up at 3.0 mg/100 ml of RPMI and diluted to 6μg/ml in each well for determination of the effects of mitogenic stimulation on lymphocytes. Nitric oxide was measured indirectly by monitoring its spontaneous oxidation products NO_2^-, and NO_3^-. One hundred microliters each of 0.8% sulfanilic acid and 0.5% N, N- dimethyl-alpha-naphthylamine, both in 5 N acetic acid were added to one hundred microliters of the cell sample. Nitrates concentrations were determined colorimetrically at 570 nm after incubation of the mixture for ten minutes at 37°C. Cell proliferation was determined manually with a hemacytometer and all counts are base upon one hundred microliter samples. The CD4/CD8 ratio was determined with the Becton Dickenson FACScan and CellFit software. Antibodies to CD4 labelled with fluorescein isothiocyanate and CD8 antibodies labelled with phycoerythrin as well as a non specific control were used in counting cells with CD4 and CD8 receptors. Blood samples were taken prior to HBO therapy (predive), during the final minutes of the third thirty minute oxygen inhalation period (oxygen), immediately following the first dive (post dive) and following the third dive 7/21/93 (post therapy). Another sample was taken before and after HBO therapy five days following the final treatment of the previous week

The results are listed in table 1 below.

Table 1

Blood Sample	Cell Count (n=3)	nitrate (48 hours) (n=3)	Cell Count (48 hours) (n=3)	CD4/CD8 ratios (n=2)
Con A + predive	$7.5x10^4$	21±4.9	$20x10^4$	
predive	$7.5x10^4$	-0.6±1.2	$6.5x10^4$	1.165, 1.125
Con A + oxygen	$7.0x10^4$	5.3±0.5	$9.8x10^4$	
oxygen	$7.0x10^4$	-0.8±2.4	$4.6x10^4$	
Con A + post dive	$7.2x10^4$	0.67±0.8	$5.3x10^4$	
post dive	$7.2x10^4$	-4.3±2.25	$3.9x10^4$	1.216, 1.308
Con A + post therapy	$8.0x10^4$	na	$7.4x10^4$	
post therapy	$8.0x10^4$	na	$4.4x10^4$	

Rats were used to determine the effects of microwave and hyperbaric oxygen therapy on mitogenic stimulation of lymphocytes and macrophage nitric oxide production. Concanavalin A stimulation and nitric oxide determinations were carried out as described for human lymphocytes. Tables 2 & 3 list data obtained from four control rats.

Table 2

Date	7/23		7/26		7/27		7/28	
Treatment	start	end	start	end	start	end	start	end
Control	81	90	46	30,23	47	33	42	36
con A control	81	89	46	27,28	47	35	42	3
Shock	81	28	44	23,23	43	34	53	48
con A shock	81	57	44	19,54	43	13	53	4
Death	81	55	45	26,2325	59	26	45	32
con A death	81	59	45	35,6400	59	7	45	2

Table 3

Date	7/23	7/26	7/27	7/28
Treatment	nitrate	nitrate	nitrate	nitrate
Control	14	9	4	4
con A control	7	11	4	17
Shock	10	10	0	2
con A shock	5	12	2	12
Death	16	14	50	4
con A death	11	22	56	14

One rat (acute) was cannulated and then entered hyperbaric oxygen therapy prior to receiving microwave exposure (Table 4). Four rats received hyperbaric oxygenation treatments for four days prior to cannulation and exposure to microwave Table 5 & 6). Hyperbaric oxygenation treatments consisted of 60 minutes at 3 atmospheres absolute at 100% oxygen. Microwave exposures consisted of 350 GHz irradiation until the animal exhibited signs of shock as indicated by the blood pressure falling below a predetermined level. Lymphocyte stimulation by concanavalin A and nitrate production by macrophages were monitored as described previously. Time to shock from beginning of microwave exposure and time to death following shock were measured for sixty control rats, one acute rat and four chronic rats (Table 7)

Table 4

single hyperbaric treatment prior to microwave exposure

	8/4	8/6	nitrate	
Treatment	start	end	8/6	8/9
Control	80	35	161	153
con A control	80	20	78	183
Shock	80	31	98	129
con A shock	80	10	75	134
Death	80	27	57	103
con A death	80	22	78	89

Table 5
Four hyperbaric treatments prior to microwave exposure

	Plate A			Plate B		
	8/5 start	8/9 end	nitrate	8/5 start	8/9 end	nitrate
Control	68	7	756	73	4	606
Con A control	68	8	1016	73	6	1421
Shock	53	5	1397	20	7	249
Con A shock	53	3	1068	20	3	214
Death	65	15	1387	104	26	231
Con A death	65	18	1706	104	23	677

Table 6

	Plate A			Plate B		
	8/6 start	8/9 end	nitrate	8/6 start	8/9 end	nitrate
Control	68	32	199	73	17	322
Con A control	68	17	203	73	19	312
Shock	53	17	199	20	15	193
Con A shock	53	10	195	20	12	221
Death	65	32	207	104	23	51
Con A death	65	37	84	104	18	50

Table 7

Treatment	Time to shock	Time to death following shock
Control (n=60)	36.68±0.72	13.44±2.49
Acute (n=1)	29	12
Chronic (n=4)	28.45±7.46	7.75±4.95

The stimulation of lymphocyte cell growth was identified through propidium iodide staining of cell nuclei following fixing with 70% ethanol for four hours. Nitroprusside (NP) was used as a nitric oxide emitter and 3-aminotyrosine (3AT) and glucose were also used to determine their effects on the lymphocyte cell cycle. The percentage of cells which were in the G2 phase are listed in Table 8.

Table 8
The percent of cells in the G2 phase

	30 min. PI stain		4 hrs PI stain	
Cell Count:	50,000	50,000	200,000	200,000
Model:	RFIT	SFIT	RFIT	SFIT
Control	3.1	3.8	5.2	4.8
30 mM NP	0.2	0.0	0.1	0.0
300 μM NP	4.2	2.8	4.6	3.7
2 mM 3AT	0.1	0.2	0.0	0.0
0.5 mM 3AT	0.4	0.6	0.5	0.0
300 mg% glucose	3.7	2.8	3.3	3.7

CONCLUSIONS

These studies demonstrate the complexity of the interaction between nitric oxide, oxygen derivatives of nitric oxide and oxygen, and enzymes which modulate the concentrations of these molecules. The results from these studies indicate the need for further investigation of the enzymes associated with the modulation of oxygen and nitric oxide concentrations and the physiological effects associated with the effects of the interactions between oxygen, nitric oxide and associated enzymes. In addition, these studies indicate that the functionality of the immune and inflammatory response is altered directly as a result of this complex interaction. These studies demonstrate that functionality of the immune system may not be identified solely through monitoring the expression of leukocyte surface proteins. The overall effect of the response of immune function may not solely be correlated to macrophage production of nitric oxide and the corresponding toxicity to lymphocytes.

Continued studies must be made into the effects of hyperbaric oxygenation on the function of the immune response and inflammatory response. The global effects which occur during an initial exposure may be different than those encountered following repeated exposures as demonstrated by this study.

ENTROPY GENERATION OF THE CARDIOVASCULAR CYCLE

Richard D. Swope

Professor

Department of Engineering Science

Trinity University

715 Stadium Drive

San Antonio, Texas 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.,

Trinity University

and

North Dakota State University

September 1993

ABSTRACT

The primary aim of this research is to develop mathematical descriptions of the entropy generation of the cardiovascular system. To accomplish this aim, the cardiovascular cycle is modeled by four discrete thermodynamic components - left and right heart, systemic and pulmonic circulation beds. Using the first two laws of thermodynamics, mathematical expressions for the irreversibility of the components are obtained. Blood flow pressure drop and extraction of metabolic fuel largely contribute to the irreversibility of the cardiovascular cycle. For the left and right heart the irreversibility is found to be due to compression losses in the ventricles, metabolic losses, and fluid flow pressure drop. In the systemic circulation, irreversibility was primarily due to fluid flow pressure drop and oxygen and fuel exchange with the tissue. Pulmonary circulation irreversibility was comprised of fluid flow pressure drop and oxygen exhange.

INTRODUCTION

Purpose:

The objective of this research was to describe, in mathematical terms, the entropy generated by the cardiovascular cycle.

Background:

The primary question that motivated this research was: how is the heart hydraulically coupled to its vascular load? Hydraulic coupling theories based on maximum first law thermodynamic efficiency or maximum external work transfer have been proposed (Suga et al, 1985). However, it appears that these approaches are not sufficient to describe the hydraulic coupling in every situation (Hayashida et al, 1992). Another approach seemed warranted.

In an excellent review, Gibbs and Chapman (1979) detailed early thermodynamic approaches to cardiac and skeletal muscle tissue. Much useful work has been done in the biothermodynamic analysis at the tissue level. However, the authors performed multiple automated and manual literature searches for previous work in which the cardiovascular cycle (CVC) was analyzed for entropy generation, lost available work, irreversibility, exergy or second law views, but found none.

In contrast, much second law analysis has been performed on engineering thermodynamic cycles such as Otto, Brayton, and vapor compression. It is a well-accepted approach for obtaining additional information about the performance of such systems. (Black and Hartley, 1991, Van Wylen and Sonntag, 1991) Therefore, it is puzzling that this approach has not been used to analyze the CVC. One good reason for studying the CVC from an entropy generation viewpoint is that to minimize *irreversibility* in a cycle, means that the fuel obtained by the cycle is utilized as effectively as possible. A good strategy for survival is to use what energy one has as effectively as possible - given certain life supporting constraints. Other reasons supporting this approach are that:

2) second law analysis relies on a relatively simple model where a few measurements of state properties can yield much information

3) viewing CV performance from this perspective may help obtain an increase in G-tolerance of tactical aircraft pilots.

4) viewing CV performance from this perspective may help improve clinical diagnosis and therapeutic approach of the general cardiac patient.

Scope of Report:

This report will discuss the basic thermodynamic laws which form the foundation for the mathematical development. Next, the development of the CVC model will be discussed. Then, the mathematical equations describing the entropy generation of each component in the CVC will be developed. Finally, the utility and future promise of this approach will be discussed.

BASIC THERMODYNAMIC LAWS

In this section, the general laws of thermodynamics will be shown along with certain simplifying assumptions which allow for a more convenient mathematical form.

1st law of Thermodynamics for a control volume:

$$\dot{Q}_{c.v} = \frac{d}{dt}\int_V e\rho\, dV + \int_A \left(h + \frac{\upsilon^2}{2} + gZ\right)\rho\upsilon_n\, dA + \dot{W}_{c.v.}$$

Where:

$\dot{Q}_{c.v}$ ≡ rate of heat transfer into the control volume (c.v.) surface.
ρ ≡ density of fluid
e ≡ energy contained in volume, dV
h ≡ enthalpy of substance
υ ≡ velocity of substance
g ≡ gravitational constant
Z ≡ elevation of substance
$\upsilon_{r.n.}$ ≡ outward-directed normal velocity
dA ≡ area of flow, $\dot{m}$
$\dot{W}_{c.v}$ ≡ done by the control volume.

In words the first law states for a control volume ...

rate of heat transfer = rate of energy stored + rate of net energy entering + power produced

For uniform states of mass crossing the control surface, the first law becomes ...

$$\dot{Q}_{C.V.} + \sum \dot{m}_i\left(h_i + \frac{\upsilon_i^2}{2} + gZ_i\right) = \frac{dE_{C.V.}}{dt} + \sum \dot{m}_e\left(h_e + \frac{\upsilon_e^2}{2} + gZ_e\right) + \dot{W}_{C.V.}$$

Where:

$\dot{m}_i$ ≡ mass flow rate into control volume
$\dot{m}_e$ ≡ mass flow rate out of control volume

Furthermore, if one assumes steady energy content in the control volume, that the control volume does not move relative to the coordinate frame, and that the mass flux and its state do not change in time, and finally, that $\dot{Q}_{C.V.}$ and $\dot{W}_{C.V.}$ remain constant, one

obtains a steady state steady flow (SSSF) equation

$$\dot{Q}_{C.V.} + \sum \dot{m}_i \left(h_i + \frac{\upsilon_i^2}{2} + gZ_i \right) = \sum \dot{m}_e \left(h_e + \frac{\upsilon_e^2}{2} + gZ_e \right) + \dot{W}_{C.V.}$$

If a chemical reaction takes place or if there is a transformation of material, the enthalpies take the form:

$$m_i h_i = H_i = n_1 \bar{h}_1 + n_2 \bar{h}_2 + \ldots + n_n \bar{h}_n$$

Where:

$n_n \bar{h}_n$ = product of moles and molar enthalpy of species n

and

$$\bar{h}_n = \left[h_f^o + \Delta \bar{h} \right]_n$$

Where:

$h_f^o \equiv$ enthalpy of formation at reference pressure and temperature

$\Delta \bar{h} \equiv$ enthalpy change due to non reference pressure and temperature.

Thus, under SSSF conditions and neglecting kinetic energy and potential energy ...

$$\dot{Q}_{C.V.} + \sum \dot{n}_i \bar{h}_i = \sum \dot{n}_e \bar{h}_e + \dot{W}_{C.V.}$$

Certainly, the CV system does not operate under SSSF conditions over a heart beat. The ventricles produce flow intermittently and the condition of the system fluctuates. But over a "sufficiently" long time, the CV system approaches a SSSF condition. It is not necessary to limit the thermodynamic analysis to SSSF conditions, but it helps to focus on the important issues, by taking this simpler approach. Just as Poiselle's law (valid for steady flow conditions) has great utility in explaining nonsteady hemodynamics, the authors believe that a SSSF approach can be of great utility in explaining CV thermodynamics as well.

2nd Law of Thermodynamics for a control volume (general form):

$$\frac{d}{dt} \int_V s\rho dV + \int_A s\rho v_{r.n.} dA = \int_A \left(\frac{\dot{Q}/A}{T} \right) dA + \int_V \left(\frac{\dot{L}W}{T} \right) dV$$

Where:

$s \equiv$ entropy

$T \equiv$ Temperature

$L\dot{W} \equiv$ lost work due to irreversibility

In words, the second law states that for a control volume the ...

rate change in entropy + net entropy leaving = entropy due to heat transfer + entropy generation

For a SSSF process ...

$$\sum \dot{m}_e s_e - \sum \dot{m}_i s_i - \int_A \left(\frac{\dot{Q}/A}{T} \right) dA = \sigma$$

Where:

$\sigma \equiv$ rate of entropy generated

$s_e \equiv$ entropy exiting the control volume

$s_i \equiv$ entropy leaving the control volume

Continuity Equation (general form):

The continuity equation for a control volume is:

$$\frac{d}{dt}\int_V \rho dV + \int_A \rho \upsilon_{r.n} dA = 0$$

For SSSF conditions:

$$\sum \dot{m}_i = \sum \dot{m}_e$$

Where:

$$\dot{m} = \rho A \upsilon$$

Now these SSSF equations can be applied to the CV system and in the next section the CVC model will be developed.

The CV system is thermodynamically represented by four compartments. These are 1) left heart, 2) systemic circulation, 3) right heart, and 4) pulmonic circulation. The CV system schematic is shown below. A control surface is drawn around the outer surface of the heart.

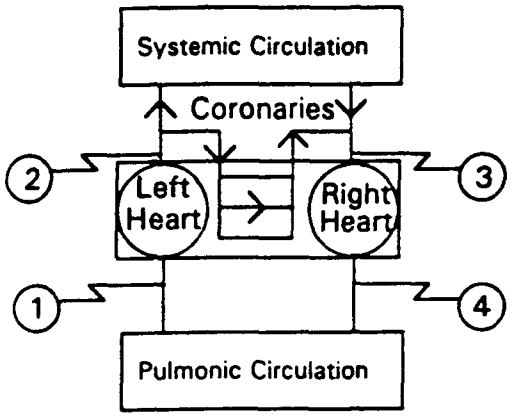


Figure 1 - CV system schematic

While the coronary flow is part of the systemic circulation, it is included with the heart as a separate flow from the "systemic circulation" compartment so that the overall thermodynamics of the heart can be determined. Note that the systemic circulation can be subdivided into individual organ systems so that the organ thermodynamics can be determined in a similar manner as the heart.

Now that the CVC has been thermodynamically represented in schematic form, it can also be represented on a fluid property diagram, such as a temperature-entropy (TS) diagram. However, unlike a common fluid such as water, blood is not truly a thermodynamically pure substance. A thermodynamically pure substance is one that has a homogeneous and invariable chemical composition and blood is actually a suspension. Also, since blood carries reacting and reacted components in it, blood cannot be considered a simple compressible substance. A simple compressible substance is one where knowing two independent variables, say pressure and specific volume, a third variable is automatically known, say temperature, through an equation of state or property chart. Unfortunately, to the authors' knowledge, an equation of state or property chart does not exist for blood. But later it will be shown, that to an engineering degree of accuracy, under certain conditions, blood may be considered a thermodynamically pure substance and a simple compressible substance. Once an equation of state or property chart for blood has been developed, the analysis suggested in this report will become much more straightforward.

LEFT AND RIGHT HEART

Consider the following model of the heart as two pumps with the energy for the pumps derived from a fuel/oxygen reaction taking place in the cardiac muscle (supplied by the coronaries).

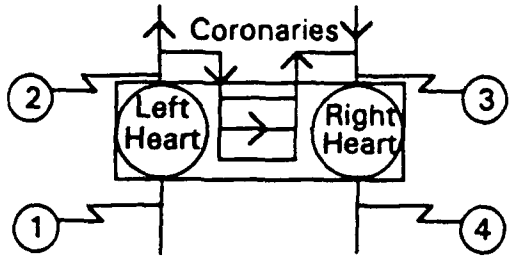


Figure 2 Left and Right Heart Schematic

Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity Equation (Heart):

Conservation of mass yields the following equations.

$$\dot{m}_1 = \dot{m}_2 + \dot{m}_C$$

$$\dot{m}_3 + \dot{m}_C = \dot{m}_4$$

$$\dot{m}_4 = \dot{m}_1, \quad \therefore \quad \dot{m}_2 = \dot{m}_3$$

1st Law of Thermodynamics:

Application of the 1st Law produces ...

$$\dot{Q} + \sum_R \dot{N}_i \left[h_f^\circ + \Delta h\right]_i = \dot{W} + \sum_P \dot{N}_i \left[h_f^\circ + \Delta h\right]_i$$

Where, R = reactants and P = products

Expanding the summation signs and assuming heat and work transfer rates are negligible across the outer surface of the heart ...

$$\dot{N}_{c_2}\left[h_f^\circ + \Delta h\right]_{c_2} + \dot{N}_1\left[h_f^\circ + \Delta h\right]_1 + \dot{N}_3\left[h_f^\circ + \Delta h\right]_3 = \dot{N}_{c_3}\left[h_f^\circ + \Delta h\right]_{c_3} + \dot{N}_2\left[h_f^\circ + \Delta h\right]_2 + \dot{N}_4\left[h_f^\circ + \Delta h\right]_4$$

where:

c_2 ≡ *coronary inf low*

c_3 ≡ coronary outflow

Substituting the following ...

$$\dot{m}_C HV = \dot{N}_{c_2}[h_f^\circ]_{c_2} - \dot{N}_{c_3}[h_f^\circ]_{c_3}$$

$$\dot{m}_C \Delta h = \dot{N}_{c_3}[\Delta h]_{c_3} - \dot{N}_{c_2}[\Delta h]_{c_2}$$

$$\dot{m}_{LH}\,\Delta h_{LH} = \dot{N}_2[\Delta h]_2 - \dot{N}_1[\Delta h]_1, \text{ where } \dot{N}_1[h_f^\circ]_1 = \dot{N}_2[h_f^\circ]_2$$

$$\dot{m}_{RH}\,\Delta h_{RH} = \dot{N}_3[\Delta h]_3 - \dot{N}_4[\Delta h]_4, \text{ where } \dot{N}_3[h_f^\circ]_3 = \dot{N}_4[h_f^\circ]_4$$

and where under SSSF conditions ...

$$\dot{m}_{LH} = \dot{m}_{RH} = \dot{m}_H.$$

Yields ...

$$\dot{m}_C HV = \dot{m}_C \Delta h_C + \dot{m}_H (\Delta h_{RH} + \Delta h_{LH})$$

At this point, it may be helpful to view the coronary flow and its thermodynamic equations by examining the schematic in Figure 3. Here Q_2 is the heat that is lost either across the outer surface of the heart or that which is lost to the blood in the atria or ventricles. Since the control surface was drawn at the outside surface of the heart and $Q_{c.v.} = 0$, Q_2 is heat transferred to the blood in the atria and the ventricles. Practically, this is probably very small compared to the heat transferred to the coronary blood, Q. W is the work performed by the heart on the blood in the atria and ventricles.

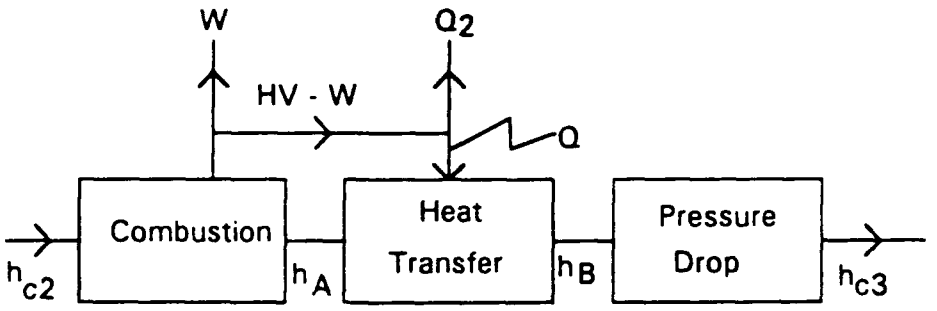


FIGURE 3 Coronary Flow Schematic

If one assumes that state point A is at the same temperature and pressure as state point c2 then $h_{c2} - h_A = HV$. It is also apparent that $HV = W + Q_2 + Q$. State point B may be at a higher temperature than A, so that $h_B - h_A = c\Delta T_{BA}$ where c is specific heat of coronary blood. Block 3 accounts for any pressure drop in the coronary flow.

Thus,

$$\Delta h_{C3,C2} = \Delta h_{A,C2} + \Delta h_{B,A} + \Delta h_{C3,B}$$

and

$\Delta h_{C3,B} = 0$, Throttling process

$\Delta u_{C3,B} = -\upsilon \Delta P_{C3,B}$

$$\Delta h_{C3,C2} = -HV + \Delta h_{B,A}$$

$$\Delta h_{C3,C2} = -HV + c\Delta T_{B,A}$$

$$\Delta h_{C3,C2} = -HV + c\Delta T_{C3,C2}$$

$$\Delta h_C = -HV + c\Delta T_C$$

$$\dot{m}_C \Delta h_C = -\dot{m}_C HV + \dot{m}_C \dot{C} \Delta T_C$$

and with substitution and rearrangement

Representing the total enthalpy increase of both left and right hearts as ...

$$\Delta h_H = (h_2 - h_1) + (h_4 - h_3)$$

$$(h_2 - h_1) = c\Delta T_{LH} + v\Delta P_{LH}$$

$$(h_4 - h_3) = c\Delta T_{RH} + v\Delta P_{RH}$$

next applying the 1st law to all the flows of the heart one obtains...

$$m_H v(\Delta P_{LH} + \Delta P_{RH}) = m_C HV - m_C c\Delta T_C - m_H \left(c(\Delta T_{LH} + \Delta T_{RH})\right)$$

Compare this equation with the equation below.

$$\dot{W} = \dot{m}_C HV - Q - Q_2.$$

This equation shows that the power output of the heart (left side of equation) is the total heating value of the fuel less some energy loss terms. These loss terms represent the net power change in the coronary flow due to heating from waste heat of combustion and pressure drop, and the net change in energy of the blood due to temperature increase of the blood as it is pressurized in the heart ventricles. If there were no loss terms, all the fuel could be converted to power output of the heart. But because of thermodynamic energy losses when energy is transformed from one form to another, such as metabolism, friction, and heat transfer, the power output is less than the energy content in the fuel.

Second Law of Thermodynamics:

$$\sigma = \sum_{out} \dot{m}s - \sum_{in} \dot{m}s - \sum_{i} \frac{Q_i}{T_i}$$

Assuming that ...

$$\sum_{i} \frac{Q_i}{T_i} = 0, \text{ because } Q_2 << Q$$

$$\sigma = \dot{m}_C(s_{C3} - s_{C2}) + \dot{m}_H(s_2 + s_4 - s_1 - s_3)$$

$$\sigma = \dot{m}_C(\Delta s_C) + \dot{m}_H(\Delta s_{LH} + \Delta s_{RH})$$

Recall, the TΔS formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_{i} \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_{i} \mu_i \Delta N_i$$

$$\sum_{i} \mu_i \Delta N_i = \Delta G)_{T,P}$$

where G is the Gibbs' free energy function...

Thus,

$$T\Delta s_{LH} = \Delta h_{LH} - v\Delta P_{LH} = c\Delta T_{LH}$$

$$T\Delta s_{RH} = \Delta h_{RH} - v\Delta P_{RH} = c\Delta T_{RH}$$

$$T\Delta s_C = \Delta h_{C3,C2} - v\Delta P_{C3,C2} - \sum_{i} \mu_i \Delta N_i = -HV + v\Delta P_{C2,C3} + \Delta G)_{T,P}$$

$$\sigma = -HV + \Delta G)_{T,P} + \upsilon\Delta P_{C2,C3} + c\Delta T_{LH} + c\Delta T_{RH} + c\Delta T_c$$

Since

$$\Delta G)_{T,P} \equiv W_{rev}$$

and

$$Wact = HV - C\Delta T_c, \text{ assuming } Q_2 \approx 0$$

$$\sigma_c = W_{rev} - W_{act} + \upsilon\Delta P_{C2,C3}$$

Lost Work (LW) = W_{rev}-W_{act}

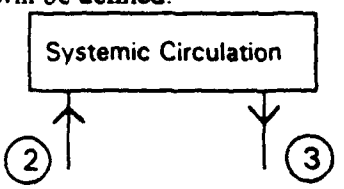
$$\sigma_c = LW + \upsilon\Delta P_{C2,C3}$$

σ_c = Lost work due to combustion inefficiency and coronary flow pressure drop.

$$\sigma = \sigma_c + c(\Delta T_{LH} + \Delta T_{RH})$$

SYSTEMIC CIRCULATION

The analysis for the systemic circulation follows closely the analysis for the coronary flow. First, a model will be defined.



Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity equation (systemic Circulation)

Conservation of mass yields the following equation.

$$m_3 = m_2$$

First Law Analysis (systemic circulation)

At this point, it may be helpful to view the systemic flow and its thermodynamic equations by examining the schematic in Figure 4. Here Q_{SKIN} is the heat that is lost to the outside tissue and Qs is the heat that is returned to the capillary across the control surface. Wc.v. = 0.

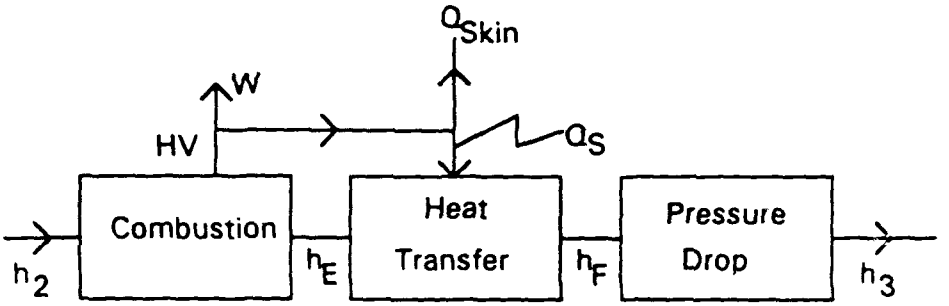


FIGURE 4 Systemic Block Diagram

If one assumes that state point E is at the same temperature and pressure as state point 2 then $h_2 - h_E = HV$. State point F may be at a higher temperature than E, so that $h_F - h_E = c\Delta T_{FE}$. Block 3 accounts for any pressure drop in the systemic bed.

Thus,

$$\Delta h_{32} = \Delta h_{E2} + \Delta h_{FE} + \Delta h_{3F}$$

$$\Delta h_{3F} = 0, \text{ Throttling Process}$$

$$\Delta u_{3F} = -v\Delta P_{3F}$$

$$\Delta h_{32} = -HV + \Delta h_{FE}$$

$$\Delta h_{32} = -HV + c\Delta T_{FE}$$

and with substitution and rearrangement

$$\dot{m}_H \Delta h_{32} = m_H (-HV + c\Delta T_{32})$$

This equation shows that the enthalpy change is the heating value of the fuel less the heat returned to the systemic circulation.

Second Law Analysis

$$\sigma = \sum_{out} ms - \sum_{in} ms - \sum_{i} \frac{Q_i}{T_i}$$

$$\sigma = m_H (\Delta s_{32}) - \frac{Q_s}{T}$$

where Q_S is the heat transferred across the circulatory system vessels.

Recall, the TΔS formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_i \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_i \mu_i \Delta N_i$$

$$\sum_i \mu_i \Delta N_i = \Delta G)_{T,P}$$

Thus,

$$T\Delta s_{32} = \Delta h_{32} - v\Delta P_{32} - \sum_i \mu_i \Delta N_i = -HV + c\Delta T_{32} + v\Delta P_{23} + \Delta G)_{T,P}$$

$$\sigma = \dot{m}_s\left(-HV + c\Delta T_{32} + v\Delta P_{23} + \Delta G)_{T,P}\right) - \frac{Q_s}{T}$$

$$\sigma = \dot{m}_s\left(-HV + \Delta G + v\Delta P_{23})_{T,P}\right)$$

PULMONARY CIRCULATION:

Shown below is the schematic of the last component of the CVC system.

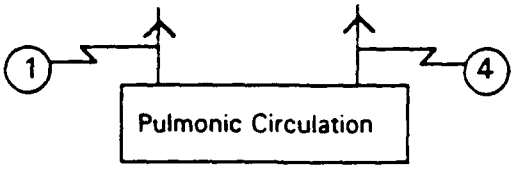
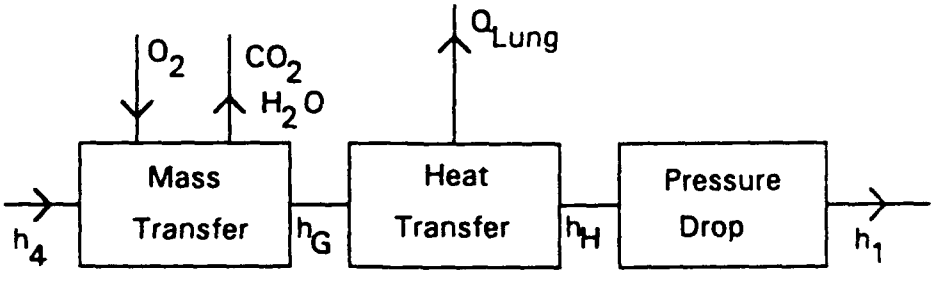


Figure 5 Pulmonary Circulation Schematic

The control surface for the control volume was drawn around the pulmonary circulation. It was assumed that the boundary work done on the pulmonary circulatory system is much smaller than the amount of heat transferred across the control surface. Given these constraints, a thermodynamic block diagram similar to the systemic circulation was used.



The first block represents products of metabolism entering with enthalpy h_4. The fluid gains oxygen (a reactant). For simplicity, it was assumed that the blood also acquires fuel at this point. While physiologically, the fuel is acquired by a number of methods and anatomical locations, thermodynamically it is a good approximation to have the fuel enter with the oxygen. Physiologically, this is termed the energy equivalent of oxygen and is about 4.8 Calories/L oxygen. Furthermore it was assumed that state 4 and state G are at the same temperature and pressure. State H is typically at a lower temperature than G and

state 1 is at the same temperature as H but a lower pressure.

Using SSSF assumptions and neglecting changes in potential energy and kinetic energy the following relations can be obtained.

Continuity equation (Pulmonary Circulation)

Conservation of mass yields the following equation.

$$m_4 = m_1$$

First Law Analysis (Pulmonary circulation)

$$\Delta h_{1H} = 0, \text{ Throttling Process}$$
$$\Delta u_{1H} = -v\Delta P_{1H} = v\Delta P_{41}$$

Thus,

$$\Delta h_{14} = \Delta h_{G4} + \Delta h_{HG} + \Delta h_{1H}$$

$$\Delta h_{G4} = HV$$

$$\Delta h_{14} = HV + \Delta h_{HG}$$

$$\Delta h_{14} = HV + c\Delta T_{HG} = HV + c\Delta T_{14}$$

and with substitution and rearrangement

$$\dot{Q}_{LUNG} = m_H \, c\Delta T_{14}$$

This assumes that most of the temperature change is due to heat-transfer with the pulmonary gases and not from metabolic waste heat from the lung tissue.

Second Law Analysis (Pulmonary Circulation):

$$\sigma = \sum_{out} \dot{m} s - \sum_{in} \dot{m} s - \sum_{i} \frac{Q_i}{T_i}$$

$$\sigma = m_H (\Delta s_{14}) - \frac{Q_{LUNG}}{T}$$

Recall, the $T\Delta S$ formulations ...

$$T\Delta s = \Delta h + v\Delta P + \sum_i \mu_i \Delta N_i$$

$$T\Delta s = \Delta u - P\Delta v + \sum_i \mu_i \Delta N_i$$

$$\sum_i \mu_i \Delta N_i = \Delta G)_{T,P}$$

Thus,

$$T\Delta s_{14} = \Delta h_{14} - v\Delta P_{14} - \sum_i \mu_i \Delta N_i = HV + c\Delta T_{14} + v\Delta P_{14} + \Delta G)_{T,P}$$

$$\sigma = \dot{m}_H \left(HV + c\Delta T_{14} + v\Delta P_{14} + \Delta G)_{T,P} \right) - \frac{Q_{LUNG}}{T}$$

$$\sigma = \dot{m}_H \left(HV + v\Delta P_{14} + \Delta G)_{T,P} \right)$$

CONCLUSIONS

The CV system was modeled as a thermodynamic system to investigate the theoretical foundations of irreversibility minimization as a control strategy for the circulatory system. Simplifying assumptions, such as SSSF conditions, were imposed. As one might suspect, irreversibility was generated in each component from fluid flow pressure drops, metabolic processes and mass transfer. Although much more work needs to be done, the foundations of the concept have been developed.

REFERENCES

Suga H, Igraashi Y, Yamada O, Goto Y: Mechanical efficiency of the left ventricle as a function of preload, afterload and contractility. Heart Vessel 1:3-8, 1985

Hayashida K, Sunagawa K, Noma M, Sugimachi M, Ando H, Nakamura M: Mechanical Matching of the left ventricle with the arterial system in exercising dogs. Circulation Research 71:481-489, 1992

Gibbs C, Chapman J: Cardiac Energetics. In: Bern RM, Sperelakis N, Geiger S, eds.: Handbook of Physiology Sec 2: The Cardiovascular system, Vol. 1: The Heart. Bethesda: American Physiology Society, 1979, pp 775-804.

Black W and Hartley J: Thermodynamics, 2nd ed, Hraper Collins, 1991

Van Wylen G and Sonntag R: Fundamentals of Classical Thermodynamics, 3rd ed, John Wiley and Sons, 1991

MODELING OF FLEXIBLE BODIES FOR THE ATB MODEL

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.

July 1993

MODELING OF FLEXIBLE BODIES FOR THE ATB MODEL

Hashem Ashrafiuon
Assistant Professor
Department of Mechanical Engineering
Villanova University

Abstract

The Articulated Total Body (ATB) is a rigid 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. In order to predict the response accurately, however, a rigid body dynamic model may not be sufficient. This is particularly true for the relatively "soft" segments such as the neck and in high speed applications. In this study, a flexible body model of the ATB is presented which incorporates small linear deformations of individual segments in the model. The finite element method is used to develop linearly elastic models of the flexible bodies. Mode shapes and structural stiffness and damping characteristics are determined using modal analysis. The information generated by the finite element models is then fed into the revised version of the ATB model. This report presents the analytical formulation of the coupling between the gross motion of the human body and small deformation of individual flexible segments.

MODELING OF FLEXIBLE BODIES FOR THE ATB MODEL

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 model treats the individual segments of the overall system (in this case the human body) as rigid bodies. However, the response predicted by the model may not be accurate particularly in high speed applications since it ignores the effects of small deformations within the individual segments (Winfry, 1971). Small deformations of some of the segments influence the dynamic response of the system to such an extent that their exclusion from the analysis may lead to substantial error in the solution. Hence, in order to have an accurate prediction of the response, it may be necessary to model the "softer" segments or segments which undergo relatively high speed motion as flexible bodies.

It is assumed that the deformation of a flexible 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. In order reduce the size and complexity of the problem, a small number of modes may be selected to approximate the displacement field using Ritz approximation (Wilson, et al. 1982). The vibration normal modes are determined using finite element modal analysis of the flexible bodies.

First, the coupled equations of motion of a flexible body undergoing gross motion are developed. Then, the constrained equations of motion of a system of interconnected flexible bodies are developed. The equations are then decoupled to have the same form as the rigid body model. This allows for convenient inclusion of flexible body equations into the rigid body formulation. Finally, an example is presented to show the effect of small deformation on the gross motion of the human body.

Kinematics of a Flexible Body

To define the gross motion of a flexible body an xyz reference frame is attached to the unreformed state of the body and its location (**x**) and orientation (**D**) is specified with respect to an inertial reference frame XYZ. The elastic displacement field of a flexible body (**u**) may be represented as a linear combination of a set of selected deformation modes determined by finite element modeling and Ritz approximation. Assuming the body is modeled with n nodes and m modes are selected, the displacement field containing nodal translations and rotations of all nodes is written as:

$$u = \Psi a \tag{1}$$

where **a** is an mx1 modal coordinate vector and Ψ is a 6nxm modal matrix. Therefore as shown in Fig. 1, the position vector of a node k on the body may be written as:

$$x_k = x + D^T s_k = x + D^T(s_{k0} + \psi_k a) \tag{2}$$

where s_{k0} and s_k are the position vectors of nodes k in the undeformed and deformed states with respect to xyz, D^T is the transposed of the direction cosine matrix, and ψ_k is a size 3xm translational modal submatrix corresponding to node k. Taking the first and second time derivatives of Eq. (2), velocity and acceleration vectors of node k are written as:

$$\dot{x}_k = \dot{x} + D^T(-\tilde{s}_k\omega + \psi_k\dot{a}) \tag{3}$$

$$\ddot{x}_k = \ddot{x} + D^T[-\tilde{s}_k\dot{\omega} + \psi_k\ddot{a} + \tilde{\omega}(\tilde{\omega}s_k + 2\psi_k\dot{a})] \tag{4}$$

where ω is the angular velocity vector about xyz axes, and "~" operating on a vector produces a skew-symmetric matrix for cross product representation in matrix algebra.

Equations of Motion of a Flexible Body

The variational equations of motion of a flexible body at time t, for a virtual displacement field that is consistent with the constraints, is (Shames and Dym, 1985):

$$-\int_V \rho \delta \boldsymbol{x}_k^T \ddot{\boldsymbol{x}}_k dV + \int_V \delta \boldsymbol{x}_k^T \boldsymbol{f}_k dV = \int_V \delta \boldsymbol{e}_k^T \boldsymbol{\tau}_k dV \tag{5}$$

where ρ is material density, $\mathbf{f}_k$ is the body force at node k, $\delta\mathbf{x}_k$ is virtual displacement of node k consistent with constraints, V is the undeformed volume, and τ_k and $\mathbf{e}_k$ are defined as stress and strain variation vectors.

Taking the variation of Eq. (2), virtual displacement of node k is written as:

$$\delta \boldsymbol{x}_k = \delta \boldsymbol{x} + \boldsymbol{D}^T(-\tilde{\boldsymbol{s}}_k \delta \boldsymbol{\pi} + \boldsymbol{\Psi}_k \delta \boldsymbol{a}) \tag{6}$$

where $\delta\pi$ is a virtual vector along the axis of rotation and its magnitude is the angle of rotation.

For small deformations, the linear strain-displacement relation may be written in terms of constant modal strain matrix **B**:

$$\delta \boldsymbol{e} = \boldsymbol{B}\ \delta \boldsymbol{a} \tag{7}$$

Also from linear elasticity, the material property matrix $\mathbf{D}_0$ relates the stress and strain vectors:

$$\boldsymbol{\tau} = \boldsymbol{D}_0 \boldsymbol{e} = \boldsymbol{D}_0 \boldsymbol{B} \boldsymbol{a} \tag{8}$$

Substituting Eqs. (4) & (6-8) into Eq. (5) and using the fact that the resulting variational equation must be valid for all admissible $\delta\mathbf{x}$, $\delta\pi$, and $\delta\mathbf{a}$, the equations of motion of a flexible body can be written as:

$$\begin{aligned} \boldsymbol{M}_{tt}\ddot{\boldsymbol{x}} + \boldsymbol{M}_{tr}\dot{\boldsymbol{\omega}} + \boldsymbol{M}_{ta}\ddot{\boldsymbol{a}} + \boldsymbol{c}_t &= \boldsymbol{u}_t \\ \boldsymbol{M}_{tr}^T\ddot{\boldsymbol{x}} + \boldsymbol{M}_{rr}\dot{\boldsymbol{\omega}} + \boldsymbol{M}_{ra}\ddot{\boldsymbol{a}} + \boldsymbol{c}_r &= \boldsymbol{u}_r \\ \boldsymbol{M}_{ta}^T\ddot{\boldsymbol{x}} + \boldsymbol{M}_{ra}^T\dot{\boldsymbol{\omega}} + \boldsymbol{M}_{aa}\ddot{\boldsymbol{a}} + \boldsymbol{c}_a &= \boldsymbol{u}_a \end{aligned} \tag{9}$$

where $\mathbf{c}_t$, $\mathbf{c}_r$, and $\mathbf{c}_a$ are the kinematic constraint forces. Elements of Eq. (9) and their lumped mass approximations are listed below (Yoo and Haug, 1986):

$$M_{tt} = \int_V \rho dV\, I_3 \approx \sum_{k=1}^{n} m_k\, I_3$$

$$M_{tr} = -D^T \int_V \rho \tilde{s}_k dV \approx -D^T \sum_{k=1}^{n} m_k \tilde{s}_k \qquad (10a)$$

$$M_{ta} = D^T \int_V \rho \psi_k dV \approx D^T \sum_{k=1}^{m} m_k \psi_k$$

$$M_{rr} = -\int_V \rho \tilde{s}_k \tilde{s}_k dV \approx -\sum_{k=1}^{n} m_k \tilde{s}_k \tilde{s}_k$$

$$M_{ra} = \int_V \rho \tilde{s}_k \psi_k dV \approx \sum_{k=1}^{n} m_k \tilde{s}_k \psi_k \qquad (10b)$$

$$M_{aa} = \int_V \rho \psi_k^T \psi_k dV \approx \sum_{k=1}^{n} m_k \psi_k^T \psi_k \qquad (10c)$$

$$u_t = \int_V f_k dV - D^T \int_V \rho \tilde{\omega}\tilde{\omega} s_k dV - 2D^T \tilde{\omega} \int_V \rho \psi_k dV\, \dot{a}$$
$$\approx \sum_{k=1}^{n} f_k - D^T \tilde{\omega}\tilde{\omega} \sum_{k=1}^{n} m_k s_k - 2D^T \tilde{\omega} \sum_{k=1}^{n} m_k \psi_k\, \dot{a} \qquad (11a)$$

$$u_r = \int_V \tilde{s}_k D f_k dV - \int_V \rho \tilde{s}_k \tilde{\omega}\tilde{\omega} s_k dV - 2\int_V \rho \tilde{s}_k \tilde{\omega} \psi_k dV\, \dot{a}$$
$$\approx \sum_{k=1}^{n} \tilde{s}_k D f_k - \sum_{k=1}^{n} m_k \tilde{s}_k \tilde{\omega}\tilde{\omega} s_k - 2\sum_{k=1}^{n} m_k \tilde{s}_k \tilde{\omega} \psi_k\, \dot{a} \qquad (11b)$$

$$u_a = \int_V \psi_k^T D f_k dV - \int_V \rho \psi_k^T \tilde{\omega}\tilde{\omega} s_k dV - 2\int_V \rho \psi_k^T \tilde{\omega} \psi_k dV\, \dot{a} - \int_V B^T D_0 B dV\, a$$
$$\approx \sum_{k=1}^{n} \psi_k^T D f_k - \sum_{k=1}^{n} m_k \psi_k^T \tilde{\omega}\tilde{\omega} s_k - 2\sum_{k=1}^{n} m_k \psi_k^T \tilde{\omega} \psi_k\, \dot{a} - \Psi^T K_s \Psi\, a \qquad (11c)$$

where I_3 is the 3x3 identity matrix, K_s is the finite element stiffness matrix, and $\Psi^T K_s \Psi$ represents the modal stiffness matrix.

The modal damping forces are also defined in a similar manner as the modal stiffness matrix:

$$f_d = \Psi^T C_s \Psi \dot{a} \tag{12}$$

and then subtracted from the LHS of Eq. (11c). Modal stiffness and damping matrices are mxm diagonal matrices written as (Clough, 1971):

$$\begin{aligned} \Psi^T C_s \Psi &= diag[2\zeta_1\omega_{n1}, \ldots, 2\zeta_m\omega_{nm}] \\ \Psi^T K_s \Psi &= diag[\omega_{n1}^2, \ldots, \omega_{nm}^2] \end{aligned} \tag{13}$$

where ω_n and ζ represent structural natural frequencies and damping ratios. Note that, Ψ contains the selected vibration modes normalized with respect to the finite element mass matrix

Kinematic Constraints

Two types of kinematic joints are considered: free or ball and socket joints and pin joints.

Free Joints: Consider flexible bodies i and j connected by a free or ball and socket joint, as shown in Fig. 2. For a free joint, the joint attachment nodes for the two bodies must coincide. Therefore, the three constraint equations are:

$$x_i - x_j + D_i^T s_i - D_j^T s_j = 0 \tag{14}$$

where

$$\begin{aligned} s_i &= s_{i0} + \Psi_i a_i \\ s_j &= s_{j0} + \Psi_j a_j \end{aligned} \tag{15}$$

and Ψ_i and Ψ_j are the modal submatrices corresponding to the translational components of joint attachment nodes of bodies i and j. Taking the first and second time derivatives of Eq. (14), results in the kinematic velocity and acceleration equations:

$$\dot{x}_i - \dot{x}_j - D_i^T \tilde{s}_i \omega_i + D_j^T \tilde{s}_j \omega_j + D_i^T \Psi_i \dot{a}_i - D_j^T \Psi_j \dot{a}_j = 0 \tag{16}$$

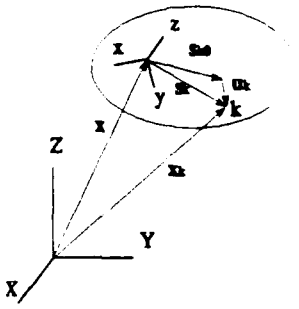


Figure 1. Kinematics of a Flexible Body

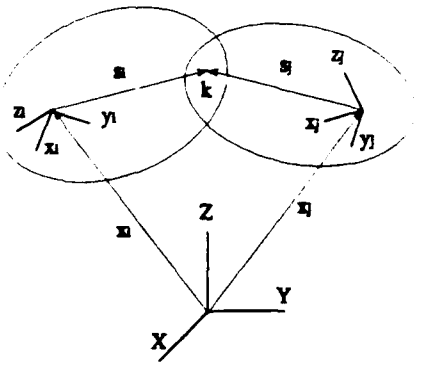


Figure 2. Free or Ball and Socket Kinematic Joint

$$\ddot{x}_i - \ddot{x}_j - D_i^T\tilde{s}_i\dot{\omega}_i + D_j^T\tilde{s}_j\dot{\omega}_j + D_i^T\Psi_i\ddot{a}_i - D_j^T\Psi_j\ddot{a}_j = - D_i^T\tilde{\omega}_i(\tilde{\omega}_i s_i + 2\Psi_i\dot{a}_i) + D_j^T\tilde{\omega}_j(\tilde{\omega}_j s_j + 2\Psi_j\dot{a}_j) \tag{17}$$

Based on the above equation the following submatrices are defined for the joint:

$$\begin{aligned} B_{11} &= [\overset{i}{I_3}, -\overset{j}{I_3}] \\ B_{12} &= [-\overset{i}{D_i^T\tilde{s}_i}, \overset{j}{D_j^T\tilde{s}_j}] \\ B_{1a} &= [\overset{i}{D_i^T\Psi_i}, -\overset{j}{D_j^T\Psi_j}] \\ \nu_1 &= -D_i^T\tilde{\omega}_i(\tilde{\omega}_i s_i + 2\Psi_i\dot{a}_i) + D_j^T\tilde{\omega}_j(\tilde{\omega}_j s_j + 2\Psi_j\dot{a}_j) \end{aligned} \tag{18}$$

Pin Joints: This type of joint introduces two more constraints in addition to the three imposed by a free joint. The direction of rotation **h** must be the same for both bodies:

$$D_i^T h_i - D_j^T h_j = 0 \tag{19}$$

where

$$\begin{aligned} h_i &= (I_3 + \tilde{\phi}_i)h_{i0} \; ; \quad \phi_i = \mu_i a_i \\ h_j &= (I_3 + \tilde{\phi}_j)h_{j0} \; ; \quad \phi_j = \mu_j a_j \end{aligned} \tag{20}$$

The terms inside the parentheses is the transposed of direction cosine matrix of the joint attachment nodes small rotations in body i and j reference frames (Greenwood, 1988). Kinematic velocity and acceleration equations are given as the first and second derivatives of Eq. (18):

$$-D_i^T\tilde{h}_i\omega_i + D_j^T\tilde{h}_j\omega_j - D_i^T\tilde{h}_{i0}\dot{\phi}_i + D_j^T\tilde{h}_{j0}\dot{\phi}_j = 0 \tag{21}$$

$$-D_i^T\tilde{h}_i\dot{\omega}_i + D_j^T\tilde{h}_j\dot{\omega}_j - D_i^T\tilde{h}_{i0}\ddot{\phi}_i + D_j^T\tilde{h}_{j0}\ddot{\phi}_j = D_i^T\tilde{\omega}_i(\tilde{h}_i\omega_i + 2\tilde{h}_{i0}\dot{\phi}_i) - D_j^T\tilde{\omega}_j(\tilde{h}_j\omega_j + 2\tilde{h}_{j0}\dot{\phi}_j) \tag{22}$$

Since there are only two independent constraint equations, three independent equations are obtained by taking the cross product of Eq. (21) with $\mathbf{h}$ and adding $\lambda \mathbf{h}\mathbf{h}^T\tau$. Note that, the component of the reaction torque in the pin direction must be zero and λ is an arbitrary number; i.e., $\lambda \mathbf{h}^T\tau = 0$. The final form of the acceleration equation is:

$$\begin{aligned}(I_3-hh^T)D_i^T\dot{\omega}_i - (I_3-hh^T)D_j^T\dot{\omega}_j - D_i^T\tilde{h}_i\tilde{h}_{i0}\mu_i\tilde{a}_i + D_j^T\tilde{h}_j\tilde{h}_{j0}\ \mu_j\tilde{a}_j + \lambda hh^T\mathbf{T} = \\ D_i^T\tilde{h}_i\tilde{\omega}_i(\tilde{h}_i\omega_i+2\tilde{h}_{i0}\dot{\phi}_i) - D_j^T\tilde{h}_j\tilde{\omega}_j(\tilde{h}_j\omega_j+2\tilde{h}_{j0}\dot{\phi}_j)\end{aligned} \tag{23}$$

Again, based on the acceleration equations the following submatrices are defined:

$$\begin{aligned}B_{22} &= [\overset{i}{(I_3-hh^T)D_i^T}\ ,\ \overset{j}{(I_3-hh^T)D_j^T}] \\ B_{2a} &= [\overset{i}{-D_i^T\tilde{h}_i\tilde{h}_{i0}\mu_i}\ ,\ \overset{j}{D_j^T\tilde{h}_j\tilde{h}_{j0}\mu_j}] \\ B_{24} &= \lambda hh^T \\ \nu_2 &= D_i^T\tilde{h}_i\tilde{\omega}_i(\tilde{h}_i\omega_i+2\tilde{h}_{i0}\dot{\phi}_i) - D_j^T\tilde{h}_j\tilde{\omega}_j(\tilde{h}_j\omega_j+2\tilde{h}_{j0}\dot{\phi}_j)\end{aligned} \tag{24}$$

System of Interconnected Flexible Bodies

Considering Eqs. (9), (17), and (23), the constraint equations of motion of a system of flexible bodies are written as:

$$\begin{aligned}M_{tt}\ddot{x} + M_{tr}\dot{\omega} + M_{ta}\ddot{a} + A_{t1}f &= u_t \\ M_{tr}^T\ddot{x} + M_{rr}\dot{\omega} + M_{ra}\ddot{a} + A_{21}f + A_{22}\tau &= u_r \\ M_{ta}^T\ddot{x} + M_{ra}^T\dot{\omega} + M_{aa}\ddot{a} + A_{a1}f + A_{a2}\tau &= u_a\end{aligned} \tag{25}$$

and the kinematic constraint equations are:

$$\begin{aligned}B_{11}\ddot{x} + B_{12}\dot{\omega} + B_{1a}\ddot{a} &= \nu_1 \\ B_{22}\dot{\omega} + B_{2a}\ddot{a} + B_{24}\tau &= \nu_2\end{aligned} \tag{26}$$

In the above equations, all system matrices are defined in terms of submatrices in Eqs. (10),

(11), (18) and (24) and

$$A_{11} = B_{11}^T \quad , \quad A_{21} = B_{12}^T \quad , \quad A_{22} = B_{22}^T \quad , \quad A_{a1} = B_{1a}^T \quad , \quad A_{a2} = B_{2a}^T \tag{27}$$

Equation (25) may be decoupled with simple manipulations to have a similar form to the equations of motion of systems with rigid bodies as presented by Fleck and Butler (1981):

$$\begin{aligned} M'\ddot{x} &+ A'_{11}f + A'_{12}\tau = u'_1 \\ J'\dot{\omega} &+ A'_{21}f + A'_{22}\tau = u'_2 \\ M_{aa}\ddot{a} &+ A'_{a1}f + A'_{a2}\tau = u'_a \end{aligned} \tag{28}$$

where

$$\begin{aligned} M' &= d_{11} - d_{12}d_{22}^{-1}d_{12}^T \\ J' &= d_{22} - d_{12}^T d_{11}^{-1}d_{12} \\ d_{11} &= M_{tt} - M_{ta}M_{aa}^{-1}M_{ta}^T \\ d_{12} &= M_{tr} - M_{ta}M_{aa}^{-1}M_{ra}^T \\ d_{22} &= M_{rr} - M_{ra}M_{aa}^{-1}M_{ra}^T \end{aligned} \tag{29}$$

$$\begin{aligned} A'_{11} &= A_{11} - M_{ta}M_{aa}^{-1}A_{a1} - d_{12}d_{22}^{-1}(A_{21} - M_{ra}M_{aa}^{-1}A_{a1}) \\ A'_{12} &= \quad - M_{ta}M_{aa}^{-1}A_{a2} - d_{12}d_{22}^{-1}(A_{22} - M_{ra}M_{aa}^{-1}A_{a2}) \\ u'_1 &= u_t - M_{ta}M_{aa}^{-1}u_a - d_{12}d_{22}^{-1}(u_r - M_{ra}M_{aa}^{-1}u_a) \end{aligned} \tag{30}$$

$$\begin{aligned} A'_{21} &= A_{21} - M_{ra}M_{aa}^{-1}A_{a1} - d_{12}^T d_{11}^{-1}(A_{11} - M_{ta}M_{aa}^{-1}A_{a1}) \\ A'_{22} &= A_{22} - M_{ra}M_{aa}^{-1}A_{a2} - d_{12}^T d_{11}^{-1}(\quad - M_{ta}M_{aa}^{-1}A_{a2}) \\ u'_2 &= u_r - M_{ra}M_{aa}^{-1}u_a - d_{12}^T d_{11}^{-1}(u_t - M_{ta}M_{aa}^{-1}u_a) \end{aligned} \tag{31}$$

$$\begin{aligned} A'_{a1} &= A_{a1} - M_{ta}^T M'^{-1}A'_{11} - M_{ra}^T J'^{-1}A'_{21} \\ A'_{a2} &= A_{a2} - M_{ta}^T M'^{-1}A'_{12} - M_{ra}^T J'^{-1}A'_{22} \\ u'_a &= u_a - M_{ta}^T M'^{-1}u'_1 - M_{ra}^T J'^{-1}u'_2 \end{aligned} \tag{32}$$

Solving Eq. (28) for the accelerations:

$$\ddot{x} = M'^{-1} (u'_1 - A'_{11} f - A'_{12} \tau)$$
$$\dot{\omega} = J'^{-1} (u'_2 - A'_{21} f - A'_{22} \tau)$$
$$\ddot{a} = M_{aa}^{-1} (u'_a - A'_{a1} f - A'_{a2} \tau)$$

Substituting Eq. (33) into Eq. (26) results in a system of linear simultaneous equations which can be solved for **f** and τ:

$$\begin{aligned} C_{11} f + C_{12} \tau &= v_1^* \\ C_{21} f + C_{22} \tau &= v_2^* \end{aligned} \tag{34}$$

where

$$\begin{aligned} C_{11} &= B_{11} M'^{-1} A'_{11} + B_{12} J'^{-1} A'_{21} + B_{1a} M_{aa}^{-1} A'_{a1} \\ v_1^* &= B_{11} M'^{-1} u'_1 + B_{12} J'^{-1} u'_2 + B_{1a} M_{aa}^{-1} u'_a - v_1 \end{aligned} \tag{35}$$

$$\begin{aligned} C_{12} &= B_{11} M'^{-1} A'_{12} + B_{12} J'^{-1} A'_{22} + B_{1a} M_{aa}^{-1} A'_{a2} \\ C_{12} &= B_{22} J'^{-1} A'_{21} + B_{2a} M_{aa}^{-1} A'_{a1} \end{aligned} \tag{36}$$

$$\begin{aligned} C_{22} &= B_{22} J'^{-1} A'_{22} + B_{2a} M_{aa}^{-1} A'_{a2} - B_{24} \\ v_2^* &= B_{22} J'^{-1} u'_2 + B_{2a} M_{aa}^{-1} u_a - v_2 \end{aligned} \tag{37}$$

Example

Figure 3 shows the fifteen segment ATB model that is commonly used in car crash and aircraft ejection simulations. The model has 14 joints all of the free type except for the knees (J_6 & J_9) and the elbows (J_{12} & J_{14}) which are pin joints (Obergefell, et al 1988).

The basic sled test simulation presented by obergefell et al. (1988) is used here as an standard to show the effect of small deformation on gross motion of human body. The two lower leg segments (7 &10) are treated as flexible bodies. Due to lack of time and availability of a finite element software, the two bodies are modeled as identical simple cantilever beams fixed at the

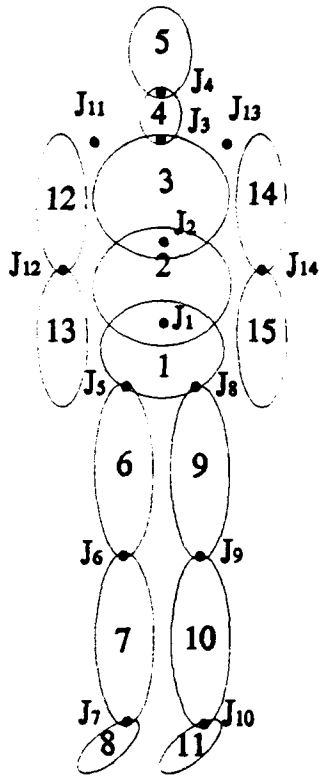
knee ends and free at the ankles. A simple finite element program was developed to determine the mode shapes, stiffness, and damping characteristics of the beam using fourteen 3D beam elements (15 nodes).

The inertial reference frame is selected such that Z-axis points downward, X-axis points forward, and Y-axis is the lateral direction by the right hand rule. Therefore, the forward motion of the vehicle is in the positive X-axis. Figures 4 through 8 show the comparison between the rigid and flexible lower leg models. Figures 4 and 5 are the plots of the right lower leg's center of mass forward velocity and acceleration. Figure 6 shows the right lower leg's angular velocity about the lateral axis. Figures 7 and 8 are the plots of right knee joint reaction force in forward direction and reaction torque about the lateral axis.

It can be seen from the figures that the segments may experience accelerations and of larger magnitude than those predicted by the rigid body model. Similarly, joints may also experience larger forces than those predicted by the rigid body model.

Conclusions

The Articulated Total Body (ATB) model has been revised to incorporate the small elastic deformation effects of its individual segments. The finite element method has been employed to determine the mode shapes and modal stiffness and damping characteristics of the flexible segments. It is shown through a simple example that a flexible body model may be necessary to accurately predict the gross motion of the ATB particularly its segment accelerations and joint forces.



Joint j connects segment JNT(j) with segment j+1

JNT(j) = 1 2 3 4 1 6 7 1 9 10 3 12 3 14

Figure 3. Fifteen Segment Body Configuration

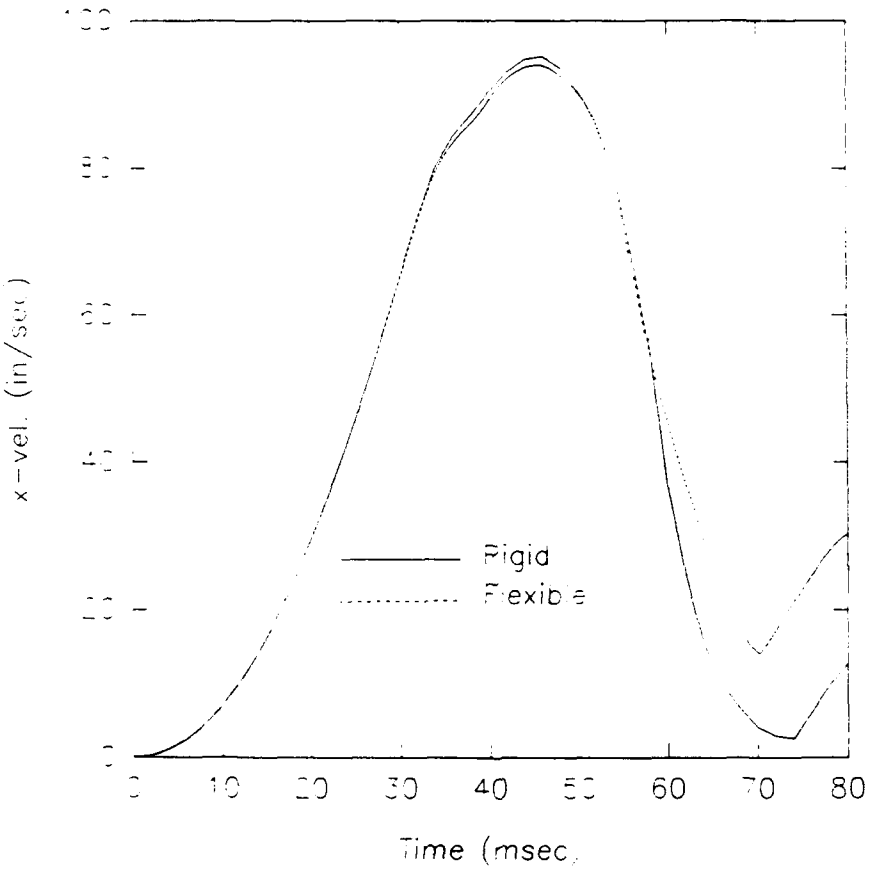


Figure 4. Right Lower Leg Velocity Relative to the Vehicle in X-Direction

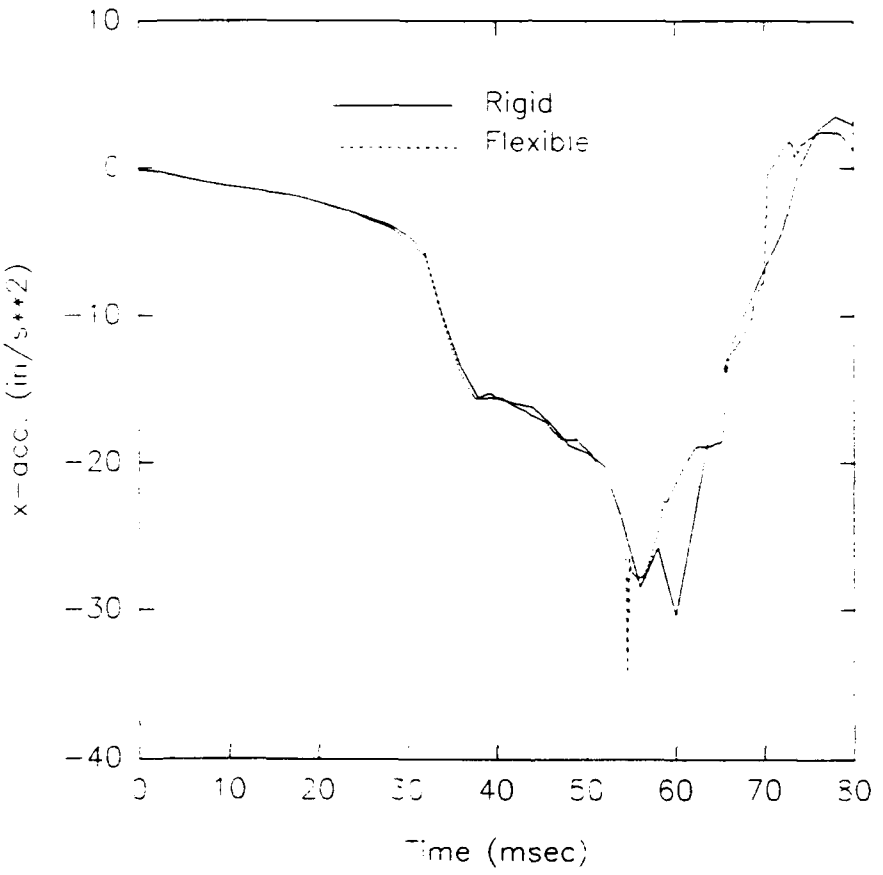


Figure 5. Right Lower Leg Acceleration Relative to the Vehicle in X-Direction

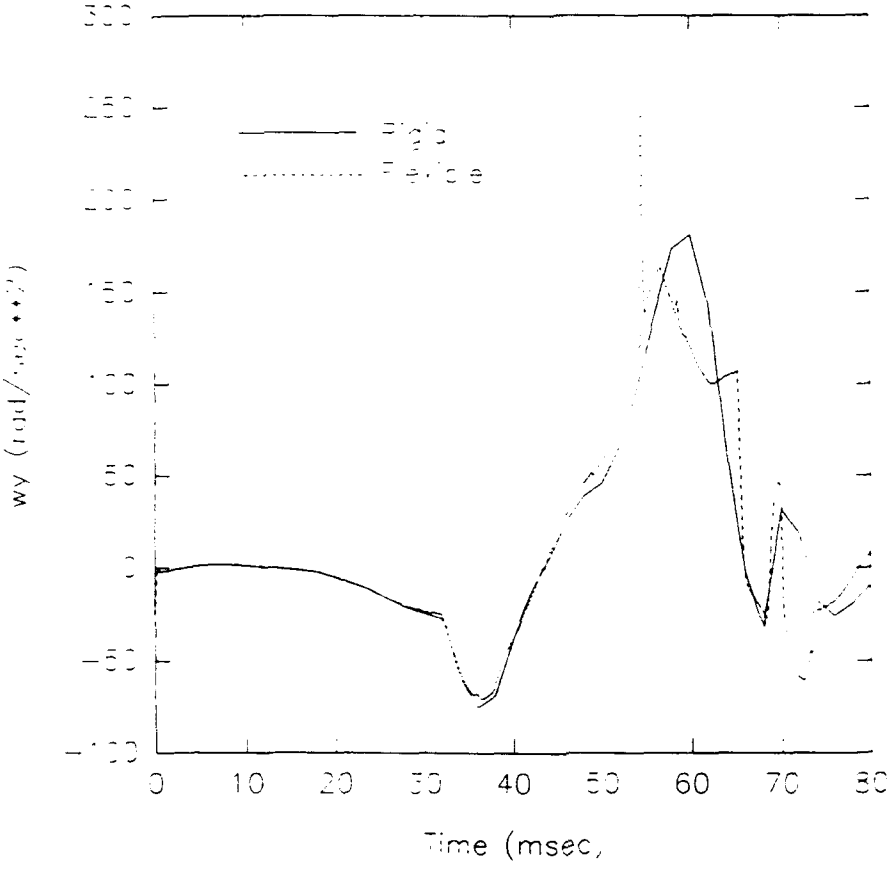


Figure 6. Right Lower Leg Angular Acceleration about Y-Axis

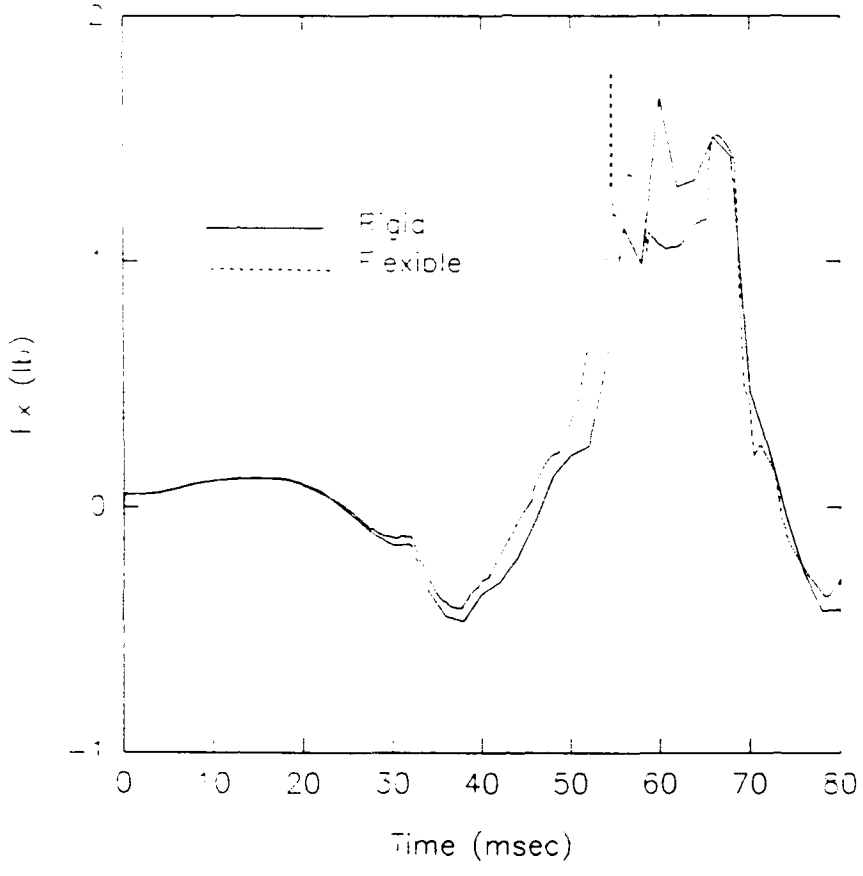


Figure 7. Right Knee Joint Force in X-Direction

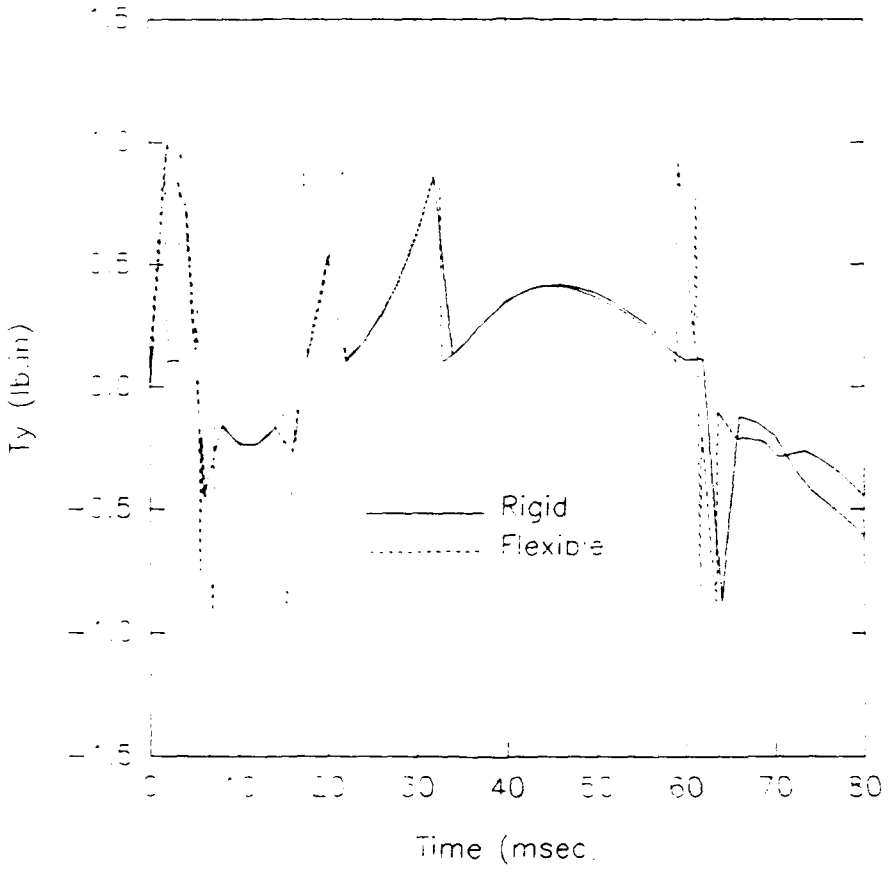


Figure 8. Right Knee Joint Torque about Y-Axis

REFERENCES

Clough, R.W., 1971, "Analysis of Structural Vibrations and Dynamic Response," in Recent Advances in Matrix Methods of Structural Analysis and Design, R. H. Gallagher, Y. Yamada, and J. R. Oden, editors., University of Alabama Press, Huntsville, Alabama.

Fleck, J.T., and Butler, F. E., 1981, "Validation of the Crash Victim Simulator, Volume 1, Engineering Manual - Part I: Analytical Formulation," Report No. ZS-5881-V-1, National Technical Information Service, Springfield, Virginia.

Greenwood, D. T., 1988, Principles of Dynamics, Second edition, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Obergefell, L. A., Gardner, T. R., Kaleps, I., and Fleck, J. T., 1988, "Articulated Total Body Model Enhancements, Volume 2: User's Guide," Technical Report AAMRL-TR-88-043, National Technical Information Service, Springfield, Virginia.

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.

Winfrey, R. C., 1971, "Elastic Link Mechanism Dynamics," Journal of Engineering for Industry, Vol. 93, pp. 268-272.

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

Yoo, W. S., and Haug, E. J., 1986, "Dynamics of Flexible Mechanical Systems Using Vibration and Static Correction Modes," Journal of Mechanisms, Transmissions, and Automation in Design, Vol. 108, pp. 315-322.

MULTIMODAL MEASURES OF MENTAL WORKLOAD DURING COMPLEX TASK PERFORMANCE

Richard W. Backs

Assistant Professor

Department of Psychology

and

Arthur M. Ryan

Graduate Student

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, Washington, D.C.

September, 1993

MULTIMODAL MEASURES OF MENTAL WORKLOAD DURING COMPLEX TASK PERFORMANCE

Richard W. Backs
Assistant Professor
Department of Psychology

and

Arthur M. Ryan
Graduate Student
Department of Psychology

Wright State University

Abstract

Central and autonomic nervous system measures of mental workload were examine concurrently during tasks that varied in their perceptual/central and physical demands. A cognitive arithmetic and continuous manual tracking task were performed singly and together. The perceptual/central demand of the cognitive arithmetic task was manipulated by varying the number of addition and subtraction operations required to solve a problem. The physical demand of a single-axis, second-order compensatory tracking task was manipulated by varying the amount of force operators had to apply to the joystick. Multiple psychophysiological responses were recorded during task performance including: electroencephalographic, cardiovascular, pulmonary, and eye blink measures. Data will be collected from twenty-four subjects, but only preliminary analyses on a subset of responses from selected subjects are available at this time.

MULTIMODAL MEASURES OF MENTAL WORKLOAD DURING COMPLEX TASK PERFORMANCE

Richard W. Backs and Arthur M. Ryan

Introduction

Mental workload (MWL) has been described as an intervening variable that reflects the extent to which the information processing abilities of an operator are actively engaged during task performance (Gopher & Donchin, 1986). MWL is a multidimensional construct that has been assessed using performance, subjective, and physiological measures that have often been observed to dissociate during task performance (see O'Donnell & Eggemeier (1986) for a review). Some investigators have emphasized that these between-class dissociations are important sources of information regarding the structure of cognitive resources underlying complex task performance (Wickens, 1990; Yeh & Wickens, 1988). Measure dissociation can also occur within each of the performance, subjective, and physiological classes that may elucidate cognitive resource structure. Of interest in the present study is the pattern of association and dissociation observed among physiological measures, specifically the relation between measures of central nervous system activity (i.e., the electroencephalogram and event-related potential) and measures of autonomic nervous system activity (i.e., the electrocardiogram, impedance cardiogram, and respiration).

In the present study, MWL is assumed to result from the consumption of limited capacity cognitive resources, where increased resource utilization leads to increased MWL. The structure of these resources is assumed to conform to Wickens's (1980; 1984) multiple resource model. According to this view, attentional capacity is considered to be limited within three independent resource dimensions: processing stages (perceptual/central, response); processing codes (verbal, spatial); and input/output modalities (auditory and visual input, manual and speech output). The present study focused on task demands that affect resources along the processing stage dimension.

Historically, event-related potentials (ERPs) have played a prominent role in demonstrating the veracity of the processing stage dimension (e.g., Isreal, Chesney, Wickens, & Donchin, 1980; Isreal, Wickens, Chesney, & Donchin, 1980; Sirevaag, Kramer, Coles, & Donchin, 1989). For example, manipulations of tracking system order increase the demand upon perceptual/central resources and reduce the amplitude of the P300 to a secondary task, while manipulations of disturbance bandwidth increase the demand upon response resources and do not affect the P300 to a secondary task. The effects of these manipulations on cardiovascular measures is unclear. On the one hand, many cardiovascular measures (e.g., heart rate: Jennings, 1986ab; heart rate variability: Mulder & Mulder, 1981; Grossman & Svebak, 1987) are thought to index central processing and would be expected to correlate positively with ERP measures. On the other hand, Backs, Ryan, and Wilson (in press) did not find that heart period or heart rate variability were sensitive to tracking system order.

A cognitive/energetic approach to information processing may be useful in understanding central and autonomic dissociations. According to Kahneman's (1973) energetic resource model, total information processing capacity can vary, with the capacity limit determined by the operator's state of arousal. Autonomic measures are considered to reflect the regulation of the capacity limit. As tasks increase in difficulty they consume more resources; however, if the resources available are sufficient to perform the task the capacity limit will not increase with increased task difficulty. In the case of perceptual/central resources, the ERP measures would be expected to vary with task difficulty, but not the autonomic measures. Only when perceptual/central resource availability is not sufficient to perform the task, and more resources are needed, will the limit increase. In this case, both the ERP and the autonomic measures would be expected to vary with task difficulty.

Alternatively, central and autonomic nervous system measures may dissociate as a result of the dual innervation of the autonomic nervous system. Effector responses such as heart period are determined by a combination of sympathetic and parasympathetic inputs that may cancel each other under some circumstances. Thus, when cardiovascular measures based upon heart period

dissociate from the ERP it may not be because the heart was unaffected by the by the task. but because of the autonomic mode of control operating during task performance (Berntson. Cacioppo, & Quigley, 1991). Only when the sympathetic and parasympathetic inputs are separated will task effects be apparent.

A difficulty with the use of autonomic responses, and with heart period in particular, is that MWL effects upon the response tend to be small in relation to the primary, metabolic, function of the physiological system. Effects of MWL may be masked by homeostatic adjustments of the autonomic response to variables such as the physical workload required to maintain task performance and the level of arousal or stress, which may or may not be associated with task performance. Thus, the signal-to-noise ratio of the MWL effects on the autonomic response may decrease as background activity due to variables such as physical workload increases. However, heart period responses to MWL may still be able to be separated from responses to physical workload by their autonomic mode of control.

In the present study, central and autonomic measures were examined concurrently while perceptual/central processing resource demands of a cognitive arithmetic task and the physical demand of a second-order compensatory manual tracking task were varied. Participants performed the cognitive arithmetic task and the tracking task singly and together. The central measures were expected to be sensitive to the perceptual/central demand of the arithmetic task. but not to the physical demand of the tracking task. The sensitivity of the cardiovascular measures was assessed separately for the individual effector responses and for the sympathetic and parasympathetic inputs as determined statistically across the measures (Backs, in press). Specifically, perceptual/central demand was expected to elicit primarily sympathetic activity consistent with the central measures, while physical demand was expected to elicit primarily parasympathetic activity independent of the central measures.

Method

Subjects

Sixteen subjects (9 female) from the subject pool maintained by Armstrong Laboratory participated have participated in the experiment. A total of 24 subjects will be run.

Apparatus

The electrocardiogram (ECG), electro-occulogram (EOG), and electromyogram (EMG) data were amplified by Grass amplifiers (Model P511K) and sampled at 1000 Hz by the Psychophysiological Assessment Test System (Wilson & Oliver, 1991). The impedance cardiogram (ZCG) was collected using a Minnesota Impedance Cardiograph (Model 304B), and stored on magnetic tape for later analysis. Respiration was collected using a Respitrace (Model 10.9000). The electroencephalogram (EEG) was collected at 100 Hz with a Biologic Brain Atlas topographic mapping system.

The task was controlled by a microcomputer. The joystick was a Measurement Systems (Model 446) force joystick.

Tasks

The cognitive arithmetic task was the Math Processing task from the Criterion Task Set (Shingledecker, 1984). Perceptual/central processing was varied across two levels using the medium and high difficulty Math Processing tasks. These tasks consist of two (medium) or three (high) operation addition and subtraction problems, where the operator's task is to decide whether the answer is greater than or less than five. The Criterion Task Set version of the task was modified for use in the present study so that the proportion of addition and subtraction operations and the proportion of problems that required carry operations was balanced within each task. Thirty-six problems were visually presented at a fixed interstimulus interval of 5 s in each 3 min. trial. All operators responded with the first (greater than) and second (less than) fingers of their left hand. Problems were terminated after the button push or 4 s, whichever came first. Participants were instructed to respond as fast as possible while maintaining 100% accuracy, and

given performance feedback after each trial.

The physical demand of a single-axis, second-order compensatory manual tracking task was varied at two levels. The input to the tracking system was scaled so that either 4 or 8 lbs. of force was required for full range movement of the joystick. The tracking cursor changed color when operators applied too much force to the joystick. Operators performed the tracking task with their right hand. Performance feedback regarding tracking error and whether too much force was applied was given at the end of each trial.

Procedure

Operators participated for four sessions distributed over successive days. The first three sessions (1.5 hrs. each) were practice; participants received 20 3-min. trials on each day. A total of 20 single task trials were presented: 4 medium and 6 high difficulty cognitive arithmetic; 10 tracking, 5 each at 4 lbs. and 8 lbs. of force. A total of 40 dual task trials were presented: 9 medium and 11 high difficulty cognitive arithmetic with each tracking force condition. Participants wore the physiological transducers on Day 3 for adaptation purposes. On the test day, participants performed one trial of each of the eight task conditions, where condition presentation order was balanced across participants with a Latin square. The block of task conditions was preceded and followed by a 3 min. eyes-open resting baseline.

Data Reduction

Performance. Reaction time in ms for correct trials was measured for the cognitive arithmetic task. Probability correct was examined as a manipulation check to make sure that no speed-accuracy trade-off occurred. Root mean square (RMS) error was measured for the tracking task. Stick RMS was measured as a manipulation check on the force manipulation.

Subjective. The NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988) was collected after each test trial. The NASA-TLX contains six subscales that are each given rated on a scale from 0 to 100: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The mean of the six subscales was used for analysis, where higher scores mean

greater MWL. The physical demand scale was examined separately as another manipulation check on the tracking force level.

EEG. EEG was filtered to pass between 1 to 30 Hz with the 60 Hz notch filter in. The middle minute of the trial was submitted to power spectral analysis.

ERP. Artifact-free trials were converted to source derivations (Hjorth, 1975, 1980) and averaged to obtain the ERP.

ECG. ECG was conditioned with a gain of 2,000 and a half-amplitude bandpass of 10 - 100 Hz. Heart period was defined as the time between successive heart beats measured in ms from the R wave-to-R wave interval (i.e., the interbeat interval (IBI)). Mean IBI across the trial was analyzed. Two measures of heart rate variability were derived from the IBIs across the trial, the Traube-Hering-Mayer (THM) wave (0.06 - 0.14 Hz) and respiratory sinus arrhythmia (RSA, 0.15 - 0.4 Hz).

ZCG. Systolic time intervals were derived from ensemble-averaged ZCG (Kelsey & Guethlein, 1990). Pre-ejection period (PEP) was used as a measure of cardiovascular sympathetic activation.

Respiration. Respiration rate in breaths-per-minute and tidal volume in mL were measured across the 3-min. trial.

EMG. EMG was conditioned with a gain of 10,000 and a half-amplitude bandpass of 1-1000 Hz. Total power (uv^2) from 1-500 Hz was summed for analysis.

Results

Due to equipment failure, only a subset of data from selected subjects was currently available for analysis.

ERP

Twelve subjects were available for analysis, but not all are represented in the data for each experimental condition: seven subjects for the single task medium difficulty cognitive arithmetic

task; ten subjects for the high difficulty cognitive arithmetic task; eleven subjects for the 4 lbs. force-medium cognitive arithmetic dual task; eleven subjects for the 4 lbs. force-high cognitive arithmetic dual task; ten subjects for the 8 lbs. force-medium cognitive arithmetic dual task; and eleven subjects for the 8 lbs. force-high cognitive arithmetic dual task. Since we were interested in a central nervous system measure of perceptual/central processing, we focused our analyses on the P300 component. Topographic maps of the P300 are presented in Figure 1 for the single tasks and in Figure 2 for the dual tasks.

Because the P300 is most prominent over the parietal cortical area, we selected the P_3, P_z, and P_4 electrode sites for closer inspection. Group mean Hjorth-corrected amplitudes from the three sites in each of the six tasks are presented in Figure 3. Several trends are apparent in the figure. First, P300 amplitude is greater in high difficulty cognitive arithmetic (MH) than in medium difficulty cognitive arithmetic (MM), consistent with the interpretation that more perceptual/central resources are required to process more difficult arithmetic problems. The second trend is that P300 amplitude to the arithmetic problems is reduced during the dual tasks compared to the single tasks. This reduction in P300 amplitude was greatest for the high difficulty cognitive arithmetic dual tasks (4H and 8H for the 4 and 8 lbs. force conditions, respectively) compared to the cognitive arithmetic single task (MH); the reduction for the medium difficulty dual tasks (4M and 8M for the 4 and 8 lbs. force conditions, respectively) compared to the cognitive arithmetic single task (MM) appeared to be smaller than for the difficult cognitive arithmetic task. This pattern of results is also consistent with the interpretation that the tracking task also requires perceptual/central processing resources, and when tracking and cognitive arithmetic are performed currently they compete for these resources. Further, resource competition was greatest for the high difficulty cognitive arithmetic task. Even though the cognitive arithmetic task was treated as having the highest priority, there appears to be insufficient perceptual/central capacity to allocate the single task level of resources to the arithmetic problem in the dual task.

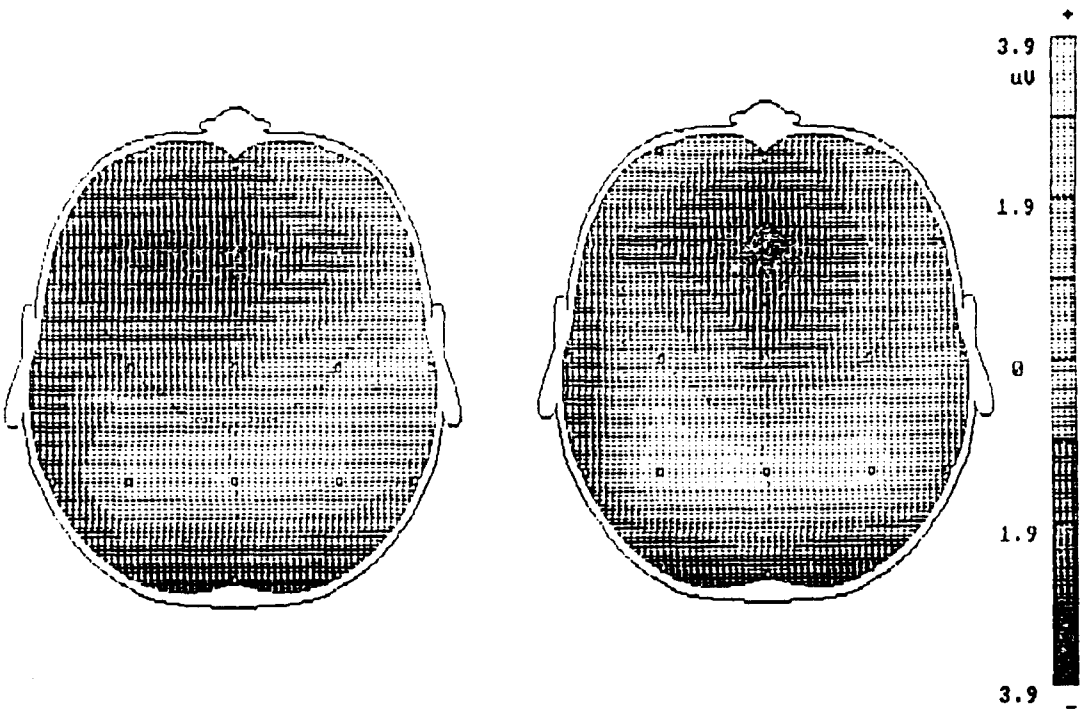


Figure 1. Topographic maps at 300 ms after the onset of the arithmetic problem for the medium (left) and high (right) difficulty cognitive arithmetic single tasks.

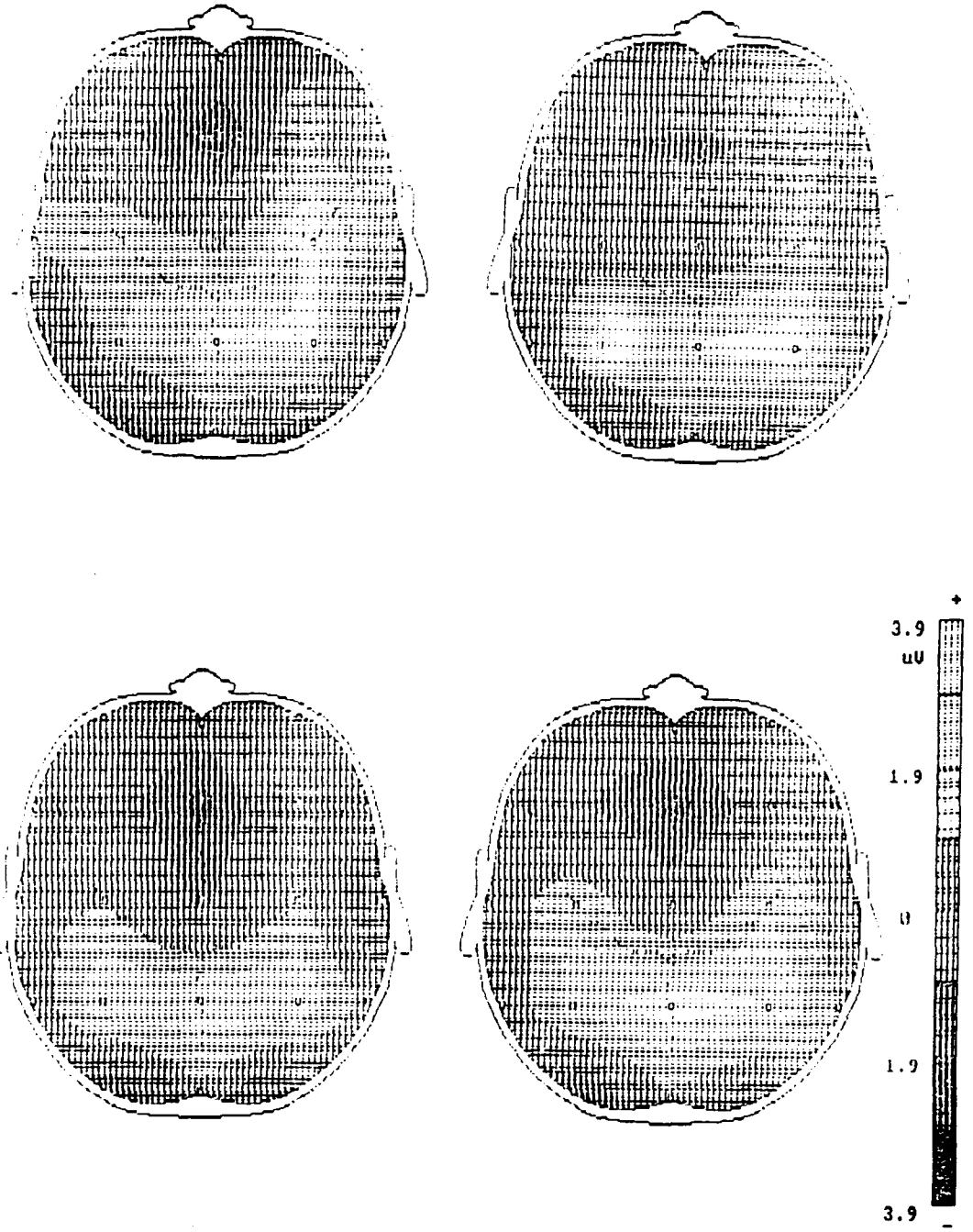


Figure 2. Topographic maps at 300 ms after the onset of the arithmetic problem for the cognitive arithmetic dual tasks. The top row is medium and the bottom row is high difficulty cognitive arithmetic. The left column is 4 lbs. force and the right column is 8 lbs. force tracking.

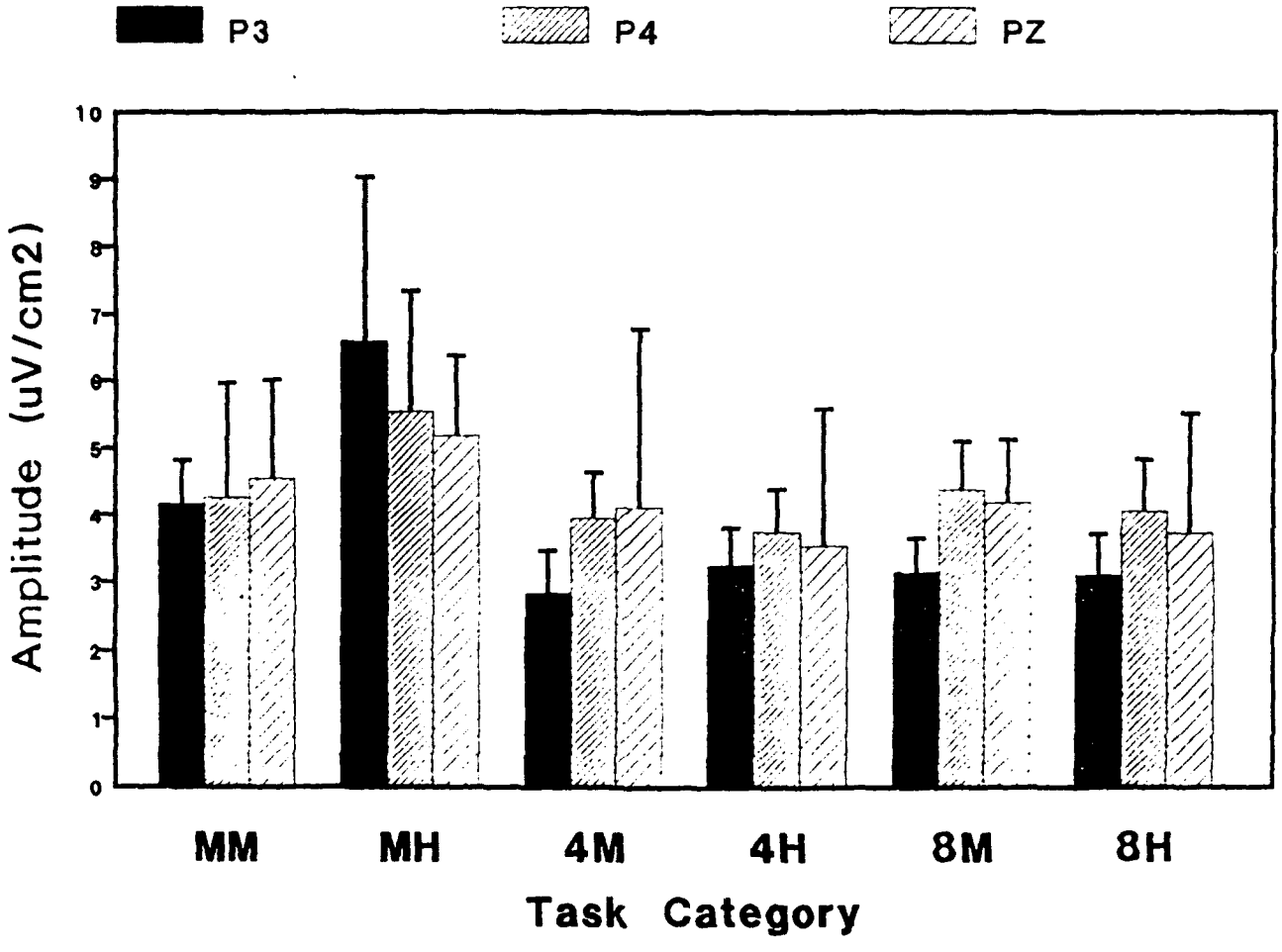


Figure 3. P300 amplitudes across the parietal electrode sites for the six experimental conditions: MM = medium difficulty cognitive arithmetic single task; MH = high difficulty cognitive arithmetic single task; 4M = medium difficulty cognitive arithmetic with 4 lbs. force tracking dual task; 4H high difficulty cognitive arithmetic with 4 lbs. force tracking dual task; 8M = medium difficulty cognitive arithmetic with 8 lbs. force tracking dual task; and 8H = high difficulty cognitive arithmetic with 8 lbs. force tracking dual task.

ECG

Heart period from the six experiment conditions in which P300s were collected was available five subjects. As can be seen in Figure 4, group mean heart period is shorter (i.e., faster heart rate) in the high (MH) than in the medium (MM) difficulty cognitive arithmetic single tasks consistent with the P300 results. Also like the P300, heart period was faster for the dual tasks than for the single tasks. The results for heart period across the dual tasks were different from the P300 results, however. While heart period was faster for the high than the medium difficulty cognitive arithmetic task, it was also faster for 8 lbs. force (high physical demand) than for the 4 lbs. force (low physical demand) condition. Further, the difference between high and medium difficulty cognitive arithmetic tasks was considerably reduced in the 8 lbs. force than in the 4 lbs. force conditions. This interaction illustrates the difficulty of using heart period to isolate MWL effects from background physiological activity driven by physical demand. Since heart period is sensitive to both mental and physical demand, the size of the MWL effect is reduced when physical demand is high.

Conclusions

Even with this small sample, it can be concluded that the experimental manipulations effectively established conditions in which our questions regarding central/autonomic nervous system interactions can be answered. Future analyses of the ZCG, heart rate variability, and other measures will provide a greater understanding of when and why psychophysiological MWL measures dissociate. This study will also help to clarify theoretical issues with regard to the effects of environmental variables such as physical workload upon cognitive processes. These effects are especially critical in operational settings where the environment is constantly changing. To be effective in the field psychophysiological measures of MWL must be robust across dynamic environmental conditions. This study will provide important information about how to increase the sensitivity and diagnosticity of psychophysiological MWL measures.

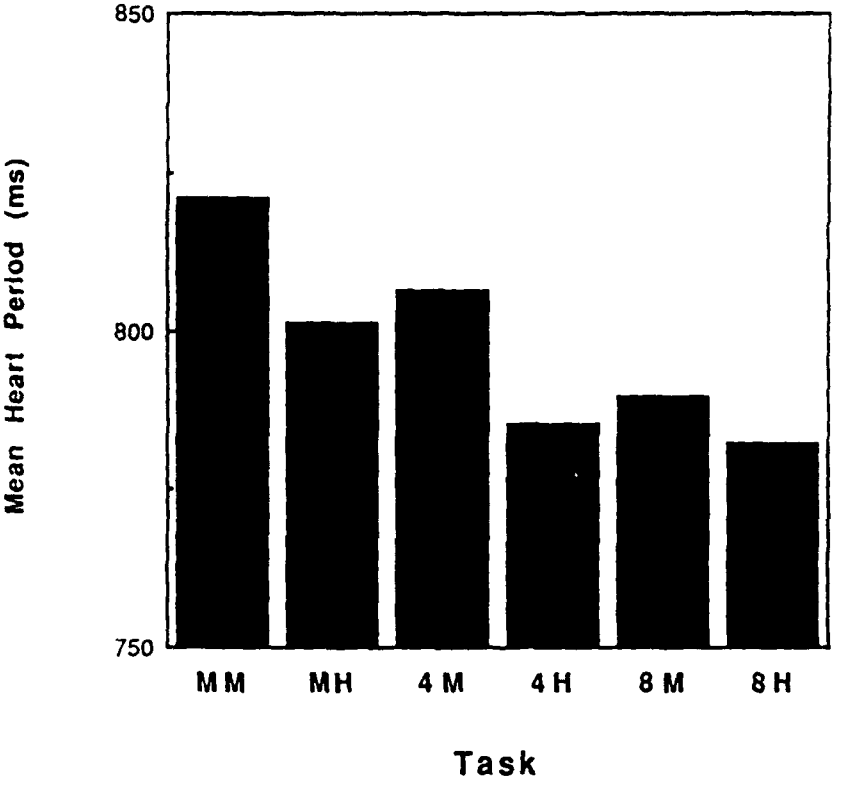


Figure 4. Mean heart period for the six experimental conditions in which P300 data were collected: MM = medium difficulty cognitive arithmetic single task; MH = high difficulty cognitive arithmetic single task; 4M = medium difficulty cognitive arithmetic with 4 lbs. force tracking dual task; 4H high difficulty cognitive arithmetic with 4 lbs. force tracking dual task; 8M = medium difficulty cognitive arithmetic with 8 lbs. force tracking dual task; and 8H = high difficulty cognitive arithmetic with 8 lbs. force tracking dual task.

References

Backs, R. W. (in press). Going beyond heart rate: Autonomic space and cardiovascular measures of mental workload. *International Journal of Aviation Psychology*.

Backs, R.W., Ryan, A.M., & Wilson, G.F. (in press). Psychophysiological measures of workload during continuous manual performance. *Human Factors*.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, **98**, 459-487.

Gopher, D., & Donchin, E. (1986). Workload-An examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance. Volume 2*. New York: John Wiley.

Grossman, P., & Svebak, S. (1987). Respiratory sinus arrythmia as an index of parasympathetic cardiac control during active coping. *Psychophysiology*, **24**, 228-235.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), Human Mental Workload, pp. 139-183. Amsterdam: North-Holland.

Hjorth, B. (1975). An on-line transformation of EEG scalp potentials into orthogonal source derivations. *Electroencephalography and Clinical Neurophysiology*, **39, 526-530.**

Hjorth, B. (1980). Source derivation simplifies topographical EEG interpretation. *Journal of EEG Technology*, **20**, 121-132.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual task performance. *Psychophysiology*, **17**, 57-70.

Isreal, J.B., Wickens, C.D., Chesney, G.L., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human Factors*, **22**, 211-224.

Jennings, J.R. (1986a). Bodily changes during attending. In M.G.H. coles, E. Donchin, & S.W. Porges (Eds.), *Psychophysiology: Systems, Processes, and applications*. New York: Guilford.

Jennings, J.R. (1986b). Memory, thought, and bodily response. In M.G.H. coles, E. Donchin, & S.W. Porges (Eds.), *Psychophysiology: Systems, Processes, and applications*. New York: Guilford.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kelsey, R. M., & Guethlein, W. (1990). An evaluation of the ensemble averaged impedance cardiogram. *Psychophysiology*, **27**, 24-33.

Molen, M. van der, Somsen, R.J.M., Jennings, J.R., Nieuwboer, R.T., & Orlebeke, J.F. (1987). A psychophysiological investigation of cognitive-energetic relations in human information processing: A heart rate/additive factors approach. *Acta Psychologia*, **66**, 251-289

Mulder, G., & Mulder, L.M.J. (1981). Information processing and cardiovascular control. *Psychophysiology*, **18**, 392-402.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance, Volume 2*. New York: John Wiley.

Shingledecker, C. A. (1984). A task battery for applied human performance assessment research. AFAMRL-TR-84-071. Dayton, OH: Air Force Aeromedical Research Laboratory

Sirevaag, E.J., Kramer, A.F., Coles, M.G.H., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis, *Acta Psychologia*, **70**, 77-97.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and Performance VIII*. Hillsdale NJ: Erlbaum.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman, & D.R. Davies (Eds.), *Varieties of attention*. Orlando, FL: Academic.

Wickens, C.D. (1990). Applications of event-related potential research to problems in human factors. In J.W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic Issues and applications*. New York: Oxford.

Wilson, G. F., & Oliver, C. G. (1991). PATS: Psychophysiological assessment test system. In E. Farmer (Ed.), *Proceeding of the XVIII WEAAP Conference: Volume II*. Aldershot, UK: Avebury Technical.

Yeh, Y.Y., & Wickens, C.D. (1988). Dissociation of performance nd subjective measures of workload. *Human Factors*, **30**, 111-120.

A REVIEW OF FLAIL AND WINDBLAST INJURIES

C.J. Charles Chuong

Associate Professor

Biomedical Engineering Program

University of Texas, Arlington

501 W 1st Street

Arlington, TX 76019

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 1993

A REVIEW OF FLAIL AND WINDBLAST INJURIES

C.J. Charles Chuong
Associate Professor
Biomedical Engineering Program
University of Texas, Arlington

Abstract

During an open seat ejection escape from a disabled aircraft, a crew member is exposed to high velocity airstream abruptly. Severe limb flail (windblast) injuries can occur as a result of this exposure. Typical injuries are described as muscle tears, joint dislocation and/or long bone fracture. This report reviews various aspects of limb flail injury from available literature and reports. Topics include injury description, statistics, mechanisms, and measurements of flail loading. To fully understand the transient physical event of flail injuries, a finite element modeling effort is planned. Mechanical properties of relevant biological soft tissues and hard tissues are also reviewed and summarized in this report and they will be used as the database for continuing modeling effort.

A REVIEW OF FLAIL AND WINDBLAST INJURIES

C.J. Charles Chuong

Introduction

During an open seat ejection escape from a disabled aircraft, a crew member is exposed to high velocity airstream abruptly. Severe limb flail (windblast) injuries can occur as a results of this exposure and typical injuries are described as joint dislocation and/or long bone fracture.

The primary reason for flail injury is that the limbs are subjected to a much higher deceleration force than the body torso and that they are moved outward and backward with respect to the torso at high velocities from their stowed or initial position. The resultant relative movement of the limbs can quickly exceed the normal ranges of motion leading to tears and permanent injuries of muscle groups, ligament, tendon, and/or dislocation of joints in severe cases. The fast and freely swing motion of limbs are only stopped by a collision with seat structure or other body parts, which commonly results in long bone fracture.

The probability for flail injury increases as the airspeed becomes higher [3]. Under combat conditions, almost half of all the ejectees suffer flail injury or death [24]. The performance and capability of ejection seats are thus limited at higher speeds by the possible occurrence of flail injuries.

To improve the protective equipment and restraint systems for open seat ejection it is important to analyze and understand the mechanism of flail injuries in quantitative terms. With this objective in mind, a review of various aspects of flail injuries is

made in this report. Reviewed topics include injury description, statistics, mechanisms, and loading measurements. To fully understand the transient physical event of flail injuries, a finite element modeling effort is planned. Mechanical properties of relevant biological soft tissues and hard tissues are also reviewed and summarized in this report and they will be used as the database for continuing modeling effort.

Injury description

Typical flail injuries for head-neck, upper and lower extremities are summarized in the following table:

TABLE 1 A summary of flail injuries at different part of the body

Body parts	Injury descriptions
Head-neck	Mandible fracture
	Sprain at neck
	Laceration and/or contusion at head
	Cervical cord transection, subluxation
	Cervical disk rupture, hernia, crushing
	Tearing of paracervical muscle/ligament
	Fracture of spinous process
Upper extremity	Shoulder dislocation
	Clavicle fracture, scapula fracture
	Rib fracture, ribs separation
	Fracture at humerus, ulna, and/or radius
	Elbow fracture, dislocation
	Disrupted or torn brachial plexus
Lower extremity	Pelvis fracture
	Fracture at femur, fibula, and/or tibia
	Dislocation at hip joint, knee, or ankle
	Fracture at anterior malleolus, metacarpals

Preceding or parallel to the occurrence of trauma at skeletal structures, over-stretch or even tears of soft tissues, namely, muscle groups, nerves, ligament, and tendon etc. have also been reported. Depending on the extent of severity, injury may range from soft tissue contusion or laceration to major debilitating long bone fracture, or joint dislocation. Commonly, multiple occurrences of these injuries were reported.

Spinal injury and vertebrae fracture at cervical and thoracic levels have also been documented [e.g. 10,11,12,13]. These trauma have been attributed to the rapid acceleration loading during the ejection phase. It is very likely that the windblast loading on the body torso and the rapid flailing movement of the limbs can easily worsen the spinal injury and vertebrae fracture.

Buschman and Rittgers [3] analyzed the records of all the non-combat ejection escapes occurring during the 1957-70 period and found the predilection of injuries at the proximal portion of extremities, particularly, the knees and legs. In another study, Ring et al [27] reported 1) an absence of major head neck flail injury, 2) a predominance of proximal over distal injury, and 3) a slight predominance of upper over lower extremity flail injury, which is in contrast to the former study. The difference may be attributed to several reasons, including slightly different injury criteria and classification, different restraint system, aircraft, and ejection seat used.

Injury statistics data

The occurrence of injury is dependent upon many flight parameters, including types of aircraft, types of seat, airspeed at

ejection, and altitude, among others. Payne and Hawker [23] attempted to quantify the relationship between flail injury and escape speed in probablistic terms. They analyzed the incidence of flail injury for USAF non-combat open seat ejection that occurred during 1964-70 period and compared their data with early British experience. Their results showed that the probability of flail injury is normally distributed by the square of the indicated airspeed (Fig. 1) and that the probability of leg flail was about the same as arm flail injury. Additional statistics of USAF non-combat ejection experience during the period 1968-73 showed the incidence rose dramatically above 300 KIAS (knots indicated air speed), clearly indicating that flail injury is a significant problem at higher airspeeds [27].

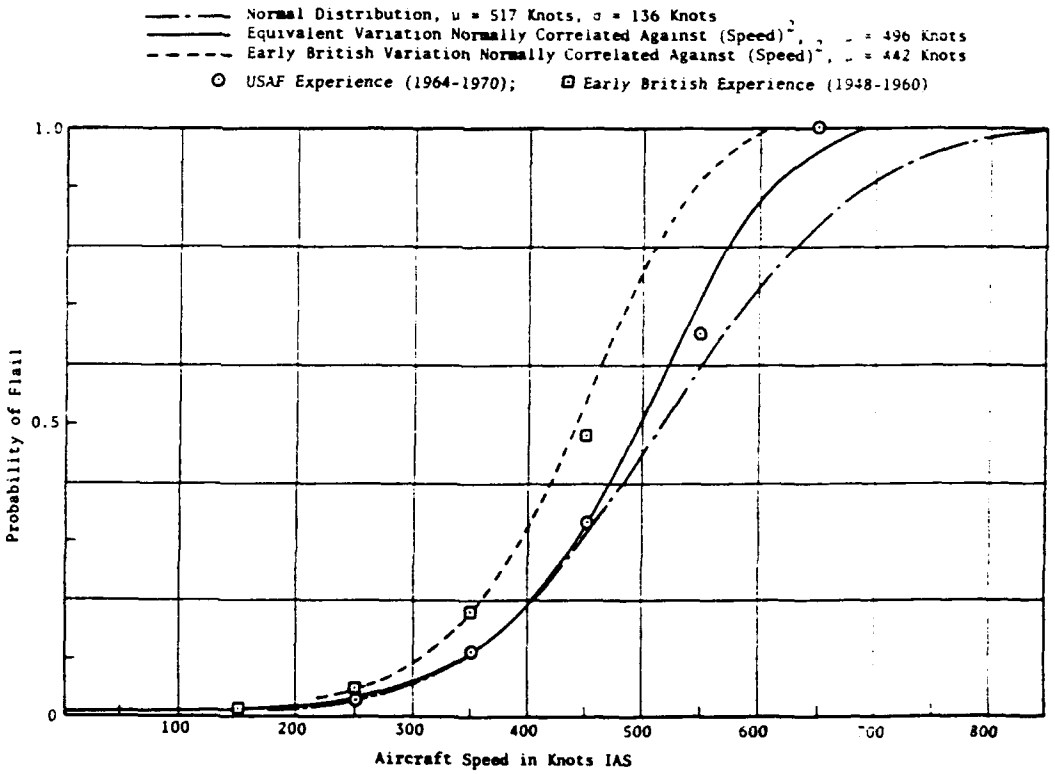


Fig. 1 Probability of flail as a function of airspeed at the time of ejection, from ref [23].

Others have reviewed the occurrence of flail injury from combat or POW experience for both USAF and the US Navy [7,8,14,15,16,30,32] and these results showed a much higher injury risk probability and a higher degree of severity in resulting injuries [24].

Belk studied the limb flail injuries and the effect of extremity restraints from a database covering the duration of 1971-78 [1]. He also used the probability approach to correlate the incurred injuries with the types of aircraft, KIAS on ejection, and the method of initiation. Comparisons of arm flail probabilities showed little difference between seats activated with D-ring and side arm controls. Seats with active leg restraints, however, were found to have a significant reduction in the frequency of lower extremity injury as well as an apparent reduction in the severity of injury.

Delgado [5] used the same criteria as Belk to analyze the records of injury incurred during the period of 1979 -1985. The effect of aircraft type, seat type, and airspeed on the probability of occurrence was studied but his results showed a lower mean airspeed for injury, possibly due to the use of ACES-II seats.

Mechanisms

When the ejection sequence starts, the seat occupant is exposed to high velocity airstream abruptly and large drag forces develop instantaneously. Throughout the deceleration phase, the seat occupant is subjected to the combined loading of forward acting inertial force and resisting drag forces, which varies in magnitude and direction through the whole sequence. Limb flailing injury is a consequence of different and inappropriate decelerations experienced by different body components during this time interval.

The initiation of flailing movement arises from the intrinsic instability of the occupant/seat system, which includes body torso, seat structure to which the torso is strapped, limbs, and the helmeted head. Because different components in the system have different geometric shapes and (surface area)/mass ratio, the aerodynamic drags developed are of different characteristics (magnitude, direction, and time histories). Further, because of varying stiffness and mass distribution, these body components have different dynamic response characteristics, e.g., frequency responses, structural damping, etc. Limb flailing movement, arising from these intrinsic instabilities, is easily triggered by the relative competition between the forward acting inertial force and the aerodynamic drag forces. In fact, high frequency oscillating forces acting at the limbs and the head have been reported based on laboratory experiments [20].

In theory, the drag forces tend to decelerate the occupant/seat system. The resulting deceleration can be written as:

$$Deceleration \quad \frac{Drag\ force}{Mass} \quad C_d \cdot \left(\frac{Area}{Mass}\right) \cdot \frac{1}{2}\ \rho\ V^2$$

where C_d is the drag coefficient, which depends on the geometric shape of the body components; and $1/2\ \rho\ V^2$ denotes the dynamic pressure. Payne first attributed the flailing movement of the limbs to their higher deceleration than the torso for their larger (Area/Mass) ratios [21,22]. Because of the differences in deceleration, limbs are pushed outward and backward with respect to the body torso at a relatively high speed. Using a potential flow solution, Schneck estimated the pressure gradient sufficient to lift the lower arm and to push it outward when a seated man is subjected

to windblast loading instantaneously [28]. Later, with the addition consideration of shear forces, he showed a possible formation of double vortex sheet on the down stream side of the limbs, which can result in oscillating forces tending to cause lateral vibration of the limbs [29].

Once the limb flailing motion starts, the roles of anatomical mobility and constraints of body components (joint, muscles, ligament, etc.) should be taken into consideration, as they affect the subsequent dynamic response and deceleration history of the limb structure. Through the limb flailing motion, besides the translation of limb structure and the rotational movement at joints, a limb structure also experiences bending moment, torsion, and axial forces, which cause deflection, angular twist, and axial shortening/elongation. The resultant motion and deformation of a limb structure can quickly exceed its normal ranges of motion and deformation limits, lead to overstretch and/or tears of muscle groups, ligament, tendon, and joint capsule dislocation in severe cases.

The motion set off can only be stopped by a collision with the relatively rigid seat structure or with other body parts leading to long bone fracture. During the impact, forces of large magnitude develop at the impacting site since there is a rapid transfer of kinematic energy from the fast moving limb to the target structure. This impacting force, transmitted into the long bone, can lead to pronounced bending and torsion loading to cause instantaneous bone fracture in tensile mode and shear mode.

In summary, the occurrence of limb flail injury are subjected to two constraints: anatomical and external [27]. Anatomical

constraint means the resistance of muscle groups (including both passive stiffness and active forces), ligament, tendon, as well as the limitation on the degree of freedom of a joint capsule. External constraint means the impact and subsequent contact with the seat structure after the anatomical constraints have been exceeded.

In qualitative terms, Kazarian and Smith-Lagnese [12,31] attempted to classify extremity wind flail injuries from a biomechanical perspective and to correlate such mechanisms with probable causal factors existing during the ejection sequence. They found that the injury site, pattern, and severity depend heavily on the initial limb position, airspeed at the time of ejection, and angle of attack at which the body/seat enter the airstream. A quantitative assessment, however, is needed before we can fully understand the physical event of flail injuries, including the interaction among the forward acting inertial forces, aerodynamic drag, occupant/seat motion, limb flail movement, and the excess mechanical stresses developed at the limb structures. With the understanding of the entire physical event and injury mechanism, we can then effectively improve the protection equipment and restraint system for open seat ejection.

Measurements of flailing loading

Another important aspect to the understanding of flail injury is the direct measurement of aerodynamic loading on the crew/seat system during ejection escape sequence. Forces acting on the arms and legs of test subjects seated in the F-105 and ACES-II ejection seats have been measured over a range of speeds up to 183 ft/sec in a wind tunnel [25,26]. Because of the difficulty in determining a

significant area to specify wind flailing loading on a forearm or knee, a term "force area" defined as the product of drag coefficient C_d and effective area was used:

$$(Force\ area) \quad C_d \cdot (Effective\ area) \quad \frac{Total\ drag\ force}{\frac{1}{2}\rho V^2}$$

Note that the force area is essentially the total drag force per unit dynamic pressure which depends on the square of velocity. Similarly, the moment is defined as the product of force area and distance. With these conventions, both the outward force area and moment area at the knee were found to vary with yaw angles up to 30°. Payne and his associates also measured flailing forces at the feet, hands, and lift forces at the helmet. Extensive data were compiled in graphics form in ref [25,26].

At small pitch angle, the occupant of an ACES-II seat was found to experience about 80% of the total drag force acting on the occupant/seat assembly [9]. This percentage was found to decrease at increasing pitch. On the other hand, it was found that the hand-up forces decrease and the feet-up forces increase as the pitch angle was increased up to 60°.

To measure the aerodynamic forces and limb flailing forces acting on a crew member at transonic speeds and to learn the effect of close proximity to a cockpit, a 1/2 scale crew/seat model was integrated to the forebody of an F-16 and wind tunnel measurements were taken for the Mach number range of 0.4 - 1.2 [2,18]. Flail potential forces of the upper arms were found to be doubled by the presence of the fuselage. Similarly, there is a magnification of upper leg flail force in the lift direction, being approximately 600

lb per leg at a Mach number of 0.6. Through these measurements, it was also found that a flailing leg and arm can cause a significant increase in yaw moment, even though the change in limb position is small. This effect can lead to destabilization of the occupant/seat system and an enhanced limb flail loading. Magnitudes of flailing force were found to depend on the Reynolds number [2], primarily through the changes in local boundary layer flow and flow separation over the surfaces of the arms, legs, and helmet/head.

Using a wind tunnel facility, these afore-mentioned studies measured the steady state loading on limbs and characterized the effects of pitch, yaw angles, Mach number, Reynolds number, etc. Efforts to measure the transient loads during an ejection escape have been underway by the AL/CFB personnel using an instrumented anthropomorphic manikin (ADAM) ejected from an accelerated rocket sled at the test facility of Holloman AFB, New Mexico. A systematic integration of these data should allow us to assess the transient aerodynamic loading surrounding the occupant/seat system during the ejection sequence. Incorporation of these measurements into a computational model for the limb structure should help us to understand the transient interaction between limbs and external loading, as well as to understand the underlying injury mechanisms.

Measurements of force-motion relation for limbs and tissue properties

To fully understand the transient physical event of flail injuries, a finite element modeling effort is planned. Force and moment responses of arms and legs to mechanical loading from young, health subjects, measured by Engin [6], will be used to calibrate this model. Extensive data include both passive and active force

components were compiled in graphics form. A value of ~ 200 Newton was found for upper extremity and a value of 400 Newton for lower extremity. These data measure the load carrying capacity of human limbs under normal condition and will be used to calibrate the model to be developed.

Mechanical properties, including elastic property and failure strength, for relevant biological soft and hard tissues [4,17,33] have been reviewed. Stress-strain response for wet bone is known to be strain-rate dependent (Fig. 2). Both compact and trabecular bone are known to fail at higher stress levels when loaded at higher strain-rates (Fig. 3). Evidence of long bone fracture after flail incidence suggests different possible failure modes (tensile, compressive, or shear), depending on the instantaneous loading. Elastic modules and failure strength, strain for wet compact bones are summarized in Table 2. Soft tissues, including muscles, ligament, and tendon, are known to be nonlinear viscoelastic with a relative insensitivity to strain rates. Table 3 summarizes mechanical properties for relevant soft tissues.

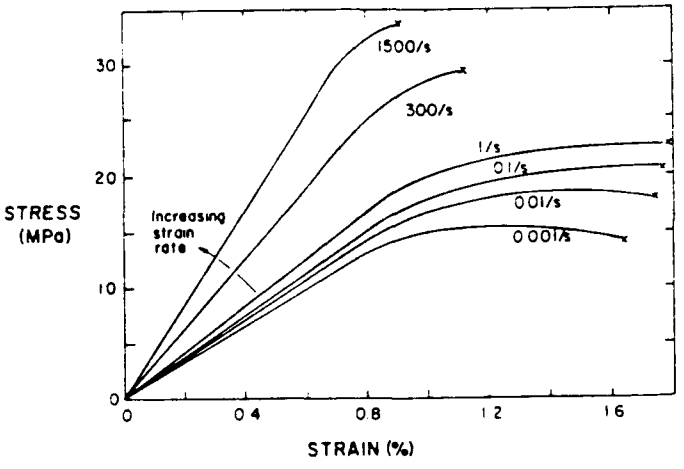


Fig. 2 Strain-rate sensitivity in stress strain responses of human compact bone, from [17].

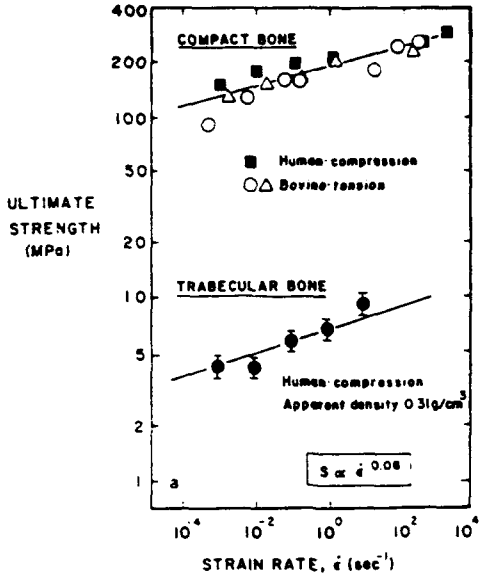


Fig. 3 Influence of strain-rate on the ultimate strength of compact and trabecular bone tested without marrow in vitro, from [4].

Suggested future work

To prevent the occurrence of limb flail injuries during ejection escape it is important to understand the physical event through the entire escape sequence and to know the interaction between external forces and body structures. This understanding is an essential step toward improvement of in the design of protective equipment and restraint systems. With this objective, a finite element modeling for the limb flail injury is planned. This model should simulate the effects of flail loading, resultant limb motion, tissue deformation, and material/structural failure when the developed stresses in tissue components exceed ultimate strengths.

TABLE 2. ELASTIC MODULI AND ULTIMATE STRENGTH OF LONG BONES

		Direction	Tension		Compression		Shear	
			Modules of Elasticity (GPa)	Ultimate strength (MPa)	Modules of Elasticity (GPa)	Ultimate strength (MPa)	Torsional rigidity (GPa)	Ultimate strength (MPa)
Arms	Humerus	Longitudinal	17.2	130	132	5	--	--
	Radius	Longitudinal	18.6	149	114	5	--	--
	Ulna	Longitudinal	18.0	148	117	5	--	--
Legs	Femur	Longitudinal	17.2	121 (1.4%)	167	5 (1.85%)	3.2	54
	Tibia	Longitudinal	18.1	140 (1.5%)	159	5	--	--
	Fibula	Longitudinal	18.6	146	123	5	--	--
Skull		Tangential	--	25	--	--	--	--
		Radial	--	--	97	--	--	--

1. Numbers in parentheses are percentage elongation or shortening.

2. Data condensed from ref [33].

TABLE 3 TENSILE STRENGTH AND CORRESPONDING % ELONGATION FOR RELEVANT SOFT TISSUES

	Tensile failure strength (MPa)	% elongation at failure (%)
Skin	7.6	78.0
Tendon	53.0	9.4
Ligament	60 - 80	10 - 15
Cartilage	3.0	30.0
Skeletal muscles	8 - 20	45 - 80
Cardiac muscles	0.11	64.0

* Data condensed from ref [33].

References

1. Belk WF, Limb flail injuries and the effect of extremity restraints in USAF ejections: 1971-78, SAFE Journal, vol 10, no 2, pp 18-21.

2. Braddock WF 1984), The effect of Reynolds number variation of flail initiating forces acting on a crewman during emergency escape, AFAMRL-TR-84-67, Wright-Patterson AFB, Ohio.

3. Buschman DL and Rittgers SE (1972), Injuries induced by high speed ejection: An analysis of USAF noncombat operational experience, SAFE Proceeding of the 10th Annual Symposium, Phoenix AZ, Oct 1972.

4. Carter DR and Hayes WC (1976), Bone compressive strength, the influence of density and strain-rate, Science, 194, 1174-1176.

5. Delgado RC (1986), Limb flail injuries in USAF ejections: 1979-85, Proceeding of the 24th Annual SAFE Symposium.

6. Engin AE (1982), Long bone and joint response to mechanical loading, Final report AFOSR-TR-82-0013.

7. Every MG and Parker JF, Jr (1973), A review of problems encountered in the recovery of Navy air crewmen under combat conditions, Report contract N00014-72-C-0101, Office of Naval Research, Washington, DC, June 1973.

8. Every MG and Parker JF, Jr (1974), Aircraft escape and survival experiences of Navy prisoners of War, Report contract N00014-72-C-0101, Office of Naval Research, Washington, DC, August, 1974.

9. Hawker FW and Euler AJ (1976), Wind tunnel measurements of vertical acting limb flail forces and torso/seat back forces in an ACES-II ejection seats, AMRL-TR-76-3., Wright-Patterson AFB, Ohio.

10. Hearon BF, Thomas HA, and Raddin JH (1982), Mechanism of vertebral fracture in the F/FB-111 ejection experience, Aviat. Space Environ. Med. 53(5):440-448.

11. Hearon BF, Brinkley JW, Luciani RJ, and von Gierke HE (1981), F/FB-111 Ejection experience (1967-80), part 1, evaluation and recommendation, AFAMRL-TR-81-113, Wright-Patterson AFB, Ohio.

12. Kazarian LE (1978), Identification and classification of vertebral fractures following emergency capsule egress from military aircraft, Aviat. Space Environ. Med. 49(1):150-157.

13. Kazarian LE, Beers K, and Hernandez J (1979), Spinal injuries in the F/FB-111 crew escape system, Aviat. Space Environ. Med. 50:948-957.

14. Kinneman RE, Every MG, and Parker JF, Jr (1973), Specific

biomedical issues in the escape phase or aircombat mishaps during Southeast Asia Operations, AGARD conference Proceeding No. 134, Escape problems and maneuvers in combat aircraft, September 1973.

15. Kittering JW and Yuill JH (1974), Southeast Asia escape, evasion and recovery experience of returned prisoners of war, thesis Air War College, Maxwell AFB, Alabama, May 1974.

16. Lewis ST (1974), Ejection injuries in Southeast Asia prisoner of war returnees, Proceedings of the 12th Annual SAFE Symposium, N Hollywood, California 1974.

17. McElhaney JH (1966), Dynamic response of bone and muscle tissue, J Appl Physiol, 21, 1231-1236.

18. Newhouse HL, Payne PR, Brown JP (1979), Wind tunnel measurement of total force and extremity flail potential forces on a crew member in close proximity to a cockpit, AMRL-TR-79-110., Wright-Patterson AFB, Ohio.

19. Obergefell LA and Kaleps I (1988), Simulation of body motion during aircraft ejection, Math Comput. Modelling, 11:436-439.

20. Payne PR (1971), Notes on the initiation of limb flailing, ARML-TR-71-45, Wright-Patterson AFB, Chio.

21. Payne PR (1973), *Some studies relating to* "limb flailing" after an emergency escape from an aircraft, AMRL-TR-73-24, Wright-Patterson AFB, Ohio.

22. Payne PR (1974), On the avoidance of limb flail injury by ejection seat stabilization, AMRL-TR-74-9, Wright-Patterson AFB, Ohio.

23. Payne PR and Hawker FW (1974), USAF experience of flail injury for noncombat ejections in the period of 1964-1970, AMRL-TR-72-111, Wright-Patterson AFB, Ohio.

24. Payne PR (1975), On pushing back the frontiers of flail injury, AGARD Aerospace Panel Specialists Meeting, Toronto, Canada, May 1975.

25. Payne PR (1975), Low speed aerodynamic forces and moments acting on the human body, AMRL-TR-75-6, Wright-Patterson AFB, Ohio.

26. Payne PR, Hawker FW, and Euler AJ (1975), Stability and limb dislodgement force measurements with the F-105 and ACES-II ejection seats, AMRL-TR-75-8., Wright-Patterson AFB, Ohio.

27. Ring WS, Brinkley JW, and Noyes FR (1975), USAF non-combat ejection experience 1968-1973 incidence, distribution, significance and mechanism of flail injury, 1975 AGARD, pp. B1 1-8.

28. Schneck DJ (1978), Aerodynamic forces exerted on an articulated human body subjected to windblast, Aviat. Space Environ. Med.

49(1):183-190.

29. Schneck DJ (1980), Studies of limb-dislodging forces acting on an ejection seat occupant, Aviat. Space Environ. Med. 51(3):256-164.

30. Shannon RH (1971), Operational aspects of forces on man during ejection/extraction escape in the USAF, Jan 1968 - Dec 1970, AGARD Proceeding No. 88, Linear accelerations of Impact Type, AL, June 1971.

31. Smith-Lagnese SD and Kazarian LE (1982), The correlation and description of windflail injury mechanism in the windblast environment, Proceedings of the 20th Annual SAFE Symposium, Las Vegas, NV 1982, pp. 293-296.

32. Till AN and Shannon RH (1971), Combat use of life support systems in Southeast Asia, Jan 1967 - June 1970, Proceedings of eight Annual SAFE Symposium, N Hollywood, CA, Sept 1971.

33. Yamada H (1970), Strength of biological materials, Williams and Wilkins, Baltimore, 1970.

AUTOMATE: A RESEARCH PARADIGM TO STUDY COLLABORATION IN MULTIDISCIPLINARY DESIGN TEAMS

Maryalice Citera

Assistant Professor

Department of Psychology

Wright State University

Final Report for:

AFOSR Summer Faculty Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D. C.

September 1993

AUTOMATE: A RESEARCH PARADIGM TO STUDY COLLABORATION IN MULTIDISCIPLINARY DESIGN TEAMS

Maryalice Citera
Assistant Professor
Department of Psychology
Wright State University

Abstract

The multidiciplinary design team approach has several advantages over conventional approaches to design. Yet is also has several potential drawbacks. These drawbacks include: miscommunication, lack of coordination, and misanalogies. A research paradigm was developed that examined these issues. The paradigm was based on the Stasser's (1992) hidden profile research. A fictitious design rationale is presented to subjects and contains the shared information. Each individual also receives a guidebook based on information from their own specialty or discipline. This guidebook contains the unique information. Information sharing can be assessed by examining the design rationales subjects provide for their design choices. Future research questions that can be addressed with this paradigm are discussed.

AUTHOR NOTES: The author would like to thank Mike McNeese, Cliff Brown, Jon Selvaraj, Randy Whitaker, and Brian Zaff for their help and support on this project.

AUTOMATE: A RESEARCH PARADIGM TO STUDY COLLABORATION IN MULTIDISCIPLINARY DESIGN TEAMS

Maryalice Citera

Introduction

Redesigns and long product life cycles can be very costly in product design. According to Kochan (1991), United States and European designers spend, on average, 50% of their time in redesign activities. In contrast, their Japanese counterparts spend only 10% of their time on redesign. Similarly, the typical product development lead time in Europe is 63 months, as compared to 43 months in Japan.

Concurrent engineering has been suggested as a way to get the product to market quicker with fewer costly redesigns (Kochan, 1991; Sobieski, 1990). In concurrent engineering, a multidisciplinary team of designers works together to integrate knowledge from the various specialties (e.g., industrial engineering, mechanical engineering, human factors, production manufacturing) early in the design process. This approach shortens the product life cycle (product development from inception to market) because tasks can be done simultaneously. Fewer redesigns are necessary because designs are scrutinized from multiple viewpoints prior to fixing or constraining the design.

In contrast, more traditional approaches generally involve design tasks being completed sequentially and independently. Kochan (1991) refers to this as "over-the-fence" engineering. Once a group of designers from a particular discipline completes their "contribution" to the design, they pass it on (throw it over the fence) to the next

department or group of designers.

The concurrent engineering approach is believed to be an improvement over traditional approaches because: the product matches customer's needs more precisely, it shortens the time to market, it reduces the need for changes late in the design process reducing developmental costs, and results in designs that are simpler and cheaper to manufacture (Kochan, 1991).

As the popularity of the concurrent engineering approach and multidisciplinary design grows, the need for better understanding of collaboration among design team members is needed. One theory that can be applied to the understanding of multidisciplinary design teams is situated problem solving. This perspective posits that problem solving takes place within a particular context (i.e., the situated context) and that this context shapes how the problem is approached and solved. A situated context is one that involves social interdependence between actors who share responsibility for solving the problem. Through social interaction, the individuals construct a joint understanding of the problem, develop shared values and norms, and become involved in a reciprocal exchange of knowledge (Lave, 1991). In this sense, the situated context is not static, but always evolving and changing as the problem and its solution unfold.

According to Young and McNeese (in press), there are 10 factors that characterize situated problem solving. First, solving the problem must require the use of multiple cognitive processes and multiple paths to the solution. Second, the context in which the problem occurs will be complex and provides a wide array of

perceptual cues that inform those involved about the possibilities that the situation affords. Third, solving the problem involves identifying attributes of the problem and separating the irrelevant from the relevant information. Fourth, several competing solutions can be generated for the problem. Fifth, given the uncertainty involved in solving the problem and its ill-structured nature, the problem is best approached by generating sub-problems. Sixth, the context is interpersonal and involves social interaction. This social interaction defines the roles of the actors and the meaning of their actions. Seventh, as the group works out the problem, they build a shared perception or understanding of the problem and its potential solutions. Eighth, situated problem solving involves the integration of distributed knowledge. Distributed knowledge means that each individual brings to the situation unique information based on past experiences and learning opportunities. Across the group, this information may reflect a variety of domains and specialties. The group must then share and integrate each member's unique information. Ninth, as the problem unfolds, the group establishes a pattern of interaction. This pattern or developmental history becomes part of the context and heavily influences later interactions and decisions. Finally, in situated problem solving, the context involves values, intentions, and goals that have personal and social significance. Thus, the situated context has important implications for the individual's identity.

Multidisciplinary design is similar to situated problem solving because it can be characterized as "a goal oriented, constrained, decision-making exploration, and learning activity that operates within a context that depends on the designer's

perception of the context" (Gero, 1990, p. 28). The design develops through the social interaction of the team members. Because they come from a variety of disciplines, knowledge is distributed across team members. Members of the team work on the design task relying on their own unique information and other team members to educate them.

Unfortunately, multidisciplinary design teams do not effectively share distributed knowledge. According to Stasser (1992; Stasser & Titus, 1985; 1987), groups often focus on commonly shared information and neglect to discuss unique or unshared information.

One reason why team members may have difficulty sharing information is because members of the team do not share a common framework. Often individuals from different disciplines will speak at cross-purposes to one another. They misunderstand and misinterpret what each other says. In other words, they speak different languages (Boff, 1987). Each discipline has its own jargon, metaphors, acronyms, definitions for words, labels etc. (Bucciarelli, 1988). In some instances, the same label may be used to explain exact opposite things (Fotta & Daley, in press). Without a common framework, the advantages of the multidisciplinary team approach cannot be realized.

Second, designers often draw heavily on their previous design experience. So instead of generating and carefully evaluating all possible design alternatives, designers rely on a case-based design strategy (Gero, 1990; Klein, 1987). In case-based design, designers draw analogies between their present design and previous

ones. Case-based designs are efficient because features of a previous design can be incorporated into the new design project. Unfortunately, unnecessary components of the original design may also get incorporated. Because these features are spuriously associated with essential ones, they needlessly restrict the design options considered. Designers may have trouble differentiating and articulating which features are essential and which are not. The result may be misanalogies. Misanalogies occur when previous learning is applied to a situation where it may be inappropriate or conflict with other aspects of the design. Because misanalogies are difficult to articulate, they add to the communication problems multidisciplinary design teams face.

In addition, cross-disciplinary team members may work on different schedules and be separated by physical distances. These disconnects may aggravate communication and coordination problems. For example, anyone who has played telephone tag, understands the frustration that can result from being unable to communicate with another team member. Add to this time constraints and you have a potentially volatile situation. The result may be that team members fail to communicate with one another and decisions are made based on incomplete information.

Although the multidisciplinary team may share a common goal, team members may identify more strongly with their own discipline than with the team as a whole. Bucciarelli (1988, p. 168) observed that "decisions made across disciplines are best seen as negotiations among parties who, while sharing a common goal at some level,

hold different interests." If communication is difficult, both a shared perception and a shared identity may be that much more difficult to build. Because of the importance of collaboration in building a shared identity among team members, it is important to examine the nature of collaboration and tools to aid collaboration (e.g., computer-supported cooperative work tools, groupware). To pursue this goal, I set out to design a paradigm that can be adapted to examine design team collaboration.

Paradigm Development

The focus of the remainder of this report is on the development of a paradigm to study collaborative design in an experimental research setting. Because one often cited drawback of experimental research is the lack of contextual realism, particular attention was paid to this issue.

Preliminary steps toward the development of a design paradigm were completed by Citera and Selvaraj (1992). A research design problem was chosen based on an extensive review of the literature and from responses to postings on several computer bulletin boards. The problem, designing a component of an on-board navigation system for an automobile, was selected because it: 1) was based on a "real world" design problem, 2) was multidisciplinary in nature, 3) could be completed in an individual or team context, and 4) would be familiar to both university students and design professionals.

Citera and Selvaraj (1992) elicited knowledge about potential tradeoffs involved in designing a navigation system for an automobile from design experts. This information was collected to establish a realistic base of information on which to build

the task. To elicit knowledge from design experts we used a technique called Concept Mapping (McNeese, Zaff, Peio, Snyder, Duncan, & McFarren, 1990; Novak & Gowin, 1984). Concept mapping is an interactive interview methodology. During the exchange, an interviewer converted the elicited information into a graphical network of concepts nodes (i.e., a concept map). Thirteen design professionals were interviewed. The design professionals included human factors psychologists and engineers, a software specialist, a display hardware specialist, an electrical engineer, and an industrial engineer. We felt it was necessary to understand the concerns and requirements of different perspectives before building the task. As part of this data collection, three key disciplines likely to be represented in a multidisciplinary design team were identified: human factors, computer programming/engineering, and marketing/business. The task was built around these three disciplines. The elicited information focused on potential tradeoffs involved in designing an automobile navigation aid within a multidisciplinary team context and was useful in identifying issues and conflicts that designers of such an aid would face.

After examining the information provided by the experts, the design problem was narrowed down to the issue of designing an auditory display to accompany the navigation aid. The problem has been named AutoMate and a description of it is included in Appendix A.

Task Rationale

The task is based on Stasser's (1992; Stasser & Titus, 1985; 1987) work on information sharing. In Stasser's paradigm, groups of subjects are given information

about a problem (i.e., information about job candidates). **Shared** information is given to all subjects. **Unique** information is given to individual subjects. Built into the information is a hidden profile. The shared information favors one decision, whereas, the unique information favors another. In the AutoMate task, unique information derives from information specific to each discipline. Thus, subjects get unique information relevant to their area of expertise. The shared information is given to all subjects in the form of design rationale for a past design case.

Design Rationale--Shared Information

As previously stated, design is often approached from a case-based perspective. Unfortunately, subjects in the experimental situation come from a variety of backgrounds and there is no way to assure that they have equivalent prior experiences in design. One way to establish a common reference point for all group members is through the use of design rationale. Design rationale is a means of preserving the reasoning behind design decisions (Candy, 1993). It is not only a record of what final design choices were made, but also of what alternative options were considered and the criteria used to choose between the options. Design rationale can serve as an archival, community memory. By recording design rationale, collaboration can take place across time (or history). By this I mean that design rationale for one project can be used by an entirely new set of designers as a case-based reference.

We chose to build the design rationale on a related but different problem. The design problem was developing an auditory warning display for a cockpit. The past

design case, contains some information that is directly applicable to the AutoMate (i.e., the transfer or new) design problem and some that is not. (For an example see Appendix B). The applicable information can be considered direct analogies for the new design problem. The non-applicable information can be considered misanalogies. Since all members of the design team will receive the case information, this information is shared. The team must attempt to separate the relevant parts of the case from the irrelevant parts based on the unique information.

Specialized Expertise--Unique Information

The unique information subjects receive will be based on their expertise in their own field of specialization (i.e., human factors, computer programming/engineering, marketing/business). Each team member will be given a guidebook relevant to his/her own perspective. The guidebook will contain information based on that perspective that identifies a misanalogy in the design case. (See Appendix B for excerpts from this.)

Dependent Measures

Ideally, one would want to measure the amount of information shared in the group discussion. Stasser, in his original work, used a measure of recall to assess shared information. Subjects were given a free-recall test prior to discussion and following the group discussion. The amount of information gained from pre- to post-test recall was considered the amount of information shared. Similarly, subjects in the AutoMate task make initial decisions based on their information packets (prior to group discussion) and record their design rationales. Following team discussion,

individuals again indicate their design choices and record their design rationale. Information sharing would be assessed by the change in information used in the design rationale. In other words, the gain in information is indicated by changes in the decision rationale. Shared or joint understanding of the problem would be indicated when the design rationales converge (become more similar) across team members. Team decisions could also be examined to determine whether unique information was shared and misanalogical parts of the problem identified and correctly addressed. Also, the confidence subjects have in their decisions might indicate the amount of trust individuals have in their team's ability to adequately share its expertise.

Applications

This approach to studying collaboration can be used to examine a variety of issues. For example, in distributed teams (i.e., teams that are physically separated by space whose members predominantly communicate through electronic means) does more or less unique information get shared? On the one hand, team members may be less likely to compete for limited air space and feel less evaluation apprehension and thus, share more information. On the other hand, these team members have few feedback cues available (because of lack of nonverbal cues such as head nods) that indicate what has been understood, and thus, may share less information.

Also, of interest would be to examine how the distributed context affects the development of trust and confidence in other team members. Furthermore, computer-supported collaborative aids might be tested to see if they improve the sharing of information and the joint construction of meaning.

References

Boff, K. R. (1987). The tower of Babel revisited: On cross-disciplinary chokepoints in system design. In W. B. Rouse and K. R. Boff (Eds.) System design: Behavioral perspectives on designers, tools, and organizations (pp. 83-96). New York: Elsevier Science Publishing.

Bucciarelli, L. L. (1988). An ethnographic perspective on engineering design. Design Studies, 9, 159-168.

Candy, L. (1993, August). Hypothetical design in the perfect world: Observations on design rationale in knowledge support system development. Poster presented at HCI International '93. Orlando, FL.

Citera, M. & Selvaraj, J. A. (1992). Development of a research paradigm to study collaboration in multidisciplinary design teams. Washington, DC: Air Force Office of Scientific Research.

Gero, J. S. (1990). Design prototypes: A knowledge representation schema for design. AI Magazine, 11(4), 26-36.

Fotta, M. E., & Daley, R. A. (in press). Improving interpersonal communications on multifunctional teams. In H. R. Parsei & W. G. Sullivan (Eds.), Handbook of concurrent engineering. London: Chapman and Hall.

Klein, G. A. (1987). Analytical versus recognitional approaches to design decision making. In W. B. Rouse and K. R. Boff (Eds.) System design: Behavioral perspectives on designers, tools, and organizations (pp. 175-186). New York: Elsevier Science Publishing.

Kochan, A. (1991). Simultaneous engineering puts the team to work. Multinational Business, 1, 41-48.

McNeese, M. D., Zaff, B. S., Peio, K. J., Snyder, D. E., Duncan, J. C., & McFarren, M. R. (1990). An advanced knowledge and design acquisition methodology: Application for the pilot's associate (AAMRL-TR-90-060). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division, Air Force Systems Command.

Novak, J. D. & Gowin, D. B. (1984). Learning How to Learn. New York: Cambridge University Press.

Sobieski, J. (1990, December). Multidisciplinary design optimization. Aerospace America, p.65.

Stasser, G. (1992). Pooling of unshared information during group discussions. In S. Worchel, W. Wood, and J. Simpson (Eds.) Group Processes and Productivity (pp. 48-67). Newbury Park, CA: Sage.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. Journal of Personality and Social Psychology, 48, 1467-1476.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. Journal of Personality and Social Psychology, 53, 81-93.

Appendix A

Design Specifications for the AutoMate Audio Display

This report summarizes the technical details of the audio display design for the AutoMate system.

Background:

Several major automobile manufacturers have started to develop in-vehicle navigation aids. These aids assist drivers in planning and driving to a destination. Many of these systems include moving map displays that visually display a map of the area and show the automobile's location and heading. The automobile's location is determined from distance and heading, satellite technology, or a combination of the two. Location is generally checked against the map database to assure that the information is accurate. Another feature of some systems is route planning. The driver simply inputs into the system his/her destination and the computer calculates the quickest route. The driver is sometimes given the option to select the most scenic route, the route without highways or tolls. Some systems offer a guidance display that gives turn-by-turn directions to the driver about how to get to a particular destination. Some systems feature traffic updates that are broadcast from a central location. These traffic updates can warn the driver of potential traffic delays along their route and may even be able to reroute the driver around heavily congested areas.

AutoMate:

AutoMate is a computerized navigation aid designed for a mid-size automobile (in the price range of $22,000 to $25,000). The market for this automobile is a combination of private owners, business sales fleets, and car rental companies. Typical usage varies according to the type of owner. With car rental companies, out-of-town visitors use the system to locate hotels, restaurants, and business locations. These users will be unfamiliar with the area and will rely on the system to give them directions. Since they are already familiar with the area, private owners rely less on the guidance

feature and place a greater value on the traffic update information. Business sales fleets will use the system to locate customer/client addresses and to avoid traffic congestion.

In the original AutoMate version (V1), the following features were included: a visual moving map display, route planning, directions and guidance visual displays, traffic information updates, and a menu of local hotel and restaurant addresses. The AutoMate system consisted of a computer equipped with a 386 Central Processor Unit (CPU), a 200 mg hard drive, a GPS (Global Positioning Satellite) board, a map data base of the purchaser's local area (e.g., the Greater Dayton area), and a CRT (Cathode Ray Tube) color monitor with a touch screen for input. The map data base listed the longitude and latitude coordinates and names for every highway, street, and alley. Also, stored in the database were the exact locations of each intersection and highway exit. In addition, a GPS antennae, a compass, and wheel sensors were included to determine the location of the automobile.

Problem Description:

AutoMate (V1) is being redesigned to resolve some of the weaknesses identified with the original version. One deficiency is AutoMate's (V1) lack of an auditory display. An auditory display, in this case, would involve presenting the route information and directions verbally to the driver. The display might also be useful for giving the driver information about current location and heading. Although an auditory display was considered for the original AutoMate, it was not implemented.

A multidisciplinary design team is being formed to develop an audio display for the updated version. You have been selected to be part of this team. This team includes a human factors specialist, a computer engineer, and a marketing specialist.

Appendix B

Example AutoMate Decision: Choice of a Speech Processing Board

Cased-based information:

Design Rationale for the Auditory Warning Display in the C152 Cockpit

1.1 Requirement: Speech Processing Board

The speech processing board should be able to produce speech that is intelligible to the pilot, operate under normal conditions found in the cockpit, and draw no more than 50mA of current.

1.2 Options:

Three different speech processing boards were considered: a 12 kHz board manufactured by Sound Products (SPR112), a 12 kHz board manufactured by TechSound (TS312), and a 10 kHz board manufactured by Vocoder (VOC510).

1.3 Criteria

The SPR112 was chosen based on the following criteria:

1.31 The sampling rate (12kHz) was the best available on the market at the time. Sampling rate represents the number of samples taken per second. High quality sound depends on the rate at which sound is sampled. The higher the sampling rate the better the sound quality. For example, CD (Compact Disk) quality sound is sampled at a rate of 44.1 kHz. The VOC510 was only a 10kHz board.

1.32 The board had to withstand extreme operation temperatures. In the summer, a

cockpit sitting on the tarmac in the sun all afternoon can reach temperatures of up to 150° F. In the winter, cockpit temperatures can drop to as low as -30° F. The TS312 could not withstand this range in temperatures. It had an operational temperature range of only -10° F to 130° F.

1.33 The specifications indicated that the board could draw no more than 50 mA of current. The TS312 was disqualified because it drew 100 mA of power.

Information from the Human Factors Guidebook

Under optimum conditions (i.e., low noise, low distractions, etc.) a message with a frequency range of 200 to 6100 Hz will be about 95 percent intelligible. Reducing this rate will produce a decrease in intelligibility, particularly of consonant sounds such as "th" which are composed largely of higher frequencies.

A good rule of thumb to follow is that sound should be sampled at about 3 times the highest sound frequency that you are interested in. The intelligibility of speech increases up to about the 22 kHz sampling frequency and then levels off. This means that higher sampling frequencies do not have a significant impact on intelligibility.

Three methods are available for measuring speech intelligibility:

a. The ANSI standard method of measurement of phonetically balanced (PB) monosyllabic word intelligibility. This method uses "phonetically balanced" lists of words, i.e., the words in each list are chosen so that all speech sounds are represented according to their frequency of occurrence in normal speech. Scoring is on the basis of percent correct.

b. The modified rhyme test (MRT) should be used if the test requirements are not as stringent or if time and training do not permit the use of the ANSI method. For the MRT the listener must correctly perceive only one phoneme in each test word and

the listener has a printed list of words so that he can check the word that he thought was correct. Although the lists of words in the MRT are not phonetically balanced to represent everyday speech, the MRT is efficient and useful because it requires perception of consonantal sounds important to intelligibility. Scoring is corrected for chance.

c. The articulation index (AI) should be used for estimations, comparisons, and predictions of system intelligibility. The AI is highly correlated with the intelligibility of speech perception tests given to a group of talkers and listeners. The AI is computed from acoustical measurements or estimates of the speech spectrum and of the effective masking spectrum of any noise which may be present along with the speech at the ears of the listener.

The Table below shows the intelligibility criteria for voice communication systems.

COMMUNICATION REQUIREMENT	SCORE		
	PB	MRT	AI
Exceptionally high intelligibility, separate syllables understood.	90%	97%	0.7
Normally acceptable intelligibility; about 98% of sentences correctly heard; single digits understood	75%	91%	0.5
Minimally acceptable intelligibility; limited standardized phrases understood; about 90% sentences correctly heard (not acceptable for operational equipment)	43%	75%	0.3

Estimated Intelligibility Results for Different Sampling Rates using the AI

Sampling Frequencies kHz	Articulation Index Percent
5	.15
10	.25
12	.3
15	.5
22	.75
32	.76
44	.76

Shared Information: Details on Speech Processors

Speech Processors

	SPR 100	VOC 525	RL 658	DM 637	VOC 522	SPR 820
Maximum Current	50mA	4mA	10mA	4mA	55mA	12mA
Sampling Rate	4 to 12 kHz Program mable	4 to 22 kHz Program mable	4 to 32 kHz Program mable	4 to 44.1 kHz Program mable	4 to 22 kHz Program mable	4 to 32 kHz Program mable
Operation Temp °F	-55 to 185	-55 to 170	-40 to 140	-40 to 185	0 to 130	-40 to 185
Speech Recognition Possible	Yes	Yes	No	Yes	Yes	Yes
Price per Chip (Quantity)	$23.75	$42.50	$47.00	$52.50	$37.50	$50.60

SOURCES AND PATTERNS OF REACTION TIME FLUCTUATION USING A CONTINUOUS SUSTAINED REACTION TIME TASK

Geoffrey L. Collier
Assistant Professor
Department of Psychology

South Carolina State University
300 College Street NE
Orangeburg, SC
29117

Final Report for:
Summer Research Extension Program.
Armstrong Laboratory

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

December 1993

SOURCES AND PATTERNS OF REACTION TIME FLUCTUATION USING A CONTINUOUS SUSTAINED REACTION TIME TASK

Geoffrey L. Collier
Assistant Professor
Department of Psychology

South Carolina State University
300 College Street NE
Orangeburg, SC
29117

Abstract

Fluctuations in attention through time was operationalized in terms of reaction times to continuously presented stimuli. Auditory and visual stimuli were presented for six or 16 minutes, at an isochronous rate varying between 800 and 2000 milliseconds. Six of the eight subjects were capable of adequately performing the task, which required high rates of continuous vigilance. Diverse trend patterns were observed, but overlaid on the trend were local fluctuations, as indexed by autocorrelational patterns. These appeared to be strategic rather than strictly endogenous. Evidence of periodicity in the 3 to 5 minute range was found, but this evidence will require replication. A PRP (psychological refractory period) effect was found, such that subjects could not keep up with fast stimulus presentation rates even when response times were consistently less than this presentation rate on blocks with slower presentation rates.

Sources and patterns of reaction time fluctuation using a continuous sustained reaction time task

Geoffrey L. Collier

General Introduction

The study of fluctuations in attention across several minutes of vigorously sustained attention is both theoretically and practically important. Practically, understanding such sustained attention performance is necessary for understanding human performance in highly stressful environments were obligatory stimuli bombard the human operator. The cockpit of a fighter jet is certainly one such environment. Theoretically, it is important to understand what consistent factors control fluctuations in performance across stretches of time. Initially, we can ask a very basic question: does performance on a given trial depend on performance on the prior trial? If so, what is the nature of that dependency? One possibility is that response speed will vary in a periodic or cyclical fashion, for example, if there are regular endogenous oscillatory mechanisms that control the fluctuation of attention. Alternatively, there might be irregular stochastic fluctuations in performance due to variability in the degree of attention (or probability of attending, in a discrete attentional model) over an extended duration.

The current experiments looked at these issues by using a continuous sustained attention task. On each trial subjects were presented with one of two stimuli, and had to respond rapidly, indicating which of the stimuli had occured (i.e., two choice reaction time task). Each block of trials consisted of a continuous train of these trials, presented isochronously (equally spaced in time). The presentation of a stimulus occured at its expected time regardless of the time of response (or none-response) to the preceding stimulus. Thus there was considerable time pressure, as subjects had no control of the onset of the stimuli. Two tasks were used, a visual task (letter identification) and an auditory task (pitch discrimination).

The requirement of the sustenance of vigilance highlights the similarity between this task and those normally used in vigilance paradigms. The critical difference, however, is that traditional vigilance tasks require a responses (usually simple reaction times) to a series of temporally unpredictible events, whereas the current task requires continuous choice responses to a series of perfectly temporally predictible events. Thus reaction times to the regular event train provide a time series. To the extent that these reaction times operationalize momentary attentional capacity, this series then has the potential of providing a continuous portrait (a

"movie") of endogenous attentional fluctuations. Furthermore, by manipulating the rate of the isochronous stimulus train (the "frame rate") it was hoped to observe the interplay between exogenous and endogenous factors in the maintainance of sustained attention.

Experiments 1 & 2

Two experiments were performed, using four subjects in each. Since the two experiments differed principally only in the length of the trial blocks, they will be for the most part analyzed and discussed together.

Design

Stimuli & equipment

Each trial consisted of a six minute isochronous pulse train of stimuli, preceded by four isochronous stimuli whose function was to prepare the subjects and entrain the rhythm.

In the visual condition the four entrainment pulses consisted of asterisks flashed in the center of a VGA monitor, in the standard typeface and size. The duration of the asterisks was always 100 milliseconds. The event rate was manipulated by varying the time between the offset of each stimulus and the onset time of the following stimulus (the off time). The target stream consisted of a series of X's or O's. Half of the total number of events were X's, the other half O's. Within this restriction, the order of the stimulus events was randomized. This randomization was performed uniquely for each trial, generated by algorithms which used the compiler's native random number generator (Borland C), seeded by the system clock.

The auditory entrainment stream consisted of four 30 ms. 1300 hz square wave beeps, generated by the internal speaker of the Gateway 486 computer on which the experiment was run. As above, the rate was manipulated by varying the offset times. The target stimuli were high (1700 Herz) and low (800 Hz.) beeps, differing from the entrainment beeps in frequency only. Randomization of stimulus presentation was the same as in the visual condition.

Subjects responded to the stimuli on a keypad. A tab was affixed to the keypad for each trial, indicating which key was the X and which was the O (in the visual condition), or which key was the high beep and which was the low beep (in the auditory condition).

In experiment 1 each block of trials was six minutes long, and four baserates (ISI's) were used: 800, 1200, 1600, and 2000 milliseconds. Since trials were kept at a constant duration of six minutes, slower baserates resulted in fewer stimulus events per trial. The totals were 450 trials (800 ms.), 300 trials (1200 ms.), 225 trials (1600 ms.), and 180 trials (2000 ms.)

In experiment 2 each block of trials was 16 minutes long, and two baserates were used: 1200 (800 trials per block) and 1800 ms (534 trials per block).

In experiment 1 the stimulus-response mapping was varied by reversing it at each new block. In experiment 2, the mapping was constant.

Subjects

Eight subjects, four for each experiment, from the Wright-Patterson AFB subject pool were used. Musical experience was examined with a musical background questionnaire. Each subject participated in four one hour sessions (except when equipment failure required an additional session), for which they were paid $5 per hour.

Design

A within-subjects design was used. In experiment 1 there were four baserates and two stimulus modalities, while in experiment 2 there were two baserates and and two modalities. Baserate was a within session factor, while modality was a between session factor. Subjects participated in four daily sessions. In experiment 1 each session included four six minute blocks, one for each baserate, resulting in a total of 16 blocks across all sessions. Each trial was preceded by a practice block of about 32 trials. Subjects could repeat the practice blocks, if desired. This was usually only necessary during the first session. In experiment 2 each session included two 16 minute blocks, resulting in a total of 8 blocks across all sessions.

In experiment 1 stimulus-response mapping was alternated between trials, to prevent the formation of automaticity between trials (and thus the confounding of automaticity with baserate within each session). At the outset of each trial, the experimenter would stick a label on the response keys. This label marked which key was "X" and which was "O", or "high" and "low" for the auditory conditions. Thus, during the practice trials, subjects would have to learn the new stimulus-response mapping, as it was reversed from the preceding trial.

In experiment 1 the order of the baserates was altered between subjects according to a latin squares design. Subjects participated in four sessions, two for each of the two modalities. The two sessions in the same modality were replications, except that the order of the baserates was reversed. The order of the sessions for two of the subjects was AVVA (A=auditory, V=visual), while for the other two it was VAAV.

In experiment two subjects performed four blocks (in two sessions) in one modality and then four blocks in the other. Two subjects performed each of the two orders. The two

baserates were presented in an arch form, that is, either 1800 1200 1200 1800, or 1200 1800 1800 1200. Two subjects had the first order mapped onto the auditory modality and the other two had the reverse mapping.

Results

Methods of data analysis.

Each block of data ranged from 180 to 450 trials, and thus presented a complex picture. Thus each block was analyzed separately. Statistics summarizing various facets of the data are presented in tables one and two. Since the data analyses were a bit complex, they described in detail here.

The data were analyzed within a particular conceptual framework. According to this framework, fluctuational patterns in reaction times could result from several conceptually distinct sources. These sources can be broadly classified into two categories: short term and (relatively) long term. The short term component is hypothesized to be comprised of local fluctuations in attention due to failures of concentration, daydreaming, etc.

In contrast, the primary components of long term changes are performance decrements due to fatigue, and performance improvements due to the development of stimulus-response automaticity. Thus, it is hypothesized that over the course of a single trial these two forces will tend to have monotonic but opposite effects. The sum of the two effects will depend on their relative efficacies, which in turn might depend on subject, condition and random factors. If only one effect is present, then a monotonic increment or decrement in RT is expected. However, if both are present, then the pattern will be more complex.

According to this two-component model the short term fluctuations are predicted to be superimposed on top of the longer term changes. In order to extract these short term fluctuations, it was necessary to look at the residuals from the long term fluctuations. However, the framework describes these latter in merely qualitative terms, yielding no parametric predictions about performance. Thus, long term fluctuations were fit nonparametrically, using Cleveland's LOWESS smoother (tension=.6), which fits the general form of the changes in reaction times, as can be seen for example in figure 1.

The residuals from the smoothed data were extracted, and subjected to autocorrelational analysis. If there were short term fluctuations in attention, resulting in changes in reaction times, the reaction times should be autocorrelated. On the other hand, if the time series are a purely noise series, no such autocorrelations should be observed. Furthermore, the extraction of trend from the smoothing operation will tend to remove spurious autocorrelation due to

trend. Thus, the autocorrelational analysis was relatively conservative: the removal of trend would tend to decrease them, so that evidence of autocorrelation in the residuals would be stronger evidence for short term fluctuations in performance ability.

In addition to the foregoing, several other summary statistics of the performances were derived, such as mean and standard deviation reaction times, percent correct, and several others. These are presented in tables 1 (experiment 1) and 2 (experiment 2), described in detail in the table notes. Since the tabled data are fairly voluminous, the key trends are summarized and discussed below.

Results - summary.

Accuracy.

Two of the eight subjects performed inadequately, so that their data will in the main not be considered. Performance levels for the remaining six were quite high. Table 3 displays grand mean percent corrects and reaction times for each subject. Mean accuracy ranges from 90% to over 99% correct. However, some b (but see the note to tables 1 and 2). Indeed, subject DMS maintained a 99.1% accuracy rate on the 7 out of 8 blocks in which performance was acceptable. Mean reaction times range from 344 to 433 ms. In general, then, these six subjects were able to maintain surprising performance levels, in spite of the difficulty of maintaining continuous attention to a repetitious task for 6 or 16 minutes.

Across subjects, there was a strong monotonic relationship between accuracy and reaction time (Spearman's R=.83), implying a strong speed-accuracy tradeoff across subjects. Thus, even though initial differences among subjects in the mean were relatively small, equating for the speed-accuracy tradeoff would probably further reduce these differences.

A lagged analysis of errors was performed, including all blocks of trials. Mean reaction times on error trials and the four trials preceding each error were calculated (excluding all cases in which one of the four prior trials was an error trial). The results are presented in figure 2 (by subject), figure 3a (by modality), and figure 3b (by baserate). The same pattern was obtained regardless of mode of analysis. The error trials themselves (denoted as lag(0) in the figures) are quite fast, as is typically observed. Interestingly, though, the reactions on correct trials leading up to the error trials were also fast. Here, too, a speed-accuracy tradeoff is probably at work. Subjects displayed local fluctuations in overall speed (see below). When they became particularly fast, the errors resulted. It is frequently observed in the literature that fast error trials are followed by a compensatory slowing down of the reaction times. Here, we

have the converse: evidence of speeding up prior to the error trials. This effect disappears at the 2000 MS baserate, as overall reaction times are slower.

Mean reaction times

Overall mean reaction times by baserate and stimulus modality are plotted in figure 4. In general, the visual task was faster than the auditory task (although the difference was small), and reaction times slowed down with presentation rate.[1] Mean reaction time across all subjects and blocks for the auditory condition was 409 ms., and for the visual condition it was 381 ms. The difference between the modalities is not surprising, given that the auditory task involved an unfamiliar (to the nonmusical the subjects) unidimensional discrimination, while the letter identification was a familiar multidimensional categorization task. The effect of baserate shows that time pressure can increase or decrease reaction times. However, subjects found the fastest rate (800 ms.) to be difficult to "keep up with". Two of the three subjects who participated in this condition (experiment 1) frequently failed to meet the deadline (i.e., the onset of the next stimulus) in this condition. As a result all but one of these blocks for these two subjects could not be analyzed. The subject who could perform adequately at this pace was the one with large amounts of musical training. Interestingly, though, she performed adequately in the visual as well as the auditory condition.

It might be inferred from the foregoing that in response to increasing time pressure, subjects decrease response times until they can no longer do so, at which point they stop being able to meet the deadline. That is, once the reaction time distribution is as compressed as is possible, further increases in the presentation rate serve to cut off the longer tail of the reaction time distribution. However, a closer analysis reveals a more complex and interesting picture. It frequently was the case that the majority of the reaction times at a presentation rate slower than 800 ms. were less than 800 ms. A good example is in the contrast between subject JPB's auditory performance at rates of 800 and 1200 ms. Excluding one response that missed the deadline, the entire distribution at the 1200 ms. rate was less than 726 ms (mean=372, s=101, accuracy=96.3 percent correct). Yet, he could not keep up with the 800 ms. presentation rate. In other words, had JPB been able to instantiate his 1200 ms performance for the 800 ms presentation rate, his performance would have been quite good.

[1] Analysis of variance was not performed, because missing cells lead to an unbalanced design, and more seriously, subjects participated in some conditions more than once. Treating these as replications would have violated the independence assumption, while the number of replications was too small to permit the treatment of subjects as a factor. Therefore, summary results are presented heuristically.

This phenomenon implies that subjects are not fully prepared to process a stimulus the moment they respond to the preceding one. This is a kind of true PRP (psychological refractory period) effect. Welford (1968) noted a similar effect in early PRP studies. Reaction times to the first of two stimuli continue to improve as the ISI between stimuli is increased, even beyond the point at which the second stimulus is occuring after the response to the first. Welford speculated that subjects need time to process the self-generated feedback from the first stimulus and the ensuing response before commencing the processing of the second one. This paradigm provides important additional data on this point.

Trend

An initial question to be addressed is of the general trends in teh reaction times across the trials in each block. That is, did performance generally remain constant, get worse (fatigue), improve (automaticity), or some combination? In fact, patterns in trend were somewhat idiosyncratic to each block. Distributions of Spearman's rho where roughly symmetrical about 0 for both experiments (figure 5).

Going beyond monotonicity, the lowess-smoothed curves that were fitted to each RT time series were inspected by eye for general patterns, but no single preferred pattern emerged. For example, 24 of the blocks showed an initial positive trend (increase in RT), followed by continued positivity, flat RT, or negativity; an almost equal number (24) showed an initial negative trend. We can only conclude that general trend patterns are fairly idiosyncratic to the block.

Periodicity

The performance of the fo egoing trend analysis was not primarily motivated by an interest in the trend patterns themselves, but rather to remove the trend so that the residuals could be analyzed. Each block of trials was subjected to a fast Fourier transform, using the residuals from the Lowess-smoothed fit. The data for the three subjects in the second experiment, who had the longest and therefore most reliable (in this context) blocks, are summarized in table 4. It was predicted that an endogenous periodicity - i.e. one not controlled by the stimuli - would manifest itself in an invariant frequency across baserates. With respect to the FFT, this predicts that when comparing the faster baserate (1200 ms) to the faster one (1800 ms), there would be a shift in energy to the higher frequencies/shorter wavelengths. This is because the time unit upon which the FFT's were performed was the event marker

(1,2,3,4 ....), not the actual time. Therefore a given periodicity represented in these event units at the slow rate would be equivalent, in real time, to a longer periodicity at the fast rate.

Table 4 represents the magnitudes of the first 15 components of the FFT, separately for each subject and rate but averaged across modality and replication, while table 5 summarizes the tendencies of the subjects. Subjects SLP and DMS show weak tendencies in the predicted direction. For subject DMS maxima are spread across 255 ,170 and 128 for the fast rate, but shift more strongly to 170 and 128 for the slower rate. However, there is another maximum for this subject at 64. The energy is spread more evenly for subject SLP, but there is a slight tendency in the predicted direction: from a maximum at 170 for the fast rate to 128 at the slow rate. On the other hand, subject HLW has maxima at 256 for both rates although there is more energy at 171 and 128 at the slow rate than there is at the fast rate, which is in the predicted direction. Here too, though, there are other maxima at faster rates: 73 and 51.

Thus we can conclude that there is a hint in the data of a consistency of cycle lengths across rates, in the range of roughly three to five minutes. However, this hint is weak and replication would be required to make this convincing, perhaps at a variety of baserates.

Fluctuations: Autocorrelation of the residuals from the lowess fit

To look at faster, more local and not necessarily periodic fluctuations in reaction time, autocorrelations were calculated for the residuals from the smooth fit. The first order autocorrelations are summarized in table 3, column 8. Across all subjects there is a tendency for positive first order autocorrelations to predominate. Twenty six of the autocorrelations were not significant, five were significantly negative, while twenty eight were significantly positive. Furthermore, the fact that trend patterns were removed before these were calculated make this a fairly conservative test of their presence.

There were some individual differences in these results, though. For example, subject SLP displayed no significant first order autocorrelations. Interestingly, there is a tendency for the presence of positive autocorrelations to be associated with accuracy: SLP was on average the least accurate subject, while the most accurate subjects showed the most significant positive autocorrelations.

Frequently, autocorrelations were significant at lags greater than one. Subject DMS showed this effect particularly strongly. In his eight blocks, positive autocorrelations were significant at anywhere from two to eleven consecutive lags. The partial autocorrelation showed significance at second or third order lags, indicating the possible inadequacy of a first

order autoregressive model to fit his behavior. However, the length of the autoregressive effect did not appear to depend on modality or baserate.

We can conclude that subjects do show consistent fluctuations in reaction time that persist across a number of consecutive trials. Several factors hint that this might be strategic:

1). If we assume that the best subjects will have the most strategic control of their performance, then the fact that they showed the strongest autocorrelations might imply that these result from strategy.

2). Secondly, if the fluctuations were due to a strictly endogenous drift process (i.e., not strategic and not stimulus driven) then we might expect that the the size of the first order autocorrelation would decrease with increasing baserate. This hypothesis is based on the assumption that the drift rate would be invariant across rates of stimulus presentation, and thus would have strengths that decrease as a function of increasing temporal distance. However, baserate was not monotonically related to the size of the first order lagged autocorrelation (Spearman's Rho= -.05).

3). Finally, we note the earlier finding that error trials tend to be preceded by trials that systematically decrease in speed. Thus the fluctuations in reaction time might be due to systematic and controlled decreases *in reaction time, up to the point that an error occurs.*

Although all of these inferences are indirect, they point towards the conclusion that the short term fluctuations in reaction time are strategic, rather than being due to a strictly endogenous fluctuation of capacity, due, say, to changes in states of arousal.

General discussion

The main conclusions of this report are summarized as follows:

1) Two of the eight subjects were unable to perform the task adequately, while the remaining six displayed performances ranging from good to outstanding. Thus, subjects with modest amounts of motivation can perform this rather difficult sustained attention task.

Overall levels of performance for the six good subjects were similar, and much of the differences in accuracy and reaction times appear to be strategic, that is, different subjects were setting different speed-accuracy tradeoff criteria. On the other hand, individual differences could be observed in the details of the performance, such as in the trends.

2) There was weak evidence of periodicity that was invariant across baserates, in the range of about three to five minutes. However, this would need to be replicated, using a broader range of baserates, in order to be convincing.

3) There were local fluctuations in reaction times, in the range of a several or more trials. Consecutive positive autocorrelations were found at lags to as far back as lag 11.

Three facts argued that these positive autocorrelations were due to strategic control of response time, rather than being a strictly endogenous phenomenon:

A). The size of the first order autocorrelation did not correlate with the baserate. This argues against the view that the fluctuations are stimulus independent local fluctuations in arousal or attention level.

B). The autocorrelation was stronger with more accurate subjects. If better subjects have better strategic control of response times, this would be consistent with the strategic view.

C). Reaction times lagged back from error trials were, like the error trials themselves, fast. This implies that subjects increase their response speed until an error occurs.

4). There was a PRP (psychological refractory period) effect. When the rate of stimulus presentation was very fast (800 ms), subjects had difficulty keeping up and responding by the deadline, that is, the onset of the next stimulus. This was true even when the deadline contained most or all of the reaction time distribution for slower presentation rates. This implies that subjects are not fully prepared to process a stimulus at the moment that they have responded to a prior one.

This has important implications for the presentation of information in stimulus displays in a man-machine environment. Estimates of the temporal stimulus density that the human operator can handle which are based on reaction time distributions will not be accurate: recovery time must be allowed for.

Tables 1 (experiment 1) and 2 (experiment 2): summary statistics for all analyzed blocks. Descriptions of the table columns:

1. Subject/block number.
2. Block# within session.
3. Number of trials. NG (no good) indicates too many late-response errors, block not analysed.
4. Modality (A=auditory, V=visual).
5. Baserate (ISI) in ms.
6. % correct, not including responses that were too late (see note below).
7. Mean reaction time.
8. SD, reaction time.
9. SD, residuals from a smooth curve fit to the data using Cleveland's LOWESS smooth estimator.
10. RMS residual from AR(1) model fit to the residuals from the Lowess fit. When the AR(1) model is not significant, this can exceed slightly the SD estimate in the preceding column due to differences in the estimation procedure.
11. Spearman's rho between trial number and RT, used as an index of monotonicity. This statistic should be interpretated with caution, since many blocks displayed more complex trends which were (in some cases) superimposed on a monotonic trend.
12. Lag(1) autocorrelation (*=significant at the .05 level).
13. Coefficient for a fit of a first order autoregressive model (AR(1)) to the residuals from the LOWESS fit (*=significant at the .05 level).

NOTES FOR TABLE ONE AND TWO.

Only blocks with high performance levels were analyzed. There were a number of blocks in which the subjects' responded after the deadline, i.e., after the presentation of the next stimulus. Due to a software bug, these were recorded as times longer than the baserate, rather than as missed trials. Thus these mistakes are not reflected in the percent corrects (column 6). However, these were noted on the basis of visual inspection of the scatterplots, and when it was deemed that the number of these trials was sufficiently large as to make analysis problematic, these blocks were dropped from the analysis. These almost invariably occured at the fast stimulus presentation rates. These blocks are marked NG (no good) in column 3 of the data tables. The blocks that were analyzed were deemed to have, for the most part, sufficiently few of the outlying trials that the computed summary statistics would not display qualitative changes with the omission of these trials.

In experiment 2, subject MLB's data were analyzed, but most of the blocks displayed unacceptable performance levels. Also in this experiment, subject HLW was unavailable for two of the blocks.

• $P<.05$, •• $P<.01$.

§ Singular Hessian; standard errors not computable (KAB15, experiment 1).

§§ First auditory block for this subject (JPB5, experiment 1), high error rate and large reaction time variability.

Table 1. summary of experiment 1.

1	2	3	4	5	6	7	8	9	10	11	12	13
JPB1	1	180	V	2000	100.0	519.0	103.5	93.8	94.0	.43••	0.012	0.010
JPB2	2	300	V	1200	98.7	457.4	91.1	85.9	84.1	-.17••	0.212•	0.210•
JPB3	3	225	V	1600	99.2	476.4	106.9	78.2	69.2	.59••	0.465•	0.470•
JPB4	4	NG	V	800								
JPB5	1	88	A	1600	79.8	367.2	203.0	202.3	202.5	-.10	0.023	0.020
JPB6	2	NG	A	800								
JPB7	3	300	A	1200	96.3	372.4	101.5	94.2	94.2	.26••	-0.054	-0.050
JPB8	4	180	A	2000	94.4	399.4	092.8	85.0	85.2	.23••	-0.027	-0.030
JPB9	1	180	A	2000	99.4	353.6	073.5	73.1	73.2	.00	0.054	0.050
JPB10	2	300	A	1200	98.0	334.3	066.6	66.2	65.8	.03	0.120•	0.120•
JPB11	3	NG	A	800								
JPB12	4	225	A	1600	98.2	336.6	067.4	67.4	67.4	.01	0.066	0.070
JPB13	1	NG	V	800								
JPB14	2	225	V	1600	99.2	375.7	062.9	62.3	62.3	-.13	0.058	0.060
JPB15	3	300	V	1200	96.7	408.7	077.6	74.4	72.0	-.25••	0.256•	0.260•
JPB16	4	180	V	2000	96.7	395.2	058.4	55.8	55.8	.18•	0.058	0.060
KAB1	1	450	A	800	100.0	343.5	055.3	45.4	36.9	.17••	0.590•	0.583•
KAB2	2	180	A	2000	99.4	473.7	038.9	37.5	37.7	-.26••	0.010	0.012
KAB3	3	225	A	1600	100.0	459.8	071.4	67.1	60.3	.51••	0.440•	0.441•
KAB4	4	300	A	1200	99.7	405.5	73.54	71.7	71.7	-.22••	-0.040	-0.044
KAB5	1	300	V	1200	94.0	320.4	062.0	60.3	59.2	-.10	0.190•	0.193•
KAB6	2	450	V	800	83.8	328.6	159.7	159.0	151.3	.15••	-0.310•	-.312•
KAB7	3	180	V	2000	100.0	426.5	075.4	68.1	64.9	-.14	0.310•	0.312•
KAB8	4	225	V	1600	99.6	367.8	063.7	63.4	60.9	-.03	0.280•	0.279•
KAB9	1	225	V	1600	97.5	361.4	105.2	105.0	102.2	.14•	-0.240•	0.237•
KAB10	2	180	V	2000	99.4	420.7	080.0	77.5	71.8	.02	0.390•	0.378•
KAB11	3	450	V	800	93.6	323.6	086.5	85.9	85.6	.17••	0.090	0.087
KAB12	4	300	V	1200	99.7	374.0	059.3	58.7	53.7	.03	0.410•	0.405•
KAB13	1	300	A	1200	100.0	376.3	045.3	42.7	38.1	.19••	0.450•	0.454•
KAB14	2	225	A	1600	100.0	578.2	063.6	50.5	44.2	.39••	0.490•	0.488•
KAB15	3	180	A	2000	100.0	534.3	150.8	146.4	8	.09	0.570•	0.565•
KAB16	4	450	A	800	98.4	328.2	069.4	63.7	54.5	.08	0.520•	0.520•
KCG1	1	300	A	1200	95.7	402.2	120.1	119.6	117.9	.02	-0.177•	-0.18•
KCG2	2	225	A	1600	98.7	399.5	109.9	110.0	110.1	-.14•	0.050	0.060
KCG3	3	180	A	2000	90.0	515.9	190.8	189.5	188.5	.26••	0.127	0.130
KCG4	4	NG	A	800								
KCG5	1	225	V	1600	97.9	379.5	072.1	70.1	68.3	-.14•	0.232•	0.230•
KCG6	2	180	V	2000	97.2	351.7	070.1	66.6	65.4	.00	0.203•	0.20•
KCG7	3	NG	V	800								
KCG8	4	300	V	1200	94.0	337.8	075.1	75.0	74.4	-.07	0.145•	0.15•
KCG9	1	300	V	1200	97.7	328.3	053.7	53.2	52.2	-.10	0.201•	0.20 •
KCG10	2	450	V	800	87.8	310.9	099.1	98.2	98.2	.20••	-0.034	-0.030
KCG11	3	180	V	2000	96.7	335.0	061.0	60.0	58.7	.02	0.219•	0.22•
KCG12	4	240	V	1600	97.5	322.5	057.0	53.5	53.3	-.32••	0.112	0.120
KCG13	1	NG	A	800								
KCG14	2	180	A	2000	98.3	412.7	130.5	131.1	131.1	-.11	0.079	0.080
KCG15	3	226	A	1600	98.7	396.5	102.0	100.8	100.8	-.05	0.060	0.060
KCG16	4	300	A	1200	95.7	359.4	094.8	94.7	94.7	.08	-0.056	-0.060

Table 2. Summary of experiment 2. Columns are as in table 1.

1	2	3	4	5	6	7	8	9	10	11	12	13
DMS1	1	534	A	1800	99.8	496.7	120.1	112.1	106.5	-0.10•	0.315•	0.32•
DMS2	2	800	A	1200	99.6	483.3	118.4	111.9	102.4	-0.28••	0.403 •	0.40•
DMS3	3	NG	A	1200								
DMS4	4	534	A	1800	99.8	413.5	110.0	107.5	102.0	-0.31••	0.317•	0.32•
DMS5	1	800	V	1200	98.0	405.2	93.5	93.3	92.6	0.05	0.126 •	0.13•
DMS6	2	534	V	1800	99.6	440.7	97.5	97.3	93.3	-0.02	0.286•	0.29•
DMS7	3	534	V	1800	99.3	404.3	88.6	86.7	83.1	-0.13••	0.283•	0.29 •
DMS8	4	800	V	1200	97.8	389.4	84.3	83.2	79.6	0.18••	0.293 •	0.29•
MLB1	1	5 BAD	V	1800	88.8	422.5	203.0	202.7	202.8	-0.06	-0.027	-0.03
MLB2	2	NG	V	1200								
MLB3	3	NG	V	1200								
MLB4	4	534	V	1800	92.5	404.6	138.7	138.5	138.4	0.09•	-0.059	-0.06
MLB5	1	NG	A	1200								
MLB6	2	NG	A	1800								
MLB7	3	some bad trials	A	1800	92.9	366.4	232.5	231.8	s	0.05	0.013	0.01
MLB8	4	NG	A	1200								
HLW1	1	534	A	1800	98.9	474.0	133.2	133.2	127.0	-0.10•	0.303•	0.30 •
HLW2	2	NG	A	1200								
HLW3	3	800	A	1200	93.5	341.2	126.6	123.8	122.0	0.14••	0.175•	0.17 •
HLW4	4	534	A	1800	99.8	452.6	151.2	148.0	142.5	-0.07	0.274•	0.27•
HLW5	1	not run										
HLW6	2	not run										
HLW7	3	534	V	1800	96.8	426.7	115.2	111.1	111.2	0.33••	0.022	0.02
HLW8	4	800	V	1200	91.6	365.0	97.3	97.0	95.1	0.06	0.203•	0.20•
SLP1	1	800	V	1200	86.6	333.8	108.3	107.6	107.6	-0.11 ••	-0.008	-0.01
SLP2	2	534	V	1800	90.6	340.8	78.7	78.3	78.1	0.08	0.091	0.09 •
SLP3	3	534	V	1800	95.3	356.3	103.3	101.0	100.9	0.20••	-0.068	-0.07
SLP4	4	800	V	1200	91.6	349.5	97.2	96.8	96.9	-0.08•	0.002	0.00
SLP5	1	534	A	1800	89.5	340.4	151.9	151.6	151.5	-0.04	-0.066	-0.07
SLP6	2	NG	A	1200								
SLP7	3	NG	A	1200								
SLP8	4	534	A	1800	91.0	340.5	135.3	134.9	134.6	-0.07	-0.073	-0.07

Table 3. Summary of the performances of the six good subjects, based on data in tables one and two.
NOTE: Numbers in columns 5 and 8 are positive, negative, non-significant.

Experiment	Subject	# blocks run	# good blocks	Trend (rho): +/-/NS	Mean PC on good blocks	Grand mean RT	# positive autocorrelations: +/-/NS
1	KCG	16	13	2/3/8	95.8	373	2/3/8
1	JPB	16	12	5/2/5	96.4	400	4/0/8
1	KAB	16	16	7/2/7	97.8	401	11/2/3
2	SLP	8	6	1/2/4	90.8	344	0/0/6
2	HLW	6	5	2/1/2	96.1	412	4/0/1
2	DMS	8	7	0/5/2	99.1	433	7/0/0

Table 4. Mean magnitude components of Fourier transforms of the time series for experiment 2. Means are averaging across replications and modality, but within subject and baserates.

SUBJECT:	SLP		DMS		HLW	
Wavelength	1200	1800	1200	1800	1200	1800
∞	5.5	12.2	6.2	11.5	13.8	21.0
512.8	3.2	3.6	6.7	6.6	6.9	5.5
255.8	1.4	5.8	11.4	7.9	10.0	12.8
170.6	6.0	4.7	10.6	10.2	8.3	11.7
128.0	1.8	7.0	11.2	13.7	5.4	9.7
102.4	3.8	4.7	2.6	5.1	4.7	5.3
85.3	3.7	7.4	4.8	5.2	4.5	7.4
73.2	3.4	3.3	6.4	8.0	7.8	11.4
64.0	4.7	4.0	10.3	7.5	4.4	6.0
56.9	5.3	4.6	4.6	5.0	6.8	7.0
51.2	3.6	5.1	8.3	6.9	6.9	10.4
46.6	4.5	2.4	7.4	6.5	4.9	2.0
42.7	3.9	2.8	6.7	3.1	4.8	9.3
39.4	3.9	1.9	7.9	5.4	5.6	10.7
36.6	4.3	4.5	8.4	6.6	5.7	4.8

Table 5 Summary of the preferences (magnitude maxima) for the subjects in experiment 2

			1200 ms		1800 ms	
Subject	1200 prefs	1800 prefs	secs	mins	secs	mins
DMS	255		306.0	5.1	0	0
DMS	170	170	204.0	3.4	306	5.1
DMS	128	128	153.6	2.56	230.4	3.84
SLP	170	128	204.0	3.4	230.4	3.84
HLW	256	256	307.2	5.12	460.8	7.68

Figure 1. Times series plot of the data for KAB. auditory 800 ms. condition. Curve fit uses Cleveland's lowess smoother. The block is representative, except that most blocks did not display as many changes (dy/dt=0) in the fitted curve.

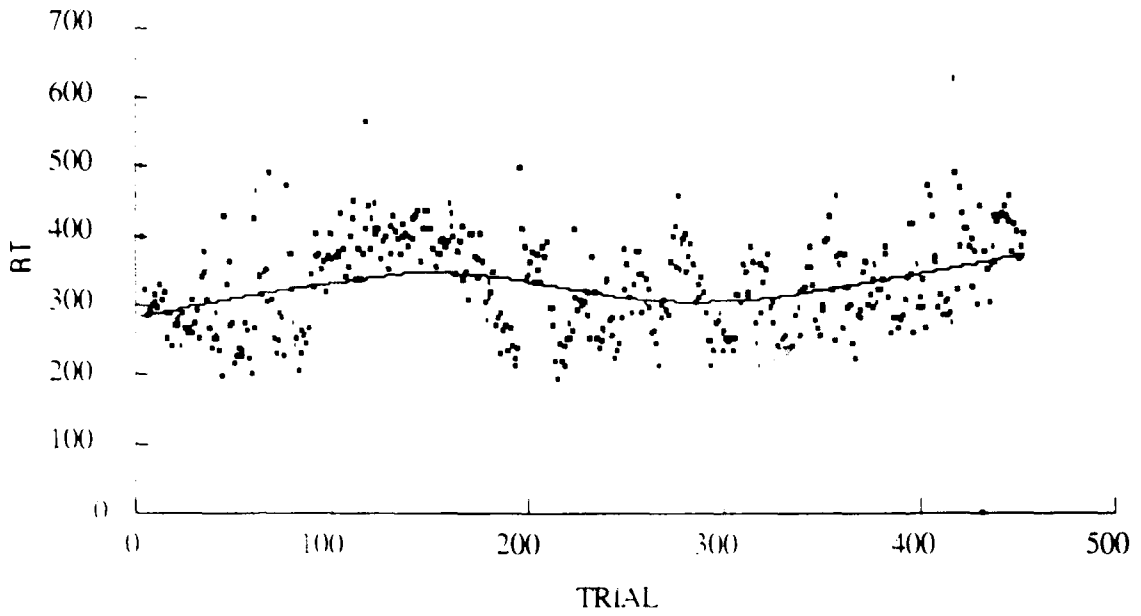


Figure 2. Mean reaction time for error trials (lag 0) and correct trials lagged back from the error trials, by subjects.

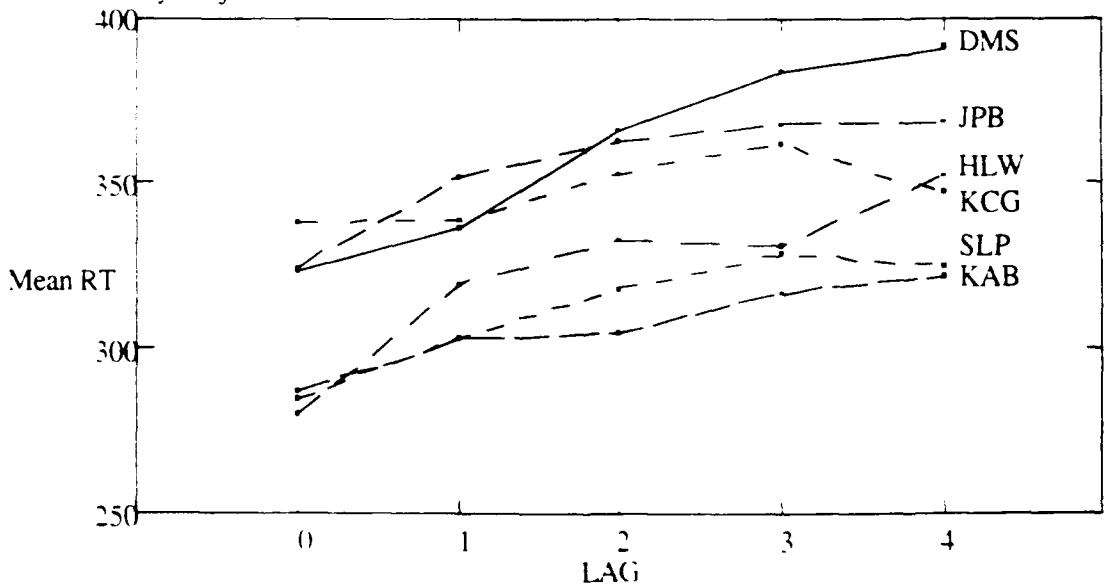


Figure 3 A & B. Mean reaction time for error trials (lag 0) and correct trials lagged back from the error trials, by modality (A) and baserate (B).

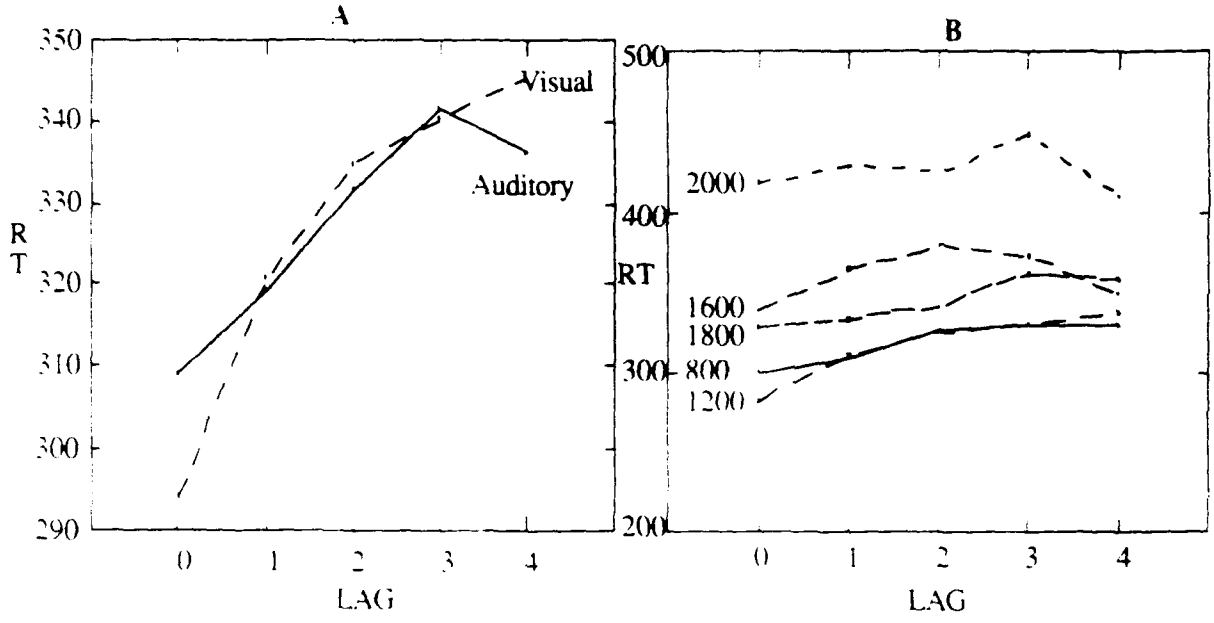


Figure 4. Mean reaction times by baserate and sensory modality. Includes error trials.

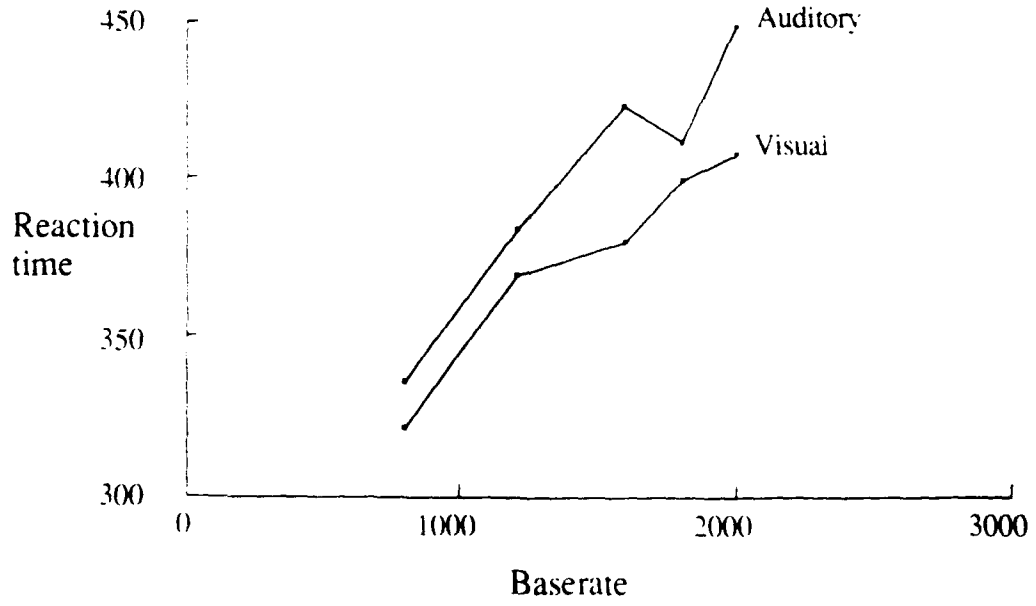
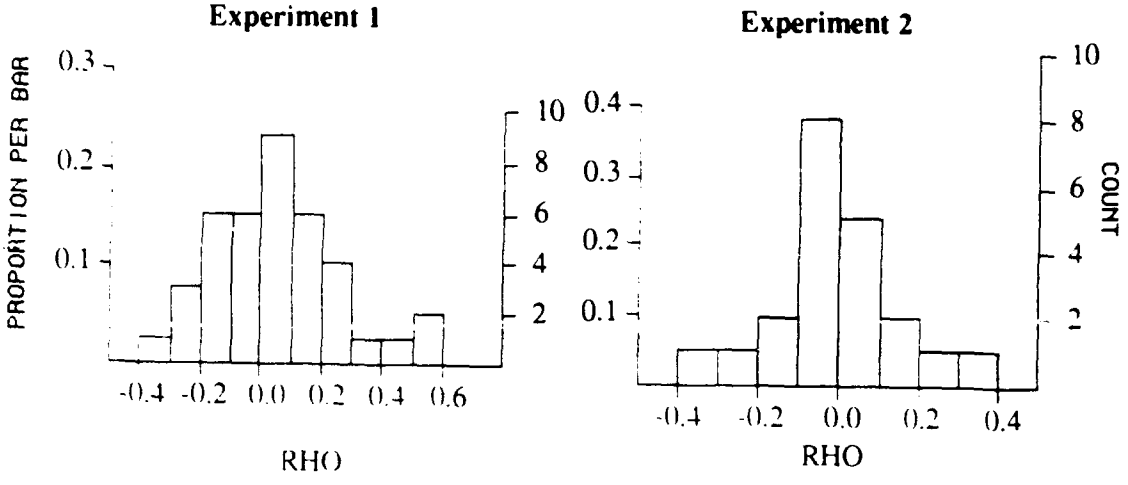


Figure 5 Distribution of Spearman's rho between event and reaction times. Negative numbers indicate improvement (decreasing RT's), negative numbers, increasing RT's.



REFERENCE

Welford, A.T. 1968 Fundamentals of Skill. London:Methuen.

SPEAKER IDENTIFICATION USING BESSEL FUNCTION EXPANSION OF SPEECH SIGNALS

Kaliappan Gopalan

Associate Professor
Department of Engineering
Purdue University Calumet
Hammond, IN 46323

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

August 1993

SPEAKER IDENTIFICATION USING BESSEL FUNCTION EXPANSION OF SPEECH SIGNALS

Kaliappan Gopalan

Associate Professor
Department of Engineering
Purdue University Calumet
Hammond, IN 46323

Abstract

Identification of speakers using Bessel function representation of speech signals was studied. Coefficients in the Fourier-Bessel expansions of frames of speech with (a) $J_0(t)$, and (b) $J_1(t)$ as basis functions were used as feature vectors. In both cases it was determined that at least 20 coefficients that have the largest magnitude in the expansion were needed to obtain a reasonable quality of synthesized speech. Of the different feature vectors obtained from the expansion coefficients, higher scores for speaker recognition resulted with $J_1(t)$ than with $J_0(t)$. In addition, for the same dimensionality, $J_1(t)$ expansion showed better speaker identification scores than reflection coefficients from linear predictive analysis.

Introduction

Decomposition of speech signals using nonsinusoidal basis functions has the advantage that the resulting representation may have less dimension than the one using the periodic sinusoidal functions. Since each of the Bessel functions of the first kind, $J_n(t)$, $n = 0, 1, ..$ is quasiperiodic with the interval between successive zero-crossings approaching π, they have the structural similarity to short-time voiced frames of speech. Based on this observation, representation of speech and identification of speakers using $J_0(t)$ and $J_1(t)$ can be achieved. The efficiency of these applications may prove the viablity of the Bessel function decomposition as an alternative to the common spectral domain methods.

Fourier-Bessel Expansion

Expansion of an arbitrary function $x(t)$ in the interval $0 < t < a$ using $J_n(t)$ (for a given n) is given by the Fourier-Bessel series expansion [1]

$$x(t) = \Sigma C_m . J_n([x_m/a].t), \qquad (1)$$

where $\{ x_m, m = 1, 2, .. \}$ are the roots of $J_n(t) = 0$.

Using the orthogonality of $J_n([x_m/a].t)$ for a given n, the coefficients C_m in the expansion are given by

$$C_m = (2/[a^2 J_n'(x_m)]) \int_0^a t.x(t).J_n([x_m/a].t)dt \qquad (2)$$

Of particular interest are the zeroth and the first order functions, $J_0(t)$ and $J_1(t)$, because of their relative simplicity in computation.

The coefficients using these two functions are, from Eq. (2) and the Bessel function identities,

$$C_{m0} = \{2/[a^2 J_1(x_m)]\} \int_0^a t.x(t).J_0([x_m/a].t)dt \qquad (3)$$

for $J_0(t)$, and $J_1(x_m) = 0$, $m = 1, 2, \ldots$

and

$$C_{m1} = \{2/[a^2 J_0(x_m)]\} \int_0^a t.x(t).J_1([x_m/a].t)dt \qquad (4)$$

for $J_1(t)$, and $J_1(x_m) = 0$, $m = 1, 2, \ldots$

The coefficients in the expansion, which are the peak amplitudes of the corresponding Bessel functions, are clearly unique for a given x(t). Therefore, speech signal may be represented using $J_0(t)$ or $J_1(t)$ in accordance with the above equations. Such a representation is useful in waveform coding for efficient storage and transmission, speech recognition, and speaker identification and verification applications.

Speech Signal Representation

As with sinusoidal representation, description of the nonstationary speech signal is performed on a short interval of speech to preserve the time-varying nature of the signal. After experimenting with different intervals for speech frames, with and without overlap, it was found that using 200 samples per frame, or segment, with 100 sample overlap was computationally optimum. Each frame of speech data was then multiplied by a 200-point Hamming window to preserve the spectral characteristics of the signal.

Using the discretized versions of Eqs. (3) and (4) the coefficients C_{0m} and C_{1m} (with basis functions $J_0(t)$ and $J_1(t)$) were obtained for each of the windowed frame of sampled speech data. Although Eq. (1) is an infinite series, it approches x(t) rapidly with a finite summation using large index m. Fig. 1 shows the 3-D plots of the first 20 coefficients, C_{0m} and C_{1m}, m = 1, .. 20, as a function of time (corresponding to frame index) and the index m for the same utterance of a speaker. The coefficients in these plots represent the relative (peak) amplitudes of the constituent Bessel functions that decompose the speech waveform, similar to the spectral energy displayed in a spectrogram.

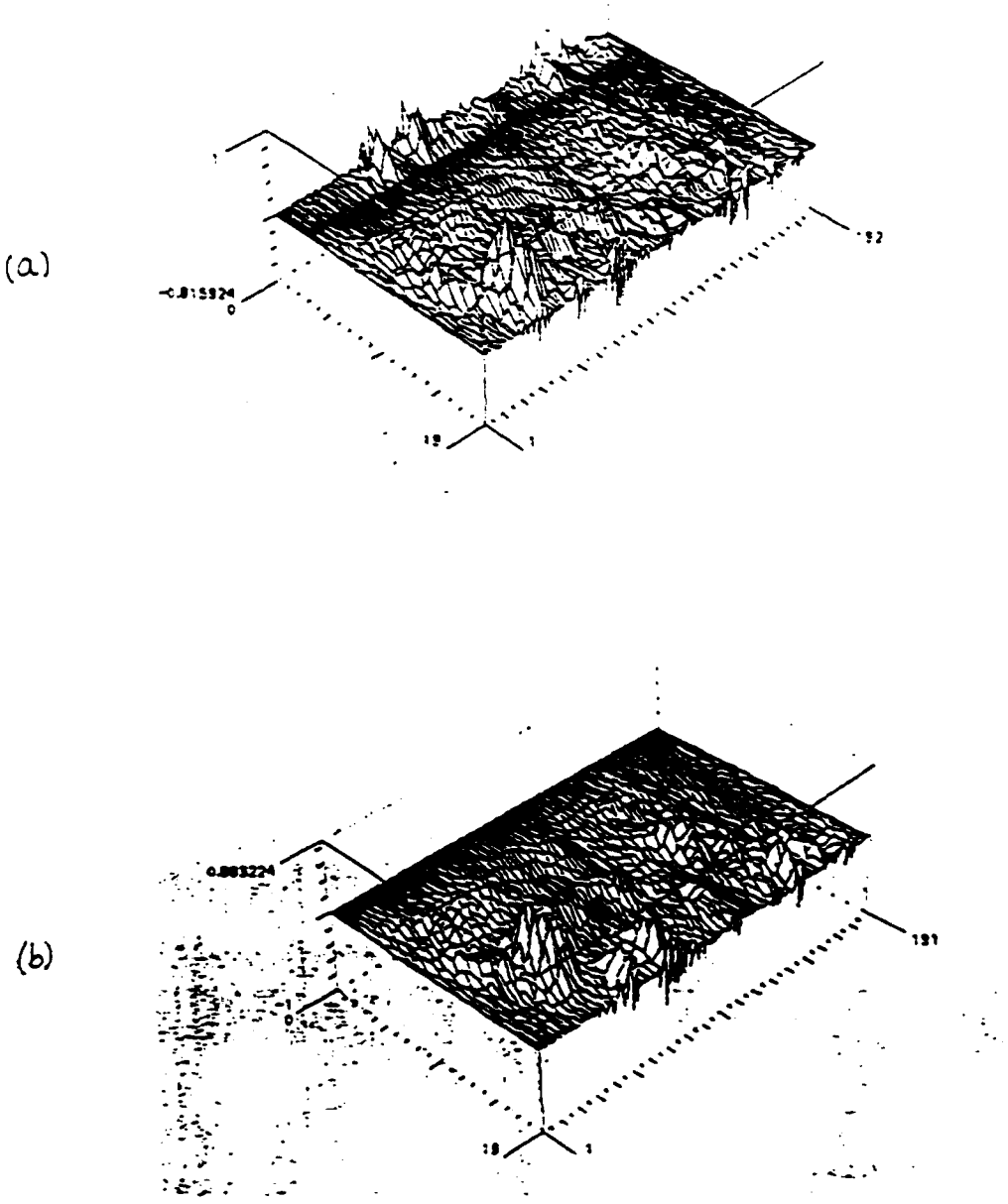


Fig. 1 Plots of the first 20 coefficients for the utterance "We were away a year ago" by a male speaker: (a) C_{1m} (b) C_{0m}

Data Reduction

The effectiveness of the Bessel function representation of speech was studied using different number of coefficients in the summation given in Eq. (1). When 100 to 150 coefficients, from either $J_0(t)$ or $J_1(t)$, were used in the summation, all the characteristics of the original speech, including the identity of the speaker and the quality of synthesized speech, were observed to be present. This was also confirmed by the spectrogram of the synthesized speech which was close to that of the original speech. The speech quality began to deteriorate for M, the upper limit of m, between 30 and 50; the message quality was still acceptabale, however. At the low end of M, particularly at 20, the identity or the gender of the speaker was difficult to determine, although the message was still discernible with background noise. Synthesized speech using only the first 10 terms had very little message quality with only the low frequency components present.

The above observations were correlated by the spectral characteristics of $J_0(t)$ and $J_1(t)$. Since both the Bessel functions are bandlimted to

$$w_{max} = x_{0m}/a, \quad J_0(x_{0m}) = 0, \; m = 1, 2, \ldots$$

and

$$w_{max} = x_{1m}/a, \; J_1(x_{0m}) = 0, \; m = 1, 2, \ldots$$

respectively, a finite number of terms, m = 1, 2, .. M in the

summation in Eq. (1) limits the spectrum of the synthesized speech to x_{im}/a, i = 0,1. Therefore, to retain the frequency components up to 3 kHz, for example, upper limit for M such that $x_{iM} = 2\pi(3000)a$ must be used for both $J_0(t)$ and $J_1(t)$. For the original speech sampled at 11000 samples per second, using 200 samples per frame gives the range a as 18.2 ms. Hence, $x_{iM} = 342.72$. Since the 109th roots of $J_0(t)$ and $J_1(t)$ are close to this value, the finite summation must have up to 109 terms to retain the spectral components up to 3 kHz. For M < 109, the synthesized speech loses the high frequency information. At M = 50, for example, the spectral energy of the reconstituted speech is limited to below 1500 Hz, which may be below the second formant for certain speakers and/or phonemes; it certainly cannot have the original speech quality of fricatives. The message content may still be preserved, albeit with noise, depending on how large M is; hence, a lower value of M (below 50) may be acceptable for storage and transmission of speech using only the first M coefficients.

Improvement in speech quality, particularly for fricatives, requires certain amount of high frequency information which correspond to higher value of M. If the high frequency components are predominant in any frame in the original speech waveform, they manifest in the form of large coefficients C_m (in magnitude) for large index m in the Bessel coefficient domain. Therefore, by selectively choosing the coefficients in each frame based on their relative magnitudes, spectral energy at both low and high frequencies may be retained in the synthesized speech. Thus, at a little additional processing (arising from sorting of C_m

in each frame), higher speech quality can be achieved at lower number of coefficients. This assertion was verified using the top 50, 30, 20 and 10 coefficients from the $J_i(t)$ expansion. It was found that a minimum of 20 sorted coefficients were needed for an acceptable quality of speech without missing the identity of the speaker. This compares favorably with the requirement of more than 30 unsorted coefficients for low noise in synthesized speech [2].

Speaker Identification

Since the coefficients C_n uniquely describe the peak amplitudes of the constituent Bessel functions $J_i(([x_/a].t)$, i = 0 or 1, their set for an entire utterance may be considered a feature vector [3]. Clearly, the efficacy of a feature vector depends on how well it can characterize or synthesize the original speech. Having established that a minimum of 20 coefficients - ordered or otherwise - were needed for useful, synthesized speech, speaker recognition was performed using different forms of the coefficients as feature vectors. In each of the testing described below five utterances of each speaker were used to form reference feature vectors. A vector-quantized codebook was created for each speaker using the set of five reference vectors. A final codebook was obtained for the set of five speakers. Feature vectors derived from the remaining utterances of the speakers were used in a clustering process with the codebook of vectors. The unknown speaker was identified as the one whose centroid in the codebook is the closest to the unknown vector. A total of 29 unknown utterances - 18 male and 11 female - were tested using the codebook.

(a) First 20 coefficients using $J_1(t)$ expansion: The collection of the first 20 coefficients in each frame of speech as reference vector was computationally fast and simple. The following recognition scores were obtained using this vector.

Speaker	m1	m2	m3	f1	f2	Score
m1	1	-	-	1	4	1/6
m2	-	5	-	-	-	5/5
m3	-	5	2	-	-	2/7
f1	-	-	-	5	-	5/5
f2	-	-	-	-	6	6/6
Overall score						19/29

With coefficients using $J_0(t)$, the overall score was 15/29.

(b) Sum of groups of five squared coefficients using $J_1(t)$ expansion: This is analogous to the spectral energy in different discrete frequency bands. With 150 coefficients in each frame, the dimension of feature was reduced to 30 per frame. The following is the score table.

Speaker	m1	m2	m3	f1	f2	Score
m1	6	-	-	-	-	6/6
m2	-	5	-	-	-	5/5
m3	-	7	0	-	-	0/7
f1	-	-	-	5	-	5/5
f2	-	-	-	1	5	5/6
Overall score						21/29

Sum-square of $J_0(t)$ coefficients yielded the same overall score.

(c) Twenty five selected coefficients: To include coefficients corresponding to high and low frequency variations of the signal, five sets of five coefficients each were chosen in each frame in the following order: C_{10} to C_{14}, C_{40} to C_{44}, C_{70} to C_{74}, C_{100} to C_{104}, and C_{130} to C_{134}. These coefficients cover five bands of frequencies centered at approximately 350 Hz, 1200 Hz, 2000 Hz, 2800 Hz, and 3600 Hz, each with a bandwidth of about 110 Hz. Because of the inclusion of higher frequencies, the feature vector of 25 selected coefficients yielded a higher score as shown below.

Speaker	m1	m2	m3	f1	f2	Score
m1	5	-	1	-	-	5/6
m2	-	5	-	-	-	5/5
m3	2	-	5	-	-	5/7
f1	-	-	-	5	-	5/5
f2	-	-	-	-	6	6/6
Overall score						26/29

In the case of $J_0(t)$, the overall score was 15/29.

To study the effect of voicing information on the identification of speakers, each of the above three feature vectors was combined with the trajectory of voicing; only the

feature vectors corresponding to the frames having a specified voicing probability of p or above were used as final feature vectors. Test results for probability p = 0.75 and p = 0.9 showed virtually no change in the overall identification scores for both $J_2(t)$ and $J_1(t)$. This is to be expected since the entire sentence used contains mostly voiced speech.

For comparison with the widely used linear predictive (LP) analysis, reference vectors based on the reflection coefficients of a 20th order LP model were used to form a vector-quantized codebook. Speaker identification results based this codebook are given in the following table.

Speaker	m1	m2	m3	f1	f2	Score
m1	6	-	-	-	-	6/6
m2	-	5	-	-	-	5/5
m3	3	-	4	-	-	4/7
f1	-	-	3	2	-	2/5
f2	-	-	-	-	6	6/6
Overall score						23/29

For the same dimensionality as the LP vectors, five groups of four coefficients each, namely, C_{10} to C_{13}, C_{40} to C_{43}, C_{70} to C_{73}, C_{100} to C_{103}, and C_{130} to C_{133} from the $J_1(t)$ expansion were used as feature vectors. Identification scores are as given below.

Speaker	m1	m2	m3	f1	f2	Score
m1	6	-	-	-	-	6/6
m2	-	5	-	-	-	5/5
m3	3	-	3	1	-	3/7
f1	-	-	-	5	-	5/5
f2	-	-	-	-	6	6/6

Overall score 25/29

Clearly, the selected set of 20 coefficients from the expansion using $J_1(t)$ show better speaker identification results than the LP reflection coefficients of the same size.

Conclusion

Based on the results for the small size of the data base, the following preliminary conclusions may be reached. (a) A reasonable quality of speech retaining most of the spectral characteristics of the original speech can be synthesized using 20 - 25 selected or sorted Fourier-Bessel expansion coefficients with $J_1(t)$ as the basis function. With $J_0(t)$, larger number of coefficients is needed for the same speech quality. For improved quality in synthesized speech, coefficients with higher indices must be included. (b) Expansion coefficients using $J_1(t)$ serve as better feature vectors for identification of speakers than those using $J_0(t)$. Based on (a), coefficients bearing wider spectral information must be used for higher identification scores. In particular, selected groups of short range of coefficients, covering discrete bands of frequencies over the entire Nyquist range, yield better identification results.

(c) Since the test sentence has mostly voiced sounds, it appears that feature vectors using only the voiced frames in any utterance, rather than the entire speech, are required for speaker identification. Additionally, with voiced frames, only certain coefficients in each frame, or their indices may be necessary to represent a speaker - analogous to spectral energy in the formants, or the frame to frame trajectory of formants.

Further work relating fundamental frequency and formants of a speaker in terms of the indices of $J_i(t)$ expansion coefficient may lead to a compact representation similar to the codes in a vector-quantized codebook. The indices-based representation will be useful in secure speech transmission. Processing complexity and the efficiency of representation need to be investigated.

Characterization of phonemes using the indices of the peak coefficients may be useful in speech synthesis.

References

[1]. I.H. Sneddon, Fourier Transforms, New York: McGraw_Hill, 1951.

[2]. C.S. Chen, K. Gopalan and P. Mitra, "Speech Signal Analysis and Synthesis via Fourier-Bessel Representation, Proc. IEEE ICASSP, Tampa, FL, March 1985, pp 497-500.

[3]. K. Gopalan, "Discrete Utterance Recognition using Fourier-Bessel Expansion," Proc. 28th Midwest Symp. on Circuits and Systems, Aug. 1988, St. Louis, pp. 1086-1088.

EFFICIENT COORDINATION OF AN ANTHROPOMORPHIC TELEMANIPULATION SYSTEM

Ming Z. Huang
Associate Professor
Department of Mechanical Engineering
and Robotics Center

Florida Atlantic University
500 NW 20th Street
Boca Raton, FL 33431

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

Florida Atlantic University

July 1993

EFFICIENT COORDINATION OF AN ANTHROPOMORPHIC TELEMANIPULATION SYSTEM

Ming Z. Huang
Associate Professor
Department of Mechanical Engineering
and Robotics Center
Florida Atlantic University

Abstract

This report documents the development of coordination algorithms for control implementation of an anthropomorphic telemanipulation system presently at Wright-Patterson Air Force Base. The telemanipulation system, which is to be used as a research platform in facilitating studies on human sensory feedback, comprises a 7 d.o.f, force-reflecting, exoskeleton master and a 6 d.o.f. articulated slave robot. The approach taken in the development emphasizes on the practical issue of computation efficiency -- a primary concern for satisfactory real-time operations. The algorithms presented here have been fully tested and implemented on the system. Implementation results indicate at least a five-fold improvement on the control sampling rate has been achieved (from 11 Hz to 62 Hz on a 68020-based VME board). Other practical issues of implementation are also discussed in this report.

EFFICIENT COORDINATION OF AN ANTHROPOMORPHIC TELEMANIPULATION SYSTEM

Ming Z. Huang

I. Introduction

Recently, studies in telepresence , which involve "man-in-the-loop" control of sensory-rich, remotely operated robotic systems, have emerged as a new critical area of research and development. This is due to the increasing recognition and acceptance of telerobotic manipulation technology as being a viable solution for remote operations in unstructured and/or hazardous environments, such as space, undersea, or nuclear sites [1-3].

For years, man-machine interface research has been a main thrust of the efforts in the Air Force to improve performance and effectiveness of its crews. Under Crew System Directorate of the Armstrong Laboratory at Wright-Patterson AFB, the Human Sensory Feedback (HSF) research program is charged with the mission to investigate telepresence and its related issues. Among the on-going research activities of the HSF program, main emphasis is currently concentrated on characterization of the role of human sensory feedback in the following three key aspects: namely, coarse positioning (large scale motion associated with the human arm and wrist), fine manipulation (small scale motion associated with the human hand), and tactile feedback.

A unique telemanipulation test platform has been designed and built to support the research in the coarse manipulative human sensory feedback. The platform is unique in that it uses a custom-built, anthropomorphic exoskeleton capable of force reflection as the master control arm to command a kinematically dissimilar slave (a revolute-jointed industrial type) robot. The Force-REFLecting EXoskeleton ('FREFLEX') master is a seven degree-of-freedom, cable-driven robot that was designed specifically to provide mobility and range of motion similar to that of a human arm (Fig. 1). The use of such an anthropomorphic exoskeleton as the master controller enables the operator to generate motions and to react to forces encountered during manipulation in a natural way, a key functional requirement in human sensory feedback studies.

Force reflection on the FREFLEX is achieved by its controller generating appropriate antagonistic actions through cables ('tendons') and pulleys, driven by brushless DC motors.

While use of cables and pulleys make it possible to drive distal links from actuators mounted at the base, significantly reducing overall weight of the arm, such an actuation scheme, results in severe cross-coupling between motions of the joints and, consequently there is coupling between joint torque commands. Given a certain force/moment to be reflected by the FREFLEX it is necessary to identify the coupling relationship in order to compute actual torque commands at the actuators.

When performing telemanipulation it is desirable that the master controller appears as being 'weightless' to the operator. Besides the obvious benefit of reducing operator fatigue, gravity and inertia compensation increases the fidelity of manipulative interaction and hence the overall system performance. Update rate in the control system is another factor which also affects the overall system performance. The system will feel sluggish or even become unstable if the cycle time is too large. In general, it is necessary to have an update rate of 20 to 200 Hz to ensure satisfactory real-time performance.

In this report, we will present analyses and solutions to the problems of coordination concerning the control implementation of the telemanipulation system as described above. Specifically, kinematic models for both the FREFLEX and the slave (Merlin^{TM} 6500 by American Robot Corp.) and their related kinematics solutions (both position and velocity) will be developed. In addition, algorithms for FREFLEX to compensate gravity loads and compute joint torques, including identification of the kinematic coupling relationship relating the actuators and the joints, will also be presented. It is noted that throughout the following development of algorithms we have taken special care to optimize computational efficiency. As a result, an five-fold improvement of the overall system update rate has been achieved with their implementation.

II. Coordination of the FREFLEX Master Robot

Coordination for FREFLEX can be divided into the following stages of computation: forward position kinematics, Jacobian, gravity compensation, kinematic coupling, and joint torque decomposition. Note that for a single microprocessor system all of these computations must be completed before next update to the controller can be made; in other words, the system sampling rate is dictated by the overall efficiency of these computations. Consequently, when formulating the solution to each stage, a sensible guideline would be to develop the algorithms in such a way that all items should be computed only once and any computation which occurs in the later stages should take maximum advantage of what has been computed before.

Forward Position Kinematics

Forward position kinematics refers to the problem that, given a set of measured joint positions, compute the position and orientation of the FREFLEX hand grip (or any conveniently chosen point of interest) in the Cartesian space. The resulting pose (combined position and orientation) is then used in the inverse kinematics solution for the slave robot (to be described later) to yield the corresponding joint commands for the slave to be driven to that same pose. Note that this transformation of position command in Cartesian space is necessary whenever the master and the slave robots are of different geometries.

To establish a kinematic model for the FREFLEX, we adopt the so-called Denavit-Hartenberg (D-H) modeling convention with the frame assignment scheme similar to that adopted in [4] and [5]. Figure 2 shows a schematic of FREFLEX with definitions of all the kinematic frames. Based on the above D-H convention, a 4x4 homogeneous transformation, denoted as ${}^{i-1}T_i$, can be derived to represent the position and orientation of frame i relative to the frame (i-1); refer to [4] or [5]. It can be easily shown that the coordinate transformation representing position and orientation of the last frame (frame 7 for FREFLEX) relative to the base frame (frame 0) can be obtained as:

$$ {}^0\mathbf{T}_7 = \prod_{i=1}^{7} {}^{i-1}\mathbf{T}_i = {}^0\mathbf{T}_1\,{}^1\mathbf{T}_2 \cdots {}^6\mathbf{T}_7 \qquad (1) $$

or equivalently, in terms of ${}^{i-1}\mathbf{R}_i$ and ${}^{i-1}\mathbf{q}_i$, as:

$$ \begin{aligned} {}^0\mathbf{R}_7 &= {}^0\mathbf{R}_1\,{}^1\mathbf{R}_2 \cdots {}^6\mathbf{R}_7 \\ {}^0\mathbf{q}_7 &= {}^0\mathbf{q}_1 + {}^0\mathbf{R}_1\,{}^1\mathbf{q}_2 + {}^0\mathbf{R}_2\,{}^2\mathbf{q}_3 + \cdots + {}^0\mathbf{R}_6\,{}^6\mathbf{q}_7 \end{aligned} \qquad (2) $$

where

$$ {}^{i-1}\mathbf{T}_i = \begin{bmatrix} c\theta_i & -s\theta_i c\alpha_i & s\theta_i s\alpha_i & a_i c\theta_i \\ s\theta_i & c\theta_i c\alpha_i & -c\theta_i s\alpha_i & a_i s\theta_i \\ 0 & s\alpha_i & c\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} {}^{i-1}\mathbf{R}_i & {}^{i-1}\mathbf{q}_i \\ 0\ 0\ 0 & 1 \end{bmatrix} $$

It is remarked that, while both equations (1) and (2) are completely equivalent, the latter, Eq. (2), results in better computation efficiency by separating the computations into rotational and translational parts. By the author's own experience, the two-part form of (2) has also been found to be more amenable to manipulate symbolically, particularly in deriving the inverse

kinematics solution. For example, in FREFLEX, from its parameter table we have: ${}^0\mathbf{q}_1 = {}^1\mathbf{q}_2 = {}^5\mathbf{q}_6 = \mathbf{0}$. This immediately leads to a simplification on the translational part of Eq. (2) to: ${}^0\mathbf{q}_7 = {}^0\mathbf{R}_2\,{}^2\mathbf{q}_3 + {}^0\mathbf{R}_3\,{}^3\mathbf{q}_4 + {}^0\mathbf{R}_4\,{}^4\mathbf{q}_5 + {}^0\mathbf{R}_6\,{}^6\mathbf{q}_7$ which is clearly easier to manipulate, although the rotational part remains unchanged. The inverse kinematics solution for the slave robot (MERLIN) to be included later, was also arrived at based on the above procedure.

Note that if, instead of the last frame (frame 7), the position and orientation of frame k is to be computed, both of the above equations still hold with only a change in the upper index from 7 to k needed. It can be seen that an efficient way to compute the pose of any link frame is to do so sequentially by starting with k=1 (from the base) and then progressing outward. In fact, it is also desirable in practice to facilitate the forward kinematics computations so that the positions and orientations of all the link frames (not just the hand grip) are easily accessible should they be needed in subsequent computations. Based on Eq. (2), an outward iteration algorithm which computes the pose of each link frame starting with the base frame can be implemented using the following recursive relationships (i = 1 to k):

$$\begin{aligned} {}^0\mathbf{q}_i &= {}^0\mathbf{q}_{i-1} + {}^0\mathbf{R}_{i-1}\,{}^{i-1}\mathbf{q}_i \\ {}^0\mathbf{R}_i &= {}^0\mathbf{R}_{i-1}\,{}^{i-1}\mathbf{R}_i \end{aligned} \qquad (3)$$

Knowing the position and orientation of each link frame (relative to the base), it is now straightforward to obtain the absolute position of any given point in any link. Let ${}^i\mathbf{p}$ be the position vector of a point in link frame i, then its corresponding position in the base frame, ${}^0\mathbf{p}$, is given by:

$${}^0\mathbf{p} = {}^0\mathbf{R}_i\,{}^i\mathbf{p} + {}^0\mathbf{q}_i \qquad (4)$$

In FREFLEX implementation, a forward kinematics solution has been developed using the above recursive scheme. The solution for the pose of each link frame is obtained in the form of explicit analytic expressions for optimal efficiency (see Appendix). It is recognized that one may achieve reasonable efficiency by direct numerical computations using the same recursive scheme with a carefully coded algorithm. However, to guarantee optimal efficiency one must ensure all necessary terms are computed once and only once, which requires being able to identify and eliminate redundant, repetitious computations of those common terms which may be embeded in more than one expression. This can be accomplished only by going through the process of analytic derivation. It is recommended that this be done whenever efficiency is a critical factor in system performance.

We have evaluated the computation efficiency of our algorithm in terms of the required operation counts for addition/subtraction (A), multiplication/division (M) and trigonometric function calls (F). For FREFLEX, the forward kinematics algorithm calls for a total of (64A+133M+14 F), as compared to the (162A+ 216M+14F) needed if implemented with direct numerical computation using Eq. (3).

Jacobian Formulation

As is well known in robotics, Jacobian is a transformation matrix which relates differential motions (linear and angular) of the robot end effector in the Cartesian space, to the corresponding differential displacements at the joints. It is also known that in statics consideration the same transformation can be used to relate the external force and moment applying at the end-effector to the torques (or forces, if prismatic) at the joints. Mathematically, the above statements can be expressed using the following equations:

$$\mathbf{J}\,\dot{\theta} = \begin{bmatrix} \omega \\ \mathbf{v} \end{bmatrix} \tag{5}$$

$$\mathbf{J}^{T} \begin{bmatrix} \mathbf{M} \\ \mathbf{F} \end{bmatrix} = \tau \tag{6}$$

where

$\mathbf{J}$: the manipulator Jacobian

ω, and $\mathbf{v}$: the end effector angular and linear velocities, respectively

$\mathbf{M}$ and $\mathbf{F}$: the resultant moment and force by the end effector, respectively

$\dot{\theta}$: nx1 vector of joint rates (n = manipulator's d.o.f.)

τ: nx1 vector of joint torques/forces

and superscript T indicates the matrix transpose operation.

Although the above relationships may have been seen in many texts, it would be helpful to clarify a few points when considering their applications. First and foremost, care must be taken to ensure that all vector quantities involved be formulated with respect to the same coordinate frame of reference. In addition, point-specific vectors, such as linear velocity $\mathbf{v}$ and moment $\mathbf{M}$, must be given such that they are all relative to the same point of reference which was used to formulate the Jacobian. In other words, the Jacobian may take on various different forms, depending upon the reference frame and the reference point chosen (to describe $\mathbf{v}$ or $\mathbf{M}$) during its formulation.

To emphasize the distinction between the various available forms, we shall denote the Jacobian with ${}^{m}\mathbf{J}_{k}$, in which the leading superscript, m, specifies the frame and the trailing subscript, k, gives the reference point. In general, ${}^{m}\mathbf{J}_{k}$ for a robot of n d.o.f. can be formulated using the following:

$$ {}^{m}\mathbf{J}_{k} = \begin{bmatrix} \mathbf{u}_{i} \\ \rho_{i} \times \mathbf{u}_{i} \end{bmatrix}; \; i = 1,2,\ldots,n \qquad (7) $$

in which $\mathbf{u}_{i}$ is the unit directional vector of joint axis i, and ρ_{i} is the position vector of axis i with respect to the reference point k; of course, both vectors are expressed in frame m. Note that the above form of column vector in Eq. (7) only applies to robots with all revolute joints. In the case for robots with prismatic joints, say joint j, then the j-th column should be replaced with $\left[(0,0,0); \mathbf{u}_{j}^{T}\right]^{T}$.

For FREFLEX, we have chosen to use ${}^{0}\mathbf{J}_{0}$, which is to formulate the Jacobian in the base frame (frame 0) with its origin as the reference point. Again, the main consideration here is to minimize computation cost. The reason to use ${}^{0}\mathbf{J}_{0}$ is two-fold: first, origin of frame 0 is the point of concurrency of axes 1, 2, and 3; and second, all the terms required to formulate the Jacobian are readily available from the forward kinematics procedure with no further manipulation necessary. As a result, we arrived at the following 6x7 matrix as the Jacobian for FREFLEX:

$$ {}^{0}\mathbf{J}_{0} = \begin{bmatrix} \hat{\mathbf{z}}_{0} & \hat{\mathbf{z}}_{1} & \hat{\mathbf{z}}_{2} & \hat{\mathbf{z}}_{3} & \hat{\mathbf{z}}_{4} & \hat{\mathbf{z}}_{5} & \hat{\mathbf{z}}_{6} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & {}^{0}\mathbf{q}_{3} \times \hat{\mathbf{z}}_{3} & {}^{0}\mathbf{q}_{4} \times \hat{\mathbf{z}}_{4} & {}^{0}\mathbf{q}_{5} \times \hat{\mathbf{z}}_{5} & {}^{0}\mathbf{q}_{6} \times \hat{\mathbf{z}}_{6} \end{bmatrix} \qquad (8) $$

where, referring to Fig. 2, $\hat{\mathbf{z}}_{i-1} = \mathbf{u}_{i}$ (i= 1 to 7) which is the third column of ${}^{0}\mathbf{R}_{i}$, and ${}^{0}\mathbf{q}_{i}$ is, as defined before, the position of the origin of frame i on axis (i+1); all of which can be obtained directly from the results of the forward kinematics computation.

The Jacobian formulated here will be used in Eq. (6) to compute the necessary joint torques in order to 'reflect' a certain external load (force and moment). Typically, the point at which the load is to be reflected will not be the same as the reference point used by the Jacobian. Therefore, before Eq. (6) can be applied it is necessary to transform the load wrench to the same reference point, which is at the base frame origin. For discussion sake, let the load reflection point be at the wrist center of FREFLEX - typically this will be the case since the

external load will usually be measured using a force/torque sensor mounted at the wrist of the slave robot. Then the following transformations are required:

$$^{0}\mathbf{F}_{0} = {}^{0}\mathbf{R}_{s}\,{}^{s}\mathbf{F}_{w} \tag{9}$$

$$^{0}\mathbf{M}_{0} = {}^{0}\mathbf{R}_{s}\,{}^{s}\mathbf{M}_{w} - {}^{0}\mathbf{p}_{w} \times {}^{0}\mathbf{F}_{0} \tag{10}$$

where ${}^{s}\mathbf{F}_{w}$ and ${}^{s}\mathbf{M}_{w}$ denote the measured force and moment in the sensor frame (s), ${}^{0}\mathbf{R}_{s}$ is the coordinate transformation matrix from the sensor to the base frames, and ${}^{0}\mathbf{p}_{w}$ is the position of the load reflection point, namely wrist center, relative to base frame. One can easily generalize the above relationships to cases where the point of reflection is different, simply by using the corresponding position vector of the new load reflection point and, if necessary, the appropriate coordinate transformation matrix.

Gravity Compensation

To facilitate better utility and reduce operator fatigue, FREFLEX needs to support its own weight as it is being moved about providing position commands to control the slave manipulator. In general, the capability of gravity compensation is necessary for any master robot used in telerobotic systems. During each sampling period, this requires computing first the gravity loads at all joints and subsequently the torques needed from the motors to statically counteract those gravity loads. In this section we will only address the former, which is the computation of the gravity loads (torques) at the exoskeleton joints. The discussion of the latter, namely the decomposition of these computed joint torques into the actual motor torques, will be deferred to the following sections.

Consider the link connection at a revolute joint in a serial chain. The following observations can be made: As a result of the serial chain configuration, the gravity load seen by that joint will be due to sum of moments generated by weights of all the links outboard (farther from the base) of it. And the load will vary as relative positions of links change from one configuration to another; in other words, it is position dependent.

Let $\mathbf{h}^{i}_{G,j}$ represent position vector of the center of gravity (CG) of link j relative to joint axis i and the link mass be given as m_j. Then, with origin of frame (i-1) as the point of reference, we can write the moment at joint i due to weight of link j as:

$$\mathbf{n}_{i,j} = \mathbf{h}^{i}_{G,j} \times m_j \mathbf{g} \tag{11}$$

Thus the total moment due to all outboard links ($j \geq i$), is:

$$\mathbf{n}_i = \sum_{j=i}^{7} \mathbf{n}_{i,j} = \sum_{j=i}^{7} \left(\mathbf{h}^{i}_{G,j} \times m_j \mathbf{g} \right) \tag{12}$$

And the corresponding joint torque (τ_i) is just the component of total moment along the joint axis:

$$\tau_i = \mathbf{n}_i \cdot \hat{\mathbf{z}}_{i-1} \tag{13}$$

It can be easily seen that computing $\mathbf{h}^{i}_{G,j}$ is the key step in the gravity torque computation procedure. The way in which they are computed directly affects the efficiency of this algorithm. A logical consideration here is to take advantage of results available from the previous forward kinematics procedure. This suggests that the computations should be made in the base frame. We compute $\mathbf{h}^{i}_{G,j}$ using the following relationship:

$${}^{0}\mathbf{h}^{i}_{G,j} = \left({}^{0}\mathbf{q}_j - {}^{0}\mathbf{q}_{i-1} \right) + {}^{0}\mathbf{R}_j \, {}^{j}\mathbf{r}_{G,j} ; \qquad j = i, \ldots, 7 \tag{14}$$

where ${}^{j}\mathbf{r}_{G,j}$ is the given position of CG for link j with respect to its local frame j, and ${}^{0}\mathbf{q}_j$ and ${}^{0}\mathbf{R}_j$ are, as defined before, position and orientation matrices of frame j, both of which are readily available from previous results.

To pursue the issue of efficiency a little further, computing in the base frame coordinate has associated with it another added benefit. For the cross product calculation in Eq. (11) it is only necessary to compute two out of the three components. This is because, when expressed in the base frame, the gravity vector will typically have a nonzero component along only one coordinate axis. For example, in terms of the FREFLEX base frame, $\mathbf{g} = [-g, 0, 0]^T$, and thus the cross product which results only has nonzero y and z components. As a result, Eqs. (12) and (13) for FREFLEX can be evaluated using the simplified component forms:

$$\mathbf{n}_i = \begin{cases} n_{i,x} = 0 \\ n_{i,y} = -g \sum_{j=i}^{7} \left(h^{i}_{j,z} m_j \right) \\ n_{i,z} = g \sum_{j=i}^{7} \left(h^{i}_{j,y} m_j \right) \end{cases} \tag{15}$$

and

$$\tau_i = n_{i,y} u_{i,y} + n_{i,z} u_{i,z} \qquad (16)$$

Implementation of the above equations, Eqs. (14) to (16), has resulted in a highly efficient gravity torque computation algorithm for FREFLEX which only requires (126M + 133A + 0F) in its calculations.

Based on the above computed results, ideally one should be able to completely compensate the weight of all the links. This of course assumes a-priori that the data on the inertia properties (link masses and CG locations) of the system are correct, which unfortunately is typically not the case. A practical method to deal with this problem is to experimentally 'scale' the values of computed gravity torques. Such constant scaling factors can be found one joint at a time by adjusting from the last one inward until the link(s) supported by that joint in effect begin to 'float'.

Kinematic Coupling Relationship

As alluded to in the introduction, the cable and pulley actuation scheme gives rise to a complication of cross-coupling between motions at the joints on the FREFLEX and its actuators at the base. Figure 3 illustrates the cable system arrangement used by the FREFLEX [6]. In the present scheme, cables to drive a distal joint are routed via pulleys (idlers) through all inboard joints which are more proximal to the base. As a consequence, when the motor for the distal joint (the joint at which the cable terminates) is actuated, torque is transmitted not only to that distal joint, but to each of the inboard joints of the arm, which in turn causes motions at all those proximal joints. This is in contrast to the conventional actuation scheme in which the motion of a joint is dependent only on a single actuator driving that joint.

To control FREFLEX, it is necessary to identify the above coupling relationship in order to 'decouple' desired joint motions or torques into appropriate actuator commands. It should be noted that the coupling as described here is kinematic in nature in that it arises solely from kinematic constraints due to cable routing, and should not to be confused with the inertia coupling as seen in dynamics. Let $\Delta\theta_{motor}$ and $\Delta\theta_{joint}$ represent, respectively, the 7x1 vectors of displacements at the motors and its corresponding displacements at the FREFLEX joints, then they can be related by a (constant global) coupling matrix of 7x7, **A**, as follows:

$$\Delta\theta_{joint} = \mathbf{A} \cdot \Delta\theta_{motor} \qquad (17)$$

The above coupling matrix is constant throughout the FREFLEX workspace, since it is only a function of geometric attributes such as relative locations of the joints and pulleys, numbers of pulleys used, and radii of pulleys [7]. Ideally, such a matrix can be expected to be lower triangular.

In theory one can attempt a direct approach to derive the coupling matrix based on the aforementioned geometric attributes. However, practical considerations--such as errors induced by uncertainties as well as inaccuracies associated with geometric data, suggested that in fact it would be more reliable to identify the coupling relationship indirectly, using actual data of joint and motor displacements measured experimentally.

Referring to Eq. (17), it can be easily seen that if sufficient measurements of $\Delta\theta_{motor}$ and $\Delta\theta_{joint}$ are available (in this case the minimum number is 7), then identification of the coupling matrix **A** amounts to solving a system of linear equations with its elements as the independent variables. Furthermore, one can minimize the effect of measurement noises by using more measurement data than the minimum and computing the generalized inverse solution of **A**. In fact, such a solution, to be given below, represents physically the 'optimum' or 'best fit' set of elements for matrix **A** in a least squares sense.

Let the number of measurement be, say, n ($n \geq 7$). By using the n measured joint displacement vectors as columns, a 7xn joint measurement matrix, defined as **C**, can be formed. Similarly, let **B** represent the corresponding 7xn actuator measurement matrix comprising the correspondingly ordered actuator displacement vectors as columns. Then we can obtain the least-squares solution of **A** by:

$$\mathbf{A} = \mathbf{C}\,\mathbf{B}^{T}\left(\mathbf{B}\mathbf{B}^{T}\right)^{-1} \tag{18}$$

To acquire the necessary joint and actuator displacement data, the following process was carried out on the FREFLEX. The FREFLEX was moved to take on different configurations within its workspace. In the meantime, actual positions readings from potentiometers at the joints and resolvers at the actuators were recorded at a constant sampling interval. Having acquired the joint and motor position readings, the difference between each consecutive set of data was computed to give the corresponding displacement vectors which can then be used in Eq. (18) to solve for the coupling matrix. It should be noted that during these data collection movements, it is important that care should be taken not to run any joint up to its mechanical limits to avoid artificial bias of joint data.

The following gives the result of the coupling matrix we have identified for the FREFLEX from a total of 16 position measurements, namely, 15 sets of displacement vectors.

$$A = \begin{bmatrix} .9617 & .0548 & -.0169 & .0194 & .0065 & -.0024 & -.0255 \\ .9632 & .9711 & .0039 & .0001 & .0045 & -.0029 & -.0046 \\ .0246 & -.4294 & .9672 & .0257 & -.0231 & .0177 & -.0305 \\ .0292 & .1976 & .7246 & -.7250 & -.0278 & .0310 & -.0408 \\ .0075 & .4336 & .5890 & .3732 & -1.1949 & -.0052 & -.0032 \\ -.0052 & .3139 & -.2185 & .2192 & 1.2706 & -1.5005 & -.0062 \\ .0539 & .4949 & -.3120 & .3114 & .0092 & 1.4494 & -1.4256 \end{bmatrix}$$

As can be seen from the result obtained here, the coupling matrix is in general agreement with that expected from theory that elements in its upper half of the matrix above the diagonal, while not ideally all zeros, all have relatively small magnitudes compared to those in the lower half. Furthermore, it can be seen that strong cross coupling exists between joints 5, 6 and 7 (as can be evidenced from unusually large values of (6,5) and (7,6) elements), a characteristics which has been regularly observed from actual behaviors of the FREFLEX.

Joint Torque Decomposition

The same kinematic coupling matrix developed in the previous section can also be used to decouple the torques required at the FREFLEX joints into those needed to be applied at the motors. By principle of virtual work, a corresponding static coupling relationship can be arrived at in a straightforward way. This is illustrated as follows. Based on the input-output relation of (virtual) work, we must have

$$\tau_{motor} \cdot \Delta\theta_{motor} = \tau_{joint} \cdot \Delta\theta_{joint} \tag{19}$$

And through the use of Eq. (17), one can easily arrive at:

$$\tau_{motor} = A^T \tau_{joint} \tag{20}$$

With Eq. (20), it is now possible to compute the torque commands which are necessary to be applied at the motors so as to (1) counteract the gravity loads due to its own weight, and (2) to generate the external load (force and/or moment) to be reflected to the operator. Designating such motor torques in vector form as $\tau_{motor,total}$, we can combine Eqs. (6), (16), and (20) to yield:

$$\tau_{motor,total} = \mathbf{A}^T \left(-\tau_{gravity} + \mathbf{J}^T \begin{bmatrix} \mathbf{M}_{ext} \\ \mathbf{F}_{ext} \end{bmatrix} \right)$$

For clarity the subscript 'gravity' has been added to the first term in the parentheses to delineate the joint torque vector due only to gravity, which is available from Eq. (16) directly. Notice that this result is then negated to give the joint torques needed to equilibrate the gravity loads before its decomposition into corresponding motor commands. Similarly, the subscript 'ext' in the second term represents resultant load of force and moment (as given by Eqs. (9) and (10)) to be generated by the motors, in addition to the gravity loads.

III. Position Coordination for the Slave Manipulator

A MERLIN 6500 robot (by American Robotics Corp.) is used as the slave manipulator which is controlled by the operator through the FREFLEX to interact with the environment. The robot is a six degrees of freedom, revolute-jointed industrial robot with stepper motor drives and a geometry similar to that of the commonly known PUMA manipulator. Typically the MERLIN can be programmed to operate using its own specific programming system, AR-SMART [8], provided by the manufacturer. In order to achieve necessary speed of response, the programming system is bypassed and a direct communication link was implemented between the FREFLEX and the MERLIN controllers through a high speed interface protocol [8].

During system operations, Cartesian position and orientation ('pose') data of the FREFLEX master is sent to the MERLIN controller which in turn must drive the slave robot to the same input position and orientation (which, of course, is now relative to the slave base frame). Obviously, for the MERLIN controller, it is necessary to compute the corresponding joint angles for a given pose command. Again, since these joint angles must be updated at every sampling time as new commands to joint servos, the computations must be made as efficient as possible.

In the following, we present the results from the inverse kinematics analysis developed for the MERLIN to facilitate efficient computations. For reason of space limitation, details of analysis in arriving at the solutions will not be reported here (see [9]). We note that the approach taken follows the one alluded to in earlier discussion, in which the kinematic equation is separated into translational and rotational parts. For MERLIN, based on the

frames and D-H parameters defined in [9], we have obtained in explicit closed form the resulting set of joint angle solutions. The complete solutions are given as follows.

Let the desired position transformation be given by:

$$ {}^{0}\mathbf{T}_6 = \begin{bmatrix} \mathbf{Q} & \mathbf{r} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & r_1 \\ q_{21} & q_{22} & q_{23} & r_2 \\ q_{31} & q_{32} & q_{33} & r_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

And define:

$$ \mathbf{h} = \begin{bmatrix} r_1 - q_{13}d_6 \\ r_2 - q_{23}d_6 \\ r_3 - q_{33}d_6 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} $$

Then

$$ \theta_1 = \operatorname{atan2}(h_2, h_1) - \operatorname{atan2}\left(\frac{d_2}{\eta}, \sigma_1 \sqrt{1 - \left(\frac{d_2}{\eta} \right)^2} \right) $$

where $\eta = \sqrt{h_1^2 + h_2^2}$; $\sigma_1 = \pm 1$

$$ \theta_3 = \operatorname{atan2}(s_3, c_3) \qquad \text{where } f = h_1 \cos\theta_1 + h_2 \sin\theta_1 ; \quad \sigma_2 = \pm 1 $$

$$ s_3 = \frac{f^2 + h_3^2 - a_2^2 - d_4^2}{2a_2 d_4} \quad ; \quad c_3 = \sigma_2 \sqrt{1 - s_3^2} $$

$$ \theta_2 = \operatorname{atan2}(s_2, c_2) \qquad \text{where} $$

$$ s_2 = \frac{f d_4 c_3 - h_3(a_2 + d_4 s_3)}{h_3^2 + f^2} \quad ; \quad c_2 = \frac{h_3 d_4 c_3 - f(a_2 + d_4 s_3)}{h_3^2 + f^2} $$

Next form:

$$ \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_1 c_{23} & s_1 c_{23} & -s_{23} \\ -s_1 & c_1 & 0 \\ c_1 s_{23} & s_1 s_{23} & c_{23} \end{bmatrix} \mathbf{Q} $$

Then:

$$ \theta_4 = \operatorname{atan2}\left(\frac{\sigma_3 b_{23}}{\sqrt{b_{13}^2 + b_{23}^2}}, \frac{\sigma_3 b_{13}}{\sqrt{b_{13}^2 + b_{23}^2}} \right) \quad ; \quad \sigma_3 = \pm 1 $$

$$\theta_5 = \text{atan2}\left(-b_{13}c_4 - b_{23}s_4 \,,\, b_{33}\right)$$

$$\theta_6 = \text{atan2}\left(\left(b_{21}c_4 - b_{11}s_4\right),\left(b_{22}c_4 - b_{12}s_4\right)\right)$$

It is important to note that the computations should be carried out in the same order as presented above.

It might be helpful to make a few general comments concerning the practical implementation of the inverse kinematics solution. Note that, although in theory multiple solutions are possible to achieve a given end effector position (e.g., a total of eight solutions for the MERLIN geometry), restrictions from physical and/or operational constraints will often render fewer feasible solutions in reality. As a case in point, the MERLIN used here in this study was physically constrained to maintain a 'left shoulder' configuration, stipulating σ_1 being always positive. If, in addition, an 'elbow-up' constraint (σ_2 negative) is also imposed, the number of feasible solutions then reduces to two, given by the remaining two values of σ_3 corresponding to different configurations of the wrist assembly. Which one out of the remaining two to pick can depend on operational consideration such as to select the one closest to the current positions, namely the one which results in the minimum joint motions.

Another seemingly trivial but nevertheless useful note concerns the way in which the joint angle command is expressed. As a rule of good practice, one should express the angles returned from trigonometric operation (e.g., atan2) in a sign consistent manner, i.e., between +/- 180 degrees, particularly when they are to be used as joint commands. This arises from the fact that most joint servos are directional sensitive to the signs of angles; angular joint commands of, say, 150 degrees and -330 degrees are likely to cause the motor to turn in opposite directions, although they both will reach the same angular position eventually.

IV. Discussions and Conclusion

We have presented the development of coordination algorithms for control implementation of a telemanipulation system consisting of an exoskeleton master and a articulated slave manipulator. Physical modeling to analyze mechanics of each subsystem was established and algorithms for its coordination was formulated subsequently. Throughout the

development. special emphasis was placed on computation efficiency due to its role on system performance.

The algorithms presented in this report have been fully tested and implemented on the robot subsystems and, as a result, significant improvements on the overall system performances have been achieved. In an initial implementation on FREFLEX using the same processor (68020), over five-fold speed increase on computation cycle time (from 90 down to 16 msec) was achieved using the present algorithm as compared to an earlier one. One can further improve the sampling rate by (1) using a faster processor, and (2) updating the gravity torque commands (the most computationally intensive part) less frequently. The latter is justified in light of slowly changing configurations of the FREFLEX when being controlled by the operator. Both of the above steps have been done and, as of this writing, a FREFLEX motor torque update rate of 287 Hz has been attained using a newly replaced 68030 processor and updating gravity torque commands once every 10 cycles. The MERLIN inverse kinematics solution is computed on a 68030 processor in less than the minimum 4 ms MERLIN controller update period

Acknowledgement

The support of this work as part of Summer Faculty Research Program sponsored by AFOSR is gratefully acknowledged. The author also wishes to express his appreciation for help received from the AL/CFBA at WPAFB. Special thanks go to members of the HSF group: Capt. P. Whalen, Capt. D. Nelson, Lt. C. Hasser, Mr. M. Crabill and Mr. T. Mosher.

References

1. T. B. Sheridan, Telerobotics, Automation and Human Supervisory Control, MIT press, 1992.
2. A.K. Bejczy and Z. Szakaly, "Universal Computer Control System for Space Telerobotics," Proc. of IEEE Conference on Robotics and Automation, Vol. 1, Raleigh, NC, 1987.
3. G. Hirzinger, J. Dietirch, and B. Brunner,"Multisensory Telerobotic Concepts for Space and Underwater Applications," Proc. of the Space and the Sea Colloquium, Paris, 1990.
4. R. P. Paul, Robot Manipulators, MIT Press, 1981.
5. K. S. Fu, R.C. Gonzalez, and C.S.G. Lee,Robotics: Control, Sensing, Vision, and Intelligence, McGraw Hill, 1987.
6. Odetics, Inc., "Exoskeleton Master Arm, Wrist, and End Effector Controller with Force-Reflecting Telepresence," U.S. DoD, SBIR Program, Phase II Final Report, December 1992.

7. J.-J. Lee and L.-W. Tsai, "The Structure Synthesis of Tendon-Driven Manipulators Having a Pseudotriangular Structure Matrix," The International Journal of Robotics Research, Vol. 10, No. 3, June 1991.
8. Merlin™ System Operator Guide, Version 3.0, American Robot Corp., June 1985.
9. M. Z. Huang, "Kinematic Solutions for Position and Velocity Coordination of MERLIN," Technical Report, AL/CFB, Air Force (in preparation).

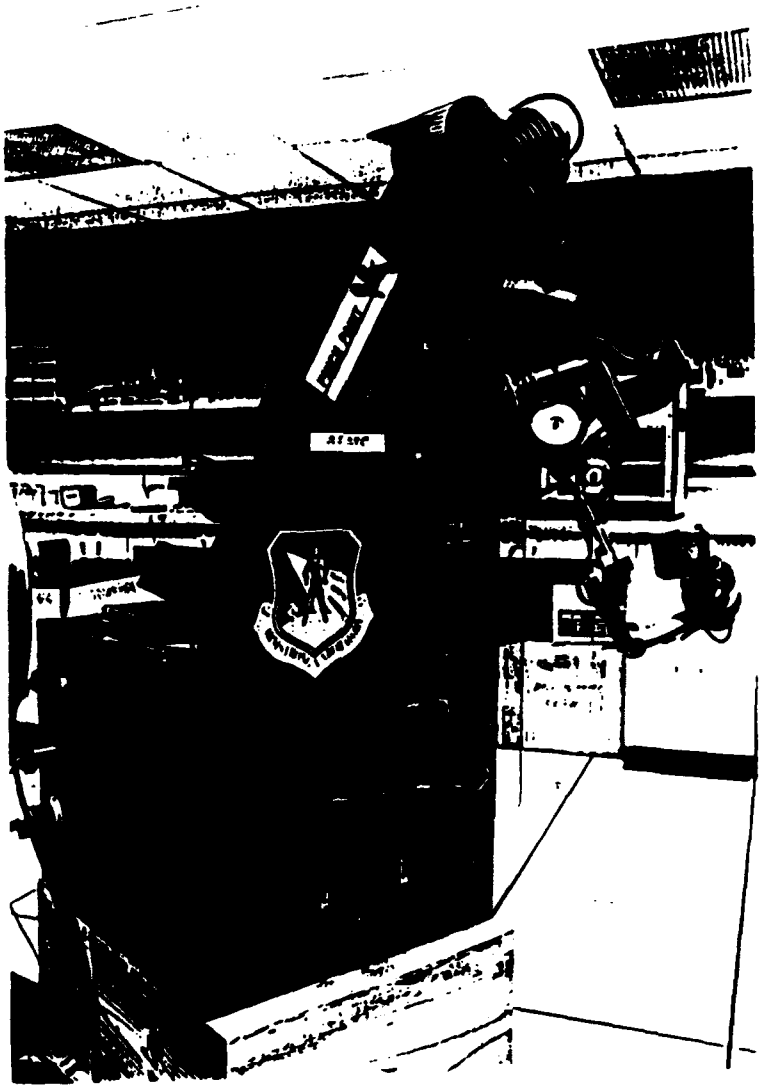


Fig. 1: The Force REFLecting EXoskeleton (FREFLEX) master robot

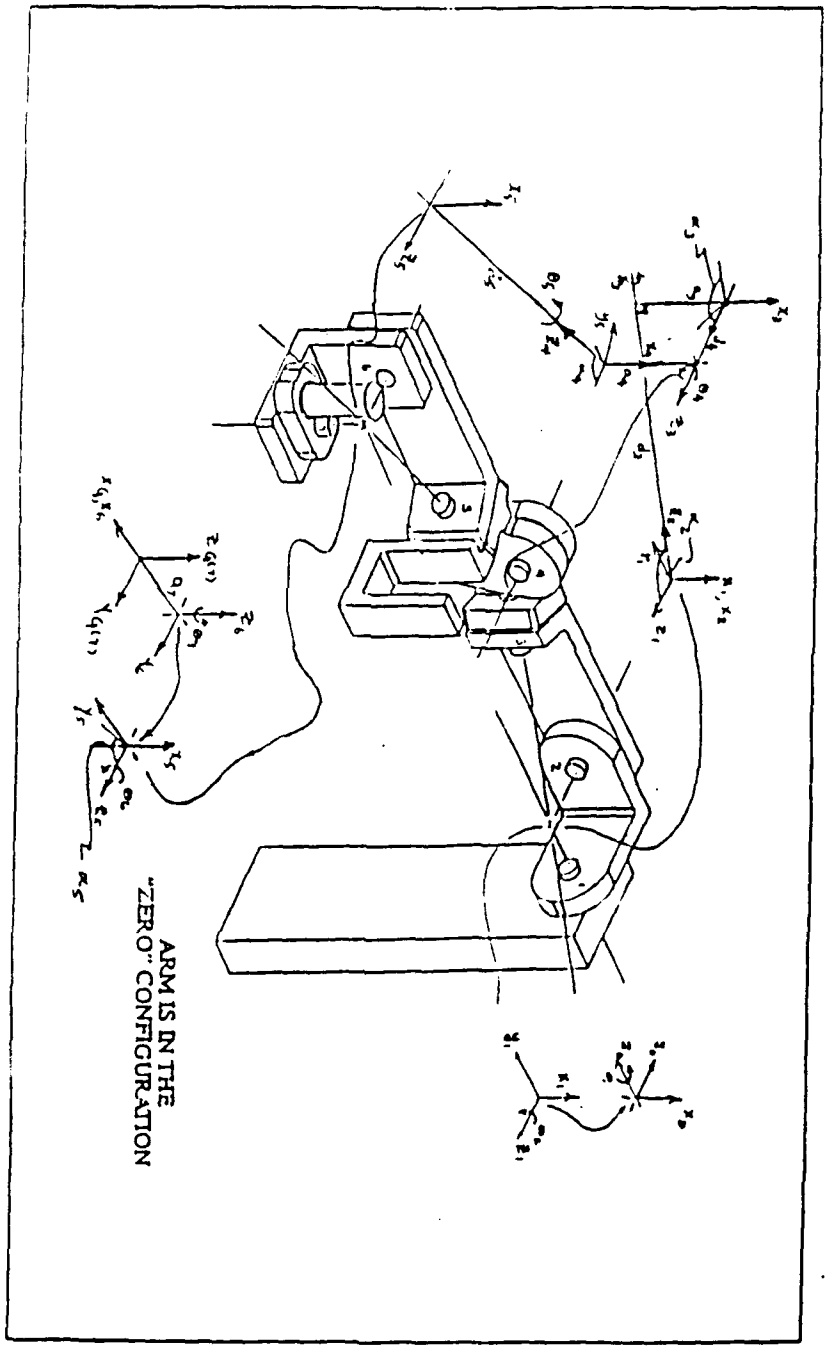


Fig. 2: Schematic of FREFLEX and definitions of kinematic frames

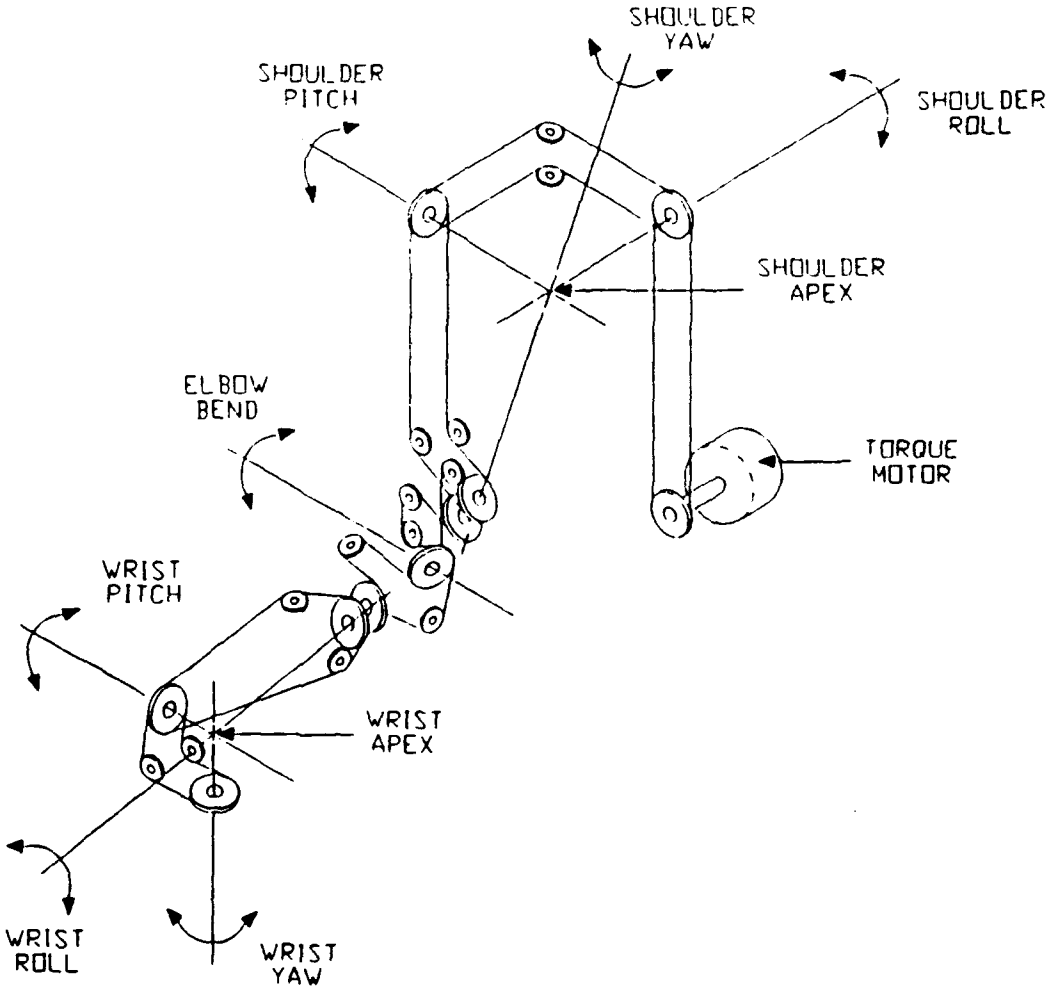


Fig. 3: Cable drive system of the FREFLEX

An Exploratory Study of Weighted Fuzzy Keyword Retrieval for the CASHE:PVS System

Donald H. Kraft
Professor
Department of Computer Science

Louisiana State University
Baton Rouge, LA 70803-4020

Final Report for:
Summer Faculty Research Program

Armstrong Laboratory

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

August, 1993

An Exploratory Study of Weighted Fuzzy Keyword Retrieval for the CASHE:PVS System

Donald H. Kraft
Professor
Department of Computer Science
Louisiana State University

Abstract

The purpose of this project was to explore the possibilities of providing improved retrieval capabilities to the CASHE:PVS system. That system currently provides access to human engineering studies and allows users to navigate from one entry to another and to simulate ergonomic experiments in order to provide understanding and improved design. The use of keywords, weights, and fuzzy Boolean logic have been explored in order to determine the feasibility of this approach, based in large part on the series of sample queries constructed for CASHE:PVS and known as the Design Checklist. Future steps necessary to continue to demonstrate the feasibility of this approach and to integrate this approach into the CASHE:PVS system are presented.

An Exploratory Study of Weighted Fuzzy Keyword Retrieval for the CASHE:PVS System

Donald H. Kraft

Introduction

The Computer Aided Systems Human Engineering: Performance Visualization System (CASHE:PVS) has been developed and is being prepared to be marketed for use. CASHE:PVS consists of a "multi-media ergonomics database containing the complete" *Engineering Data Compendium* (EDC) "and MIL-STD-1472D, plus a unique visualization tool," the Perception and Performance Prototyper (P^3). Through cooperation with NASA and NATO, plus Tri-Services, the EDC was produced to define approaches to communicate human factors data to designers (levels of technical content, presentation format and style, and terminology) and to provide access to specific technical information relevant to design problems. The MIL-STD-1472D is the military standard for human engineering design criteria, including systems, equipment, and facilities. This standard includes over eighty figures and tables, and is used to insure human-systems integration as well as efficiency, reliability, safety, trainability, and maintainability in acquired systems [Boff91]. CASHE:PVS's goal is to enable "ergonomics to be supported as a 'full partner' among other design disciplines within a computer-aided environment. By fully integrating ergonomics into the systems design process, more effective human-system designs can be visualized."

Just over a decade ago, the Integrated Perceptual Information for Designers (IPID) project was underway to begin to aid the accessibility and use of ergonomics data in design. This included the identification, collection, and consolidation of human performance data; the representation and presentation of such data to designers (human engineering); the training of designers in the use of ergonomics data; and the definition and evaluation of integrated media options to allow designers to access, interpret and apply such data. Out of these efforts has come the CASHE:PVS system, which will soon be marketed as a commercial product.

The heart of the CASHE:PVS system, which must reside on an Macintosh II computer, is the Bookshelf, which consists of several databases, including the previously mentioned EDC and the MIL-STD-1427D. The Bookshelf also contains User Project files, where individual users can customize the system, adding annotations (perhaps to the research articles contained in the EDC entries), storing files, and otherwise augmenting the reference databases. The Bookshelf also contains the P^3, the test bench simulator mentioned above. A user interested in what types of alarms would be best in a noisy environment, such as the cockpit of an airplane, can look up the appropriate article(s) in the EDC and peruse it (them). However,

looking at charts and tables of data generated by human factors experts may not provide the full understanding of what the numbers mean for a design engineer. Thus the P^3 allows the user to specify a given level of noise and type of alarm and then hear for him/herself what the numbers "sound like." Moreover, users can construct additional simulated tests to go beyond the entries for better design [Cona93].

A sample of an EDC entry [Boff88a] is in Appendix I of the set of Appendices for this Final Report. The appendices are not attached to this Final Report, but they are available from the contact person for this project, whose name and address are given below:

Donald L. Monk, Program Manager
Armstrong Laboratory
Crew Systems Directorate
Human Engineering Division (AL/CFHD)
Wright Patterson AFB, OH 45433-6573
(513) 255-8814
dmonk@falcon.aamrl.wpafb.af.mil

There are up to ten sections of an EDC entry. These include a *title* with a concise description of the entry content (including a number based on a topic outline); *key terms* to verify entry content and to provide access points to the entry via index search; a *general description* to summarize the entry content (findings, results, conclusions, models, laws, or principles); an *applications* section to describe areas of application for the entry (e.g., types of displays); a *methods* section to describe how data was collected; an *experimental results* section that may contain graphics and tables with details on the results of data analysis; an *experimental validation* section to show how a model, law, or principle was verified; a section on *constraints* to show limitations in the application of the entry results (e.g., criteria that must be met for proper application); a set of *key references* with bibliographical citations to original literature with more detailed information on the topics of the entry; and *cross-references* to other entries on the topics in the entry [Boff91].

The user can use the TextViewer component of CASHE:PVS to view the text entry from the EDC, once an entry has been selected. Users can browse through any of the fields mentioned above. Users can also navigate to other EDC entries. In addition, FigureViewer and TableViewer components of CASHE:PVS allow the user to view the entry graphs and tables, respectively [Boff91]. In addition, as mentioned above, CASHE:PVS users can simulate situations such as noise or vibration to sense directly meaningful situations of the data in action. However, one component needing additional capability is the provision of initial navigation aids so that a user can begin to use the system properly.

The Weighted (Fuzzy) Retrieval Approach

In effect, CASHE:PVS is a multi-media system that allows a user to search through the EDC and MIL-STD for ergonomic information that will be helpful in design. Users can navigate from one entry to another via a number of paths, such as an index, a set of keywords and cross-references to other entries.

There are also links to the appropriate portions of the military standard (MIL-STD-1472D). Appendix II contains the table of contents for the military standard, plus a sample entry from the standard. Appendix III contains a portion of the back-of-the-book index for the EDC. This particular project is intended to explore means of employing fuzzy set theory and other information retrieval mechanisms to better to allow a user to specify one or more EDC entries of interest. The basis of this research project is the employment of keywords and weights for those keywords in order to provide an appropriate match between a user request and the EDC collection. Moreover, the focal point was the use of a Design Checklist, a series of queries hierarchically put together as sample user queries.

Let us first generate a model of information retrieval in order to better understand the background from which this summer project has been developed. Information retrieval, unlike standard database systems, is in general concerned with the imprecise nature of determining which textual records are relevant to users in response to queries. Previous research has involved document representation (e.g., indexing) to determine which terms to use for topicality. Of course, it is well known that relevance itself is an imprecise concept, incorporating many factors other than topicality. Weights specifying a degree of "aboutness", with a fuzzy set interpretation, can be incorporated to induce a document ranking mechanism. Weights on terms in the user query can also be added to specify a degree of relative term importance. Let us consider a retrieval system as a set of records that are identified, acquired, indexed, and stored, plus a set of user queries for information that are matched to the index to determine which subset of the stored records should be retrieved and presented to the user.

We can begin to model the retrieval system by the following [Kraft85, Kraft83, Kraft93]. Consider D to be a set of documents, or textual records, from which we wish to retrieve subsets of documents in response to users. In order to match documents to user requests, we need to describe the contents of the documents in some manner. A most common mechanism for doing this, especially in terms of doing this automatically by computer, is through the use of keywords [Salton89]. Thus, consider T to be a set of index terms, i.e., keywords. It may be the case that some of the members of T are phrases (e.g., "information retrieval"), but we will consider only single-word keywords.

The assignment of keywords to documents is a process often known as indexing. It is a difficult, complex task, but can be represented algebraically as:

F: the indexing function, where F: D x T ---> {0,1}.

This implies that one takes a given document d and a given term t, indexes d with regard to t through the function F that produces a 1 (t is used to describe d) or a 0 (term t is not used to describe d). Thus, for the EDC entry in Appendix I, whose title is "Measurement of Radiant and Luminous Energy," we might use

the term "illumination" (an F of 1) but not use the term "auditory" (an F of 0).

However, the notion of weights, i.e., fuzziness, has long been considered not only feasible but desirable [Kraft85]. Thus, we can let the range of F be the continuous unit interval [0,1], so that partial indexing is allowed. Thus, we might decide that the document in question is about the concept "irradiance" 0.85, while it is about "flux" 0.45. That is to say, the indexing function F maps a given document d and a given indexing term t to a number between 0 and 1 (0 implies that the document is not at all about the concept(s) represented by term t and 1 implies that the document is perfectly represented by the concept(s). Thus, we have a fuzzy set with F being the membership function, mapping the degree to which document d belongs in the set of documents "about" the concept(s) represented by term t.

There are several means of estimating F for a given situation. Salton [Salton89] suggests a measure, the inverted document frequency (IDF), defined as

$$f_{i,t} \times \log[N / N_t],$$

where $f_{i,t}$ is the number of times term t occurs in document i, N is the number of documents in the collection (database), and N_t is the number of documents in the collection in which term t occurs at least once. We note that Salton's measure gives larger weights to terms that occur frequently in the document in question but do not occur at all in most documents, and furthermore that Salton's IDF needs to be normalized to be contained in the interval [0,1].

Most retrieval systems employ a stop list of common words (e.g., "a", "an", "the", and "moreover") to be ignored, since these words are too common and convey little if any topicality information about documents that contain them. We also note that if one restricts F to the set {0,1}, one has the case of classical indexing, where a specific term is either attached to the document (1) or it is not (0). By allowing F to also take on values in the open interval (0,1), one can weight terms according to their importance or significance in describing the content of a document in order to better retrieve it for users who want it. In this latter case, F has been called an index term weight.

Now, let us define Q as the set of user queries for documents. This leads to the the need to represent the query with a set of keywords. This is stated algebraically as

$$a: Q \times T \dashrightarrow \{0,1\} = a(q,t) = \text{the importance of term t in describing the query q}$$

We note that the function a assigns a term t to the query (a = 1) or it leaves out the term t (a = 0). However, again we may wish to generalize this by incorporating weights on the terms to indicate relative importance; thus we again let the range become the unit interval [0,1]. Clearly, not all terms are equally important in a given query. In determining which houses to consider from listings in the real estate section of a newspaper, terms describing whether there is a fireplace or not may not be as important as terms describing the

cost of the house.

It is here that one begins to introduce problems in terms of maintaining the Boolean lattice [Kraft83]. Because of this, certain mathematical properties can be imposed on F, but more directly on a and on the matching procedure [Kraft85]. To describe the matching procedure, consider the function g, defined as

g: [0,1] x [0,1] ---> [0,1],

We have g(F,a) as the retrieval status value (RSV) for a query q of one term (term t) with query weight a in terms of document d, which has index term weight F(d,t) for the same term t. That is to say, g is the evaluation of document d in terms of its estimated relevance with respect to this one-term query q. Kraft, et al. [Kraft93] consider a variety of forms for g, striving to allow the function to reflect the semantics of the query. The function g can be interpreted as the evaluation of the document in question along the dimension of the term t if the actual query has more than one term. While many researchers have considered models that do not use Boolean logic for the queries [Salton89], virtually all commercial systems use Boolean logic for multi-term queries. For this case, let

e: [0,1]* ---> [0,1]

be the RSV for a Boolean query of many terms, where each term in the query is evaluated as a single-term query against the document and then the total is evaluated using fuzzy Boolean logic (e.g., AND uses a max function, OR uses a min function, and NOT uses a one-minus function). This notion of allowing e to be a function of the various g values is based on the criterion of separability [Kraft83].

Evaluation of a retrieval system is often based on two factors, recall and precision. Recall is the percentage of the relevant documents that were retrieved, while precision is the percentage of the retrieved documents that were relevant. Of course, other factors, such as cost, can enter into the evaluation of a retrieval system. Moreover, it is well known that topicality is but one key facet of relevance; users are influenced by other factors such as language, timeliness, and appearance.

Thus, the notion of using keywords to determine a small subset of the EDC entries is the object of this summer project. It makes a good deal of sense to allow users to form a query and get a small handful of entries to explore. The hypertext navigational aids of CASHE:PVS will then allow the user to go from there and find the proper design answers needed.

Methodology - Obtaining and Formatting the Data

The EDC entries are organized along a hierarchical outline as presented in the EDC Table of Contents. The numbering system as part of the title field reflects that hierarchical ordering. For example, the entry in Appendix I is numbered 1.104; the category 1.0 reflects "visual acquisition of information", and the subcategory 1.1 reflects "measurement of light"; hence the entry on "measurement of radiant and luminous

energy" falls in this category and subcategory. The entire outline is presented in Appendix IV.

In order to make CASHE:PVS more useful to designers, a series of queries were generated. These sample queries are there to serve as examples of the kinds of questions designers might ask. Each query consists of one or more questions, such as "For extended light sources to be treated, with small error, as point sources, how much greater than their diameter must their distance be from illuminated surfaces?". Moreover, each query has a list of EDC entry numbers referring to those EDC entries that could answer the question(s) in the query (e.g., 1.104). There are 1,069 queries in the set, which is known as the Design Checklist. The Design Checklist is ordered hierarchically by subject; however, this hierarchy is independent of the EDC hierarchy. The hierarchical arrangement, with the query topic headers, is given in Appendix V.

It was decided early on that the best approach, given the time constraint for this project among other limitations, was to generate a keyword mechanism for access to the Design Checklist queries and then to link the retrieved queries to the EDC entries. In order to understand the reasons for the approaches taken with this project, we now need to consider the environment in which we were working.

The Design Effectiveness Technology Laboratory (DEFTech Lab) in the Paul M. Fitts Human Engineering Building is primarily an Apple Macintosh shop. Thus, it is not surprising that CASHE:PVS has been built to run on a Macintosh platform. Due to scarcity and personal preference, we were given access to a Zenith microcomputer running 3.3 DOS. It was connected, via the software system known as Kermit, to the VAX computer (from Digital Equipment Corporation or DEC) running VMS (rather than the preferred Unix operating system). The name given to the Vax was Falcon. In addition, we were able to connect remotely from Falcon to one of the Louisiana State University (LSU) Department of Computer Science Vaxstations running Ultrix, a version of Unix. The Zenith was used with Turbo Pascal 4.0 from Borland to generate and test several Pascal programs to accomplish most of the tasks in demonstrating the feasibility of keyword retrieval for CASHE:PVS. However, the limitations on size and occasionally on time caused some difficulties. The LSU Vaxstation, known at LSU by the name "bit" so that formal access was through the electronic address of "bit.csc.lsu.edu" under the login id of "kraft", was used to do some sorting, and some retrieval analysis as will be discussed below. The Falcon Vax was used for intermediary storage, word processing, and as an intermediary between the Zenith and the LSU Vaxstation. All files stored on the Falcon computer are in the directory guest$disk:[dkraft.cashe].

The first step was to get a copy of the Design Checklist, originally on one of the Apple Macintoshes, for use in our retrieval experiments. After some difficulty, it was possible to transfer a copy to Falcon and to store it in the file design.old. This file had portions of the topic headers given in the hierarchical

classification of the queries, as well as the queries themselves. However, there were problems in that some of the queries were over 255 characters in length, but the standard editor truncated those to the first 255 characters, thus losing data. Moreover, the standard editor did not provide a wrap-around feature so that only the first eighty or so characters could be viewed. Thus, we had to edit the file, using the hardcopy version of the Design Checklist to provide the lost data, so that the new file would have no line of more than eighty characters. Care was taken to break the lines at reasonable points so that further processing could take place such as recognition of queries versus topic headers. The new version of the Design Checklist file that we used in subsequent processing is stored on Falcon in design.txt. A page providing a listing of the first few entries in design.txt is in Appendix VI.

Because of the size of the design.txt file, it was decided to not transfer it to the Zenith. Thus, since Falcon did not have a Pascal compiler, a copy of design.txt was electronically sent to the LSU Vaxstation to be processed there. A Pascal program, a copy of which is stored on Falcon in header.pas, produced a file of the topic headers from design.txt. The copy of this header file is stored on Falcon in header.txt; moreover, a page providing a listing of the first few entries in header.txt is in Appendix VII.

In addition, two other Pascal programs were run on the LSU Vaxstation to generate the queries from the Design Checklist stored in design.txt. Copies of these two programs are stored on the Falcon in quest.pas and quest1.pas. These two programs both produce the set of queries. However, quest.pas generates the queries without query numbers and stores the result in a file (a copy of which is on Falcon under the name quest.txt). A listing of the first few entries in quest.txt can be seen in Appendix VIII. Moreover, quest1.pas was created and used because it produces the more useful file of queries with query numbers which can be used as primary keys for the query file. A copy of this second file is found on Falcon under the name quest1.txt, and a listing of the first few entries in this file can be found in Appendix IX. To illustrate, consider the quest1.txt entry:

> Q123 For extended light sources to be treated, with small error, as point sources, how much greater than their diameter must their distance be from illuminated surfaces? (1.0104)

We note that the query number, Q123, is present, as well as the question (note that this query has but one question). In addition, the location of the EDC entry where the answer can be found is given (1.0104). Further, we note that this query indicates only one EDC entry is to be sought, while some queries have more than one. In addition, the format of the EDC entry number has been modified in the design.old file, and hence the query1.txt file, to have a four digit number (rather than three digits) to the right of the decimal point.

Because the Design Checklist file had only partial headers, indicating only that a new subsection portion of the full header, we had to expand the header.txt file to include the full header for each header. For

example, the first header is:

| 'I. Displays | A. Visual displays | 1. Image quality | a. Resolution |
1) Imaging properties of the eye | General'

while the second is but

| 'Visual optics'

Thus, we had to create the file of full headers, stored on Falcon in the file header1.txt. The first full header is now

| 'I. Displays | A. Visual displays | 1. Image quality | a. Resolution |
1) Imaging properties of the eye | General'

and the second full header is now

| 'I. Displays | A. Visual displays | 1. Image quality | a. Resolution |
1) Imaging properties of the eye | 'Visual optics'

A listing of the entire header1.txt file is in Appendix X.

Next, the full headers had to be edited to remove unneeded punctuation, such as single reverse quotes (`), and the outline indicators (e.g., I. or A. or a. or 1)) In addition, the separator symbol (|) was deleted from all headers except at the very beginning in order to establish that this was the beginning of each new header. A Turbo Pascal program, stored in header12.pas on Falcon, was run on the Zenith, to edit the headers. A copy of the resulting file is stored on Falcon in the file header2.txt, and a listing of the first page of this file is in Appendix XI. To illustrate, the edited full header for the first entry in this file is:

| Displays Visual displays Image quality Resolution
Imaging properties of the eye

Methodology - The Inverted File

Now, the next task was to generate an inverted file or index of the keywords in the headers and in the queries under those headers. This meant that a merged file of edited full headers and queries would be needed. In order to accomplish this, we had to generate a file in which the last query number under each edited full header is presented. This particular file is stored as limits.txt on Falcon and is shown in Appendix XII. For example, the last query listed under the first header is Q8, the eighth query in the file.

One additional problem cropped up in that the query numbering scheme had a few inconsistencies. Queries Q468, Q974, and Q1012 were "See Also" references. It was decided to add the keywords from those references to the header under which they occurred. Query Q704 occurred twice, once as a "See Also" reference and once as a real query. The terms in the "See Also" reference would have to be added to the appropriate header, and the real query was treated like all other queries. Queries Q433, Q1064, and Q1065 were listed under the wrong headers in the database when compared to the hardcopy version of the

Design Checklist, and would have to be moved. Finally, there were two queries with the number Q1012, each under a different header, and that problem had to be resolved. Thus, limits.txt and header2.txt had to manually edited to account for some of these inconsistencies.

A Turbo Pascal program was created, now stored as header1.pas on Falcon, and run on the Zenith that inputed the headers in header2.txt, the queries with query numbers in quest1.txt, and the limits in limits.txt. This program then produced a list of queries followed by the edited full header under which that particular query was found in the original Design Checklist file. This new list was stored in hq.dat on Falcon, and a sample of the first few entries is found in Appendix XIII.

We then transferred a copy of hq.dat to the LSU Vaxstation, where a canned program was employed to produce a list of question numbers, each followed by a set of keywords found in the appropriate header and/or query. This canned program was written in C under Unix for the chapter on stop lists in [Frakes92; Fox, C., "Lexical Analysis and Stop List,", Chapter 7]. For a copy of this program and of other retrieval programs, one should contact:

Professor William Frakes
Department of Computer Science
Virginia Polytechnic Institute and State University
Northern Virginia Graduate Center
2990 Telestar Court
Falls Church, VA 22042
703-698-4712 voice
703-698-6062 fax
frakes@sarvis.cs.vt.edu

The authors of [Frakes92] provide a caveat in that their codes are intended to illustrate the concepts discussed in their book. Although the codes have been thoroughly tested, they are meant to be instructional material only, and are not guaranteed to perform as intended by the authors. Thus, no warranty of the software is stated or implied by the authors, editors, testers, publishers, or suppliers of the code. We used it solely to experiment with and to demonstrate the viability of keyword retrieval for the CASHE:PVS system. The list is stored on Falcon in the file hq.sto, and a sample of the first few entries are presented in Appendix XIV. As an example, the first entry is:

```
q1
human      eye          displays          visual      displays
imagequality            resolution        imaging           properties   eye
```

In order to better group these entries, a Turbo Pascal program was created and run on the Zenith. This program, a copy of which can be found on Falcon in header2.pas, formatted the hq.sto file in order to produce a list of terms followed by the query number. The result is a file, stored on Falcon as hq2.sto. A sample of the first few entries is found in Appendix XV. For an illustration, the first few entries are

```
human q1
eye q1
```

displays q1

We then sorted the hq2.sto file in ascending, alphabetic order on the basis of the keywords, eliminating duplicate words with the same query numbers. This result is stored in the file hq2.srt on Falcon, and a sample of the first few entries is presented in Appendix XVI.

A Pascal program, a copy of which is stored in inv.pas on Falcon, was created and run on the LSU vaxstation. This program incorporated linked lists, and it was the size of these lists that necessitated the use of the LSU Vaxstation, rather than the Zenith, in order to generate the inverted file. This inverted file, or index, is stored in inv.txt on Falcon. It is sorted on the keywords, and each entry consists of a keyword followed by a list of query numbers in order. Moreover, asterisks indicate how many times each keyword (or term) is repeated in the query and/or the header under which the query occurs. This method allows one to see how often each term is used as a measure of the importance of the term in indicating a relevant query. A sample of the first few entries is given in Appendix XVII. As an illustration, consider the entry:

acceleration q376 q405 q418 q695 q703 q705 **q706 *q707 *q708 *q709 *q710
**q711 *q712 *q713 q714 q715 q915 q1045 q1050

Methodology - Querying the Query Index

Now, with the data acquired and formatted and the inverted file created and stored, we were ready to generate a mechanism for users to enter requests for the appropriate Design Checklist query or queries. A Turbo Pascal program was then created, which is stored on Falcon as query.pas. This program asks the user for a single keyword (or term). The user can then decide whether or not he/she wishes the search to use the term as it is or to stem it when making comparisons in the index. The stemming procedure was based on the work of C. Paice, as encoded in a Pascal routine along with a set of rules to determine how to strip off suffixes (e.g., "s" from "pixels" to get the stem "pixel"). A copy of Paice's rules is stored on Falcon in rules.txt. Paice's stemming procedure is incorporated in its entirely in the query.pas computer program. However, one could use other algorithms and sets of rules, such as the code that accompanies [Frakes92; Frakes, W., "Stemming Algorithms," Chapter 8].

Once the system has found a match (or matches if stemming is used) between the request term and the keywords in the inverted file (index), the system then searches through the list of query numbers and matches them with the query numbers in the query file (quest1.txt). This is a very slow process, due mainly to the fact that Turbo Pascal's file handling capabilities do not allow for efficiencies in file searches; a simple term with but a few query entries in the index may take up to a few minutes to process. However, it was the fact that the keyword search could be accomplished successfully that was important in this experimental procedure, not its efficiency (which will have to be improved in future generations of CASHE:PVS).

The user can have the output directed to the screen or to a file. To simplify matters here, the experimental program always puts the output file in the Turbo Pascal directory under the name output.dat. Obviously, this can and should be generalized. Based on experiments with the query procedure, a colleague suggested that the user should be able to get a list of the EDC entry numbers attached to the queries deemed relevant to the user's request term. Thus, the Turbo Pascal program, stored on Falcon in edc.pas, was created and run on the Zenith to process the query output file output.dat and produce another output file, by default entitled edcno.dat. Asterisks are indicated to show how often a given EDC entry number is repeated in the output.txt file.

Preliminary Results

In order to test the system, some sample user requests were performed, using the query.pas program on the Zenith. The first term was "pixels" with stemming (producing the individual terms "pixel" and "pixels"). The results of that search yielded six queries from the Design Checklist (Q440, Q517, Q526, and Q1002 for "pixel", and Q428 and Q525 for "pixels"). A copy of the output file is in output.txt on Falcon and is presented in Appendix XVIII. The use of edc.pas on this file produced a list of three EDC entries for this request (10.0415 - which was repeated another three times, 11.0114 - mentioned once, and 11.0207 - mentioned once). The output from edc.pas for this request is stored in edcno.txt on Falcon and is presented in Appendix XIX.

Since the major objective is to eventually be able to link the retrieved Design Checklist queries to the desired EDC entries, another test was run. For the stemmed request term "pixels," as discussed above, the three EDC entries were obtained in machine-readable form, as listed in edc.txt on Falcon. A listing of edc.txt is found in Appendix XX. This file was then transferred to the LSU Vaxstation, where a list of keywords in these three EDC entries were obtained via the same stop list program as used before. A sorted version of the keywords, with duplicates eliminated, is found in edc.sto on Falcon, with a listing in Appendix XXI. The stemming routine that accompanies [Frakes92] was used on edc.sto, the output of which was again sorted to remove duplicates, and is presented in both edc.stm on Falcon and in Appendix XXII.

A fortunate happenstance then occurred in that a colleague needed a search carried out for the unstemmed term "auditory" for work she was doing in the DEFTech Lab. The results of that search are stored in outputa.txt on Falcon, and a listing of the first few entries is presented in Appendix XXIII. We note that "auditory" is found as a term in some of the headers as well as in some of the queries, and as such, is much more common than the term "pixels". Thus, there are many more queries with the term "auditory", making the file outputa.txt much larger than output.txt.

Methodology - Adding (Fuzzy) Weights

In order to improve retrieval performance, weights should be added reflecting the relative frequency of occurrence of keywords in the queries and/or query headers. As mentioned above, this is a long-standing view, and has the advantages of being able to represent partial relevance and also being able to rank the output. It was decided to incorporate Salton's inverted document frequency (IDF) formula, as presented above, for the weights. A Pascal program was written to work as does the inv.pas program discussed above, except that IDF weights are incorporated instead of asterisks. The program is stored in winv.pas on Falcon, and the weighted index (inverted file) is stored in winv.txt on Falcon. A listing of the first few entries in winv.txt is found in Appendix XXIV.

With the weighted index constructed, the Turbo Pascal program stored in wquery.pas on Falcon was written and run on the Zenith in order to perform single term requests. This program parallels the unweighted query program query.pas except that the weighting mechanism has been incorporated. The results of the stemmed search for "pixels" is stored in wouput.txt on Falcon, and the listing for woutput.txt is in Appendix XXV. One notes that the listing provides the same results in terms of the queries that were found in output.txt, the unweighted search result. Further, the queries are not ranked as presented; this feature should be added with the choice of either being in order of query number or of weight being given to the user.

In addition, the results of the unstemmed request "auditory" are given in the file woutputa.txt on Falcon, and a listing of the first few entries in the results file are presented in Appendix XXVI. When our colleague saw these results and analyzed them to an extent, she requested another search be done on a term that occurred frequently in queries that had relevant EDC entries, that term being "intelligibility". The results of that search are shown in woutputi.txt on Falcon, with a listing of the first few entries being presented in Appendix XXVII.

In order to generate the weighted list of EDC entries mentioned in the Design Checklist query output, a Turbo Pascal program was written that paralleled the unweighted mechanisms in edc.pas. This new program, stored in wedc.pas on Falcon, produces a ranked list of EDC entry numbers, ranked in EDC entry number order. The results of this program for the stemmed term "pixels" is found in wedcno.txt on Falcon, with a listing in Appendix XXVIII. The results for the unstemmed term "auditory" and for the unstemmed term "intelligibility" are found in wedcnoa.txt and wedcnoi.txt, respectively; both on Falcon. In addition, a listing of the first few entries of wedcnoa.txt is seen in Appendix XXIX; while a listing of wedcnoi.txt is found in Appendix XXX.

In order to verify the observations of our colleague further in terms of correlations among terms, we generated the list of keywords in the file woutputa.txt, the results of the weighted search for the unstemmed request "auditory". The stop list program accompanying [Frakes92] run on the LSU Vaxstation was again employed to accomplish this, the file was then sorted, and the net result is stored in the file saud.txt on Falcon. Moreover, the first few entries are listed in Appendix XXXI. A Turbo Pascal program, stored as count.pas on Falcon, was written and run on the Zenith to produce a count of how often each term in woutputa.txt was seen. That output is listed in the file saudc.txt, stored on Falcon, with the first few entries listed in Appendix XXXII. The file saudc.txt was then sorted in descending order on the basis of the count to see which terms co-occurred with "auditory". The results of that sort are stored in the file saudcs.txt on Falcon, with the first few entries presented in Appendix XXXIII. It can be seen that the most common co-occurring term is "displays", which occurs as a very frequent header term, followed by "speech", "auditory", "signals", "signal", "noise", and then "intelligibility". This confirms our colleague's contention that "intelligibility" would be a good term under which to do further searching of the Design Checklist file.

Based on another suggestion by our colleague, we decided to improve the output generated from the weighted search of the Design Checklist queries. We decided to add to the output the topic header under which each retrieved query was listed in the Design Checklist. This gives each retrieved query a context in terms of the outline of the overall Design Checklist, aiding the user in terms of understanding just what has been found for him/her. Thus, another Turbo Pascal program, stored in wqh.pas on Falcon, was generated to produce the list of queries from woutput.dat (or its equivalent) merged appropriately with the headers attached to those queries as found in hq.dat. The results for the stemmed request "pixels" is found in qh.txt on Falcon, a listing of which is seen in Appendix XXXIV. The results for the unstemmed requests "auditory" and "intelligibility" are found in qha.txt and qhi.txt on Falcon, respectively. The listing of the first few entries of qha.txt and of qhi.txt are in Appendix XXXV and in Appendix XXXVI, respectively.

Methodology - Adding Boolean Request

Finally, it was decided to demonstrate that it was possible to handle more complex requests of many terms via Boolean logic. The constraints on time and computer capacity forced us to consider only requests of two terms, using only the Boolean connectives AND and OR. The use of fuzzy logic was employed for the weights [Kraft83, Kraft85, Kraft93]. By this we mean that if two terms were ANDed in the request, the minimum of the weights of each term as assigned to a given Design Checklist query was used to evaluate that query. Moreover, the maximum of the weights of each term was used when the terms were ORed in the request. Of course, a term that did not occur at all in the Design Checklist query or appropriate header was assumed to have a weight of zero.

The Pascal program, stored as bool.pas on Falcon, was generated to produce the output queries from a simple Boolean request of two terms linked by AND or by OR. The output for the unstemmed request "pixels AND symbol", which was processed on the Zenith microcomputer is stored in the file bools.tst on Falcon. The output for the unstemmed request "auditory AND intelligibility" is stored in the file bools.txt, also on Falcon. This latter request had to be processed on the LSU Vaxstation due to the size of the output. The listing for bools.tst is in Appendix XXXVII, while the listing of the first few entries in bools.txt is in Appendix XXXVIII.

The final computer program, a Turbo Pascal program stored in wbedcno.pas on Falcon, was generated. This parallels the wedc.pas program to produce a weighted list of EDC entry numbers, but with Boolean requests. The output for the unstemmed request "pixels AND symbol" is in the file wbedcno.tst on Falcon. The output for the unstemmed request "auditory AND intelligibility" is stored in the file wbedcno.txt on Falcon. The listing of wbedcno.tst is in Appendix XXXIX, while the listing of the first few entries of wbedcno.txt is in Appendix XL.

The output files wbedcno.tst and wbedcno.txt are sorted in EDC entry number sequence. A version of wbedcno.txt sorted on question weight can be found in the file wbedcno.srt on Falcon, and a listing of the first few entries is in Appendix XLI.

Summary, Conclusions, and Future Work

It is clear that keyword retrieval can and does work well in a variety of textual database environments. It has been shown that the possibility for such a system for CASHE:PVS *definitely does exist.* This possibility is a strong one, and a system could be constructed that would be able to let users quickly and easily find and use a few relevant entries.

However, a caveat must be expressed. The programs developed here were done to quickly demonstrate the feasibility of such a system. The step-by-step development outlined above in a time sequence needs to be submitted to a much more rigorous software engineering approach. A much more integrated system should system be developed with more general capabilities along with appropriate systems and user documentation. For example, the request processing mechanism needs to be able to handle requests of arbitrary complexity, including negation (NOT) and as many terms as the user desires. In addition, the user should be able to refine a request, adding terms to add queries to or to delete queries from a given search output.

Moreover, more research is needed. There are a variety of access points available to the CASHE:PVS user. One can use the Design Checklist header outline, the Design Checklist queries, the EDC Table of Contents outline, the EDC back-of-the-book index, and the impending keyword index of the

EDC entries that is currently being developed. One needs to correlate these to see which ones are most effective and how they can be integrated into a whole system for easy and efficient access for users. Furthermore, one needs to confront the entire CASHE:PVS system with real end-users, i.e., the designers for whom it is intended.

Acknowledgements

The author wishes to thank Donald L. Monk for his generous assistance, support, and guidance with this project. Discussions with him led to an appreciation and understanding of the current state and needed improvements in the system. In addition, provision of the basic data and computing facilities made this project possible. We would also like to thank Alan Straub for his help and for additional access to a copy of the EDC database. In addition, Dr. Maryalice Citera (another summer faculty research associate), Dr. Michael McNeese, the Logicon contractor staff in the DEFTech Lab (Thomas R. Cona, Rebecca Donovan, Roy Livingston, Glenn Johnson, and Keith Adams), and Eleanor Eleanor Chuang (a high school apprentice in the 1993 summer research program, also sponsored by AFOSR), all provided invaluable help and suggestions for portions of the project. Finally, the staff of the Human Engineering Division of the Armstrong Aerospace Medical Research Laboratory, in the Paul M. Fitts Human Engineering Building, provided needed support.

In addition, the U.S. Air Force must be thanked for its support in allowing us to maintain our professional responsibilities while working away from our home institution. These responsibilities included serving as the Editor of the *Journal of the American Society for Information Science* (JASIS), a Council member for the North American Information Processing Society (NAFIPS), chair of the American Society for Information Science (ASIS) Award of Merit Jury (ASIS's highest award), and an evaluation team chair for the Computer Science Accreditation Commission (CSAC). In addition, while serving as a Summer Faculty Associate, we were invited to serve on two panels and had a paper accepted at the upcoming ASIS meeting (to be held October 24-16, 1993 in Columbus, OH) and had another paper accepted at the upcoming NAFIPS meeting (to be held August 21-24, 1993 in Allentown, PA). Moreover, we visited Miami University in Oxford, OH, Wright State University in Dayton, OH, and the Air Force Institute of Technology (AFIT) in Wright-Patterson AFB, OH, and met with a colleague from Mead Data Central, where we visited with colleagues and informally discussed the fuzzy retrieval project for CASHE:PVS.

Finally, work by colleagues, including the software accompanying [Frakes92] and the software for stemming developed by Paice, must be properly acknowledged.

References

[Boff88a]
Boff, K. R. and J. E. Lincoln (eds.), *Engineering Data Compendium: Human Perception and Performance*, vols. I, II, and III, Wright-Patterson Air Force Base, OH: Human Engineering Division, Harry G. Armstrong Medical Research Laboratory, 1988

[Boff88b]
Boff, K. R. and J. E. Lincoln (eds.), *User's Guide: Engineering Data Compendium, Human Perception and Performance*, Wright-Patterson Air Force Base, OH: Human Engineering Division, Harry G. Armstrong Medical Research Laboratory, 1988

[Boff86]
Boff, K. R., L. Kaufman, and J. P. Thomas, *Handbook of Perception and Human Performance*, vols. I and II, New York, NY: John Wiley & Sons, Inc., 1986

[Boff91a]
Boff, K. R., D. L. Monk, and W. J. Cody, "Computer Aided Systems Human Engineering: A Hypermedia Tool," presented at SOAR '91, Space Operation Applications and Research, NASA-Houston, July, 1991

[Buckles92]
Buckles, B. P. and F. E. Petry (eds.), *Genetic Algorithms*, Washington, DC: IEEE Computer Society Press, 1992

[Boff91b]
Boff, K. R., D. L. Monk, S. J. Swierenga, C. E. Brown, and W. J. Cody, "Computer-Aided Human Factors for Systems Designers," presented at Human Factors Society annual meeting, San Francisco, CA, September, 1991

[Cona93]
Cona, T. R. and D. L. Monk, "Bringing Human Performance Data to the Design Table," to be presented at Human Factors and Ergonomics Society annual meeting, Seattle, WA, October, 1993

[Frakes92]
Frakes, W. B. and R. Baeza-Yates (eds.), *Information Retrieval: Data Structures and Algorithms*, Englewood Cliffs, NJ: Prentice Hall, 1992

[Klir88]
Klir, G. J. and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*, Englewood Cliffs, NJ: Prentice Hall, 1988

[Kraft85]
Kraft, D. H., "Advances in Information Retrieval: Where is That /#*%@^ Record?," in Yovits, M. (ed.), *Advances in Computers*, v. 24, Academic Press, New York, 1985, pp. 277-31o

[Kraft93]
Kraft, D. H., G. Bordogna, and G. Pasi, "An Extended Fuzzy Linguistic Approach to Generalize Boolean Information Retrieval," presented at the First Fuzzy Theory and Technology (FTT) '92 conference, Durham, NC, October, 1992, submitted to *Information Sciences*, 1993

[Kraft83]
Kraft, D. H. and D. A. Buell, "Fuzzy Sets and Generalized Boolean Retrieval Systems," *International Journal of Man-Machine Studies*, v. 19, 1983, pp. 45-56; reprinted in Dubois, D., H. Prade, and R. Yager (eds.), *Readings in Fuzzy Sets for Intelligent Systems*, San Mateo, CA: Morgan Kaufmann Publishers, 1993

[Monk92]
Monk, D. L., S. J. Swierenga, and J. E. Lincoln, "Developing Behavioral Phenomena Test Benches," presented at Human Factors Society annual meeting, Atlanta, GA, October, 1992

[Miyamoto90]
Miyamoto, S., *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Boston, MA: Kluwer Academic Publishers, 1990

[Salton89]
Salton, G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Reading, MA: Addison-Wesley Publishing Co., 1989

A DESIGN FOR A SMALL, FORCE REFLECTING, TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.
Associate Professor
Department of Manufacturing Engineering

Central State University
Wilberforce, OH 45384

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 1993

A DESIGN FOR A SMALL, FORCE REFLECTING, TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.
Associate Professor
Department of Manufacturing Engineering
Central State University

Abstract

A force reflecting, two degree of freedom joystick was developed at AAMRL, using servomotors as the torque actuators. This was a second generation joystick, which evolved from a similar, one degree of freedom joystick using a pneumatic actuator. The current design is a portable version of the first design. However, further miniaturization is desired to improve the reliability and durability of the joystick during transport. The focus of this work was to analyze the current design and make recommendations on the design of a smaller version of the same joystick. Details are given for a one half scale joystick utilizing the same mechanism, actuated by smaller servomotors, and interfaced to a microcomputer.

A DESIGN FOR A SMALL, FORCE REFLECTING, TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.

Introduction

A dual axis force reflecting joystick was constructed at the Armstrong Aerospace Medical Laboratory at Wright-Patterson Air Force Base, Ohio in 1991. This is the second generation of a similar device designed in the early 1980's [1]. At that time, the joystick was a single axis controller and the force reflection was accomplished through a pneumatic actuator.

It was discovered in the 1980's that by controlling the force reflection envelope felt by the human at the controller, an improvement in manual control efficiency was possible. For a compensatory tracking task with a sum of sines disturbance input, it was found that this first generation joystick allowed the human operator to behave as an optimal controller [2]. This prompted the Armstrong Laboratory to construct a smaller, portable two degree of freedom joystick interfaced to a microcomputer [3]. Since the construction of this device, studies have been done to determine if such a controller could also be useful to aid spastic individuals in controlling their environment more efficiently [4]. A direct application of this technology would be the use of such a joystick to control a motorized wheelchair to give spastic individuals safe,

independent transportation.

This has resulted in a need to further reduce the size of the current design. Factors to be considered for this task would include mechanism design, alternative materials selection, and actuator selection. The use of a notebook computer or an embedded digital processing system should also be considered. However, experimental studies having spastic individuals use the new design would be required to determine if the same improvements in performance would result.

Approach

An analysis of the components of the current design was done with respect to other design and technology considerations. The components looked at in detail were joystick mechanisms, actuators, computer interfacing, sensors, and control design. Based on the results of the survey, recommendations were given toward the construction of a smaller prototype to be tested.

Joystick Mechanism

The mechanism used for the current controller was very similar to joystick mechanisms used for microcomputer video games. A gimbal mechanism allowed the joystick to be moved in the roll and pitch directions with independence. Either end along both axes of the gimbal provided a coupling for a servomotor or a position and velocity sensor. The mechanism provided approximately $\pm$ 45

degrees for roll and approximately $\pm$ 60 degrees for pitch. The joystick was constructed primarily from aluminum with some steel components.

The mechanism was a sound design; however, tolerances and alignment were rather restrictive. A stress analysis was done on critical sites of the joystick. It was found that the strength of the mechanism was more than adequate for the torque rating of the servomotors used. However, evidence of wear and failure was found. The wear was primarily due to errors on the tolerance levels. Failure has occurred at a location where the shaft of a servomotor was coupled to the joystick mechanism. The problem was corrected. The failure was probably caused by large impulsive forces produced in the mechanism when the servomotors forcefully drove the joystick to the physical limits of the mechanism. Soft mechanical stops or electrical limit switches would easily eliminate this problem.

If the present mechanism were reduced to one half scale, the strength of it would still be adequate for the same torque limitations. The availability of aluminum would still make this the material of choice. Nevertheless, the use of composite materials could be used on certain components without loss of strength while reducing the weight of the joystick.

Actuator Selection

The current joystick used two torque servomotors with a peak torque rating of 4.8 ft-lbs (Inland model QT-3802-B) [5]. The motors provided an adequate amount of torque; however, they would take up too much volume for a scaled down version.

Alternative actuators for the smaller joystick were surveyed. Typical alternatives include pneumatic and hydraulic analogs. Even though a smaller pneumatic controlled joystick could easily be designed, nonlinearity and noise considerations did not make this selection an attractive one. Hydraulics were much better suited for this application; however, the servicing of hydraulic systems could become messy and dangerous. As a result, other methods of actuation were sought.

Stepper motors and electromechanical brakes were surveyed for the application. The low cost and availability of stepper motors could make this an attractive avenue. The disadvantages of using stepper motors would be the limits on the slip torque and the complex control system that would have to be developed. Electromechanical brakes also could be a good alternative. The drawback in this case would be the lack of a means to give the joystick a restoring spring force to center it.

Electrorheological fluids showed real promise of being an alternate to the servomotor as the torque actuator. More

theoretical and application studies were done using this technology over the last few years. Electrorheological fluids have the potential of providing all the impedance characteristics of the joystick. However, until the technology in this area matures more, this alternative would not be considered as a possible torque actuator.

The results of the survey of force or torque actuators suggested that torque servomotors should continue to be used as the actuators for the smaller joystick. Linearity, ease of programmability, and easy maintenance were major factors in the decision. This conclusion was strengthened by the selection of brushless, servomotors used to actuate a 6 degree of freedom handcontroller, designed by Cybernet Systems Corporation, of Michigan [6,7].

Computer Interfacing

The current joystick is interfaced to a PC compatible microcomputer by means of an I/O board. For the use of experimentation, where data collection and storage are necessary, this standard system would continue to prove adequate for the new design. However, for the issue of ease of transport, the use of a notebook computer could provide the same flexibility and power of a standard desktop. The difficulty would be in finding an I/O board for use with a notebook computer. The real possibility exists that an I/O board for this application would have to be

designed.

A more attractive alternative would be to embed the entire system with the use of a digital signal processing chip (DSP). There are vendors available that have low cost DSP development systems and kits in which the developed software could be downloaded into an embedded system quite easily.

Sensors

Potentiometers and tachometers were used on the current design to measure the position and velocity of the joystick. These signals were used to modulate the force reflection of the joystick. For a smaller joystick, these same sensors could be used. Only a potentiometer could be used, but the velocity would have to be derived by taking the derivative of the position signal. There would be limits to the fidelity of this computed signal. In addition, smoothing filters would be required to reduce the noise produced. Only a tachometer could be used, but the position signal would have to be derived by integrating the velocity signal. In addition, a reference switch must be employed and software developed in order to measure absolute position.

In some instances, the measurement of forces applied to the joystick would be a useful signal to consider in modulating the force reflection. The current design did not employ any force sensors because the linear relationship between the motor current

and torque produced was adequate to control the complete system. Particularly in some systems which employ gear reduction measures, controlling motor torque through a feedback system is mandatory to keep the system stable. The reason for this is that gear reduction could introduce large time constant delays due to the increased effective moment of inertia of the joystick.

Control Design

The architecture for the current joystick is shown on Figure 1. This would be very similar to the architecture for the new design. The control laws for both axes were programmed on the computer. For this system, all fundamental mechanical characteristics of the joystick could be realized through a standard PID controller. However, there is enough flexibility for more complex control laws, including neural and fuzzy control.

Two additional I/O channels would be required if force feedback were to be included. In this case, it would be preferred to employ a state space approach. More flexibility and modularity in the software development would result.

Conclusions

The results indicate that scaling down the current joystick design by one half is plausible for all components except for the actuators. In order to employ servomotors of the same quality,

size of the motor does not necessarily follow the reduction of peak torque. Other means of actuation would be possible. However, at this stage of development, the simplicity and reliability of servomotors would be greater considerations.

Even though other materials could be used for the gimbal mechanism, the cost of purchasing and processing these materials may fall outside this particular phase of development. As an end result, the use of these alternative materials could provide equal or greater strength with a substantial reduction in mass. However, for the interim, aluminum is readily available and would be highly durable for a prototype joystick under rigorous testing conditions.

As long as rigorous testing would be performed on the scaled down joystick, the need for interfacing it to a notebook computer or an embedded DSP system would not be practical. The use of a standard microcomputer would suffice. However, if the joystick would be transported with any regularity, the use of a notebook computer would tremendously ease the transport and reduce the risk of damaging the system. If certain control laws would be proven highly desirable, the use of an embedded DSP system would be justified.

If the force reflection of the joystick is provided by direct drive servomotors, only kinematic variables need to be measured.

The present design measures only position and velocity. The performance of a smaller joystick under these conditions should yield similar results.

However, if the new design includes gear reduction, analysis would need to determine if the system would remain stable measuring only position and velocity. If the system time delay is too large, force feedback would be required and a force sensor would have to be developed to measure the applied force on the joystick. A direct drive system would simplify the control system where more attention would be given to the specifics of the man-stick interface.

Recommendations

As a result of the research effort this summer, it has been demonstrated that a need exist to initiate studies with a smaller, more portable force reflecting joystick. Although many design approaches were possible, the following recommendations are given considering ease of synthesis, reliability, and cost.

1. The joystick gimbal mechanism should be reduced to one half scale, constructed from aluminum.
2. Direct drive servomotors should be employed, realizing that the torque rating required may result in motors that are disproportionately larger than the joystick gimbal.

3. A standard, desktop computer system should be used for careful experimentation; however, a notebook computer should be employed if transportation of the device is required.
4. Potentiometers and tachometers should remain the sensors used to measure the stick position and velocity.
5. Force feedback in the control system should not be considered unless a gear reduction drive is employed.
6. All of the above items are within the budget limits of the Summer Research Extension Program, and a proposal to continue the work is recommended.

References

1. Repperger D.W., Frazier J.W., van Patten R.E. A Smart Stick Controller Based on a Static Equilibrium Model. Proceedings of the 1984 National Aerospace and Electronics Conference (NAECON), 810-817.

2. Morris A.,Jr., Human Operator Performance in a Regulatory Tracking Task with an Active versus a Passive Control Stick. Ph.D. Dissertation, Wright State University, 1988.

3. Repperger D.W., Scarborough E.L., Chelette T.L. Construction of a Dual Axis Force Reflection Stick and Test Station. Air Force Tech Report AL-TR-1992-0041, November 1991.

4. Trimble J., Repperger D.W., Phillips C.A. Parameter Determination for Development of Spastic Tolerant Stick Controllers. Proposal to the Veterans Administration, accepted Dec 27, 1991.

5. Inland Motor Specialty Products Group: Direct Drive DC Motors, May 1985, 3rd printing, General Motors Corp., Detroit, MI.

6. Jacobus H.N., Riggs A.J., Jacobus C.J., Weinstein Y. Implementation Issues for Telerobotic Handcontrollers: Human-Robot Ergonomics. Human-Robot Interaction editors M. Rahimi and W. Karwowski. Taylor and Francis, 1992.

7. Jacobus H.N., Personal Communication, July, 1993.

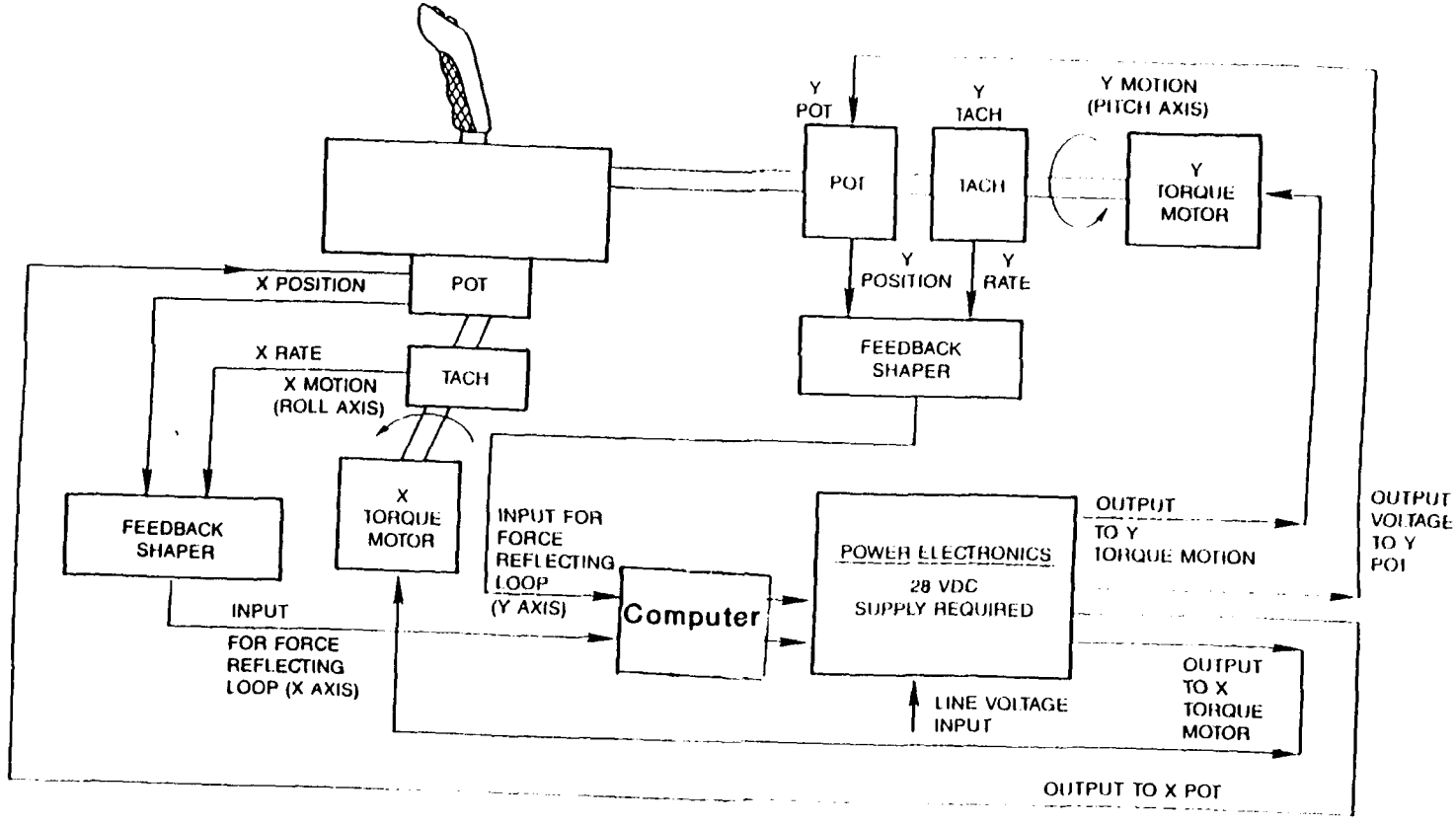


Figure 1 Joystick Architecture

POST-PROCESSING OF CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre

Assistant Professor

Department of Electrical & Computer Engineering

Ohio University

Athens, Ohio 45701

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 1993

POST-PROCESSING OF CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre
Assistant Professor
Department of Electrical & Computer Engineering
Ohio University

Abstract

Full field surface data of cylindrically shaped objects, such as a human's head, can be acquired by rotating a triangulated laser and imaging system about the subject. The method of acquisition is imperfect and requires post processing of the data obtained. Some of the problems that must be addressed by post-processing include: spikes, rough surface data, irregular surfaces and missing data points. The problems require a variety of different processing tools, many of which are already used in the research community. However, the cylindrical nature of the data presents problems with implementation of these tools. This report presents recently developed software tools that are now available for editing and analyzing head scan data as well as some likely applications.

POST-PROCESSING OF CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre

I. Introduction

The development of helmet systems has continued to evolve to meet the challenges of even more sophisticated aircraft. Better performing helmets, in terms of protection and head mounted sensory devices, require exact three dimensional anthropometry. The latest research in this area has employed computer vision and laser scanning of heads to achieve full field three dimensional range data.

To acquire surface data, a laser light is projected along a known trajectory and is located on the surface with an imaging camera. The search for the laser illumination proceeds from a maximum allowable radial diameter to zero. As the subjects are scanned, stray external light sources, unexpected surface reflections and suspended dust particles result in spikes and rough surface data. These sources of error, as well as intentionally placed black markers used for landmarking, c o cause missing data points. Latex bald caps are worn by each subject in an effort to capture cranium shape (hair does not reflect well) which quite often generates irregular surfaces. A subject's unintentional movement during the scanning process (approximately 15 seconds) can also generate irregular data.

Post-processing range data to compensate for unwanted defects has been an area of research for many years. The topic is usually included in the field of image processing and many techniques are available. One unique feature of the head scan data is its cylindrical coordinate system. This requires special software design techniques. The Computerized Anthropometric Research & Design Laboratory at the U.S. Air Force's Armstrong Laboratory has been developing software to manipulate and display the head scan data. The main program called INTEGRATE has a wide variety of functions, implemented in a modular programming fashion. New modules were added to this program to address the data problems mentioned above.

In this report, the new data processing modules incorporated into INTEGRATE will be described in Section II. Section III will discuss some explicit examples of applications appropriate for the new modules. Section IV will then present a discussion.

II. Post-processing tools

Several new functions have been added to the INTEGRATE repertoire for dealing with the head scan data. The first group of functions described are general purpose image processing techniques, useful for performing experiments and documenting results. Given below are descriptions of the commands and examples of their use. All the examples presented in this Section refer to Object 1. Object is the term used by INTEGRATE to refer to a head scan. Object 1 is shown in Figure 1. Additional information about the following group of general functions can be found in several image processing texts[1][2] or by referring to the source code.



Figure 1. The Object 1 head scan, used in all of the examples.

COMMAND: *AND reference_object replace_object*

This command performs a logical AND operation on two objects. Points with a value of zero in the objects' data are equated to binary zeros, while non-zero values are considered binary ones. A reference object and replace object, of the same size, must be specified. The values of the reference object are stored in the replace object, wherever the two objects AND to a binary one.

Example: At the INTEGRATE command prompt, the user may input:

AND 1 2

In this example Object 1 (given in Figure 1) remains unchanged. Object 2 is made up of two smooth surface patches as shown in Figure 2. After the *AND* operation, Object 2 now contains the nose and ears of Object 1, as shown in Figure 3.

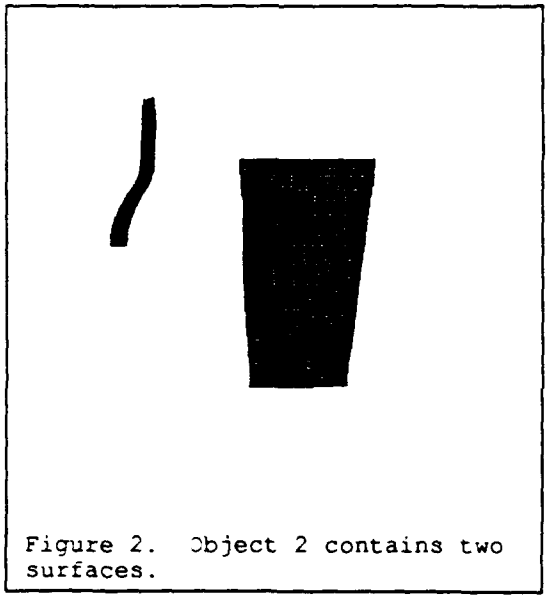


Figure 2. Object 2 contains two surfaces.

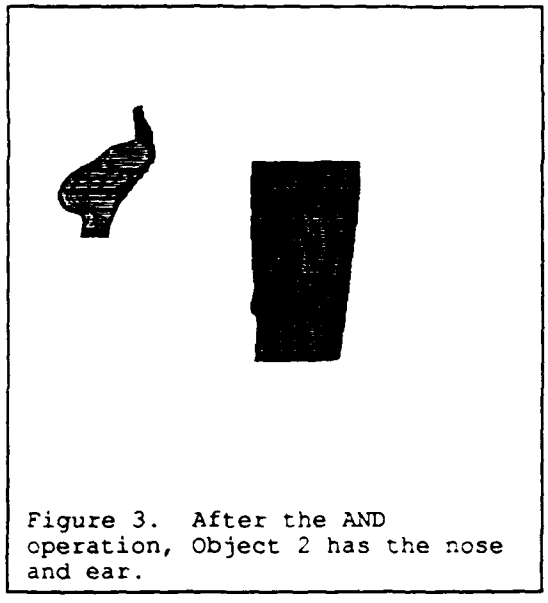


Figure 3. After the AND operation, Object 2 has the nose and ear.

COMMAND: *ABSSUB reference_object replace_object*

This command performs an absolute subtraction on two objects. A reference object and replace object of the same size, must be specified. The results of subtracting these two objects are placed in the replace object.

Example: At the INTEGRATE command prompt, the user may input:

ABSSUB 1 2

Once again Object 1, in Figure 1, remains unchanged. Object 2 will be the resultant of the *AND* example as shown in Figure 3. After the *ABSSUB* operation, the nose and ears have been subtracted from Object 1, as depicted in Figure 4.



Figure 4. The absolute subtraction of Object 1 from Object 2.

COMMAND: *HISTOGRAM object interval*

This command creates a histogram of an object. The histogram is limited to ten equally spaced, user specified intervals. The histogram is placed in a file called 'histogram.dat'. The software module then generates a system call which activates 'jot', Silicon Graphics' window based full-screen text editor. 'jot' will display the histogram in a new window, making it available for editing.

Example: At the INTEGRATE command prompt, the user may input:

HISTOGRAM 1 10

Presented in Figure 5 is the histogram generated for Object 1 at 10 mm intervals.

INTERVAL	POINTS	ALL POINTS	NO ZEROS
0.00 - 0.00	0	0.00%	****
0.00 - 10.00	0	0.00%	0.00%
10.00 - 20.00	0	0.00%	0.00%
20.00 - 30.00	0	0.00%	0.00%
30.00 - 40.00	33	0.04%	0.04%
40.00 - 50.00	2470	3.30%	3.30%
50.00 - 60.00	8193	10.96%	10.96%
60.00 - 70.00	12882	17.23%	17.23%
70.00 - 80.00	14652	19.60%	19.60%
80.00 - 90.00	14409	19.28%	19.28%
90.00 - 100.00	10503	14.05%	14.05%
Above > 100.00	11610	15.53%	15.53%

The mean (excluding zeroes) is: 79.993835
The standard deviation (excluding zeroes) is: 17.725634

Figure 5. Histogram of Object 1.

COMMAND: *RUIN object object_copy*

This command randomly creates void patches in an object. The command requires both the object and a copy of the object to operate. After execution, the copy of the object will contain the data of the newly created voids.

Example: At the INTEGRATE command prompt, the user may input:

RUIN 1 2

Before executing this command, it was assumed that both Object 1 and Object 2 are the head scan present in Figure 1. The results of the *RUIN* are shown in Figures 6 and 7. The results differ each time *RUIN* is executed.

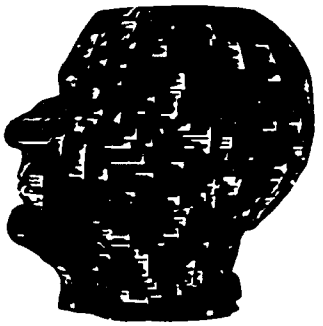


Figure 6. Object 1 with randomly inserted voids.

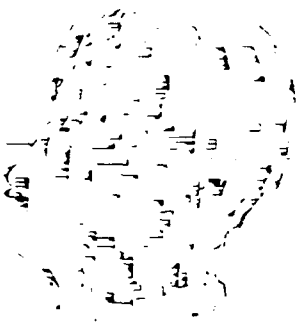


Figure 7. Object 2 now contains data points voided from Object 1.

COMMAND: *THRESHOLD object value {ABOVE|BELOW}*

This command performs a threshold operation on an object. The object and threshold values are specified by the user. The qualifiers 'above' or 'below' refer to the values to be zeroed.

Example: At the INTEGRATE command prompt, the user may input:

THRESHOLD 1 82 BELOW

Figure 8 shows the results of thresholding Object 1 where all points below 82 mm were set to zero.

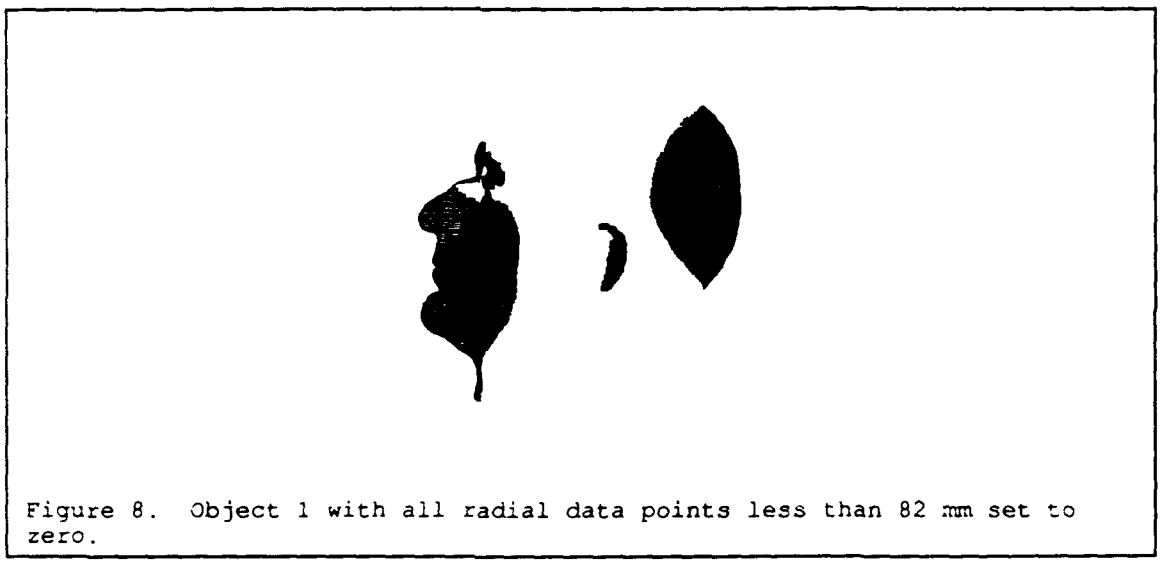


Figure 8. Object 1 with all radial data points less than 82 mm set to zero.

The next group of commands, now available in INTEGRATE refers to an image processing technique known as morphology. Mathematical Morphology is the probing of an image shape with a specified structuring element. The geometry of the structuring element implies certain geometric characteristics about the image being analyzed. For the morphological operations described below, the structuring element used is shown in Figure 9. The element is the intersection of a sphere at a specified radius, r_s, and a cylinder at a specified radius, r_c. The origin of the structuring element is located at a distance of r_c from the top of the sphere, along the cylinder's center axis.

Two morphological operations, erosion and dilation will be briefly

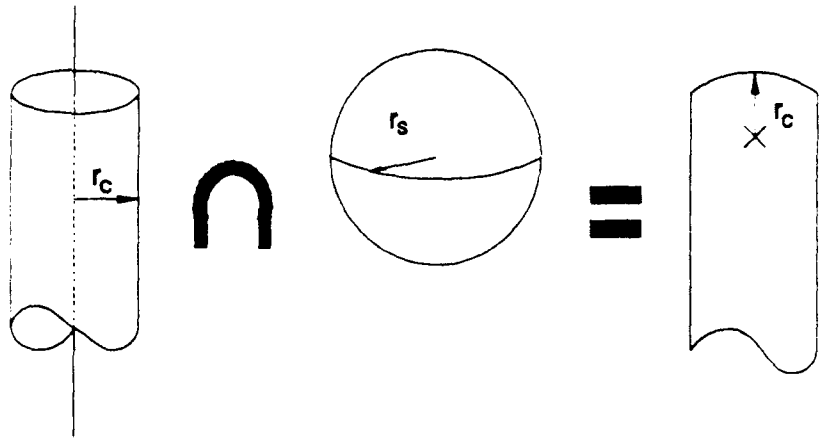


Figure 9. The definition of the structuring element used for morphological operations.

introduced below. Further information on morphological techniques and the theory of Mathematical morphology can be found in publications by Dougherty[3][4] and Serra[5].

As stated earlier, morphology concerns itself with the probing of a signal with a geometric structuring element. To find the erosion of a signal by a structuring element, the element is placed below the signal and forced up to the highest elevation which will touch but not cross the signal. The origin of the element becomes the new signal value. The process is repeated throughout the length of the original signal. During erosion, the structuring element was always placed virtually below the smallest radial value for an affected longitude in the scan. The element was pushed up from this point. Because of the cylindrical coordinates, fewer points have an influence on the structuring element as it moves up.

COMMAND: *ERODE sphere_size cylinder_size (MINUS)*

This command performs a morphological erosion on an object. The structuring element used is a cylinder with a spherical top. The user specifies the radius of both the cylinder and the sphere. The default erosion operation is positive. A negative erosion can be performed by

adding the modifier "minus".

Example: At the INTEGRATE command prompt, the user may input:

ERODE 10 10

Figure 10 shows the results of performing this erosion on Object 1. The structuring element has $r_s = 10$ and $r_c = 10$.

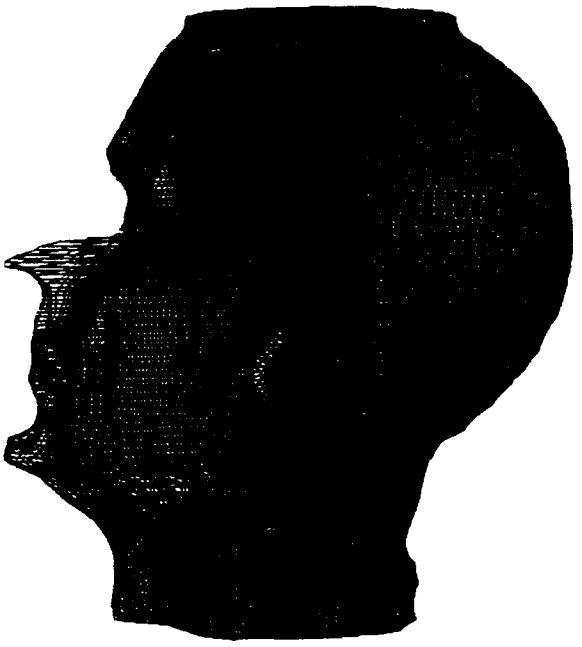


Figure 10. The erosion of Object 1.

Dilation is the dual operation to erosion. When dilating a signal with a structuring element, the signal becomes a path for the origin of the element. As the element translates its path, a new signal is created which is the minimum signal needed to bound the structuring element. A dilation can also be found by reflecting the structuring element and finding the minimum signal above the original signal in which the reflection will fit.

COMMAND: *DILATE sphere_size cylinder_size (MINUS)*

This command performs a morphological dilation on an object. The

structuring element used is again a cylinder with a spherical top. The user specifies the radius of both the cylinder and the sphere. The default dilation operation is positive. A negative dilation can be performed by adding the modifier "minus".

Example: At the INTEGRATE command prompt, the user may input:

DILATE 5 3

Figure 11 shows the results of performing this dilation on Object 1. The structuring element has $r_1 = 5$ and $r_2 = 3$.



Figure 11. The dilation of Object 1.

The third group of commands, now available in INTEGRATE perform standard filtering operations. These commands replace the cumbersome convolution command sequence previously available. Both commands making up this group perform filter creation and convolution in one step.

The primary difference between these two commands is their technique for handling boundary points. On some occasions, it is desirable to smooth only

a portion of the head scan data. This is accomplished in INTEGRATE by trimming the data. The outer most points of the trimmed segment make up the boundary. For the *FILTER* command, boundary points are smoothed by using adjacent data points, outside of the trimmed area. This may result in discontinuities when the segment is re-integrated with the head. To help alleviate this problem, the *FILTSEG* command replicates the boundary point value and uses that value in place of data points outside of the trimmed segment. This helps keep the boundary points stationary and, hence, continuous with the rest of the head scan.

COMMAND: *FILTER {GAUSS|DISCRETE|GREEN} scale {LON|LAT|BOTH}*

This command filters the data with one of the currently available smoothing filters. The user selects the type of filter and a scale factor to determine the smoothing strength of the filter. (The larger the scale, the larger the number of adjacent points involved in the filter function). Options are: GAUSSIAN, DISCRETE, or GREEN filters. GAUSSIAN refers to the standard analog Gaussian filter. DISCRETE refers to a discrete formulation of the Gaussain filter.[6] GREEN refers to the Green filter. The filter may then be applied latitudinally, longitudinally or in both directions. Note: When using *FILTER* on a trimmed area, points outside of the area are used in calculations.

Example: At the INTEGRATE command prompt, the user may input:

FILTER GAUSS 1.2 BOTH

Figure 12 shows the results of smoothing Object 1 with a Gaussian filter.

COMMAND: *FILTSEG {GAUSS|DISCRETE|GREEN} scale {LON|LAT|BOTH}*

This command is identical to the *FILTER* command except that the edge of a trimmed area is replicated and used in place of outside data points. This helps to prevent discontinuity from the remaining head scan points.

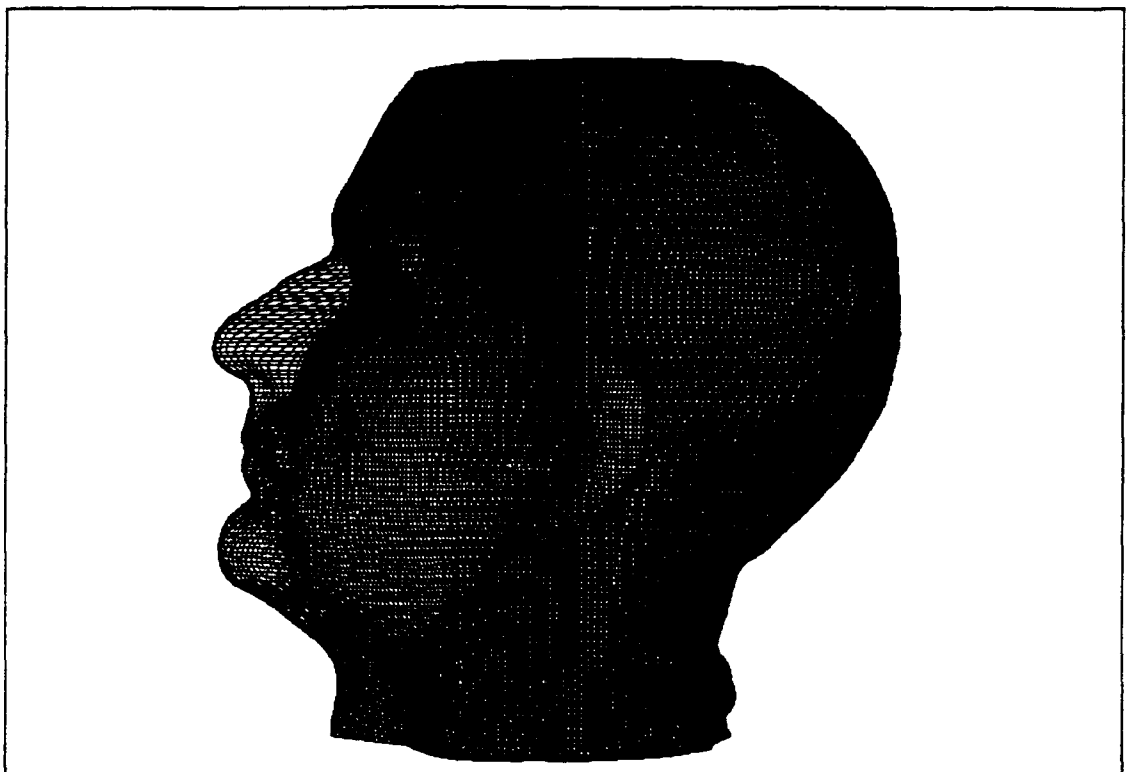


Figure 12. The results of smoothing Object 1 with a Gaussian filter.

Finally, one new command now available in INTEGRATE, can be classified by itself. The *INTERPOLATE* command uses a regularization technique to fill voids.[7] The results of the interpolation vary, depending on the filter used. Currently three filters are available. The first filter option, LINEAR, generates a linear interpolation. Missing points are placed in a straight line fashion between bounding endpoints. The *DO FILL* command already present in INTEGRATE performs the same function in a slightly less robust manner. However, *DO FILL* is much faster than *INTERPOLATE* and will provide adequate results in most cases, when linear interpolation is desired. Use of the SMOOTH filter generates a spline interpolation. This technique has been shown to provide better estimates of missing data points than linear interpolation.[8] The third filter, GAUSS, is experimental and not recommended for use at this time.

COMMAND: *INTERPOLATE (LINEAR|SMOOTH|GAUSS)*

This command replaces voids with approximated values based on surrounding points. Three alternative methods are available for filling the voids: LINEAR, SMOOTH or GAUSS. LINEAR interpolation fills in a straight line between surrounding points. GAUSS interpolation provides a very nice looking fill. SMOOTH interpolation provides the most accurate data point estimation.

Example: At the INTEGRATE command prompt, the user may input:

INTERPOLATE SMOOTH

To appreciate the *INTERPOLATE* command, the Object must have random voids. For this example, Object 1 was processed by *RUIN* before interpolation. Figure 13 shows the results of estimating the missing data.



Figure 13. Results from interpolating Object 1 after the *RUIN* operation.

III. Specific Applications

One particularly troublesome flaw discussed in the Introduction is

impulse noise or spikes. To remove these spikes and maintain the integrity of the head scan data, a morphological opening can be performed. An opening operation is defined as an erosion operation followed by a dilation. Using the structuring element presented earlier has been shown to be effective at removing spikes. The technique can be justified based on skin thickness and the topology of a human head.[9]

A radial value of 2 mm for both the spherical and cylindrical radius effectively removes intermittent spikes from the head scans. A large group of head scans, taken at the Eglin Air Force Base, was severely affected by an external light source. To remove the large continuous spikes that were generated, a cylindrical radius of 4 mm should be used. The opening should be limited to the back of the head where most of these spikes occurred.

The group of surface filtering commands can be used to smooth rough surfaces and discontiguous surface data. Preliminary results show that a *DISCRETE* filter with a scale factor between 1 and 2, results in a smooth scan without compromising the head geometry.[10] DISCRETE which is a discrete formulation of the Gaussian filter performs better than GAUSSIAN, its analog counterpart. GREEN, the Green filter whose properties include preserving corners in surface data, is still under investigation.

The use for interpolation is self evident. The structure of the *INTERPOLATE* function allows different mathematical techniques for filling voids to be tested without modifying the software. This flexibility is demonstrated by the three currently available options.

IV. Conclusions

New processing tools have been added to the range data handling software. The tools have already been proven to be effective at eliminating spikes, interpolating data voids and smoothing rough surfaces. Because of the modularity and flexibility of the commands, future data processing applications are certain to arise. With these enhanced features, the INTEGRATE software will continue to be at the leading edge of cylindrical scan data analysis.

References

1. B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, 1986.

2. E.L. Hall, *Computer Image Processing and Recognition*, Academic Press, New York, 1979.

3. E.R. Dougherty and C.R. Giardina, *Matrix Structured Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

4. E.R. Dougherty, *An Introduction to Morphological Image Processing*, SPIE Optical Engineering Press, Bellingham, Washington, 1992.

5. J. Serra, editor, *Image Analysis and Mathematical Morphology*, Vol. 2, Academic Press, New York, 1988.

6. T. Lindeberg, "On the construction of Scale-Space For Discrete Images", Dept. Num. Anal. Comput. Sci., Royal Institute Tech., Stockholm, Sweden, Internal Rep. TRITA-NA-P8808.

7. J.H. Nurre, "Post Processing Cylindrical Surface Data of a Subject's Head", *SPIE Biomedical Image Processing and Biomedical Visualization*, Vol. 1905, pp. 872-878, 1993.

8. J.H. Nurre, "Interpolating Missing Data Points for Head Scan Data", Technical report, In preparation.

9. J.H. Nurre and J. Whitestone, "Removing Spikes from Head Scan Data Files", Technical report, In preparation.

10. H. Fang and J.H. Nurre, "Smoothing Head Scan Data with Generalized Cross Validation", *SPIE Tools for Manufacturing*, Boston, Massachusetts, September, 1993.

EXAMINING THE ROLE OF JUDGMENT AND DECISION MAKING IN THE EMERGING FIELD OF COGNITIVE ENGINEERING

Thomas E. Nygren
Associate Professor
Department of Psychology

Ohio State University
1885 Neil Avenue Mall
Columbus, Ohio 43210

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

August, 1993

EXAMINING THE ROLE OF JUDGMENT AND DECISION MAKING IN THE EMERGING FIELD OF COGNITIVE ENGINEERING RESEARCH

Thomas E. Nygren
Associate Professor
Department of Psychology
Ohio State University

Abstract

Cognitive engineering is a term that has been applied to a new approach to studying the influence of information technology on the human-machine (i.e., computer) work system. It suggests a research focus that incorporates but also goes beyond contemporary research in human factors, computer science, and cognitive psychology. As a subfield of cognitive psychology, the study of judgment and decision processes has, over the past two decades, shown itself to be an important component of the overall study of cognitive functions. I will attempt to show in this paper that the study of decision processes should be a significant component of the new field of cognitive engineering as well. In this paper findings from a number of research areas in the field of judgment and decision making are summarized and their relevance to cognitive engineering is described. Some general proposals for additional research integrating decision making and cognitive engineering are suggested.

Introduction

Cognitive engineering is a term that has been used to describe a research focus that incorporates but also goes beyond contemporary research in human factors, computer science, and cognitive psychology. I will restrict for the most part the term *cognitive engineering* to the idea of describing and studying the multi-faceted influence of information technology on the man-machine work environment. In this paper, I will treat the field of cognitive engineering as primarily including, although not limited to, the study of man-machine interfaces and the cognitive functions that produce and allow for necessary interactions between the human operator and a machine/computer that is capable of exhibiting some cognitive behaviors. Such a definition, however, is in many respects not far removed from that of traditional human factors work in the sense that it suggests a primary focus on the human-work system or work environment interaction. But it extends beyond this human-system interaction in that there is also the fundamental emphasis on both human and machine information processing, components of cognition, and the relationships among the specific component cognitive processes.

Initial conceptualizations of cognitive psychology (Norman & Rumelhart, 1975; Bourne, Dominowski, & Loftus, 1979) focused primarily on such processes as perception and acquisition of new information, storage and retrieval of information, short and long-term memory, language acquisition, concept formation, and reasoning. Higher order processes of judgment and decision making (J/DM) were largely ignored, despite having a clear overlap with retrieval, concept formation and reasoning. This is not surprising since at the time of the "cognitive revolution", judgment and decision making research, other than the prescriptive or normative modeling found in formal decision analysis, was at its infancy. Largely because of the important initial work of Kahneman and Tversky (1973; Tversky & Kahneman, 1973, 1974), judgment and decision research has, over the past two decades, now become a prominent and established component of cognitive psychology and cognitive science.

Basic and applied research on human decision making has focused on three broad issues: (1) ***process*** or how people actually make decisions, (2) ***quality*** or how good their decisions are, and (3) ***improvement*** or how

decision behavior can be made better. Each of these research domains has produced important work that has implications for the new field of cognitive engineering. In the next sections I will illustrate the importance of this research with a discussion of seven specific sub-areas of decision making research that cut across each of the three broad topics mentioned above and that are relevant to the emerging field of cognitive engineering. In each case a brief summary of the current research is provided followed by a suggested research program relevant to cognitive engineering. These topics are, of course, not meant to be exhaustive; they are merely an attempt to illustrate the need for understanding the contribution of decision making processes as well as perception, learning, memory, retrieval, language, and reasoning in designing and developing optimal man-machine systems.

1. **Biases and Heuristics in Judgment and Decision Making.**

There is now a volume of published literature (see Kahneman, Slovic, & Tversky, 1982) that clearly demonstrates that people often introduce biases in each stage of the decision making process. These include biases in the actual generation of hypotheses about a risky decision making situation, in the perception of the likelihood of the plausibility of the hypotheses, and in the perception of likelihoods of possible outcomes associated with a chosen course of action. For example, because of confirmation bias, when generating potential hypotheses relevant to a complex decision situation, people often will show a far greater confidence or estimated probability for the most likely hypothesis than is actually warranted, and they generate only a small number of the feasible problem-solving actions that may be appropriate for the given situation (see Kahneman, Slovic, & Tversky, 1982 for a review of this literature).

Risks are also often inaccurately assessed throughout the decision making process. Studies have shown that people do not make good probability estimates (Fischhoff, 1977; Slovic, Fischhoff, & Lichtenstein, 1980, 1982; Slovic & Lichtenstein, 1983). People overestimate low and underestimate high probabilities and they ignore base-rate information (Kahneman & Tversky, 1973; Bar-Hill, 1980); they revise opinions too conservatively (Edwards, 1968); and they often indicate excessive confidence in their judgment (Fischhoff & Slovic, 1980), and they are known to be subject to hindsight biases and the "I knew it all along" effect (Fischhoff, 1975).

Importance to Cognitive Engineering. Both laboratory and field research have described many interesting situations where one

expects human decision makers to fail in making optimal choices among a set of courses of action. Yet, despite this large research base, Hogarth (1981) has argued that "a serious criticism [of this research] is the failure to specify ***conditions*** under which people do or do not perform well. There are many interesting phenomena, but few attempts at theoretical integration" p. 198). This statement is as true today as it was a decade ago; we have identified many specific contexts in which decision biases can be robustly demonstrated, but we have not generated an acceptable and cohesive theoretical explanation that would allow us to generalize these phenomena in such a way as to be useful to classes of man-machine interactions or work environments.

If cognitively-based system design is going to lead to better human machine/computer interaction, then it is critical to those components of the human-machine interface that involve dynamic decision making with serious time constraints, that we know more about where biases may be expected to be used and/or are detrimental to performance. This implies the need for models that can theoretically relate cognitive design with dynamic decision making. Edwards (1988) has correctly noted that current decision models remain, for all practical purposes, static models. Despite some early initial efforts, most notably on his part, (see Edwards, 1962), adequate dynamic decision making models are simply nonexistent, and useful methodological tools for studying dynamic decision making have been very slow in developing. The advent of inexpensive and easily programmable microcomputers, however, makes it possible to change this and for the first time to develop laboratory studies that can model relatively complex cognitive environments for which a major component of the human-machine interaction in a dynamically changing judgment and decision making task.

Such studies are arguably an important component of cognitive engineering work. For example, from a cognitive engineering perspective, two important questions related to understanding use of heuristics decision bias in man-machine environments are " How can systems be designed to take advantage of individuals' use of some heuristics as decision aids?", "How ***should*** human operators (or expert system interfaces that they must rely on) relate uncertainties to observations of events so that biases can be minimized?" and "What mental processes actually control the use of heuristics and biases in operators' choices and judgments of uncertainties?" As Tversky and Kahneman (1974) pointed out twenty years ago, heuristics are "highly economical and usually efficient, but they lead to systematic and predictable errors. A better understanding of the heuristics and the

biases to which they lead could improve judgments and decisions in situations of uncertainty (p. 1131)."

Although the use of heuristics and biases is now clearly established in the J/DM literature, a next logical step is to study their use in actual work environments. A goal should be the training of human operators to become better decision makers in the sense of eliminating or reducing decision biases and optimally using simplifying heuristics as valuable decision aids.

2. Framing in Judgment and Decision Making.

By far the most significant work of the past several decades in formulating a ***descriptive*** model of decision making, is Kahneman and Tversky's prospect theory model (1979, Tversky & Kahneman, 1992). They have shown that individuals can actually make opposite overt choices between a pair of competing alternative if the situation is merely framed differently. This idea of "framing" has since been widely studied and has become recognized as an integral component of the initial phase of many decision situations (Hersey & Shoemaker, 1980; Slovic, Fischhoff, & Lichtenstein, 1982; Redelmeier & Tversky, 1992). It is proposed as a cognitive operation that acts upon the components of the choice alternatives (i.e., their outcomes and probabilities) so as to often produce observable preference reversals.

Specifically, Kahneman and Tversky and others have repeatedly found that when choice alternatives are framed in terms of what could be gained, people are generally risk-averse. That is, they will often choose a sub-optimal sure gain in order to avoid or reduce risk. However, when the same situation is framed in terms of what could be lost, people are often risk-seeking -- that is, willing to take a chance on a sub-optimal gamble in order to avoid a sure loss.

Importance to Cognitive Engineering. Although the robustness of this framing effect is now well-documented, it has also become clear that not all individuals are susceptible to it, at least not to the extent of actually producing observable choice inconsistencies (Tversky & Kahneman, 1988). This implies that susceptibility to contextual framing of choice situations may be at least partially controlled by some predispositional factors within the individual decision maker. It is somewhat surprising, then, as Schoemaker (1993) notes, that individual differences variables like propensity to be risk-seeking or risk-averse, to be loss-averse, or to be impulsive have not been incorporated into information processing models explaining risky decision making behavior. Indeed, despite their impressively extensive

work, Tversky and Kahneman 1988), acknowledge that "we have identified several common rules of framing, and we have demonstrated their effects on choice, but we have not provided a formal theory of framing (p. 186)."

It would seem that a formal theory of framing would be extremely valuable to those cognitive engineers who must design human work system interfaces where successful performance is dependent on complex cognitive processing and dynamic decision making. Questions of interest are whether some types of system interfaces might actually promote or enhance the potential for sub-optimal framing effects in judgment and choice tasks, and, if so, can these interfaces be designed to reduce operator framing effects? Secondly, can training or experience reduce or minimize framing effects and lead to better task performance? Finally are there predictor variables that can be developed to identify susceptibility to framing effects?

3. **Reference Points in Judgment and Decision Making.**

The dependence of the decision making process on **dynamic** components has, until recently, also been largely ignored in current utility-based models of decision making. There is, of course, a recognition in SEU theory and newer models that are also based on some aspect of the maximization of expected utility principle that the human decision maker does, in fact, learn from his or her past experiences and is affected by recent changes in his or her current state (e.g., gains and losses). This suggests that at any given point in time in a dynamic or sequential decision making task the decision maker must have a psychological ***reference point***. Research based on Kahneman and Tversky's (1979, 1992) prospect theory model, has shown that this reference point is critical; it defines the point at which, psychologically, perceived gains are separated from perceived losses.

Importance to Cognitive Engineering. Clearly, this separation or location of the reference point has been shown to have a strong influence on overt choices. However, this model of a dynamically changing psychological reference point has been developed and tested largely in the context of laboratory-based gambling scenarios. As such it too has a weak theoretical basis for describing how shifts in the reference point might occur in other types of decision making under uncertainty, time demands, or stress, and how such changes might affect performance in sequential decision making tasks or environments. These are the contexts of particular relevance to the field of cognitive engineering. Nevertheless, the extensive robustness of framing and reference point effects in the decision making literature suggests their

potential relevance to cognitive engineering and human/work system interactions. For example, an obvious extension of the framing literature might ask whether performances gains and setbacks or losses in an interactive decision making tasks can produce similar framing effects, and if so, how could feedback be used effectively to minimize sub-optimal behavior due to framing? Second, what are the mechanisms by which an operator might initially set or dynamically change his or her reference point? Third, can such reference point shifts be functions of positive or negative affective components of the work environment or machine interface? If negative, how can the system be designed to minimize changes in reference points and framing effects?

4. Loss-aversion and Individual Differences in Decision Making.

Recent studies in which individual utility functions have been derived empirically (Isen, Nygren, & Ashby, 1988; Nygren & Morera, 1988), indicate that large individual differences do occur in the subjective utilities that decision makers assign to outcomes and that these differences are reflected in individuals' utility curves -- primarily in the degree of ***loss-aversion*** shown in the functions. By loss-aversion we mean that losses are psychologically more aversive than are comparable gains. In gambling, the psychological discomfort felt in losing $10 relative the elation felt in winning $10 is typically much greater; in work, the psychological discomfort felt in poor performance relative the elation felt in good performance is typically much greater.

Although most individuals' utility functions do exhibit a common "S-shaped" concave/convex form (i.e., concave in the positive or gains end indicating risk-aversion and convex in the negative or losses end indicating risk-seeking), they vary significantly in the steepness of the losses end of the function relative to the gains end. Some people are much more loss-averse than others; that is, psychologically, losses are more aversive for some individuals than for others.

Importance to Cognitive Engineering. This suggests that loss-aversion is more than an observable and measurable property of a utility function. In a recent study, Nygren, Taylor, and Dulin (1992) found that loss-aversion, and not risk-aversion/risk-seeking, seemed to describe a relatively stable psychological trait that lead individuals to remain generally more (or less) sensitive to the negative affective impact of real or psychological losses across a wide range of decision making situations involving risk, uncertainty, or ambiguity. Loss-aversion was proposed to act primarily as an *affective* rather than *cognitive* influence on pre-choice valuation of subjective judgments

concerning the outcomes and likelihoods associated with the choice alternatives.

Alternatively, as Schoemaker (1993) notes, the work of Kahneman and Tversky (1979) and others on the framing effects described earlier has led to the predominant view in the behavioral decision theory literature that risk-seeking and risk-aversion tendencies are not affective in mature but are "mostly a function of the task, peoples' decision frames, and their information processing strategies, rather than a function of individual predispositions" (p. 52). Risk-taking and risk-aversion tendencies are the result of more context or task-specific cognitive influences, particularly on subjective weightings of likelihoods of relevant outcomes, while loss-aversion, then, is more of an intrinsic predispositional trait that can be observed and measured. An individual who has a strong propensity toward loss-aversion may, nevertheless, through the influence of cognitive framing, actually exhibit in different contexts either relatively strong risk-seeking or risk-averse behavior.

Hence, the study of loss-aversion and risk-aversion are important constructs in cognitive engineering research. Both loss-aversion and risk-aversion could strongly influence decision making behavior in dynamic decision tasks. These constructs are critical to linking both cognitive and affective components of decision making to man-machine interaction problems and need to be explored further. Of particular interest are individual differences questions about the role of propensity for loss-aversion or risk-aversion/risk seeking as an influence on decision making in complex work environments.

5. Linking Emotional and Cognitive Components of Decision Making.

The previous discussion of loss-aversion and risk-aversion suggest an affective-cognitive duality in human decision making that has been largely unexplored. Some other affective-cognitive linkages, however, have been proposed and are relevant to cognitive engineering as well. For example, Zajonc and his colleagues (1980; Zajonc & Marcus, 1982) have suggested that decision representations are typically encoded by decision makers both affectively and cognitively, and that these encodings may be largely independent of one another.

Support for this idea comes from compelling evidence that decision making processes can be influenced by such variables as one's affective mood state. For example, Isen, Nygren, and Ashby (1988) have shown that a positive mood state can lead decision makers to exhibit conservatism in risky choice situations. They become overly sensitive to potential

losses and make decisions in such a way as to avoid losses. People in negative moods also exhibit a cautious shift toward risk-aversion in their actual choices, but they do so apparently through a different mechanism. These individuals tend, when evaluating alternatives, to focus almost exclusively on negative outcomes and give them disproportionately more weight in the decision making process.

A growing body of literature suggests that a positive or negative mood state can differentially influence other components of the cognitive processes associated with decision making under risk or uncertainty as well. For example, subjects in a positive mood have been found to make significant over-estimations of likelihoods of positive events and under-estimations of negative events (Mayer, Gaschke, Braverman, & Evans, 1992; Wright & Bower, 1992) and subjects in a negative mood have been found to show just the opposite trend (Johnson & Tversky, 1983; Mayer et al., 1992; Wright & Bower, 1992). Finally, evidence is also strong that a positive affective state influences risk-taking attitudes as well as actual choice behavior (Isen, Nygren & Ashby, 1988; Isen & Geva, 1987; Isen and Patrick, 1983).

Importance to Cognitive Engineering. Despite these and other consistent findings on the influence of affect on decision making, it is still not clear which components of either the pre-choice or overt choice processes are actually being influenced or altered by mood state in a risky decision making task. For example, it is not clear whether the increased conservatism in betting behavior often shown by positive affect subjects is attributable to an increased sensitivity to either risk-aversion (i.e., to a greater weighting of perceived probabilities of losses relative to probabilities of gains) or to loss-aversion (i.e., to a greater weighting of perceived utilities of potential losses relative to gains), or both.

Positive affect may also impact on the subjective utilities of perceived gains and losses differently. Isen et al. (1988) found that in a risky decision situation where real loss was possible (subjects were gambling with their credit hour for participating in the experiment), positive affect, in comparison with a control state, was associated with an increased tendency to avoid losses.

An obvious issue relevant to cognitive engineering research is whether these findings for effects of mood state on a gambling task generalize to performance concerns in more complex dynamic decision tasks. One would expect so as long as the task was one involving significant perceived loss or risk. For example, we might ask whether, as possible loss in performance on a task increases, would a positive

affective state accentuate an aversion to choosing "riskier" choice alternatives or courses of action? Would operators in negative affective states focus more attention on losses or inflated probabilities of potential losses or on consequences of poor performance than is warranted? Furthermore, could a change in mood tunnel an individual operator's attention may be almost exclusively focused on losses as the Isen et al. (1988) study suggests with potential impact on performance? Would such behavior negatively impact on the operator's situation awareness of the work environment? Finally, a comparable set of concerns could be raised about the effects of negative affect caused by natural mood, mental or physical workload, psychological stress, or environment. All of these concerns are important to incorporating decision process in cognitive engineering work.

6. Anchoring and Adjustment of Probability Weights in Decision Making.

A recently proposed model that incorporates the influence of a loss-aversion trait in the pre-choice judgment process is an extension of Hogarth and Einhorn's (1989, 1992) venture theory. In venture theory, a decision maker is assumed to form subjective probability-like *decision weights* by first establishing some initial anchoring estimates about the perceived likelihood of occurrence of relevant events and then mentally adjusting those estimates either upward or downward. Hogarth and Einhorn (1989) argued that the net adjustment made to a probability anchor for an event ***A*** will be positive if more values above the anchor are recalled, mentally evaluated, or "weighted in imagination" than those below it. Mentally evaluating more values below the anchor should lead to the opposite result, a negative net adjustment.

Hogarth and Einhorn (1989) suggested that for most of us there is, in fact, a generally greater affective strength for our experiences of losses or bad outcomes than there is for gains or good outcomes. Hence, probability estimates *greater* than an anchor should be *more likely* to be imagined and weighted in the adjustment process for losses for most individuals, and *less likely* to be imagined and weighted in the adjustment process for gains. The end result of this anchoring and adjustment process for most individuals is expected to be two distinct decision weight functions, one which regularly overweights probabilities for perceived psychological losses or "bad" outcomes and one that regularly underweights these probabilities for psychological gains or "good" outcomes.

Importance to Cognitive Engineering. This duality of probability

weighting was made explicit in Hogarth and Einhorn's 1989) venture theory and has now been incorporated into the newest version of prospect theory as well (Tversky & Kahneman, 1992). If individuals do in fact have a tendency to judge and evaluate the likelihoods of "good" or positive outcomes and "bad" or negative outcomes differently, then such behavior could have very important implications for the design of work environments where probability assessment is a critical component. Of interest is whether critical events can be presented to the human operator in such a way as to reduce or eliminate good or bad connotations. Also of relevance is whether depersonalizing likelihood judgments has an impact on judgment. For example, some of our current work suggests that when individuals make impersonal likelihood judgments they are less prone to an overconfidence bias than when the judgment is a personal assessment. A continuation of this line of research in more complex work environments would be important to the field of cognitive engineering.

7. **Workload and Decision Making.**

The final topic concerns issues of workload and decision making. Human factors and ergonomics researchers have recognized for some time the increasing importance of understanding the role of the construct of ***mental workload*** in man-machine research. Current models of mental workload suggest that it is a multidimensional and complex construct, but it is one that has proved difficult to precisely define. Because of this difficulty, emphasis has often been placed on using direct reports through subjective measures such as rating scales to assess levels of mental workload. The Subjective Workload Assessment Technique, SWAT, (Reid and Nygren, 1988) was developed to provide a multidimensional model of mental workload that is based on an additive representation of three dimensions -- perceived Time load, Mental Effort load, and Psychological Stress load. An aspect of SWAT that makes it uniquely different from other subjective measures of workload is that it is based on an explicitly testable psychological model of how judgments of mental workload are formed and on a sound formal measurement system, conjoint measurement theory, that allows for meaningful interval scales. The NASA Task Load Index (NASA/TLX, Hart and Staveland, 1988) has also been shown to be a highly reliable and sensitive measure of perceived mental workload.

The empirical use of subjective workload measures has largely been to provide estimates of the ***cognitive*** components of the actual mental workload required for a task. However, some recent research (Nygren,

1990) suggests that these measures may, in fact, have potential in accurately assessing the **affective** components of workload as well. If we assume that subjective measures of workload actually assess the affective components of workload as well as the cognitive components, then these measures could prove useful in relating workload to decision making behavior like risk-taking tendencies. For example, SWAT has dimensional components that would allow one to measure the affective nature of an individual's subjective experience of workload, either positive or negative. SWAT can also be used just as easily to establish individuals' preferences for workload situations as it can for assessing their actual experiences of workload. One might expect that both experiences of workload levels and preferences for ideal workload levels should be important aspects of the complex relation between workload, decision behavior, and performance.

Importance to Cognitive Engineering. Thus, one can argue that subjective workload measures should be explored as potential indicators of affective biases that may actually influence changes in the dynamic decision making strategies used by individuals in low and high workload environments, and individual workload preferences and tolerances for the different workload factors that may affect actual operator performance. For example, TLX and SWAT may be more likely to assess the positive and negative feelings associated with varying workload levels, which in turn may potentially influence the decision making behavior that directly bears on performance and safety issues. Pilots, for example, are often called upon to complete many complex tasks that are high in mental effort, stress, and frustration, and that have significant dynamic decision making components -- often ones that involve risk as well.

There has been little systematic research investigating the potential relationship between changes in workload and decision making behavior, particularly risk-taking behavior. A major component of incorporating decision making in cognitive engineering research ought to be a systematic examination of the relationship between perceived workload and decision making behavior. If subjective measures of workload actually assess the affective components of workload as well as cognitive components, then these measures should prove useful in relating workload to a number of decision making issues raised here including: use of heuristics, framing and reference points, positive and negative affect, stress and frustration, probability assessment, and risk-taking/risk-avoidance.

Conclusions: Decision Making Research in Cognitive Engineering.

The purpose of this paper was to begin to tie some current work in the field of judgment and decision making to the emerging field of cognitive engineering. Several subareas of J/DM research were examined. A recurring theme of the paper is the need for J/DM research to begin looking at such relevant issues as decision bias, framing and reference points, affect and stress, risk-aversion and loss-aversion, probability estimation, and personality and individual differences in new ways. We need to get J/DM research out of the framework of gambling and similar static decision making contexts and into the framework of more complex dynamic decision making contexts with real-time constraints. Such tasks although more complex and difficult to study than gambling tasks, are, with the aid of microcomputers, now quite possible to design. Studies of this type offer a great potential for linking the study of human decision making to the cognitive engineering and use of efficient and effective computer-based work systems.

Decision making is and will continue to remain a critical component of the cognitive engineering of efficient and effective man-machine systems. Fortunately, over the past twenty years there has been an explosion of judgment and decision research that has added significantly to our knowledge base in cognitive psychology. It is now possible and important to utilize this work and build on it in new ways in our efforts to create and design better work environments.

References

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

Bourne, L. E., Dominowski, R. L., and Loftus, E. F. (1979). ***Cognitive processes***. Englewood Cliffs, N. J.: Prentice-Hall.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. ***Human Factors, 4,*** 59-73.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), ***Formal representation of human judgment***. New York: Wiley, pp. 17-52.

Edwards, W. (1990). Unfinished tasks: A research agenda for behavioral decision theory. In R. Hogarth (Ed.), ***Insights in decision making: A tribute to Hillel J. Einhorn***. Chicago: University of Chicago Press, pp. 44-65.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. ***Journal of Experimental Psychology: Human Perception and Performance, 1,*** 288-299.

Fischhoff, B. (1977). Perceived informativeness of facts. ***Journal of Experimental Psychology : Human Perception and Performance, 3,*** 349-358.

Fischhoff, B., & Slovic, P. (1980). A little learning ...: Confidence in multicue judgment tasks. In R. E. Nickerson (Ed.), ***Attention and performance (Vol. 8)***, pp. 779-800. Hillsdale, NJ: Erlbaum.

Hart, S. G. and Staveland, L. (1988). Development of the NASA Task Load Index (TLX): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) ***Human mental workload*** pp. 139-183). Amsterdam: North Holland.

Hershey, J.C. & Shoemaker, P. J. H. (1980). Prospect theory's reflection hypothesis: A critical examination. ***Organizational Behavior and Human Performance***, 25, 395-418.

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

Hogarth, R. M., & Einhorn, H. (1990). Venture theory: A model of decision weights. ***Management Science, 36,*** 780-803.

Hogarth, R. M., & Einhorn, H. (1992). Order effects in belief updating: The belief-adjustment model. ***Cognitive Psychology, 24,*** 1-55.

Isen, A. M., & Geva, N. (1987). The influence of positive affect on acceptable level of risk and thoughts about losing: The person

with a large canoe has a large worry. ***Organizational Behavior and Human Decision Processes, 39,*** 145-154.

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

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

Johnson, E.J., & Tversky, A. (1983). Affect, generalization, and the perceptions of risk. ***Journal of Personality and Social Psychology,*** 45, 20-31.

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

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

Kahneman, D., and Tversky, A. (1979). Prospect Theory: An Analysis of Decisions Under Risk. ***Econometrika, 47,*** 263-291.

Mayer, J., Gaschke, Y., Braverman, D., & Evans, T. (1992). Mood-congruent judgment is a general effect. ***Journal of Personality and Social Psychology,*** 63, 119-132.

Norman, D. A., and Rumelhart, D. E. (1975). ***Explorations in cognition.*** San Francisco: W. H. Freeman.

Nygren, T. E. (1990). Assessing individual differences in perceived mental workload. Paper presented at the 62nd Annual Meeting of the Midwestern Psychological Association, Chicago, IL, May, 1990.

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

Nygren, T. E., Taylor, P. J., & Dulin, J. (1992). How does positive mood alter risky decision making behavior?. Paper presented at the Society for Judgment and Decision Making meetings, St.Louis, MO, November, 1992.

Redelmeier, S. L., Tversky, A. (1992). On the framing of multiple prospects. ***Psychological Science, 3,*** 191-193.

Reid, G. B., and Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.) ***Human mental workload*** (pp. 185-218). New York: North Holland.

Shoemaker, P. J. H. (1993). Determinants of risk-taking: Behavioral and economic views. ***Journal of Risk and Uncertainty***, 6, 49-73.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1980). Facts and fears: Understanding perceived risk. In R. Schwing & W. A. Albers, Jr. (Eds.), ***Societal risk assessment: How safe is safe enough?*** pp. 181-216). New York: Plenum.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Response mode, framing, and information-processing effects in risk assessment. In R. Hogarth (Ed.), ***New directions for methodology of social and behavioral science: No.11. Question framing and response consistency*** (pp. 21-36). San Francisco: Jossey-Bass. (11).

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. ***American Economic Review, 73,*** 596-605.

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

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

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

Tversky, A. & Kahneman, D. (1988). Rational choice and the framing of decisions. In D.E. Bell, H. Raiffa & A. Tversky (Eds.), ***Decision making.*** New York: Cambridge University Press.

Tversky, A. & Kahneman, D. (1991). Loss aversion in riskless choice: A reference dependent model. ***Quarterly Journal of Economics,*** 107(4), 1039-1061.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. ***Journal of Risk and Uncertainty, 5,*** 297-323.

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessments. ***Organizational Behavior and Human Decision Processes, 52,*** 276-291.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. ***American Psychologist, 35,*** 151-175.

Zajonc, R. B., and Markus, H. (1982). Affective and cognitive factors in preferences. ***Journal of Consumer Research, 9,*** 123-131.

MODELING OF THE EJECTION PROCESS

A. G. Ramm
Professor
Department of Mathematics

Kansas State University
Cardwell Hall
Manhattan KS 66506-2602

Final Report For:
Summer Research Extension Program
Armstrong Laboratory

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

and

Kansas State University

July 1993

MODELING OF THE EJECTION PROCESS

A. G. Ramm
Department of Mathematics
Kansas State University

Abstract

A pilot is ejected from the aircraft. The equations modeling this process are written and solved analytically. Recommendations are given for the safe ejection on the basis of the above analysis. Suggestions are made concerning the choice of the optimal parameters for the safe ejection.

Modeling of the Ejection Process

I. Introduction

Consider a pilot in an aircraft of length L moving with constant velocity V at a certain constant height. This means that no vertical movement of the aircraft takes place (or, if it occurs it is assumed to be negligible). The pilot is rigidly fixed in a seat. The upper part of the body of the pilot is assumed to be connected with the lower part of the body (which is fixed in the seat) by an elastic-damping mechanism. Physiologically, this connection is done by the spine of the pilot. An ejection process we model consists of the following: an acceleration $a(t)$, where t is time, is applied vertically to the pilot's seat during the time period $[0,t_0]$, $a(t)=0$ for $t>t_0$. As a result, the pilot and the seat are moving vertically during the period $[0,t_0]$ and afterwards their vertical motion is governed by the gravitational acceleration and the initial conditions at $t=t_0$. Their horizontal movement is governed by the initial velocity at $t=0$ and the air-resistance force. The condition for the safe ejection of the pilot is that by the time the pilot moves horizontally to the position at which the tail of the aircraft is, the height at which the pilot's seat is should exceed the height h of the tail of the aircraft. The other condition we are interested in, is the condition which will allow to choose acceleration a(t) in such a way that the maximal deformation of the pilot's spine is minimized. This second condition leads to physiologically optimal conditions of the ejection process. In section II the basic notations are introduced and the governing equations are derived and solved. In section III the limiting case of the very short-time acceleration pulse is discussed. In

section IV the conditions for the safe ejection are obtained. In section V the maximal drift time of the pilot is calculated. In section VI the results are summarized.

There is an extensive literature of the subject [1-5].

II. Basic Equations and Their Solutions

1. Let us fix the notations:

m_1 - Mass of the upper part of the pilot's body

m_0 - Mass of the lower part of the pilot's body and the seat

$x_1(t)$ - Vertical coordinate of m_1 (say the coordinate of the shoulders of the pilot)

$x_0(t)$ - Vertical coordinate of the pilot's seat

l - The length of the pilot's spine at rest

$\Delta(t) := x_1(t) - x_0(t) - l,\ \Delta(0) = 0$

t - Time

$a(t)$ - Vertical acceleration applied to the seat during the time $[0, to]$, $a(t)=0$ for $t>to$, $a(t)\geq 0$, $a(0)=0$

k, c - Elastic and damping parameters of the pilot's spine

$\beta_j = \frac{c}{m_j},\ \omega_j^2 = \frac{k}{m_j},\ \alpha_j = \frac{\beta_j}{2},\ j=0,1$

$\varpi^2 = \omega_0^2 + \omega_1^2$

$\bar{\beta} = \beta_0 + \beta_1$

$\bar{\alpha} = \frac{\bar{\beta}}{2}$

$\Omega_1 = (\omega_1^2 - \alpha_1^2)^{\frac{1}{2}}$

$\Omega = (\omega^2 - \tilde{\alpha}^2)^{\frac{1}{2}}$

$\dot{x} = \frac{dx}{dt}$

t_s - Time for safe ejection

t_c - Critical time

t_0 - Time after which the acceleration $a(t)$ vanishes

b - The coefficient in the drift equation

A_s - The threshold

2. Let us derive the basic equations for $0 \le t \le t_0$. The equation of motion of the seat for $0 < t < t_0$ is

$$\ddot{x}_0 = a(t),\ x_0(0) = \dot{x}_0(0) = 0,\ 0 < t < t_0 \tag{1}$$

Therefore

$$x_0(t) = \int_0^t (t-\tau)a(\tau)d\tau,\ 0 < t < t_0 \tag{2}$$

$$x_0(t_0) = \int_0^{t_0} (t_0-\tau)a(\tau)d\tau := x_{00} \tag{3}$$

$$\dot{x}_0(t_0) = v_0(t_0) = \int_0^{t_0} a(\tau)d\tau := x_{01} \tag{4}$$

The equation of motion of m_1 (the upper part of the pilot's body) is

$$m_1\ddot{x}_1 = -m_1 g - c[\dot{x}_1(t) - \dot{x}_0(t)] - k[x_1(t) - x_0(t) - l] \tag{5}$$

$$x_1(0) = l,\ \dot{x}_1(0) = 0 \tag{6}$$

The term $-m_1 g$ is the gravitational force, the term

$-k[x_1(t) - x_0(t) - l] = -k\Delta(t)$ is the elastic force, the term

$-c[\dot{x}_1(t) - \dot{x}_0(t)] = -c\dot{\Delta}(t)$ is the damping force. For $\Delta(t)$ the problem (5)-(6) takes the form

$$\ddot{\Delta} + \beta_1\dot{\Delta} + \omega_1^2\Delta = -a(t) - g,\ \Delta(0) = \dot{\Delta}(0) = 0 \tag{7}$$

This problem can be solved explicitly:

$$\Delta(t) = -\frac{g}{\alpha_1^2 + \Omega_1^2} + \frac{g\exp(-\alpha_1 t)[\Omega_1\cos(\Omega_1 t) + \alpha_1\sin(\Omega_1 t)]}{\Omega_1(\Omega_1^2 + \alpha_1^2)} \tag{8}$$

$$-\int_0^t \exp[-\alpha_1(t-\tau)]\frac{\sin[\Omega_1(t-\tau)]}{\Omega_1}a(\tau)d\tau,\ 0 \le t \le t_0$$

Note that

$$\int_0^t e^{-\alpha_1 y}\sin(\Omega_1 y)dy = \frac{\Omega_1 - e^{-\alpha_1 t}[\Omega_1\cos(\Omega_1 t) + \alpha_1\sin(\Omega_1 t)]}{\alpha_1^2 + \Omega_1^2} \tag{9}$$

Using (8) one calculates

$$\Delta_0 = \Delta(t_0) = -\frac{g}{\Omega_1}\int_0^{t_0} e^{-\alpha_1 y}\sin(\Omega_1 y)dy$$
$$-\frac{1}{\Omega_1}\int_0^{t_0} e^{-\alpha_1(t_0-\tau)}\sin[\Omega_1(t_0-\tau)]a(\tau)d\tau, \tag{10}$$

where the first integral in (10) is calculated in (9), and

$$\Delta_1 = \dot{\Delta}(t_0) = -\frac{g}{\Omega} e^{-\alpha_1 t_0} \sin(\Omega_1 t_0) - \frac{1}{\Omega_1} \int_0^{t_0} \{-\alpha_1 e^{-\alpha_1 (t_0-\tau)} \sin[\Omega_1 (t_0-\tau)] + \Omega_1 e^{-\alpha_1 (t_0-\tau)} \cos[\Omega_1 (t_0-\tau)]\} a(\tau) d\tau \tag{11}$$

Let us derive the basic equations for $t>t_0$. If $t>t_0$ then $a(t)=0$ and the equations of motion are

$$\ddot{x}_1 + \beta_1 [\dot{x}_1(t) - \dot{x}_0(t)] + \omega_1^2 [x_1(t) - x_0(t) - l] = -g \tag{12}$$

$$\ddot{x}_0 - \beta_0 [\dot{x}_1(t) - \dot{x}_0(t)] - \omega_0^2 [x_1(t) - x_0(t) - l] = -g \tag{13}$$

Subtract (13) from (12) to get

$$\ddot{\Delta}(t) + \tilde{\beta} \dot{\Delta}(t) + \tilde{\omega}^2 \Delta(t) = 0 \tag{14}$$

$$\Delta(t_0) = \Delta_0 \,, \quad \dot{\Delta}(t_0) = \Delta_1 \tag{15}$$

where the initial values Δ_0 and Δ_1 are given in (10) and (11). The solution to (14), (15) is

$$\Delta(t) = e^{-\tilde{\alpha}\tau} [\Delta_0 \cos(\tilde{\Omega}\tau) + (\Delta_1 + \tilde{\alpha}\Delta_0) \frac{\sin(\tilde{\Omega}\tau)}{\tilde{\Omega}}] \tag{16}$$

where

$$\tau := t - t_0 \geq 0 \tag{17}$$

The function $x_0(t)$ for $t>t_0$ can be found from equation (13). Let us write the equation as

$$\ddot{x}_0 = -g + \beta_0 \dot{\Delta}(t) + \omega_0^2 \Delta(t) := f(t), \quad t > t_0 \tag{18}$$

where $\Delta(t)$ is given by the formula (16) for $t \geq 0$, so that the function $f(t)$ in (18) is known explicitly. One also has (see formulas (3), (4)) $x_0(t_0) = x_{00}$ and $\dot{x}(t_0) = x_{01}$. Solving equation (18) with the initial conditions (3), (4) one gets

$$x_0(t) = x_{00} + \tau x_{01} + \int_0^\tau (\tau - s) f(s) ds, \quad \tau = t - t_0 \geq 0 \tag{19}$$

III. Pulse-type acceleration

Let us discuss the case $t_0 \ll 1$, that is, accelerations acts during a very short time. Assume, in the limiting case, that $a(t) = a_0 \delta(t)$, where $\delta(t)$ is the delta-type function, $\int_0^\epsilon \delta(t) dt = 1$ for any $\epsilon > 0$, $\int_\epsilon^b \delta(t) dt = 0$ for any $0 < \epsilon < b$, $a_0 = const > 0$. In this case to $t_0 \ll 1$, and formulas (3), (4) yield

$$x_{00} = 0, \quad x_{01} = a_0 \tag{20}$$

while formulas (10) and (11) yield

$$\Delta_0 = 0, \quad \Delta_1 = -a_0 \tag{21}$$

Formulas (16) and (21) yield:

$$\Delta(t) = a_0 e^{-\varepsilon\tau} \frac{\sin(\tilde{\Omega}\tau)}{\tilde{a}}, \quad \tau := t - t_0 \geq 0 \tag{22}$$

Thus, since $t_0 \approx 0$,

$$|\Delta(t)| = a_0 e^{-\alpha t} \left| \frac{\sin(\tilde{\Omega} t)}{\tilde{\Omega}} \right|, \quad t>0, \quad t_0 \approx 0 \tag{23}$$

It is clear from this analysis that for $t_0 \ll 1$ the function $|\Delta(t)|$ is determined basically by the initial velocity x_{01} at the time t_0. If $0 \le t_s \le t_0$, $t_0 \ll 1$, then $|\Delta(t)|$ is monotonically growing on the interval $0 \le t \le t_s$. Therefore $\max_{t>0} |\Delta(t)|$ is attained for $t > t_0$.

IV. Conditions for Safe Ejection

1. The first condition for the safe ejection is given by the inequality

$$x_0(t_s) > h, \quad 0 < t_s < t_c \tag{24}$$

where t_s is the safe ejection time, and t_0, the critical time, is calculated in section V. This is the time in which the pilot moves in the horizontal direction during the ejection process to the tail of the aircraft. This drift of the pilot is due to the air resistance. Clearly the condition (24) should be satisfied, otherwise the tail of the aircraft might hit the pilot.

Consider two cases. We always assume $t_0 < t_c$. First, assume that that $0 < t_s \le t_0$. In this case condition (24) and formula (2) yield

$$\int_0^{t_s} (t_s - \tau) a(\tau) d\tau > h \tag{25}$$

Since $a(\tau) \ge 0$, the left-hand side of (25) is monotonically increasing function of t_s.

Let $a(t) = Aw(t)$, where A = const and $w(t)$ is the normalized acceleration which is a continuous function satisfying the following conditions:

$$0 \le w(t) \le 1 \text{ (normalization condition)} \tag{26}$$

and

$$w(0) = 0 \tag{27}$$

Condition (26) means that

$$A = \max_{t>0} a(t) \tag{28}$$

Condition (25) is satisfied if and only if

$$A > \frac{h}{\int_0^{t_s} (t_s - \tau) w(\tau) d\tau}, \quad t_s \le t_0 \tag{29}$$

Let us discuss the following question: how does one find among all $w(t)$ satisfying restrictions (26) and (27) the one for which A is minimal and the safe ejection condition (29) is satisfied? What is A_{min} the minimal A? To answer these questions let us note that A_{min} equals to

$$A_{min} = h(\sup_w \int_0^{t_s} (t_s - \tau) w d\tau)^{-1} \tag{30}$$

Since w satisfies conditions (26) and (27) one has

$$\sup_w \int_0^{t_s} (t_s - \tau) w(\tau) d\tau = \int_0^{t_s} (t_s - \tau) d\tau = \frac{t_s^2}{2} \tag{31}$$

This supremum is not attained in the class of continuous functions satisfying the restrictions (26), (27): the function $w(t) = 1$ does not belong to this class. However, the function:

$$w_m(t)=\begin{cases}\frac{t}{t_m}, & 0\le t\le t_m\\ 1, & t_m\le t\le t_s-t_m'\\ \frac{1-t}{t_m'}, & t_s-t_m'\le t\le t_s\end{cases} \tag{32}$$

Where t_m and t_m' are small numbers, $0<t_m$, $t_m'\le t_s/2$ allows one to approach the supremum $\frac{t_s^2}{2}$ as close as one wishes by choosing t_m and t_m' sufficiently small. Indeed, a simple calculation (which is left to the reader) shows that in the case $t_m=t_m'$ one has

$$I_m=\int_0^{t_s}(t_s-\tau)w_m(\tau)d\tau=\frac{t_s^2}{2}-t_st_m+\frac{t_m}{2} \tag{33}$$

Therefore $I_m\to\frac{t_s^2}{2}$ if $t_m\to 0$, as claimed. The minimal value A_{min} is given by the formula

$$A_{min}=\frac{2h}{t_s^2} \tag{34}$$

Let us give for comparison the values of A for the following functions:

$w(t)=\sin\frac{\pi t}{t_s}$ and $w(t)=w_m(t)$ for $t_m=t_m'=\frac{t_s}{2}$. For the first function

$w = \sin\frac{\pi t}{t_s}$, a simple calculation yields $A = \frac{\pi h}{t_s^2}$. For the second function, which is a triangular pulse, one has $A = \frac{3h}{t_s^2}$.

One can see that the gain in choosing optimal A is about 50%.

2. If one wishes to find among all continuous functions $a(t)$, satisfying the conditions $a(t) \geq 0$, $a(0) = 0$, the one for which

$$\max_{t>0} |\Delta(t)| = \inf \tag{35}$$

One can use the results in section II to solve this optimization problem numerically. Namely, the function $\Delta(t)$ is calculated analytically by formulas (8) and (16) for all $t > 0$. Therefore, given a finite set of functions $a(t)$ one can calculate the graphs of the corresponding functions $\Delta(t)$, find $\max_{t>0} |\Delta(t)|$ and choose $a(t)$ for which this quantity is minimal.
Analytical solution of the minimization problem (35) does not seem to be feasible.

3. If $t_s > t_0$, then the basic condition (24) and formula (19) lead to the following condition for the safe ejection:

$$x_{00} + \tau_s' x_{01} + \int_0^{\tau_s'} (\tau_s' - \tau') f(\tau') d\tau' > h, \quad \tau_s' = t_s - t_0 \tag{36}$$

In (36) the constants x_{00} and x_{01} and the function $f(\tau')$ are analytically calculated by formulas (3), (4) and (18), given $a(t)$. If $a(t) = Aw(t)$ and $w(t)$ satisfies conditions (26) and (27) then condition (36) is satisfied if

$$A > A_s , \tag{37}$$

where A_s is calculated by the following method. Let us rewrite (36) as

$$x_{00} + \tau_s x_{01} + \int_0^{\tau_s} (\tau_s - \tau') f_1(\tau') d\tau' > g\frac{\tau_s^2}{2} + h \tag{38}$$

where (cf(18))

$$f_1(t) := f(t) + g \tag{39}$$

Note that x_{00}, x_{01} and f_1 (but not f) are linear functions of $a(t)$. Therefore

$$A_s := (\frac{g\tau_s^2}{2} + h)/B \tag{40}$$

where B is, by definition, the left-hand side of (38) calculated for $a(t) = w(t)$ (that is, for the normalized acceleration). Note that for $t_s > t_0$ the threshold A_s is not proportional to h (in contrast with the case $t_s < t_0$).

V. Calculation for the Critical Time

In this section the critical time t_c is calculated. Assume that the motion of the pilot in the horizontal direction is governed by the equation

$$\dot{v} = -bv^2, \; v(0) = V \tag{41}$$

where v is the aircraft velocity, $b > 0$ is the coefficient proportional to the area of the projection of the pilot onto the plane perpendicular to the direction of the horizontal velocity

v, the velocity of the drift of the pilot. Problem (41) is easy to solve analytically

$$v(t) = \frac{V}{1+bVt} \tag{42}$$

Let us calculate the critical time as the time needed for the pilot to travel the length L relative to the aircraft. It was explained in the introduction that the inequality $t_s < t_c$ is necessary for the pilot to clear the tail of the aircraft in the course of the ejection. The relative velocity of the pilot in the coordinate system fixed at the aircraft is

$$v - V = -\frac{bV^2t}{1+bVt} \tag{43}$$

Therefore the critical time t_c can be found from the equation

$$L = \int_0^{t_c} \frac{bV^2t\,dt}{1+bVt} = V\left[t_c - \frac{1}{bV}\ln(1+bVt_c)\right] \tag{44}$$

Define the time needed for the aircraft to travel distance L

$$T = \frac{L}{V} \tag{45}$$

Then t_c is the positive root of the equation

$$T = t_c - \frac{1}{bV}\ln(1+bVt_c) \tag{46}$$

This equation can be easily solved numerically by the Newton method. Denote

$$bVt_c = Z,\ bVT = Z_T \tag{47}$$

$$\psi(Z) := Z - \ln(1+Z) - Z_T = 0 \tag{48}$$

One can easily check that the unique positive root of the equation (48) lies on the interval $(Z_T, 2Z_T)$, and that the function $\psi(z)$ is monotonically growing and convex $(\psi'' > 0)$ on this interval. Therefore the unique root of the equation (48) can be found by the Newton method

$$Z_{n+1} = Z_n - \frac{\psi(Z_n)}{\psi'(Z_n)},\ Z_0 = 2Z_T \tag{49}$$

$$Z = \lim_{n \to \infty} Z_n \tag{50}$$

The scheme (49) can be written as

$$Z_{n+1} = Z_n - (1+Z_n)\left[1 - \frac{\ln(1+Z_n)}{Z_n}\right],\quad Z_0 = 2Z_T \tag{51}$$

The iterative process (51) converges rapidly and allows one to calculate Z by formula (50). If Z is found then t_c is easy to calculate

$$t_c = \frac{Z}{bV} \tag{52}$$

VI. Summary of the Results

Given $a(t)$ one calculates $x_0(t)$ and $\Delta(t)$ by formulas (2), (8), (15) and (19) and checks the conditions (25) or (36) for the safe ejection.

If $a(t)=A\,w(t)$ and $w(t)$ is the normalized acceleration, then the safe ejection is achieved for minimal amplitude A of acceleration if A is chosen from the conditions (29) or (37). If $t_s \le t_0$ then A_{min} is given by formula (34). A numerical procedure is suggested for finding $a(t)$ which solves optimization problem (35). The critical time t_c is calculated by formula (52). The safe ejection time t_s has to be less than t_c.

References

1. J. Brinkley, Personnel Protection Concepts for Advanced Escape System Design, Aerospace Medical Panel Symposium Paper, Williamsburg, VA, 30 April-4 May, 1984

2. M. Kornhauser, A. Gold, Application Of The Impact Sensitivity Method To Animate Structure, Nat. Res. Council, Broons AFB, November 27-29, 1961 conference paper.

3. F. Latham, A Study in Body Ballistics: Seat Ejection, Proc. Roy. Soc. London, Ser. B, 147, (1957), 121-39.

4. P. Payne, Injury Potential of Ejection Seat Cushions, Journal of Aircraft, 6, N3, (1969), 273-278.

5. S. Ruff, Brief Accelerations: Less Than One Second, German Aviation Medicine in WWII, vol 1, Chap VI-C, Dept. of The Air Force.

During my tenure at the Armstrong Laboratory I had many discussions with technical staff, gave consultations on several topics: computational aspects of the least squares method, diagonalization of the inertia tensor Perturbation Theory for inertia moments, stability of the motion of rigid bodies with respect to small perturbations of their inertia moments, stable calculations of derivatives and stable inversion of ill-conditioned matrices.

I have read a number of technical reports and papers in the area of safety of aircraft escape systems and commented on some of them.

DETERMINATION OF THE REDOX CAPACITY OF SOIL SEDIMENT BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson
Professor
and
Tashia V. Sullins
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
University of Georgia

September, 1993

DETERMINATION OF THE REDOX CAPACITY OF SOIL SEDIMENT BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson
Professor
and
Tashia V. Sullins
Graduate Student
Department of Chemistry
University of Georgia

Abstract

The oxidative redox capacity was determined for size-fractionated soil sediment samples by the method of spectroelectrochemical coulometric titration. This method involves the measurement of absorbance of sediment particle slurries at the wavelength absorption maxima 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. Titrations were carried out on diluted samples of gravitationally sedimented particle fractions containing particles smaller than 2 micrometers average diameter. The fraction containing particles of size < 2 micrometers was 0.115 % by weight of the initial sample slurry, which was 4.3 % solids by weight. The total organic content of the < 2 micromenter solids was 3.5 % organic carbon by weight. Titration was carried out at a diluted sediment particle concentration of 0.0128 % by weight. Resorufin was reduced first, followed by an irreversibly reducible sediment component which was consistently observed to titrate between resorufin and methyl viologen, and finally methyl viologen. The reducible component, which was absent from titration blanks, was not reoxidized when the methyl viologen and resorufin were electrochemically reoxidized. The sediment fraction studied had an oxidative redox capacity of 15 $\pm$ 2.5 millicoulombs, corresponding to 0.65 milliequivalents per gram of sediment. The heterogeneity of the original sample was evidenced by the observation that the whole sediment slurry became reducing, whereas the fractionated < 2 micrometer particle slurry remained oxidizing.

DETERMINATION OF THE REDOX CAPACITY OF SOIL SEDIMENT BY SPECTROELECTROCHEMICAL COULOMETRIC TITRATION

James L. Anderson

Tashia V. Sullins

Introduction

The redox capacity of a sediment is a measure of the number of electron equivalents which it may donate or accept to reduce or oxidize a substance. This has become a topic of interest recently because of the large number of groundwater and aquifer sites which have been contaminated by pollutants. There is a growing need to clean up and prevent further contamination in such sites. The redox capacity is considered because of the possibility of biological or chemical degradation of pollutants by the sediment(1). Barcelona and Holm claim that "oxidation-reduction processes were mediated by natural microbial populations that catalyze electron-transfer reactions"(1). While both biological and chemical processes appear to be involved in sediment redox processes, evidence suggests that many of these processes are catalytic. Relatively large concentrations of electron donors and acceptors may be present, but their reactions with the pollutant species are sluggish relative to the catalytically active species. In this manner, the catalytic species can be recycled between reduced and oxidized states many times, until the supply of electron donors or acceptors is exhausted. The extent of possible cleanup or other transformation of the pollutant is limited by the overall redox capacity, which determines how many times the catalytic species can be turned over. If the sediment is capable of naturally degrading the pollutant, the following questions then arise. 1) Is it a biological or chemical process? 2) If it is biological, what organisms are involved? 3) Is the sediment a reducing or oxidizing agent? and 4) How much can it reduce or oxidize? Of course all of these questions cannot be answered by one experiment, but by studying the redox capacity, the last two questions can be answered, and give a good indication of the course to take to answer the others. We present here the results of a study of the oxidative redox capacity of a size-fractionated sediment sample.

Methodology

The method of determining the redox capacity in this study was by spectroelectrochemical coulometric titration. This method is an attractive one because low micromolar concentrations of spectrally visible and invisible species can be studied, and reliable quantitative data can be obtained(2). In a conventional chemical titration, acids or bases are used as the titrant. In an electrochemical titration, electrons (measured by the charge passed) are used as the titrant. In this experiment, charge was added

in fixed increments at a specific applied potential for both reduction and oxidation steps. The reduction potential was -0.700V vs. a Ag/AgCl/1M KCl reference electrode with SnO_2 as the working electrode, and the oxidation potential was +1.25V with Pt wire as the working electrode. Charge increments ranged from 0.25 millicoulombs to 5 millicoulombs.

The progress of the titration was monitored spectrophotometrically after each addition of charge. Two mediator titrants were used to help monitor the progress of the titration. Resorufin (concentration 10 micromolar) was used primarily as a reporter molecule, and methyl viologen (concentration 0.5 millimolar) was also used as a primary titrant to drive the reaction. The solutions of resorufin and methyl viologen were prepared in Milli-Q deionized water and a pH 7 buffer solution of KH_2PO_4 and Na_2HPO_4 (ionic strength 0.1). The buffer also served as an electrolyte, and a pH of 7 was chosen to maintain a neutral solution. Both of these titrants exhibit different colors in their oxidized and reduced states. Resorufin is pink in the oxidized state and colorless in the reduced state, while methyl viologen is colorless in the oxidized state and a deep blue in the reduced state. Because both of these mediators can be spectrally detected, the absorbances of these substances can be studied during the titration. Resorufin exhibits an absorbance peak at 572 nm, and methyl viologen exhibits two major absorbance peaks at 396 and 600 nm. The titrations began with both species in the oxidized state. The negative applied potential of -0.700V was applied, and the reduction phase of the titration was begun. Once the reduction was complete (when the absorbance peak of resorufin is extinguished, and the peaks for the viologen appear), the applied potential was changed to +1.25 V and the oxidation begun. Oxidation was complete when the viologen peaks were extinguished and the resorufin peak appeared and was restored to its initial amplitude. Examples of spectra obtained during these titrations can be seen in Figures 1 and 2. Figure 1 represents the reduction step and Fig. 2 represents the oxidation step. "Blank" titrations involving only the resorufin and viologen (no sediment) were performed in order to determine the charge taken up by the resorufin and viologen, and to compare the titrations with and without sediment. The treatment of the sediment before the titration is explained in the following paragraph.

The sediment was collected from the Beaver Dam site in Athens, GA. The samples were collected in Mason jars, wet-sieved in air with the surface pond water by a 1 mm mesh brass sieve to remove large debris particles, and stored in sealed Mason jars. The sample to be studied was placed in an anaerobic glove box under N_2, CO_2, H_2 atmosphere where it was then fractionated by gravitational sedimentation to allow for the collection of particles smaller than 2 micrometers in diameter. The fractionation method was adopted from "Methods of Soil Analysis" according to the equation $t = 18\,\eta\,h / [g\,(\rho_s - \rho_l)x^2]$ where t = time, η = viscosity (poise, g cm^{-1} s^{-1}), h = height (cm), g = 980 cm/s^2 (acceleration due to gravity), ρ_s = 2.6 g/cm^3, ρ_l = 1.0 g/cm^3 (density of sediment particles and the aqueous medium), x = effective

particle diameter of the largest particle to clear (3,4). The settling time t was the variable solved in the equation in order to determine how long the sediment should settle to retain only particles of size <2 micrometer in the top 10 cm of the jar. This time was determined to be 8.73 hr.

Sample containers were kept closed except during transfer. After the fractionation, samples were collected from the top 10 cm of the jar by siphoning and placed into three smaller jars which were crimped with a butyl rubber stopper and an aluminum cap. Samples were stored in the glove box until ready for use. Final solution preparation was also carried out in the glovebox. For the titration, sediment samples from these smaller jars were diluted with Milli-Q water in a 1:9 sediment to water ratio. This ratio was chosen due to the extensive light scattering of the sediment particles. The diluted sediment was mixed in a 1:1 ratio with the resorufin and methyl viologen solution described earlier, transferred anaerobically to a degassing bulb and the cell on a nitrogen/vacuum train outside the glove box, degassed, and titrated.

The total organic carbon (TOC) content of < 2 micrometer sediment slurries was determined using a Shimadzu TOC-5000 Analyzer and ASI-5000 Autosampler. A TOC calibration curve was obtained with KHP standards before the sediment samples were determined. Samples included with three < 2 micrometer fractionated sediment samples were a Milli-Q water sample and a known 50 ppm total carbon sample of KHP. The water sample showed a total organic carbon value of 0 ppm, the 50 ppm KHP sample showed a total carbon value of 54.86 ppm, and the three sediment samples showed values of 44.73 ppm, 42.77 ppm, and 42.49 ppm (average value of 43.33 +/- 1.2 ppm) for the sediment samples corresponding to 3.5% of organic content referred to dry sediment weight.

The weight percent of solids was found to be 4.3% for the initial sediment slurry, and 0.115% for the fraction with particle size < 2 micrometers. These values were determined by measuring the mass of a 5 mL aliquot of undiluted sediment, placed in a previously weighed container, before and after overnight drying in an oven at 100 °C.

Samples were thoroughly degassed before the titration(to eliminate O_2) and to keep the solution as anaerobic as possible. This procedure will be explained in explicit detail in the apparatus section, because the procedure involves an explanation of the apparatus first.

Apparatus

The most important piece of apparatus is the vessel in which the titration takes place: the electrochemical cell. A diagram of the cell is shown as Figure 3(2). The main body and the electrodes are composed of Pyrex glass. The Pyrex valve at the top of the cell allows for filling and for closure after the cell has been

filled with solution. The main chamber of the cell excluding the liquid in the sidearms, and at the tip of the valve holds 1.864 ± 0.0016 mL of solution. The reference electrode holds 0.2712 ± 0.0003 mL, and the auxiliary electrode holds 0.4547 ± 0.0017 mL of solution. The magnetic stirrer is a ca. 7 mm long piece of steel paper clip encased in Pyrex glass and flame-sealed under vacuum. The Pt potentiometric electrode is fused inside the cell in order to serve as a potentiometric or working electrode; this electrode was used as the working electrode for the oxidation steps of the titration. The sidearms that house the electrodes are also composed of Pyrex and incorporate porous frits that connect to the main body of the cell. The working electrode for the reduction steps is a 2.5 cm square piece of SnO_2 glass that has been epoxied to the bottom of the main chamber of the cell with Devcon 2-Ton clear epoxy. The working electrode is also used to align the cell in a positioning recess in the optical train of the spectrophotometer. The reference and auxiliary electrodes are both made of Pyrex glass and join to the sidearms of the cell by ground-glass fittings(size 10/30). The electrodes are filled with 1M 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. Solution contact is made via porous Vycor frits epoxy-sealed into the ends of the 10/30 joints. Light was passed through the main chamber of the cell to determine the absorbance values. Because the cell is composed of Pyrex, light is detected mostly in the visible region of the spectrum.

A schematic of the overall apparatus is shown as Figure 4. The electrochemical cell previously described was placed in the spectrophotometer on top of a water driven magnetic stirrer and a plexiglass platform that was designed to clamp the working electrode of the cell into place in order for the light beam to pass through the main body of the cell. The water was circulated through a constant-temperature bath to help thermostat the cell. A momentary contact switch connected a high-impedance digital voltmeter to the cell when the potentials of the working and potentiometric electrodes were periodically measured during the course of the titration. The spectrophotometer was a Perkin-Elmer Lambda Array 3840 UV/VIS spectrophotometer and was interfaced to a Perkin-Elmer 7000 Series computer. Spectral data were converted to ASCII format and transferred to a Sun Sparcstation for data manipulation using MATLAB software. The cell was also attached to a BAS-CV1B potentiostat which was responsible for applying the potentials during the oxidation and reduction steps. In order to determine the charge applied in each step, an absolute value amplifier converted the potentiostat current output to a positive value and fed it to a voltage to frequency converter . This conversion device was constructed in the laboratory of the investigators (2). In order to count the frequency pulses, the converter was connected to an events counter which allowed the frequency to be converted to counts of charge in millicoulombs with a conversion factor of 100,000 counts per millicoulomb (Data Precision Model 5740).

The degassing procedure is described in the following paragraph. The electrochemical cell was attached by a ground glass fitting to a cell adapter which was then attached to the degassing assembly. The degassing assembly consisted of a Ridox catalyst, a nitrogen gas line, a water-filled bubbler to saturate the gas with water, and a vacuum line trapped with liquid nitrogen. The cell could be either pressurized with nitrogen or evacuated via a two-way valve on the degassing assembly. An O_2 trap was placed on the copper-tubing gas line between the nitrogen cylinder and the degassing assembly as an extra precaution against leakage. All pieces of the degassing assembly were joined by ground glass fittings greased with Apiezon N. The two-way valve was connected to a vacuum system which consisted of a liquid nitrogen trap, a vacuum flask, and a mechanical vacuum pump. These were connected by butyl rubber tubing. Liquid nitrogen was placed in the trap and the valves and pump turned on. The two-way valve was switched to evacuate the solution degassing bulb/cell adapter and the cell for at least 15 min. The cell was then purged alternately with N_2 and the vacuum again for five cycles of 1-2 min. per cycle. The cell was then left under N_2 , purged for 2 min. and the cell stopcock closed off. The degassing bulb was placed on vacuum for 2 min. and closed off as well. A previously prepared solution of 1M KCl was degassed for 15 min. by passing N_2 from the degassing assembly through a ground glass adapter made for that purpose. Both the reference and auxiliary electrodes were degassed via a needle inserted through the septum cap while attached to the cell using this same adapter. They were evacuated and filled with nitrogen alternately for 2 min. for 5 cycles. After the electrodes had been degassed, they were filled with the KCl solution by a 1 mL microliter syringe. If a bubble appeared in the electrode, the electrode was emptied by the vacuum side and taken again through the procedure just described. Once the electrodes were filled, the solution to be studied was introduced into the degassing bulb. The amount of solution placed in the degassing bulb was 5.46 mL(half being diluted sediment, the other half being the resorufin/viologen solution). Because the degassing bulb had been placed on vacuum before closing it off, it was possible to place the solution in the side arm of the degassing bulb, and open the valve in air to transfer the solution into the bulb. The valve on the degassing bulb was then closed and connected once again to the two-way valve and the cell. A magnetic stirrer placed beneath the degassing bulb activated a magnetic stirrer in the latter's solution chamber to facilitate degassing. The two-way valve of the degassing assembly was then placed on vacuum for 10 min. and the solution was degassed before putting it into the cell. To put the solution into the cell, the valve between the degassing bulb and the cell was opened and the stirrer removed from beneath the bulb. The cell and degassing bulb were tilted downward and the cell was filled by purging both units with N_2. To minimize bubble formation, the vacuum was applied for 5 sec. and the apparatus purged with N_2 again. This procedure was continued until no bubbles were visible inside the cell.

Results/Discussion

After the titrations were performed, plots were constructed using the absorbance data for peaks at 572 nm and 396 nm versus the total charge needed for the completion of the titration. Examples of these plots are shown as Figs. 5-8. The plot of the absorbance at 572 nm versus the total charge represents the titration of the resorufin while the plot of absorbance at 396 nm represents the titration of the methyl viologen. Figure 5 represents the reduction of a "blank" (no sediment). The applied potential was -0.700V and the working electrode was SnO_2. This plot shows that the blank took up 13.5 millicoulombs of charge before the absorbance of resorufin was extinguished. A significant fraction of this charge appears to be due to residual oxygen. This plot also shows that once the resorufin was reduced, the viologen began to reduce with no significant break in between the two. Figure 6 represents the oxidation of the same "blank" solution. Here, the Pt wire was used as the working electrode with an applied potential of +1.25V. This plot shows that it took approximately 1.75 millicoulombs to oxidize the methyl viologen, but once the viologen was oxidized, the absorbance for resorufin began increasing. Figures 7 and 8 represent the titration of the sediment and the mediators. Fig. 7 represents the reduction step and Fig. 8 represents the oxidation step. The reduction plot of the sediment looks very different from the reduction plot of the blank. This suggests that something in the sediment is being reduced. In the reduction of the sediment, approximately 15 millicoulombs is taken up in the beginning of the titration. This can be explained by residual O_2 which most likely occurred as a result of insufficient degassing. A similar quantity of charge was observed for the blank. It can also be seen, however, that the charge consumed during reduction of resorufin is increased relative to the blank, and 15 millicoulombs is taken up between the reduction of resorufin and the reduction of methyl viologen. This significant charge uptake between titration of resorufin and methyl viologen is attributable to a component of the sediment, since this region is absent from a blank titration plot. Based on the relatively sharp transitions of the plots at the end of the titration of resorufin [$E^{\circ'}$ = - 40mV vs. normal hydrogen electrode (NHE) at pH 7], and the beginning of the titration of methyl viologen ($E^{\circ'}$ = - 446 mV vs. NHE), it appears that the nominal reduction potential of the titrable sediment component lies approximately midway between these values. The oxidation of the sediment looks much like the oxidation of the "blank". Two millicoulombs were used to oxidize the viologen, and at that point, the peak for resorufin began to appear. This implies that the component in the sediment that was reduced was irreversibly reduced. To reinforce this idea, a second reduction was performed on the sample after the oxidation. This reduction step proved to look just like the reduction step for the "blank". Therefore, whatever component was reduced in the sediment was reduced only once and did not oxidize or rereduce. Figures 9-13 are three-dimensional representations of the data of Figs. 5-8.

The three dimensional plots were constructed to see the relationship between the three variables used for the titrations (absorbance, wavelength (nm),and charge (mcoul)). In Fig. 9 (reduction of the blank), the peak at 572 nm is reduced with the addition of charge. Once this peak is completely reduced, the peaks at 396 and 600 nm begin to appear representing the viologen. In comparison, Fig. 11 (reduction of the sediment) looks different because there is a space of about 15 millicoulombs between the reduction of the resorufin and the appearance of the viologen. Figures 10,12, and 13, corresponding to oxidation of the blank, oxidation of the sediment, and rereduction of the sediment respectively, exhibit behavior analogous to the blank of the reduction. This implies that very little or no charge was taken up by the sediment during the oxidation or rereduction process.

A preliminary examination of the heterogeneity of the sample was carried out in an anaerobic glove box by adding an aliquot of the resorufin/methyl viologen/pH7 phosphate buffer medium to an aliquot of the original sediment slurry after storage for several weeks following removal of the < 2 micrometer particle fraction. The volumetric proportions were comparable to those previously used for the < 2 micrometer fraction. The resorufin was rapidly decolorized in the whole sediment fraction, whereas the resorufin retained its pink color when added under identical conditions to the < 2 micrometer fraction. Numerous factors may contribute to this observation. First, the total content of solids of the whole sediment was much greater than the solids content of the < 2 micrometer fraction. However, it is possible that the active species in the sediment may preferentially accumulate in particles of certain size. Further studies are being initiated to assess the origin of this particle size-based differentiation.

Conclusions

In the studies of the redox capacity of the < 2 micrometer particle size fraction of Beaver Dam sediment by spectroelectrochemical titration, it was determined that the sediment had an oxidative redox capacity of 15 ± 2.5 millicoulombs for a sample that was 0.118 % solids by weight in the sample diluted to final concentration in a titration volume of 1.864 mL. It was not a reversible process because the component could not be oxidized or rereduced. Barcelona and Holm reported in their study that oxidative capacities averaged on the order of 0.3 to 0.4 mequiv/g of solids(1). The results in this report indicate an oxidative redox capacity of the < 2 micrometer particle sediment to be 0.65 mequiv/g of solids. The differences in these two numbers possibly arise from different types of sediment studied. Barcelona and Holm worked with aquifer material which was very low in organic carbon content (ca. 0.1 % by weight), while this study was done on pond material with a higher organic carbon content (ca. 3.5% by weight). This relatively high value of 0.65 mequiv/g indicates a considerable capacity of sediment material to drive

oxidative reactions. Although much work remains, the spectroelectrochemical titration method has been demonstrated to be a viable one for the determination of the redox capacity. The approach should be useful in evaluating the feasibility of possible remediation processes. The method studied in this project can give an idea of the quantity of pollutant that can potentially be transformed by redox processes in the sediment, as well as the potential range in which such processes may be expected to occur.

Acknowledgments

Armstrong Laboratory

Dr. David Burris
Dr. Howard Mayfield
Eila Burr
Tyndall Air Force Base, Panama City , FL

Dr. Kevin Novo-Gradac
Department of Chemistry
University of Georgia

Dr. William MacIntyre
Virginia Institute of Marine Science

References

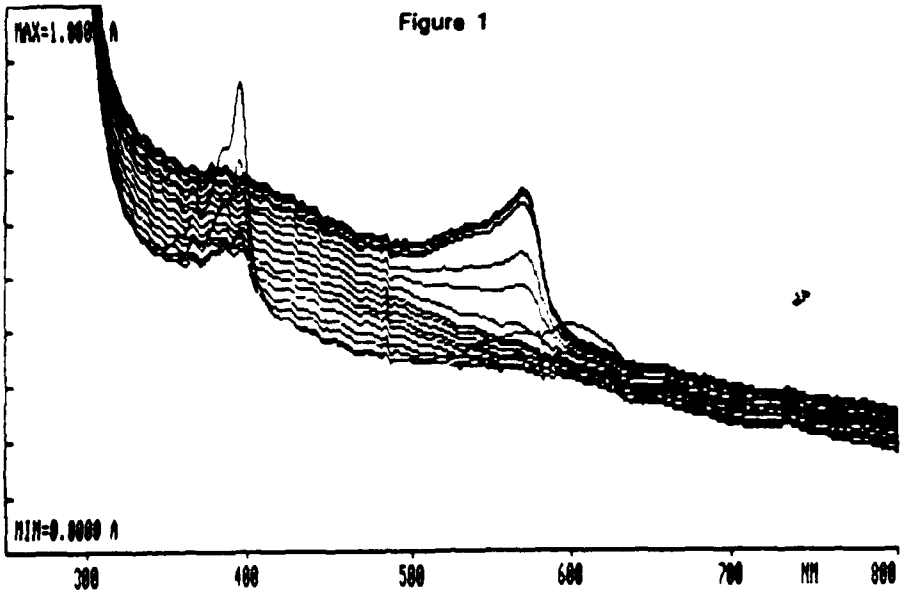
(1) Barcelona, M; Holm,T., *Environ. Sci. Technol.*, **1991**, 25,1565-1572.

(2) Anderson, James L., Dept. of Chemistry, University of Georgia, Final Seminar, Tyndall Air Force Base, Panama City, FL, Aug. **1993**.

(3) Day, P.A., "Methods of Soil Analysis. Part 1. Physical and Mineralogical Properties...", Agronomy9, Ch. 43, **1965**, 545-566.

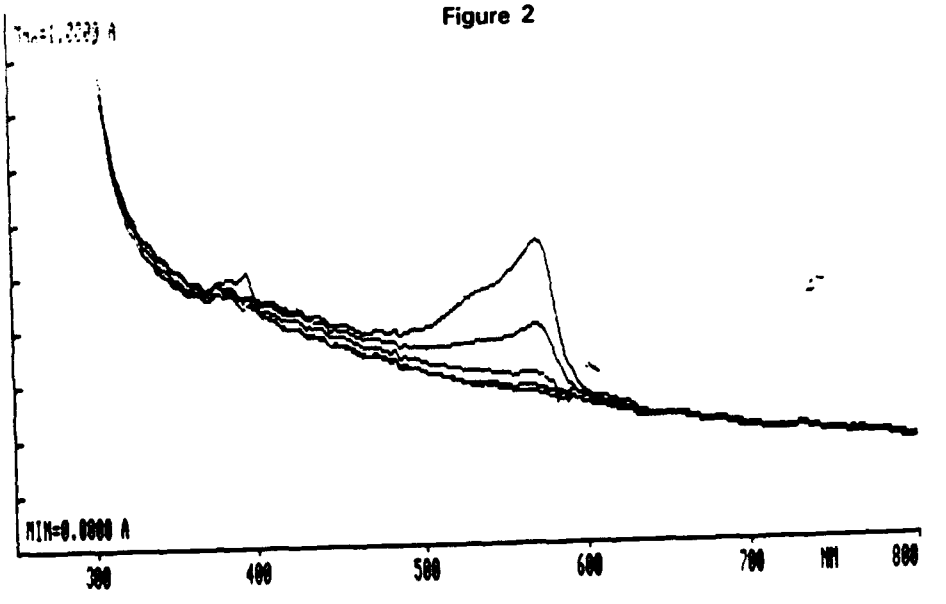
(4) Jackson, M.L., "Soil Chemical Analysis-Advanced Course", Second Edition, 10+L printing, **1975**, 100-166.

Figure 1



REDUCTION OF BEAVER DAM SEDIMENT
8/12/93

Figure 2



OXIDATION OF BEAVER DAM SEDIMENT
8/12/93

Spectroelectrochemical Cell Used in
Anaerobic Sediment Studies

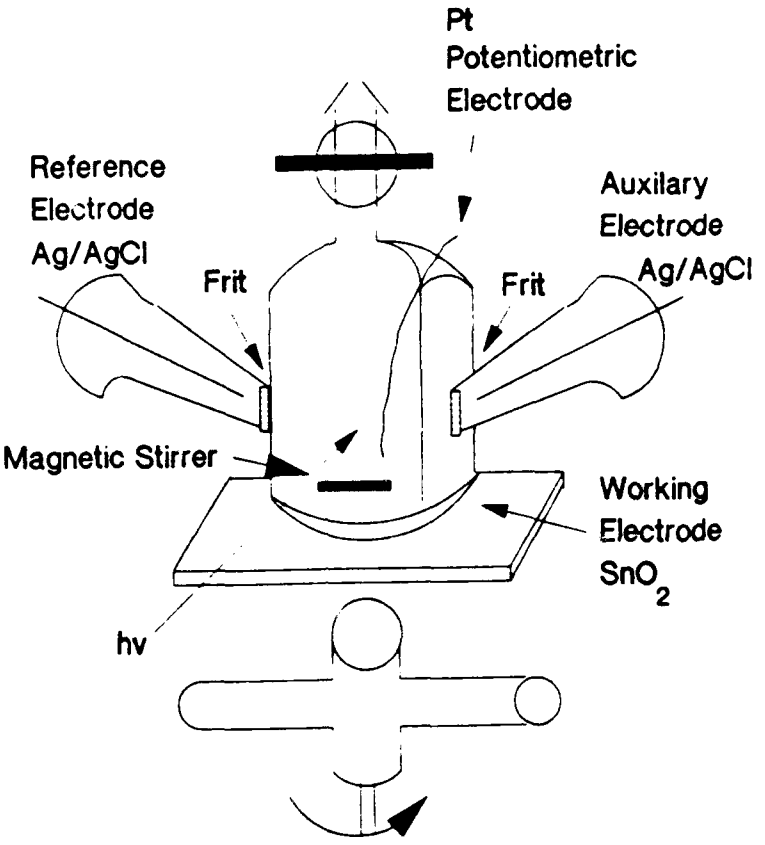


Figure 4

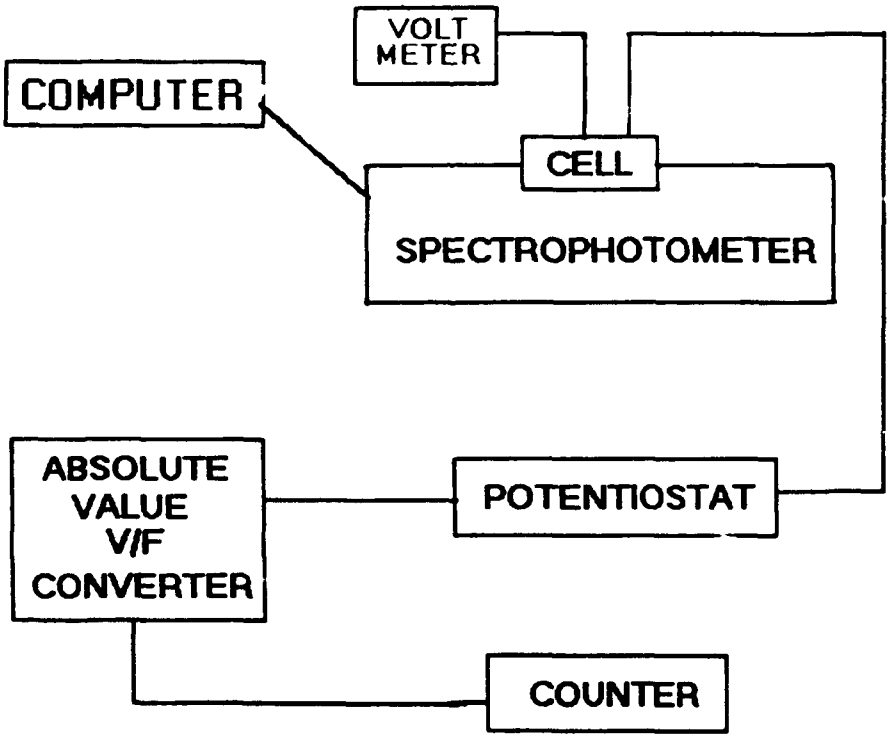


Figure 5

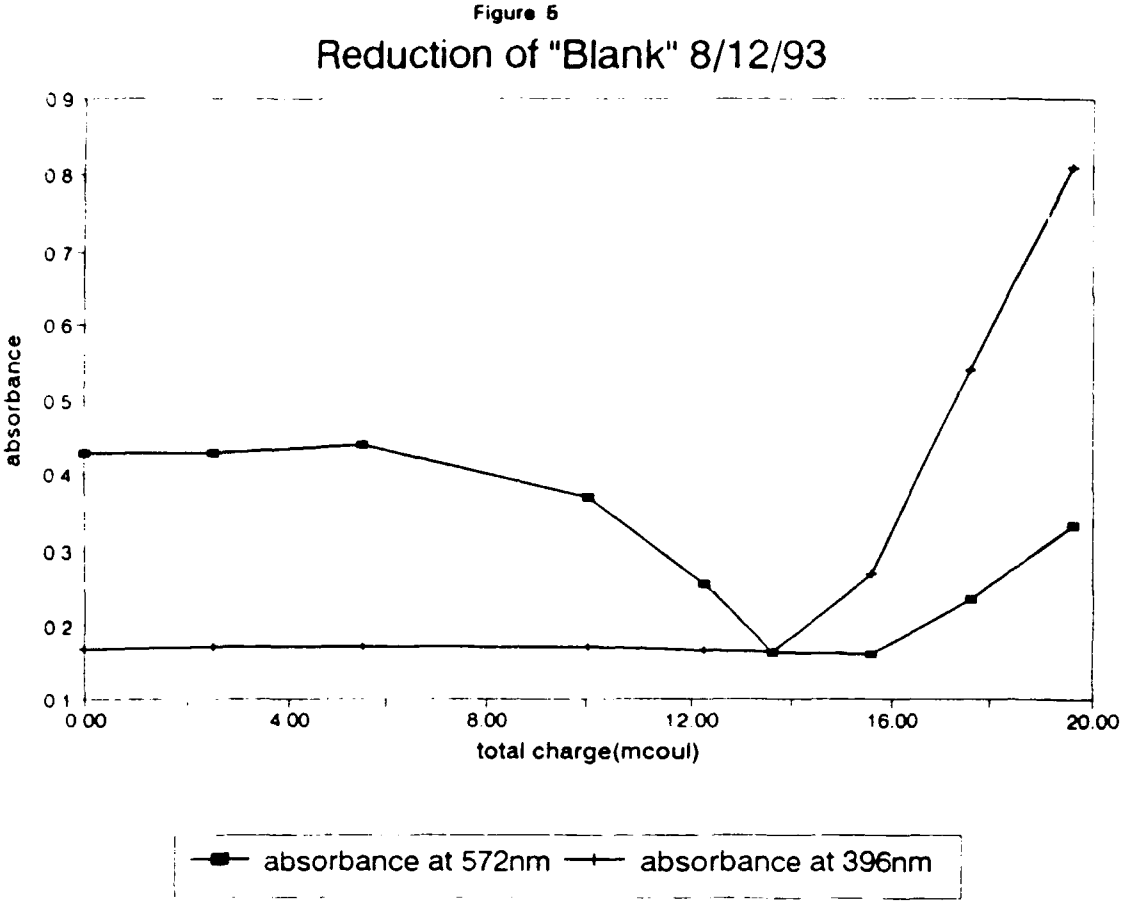
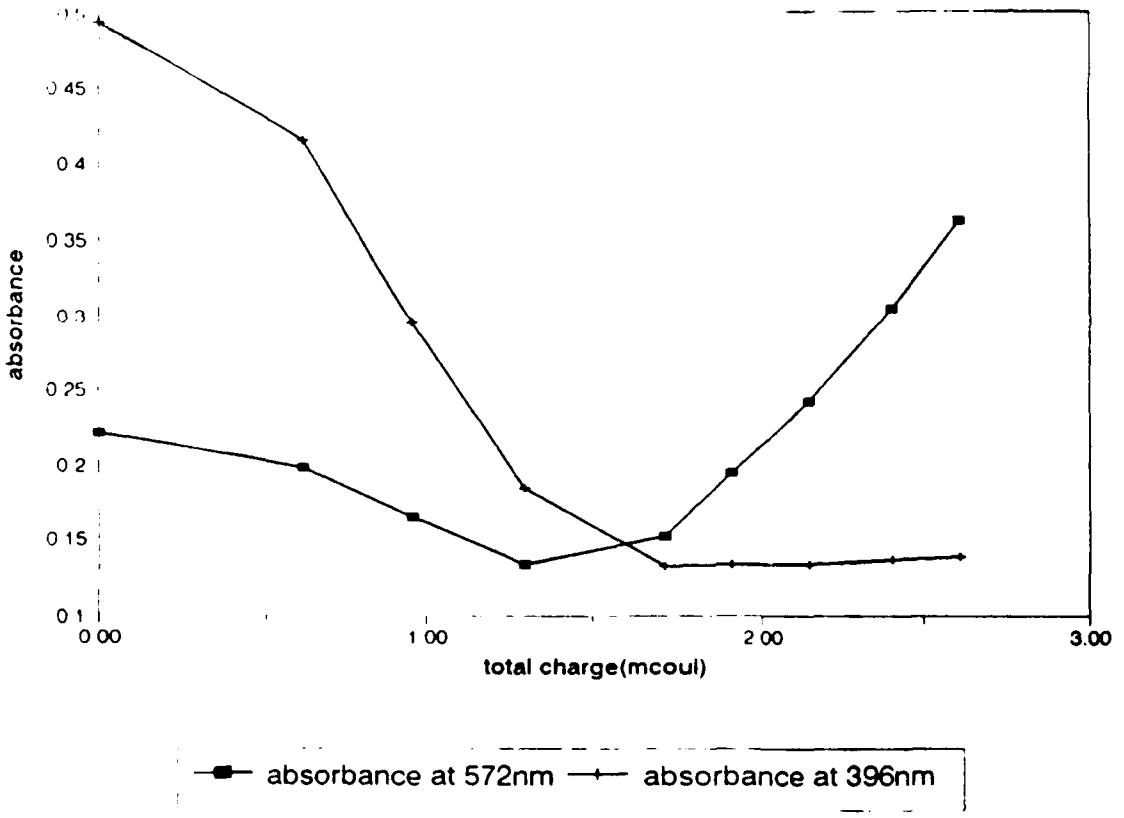


Figure 6



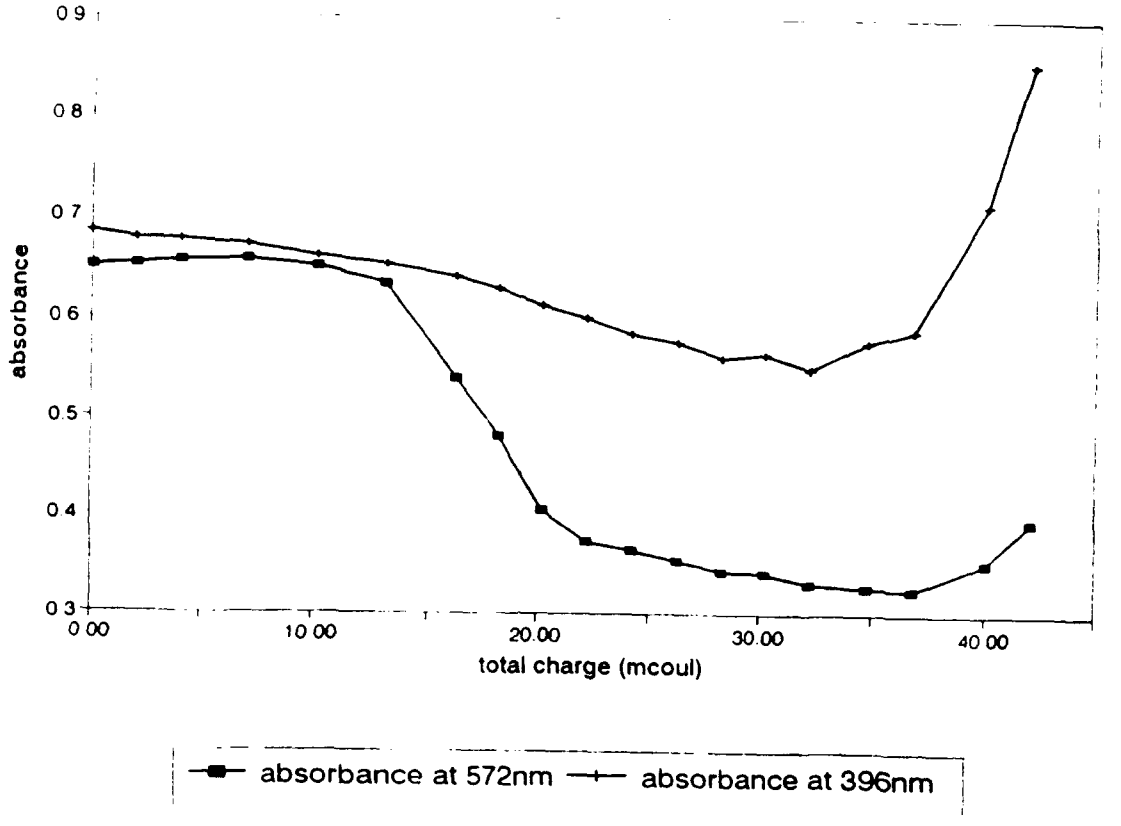

Reduction - Sediment 8/12/93
absorbance
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.00
10.00
20.00
30.00
40.00
total charge (mcoul)
absorbance at 572nm
absorbance at 396nm

Figure 8

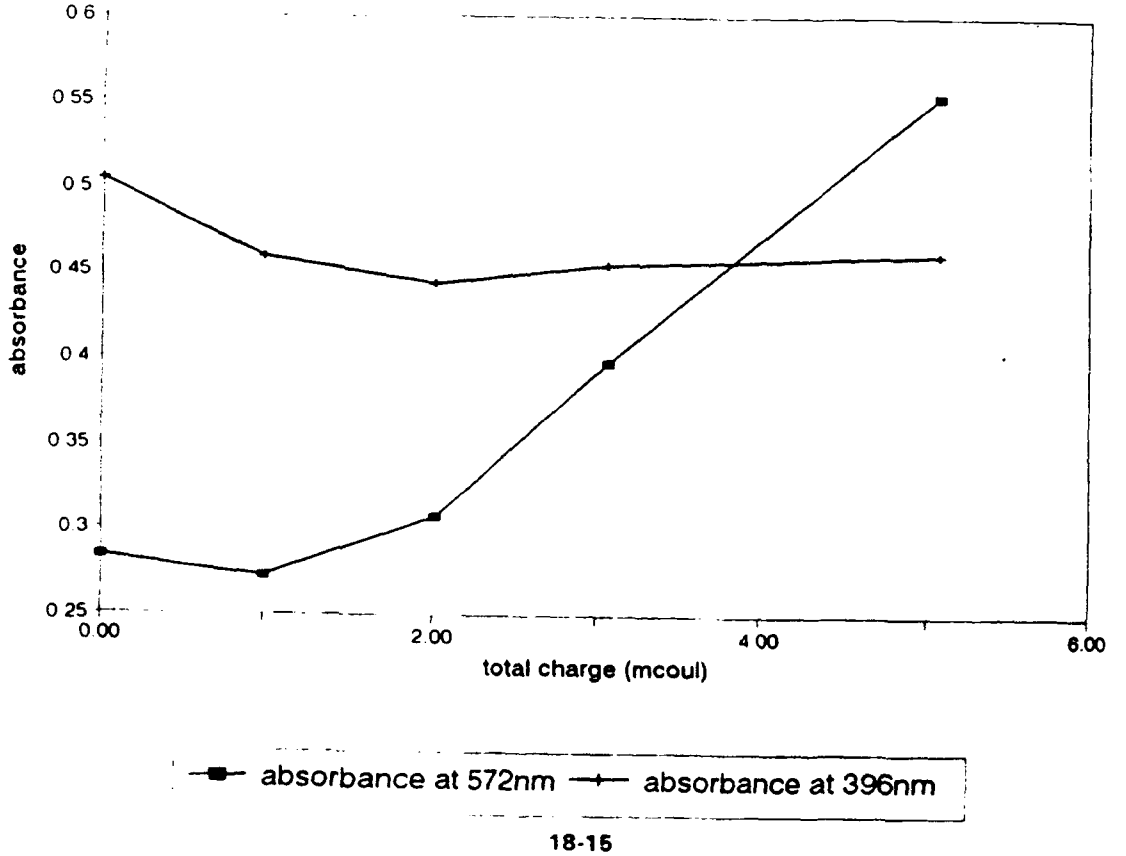

Oxidation - Sediment 8/12/93
absorbance
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.00
2.00
4.00
6.00
total charge (mcoul)
absorbance at 572nm
absorbance at 396nm

Figure 9

Reduction of "Blank"

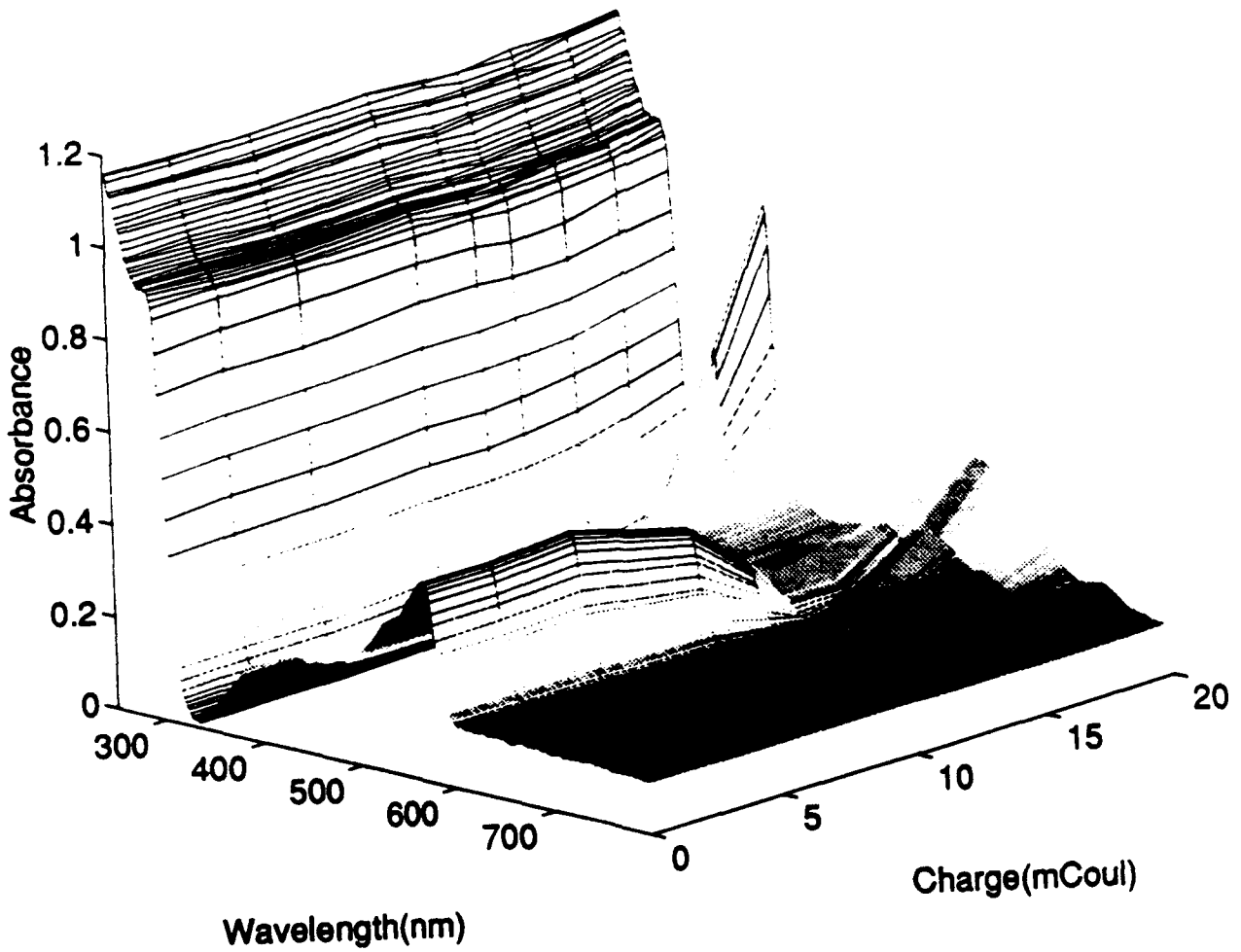


Figure 10

Oxidation of "Blank"

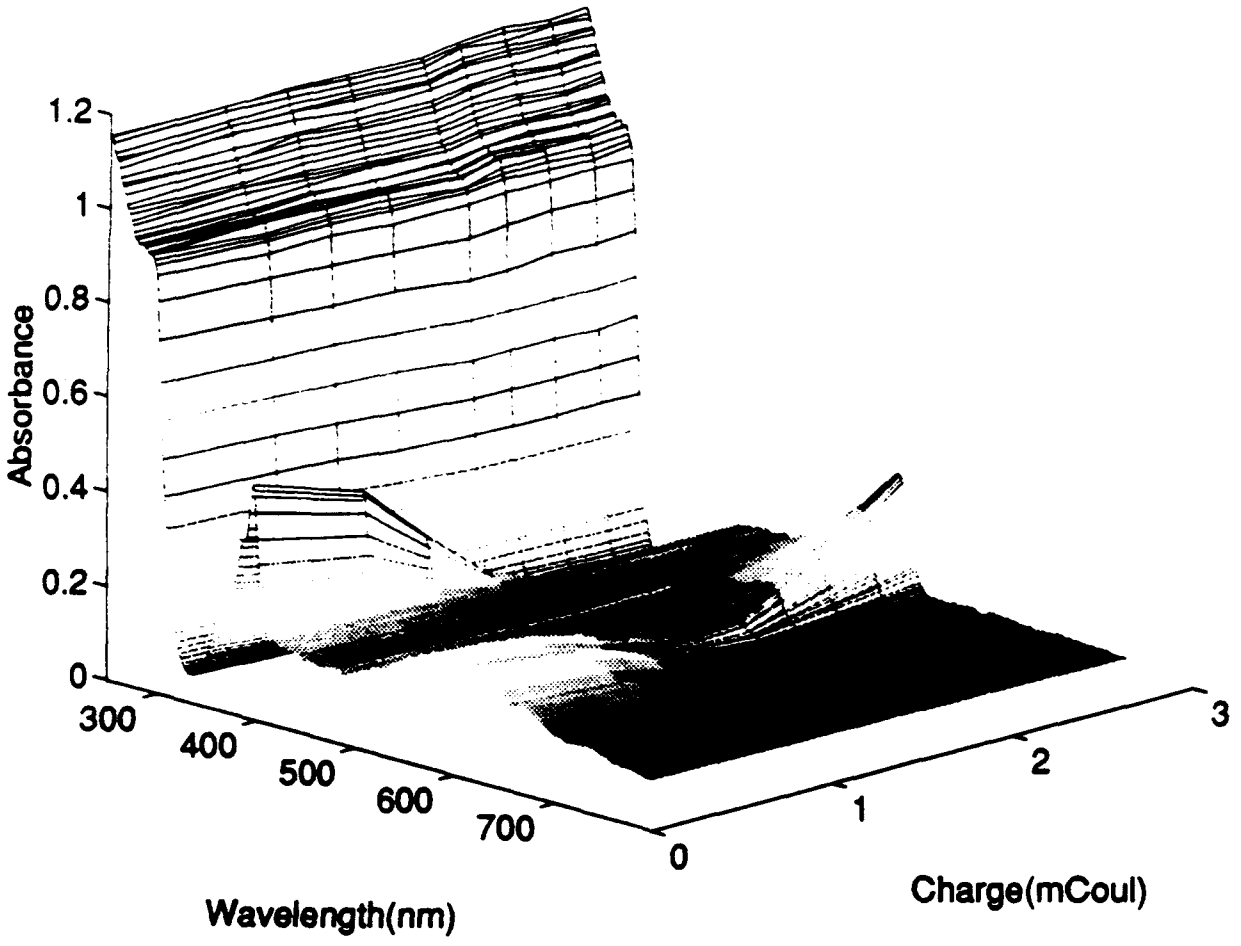


Figure 11
Reduction of Sediment

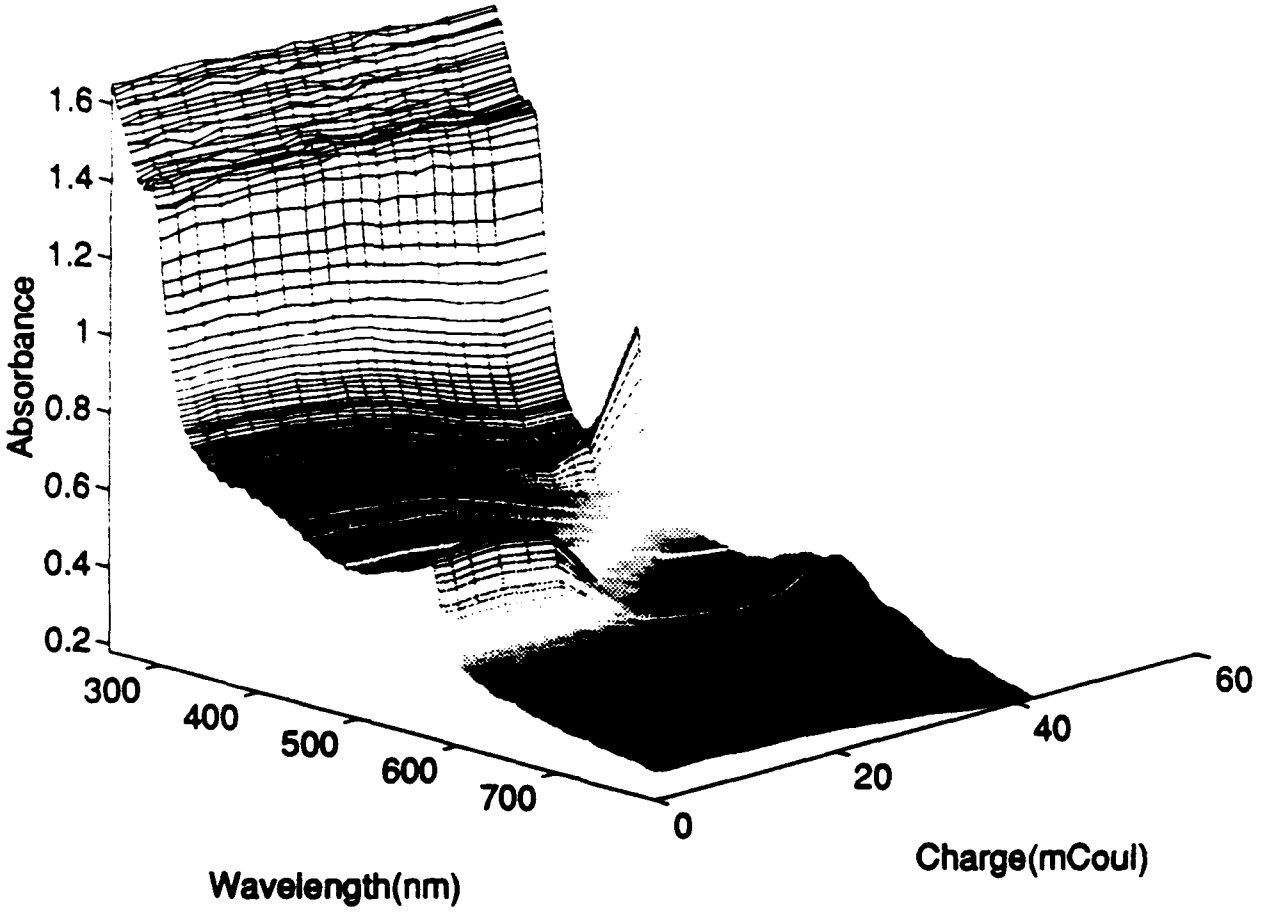


Figure 12

Oxidation of Sediment

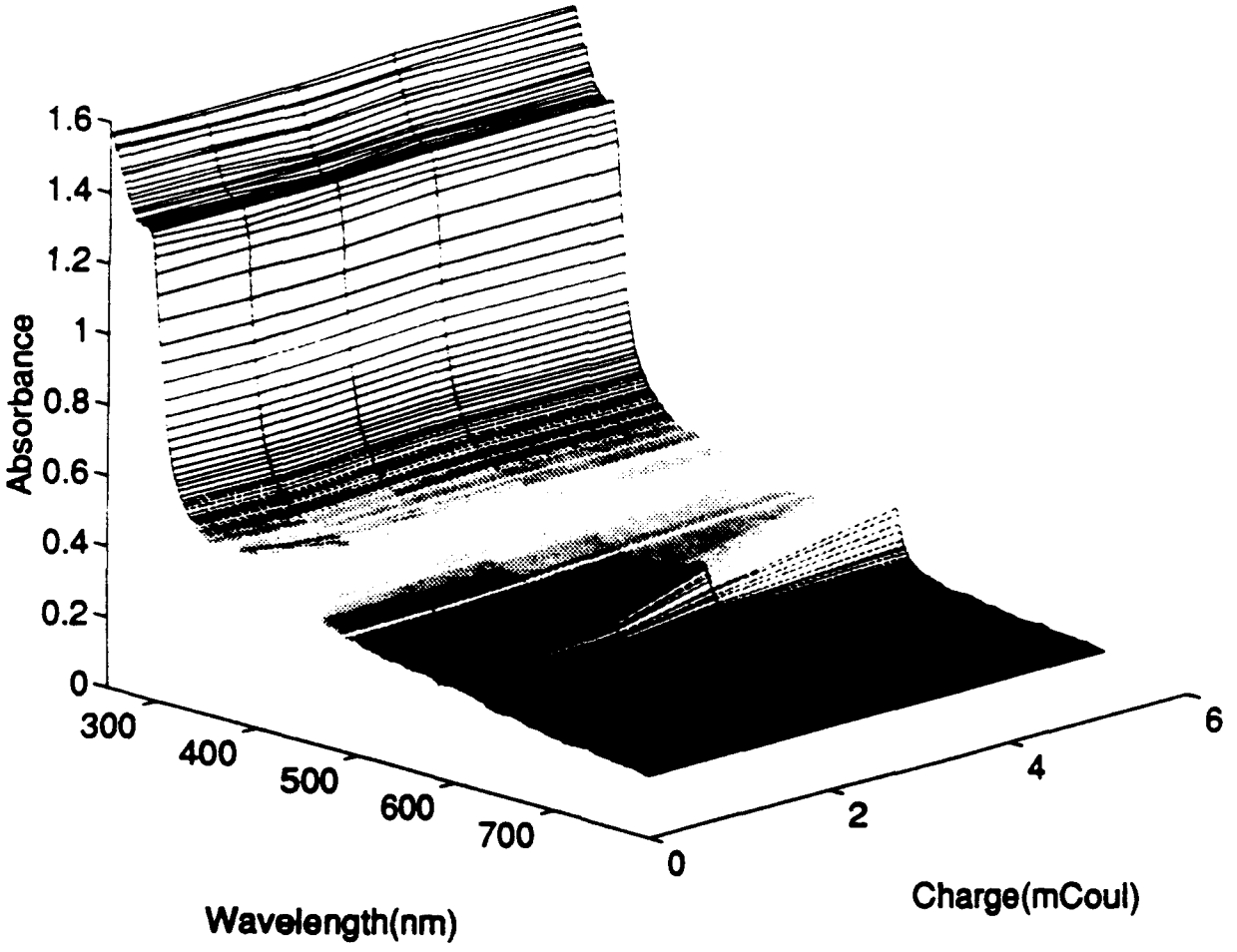
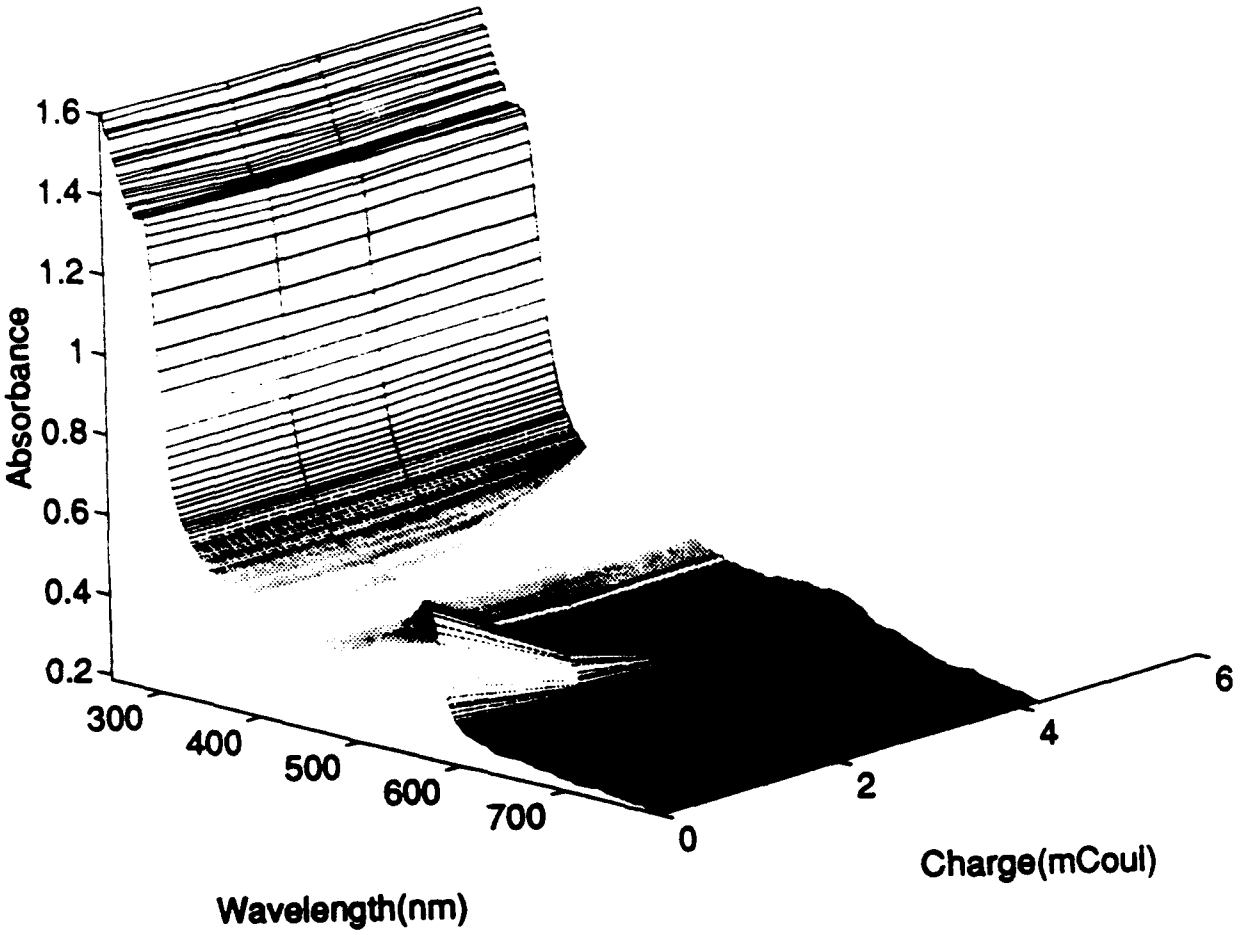


Figure 13

Second Reduction of Sediment



Styela plicata AND *Molgula occidentalis* (UROCHORDATA: ASCIDACEA: STOLIDIFERA) IN ST. ANDREW SOUND, FLORIDA

Sneed B. Collard
Professor
Department of Ecology and Evolutionary Biology

The University of West Florida
11000 University Parkway
Pensacola, FL 32514

Final Report for:
Summer Research Program
AL/EQ-OL

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

August 1993

Styela plicata AND *Molgula occidentalis* (UROCHORDATA: ASCIDACEA: STOLIDIFERA) IN ST. ANDREW SOUND, FLORIDA

Sneed B. Collard
Professor
Department of Ecology and Evolutionary Biology
The University of West Florida

Abstract

Stolidiferan sea squirts, *Styela plicata* and *Molgula occidentalis* collected in the eastern arm of St. Andrew Sound, northern Gulf of Mexico, had filtration rates of about 69 L d^{-1}/animal in phytoplankton concentrations of 4-9 X 10^6 cells L^{-1}. Filtration rates and efficiencies were calculated by measures of water clarity changes over time using time-lapse photography and the presence/absence of bioluminescence in experimental v. control aquaria. Fluorescein and rhodamine B dye observations confirmed filtering activity. Microscopic examination of tunicate-filtered water for the presence of living phytoplankton cells and sediments confirmed that filtration was efficient, and that increased water clarity in experimental aquaria over time was not attributable to the settlement of suspended sediments.

The number of *S. plicata* and *M. occidentalis* in the sampling area was estimated to be 0.7 individuals m^{-2}, with a total population of about 846,000 of each species. Given obviously unrealistic assumptions of uniform phytoplankton distributions and availability; uniform distributions of tunicates; stable and continuous feeding rates and efficiencies; and invariable environmental conditions, the 2 X 10^6 m^3 (2 X 10^9 L) of water in the sampling area could be cleared by the two species in about 17 d. In spite of probable order-of-magnitude errors in estimates, the two species of sea squirts clearly play an important trophodynamic role in the ecology of St. Andrew Sound, and likely do so in other favorable, high salinity coastal ecosystems.

High and efficient filtering rates coupled with the known ability of ascidians to assimilate heavy metals (up to 10^9 above ambient concentrations), strongly suggests that sea squirts may be usefully employed as biological indicators/ monitors in anthropogenically impacted marine ecosystems. Ascidians may also provide a low-cost alternative to microbiologically-mediated environmental bioremediation. Certain species may be harvested for the extraction of strategic metals.

The major, possibly only significant predator of *S. plicata* in St. Andrew Sound is the Florida crowned conch, *Melongena corona*, which also feeds upon the tunicate's possibly mutualistic mussel symbiont, *Musculus lateralis*. The relationship between sea squirt and mussel may be reciprocally beneficial in terms of facilitating the disruption of an animal-water interface boundary layer, which in non-turbulent conditions may impede the acquisition of food and oxygen to the species pair.

Styela plicata AND Molgula occidentalis (UROCHORDATA: ASCIDACEA: STOLIDIFERA) IN ST. ANDREW SOUND, FLORIDA

Sneed B. Collard

INTRODUCTION

Ascidiacean tunicates are sessile, suspension/filter-feeding planktivores usually referred to as "sea squirts". As a group, the ascidians are cosmopolitan, opportunistic colonizers of subtidal marine communities containing hard substrata. They are found, for example, on rocks, shell hash, compacted sands, and on the bodies of sessile plants and animals (e.g., seagrasses, mangroves, corals, oysters, other ascidians), and on some motile species such as scallops and non-burrowing gastropods. Sea squirts may dominate fouling communities on wood, plastic, teflon, glass and other artificial substrata, such as the hulls of ships. The planktonic larvae of some ascidians respond to specific environmental and/or chemical cues before settlement on favorable substrata, while others appear to be opportunistic, settling on a wide variety of substrate types. Many of the 2000 or so known species of ascidians form encrusting colonies of hundreds or thousands of small zooids. In contrast, the large and speciose order Stolidifera contains non-colonial (simple, as opposed to compound) species which develop as solitary individuals.

A rich literature exists on the structure and physiology of ascidians (e.g., Berrill, 1950; Goodbody, 1974; authors cited in Morris et al., 1990). In addition to the many unique features of adult sea squirts, their tadpole-like larvae exhibit important characteristics of chordates with which tunicates are phylogenetically allied. The recent development of a host of sophisticated biochemical and immunological tools and techniques has stimulated a resurgence of interest in the group. Current research on the tunichromes; on tunicin; and on and the mechanisms by which ascidians and their pelagic thaliacean and larvacean relatives accumulate and sequester heavy metals, for example, should add significantly to our understanding of the group (Michael and Pattenden, 1993).

While information on the molecular-cellular biology of tunicates continues to accumulate rapidly, knowledge of the functional, community-level ecology of the group has not kept pace. The primary objective of the present work was to gain additional knowledge about the ecological roles and functions of wild populations of Styela plicata and Molgula occidentalis in St. Andrew Sound, an undescribed, nearly pristine marine lagoon located in the central panhandle of Florida. Space and formatting limitations permit no more than a brief and incomplete discussion of new information obtained on these species, and preclude the inclusion of a literature review. Observations on ascidian inquilines, successional seres and community structure; phytoplankton species, seagrass meadow ecology, speculation that the Sound may be copper, zinc, iron and lead-enriched, significant physicochemical and biological changes in the Sound from 1992-1993, and other information will be reported elsewhere.

Study area -

St. Andrew Sound is a shallow, microtidal, marine lagoon about 14.5 km long and 0.2-2.0 km wide lying between Tyndall AFB to the north and Crooked Island to the south (Collard, 1992). The Sound communicates with the Gulf of Mexico through "Hurricane Pass", a blowout near the center of Crooked Island created by the passage of Hurricane Eloise in 1975 (Figs. 1, 2)[1].

St. Andrew Sound (characterized by Collard, 1992) is a seagrass-based ecosystem used as a foraging area for juvenile and adult green, loggerhead and Kemp's ridley sea turtles. The seagrass meadows support a diverse juvenile and

[1] Erosional longshore drift to the west and a winter storm in 1993 increased the width of the pass from about 0.4 km in summer 1992, to 0.8 km in summer 1993, and decreased its depth from about 3-6 m (across its width) in 1992, to 2-3 m across a <50 m-wide, unstable channel in 1993. It is possible that continued shoaling may impound and transform St. Andrew Sound into a hypersaline, eutrophicated lake.

adult ichthyofauna (Collard, 1992; Reisinger, 1993), as well as shrimp, blue crabs, scallops and other important fisheries species. Large numbers of Stvela plicata and Molgula occidentalis live on seagrasses and unvegetated sediments -- in single-species clumps (pseudocolonies) of adherent, but physiologically independent (?) individuals, or in mixed-species clumps of various sizes. The tests (tunics) of both species provide habitats for a variety of epibionts and inquilines, including other ascidians.

Figure 1: Location of St. Andrew Sound

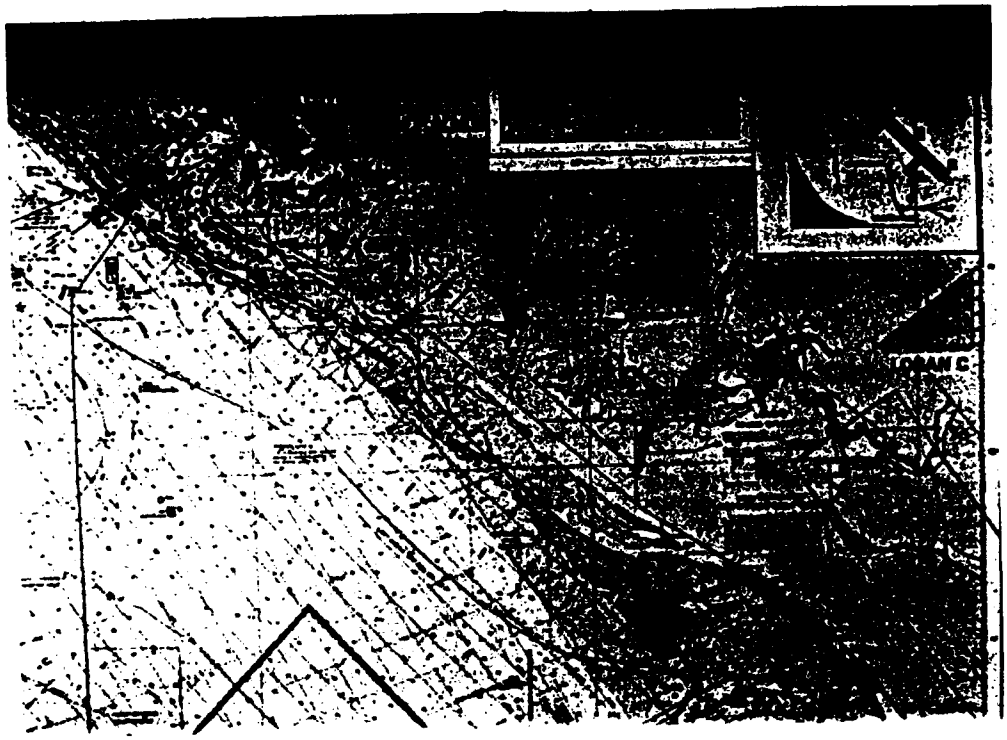
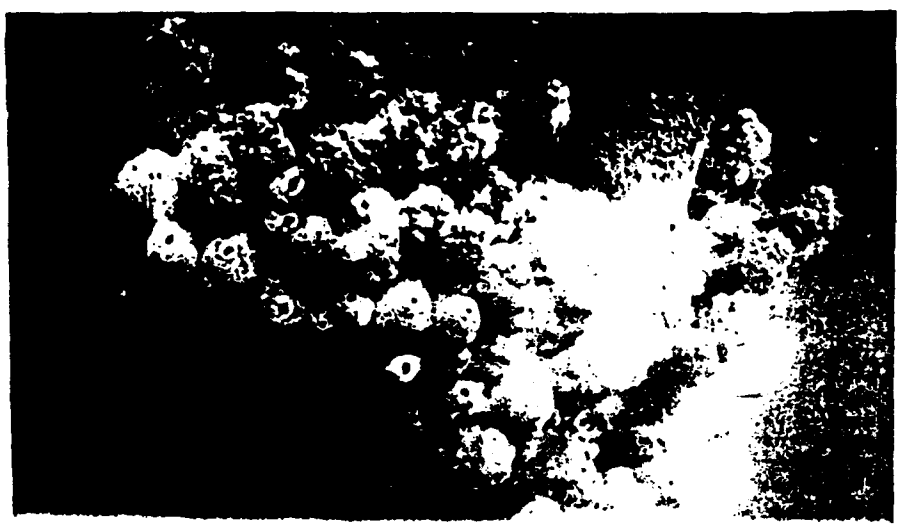


Figure 2: Photograph of the sampling area in St. Andrew Sound



Figure 3: Underwater view of Styela plicata on pier piling



MATERIALS, RESULTS AND DISCUSSION

Calculations of some dimensions and quantities were based on scant information (i.e., tidal prism, volume, surface area and bathymetry of the Sound), and estimates were made with unknown, but assumed errors of as much as ± 25% in some cases. Phytoplankton cell counts and the filtration rates of Styela plicata are reliable within the limits of natural variation (e.g., differences in the size and physiological condition of individuals and clumps, water temperature, time of day, etc.). The filtration rate of Molgula occidentalis was usually observed to be greater than that of S. plicata, but fewer measurements/observations were made, and the number reported is a conservative estimate. For purposes of the present work, the magnitude of probable error is considered acceptable.

Dimensions and calculations -

Using charts (Fig. 1), aerial photographs (Fig. 2), and known distances between permanent land features (radio towers, roads, etc.) the sampling area (Davis Beach to the western end of Raffield Peninsula) was calculated to be 5.1 km long and 0.138-0.740 km (average 0.46 km) wide, with an average submerged surface area of 2.3×10^6 m^2 (2.3 km^2).

The tidal prism was estimated using water marks and the vertical zonation of barnacles, oysters and tunicates on pier pilings. The vertical distance between the upper live barnacle horizon and the lowest zone of live, adult oysters (50 cm) was used as the average width of the intertidal zone, and the vertical distance between the uppermost zone of live barnacles and sea squirts (80 cm) was considered to approximate the maximum (non-storm) tidal range.

The bathymetry of St. Andrew Sound has not been charted. Average water depth in the sampling area was estimated from several hundred measurements taken with meter sticks, stadia rods and sounding leads on transects across the Sound during June-August 1992 and June-August 1993. Water depths ranged from 0.2-5.0 m, but most measurements (with respect to a tidal range of 0.8 m) indicated depths of 0.5-1.0 m. Using an average of these most commonly

measured depths, the sampling area was estimated to have a mean water depth of about 0.75 m. Given a mean depth of 0.75 m and an average tidal range of 0.8 m, a subtidal area of 1.2 X 10^6 m^2 was estimated to be the total area available for tunicate colonization. Based on these estimates, the average (zero-tide) volume of water in the sampling area was calculated to be about 2 X 10^6 m^3 (2 X 10^9 L).

Phytoplankton collections -

Phytoplankton collections were made throughout the sampling period (mid-June to late August 1993) from near the surface with a 5 gal bucket, and from the water column (oblique tows from the surface to near the bottom in 1-3 m water) with a 0.1 m, 63 nm mesh phytoplankton net. Live and preserved samples were sent to K. Steidinger (FDEP Marine Research Laboratory) for identification. Cell counts were made at 400X magnification with a Nikon microscope using a hemacytometer, Sedgwick-Rafter cells, and a "micro-aquarium" method suggested by Steidinger (pers. comm.). In the latter method, a 22 cm^2 #1 glass coverslip was "tacked" with stopcock grease to a microscope slide, and a measured volume (1-3 ml) of bucket-collected water containing phytoplankton was pipetted into the resulting chamber. Cell counts using this method were consistent in replicate sub-sample comparisons. The number of cells L^{-1} ranged from 4.41 X 10^6 on 17 August (4:1 dinoflagellates) to 7.28 X 10^6 on 22 July (3:2 diatoms:dinoflagellates) (Collard and Steidinger, in prep.).

Tunicate collections -

Most observations and collections of tunicates were made in the eastern portion of the Sound: from a small islet west of Davis Beach, to the western end of Raffield Peninsula. Wild Goose Lagoon (Fig. 2) was not adequately sampled. S. plicata collected from boathouse pilings (Fig. 3) were used in some filtration rate experiments, and for heavy metals analyses (Collard, Cornette and Kirshharr, in prep.). Collections were made from a 16-ft boat using a 10-ft otter trawl with a 1½ in stretch-mesh net (no codend liner), and by hand, using snorkeling gear. Very turbid water (secchi depths <0.6 m)

coupled with observations of large alligators and numerous reports of sharks in the sampling area discouraged extensive in-the-water collections. One "large" shark was seen at the boathouse by the author at 0700 on 24 July 1993 while retrieving an artificial substrate for inspection.

Twenty-one timed otter trawl tows were taken at irregular intervals in July and August to estimate tunicate catch per unit effort. Permanent land features were used to estimate the total distance trawled per unit time. Tows were made in representative vegetated and unvegetated, shallow- and deep-water regions throughout the sampling area. S. plicata and M. occidentalis from each tow were counted, as was the species composition and number of individuals occurring in clumps. The number of S. plicata (the target species) was calculated to be 0.7 individuals m^{-2}, with a total population of about 846,000 in the sampling area. M. occidentalis (not targeted) was undersampled, and its distribution was patchier than that of S. plicata (0-59 m^{-2}, in 36 m^2 transect-plot collections; 0->100 m^{-2} in trawl collections; see Fig. 4). The habitat in which M. occidentalis were most numerous (partially buried in, and loosely attached to fine sediments in shallow water) was inefficiently sampled with the otter trawl. These factors precluded an acceptably accurate (± 25%) estimate of the total number of M. occidentalis in the sampling area. Several hundred hours of in-the-water observations suggested, however, that M. occidentalis is at least as abundant as S. plicata in the sampling area except on above-bottom substrata (pier pilings, PVC posts, crab traps) where M. occidentalis was rarely found. While the two species often co-occurred, S. plicata was not found in shallow muddy substrates or in Halodule wrightii meadows.

Figure 4: Molgula occidentalis from an otter trawl collection



Figure 5: Boathouse laboratory



Boathouse laboratory and aquaria -

S. plicata and M. occidentalis were placed in aquaria located in the boathouse immediately after collection (Fig. 5). To determine the influence of crowding on filtering rates/efficiencies, from 1-72 test animals were placed in aquaria which, when filled to the level most often used, contained the following volumes of water: 0.7, 1.4, 18, 23, 43, 65, 69, 79, 150, 430 and 6,051 L. With the exception of a 65 L plastic tub, aquaria from 0.7-79 L were made of glass. Larger aquaria were plastic wading pools. Aquaria were filled with ambient sea water collected from the boathouse docks by hand with clean 5 gal buckets. None of the aquaria were equipped with filters or circulation pumps. An air stone was occasionally used when it was necessary to hold a large number of tunicates in a too-small aquarium for short periods of time. Water pumps were initially employed to reduce the labor involved in changing water in aquaria (usually from 900-2,000 gal or more per day). Pumped water resulted in mass mortalities when large (> 400 L) aquaria were used, or when only one or two tunicates were held in aquaria smaller than 23 L. With the types of pumps used, water was forced to pass through impeller blades, and/or was heated by the pump's motor, killing most of the plankton. It was concluded that when the volume of water in an aquarium exceeded the rate at which tunicates could filter out dead plankters prior to their decomposition by microbes, rapid reproduction of the latter reduced water quality and killed the tunicates. In aquaria containing a volume of water that could be completely filtered by tunicates in 2-4 hr, no mortality was observed using pumped water.

Figure 6: Styela plicata with siphons open and producing fecal pellets



Tunicate filtering rates -

Tunicates were placed in aquaria with fresh, unfiltered sea water and their behavior was observed and recorded after varying periods of acclimation to aquarium conditions (glass substrates, fluorescent lights, often shallower depth and still water). Animals were judged to be fully acclimatized when most of the individuals in a given aquarium were observed to be actively filter-feeding and producing fecal pellets (Fig. 6). Newly captured animals fully acclimated in 2-6 hr, while those "used to" aquarium conditions began filtering immediately after water changes. Acclimatization of Styela species to aquarium conditions has been reported to take as long as several days; in the boathouse environment used here (roofed, but otherwise exposed to ambient environmental conditions), this period was greatly reduced. Filtering rates were calculated as follows.

I. Water clarity comparisons -- Aquaria containing tunicates were filled with freshly collected, turbid, phytoplankton-rich sea water (Forel-Ule color XIII -- pea green), and observed at varying intervals of time until it appeared (compared to control aquaria with no tunicates) to be crystal clear. In the most frequently used "triple aquaria" each having a filled volume of 23 L each (Fig. 6), water was (on average) completely cleared by 10-12 S. plicata within 3 hr; by 6-8 S. plicata within 3-4 hr; and by 2-4 S. plicata within 4 hr. Non-linear differences in clearing rates/number of animals are attributed to variation in the groups of test animals (e.g., sizes, ages, nutritional/physiological/reproductive condition), and somewhat lowered efficiency due to competition between individuals in larger clumps.

In three trials 72 individuals, the largest clump of S. plicata collected, were unable to clear water in a 69 L aquarium in 24 hr (after which the water was replaced). This unexpected result was tentatively attributed to fouling of the water caused by the breakup and resuspension of fecal pellets by the group's atrial siphon currents. Subsequent videotaped dye experiments documented turbulence in the aquarium for several hours after water exchanges, followed by decreased branchial siphon activity (i.e., animals closed their

incurrent siphons for increasingly longer periods of time). Turbulence throughout the aquarium, and fewer than the expected ("normal") number of formed fecal pellets on the bottom of the aquarium supported this explanation. In poor water quality conditions S. plicata were repeatedly observed to reduce their filtering activity, then close their siphons until they died. The clump of 72 Styela was narcotized with MS-222, and following preservation, examined under a dissecting microscope. Individuals in the clump were heavily infested with the nest-building commensal mussel, Musculus lateralis (Figs. 9, 10), which suggested that the initial explanation (turbulence-induced fouling prevented Styela from clearing the water) was inadequate (discussed below).

Water clarity in experimental and control aquaria was photographed at the beginning and end of each trial using a Minolta Maxxim 7000 and/or a Nikonos V camera and Scotch Chrome ASA 100 color film (Fig. 7). Photographs were taken from above, and through the walls of aquaria, under differing light conditions, with and without polarizing and ultraviolet filters. Six 4-8 hr time-lapse videotape recordings were made of tunicates to document their filtering behavior and changes in water clarity. Photographs and unedited tapes are available from the author.

The bottoms of control and experimental aquaria lacked detectable accumulations of sediments, confirming that turbidity in test water was the result of dense phytoplankton accumulations.

II. Dye experiments -- Fluorescein or rhodamine B dyes dissolved in sea water were introduced via syringe and needle to aquaria containing either starved or freshly collected tunicates to obtain unequivocal evidence that open branchial siphons coincided with filtering activity. As clearly documented on real-time videotapes, when a slow (low pressure) stream of dye was allowed to approach within < 2 cm of the branchial siphon, a steady, non-turbulent stream was drawn into the animal by ciliary action. Dyed water was quickly expelled through the atrial siphon in 1-3 s, in a broad, turbulent, continuous high-velocity stream. The velocity of the exhalent stream was not measurable (Vogel, pers. comm.), but roughly corresponded to the distance that fecal

pellets were ejected, always from 8-12 cm away from the animal. Strong concentrations of dyes were used, and test animals allowed to remain in the dye solution without water changes or aeration for >48 hr showed no detectable ill-effects.

III. Bioluminescence -- It was noted earlier that dinoflagellate bloom conditions occurred in the Sound throughout the summer months. These dinoflagellates were bioluminescent. By chance it was discovered during an early morning (0200) visit to the laboratory, that visibly clear water showed no bioluminesce when mechanically agitated, and that control aquaria remained brilliantly bioluminescent throughout and after test animals had presumptively filtered out all or most of the phytoplankton. Samples of water taken from cleared experimental aquaria (containing respectively, 7 and 11 S. plicata) were examined under 400 and 1000X magnification and found to be devoid of diatoms and dinoflagellates. Water taken from the bottom of these aquaria revealed low numbers of empty diatom frustules, a few live diatoms, and no dinoflagellates or sediments. Water had been changed in the aquaria described at approximately 2300, three hours earlier.

Styela plicata aggregate filtering rates -- Given a population of about 8.5 X 10^5 S. plicata in a sampling area containing 2 X 10^9 L of water; an effective filtration rate of 69 L^{d-1} (18.2 gal) per individual; and an average phytoplankton density of 5.5 cells L^{-1}, the standing Styela population should clear the water column in about 34.5 d (58 X 10^6 L/day). This number translates into 16 X 10^6 phytoplankton cells hr^{-1} $tunicate^{-1}$. In other words, a population of about 1.1 X 10^8 S. plicata could theoretically clear the Sound of phytoplankton in one day assuming (unrealistically) that tunicates and phytoplankton were uniformly distributed, that tunicate filtering rates were constant, and that filtering was 100% efficient. Further speculating that Styela plicata and Molgula occidentalis occur in roughly equal numbers in the Sound, and that the species have comparable filtering rates, solitary tunicates might clear waters of the Sound in as little as 17 days. Should this astonishing number be in error by 1-2 orders of magnitude (or more), which is

possible, the two species of ascidians nonetheless clearly play, in addition to their other ecological functions, an important trophodynamic role in St. Andrew Sound.

Observations --

Response of S. plicata to currents

Numerous authors have reported that Styela species orient with respect to, and thereby feed more efficiently (actively?) in currents. Observations made in still-water aquaria and on clumps of S. plicata living on pier pilings (in currents), showed no detectable differences in their response to the uptake of dyes, or in branchial siphon-open intervals. These observations were unnecessary, however, for it is obvious that tunicates living in densely packed clumps can not significantly change the direction of their branchial siphons, and can not orient themselves to take advantage of ostensibly favorable currents. S. plicata is often found, however, on unvegetated bottom sediments, where current scour is evident. Given that orientation is not possible, a possible adaptive advantage to living in clumps may be that aggregated tunicates are less likely than individuals to be displaced or stranded by currents.

The above observations notwithstanding, an experiment was done to determine whether the behavior of S. plicata in aquaria with sand on the bottom differed from that observed on the bottom of clean glass aquaria. Results recorded on videotape showed that aggregates of five or more animals remained in place. The limited movement of siphons suggested that clumps of this size do not orient in response to environmental stimuli. Individuals located in unfavorable positions on a clump die.

Observations made in the field suggested that two or three S. plicata attached to each other occurred far more commonly than larger aggregates or solitary animals. When placed in aquaria with sand, it was seen that the strong jet of water produced during fecal pellet ejection caused the entire clump to turn, thus re-orienting each member of the group, which were in turn exposed to a different aspect of ambient currents. Orientation then, may occur

Figure 7: Water clarity comparisons in experimental and control aquaria

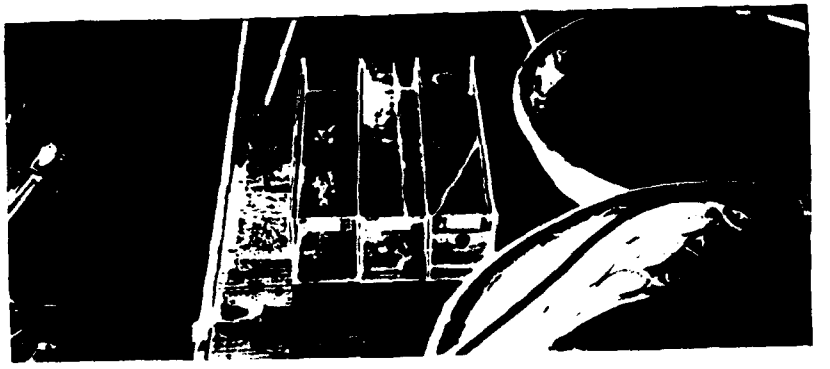


Figure 8: Rotation of clump due to atrial siphon currents

Figure 9: Musculus lateralis in tunic of Styela plicata





as a by-product of optimal clump size which, in part, may explain the adaptive advantage and commonness of aggregates of less than five or so individuals.

Inquilines and commensals of Styela plicata

The presence of toxic metals in ascidian tunics are believed to prevent or discourage the settlement of fouling organisms on many species (e.g., Stoecker, 1980). In spite of above-ambient concentrations of vanadium, zinc, copper, lead, iron, niobium and tantalum in S. plicata, and other metals in M. occidentalis (reported elsewhere), the tunics of both species provided a substrate for a diverse assemblage of epibionts including diatoms, ciliates (Folliculina sp.) entoprocts, bryozoans, nematodes, polychaetes, at least four species of colonial tunicates, and the filter-feeding mussel, Musculus lateralis. Although this community will be described elsewhere (Collard, in prep.), it is suggested here that Styela plicata and M. lateralis are mutualistic symbionts whose relationship is adaptively advantageous to both species. [Other aspects of the relationship between S. plicata and M. lateralis are described in an excellent paper by Bertrand, 1971.]

As shown in Fig. 9, numerous M. lateralis (often > 50/host) lie embedded in the tunic of S. plicata with their valve margins at the surface of the host. These valves of these little mussels were most frequently observed to be open with the proboscis extruded (i.e., suspension/filter-feeding) when the branchial siphon of S. plicata was also open.

In the frequently still-water habitat (e.g., of S. plicata in St. Andrew Sound), sessile, suspension-feeding species able to generate only weak food-gathering currents are likely to become oxygen and food-limited due to the formation of non-turbulent boundary layers (Vogel, 1981). In the absence of water currents (e.g., Bretz, 1972), turbulence sufficient to disrupt such boundaries must be generated by the affected organisms. It is suggested that the presence of numerous suspension-feeding, current-producing Musculus lateralis on the surface of S. plicata may facilitate the disruption of boundary layers surrounding the bodies of both species. Exhalent currents produced by S. plicata may, in like fashion, enhance disruption of a boundary

layer that might otherwise impede or prevent the effective transport of small prey organisms and oxygen to M. lateralis. While it is far from clear that boundary layer problems favor the observed relationship between S. plicata and M. lateralis, the speculation has interesting, more general ecological implications, and deserves further investigation. Observations of a "halo" of clear water surrounding clumps of S. plicata in large aquaria add support to this view. Unfortunately, photographs of the "halo effect" did not confirm its presence.

Predators and the distribution of Styela plicata

Dalby (1989) provided weak evidence that Melongena corona, the most abundant large gastropod in St. Andrew Sound, preyed upon S. plicata in coastal areas of northwest Florida. Experimental evidence gathered during the present investigation confirmed this, and added Musculus lateralis to the known list of M. corona food items. Observations in the field and in aquaria showed that Melongena (even when completely buried in the sand) can rapidly detect the presence of S. plicata at distances of 10 m or more. After detection, snails approached S. plicata and either stopped when a few cm away, or "crawled" onto the test(s) of the solitary or clumped tunicates. In the former case, M. corona positioned itself in front (downstream) of the siphons, remained still until the one of the two siphons opened, then rapidly thrust its proboscis into the branchial basket of S. plicata (Fig. 11). S. plicata closed its siphon which more firmly held the snail in place. After a few minutes tunicates appeared to relax (presumably narcotized by a substance produced by its attacker), and the snail consumed the visceral mass of the sea squirt. It took several hours for M. corona to consume the viscera, but when the process was finished S. plicata tests appeared as flattened parchment.

When M. corona crawled onto the surface of S. plicata it moved slowly over a period of several hours, then left the tunicate. In all cases observed, the tunicate itself was neither attacked nor damaged. Examination of tunicates after such an event showed that a majority or all M. lateralis were eaten by Melongena (Fig. 11).

The non-overlapping distributions of Melongena corona and Styela plicata in St. Andrew Sound, when considered in light of the foregoing observations, strongly suggests that the distribution and local abundance of S. plicata is in large part determined by the distribution and abundance of its only observed predator, the Florida crowned conch.

Figure 10: Craters of Musculus eaten by M. corona



Figure 11: Melongena feeding on S. plicata



CONCLUSIONS AND COMMENTS

Populations of the tunicates, Styela plicata and Molgula occidentalis, consume an enormous number of phytoplankton each day, and play an important trophodynamic role in the ecology of St. Andrew Sound. In addition to the removal of excess nutrients through the consumption of primary producers, they clarify the water column, which favors the growth of seagrasses and the populations of fishes and invertebrates which depend upon them for forage and refuge.

The ability of sea squirts to concentrate heavy metals may permit them to be used as sensitive bioindicators of metals-enriched coastal ecosystems. If the abundance of Styela, Molgula or other species of widely distributed eurytopic tunicates was increased using established, low-cost mariculture techniques (Calinski, pers. comm.), it should be possible to improve the water

quality of anthropogenically impacted estuaries and lagoons. Thus, an under-investigated alternative to microbially mediated environmental bioremediation seems not only feasible, but practical. Further, it may be possible and cost-effective to harvest tunicates during or after bioremediation and extract metals of strategic value from them.

S. plicata thrives in periodically stagnant ecosystems, supports a diverse community of inquilines, and has few known predators. The biology of the symbiotic relationship between S. plicata and Musculus lateralis, as well as the influence of boundary layers on the filtration efficiencies of both species warrants additional research.

ACKNOWLEDGEMENTS

While it is not possible here to acknowledge by name all of the many individuals and organizations who contributed to the work, good fortune allowed me to let them know of my appreciation for their efforts in person. Among those deserving special thanks are AFCESA Research Librarian Janet Davis who proved that wizards still exist. Ray Hauk encouraged our use of the TAFB boathouse facilities; and Jimmy Cornette provided unrestricted use of his remote-site laboratory. Without these facilities the work could not have been accomplished. Dwain Fletcher Co. photography - graphics laboratory personnel expertly processed the hundreds of photographs which were crucial to the work. I thank TSgt. Terry Junn and Mr. Bob Donn for logistical support; SMsgt Daniel Sheldon (OSW/CWSO) for base weather summaries; and Wildlife Biologist Stephen Shea for providing information on sea turtles. My thanks to Dr. Richard Brewer and Ms. Cecilia Dietz (ASI) for providing the aircraft from which we photographed the sampling area. Suzanne and Tyler Collard spent many hours collecting information and animals in the field, and they have my thanks. I thank The University of West Florida for providing the use of their boats and sampling gear; and for funding the cost of the cameras used to document results. I am grateful to Dr. Jimmy Cornette and Col. Niel Lamb for supporting and facilitating funding for the work through AFOSR. Last, and most importantly, I thank Joyce Riesinger ("J.R.") for her professionalism, and her enthusiastic and invaluable contributions to all phases of the work.

Styela plicata AND Molgula occidentalis
(UROCHORDATA: ASCIDACEA: STOLIDIFERA) IN ST. ANDREW SOUND, FLORIDA

Sneed B. Collard

ERRATA

Page 19-14 should read:

"**III. Bioluminescence** -- It was noted earlier that dinoflagellate bloom conditions occurred in the Sound throughout the summer months. These dinoflagellates were bioluminescent. By chance it was discovered during an early morning (0200) visit to the laboratory that water presumptively cleared of phytoplankton by tunicates showed no bioluminescence when mechanically agitated, and that control aquaria remained brilliantly bioluminescent. Water from experimental aquaria containing, respectively, 7 and 11 S. plicata was examined under 400 and 1000X magnification and found to be devoid of diatoms and dinoflagellates. Water taken from the bottom of these aquaria revealed low numbers of empty diatom frustules, a few live diatoms, and no dinoflagellates or sediments. Water had been changed in the aquaria described at approximately 2300, three hours earlier."

"**Styela plicata aggregate filtering rates** -- Given a population of about 8.5 X 10^3 S. plicata in a sampling area containing 2 X 10^9 L of water; an effective filtration rate of 69 L d^{-1} (18.2 gal) per individual; and an average phytoplankton density of 5.5 X 10^6 cells L^{-1}, the standing Styela population should clear the water column in about 34.5 d (58 X 10^6 L/day)."

AN SFE-GC METHOD FOR MONITORING THE WEATHERING OF JET FUELS IN SOIL

Larry E. Gerdom
Associate Professor of Chemistry
Department of Physical Sciences

University of Mobile
P.O. Box 13220
Mobile, Alabama 36663

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory
Environics Directorate

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

July 1993

AN SFE-GC METHOD FOR MONITORING THE WEATHERING OF JET FUELS IN SOIL

Larry E. Gerdom
Associate Professor of Chemistry
Department of Physical Sciences
University of Mobile

ABSTRACT

An SFE-GC on-line procedure for the analysis of jet fuels in a soil matrix was developed. The method has been found to quantitative for the volatile components of jet fuel at the parts per million level. The on-line procedure directly traps the analytes from the supercritical fluid extraction onto the head of the GC column. This is accomplished by using a heated transfer line to transfer the SFE analytes directly to the injection port of a GC during the dynamic extraction of a soil sample. The operating parameters and instrumentation are detailed within this report. GC chromatograms for prepared standard samples are presented.

AN SFE-GC METHOD FOR MONITORING THE WEATHERING OF JET FUELS IN SOIL

Larry E. Gerdom

INTRODUCTION

Supercritical fluids are gases under high pressures(typically 100 atm or more) and temperatures above their critical temperature. The unique solvent properties of supercritical fluids stem from the fact that these fluids have densities comparable to liquids and mass transfer properties similar to gases[1]. The supercritical fluid of choice for these studies is carbon dioxide because it is nonflammable, chemically inert, there are no toxicity/pollution considerations, and it is readily available as a high purity liquid specially formulated for supercritical fluid extraction work[1].

The type of analysis described in this paper involves three basic steps[2]. First is the extraction step. This involves using the supercritical fluid to dissolve and remove the analytes of interest from the soil matrix. It is important that in this portion of the analysis all of the analyte is extracted from the soil. The soil sample is placed into an extraction cell that is contained in a temperature controlled oven. Extraction cells may have a volume of from 0.5 to 50 milliliters. Supercritical fluid is pumped into the extraction cell at a controlled pressure and temperature. The efficiency of extraction is directly related to density and hence both temperature and pressure are important factors in the extraction process. Extractions for this work were performed at 400 atmospheres and 80 °C. Standard methods usually begin with a static extraction of the cell. During this process supercritical fluid is pumped into the extraction cell but no material is allowed to leave the cell. This step allows the supercritical fluid to permeate the sample and to dissolve the analytes. The staic extraction is followed by a dynamic extraction in which more supercritical fluid is pumped through the cell removing the analytes from the soil matrix. The analytes pass from the cell into a restrictor line. The internal diameter of the restrictor line (5-50 μL) allows for enough back pressure that the fluid passing through it remains supercritical until the tip is reached. Then the

carbon dioxide and analytes assume their standard physical states at room temperature. The entire supercritical fluid extraction typically requires from 10 to 45 minutes.

The second step is the collection/concentration of the analytes. As the analytes leave the end of the restrictor tubing they must be trapped for further analysis. There are three standard methods commonly used for trapping. The analytes can be trapped by placing the end of the restrictor into a solvent such as methylene chloride that will dissolve the analytes similar to impinger methods for trapping impurities from air. Since the carbon dioxide is no longer a supercritical fluid after leaving the restrictor but a gas, it is allowed to bubble through the trapping solvent and escape. The trapping solvent chosen should readily dissolve the analytes of interest which are generally not as volatile as the carbon dioxide; and hence, these compounds become trapped in the solvent. A second method of collection that can be employed involves using a cold surface as the analyte trap. The restrictor tubing is feed into an empty vial that is submerged in liquid nitrogen. The analytes become trapped on the sides of the vial along with some solid carbon dioxide. After the extraction is complete the vial is warmed to room temperature and the carbon dioxide is allowed to escape leaving the less volatile analytes adhering to the sides of the vial. A small amount of a suitable solvent (1-2 milliliters) is added to the vial to dissolve and concentrate the analytes for further analysis. Both of these methods require the analytes to be relatively none volatile which is not the case when analyzing jet fuels. Furthermore, the analytes are trapped into a solvent with a volume of at least 1 milliliter but only 1 microliter of this sample will be used in a typical GC analysis. Hence, only one thousandth of the sample extracted is analyzed. A third technique generally referred to as on-line extraction involves cryogenically trapping the extract directly onto a GC column[3]. The extraction tube is attached directly to the injection port of a gas chromatograph and the injection port is heated to minimize the Joule-Thompson cooling which occurs when the supercritical fluid is decompressed. During the dynamic portion of the extraction process the helium carrier gas is switched off and the carbon dioxide from the extractor is allowed to pass through the GC as the carrier gas. The GC oven can be cooled using liquid nitrogen to as low as -50 °C and the analytes of interest are trapped at the head of the capillary column in the GC. When the dynamic extraction is complete the helium carrier gas is turned back on and the GC analysis is performed. Coupled (on-line)

supercritical extraction with capillary gas chromatography (SFE-GC) was pursued because: there is no sample handling required between the extraction and GC analysis; the entire extract is injected into the GC which enhances the sensitivity of the analysis; volatile analytes should be trapped during the extraction with little or no loss of sample; and there are several reports that extraction, concentration, and GC separation can be completed in approximately an hour[4].

The third portion of the analysis involves the GC separation of the analytes. Gas chromatography uses a non-reactive gas as the solvent phase which moves analytes through a column lined with either a polar or non-polar compound referred to as the stationary phase[5]. The time it takes a compound to move through the column is termed its retention time. The GC is fitted with a detector at the end of the column that indicates that a compound is exiting the column. It usually can not identify the compound but simply indicates a compound is present. However, retention times are relatively constant given that the column and chromatographic parameters are kept constant and the detector signals are proportional to concentration. Using a GC/MS system the retention times for compounds of interest can be identified and then a similar GC system can be used to quantitatively identify these substances. Moreover, several different compounds can be analyzed simultaneously within one GC separation provided that calibration curves have been formulated for each compound of interest.

METHODOLOGY

Five standard solutions were prepared by adding benzene, *n*-octane, *p*-xylene, and *n*-hexadecane (5 µL, 25 µL, 50 µL, 75 µL, and 100 µL) with 100 µL of 2,2,4,4,6,8,8-heptamethylnonane to a 10.00 mL volumetric flask and diluting to volume with carbon disulfide. Carbon disulfide was used because it has a low response in the FID detector being used in the GC analysis and it has a low boiling point which means that it will come off the column ahead of the sample of interest. The nonane compound was used as an internal standard and the peak areas of the other four compounds would be compared to it to determine if the method was quantitatively accurate for these standard solutions. The nonane compound was selected because it will probably not be significantly present in most jet fuels

that might be analyzed by this method and because it has a relatively high boiling point it should not be lost during the trapping procedure. The other standard compounds were chosen because they have retention times that do not overlap with a range of from 6.5 to 21.5 minutes in the GC analysis and hence we can determine if the method is quantitative for a wide range of compounds (aromatic and aliphatic) with varying boiling points from 80 °C to 287 °C. These standard samples were used to prepare calibration curves for benzene, *n*-octane, *p*-xylene, and *n*-hexadecane. A sixth standard was prepared which consisted of 100 µL of the nonane compound diluted to volume with carbon disulfide for use in the JP-4 studies.

The soil used in this study was a sand as described in an earlier report[6]. These experiments were performed using extraction cells with a total internal volume of 3.5 milliliters. A small amount of Whatman GF/C filter paper is used on each end of the extraction cell to avoid direct contact between the extraction cell frits and the soil sample to insure that the sand from the samples did not seep into the extractor fluid lines.

A Suprex SFE/50 supercritical extraction unit equipped with an optional 1.0 µL sample injector and an optional SFE-GC Sample Transfer Line Kit was used to perform the soil extractions. Just before each extraction was initiated the helium carrier gas to the GC was shut off. The extraction method consisted of a 20.00 minute static extraction at 400 atm and 80 °C. As soon as this static extraction was commenced a 1.0 µL standard sample was injected into the carrier gas line for the extractor. This static extraction was followed immediately by a dynamic extraction for 16.00 minutes at 400 atm and 80 °C. The supercritical fluid extract from the extraction cell was transfered directly to a GC injection port by using an SFE-GC sample transfer line which maintained a constant temperature of 100 °C along the transfer line. This transfer line was connected directly to the extraction cell. The other end of the transfer line was inserted 35-40 mm into the split capillary injection port of the GC. The GC was programmed to keep the injection port temperature at 270 °C. The GC end of the transfer line was modified to control the rate of flow of fluid through the transfer line. A pair of needle nose pliers were used to c e line which slowed down the flow of carbon dioxide into the GC injection port durin[illegible]e dynamic extraction of the sample. This flow rate was measured with a gas bubble flow meter and found to be 115

mL/min which was in the range of values suggested by the technical staff at Suprex Corporation. At higher flow rates the extracted analytes were not efficiently trapped on the GC column. The purge valve on the split injection port was set to activate at 20 psi. With the helium flow shut off at the beginning of the extraction the pressure differential drops to zero in the injector port before the dynamic extraction begins. During the dynamic extraction the pressure inside the injection port increases rapiidly to 20 psi. At the completion of the dynamic extraction the GC analysis is ready to be started. The helium carrier gas is turned on immediately after the conclusion of the extraction and the GC data acquisition is initiated.

A Hewlett/Packard 5890 Gas Chromatograph with optional cryogenic cooling ability and an FID detector was used to perform the GC separation. The GC was equipped with a capillary column from J&W Scientific. The column was a 20 meter fused silica column with an internal diameter of 0.18 mm. The stationary phase was a 0.4 μm thick coating of DB-5. The GC oven was programmed with an initial temperature of -50 °C and it maintained this temperature for three minutes to allow the helium gas to be restored in the column following the completion of the supercritical fluid extraction of the sample. At this point the GC was programmed to increase the temperature to 40 °C at a rate of 50 degrees/minute. These procedures were followed by a gradual temperature increased to 300 °C at a rate of 10 degrees/minute. The final temperature 300 °C was maintained for 20.00 minutes to insure that all of the analytes were purged the GC column. The entire GC analysis required 50.80 minutes to complete.

This SFE-GC method requires a minimum of 86.80 minutes per analysis which is a relatively rapid procedure for the investigation of compounds in soils. The above procedures were based on similar reports in the literature[7,8,9]. However, this procedure is unique in its ability to effectively extract, trap, and analyze volatile compounds generally found in Jet fuels.

PROCEDURE FOR ANALYSIS

1. Prepare the injector on the extractor for the next run.
   a. Put the injector into the **load** position.
   b. Flush the injector with at least 500 μL of carbon disulfide.

c. Switch valve one from **load** to **inject** with no extraction cell in place and allow carbon dioxide to escape until the tubing tip just begins to freeze then switch the valve back to the **load** position.

d. Flush the injector with 100 µL of the standard solution.

e. Carefully load a syringe with another 80-100 µL of standard and place into the injector port but do not inject it into the port.

2. Load the extraction cell and immediately attach it to the SFE system.

3. Turn off the helium flow to the GC.

4. Press **EQUIL** on the SFE control panel and wait for the ready indicator to appear.

5. When the system is ready press **RUN** on the SFE control panel.

6. Immediately inject the standard sample into the carrier gas line. While slowly pressing down on the syringe press the switch to **INJECT** on the injector control. Make sure that no gas bubbles are in the injection valve when making this injection.

7. During the static injection use the LDS system to set up the assign file for the GC. Also check the pump flow rate on the SFE. It should drop to about 0.200 milliliters/minute or there is a leak in the system. If there is a leak, immediately stop the extraction procedure to avoid the unnecessary loss of supercritical fluid from the extraction system.

8. At the 15 minute mark in the static extraction turn on the cryogenic control for the GC oven. The oven must cool to -50 °C **before** the dynamic extraction begins for the SFE. The head pressure on the GC column **must** also be at zero when the dynamic extraction begins which is usually the case.

9. When the dynamic extraction is complete, immediately turn on the helium flow to the GC and press **START** on the GC control panel.

10. Press **STOP** on the SFE control panel and turn off the oven control unit to allow the SFE oven to cool before the next run.

11. Remove the extraction cell from the SFE system and set the injector control to **load.**

12. To save on liquid nitrogen set the cryogenic control to **off** for the GC oven after the oven temperature reaches 5 °C and before the end of the GC analysis.

RESULTS AND DISCUSSION

This work is an extension of a project that was begun in the summer of 1992. These studies were aimed at developing an efficient method for extracting jet fuels from a soil matrix for qualitative and quantitative analyses. Although a supercritical fluid extraction procedure was developed last summer the sensitivity of the overall method suffered due to the loss of analytes during the trapping procedure. A solvent capable of efficiently trapping these analytes was not be found. In subsequent work the solvent trap was replaced with a cryogenic trap utilizing liquid nitrogen. However, the volatile components of jet fuel still were not being quantitatively trapped by this method. The major problem appeared to be that solid carbon dioxide is also trapped in the vile along with the analytes. When the vile is heated to room temperature the carbon dioxide escapes from the vile and the volatile components from the jet fuel are swept out of the collection vile as well. Given these results, it was decided to investigate the possibility of developing an on-line SFE-GC method in which the analytes would be collected directly onto a cryogenically cooled GC column as the trapping procedure.

It is not possible to give details of every step involved in the development of this method within the space allocated this report. However, it should be noted that the major obstacle to developing this on-line method was the proper interfacing of the out flow from the SFE instrument with injector port of the GC system. The flow rate through the restrictor line from the SFE into the injector port, the flow rate of gas through the purge valve of the GC injector port, and the flow rate of Helium carrier gas for the GC can all significantly effect the sensitivity of this method. Until these parameters were properly adjusted it was not possible to get a quantitative response for these analyses.

The overall efficiency of the SFE-GC procedure was evaluated by preparing standard solutions that contained four of the components found in JP-4 jet fuel as outlined in the methodology section. Figure one is the GC chromatogram produced by this method for a standard sample . This analysis was performed by injecting a 1.0 μL sample into the SFE

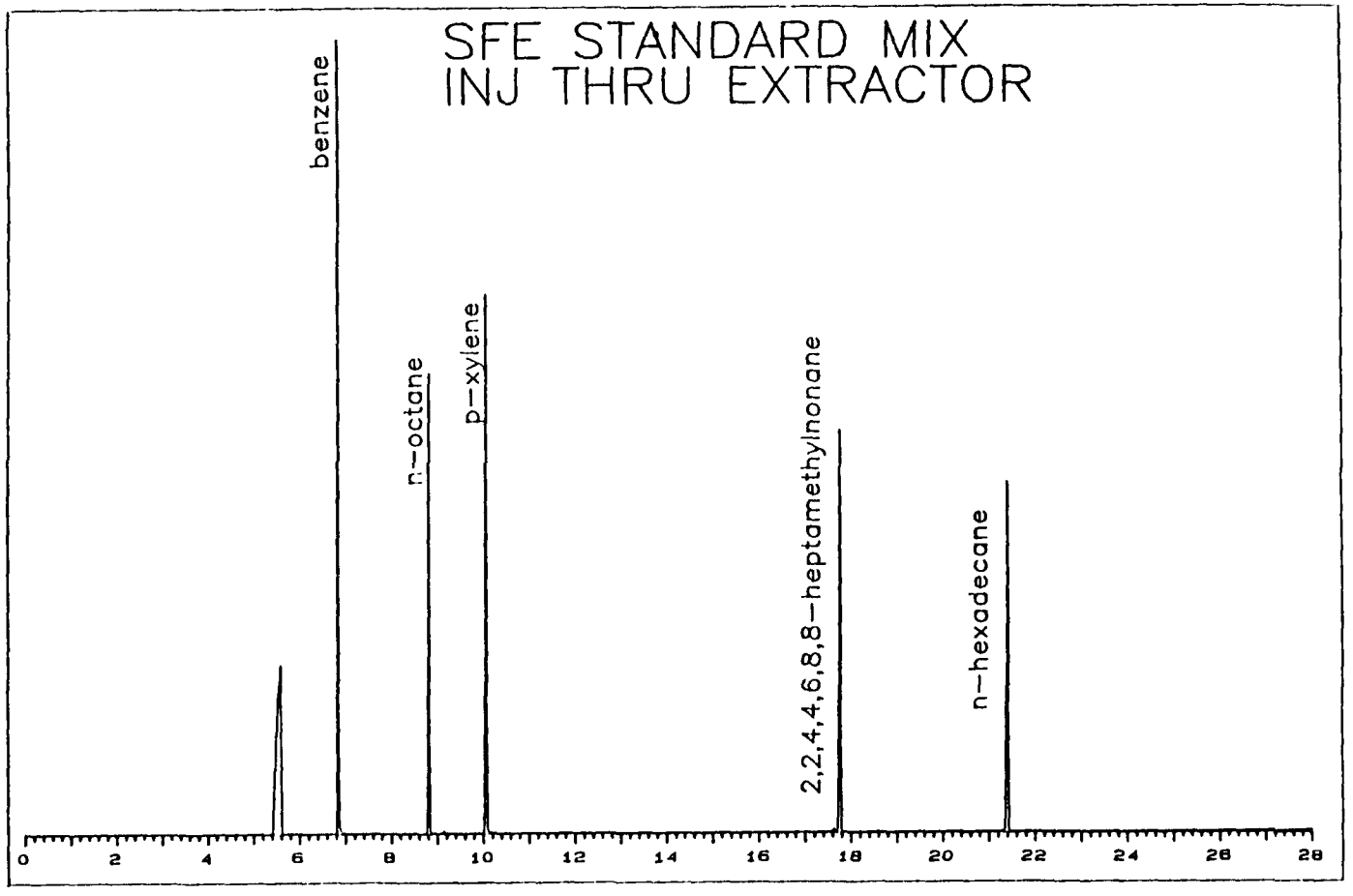


FIGURE 1. GC chromatogram of a 1.0 µL sample injection into the SFE system of standard 100 (100 µL/10.00 mL) utilizing an extraction cell filled with previously extracted sand.

system through the injector loop. The extraction procedure was followed as previously described and the analytes were trapped using the on-line procedures developed. This methodology was followed because the components in jet fuels are so volatile that it was thought that a prepared sand sample would undergo sample loss from evaporation and this would make it nearly impossible to verify the extraction procedure.

The FID detector for the GC gives a signal output that is proportional to concentration. In GC methodology the peak area for a given component is measured and this area is used in formulating calibration curves. Five solutions were prepared with 5-100 μL of each component and 100 μL of the internal standard as described in the methodology section. Figure two a plot of the detector response ratio verses volume of component for each of the four components. This graph shows that over this concentration range there is a linear response between peak height and sample volume. Figures three and four show calibration curves for *n*-octane and *p*-xylene that can be determined from this data.

As stated earlier any on-line technique should be about 1000 times more sensitive than other trapping methods because the entire extract is used in the GC analysis. The sensitivity of this method is illustrated in figure five. These are the GC chromatograms for two different sand samples. In earlier studies there were no organic components that could be detected in the sand samples that had not been spiked with jet fuel. Figure five shows that sand that has not been spiked will contain some organic compounds and that it will be necessary to consider background interferences when using this method of analysis. Furthermore, in studies that were conducted in the summer of 1992 it was not possible to measure fuel components in spiked sand samples that had been exposed to weathering conditions after a period of three days. Using this SFE-GC on-line method it was possible to get a detector response for these fuel components after one year of weathering as evidenced by the top GC chromatogram in figure five.

The overall goal of this project is to study the effects of weathering on jet fuel in soil. Figure six is the GC chromatogram for JP-4 injected directly into the GC sample port. In comparison figure seven is the GC chromatogram for a 1.0 μL sample that was injected into the SFE system through the sample loop normally used to add the internal standard for an extraction run which contained jet fuel and the internal standard. Although there are some

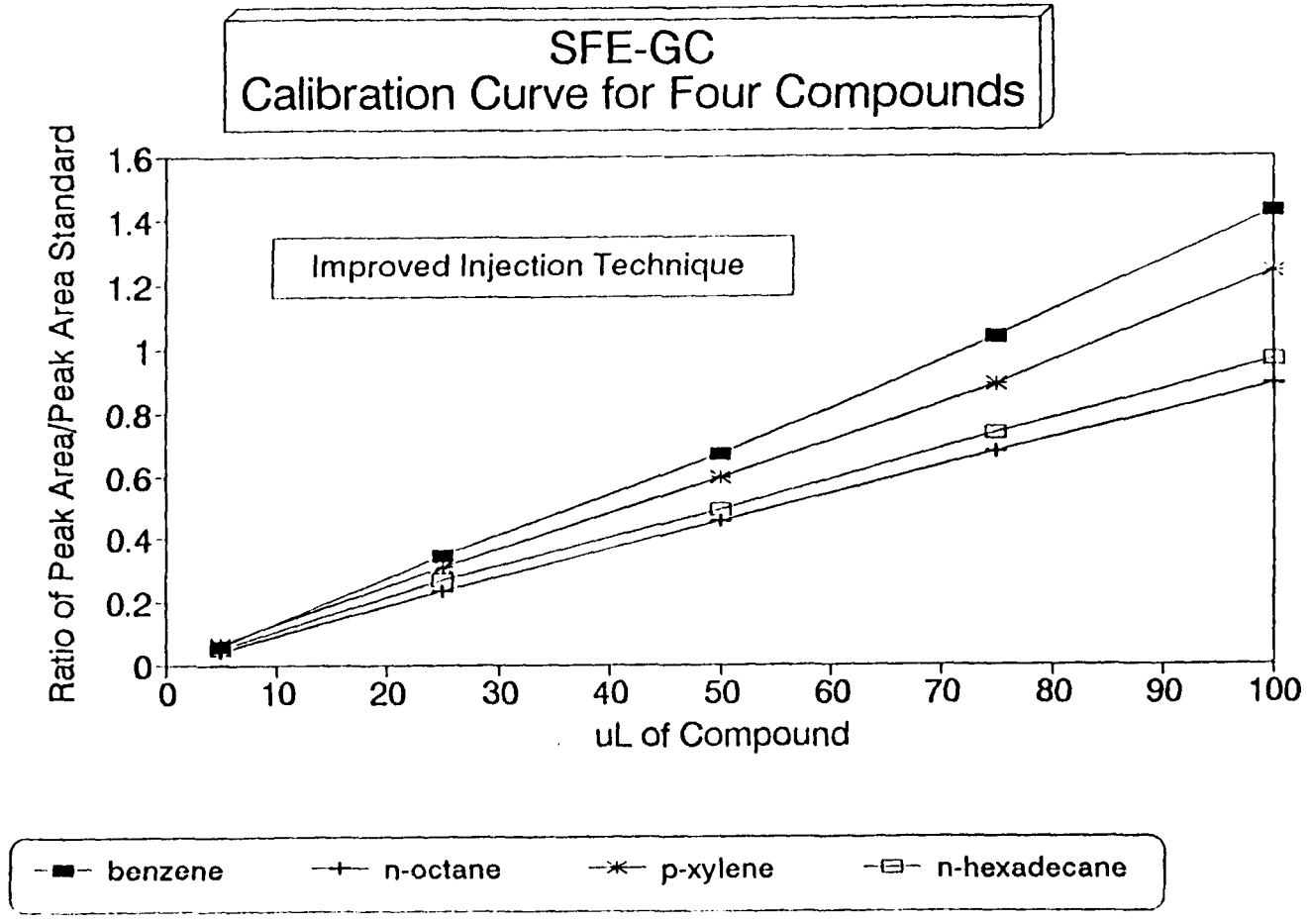


FIGURE 2. Graph of five standard solutions containing 5, 25, 50, 75, and 100 µL of each of four components found in JP-4 fuel.

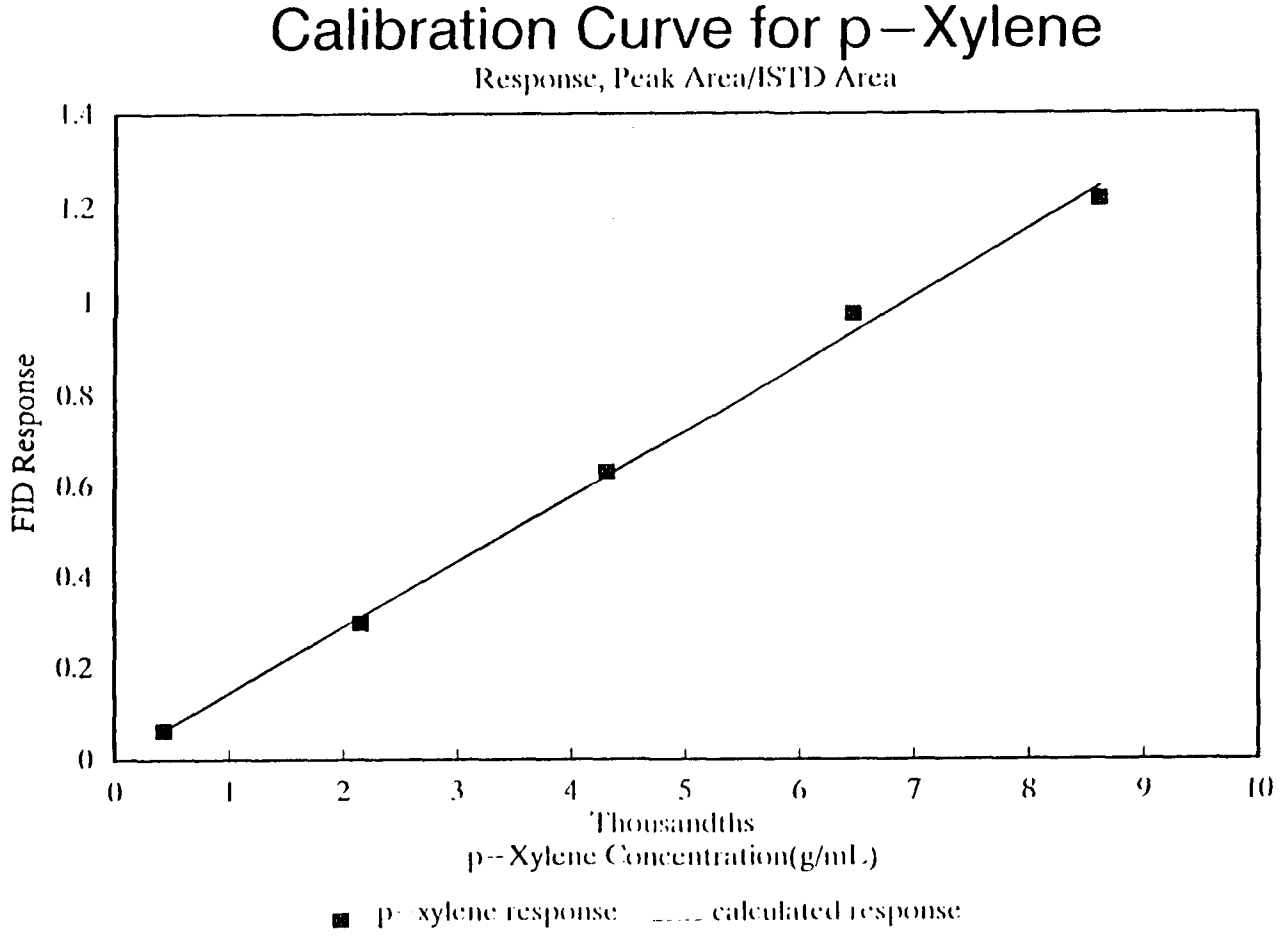


FIGURE 3. Calibration curve for *p*-xylene in grams of analyte per mL of solution.

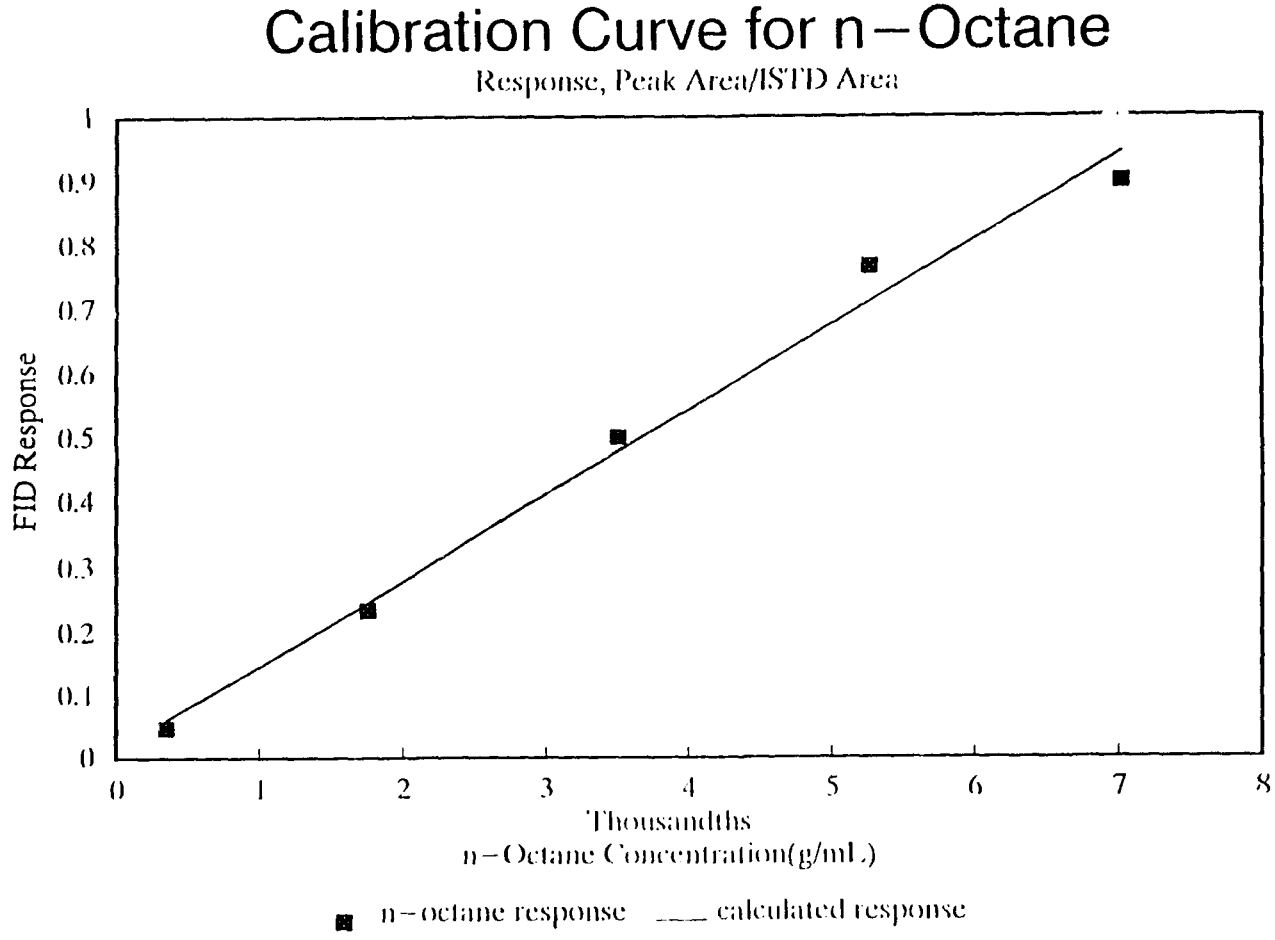


FIGURE 4. Calibration curve for *n*-octane in grams of analyte per mL of solution.

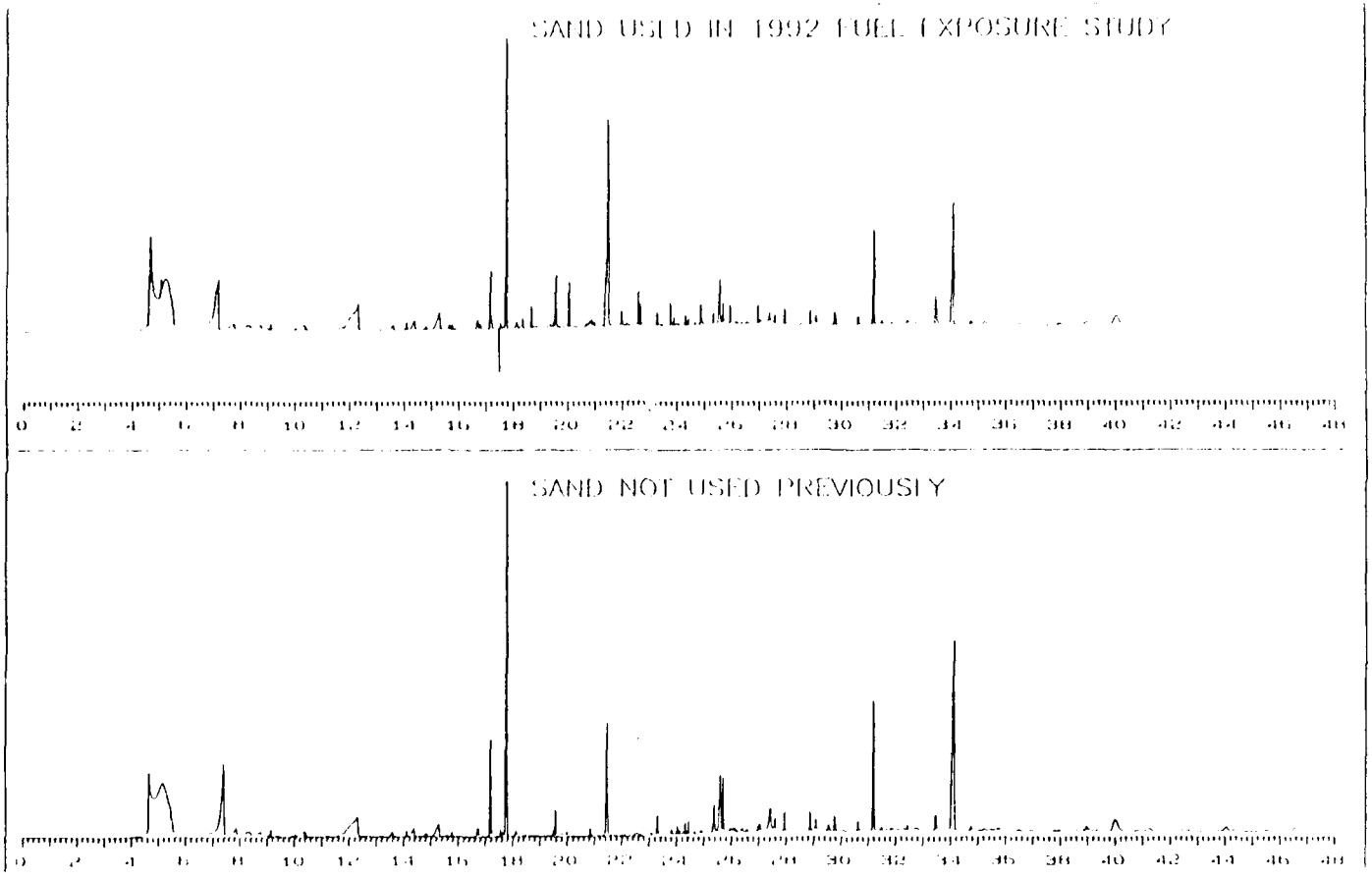


FIGURE 5. The GC chromatograms for sand that had not previously been extracted but had been prepared for use in jet fuel weathering studies (bottom chromatogram) and sand previously injected with 0.5 mL/KG of jet fuel.

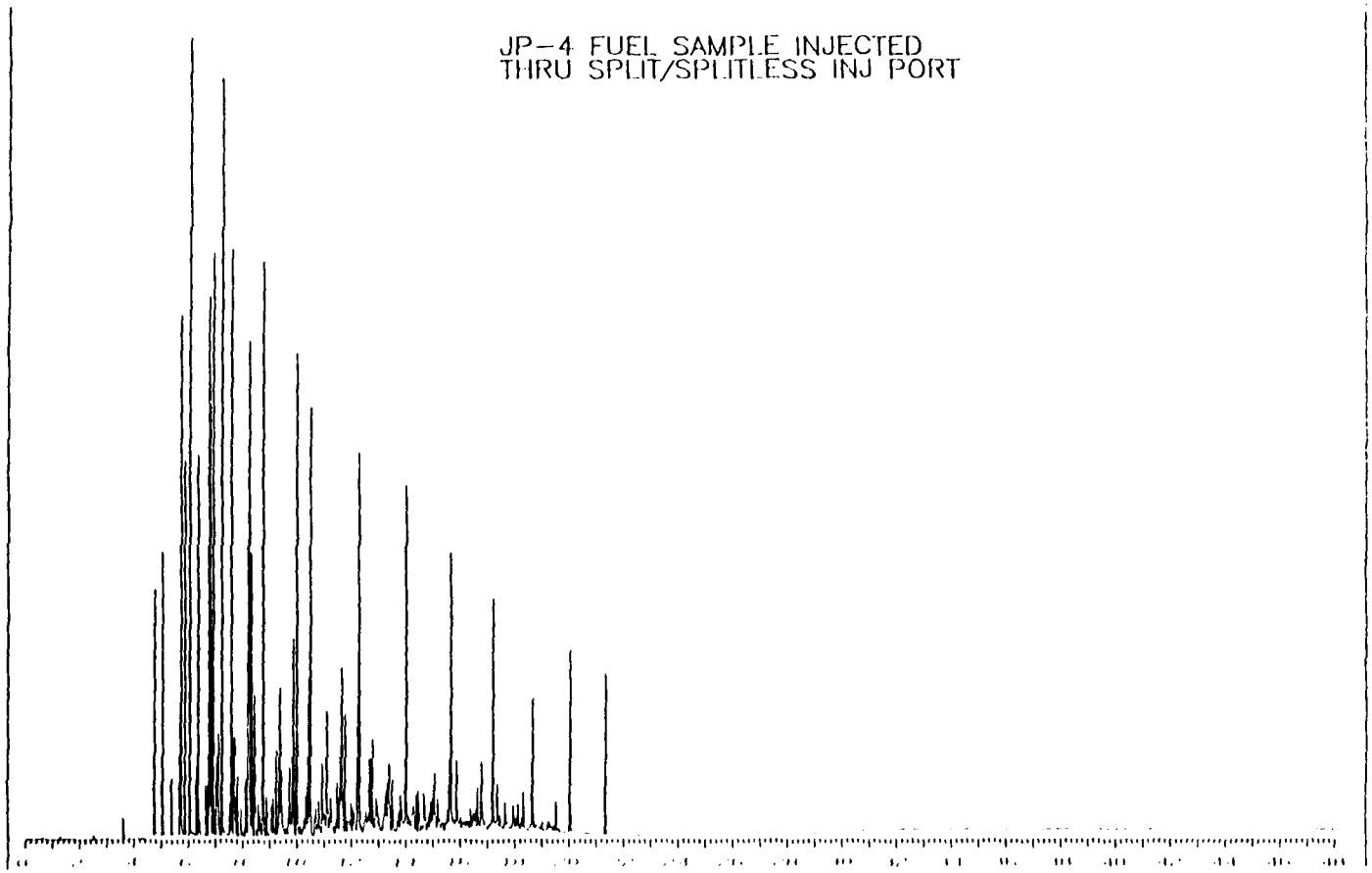


FIGURE 6. GC chromatogram of JP-4 injected directly into the GC for analysis.

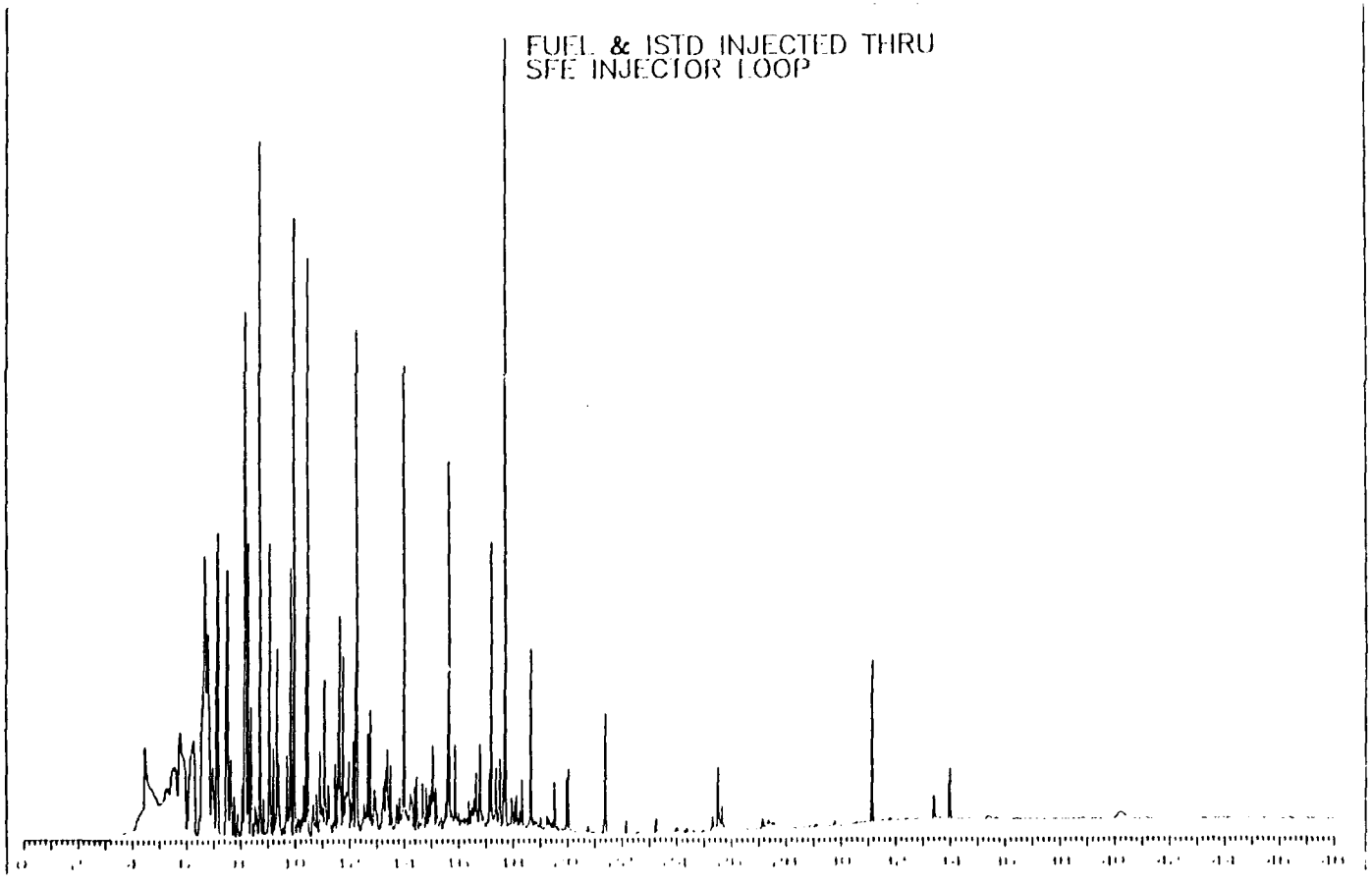


FIGURE 7. GC chromatogram of JP-4 sample injected into the SFE system.

differences in peak heights it is obvious that most of the JP-4 components; if not all, are being detected in the GC analysis after the extraction procedure. Figure eight is the GC chromatogram for an extracted sand sample that has been spiked with a solution of JP-4 and internal standard. This preliminary analysis indicates that the SFE-GC on-line method developed can be used to qualitatively and quantitatively identify jet fuel components in a soil matrix.

The slight differences in these three chromatograms may be as a result of not adjusting the flow rate of helium gas to the GC system. The flow of gas through the purge valve of the injection port is much higher when the helium gas flow is resumed after the dynamic extraction. Since the flow of gas through the purge valve is not the same when comparing the SFE-GC analysis to a simple GC analysis, it not unexpected that there will be slight differences in the chromatograms produced. There simply was not enough time this summer to lower the flow rate of the helium gas and to repeat the calibration experiments. The method could also be significantly improved by adding the ability to automatically control the flow of helium gas to the GC and to synchronize the start of the GC analysis with the end of the SFE dynamic extraction procedure. This would allow the method to be fully automated and it would decrease the amount of operator time required for the analysis.

The final results obtained from this study indicate that an SFE-GC on-line procedure can be used to qualitatively and quantitatively identify jet fuels from a soil matrix. This method will be used to collect data to be used in constructing a model of the effects of weathering on jet fuels in a soil matrix.

ACKNOWLEDGEMENTS

I wish to thank Cadet Allen Stanley for his help in the earlier portions of this study. Dr. Howard T. Mayfield was my laboratory focal point and he served as my mentor while developing these procedures. I am indebted to Armstrong Laboratories-Environics Directorate and the Air Force Office of Scientific Research for funding this project and supporting me through the Summer Faculty Research Program (SFRP).

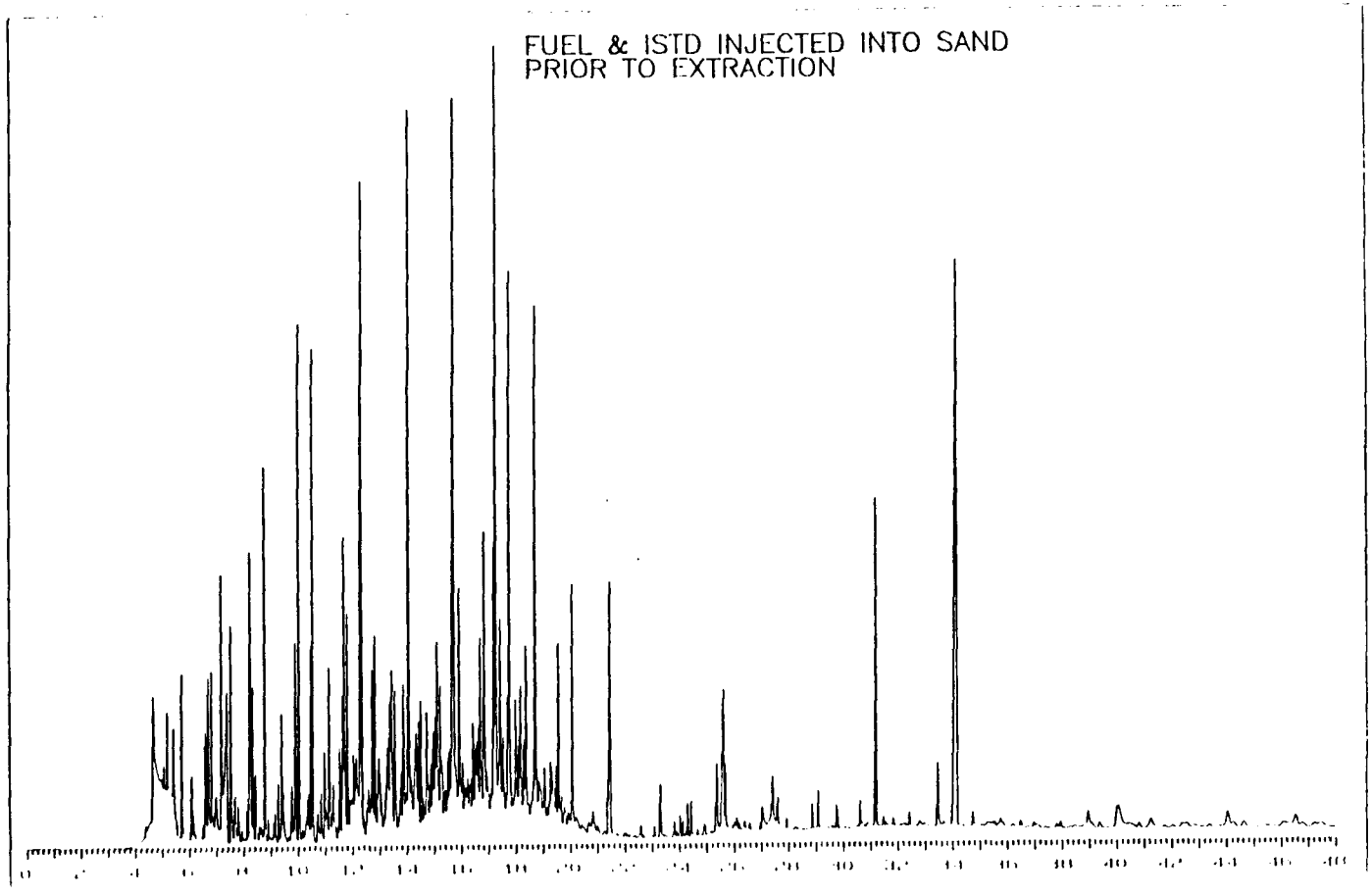


FIGURE 8. GC chromatogram of JP-4 sample injected onto sand and then analyzed by the SFE-GC on-line method.

BIBLIOGRAPHY

1. Larry Taylor. The Supercritical Fluid Extraction and Analysis of Aged Single-Based Propellants. Am. Lab., 25(8), 22-26 (1993).

2. Steven B. Hawthorne. Coupled Supercritical Fluid Extraction-Capillary Gas Chromatography (SFE-GC) in Analysis with Supercritical Fluids: Extraction and Chromatography. pages 61-73. Springer-Verlag: Berlin (1992).

3. J.M. Levy and A.C. Rosselli. Quantitative Supercritical fluid Extraction Coupled to Capillary Gas Chromatography. Chromatographia, 28(11/12), 613-617 (1989).

4. Steven B. Hawthorne, David J. Miller and John J. Lagenfeld. Quantitative Analysis Using Directly Coupled Supercritical Fluid Extraction-Capillary Gas Chromatography (SFE-GC) With a Conventional Split/Splitless Injection Port. J Chrom. Sci., 28, 2-8 (1990).

5. Gary A. Eiceman, R.E. Clement and Herbert H. Hill, Jr. Gas Chromatography. Anal. Chem., 64, 170R-180R (1992).

6. Larry E. Gerdom. A Preliminary study of the Weathering of Jet Fuels in Soil Monitored by SFE with GC Analysis. SFRP Reports. Volume 6, report 10. Air Force Office of Scientific Research, AFOSR-TR-93-0116, Bolling AFB, DC.

7. Steven B. Hawthorne and David J. Miller. Directly Coupled Supercritical Fluid Extraction-Gas Chromatographic Analysis of Polycyclic Aromatic Hydrocarbons and Polychlorinated Biphenyls from Environmental Solids. J. Chrom., 403, 63-76 (1987).

8. Tsuneaki Maeda. Introduction of Directly Coupled SFE/GC Analysis in Hyphenated Techniques in Supercritical Fluid Chromatography and Extraction ed. by K. Jinno. Journal of Chromatography Library Series, Vol. 53, pages 255-274. Elsevier Science Publishers (1992).

9. Steven B. Hawthorne, David J. Miller and Mark S. Krieger. Rapid Extraction and Analysis of Organic Compounds from Solid Samples Using Supercritical Fluid Extraction/Gas Chromatography. Fresenius Z Anal Chem., 330, 211-215 (1988).

NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray
Associate Professor
Department of Civil and Environmental Engineering

West Virginia University
Morgantown, WV 26506-6101

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 1993

NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray
Associate Professor
Department of Civil and Environmental Engineering
West Virginia University

Abstract

Public domain computer programs were used to model the tritium plume observed during Macrodispersion Experiment 2 (MADE-2), a field scale natural gradient experiment conducted at Columbus Air Force Base, Mississippi. The finite difference program MODFLOW was used to simulate the flow of groundwater through a 330 m x 105 m computational domain. The grid had 66 rows, 21 columns, and 9 layers - a total of 12,474 cells. The 468 day experiment was simulated on a Sun Sparcstation 2 in 37 minutes, assuming uniform hydraulic conductivity, and in less than 6 hours with a more realistic, highly heterogeneous conductivity field derived from 67 measured conductivity profiles. Both solutions had small mass balance errors and appeared reasonable, but there was insufficient time to perform satisfactory calibrations.

The mixed Lagrangian-Eulerian finite difference program MT3D was employed to solve the contaminant transport equation using the MODFLOW-predicted flow field. Dispersivities in the longitudinal, transverse horizontal, and transverse vertical directions were assumed based on a previous analysis of the MADE-2 data. Computation times were excessive: to simulate 80 days required 6.4 hours for the uniform conductivity model and 17 hours for the heterogeneous conductivity model. In both solutions there were a few cells with negative concentrations, and the mass balances varied erratically. Both models reproduced the general features of the observed plume on simulation day 42, but the heterogeneous conductivity plume was definitely more realistic. Both models exhibited excessive upstream dispersion.

Further work is needed to establish a properly calibrated model for the MADE-2 experiment, and to develop practical modeling tools for the generic plume prediction problem.

NUMERICAL MODELING OF GROUNDWATER FLOW AND TRANSPORT AT THE MADE-2 SITE

Donald D. Gray

INTRODUCTION

In order to more effectively and economically remediate groundwater pollution on its properties, the Air Force has established a substantial program of research on groundwater flow and contaminant transport. A major component of this program has been participation, together with the Electric Power Research Institute and the Tennessee Valley Authority, in Macrodispersion Experiment 2 (MADE-2). MADE-2 was a field-scale natural gradient experiment conducted in 1990-91 at Columbus Air Force Base in Columbus, Mississippi.

As the numerical suffix implies, there was an earlier macrodispersion experiment conducted at the same location. The results of MADE-1 have recently been documented in a series of journal articles (Boggs and others, 1992; Adams and Gelhar, 1992; Rehfeldt, Boggs, and Gelhar, 1992; Boggs and Adams, 1992). The data gathered in MADE-1, especially those which characterize the site, were valuable in planning MADE-2 and in analyzing the results.

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 upper layer of soil was a shallow alluvial terrace containing an unconfined aquifer about 11 m thick. This was underlain 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.

Heterogeneity was found to be the outstanding characteristic of the subsurface environment. The borehole flowmeter technique was used to measure the hydraulic conductivity every 15 cm in numerous test wells scattered irregularly over the site. Conductivity variations of up to four orders of magnitude were found in individual profiles. Based on 49 profiles with a total of 2187 measurements, Rehfelt, Boggs, and Gelhar (1992) computed statistical measures of heterogeneity and compared them with values from other natural gradient macrodispersion experiments. As Table 1 shows, the MADE-2 site is much more heterogeneous than the others. Both the variability of the conductivity and the scales over which it is correlated are comparatively large.

Table 1. Statistical parameters of the hydraulic conductivity field at MADE-2 and other macrodispersion experiments.

K = hydraulic conductivity [cm/s]
L_h = horizontal correlation scale [m]
L_v = vertical correlation scale [m]

location	variance(ln(K))	L_h	L_v
MADE-2	4.5	12.8 m	1.6 m
Borden	0.29	2.8 m	0.12 m
Cape Cod	0.26	5.1 m	0.26 m
Twin Lakes	0.031	3.0 m	0.91 m

The purpose of MADE-2 was to better understand the fate and transport of dissolved organic compounds typical of jet fuels and solvents. To achieve this, 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 (essentially a passive tracer), benzene, naphthalene, p-xylene, and o-dichlorobenzene. The three dimensional spread of the plume 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. The isopleths of concentration at 59.5 m in Figure 1 shows that the tritium plume spread in an essentially linear fashion during MADE-2. Figure 2 discloses a complex vertical structure along the plume axis.

By numerically integrating the observed tritium activity distributions, Boggs and others (1993) estimated that the ratios of observed mass to injected mass in the first 4 snapshots were 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 caused by the migration of the leading edge of the plume past the farthest downstream samplers. There was no attempt to define the entire tritium plume in snapshot 5.

My objective as a 1993 Summer Faculty Fellow was to assess how well the MADE-2 tritium plume could be predicted using available computer codes for groundwater flow and contaminant transport. A second goal was to determine the degree of accuracy which could be achieved in such a heterogeneous aquifer if extensive measurements of hydraulic conductivity were not available. The results described here build upon my work as a 1992 Summer Faculty Fellow (Gray, 1992).

FLOW MODELING

The density of groundwater depends on solute concentrations and on temperature. Density variations cause buoyancy forces which affect the pattern of flow and hence the spread of the solutes. If buoyancy effects are important, the groundwater flow and solute transport equations are coupled and must be solved simultaneously. But if the effects of density variations are small, the flow equation can be solved first, without reference to the concentration field. Based on this velocity field, the transport equation can then be solved for the concentrations. This forced convection approximation greatly simplifies the calculations and was adopted here. The computer programs used in this work were MODFLOW for the flow problem and MT3D for the transport problem.

MODFLOW (McDonald and Harbaugh, 1988) is a U. S. Geological Survey (USGS) public domain program for the solution of the groundwater flow equation. Coded in Fortran 77, MODFLOW's name refers to its modular structure which facilitates the writing of new subroutines to handle specific tasks. Since 1988 several new modules have been published which greatly extend the types of problems which can be treated, but they were not obtained in time for use in this study.

The MODFLOW used here is essentially the version described in the 1988 report, which 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. MODFLOW is noted for its flexibility. It can solve either steady or transient cases and provides options for recharge, wells, streams, and other hydrologic features. Both confined and unconfined aquifers can be modeled. The block centered flow module used here (BCF1) allows the dewatering of layers during periods of water table decline, but cannot handle rewetting due to a rising water table. BCF2, one of the newer modules (MacDonald, Harbaugh, Orr, and Ackerman; 1991), does allow rewetting. Flexibility, robustness, clarity of coding, and outstanding documentation all contributed to making MODFLOW a natural choice for this project.

To run MODFLOW, the user must specify the grid geometry, boundary and initial conditions, values related to the principal hydraulic conductivities for each cell, storage coefficients for each cell, and any sources of water.

The first step in applying MODFLOW to any problem is to define a suitable grid. Given the heterogeneity of the site and the observations of the plume, it was clear that a uniform three dimensional grid was needed. The chosen grid consists of 9 layers, each containing 66 rows and 21 columns of 5 m x 5 m cells, for a

total of 12,474 cells. The 105 m and 330 m sides of the computational domain parallel 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 the 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.

The computational domain is bounded below by an impermeable plane at 51.0 m MSL, and the lower 8 layers are each 1 m thick. The top layer, whose base is at 59.0 m MSL, is unconfined and has an upper boundary which fluctuates with the water table. In some cases cells in the top layer were as much as 6 m thick. This is undesirable from the standpoint of accuracy; but it was unavoidable due to the limitations of BCF1, which required that the lower boundary of the top layer be low enough to insure that dewatering never occurred.

Piezometric heads were recorded continuously in 15 monitoring wells. In addition, 17 manual surveys were made at intervals of about one month during MADE-2. These surveys included up to 48 wells scattered irregularly over and near the computational domain. 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 to accurately simulate MADE-2.

The piezometric head surveys were krigged using the commercial program SURFER in order to obtain water table elevations at each node in the top layer. The results from the first survey were used to establish the initial heads at every node. The krigged results from the later surveys were used to fix the boundary node heads for each stress period using MODFLOW's General Head Boundary module. In assigning the initial and boundary conditions, it was assumed that the head was constant with depth. The krigging procedure used a linear variogram based on the nearest points (up to 10 points) within 100 m in each octant. These parameters cannot be rigorously justified. Figure 3 shows water table contours from the first two surveys (June 19 and July 23, 1990).

The gain or loss of water through recharge was estimated from meteorological data. Daily temperature and precipitation 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, was 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 net recharge was calculated for each day as the difference between the precipitation and 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. 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 2 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 into the cell at row 61, column 11, and layer 2. A constant time step of 2 days was used in the MODFLOW simulations reported here.

Table 2. 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 inject.	"	"	-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				

Hydraulic conductivity profiles were available from 67 wells scattered irregularly over and near the computational domain. The data points represented averages over successive 15 cm layers except for gaps where the well screens were jointed. The height profiled and the layer boundaries varied from well to well. The gaps were filled in using the values immediately above and below in an alternating manner. The profiles were extended upward to 60.0 m or to the next higher integer elevation using the conductivity of the highest layer, and downward to 51.0 m using the smallest detectable conductivity (8.64×10^{-3} m/day).

The extended profiles were averaged arithmetically over each MODFLOW layer to generate horizontal conductivities. Assuming 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 conductivities. Exceptions were made for the top layer in which the average extended to the top of the extended profile, and for the bottom layer where the vertical conductivity vanished by virtue of the underlying impermeable plane.

SURFER was used to interpolate and extrapolate these profiles horizontally so as to obtain horizontal and vertical conductivities at each node. The measured values were log-transformed, krigged, and transformed back. The transformation was necessary to avoid negative values. The krigging was done using linear variograms based on the nearest points (up to 10) in each octant. These parameters cannot be rigorously justified, but they gave reasonable-looking results. The inverse square gridding procedure was also tried, but produced lumpier distributions. Figure 4 shows distributions of horizontal and vertical conductivity in the top layer. In general, the conductivities in the lower layers were much smaller.

For the top layer, MODFLOW calls for the horizontal transmissivity, which is the product of the conductivity and the saturated thickness. Similarly, the program calls for the vertical leakance, which is the vertical conductivity divided by the thickness between adjacent nodes. Since the actual thickness of the top layer was unknown and varied with time, a value of 1 m was arbitrarily used in both cases.

In few practical cases would there be such extensive data on hydraulic conductivity. Thus it is of interest to see how a simulation based on fewer data would compare. At the MADE-2 site a traditional pump test (AT-2) provided the data needed to investigate this question. The elliptical drawdown contours observed in AT-2 indicated that the horizontal conductivity was heterotropic

with principal axes approximately aligned with the MADE-2 coordinate system. Boggs and others (1990) reported principal conductivities K_x= 17.28 m/day, K_y = 44.9 m/day, and K_z = 2.42 m/day. These conductivities were used for every cell in the uniform conductivity simulations. As there were no other measurements of specific yield, the AT-2 value of S_y = 0.10 was used as the primary storage coefficient for every top layer cell in both the uniform and variable conductivity simulations. For the lower layers in both simulations, a primary storage coefficient of 0.0001 was assumed, based on textbook values for specific storage (Anderson and Woessner, 1992).

The 468 day experiment was simulated on a Sun Sparcstation 2 using the MODFLOW Preconditioned Conjugate Gradient solver option (PCG2). The uniform conductivity model required 37 minutes to execute. The maximum volumetric rate discrepancy in any time step was -3.05 %. The maximum cumulative volumetric discrepancy was -0.45 %, and the final value was -0.35 %. The variable conductivity model ran in less than 6 hours. The maximum volumetric rate discrepancy in any time step was -7.02 %. The maximum cumulative volumetric discrepancy was +1.93 %, and the final value was -1.93 %. These values indicate acceptable levels of internal consistency, but there was not enough time to calibrate the models.

Figure 5 presents the water table contours on July 23, 1990, (simulation day 42) for both models. Both predicted water tables indicate a predominant flow in the positive y-direction, bending toward the negative x-direction at large values of y. The models generally agree within 0.1 m with each other and with the krigged observations shown in Figure 3. The biggest disagreement is in the shape of the 61.6 m and 61.7 m contours. The constant head contours in vertical planes indicate that the flow is predominantly horizontal.

TRANSPORT MODELING

MT3D is a public domain program developed for the U. S. Environmental Protection Agency (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 three dimensional, rectangular grid, block-centered, finite difference 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 1.1 was used in this research.

Because of the well known computational difficulties of numerical dispersion and oscillation in advection-dominated flows, MT3D incorporates several options for calculating the advection term. The three preferred methods are Lagrangian particle tracking schemes. The first, called the Method of Characteristics (MOC), tracks a large number of imaginary tracer particles forward in time. A second option, 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 successful 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 option is also 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, an option which distinguishes horizontal and vertical transverse dispersivities is provided. The use of an explicit finite difference 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 solving the transport equation.

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. First order irreversible rate reactions are those in which the rate of mass loss is linearly proportional to the mass present. Typical of this class are radioactive decay and certain types of biodegradation. MT3D allows different decay rates for the dissolved and sorbed phases.

MT3D requires information beyond that needed for and calculated by MODFLOW. The porosity of each cell must be specified in order to calculate seepage velocities. Unfortunately porosities were measured at the site in only 4 core holes. The 84 samples had a mean porosity of 0.32, and so this value was assigned to every cell in the grid. Based on the MADE-2 observations and an assumed two dimensional analytical model for the plume, Boggs and others (1993) estimated the longitudinal dispersivity to be 10 m and the transverse horizontal dispersivity to be less than 2.2 m. The dispersivities used in MT3D 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, the top layer was specified to have a uniform thickness of 4 m, which is too large for the period which was actually simulated.

MT3D was applied to predict the tritium plume only. 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/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 was 2×10^{-6} Ci/m^3. This background was assigned to all natural sources, including recharge. Water leaving the domain carried the concentration of the cell it last occupied. Sorption does not affect tritiated water, but tritium is radioactive and decays with a 12.26 year half-life.

MT3D ran so slowly compared to MODFLOW that it was not practical to simulate the entire MADE-2 experiment. In fact only the first 4 stress periods (80 days) were computed. Due to a blunder in preparing the input file, only the first three stress periods (52 days) are meaningful. The uniform conductivity model, using the HMOC algorithm, ran in 6.4 hours and had a maximum cumulative mass discrepancy of -6.07 %. The variable conductivity case, using the MOC method, required 17 hours to simulate 80 days. It had a maximum cumulative mass discrepancy of +6.43 %. In both cases a small number of cells experienced negative concentrations. Although these simulations were rerun using several algorithms and convergence parameters, the results described were the best achieved. Of course, many other choices could have been tried, and some might have been more satisfactory. The mass discrepancies seem rather large, and varied erratically, but that stems from the discrete nature of the particle tracking methods. Dr. Chunmiao Zheng, the author of MT3D, has told the present writer that discrepancies of up to 15 % are considered acceptable.

Figure 6 shows the contours of relative concentration in the top layer, and Figure 7 shows the vertical profiles along column 11 for both models for simulation day 42. The plume has a greater spread and dilution in the uniform conductivity model. Comparing these predictions with the snapshot 1 observations in Figures 1 and 2 shows that the variable conductivity model is much closer to the real plume. Nevertheless, it shows too much spread, especially in the upstream direction.

CONCLUSIONS

1. Modelers should be involved in planning the monitoring network for field experiments. The problems of defining boundary conditions and property distributions could be reduced by considering the needs of numerical models.

2. Three dimensional groundwater flow simulations using MODFLOW are practical and self consistent. The computational burdens of the flow calculation were large but bearable.

3. The rewetting capability of the newer BCF2 module would add significantly to the accuracy of modeling flow near a fluctuating water table using MODFLOW. This capability should be exploited in future simulations of MADE-2.

4. Despite the satisfactory performance of MODFLOW, the model has not been calibrated for MADE-2. Given the uncertainty of many of the properties and boundary conditions, it is essential to investigate the sensitivity of the solution to their values. The effect of mesh size should also be tested.

5. Three dimensional transport simulations using MT3D Version 1.1 do not seem to be practical and may not be sufficiently self consistent. The transport simulations required too long to run and exhibited too many anomalies to be considered satisfactory. These problems may reflect the limitations of the modeler rather than the model; but, if so, it is still fair to conclude that MT3D is too difficult to use properly in applications of this type. Obviously, the transport model has not been calibrated for MADE-2.

6. A fully satisfactory transport model should be able to simulate several contaminants simultaneously.

7. Using a uniform conductivity to model a very heterogeneous site appears to be unsatisfactory if transport predictions are needed.

8. The ability to visualize and interpret three dimensional data and predictions still lags far behind our ability to generate numbers.

ACKNOWLEDGEMENTS

I am pleased to thank Dr. Tom Stauffer (who made this work possible), Dr. Howard Mayfield, and Mr. Chris Antworth of Armstrong Laboratory for their assistance. I am also grateful to Dr. V. S. Manoranjan of Washington State University, Dr. Sungkwan Kang, and Dr. Kirk Hatfield of the University of Florida for their generous help and advice.

REFERENCES

E. E. Adams and L. W. Gelhar, 1992. Field Study of Dispersion in a Heterogeneous Aquifer 2. Spatial Moments Analysis, Water Resources Research, 28, 3293-3307.

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 and E. E. Adams, 1992. Field Study of Dispersion in a Heterogeneous Aquifer 4. Investigation of Adsorption and Sampling Bias, Water Resources Research, 28, 3325-3336.

J. M. Boggs, L. M. Beard, W. R. Waldrop, T. B. Stauffer, W. G. MacIntyre, and C. P. Antworth; 1993. Transport of Tritium and Four Organic Compounds During a Natural-Gradient Experiment (MADE-2), EPRI draft report.

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.

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. MacDonald, 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.

C. Zheng, 1990. MT3D, A Modular Three-Dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems, prepared for U. S. EPA Robert S. Kerr Environmental Research Laboratory, Ada, Oklahoma, by S. S. Papadopulos & Associates, Inc., Rockville, Maryland.

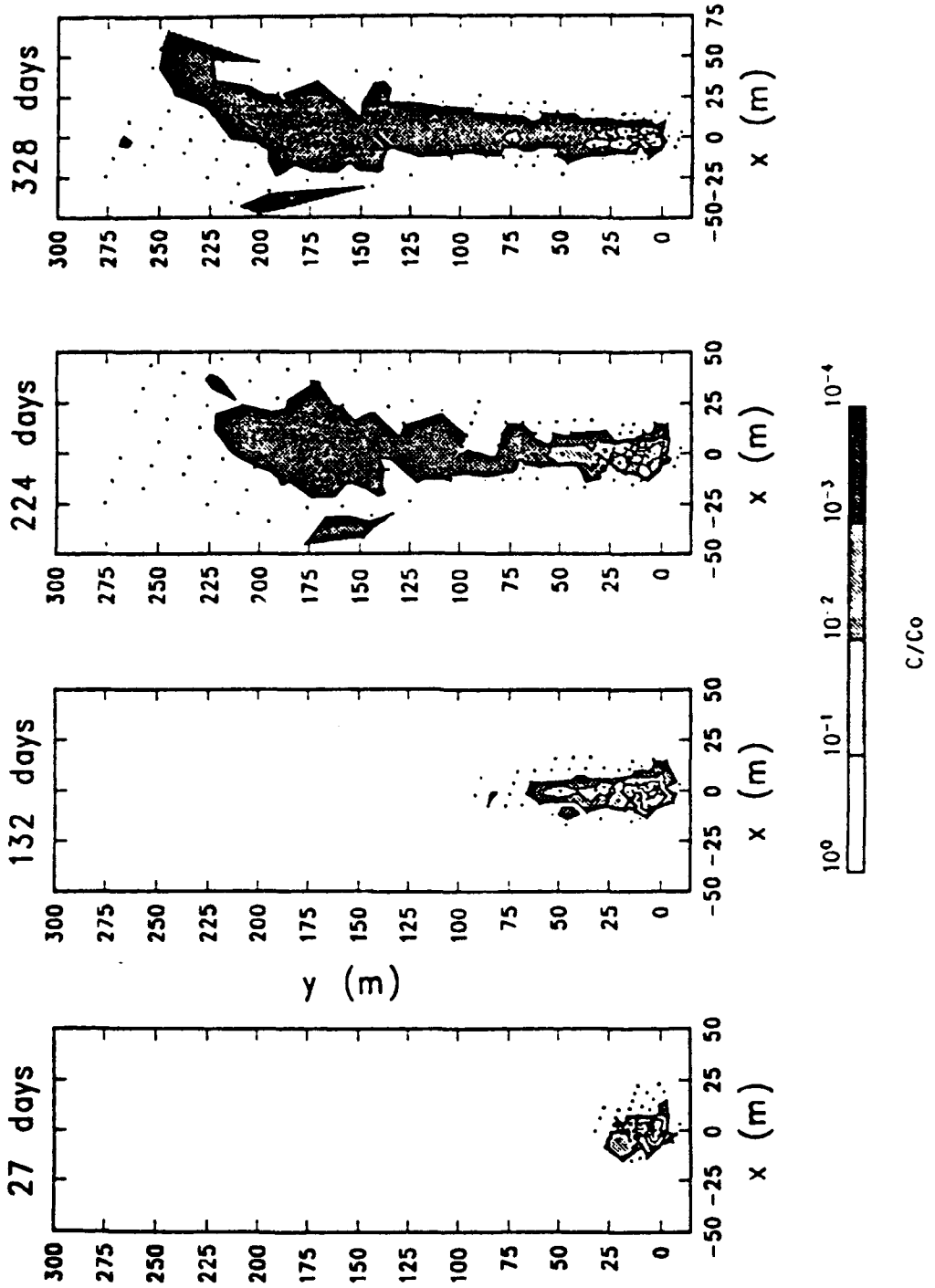


Figure 1. Tritium relative activity at elevation 59.5 m during MADE-2. Times are given in days after start of injection: 27 days after injection = simulation day 42, etc. Source: Boggs and others, 1993.

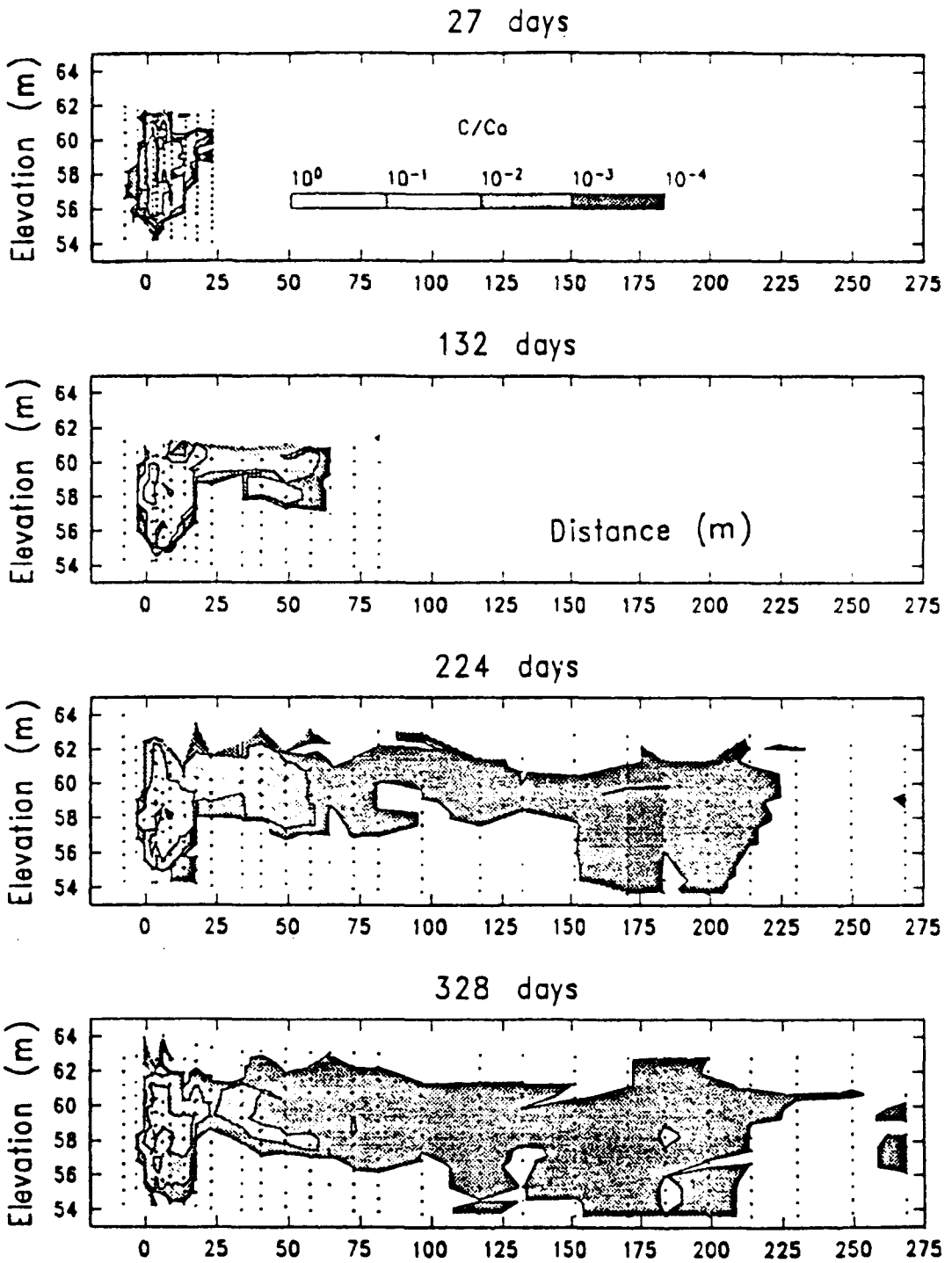


Figure 2. Tritium relative activity profiles along the plume axis during MADE-2. Times are days after start of injection: 27 days after injection = simulation day 42, etc. Source: Boggs and others, 1993.

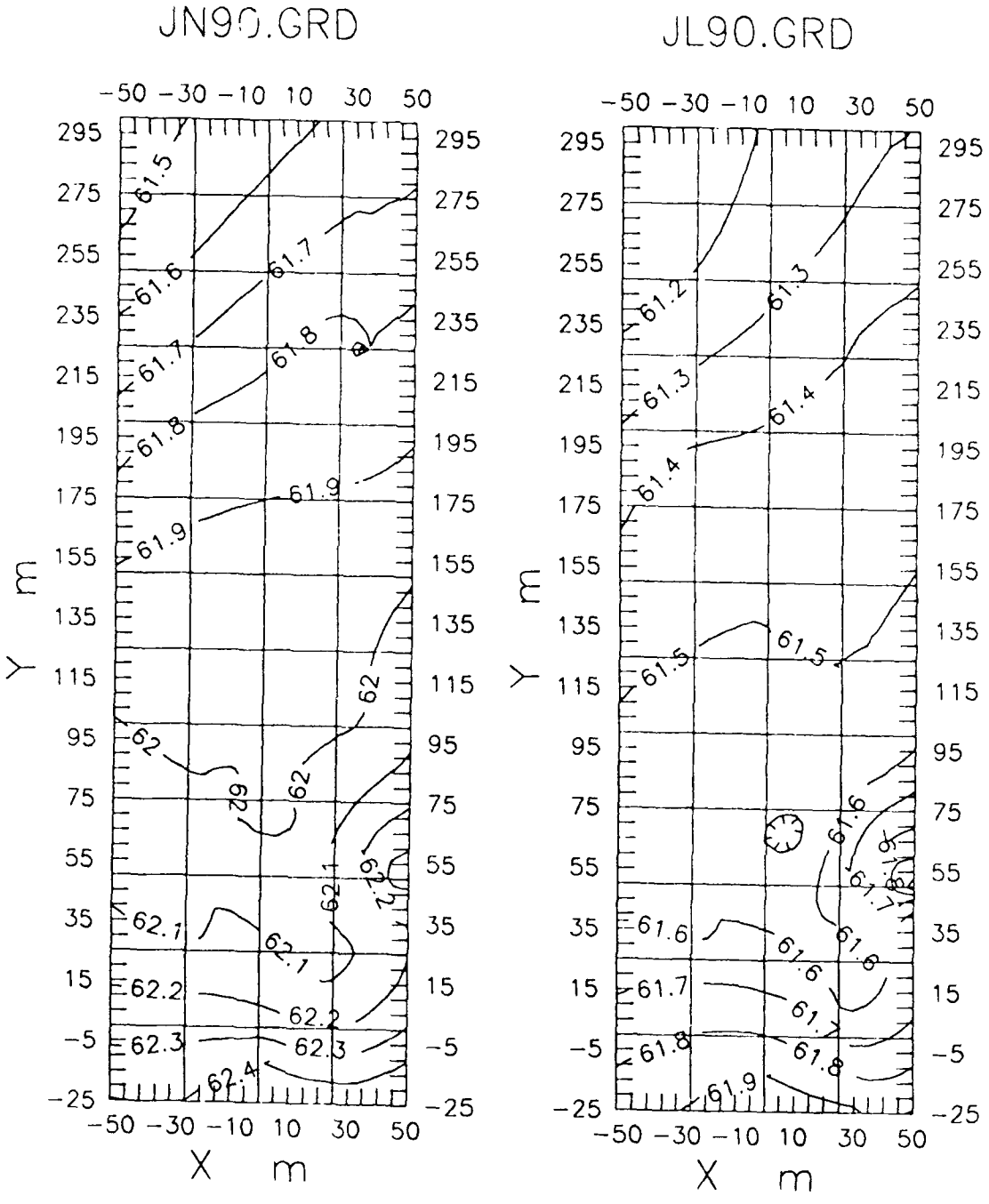


Figure 3. Krigged water table contours in meters for the first two piezometer surveys. Left: simulation day 8 (June 19). Right: simulation day 42 (July 23).

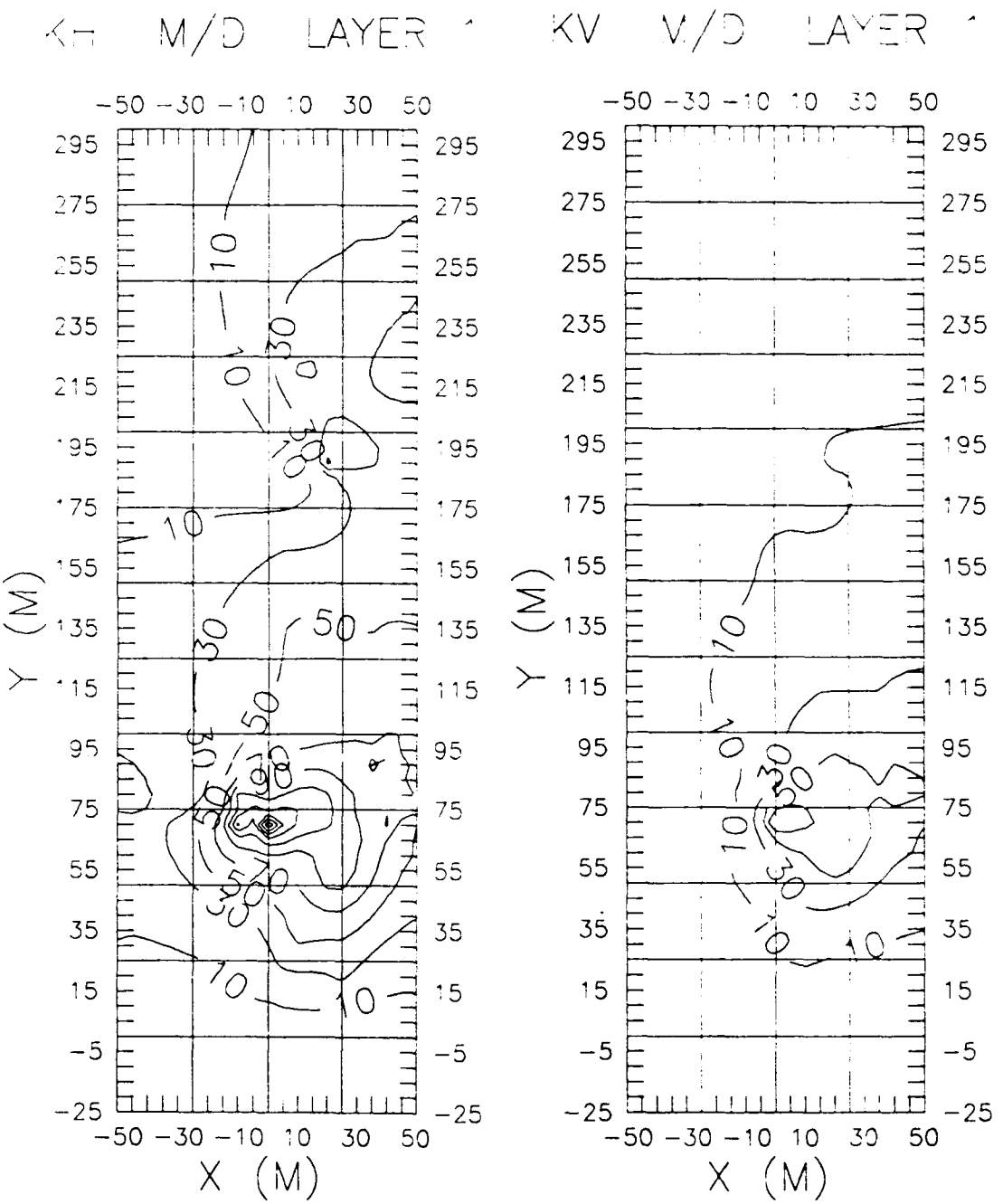


Figure 4. Horizontal (left) and vertical (right) hydraulic conductivity in m/day for the top layer of the variable conductivity model.

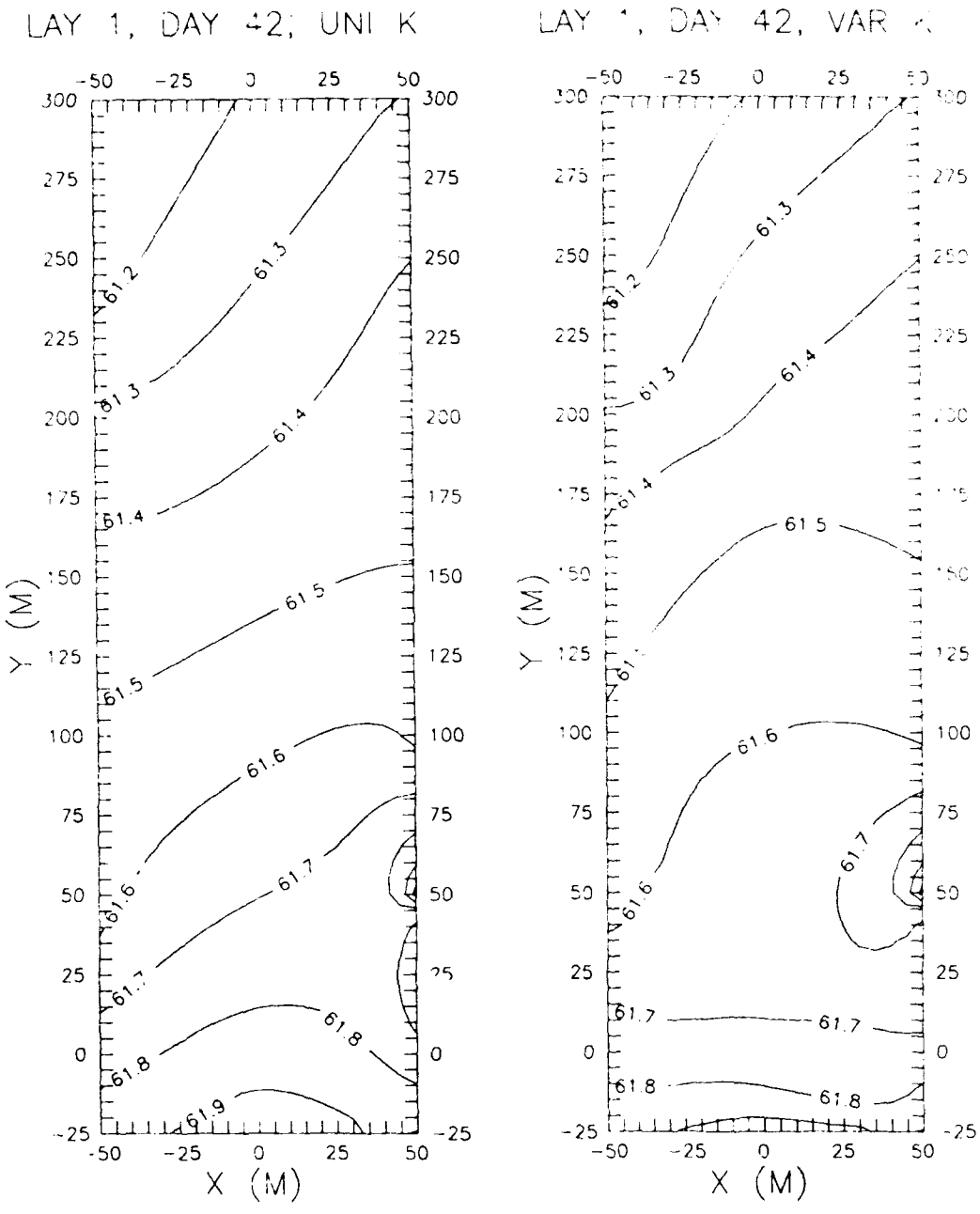


Figure 5. Predicted water table contours in meters on simulation day 42. Left: uniform hydraulic conductivity model. Right: variable hydraulic conductivity model.

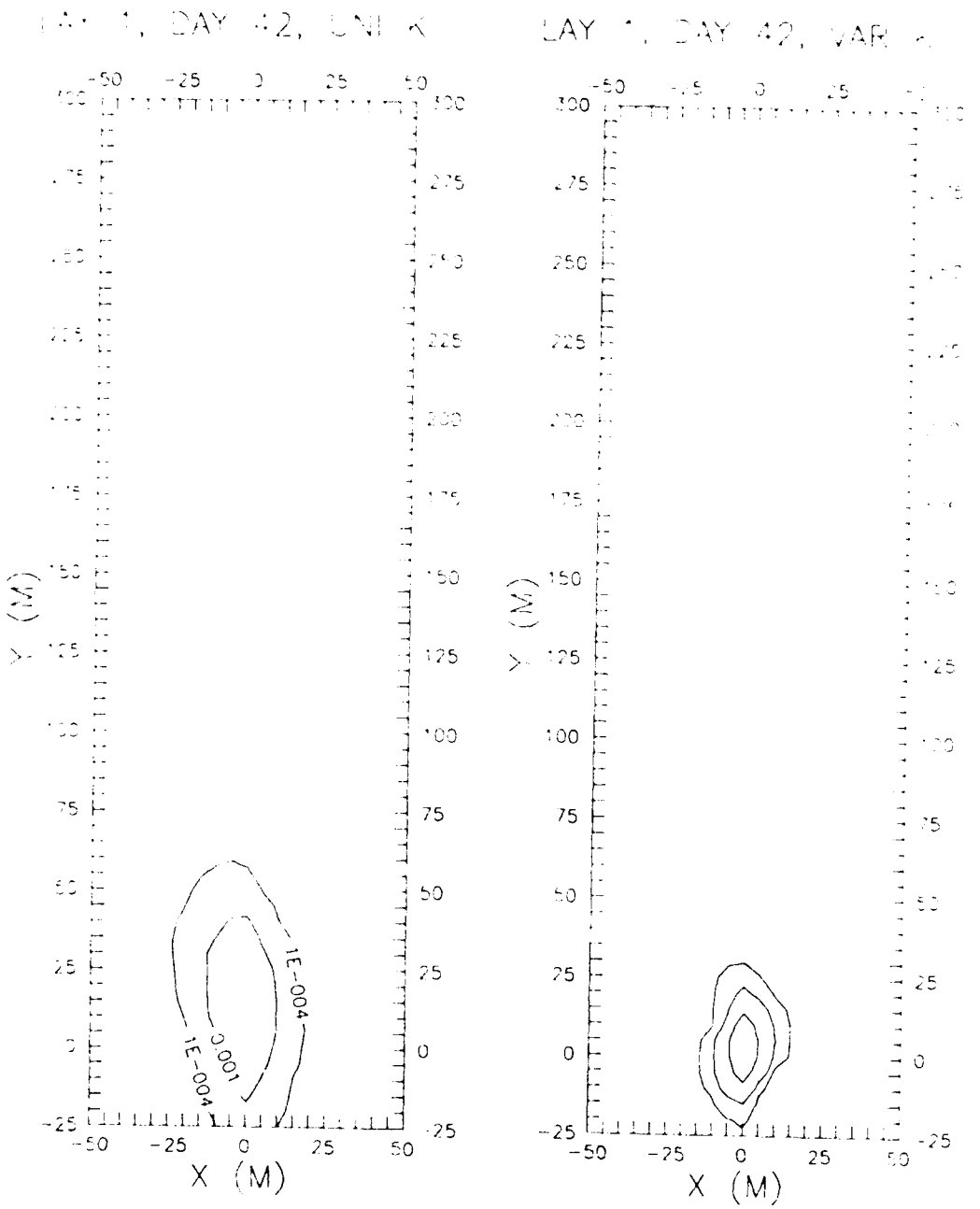


Figure 6. Tritium relative activity in the top layer on simulation day 42. Contours are 0.0001, 0.001, 0.01, 0.1. Left: uniform hydraulic conductivity model. Right: variable hydraulic conductivity model.

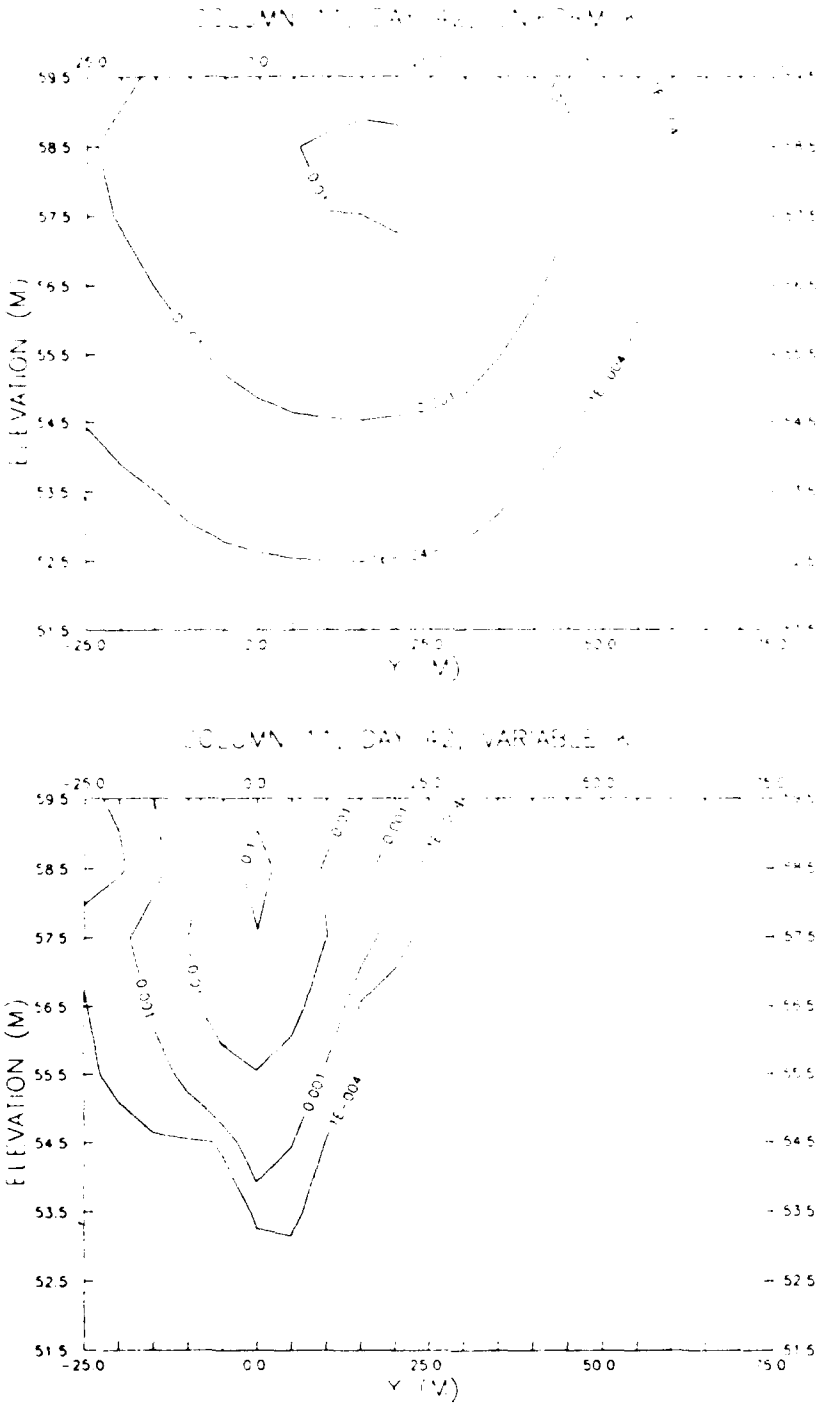


Figure 7. Tritium relative activity profiles along column 11 on simulation day 42. Top: uniform hydraulic conductivity model. Bottom: variable hydraulic conductivity model.

HYDRAULIC CONDUCTIVITY VARIABILITY: CO-KRIGING

Valipuram S. Manoranjan
Associate Professor
Department of Pure and Applied Mathematics

Washington State University
Pullman, Washington 99164-3113

Final Report for:
Summer Research Program
Tyndall Air Force Base Civil Engineering Laboratory

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

September 1993

HYDRAULIC CONDUCTIVITY VARIABILITY: CO-KRIGING

Valipuram S. Manoranjan
Associate Professor
Department of Pure and Applied Mathematics
Washington State University

Abstract

The spatial variability of hydraulic conductivity is studied using the grain-size data collected at a groundwater tracer test site located at Columbus Air Force Base in Mississippi. Following the initial study which involved vertical kriging and the construction segmented trend surfaces, a co-kriging analysis is carried out. It is observed that the total organic carbon content (toc) data do not have a great influence on the variability pattern of hydraulic conductivity. The sensitivity of the spatial pattern to certain corehole data is also observed.

HYDRAULIC CONDUCTIVITY VARIABILITY: CO-KRIGING

Valipuram S. Manoranjan

1. INTRODUCTION

The contamination of soil and groundwater systems and the adverse effects of the contamination of these resources have become a matter of great concern. For proper management of soil and groundwater, it is necessary to have tools, such as mathematical and/or physical models, with which the effects of contamination can be assessed. Using such assessments, predictions can be made for example, concerning transport of contaminants, when measurements are environmentally undesirable or difficult to conduct. Contaminants preferentially migrate along paths of interconnected units of higher hydraulic conductivity. So, it is of utmost importance to understand the spatial variability pattern of hydraulic conductivity. However, yet, there is no proved and tested method of analyzing spatial variability of hydraulic conductivity. We focus on this problem and carry out a study on the spatial variability of hydraulic conductivity employing ideas such as vertical kriging, segmented trend surfaces and co-kriging. The study uses the (grain-size analysis) data collected at a tracer test site located at Columbus Air Force Base in Mississippi. Already, we have conducted a preliminary investigation and have some interesting results [Manoranjan, 1992]. The co-kriging analysis is performed combining both the hydraulic conductivity data and the total organic carbon content data.

The standard approach of tackling a problem of the kind proposed here is to introduce the concept of regionalized variable (ln[hydraulic conductivity] in our case), and to do kriging, a local estimation technique which provides the best linear unbiased estimator, on the regionalized variable over the site (a three dimensional domain) of interest. We feel, sometimes, such a three dimensional kriging study might obscure any simple behavioral patterns the regionalized variable might have in one particular direction or the other. In the earlier study [Manoranjan, 1992], we took a different approach and introduced segmentation in the

data. However, for the study reported here, we carry out co-kriging analyses in the vertical direction and on horizontal benches.

2. HYDRAULIC CONDUCTIVITY ESTIMATES

The experimental data for this study are derived from core samples taken from the saturated zone at the Macro-Dispersion Experiment (MADE) test site at Columbus AFB, Mississippi. Six-inch diameter core samples were collected over the entire saturated depth at seventeen irregularly spaced locations through the center of the tracer plume path. Each core was subdivided into either six or twelve intervals for subsequent chemical and physical analyses. The solid materials in each interval were dried and sieved using U.S. Standard sieves to determine particle size distributions. The data used here are new and are different from the data obtained at the MADE site by [Rehfeldt et al., 1992]. An empirical formula which relates hydraulic conductivity to grain size is used to estimate hydraulic conductivity values for these sieved soil samples. We use the following relationship in the form presented in [Seiler, 1973]:

$$K = fd^2$$

where, K is the hydraulic conductivity, d, the representative grain diameter and f, a proportionality factor. The factor f is a function of the uniformity coefficient U given as,

$$U = d_{60}/d_{10}$$

where, d_m is the grain diameter such that $m\%$ of the sample by weight is of diameter less than d_m. Following [Seiler, 1973], when $U \geq 5$ the hydraulic conductivity estimates are based on the formulae

$$K = f(U)\ \ d_{10}^2, \quad 5 \leq U \leq 17$$

and

$$K = f(U)\ \ d_{25}^2, \quad U > 17.$$

The values for $f(U)$ can be obtained from either [Seiler, 1973] or [Rehfeldt et [illegible]992]. However, when $U < 5$, we use the formula

$$K = d_{10}^2$$

due to [Hazen, 1892].

3. KRIGING\CO-KRIGING OVER DEPTH

In order to analyze the spatial variability of the hydraulic conductivity estimates obtained using the empirical formulae above, we introduce the concept of regionalized variable. A regionalized variable may be thought of as an intermediate variable between a truly random variable and a completely deterministic one. Unlike random variables, a regionalized variable will have continuity from point to point, but usually it will not be possible to know a regionalized variable's value everywhere, as one could for a deterministic variable. Here, the obvious candidate for the regionalized variable is the hydraulic conductivity, K. However, geostatistics requires that the regionalized variable be normally distributed, and our hydraulic conductivity estimates are not normally distributed. Therefore, a better choice for the regionalized variable is $ln(K)$. Usual practice is to assume stationarity of the regionalized variable and to find the best estimator (with minimum variance) of the mean value of the regionalized variable (punctual kriging).

The first step in the kriging process involves calculating the semi-variogram associated with the experimental data. At this point, the basic assumption is that the log hydraulic conductivity field is stationary. The semi-variogram γ describes the expected difference in value between pairs of samples with a given relative orientation. If we denote the regionalized variable $ln(K)$ by R, the formula that is used to calculate the semi-variogram is given by

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} [R(z_i) - R(z_i + h)]^2$$

where, z_i and $z_i + h$ are two different depths at a distance h apart and n is the number of sample pairs separated by this distance h, known as the lag distance. The graph γ *vs.* h is the semi-variogram. The computed γ values can be fitted by various models and in most studies the use of linear models has been found adequate [Burgess and Webster, 1980]. However, in this study we find that the semi-variograms are better fitted by nonlinear models.

Once the form of the semi-variogram is determined, the next step is kriging. Kriging provides the estimated value for the regionalized variable R at any depth at a given location. For a detailed account of kriging, the interested reader should consult [Journel and Huijbregts, 1991]. For the sake of completeness, we will briefly describe the idea of kriging in a simple fashion. At a given location, let the number of points z_i employed for kriging be m, and $R_i, (i = 1, 2, \cdots, m)$ be the values of some quantity R at these points. Then, the estimate value $\hat{R}_p$ at any point z_p will be given by

$$\hat{R}_p = W_1 R_1 + W_2 R_2 + \cdots + W_m R_m.$$

(i.e. a weighted sum of the values $R_1, R_2, \cdots, R_m$).

The weights W_i's are not known at this stage and they will be evaluated so that the error associated with the estimate is less than that for any other linear sum involving $R_1, R_2, \cdots, R_m$. It is required that this estimate be *unbiased*. So, we want $\hat{R}_p$ to be the same as the expectation of the regionalized variable R at location z_p. This amounts to the condition that the weights W_i's should sum to 1:

$$\sum_{i=1}^{m} W_i = 1.$$

The estimation variance at z_p can be obtained as,

$$-\sum_{i=1}^{m}\sum_{j=1}^{m} W_i W_j \gamma_{ij} + 2\sum_{j=1}^{m} W_j \gamma_{pj},$$

where, γ_{ij} is the value of the semi-variogram of pairs of points at a lag distance $|z_i - z_j|$ apart. γ_{pj} is defined similarly. For the estimate we obtain to be the best linear unbiased estimate, W_i's should be found by minimizing the estimation variance subject to the constraint $\sum_{i=1}^{m} W_i = 1$.

This will result in finding partial derivatives with respect to W_i's, introducing the Lagrange multiplier λ and finally showing that the minimum variance is obtained when

$$\sum_{j=1}^{m} W_j \gamma_{ij} + \lambda = \gamma_{ip}.$$

The above equation can be written in the matrix form as

$$AU = b,$$

$$\text{where } A = \begin{bmatrix} \gamma_{11} & \gamma_{12} \cdots\cdots \gamma_{1m} & 1 \\ \gamma_{12} & \gamma_{22} \cdots\cdots \gamma_{2m} & 1 \\ \vdots & & \\ \gamma_{1m} & \gamma_{2m} \cdots\cdots \gamma_{mm} & 1 \\ 1 & 1 \cdots\cdots\cdots 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} \gamma_{1p} \\ \gamma_{2p} \\ \vdots \\ \gamma_{mp} \\ 1 \end{bmatrix}$$

$$\text{and } U = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_m \\ \lambda \end{bmatrix}.$$

The minimum estimation variance will be given by $b^T U$.

In theory, one can obtain more accurate results with more data, (i.e. with a larger m value in the above equation). However, if m is small, and it is felt that the grain-size data is not sampled enough, then it is possible to combine the grain-size data with some other sampled quantity and perform kriging. Such a kriging process is known as co-kriging, and we carried out such an analysis using total organic carbon content (toc) data as the other sampled quantity. In this case, one has to keep track of two regionalized variables, *ln* (hydraulic conductivity) and *ln* (toc).

In order to perform co-kriging, instead of a semi-variogram one would have a cross-variogram. Generally, the cross-variogram can be written as a linear combination model of the form.

$$\gamma_{12} = \alpha_1 \text{ (function 1)} + \alpha_2 \text{ (function 2)},$$

where function 1 and function 2 are independent functions with α_1 and α_2 being scalar parameters. The functions 1 & 2 are chosen such that for particular choices of α_1 and α_2 the combination model should give the individual semi-variograms corresponding to *ln* (hydraulic conductivity) and *ln* (toc). Now, we will also have an additional set of weights W_i's$(i = 1, \cdots, m)$ corresponding to the regionalized variable *ln* (toc). Whereas, W_i's corresponding to *ln* (hydraulic conductivity) satisfy the condition,

$$\sum_{i=1}^{m} W_i = 1$$

$\tilde{W}_i$'s will satisfy the condition,

$$\sum_{i=1}^{m} \tilde{W}_i = 0.$$

Initially, we co-kriged the data over depth (in the vertical direction). The corresponding semi-variograms and cross-variogram are presented in Fig. 1. The co-kriged estimates for hydraulic conductivity over depth is given pictorially in Fig. 2. When these co-kriged estimates were compared with direct estimates for hydraulic conductivity (i.e. without involving toc data), both sets of estimates were found to be almost the same. This means that either toc data has no great influence on hydraulic conductivity variability or grain-size data sampling is good. Although more work is needed to clarify this situation, we tend to think that the former is true. Also, we performed co-kriging on horizontal benches at selected depths and a typical variogram and the spatial variability pattern of hydraulic conductivity are presented in Fig. 3. In order to understand the sensitivitiness of the estimated variability pattern on certain corehole data, we carried out a simple study where certain corehole data were removed from the kriging process. Figs. 4 and 5 give the estimated variability pattern over depth when some corehole data are missing. In the figures, '*Less two*' means two coreholes are missing and '*All*' means all the coreholes are included. As one can see almost all the patterns are qualitatively similar except for the spike value at the depth of 20 feet. The only exception is the pattern at the bottom of Fig. 5. Here, just removing one corehole changed the pattern drastically, implying the influence of that corehole data on the variability pattern. Again, more systematic work is needed to identify the influential coreholes. Such a study can provide important information to do good sampling of data.

4. CONCLUSIONS

In this study we demonstrated the use of co-kriging analysis in understanding variability pattern of hydraulic conductivity. The preliminary results indicate the non-influence of toc data on hydraulic conductivity estimates. It is also seen how a single corehole data can have a major influence on the estimated variability pattern.

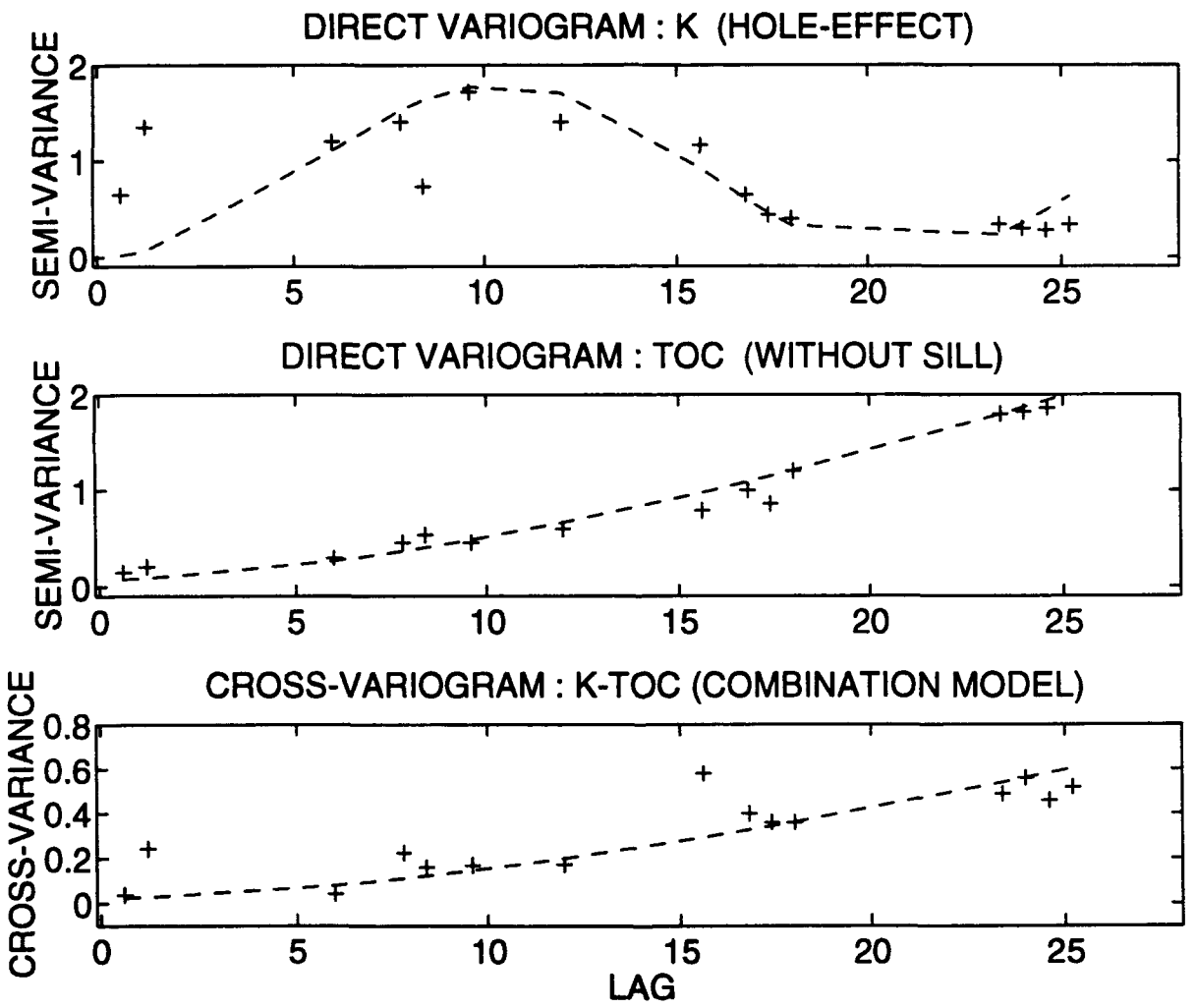

DIRECT VARIOGRAM : K (HOLE-EFFECT)
SEMI-VARIANCE
0
1
2
0
5
10
15
20
25
DIRECT VARIOGRAM : TOC (WITHOUT SILL)
SEMI-VARIANCE
0
1
2
0
5
10
15
20
25
CROSS-VARIOGRAM : K-TOC (COMBINATION MODEL)
CROSS-VARIANCE
0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
LAG

Fig. 1

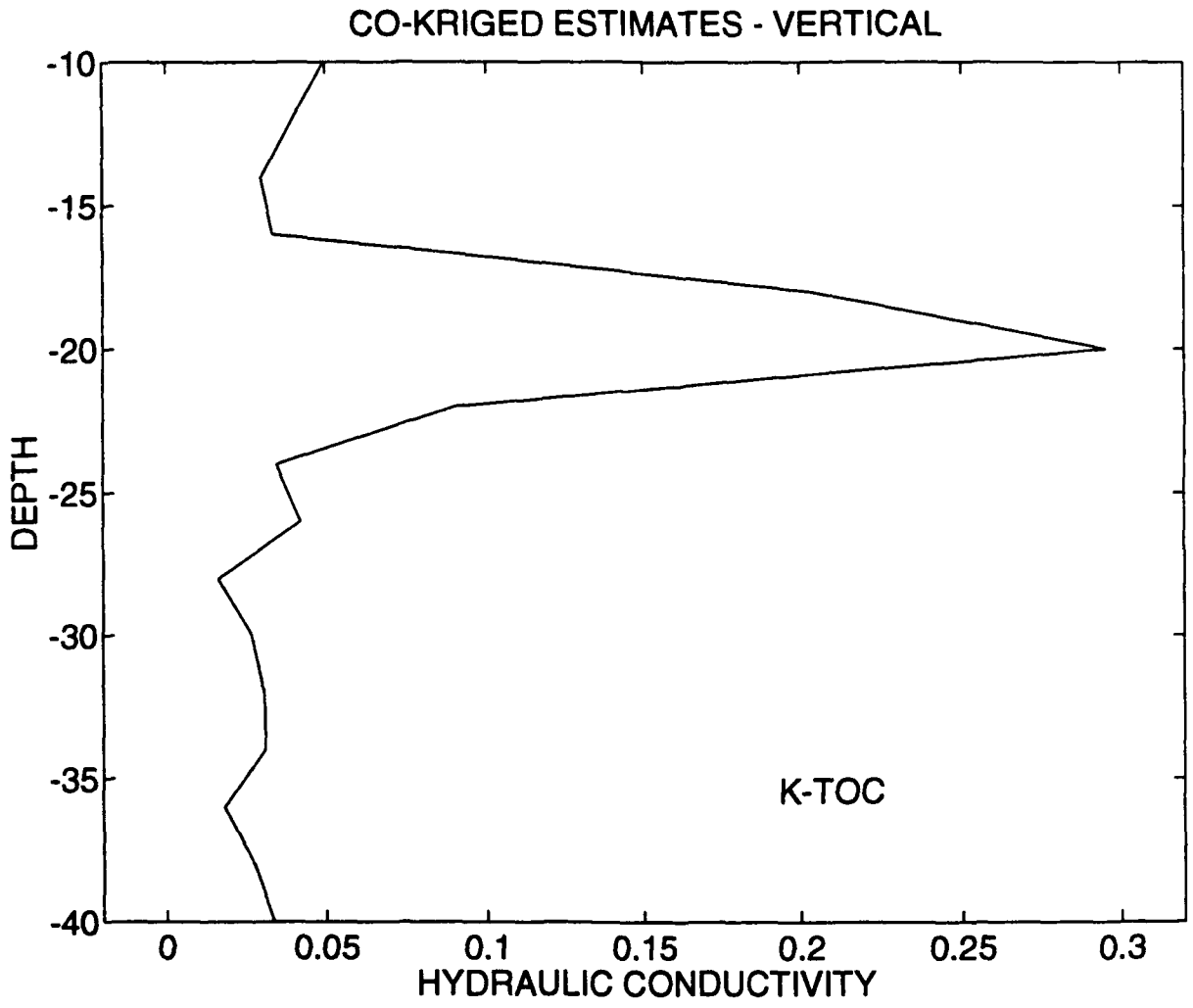

CO-KRIGED ESTIMATES - VERTICAL
-10
-15
-20
-25
-30
-35
-40
DEPTH
K-TOC
0
0.05
0.1
0.15
0.2
0.25
0.3
HYDRAULIC CONDUCTIVITY

Fig. 2

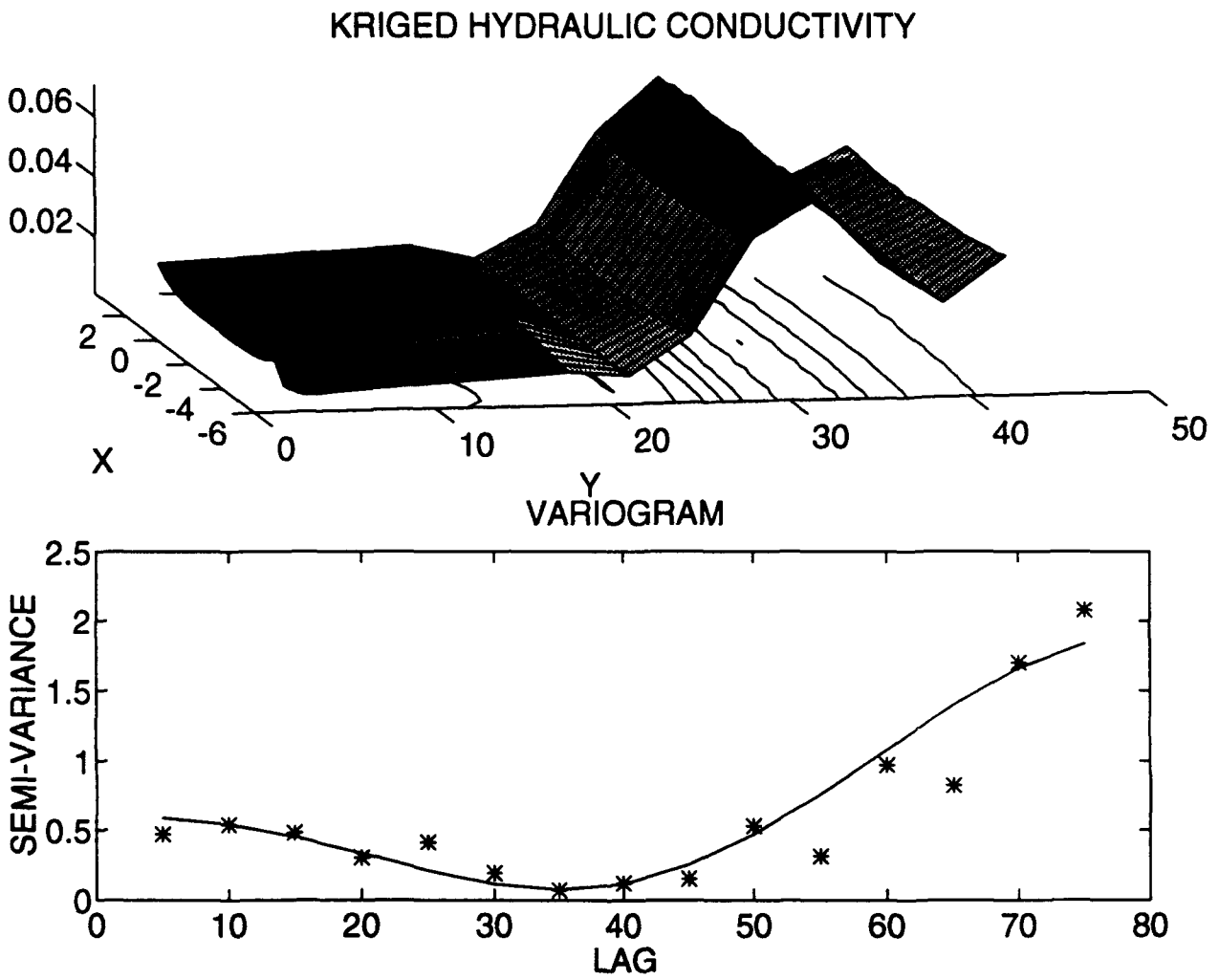

KRIGED HYDRAULIC CONDUCTIVITY
0.06
0.04
0.02
2
0
-2
-4
-6
X
0
10
20
30
40
50
Y
VARIOGRAM
SEMI-VARIANCE
2.5
2
1.5
1
0.5
0
0
10
20
30
40
50
60
70
80
LAG

Fig. 3

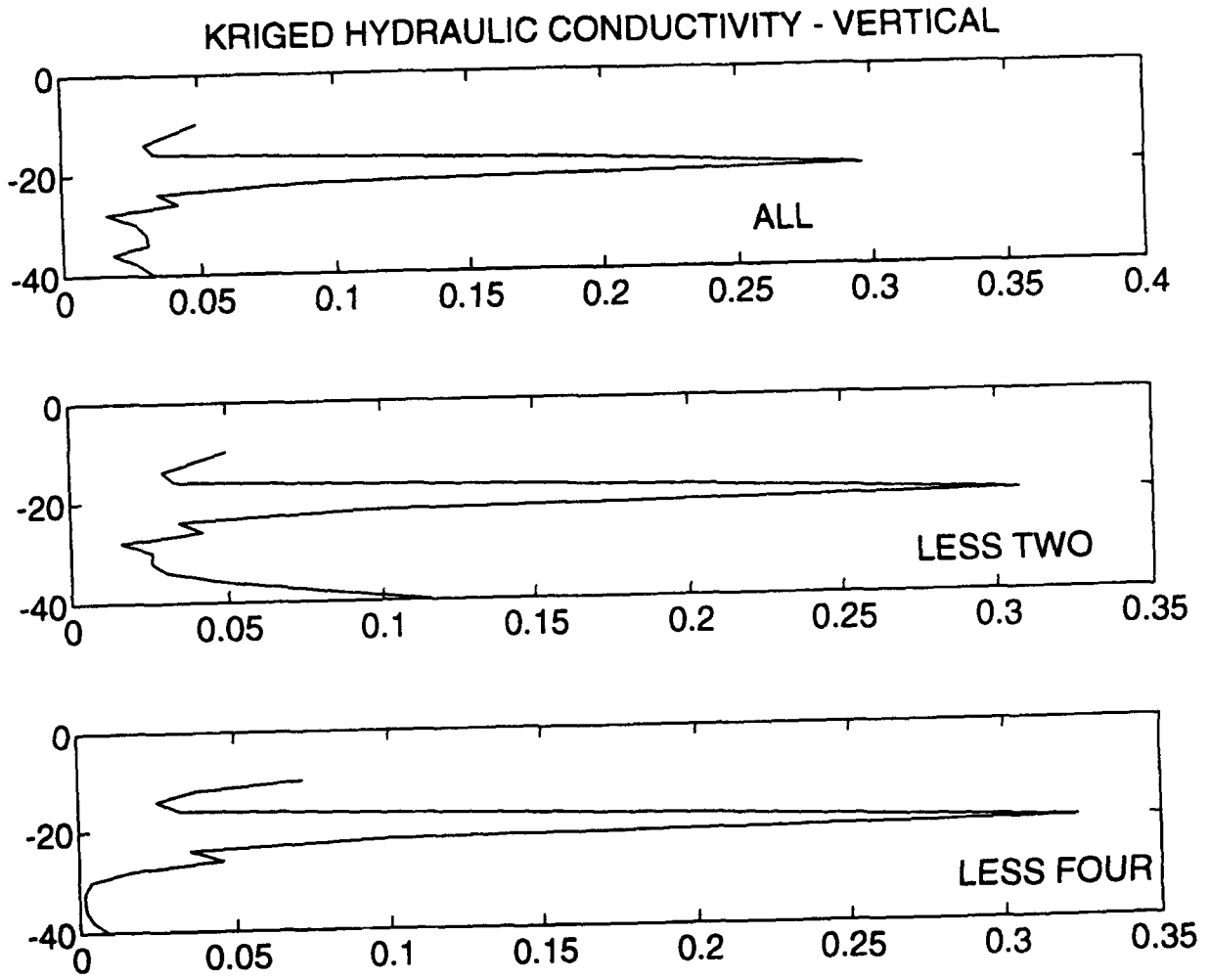

KRIGED HYDRAULIC CONDUCTIVITY - VERTICAL
0
-20
-40
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
ALL
LESS TWO
LESS FOUR

Fig. 4

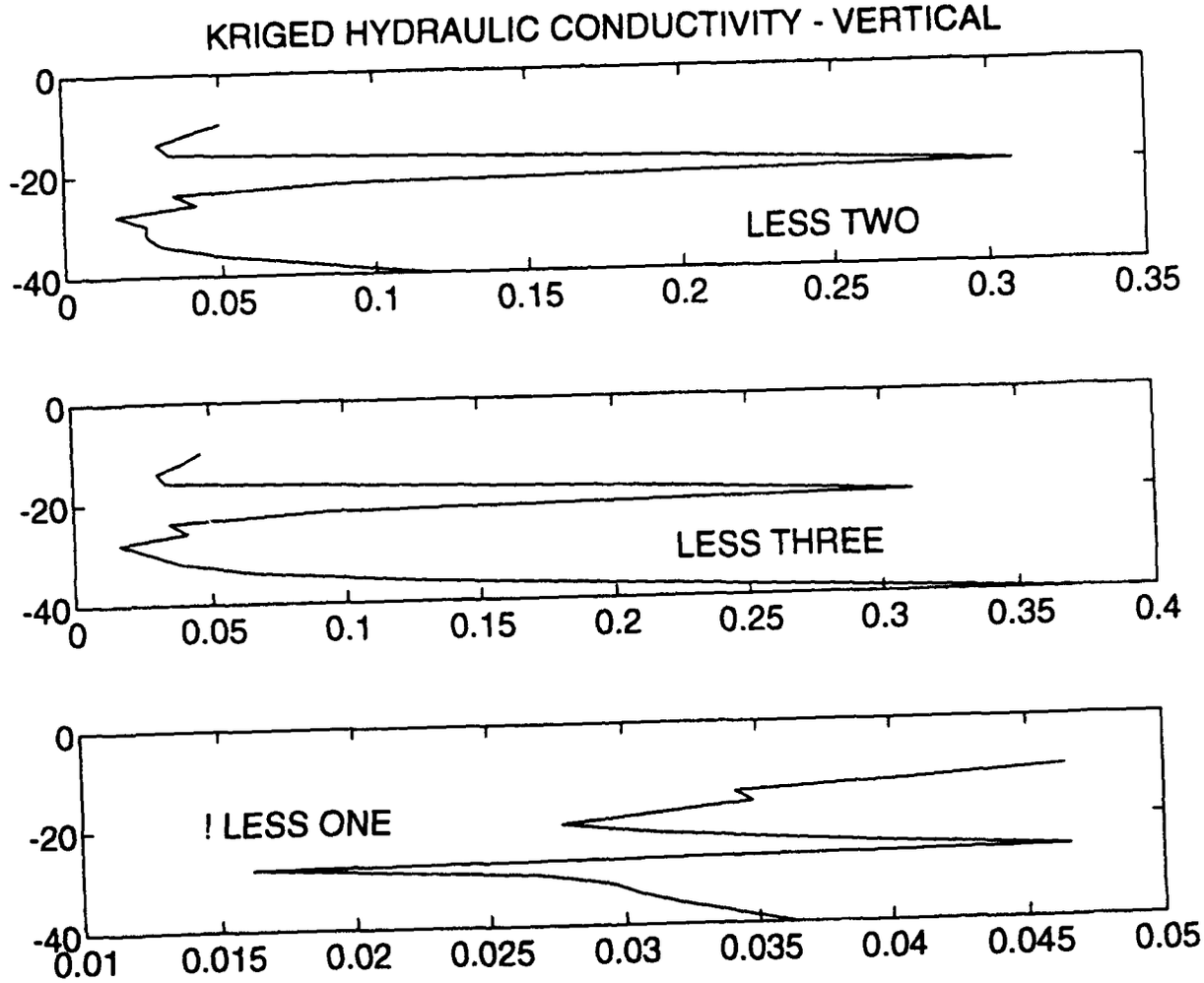

KRIGED HYDRAULIC CONDUCTIVITY - VERTICAL
0
-20
-40
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
LESS TWO
0
-20
-40
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
LESS THREE
0
-20
-40
0.01
0.015
0.02
0.025
0.03
0.035
0.04
0.045
0.05
! LESS ONE

Fig. 5

REFERENCES

1. Burgess TM, Webster R. Optimal interpolation and isarithmic mapping of soil properties: 1. The semi-variogram and punctual kriging. J. Soil Science 1980; 31 315-31.

2. Hazen A. Experiments upon the purification of sewage and water at the Lawrence experiment station, Mass. State Board of Health. 23rd Annual Report. 1892.

3. Journel AG, Huijbregts ChJ. Mining Geostatistics. New York: Academic Press. 1991.

4. Manoranjan VS. Hydraulic conductivity variability, kriging, trend surfaces and travelling waves with nonlinear, nonequilibrium adsorption. Final Report - AFOSR Summer Research Program, 1992.

5. Rehfeldt KR, Boggs JM, Gelhar LW. Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. Water Resources Research 1992; 28:3309-24.

6. Seiler KP. Durchlässigkeit, Porosität and Kornverteilung quartärer Kies-Sand- Ablugerunger des bayerischen Alpenvorlandes. Gas-und Wasserfach 1973; 114:353-400.

EVALUATION OF AN IMMOBILIZED CELL BIOREACTOR FOR DEGRADATION OF META- AND PARA-NITROBENZOATE

Steven W. Peretti[1] and Stuart M. Thomas[2]
[1]Associate Professor, [2]Research Associate
Department of Chemical Engineering

North Carolina State University
Raleigh, NC 27695

Final Report for:
[1]Summer Faculty Research Program
[2]Graduate Student Research Program
AL/EQC

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

September 1993

EVALUATION OF AN IMMOBILIZED CELL BIOREACTOR FOR DEGRADATION OF META- AND PARA-NITROBENZOATE

Steven W. Peretti[1] and Stuart M. Thomas[2]
[1]Associate Professor, [2]Research Associate
Department of Chemical Engineering
North Carolina State University

Abstract

Meta- and para-nitrobenzoic acid (m-NBA, p-NBA) are pollutants found in waste streams from metal-stripping processes utilizing cyanide-free solvents. The Kelly AFB Industrial Waste Treatment Plant (IWTP) is currently incapable of removing these compounds from the waste water it receives because of (1) the presence of significant quantities of ethylenediamine, a preferred substrate, and (2) an upper limit of 4.5 hours on the hydraulic residence time in the IWTP. This work describes the enrichment and preliminary characterization of a microbial consortium capable of utilizing both m-NBA and p-NBA as sole carbon sources. Experimental results indicate that m-NBA degradation involves an oxidation pathway, while p-NBA degradative proceeds through a reductive pathway. This consortium was immobilized by entrapment in alginate beads and grown in a continuous-flow airlift reactor. Single substrate and mixed substrates were fed to the reactor. Conditions were varied to simulate different waste treatment scenarios: switching from one stripping solvent batch to another, starting up of the metal stripping process, mixed solvent batches, and changing the loading rate of substrate to the bioreactor. Results indicate that the nitrobenzoate fraction of the metal stripping waste can be effectively treated in a continuous-flow, immobilized-cell bioreactor with a hydraulic residence time well below 3 hours. Furthermore, the process can be operated over long periods (>250 hours) with little diminution of performance and responds rapidly to changes in substrate.

EVALUATION OF AN IMMOBILIZED CELL BIOREACTOR FOR DEGRADATION OF META- AND PARA-NITROBENZOATE

Steven W. Peretti and Stuart M. Thomas

INTRODUCTION

Biotreatment

Proliferation of toxic organic compounds that are resistant to photochemical or biological degradation (xenobiotic) is a burgeoning threat to human health. Complications in the treatment of xenobiotics arise due to their chemical composition and the difficulty of their isolation from other compounds. Process waste streams, for example, can be isolated and their chemical composition well defined. On site treatment of these wastes is constrained by the chemical properties of the compounds present (mixed organics, extreme pH, high salinity) in the process streams. Recently, a large body of literature has accumulated regarding the use of microorganisms to degrade a wide variety of halogenated, polycyclic, and multi-substituted organic molecules heretofore thought not to be susceptible to biodegradation[1-4]. While in a few cases, a single organism has been found which can mineralize a specific xenobiotic, most often a consortium of species is necessary to completely detoxify mixtures [5,6]. Exploiting the capability of these organisms is problematic, and care must be taken to establish appropriate bioreactor design and operation. Of primary importance for these efforts will be the long-term maintenance of a population which preserves efficient biodegradative activity.

Many considerations affect the design of bioreactors for waste treatment applications. The compounds involved are often present either at low concentrations that do not support rapid growth or at high concentrations that inhibit growth. Substrate inhibition often results in growth lags, the length of which are directly proportional to the concentration of the inhibitory compound. The composition of the waste mixture may vary over time, as might the loading rates and the volumetric flow of the waste stream. Multiple organisms are usually required to effect the bioremediation of a stream containing several organic compounds.

These considerations mitigate against the use of suspended cell reactors. Cell suspensions can be grown to a limited cell density, restricting the degradative activity per unit volume. This implies that for either batch or continuous operation, the reactor must be sized such that the residence time of the fluid in the reactor is sufficient to allow degradation of the compounds. The limitation of cell growth by

high and low concentrations of the compounds will tend to cause suspended cell reactors to be rather large. Continuous flow reactors would enable a higher throughput of material than batch reactors, but the flow rates that can be accommodated are limited by the growth kinetics of the organisms. Mixed cultures can be maintained in batch reactors, but due to differences in growth rates, yields and dependence on environmental conditions, it is extremely difficult to maintain two populations in a continuous flow suspension culture, and virtually impossible to maintain three or more.

The use of immobilized cell reactors can alleviate each of the problems presented above. Immobilized cell reactors have significantly higher cell densities than suspended cell reactors, and the volumetric productivity reflects the ten- to one hundred-fold increase in cell density that can be achieved. Consequently, smaller reactors can be used. Since cell densities are higher and the cells remain in the reactor, degradation rate and cell growth rate are to a large degree uncoupled. This lessens the negative effects of the extremes of substrate concentration. Low concentrations can be fed rapidly since the cell population does not need to grow sufficiently quickly to avoid being washed out of the reactor. Substrate inhibition is also relieved. Immobilizing cells introduces mass transfer resistance due either to the immobilization matrix or the cell layers. This resistance lowers the effective substrate concentration for a fraction of the cells in the reactor, enabling them to actively degrade the compounds without a lag time.

Problem Summary

For many years the United States Armed Forces has utilized cyanide-containing stripping compounds in their metal-refinishing processes. The U.S. Air Force has undertaken a program to completely remove cyanide from these processes, and the effort has reached the demonstration stage in the plating shop of Kelly AFB. The successful demonstration of efficacy of the cyanide-free metal stripping compound CLEPO 204 led to questions surrounding treatment of the stripping wastes in the Kelly AFB Industrial Waste Treatment Plant (IWTP). The composition of spent CLEPO 204 is 33% ethylenediamine, 10% sodium nitrobenzoate and an "unidentified" red compound. Experiments were performed under the direction of the Idaho National Engineering Laboratory to determine the biodegradability of ethylenediamine(EDA) and nitrobenzoate (NBA) using sludge from the Kelly AFB IWTP.

Shake flask tests and continuous flow, bench-scale bioreactor tests were conducted using EDA or spent CLEPO 204 as the substrate. It was found that the shake flask cultures completely degraded EDA when it was the sole substrate. However, using spent CLEPO 204 as the substrate caused a reduction in EDA degradation and less than 20% degradation of NBA within 48 hours.

Continuous-flow tests with a hydraulic residence time of 5.3 hours (similar to that of the IWTP) gave only 8.8% degradation of ethylenediamine when it was the sole carbon source. Increasing the residence time to 8.3 hours led to an 88% removal of EDA but also caused ammonia levels to jump to well over 100 ppm. Spent CLEPO 204 components were not removed in the continuous-flow bioreactor with a residence time of 5.3 hours. Increasing the hydraulic residence time to 8.3 hours led to 100% removal of EDA within 30 hours and NBA within 150 hours. Concomitant with the degradation of these compounds were increases in effluent ammonia and nitrite.

The IWTP is running at or above its designed capacity with a hydraulic retention time of 4.8 hours. The residence time in the plant cannot be further increased. The preliminary conclusion drawn from these results is, therefore, that biological treatment of spent CLEPO 204 using the Kelly AFB Industrial Waste Treatment Plant is not currently feasible.

Research Objective

The objective of this work is to evaluate the feasibility of using a continuous-flow, immobilized cell reactor to remove m-nitrobenzoate and p-nitrobenzoate from aqueous streams. The effects of substrate loading rate, hydraulic residence time, substrate concentration and substrate composition on the fractional removal of nitrobenzoate have been studied. The response to starvation and rapid changes in substrate have also been investigated. These studies indicate the degree of stability of this waste treatment scheme to operational upsets and suggest other variables that might have a significant impact on the performance of this reactor system.

MATERIALS AND METHODS

Isolation of Bacteria

The organisms used in these studies were isolated from an activated sludge sample taken from the Industrial Waste Treatment Plant at Kelly Air Force Base in Austin, Texas. Inocula were diluted 1:10 v/v in Spain's minimal salts medium (SMSB)[7] supplemented with 100 ppm m-nitrobenzoate and grown in 250 mL flasks. The

cultures were diluted 1:10 following an observable change in turbidity, which took 1-2 days following inoculation. A consortium containing at least two distinguishable species resulted from serial cultivation on m-NBA, and these species were designated Kelly 4 and Kelly 7. Kelly 4 grows rapidly on m-NBA concentrations as high as 200 ppm, following a lag period of approximately 18 hours. Nitrite is released. Kelly 4 was not observed to grow on p-NBA. Kelly 7 grows extremely slowly on p-NBA, with concomitant release of ammonia. Strains were maintained on plates containing MSB[8], 18 g/L Bitek agar and 100 ppm m-nitrobenzoate.

Bacterial Immobilization

A suspended cell culture of the consortium was grown in SMSB containing 100 ppm m-NBA at 30° C. A previously autoclaved solution of 4% Na-alginate (Kelco, Manugel "GHB") dissolved in 200mM NaCl was mixed with log phase cells to give a final concentration of 3% alginate. Immobilized cell beads were formed by passing this solution through an 18-gauge needle, allowing individual droplets to fall into the airlift reactor vessel which contained 400 mL of 100mM SrCl at room temperature. *Strontium* was used as the cross-linking agent because it yields stronger gel beads than those formed using calcium[9]. The airlift and SrCl solution had been autoclaved and sealed so that the process was carried out under sterile conditions. The beads were allowed to cure for 4 hours with no agitation or air sparging. After curing, the reactors were flushed with SMSB to remove excess SrCl. The beads were fluidized by air sparging, and the culture allowed to grow in batch overnight in SMSB containing 100 ppm m-NBA.

Cell Culture

Suspended cell cultures, both batch and continuous, were analyzed for the more rapidly growing of the two microorganisms isolated, Kelly 4. Erlenmeyer flasks (250 mL) were used in studies to determine the maximal growth rate of Kelly 4 at 30° C in SMSB containing 100 ppm m-NBA. The concentration of m-NBA was also monitored to quantify its rate of degradation. Chemostat cultivation of Kelly 4 was also undertaken as a means to determine the dilution rate at which suspended cells would be washed out of a well-mixed reactor. SMSB amended with 100 ppm m-NBA was fed to a New Brunswick Bioflo reactor with a working volume of 500 mL. Temperature was maintained at 30° C, pH at 7.0.

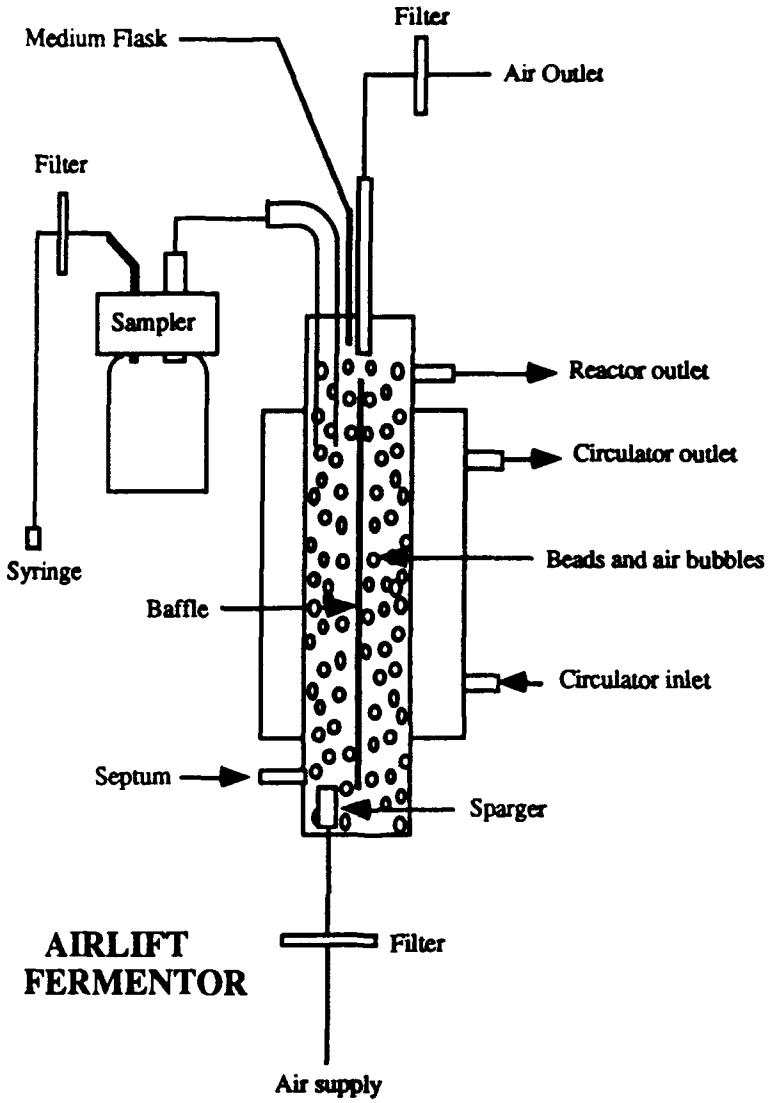


FIGURE 1

Immobilized consortium fermentations were conducted at 30°C in a pair of Kontes airlift vessels with a single central vertical baffle to promote mixing. Figure 1 is a schematic of the reactor used and its operation. Airflow was set so that the beads in the reactor (bead volume between 150 and 200 mL) were well mixed and the baffle remained free of clogging. The working volume in each reactor was maintained at 600 mL by fluid overflow. The original gel bead volume in reactor 1 was 172 mL and in reactor 2 was 185 mL. Final bead volumes were 106 mL and 115 mL, respectively. Medium and bead samples were periodically withdrawn aseptically.

The cell concentration in the medium was determined by measuring absorbance at 600 nm and using a correlation. Both reactors were fed SMSB-based medium with the pH adjusted to 7.5. After batch cultivation to establish culture growth, each reactor was subjected to a different substrate and dilution rate profile. The reactor fluid volumes used to determine dilution rates are those of the fluid volume in the reactor at the time. All dilution rates reported have the units of reciprocal hours (hr^{-1}).

Reactor 1 was subjected to the following sequence: continuous feed of 100 ppm m-NBA at dilution rates of 0.4 and 0.6; a step change to continuous feed of 50 ppm m-NBA at a dilution rate of 1.2; batch cultivation to complete depletion of m-NBA; a step change to continuous feed of 100 ppm m-NBA at dilution rates of 0.67 and 0.33; a step change to continuous feed of 48 ppm p-NBA at dilution rates of 0.33 and 0.1; and a step change to 50 ppm m-NBA at a dilution rate of 0.31 hr^{-1}.

Reactor 2 was subjected to the following sequence: continuous feed of 100 ppm m-NBA at dilution rates of 0.7 , 1.4, and 0.34; a step change to continuous feed of 50 ppm m-NBA at a dilution rate of 0.74 and 0.34; a step change to continuous feed of medium containing 40 ppm of both m-NBA and p-NBA at a dilution rate of 0.32; and a step change to continuous feed of medium containing 80 ppm of both m-NBA and p-NBA at a dilution rate of 0.31 hr^{-1}.

Analytical Methods

When a single substrate was fed to the reactor, the concentrations of m-NBA and p-NBA in the reactor samples were determined using the A_{266} for m-NBA and A_{275} for p-NBA. Calibration curves for both compounds were established in SMSB at pH 7.0. For those reactor runs containing mixed substrate feed, high-pressure liquid chromatography (HPLC) was performed using a mBondapak C_8 column (3.9 mm by 30 cm; Waters Associates, Inc., Milford MA). A linear gradient was run using methanol-water (acidified with trifluoroacetic acid) as the mobile phase. The MeOH:water composition was 50:50 at the start, changing linearly to 40:60 after 3 minutes, then changing immediately to 30:70 for the next 5 minutes. This gave a peak separation time of 40 seconds and no overlap of peaks. Compounds were detected by their absorbance at 270 nm with a Hewlett Packard diode array detector. Concentrations were quantified using peak areas and calibration curves established using pure components and mixtures of m-NBA and p-NBA.

Alginate bead samples were removed from the reactor at specified times and stored at 4° C for no more than a week prior to processing. The volume of the

sample was determined by decanting the residual medium, and estimating the packed bead volume using a graduated conical centrifuge tube. Approximately 1 mL of beads from each sample were dissolved with buffered saline solution[9]. The cells were precipitated by centrifugation in a Sorvall SS34 rotor at 8,000 rpm for 15 minutes. The cells were resuspended in 990 mL of 20 mM sodium phosphate pH 7.0, and lysed by adding 10 mL of 5 M NaOH and placing the samples at 100 °C for 15 minutes. A Pierce BCA total protein determination kit was used to analyze the samples. Total bead volume collected was noted throughout the course of the experiments, so that an estimate of total protein present in the reactor at any given time could be made.

RESULTS

Cell Growth

Figure 2 indicates that the maximal growth rate of Kelly 4 on 100 ppm m-nitrobenzoate in SMSB is approximately 0.36 hr^{-1}. The degradation rate of the m-NBA under these conditions is 186 mg hr^{-1} L^{-1} OD_{600}^{-1}. This implies that the washout dilution rate in a chemostat should be roughly 0.36 hr^{-1}. Figure 3 shows the results of a chemostat fermentation of Kelly 4 at 30 °C, pH 7.0. The culture is washed out at a dilution rate close to 0.2 hr^{-1} rather than the expected 0.36 hr^{-1}. The discrepancy between the performance in batch and continuous suspension culture is surprising. One possible explanation is substrate inhibition by m-NBA.

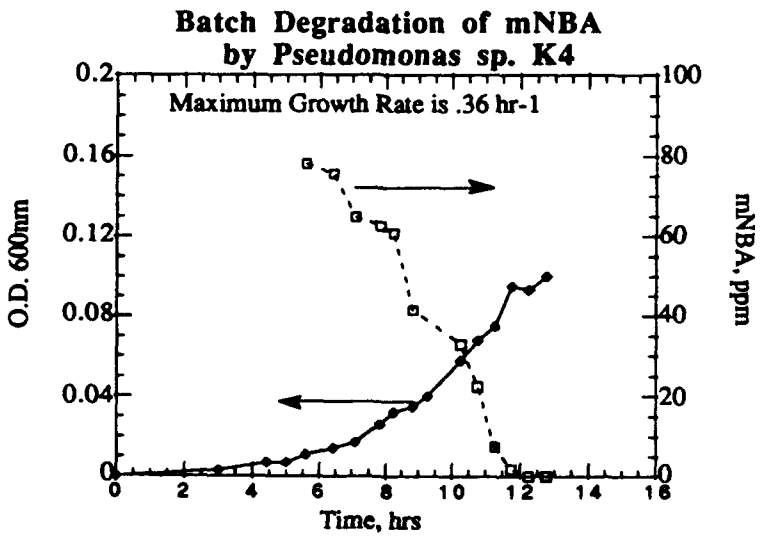


FIGURE 2

The maximal growth rate in shake culture was achieved at m-NBA concentrations in the range of 40 to 65 ppm. This hypothesis is supported by the observation that cells actively growing in shake culture exhibit some lag in growth upon sub-culturing into 100 ppm m-NBA.

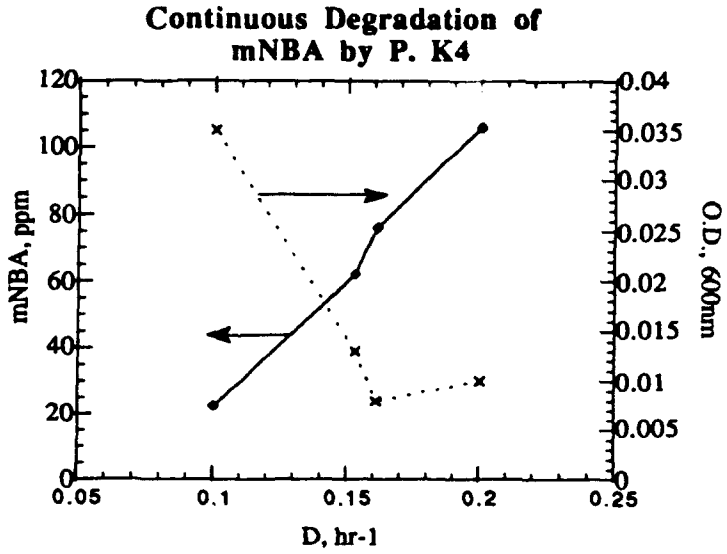


FIGURE 3

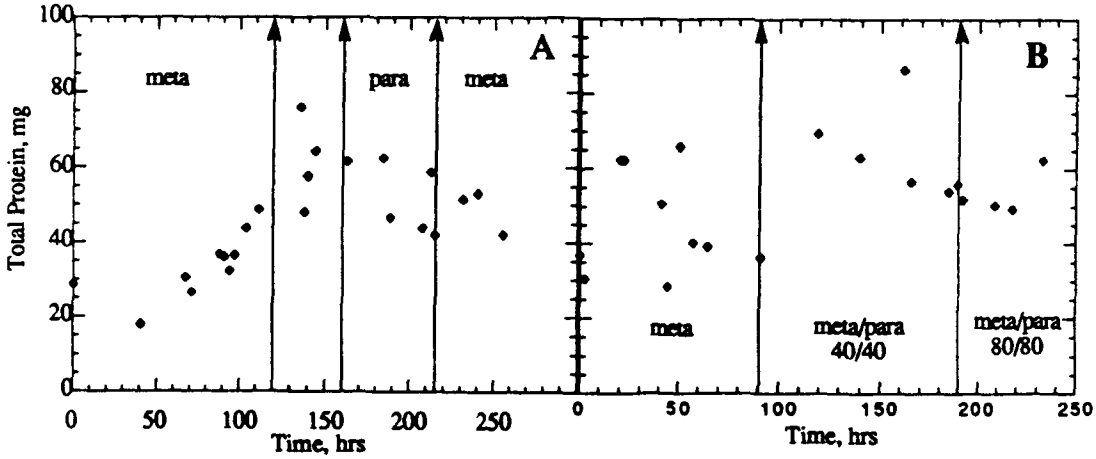


FIGURE 4

Total Protein

Figures 4A and B indicate the total protein in reactors 1 and 2, respectively, as a function of time. The results for reactor 1 indicate that it took nearly 150 hours to fully load the beads with biomass. This is supported by the observation that before

about 120 hours, the A_{600} of the fermentation broth was essentially zero, indicating no cells released from the gel beads. The gradual decline in total protein in reactor 1 after 150 hours can be accounted for by the volume of beads removed during sampling. Initially, the biomass in reactor 2 grew more rapidly, probably because it was the first reactor in which beads were formed and to which growth medium was introduced. As with reactor 1, the subsequent decline in total protein in the reactor can be attributed to gel bead removal.

Reactor 1 Performance

Starvation

Figure 5 illustrates the degradative performance of the consortium in reactor 1 following a period of starvation. The solid line and diamonds represent the loading rate of m-NBA into reactor 1 as a function of time. A 15 hour period of starvation is followed by rapid pumping of m-NBA into the reactor. The outlet concentration of m-NBA increases transiently, followed by complete degradation of the compound. The transient increase in m-NBA in the reactor is due solely to mixing. There is no indication of a loss of degradative activity or of a lag phase following this period of starvation. The pH in the reactor remained roughly constant during the starvation and post-starvation periods at 6.9

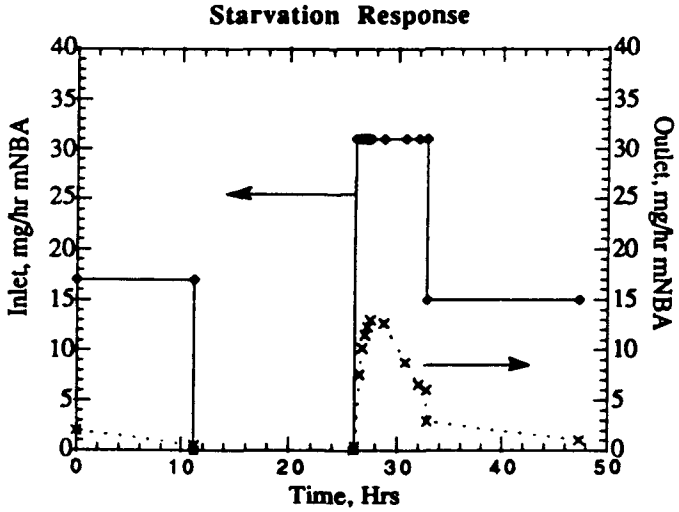


FIGURE 5

Shift from m-NBA to p-NBA

The inlet and outlet concentrations of p-NBA following a shift from m-NBA feed to p-NBA feed are shown in Figure 6. This culture was previously exposed to only m-NBA as a carbon source. The outlet concentration of p-NBA following the switch from m-NBA indicates the immediate degradation of some of the p-NBA. Subsequently, this reactor degraded about 60% of the p-NBA fed. The pH in the reactor rose to 7.5, probably due to the accumulation of ammonia released during the metabolism of p-NBA.

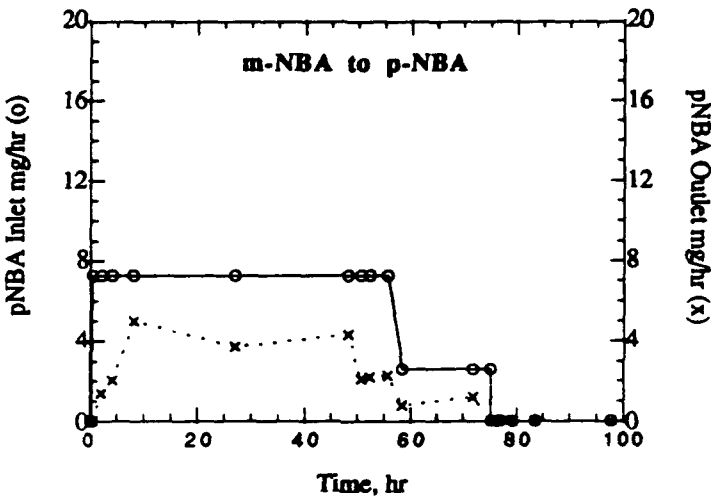


FIGURE 6

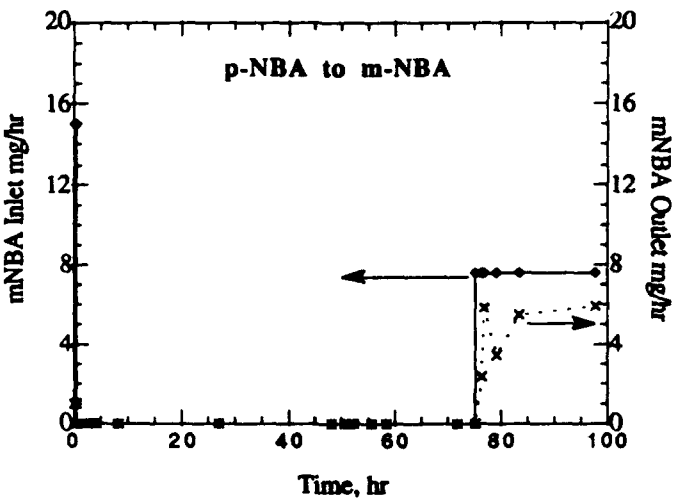


FIGURE 7

Shift from p-NBA to m-NBA

Figure 7 shows the concentration of m-NBA in the inlet and outlet streams following a switch in feed from p-NBA to m-NBA. The m-NBA feed had a pH of 7.1 instead of 7.5 due to the elevation of reactor pH caused by p-NBA degradation. Only approximately 20% of the influent m-NBA was degraded, compared to the greater than 90% degradation exhibited 50 to 100 hours earlier. The apparent inability of the reactor to exhibit the degradative performance it displayed at an earlier time may be due to the ammonia produced during the metabolism of p-NBA. The pH of the reactor was 7.5 immediately prior to the switch in substrate and took over 20 hours to drop to 6.8.

Reactor 2 Performance

Degradation of m-NBA

The carbon and energy sources provided to reactor 2 were m-NBA and mixtures of m-NBA and p-NBA. It is evident from the outlet concentrations for m-NBA displayed in Figure 8A for the straight m-NBA feed that the consortium is capable of degrading nearly 100% of the m-NBA at contact times below 3 hours. The addition of a moderate level of p-NBA (40 ppm) to the feed has essentially no effect on the degree of degradation of m-NBA. However, increasing both the m-NBA and p-NBA concentrations to 80 ppm in the feed results in a loss of the majority of the degradative activity. This is indicated by the rapid and sustained elevation in outlet m-NBA concentration following the increase in feed concentration.

Degradation of p-NBA

Figure 8B displays the inlet and outlet concentrations of p-NBA for reactor 2 during the introduction of mixtures of p-NBA and m-NBA to the vessel. Despite being fed only m-NBA for over 100 hours at the time of introduction of p-NBA to the reactor, the p-NBA was rapidly and completely degraded following a short mixing period. This performance continued unabated in the face of a concomitant doubling of m-NBA and p-NBA concentrations in the feed.

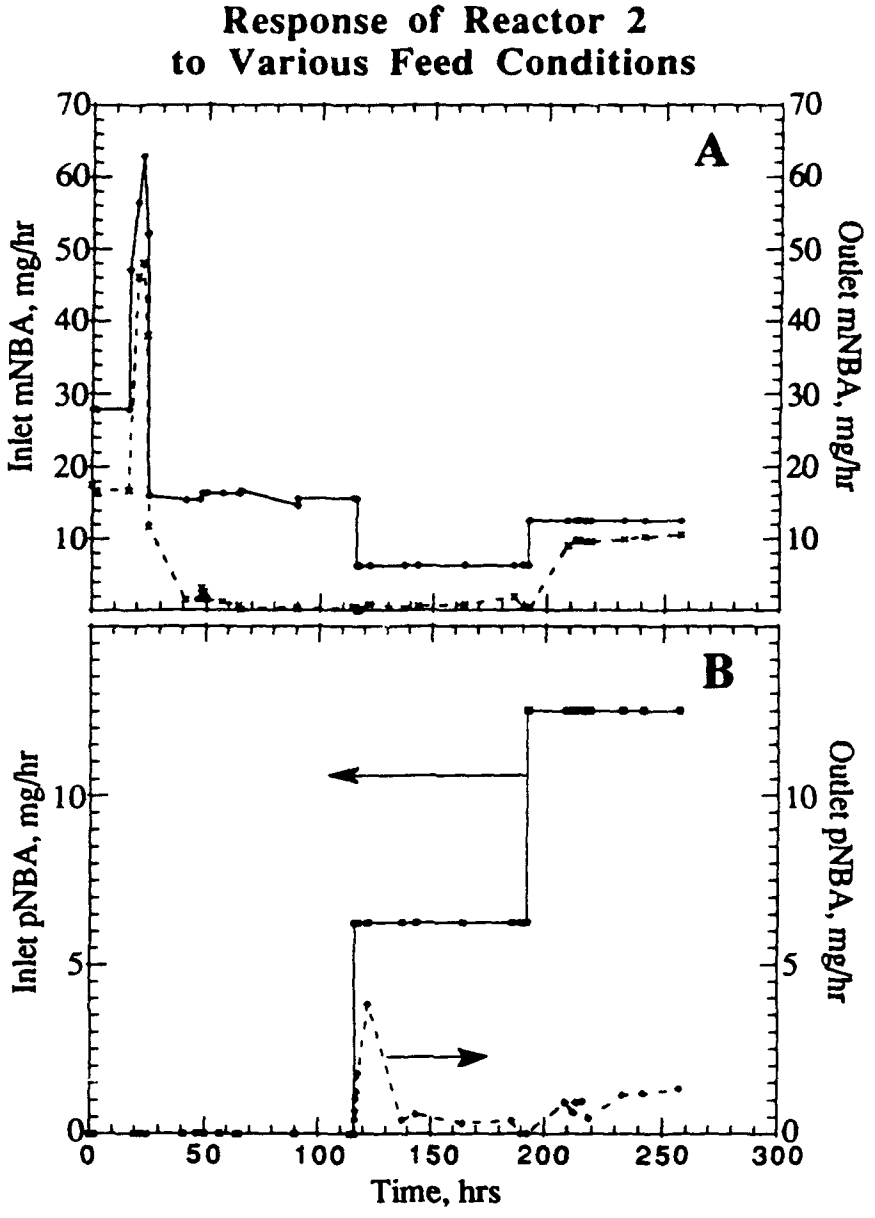


FIGURE 8

DISCUSSION

These studies indicate that the combination of organisms evaluated and the immobilized cell bioreactor represents a feasible option for the treatment of aqueous

streams containing mixtures of m- and p-nitrobenzoate. The system exhibits the capability to degrade both pure substrate and mixed substrate feeds. The system responds rapidly to shifts in substrate, exhibits stable activity over time, is insensitive to moderate starvation times (15 hours), and is effective over a wide range of loading rates.

Comparing the degree of p-NBA degradation in reactor 1 under single substrate conditions with that for reactor 2 under mixed substrate conditions suggests that p-NBA metabolism is enhanced by concurrent metabolism of m-NBA. Only about 60% of the p-NBA was degraded in reactor 1 in the absence of m-NBA, versus greater than 95% degradation in reactor 2 when both m-NBA and p-NBA were fed at a concentration of 40 ppm. Conversely, the metabolism of m-NBA appears relatively unaffected by p-NBA degradation at lower p-NBA concentrations (40 ppm) but negatively affected at higher levels (80 ppm).

It is hypothesized that each isomer is metabolized by a different organism, though Kelly 7 is expected to degrade both isomers. The interaction between Kelly 4 and 7 is probably mediated by an extra-cellular product. A plausible hypothesis for the beneficial affect of m-NBA metabolism on p-NBA metabolism involves the pH inside the gel beads. We have established that p-NBA metabolism is reductive, releasing ammonia, while m-NBA metabolism is oxidative, releasing nitrite. The medium was initially at a pH of 7.5. In the absence of nitrite generation, the release of ammonia would cause the pH within the beads to rise significantly during p-NBA metabolism, inhibiting further cell growth or p-NBA metabolism. The rate of p-NBA metabolism would then be limited by the diffusion of ammonia out of the gel beads. If nitrite were being generated as well as ammonia, the pH change would depend upon the relative rates of generation of nitrite and ammonia. Calculations based on the diffusivity of ammonia in alginate and a bead diameter of 3 mm suggests that the pH inside the bead could rise above 8.0 given the observed rates of p-NBA metabolism.

This mechanism of interaction is consistent with the observation regarding the inhibition of m-NBA metabolism at higher p-NBA levels. In this medium, ammonia is the nitrogen source for the m-NBA degrader. At low relative rates of p-NBA metabolism, the greater availability of a nitrogen source inside the bead would be advantageous to the m-NBA degraders and would alleviate the acidic inhibition of growth that is often caused by nitrite generation. As the rate of generation of ammonia increases, the pH in the beads would rise, possibly to an inhibitory level.

This would explain the observation that the rate of m-NBA degradation per unit protein in reactor 2 drops 60% when the feed concentration goes from 40:40 to 80:80.

The assertion made regarding substrate inhibition is supported by certain details of Figure 9, which shows the percentage removal of m-nitrobenzoate as a function of reactor loading rate. Squares denote a feed concentration of 50 ppm m-NBA while circles denote 100 ppm. Data are taken from both reactors during the initial periods when only m-NBA was fed. Two trends are indicated. First, percentage removal drops as loading rate increases. Second, lower feed concentration leads to higher percentage removal. Of particular interest are the two data points connected by the arrow. These points were achieved sequentially. In going from point 1 to 2, the concentration of m-NBA in the feed was changed from 100 ppm to 50 ppm and the flow rate was doubled. The jump in percent removal indicates that the degradative activity in the reactor was inhibited by the higher m-NBA concentration.

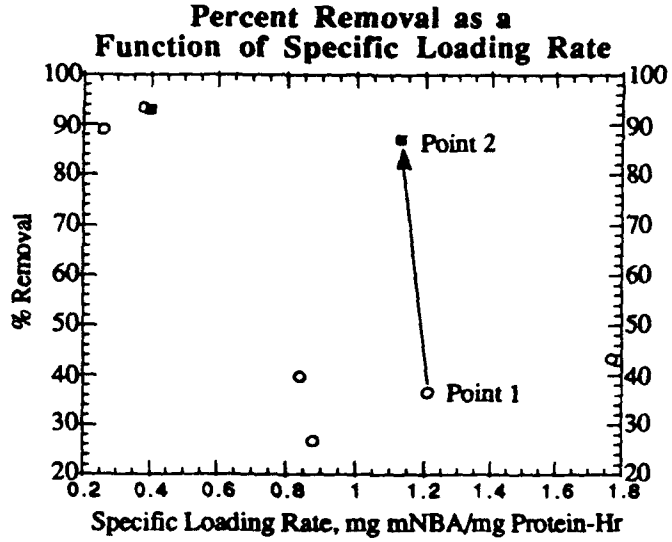


FIGURE 9

Another factor likely to contribute significantly to the behavior exhibited is the composition of the population in the reactor. As shown in Figure 6, the p-NBA concentration in reactor 1 decreases continually over time. The trend might indicate the replacement of m-NBA degraders by p-NBA degraders in the reactor. The transient is of too great a duration to be explained simply by induction of the p-NBA pathway. Population shifts could also explain the changes in degradative activity under mixed substrate feed conditions. The decline in m-NBA degradative activity

following the shift from 40:40 to 80:80 could be caused by a decline in the population most active in m-NBA metabolism.

CONCLUSION

Meta-nitrobenzoate and para-nitrobenzoate can be degraded rapidly and completely using a consortium initially isolated from the IWTP at Kelly AFB. The degradation occurs when either isomer is the sole carbon source or when both are present. Preliminary metabolic characterization of the consortium indicates that m-NBA degradation involves an oxidation pathway, while p-NBA degradative proceeds through a reductive pathway. Under certain cultivation conditions in an immobilized cell reactor, concurrent metabolism of both isomers is synergistic, leading to more rapid and complete degradation of each compound than observed in single-substrate studies. Under other conditions, the interaction appears inhibitory. Clarification of the mechanisms underlying these observations is necessary before reactor performance can be maximized. The pH sensitivity of the organisms involved is probably an important determinant and must be investigated more fully. Combined with measurements of the pH gradient inside the gel beads under different operating conditions, this information would help establish the effect of pH, nitrite and ammonia generation on degradation rates. In addition, enumeration of the population size of each species in the reactor, and elucidation of the spatial distribution of each species in the gel beads, are necessary to more clearly define the local environments these organisms experience. With such information, realistic models of reactor performance can be formulated and used to design and operate waste treatment systems effectively.

REFERENCES

1. Zeyer, J., H. P. Kocher and K. N. Timmis (1986) *Appl. Environ. Microbiol.*, **52**: 334-339.
2. Saber, D. and R. L. Crawford (1985) *Appl. Environ. Microbiol*, **50**: 1512-1518.
3. Brunner, W., F. H. Sutherland and D. D. Focht (1985) *J. Environ. Qual.*, **14**: 324-328.
4 Focht, D. D. and W. Brunner, (1985) *Appl. Environ. Microbiol.*, **50**: 1058-1063.
5. Dwyer, D. F., M. L. Krumme, S. A. Boyd and J. M. Tiedje, (1986) *Appl. Environ. Microbiol.* , **52**: 345-351.
6. Mikesell, M. D. and S. A. Boyd (1986) *Appl. Environ. Microbiol.*, **52**: 861-865.
7. Spanggord, R. J., J. C. Spain, S. F. Nishino, and K. E. Mortelmans (1991) *Appl. Environ. Microbiol.*, **57**: 3200-3205.
8. Stanier, R. Y., N. J. Palleroni, and M. Doudoroff (1966) *J. Gen. Microbiol.*, **43**: 159-271.
9. Kuhn, R. H., S. W. Peretti and D. F. Ollis, *Biotechnol. Bioeng.*, **38**: 340-352.

SOIL COLUMN STUDIES WITH A FIBER-OPTIC LASER SPECTROMETER

Brian S. Vogt
Professor
Department of Chemistry

Bob Jones University
Greenville, SC 29614

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

August, 1993

SOIL COLUMN STUDIES WITH A FIBER-OPTIC LASER SPECTROMETER

Brian S. Vogt
Professor
Department of Chemistry

Abstract

A fiber-optic laser spectrometer was used in soil column experiments to study transport of contaminants through soils. A unique fiber-optic laser probe was adapted into a modular stainless-steel column system so that the probe could be used to study the transport of fluorescent contaminants through soils in the column. It was verified that the probe performed well when sealed inside a soil column. Retention characteristics of naphthalene and amino G acid (7-amino-1,3-naphthalene disulfonic acid) on washed sand in columns were determined: amino G acid passed through the column quickly while naphthalene was retained for a significant period of time. A procedure for ensuring accurate calibration of the laser spectrometer was delineated. It was found that low levels of suspended solids do not interfere significantly with the fluorescence of amino G acid. Moderate levels of suspended solids were compensated for by using turbidity measurements made with the laser probe.

SOIL COLUMN STUDIES WITH A FIBER-OPTIC LASER SPECTROMETER

Brian S. Vogt

INTRODUCTION

The Environics Directorate of Armstrong Laboratory (AL/EQ) conducts research dealing with many environmental issues. Among these is the development of new technology to monitor and remediate contaminated Air Force sites. Another area of research at AL/EQ is the use of model aquifers to study the fate and transport of contaminant plumes. Data from model aquifers are mathematically modeled, permitting a better understanding of how contaminants are distributed when released into the environment. The work reported here involves both monitoring technology (fiber-optic laser fluorescence spectroscopy) and fate and transport research (soil column studies).

Traditional site evaulation entails collecting soil and groundwater samples from wells drilled into the soil and then analyzing those samples in a laboratory with methods such as liquid chromatography (LC), gas chromatography (GC), or liquid scintillation counting (LSC). The acquisition, transport, and analysis of samples by these methods are moderately time-consuming. Fiber-optic laser fluorescence spectroscopy is an alternative method of analysis that permits rapid *in situ* measurements of fluorescence coming from aromatic hydrocarbons, which are components of many fuels and solvents. This method relies on optical fibers to convey excitation light from a laser to a sample. Analyte fluorescence induced by the laser light is carried from the sample back to optics and electronics for analysis and display. Analysis of toluene in monitoring wells at Tinker AFB, OK, has been performed with a transportable fiber-optic spectrometer using a tunable dye laser as an excitation source[1]. Results of the fluorescence analysis in that study compared favorably to results of samples analyzed by GC.

Additional time and cost savings in site evaluation can be realized by using cone penetrometry. The SCAPS (Site Characterization and Analysis Penetrometer System) program is a collaborative effort of the Air Force, Army, and Navy initiated to develop efficient site characterization methods. A cone

penetrometer is a truck-mounted system that hydraulically pushes a hardened steel tube into the ground at a rate typically on the order of one meter per minute. The tube contains a variety of sensors or soil and groundwater sampling tools. Adaptation of a fiber-optic probe into a cone penetrometer permits rapid spectroscopic characterization of a contaminated site without drilling monitoring wells. Results of field measurements with fiber optics configured in cone penetrometers using a nitrogen laser[2] or a Nd:YAG-pumped dye laser[3] for fluorescence excitation have been reported. A comparison of instruments of these types was recently made[4].

APPARATUS

The laser system was designed and installed at AL/EQ by G.D. Gillipsie and his research group[5]. The principles of operation have been described elsewhere[6]. Further details may be found in the full report filed at AL/EQ[7].

Figure 1 is a cut-away view of the end of the probe showing how it is constructed[8]. This probe is a modification of that designed and tested during the Summer of 1992[9]. In summary, one end of a five-meter bundle of seven fibers is glued inside a foot-long stainless-steel tube for support. The

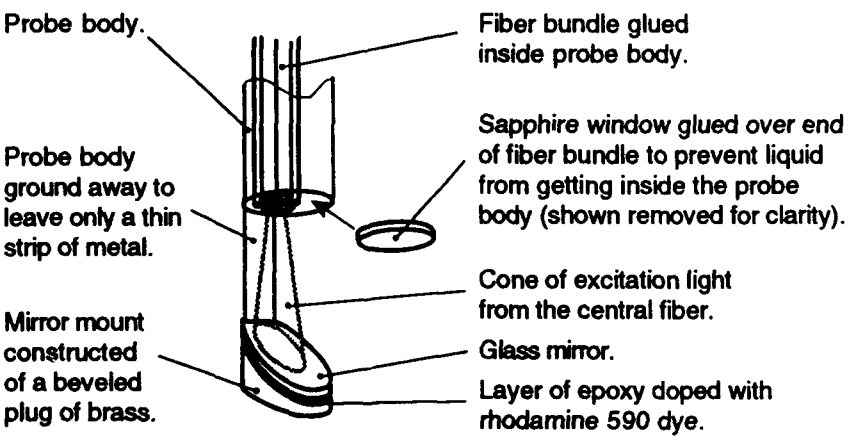


Figure 1. Construction of the Fiber-Optic Probe.

other end of the bundle is interfaced to laser and monochromator optics. The probe is inserted into a sample to be analyzed. Laser light enters the solution from the central fiber in the bundle and excites fluorescence from the analyte in the solution. Five of the surrounding fibers are used to

collect fluorescence and transmit it to the monochromator. Without the mirror attached to the probe end it was found that some of the light signals measured were sensitive to the position of the probe inside the vessel containing the sample. This phenomenon was attributed to the reflection of light off the walls of the container. The mirror presents a constant, controlled optical view to the collection fibers and minimizes problems due to light scattering and reflection. The mirror is constructed of microscope slide glass and is angled to reflect UV laser light out of the view of the collection fibers. UV light that is neither absorbed by the sample nor reflected by the mirror enters into the mirror glass. Most of the UV light that enters the mirror is absorbed instead of being reflected back into the optical path.

Water in monitoring wells in both real and model aquifers often contains suspended solids that render the water turbid. Turbid solutions can scatter both excitation light and analyte fluorescence. Light scattering leads to erroneously low signals. It is necessary, therefore, to quantify sample turbidity and make signal corrections. Ideally, measurements of both fluorescence and turbidity measurments would be made with a single probe. This is easily accomplished with the probe used in this study. Only five of the fibers surrounding the central fiber are used for collection of analyte fluorescence. The remaining surrounding fiber is used to tramsit a portion of the visible light from the Nd:YAG laser (532 nm) to the probe end, where it passes through the solution and glass mirror. Once the visible 532 nm light travels through the mirror, it excites fluorescence from the rhodamine 590 dye embedded in the epoxy. Some of this fluorescence is scattered into the collection fibers. Any solids suspended in the sample will attenuate both the amount of excitation light reaching the rhodamine 590 and the amount of rhodamine 590 fluorescence reaching the collection fibers. The greater the turbidity, the lower the rhodamine 590 fluorescence signal detected will be. A solution with no suspended solids is used as a reference. This approach assumes that the suspended solids act as a neutral density light filter. As will be seen, it appears that this assumption was valid in these experiments. Further details about how the probe is used for both fluorescence and turbidity measurements are available[10].

The compounds studied in this research were naphthalene (Aldrich Chemical Company, purified by sublimation[11]) and 7-amino-1,3-naphthalene disulfonic acid (monopotassium salt) (Eastman Fine Chemicals, used as received). The latter is sometimes called amino G acid (abbreviated AGA hereafter). Naphthalene exhibits UV fluorescence and was monitored at 335 nm instead of its peak maximum of 323 nm to avoid a water Raman signal. AGA exhibits blue fluorescence and was monitored at 445 nm, which corresponds to the maximum in its fluorescence spectrum. Rhodamine 590 was monitored at 590 nm. The spectra of the analytes were so different from the spectrum of rhodamine 590 that analyte fluorescence did not pose any significant interference in the measurement of rhodamine 590 fluorescence. If such an interference had been a problem, it could have been easily eliminated by inserting a shutter into the optical path to prevent 287 nm UV excitation light from entering the probe and exciting the analytes. Changing excitation wavelengths with the dye laser used in this study is accomplished by rotating a mirror on a mechanical stage. The dye laser remained tuned at 287 nm for both AGA and naphthalene excitation. This gave adequate results and eliminated the need to change dye laser output wavelengths.

The probe was adapted into a stainless-steel column and pump apparatus[12]. Two LC pumps (Waters 510) were connected to the inlet end of the column with a six-port LC sample loop injector (Valco C6U). One pump was used to pump background solution through the column; the other was used to pump contaminant plumes. The pumps and column were connected to the sample loop injector in such a way that the flow from one pump was directed to waste while the flow from the other was directed to the column. A liquid handler (Gilson 212B) was connected to the effluent end of the column to permit fractionation of the effluent for subsequent analysis. The background solution consisted of distilled water containing 0.005 M $CaSO_4$ (calcium sulfate) and 0.02 w/w % NaN_3 (sodium azide). The $CaSO_4$ serves to raise the ionic strength of the solution; the NaN_3 was added to prevent microbial and algal growth. This mixture has been used successfully in batch, column, and model aquifer sorption studies[13]. All calibration standards and contaminant plumes were made up in background solution.

Initial attempts to adapt the probe into a glass column were abandoned because the glass was too fragile. It was undertaken to use a stainless-steel column instead, details of which are shown in Figure 2[14]. Different columns lengths were easily achieved simply by replacing the upper and lower column sections with tubing of different length. The length of the upper column section was 7.5 cm in all experiments. The time required for any given

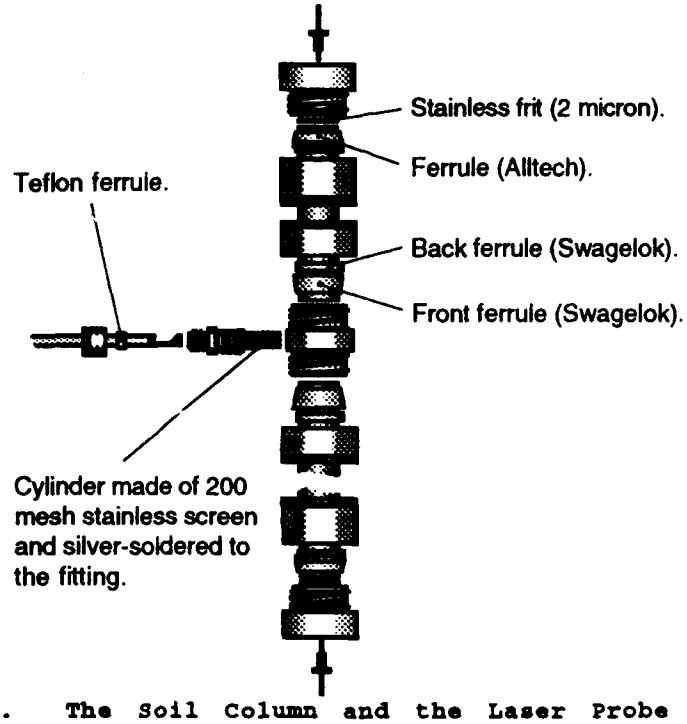


Figure 2. The Soil Column and the Laser Probe Interface.

experiment was a function of both column length and pump rate. Initially the length of the lower section was 47.8 cm. It was, however, cut to 15.0 cm and then 7.5 cm to shorten experiment times. The column was filled with a slurry of washed sand in background solution.

RESULTS

It has been reported[15] that the fluorescence of a given concentration of AGA is constant between pH 6.3 and 9.4. The effect of sand on the pH of water was checked. It was observed that addition of 15 g of unwashed sand to 60 mL of distilled water changed the pH to about 6.2 after stirring for two minutes.

Monitoring the pH for approximately 35 minutes while continuing to stir showed that the pH gradually rose to about 6.5. It was unknown whether the pH of the background solution would remain between 6.3 and 9.4 during column studies. The pH of a solution of AGA was varied between pH 5.3 and 8.9. The fluorescence of this solution was monitored with the laser probe and found to be constant. Consequently, it was decided that buffering the background solution was unnecessary.

The dye laser used to excite fluorescence in this study exhibited some variation in output power with respect to time. Fluorescence intensity is a function of excitation power. Consequently, the fluorescence signal from any given solution varied with time. A photodiode was positioned in the optical system to monitor the amount of UV light scattered off the face of the laser probe excitation fiber. The signal from the photodiode was recorded whenever a fluorescence measurement was made. Each fluorescence signal was subsequently corrected by normalizing to a photodiode signal of 50 mV. A value of 50 mV was chosen because it was typical of photodiode signals. All fluorescence readings, including those from calibration standards, were corrected in this fashion. It was initially assumed that the photodiode signals were sufficiently reproducible from day to day so that daily calibration would be unnecessary. However, after several experiments it was found that AGA calibration behavior exhibited some variation from day to day that could not be corrected for with photodiode readings. It was suspected that the AGA standards were decaying, but spectrophotometer absorbance measurements indicated that this was not so. At one point, however, the photodiode had to be replaced because its signal had degraded so far as to give inadequate response. It is believed that the UV light incident upon the photodiode was damaging it, causing its response to a given level of dye laser power to change over moderate periods of time. It is felt that the normalization of fluorescence signals based on photodiode signals in this experimentation was effective during a given day but not necessarily from day to day. It is also possible that the spatial characteristics of the excitation beam varied. This may have changed the effective volume of the cone of light entering the solution from the laser probe, which in turn would

change the amount of AGA excited. Consequently, it was decided that calibration would be performed every day the laser was used for analysis.

The limit of detection was estimated using the IUPAC method[15] and found to be 0.17 ppb AGA from typical calibration data. Calibration data for AGA were broken down into two sections with a linear fit for the low concentration data and a second-order polynomial fit for the high concentration data. This gave greater accuracy when estimating AGA concentrations from signal measurements than using a single equation for all the calibration data because no single equation describes all of the data adequately.

Figure 3 shows breakthrough curves for AGA on a sand column with a 47.8 cm lower column section. Pumping rate was 1 mL/min. The contaminant plume

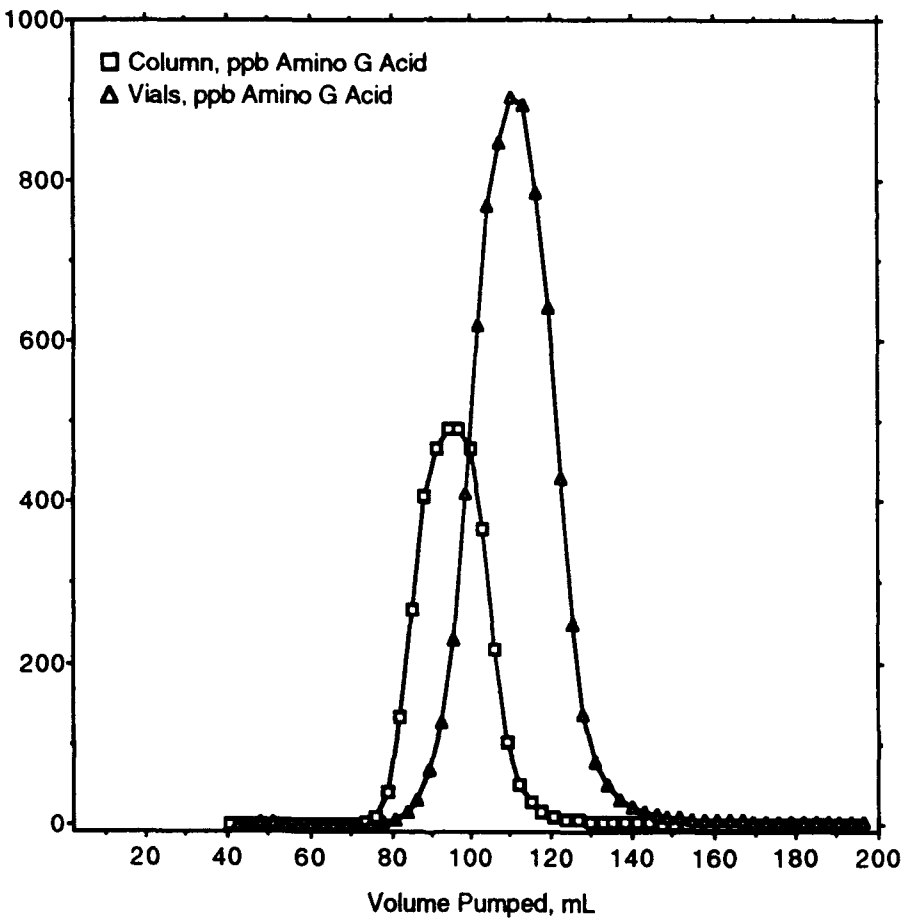


Figure 3. Breakthrough of AGA on Sand (1 mL/min Pump Rate).

was a 19.8 mL pulse of 1200 ppb AGA. The left-hand, lower-volume profile in Figure 3 represents AGA concentrations detected by the probe while it was embedded in the column. Each point on the other profile is the concentration of AGA detected in a vial containing effluent dispensed into it by the liquid handler. The probe was then removed from the column to perform analysis on the vial contents. The vial profile occurs at higher volume (later in time) than the column profile because the contaminant plume must travel past the probe and through the upper 7.5 cm column section before reaching the vials.

The mass of AGA recovered in each vial was calculated by multiplying the AGA concentration by the volume of effluent in the vial. The sum of the masses recovered in all vials was 22.1 µg, which is 93.1% of the 23.8 µg in the plume. This experiment was performed before it was known that daily calibration would be required. Consequently, the AGA concentrations in this experiment were estimated from calibration data acquired on a different day. Inaccuracy in AGA concentrations may account for the less than 100% recovery. Trapezoidal integration of the column profile gave an area of only 10.6 µg, or only 44.7% recovery of the mass contained in the plume. A simple hypothesis was constructed to account for this. The cylinder housing the end of the probe in the column consists of very fine, 200 mesh stainless steel (see Figure 2). It is conceivable that the highly fractured edges of the sand particles blocked some of the openings in the screen. At high pump rates, one would expect partial blockage to present a great enough resistance to liquid flow that some channeling around the mesh cylinder would occur. In that event not all of the analyte would pass through the optical path of the probe and the concentrations detected by it would be low, giving the behavior evident in Figure 3. To test this hypothesis, the sand was removed from the column and replaced with a mixture of relatively large glass beads (4-6 mm in diameter). The lower column section was shortened to 150 mm in order to shorten experiment time. The experiment was repeated with a 4.98 mL plume of 1200 ppb AGA and a pump rate of 1 mL/min.

Analysis of the vial contents indicated a recovery of 5.81 µg, or 97.9% recovery. This recovery was better than in the previous run because a single-point calibration of the 1200 ppb plume was used to corrrect the calibration data of a previous day. This correction was not entirely effective because

correcting a nonlinear calibration plot with a single point is subject to a moderate amount of error. Integration of the column profile gave 7.57 µg of AGA, or 127% recovery. The fact that the recovery was not low supports the idea that channeling occurred in the previous experiment. This led to lower pump rates in subsequent experiments. That the recovery was signficantly higher than 100% merits some explanation. Putting large particles in a column results in higher porosity at the column walls than in the column center because particles are packed more tightly in the center. Consequently, fluid flow is radially heterogeneous. In this experiment it appears that the AGA concentration was also radially heterogeneous. Higher AGA concentrations at the center of the column than at the edges explain the observed behavior.

Figure 4 shows the results of an AGA column experiment performed with a

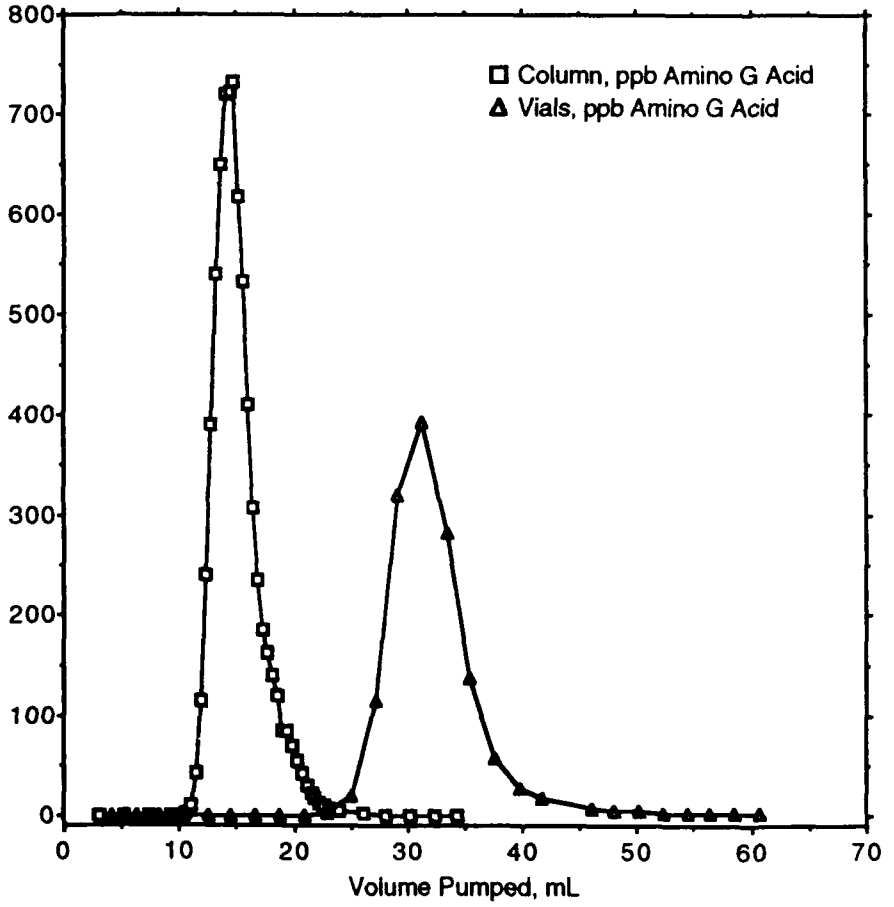


Figure 4. Breakthrough of AGA on Sand (0.2 mL/min Pump Rate).

low pump rate (0.2 mL/min) to determine whether channeling around the screen cylinder could be avoided. This and all subsequent column experiments were performed with a 7.5 cm lower column section. A 2.82 mL plume of 1020 ppb AGA was used. Recovery in the vials was 2.94 µg (102% of the mass in the plume).

Integration of the column profile yielded an area of 3.09 µg (107% recovery). These recoveries are good. The deviation from 100% is not hard to explain. The value of the volume for any given point in the breakthrough curves presented is calculated from the pump rate and the time of pumping. For this experiment the flow rate was 0.21 mL/min. If it had been 0.20 mL/min instead, the column recovery would come out to 102% and the vial recovery 97.3%. Consequently, a small amount of inaccuracy in the pump rate determination explains all of the error. In this and prior experiments the flow rate was obtained by dividing the volume of effluent collected in a 5 mL graduated cylinder by the time it took to collect it. In subsequent experiments the volume collected was calculated by weighing the effluent collected in a given period of time and dividing it by the density of the effluent, which was taken to be the density of the background solution. The density (0.9971 g/mL) was determined by weighing 50 mL of solution in a volumetric flask. Error in timing and any inaccuracy inherent to the analysis will also affect calculated recoveries.

The fact that the areas of the two curves in Figure 4 are essentially identical suggests that no channeling was occurring around the laser probe. Clearly, the probe performs well in the soil column environment. It should be straightforward to fit a screen cylinder around the end of the probe to permit measurements to be made with the probe embedded in model or real aquifers. One would expect fluorescence measurements from such environments to be representative of actual analyte concentrations only if the rate of water flow is comparable to or lower than the 0.2 mL/min pump rate used in this experiment. Estimates of groundwater flow rates are often reported as linear velocities, which are in units of distance/time rather than volume/time. The estimation of linear velocity in this experiment was made as follows. First, it was assumed that AGA was not adsorbed to any appreciable extent on the sand. One study reported that losses of AGA due to adsorption on soil ranged

from only 0.5% up to 80%, depending on soil type[17]. Second, it was assumed that the pore volume of the lower column section was 15 mL, which is the volume corresponding to the maximum of the column profile in Figure 4. At a flow rate of 0.2 mL/min, the contaminant plume would reach the laser probe in (15 mL)/(0.2 mL/min) or 75 min. In the 75 mm column that was used, this translates to a linear flow velocity of 1 mm/min or 1.4 m/d. Making different assumptions in this estimation would result in a value higher than 1.4 m/d. Reported average linear groundwater velocities for a sand aquifer in Ontario, Canada, vary between 0.076 m/d and 0.091 m/d, depending on the method of estimation used[18]. The linear groundwater velocity at Columbus Air Force Base, Mississippi, USA, can be estimated to be no greater than 0.4 m/d by using published tritium plume data[19]. These groundwater velocities are significantly lower than those employed in these column studies. This suggests that, barring other complications, fluorescence measurements made with a screen-protected laser probe embedded in soils could be used to reliably estimate contaminant concentrations. The same pertains to model aquifer studies employing appropriate flow conditions.

In another effort to shorten experiment time, the experiment represented by Figure 4 was repeated with a 2.937 mL plume of 1020 ppb AGA and a background pump rate of 0.3 mL/min. Recoveries were 2.99 mg (99.8% recovery) for the column and 2.98 mg (99.5% recovery) for the vials. Comparison of the 0.3 mL/min data to those in Figure 4 showed that the results from the two experiments are essentially identical, indicating that 0.3 mL/min was low enough to prevent channeling around the laser probe. Consequently, all subsequent experiments were performed with a pump rate of 0.3 mL/min.

Naphthalene is of interest in the monitoring of Air Force sites because it is a key component of jet fuels. Several experiments involving naphthalene were performed. The limit of detection was estimated to be 5.63 ppb from typical calibration data.

A column experiment was performed with a 2.944 mL plume of 1530 ppb naphthalene. Integration of the profile gave an area of 4.24 µg (94.1% recovery of the mass in the plume). Fractionation of the effluent into many vials was not performed in this or subsequent experiments because the AGA experiments demonstrated that integration of the column and effluent profiles

gives identical results when experimental conditions are properly controlled. However, the effluent was collected in flasks for analysis to check naphthalene mass balance. Analysis of the effluent showed a recovery of 4.34 µg of naphthalene (96.4% recovery), which agrees well with the the profile integration.

The last column experiment was performed with 2.958 mL of contaminant plume containing 1030 ppb AGA and 1530 ppb naphthalene. Fluorescence measurements at the analytical wavelengths for both were taken alternately as a function of time. The results are shown in Figure 5. Recovery estimated by

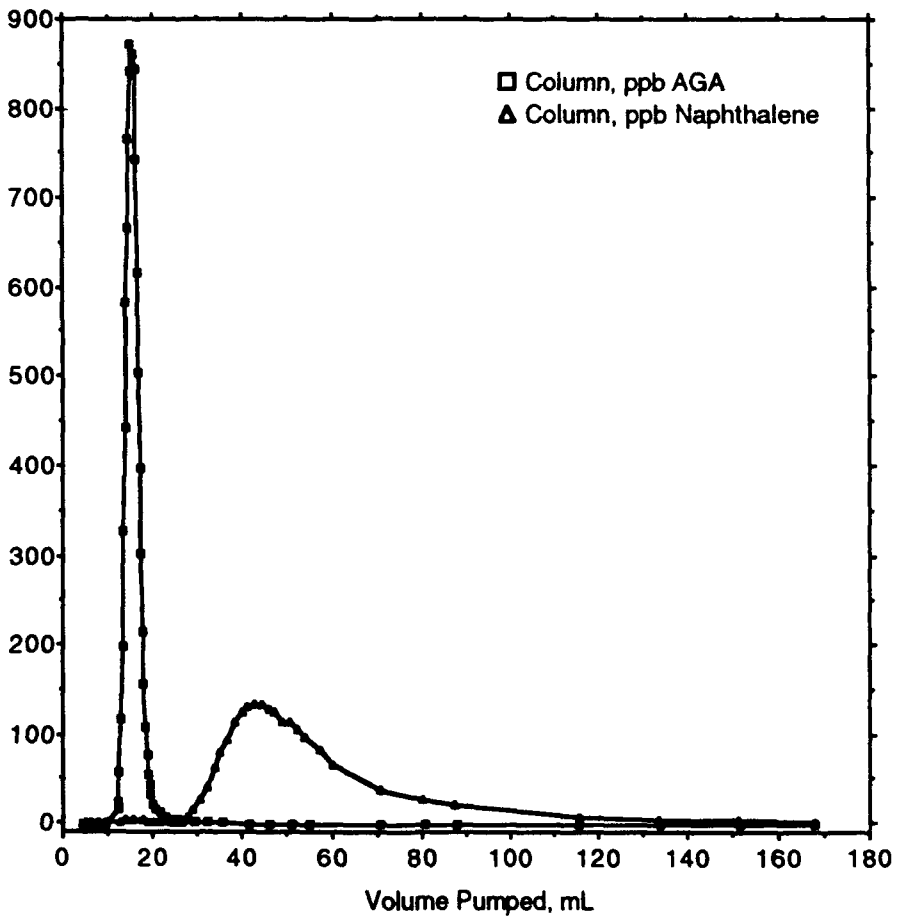


Figure 5. Breakthrough of AGA & Naphthalene on Sand.

integration of the profiles was 4.64 µg (103% recovery) for naphthalene and 3.04 µg (99.6% recovery) for AGA. The effluent was collected in three flasks

and analyzed. Recovery was 2.90 µg of AGA (95.1% recovery) and 3.75 µg of naphthalene (82.8% recovery). All of the 3.75 µg of naphthalene recovered was present in the same effluent recovery flask. Assuming that the remainder of the naphthalene was present in the following flask, calculations show that it would have been present at a concentration below the limit of detection. This explains why it appears that some naphthalene is missing.

Sorption coefficients are constants that describe the distribution of a contaminant (AGA or naphthalene) between the liquid phase (background solution) and the solid phase (sand). They are used in characterizing the interaction between contaminants and soils. The sorption coefficient has the same form as an equilibrium constant. Apparent sorption coefficients may be estimated even in systems that do not reach equilibrium. At sufficiently high sorption rates, the adsorption process does appear to reach equilibrium even in flowing systems[20]. The determination of sorption coefficients in columns depends upon comparing the volume pumped when the solvent in the contaminant plume passes a given point in the column to the volume pumped when the contaminant passes that same point. Consider the naphthalene profile in Figure 5. The long tail indicates that naphthalene travel through the column is retarded by the sand. This is qualitatively similar to previously observed naphthalene behavior[21]. However, the water in the contaminant solution is not retained by the sand. If one could follow the progress of that water through the column, one would observe it passing the laser probe much sooner than the naphthalene does. One could then estimate the sorption cofficient of naphthalene on the sand by comparing the breakthrough behavior of naphthalene to that of water. Monitoring the contaminant plume water is possible when collecting the effluent in fractions. One can add tritiated water to the plume and analyze the effluent for it with LSC. That process, however, introduces additional sample handling and analysis time. It would be much more convenient to be able to monitor the contaminant solvent plume with the laser probe instead of having to collect samples for separate analysis. The AGA profile passes the laser probe at a relatively low volume and is comparatively sharp and symmetrical, indicating that it is retarded by the sand minimally or not at all. If it is completely unretarded by the sand, AGA could be used to monitor the contaminant solvent plume fluorimetrically

instead of by LSC with tritiated water, thus permitting complete characterization of the adsorption process by in situ fluorescence measurements. Although they have yet to be performed for AGA on sand, batch equilibration studies can be used to determine the degree of adsorption on any given matrix. Even if AGA is adsorbed to some extent, one could perform a preliminary experiment with tritiated water and AGA to determine by how much the AGA profile is displaced from the profile of the tritiated water. Subsequent experiments could then rapidly determine the relationship of various contaminants to AGA, and by using a combination of the data determine sorption coefficients. This would provide a signficant time and cost savings if several column experiments were performed.

As previously mentioned, samples from model and real aquifers often contain suspended solids. For this reason the effect of suspended solids on AGA fluorescence was investigated. A slurry of unwashed sand in background solution was made. The large solids were permitted to settle out and the supernatant, which was turbid due to the solids it contained, was poured off. 10 mL aliquots of this turbid suspension were evaporated in an oven and the residues weighed to determine the mass of dissolved and suspended solids contained in the aliquots. The same was done with 10 mL aliquots of background solution. Subtraction indicated that the sand suspension contained 550 mg/L of suspended solids. A normal set of AGA calibration standards in background solution was prepared in volumetric flasks. A duplicate set, identical to the first except that a fixed volume of suspension was added to each flask, was also prepared. The suspension was maintained during the preparation of these solutions by stirring with a magnetic spin vane. Each pair of solutions was then titrated together to obtain fluorescence readings for a fixed concentration of AGA with varying amounts of suspended solids. For example, a normal 25.7 ppb AGA solution was titrated with a 25.7 ppb AGA solution containing suspended solids. These solutions were stirred with magnetic spin vanes throughout the process. The resulting data are presented in Figure 6. Each curve represents a calibration curve for AGA in the presence of a fixed amount of suspended solids. The trend is not surprising: AGA fluorescence drops as turbidity increases. Visual comparison of the 200 mg/L solution to photographs of solutions obtained from monitoring wells at

Tinker Air Force Base suggested that the turbidity of the 200 mg/L solution used in this study approximated a worst case scenario from a real site.

The fluorescence signal from the rhodamine 590 was measured at 590 nm immediately after each 445 nm AGA measurement. Each AGA fluorescence measurement was corrected by dividing it by the fraction of rhodamine signal left in the turbid solution. For example, if the rhodamine signal in the turbid solution was 90% of what it was in the corresponding solution with no

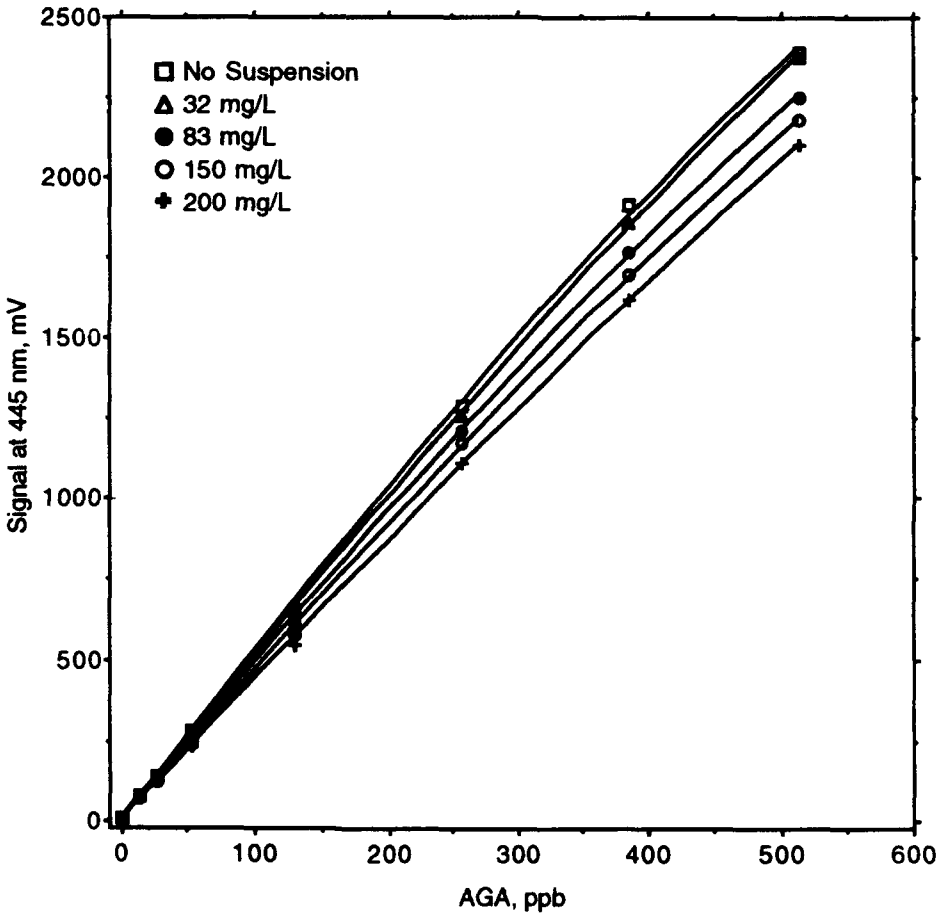


Figure 6. Affect of Turbidity on AGA Fluorescence.

suspended solids, the AGA fluorescence signal was corrected by dividing it by 0.90. Figure 7 shows the results of these corrections for the 150 mg/L suspensions. The corrected 150 mg/L calibration curve is essentially collinear with that corresponding to the solutions with zero suspended solids,

indicating that the correction worked well. It also suggests that 445 nm light and 590 nm light were scattered to the same extent by the suspended solids used in this experiment. This is not necessarily true with solids from other sources, such as clay-containing soils. Experiments need to be performed with other fluorophores and soil types to determine the scope of applicability of the process used in this study. The idea of being able to make in situ fluorescence measurements and correct them for turbidity in the sample by making measurements with the same probe at essentially the same time is very attractive.

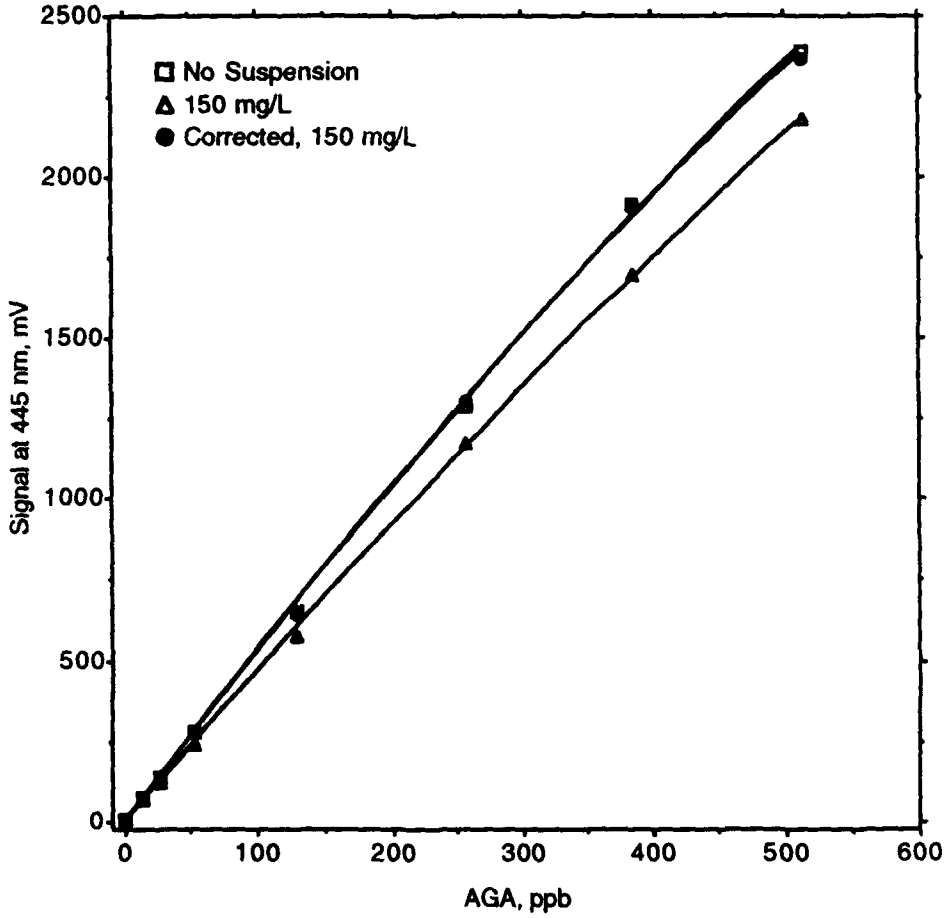


Figure 7. Correction of AGA Fluorescence Data for Turbidity.

All of the solutions containing turbidity were treated as unknowns. The concentration of AGA was estimated from both uncorrected and corrected

fluorescence readings. The results are given in Table 1. Additional calculations demonstrating the effectiveness of the correction process were done; ways to work around the limitations of the process were discussed[22].

Table 1. Average Errors in Calculated AGA Concentrations.

Sand, mg/L	Average Percent Error Without Turbidity Correction	Average Percent Error With Turbidity Corrrection
32	2.8	2.4
83	6.6	2.6
150	11	2.8
200	13	4.2

CONCLUSION

The soil column experiments performed in this research demonstrate not only the feasibility of using a fiber-optic laser spectrometer for such studies, but also show that such an approach is accurate. Measurements made with the probe adapted into the column permit rapid, in situ analysis to be performed, obviating the need to collect samples for subsequent analysis. The practicality of making both fluorescence and turbidity measurements with a single probe has not only been demonstrated, but shown to be highly successful.

Work currently in progress centers on several areas. First, the probe is being miniaturized from one-fourth inch in diameter to one-eighth inch in diameter. Second, efforts are being made to make the probe more durable. The third area is probe multiplexing to permit several probes to be connected to a single optical system. This will permit the semi-simultaneous monitoring of several locations within aquifers. Last, it is now possible for a model aquifer study to be done as soon as the apparatus for such a study is ready.

REFERENCES

1. G.D. Gillispie and R.W. St. Germain, "In-Situ Tunable Laser Fluorescence Analysis of Hydrocarbons," *Environmental and Process Monitoring Technologies*, Tuan Vo-Dinh, Editor, *Proc. SPIE 1637*, 151-162 (1992).
2. S.H. Lieberman, G.A. Theriault, S.S. Cooper, P.G. Malone, R.S. Olsen, and P.W. Lurk, "Rapid, Subsurface, In Situ Field Screening of Petroleum Hydrocarbon Contamination Using Laser Induced Fluorescence Over Optical Fibers," *Proc. of the Second International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals*, 57-63 (1991).

3. W.L. Bratton, J.D. Shinn, S.M. Timian, G.D. Gillispie, and R. St. Germain, "The Air Force Site Characterization and Analysis Penetrometer System (AFSCAPS); Laser Induced Fluorescence Cone Penetrometer. Volume I-System Development and Evaluation," Air Force technical report AL/EQ-TR-1993-0009 (1993).
4. G.D. Gillispie, R.W. St. Germain, and J.L. Klingfus, "Subsurface Optical Probes: Current Status and Future Prospects," *Proc. of the Third International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals,* in press (1993).
5. G.D. Gillispie, North Dakota State University, Department of Chemistry, Ladd Hall, Fargo, ND 58105-5516.
6. G.D. Gillispie and R.W. St. Germain, "In-Situ Tunable Laser Fluorescence Analysis of Hydrocarbons," *Environmental and Process Monitoring Technologies,* Tuan Vo-Dinh, Editor, *Proc. SPIE 1637,* 151-162 (1992).
7. This report is abbreviated due to length requirements. A full report of the same title was filed at AL/EQ.
8. The probe used in this study was constructed by Dr. Terence Tipton, AL/EQC, 139 Barnes Drive, Suite 2, Tyndall AFB, FL 32403-5323 PO Stop 37. The author appreciates the probe construction and optical configuring and tweaking performed by Dr. Tipton.
9. B.S. Vogt, "Application of Fiber-Optic Laser Fluorescence Spectroscopy to Environmental Monitoring," *USAF Summer Research Program--1992, Summer Faculty Research Program (SFRP) Reports, 6,* Air Force technical report AFOSR-TR-93-0116 (1992).
10. See reference 7 for further details.
11. This purification was performed by C1C Anthony L. Mitchell, who was involved in this research for several weeks due to his participation in the Air Force Academy Cadet Summer Research Program. The author appreciates Cadet Mitchell's assistance during his visit to Tyndall AFB.
12. See reference 7 for further details.
13. W.G. MacIntyre, T.B. Stauffer, and C.P. Antworth, "A Comparison of Sorption Coefficients Determined by Batch, Column, and Box Methods on a Low Organic Carbon Aquifer Material," *Ground Water, 29,* 908-913 (1991).
14. See reference 7 for further details.
15. P.L. Smart and I.M.S. Laidlaw, "An Evaluation of Some Fluorescent Dyes for Water Tracing," *Water Resources Research, 13,* 15-33 (1977).
16. J.D. Winefordner and G.L. Long, "Limit of Detection: A Closer Look at the IUPAC Definition," *Analytical Chemistry, 55,* 712A-724A (1983).
17. S.T. Trudgill, "Soil Water Dye Tracing, with Special Reference to the Use of Rhodamine WT, Lissamine FF, and Amino G Acid," *Hydrological Processes., 1,* 149-170 (1987).
18. D.M. Mackay, D.L. Freyberg, and P.V. Roberts, "A Natural Gradient Experiment on Solute Transport in a Sand Aquifer: 1. Approach and Overview of Plume Movement," *Water Resources Res., 22,* 2017-2029 (1986).
19. J.M. Boggs, L.M. Beard, W.R. Waldrop, T.B. Stauffer, W.G. MacIntyre, and C.P. Antworth, *Transport of Tritium and Four Organic Compounds during a Natural-Gradient Experiment (MADE-2),* Electric Power Research Institute technical report TR-101998 (project 2485-05), 5-7 (1993)
20. W.G. MacIntyre, T.B. Stauffer, and C.P. Antworth, op cit.
21. Ibid.
22. See reference 7 for further details.

TRANSPORT DELAY MEASUREMENT IN FLIGHT SIMULATORS

Hussein F. Almallahi

Department of Electrical Engineering Technology

Prairie View A&M University

P.O. Box 308

Prairie View, TX 77446

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

August 1993

TRANSPORT DELAY MEASUREMENT IN FLIGHT SIMULATORS

Hussein F. Almallahi
Department of Electrical Engineering Technology
Prairie View A&M University

Abstract

Transport delay measurement has been studied over the past few years in both methods the time domain and the frequency domain techniques. This report explain the phase-lock loop method to measure the delay. The phase lock acquires lock after certain time, this time is correlated to the transport delay time. Also the loop error decays to zero as the loop track the incoming signal as soon as the phase lock loop(pLL)is in complete lock. This report will discuss the delay in general and how the phase-lock loop will be able to track the incoming signal and measure the delay.

TRANSPORT DELAY MEASUREMENT IN FLIGHT SIMULATORS

Hussein F. Almallahi

Introduction

The determination of simulation delay in the flight simulator is essential in rendering the utility of that simulator useful. Time and frequency domain analysis or the combination of both have been used in the past to estimate this transport delay. The delay in question is attributed to the combined effect of the processors (software and hardware and the physical channel) on the signal originated by the pilot at the cockpit. the delay here is defined as the time elapsed from the point of starting an action by the pilot until a response (due to this action) is visible on the visual computer.

In digital and analog communications, the need to estimate the phase of the carrier is essential in extracting the information from the incoming signal. It is customary in dealing with these applications that a linearized phase detector is used as a part of the receiver to extract the phase information (delay due to the Channel or otherwise) of

the carrier. This arrangement is known as the phase-locked loop. The utility of this phase-lock loop is to synchronize the transmitter to the receiver. In other words the phase lock loop is going to serve as a clock where both the transmitter and the receiver tune to. This timing circuitry may well be used in determination of the time delay associated with the Flight Simulator. Even the delay associated with the real time flight may also be assessed by the phase lock loop.

Phase Lock Loop Method

Phase lock loop are servo-control loops, whose controlled parameter is the phase of locally generated replica of the incoming carrier signal. Phase lock loop consists of three basic component, a phase detector which is a device that produce a measure of the difference in phase between the incoming signal and the local replica. The incoming signal and the local replica change with respect to each other, will cause a loop error. The loop error decays to zero as the loop tracks the input signal. The phase difference will become time varying signal into the loop filter. Then the other component is the loop filter. The loop filter is a low pass filter used to remove the second harmonic of the carrier (control the phase lock loop response to the variation in the error signal). Then, the

last component is the Voltage Controlled Oscillator VCO. The VCO is a device that produces the carrier replica. Figure 1 shows the phase detector as a multiplier, the loop filter is described by its impulse function, f(t), and the voltage controlled oscillator VCO is also shown. r(t) is the incoming signal, e(t) is the phase error between the incoming signal and the output of the VCO, and y(t) is the input of the VCO.

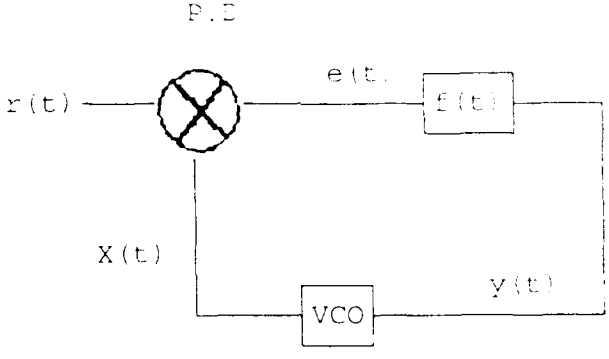


Figure 1 Schematic diagram of phase lock loop

The general principle of operation of a phase lock loop may be explained in the following fashion. A phase modulation scheme known as Binary Phase Shift Keying (BPSK) is used to illustrate the operation. In this signaling scheme the phase of the cosinusoid is varied as the data modulates the

carrier. The phase of the carrier takes on +pi or zero depending on whether a bit one is transmitted or zero (high or low logical levels) as shown below:

$S(t) = A \cos (wt + phi)$

where

phi = 0 or pi

A reprents the amplitude of the carrier

W reprents the carrier frequency

t corresponds to time

and s(t) denotes the noise free phase modulated signal at he output of the transmitter.

As the signal passes through the Channel, noise and filtering (Linear or non linear) corrupts the transmitted signal and it undergoes through a delay due to the filtering process. This manifests in the phase of the carrier receive waveform) as given below.

$R(t) = A \cos (wt + phi + Theta) + n (t)$

where

R(t), is the received noisy and delayed signal. Theta is the phase of the carrier due to the combined effect of the channel and otherwise.

n(t) is the noise introduced by the channel and the various components of the transmitter or the receiver.

The parameter theta is modeled as arandom variable between +/-pi. This random variable is guided by a uniform probability density function 1/2pi.
The phase lock loop is designed to track this phase (which actually translates into delay) so the receiver can compensate and hence retrieve the data (phi).

Types of Delays

The time delay can be found when the vehicle is straight and level and a rapid input is applied, while the frequency delay is when the vehicle is straight and level but a continuous sinewave input in assumed. The delay can be divided in three parts. First is the delay of the computer and the graphics, second is the delay due to the dynamics (mechanical, real delay) and finally the delay due to the visual computer with its software and hardware.

Simulation Setup

For delay analysis purposes, Figure 2 shows the simulation setup with the computer software/hardware, visual computer and graphics. The simulation computer's job can be further subdivided into computations and input/output operation. Also the simulation was designed to find the disturbances, the linearing and the nonlinearity elements (i.e filters, saturation and delay). The phase lock loop has been

controlling all those parts and locking the incoming signal, where after some time both the incoming signal and the replica will track each other and will be on line. When that happen it is the time of interest (total time delay) due to the previous parts.

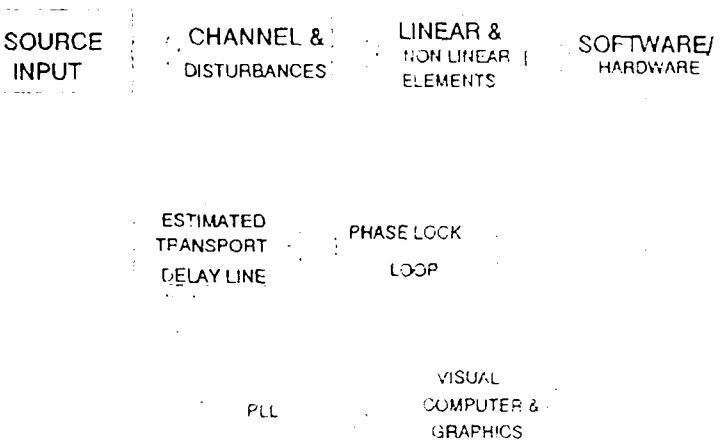


Figure 2 shows the block diagram illustrating the overall system with aphase lock loop to track the the transport delay.

Conclusion

This method of phase lock loop has been studied and analyzed. The actual results and measurements of the delay hasn't been done due to the short time of research in a few weeks. In the previous method, the total delay in the time domain was found to be in the range of 297ms-396 ms. The frequency domain was found to be less and in the range of

177 ms. If the delay in the phase lock loop will be less than those delays, then compensation techniques for the delay will be researched and applied to minimize the total transport delay.

Finally, the work of this methods, the search for the result has been stopped, and will be continued in the near future to find the calculated delay. If it is of interest, then the measured delay will be researched and worked out. Different method of compensation techniques will be applied toward different parts or the whole simulation in the problem.

References:

1. Principle of communication - systems, modulation, and noise. Third edition by R.E. Ziemer and W.H. Tranter

2. The Flight Simulator Time Delay Problem
by Don R. Gum & Edward Martin
AIAA Simulation Technology Conference
Session 23 - System
System Delay- Problem and effect
17-19 August-1987

An Intelligent Tutoring System Architecture for Schematic Knowledge

August 12, 1993

L. Karl Branting
Assistant Professor
Department of Computer Science
University of Wyoming
P.O. Box 3682
Laramie, WY 82071

Final Report for:
AFC :mmer Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Brooks Air Force Base
San Antonio, Texas

August 1993

An Intelligent Tutoring System Architecture for Schematic Knowledge

L. Karl Branting

Assistant Professor

Department of Computer Science

University of Wyoming

Abstract

This paper describes a general tutorial architecture for schema-based expertise. Development of tutorial architectures applicable to classes of cognitive tasks is an important step in reducing the development costs of ITS's. Schema-retrieval is a category of cognitive task described in the psychological literature and identified in protocol analyses of problem solving in such varied domains as landlord/tenant law, grasshopper infestation control, and fire fighting.

This paper describes a three-step model of schema-based problem solving derived from protocol analyses. A general architecture for tutoring schema-retrieval skills follows directly from this model. The expert model consists of the elements necessary for the schema-retrieval process, the student model is a subset of the expert model, and a curriculum is an ordered partition of the expert model together with ancillary presentation information.

The general architecture was implemented in a prototype domain-independent tutorial shell called TASK. A tutor for the domain of grasshopper infestation control was partially implemented in TASK. While time limitations precluded completion of the tutor, the initial implementation indicates the feasibility of a reusable architecture for tutoring schema-retrieval skills.

An Intelligent Tutoring System Architecture for Schematic Knowledge

L. Karl Branting

1 Introduction

There is abundant evidence that instruction is most effective when it is tailored to the particular aptitudes and knowledge level of each individual student [RS92]. Unfortunately, it is seldom feasible to provide individual students with separate human tutors. However, intelligent tutoring systems (ITS's) provide a potential mechanism for delivering individualized instruction through a computer.

The ability of an ITS to provide individualized instruction depends on the existence of an explicit student model (*i.e.*,, a representation of student's knowledge and problem-solving ability) together with an explicit model of expertise in the domain. Explicit student and expert models make it possible to select the instructional actions, such as presenting exercises or expository text, that will lead to the greatest improvement in the student's understanding and ability, *i.e.*, cause the greatest reduction in the difference between the student model and the expert model.

However, the complexity of expert and student models for even a relatively simple domain leads to very high development costs for ITS's. ITS's typically require years of development by teams that include psychologists, educators, and computer scientists. Widespread use of ITS's is unlikely unless these development costs can be reduced.

One approach to reducing ITS development costs is to reuse part or all of an ITS for training in several different skills or domains. Amortizing a single development effort over multiple applications can reduce the development cost per application. However, a single ITS architecture is likely to be useful only

for members of a set of closely related applications. For example, an ITS for diagnosis of carburetor disorders might be easily modifiable for use in training for diagnosis of fuel injector problems, but would probably be poorly suited for algebra word problems. ITS reuse therefore depends on identifying classes of related tasks for which a single ITS architecture would be appropriate.

A promising effort at distinguishing classes of related tasks is the *Initial Task Taxonomy and Criterion Tasks to Support Stamp* [HLF93] (hereinafter *Task Taxonomy*). The Task Taxonomy distinguishes a number of classes of perceptual, cognitive, motor, and task-control tasks. If an appropriate ITS architecture could be identified for each of these classes, then the development costs of new ITS's could be significantly reduced. Developing an ITS for a given domain would require only determining the class to which the skill belongs, fetching the ITS architecture appropriate for that class, and instantiating the general architecture with the knowledge specific to that domain. Devising an appropriate ITS architecture for each class of task identified in the Task Taxonomy is therefore an important step in achieving the goal of reducing ITS development costs.

One of the tasks identified in [HLF93] is *schema retrieval*:

> Sometimes the object to be retrieved is not so much a class as an entire schema for dealing with situations. The task performer must learn to utilize appropriate cues for retrieving a schema, and then learn how to execute the schema once it is retrieved. [HLF93] at page 9.

Despite a growing recognition of the ubiquity of schema retrieval as a problem solving method, there have been no published descriptions of ITS's for tasks falling into this category.

The purpose of the project described in this paper is to develop a general ITS architecture for expertise consisting of schema retrieval. This architecture was partially implemented in a program called *TASK*, an acronym for "Tutor for Acquisition of Schematic Knowledge."

The next section describes in detail a model of problem solving by schema retrieval. Section three describes a general ITS architecture that follows straightforwardly from the model of schema-based problem solving. TASK, a

partial implementation of this architecture, and the Grasshopper Infestation Tutor, an application constructed using TASK, are described in section four.

2 Problem Solving by Schema Retrieval

An essential element of intelligent behavior is the ability to use past problem-solving experience to assist in the solution of new problems. One approach to using past experience is to perform induction over past experience to form general rules useful for future problem solving [Die90]. An alternative approach is to use specific past problem-solving episodes to assist with new problems. This process is termed *reasoning by analogy* [Gen83] or *case-based reasoning* [Bar91].

Although there have been various characterizations of analogy and case-based reasoning, most include the following steps:

1. *Indexing and Retrieval.* Given a description of a new problem, determine the case or analogue most relevant to the solution of the new problem.
2. *Elaboration.* Some facts not explicitly stated in the problem description may have to be inferred to assist in the comparison between the new problem and the old.
3. *Adaptation.* The solution to the old problem may have to be adapted to apply to the new situation, in view of the differences that remain after elaboration.
4. *Learning.* Optionally, the new problem together with its solution may be added to the case library and indices to the new case added to the retrieval mechanism.

Most research in analogy and case-based reasoning has assumed that the analogues or cases being retrieved consist of specific past problem solving episodes or exemplars, *e.g.*, [Gen83, Kot88, Sim85, Ham89, BP91, Ash88], although some case-based reasoning systems merge similar cases to form generalized schemata [PBH90, Bra87]. There is empirical evidence that in human problem solving the choice between specific past cases and generalized

schemata depends on the number of cases that the human problem solver has encountered. For example, a protocol analysis of firefighters revealed the following:

> [A]nalogues (cases that retain the episodic nature of the original experience) disappear under the weight of cumulative experiences for experienced personnel. The fireground commanders we interviewed averaged 23 years of service and frequently responded that an incident reminded them of scores of previous incidents rather than any particular one. Individual incidents have blended together in their memory into what may be considered a prototype or schema representation, and these carry important information about typicality that is not represented when people recall a single analogue [KC88].

Among the few published empirical analyses of schema retrieval problem solving is *A Process Specification of Expert Lawyer Reasoning* [O'N87], which sets forth a protocol analysis of problem solving by experts in landlord-tenant law. This analysis showed that experts in landlord-tenant law seldom solve problems from first principles, but instead draw from a "library of eighteen to twenty prototypical stories" based on previous problem-solving experiences. The lawyer uses the prototypical story that seems most relevant to the current case for

> gathering information to develop theories about the case, gauge what typically fits within expectations, notice anomalies, and generate mentally simulated predictions about how the new case may be concluded. *id.* at page 57.

A more detailed analysis of schema-based problem solving was performed in the context of a project at the University of Wyoming to develop an expert system for agricultural pest management. The objective of this project is to model the process whereby entomologists advise ranchers on the most economical response to grasshopper infestations.

A protocol analysis of this process revealed that entomologists use a schema-based approach to part of this process: predicting the likely proportion of

forage consumed. The entomologists appear to have a relatively small ($\approx$ 10) collection of prototypical grasshopper infestation cases. For example, one prototypical case concerns a cool, rainy spring. Typically, a cool, rainy spring leads to increased growth of bacterial and protozoan pathogens that cause diseases among newly-hatched grasshoppers. This typical situation is neither a general rule nor a specific case, but is instead a prototypical case or schema.

When presented with the facts of a new case (*e.g.*, when contacted by a rancher in need of advise), entomologists appear to perform the following steps:

- Infer any relevant case features not explicitly provided, *e.g.*, species, developmental stage, and density of grasshoppers.
- Determine the schema that best matches the relevant features of the given case.
- Apply the schema to the new case, adapting the schema if there are any relevant differences between the schema and the new case.

Consider, for example, the following example (a simplification of an actual problem-solving episode). Suppose that the entomologist is presented with the following facts about ranch A: The date is June 5, the weather has been hot and dry, the productivity of the ranch (*i.e.*, the amount of forage produced per acre) is moderate, and the rancher says that with every footstep he sees approximately 10 yellow or green grasshoppers hop. The entomologist begins by inferring several important features of the case not provided by the rancher:

- The yellow or green color of the grasshoppers suggests that they are slant-faced grasshoppers (Gomphocerinae), a species that eats both forbs and grass and therefore may cause significant forage loss.
- The number of grasshoppers that jump with each footstep is typically about 10% of the number of grasshoppers per square foot. There are therefore about 60 grasshoppers per square foot, an extreme density.

These facts remind the entomologist of the stereotypical infestation likely to cause significant forage loss:

Feature	Current Case	Serious Infestation Schema
Date:	June 5	June 1
Species:	slant-face	slant-face or spur-throat
Precipitation:	low	normal/low
Temperature:	high	normal/hot
Range value:	moderate	moderate-high
G. Density:	extreme	high
Forage loss	unknown	high

The forage loss in the prototypical case is high. If the facts of the current case were identical to those of the prototypical case, one would predict that high consumption in the current case as well. However, the current case differs relevantly in that the grasshopper density is extreme, rather than just high. Higher grasshopper densities tend to lead, *ceteris paribus*, to higher forage loss. Thus, the loss in the current case is likely to be extreme, rather than merely high.

The three stages of the entomologist's problem solving were therefore

1. Inferring missing case features, *i.e.*, the species and density of the grasshoppers. Note that this is a form of case elaboration.

2. Retrieving a relevant schema, *i.e.*, the prototypical infestation likely to cause significant forage loss.

3. Adapting the schema to fit the current case, *i.e.*, concluding that a grasshopper density incrementally higher than in the prototypical case will lead to forage loss incrementally higher than in the prototypical case.

A retrieved schema contains more than a set of case features and a solution, *e.g.*, a prediction about forage loss. In addition, a schema contains stereotypical explanatory information. For example, the prototypical cool, rainy spring schema contains the following stereotypical explanation:

> Rainy spring → high pathogens → high attrition among young grasshoppers → decreased grasshopper density → decreased forage loss.

This explanatory information is useful both for justifying the entomologist's conclusions to a possibly skeptical rancher and for matching and adapting schemata to a new case. For example, if a given case involved cool wet weather which began after most of the grasshoppers had reached maturity, an entomologist could conclude that the match to the prototypical rainy spring schema was poor, because the explanation underlying that schema doesn't apply if the wet weather occurs after the grasshoppers mature.

In summary, there is empirical evidence for schema-based reasoning in domains in which experts are exposed to large numbers of similar cases. Protocol analyses have been performed in the domains of landlord/tenant law and grasshopper infestation management. These protocol analyses indicate that schema-retrieval problem solving is a three-step process:

- Infer any relevant case features not explicitly given.
- Determine the schema that best matches the relevant features of the given case.
- Apply the schema to the new case, adapting the schema if there are any relevant differences between the schema and the new case.

3 An Architecture for Tutoring Schema Retrieval

An architecture for intelligent tutoring of a given skill requires a model of expertise in that skill, a student model that can represent the degree to which a student has acquired the expert model, and a curriculum consisting of an ordered presentation of the content of the expert model.

3.1 Modeling Schema-Based Expertise

The previous section outlined the process of schema-retrieval problem solving observed in protocol analyses. The knowledge required to perform this process consists of the following elements:

- A language for expressing cases, *e.g.* a set of case features.

- A collection of schemata.
- A set of rules for
  - inferring relevant case features, and
  - schema adaptation.
- The three step procedure for schema-based problem solving.

Given an adequate case-description language, schema library, and rules, a system applying the procedure for schema-based problem solving should be capable of expert performance.

3.2 Student Models

Student models are customarily based directly upon the expert model, typically consisting of the expert mod· plus a collection of differences [Van88]. These differences are of two types: (1) missing elements, and (2) additional, incorrect elements. *id.* Student models that include only differences of the first kind, termed *overlay models*, are equivalent to subsets of the expert model. Differences of the second kind are termed *bugs*, by analogy erroneous statements in computer programming languages. Constructing a library of possible bugs is typically a very difficult process. Techniques for bug library construction include searches of educational literature, protocol analyses of student problem solving, and predictions based on a learning theory in the particular domain [Van88].

An overlay student model in a domain in which expertise consists of schema retrieval would consist of a subset of the elements of the expert model listed above. Typically, each element in an overlay model is annotated by the degree of belief that the student has actually acquired that element [ABCL90]. This degree of belief is adjusted based on the student's performance on exercises posed by the tutor. Given an overlay student model of schema-retrieval expertise, curricular actions to be executed by an ITS can be chosen to minimize the differences between the student model and the expert model, *i.e.*, to identify and supply the elements of the expert model that the student has not yet acquired.

The simplicity of the procedural component of the expert and student models means that *model tracing* [ABCL90], monitoring the performance of the student in the process of solving problems, is straightforward. The procedure of schema-retrieval under these models is a three-step process. The only nondeterminacy is in the order in which feature inference rules are applied in the first step and in the order in which adaptation rules are applied in the third step. Because nondeterminacy in the first step has no effect on the choices available in the third step, there is no combinatorial growth in solution paths. Thus, unlike the case of Anderson's tutors [ABCL90], in which there is a proliferation of possible solution paths, verifying whether a student is on a valid problem-solving path is very inexpensive.

3.3 Curriculum Structure

Given the expert and student models described above, a *curriculum* consists of an ordered partition of the expert model, *i.e.*, an ordered sequence of the case features, schemata, and rules constituting the expert model. In addition to one or more elements constituting expertise in the domain, each element of this sequence, or *lesson*, must include any ancillary information necessary for the student to understand the expert model elements. This ancillary information can include the following:

- A block of expository text explaining the elements in the lesson.
- A collection of examples presented as
  - solved problems, or
  - exercises.

For example, in the domain of grasshopper infestation management, the expert model would consist of:

- A case description language including such features as date, species, grasshopper density, and precipitation.
- A collection of infestation schemata, such as the prototypical infestation leading to significant losses and the prototypical rainy spring schemata.

- A set of rules for inferring relevant case features, such as the rule that grasshopper density is six times the number of grasshoppers that jump at a single footstep, and adaptation rules, such as that greater grasshopper density leads, *ceteris paribus*, to greater losses.

A curriculum for this domain would consist of an ordered partition of these elements, together with expository text and examples for each lesson in the partition. For example, one lesson might concern the effect of grasshopper species on forage loss. This lesson might consist of the following elements:

- A prototypical case of infestation by bandwing grasshoppers, which die before most forage growth occurs and therefore do not cause significant forage loss.
- Rules for inferring whether grasshoppers are bandwing, or are instead slant-faced or spur-throat grasshoppers that often cause significant forage loss.
- A block of expository text setting forth these elements in well-structured English:

  > The species of grasshopper found on a rangeland is important for determining the likelihood of significant loss of forage and possible treatment options. There are three important types of grasshoppers in Wyoming: bandwing (Oedipodinae), spur-throat (Melanoplinae), and slant-faced (Gomphocerinae) ...

- A worked-out example problem illustrating how the rules and schema, together with previously presented schemata, can be used in problem solving.
- Several exercises requiring use of the rules and schemata.

In summary, the proposed architecture for tutoring schema-retrieval skills consists of the following elements:

- An expert model containing

  - A language for expressing cases, *e.g.* a set of case features.
  - A collection of schemata.
  - A set of rules for
    * inferring relevant case features, and
    * schema adaptation.
  - The three step procedure for schema-based problem solving.

- An overlay student model consisting of a subset of the expert model, annotated with the strength of belief that each element of the expert model has been acquired by the student. Optionally, the student model may include buggy rules or schemata, or a buggy procedure for schema-based reasoning. However, since such bugs are domain-specific, they are unspecified by the architecture.

- A curriculum consisting of a sequence of lessons, each of which contains one or more case features, schemata, or rules with ancillary information such as

  - A block of expository text explaining the elements in the lesson.
  - A collection of examples presented as
    * solved problems, or
    * exercises.

3.4 Strengths and Weaknesses of the Architecture

The most important advantage of the proposed architecture is that it is at the same level of generality as the schema-retrieval category of cognitive task specified in the Task Taxonomy. The architecture specifies the structure of the expert and student models and of a curriculum independent of the particulars of the domain. No aspect of the architecture itself is dependent on the choice of domains. Thus, the architecture represents exactly those aspects of schema-retrieval tutoring that should be reusable across skills or domains within this cognitive task category.

The price of this generality is that many aspects of a tutor for a particular domain are unspecified within this architecture. Additional elements that must be specified in any tutor implemented within this architecture include the following:

- **Curriculum.** The architecture specifies the nature of a curriculum as a partition of the expert model together with ancillary information. However, the construction of a particular curriculum requires extensive domain-specific pedagogical knowledge.
- **Remediation.** Much of the intelligence of an ITS stems from its ability to diagnose and remediate student errors. The architecture stipulates that such errors consist of missing elements of the expert model, but does not specify how such errors are to be identified or remediated.

4 TASK: A Partial Implementation of the Architecture

TASK (Tutoring for Acquisition of Schematic Knowledge) is a partial implementation of the architecture for schema-based tutoring constructed during my visit to Armstrong Laboratory. There were two primary purposes for constructing a prototype implementation of the architecture. The first was to develop a set of data structures and algorithms and an interface adequate to support the architecture. The second was to demonstrate the feasibility of implementing a tutor for a particular domain using the architecture.

In keeping with the overall goal of reusability, the design of TASK maintains a clean separation between the domain-independent architecture and the domain-specific content of the particular tutor. An important requirement for this separation is a declarative representation of all domain specific knowledge. In order to achieve this declarative representation, TASK was written in LISP using CLOS, the common LISP object system. To make the system executable on the PC's available in Armstrong Laboratories testing labs, the particular

dialect of LISP selected was Allegro CL/PC. Allegro CL/PC runs under Microsoft Windows, and permits windows operations using CLOS objects and windows.

4.1 TASK Data Structures

Any implementation of the architecture for tutoring schematic knowledge must provide data structures to represent the central elements described above: a case-description language, expert and student models, and a curriculum. In TASK, a case description is a CLOS object having as slots the possible case features. A schema consists of a case description that includes a solution (*e.g.*, the expected total forage loss) and any associated explanatory information (*e.g.*, Rainy spring → high pathogens → high attrition among young grasshoppers → decreased grasshopper density → decreased forage loss). Similarly, a rule is CLOS object having slots for antecedents, consequent, and name. The expert model consists of a set of schemata and rules together with the procedure for applying them to new cases.

A curriculum is a list of lessons, each of which is an object having as slots a list of expert model elements (*i.e.*, schemata or rules), expository text, solved examples, exercises, and the strategy necessary to solve the exercises. The strategy is a subset of the three step procedure of schema-based reasoning; although the procedure is simple, it may be pedagogically desirable to present it in stages. In addition, the lesson has some bookkeeping information, such as the name of the lesson and a short description of its purpose.

A student model is an object having slots for the students name and other identifying information, a pointer to the current lesson in the curriculum to be presented next to the student, and a *presentation record*. The presentation record sets forth each element of the expert model that has been presented to the student and records the strength of belief that this element has been acquired.

A critical TASK data structure is the *explanation*. An explanation represents the reasoning steps whereby a domain conclusion is justified in terms of the elements of the expert model. There are three types of explanations in TASK: observations, rule explanations, and match explanations. An observa-

tion is an object representing that a given conclusion has been observed to be true. A rule explanation is an object representing that a conclusion follows from a given rule because the antecedents of the rule all have explanations. Rule explanations are used to represent the feature inference and adaptation steps of schema-based problem solving.

The most important form of explanation is the match explanation, which represents that a conclusion applies in a given case because the case matches a schema. A match explanation records not only the conclusion, case facts, and schema, but also the relevant differences between the case and the schema together with the adaptation steps necessary to apply the schema to the case.

The solved problems and exercises in a lesson both consist of match explanations. For solved problems, TASK's presentation module presents the facts of the case to the student and replays each step of the schema-retrieval process recorded in the match explanation. For exercises, TASK presents the facts of the case and then elicits from the student each step in the explanation.

A key design property of match explanations in TASK is that the data structure used by the presentation module is identical to the data structure produced by the expert module. This is important for several reasons. The first is that it promotes reusability by specifying a single data structure common to the expert module in any domain and the presentation module. Second, using identical data structures makes it possible for the system to use the expert module to answer students' hypothetical variations on the facts presented in an exercise. Although this feature has not yet been implemented in TASK, the use of consistent data structures makes doing so feasible. Finally, constructing a curriculum is easier if the curriculum builder can use the direct output of the expert module for solved problems and exercises, rather than having to represent the full complexity of the explanations manually.

4.2 The Grasshopper Infestation Tutor

The domain-specific content of the initial implementation of TASK consisted of a set of lessons in the domain of grasshopper infestation treatment. The source of this domain knowledge was the protocol analysis of entomologist problem solving described above.

A curriculum for a portion of the expert model identified through the protocol analysis was constructed. This curriculum consisted of three lessons, comprising three schemata, three rules, and the three-step schema-retrieval strategy. Each lesson contained several paragraphs of text, and the last two lessons contained both solved problems and exercises.

Unfortunately, there was insufficient time during the course of the visit to Armstrong Laboratory to complete the Grasshopper Infestation Tutor. However, the initial portion of the implementation indicates that it is feasible to construct a complete tutor using the general architecture as implemented in TASK.

5 Conclusions and Future Work

This paper has described a general tutorial architecture for schema-based expertise. Development of tutorial architectures applicable to classes of cognitive tasks is important for reducing ITS development costs by amortizing development efforts over multiple applications. Schema-retrieval is a category of cognitive task identified in the Task Taxonomy document and identified in protocol analyses of problem solving in such various domains as landlord/tenant law, grasshopper infestation control, and fire fighting.

This paper has described a three-step model of schema-based problem solving derived from protocol analyses. A general architecture for tutoring schema-retrieval skills follows derived directly from this model. The expert model consists of the elements necessary for the three step process, the student model is a subset of the expert model, and a curriculum is an ordered partition of the expert model together with ancillary presentation information.

The general architecture was implemented in a prototype domain-independent tutorial shell called TASK. A tutor for the domain of grasshopper infestation control was partially implemented in TASK. While time limitations precluded completion of the tutor, the initial implementation indicates the feasibility of a reusable architecture for tutoring schema-retrieval skills.

Several important pieces of future work remain. The most immediate is the completion of the TASK shell and construction of a complete tutor in

TASK. In the long run, it is important to obtain empirical validation for the schema-retrieval model of problem solving described in this paper. Open questions concerning the modeling include:

- The degree of variation in the process of schema retrieval across domains.
- The extent to which the content of schemata consists of procedural knowledge (*e.g.*, the steps to take when a certain situation arises) or declarative knowledge (*e.g.*, the explanation for the probable consequences of a given set of facts).
- Empirical data on acquisition of schema-retrieval ability.

Answering these oper questions will require testing hypotheses about the acquisition and use of schematic knowledge. The general architecture described here and the TASK implementation of that architecture shows promise as an environment in which to empirically explore these hypotheses.

References

[ABCL90] J. Anderson, C. Boyle, A. Corbett, and M. Lewis. Cognitive modeling and intelligent tutoring. *Artificial Intelligence Journal*, 42(1), February 1990.

[Ash88] K. D. Ashley. *Modelling Legal Argument: Reasoning with Cases and Hypotheticals.* PhD thesis, The University of Massachusetts, 1988.

[Bar91] R. Bareiss, editor. *Proceedings of the Third DARPA Case-Based Reasoning Workshop.* Morgan Kaufmann, May 1991.

[BP91] L. K. Branting and B. W. Porter. Rules and precedents as complementary warrants. In *Proceedings of Ninth National Conference on Artificial Intelligence*, Anaheim, July 14–19 1991. AAAI Press/MIT Press.

[Bra87] G. Bradshaw. Learning to recognize speech sounds: The nexus project. In *Proceedings of the Fourth International Workshop on Machine Learning*, pages 1–11, 1987.

[Die90] T. G. Dietterich. Machine learning. *Annual Review of Computer Science*, (4), 1990.

[Gen83] D. Gentner. Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170, April-June 1983.

[Ham89] K. J. Hammond. *Case-Based Planning: Viewing Planing as a Memory Task.* Academic Press, Inc., San Diego, California, 1989.

[HLF93] E. Hunt, A. Lesgold, and D. Fisk. Initial task taxonomy and criterion tasks to support stamp. Armstrong Laboratory Technical Report, 1993.

[KC88] Gary A. Klein and Roberta Calderwood. How do people use analogues to make decisions? In *Proceedings of the DARPA Workshop on Case-based Reasoning*, pages 209–218, Clearwater, Florida, May 1988. Morgan Kaufmann.

[Kot88] P. Koton. *Using Experience in Learning and Problem Solving.* PhD thesis, Massachusetts Institute of Technology, 1988. Department of Electrical Engineering and Computer Science.

[O'N87] Peter O'Neil. A process specification of expert lawyer reasoning. In *Proceedings of the First International Conference on Artificial Intelligence and Law*, pages 52–59, Boston, Massachusetts, May 27–29 1987.

[PBH90] B. W. Porter, E. R. Bareiss, and R. C. Holte. Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence Journal*, 45(1–2), 1990.

[RS92] W. Regian and V. Shute. Automated instruction as an approach to individualization. In W. Regian and V. Shute, editors, *Cog-*

nitive Approaches to Automated Instruction. Lawrence Erlbaum Associates, Hillsdale, N.J., 1992.

[Sim85] R. L. Simpson. *A Computer Model of Case-based Reasoning in Problem Solving: An Investigation in the Domain of Dispute Mediation.* PhD thesis, Georgia Institute of Technology, 1985.

[Van88] K. VanLehn. Student modeling. In M. Polson and J. Richardson, editors, *Intelligent Tutoring Systems.* Lawrence Erlbaum Associates, Hillsdale, N.J., 1988.

INTEGRATED COMBAT TRAINING

Gerald P. Chubb

Assistant Professor, Department of Aviation

The Ohio State University

164 W. 19th Avenue

Columbus, Ohio 43210

FINAL REPORT FOR SUMMER FACULTY RESEARCH PROGRAM

Armstrong Laboratory (AFMC)

Aircrew Training Resources Division

Williams AFB, AZ

Sponsored by

Air Force Office of Scientific Research

Bolling AFB

Washington, DC

September 1993

INTEGRATED COMBAT TRAINING

Gerald P. Chubb, Assistant Professor
Department of Aviation, The Ohio State University

Abstract

This report examines the needs and opportunities for using simulation in the service of integrated combat training. Two separate but related needs are foreseen: 1) continuing combat unit training in a decreasing resource environment, and 2) possible surge training to meet presently unforseen contingencies.

Clearly, simulation relates to simulators: devices used for training, but it does not stop there. Training simulation, properly conducted can support other important needs: 1) modeling of crew activities, 2) identification of operational capabilities and limitations in new uses of old systems, 3) validation of future weapons systems operational requirements, 4) evaluation of future designs, either for modernization of existing systems or acquisition of new systems, and 5) support of preparatory analyses for instructional systems development.

No single training environment is by itself ideal or self-sufficient for assuring combat readiness. The most effective training regimen combines a variety of experiences, each of which provides its own unique contribution to the overall training objective. To make this mix of training experience truly effective, the individual must be able to integrate that experience in a coherent manner and relate it to wartime combat requirements.

The utility of using constructive models for training applications was examined. AASPEM allows user-specified Pilot Decision Logic. TAC BRAWLER (TB) models pilots' situational awareness and value-driven decision making, using a production system approach. TB seems especially well-suited to showing the impact cognitive skill training can have on system performance and mission outcomes. TB could: 1) demonstrate training goals, 2) illustrate what factors are important and to what degree, 3) reinforce the importance of learning task priorities, and 4) show the potential results of disciplined practiced.

To better tailor such production systems to training applications, is recommended that: 1) the mental model's logic structure be documented in terms of a concept map, 2) the task management strategy be explicitly captured, 3) areas for incorporating skill progression-regression into the model be identified, and 4) workload considerations be expanded.

INTEGRATED COMBAT TRAINING

Gerald P. Chubb

1.0 Introduction

Poorly trained military personnel are not only at risk themselves but they put others who must depend upon them at risk as well. Appropriate training is clearly necessary for combat success. Training combat qualified personnel has a long tradition in the military. Changing any part of that tradition has serious implications, and many changes can be anticipated in years ahead.

Current declines in defense related funding impose considerable pressure to reduce operating costs, which in turn forces reductions in training budgets as well. There is a need to be concerned about the quality of the residual war fighting capability as the force structure itself is downsized, but there is also a need to be concerned about how that force structure can be increased should the need arise to do so, and do so rapidly.

Poltorak (1992) examines prior history to provide a basis for estimating some of the possible problems one can expect in the future. He emphasizes the need for collective combat training, especially for joint operations. Integrated combat training requires adding a concern for coordination, communication, and

synchronization of effort until all participants act as a single strike force, not a collection of independent elements.

No single device or training experience provides a completely realistic representation of combat. A combination of training experiences can better prepare an individual, but only if those experiences can be appropriately integrated into decision making and operating capabilities of individuals and their combat units, as well as their command authorities and support units: the total combat team.

Integrated Combat Crew Training therefore has at least two dimensions: a) horizontal integration of skills in each individual to achieve proficient performance (automaticity: efficient production of the right response), and b) vertical integration of the combat team, effectively using multiple combat elements in joint force operations.

In an austere budget environment, distributed, seamless networking is recognized as an important asset (Thorpe, 1987; Houck, et al., 1989; Bankes, 1991; and Downes-Martin, 1992). Less emphasis has been given to exploiting constructive simulation models for training. This report therefore focuses on how constructive models for studies and analysis might be used to support training, while also improving the models themselves.

2.0 Statement of the Problem

There are two co-associated problems: 1) meeting the need for continuation training for a decreasing force structure, and 2) being prepared to train new combat ready units in a rapidly expanding force structure, should the need arise. The residual force structure must itself spawn the training cadres needed to meet the surge requirement by training the trainers.

There are two clearly identifiable risks: 1) the residual force structure could become too small to meet immediate combat needs and concurrently support the training cadres needed to increase the force structure, and 2) the time needed to get the training accomplished exceeds the time available to create an effective force structure. The time needed to create increases in the force structure depends on at least two major ingredients: 1) how many personnel are needed, and 2) how much (and what kind of) training is required? Training resources become the bottleneck in this scenario, especially those of limited number or capacity.

In order to effectively support integrated combat training, both dimensions of that training have to be addressed: 1) getting individuals ready to perform in units, and 2) getting multiple units to act as an integrated combat team. The first step is achieved incrementally as a three-stage process (Fitts, 1964; and

Anderson, 1983): a) declarative, b) associative , and c) autonomous. The second step requires the interaction of all combat units involved in the joint operation.

In this second step of integrated combat training, every unit is presumed to be proficient in the performance of its assigned role. The training problem to be solved is to develop automatized handling of the interactions among the team elements. This can be accomplished by imposing increasingly demanding scenario scripts representative of the real or expected combat environment. Demands are generated either by altering the tempo (rate of events) or adding unexpected events to the scenario. The rate of events can be manipulated several ways: a) varying the number of participants (especially adversaries), b) varying the type or characteristics of the systems (especially adversary systems), and c) varying the dynamic behavior of the adversary systems.

There are some key requirements in order to accomplish the training objective. First, one must be able to quantify performance, so feedback can be given with respect to which responses were or were not appropriate. This is never easy, and it becomes more difficult with multiple participants in the exercise. Second, one must be able to define (and measure or control) what needs to occur in the training scenario in order to meet some training objectives.

Defining what needs to happen in the scenario implies answers to two other questions: a) what are the training objectives and how are they met by manipulating scenario variables, and b) what impact will that manipulation have (can control of the scenario be maintained once variations are introduced)? This suggests that two kinds of unexpected events can and will occur in multi-unit integrated combat training: a) those unexpected events intentionally introduced by the trainer, and b) those unexpected events which result from the reactions of the participants. The latter pose the biggest problem for the trainer, because they are the least easily controlled, and the effects of their introduction might compromise the achievement of the training objective.

To support the first step of integrated combat training, constructive models of pilot performance could demonstrate how individuals influence system and mission outcomes. This can be especially important in learning what to ignore and what cannot be ignored in the stimulus-rich arena of air-to-air engagements. Sufficiently detailed models of situation awareness and decision making could show how the weighting of information influences not only actions, but their consequences: combat outcomes. Constructive models could show not only how to succeed (formulating an appropriate goal-image) but how to survive in combat (a compelling motive). Coupling constructive models to the training enterprise therefore appears to be worth pursuing.

3.0 Methodology

A DTIC Bibliography search identified documents associated with SIMNET (e.g., Rogers, et al., 1990; Orlanski and Thorpe, 1992), and enhanced training (Angler, Alluisi, and Horowitz, 1992).

ASC(XRE) at Wright-Patterson AFB was contacted to make arrangements to access and review the AASPEM and TAC BRAWLER documentation. These documents were reviewed and the training implications of AASPEM and TAC BRAWLER were assessed in terms of the instructional model proposed by Kibler, et al. (1974). Recommendations were made for tailoring TAC BRAWLER's generic pilot model for air-to-air engagement training applications.

Behavioral science literature on training, skill acquisition, and cognitive psychology was reviewed in areas believed to be directly related to integrated training and pilot performance modeling. This review was highly selective, focusing on production system models of problem solving and related cognitive activities.

Finally, visits were made to the Simulator for Air-to-Air Combat (SAAC) at Luke Air Force Base (to better appreciate the simulator's capabilities and limitations) and to DSA, Inc. the developers of TAC BRAWLER (to discuss its architecture further).

4.0 Results

Results include assessments of the pilot decision logic of AASPEM and the mental model of pilot decision making incorporated into TAC BRAWLER. A review of SOAR's capabilities was made to determine whether it would replace TAC BRAWLER as a production system model of the pilot.

4.1 The Nature of Air-to-Air Combat and Training Simulators

The nature of aerial combat is described in Shaw (1985). The final stages of fighter pilot training occur in the SAAC. SAAC includes a number of important features for realistic training: 1) it cues the pilot when the g-load is approaching tolerance limits, 2) it provides buffeting vibrations indicative of aircraft handling qualities, and 3) it provides other cues related to maneuvering that augment one's sense that the aircraft is responding to commanded inputs.

It is also clear that the SAAC has very definite limitations in what kinds of visual cues can be provided to pilots, particularly those associated with attitude when the aircraft is observed at longer ranges. Networking will not solve that problem,

but has improved SAAC training capabilities in other areas. Networking and distributed simulation is not a panacea.

The SAAC limitations mean that a model of combat decision making behavior may not perfectly match behavior measured in a flight simulator. It will be important to document and consider the known differences in the stimulus ensembles, the operating environments, and the nature of the actions which can be taken by the pilot(s). Otherwise, the model may be judged to be a poor predictor when the simulator is the problem, not the model. Since the simulator is difficult to modify quickly or cheaply, the model of pilot performance may need to be "tuned" to the behaviors which can and will occur in the simulator if performance comparisons are to be made.

4.2 Air-to-Air System Performance Evaluation Model (AASPEM)

AASPEM was clearly designed to be a flexible tool for comparing future weapons systems to current ones. That places a considerable burden on the user who prepares the model inputs to represent a particular scenario. That extends especially to the preparation of inputs describing the pilots' decision logic (PDL).

The AASPEM portrayal of tactical posture (and maneuvering priorities as a consequence of posture) are well-suited to modeling, not training. The tracking angle plots used to describe what influences maneuver priorities are not easily interpreted, although they condense considerable useful information about the relative position of combatants and nicely categorize regions of interest. This representation appears difficult to use under stress but is convenient for analysis purposes.

AASPEM includes a consideration of information acquisition and denial as well as simply positioning the aircraft for an attack. The model leaves to the user the specification of a priority structure that captures the relative importance of these three goals, which may sometimes be mutually exclusive.

Large portions of the PDL are opaque to a reviewer of the Analyst's Manual (Anonymous, 198?). Subsequent review of the four Programmers Manuals (Porter and Perry, 1990) revealed that full understanding will require access to the source code itself.

The model lets the user enter a number of assumed pilot capabilities. For example, certain reaction times are allowed, but the format for entering these data have operational, not psychological significance. There is a disconnect between the

semantics of terms being used by the modelers and those commonly found in the literature on human behavior.

The tactics of individual aircraft is separately specifiable. Opportunities for cooperative combat tactics are represented by selecting appropriate tactics. However, the communication and coordination required to effect such operations does not appear to be modeled. There is no real representation of the pilot's mental model of the situation: it is presumed.

4.3 TAC BRAWLER

TAC BRAWLER (Bent, et al., 1991) appears less well-suited to advanced weaponry evaluations, but much better suited to contemporary tactics evaluations than AASPEM. It includes a very explicit and detailed representation of the pilot's mental model and the communication processes which occur, not only between players, but between man and machine.

A three-tiered hierarchy of decisions is captured, higher levels constraining the lower, much as AASPEM does. However, the boundaries among these layers are different, and the partitioning of the missions, phases, and other tactical terminology does not exactly match, again suggesting the importance of semantics in

interpreting these models. The lowest level of decision making captures basic flight maneuvering and weapons release decisions. The middle layer is the one which is most directive, and decisions made at this level typically have a corresponding (nested) link to a particular class of decisions at the next, more detailed level.

Visual workload considerations drive scanning displays and sectors of airspace external to the aircraft. Decisions (including where to look) are value-driven, based upon the mental model which represents the pilot's situational awareness. The variables in that mental model are updated by the production system representing cognitive information processing activities. These activities include not only the acquisition of information, but making estimates of particular derived values (like ranges, speeds, etc.) as well as making judgments (like "surrogate probabilities:" subjective estimates of attack and survival probabilities) which influence the decisions being made.

Attempts to represent pilot proficiency (ACE, PILOT, and ROOKIE) center around the number of entities which the pilot can track in the mental model, as well as some timing characteristics. Hand movements are not captured, and "switchology delays" are incorporated only for one kind of radar mode change. Certain estimates internal to the mental modeling are based on Kalman

filtering theory rather than on psychological theory about information processing and integration.

Further DSA discussions confirmed that: 1) workload considerations are treated only for visual tasks, 2) task management is essentially a linear, sequential procedure, and 3) resource utilization considerations are not incorporated into the mental model, but could be (they had a specific approach in mind).

4.4 The SOAR Architecture of Cognitive Processing

Laird, Rosenbloom, and Newell (1989) describes the SOAR architecture. Under funding from the Office of Naval Research, Rosenbloom (1992) is applying this approach to the design of intelligent adversaries in distributed simulations. Under DARPA funding, others are reportedly using SOAR to specifically examine one-on-one engagements using missiles.

Everything SOAR does is within the context of a well-defined problem space, consisting of a set of states and productions (condition-action pairs) which change those states. These are taken as "given," and completely characterize the task being performed. Subgoaling is accomplished when attempts to solve a

problem are blocked. Chunking occurs in reorganizing information to make its retrieval and use more efficient.

The identification of states is based on the task characteristics: what is to be done or achieved. The productions relate to actions which will achieve (or contribute to achieving) the task goal. The application of SOAR typically ignores: 1) sensation and perception (beginning with internally coded concepts), and 2) motor implementation (ending with what might be regarded as response selection). Interaction with the environment has been implied in the modeling, while cognitive processing of perceived information is explicitly treated in minute detail. Since SOAR does not typically treat the external world explicitly, some kind of representation of aircraft dynamics and the relative geometry related to the posturing and maneuvering has to be added to or interfaced with SOAR for air combat modeling applications.

While verbal protocols may be useful for identifying possible productions at some particular level of abstraction, the subgoaling and chunking which occur in SOAR to partition (getting more detailed) or to abstract (to reduce complexity) make the model "a unique individual" but do not assure correspondence with any particular person (real or stereotypical). How these models can be compared to pilot performance (and to TAC BRAWLER) will have to be determined once the models are available for review.

5.0 Conclusions

Integrated combat training will require a combination of training experiences ranging from table-top exercises to joint operation exercises. No one simulation or training environment will or can meet every training need.

Constructive modeling can be a useful adjunct to real-time, simulations by providing a theoretical basis for explaining empirical results. As a training aid, constructive modeling could: 1) show students what is desired, 2) illustrate what is necessary to achieve the desired result, 3) demonstrate how decisions and actions might be affected by failure to consider all relevant factors, 4) predict what will happen if improper emphasis is placed on the relevant factors, 5) suggest how often certain activities need to be done, and 6) answer "What if ... ?" types of questions students want to raise.

Simulation data collection, reduction, and analysis aids are needed to assist training and modeling communities use the information obtainable from these simulations. The TAC BRAWLER model may provide help on two levels: 1) given some students' expressed value structure (action preferences and task priorities), predict the implications in terms of observable behaviors, and 2)

for a given observed behavior, provide a tool for determining at least a plausible explanation.

Since the model makes assumptions about human estimates and judgments which influence decisions, the accuracy of those assumptions can be empirically determined for an individual subject and used as model inputs to show how individual differences can affect decisions, actions, and outcomes. Moreover, the model can be used to determine the impact estimation and judgment reliability and accuracy have on system performance and mission outcomes.

The immediate need is to: 1) document the generic structure of the TAC BRAWLER pilots' mental model, 2) identify the nature of its task management and workload assessment features, 3) determine where and how skill acquisition features might be added, and 4) outline the research which would be needed to apply such modeling to commercial and general aviation skill training.

A Research Initiation Proposal (RIP) is planned to explore generic, value-driven decision modeling in association with studies of learning to land: deciding when to flare and land and when to "go-around." This will include the development of a data collection system for measuring aircraft performance as well as developing the concept maps of the decision process, based on or in-lieu of additional TAC BRAWLER documentation analyses.

REFERENCES

Anderson, John R. (1983), "6 Procedural Learning," in The Architecture of Cognition, Harvard University Press, Cambridge, MA, 215-260.

Angler, Bruce N., Earl A. Alluisi, and Stanley A. Horowitz (1992), Simulators and Enhanced Training, IDA Paper P-2672, Institute for Defense Analysis, Alexandria, VA.

Anonymous (1982), Boeing Improvements to the Piloted Air Combat Analysis Model, Analyst's Manual, D-180-28789-1, Boeing Military Airplane Company, Seattle, WA.

Bankes, Steven C. (1991), Issues in Developing the Potential of Distributed Warfare Simulation, R-4131-DARPA, Rand Corporation, Santa Monica, CA.

Bent, N. E. et al. (1991), The TAC BRAWLER Air Combat Simulation Analyst Manual (Rev. 6.1), Decision-Science Applications, Inc., Arlington, VA.

Downes-Martin, Stephen (1992), Seamless Simulation Literature Survey, IDA Document D-1109, Institute for Defense Analysis, Alexandria, VA.

Fitts, Paul M. (1964), "Perceptual-Motor Skill Learning," in Arthur W. Melton, editor, Categories of Human Learning, Academic Press, New York, 243-285.

Houck, Michael R., G. S. Thomas, and Herbert H. Bell (1989), "Training Potential of Multiplayer Air Combat Simulation," Proceedings of the 33rd Annual Human Factors Society Meeting, Santa Monica, CA, 1300-1304.

Kibler, Robert J., Donald J. Cegala, Larry L. Barker, and David T. Miles (1974), Objectives for Instruction and Evaluation, Allyn and Bacon, Inc. Boston, MA.

Laird, John, Paul Rosenbloom, and Allen Newell (1986), Universal Subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies, Kluwer Academic Publishers, Boston, MA.

Orlansky, Jesse and Jack Thorpe (1992), 73 Easting: Lessons Learned from Desert Storm via Advanced Distributed Simulation Technology,

IDA Document D-1110, Institute for Defense Analysis, Alexandria, VA.

Poltorak, Gerald J. (1992), Effective Joint Training Assessment, Masters Thesis, ATZL-GOP-SE, Ft. Leavenworth, KS.

Porter, Richard F. and Doug D. Perry (1990), Programmer's Manual for the Advanced Air-to-Air System Performance Evaluation Model (AASPEM), Eglin Version 3.3, Vol. I, Vol. IIA, Vol. IIB, and Vol. III, Battelle Memorial Institute, Columbus, OH.

Rogers, Brian K., Clarence W. Stephens, and Alan B. Oatman (1991), "12th Interservice / Industry Training Systems Conference 1990: SIMNET Fighter Aircraft Application," ITECH Proceedings, also identified as AL-TR-1992-0020, Aircrew Training Research Division, Human Resources Directorate, Williams AFB, AZ.

Rosenbloom, Paul S. (1992), Towards Intelligent Automated Forces for SIMNET, Semi-Annual Report, Office of Naval Research, Washington, DC. AD A252 267.

Thorpe, Jack A. (1987), "The New Technology of Large Scale Simulator Networking: Implications for Mastering the Art of Warfighting," Proceedings, 9th ITECH, 492-501.

ENTERPRISE INTEGRATION IN IICE WITH ONTOLOGY CAPTURE

Asesh Das
Department Statistics and Computer Science
West Virginia University
Morgantown, WV 26506

Final Report for:

Summer Faculty Research Program
Armstrong Laboratory : Human Resources Directorate
Logistics Research Division

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

July 1993

ENTERPRISE INTEGRATION IN IICE WITH ONTOLOGY CAPTURE

Asesh Das
Department Statistics and Computer Science
West Virginia University
Morgantown, WV 26506

Abstract

The task of improving the performance of large complex processes by managing the interactions among participants leads to integration of enterprises. Within the context of Information Integration in Concurrent Engineering (IICE) practices, ontology sharing by different enterprises is a means of achieving integration. To share ontology is to capture it while in transition from As-Is to To-Be. This can be done by using a functional form for ontology given by the expression <Form, Structure, Intention> ---> Implication at t1.

ENTERPRISE INTEGRATION IN IICE WITH ONTOLOGY CAPTURE

Asesh Das

Introduction

IICE, i.e., Information Integration in Concurrent Engineering is a multiyear research and development method sponsored by Armstrong Laboratory (AL) logistic Research Division (HRG) at the Wright-Patterson Air Force Base, Dayton, Ohio. Concurrent Engineering has been defined as a systematic approach to the integrated, concurrent design of products and their related processes, including manufacture and support. This approach is intended to cause the developers, from the outset, to consider all elements of the product life cycle from conception through disposal, including quality, cost, schedule and user requirements.

IICE effort has been divided into eight distinct thrust areas [1]. These are,

(1) Integrated Systems Theory thrust,
(2) Ontology Thrust,
(3) Methods Engineering thrust,
(4) Experimental Tools thrust,
(5) Three-Schema architecture thrust,
(6) Application thrust,
(7) Frameworks thrust,
(8) Technology Transfer/Transition thrust.

To study all these thrusts and to formulate a unified discipline for dealing them, ontology- driven methods are very important [2]. Such methods directly leads to the integration of enterprises, which has been defined as the task of improving the performance of large complex processes by managing the interactions among the participants [3]. It is clear that enterprise integration is basically a concurrent engineering activity. This research aims at that within the IICE environment.

What is ontology?

Ontology is the task of acquiring the terminology, statement structure, and sanctions inferences of a given engineering, manufacturing, business, or logistical domain and storing it in an usable representational medium [4]. It can also be stated as the discipline giving a computational model of a system's general view of the world, and that includes symbolic representations of the most general concepts (objects, relations, events, attributes) and serves as a detailed index for the attachment of the various domain models [5].

The Nature of Enterprise Integration

The integration of enterprises is primarily based up on distributed computing environment. Ideally in this environment different computers exist and communicate with different types of software packages. The environment also maintains an overall strategy in making one-person or group decisions making. The decisions may be on a product developed in collaborative efforts, or on multiproject planning/control, scheduling, cost-estimating, performance measurement, progress reporting, resource allocations, etc. Enterprise integration also needs well-developed and uniformly accepted modeling methodology with standardized metrics for performance study, methodologies on process study and domain model elucidation is also needed.

How Ontology Capture Helps in Enterprise Integration?

Within the IICE environment, to capture ontology the following relation is proposed:

$$\{\langle \text{Form, Structure, Intention} \rangle \dashrightarrow \text{Implication at t1}\}. \qquad \text{(Eq.1)}$$

Here it is intended that Form, Structure and Intention are abstractions of a class of ontologies, and they ultimately converge to a desired implication at a particular time. Intention works as constraints on ontology derived from 'Form' and 'structure'. Instead of giving a schema (a weak ontology) at a particular time,

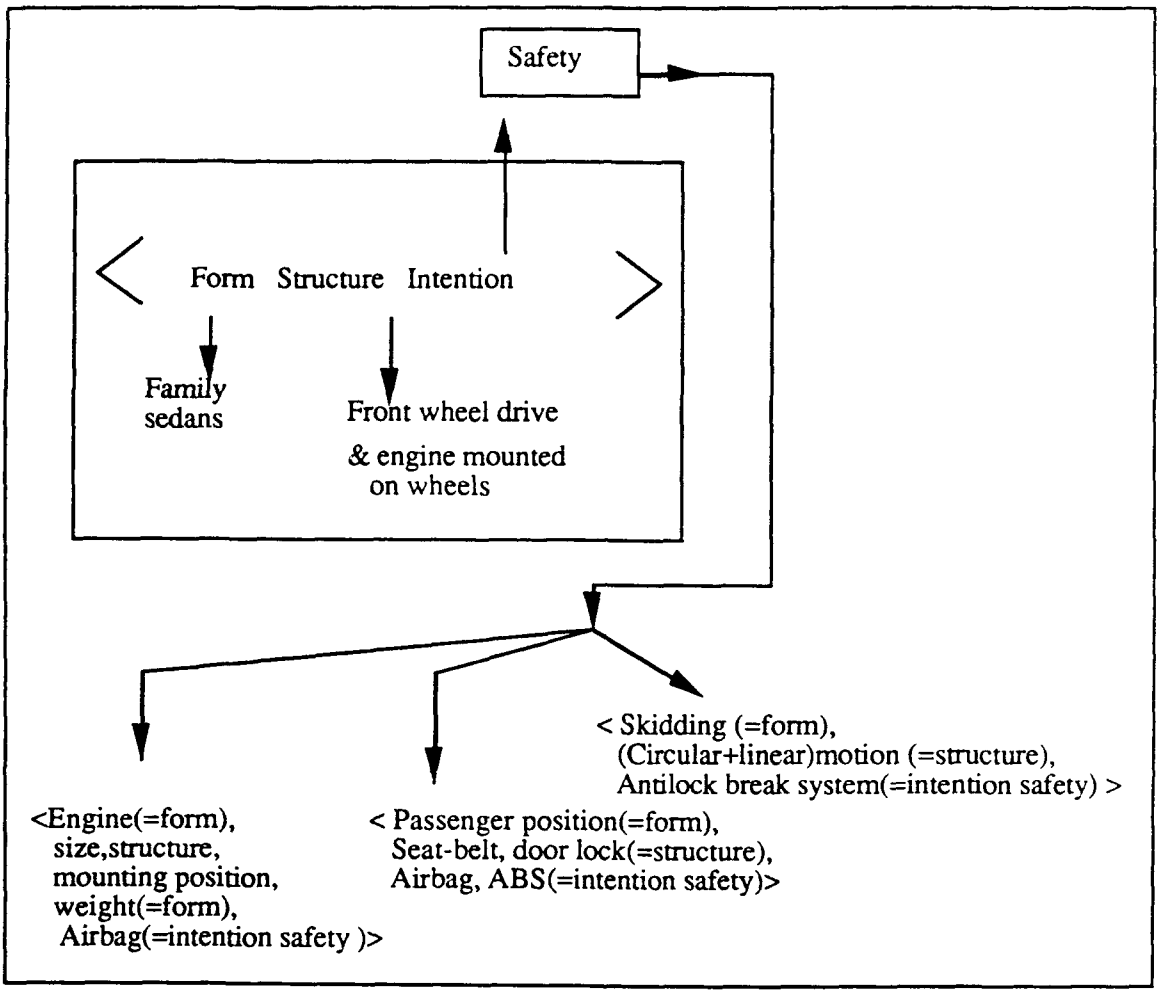
$$\dashrightarrow \text{Implication at a particular time t1}$$

gives a computerized domain model engineered over a band of strong and week ontologies. Also, depending on situations and conditions, a weak ontology at a time t1 may become a strong ontology at another time t2.

Say now, that we have an enterprise A with a problem on safety associated with a car driver at the time of front-end collision. It is proposed by enterprise A in this fashion:

> Find a technique or a new design for cars to prevent drivers from steering-impact related injury on the chest.

Assume that several other enterprises have become interested in this problem, and wish to integrate their interest (ontologically) with enterprise A. Consider that enterprise B has come close to the solution by having shared ontologies with enterprise A. Enterprise B presents itself to enterprise A in a fashion described in Figure 1 below.



Ontology capture with constraints

Figure 1

Here form ontology (family sedans) shares structure ontology with front wheel drive engine structure. The intention ontology, when shared with form and structure sets up constraints. The intention is to have safety features. The intention is broken down into ontological forms as shown in the Figure 1. At this point enterprise B *could not* integrate nicely with enterprise A, because its ontologies are not strong enough to satisfy enterprise A's demand that it needs safety features saving chest from steering-impact. Even though antilock break system is there with airbag system, chest injury in severe front-end collision is not guaranteed. Enterprise integration demands stronger ontology to be shared, where mechanisms for engine mounting and crumple zone creation needs to be revisited, computerized sensors needs to be activated, and in front-end collisions such arrangements

should be able to pull the engine down taking the steering with the engine. The driver will be protected from chest injury by the steering. If another enterprise C can come out with this plan that satisfies enterprise A, then with ontology sharing C integrates to A. The next phase is to invoke concurrent engineering practices to have the actual design implemented.

Enterprises and ontology

An integration of enterprises within the context of IICE environment is largely a task of effective and intelligent communication between different components of enterprises. This can be achieved by maintaining three disciplines [6]:

1. Communication between inter-disciplinary team : the *Low Road.*
2. Quality function deployment, statistical process and quality management/control,: the *Middle Road.*
3. Cooperative product development by networked teams, knowledge integration: the *High Road.*

On the issue of ontology capture four fundamental *engineering needs* are in order [7]. These are:

(i) Information access : Ontology on corporate networks should provide all types of information *efficiently*, requiring less time, money and other resources. The accessibility tools should be more responsible to changes in demand and opportunities. Sharing of ontologies should provide better quality and should be able to *customize* the product for small groups of customers, perhaps at the individual level.

(ii) Monitoring and automation : Ontology share permits notifying people and software of decisions that affect them. This permits effective action and future planning.

(iii) Cooperative work : In space and time people need to work in a world of shared knowledge and information.

(iv) System Integration : Different software packages must be plug-compatible to the enterprise allowing seamless interoperability.

For the IICE environment a general framework behind above experienced needs have been sketched under an overall philosophy of integrated system theory thrust [2]. It suggests common data sharing, work flow support, change propagation (modification of one artifact propagated to related artifacts) support and flexibility with respect to change.

Whatever may be the proposed systematics for enterprise integration, it is obvious that a strong ontology-driven mechanisms of integration *prescribes a mediator with good*

knowledge engineering support. From this angle an *enterprise* in a function form as given in Eq.(2) below.

Enterprise = f (S { <Form, Structure, Intention>
---> implications at different times}), (Eq.2)

where f stands for "function of". S stands for summation over ontology forms at different times of ontology capture. The simplest form of the above relation is

Enterprise = function of strong ontology . (Eq.3)

For a simple illustration in Eq.(3) these prescriptions may be taken:

Form : Business modeling process modeled with generic ontology, business domain-specific ontology and business models.

Structure : Model engineering. It can be knowledge-based simulation models (Das et al,). It can be structured analysis and design techniques (SADT) approach to information systems.

Intention : Functional matching of participating (shared) ontology as done in knowledge-acquisition process in domain-expert interactions, or as done in DES/STEP frame work. The intention is to get ideas on changes, extensions, recommendations, etc. for ontology capture.

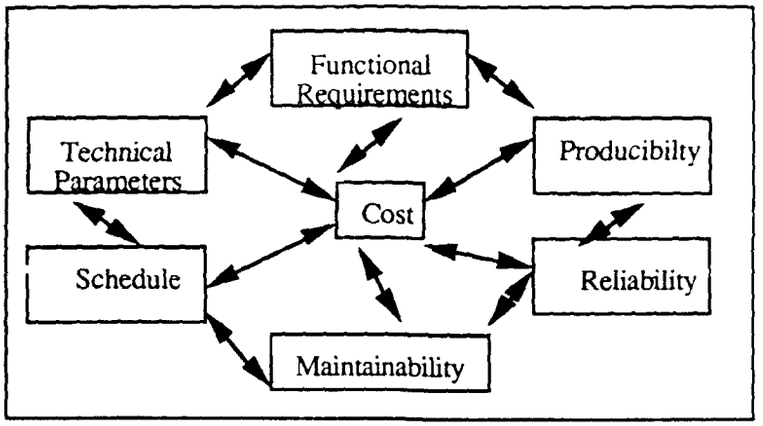
Implication

at t1: Enterprise pictured as process model with shared ontology.

Then Eq.() becomes in short,

Enterprise = function-of
(business matching process, model engineering,
functional matching of shared ontology)
---> shared ontology engineering at t1. (Eq.4)

The functional matching of shared ontology focuses some trade-off to be implied at time t1 (time when the enterprise has been interrogated). This trade-off can be represented in the Figure 2 shown below. ◄► stands for sharing.



Ontology sharing

Figure 2

The Figure (3) represents the enterprise (Eq.4) in diagram form.

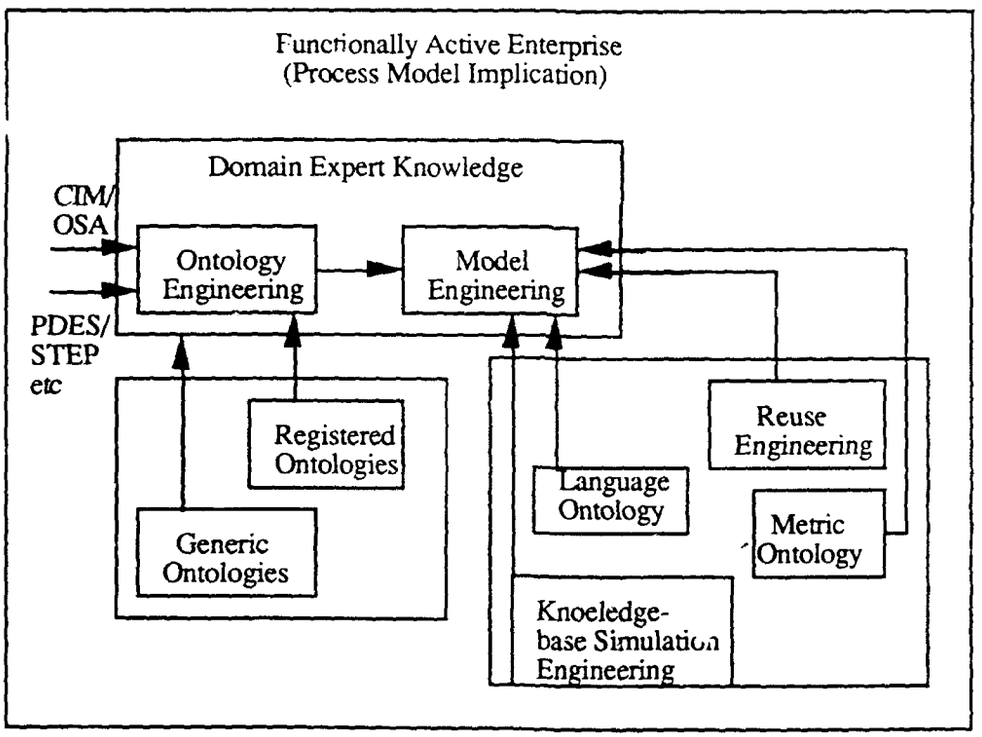


Figure 3 : *Ontology engineering with the enterprise*
(adopted from working group 3 ICEIMT workshop I, see Ref.3)

The enterprise model with shared ontology at a time t1 will have different appearance at time 2, if different ontology is captured and shared. This leads to the study

of SHADE : Shared Dependency Engineering or PACT :Palo Alto Collaborative Testbed [8, 9]. Here a distributed framework for enterprise integration is administered in which explicitly represented knowledge serves as a medium for communication among people and software. Each agent informs a shared knowledge base about its internals and capabilities, and about the existence of information it is willing to share. The shared information ranging from formal CAD models to unstructured drawings, is encapsulated as shared, persistent objects. Agents declare the existence of such objects and their attributes and relationships using predicate calculus. Automated reasoning techniques use this metaknowledge to answer queries, route information. The framework is rooted in two key ideas that appear to have broad applicability for enterprise integration : (1) a minimal shared representation, and (2) a federation architecture. The representation is minimal because most of the knowledge resides in the internal representations of participating agents; only a subset of the knowledge is explicitly exported as shared objects, and only selected attributes and relationships are made explicit. The federation architecture enables any agent to post information queries and assertions, and automatically route them to appropriate recipients. It creates the illusion of a centrally shared knowledge base, in a heterogeneous multi-agent environment.

For a mediator-based enterprise integration system with current technology these following points from IDSE are noteworthy [10]:

(1) Many people will be involved in the process of having enterprises integrated with the mediator and a substantial part of them will not be computer scientists.

(2) Vast amounts of documents will appear to a vast number of users as poorly documented.

(3) System maintenance for the mediator will appear as very costly.

(4) The mediator will be in a partially-automated environment. There will be "islands of automated assistance", where data once entered/created would have be reentered/created for use in later phase.

(5) The lessons learned and component created in one project are seldom carried over to others.

(6) The existing project management tools do not address the problems of task assignment and accountability with any degree of completeness.

(7) The system produced often do not satisfy the customer requirements.

(8) No one person can understand all the details of a system because the size of the code in even one now involves a very large amounts of line.

This is what is needed [10]:

(1) A means to capture, represent, present, manipulate, and integrate information about the system definition and design, as well as the design of process used to produce it and the modification history of resulting artifact.
(2) An integrated set of tools to be used by the individuals involved in the development process. The integration supported by these tools should enable a process of evolvable, tailorable, and universally automated tool integration. In the environment, integrating and accessing automated tools should be accomplished without extensive work. These tools must address all phases of the system life cycle.
(3) A means for controlled sharing and tracking of design information. This will require design databases that allow the linking of the design information and their interlinking to requirements information.
(4) A means for tracking, monitoring and controlling design dependencies and change, as well as propagating their effects. The data storage facilities must provide for the linking of the system needs and requirements to designs and design decisions.

Ontology Engineering

The integration of enterprises with ontology share needs thorough knowledge engineering activities with well-specified guide lines. Within the IICE environment all whatever narrated before can be organized into the following discipline format:

Name

Enterprise Integration

Header

Ontology capture

Constraints

Form .+. Structure ---> Intention
Intention ---> Implication
Implication at t1
Implication at t2
..
..
Implication at tn.
Find the most effective
set of constraints.
Validate them on test models
indexing the ontology

Precondition

World-state <---> History
---> Objects
---> Relation Functions
---> Domains

Effects

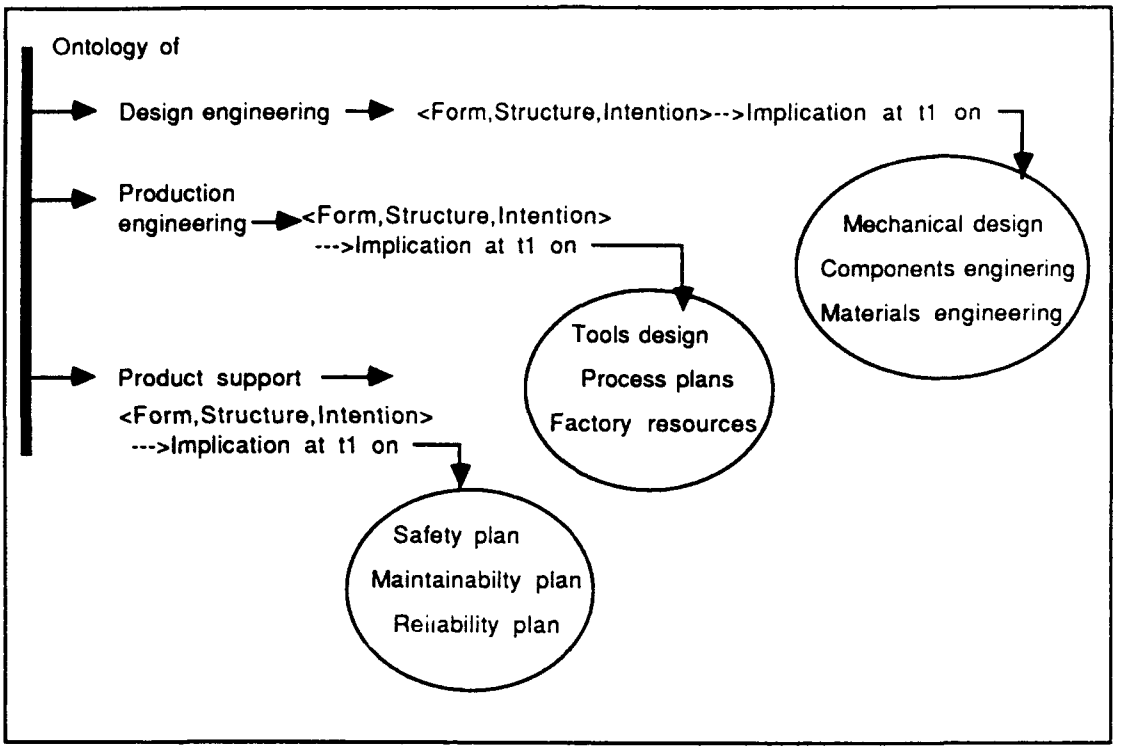
Know how

Decomposition

Indicate shallow level ontology capture with
<Form, Structure> ---> a-kind-of.
Execute <Form, Structure, Intention>.
Identify correlation between states with
shared ontology.
Classify them.
Evaluate them.
Integrate them.
Where
Integration rationale explains why
<Form,Structure,Intention>->Intention at t1
is always valid.

One major performance of the ontology engineering would to capture the current process in As-Is form. Such processes would index all other processes with the ontology expression <Form, Structure,Intention>--->Implication at t1. There after, ontologically engineered process changes would be subsumed to To-Be scenario [13]. The As-Is to To-Be transition could be effective with integrated enterprises, ontologically integrated. Figure 4 shows how To-Be product development depends on ontology capture.

In this figure it is shown what factors could contribute to ontology capture for the To-Be stage. By no means it is complete, immediately contributing ontology will come from quality engineering (Tools/process certifications, Statistical process control, etc), and procurement engineering (materials, purchasing, subcontracts, etc.), test engineering (equipments, recruitments, softwares, etc). The basic premise on which ontology capture using Eq.(1) would depend is clear : it is how much concurrent engineering practices would be administered by it.



Ontology capture for the To-Be

Figure 5

Conclusion

What lessons are learned from ontology captures for enterprise integration? In IICE it is mainly accepting a *design rationale* among all the participating enterprises, those need to be integrated. What does design rationale do to ontology engineers has been discussed in the AAAI '92 workshop on design rationale (July 15, 1992, Palo Alto, California). To summarize, to the ontology engineers a shared design rationale among enterprises will provide explanation to answers on why a product is as it is produced from As-Is to To-Be.

Acknowledgment:

All the members at the Armstrong Laboratory (AL) Human Resources Directorate Logistic Research Division (AL/HRG), Wright-Patterson Air Force Base, Dayton, Ohio, specially, Capt. JoAnn Sartor, provided immense help in completing this summer project. It is gratefully acknowledged. I am also thankful to Lt. Brian Strawn for his patience and cooperation in all form while the work was in progress. I am also thankful to RDL, California, for financial assistance via the of summer faculty research program.

References

[1]. Painter, M. (1991). Information Integration for Concurrent Engineering. Armstrong Laboratory (AL) Human Resources Directorate Logistic Research Division (AL/HRG), Wright-Patterson Air Force Base, Dayton, Ohio. 1-21

[2] Mayer, R. J., et al. [1992]. Information Integration for Concurrent Engineering (IICE): Philosophy, Concepts, and Theoretical Foundation Report. AL/HRGA CDRL Sequence No. AO16.

[3] Petrie,Jr.,C. J. [1992]. (Ed): Enterprise Integration Modeling. Proc. of the First International Conference. Cambridge, (Mass): The MIT Press.

[4] Mayer, R. J., Menzel, C. P., Painter, M. and Benjamin, P. (1993). The Role of Ontology in Enterprise Integration. *Proc. of the IDEF Users Group Conference.* 1993. 7p.

[5] Nirenburg, S. et al. [1992]. *Machine Translation: A knowledge Based Approach.* San Mateo, (CA): Morgan Kaufmann Pub., Inc.

[6] Cleetus, K. J. and Reddy, Y. V. [1992]. Concurrent Engineering Transactions. In the Proc. of CALS/CE Symposium. Washington, (DC): 17-25.

[7] Pan, J. Y. C. and Tenenbaum, J. M. [1991]. An Intelligent Framework for Enterprise Integration. IEEE Transaction on Systems, Man, Cybernatics. Vol. 21. 1391-1408.

[8] Cutkosky, M.R. et al. [1993]. PACT: AN Experiment in Integrating Concurrent Engineering Systems. Computer. Vol. 26. 28-37.

[9] Tenenbaum, J. M., Weber, J. C. and Gruber, T. [1992]. Enterprise Integration: Lessons from SHADE and PACT. In Petrie,Jr., C. J. (Ed): Enterprise Integration Modeling. Proc. of the First International Conference. Cambridge, (Mass): The MIT Press.

[10] Mayer, R. J., Wells, M. S. and Painter, M. K. [1992]. Integrated Development Support Environment (IDSE). Tech. Report No. AL-TP-1992-0052. AL/HRG, Wright-Patterson Air Force Base, Dayton, Ohio.

Note:

A detailed and rigorous discussion of the subject treated in this report has been done under the title, MEDIATORS OF INTELLIGENT DECISIONS (under the IICE ontology thrust), *and has been submitted for publication as a technical report to the AL/HRG, WPAFB, Dayton, Ohio. The report can be obtained from the division chief Bert Cream.*

USING A NEGOTIATION SUPPORT SYSTEM TO INTEGRATE INTERESTS

Kenneth A. Graetz

Assistant Professor

and

Jason Seifert

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.

September 1993

USING A NEGOTIATION SUPPORT SYSTEM TO INTEGRATE INTERESTS

Kenneth A. Graetz
Assistant Professor

Jason Seifert
Graduate Research Assistant

Department of Psychology
The University of Dayton

Abstract

The current study used a variable-sum negotiation task to determine the degree to which computer-assisted dyads are better than manually assisted and unassisted dyads at achieving integrative bargaining agreements. Male and female dyads engaged in both a four-issue and an eight-issue negotiation during a single experimental session. While computer assistance did not improve performance for females, computer assisted males obtained a significantly higher proportion of the integrative total on the four-issue task than did unassisted and manually assisted males. In addition, while computer assistance did not appear to improve interest estimation, significant positive correlations were obtained between estimation accuracy and the outcome measure for both tasks.

USING A NEGOTIATION SUPPORT SYSTEM TO INTEGRATE INTERESTS

Kenneth A. Graetz and Jason Seifert

"There is no shortage of disputes."

Howard Raiffa, 1982

The first sentence of Raiffa's (1982) volume, *The Art and Science of Negotiation* remains a humorous understatement. Whether discussing vacation plans or negotiating a contract, social conflict and social interaction often appear synonymous. The current study explores a potential technological solution to conflicts involving *scarce resource competition* (Aubert, 1962; Druckman & Zechmeister, 1973), the allocation of a limited pool of desirable resources (e.g., money, commodities, services, information). *Negotiation* is often viewed as one subtype of scarce resource conflict (Thompson and Hastie, 1990). In labor-management disputes, for example, parties attempt to reallocate tangible rewards (e.g., money and working hours) in a mutually satisfactory fashion. In some cases, the interests of both parties are diametrically opposed or *zero-sum*. On these occasions, the optimal joint outcome is the *distributive* or compromise solution (i.e., the negotiators split the resources in half). While this may be an optimal strategy for maximizing joint profit in zero-sum negotiations, it yields a suboptimal solution when applied to negotiations that include an *integrative* potential. An integrative solution is one that affords both negotiators a higher outcome than the compromise solution. Research in the area of integrative bargaining suggests that such solutions, while frequently present, are not as readily attainable as they might appear. Evidence that negotiators arrive at suboptimal, distributive solutions in negotiations that present an opportunity for interest integration has been found in a variety of case and field studies (Raiffa, 1982; Lax & Sebenius, 1986; Howells & Woodfield 1970; Howells & Brosnan, 1972) as well as in controlled laboratory environments (Pruitt & Rubin, 1986; Bazerman & Neale, 1983).

Why do two negotiators fail to integrate their interests when doing so would result in a better outcome for both? During a typical, face-to-face negotiation, participants are confronted by a milieu of stimuli and tasks that demand their attention. Assuming that interest integration requires more time and cognitive effort than simply settling for the compromise solution, negotiators who are under pressure to reach an agreement may be aware that interest integration is possible, but may not have time to achieve it. Individual negotiators may also be under strict instructions from a constituency or supervisor to *win* the negotiation, either by maximizing the difference between their own outcomes and those of their opponent or by simply obtaining a higher outcome than their opponent.

Another possibility is that one or more of the negotiators may fail to realize the existence of the integrative solution. While a variety of factors are presumed to obstruct and /or delay this insight (Foroughi & Jelassi, 1990), the most pervasive seems to be the erroneous perception that one's own interests and the interests of the other negotiator are diametrically opposed. This has been referred to as the *fixed-pie perception* (FPP) and seems to originate from preconceived beliefs about the social act of negotiation. For example, subjects who were asked to play the role of a negotiator tended to bargain more competitively and obtained lower outcomes than subjects who were not asked to play the role (Enzle, Harvey, & Wright, 1992). One interpretation of these findings suggests the existence of a scripted set of role obligations which place a premium on being a *hard* negotiator and which bias individuals toward competitive behavior. The negotiator who falls victim to this bias may simply opt for the distributive outcome, perhaps after selfishly attempting to obtain a higher outcome at her opponent's expense. The thought of working together to obtain a higher joint outcome is not considered.

It is with these cognitive limitations and biases in mind that the Bargaining Analysis and Resolution Builder (BARB) was designed. This software joins other negotiation support systems (NSS) in an effort to make negotiation problems more manageable and comprehensible for negotiators. BARB resembles existing NSS packages

in its commitment to providing simple tools that do not interfere with the social dynamics of the typical face-to-face negotiation. It is unique in that behavioral prescriptions based on normative decision making models are purposely absent. Also, BARB's session-oriented information displays track and organize negotiator *behavior*, rather than subjective beliefs and/or expectations.

BARB is expected to be a full-featured NSS capable of assisting individual negotiators both in and out of the conference room. The tools targeted for evaluation in the current study are *session-oriented*, that is, designed for use during the face-to-face portion of the negotiation. These tools are part of a database management system (DBMS) written using Borland's Paradox for Windows™ and it's ObjectPal™ object-oriented programming language. The general goals of BARB's session-oriented tools are as follows: (a) to provide information that a human negotiator would normally calculate, (b) to provide information that a human negotiator is incapable of calculating, (c) to generate simple information displays based on the behavior of the negotiators, (d) to minimize the amount of interaction between the negotiator and the computer, and (e) to refrain from providing computer-generated advice based on subjective probability estimates. These tools are designed to be placed on the file server of a local area network. Negotiators in the conference room access the tools using notebook computers. Every attempt is made to minimize the impact of the technology on the social processes at work during the face-to-face portion of the negotiation.

While BARB will eventually be used to handle real negotiation situations, thereby requiring the users to input the negotiation issues, the levels of each issue, and their subjective utilities for each level, the current study presented negotiators with two preloaded variable-sum negotiation tasks (Pruitt & Lewis, 1975; Kelley, 1966; Thompson & Hastie, 1990; Thompson, 1991). While negotiating agreements to these tasks, some individuals had access to a computerized *offer screen*. This screen has three general components: (a) a *pushbutton interface* representing the negotiation problem space, (b)

two *point calculators*, and (c) a *vertical bar chart*. The negotiator is able to view her problem space on the screen, including all negotiation issues, all levels per issue, and her outcome (points) for each level. The negotiator can then pick a level for each issue by clicking the left mouse button (LMB) on the pushbutton representing that level. As the negotiator selects levels with the LMB, the total points per issue are displayed as green, vertical bars on the bar graph. The bars can be compared with existing black bars representing the maximum points the negotiator could win or lose for each issue. A third set of yellow bars are displayed when the negotiator uses the right mouse button (RMB) to select an issue level. In this way, negotiators can simultaneously track two potential offers, a *green offer* and a *yellow offer*. In addition to updating the vertical bar graph, clicking the LMB on an issue level changes the value displayed in the green point total box. This is the total number of points that the negotiator would receive if the final agreement included all of the levels selected using the LMB. Similarly, clicking the RMB on an issue level updates a second yellow point total box. In this way, negotiators receive instantaneous and simultaneous graphical and numerical information comparing the outcomes of two potential offers. Negotiators are trained to track their own offers with the LMB and their opponent's offers with the RMB. It is anticipated that this combination of the interactive problem space with the automatic point calculation and display capabilities will alleviate the problems caused by preconceived beliefs and poor strategy choice. Attentional resources will be redirected away from mathematical calculation of outcomes and toward other tasks. The bar graph will provide the user with a quick visual ranking of each issue's relative importance as well as displaying the issues on which the opponent is approaching (or not approaching) a satisfactory distribution of outcomes.

An experimental lab study was designed to evaluate the degree to which these session-oriented tools help negotiators to integrate their interests. It is hypothesized that computer-assisted dyads will differ from manually assisted and unassisted dyads on a number of measures relevant to bargaining ability. Computer-assisted negotiators will

obtain higher joint outcomes and will achieve higher interest estimation accuracy. Furthermore, these main effects will interact with negotiation complexity, the benefits of computer assistance being more pronounced for highly complex negotiations.

Method

Overview

The current study used a variable-sum negotiation task to determine the degree to which computer-assisted dyads are better than unassisted and/or manually assisted dyads at achieving integrative bargaining agreements. Dyads engaged in both a four-issue and an eight-issue negotiation during a single experimental session. Two between-subjects independent variables were task order and level of assistance. Dependent measures were constructed using the number of points obtained by the dyad for each of the two negotiated agreements as well as each negotiators' estimate of her opponent's interests.

Subjects

Sixty males and 60 females served as subjects in this study. All were students and staff from a midwestern liberal-arts college and were recruited using advertisements placed at various campus locations.

Independent variables

There was one major independent variable in the current study: *type of assistance*. This between-subjects variable had three levels: unassisted, manually assisted, and computer-assisted. The order of presentation of the four-issue and the eight-issue tasks was also manipulated, with the four-issue task occurring first for half of the dyads. Finally, an equal number of male and female dyads participated in the study in order to assess potential gender effects.

Procedure

Subjects arrived in same-sex pairs. During the recruitment phase, experimenters ensured that the dyads were unfamiliar with one another. Upon arrival, one of two experimenters escorted each subject to a private desk to minimize pre-experimental

contact and discussion. After obtaining informed consent, an experimenter introduced the first negotiation task, distributed the payoff schedules, and described the method of payment. Participants were awarded $10 for their participation in the study. The chance to earn a $100 *bonus prize* was used to further motivate the participants during the negotiation phase. The experimenter informed each individual that the person who earned the most points would receive the $100 prize. Subjects were told that the winner would be determined upon completion of the *entire study*. They were also told that the probability of a *tie* was quite high and, in the case of a tie, the recipient would be randomly selected from among the top point earners[1].

During the introduction, the experimenter oriented each subject toward her role in the first negotiation. The current study included two negotiations: *program analyst versus systems engineer* and *salesperson versus buyer*. For each negotiation, an experimenter randomly assigned subjects to one of the two roles. The experimenter then read a role description aloud, while the subject reviewed his payoff schedule. Each protocol also included a short statement describing the opposing negotiator. A *briefing sheet* distributed to each negotiator contained information concerning the negotiation issues and her *supervisor's* interests for each issue. These statements were simply phased to reflect the point distributions *within* each negotiation issue. No explicit comparisons were made *across* issues at any point during the introduction.

Next, subjects completed a short, paper-and-pencil exercise designed to measure their understanding of their first payoff schedule. The exercise required that the subject determine the number of points he would obtain given a particular negotiated outcome. Subjects in the computer-assisted condition used a pointing device (mouse) to select levels within each negotiation issue from the offer screen. The computer automatically calculated

---

[1]This protocol was used to avoid implying that the negotiators were to compete with one another during the upcoming negotiations.

the points earned by the negotiator given his selections and displayed this total on the screen. The experimenter provided subjects in the manually assisted condition with a simple electronic calculator to use in completing the exercise. Subjects in the unassisted condition were given note paper and a pencil. All subjects were expected to achieve 100% accuracy on the exercise before continuing with the procedure. This pre-negotiation phase of the study (the introduction, role orientation, and exercise) took approximately 20 minutes to complete.

Subjects in the computer-assisted condition then participated in a private, 10-minute training session with an experimenter. This training session was designed to instruct participants on the use of the computer interface. The experimenter described the information displayed on the computer screen, and emphasized the computational method and the basic meaning of each piece of information.

The experimenter then escorted each subject to a conference room where each sat across the table from her opponent. In the unassisted and manual condition, the computers were replaced with cardboard screens of about the same physical size, designed to shield from view any notes taken during the negotiation. Subjects in the unassisted and manual condition received blank sheets of note paper and were allowed to bring their briefing sheets and the payoff schedules into the conference room. Also, those in the manual condition were allowed to use their calculators during the negotiation.

Next, an experimenter introduced the subjects to one another by role (e.g., "This is the program analyst."). The experimenter advised the subjects as to the 15 minute time limit and the consequences of not reaching an agreement. Subjects were told that, if they failed to reach an agreement within 15 minutes, each negotiator would receive zero points for that negotiation. Actually, the experimenters allowed each dyad an extra 5 minutes for negotiation if needed. During each negotiation, dyads were also given a *two minute warning* signal at the 13 minute mark.

The experimenter informed the subjects that, while each had a copy of his own payoff schedule, they were never to show the schedule to their opponent during the negotiation. If this occurred, the experimenter terminated the negotiation and recorded the last offer as the final agreement. Otherwise, there were no limits placed on between-subject communication. Before leaving the room, an experimenter instructed the dyad to knock on the door of the conference room after reaching a final agreement.

The experimenters observed the negotiation from behind a one-way mirror, timing the duration of the negotiation with a stopwatch. Following the final agreement, the participants were escorted back to their private desks and given an *estimation sheet*. The main purpose of this sheet was to obtain subjects' estimates of their opponents' point distributions across issue levels. The estimation sheet was simply a payoff schedule with the point values missing. The issue levels were arranged in a manner identical to the negotiators own payoff schedule. An experimenter asked the subject to fill out the schedule based on her impressions of her opponent's interests.

The second negotiation proceeded much the same as the first. An experimenter oriented each subject to his new role. Subjects received new payoff schedules and completed a second exercise. The negotiators followed the same rules as were used in the first negotiation. After a final decision was reached, the subjects completed a second estimation sheet. An experimenter then paid and debriefed the subjects.

Payoff schedules.

The variable-sum experimental task used in the current study is similar to that used in previous research (Pruitt & Lewis, 1975; Kelley, 1966; Thompson & Hastie, 1990; Thompson, 1991). Each subject is privately presented with a payoff schedule consisting of a number of negotiation issues and a number of potential levels within each issue. The schedule also includes the point values associated with each issue level. Each negotiator is privy to her own payoff schedule only. At no time prior to reaching a negotiated agreement is either negotiator given access to her opponent's schedule.

The eight-issue payoff schedule presented to the program analyst role-player and illustrated in Figure 1, represents a negotiation between two United States Air Force employees: a *program analyst* and a *systems engineer*.

Figure 1

Payoff Schedule - Program Analyst

Number of Maintenance Persons	Scope of Developmental Testing	"Skin" Material Type	Average Time to "Turn"
2 (4000)	2000 hrs (1600)	Type A (2400)	20 min (0)
4 (3000)	4000 hrs (1200)	Type B (1800)	30 min (-600)
6 (2000)	6000 hrs (800)	Type C (1200)	40 min (-1200)
8 (1000)	8000 hrs (400)	Type D (600)	50 min (-1800)
10 (0)	10,000 hrs (0)	Type E (0)	60 min (-2400)

Foreign Interoperability	Average Missions Between Failures	Time to First Delivery	Maximum Speed
13 countries (3200)	10 missions (0)	1 yr (800)	500 mph (0)
10 countries (2400)	8 missions (-1500)	2 yrs (600)	550 mph (-400)
7 countries (1600)	6 missions (-3000)	3 yrs (400)	600 mph (-800)
4 countries (800)	4 missions (-4500)	4 yrs (200)	650 mph (-1200)
1 country (0)	2 missions (-6000)	5 yrs (0)	700 mph (-1600)

Both are assigned to the same program management team and are given the task of determining the desired characteristics of a new unmanned reconnaissance aircraft. The following issues are open for negotiation: (1) *number of maintenance personnel*, (2) *scope of developmental testing*, (3) *skin material type*, (4) *average time to turn*, (5) *foreign interoperability*, (6) *average time between failures*, (7) *time of first delivery*, and (8) *maximum speed*. Each negotiation issue has five levels. Each level is worth a certain

number of points to the negotiator. Each negotiator's total points are calculated by summing across all eight issues following the dyad's final agreement.

In the eight-issue task, two of the issues have identical point distributions for both the program analyst and the systems engineer (*average time to turn* and *foreign interoperability*). Two of the issues have zero-sum or diametrically opposed distributions (*skin material type* and *maximum speed*). The other four are logrolling issues. While the two negotiators have conflicting interests, each issue is less important to one negotiator than it is to the other (i.e., the within-issue point sum is less for one party than it is for the other). *Number of maintenance personnel* and *average time between failures* are logrolling issues that are more important to the program analyst than to the systems engineer. *Scope of developmental testing* and *time to first delivery* are logrolling issues that are more important to the systems engineer than to the program analyst.

The four-issue payoff schedule presented to the buyer and illustrated in Figure 2, represents a negotiation between a United States Air Force *buyer* and a *salesperson* from a large defense contractor.

Figure 2

Payoff Schedule - Buyer

Unit Price	ASAT Defense	Time to First Delivery	Orbits to Change Altitiude
7 billion (1600)	All three (0)	1 yr (2400)	1 orbit (4000)
8 billion (1200)	"S" & "T" (-600)	2 yrs (1800)	1.5 orbit (3000)
9 billion (800)	"T" only (-1200)	3 yrs (1200)	2 orbits (2000)
10 billion (400)	"S" & "D" (-1800)	4 yrs (600)	2.5 orbits (1000)
11 billion (0)	"S" only (-2400)	5 yrs (0)	3 orbits (0)

The buyer is interested in purchasing a communications satellite. Four issues are open for negotiation: (1) *unit price*, (2) *anti-satellite defenses*, (3) *time of first delivery*,

and (4) *orbits to change altitude.* Again, each negotiator's total points are calculated by summing across all four issues following the dyad's final agreement.

In the four-issue task, one of the issues has an identical point distribution for both the buyer and the salesperson (*anti-satellite defenses*). One of the issues has a zero-sum or diametrically opposed distribution (*time of first delivery*). The other two are logrolling issues. *Orbits to change altitude* is a logrolling issue that is more important to the buyer than to the salesperson. *Unit price* is a logrolling issue that is more important to the salesperson than to the buyer.

Dependent variables.

There were two major dependent variables in the current study, the proportion of the integrative total obtained by the dyad for each negotiation and the accuracy of the dyad's point estimates. The total number of points obtained by the dyad was divided by the integrative total (10,400 for the four-issue negotiation and 15,200 for the eight-issue negotiation). The estimation-related variable consisted of a measure of the degree to which each member of the dyad realized that the opposing member's points were distributed such that one issue of a logrolling pair was less important than the other. This measure was obtained using the point values from the estimation tasks. For each logrolling pair, a difference score was calculated by subtracting the absolute value of the sum of the estimated point values (across levels) of the less important issue from the absolute value of the sum of the estimated point values of the more important issue. This difference was then divided by the actual difference between the two sums as stated on the opposing player's payoff schedule. For the eight-issue task, a total score for each player was obtained by averaging the scores for the two logrolling pairs. For both the four-issue and the eight-issue negotiation, the score for the dyad was the arithmetic mean of the scores of the two players. Using this measure, a score of zero indicates that both members of the dyad mistakenly estimated the logrolling issues to be of the same importance (i.e., worth the same number of points) to the opposing negotiator. A score of +1.0 indicates that both

members of the dyad correctly guessed the preference ranking and the magnitude of the difference. A score of -1.0 indicates that both members of the dyad correctly guessed the magnitude of the difference but mistakenly reversed the preference order. Scores greater than ±1.0 are possible using this measure. Such a score would indicate an overestimation of the magnitude of the difference between issues in a logrolling pair.

Units of analysis.

Because the experimental task involved an interaction between two negotiators, the dyad was used as the unit of analysis. Ten male and 10 female dyads were run in each cell of the current 2 (simple task first versus complex task first) x 3 (unassisted versus manually assisted versus computer assisted) design.

Results

Proportion of integrative total

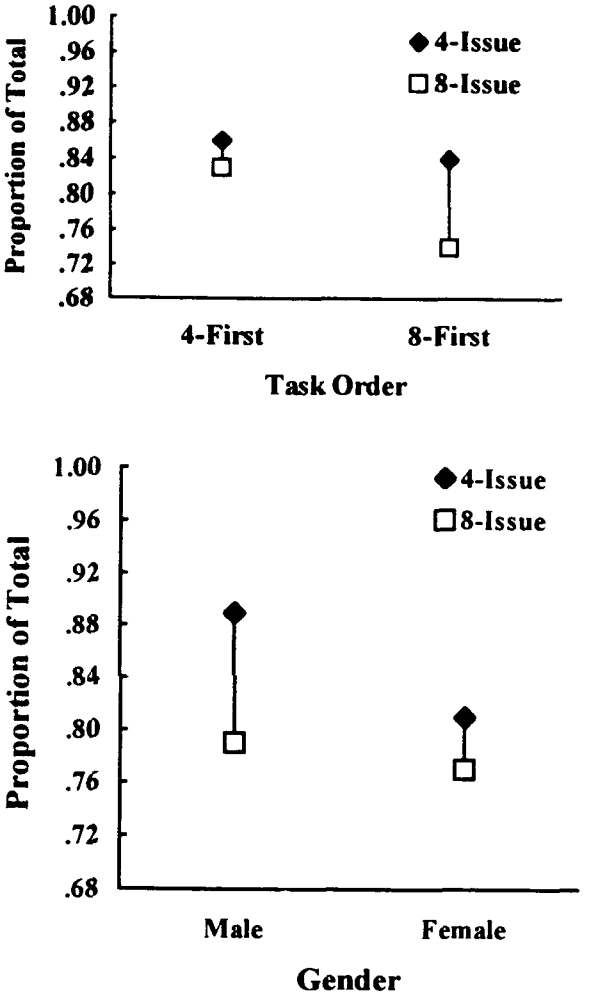
The proportion of the integrative total obtained by dyads in the three different assistance conditions is presented in the first row of Table 1.

Table 1

	Level of Assistance					
	Unassisted		Manual		Computer	
Dependent Variable	Male	Fem.	Male	Fem.	Male	Fem.
Four-Issue Negotiation						
Prop. of Integrative Total	.85	.85	.88	.80	.95	.78
Estimation Accuracy Index	-.39	-.72	-.23	-.59	-.01	-.67
Eight-Issue Negotiation						
Prop. of Integrative Total	.78	.81	.74	.75	.86	.77
Estimation Accuracy Index	-.69	-.42	-.20	-.65	-.20	-.60

A repeated measures analysis of variance (ANOVA), using the proportions obtained in the four-issue and the eight-issue negotiations as two repeated measures and task order, gender, and level of assistance as independent measures, obtained a significant difference overall, $F(1,48) = 29.56$, $p < .001$. Dyads achieved a significantly higher proportion of the integrative total on the four-issue negotiation than on the eight-issue negotiation. This effect was qualified by significant interactions with task order, $F(1,48) = 7.08$, $p < .05$, and with gender, $F(1,48) = 6.08$, $p < .05$. These interactions are illustrated in Figure 3.

Figure 3

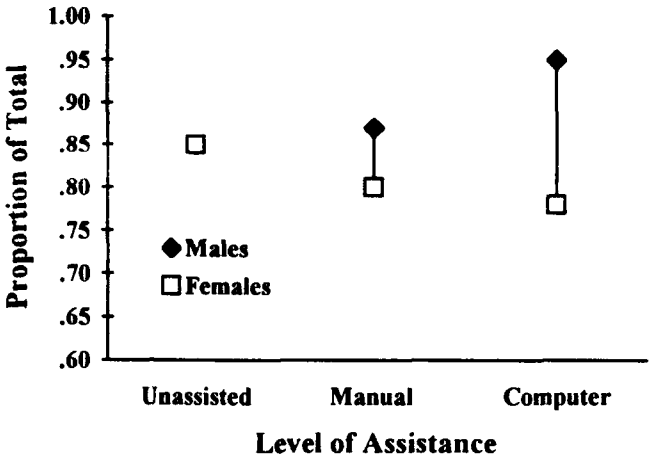


The univariate ANOVA using the proportion of the integrative total obtained in the four-issue task as the dependent measure revealed a significant main effect for gender, $F(1,48) = 11.69$, $p < .01$, qualified by a significant gender by level of assistance interaction, $F(2,48) = 4.37$, $p < .05$. In addition, a significant interaction was obtained between order and level of assistance, $F(2,48) = 4.61$, $p < .05$. The gender by level of assistance interaction is illustrated in Figure 4. While males in the computer condition obtained a significantly greater proportion of the total profit than did males in both the unassisted condition, $t(48) = 2.46$, $p > .01$, and the manual condition, $t(48) = 1.72$, $p <$

.05, female negotiators' performance did not improve with assistance. With respect to task order, negotiators achieved a higher proportion of the integrative total on the four-issue task when this task came first rather than second.

A similar ANOVA was conducted using the proportion of the integrative total obtained in the eight-issue task as the dependent measure. A significant order effect was obtained, $F(1,48) = 15.50$, $p < .001$, with dyads reaching a higher joint profit on the eight-issue task when this task occurred second in the task order rather than first (means = .83 and .74 respectively). A significant main effect for assistance was also obtained, $F(1,48) = 3.19$, $p < .05$. While not significantly different from each other, the proportion of joint profit in both the unassisted (mean = .80) and the computer assisted (mean = .81) conditions was significantly greater than that obtained in the manually assisted condition (mean = .74).

Figure 4



Interest estimation

The estimation accuracy values obtained by dyads in the three different assistance conditions is presented in the second row of Table 1. A repeated measures analysis of variance (ANOVA), using the values obtained in the four-issue and the eight-issue negotiations as repeated measures obtained no significant repeated measures effects. Separate univariate ANOVAs revealed significant main effects for gender on both the four-issue task, $F(1,41) = 6.84$, $p < .05$, and the eight-issue task, $F(1,41) = 5.93$, $p < .05$. On both the four-issue and the eight-issue negotiation, females received lower accuracy scores than males.

Significant Pearson product-moment correlation coefficients were obtained between the estimation accuracy index and the proportion of the integrative total for both the four-issue, $r(54) = .41$, $p < .01$, and the eight-issue negotiation, $r(56) = .32$, $p < .05$.

Discussion

The results of this first study are mixed. When applied to the four-issue task, computer assistance did not appear to improve females' ability to obtain a higher joint profit and, on the eight-issue task, accomplished little for either gender. In addition, although estimation accuracy correlated with interest integration on both tasks, there was little evidence that this brand of computer assistance improved negotiators' ability to estimate the point values in their opponents' payoff schedules. On the other hand, interest integration by male dyads in the four-issue task was clearly enhanced by computer assistance.

References

Aubert, V. (1962). Competition and dissensus: Two types of conflict and conflict resolution. Journal of Conflict Resolution, 7, 26-42.

Bazerman, M.H., & Neale, M.A. (1983). Heuristics in negotiation: Limitations to effective dispute resolution. In M.H. Bazerman & R.J. Lewicki (Eds.) Negotiation in organizations, Beverly Hills: Sage.

Druckman , D., & Zechmeister, K. (1973). Conflict of interest and value dissensus: Propositions on the socilogy of conflict. Human Relations, 26, 449-466.

Enzle, M.E., Harvey, M.D., & Wright, E.F. (1992). Implicit role obligations versus social responsibility in constituency representation. Journal of Personality and Social Psychology, 62, 238-245.

Foroughi, A., & Jelassi, M.T. (1990). NSS solutions to major negotiation stumbling blocks. Proceedings of the 23rd Annual Hawaii International Conference on System Sciences, Vol. IV: Emerging Technologies and Applications Track, Kailua-Kona, Hawaii, January 2-5, pp. 2-11.

Howells, J.M., & Brosnan, P. (1972). The ability to predict workers' preferences: A research exercise. Human Relations, 25, 265-281.

Howells, J.M., & Woodfield, A.E. (1970). The ability of managers and trade union officers to predict workers' preferences. British Journal of Industrial Relations, 18, 237-251.

Kelley, H.H. (1966). A classroom study of the dilemmas in interpersonal negotiation. In K. Archibald (Ed.), Strategic interaction and conflict, Berkely: Institute of International Studies, University of California.

Lax, D.A., & Sebenius, J.K. (1986). The manager as negotiator. New York: Free Press.

Pruitt, D.G., & Lewis, S.A. (1975). Development of integrative solutions in bilateral negotiation. Journal of Personality and Social Psychology, 31, 621-630.

Pruitt, D.G., & Rubin, J.Z. (1986). Social conflict: Escalation, stalemate, and settlement. New York: Random House.

Raiffa, H. (1982). The art and science of negotiation. Cambridge, MA: Harvard Univ. Press.

Thompson, L.H. (1991). Information exchange in negotiation. Journal of Experimental Social Psychology, 27, 161-179.

Thompson, L.H., & Hastie, R. (1990). Social perception in negotiation. Organizational Behavior and Human Decision Processes, 47, 98-123.

LOCALLY DERIVED STUDENT MODELS CAN PREDICT PERFORMANCE IN COMPUTER-BASED DRILLS

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 1993

LOCALLY DERIVED STUDENT MODELS CAN PREDICT PERFORMANCE IN COMPUTER-BASED DRILLS

Thomas E. Hancock
Assistant Professor
Educational Psychology
Grand Canyon University

Abstract

The construction of student processing models was tested. Response certitude judgments, latencies, and response correctness histories were recorded from each of 48 subjects in about 4 hours of interactive computer drills. Results indicate that these easily collected data, along with various instructional manipulations serve as statistically significant predictors of performance, both for the grouped data and on a subject by subject basis. In the future such models could be fit for each trainee and enable more robust administration of instruction in Intelligent Tutoring Systems and other forms of computer managed learning.

LOCALLY DERIVED STUDENT MODELS CAN PREDICT PERFORMANCE IN COMPUTER-BASED DRILLS

Thomas E. Hancock

Introduction

The focus of this line of research is the improvement of student model construction. This will enable instructional manipulations such as post-response information (or instructional feedback) and delay of re-p entation to be matched to the changing cognitive needs of individual earners at each frame of instruction.

The models of concern here are locally determined ones or student processing models, rather than student knowledge models (e.g. Johnson & Norton, 1992). We are attempting to tap the potential of the computer to construct models for each student based on the continuous recording of several measures of a student's processing. A computer drill program records learner metacognitive judgements, response latencies, and post-response information study times, and their variance in the presence of several types of post-response information, various delay presentation rates and between two levels of response correctness. From these variables the intent is to form a best fitting model which is unique to each student, and which could be dynamically updated.

Previous Research Along This Line

It has been previously demonstrated that in addition to response correctness history, the use of the learner's metacognitive judgements of response certitude (postresponse judgments by S about the correctness of responding--"I am 0% certain of my response", "I am 75% certain of my response", etc.) is predictive of postresponse behaviors of instructional information processing and posttest performance (Hancock, Hubbard, & Thurman, 1992a, 1992b; Hancock, Stock, Kulhavy, 1992; ; Kulhavy & Stock,

1989; Kulhavy, Stock, Hancock, Hammrich, Swindell, 1989). For example, in Hancock, Stock, and Kulhavy (1992) it was reported that the conditional probability of subsequent correct responding was related to a cognitive discrepancy index formed from combining certainty ratings with initial response correctness at each trial. These results give us an indication of trends that could be useful on a frame by frame basis. However, if such processing variables are to be useful in computer based training, it would be desirable if the likelihood of the response correctness on the next trial were predicted. And furthermore if such modeling is to be used then it should hold up on a subject by subject basis, not just for the group mean.

As a step in that direction, the data from the Hancock et al (1992) study were reanalyzed (Hancock, Hubbard, & Thurman, 1992c), which work was the precursor for the present study. A logistic regression analysis was used with type of PRI (verification or elaboration), and initial response correctness (right or wrong), and certitude (five levels) as categorical main effects. Certitude and response correctness measures were used from the first two trials for each item. The predicted response was the correctness of the third trial of each item. In addition, the effectiveness of three continuous variables as covariates was investigated. These were response times (initial and current) and post-response information study time. The maximum likelihood Chi-square goodness of fit test on the combined data was not significant indicating that the model adequately fit the data (i.e. the null hypothesis that the model fit the data could not be rejected). All of the terms in the initial model were significant, as indicated by the Wald statistics, beyond the traditional level of significance (alpha = 0.05). Similar analyses were conducted separately for each subject. The model was an adequate fit for 21 of the 26 subjects. The significant predictors varied between the subjects--the most common (for 10

subjects) being the initial test correctness; but every one of the other seven factors was a significant predictor for at least two of the subjects.

Thus the intent of the current study was to determine if similar variables would yield similar results in a different domain. In the 1992 study, multiple choice items on science and history facts were used. While in the current study a domain was chosen for which the students would have no prior knowledge. In general, the purpose was to generate further evidence of the predictive value of our measures, but in particular our interest was whether these variables could be used in adequate model construction on a subject by subject basis. It was reasoned that such results should provide support for the domain independent use of these variables as local measures in student model construction.

Experiment 1

Method

Design

The basic experimental design was a mixed factorial with one between subjects factor--5 levels of PRI (post-response information, formerly called instructional feedback), and three within subjects factors--5 levels of certitude estimate, 2 levels of response correctness, and 4 levels of delay of re-presentation of drill items--with response latency, certitude latency, and PRI frame latency treated as continuous covariates. The primary predicted value of concern was the response correctness on the subsequent trial of each item.

Subjects

Fifty-four university undergraduates participated for partial course credit. The subjects were randomly assigned to the experimental conditions.

Six subjects data were not included in the final analyses: one whose responses appeared to be random; two who completed only one of the

experimental sessions; three whose equipment malfunctioned. The subjects were male and female—with the majority being female; and their ethnic origins were Caucasian, Hispanic, Native American, and Oriental—with the majority being Caucasian.

Equipment

MacIntosh Plus computers were programmed to present the task and collect data. The monitors had a screen resolution of 512x342 monochrome pixels. Timing for stimulus presentations and response latency recording was accurate to within 16 milliseconds. The stimuli were presented via Hypercard, an authoring system which allows graphical as well as auditory display of information and can accommodate subject responses (Apple Computer Inc., 1990).

Stimuli

The stimuli consisted of 27 separate items each presented as a screen of information (See Figure 1). 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 display item by clicking on the appropriate name with the computer's mouse.

All assignment of stimulus items, blocking of items, and order of presentation were randomized separately for each subject. Items were blocked into groups of 4, with one of four ISD (inter-stimulus delay) conditions assigned to each item. Each block was presented once with two other blocks following (with the same ISD assignment for each item but item orders re-randomized); the three blocks were cycled four times during one session (at day one and again at day 3 and 5); another twelve items were displayed,

again for eight trials each (at days 2 and 4). Three items were randomly assigned as ISD position fillers, so that the ISD could remain constant.

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.

The five types of post-response information, PRI, were ordered according to the amount of available information (See Figure 2 for PRI 4): -PRI 1, Basic: included basic response sensitive verification information (Kulhavy & Stock, 1989): "No, (or yes) the correct response was (name)", plus the re-display of the graphic and iconic information.

-PRI 2, Basic + name: included type 1 information and the re-display of the incorrectly chosen name and the correct name, both in their usual spatial location on the screen.

-PRI 3, Basic + verbal elaboration: included type 1 information and elaborative information labeling each of the three stimuli in the correct and incorrect responses.

-PRI 4, Basic + verbal elaboration + name: included type 2 and type 3 information.

-PRI 5: Basic + visual elaboration with name button enabled: included type 2 information and the name buttons were enabled so that the subject had under his/her control the re-display of the the graphic, iconic and also the echoic information.

Procedures

At the beginning of the experiment, subjects were given oral instructions, specifying the general nature of the task and the type of responses required. Subjects were informed that they would be required to participate in one experimental session per day for 5 consecutive days.

Samples of the screens were then displayed and each subject practiced the certainty scale. Finally the subjects were instructed to use headphones during the sessions, and to report any malfunctions to the lab monitor. Subjects proceeded at their own pace with each session being about 40 minutes.

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 PRI screen, and press a "continue";

-4. view the next stimulus item, etc.

In the normal case, there was one experimental session per day (except for 7 subjects) for 5 consecutive days. However, twelve of the subjects continued on for an extra session; their extra data were included in the analyses. Fourteen subjects did not complete the last session. Two subjects did not complete the fourth and fifth sessions. These inconsistencies are of little concern, since the goal is prediction on a trial by trial basis.

There was a debriefing session one week after the last session was completed.

Results

Data compilation was accomplished by downloading each subjects' response and latency records into text files. These files were transferred to a VAX and data were analyzed with logistic analyses using the CATMOD Procedure of SAS (SAS Institute, 1990). The PRI frame latencies were recorded in "tics":1 tic = a 60th of a second. log2 PRI times seemed to fit the data best, based on the residual analyses, so for all analyses log2 times were used.

The data set was sorted by subject, item, and trial for each item. The

trials were tagged as either short-term recall items (a delay of 1,2,4, or 8 items since the previous exposure of that item) or as longer-term recall items (a delay of at least 20 interposing items since the last exposure of that item). Unless otherwise indicated significance of statistical was set at an alpha level of .01.

Preliminary Analyses

We first analyzed prediction of long-term recall: the predicted variable being the 10 celled, certitude by correctness matrix for responses at a trial interval of about 20 or more intervening items (see Agresti, 1990). The predictors were certitude, response correctness, and post-response information type. According to the goodness of fit test [Chi-Square (144) = 149.56, p = .359] the model is a good fit. Each of the main effects and the certitude by correctness interaction were significant. With the addition of post-response information latencies to the model, the Chi-Square (140) became 133.92, and the log2 latency was significant.

We attempted to fit the certitude by correctness matrix at the short-term recall as well by subject at both short and long term recall. There were too many empty cells to complete a meaningful analysis.

Before proceeding to the subject by subject analyses, we fit a model for the prediction of response correctness at the short term delay: Chi-square (11156) = 10326.20, p = 1.000. The significant predictors were response correctness, certitude, the correctness by certitude interaction, delay, and response latency, certitude rating latency, and PRI latency.

Separate Analyses for Each Subject

At this point, separate logistic analyses were conducted for each subject, with the predicted being the response correctness at the next short-term delay trial. The predictors were response correctness, certitude, delay, log2 PRI, certitude rating latency, and also log2 response latency,

with the three two-way interactions for response correctness, certitude and delay. According to the Chi-square goodness of fit test 41 of the 48 subjects had a model that was a good fit, that is $p > .05$; and of these 21 had $p > .50$. The number of subjects with each of the significant main effects were as follows: response correctness, 11; certitude, 6; delay, 35; response latency, 7; post-response information latency, 9; certitude latency, 4; and the interactions—delay/response correctness, 11; delay/certitude, 6; response correctness/certitude, 5.

Another set of analyses were conducted for every subject's long-term response correctness. The predictors were response correctness, certitude, delay, log2 post-information latency, and also log2 response latency. There was a good fit in predicting the short-term delay items for most of the individual subjects. According to the Chi-square goodness of fit test 36 of the 48 subjects had a model that was a good fit, that is $p > .05$; and of these 28 had $p > .50$. The number of subjects with significant (.05) main effects were as follows: response correctness, 19; certitude, 12; delay, 23; response latency, 4; post-response information latency, 12.

Discussion

The data from the preliminary analyses demonstrate that the processing measures and other instructional variables (delay and PRI type) are adequately related to response correctness and to the 10 celled certitude by correctness matrix– a "degree" of correctness measure. Significant relationships were demonstrated in predicting the correctness by certitude matrix at long-term recall, as well as the response correctness at both short-term and long-term recall. This indicated to us that it is meaningful to use these variables and to look for further relationships between these variables.

More importantly, our analyses demonstrate that the likelihood of a

correct response in a computer-based training environment can be predicted with locally derived measures which are fit into logistic regression models. That is, measures of a student's processing at each frame of instruction could be taken: metacognitive judgments of response certainty, certainty latencies, post-response information latencies, and response latencies. These processing measures could be combined with measures of response correctness, delay of trials and type of PRI in a separate model for each subject.

In our results, there were not as many significant predictors for each of the individual subjects as for the grouped data. However, every one of the factors in the original model was significant for at least some of the individual subjects. Thus, to use a single group-based model in order to predict the probability of a subsequent correct response may not be as effective as using a separate model for each subject which is fit with his/her significant predictors.

As discussed in the introduction, these results are in line with those of the Hancock, Hubbard & Thurman paper (1992c) where similar measures were used in forming logistic models in an entirely different domain of learning

Experiment 2

In order to build a stronger case for the use of processing models in predicting performance we chose to attempt to replicate our results of Experiment 1. However, the instructional stimuli were changed slightly. Since it is possible that metacognitive judgments made subsequent to studying PRI may have greater predictive validity (see Nelson & Dunlosky, 1991), the location of the certitude estimate was moved-in this case to the frame immediately following the PRI. And the PRI was limited to one type—one not used in Experiment 1: Type 4, but with the name buttons enabled as in Type 5. We thus hoped to test the modeling in the specific training

environment where the learner has control over the display of various types of information.

Method

Design

The basic experimental design was a completely within subjects design--5 levels of certitude estimate, 2 levels of response correctness, and 4 levels of delay of re-presentation of drill items--with response latency, certitude latency, and post-response information (PRI) frame latency treated as continuous covariates. The predicted value in the logistic regression analyses was correct or incorrect.

Subjects

Twenty-six university undergraduates participated for partial course credit. These subjects were from the same academic program as those in Experiment 1.

Three subjects' data were not included in the final analyses: two who did not complete all of the experimental sessions; one whose equipment malfunctioned.

Equipment

The equipment was the same as used in Experiment 1.

Stimuli

The stimuli were identical to what was used in Experiment 1 except the certitude estimate and the learner control over the post-response information.

The certitude estimate occurred after the studying of the PRI rather then before receiving PRI. The question was "How sure are you that you will get this item correct the next time you see it?" Responses were with percentages of certainty as in Experiment 1.

The PRI was the same as condition 4 from Experiment 1 (See Figure 2),

except that the name buttons were enabled. Thus the subjects saw a basic response verification, verbal elaborative information about both incorrects and the correct response, in addition to names for both the correct and incorrect which were enabled to display visual and echoic elaborative information—a redisplay of the task stimuli.

Procedures

The procedures were identical to Experiment 1, except that the viewing of certainty rating scale occurred after the viewing of the PRI screen, and all of the subjects had the option of selecting visual and echoic PRI.

Results

This data set was sorted as in Experiment 1.

Predictors of Response Correctness

We attempted an analysis on the 10 celled, certitude by correctness matrix over a long-term trial, as in Experiment 1, but were not able to run analyses due to low N in some of the cells. However, we were able to test a logistic regression model with response correctness at the short-term delay as the predicted variable. According to the goodness of fit test [Chi-Square (5464) = 5307.85, p = .933] the model is a good fit. The following factors obtained statistical significance: certitude, response correctness, delay, log2 certitude latency, log2 PRI latency.

Separate Analyses for Each Subject

As in Experiment 1, separate logistic analyses were conducted for each subject, with the predicted variable being the response correctness at the next short-term delay trial. The predictors were the same. There was a good fit in predicting the short-term delay items for most of the individual subjects. According to the Chi-square goodness of fit test 20 of the 23 subjects had a model that was a good fit, that is $p > .05$; and of these 10 had $p > .50$. The number of subjects with significant (.05) main effects were

as follows: response correctness, 6; certitude, 3; delay, 13; response latency, 3; post-response information latency, 2; certitude latency, 1; and the interactions—delay/response correctness, 3; delay/certitude, 2; response correctness/certitude, 3.

Discussion

Again our predictions were confirmed, and the results of Experiment 1 were replicated. That is, logistic models were used, in computer-based instructional environments, to predict response correctness for both grouped data and for individual subjects. And once again the predictors vary from subject to subject; but each of the factors in the original model is helpful for predicting response correctness for some of the subjects.

The stimuli changes from Experiment 1 did not seem to make much difference; the fit of the models in Experiment 1 was about the same as those in this second experiment. Thus, the change in the location of the certitude estimate from before the PRI to after the PRI does not seem to help or hinder prediction. And regarding the learners having control over some of the PRI, we can see that this modeling can be successful in such a training task with some learner control.

General Discussion

It has been demonstrated in three experiments, including the Hancock, Hubbard, & Thurman study (1992c), that response correctness can be predicted by measures which are easily gathered in a computer-based training environment. These measures go beyond the usual response correctness record—they include learner metacognitive judgments and various processing latencies; they are locally derived and hence are an immediate indication of the learning of the trainee. In the future such models could be fit for each trainee and reduce the uncertainty that has previously been evident in forming student models (see Katz, Lesgold, Eggan, & Gordin, 1992).

As research along this line continues we should be concerned about prediction of the likelihood of a correct response (as in the present study) and perhaps degrees of correctness, but we should also improve our understanding of the processing underlying each response. The certitude estimates and various latencies are concomitant with the learner's processing and thus they have been called processing measures, but the theoretical framework for this processing has not been specified. If PRI is to be administered in the most intelligent manner, it must not be based on simply an empirical prediction, but it will need to be theoretically rooted and guided as well. It is the author's belief that the processing which underlies the prediction in these experiments can be explained in terms the human as a hierarchical control system (e.g. Powers, 1973, 1978). In brief, the certitude ratings, correctness states, and latencies are indicative of the amount and type of cognitive discrepancy in each subject. As a control system, the subject's behaviors and performance can be explained as his/her attempt to control perceptions by opposing discrepancy (e.g. Hancock, 1992; Hancock, Thurman, Hubbard, 1993). The data in the present paper is soon to be analyzed in terms of perceptual control theory.

Thanks to my peers, 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

Agresti, A. (1990). Categorical data analysis. New York: Wiley.

Apple Computer Inc. (1991). Hypercard. Cupertino, CA.

Hancock, T. E. (1992). Control Theory and Instructional Feedback Efforts. A report submitted to the University of Dayton Research Institute under contract F33615-90-C-0005.

Hancock, T. E., Hubbard, D. C., & Thurman, R. A. (1992a). Correct Responses with Longer Latencies are Associated With Subsequent Incorrects. A paper presented at the Annual Convention of the American Psychological Association, Washington D.C.

Hancock, T. E., Hubbard, D. C., & Thurman, R T (1992b). 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. A. (1992c). Modeling student performance and cognition using multiple CBT data sources (for more intelligent feedback). A paper presented at the Annual Convention of the American Psychological Association, Washington D.C.

Hancock, T. E., Thurman, R. A., & Hubbard, D. C. (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, Ill.

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.

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.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (1992). Modeling the Student in Sherlock II. Journal of Artificial Intelligence in Education, 3(4).

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.

Nelson, T. O., & Dunlosky, J. (1991) When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL Effect". Psychological Science, 2, 267-270.

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.

SAS Institute, (1985). SAS User's Guide: Statistics. Cary, NC: SAS Institute Inc.

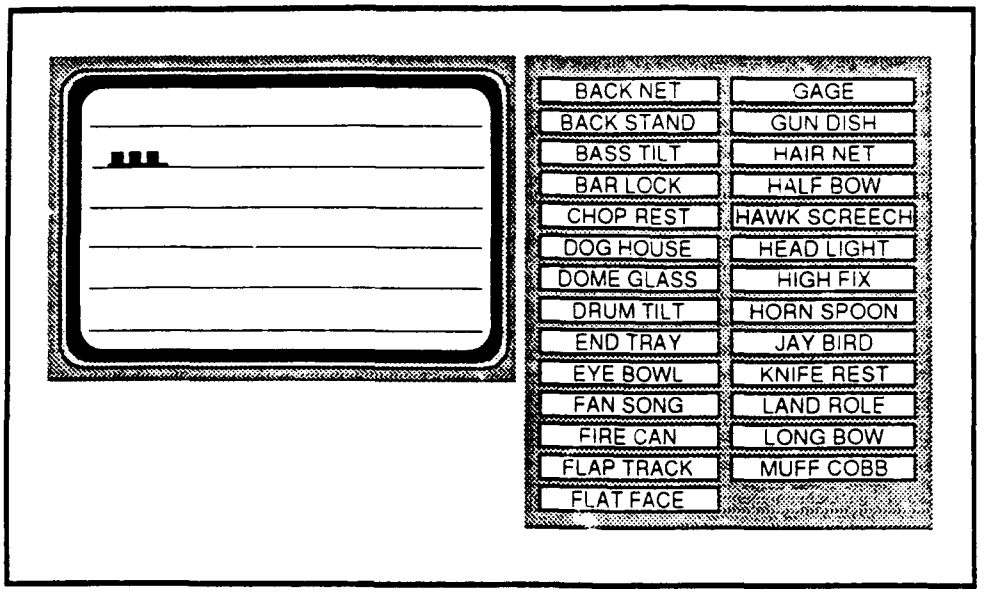


Figure 1. Sample screen display for the drill task: Response.

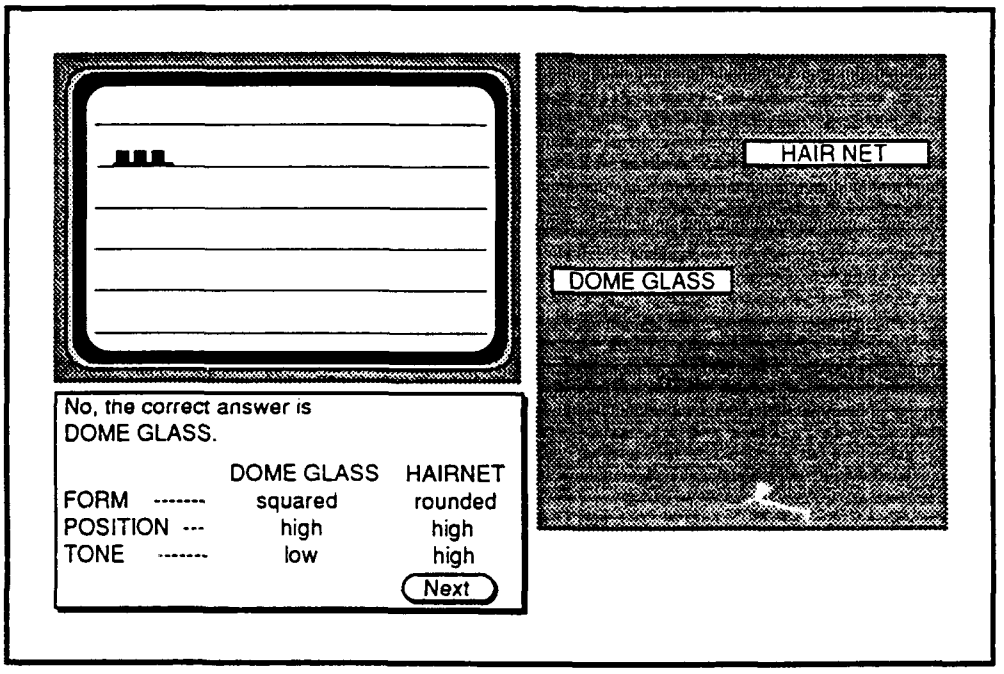


Figure 2. Sample screen display for the drill task: Post-response information.

THE COGNITIVE IMPLICATIONS OF COMPUTER-BASED LEARNING ENVIRONMENTS: A CONCEPTUAL FRAMEWORK

Michael J. Hannafin

Professor

Department of Educational Research

Florida State University

305 Stone Building, B-197

Tallahassee, FL 32306

Final Report for:

Summer Research Extension Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

July 1993

THE COGNITIVE IMPLICATIONS OF COMPUTER-BASED LEARNING ENVIRONMENTS: A CONCEPTUAL FRAMEWORK

Michael J. Hannafin
Professor
Department of Educational Research
Florida State University

ABSTRACT

Interest has emerged in the design of alternative learning systems, characterized collectively as "learning environments." Learning environments are comprehensive, integrated systems that promote cognitive engagement through learner-centered activities, concrete manipulation, and guided exploration. In this paper, a conceptual framework for learning environment research and development is presented. The purposes of this paper are to briefly summarize research and theory related to learning environments, identify similarities and differences between learning environments and conventional training and instruction, describe the underlying foundations and assumptions of learning environments, and describe the cognitive consequences of such systems.

THE COGNITIVE IMPLICATIONS OF COMPUTER-BASED LEARNING ENVIRONMENTS: A CONCEPTUAL FRAMEWORK

Michael J. Hannafin

Considerable interest has emerged in the design of learning environments. Some have abandoned traditional notions of ISD (Kember & Murphy, 1990), claiming that such models are inherently restrictive and incompatible with current views of teaching and learning. Others have extended (e.g., Hannafin, 1992) or adapted (e.g., Merrill, Li, & Jones, 1990) ISD to better reflect contemporary research and theory. However, the required changes and the theory base supporting or contradicting these extensions and adaptations have not been fully explored. The purposes of this paper are to briefly summarize research and theory related to learning environments, identify similarities and differences between learning environments and conventional views of training and instruction, describe the underlying foundations and assumptions of learning environments, and propose a framework for examining their associated cognitive implications.

AN OVERVIEW OF LEARNING ENVIRONMENTS

The phrase "learning environments" has been used to characterize everything from classroom climate to specific learning technologies. In the present context, learning environments are comprehensive, integrated systems that promote engagement through user-centered activities, manipulations, and explorations (Hannafin, Peck, & Hooper, in press). The concept is not new, nor is it uniform in meaning. Dewey (1933), for example, envisioned schools as places where learners could be guided in their pursuit of knowledge and provided hands-on opportunities to acquire insight through first-hand experience. Papert's (1980) view of "microworlds" as incubators of knowledge that illuminate and facilitate the *process* of learning are also consonant with learning environment. Others have expressed interest in learning where knowledge is situated in authentic contexts (e.g., Brown, Collins, & Duguid, 1989; Cognition and Technology Group at Vanderbilt, 1992) rather than disembodied from natural referents. Learning environments emerged not as the product of a singular psychological theory or learning paradigm as an effort to empower learners by creating systems which are principally learner- focused.

Learning Environments vs. Traditional Views

A fundamental distinction between learning environments and conventional approaches is found in their assumptions and defined purposes. The contrasting assumptions are shown in Table 1.

Traditional Instructional Methods	Learning Environments
• Instruction is a directed activity requiring the advanced specification of explicit learning objectives, the development and validation of activities to reach the objectives.	• Open-ended, user-centered systems required to support varied types of learning.
• Instruction emphasizes the transmission of domain and content knowledge.	• Activities must focus on underlying cognitive processes, not solely the products of learning.
• Instruction comprises discrete pieces of knowledge and skill organized hierarchically or sequentially--the whole is equal to the sum of the parts.	• Learning is continuous and dynamic and represent states of knowing that are continuously redefined--the whole is greater than and different from the sum of the parts.
• Instruction is principally externally directed and managed based upon the judgments of designers as to level, sequence, pace, etc.	• Individuals must assume greater responsibility for their own learning.
• Instruction emphasizes the role of the designer in either imposing or ensuring that learner options "protect the learner from himself."	• Learners can make, or can be guided to make effective choices, but need to be empowered and aided in the transition from external to internal attribution.
• Instruction is most efficient when it restricts its focus to those aspects of immediate relevance to the objectives.	• Learners learn perform best when rich and varied methods and activities are provided.
• Instruction emphasizes the breaking down of to-be-learned knowledge, many of which can be taught efficiently as verbal information via traditional teaching methods.	• Learning is best when rooted in original experience.
• Instruction tends to decontextualize learning by separating knowledge and skills from the contexts in which they derive meaning.	• Learning is most meaningful when rooted in relevant contexts.
• Instruction is probably best when efficiency and preciseness of learning are required.	• Learning environments are best for abstraction and far-transfer tasks, fuzzy or ill-defined domains, "performance in context" tasks, and problem-solving.

Table 1. Distinctions between learning environments and traditional methods.

Four summations can be extracted from these assumptions: 1) Learning environments are user-centered in terms of both the locus of control and the organization of the environment; 2) learning environments emphasize construction over transmission; 3) learning environments view the learner as integral to the system ecology; and 4) learning environments emphasize learning as a holistic process. User-centeredness requires that the system not merely accommodate user interests and preferences, but engages the learner *through* his or her perspectives; it does not merely permit, but encourages, inquiry and manipulation (Hannafin & Land, 1993). Learning environments provide tools that encourage discovery through manipulation, not merely the display of intact structures. The learner is integral to the ecology of the system. Knowledge, in this sense, does not exist apart from the individual's experience. It is nurtured and modified through interactions within the system. Learning is a holistic process, where knowledge is greater than and different from the sum of the activities and information presented. Individuals derive personal understandings which may mirror or vary considerably from others' views. The learning process is not a succession of discrete steps designed to hone understanding through simple accretion, but one in which all aspects are continuously interpreted according to the experiences and beliefs of learners.

These distinctions can be further underscored by comparison with prevailing views of instruction. Dick (1991) refers to instruction as "...an educational intervention that is driven by specific outcome objectives...and assessments that determine if the desired changes in behavior (learning) have occurred" (p. 44). Instruction relies heavily on content-driven approaches. ISD models emphasize congruence between objectives and performance standards, hierarchical analysis of the to-be-learned lesson content, externally-determined sequencing of instructional objectives, and convergent, externally-prescribed instructional activities. Instruction is, by definition, directive in nature, focusing more on the performance to be elicited than how it is derived.

In contrast, learning environments are largely non-directive in nature, often emphasizing reasoning processes and the evolution of insight over specific learning products or outcomes (Hannafin, Peck, & Hooper, in press). The tools and resources of learning environments are not designed to impart explicit knowledge at specific times, but to enable the learner to navigate productively on his or her own terms, explore the structures and limits of available concepts, generate and test tentative beliefs, and

reconstruct understanding accordingly. Learning sequences, in effect, are *supported* by the system but *generated* uniquely by individuals.

Learning environments are divergent vs. convergent, open-ended vs. closed-looped systems, and user-centered vs. content-centered. They reflect underlying models and strategies that are different from objectivist approaches. The function of the environment is not to direct learning per se but to support the negotiation of understanding and the development of insight. Learning environments seek to capitalize on the user's knowledge, experience, and epistemic curiosity by providing varied approaches to subject matter, tools for manipulating it, and resources which enable the user to create, then pursue, their own learning agenda. System features are employed for purposes that are the learner's, not the designer's.

Traditionally, ISD has emphasized "harnessing" technology to better address the goals inherent in their approaches. The emphasis has been on automating instructional activities such as eliciting responses and providing feedback, providing response-dependent presentation sequences through embedded questions and menus, record keeping, and so on. The goal has been to "increase the horsepower" of traditional methods and models, making them more powerful and efficient in addressing their goals. Learning environments seek to unleash rather that harness the capabilities of technologies and support varied teaching and learning models. They seek to aid the user in ways that are uniquely sensible. They focus on furthering and redefining prevailing notions about thinking and learning, not adhering dogmatically to the so-called "science fiction" of human cognition (Porter, 1988). They attempt to shift the locus of learning, in meaningful ways, to the learners themselves.

Traditional views tacitly assume that learning is a discrete act that can be broken into constituent parts or events; the whole is equal to the sum of the parts. This assumption promotes instructional design principles and strategies that emphasize the attainment of discrete steps, tasks, and objectives, and a view of learning as "complete" when the sequence of objectives has been mastered. Thus, designers divide terminal into enabling objectives and sequence activities procedurally or hierarchically. In learning environments, knowledge and skills are tools for refining understanding. They evolve continuously and dynamically, being clarified, modified, and revised through usage. Knowledge and skill evolve through a progression of ins nd refinements in

understanding, not through simply being told or shown. Learners do not "receive" knowledge, they construct it; they are not "given" skill, they develop them. Learners are active systems, not passive repositories.

Foundations of Learning Environments

Learning environments have three primary foundations: psychological, pedagogical, and technological. These are summarized in Table 2.

Foundation	Description	Strategy Examples
Psychological	Emphasize how individuals process information, how knowledge becomes memorable and meaningful, how it is retrieved, and how it is ultimately applied to either perform some action or support related learning; how individuals acquire, structure, retrieve, and reconstruct knowledge.	Induce cognitive dissonance through apparent contradiction; activate prior knowledge and existing schemata by providing problem context; aid learner to define expectancies by eliciting predictions and hypotheses; help learner to restructure and reconstruct knowledge by introducing multiple perspectives
Pedagogical	Emphasize how knowledge can be conveyed or otherwise made available to learners; create the structure of the learning system itself and design activities that assist learners in acquiring knowledge.	Organize content structures into internally coherent segments; supply diverse elaborations of basic concepts; provide tools and resources to manipulate constructs concretely; amplify important linkages among concepts; provide organizing problems and themes for learning system
Technological	Emphasize the capabilities and limitations of emerging technologies; the operations they support as well as the symbol systems they employ. Technological capabilities and limitations enhance or constrain possible transactions.	Supply varied presentation stimuli and symbols via multi-media; provide natural user interfaces; link among multiple, related knowledge bases; maintain coherent audit trails; support both user-select and user-query options; provide object "capture" tools.

Table 2. Foundations of learning environments.

Learning environments draw extensively from psychological research and theory in situated cognition, authentic learning, and constructivism. They stress ecological validity with respect to both the learning process and the situated nature of knowledge.

Knowledge, and the contexts in which it derives meaning, are inextricably interwoven, i.e., knowledge cannot be separated from the contexts in which it has meaning. Learning environments seek to induce cognitive engagement by situating knowledge and skills in naturally occurring, meaningful contexts. Children, for example, experience difficulty solving mathematics word problems because they acquire the computation skills independent of authentic contexts. The computation skill, for application purposes, is "inert" (Bransford, Franks, Vye, & Sherwood, 1989) and provides little productive value to the learner. Given the identical conceptual problems situated in realistic contexts, with real referents, children can solve the problems readily. They do not lack the capacity to reason, but have compartmentalized their knowledge and skills.

Pedagogical foundations influence everything from the structure of the information to be learned to methods used to convey content. Pedagogical foundations reflect differences engendered by assumptions about the learner and the learning task. Top-level (or macro-design) strategies represent an overriding pedagogical orientation. Objectivists, for example, require extensive outcome specification; constructivists, on the other hand, require few or no imposed hierarchical structures. The natural meaning of knowledge resides not in presumed hierarchies but in the contexts in which it is manifested (Cognition and Technology Group at Vanderbilt, 1990). Micro-design level strategies empower the psychological orientation of the system. In learning environments, the focus is on representations and strategies that afford opportunities, not requirements, to understand.

Technological capabilities focus on the input, output, control, and processing capabilities of technology. Independent of the psychological or pedagogical model manifested in the learning system, technologies and their associated capabilities define the "tool kit" of the designer--not the product of the effort, but the potential of the tools with which to work (Park & Hannafin, 1993). Learning systems strive to capitalize on technological enhancements in varied ways based upon different underlying models and assumptions. They dictate which formal features and symbol systems are available, the manner in which they can be invoked, the degree to which they can be merged or mixed, the speed with which computations can be performed, and the parameters within which learner-system transactions can occur. In this sense, they define the outer limit of what is possible technologically in merging psychological and pedagogical influences. Learning environments, for example, emphasize both the *designer's capability to establish* linkages

between and among nodes within an environment and the *learner's capability to generate* connections of unique meaning. This requires capabilities that transcend simple presentation and response management. Flexible data structures are required that can be organized, updated, and reorganized continuously.

Taken interactively, these foundations influence the system features and strategies needed to invoke desired cognitive processes. To the extent that learning goals emphasize high-level troubleshooting involving vaguely specified problems, for example, research and theory related to cognitive flexibility and problem-solving in ill-defined domains (e.g., Spiro & Jengh, 1990), acquiring expertise (e.g., Derry & Murphy, 1986), and invoking related prior knowledge would be referenced. Instruction and teaching research and theory associated with the learning goals, such as top-down methods of problem analysis, chunking of lesson content, and methods for transferring knowledge would be referenced. Technological capabilities then provide the capacity to link related concepts, construct and test strategy alternatives, provide detailed diagrams, and examine solution alternatives based upon the cognitive and teaching strategy requirements. The challenge for designers is to capitalize on these capabilities while not limiting views of what is possible based on traditional notions of teaching and learning.

Implications

The cognitive implications of these foundations and assumptions are related to, but extend beyond, those of traditional instruction. Clearly, while the potential payoffs may be greater, the cognitive demands of learning environments are more complex to estimate. The demands summarized in Table 3 can be further organized into 5 categories: metacognitive, management, diversity, structure, and generative.

Learning environments require substantially greater metacognitive judgment by learners. Individuals must determine which aspects of the system to use in which order, when sufficient understanding has been attained, and whether or not ongoing comprehension is sufficient to guide learning. Tools and resources must be used effectively; the mere existence of them in no way ensures effective use. The capacity to expand understanding through multiple, sometimes contradictory perspectives, contributes further to the demands. It is not sufficient to simply learn "a way" to perform procedures or a single answer or explanation. Diverse points of view must be understood, brought to bear

Assumption	Psychological	Pedagogical	Technological
Open-endedness	Cope with limited externally supplied structure to represent knowledge.	System strategies used to empower individually-relevant structures.	Elements of system provide minimal imposed guidance but require significant judgment to define usage.
Process-based	Knowledge and skill is required to abstract underlying processes.	Must integrate learning content with relevant processes.	Manipulate processes via system-supplied cognitive tools.
Dynamic	Knowledge must be integrated, flexible, and usable for related learning.	Infer beyond literal limits of content focus in systems.	Progressively build understanding and insight using options.
Individual responsibility	Personal attribution needs to be developed and increased.	Recognize how to deploy available methods to learn.	Identify how available tools and resources can be used to manage learning.
Effective choices	Utilize and/or develop metacognitive knowledge and skill.	Recognize when learning is effective and invoke decisions needed to support it.	Determine which available tools and resources provides what kinds of information.
Rich, varied methods	Multiple perspectives on concepts and topics need to be internalized.	Use available methods to gather relevant data and points of view.	Decode and interpret varied outputs, and become successful in varied input methods.
Original experience	Meaningfulness needs to be established by individual learners.	Use available methods to manipulate, experiment, and test phenomena.	Engage the system purposefully as "phenomenaria" to experience concepts and constructs.
Relevant contexts	Anchors in authentic problems need to be created; self-referencing needed.	Recognize implications of concepts within problems and scenarios.	Immerse self in system-generated settings as proxies for external context.
High-level learning	Knowing "what" insufficient to understanding, analyzing, or doing.	Build and test solutions, hypotheses, theories using given methods.	Manipulate data, knowledge, concepts using given tools.

Table 3. Foundations and assumptions of learning environments.

on a range of problems, and reconciled collectively under varying contextual circumstances. The burden for sense-making also falls more heavily on the learner. Individuals must not simply collect and organize data in ways that are compatible with accepted external notions, they must modify their beliefs and understanding accordingly and seek new data to further confirm, disprove, or modify ongoing understanding.

A FRAMEWORK FOR ANALYZING COGNITIVE IMPLICATIONS

The cognitive implications of learning environments are significant. Unlike traditional approaches, cognitive demands are reconciled individually by learners, not externally by designers. What cognitive demands are associated with learning environments? How are the demands influenced by the various tools and resources provided? How does the nature of open-ended learning systems influence perceived and predicted cognitive demands, and how do users adapt to such demands?

The framework shown in the Figure 1, adapted from research in hypertext (Gall & Hannafin, 1993), comprises five interactive components: the individual learner, overall learner attributes, the processing or performance task requirements, the features available in a given system, and the setting. At the center, both conceptually and in terms of cognitive requirements, is the learner. The relationships between the learner and other components are bi-directional, emanating from the learner to, as well as among, the other influences. This is significant. It implies that the components are interdependent and contingent upon the experiences and perceptions of individual learners. The cognitive requirements associated with any given learning task cannot be examined independently from the moment-to-moment processing demands on the learner and the influence of context, learner differences, and system features. Stated differently, cognitive requirements do not reside in the task itself, as conventionally presumed, but in the interaction between the learner and the components of the environment. The requirements can be decreased to the extent the system components facilitate engagement and enable the learner to better manage the process, or be increased to the extent the components collectively fail to accommodate the learner's needs.

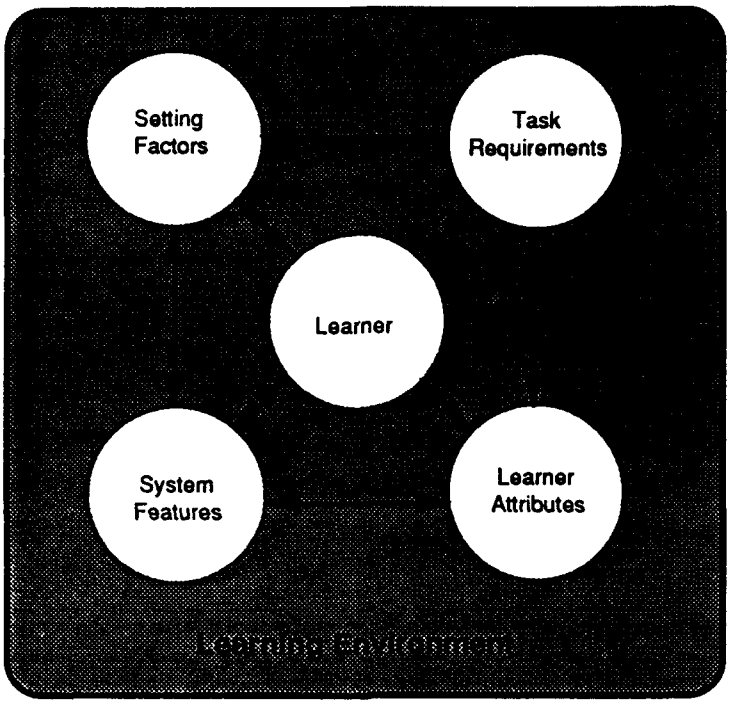


Figure 1. Learning environment framework

For clarity and simplification, the components are grouped into individually-generated (learner) and externally-stimulated influences (learner attributes, the processing task, system features, and setting factors).

Individually-Generated Influences

The learner is the center of all transactions within the environment. Perceptions are formed, meaning is defined, relationships to existing knowledge, skills, and beliefs are constructed, understanding is revised iteratively, inferences and implications are generated, needs to know are redefined, and so on. Cognition is not described in terms of discrete, static, isolated steps but as continuous, dynamic, holistic processes. Using these processes, learners do not simply collect and store, they generate and construct. The environment, therefore, must support the learner's ongoing efforts to clarify understanding, revise beliefs, define and test the limits of understanding, represent knowledge, and pursue learning activities deemed uniquely appropriate. In this sense,

learning environments need to provide system features that enable the learner to effectively mediate the internal processes of understanding.

Externally Stimulated Influences

Environmental influences represent tangible, identifiable elements of the learning system: the setting, the system's components and features, the learning task, and the characteristics and attributes of the population for whom the system is designed.

Setting

Setting refers to the contexts within which to-be-learned knowledge and skills are integrated. Factors such as authenticity, familiarity, and relevance influence both the nature of the engagement fostered and its associated processing requirements. Authentic contexts--problems organized in genuine ways and emphasizing real-life phenomena--increase the learner's ability to relate to everyday events. Pilots, for example, are known to discount training efforts which employ crude or oversimplified contexts. They limit preemptively their willingness to engage the context purposefully, deeming it contrived and unrealistic. Familiarity is also essential since the context serves to represent real-life events which must be recognized to be understood, or metaphorically to represent knowledge and skills which parallels known contexts. In the case of simulation, the context imparts cues which, if unfamiliar, go undetected and unused. When employed metaphorically, familiarity is essential since the learner will map familiar attributes to different, but parallel, concepts.

System Features

System features include specific information, data structures, symbol systems, tools, and resources provided to organize the environment and make it accessible. System features are largely technology-dependent; the specific manifestations, however, reflect the underlying teaching-learning model of the designer. Interface protocol and procedures, for example, require the investment of cognitive resources, but these requirements should be minimal compared with the substantive processing task (Norman, 1988). Undue effort to meet system requirements for secondary tasks (e.g., accessing information, navigating

within the system, inserting values into a plotting function) reduces the availability of cognitive resources for primary processing tasks (e.g., developing hypotheses, observing effects, establishing connections among concepts) (Gavora & Hannafin, 1993). In some cases, fairly complex physical and/or cognitive resources are invested to perform comparatively low-level, low-yield interactions (e.g., accessing help functions, navigating within and across multiple databases, etc.). The cognitive resources required to perform simple procedural navigation, for example, can be *vastly* disproportionate to the learning gain. In others, however, the interface permits simple, but largely superficial, engagement. Learners can peruse a database with comparative ease but have limited capacity to manipulate, connect, or otherwise engage the environment. The system must not only permit, but encourage, engagement by inducing the learner to invest cognitive resources more in significant conceptual processes rather than procedural requirements.

Learning Task

As with traditional approaches, presumed processing and response requirements must be assessed. Gagné (1985), for example, provided a method of analysis using the inherent structure of to-be-learned knowledge with "internal" and "external" events which correspond to the presumed processing and response requirements. Learning tasks are analyzed with reference to the presumed cognitive processes required and the external events likely to engender them. Once determined, the information and activities are structured to maximize the likelihood that defined knowledge and skills will be effectively transmitted, or "learned."

Unlike conventional approaches, however, learning environments do not emphasize the inherent structure of to-be-learned concepts, but the creation of *enabling representations*. Enabling representations are vehicles through which understanding can be derived at varied levels based upon different intents and motivations. The structures do not forge particular interpretations, but support the learner in his or her individual quest for meaning and understanding. The representations do not represent "ideal" structures for organizing knowledge, but provide mechanisms that support the learner's sense-making. The purpose is not so much to depict how knowledge should be represented but to organize information in ways that facilitate the individual's access to, and use of, the information.

Population Attributes

Population attributes reflect group factors and variables that serve as "givens" in design. Learners, for example, may be poorly motivated or reluctant to engage the environment, requiring additional focus on personalizing the system and affective concerns. They may be ill-equipped to pursue and/or manage their own learning, possess limited experience with the technology employed, or be unfamiliar with their roles in open-ended learning systems. They may, as a group, possess serious limitations or unusual strengths that need to be accounted for. Each learner ultimately negotiates his or her individual progress, but the system must support the homogeneity or diversity of its users.

PROBLEMS, ISSUES, AND UNRESOLVED QUESTIONS

The proposed framework does not, by itself, define the cognitive requirements of learning environments. Instead, it provides a perspective for organizing the complex, interrelated factors endemic to learning environments using foundations and assumptions about what they are, what they are presumed to do, how they are similar to as well as different from conventional instruction, and what learners must do in order to profit from them. It is a framework for defining problems, studying complex questions, and organizing answers, not an answer by itself.

The cognitive load of learning environments must be examined further. Learning environments allow individuals to manage their learning (and presumably the corresponding cognitive load and pace). Learners can remain focused on given topics or problems until they feel ready to proceed rather than attempting to maintain the pace--fast, slow, or appropriate--established externally. Learners are assumed to be productive in their quest for understanding without the need for external metering. The problems immerse learners in domains where they have acquired some sense, but often lack sufficient knowledge to be productive--the environment is designed to engage the learner in such a way that the pursuit of knowledge is rationally tied to the individual's need to know. In effect, unlike "basics first," bottom-up approaches, the environment creates contexts for knowing. The cognitive demands of these methods, for the most part, have not been studied. Little research has been reported to indicate how (or if) individuals make the

needed adjustments, what kinds of problems they encounter, and how their understanding is either enhanced or limited in problem-based learning.

While the negotiation of individual meaning and understanding is essential, much learning is routinely referenced to external standards--the so-called accountability-based learning. Certain formal knowledge and skills exist, it is argued, that must be understood absolutely according to common versus unique criteria. The issues, therefore are fourfold: 1) Left to their own devices, will learners ultimately negotiate meanings that are consistent with external standards?; 2) Do differences between external and unique understandings represent substantive knowledge gaps or differences in interpretation and representation?; 3) If differences exist, is it necessarily true that external standards are superior to those generated by the learner?; and 4) Are differences of sufficient consequence to impose meaning externally. It is apparent that learning environments yield a different kind of learning--both qualitatively and quantitatively. It is not yet clear, however, whether these difference constitute a weakness, alternative but valid understanding, or superior understanding.

The emphasis on process over product also requires validation. Again, learning environments promote understanding by emphasizing thinking and learning processes more so than specific product, or outcome, knowledge (Brown, 1985). Generalizable thinking processes, therefore, are presumed to be superior for learning environments. The learning of specific product knowledge, though also presumed to occur, may be quite variable. A great deal of product knowledge normally isolated in objectivist approaches is presumed embedded in high-level reasoning processes, i.e., it is assumed to develop as a natural requirement or consequence of the high-level reasoning. Still, it seems likely that specific product knowledge would be substantially more variable for learning environments than direct instruction. It is important to assess differences in process and product learning, and to examine the relative tradeoffs of gains as well as losses for each kind of learning.

Finally, the potential to automate or guide the design processes underlying learning environments requires study. Learning environments have historically lacked an identifiable design technology. Typically, they are discrete products based on varied psychological foundations, strategies of largely unverified effectiveness, and vastly different structures. Only recently have efforts been advanced to extrapolate common

structures across environments. Whereas the structures reported in this paper suggest commonalities, the specific implementations tend to be variable. Systems to automate or guide the design of learning environments must be sensitive to the diversity and flexibility requirements, but be sufficiently detailed to support design decisions in constructive and effective ways.

REFERENCES

Bransford, J., Franks, J., Vye, N., & Sherwood, R. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning.* New York: Cambridge University Press.

Brown, J.S. (1985). Process versus product: A perspective on tools for communal and informal electronic learning. *Journal of Educational Computing Research, 1,* 179-201.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-41.

Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*(6), 2-10.

Cognition and Technology Group at Vanderbilt (1992). The Jasper Experiment: An exploration of issues in learning and instructional design. *Educational Technology Research & Development, 40*(1), 65-80.

Derry, S., & Murphy, D. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research, 56,* 1-39.

Dewey, J. (1933). *How we think.* Boston: Heath.

Dick, W. (1991). An instructional designer's view of constructivism. *Educational Technology, 31*(5), 41-44.

Gagné, R.M. (1985). *The conditions of learning* (4th ed.). New York: Holt, Rinehart, & Winston.

Gall, J., & Hannafin, M. (1993). *A framework for the study of hypertext.* Submitted for publication.

Gavora, M., & Hannafin, M (1993). *Interaction strategies and emerging technologies.* Submitted for publication.

Hannafin, M. (1992). Emerging technologies, ISD, and learning environments: Critical perspectives. Educational Technology Research & Development, 40(1), 49-63.

Hannafin, M., & Land, S. (1993). *The foundations and assumptions of learning environments.* Submitted for publication.

Hannafin, M., Peck, K., & Hooper, S. (in press). *Advanced design concepts for emerging technologies.* Englewood Cliffs, NJ: Educational Technology Publications.

Kember, D., & Murphy, D. (1990). Alternative new directions for instructional design. *Educational Technology, 30*(8), 42-47.

Norman, D. (1988). *The psychology of everyday things.* New York: Basic Books.

Merrill, M.D., Li, Z., & Jones, M. (1990). The second generation instructional design research program. *Educational Technology, 30*(3), 26-31.

Papert, S. (1980). *Mindstorms.* New York: Basic Books, Inc.

Park, I., & Hannafin, M. (1993). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology Research & Development.*

Spiro, R., & Jengh, J. (1990). Cognitive flexibility, random access instruction, and hypertext: Theory and technology for non-linear and multi-dimensional traversal of complex subject matter. In D. Nix and R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163-205). Hillsdale, NJ: Erlbaum.

Professionals in Air Force Medical Bureaucracies: What Motivates Them to Manage?

Charles W. Mueller
Professor
Department of Sociology
University of Iowa
W140 Seashore Hall
Iowa City, IA 52242

Thomas W. Watson
Research Scientist
AL/HRMJ
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

7-28-93

Professionals in Air Force Medical Bureaucracies: What Motivates Them to Manage?

Charles W. Mueller
Department of Sociology
University of Iowa
W140 Seashore Hall
Iowa City, IA 52242

Thomas W. Watson
AL/HRMJ
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

Abstract

Physicians are an integral part of Medical Treatment Facilities (MTFs) in the Air Force. Not only must they provide medical care, but some are expected to take management positions in the MTFs. Arguing from the literature on "professionals in bureaucracies," a causal model is developed to explain physicians' desire to manage, their willingness to train for management and their beliefs about incentives needed in the future to motivate physicians to seek management positions. Survey data on 1593 Air Force physicians are used to estimate the model. There is considerable support that physicians are motivated by professional goals and norms, such as seeing the total health care picture, overcoming bureaucratic obstacles and retaining clinical skills. This is consistent with the desire for power in the form of autonomy, rather than power in the form of control or dominance.

Professionals In Air Force Medical Bureaucracies: What Motivates Them to Manage?

Charles W. Mueller and Thomas W. Watson

INTRODUCTION

Any organization with multiple levels of authority and control must constantly be recruiting and selecting people to fill management positions. Sometimes these people come from outside the organization, but often there is a reliance on internal labor markets (ILMs), that is, the filling of management positions with those already in the organization. The Medical Treatment Facilities (MTFs) in the Air Force are hierarchical bureaucratic structures that rely almost exclusively on filling management positions from an ILM composed of medical professionals, especially physicians (who also are officers in the Air Force).

The Air Force must recruit, select, train, and transition those they wish to move to management positions in MTFs. How this is currently done is described below. Critical to this process is motivating these medical professionals to seek and then train for these management positions. Little is known, however, about what motivates people to seek leadership positions, and in the case of physicians, the motivational question is particularly intriguing because their professional training prepares them for a vastly different role. The purpose of this paper is to develop and estimate a model of (1) what motivates physicians in MTFs to desire to manage, (2) what motivates them to be willing to train for these management positions, and (3) what produces their perceptions about future incentives necessary to make management attractive to them.

WHY DO PEOPLE DESIRE TO MANAGE?

There is a wealth of theory and research on leaders <u>after</u> they become leaders (e.g., Yukl, 1989; Bass, 1990), where the interest is in such topics as leader attributes, behavior and effectiveness. However, the psychological and management literatures have little to say about what motivates people to seek leadership and management positions in the first place. When leader motivation is discussed, it is argued that leaders are typically highly power motivated, either in a personalized or socialized sense (Bass 1990; Howell 1988; McClelland 1970; McClelland and Boyatzis, 1982). Power motivation in this context involves

the desire to influence others. The common sense reason, that the person wants to bring about change, would be included under power motivation.

Some investigators have also emphasized the importance of achievement motivation, which is defined as the need to excel or to exceed a standard, i.e., to do things better (McClelland 1961, McClelland and Winter 1969). Still other scholars have found that leaders combine high power motivation with moderate to high achievement motivation (Wainer and Rubin, 1969; Stahl, 1983).

In sum, it is assumed that there are sufficient numbers of individuals in organizations who wish to express their power or achievement motives by aspiring to management positions. These individuals can serve as the pool of leader candidates their organizations can select from, providing they either have the ability to lead, or have a potential for leadership that can be cultivated later through training or career development experiences.

There is, however, a critical problem associated with this assumption in the context of professionals in bureaucracies, in general, and physicians in MTFs, in particular. Professionals generally are NOT socialized to desire management positions and they usually do not apply for such positions. In many contexts, in fact, professionals employed by bureaucracies do not represent a pool of willing future managers seeking movement to administration and eagerly awaiting the offer from above.

The classic literature on professionals in bureaucracies suggests that professionals will not be motivated to seek management positions. Future managers, however, are precisely what the Air Force wishes to find among physicians in MTFs. Below, we briefly review the literature on professionals in bureaucracies. We then describe the current work context in MTFs and the current leader selection procedures. This is followed by a series of hypotheses about what motivates professionals in bureaucracies to seek management positions.

PROFESSIONALS IN BUREAUCRACIES

Literature on professionals in bureaucracies (Scott, 1966) portrays an inherent mismatch and conflict of interests and goals that will result in dissatisfied and uncommitted professional employees. This is especially apparent in professional bureaucracies like colleges/universities and hospitals where professors and physicians, respectively, are not just biding time until they are selected to be a dean or a director of hospital services.

Scott (1966) identifies four areas of role conflict for professionals in bureaucracies: (1) the professional's resistance to bureaucratic rules, (2) the professional's rejection of bureaucratic standards, (3) the professional's resistance to bureaucratic supervision, and (4) the professional's conditional loyalty to bureaucracy. The first three role conflicts are primarily a consequence of the professional's expectation of autonomy in work tasks and in decision making. This is consistent with Vollmer and Mills' (1966:264-5) portrayal of the norm that professional work should be controlled by the profession and its ethical standards and not by the more rigid and formal standards of the organization. Conditional loyalty refers to the professional being loyal to the organization if it provides the needed resources and the opportunities to use professional skills and knowledge. In short, the primary commitment is to the profession, and to the organization only if it facilitates professional growth. Scott also argues that legitimacy of the authority structure is important and, if the structure is perceived as legitimate, role conflict will be reduced.

This depiction has been challenged, however, and the goals and interests of professionals have been argued to be consistent with bureaucratic life (Kornhauser, 1962; Wilensky, 1964; Hall, 1968). The organization provides resources, integration, coordination, and communication linkages the professional would have difficulty providing independent of the organization. In return, the professional provides the innovation the organization would have difficulty producing with just "bureaucrats." In addition, Wallace (1993) has shown that commitment to one's profession (career) is generally not incompatible with one's commitment to the employing organization. Her meta-analysis shows a positive correlation of .45 between professional and organizational commitment.

As a consequence of these two literature streams, the prevailing view is that there is some truth to a claim of inherent mismatch of professionals in bureaucracies, but it is not as extreme as once believed. The simple fact is that, in modern industrialized societies, bureaucratic organizations are dominant, as are highly trained and specialized professionals. As a consequence, professionals and bureaucracies must coexist. Organizations must recruit, select and train professionals for management positions and professionals must accept these roles.

Even though this more moderate position now prevails, and we agree that the mismatch argument has been overstated, we believe there is sufficient merit

to the mismatch arguments to use them as the basis for hypotheses about what motivates professionals to seek management positions. As a result, this research not only will answer the basic question of what motivates professionals to manage, but will also address these more classical theoretical concerns.

THE MANAGEMENT STRUCTURE OF MTFs

Small facilities may have a Commander (SG) and a part-time Chief of Hospital Services (SGH). Being able to continue with clinical practice while assuming one of these positions is more likely at the smaller facilities. At larger facilities, these roles are likely to be full-time and take senior managers away from their clinical activities. Senior medical managers may experience a loss of pay if special pay for their clinical specialty is taken away. At larger facilities one also finds a much larger array of mid and senior management positions. Larger facilities are likely to have midmanagement positions such as Chief of Division, Department or Division Chairperson, Chief of Service and Chief of Aerospace Services. These midmanagement positions are often part-time and are "taken out of the hide" of the physician. He or she often performs these duties in addition to a full or almost full caseload. At larger facilities such as a Medical Center, the full-time position of Vice Commander (SG2) usually exists. Commanders and Vice Commanders are often in a boundary-spanning role dealing with those outside the facility: The Base or Wing Commander (non-medical line officers) community leaders, the press, the AF Surgeon General and staff at Bolling AFB in Washington DC or medical personnel managers at HQ AF Military Personnel Center at Randolph AFB. At larger facilities, the job usually requires active socializing (which includes one's spouse) and frequent temporary duty (TDY). Commanders also must relocate frequently. The SGH deals more with the day-to-day operation of the MTF, looking after staff and patient needs and concerns. It has been described as one of the most difficult and thankless jobs in the Air Force. Serving in the SGH role is usually considered a prerequisite for becoming an SG.

MANAGER SELECTION, TRAINING AND MOVEMENT

MTFs are categorized into difficulty groups from a low of 6 to a high of 10 based on their difficulty to manage. The rating depends mostly on size, but other factors, such as location and access to other facilities play a part. A Medical Services Management Screening Board meets once a year to screen candidates who have been prescreened before reaching the Board. The prescreening is required because a candidate must have the endorsement of

those in the chain of command prior to being presented to the Board. This Board is composed of senior medical managers, who meet and score individuals on the 6-10 difficulty dimension identified above. This means a candidate with an evaluation of 8 from the Board could be assigned to an MTF rated as high as 8 in difficulty to manage. When the Board members differ by more than 2 rating points on a candidate, the case is discussed, and if consensus cannot be reached, a MAJCOM can accept a candidate rejected by other commands. The candidates are usually known by at least some of the Board members and the selection process has been described as highly subjective. If you prove yourself to be an effective commander or SGH at less difficult-to-manage MTFs, you are likely to be seriously considered by the screening board for more demanding leadership roles at larger facilities.

Increasingly it is accepted that people with leadership potential can be groomed for future leadership roles. The grooming takes the form of training, as in the Air Force or US Military Academies, in Professional Military Education such as Squadron Officers School, Air Command and Staff College, and Air War College. Also, preparation for leadership is now assumed to be a developmental experience and future leaders need to be exposed to the right career developmental experiences in order to widen their repertoire of leadership skills.

THE BASIC MODEL

As suggested above, we believe the classical concern with the mismatch between professional and bureaucratic goals suggests a number of hypotheses about what motivates professionals to seek or not seek leadership positions. These, as well as other hypotheses, are stated below.

Figure 1 portrays a causal model that was developed from the argument to be presented below. It shows the physician's desire to manage to be a function of (1) current and past leadership experiences, (2) the current work setting, (3) incentives to manage associated with the current system, (4) disincentives to manage associated with the current system, (5) the physician's leadership self-image, and (6) the perceived legitimacy of the current selection process.

The model also portrays the physician' s desire to lead as directly affecting his/her willingness to undertake training to prepare for managing. This pushes the model beyond the attitudinal "desire to manage" to the more behavioral commitment to the training needed if the transition to management is to be

successful. Adding this component is critical if policy implications are to be derived. Although we expect a fairly high positive relationship between desire to manage and willingness to train, they clearly are not the same.

The final component in the causal model is the physician's stated perceptions of *future* incentives he/she feels are needed to make movement to a management position attractive. This component also is critical if we are to draw policy implications from this research. Although the model shows these incentives as causally subsequent to training willingness, we expect many of the perceived incentives/disincentives of the current system to impact on the future incentives. The analysis will allow us to assess this.

We argue that the professional's motivation to seek management positions will *increase*:

(1) the more they perceive taking a leadership position will provide work-related incentives (rewards) consistent with being a professional,

(2) the less they perceive taking a leadership position will produce work-related disincentives (costs) inconsistent with being a professional,

(3) the more they have a positive leadership self-image,

(4) the more they perceive the existing leader selection process as legitimate,

(5) the less they have held leadership positions in the past, (although this could be nonlinear such that the effect begins to weaken with experience),

(6) if they currently are <u>not</u> in a leadership position, and

(7) if they are in a work environment where there is good management (this would be in the larger MTFs; we have no hypothesis here about CONUS and nonCONUS location).

Arguments for each of these are briefly discussed below:

Incentives/Disincentives for Leadership (Rewards/Costs)

These hypotheses come from the literature on professionals and the more general literature on exchange theory about motivation and costs/rewards of work as motivators. Exchange theory applies because physicians will have to give up certain rewards associated with being a professional if they move to management; they literally will be exchanging certain work conditions for others.

The incentives (rewards) gained by becoming a manager are: being able to see a total health care picture, having a positive impact on health care, advancing in one's career and gaining more power/control. Being able to see

the total health care picture, career advancement and having a positive impact on health care are clearly consistent with professional goals of practicing medicine and, thus, should be motivators consistent with professional norms. Gaining power and control, however, is not a professional goal, except where it is expressed as power and control over one's work (autonomy). Also, enhanced power is afforded professionals in terms of what French and Raven (1959) describe as expert power.

Likewise, the disincentives may be viewed in terms of professional goals and norms. Having to face the bureaucracy, having to confront the resources/health goals mismatch and losing clinical skills should be reasons for physicians, as professionals, to not seek management positions. Loss of pay should not be a major concern of a professional, although from the perspective of economists with their major underlying assumption about economic rationality driving economic behavior, this should be a disincentive for every employee. Physicians, having spent considerable time and money for training, may be especially sensitive to pay loss as an equity issue. Finally, having to experience more frequent social disruptions, like making the family relocate, violates a more general social norm, and is not one specific to professionals. It should reduce the desire to manage for everyone.

Leadership Self-Image

Bandura (1979, 1982) speaks of the centrality of self-efficacy as a mechanism motivating human behavior. He even relates the construct directly to career interests and pursuits, arguing that perceived self efficacy (or lack of it) can facilitate or inhibit career choices. Physicians, in part, probably choose medicine as a career because they believe themselves to be more competent or efficacious in such a role. Physicians who have a positive image of themselves as a leader or potential leader would logically be more likely than other physicians to aspire to leadership and command roles. Bass (1990) in his recent revision of the Handbook of Leadership confirms that self image has a lot to do with the tendency to want to lead. He reports that leaders who see themselves as masters of their own fate also tend to see themselves as more self-efficacious. Bass also reports that the closely related concepts of self confidence and self esteem are positively related to leadership.

Legitimacy of Selection Process

An authority structure is legitimate if employees are clearly aware of its existence and also perceive it to be proper. People are more accepting of an authority structure they believe is legitimate and they become more attached (committed) to the organization when this is the case (Weber, 1947 ; Halaby, 1986). Related to this is the literature on procedural justice (Folger, 1987), where it is argued that individuals are more positive about outcomes that are perceived to have been arrived at through the application of fair procedures. From this we argue physicians will be more willing to become part of a management structure that is staffed fairly. Another possibility here, however, is that legitimacy serves as a moderator; i. ., incentives and disincentives to manage will motivate employees to manage ONLY if the procedures for selecting managers are perceived as legitimate. This moderator hypothesis is also tested.

Current and Past Leadership Experiences

Those who are currently in management positions or have been in leadership positions in the past should have a more accurate picture of whether managing allows for realizing professional goals. These experiences could have been both positive and negative, thereby making it difficult to hypothesize the effect of these on desire to manage. However, following the literature on professionals again, we expect the effects to be negative because the problems confronted would likely have been frustrating for someone socialized to uphold professional goals and norms. Any effect is likely to be nonlinear, with the effect tapering off in strength with additional experience.

Current Work Conditions

Experiences with "good" managers should positively reinforce any desires to manage and help erase concerns about the incompatibility of a professional as a manager. Because managers in the larger MTFs have had to "prove" themselves to reach those positions, we expect the larger MTFs to provide a more positive environment and thus result in a heightened desire to manage. The CONUS vs. nonCONUS distinction is included for purposes of control.

DATA AND METHODS

The Sample

The population targeted was the approximately 4000 Air Force physicians. Some physicians in student status (such as through the Air Force Institute of Technology or AFIT) were excluded from the sample since they were

more difficult to reach. The number who returned surveys were as follows: 1988 or 50%, of which 129 were in command positions (Commander, Vice Commander, and Chief of Hospital Services); 1138 were staff physicians or "other" which typically meant resident; most of the remainder (704) were in various mid-management positions such as Chief of Division, Department Chair, or in a Headquarters position. Position identification was left blank by 6 respondents and eleven provided multiple responses.

The Surgeon General's committee (see below) encouraged us to keep the number of demographic items small, and thus, we had few demographic characteristics on which to compare the physician population with the respondent sample. However, using data bases available at the Human Resources Directorate of the Armstrong Laboratory, some comparisons were possible. Overseas location was slightly over-represented (4.1%), captains were slightly under-represented (5.1%) and Lt. Colonels were slightly over-represented (3.1%).

The subsample used in the analysis reported here excludes commanders, vice commanders and chiefs of hospital services. We wanted to estimate the model for those currently not in senior management positions. After cases with any missing data are excluded, we have an N of 1593 for the analysis.

Measurement

Because this area of research is understudied, no established scales existed for the constructs in the model. An instrument, called the Physician Leadership Survey, was developed specifically for this study. This survey was developed by the second author with the extensive assistance of the members of the Surgeon General's Ad Hoc Committee on Physician Leadership Development (with the exception of the second author, all were senior medical officers (physicians) with the rank of Lt. Colonel through Maj. General), and approximately 35 Wilford Hall Medical personnel ranging in rank from Captain to Colonel. These physicians were interviewed in small groups to solicit their suggestions on earlier drafts of the survey. Colleagues at the Human Resources Directorate of the Armstrong Laboratory also reviewed drafts of the survey and suggested improvements. The second author met with the committee members on two occasions to gain information on survey development and refinement. In its final form the survey was composed of 8 parts with a total of 113 items.

Because the items in the survey were "new" and not from standard scales (but were carefully developed and evaluated, as described above), items within

the various parts of the survey were submitted to exploratory factor analysis: all of the items identified as current incentives and disincentives were factor analyzed together, all of the items identified as future incentives for management were factor analyzed together, and all of the attitude items were analyzed together. Because we expected the factors to be related, nonorthogonal rotation techniques were used. Following the eigenvalue of one criterion for extracting factors and attempting to obtain close to "simple structure" patterns, items loading under .30 were dropped as were items that loaded above .30 on more than one of the factors. Also, the substantive criterion of meaningfulness of the factors and the associated items was employed throughout. As a consequence of these criteria, several of the scales are based on only two items. Despite this, most of the internal reliability coefficients (alphas) were respectable, ranging from .56 to .88. Only four were below .60. The Appendix lists the items used for each scale and provides specific alphas.

Because we have multiple items for most of the constructs in the model, LISREL is used to estimate the causal paths. LISREL is superior to ordinary least squares regression because it corrects for unreliability in estimating the paths in the causal model and it routinely produces information useful in assessing the goodness of fit of the model to the data.

RESULTS

Table 1 presents means and standard deviations for the variables based on composites where the items for each scale are summed and divided by the number of items. For all but the current work setting and current and past leadership experience variables, the scales range from 1 to 6. As may be seen there is no overwhelming desire to lead among these physicians--the mean is 3.05. This is consistent with the argument that professionals generally do not seek management positions. The willingness to train is relatively high and there is strong support for both extrinsic and autonomy incentives being added in the future. The especially high mean for autonomy incentives is consistent with arguments about professional norms being carried to the workplace. The also high mean for extrinsic incentives is not what would be expected on the basis of these arguments, however. It should be noted that Air Force physicians are underpaid relative to their civilian counterparts. Comment pages from the surveys support the view that they expect additional extrinsic rewards for taking on the extra hassles of management.

In general, the physicians perceive the current system as providing incentives (rewards) for taking a management position, with the largest mean being for having a positive impact on health care, which clearly would be desirable as an enticement for a physician to take a management position. With the exception of resources/health goals mismatch, the physicians perceive the cost features as disincentives for taking a management position. Based on magnitudes of the means, a loss of pay is the strongest disincentive. Finally, we see that the physicians generally see themselves as having leadership potential, but they have mixed feelings about whether the leadership selection procedures are legitimate; they are skeptical of command.

These univariate data, however, can be misleading because they do not inform us about which of these variables directly impact on the physician's desire to manage. Estimation of the causal model provides this information.

Prior to the LISREL estimation of the model, we tested for the possibility of nonlinear effects of number of past military positions and number of nonmilitary leadership positions. The effects of both variables are linear on all endogenous variables (the four dependent variables). In addition, we tested the hypothesis that legitimacy of the selection process moderates the effects of the other independent variables on desire to manage. We found no support for this interaction hypothesis for any of the exogenous variables.

Figure 1 includes the path coefficients for the statistically significant paths (one-tailed at $p < .05$). Any paths excluded from the diagram should be interpreted as not being statistically significant. The four residual paths going into each of the dependent variables represent the effects of all unknown causes that have not be included in the model. Squaring these gives an indication of the amount of variance not explained. All path values may be interpreted as standardized partial regression coefficients and their magnitudes may be directly compared to assess relative impact. As an illustration, career advancement (.21) is more than twice as important as positive impact on health care (.08) in producing a willingness to train for management positions.

The model is quite successful in explaining desire to manage: 68% of the variance is explained. Considering specific determinants of *desire to manage*, we find the following significant <u>net</u> effects:

1) Those not currently in a management position are slightly more interested (.05) in management than those already managing.

2) The incentives are less important than the disincentives. Specifically, being able to understand the total health care picture (.36) motivates physicians to manage, but negative social consequences (-.20), having to confront the bureaucracy (-.24) and losing clinical skills (-.28) all reduce the desire to manage.

3) Those with a positive self-image of their leadership skills (.26) are more interested in managing.

4) Those who consider the current process of selecting leaders as not legitimate (-.15) have a greater desire to manage. We had not hypothesized this. However, as it suggests, it is possible that those who perceive it as not legitimate may want to manage in order tc change the procedures.

Considering the determinants of *willingness to train*, where 44% of the variance is explained, we found that those who want to manage (.46) are also those who are more willing to train, those who see management as a way to advance one's career (.21) are the ones more willing to train, and those who feel they can have a positive impact on health care (.08) are more willing to train. In addition, there are a number of indirect effects on willingness to train through Desire to Manage (data not shown).

Finally, we are able to assess what factors influence the physicians' suggestions for future ways to motivate them to seek management positions. Those who want to see *extrinsic incentives* instituted are those most willing to train (.29), those who see management as a means of career advancement (.23) and those who are concerned about possible loss of pay (.46). In addition, those who see managers as having to confront bureaucracy (.21) do not want extrinsic incentives instituted. Forty-nine percent of the variance in extrinsic incentives is explained.

Those who want greater *autonomy* introduced as a future incentive are those who are willing to train (.19) , see the gain of power in positive terms (.29), see the current bureaucracy in negative terms (.41), believe being a manager will result in clinical skills loss (.14) and have a positive self-image of their leadership skills (.14). Twenty-nine percent of the variance is explained.

CONCLUSIONS

Is power the major motivator as the literature on leadership suggests. If by power we mean control and major decision making in a MTF, then the answer is "no." Gaining this type of power does not affect desire to manage, nor is it an incentive for pursuing more training or asking for future extrinsic rewards. This finding is consistent with the "professionals in bureaucracies" literature.

Professionals are not supposed to seek control through management and they do not here.

If by power we mean autonomy, however, then the findings are consistent with it as a motivator. Probably the key defining characteristic of being a professional is having autonomy in decision making about the day-to-day tasks associated with one's work. Not wanting to confront bureaucracy as a manager is consistent with this. Also, those who see power more traditionally as control and major decision making want greater autonomy incentives in the future.

Other findings are also consistent with this argument that professionals will consider professional goals and norms when deciding whether to seek management positions. Physicians should want to manage if doing so allows them to better see the total health care picture and if they do not have their clinical skills deteriorate as a consequence of managing. Both effects are found. In addition, professionals should not view pay incentives as critical in their decision making. Consistent with this, there was no effect of loss of pay on the desire to manage.

In conclusion, the data marshaled in this study are quite consistent with the argument that professionals do use their professional norms and goals in evaluating whether or not to seek management positions. With only a few exceptions, the findings are remarkably consistent with such a portrayal.

Another way to think of our results is that they paint a fairly rational picture of why physicians seek management positions, why they are willing to train for such positions, and what they want as future incentives to motivate physicians to seek management positions. It is not rational in an economic rationality sense--pay (loss of) is not the key motivator as the economists would argue. Air Force physicians are accustomed to receiving less pay than their civilian counterparts and have already foregone that higher pay outside the military. What is rational for them appears to have been shaped by professional goals and norms. That is, they will desire to manage if to do so allows them to continue to pursue current professional goals (such as remaining clinically proficient) while also serving broader professional interests such as seeing the total health picture. They were also willing to train if they could have an impact on health care at a broader level and if it would advance their careers.

APPENDIX

Current and Past Leadership Positions

NONMANAGER: a dummy variable coded 1 if not currently in a management position and 0 if a manager (Chief of Division, Dept or Division Chairperson, Chief of Service, Chief of Aerospace Services, Command Staff Position).

PAST MILITARY POSITIONS: the sum of the military management positions held in the past (Commander, Vice Commander, Chief of Hospital Services, Chief of Division, Dept or Division Chairperson, Chief of Service, Chief of Aerospace Services, Command Staff Position) .

PAST NONMILITARY LEADERSHIP POSITIONS: the sum of the nonmilitary leadership positions held in the past (HS or College Athletics, HS or College Non-athletic organization, fraternity or sorority, in medical school, in religious activities, in community service, in politics, in other activities).

Current Work Setting

CONUS: dummy variable coded 1 for located in the continental U.S. and 0 if located outside the continental U.S.

Size of the MTF: coded 1 for a clinic, 2 for a hospital of less than 50 beds, 3 for a hospital of more than 50 beds and 4 for a large medical center. Those in Command Staff Positions are coded as 4.

Incentives (Rewards)

The following constructs are based on multiple-item scales using a 6-point response scale: (1) Very Great Disincentive, (2) Great Disincentive, (3) Slight Disincentive, (4) Slight Incentive, (5) Great Incentive, (6) Very Great Incentive. Don't know was coded as 3.5. Scales marked with an * use a different set of response categories: (1) Strongly Disagree, (2) Disagree, (3) Slightly Disagree, (4) Slightly Agree, (5) Agree, and (6) Strongly Agree. Don't know was coded as 3.5. An R indicates that the item was reverse coded.

See Total Health Care Picture: measured by two items: 22:getting a "big picture" understanding of Air Force Health care; 38:knowing all aspects of MTF operations. Alpha= .70.

Positive Impact on Health Care: measured by five items: 8:having influence over health care quality; 11:being in a position to mentor others; 12:the ability to influence staff or patient morale; 15:fostering teamwork or cohesion; 20:ensuring good patient care. Alpha= .84.

Career Advancement: measured by two items: 18:the opportunity to get leadership and management training; 19:the impact on one's Air Force career. Alpha= .69.

Power and Control: measured by two items: 10:have power or control; 29:being a major decision maker. Alpha= .59.

***Disincentives (Costs)* (coded so high score indicates a cost)**

Negative Social Consequences: measured by three items: 44:having to move frequently; 45:the impact on ones family; 46:the social obligation of the job. Alpha= .77.

Loss of Pay: is measured by two items: 9:the lack of additional pay; 23:the potential loss of incentive and special pay.
Alpha = .58.

Confront the Bureaucracy: measured by two items: 17:dealing with bureaucracy; 21:having to carry out the desires of more senior leaders. Alpha= .56.

Resources/Health Goals Mismatch: measured by two items: 37:the match between mission needs and resources; 43:emphasis upon the cost of medical care. Alpha= .59.

Clinical Skill Loss*: measured by two items: 106:it would be difficult to return to clinical practice after being an MTF commander; 108:MTF commanders lose their clinical skills. Alpha= .77.

Leadership Self Image*: measured by three items: 100:since becoming a physician, I have assumed positions of leadership on or off the job; 103:I am comfortable with the idea of being a leader; 104:I relate well to all kinds of people. Alpha= .64.

Legitimacy of the Selection Process*: measured by two items: 107:senior medical managers do a good job of identifying physicians with leadership and management potential; 109:young physicians who could become good future leaders or managers are frequently overlooked by senior management. Alpha= .62.

Dependent Variables

Desire to Lead*: measured by three items: 92:being an MTF commander appeals to me; 93:the disadvantages of being an MTF commander outweigh the advantages (R); 102:I would rather treat patients than be in a leadership or management position. Alpha=.77.

Training Willingness: measured by 34 items (58-91)about willingness to take additional training needed for management positions. These items were grouped on the questionnaire into five categories: (1) Type of training, (2) Duration, (3) Location, (4) Formal recognition, and (5) types of programs. The first four groupings used a six-point response scale: (1) Very Unwilling, (2) Unwilling, (3) Somewhat Unwilling, (4) Somewhat Willing, (5) Willing and (60 Very Unwilling. The fifth set of items use the following six-point response scale: (1) Very Disinterested, (2) Disinterested, (3) Somewhat Disinterested, (4) Somewhat Interested, (5) Interested, (6) Very Interested. Don't knows were coded as 3.5. The items within each grouping were summed and divided by the number of items to form five scales that were then used as the five indicators in the LISREL analysis. The actual 34 items used are available from the author upon request. Alpha= .88.

Future Management Incentives

A six-point response scale was used: (1) Very Unimportant, (2) Unimportant, (3) Somewhat Unimportant, (4) Somewhat Important, (5) Important, (6) Very Important. Don't knows were coded as 3.5.

Extrinsic Incentives: measured by five items: 47:special promotion incentives for serving in leadership positions; 48:academic degree credit for management training (e.g., Master's degree); 50:Special TDY funds for management responsibilities; 53:Ability to obtain flight pay; 55:special administrative bonuses or incentive special pay. Alpha= .72.

Autonomy Incentives: measured by three items: 51:Being able to waive regulations that interfere with one's ability to manage; 54:having autonomy to run one's MTF without outside interference; 56:being supported by, rather than "under the thumb" of, one's MAJCOM and HQ. Alpha= .70.

REFERENCES

Bandura, A. 1977. "Self-efficacy: Toward a unifying theory of behavior change." *Psychological Review* 84: 191-215.

Bandura, A. 1982. "Self-efficacy mechanism in human agency." *American Psychologist* 37:122-147.

Bass, B.M. 1990. *Bass and Stogdill's Handbook of Leadership: Theory, Research and Managerial Applications* (3rd ed). New York: The Free Press.

Folger, R. 1987. "Distributive and procedural justice in the workplace." *Social Justice Research* 2:143-159.

French, J. and B.H. Raven. 1959. "The bases of social power." In D. Cartwright (ed.), *Studies of Social Power*. Ann Arbor, MI: Institutde for Social Research.

Halaby, C.N. 1986. "Worker attachment and workplace authority." *American Sociological Review* 51:634-649.

Hall, R.H. 1968. "Professionalization and bureaucratization." *American Sociological Review* 33:92-104.

Howell, J.M. 1988. "Two faces of charisma: Socialized and personalized leadership in organizations." In J. Conger and R. Kanungo (eds), *Charismatic Leadership:The Illusive Factor in Organizational Effectivieness.* San Francisco: Jossey-Bass.

Kornhauser, W. 1962. *Scientists in Industry: Conflict and Accommodation.* Berkeley: University of California Press.

McClelland, D.C. 1961. *The Achieving Society.* Princeton, NJ: Van Nostrand.

McClelland, D.C. 1970. "The two faces of power." *Journal of International Affairs* 29-47.

McClelland, D.C. and R.C. Boyatzis. 1982. "Leadership motive pattern and long-term success in management." *Journal of Applied Psychology* 67:737-743.

McClelland, D.C. and D.G. Winter. 1969. *Motivating Economic Achievement.* New York: Free Press.

Scott, W.R. 1966. "Professionals in complex organizations." Pp. 265-275 in *Professionalism*, H.M. Vollmer and D.L. Mills, editors. Englewood Cliffs, NJ: Prentice-Hall.

Stahl, M.J. 1983. "Achievement, power and managerial motivation: selecting managerial talent with the job choice exercise." *Personnel Psychology* 36:775-789.

Vollmer, H.M. and D.L. Mills, 1966. *Professionalism.* Englewood Cliffs, NJ: Prentice-Hall.

Wainer, H.A. and I.M. Rubin. 1969. "Motivation research and development entrepreneurs: Determinants of company success." *Journal of Applied Psychology* 53:178-184.

Wallace, J.E. 1993. "Professional and organizational commitment: Compatible or incompatible?" *Journal of Vocational Behavior* 42:333-349.

Weber, M. 1947. *The Theory of Social and Economic Organization.* Glencoe, IL: The Free Press.

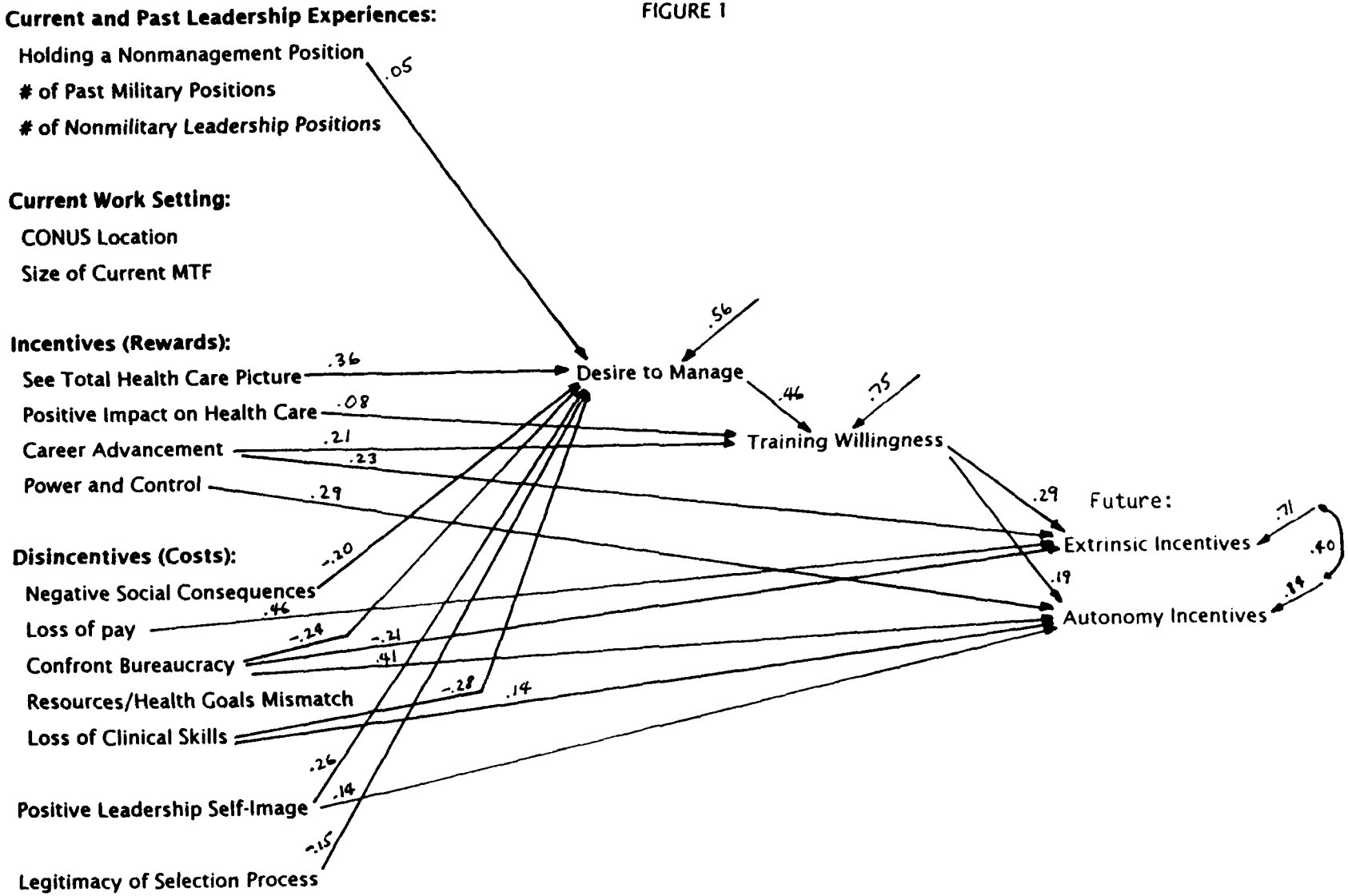
Wilensky, H.L. 1964. "The professionalization of everyone?" *American Journal of Sociology* 70:137-158.

Yukl, G.A. 1989. *Leadership in Organizations.* Englewood Cliffs, NJ:Prentice-Hall.

Table 1. Means and Std Deviations (N=1593)

Variable	Mean	Std Dev
Dependent Variables		
Desire to Manage	3.051	1.148
Training Willingness	4.118	.751
Extrinsic Incentives	4.459	.758
Autonomy Incentives	4.961	.764
Current and Past Leadership Experiences		
Nonmanagement Position	.618	.486
# Past Military Positions	1.394	1.226
# of Nonmilitary Leadership Positions	2.748	2.097
Current Work Setting		
CONUS Location	.841	.366
Size of Current MTF	2.814	1.046
Incentives (Rewards)		
See Total Health Care Picture	4.090	.814
Positive Impact on Health Care	4.761	.628
Career Advancement	4.253	.772
Power and Control	4.244	.721
Disincentives (Costs)		
Negative Social Consequences	4.416	.897
Loss of Pay	4.672	.763
Confront Bureaucracy	4.453	.861
Resources/Health Goals Mismatch	3.843	.975
Loss of Clinical Skills	4.268	1.019
Positive Leadership Self-Image	4.735	.741
Legitimacy of Selection Process	3.044	.980

FIGURE 1


Current and Past Leadership Experiences:
Holding a Nonmanagement Position
of Past Military Positions
of Nonmilitary Leadership Positions
Current Work Setting:
CONUS Location
Size of Current MTF
Incentives (Rewards):
See Total Health Care Picture
Positive Impact on Health Care
Career Advancement
Power and Control
Disincentives (Costs):
Negative Social Consequences
Loss of pay
Confront Bureaucracy
Resources/Health Goals Mismatch
Loss of Clinical Skills
Positive Leadership Self-Image
Legitimacy of Selection Process
.05
.56
.36
.08
.21
.23
.29
-.20
.46
-.24
-.21
.41
-.28
.14
.26
.14
-.15
Desire to Manage
.46
.75
Training Willingness
.29
.19
Future:
Extrinsic Incentives
Autonomy Incentives
.71
.40
.14

Aircrew Training Management Systems: A Blueprint for Design and Development

R. Ramesh

Associate Professor
Department of Management Science & Systems
School of Management
State University of New York at Buffalo
Buffalo, N.Y. 14260.

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory, Williams AFB

Sponsored by:
Air Force Office of Scientific Research
9800 Uplander Way, Culver City, CA 90230-6608

September, 1993

Abstract

Aircrew training is an important function in the Air Force, and is aimed at developing and maintaining the forces in a perpetual state of mission-readiness. The training is carried out intensively at both the formal training schools and the field units of the Air Force. This requires a considerable committment of resources and the co-ordination of several training-related activities. The growing complexity of the training programs, warfare technology and the training management strategies in the Air Force have continuously underscored the importance of system integration in the management of training programs. The concept of a Training management System (TMS) is an offshoot of this. While the interpretations of a TMS may vary, a TMS in general, can be defined as an information/decision support system for a training organization, that supports and integrates the organizational functions in an efficient and cost-effective manner. The purpose of the current study is to formalize this definition and develop a blueprint for the standardization and structured development of a TMS in general for the Air Force. This objective is accomplished by developing the specifications for a TMS at an appropriate level of detail, and illustrating the standards by adapting them to the support requirements of the 542nd CTW, Kirtland AFB. Detailed specifications for the 542nd CTW have been developed, and the next set of logical stages in system development have been outlined. The proposed standards should serve as a frame of reference for all future TMS development in the Air Force. This report is a condensation of the TMS model specifications provided in the comprehensive report submitted to the Armstrong Laboratory at Williams AFB and the 542nd CTW. The reader is referred to the comprehensive report for the details of these specifications, which may be obtained from either the Armstrong Laboratory or the 542nd CTW.

1 Introduction

The development and the maintenance of mission-ready aircrews is a primary goal of the Air Force. Aircrew training is systematically carried out at both formal training schools and the field units of the Air Force to meet this objective. The training is carried out using a variety of techniques such as in-flight, static-airframe, ground, simulator and classroom instruction, and the training programs at the formal schools are usually packaged into programmed instruction modules comprising of these techniques. The development of such training programs entails considerable human and equipment resources, and their management requires extensive planning, co-ordination and control of all the training-related activities. The complexity of these tasks, the magnitude of the resource investments and the extreme importance of achieving the training objectives together strongly underscore the need for the efficient management of training programs. As a result, the current trend within the Air Force is to design aircrew training programs as *total integrated systems*, in order to achieve the training objectives at a desired level of cost-effectiveness. *Training Management Systems (TMS)* are products of this doctrine, and are aimed at providing efficient and integrated computer support for all the major functions in a training program. In particular, a TMS is required to provide adequate information and decision support throughout a training organization by integrating the support requirements of its constituents and serving as a central facilitator of all their functions.

While the concept of a TMS addresses an important need in the current aircrew training programs, it is also relatively new. This concept has evolved from years of experience with aircrew training, and has been necessitated by the increasing complexity of the training requirements, and consequently, that of program management. Coupled with this is a concurrent shift in the Air Force to contracting out the design, delivery and support of aircrew training. These changes have introduced several new dimensions in program management, emphasizing the need for system-wide integration more than ever before. However, the relative newness of the concept and the emerging facets of program management have led to different interpretations of a TMS and its capabilities. This can be largely attributed to the diversity in the support requirements of various training organizations. Nevertheless, it is possible to extract the support elements of a TMS at a relatively abstract level. This abstraction is generalizable, because the support requirements at this level are common to most aircrew training programs. However at a logical level, training management systems could take different forms. Several such systems are in different stages of development in the Air Force, and the TMS for the C-130 training program at the Little Rock AFB is probably the most mature and well-documented system to date.

The lack of a unifying model at the abstract level has led to a multiplicity of interpretations and a proliferation of training management systems in the Air Force. Such a model, if available, would provide a macro-level view of the architecture of training management systems and their functional requirements. The architectural view can be adapted to the logical requirements of

individual training programs, and appropriate systems designed. The current study is motivated by this observation. The purpose of the current study is twofold. The first is to develop an architectural model of a TMS and its functional specifications at a maximum level of applicability to any TMS in general. The second is to show how this abstraction can be adapted to the logical requirements of an Air Force training organization. The helicopter training programs at the 542nd CTW, Kirtland AFB have been chosen for this study.

This report is a condensation of the TMS model specifications provided in the comprehensive report submitted to the Armstrong Laboratory at Williams AFB and the 542nd CTW. The reader is referred to the comprehensive report for the details of these specifications, which may be obtained from either the Armstrong Laboratory or the 542nd CTW.

1.1 The Research Objectives

The major objectives of this research are as follows.

- Develop an architectural model of a TMS at the minimum level of generalizable abstraction. The specifications of this model should serve as a blueprint for the design and development of training management systems in general.
- Assess the logical requirements for a TMS at the 542nd CTW.
- Develop specifications for the logical design of a TMS supporting the 542nd CTW operations by adapting the TMS model to the logical system requirements.
- Develop a strategy for the development and the successful implementation of a TMS at the 542nd CTW.

A summarized version of the architectural model of a TMS is provided in the following discussion.

2 The TMS Model

The TMS model is a specification of the functional requirements, system configuration and process configuration for a training management system in general. While the generalizability provides a frame of reference for designing a TMS for any aircrew training program, it also requires a certain level of abstraction due to the program diversities. The proposed model captures the functional and system commonalities among the programs at a maximum level of applicability to any TMS in general. The model ensures system flexibility and modularity, such that it can easily be adapted to any training organization. The model specifications are as follows.

2.1 Functional Specifications

The structure of training organizations can be different, and that of a given organization could change over time. However, their functional responsibilities tend to remain constant, and the organizations exhibit a certain commonality in this area. Hence, the proposed approach to TMS standardization is through the functional specifications.

The major functions of a training organization can be broadly organized into seven management areas: *Curriculum management, Course-configuration management, Scheduling management, Logistics management, Resource management, Student performance management* and *Administrative management.* The major functions in each of the management areas can be organized into a hierarchical taxonomy, constituting the functional specifications. These specifications are as follows.

1. Curriculum Management

- Courseware Design
  - Master task listings
  - Objective hierarchies
  - Course structure design
    - Hierarchy of lessons
    - Lesson/objective cross-listing
    - Lesson specifications
- Courseware Production
  - Lesson plan production
  - Lesson material
    - Documentation
    - Software/hardware
- Curricular Planning
  - Assessment of resource requirements
  - Classload planning
  - Master schedule preparation
- Audit and Evaluation
  - Task listings/objectives A& E
  - Course structure A & E
  - Courseware production tracking
  - Courseware A & E

- Resource planning A & E
- Master schedule A& E

2. Configuration Management

- Storage and Maintenance
  - Course structure
  - Lesson plans
  - Lesson material
- Information Access Support
- Configuration A & E

3. Scheduling Management

- Schedule Preparation
  - Short-term schedules
  - Daily schedule refinement
- Co-ordination
  - with on-base agencies/individuals
  - with outside agencies/individuals
- Documentation/Report Generation
- Schedule Announcement
- Scheduling Tasks
  - Flight scheduling
  - Static scheduling
  - Ground scheduling
  - Simulator scheduling
  - Academic scheduling
  - Individual scheduling
- Audit & Evaluation
  - Schedules A & E
  - Co-ordination A & E
  - Documentation/Reports A & E

4. Logistics Management

- Aircraft Maintenance
  - Configuration inventory management
  - Spare part inventory management
  - Technician resources management
  - Preventive maintenance scheduling
  - Breakdown maintenance management
- Aircraft Availability Reporting
  - Configuration availability reporting
  - Time slot availability reporting
- Aircraft Maintenance A & E
- Availability Reporting A & E

5. Resource Management
   - Human Resources Inventory Management
     - Instructor qualifications maintenance
     - student qualifications maintenance
   - Equipment Resources Inventory Management
     - Simulator configurations maintenance
     - Other equipment configurations maintenance
   - Facilities Inventory Management
     - Facilities configurations maintenance
   - Equipment/Facilities Maintenance Management
     - Preventive maintenance scheduling
     - Breakdown maintenance scheduling
     - spare parts inventory management
   - Resource Availability Reporting
     - Human resources availability
     - Equipment resources availability
   - Audit & Evaluation
     - Human resources inventory A & E
     - Equipment resources inventory A & E
     - Facilities inventory A & E

- Equipment/facilities maintenance A & E
- Resource availability reporting A & E

6. Student Performance Management

- Maintenance of Student Folders
- Student Performance Tracking
- Student Performance Analysis
- Assessment of Student Training Requirements
- Student Training Program Planning
- Student Performance Management A & E

7. Administrative Management

- Maintenance of Personnel Records
- Management of Security Clearance Requirements
- Co-ordination
  - with on-base agencies
  - with outside agencies
- Personnel Administration
- Administrative Management A & E

The specific functions in this taxonomy are fairly self-evident. The taxonomy provides a macro-level view of the functions in a training organization. In developing logical designs of TMS for specific organizations, this taxonomy should be broken down into finer task elements and organized into a larger task hierarchy. The model specifications have been derived independently of the organizational structure, and can be mapped onto the structure by identifying the functional responsibilities of the organizational units and relating them to the functional/task taxonomy. Such a mapping would provide an alternate view of the operations in terms of the organizational structure. Combining the two views, the system and process configurations can be designed.

2.2 System Specifications

A conceptual model of a TMS system configuration is shown in Figure 1. This model maps the functional specifications into a modular architecture of a TMS. A TMS is modelled as a collection of eight distinct but interacting modules, where a module corresponds to a management area in the functional specifications, and with an additional module for special reports. The special reports module is intended for nonstandard reporting which may require a consolidation of data from several

system modules. Standard reporting is designed as a built-in function of each module. This is only an option, and all the reporting can be relegated to the reports module if desired. Each module is designed to be totally self-contained, in the sense that it is entirely responsible for its databases, internal data management, input/output and interfacing. Such an architecture can be implemented within the overall environment of a Data Base Management System (DBMS) by interfacing the DBMS at the backend of each module and preserving the access privacy of its exclusive databases. The interfaces between the modules actually represent the functional interfaces in the training organization, and are shown in Figure 1. A superimposition of the functional responsibilities of the organizational units on the TMS architecture yields the total picture. This has been illustrated in Figure 1 by using the relevant organizational units of the 542nd CTW. A similar mapping can be derived for any training organization. The dashed lines represent the access to TMS modules by the organizational units.

The system configuration of each module is modelled in Figure 2. This configuration is the same for each module, and provides a blueprint for independent module development and interfacing. The identical but independent structures of the modules provide the required standardization in system development and the flexibility/independence in long-term system maintenance. The structural specifications for the system modules consist of the following: *Front-end user interface manager, Access control manager, responsible for data security/user access authorization, Interactive query processor, Process automation submodules, Decision support submodules, incorporating man-machine problem solving models, Standard-form report generator, Back-end database manager, Front-end inter-module interface manager, Module databases/access control databases/system libraries, Module controller, responsible for co-ordinating all the other submodules.*

Replacing each system module with its system configuration in Figure 1 yields the complete picture of the total TMS configuration. In the above architecture, each module is directly interfaced with the users and other system modules. Alternate approaches to this are also possible. For example, a TMS can be organized in a three-level hierarchy of modules with a TMS controller at the first level, the module controllers of all the modules at the second level, and the remaining submodules at the third level. In this case, the interface co-ordination between the modules can be routed through this hierarchy, with the TMS controller serving as the co-ordinator. This can be contrasted with the architecture in Figures 1 and 2, which is a two-level hierarchy without a TMS controller. The idea of a TMS controller is to draw out the interface management from the system modules and assign it to a central module, so that the system modules can be modelled as simply processing their inputs and transmitting their outputs. The two-level architecture is probably more suitable for smaller systems where significant processing efficiency may be achieved without introducing an additional control module. However, when systems grow in size and capabilities, the hierarchy also tends to expand, due to the need for greater system control and co-ordination.

This strongly emphasizes the need for flexibility and modularity in designing a TMS right from the beginning, so that transitions to higher levels of system management can be achieved relatively easily when systems grow.

The modular hierarchy is also determined by the physical hardware environment, the user needs and their locations. In general, the current trend is to house a TMS in a client-server environment connecting hundreds of users spread over several building areas. However, mainframe driven, multiprocessing/multitasking environments are also equally conceivable. A specification of the hardware, software and system architectural parameters to be taken into account in designing a TMS is as follows.

Hardware Parameters

1. Hardware Environment
    - Computing systems used
        - LAN, workstations, mainframes, etc.
    - Communication network
        - Topology, medium, access control protocols
2. User Location Map
    - Network path
    - Network access from locations
3. Hardware Capabilities
    - Servers/mainframes control
        - Location
        - Computing systems served
    - Servers/clients/mainframes capacities
        - Storage capacities
        - Processing speeds
        - Transmission capacities/speeds
        - Multiprocessing/multitasking capabilities
        - Specialized features (like windows, etc.)
    - Network capacities
        - Transmission capacities/speeds

- Collision control/management
- Error control/management
- Hardware support capacities

Software Parameters

1. Software System Structure
    - Modular hierarchy
    - Interfacing requirements
2. Modular Support Features
    - User access control
    - On-line query processing
    - Process automation
    - Decision support
    - Report generation
    - Database management
    - Inter-modular data traffic control
3. Modular Support Requirements
    - Database volume
    - Query traffic
    - Standard reports - volume, periodicity
    - Nonstandard reports - expected volume, expected periodicity
    - Automation - expected workload, traffic
    - Decision support - expected workload, traffic
    - Database update - expected workload, traffic
    - Intermodular traffic - expected workload

Architectural Parameters

1. Server/Client Hierarchy
    - Main server/client servers/clients architecture
2. Module Location Architecture

- Main server centralized
- Main server centralized, client server replicated
- All servers distributed with no replication
- Combination of centralized/replicated/distributed location architecture

3. Database Location Architecture

- Main server centralized
- Main server centralized, client server replicated
- All servers distributed with no replication
- Combination of centralized/replicated/distributed location architecture

4. Processing Location Architecture

- Main server processing, output transmission via client server
- client server processing, direct output transmission to client
- local client processing
- Combination of main server/client server/local processing

The architectural parameters specify the overall system configuration. This configuration should be tailored to a training organization by an assessment of the compatibility between the underlying hardware and software parameters. Such a configuration should serve the following major objectives: (i) Maintenance of appropriate database access security, currency, integrity and consistency among replicated databases, (ii) Minimization of query response times and interactive database update times, (iii) Minimization of the automation and decision support process times, (iv) Efficient management of inter-modular data traffic, (v) Flexibility in reporting, (vi) Maximization of network/system resource capacity utilization under an efficient system configuration, and (vii) Flexibility to handle changes in system architecture, modular structure and user requirements.

2.3 Process Specifications

The process specifications address the functional support features of a TMS. These specifications provide a frame of reference for logical system design, and are based on the functional tasks in the organization, resources available, and the recurrent/transient user requirements and expectations. In these specifications, a TMS is viewed as a collection of *processes* by the system modules in support of the user requirements. Typically, each component of a module is responsible for specific system processes, and there are seven such processes: ***access control, query processing, automation, decision-support, report generation, database management*** and ***internal data transfers***. Access control, query

processing and report generation are specific to the user requirements. Database management and internal data transfers are specific to the system environment. Automation and decision support derive from the functional specifications. A summary specification of these processes for a TMS in general is as follows.

Module: Curriculum Management

Major Processes:

- Analysis of task listings/objective hierarchies
- Course structure design and analysis
- Lesson plan production
- Lesson material production planning/tracking
- Curricular resource requirements analysis
- Curricular project planning/scheduling/tracking
- Courseflow structure document production
- Classflow plan production
- A & E - sampling/estimation/testing

Major Databases:

- Master task listings
- Objective hierarchies
- Course flow plans
- Lesson specifications
- Classload plans/projections

Data Interfaces:

- Resource, Logistics, Scheduling, Configuration, Reports, Student Performance modules.

Module: Configuration Management

Major Processes:

- Automated course document storage, management and access
- Access services help facility
- Course document catalogue/library management
- A & E - sampling/estimation/testing

Major Databases:

- Course structure
- Lesson plans
- Lesson material
- Course material access services library

Data Interfaces:

- Curriculum, Reports modules

Module: Scheduling Management

Major Processes:

- Production of short-term schedules
- Reporting on co-ordination requirements
- Resource availability verification
- Resource requirements specifications
- Scheduling requirements specifications
- Interactive fine-tuning of daily schedules
- Schedule deconfliction/finalization/announcement
- A & E - sampling/estimation/testing

Major Databases:

- Class schedules (short-term/daily)
- Individual schedules (short-term/daily)
- Program schedules (short-term/daily)
  - Flight, static, ground, simulator, academic, meetings
- Resource schedules (short-term/daily)
  - Aircraft, simulator, classroom, outside agencies

Data Interfaces:

- Resource, Logistics, Curriculum, Reports, Student Performance, Administrative, Reports modules.

Module: Logistics Management

Major Processes:

- Spare parts inventory monitoring/planning/procuring/management
- Preparation of preventive maintenance schedules
- Configuration availability reporting

- Time-slot availability reporting
- A & E - sampling/estimation/testing

Major Databases:

- Aircraft configurations
- Spare parts inventory - levels, management policies, models
- Maintenance Schedules
- Maintenance Logs
- Aircraft configuration/time-slot availability

Data Interfaces:

- Curriculum, Scheduling, Reports modules

Module: Resource Management

Major Processes:

- Human resources inventory maintenance/update
- Equipment resources inventory maintenance/update
- Facilities inventory maintenance/update
- Resource availability tracking/reporting
- Equipment/facilities preventive maintenance scheduling
- A & E - sampling/estimation/testing

Major Databases:

- Human resources - qualifications
- Human resources - availability
- Equipment resources - configurations
- Equipment resources - availability
- Facilities resources - configurations
- Facilities resources - availability
- All maintenance schedules

Data Interfaces:

- Curriculum, Scheduling, Administrative, Reports modules.

Module: Student Performance Management

Major Processes:

- Student folders maintenance/update

- Student performance tracking/analysis
- Training requirements assessment/program planning
- A & E - sampling/estimation/testing

Major Databases:

- Student folders

Data Interfaces:

- Curriculum, Scheduling, Administrative, Reports modules.

Module: Administrative Management

Major Processes:

- Personnel records maintenance/update
- Security/other requirements flagging
- Administrative reporting
- A & E - sampling/estimation/testing

Major Databases:

- Personnel records

Data Interfaces:

- Curriculum, Resource, Scheduling, Reports, Student Performance modules.

3 Conclusion

Aircrew training is an important function in the Air Force. The total mission-readiness of the aircrews depends on the effectiveness of the training programs at both the formal training schools and the field units of the Air Force. Training is a complex and large operation, and the management of a training school involves a considerable resource committment and co-ordination of numerous activities. The growing complexity of the training programs and the warfare technology have continuously underscored the importance of system integration in the management of training programs. Training Management Systems (TMS) are products of this doctrine, and are conceptually intended to provide this integration by efficient and cost-effective computer support of the various functions in a training organization. The goals of a TMS are to: (i) capture, store and manage all the important data generated in the day-to-day operations in a training program, (ii) provide appropriate information whenever and wherever needed in the program, (iii) provide intelligent decision support and analytical tools to facilitate the decision processes, and (iv) integrate all the major operations in the system in such a manner that the information, operations and decisions generated at an

organizational unit are apparent and accessible to all the other relevant and functionally connected units.

The current research is probably a first attempt at formalizing the objectives and functions of a TMS. The interpretations of a TMS in the Air Force are several, and this study is an attempt to unify these perceptions in a solid system framework. This is achieved by developing an architectural model of a TMS in general, and detailing all the important functional, system and process specifications for a TMS. This model is illustrated with a case study, using the helicopter training programs at the 542nd CTW, Kirtland AFB. A TMS architecture has been developed for this wing and all the relevant specifications for some of the TMS subsystems for this wing have been developed. A comprehensive strategy for the next logical steps in the development of a TMS for this wing has been developed. The proposed strategy is intended for cost-effective and successful TMS development and implementation at this wing.

This report is a blueprint for TMS design and development in general. The blueprint is illustrated with a specific case study using the 542nd CTW system, and we expect this to serve as a standard or a frame of reference for all TMS development in the future.

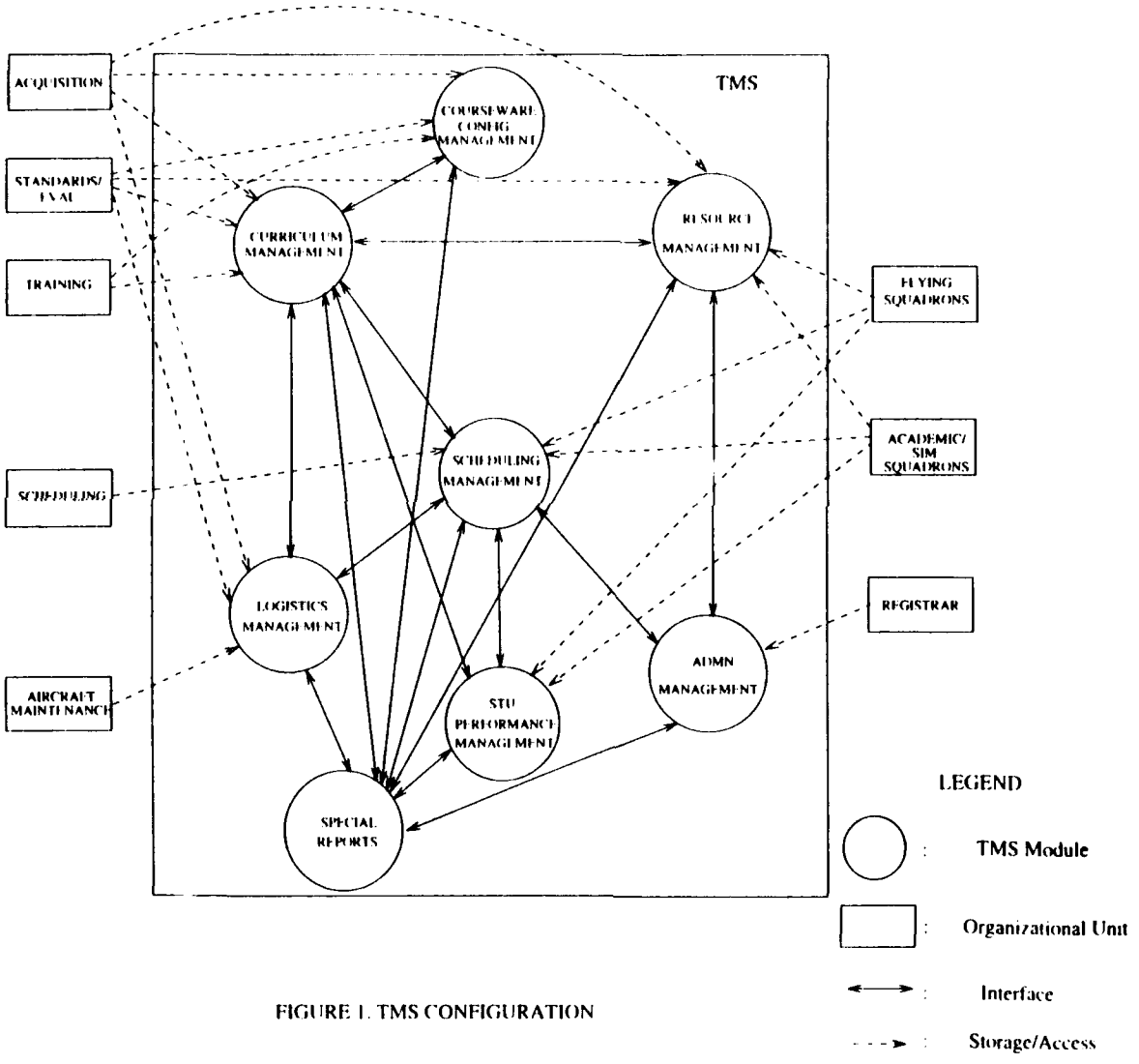


FIGURE 1. TMS CONFIGURATION

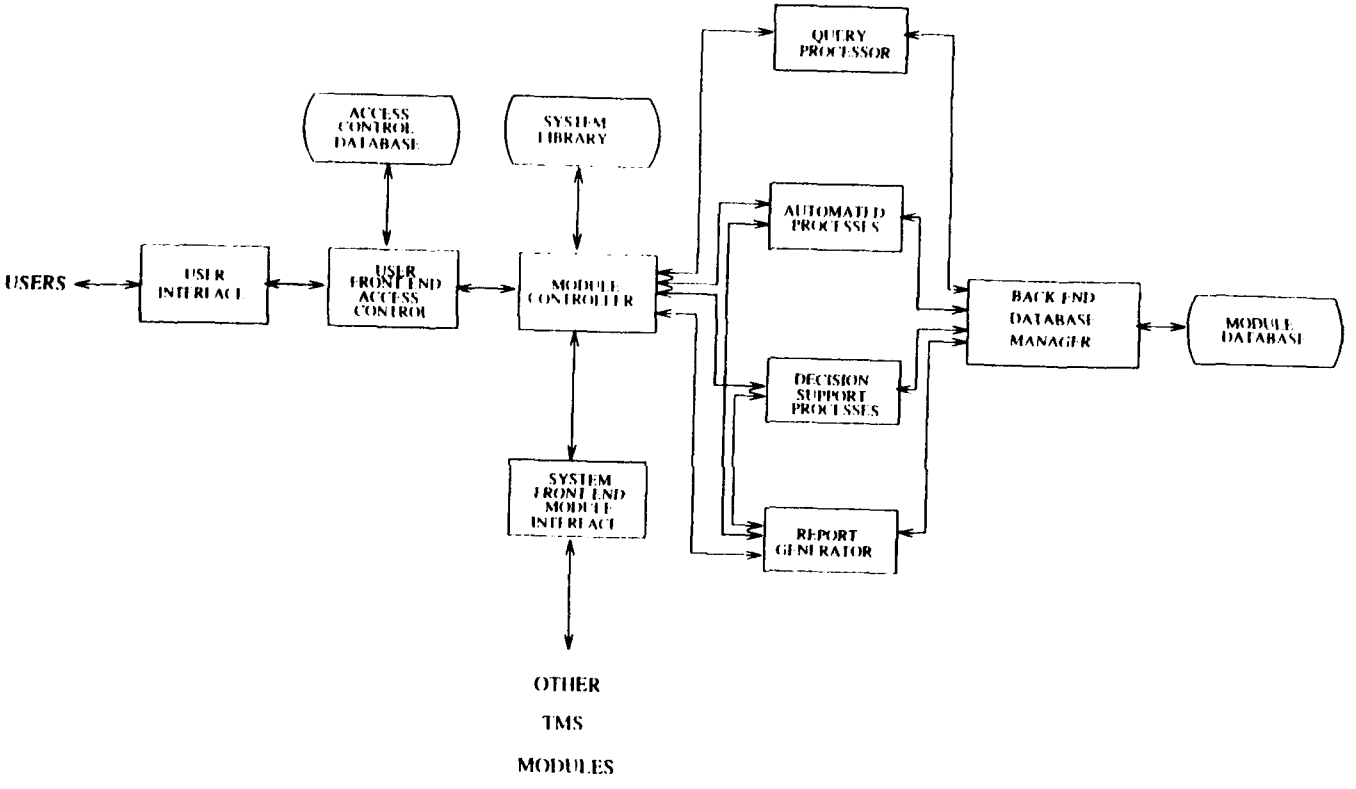


FIGURE 2. MODULE CONFIGURATION

References

[1] Bruce, P. D., Killion, T. H., Rockway, M. R. and Povenmire, H. K., "B-52 and KC-135 Mission Qualification and Continuation Training: A Review and Analysis," (AL-TR-1991-0010), Williams AFB, AZ: Human Resources Directorate, Aircrew Training Research Division, 1991.

[2] Bruce, P. D., "Aircrew Training Evaluation: B-52 and KC-135 Formal School Training," (AFHRL-TR-88-49), Williams AFB, AZ: Human Resources Directorate, Aircrew Training Research Division, 1989.

[3] Dukes, R., Rockway, M. R. and Nullmeyer, R. T., "Lessons Learned in the Development of the C-130 Aircrew Training System: A Summary of Air Force On-Site Experience," Williams AFB, AZ: Human Resources Directorate, Aircrew Training Research Division, 1991.

[4] Fishburne, R. P., Spears, W. D. and Williams, K. R., "Design Specification Development for the C-130 Model Aircrew Training System: Final Report," (AFHRL-TR-86-51), Williams AFB, AZ: Human Resources Directorate, Aircrew Training Research Division, 1987.

[5] McGann, E. J. J., Rooney, J. and Rogers, M., "Reference Book for the C-130 Aircrew Training System," Little Rock AFB, AR: 34th Tactical Airlift Training Group, 1987.

[6] Nullmeyer, R. T., Bruce, P. D. and Rockway, M. R., "Integrated Aircrew Training Management Information Systems: An Organizational Perspective," *Proceedings of the* 14th *Interservice/Industry Training Systems Conference*, 1992.

[7] Reakes, M., "Requirements of a Training Management System for Aircrew Training," *Proceedings of the* 11th *Interservice/Industry Training Systems Conference*, 1989.

[8] Rockway, M. R. and Nullmeyer, R. T., "Aircrew Contracted Instruction: Major Issues and Lessons Learned," Williams AFB, AZ: Human Resources Directorate, Aircrew Training Research Division, 1991.

PREDICTING EASE OF MOVEMENT BETWEEN AIR FORCE SPECIALTIES

Stephen A. Truhon
Associate Professor
Department of Social Sciences
Winston-Salem State University
601 Martin Luther King Jr. Drive
Winston-Salem, NC 27110

Final Report for:
AFOSR Summer Research Program
Armstrong Laboratory

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

July 1993

Running Head: Ease of Movement

PREDICTING EASE OF MOVEMENT BETWEEN AIR FORCE SPECIALTIES

Stephen A. Truhon
Associate Professor
Department of Social Sciences
Winston-Salem State University

ABSTRACT

The current study focused on the prediction of ease of movement between 43 Air Force specialties. Nine possible predictors were examined, but only four variables, two dealing with job difficulty and two dealing with job similarity, consistently explained between 30 and 32 percent of the variance. The regression equations suggest that it is easier to move to a specialty that is lower in difficulty but similar to one's current specialty. Examination of individual specialties revealed that these predictors worked best when examining ease of movement from a specific specialty to other specialties rather than from other specialties to a specific specialty. Exceptions to these findings are noted.

PREDICTING EASE OF MOVEMENT BETWEEN AIR FORCE SPECIALTIES

Stephen A. Truhon

INTRODUCTION

Personnel psychologists have long been concerned with the clustering of individuals into jobs and similar jobs into higher levels of job families (Harvey, 1986). Such clustering can be useful in many ways, including helping individuals move between jobs with minimal retraining. This issue has been especially important to the Air Force as the number of personnel decline and, consequently, specialties need to be combined.

One such study that examined the similarity of Air Force specialties (AFS's) was the Ease of Movement (EOM) study (Mayfield & Lance, 1988). As part of this study, personnel in 43 AFS's, ranging in number from five to 60 per specialty, were asked to rate the amount of retraining necessary to move from their AFS into other AFS's (TO) and into their AFS from other AFS's (FROM) on a nine- point scale (1= very small amount of retraining required; 9= very large amount of retraining required. Subjects were asked to mark "C" if they could not determine the amount of retraining required, and "X" for the amount of retraining required from their own specialty to their own specialty.). Using clustering techniques such as ADDTREE (Corter & Tversky, 1986), Truhon (1993) found support for the taxonomy of AFS's which the Air Force currently uses, with some exceptions.

Nevertheless, clustering techniques have been criticized for simply being hypothesis generators (Blashfield & Aldenderfer, 1988; Gordon,1981). Researchers have typically analyzed similarity data by cluster analysis as a heuristic to understand the structure of the original data. While there is nothing wrong with such an

approach, such cluster analysis is a form of exploratory data analysis. Clusters obtained from previous analyses should be confirmed and explained. (This argument is similar to that made in criticizing traditional factor analysis as exploratory, while viewing structural equations models as confirmatory factor analysis [e.g., Connell, 1987].)

The purpose of the current study was to gain some understanding of the factors predicting job similarity. To do this data from the EOM study was used. The similarity of each AFS to each of the 42 other AFS's was obtained for both the TO and FROM data sets. Because the similarities were asymmetrical, a 43 x 42 matrix (i.e., 1806 values) was derived from each data set. These 1806 values were treated as the dependent variables in these analyses. To achieve comparability between the data sets the matrix of FROM values was transposed.[1]

One minor change was made in the 43 AFS's from those used in the EOM study to the present study. Aircraft Electrical Systems Specialist (423X0 in the EOM study) had been reclassified as Electrical and Environmental Systems Specialist (452X5). The 43 AFS's are listed in Table 1.

PREDICTOR VARIABLES

The following nine variables were considered in predicting the perceived similarity of AFS's.

Occupational Learning Difficulty (OLD): A measure of learning difficulty (Mumford, Weeks, Harding and Fleishman, 1987) was obtained for each of the 43 AFS's. The difference between each AFS's learning difficulty and the other AFS's was

---

[1] In the TO data set, item ij refers to the ease of movement from AFS_i to AFS_j. In the FROM data set, item ij refers to the ease of movement from AFS_j to AFS_i.

TABLE 1

Forty-three Specialties Compared in the Present Study

Specialty Number	Description
113X0C	Flight Engineer
114X0	Aircraft Loadmaster
207X1	Morse Systems Operator
241X0	Safety Specialist
242X0	Disaster Preparedness Specialist
251X0	Weather Specialist
271X1	Airfield Management Specialist
272X0	Air Traffic Control Operator
276X0	Aerospace Control & Warning Systems Operator
303X2	Aircraft Control & Warning Radar Specialist
304X0	Wideband Communications Equipment Specialist
304X4	Ground Radio Communications Specialist
324X0	Precision Measurement Equipment Laboratory Specialist
452X4	Tactical Aircraft Maintenance Specialist
452X5	Electrical and Environmental Systems Specialist
454X1	Aerospace Ground Equipment Mechanic
455X2	Avionic Communications Specialist
457X0	Strategic Aircraft Maintenance Specialist
461X0	Munitions Systems Specialist
462X0	Aircraft Armament Systems Specialist
472X1	Special Vehicle Mechanic
542X0	Electrician
545X0	Refrigeration and Air Conditioning Specialist
551X0	Pavements Maintenance Specialist
551X1	Construction Equipment Operator
553X0	Engineering Assistant Specialist
571X0	Fire Protection Specialist
603X0	Vehicle Operator/Dispatcher
631X0	Fuel Specialist
645X0	Inventory Management Specialist
645X1	Materiel Storage & Distribution Specialist
645X2	Supply Systems Specialist
651X0	Contracting Specialist
661X0	Logistics Plans Specialist
702X0	Administration Specialist
732X0	Personnel Specialist
741X1	Fitness & Recreation Specialist
791X0	Public Affairs Specialist
811X0	Security Specialist
811X2	Law Enforcement Specialist
902X0	Medical Service Specialist
906X0	Medical Administrative Specialist
981X0	Dental Assistant

calculated. Because it was hypothesized that more difficult AFS's were more similar to less difficult AFS's than vice versa, the algebraic difference, rather than the absolute value of the difference, was calculated. As a result, these data were asymmetrical.

Mechanical, Administrative, General, and Electronics (MAGE): Since the mid-1950's the Air Force has classified AFS's into the categories of Mechanical (M), Administrative (A), General (G) and Electronic (E) (Weeks, Mullins, and Vitola, 1975). Each of the 43 AFS's had previously been categorized into one or more of the above. Two AFS's with the same code(s) were considered similar (i.e., a value of 1), while two AFS's with different codes were considered dissimilar (i.e., a value of 0). Fractional values were possible for AFS's with more than one code, but asymmetries were also possible. For example, if AFS_i were coded M and E and AFS_j were coded as M, then the similarity of i to j was valued at 1, but the similarity of j to i was valued at .5.

Functional Account Code (FAC) and Program Element Code (PEC): Each AFS is described by what activities personnel perform in that specialty, how many personnel perform that activity, and the amount of time spent on that activity. The difference between FAC and PEC lies in that FAC emphasizes the duty performed, while PEC emphasizes the area (e.g., the type of forces or weapon system) where the duty is performed. The overlap (in percent) between AFS's with respect to FAC and PEC were calculated. Asymmetries were possible, because the overlap value is relative to the number of FAC's (or PEC's) assigned to the members of each AFS. Thus, AFS_i and AFS_j will have a certain number of common FAC's (or PEC's), but the overlap between i and j and the overlap between j and i will differ if the total number of FAC's (or PEC's) in AFS_i and in AFS_j differ.

Diversity of Activities (DIV): The diversity of each AFS was determined by calculating the sum of FAC's and PEC's used to describe that AFS. The difference between each AFS's diversity and the other AFS's was calculated. Because it was possible that movement from more diverse AFS's to less diverse AFS's is not the same as movement in the opposite direction, the algebraic difference, rather than the absolute value of the difference, was calculated. As a result, these data were asymmetrical.

Amount of Technical Knowledge (TECH): Each AFS was categorized as requiring technical knowledge or not. All mechanical and electronic AFS's were considered technical, all administrative AFS's were considered nontechnical, and general AFS's were categorized as technical or not on an AFS-by-AFS basis. Two AFS's that were both technical (or nontechnical) were considered similar (i.e., a value of 1), while a technical and a nontechnical AFS were considered dissimilar (i.e., a value of 0). These data were symmetrical.

Number of Personnel (NUM): The number of personnel in each AFS was determined by using the total number listed in the FAC and PEC analyses. While this number may not be totally accurate, the true values should be proportional to the numbers used. The ij and ji differences were calculated for all possible pairs of AFS's. These data were asymmetrical because positive and negative values were possible.

Amount of Training (TRAIN): The average amount of resident school training (in weeks) for each AFS was determined based on the mean training times for fiscal year 1989 (close to the date of the EOM study) at the basic level (i.e, 3-skill level). There were five AFS's for which the amount of training was not available: Flight Engineer (113X0C), Airfield Management Specialist (207X1), Vehicle

Operator/Dispatcher (603X0), Materiel Storage & Distribution Specialist (645X1), and Fitness & Recreation Specialist (741X1). The missing values for these AFS's were calculated by regressing the amount of training on occupational learning difficulty, number of FAC codes, number of PEC codes, and number of personnel. The ij and ji differences were calculated for all possible pairs of AFS's. These data were asymmetrical because positive and negative values were possible.

Complexity (COMPLEX): The complexity of the 43 AFS's was rated on a nine-point scale (1 = low complexity; 9 = high complexity) by the senior occupational analyst at the U.S. Air Force Occupational Measurement Squadron, who was familiar with these AFS's. He was unable to rate the complexity of Morse Systems Operator (207X1), which is a secret AFS. The missing value for this AFS was calculated by regressing complexity on occupational learning difficulty, number of FAC codes, number of PEC codes, number of personnel, and amount of training. The ij and ji differences were calculated for all possible pairs of AFS's. These data were asymmetrical because positive and negative values were possible.

CORRELATIONAL ANALYSES

The correlations between the nine predictor variables and the two measures of ease of movement similarity (TO and FROM) were calculated, with the results presented in Table 2.

The correlations reveal that the two measures of similarity are highly correlated but are not identical ($r = .6658$). FAC and PEC are highly correlated ($r = .5030$) and are moderately correlated with DIVERSE (r's = .3533 and .3502 respectively), which is derived from FAC and PEC. It is also notable that TECH has zero correlations with five variables.

TABLE 2

Correlations of Measures of Similarity and Predictors

	TO	FROM	OLD	MAGE	PEC	FAC	DIV
FROM	.6658						
OLD	.3368	.3912					
MAGE	-.2514	-.2982	-.0016				
PEC	-.3297	-.2699	-.1156	.1012			
FAC	-.2624	-.2356	-.1289	.0918	.5030		
DIV	-.1860	-.0837	-.2787	.0313	.3502	.3533	
TECH	-.1764	-.2317	.0000	.3230	.1176	.1080	.0000
NUM	-.0723	-.1370	-.1370	.0334	.2184	.2126	.6075
TRAIN	.4031	.3475	.4892	-.0178	-.0662	-.0568	-.1747
COMPLEX	-.0055	-.1121	-.1050	.0227	.1740	.1508	.4144

	TECH	NUM	TRAIN
NUM	.0000		
TRAIN	.0000	-.2147	
COMPLEX	.0000	.3220	.0406

REGRESSION ANALYSES

The nine predictor variables were regressed on the two measures of similarity (TO and FROM). With approximately 1800 degrees of freedom, it is possible for a predictor variable to account for a statistically significant but practically small amount of variance. It is also possible for a predictor variable to account for a significant amount of variance for one measure of similarity but not for the other. To avoid these problems, predictor variables were retained if they accounted for at least one percent of variance for *each* measure of similarity. Following this guideline, four of the predictor variables were retained: OLD, MAGE, PEC, and TRAIN.

The following raw score equations were derived from the regression analyses:

TO = .04064 * TRAIN - .75409 * MAGE - .020061 * PEC + .00834637 * OLD + 5.81173

(multiple R = .56499; R^2 = .31921; adjusted R^2 = .31770; F [4, 1801] = 211.11816; p < .0001); and

FROM = .02161 * TRAIN - .78084 * MAGE - .01249 * PEC + .01199 * OLD + 6.14513

(multiple R = .55655; R^2 = .30975; adjusted R^2 = .30821; F [4, 1801] = 202.04634; p < .0001)[23]

As can be seen, the two regression equations are quite similar. The regression weights in the two equations are fairly similar and the amounts of variance accounted for are about equal (between 30 and 32 percent). When the equation for TO was applied to FROM, a correlation of .5396 occurred. When the equation for FROM was applied to TO, a correlation of .5478 occurred. In both cases the correlations are only slightly lower than predicted by the appropriate regression equations.

Considering how the variables are coded, these equations suggest that more difficult ease of movement (i.e., higher TO and FROM scores) is associated with greater differences in the amount of training and occupational learning difficulty (i.e., particularly in moving from specialties with lower levels of training and difficulty to those that are higher) and lower amounts

---

[2] The relationships in these two equations seem to be linear. A neural network analysis of these variables (cf. Wiggins, Engquist, and Looper, 1992), which can detect nonlinear relationships, found a correlation of .6395 with R^2 of .3649 for TO and a correlation of .6697 with R^2 of .3973 for FROM. In addition, including squares and interactions of these variables in a regression analysis failed to find one that accounted for at least one percent of the variance in both equations.

[3] The partial correlation coefficients (which is the correlation between each independent variable and the dependent variable when the effects of other independent variables are partialled out) were as follows: for the TO equation TRAIN .3067, MAGE -.25470, PEC -.30693, and OLD .16208; for the FROM equation TRAIN .20200, MAGE -.31192, PEC -.22894, and OLD .27317 (all p's < .0001).

of overlap in the MAGE and PEC codes. To put this in terms of ease of movement, these equations say personnel will have an easier time moving between their current AFS and a proposed AFS if the current AFS is more difficult and requires more training than the proposed AFS, and if the two AFS's are similar in MAGE and PEC codes.

EXCEPTIONS TO REGRESSION FINDINGS

Although the regression equations applied in general to the relationship between AFS's, it is possible for significant inaccuracy to occur when they are used to predict ease of movement to or from specific AFS's. Therefore, the general equations were applied to the individual AFS's to determine how well the equations fit each AFS's TO and FROM similarity values.

The correlations from each specialty to other specialties (TO) as predicted by these formulas and to each specialty from other specialties (FROM) are presented in Tables 3 and 4, respectively.

As can be seen in Table 3, the regression equations fit most of the AFS's reasonably well with the following exceptions: Flight Engineer (113X0C), Aircraft Control & Warning Systems Operator (303X2), Wideband Communications Equipment Specialist (304X0), Ground Radio Communications Specialist (304x4), Precision Measurement Equipment Laboratory Specialist (324X0), Electrical and Environmental Systems Specialist (452X5), Aerospace Ground Equipment Mechanic (454X1), and Avionic Communications Specialist (455X2).

Examination of Table 4 reveals the regression equations do not provide as good a fit as they do in Table 3. This suggests that predicting ease of movement works better if one knows the AFS one is moving *from* than the AFS one is moving *to*.

TABLE 3

Correlations between Actual and Predicted Ease of Movement from Specific AFS's to Other AFS's

Specialty	TO	FROM
113X0C	-.0012	.3437
114X0	.5287	.5801
207X1	.5629	.4705
241X0	.6494	.6422
242X0	.7028	.7119
251X0	.5143	.5939
271X1	.6960	.5129
272X0	.5224	.6439
276X0	.5333	.5093
303X2	.3528	.3608
304X0	.4038	.2367
304X4	.1666	.2242
324X0	.3422	.1572
452X4	.6165	.3523
452X5	.4010	.1343
454X1	.4113	.3525
455X2	.2650	.1113
457X0	.5019	.3224
461X0	.4791	.5018
462X0	.5785	.5294
472X1	.6580	.6103
542X0	.3122	.4363
545X0	.5739	.5471
551X0	.6231	.6482
551X1	.6377	.6563
553X0	.5116	.4972
571X0	.4162	.4994
603X0	.7475	.6702
631X0	.7241	.6485
645X0	.7419	.6629
645X1	.6518	.6596
645X2	.7997	.7050
651X0	.8351	.7758
661X0	.8156	.6956
702X0	.6671	.7024
732X0	.7065	.7278
741X1	.7366	.5249
791X0	.7212	.7592
811X0	.5419	.6089
811X2	.6838	.6581
902X0	.6402	.7029
906X0	.5851	.6609
981X0	.6255	.6609

TABLE 4

Correlations between Actual and Predicted Ease of Movement to Specific AFS's from Other AFS's

Specialty	TO	FROM
113X0C	.4234	.2888
114X0	.5530	.2215
207X1	.3018	.3489
241X0	.2542	.2782
242X0	.3376	.3368
251X0	.4332	.4770
271X1	-.1563	.0157
272X0	.4471	.4309
276X0	.6255	.4715
303X2	.6639	.8691
304X0	.7384	.8939
304X4	.7456	.8784
324X0	.6936	.8143
452X4	.6672	.6552
452X5	.7456	.7697
454X1	.7158	.7287
455X2	.6822	.8144
457X0	.5864	.6384
461X0	.6196	.3677
462X0	.6383	.7404
472X1	.4392	.6595
542X0	.7808	.5089
545X0	.6903	.5710
551X0	.5539	.4800
551X1	.5730	.3922
553X0	.4647	.4130
571X0	.2574	.2110
603X0	.3400	.3284
631X0	.3448	.3464
645X0	.2005	.3650
645X1	.2342	-.0921
645X2	.0470	.1210
651X0	.0585	.0439
661X0	.0612	.2178
702X0	.0199	.0348
732X0	.0471	.0073
741X1	-.0403	.1344
791X0	.1279	.0683
811X0	.1213	.2303
811X2	-.1474	.1463
902X0	.4648	.2008
906X0	.3250	-.1372
981X0	.4159	.1132

FINDING REGRESSION EQUATIONS FOR THE EXCEPTIONS

Because the overall regression equations failed to fit some AFS's, an attempt was made to find equations that would best fit these specialties. All nine predictors were used for possible regression analysis but only those explaining significant amounts of variability and common to both equations were retained[4].

It was noted that most of the exceptions in Table 3 are communications specialties (303X2, 304X0, 304X4, 324X0, and 455X2). These specialties have also clustered together in previous research (Truhon, 1993). Following the above guidelines only two variables were retained: MAGE and FAC. The resulting equations are listed below:

TO = -.71523 * MAGE - .02593 * FAC + 4.7265

FROM = -.48891 * MAGE - .02744 * FAC + 5.66591

These two equations suggest that similarity of AFS's is more important than difficulty in predicting ease of movement from these specialties to other specialties. The correlations of these equations to the ease of movement measures ranged from .3529 to .5769 for the TO equation and from .3015 to .4522 for the FROM equation, suggesting from comparable to better fit with these equations.

The exceptions in Table 4 formed six groups based upon previous clustering (Truhon, 1993): safety specialties (241X0, 242X0, and 571X0), vehicle specialties (603X0 and 631X0), inventory, contracting and logistics specialties (645X0, 645X1, 645X2, 651X0, 661X0), administrative specialties (702X0, 732X0, 741X1, and 791X0), security specialties (811X0 and 811X2), and medical specialties (902X0, 906X0, and 981X0). Separate regression

---

[4] Because performing several regression equations on the same data set complicates calculating significance level, the significance of the following regression equations is not reported.

analyses were done for each group of specialties, again retaining only those variables explaining significant amounts of variance in both equations.

For the safety specialties (241X0, 242X0, and 571X0) only FAC was retained, but it explained so little of the variance (less than 10 percent) that setting up a regression for these specialties was dropped.

For the vehicle specialties (603X0 and 631X0) FAC, DIV, and NUM were retained. The resulting equations are listed below:

TO = .0001061429 * NUM - .08149 * FAC - .00258125 * DIV + 3.643

FROM = .000076257 * NUM - .11987 * FAC - .00160969 * DIV + 4.4213

Not much significance should be attached to these regression equations. The correlations of these equations for Vehicle Operator/Dispatcher (603X0) and Fuel Specialist (631X0) to the ease of movement measures were, respectively, .4780 and .3815 for the TO equation and .3815 and .1018 for the FROM equation, suggesting comparable to worse fit with the these equations.

For the inventory, contracting, and logistics specialties (645X0, 645X1, 645X2, 651X0, 661X0) OLD, MAGE, TECH, and NUM were retained. The resulting equations are listed below:

TO = -.48838 * TECH - .000030402 * NUM + .00929303 * OLD - .33479 * MAGE + 5.04218

FROM = -.61552 * TECH - .0000377773 * NUM + .02125 * OLD - .76058 * MAGE + 6.36385

Only slightly better fit occurred with these equations. The correlations of these equations to the ease of movement measures ranged from .0624 to .2553 for the TO equation and from -.0594 to .3735 for the FROM equation.

For the administrative specialties (702X0, 732X0, 741X1, and 791X0) OLD, PEC, DIV, TECH, NUM, and COMPLEX were retained. The resulting equations are listed below:

TO = .13556 * COMPLEX - .88247 * TECH + .008721197 * OLD + .006413586 * PEC - .0000240542 * NUM + .0005020555 * DIV + 4.0163

FROM = .06471 * COMPLEX - 1.25562 * TECH + .01594 * OLD + .00689178 * PEC - .0000605757 * NUM + .00139283 * DIV + 5.54819

For the most part marked improvement in the correlations resulted, with the exception of the correlation between ease of movement for Fitness & Recreation Specialist (741X1) and the TO equation (r = .1761). Ignoring this exception, the correlations ranged from .4633 to .6095 for the TO equation and from .5383 to .7215 for the FROM equation.

These specialties are average to above average in complexity, are nontechnical, and are low in difficulty. The above equations suggest that personnel will have easier movement to these AFS's if their current AFS is also nontechnical, of similar or higher complexity, and is more difficult.

For the security specialties (811X0 and 811X2) only FAC was retained but it did account for significant variability. The resulting equations were as follows:

TO = -.09991 * FAC + 4.50974

FROM = -.14552 * FAC + 4.91178

The correlations of these equations for Security Specialist (811X0) and Law Enforcement Specialist (811X2) to the ease of movement measures were, respectively, .3181 and .3663 for the TO equation and .7660 and .7526 for the FROM equation, suggesting moderate to substantial improvement in fit with the these equations.

For the medical specialties (902X0, 906X0, and 981X0) OLD, FAC, DIV, and TRAIN were retained. The resulting equations were as follows:

TO = .03413 * TRAIN - .20515 * FAC + .0002794635 * DIV - .01755 * OLD + 5.83379

FROM = .01207 * TRAIN - .12643 * FAC + .0007563513 * DIV - .00906764 * OLD + 5.07649

The correlations of these equations to the ease of movement measures ranged from .5352 to .6940 for the TO equation and from .5930 to .7825 for the FROM equation, suggesting better fit with these equations. These specialties have moderate to high levels of training, are moderately diversified, but are regarded as relatively easy. The above equations suggest that personnel will easier movement to these AFS's if their current AFS required more training, is more diversified, but is similar in activities. The negative value for OLD suggests a correction on TRAIN, with which it is highly correlated.

PREDICTING OCCUPATIONAL LEARNING DIFFICULTY

Of the various measures used in the above equations, occupational learning difficulty (OLD) is the most widely used and most difficult to generate, if it is not already available. While most of the other measures can be easily determined from the Uniform Airman Record and the Technical Training- Graduates and Eliminees Report, OLD must be estimated from bench-marked task difficulty ratings. An attempt was made in this study to estimate OLD by an equation using more readily available variables.

For each of the 43 AFS's the following variables were available: OLD, the number of Functional Account Codes (NFAC), the number of Program Element Codes (NPEC), the number of personnel in the AFS (NUMBER), the number of weeks of training required (WEEKS), and the complexity of the AFS as determined by a subject matter expert (COMPLEXITY). Means were substituted for missing values. The correlations between these variables are shown in Table 5. The regression of OLD on the remaining variables was performed. All predictor variables were first entered into the equation. A predictor was subsequently removed if it did not significantly

improve the equation. As a result, NUMBER and COMPLEXITY were removed. The resulting equation is presented below:

OLD = .87859 * WEEKS - .29913 * NFAC + .242 * NPEC + 89.9291

(multiple R = .55745; R^2 = .31075; adjusted R^2 = .29913; F [3, 39] = 5.86103; p = .0021)

TABLE 5

Correlations of Occupational Learning Difficulty and Predictors

	OLD	NFAC	NPEC	NUMBER	WEEKS
NFAC	-.317				
NPEC	-.240	.975			
NUMBER	-.137	.591	.622		
WEEKS	.419	-.195	-.170	-.231	
COMPLEXITY	-.095	.405	.428	.329	.013

CONCLUSIONS

The finding that four measures can account for slightly more than 30 percent of the variance for more than 1800 estimates of similarity is most encouraging. However that means that 70 percent of the variance is still not accounted for. It is important to note that the measures of similarity are not measures of *actual* ease of movement but *perceived* ease of movement. Nevertheless the results suggest that subject matter experts can make reasonably consistent and predictable judgments of ease of movement between AFS's. Three of these variables (MAGE, PEC, and TRAIN) are readily available and the fourth (OLD) can be reasonably derived from another equation.

However, it must be noted that a single general equation does not fit all AFS's, and that several specific equations are needed to identify ease of movement to or from clusters of AFS's. The mixed success in producing these equations suggests that an alternative approach might involve producing equations for individual AFS's and then clustering similar equations (Ward, Treat, and Albert, 1984). The finding that the regression equations

work best when predicting ease of movement from a specific AFS to other AFS's rather than to a specific AFS from other AFS's suggests that when a change of AFS is requested the current AFS is more useful to work with than the possible future AFS.

REFERENCES

Blashfield, R.K. & Aldenderfer, M.S. (1988). The methods and problems of cluster analysis. In J.R. Nesselroade & R.B. Cattell ed.), Handbook of multivariate experimental psychology (2nd edition). New York: Plenum, pp. 447-473.

Connell, J.P. (1987). Structural equation modeling and the study of child development: A question of goodness of fit. Child Development, 58, 167-175.

Corter, J.E. & Tversky, A. (1986). Extended similarity trees. Psychometrika, 51, 429-451.

Gordon, A.D. (1981). Classification. New York: Chapman and Hall.

Harvey, R.J. (1986). Quantitative approaches to job classification: A review and critique. Personnel Psychology, 39, 267-289.

Mayfield, D.L. & Lance, C.E. (1988). Development of a candidate task taxonomy for Air Force enlisted specialties. Unpublished paper.

Mumford, M.D., Weeks, J.L., Harding, F.D., & Fleishman, E.A. (1987). Measuring occupational difficulty: A construct validation against training criteria. Journal of Applied Psychology, 72, 578-587.

Truhon, S.A. (1993). The similarity of Air Force specialties as analyzed by additive trees, networks, and multidimensional scaling. Paper presented at Midwestern Psychological Association, Chicago.

Ward, J.H., Treat, B.R., & Albert, W.G. (1985). General applications of hierarchical grouping using the HIER-GRP computer program (AFHRL-TP-84-42). Brooks AFB, TX: Manpower and Personnel Division.

Weeks, J.L., Mullins, C.J., & Vitola, B.M. (1975). Airman classification batteries from 1948 to 1975: A review and evaluation (AFHRL-TR-75-78, AD-A026 470). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Wiggins, V.L., Engquist, S.K., & Looper, L.T. (1992). Applying neural networks to Air Force personnel analysis (AL-TR-1991-0118). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate.

Criteria for Training Evaluation:
Conceptual and Operational Perspectives

David J. Woehr
Assistant Professor
Department of Psychology

Texas A&M University
College Station, TX 77843-4235

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

August 1993

35-1

Criteria for Training Evaluation: Conceptual and Operational Perspectives

David J. Woehr
Assistant Professor
Department of Psychology
Texas A&M University

Abstract

A general criterion construct model specific to training evaluation is developed and presented. This training evaluation criterion model builds on both the training evaluation and criterion development literature to present a broad-based conceptualization of the criterion domain for training evaluation. Relevant boundary conditions that impact the effect of training interventions on various outcomes are discussed. In addition, a distribution-based measurement approach is presented as a way to improve the utility of organization level criterion measures.

Criteria for Training Evaluation: Conceptual and Operational Perspectives

David J. Woehr

Introduction

Goldstein (1991) defines training evaluation as: "the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to selection, adoption, value, and modification of various instructional activities." (p. 557) The evaluation of any training intervention is crucial to informed decision making regarding the intervention. Of vital importance for training evaluation is the standard or criteria against which the training is evaluated. Despite this importance, relatively little systematic attention has focused on a broad-based criterion model for training evaluation. The primary goal of the present paper is to provide an integration of *the criterion construct and training evaluation literatures* to develop a conceptual framework for training evaluation criteria. In addition, the comprehensive evaluation of training interventions mandates the use of multiple criterion measures. The impact of training interventions must be assessed at different levels (e.g., person, work group, organization). Unfortunately, organization level outcome measures are often dismissed as criterion measures due to contamination by extraneous aspects of the work environment. Despite this limitation, the use of these measures is extremely important for demonstrating the utility of training interventions. Thus a second goal of the present paper is to present a measurement approach to operational measures that has the potential to improve the utility of organizational level criterion measures.

Criterion Models for Training Evaluation

The most widely cited approach to criteria for training was postulated by Kirkpatrick (1959, 1960). Kirkpatrick described 4 types of criteria: reactions (what trainees thought of the training program), learning (what trainees learned in training), behavior (trainee performance on the job), and results (organizational level outcome

measures). While this framework was originally intended as a descriptive typology, subsequent researchers have seized on the idea that the four types of criteria represent progressively higher levels such that each proceeding level is a causal precursor to the subsequent level. This hierarchical view of the different criterion measures assumes a high degree of interrelationship among different levels of criterion measures for a particular training context. Recent research (Aliger & Janak, 1989), however, indicates little if any support for this assumption. Thus, although the framework provides an important descriptive typology, it is overly simplistic with respect to the potential relationship among various criterion measures. Specifically, the framework fails to specify relevant boundary conditions that might serve to constrain or facilitate the degree of interrelationship among different criterion levels.

A number of authors have proposed more broad based (i.e., not specific to training evaluation) criterion models (see Borman, 1991 for a review). One such model is the "general criterion model" presented by Campbell, Dunnette, Lawler, & Weick (1970). The Campbell et al. model (presented in Figure 1) specifies 3 types or levels of criteria: job behavior, job performance, and organizational outcomes. In addition, the model explicitly posits a causal linkage among the 3 levels of criteria. To the extent that job behaviors are of value to the organization, they represent job performance. Job performance then determines organizational outcomes. The model also provides a preliminary framework for the identification of relevant boundary conditions. Contextual factors including organizational climate, organizational reward structure, and task demands are included as additional determinants of job behavior. Thus, the relationship between individual abilities and job behavior may be either mediated or moderated by external contextual factors. A further refinement of the general criterion model, postulated by James (1973), suggests that the general criterion model may be expanded to incorporate the multiple criterion model (Dunnette, 1963) which addresses the multidimensionality of job performance. Specifically, multiple measures of performance at each level (job behaviors, performance, and organizational outcomes) are required to appropriately reflect the dimensionality of the criterion domain. James (1973) also advocates a construct validation approach to ascertaining what has been measured by a

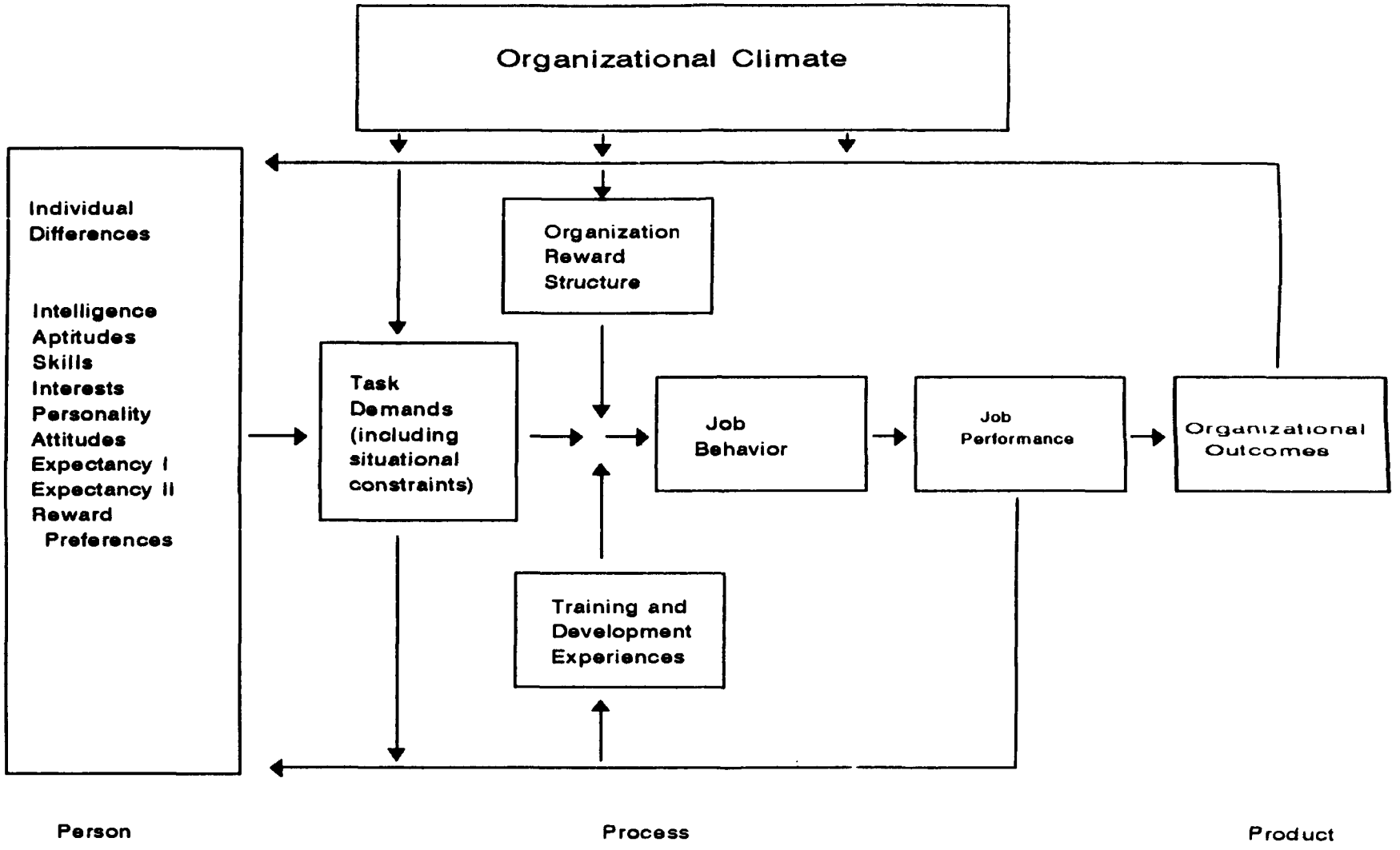


Figure 1: GENERAL CRITERION MODEL (Campbell, Dunnette, Lawler, & Weick, 1970)

criterion. Such construct validation is a critical, but often overlooked, requirement for understanding the relationship among different criterion measures.

The general criterion model is the most complete conceptualization of the criterion construct to date. However, the value of the general criterion model for training evaluation is somewhat limited. First, the model does not explicitly incorporate many aspects of the criterion construct crucial to training evaluation. For example, the learning criterion level (as identified by Kirkpatrick) is not explicitly represented in the model. Rather learning is implicitly subsumed within the individual differences and/or training and development components of the model. Another limitation stems from the proliferation of research focusing on potential boundary conditions for the training process. The literature emerging from this research identifies a number of specific components and the model needs to be updated in light of much of the current training literature.

Thus, it would appear that research focusing on training evaluation would benefit from a reformulation of the general criterion that more directly addresses the training process as well as those situational conditions that affect the impact of training. Such a model is presented below as the training evaluation criterion model.

The Training Evaluation Criterion Model

A reformulation and extension of the general criterion model specifically focusing on training evaluation is presented in Figure 2. This model incorporates much of the current training evaluation literature. Each of the major components of the model and key linkages are presented and discussed below.

Individual Differences. As with the general criterion model, individual differences are assumed to have a direct effect on job behavior. In addition, it is postulated that individual abilities and characteristics impact learning. Tannenbaum, Mathieu, Salas, & Cannon-Bowers (1991), for example, found that cognitive ability, age, gender, physical self-efficacy, and pre-training motivation all predict post-training learning. A number of other individual difference variables have also been found to impact learning. These

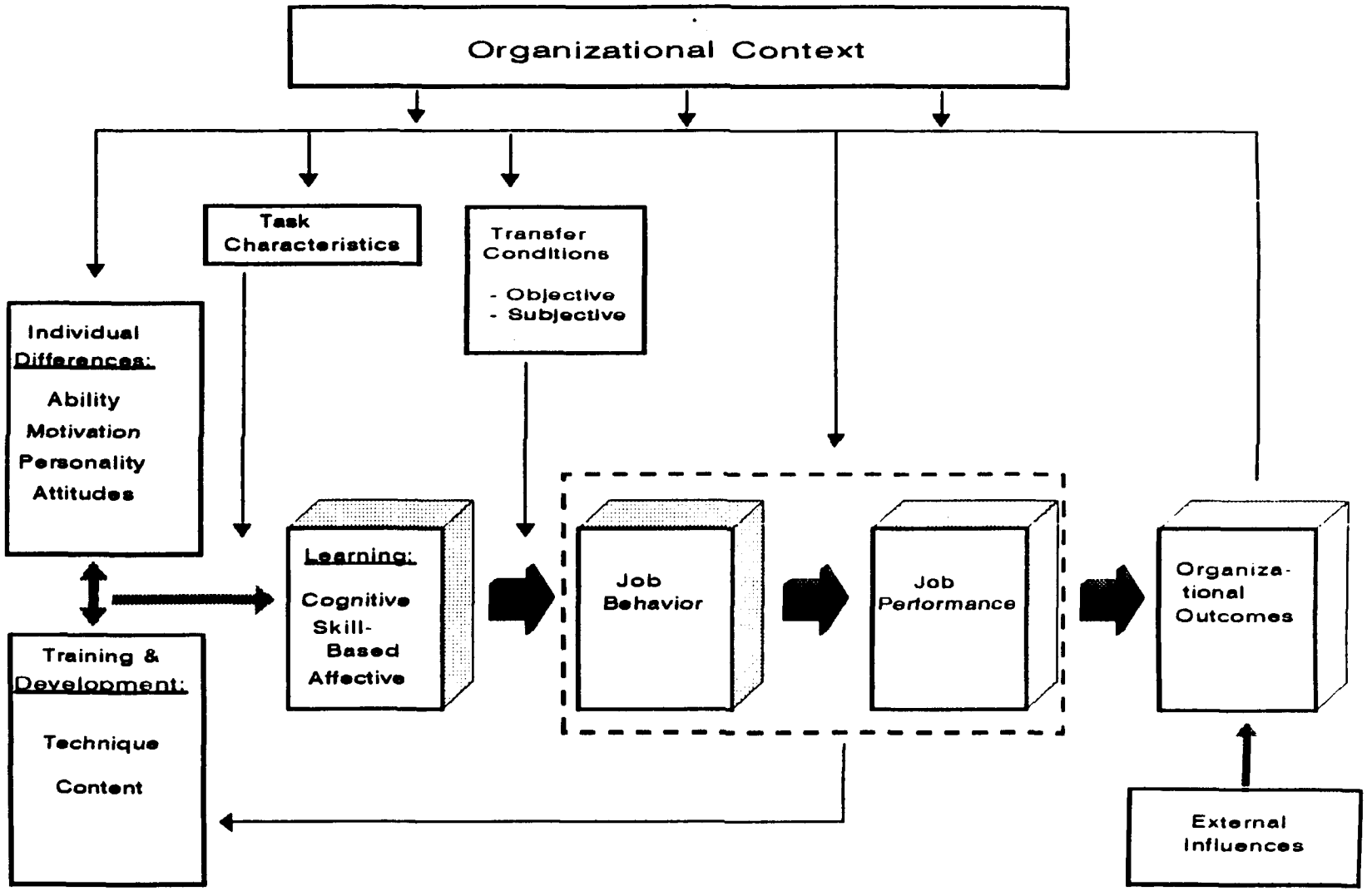


Figure 2: Training Evaluation Criterion Model

include: reactions to previous training (Baldwin & Ford, 1988; Mathieu, Tannenbaum, & Salas, 1992); levels of performance on pre-training work samples, and education (Mathieu et.al., 1992); locus of control, and career/job attitudes (Williams et.al., 1991).

Training and Development. Equally important as a determinant of learning are training and development experiences. Here, two distinct aspects of training must be considered: content and technique. Content refers to the actual knowledge, skills, and abilities conveyed in training. A key issue here, typically addressed through training needs assessment, is one of matching training content to actual job content (i.e., content validity). Ford & Sego (1990), for example, address this issue in their conceptual model for linking training evaluation to training needs assessment. To the extent that training content matches job content, training is more likely to result in a change in job performance and thus organizational outcomes. Alternately, technique refers to how relevant KSA's are conveyed in training and to a large extent determines the strength of the relationship between training and learning. Two aspects of technique must be considered in the design and implementation of training. This first aspect deals with the incorporation of learning principles such as overlearning, varied practice, fidelity, and the teaching of principles. The second aspect deals with the inclusion of meta-strategy skills. That is, in addition to specific content, training must also include general strategies to enhance learning. These strategies, referred to as "in-training transfer enhancing activities" (Thayer & Teachout, 1993), include: goal setting, relapse prevention training, and self-management training. Along these lines Cannon-Bowers, Tannenbaum, & Converse (1991) present an organizing framework for the integration of training theory and technique. While this framework primarily addresses the design and implementation of training programs, it is also directly relevant to training evaluation. Specifically, it focuses on the identification of the constructs underlying learning and subsequent behavior change. Thus, the theoretically based training design advocated by the framework facilitates the formation of conceptual models (and thus criterion measures) for training evaluation.

Learning. Learning is a key construct in the training evaluation criterion model. In the current model, learning is posited to be a required precursor to any behavioral or organizational change that occurs as the result of training. Here, learning is defined in terms of the model presented by Kraiger, Ford, & Salas (1993) which specifies 3 types of learning outcomes: cognitive, skill-based, and affective. This multidimensional perspective of potential learning outcomes has a number of critical implications for training evaluation. First, any training evaluation at this level must consider the congruence among the learning construct being evaluated and both the skills being conveyed through training and the measurement approach. Evaluation of training outcomes must carefully consider whether the appropriate learning construct is in fact being measured. Toward this end, training evaluation must proceed from a conceptually derived theory specifying under what conditions different learning constructs should be related to different behavioral change measures. It is also important to note, that this conceptualization of learning incorporates reactions as a component (or type of) learning. Thus, reactions are not precursors to learning but a separate component of learning. Here again it is important to rely on a conceptually derived theory of the relationship among the various learning components and subsequent behavioral change. One avenue for advancing this with respect to reaction measures is to consider the social psychological literature on the attitude-behavior linkage.

Task Characteristics. Recent theories focusing on ability determinants of skilled performance (Ackerman, 1988; Kanfer & Ackerman, 1989) provide considerable evidence that the relationship between ability and performance is to a large extent moderated by task demand characteristics. Ackerman (1988, 1992) demonstrates differential relationships between individual abilities and performance as a function of three task characteristics: a) consistency of information-processing demands; b) task complexity; and c) degree of task practice. Further, different individual abilities are more important for task performance at different stages of skill acquisition. This has a number of implications for training implementation as well as training evaluation. One important implication is that any KSA-performance linkage must be specified not only in

terms of the specific abilities underlying performance, but also in terms of specific task characteristics. Another implication is that both training design and evaluation must address the dynamic interaction of ability and contextual determinants. More specifically, both the determinants of performance and the required task characteristics may change across the temporal sequence of skill acquisition.

Transfer Conditions. Transfer conditions refer to those organizational or environmental conditions in the work setting that either facilitate or inhibit the expression of learning. Naylor, Pritchard & Ilgen (1980) postulate that environmental factors may impact performance by either constraining an individual's ability to perform a particular task or by prompting an affective response that influences work behavior. Peters & O'Connor (1980), for example, describe "constraints on performance" as lack of proper tools, absence of required help from supervisors or coworkers, or insufficient job information. Similarly Ford, Quinones, Sego, & Sorra (1992) propose an "opportunity to perform" construct that focuses on the opportunity to use knowledge or skills. Alternately, affective responses to the work environment such as job satisfaction or perceived task importance may also serve to either facilitate or inhibit performance. Both actual constraints on performance and affective responses are typically assessed through measures of individual perceptions. Olson & Borman (1989), for example, identified 5 "transfer condition" climate factors: a) general situational constraints; b) supervisor support; c) training/opportunity to use skills; d) job/task importance; and e) unit cohesion/peer support. Similarly, Thayer & Teachout (1993) present a "climate for transfer" model for conceptualizing and measuring transfer conditions. The key point for training evaluation is that any description of the link between learning and behavior change must consider the context in which the behavior is to occur.

Job Behavior/Job Performance. The training criterion model incorporates the job behavior/job performance distinction postulated by Campbell et al. (1970). Job behavior refers to those activities actually performed on the job, thereby defining the job (position). Job performance is defined by the subset of behaviors individuals engage in

that are judged as important for accomplishing the goals of the organization. Thus, performance reflects behaviors that lead to or detract from a position's contribution to organizational goals. This distinction is important in that it suggests that some behaviors have no valuative component. Further the valuative component is a judgment call on the part of the organization based on the perceived impact of performance on organizational outcomes and organizational values. A key point for training is identifying and targeting behaviors that reflect performance as training needs. Another crucial point is the importance of developing a conceptual framework for behavior/performance linkages.

Organizational Outcomes/External Influences. Organizational outcome measures represent global indices of job effectiveness. They typically include results-oriented measures such as productivity indices, promotion rate, salary progression or level, and turnover rates. The value of these measures as a standard for training evaluation is somewhat controversial. Two schools of thought can be found in the literature with respect to ways of conceptualizing the criterion construct. One school of thought emphasizes a conceptualization of performance as reflected in overt individual behaviors (e.g., Campbell, et al. 1971; Borman, 1983). This view focuses on the identification of behavioral regularities important to organization functioning. The other school of thought focuses on outcomes. This view emphasizes the importance of outcomes and results to organizational functioning. Recent theories of the criterion construct, however, have begun to recognize the inextricable relationship between job behaviors and outcomes. Along these lines Binning and Barrett (1989) argue: "...optimal description of the performance domain for a given job requires careful and complete delineation of valued outcomes and the accompanying requisite behaviors" (p. 486).

The detailed conceptual delineation of the relationship between job performance and outcomes is especially relevant to training evaluation. An important direction for future research is a focus on behavior/outcome linkages and generating empirical support for these linkages. Unfortunately, the operationalization of specific outcome measures generates somewhat of a dilemma for training evaluation. On the one hand, the ultimate value of training lies in its ability to impact outcomes of value to the

organization. Operational measures (eg., productivity levels, turnover rates, error rates, etc.) at both individual and aggregate levels would appear to be the ultimate criterion of interest for evaluating training interventions. On the other hand, these measures suffer from a number of problems that limit their usefulness as a standard against which to judge the impact of training.

First and foremost among these problems is the fact that these measures are typically contaminated to an undetermined extent by sources of variance over which the individual has no control. Specifically, the measured outcome is to some extent determined by factors other than performance. External influences can be separated into two broad categories: external determinants and observational/measurement deficiencies. External determinants are external conditions that contribute to or detract from organizational outcomes. For example, economic conditions may differentially impact productivity indices at different times. Observational/measurement deficiencies refer to sources of error stemming from inadequate measurement or observation. A second problem with operational measures is that they are not based on a common metric. Operational measures are often unique to particular units within an organization and thus are difficult to interpret and compare across organizational work groups or divisions. Additionally, the lack of a common metric typically precludes the meaningful aggregation of performance information across organizational units. A third problem is that operational measures only provide an indication of outcome as opposed to the process underlying the outcome. Thus these measures provide very little, if any, information about the nature of performance. Finally, the traditional use of operational measures offers little, if any, means of assessing measurement quality (i.e., how good are the measurements obtained with these measures).

Thus as a criterion against which to judge the impact of various training interventions in organizations, outcome measures have not proven as useful as criteria which are defined in terms of individual behavior. Despite this, however, the use of these measures is extremely important for demonstrating the ultimate utility of training interventions. Consequently, an important goal with respect to training evaluation is the development of ways to improve the utility of organization level criterion measures.

Toward this end, a specific measurement approach to operational criterion measures is presented below.

A Distributional Approach to Criterion Measurement

The measurement approach presented here extends the system for assessing individual performance developed by Kane (1986) to organizational level criteria measurement. It is believed that this approach may offer a partial solution to the problems associated with operational measures. The original system presented by Kane (1986), labeled Performance Distribution Assessment (PDA), is based on the distributional measurement model postulated by Kane and Lawler (1979). An important characteristic of this model is a focus on the range of performance observed. Specifically, the model stipulates that not only is the level of performance important, but the fluctuation or variance in performance must also be considered. For example, two individuals may both be appropriately characterized as "average performers"; however, if one is consistently average and the other alternates between very poor and very good, very different pictures emerge with respect to the individuals' performance. Thus performance measurement must assess the range of performance over time. Specifically, performance is defined in terms of the outcomes of job functions that are carried out on multiple occasions within a specified time span. It is expected that, due to varying levels of individual ability and motivation as well as varying levels of external constraints, these outcomes will reflect different levels of effectiveness. Performance can subsequently be represented in terms of the frequency at which various outcome levels occurred within a given time span.

Another important characteristic of the PDA approach is that it incorporates a relativistic scaling of performance information. More specifically, performance is expressed as a ratio of actual performance (as reflected in the performance distribution generated) to a maximum feasible performance distribution. This maximum feasible distribution reflects the highest level of performance attainable given the constraints under which the work occurs. This scaling process serves to express performance in terms of a relative range of potential performance. Thus, the method allows for

quantifiably excluding from consideration in the evaluation of performance the range of performance that is attributable to circumstances beyond the performer's control.

The representation of performance in distributional form along with relativistic scaling has several important advantages. First, it allows for a consideration of performance variability as well as average levels of performance. Thus it allows for an assessment of the consistency of performance and the extent to which negatively valued outcomes are avoided. In this way more information is provided regarding the idiosyncratic nature of individual performance. Second the relativistic scaling process advocated by the PDA process produces measures of the effectiveness of performance on relativized 0-100% scales with common zero and common upper limits of 100%. Thus any given percentage level remains constant in its meaning regardless of the job, division, or even the organizational level in which it occurs. At the same time, the particular outcome measures used to assess performance may be individualized to meet situational demands and organizational constraints. Specifically, if positions have appreciably different content and extraneous-constraint conditions, measures can be scaled to account for these differences.

The PDA approach was originally advocated as method for enhancing performance ratings. Specifically, it was formulated to incorporate subjective estimates of individual performance outcome frequencies (i.e., ratings of the frequency at which individuals performed at a particular level). However, it's focus on the frequency of particular performance outcomes make it particularly amenable to use with more objective outcome measures. Thus, the application of this methodology to the measurement of organizational outcomes using iterative operational measures appears to be a fruitful avenue for research and may serve to increase the utility of these measures in the training evaluation process.

Adaptation of the PDA System for Use with Operational Measures

As noted above, the PDA system appears to be well suited for the measurement and scaling of operational criterion measures. While a full description of the PDA measurement and scaling process is beyond the scope of the present paper, a brief

overview of the steps involved in the process are presented in table 1. A primary characteristic of the PDA system is a focus on the frequency of occurrence of specific outcomes. In the original PDA process (summarized in column 1 of table 1) rationally based estimates of these frequencies are required at several points. An adaptation of the PDA system for use with operational measures (summarized in column 3 of table 1) replaces these estimates with actual frequencies based on archival records of the operational measures. This revised approach would extend the beneficial characteristics of relative distribution based performance assessment to organizational level criterion measures.

Directions for Future Research

A number of directions for future research are indicated with respect to training evaluation criteria. While a good deal of research has focused on training evaluation, very little has systematically employed multiple criterion measures. Less yet has proceeded from a conceptually based formulation of the criterion construct. Consequently, evidence pertaining to the relationship among training interventions and different levels of criteria is lacking. Here the importance of a construct validation approach to training evaluation criterion measures can not be over emphasized.

In addition, organizational level outcome measures have often been dismissed as criterion measures for training evaluation as a result of the difficulties inherent in these measures. However, given the importance and the wide availability of such measures more attention is needed with respect to the development of better measurement systems. A system such as the performance distribution based one outlined above should permit better measurement of operational measures and thus increase the utility of these measures as criteria for training evaluation.

Table 1

Process	Rational (Ratings)	Empirical (Operational)
Step 1: Determine job functions, measures of these functions, and relative importance of each job function	SME's generate list of relevant measures for position	Same
	SME's scale measures in terms of relative importance	Same
Step 2: Determmine most/least effective outcomes for job function measure	SME's provide estimates of the most and least effective, possible outcomes	Use archival data to establish range of outcomes
Step 3: Determine effectiveness level values	SME's scale least effective outcome relative to most effective outcome	Same
Step 4: Determine maximum feasible performance distribution	SME's provide estimates of minimum feasible frquency at each performance level	Use archival data to establish frequency estimates
Step 5: Determine actual performance distribution	Supervisory frequency of performance ratings	Actual frequency measures

References

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Ackerman, P.L. (1992). Predicting Individual Differences in Complex Skill Acquisition: Dynamics of Ability Determinants. Journal of Applied Psychology, 77, 598-614.

Aliger, G.M., & Janik, E.A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. Personnel Psychology, 42, 331-342.

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

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. Journal of Applied Psychology, 74, 478-494.

Borman, W.C. (1983). Implications of personality theory and research for the rating of work performance in organizations. In F. Landy, S. Zedeck, & J. Cleveland (Eds.) Performance measurement and theory (pp. 127-172). Hillsdale, NJ: Erlbaum.

Borman, W.C. (1991). Job behavior, Performance, and Effectiveness. In M.D. Dunnette & L.M. Hough (Eds.) Handbook of Industrial and Organizational Psychology, vol. 2, (pp. 271-326). Palo Alto: Consulting Psychologists Press.

Campbell, J.P., Dunnette, M.D., Lawler, E.E., & Weick, K.E. (1970). Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill.

Cannon-Bowers, J.A., Tannenbaum, S.I., & Converse, S.A. (1991). Toward an integration of training theory and technique. Human Factors, 33(3), 281-292.

Dunnette, M.D. (1963). A note on the criterion. Journal of Applied Psychology, 47, 251-254.

Ford, J.K., Quinones, M.A., Sego, D.J., & Sorra, J.S. (1992). Factors affecting the opportunity to perform trained tasks on the job. Personnel Psychology, 72, 387-392.

Ford, J.K., & Sego, D.J. (1990). Linking training evaluation to training needs assessment: A conceptual model. Technical report.

Goldstein, I.L. (1991). Training in work organizations. In M.D. Dunnette & L.M.

Hough (Eds.) Handbook of Industrial and Organizational Psychology, vol. 2, (pp. 507-620). Palo Alto: Consulting Psychologists Press.

James, L.R. (1973). Criterion models and construct validity for criteria. Psychological Bulletin, 80, 75-83.

Kane, J.S. (1986). Performance distribution assessment. In R. Berk (Ed.), Performance Assessment: Methods and Applications, (pp. 237-273). Baltimore, MD: Johns Hopkins University Press.

Kane, J.S. (1987). Measure for measure in performance appraisal. Computers in Personnel, 2, 31-39.

Kane, J.S., & Lawler, E.E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. Staw (Ed.) Research in Organizational Behavior, Vol 1. Greenwich, CT.: JAI Press.

Kanfer, R., & Ackerman, P.L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. Journal of Applied Psychology, 74, 657-690.

Kirkpatrick, (1959, 1960). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 13, 21-26, & 14, 13-18,28-32.

Kraiger, K., Ford, J.K. Ford, & Salas, E (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. Journal of Applied Psychology, 78, 311-328.

Mathieu, J.E., Tannenbaum, S.I., & Salas, E. (1992). Influence of individual and situational characteristics on measures of training effectiveness. Academy of Management Journal, 35, 828-847.

Naylor, J.D., Pritchard, R.D., & Ilgen, D.R. (1980). A theory of behavior in organizations. New York: Academic Press.

Olson, D.M., & Borman, W.C. (1989). More evidence on relationships between the work environment and performance. Human Performance, 2, 113-130.

Peters, L.H., & O'Connor, E.J. (1980). Situational Constraints and work outcomes: The influences of a frequently overlooked construct. Academy of Management Review, 5(3), 391-397.

Tannenbaum, S.I., Mathieu, J.E., Salas, E., & Cannon-Bowers, J.A., (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. Journal of Applied Psychology, 76, 759-769.

Thayer, P.W., & Teachout, M.S. (1993). A climate for transfer model. Unpublished manuscript.

AN INITIAL ASSESSMENT OF THE CURRENT LIMITATIONS IN VIRTUAL MANUFACTURING TECHNOLOGY

Albert D. Baker
Assistant Professor
Electrical and Computer Engineering Department

834A Rhodes Hall - ML 30
University of Cincinnati
Cincinnati, Ohio 45221-0030

Final Report for:
Summer Faculty Research Program
Integration Technology Branch
Manufacturing Technology Directorate
Wright Laboratories

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

October 1993

AN INITIAL ASSESSMENT OF THE CURRENT LIMITATIONS IN VIRTUAL MANUFACTURING TECHNOLOGY

Albert D. Baker
Assistant Professor
Electrical and Computer Engineering Department
University of Cincinnati

Abstract

This report presents and initial assessment of the limitations in Virtual Manufacturing Technology. It provides a working definition of Virtual Manufacturing, including an enumeration of the hoped-for benefits from Virtual Manufacturing. The definition is broken into various aspects, and technologies are identified for each aspect. The maturity of each of these technologies is rated. Each aspect is ranked according to how well the identified technologies cover the Virtual Manufacturing vision, the maturity of those technologies, and the potential impact the Virtual Manufacturing Initiative could have. Further work is required to prioritize which high impact items are most desirable to pursue.

AN INITIAL ASSESSMENT OF THE CURRENT LIMITATIONS IN VIRTUAL MANUFACTURING TECHNOLOGY

Albert D. Baker

In the 1930s, the United States Air Force (then the US Army Air Corps) started using flight simulators for pilot training. These early flight simulators used pneumatic bellows, wires, and pulleys to simulate pitch, roll, yaw, and altitude disturbances. Later advances added simulated instrument readings, and eventually television combined with terrain boards provided out-the-window scenes for visual flight. Now, Air Force flight simulators standardly use computer imagery, aircraft sounds, motion, and handling simulations. Other military services have also implemented simulators for reducing the cost of training on expensive military equipment. Recent efforts have been aimed at linking these various simulators together so that now real and imaginary pilots, sailors, and soldiers can fight together in a large scale, synthetic world. Today virtual military experiences are expected to support not only multi-service training, but also military operations, force modernization, force concept exploration, force requirement definition, acquisition prototyping, and equipment test and evaluation. Simulation is now considered fundamental to the military's readiness for war.

Public knowledge about the military's virtual experiences has recently caught the imagination of movie producers, novelists, and the entertainment industry. Now, Virtual Reality is an area of fascination to the American public. These technologies are now being transferred out of the defense industries and being further developed by commercial industries for applications in such far reaching fields as medicine, education, architecture, automotive design, chemistry, biology, and recreation.

Some are asking what impact these emerging technologies can have on the American manufacturing infrastructure. The Air Force's Wright Laboratories are uniquely posed to help answer this question because of the Air Force's long term leadership in both Virtual Reality and in Manufacturing Technology. The Manufacturing Technology Directorate at Wright Laboratories has taken the lead in combining this expertise by forming the Virtual Manufacturing Initiative.

This paper presents an initial assessment of the technical limitations for the full implementation of the Virtual Manufacturing vision. It begins with a working definition of Virtual Manufacturing. This definition is elucidated by describing some of the benefits which are expected from Virtual Manufacturing. The bulk of the paper discusses the current state of Virtual Manufacturing technology.

1. WHAT IS VIRTUAL MANUFACTURING?

Virtual Manufacturing (VM) is an integrated, synthetic manufacturing environment exercised to enhance all levels of decision and control in a manufacturing enterprise.

Decision makers use the environment to answer their questions about the impact of change before making costly investments. Virtual Manufacturing enhances the decision making process by increasing the value, accuracy and validity of decisions. The long-term impact of a decision can be known before the decision is finalized. Tools used to aid decision makers become more powerful when integrated into a virtual environment. Virtual Manufacturing not only helps decision makers visualize non-existent manufacturing systems, but it provides understanding of the proposed manufacturing systems' operations, and identifies possible alternatives. Once decisions are made, models from the virtual environment are used to implement and control the new manufacturing system, insuring that predicted behavior matches actual behavior. The Virtual Manufacturing environment can only operate in a business environment which is technically responsive and proactive in accepting, validating, and implementing change.

Virtual Manufacturing enhances all levels of decision and control in the manufacturing enterprise during the whole product realization process. Decision and control in the manufacturing enterprise are supported from the shop-floor level to the corporate level and beyond. Decision and control during product realization are supported from the concept level through the product support level and eventual disposal. Knowledge and material transformation processes can be assessed using Virtual Manufacturing. Enhanced decision and control at all levels assures products are producible and affordable throughout their life.

The synthetic manufacturing environment interactively mixes real and simulated objects and activities of a new or changed manufacturing system. These objects and activities can be any in number, they can be distributed geographically, and they may be provided by many different sources. Object and activity simulations can be at a variety of detail. The synthetic manufacturing environment immerses its users into models of the new manufacturing system's products, processes, and resources. The synthetic manufacturing environment may be presented in a single location or at many locations.

2. BENEFITS OF VIRTUAL MANUFACTURING

The expected benefits from implementing the above Virtual Manufacturing vision include:

- improved military preparedness,
- the enablement of agile manufacturing,
- the enablement of lean manufacturing,
- improved product design,
- reliable analysis and reduction of time, cost, and effort,
- risk reduction,
- improved process planning,
- better manufacturing system design,
- more effective manufacturing system changes,

- enhanced manufacturing system operation,
- increased understanding of the manufacturing enterprise,
- improved product service and repair, and
- the provision of a vehicle for manufacturing education and research.

VM can improve military preparedness by allowing the virtual production of weapons systems without the expense of actual production, storage, or disposal. Once production methods have been proven, these methods can be stored in the computer, ready for deployment if needed. Manufacturing "know-how" can be preserved with near-zero production. The military complex would be prevented from repeatedly investing in technology which is rapidly becoming obsolete. It can prepare for war without actually investing in war-time production facilities or stockpiling excess, potentially environmentally dangerous munitions. By conserving resources until they are needed, and yet having the ability to produce a variety of needed weapons systems, the military can rapidly respond to changing mission requirements.

VM can assist the development of agile manufacturing enterprises. An "agile" manufacturing enterprise has been defined as a company which thrives in a competitive environment of continuous and unanticipated change [1]. A key idea in agile manufacturing is that companies temporarily form alliances to manufacture products. Some have termed such temporary alliances as "virtual enterprises" or "virtual corporations." They might be termed "temporary" or "agile" enterprises, to avoid confusion with the concept of Virtual Manufacturing described here. VM can assist with the reorganization of an agile enterprise to maximize its performance for the current products being manufactured. VM can allow decision makers to virtually make the proposed changes and test the results. Virtual markets or virtual battlefields can be linked to the enterprise to test the impact of different competitive environments or battlefield scenarios. Unified e[illegible]e models can allow easier temporary integration of manufacturing operations and eas[illegible]ery of new manufacturing systems. Unified models can speed the sharing of analysis tools and manufacturing technology especially between customer and supplier. Communication between experts in different fields (e.g. machining processes, cost accounting, scheduling, and manufacturing engineering) can be facilitated. Product models with enough detail to be useful by both customer and supplier can be integrated into Computer-Aided-Logistics-and-Support (CALS) and Electronic-Data-Interchange (EDI) standards. Customer and suppliers can thus be better integrated in an agile manufacturing framework. Companies seeking partners can more quickly evaluate different partnering relationships. VM can improve the ability to measure the performance, analyze the abilities, plan the use, and synthesize the outputs of a potential partner's factory. Suppliers can use VM to provide more reliable cost, schedule, and performance estimates when quoting work. Cycle-times and time-to-market can be determined in advance. Customers can design products and know that the product can be produced in the supplier's manufacturing system. The corporation can be more agile because it can more rapidly respond to changing customer, process, and technology needs by using Virtual Manufacturing.

VM can also be used to enable lean manufacturing. Lean manufacturing has been defined as the use of teamwork, communication, and continuous improvement to efficiently use resources

and eliminate waste so as to use half the human effort , half the manufacturing space, half the investment tools, half the engineering hours, and half the product development time of traditional mass production [2]. VM can support lean manufacturing by supporting product concept development, design, prototyping, production, manufacturing, maintenance, and disposal without actually making the product. VM can provide a more effective development environment. Producibility can be assessed and cycle-times can be determined in a virtual world. The product-life-cycle can be shortened and simplified by accelerating and integrating decisions into the VM environment. VM can help manufacturers move down the learning curve of new products or processes without actually manufacturing a product. This all can result in more quickly achieving products with higher performance and quality at lower costs.

VM can improve product design. For weapons systems, designers can test new product concepts and the effects of different manufacturing options on product function in battle. Product requirements can be better analyzed. Because options are tested in a virtual world, more options can be tried. VM can speed product design and verification. Products can be verified before production and product constraints can be identified early. Designers can immediately know if a product can be produced, and cost effective prototypes can be made. Design for manufacturing can become a reality by providing immediate design validation and feedback. New research results can be more quickly brought into production by testing various implementation scenarios. Product designs can be modified to assure they conform to given manufacturing capabilities. The virtual environment provides an environment where performance and affordability can be objectively traded off. This can result in fewer engineering change orders. It can facilitate the delivery of high quality and affordable products.

VM can be used to analyze and reduce costs, time, and effort in the engineering and production of products. It can help identify cost drivers and reduce non-recurring costs. It can reliably predict manufacturing costs, determine affordability, and help make reliable cost estimates for making new products. This cost information can be used to determine where to best make capital investments.

Because the changed manufacturing system can be better understood, non-recurring risks can be better identified and reduced using VM. New-concept risks can better be determined.

VM can be used to visualize, understand and determine process plans. This can result in more consistent, accurate, and error-free process planning. The process can be optimized and verified before production. Collision detection and avoidance can be performed on process plans. VM can assist with capturing improvements made on the shop floor to be included in future plans. Newly proposed manufacturing processes can be visualized without building necessary facilities or interrupting current manufacturing operations.

VM can be used to improve the design of manufacturing systems. Complex manufacturing systems can be visualized before being implemented. The manufacturing system concept can be proven and the design can be optimized before costly investments. Numerous options can be tried at little additional cost. Different required capabilities can be assessed, constraints can be identified, potential problems can be isolated, bottlenecks can be determined, and required capacities can be evaluated. Human interfaces to the manufacturing system can be prototyped,

analyzed, and tested before installation. Factory space can be walked through by virtual people and virtual products to make sure that adequate space is allowed in the factory's construction and layout. Floor-space utilization can be optimized before installing equipment. The design can be made right the first time. This can save costs for most manufacturers and it is essential in some applications, such as space applications, where latter modifications are not possible. Process flows can be improved in a virtual world. The manufacturing systems development process can be accelerated while decreasing manufacturing system design errors. Virtual manufacturing cells can be rapidly constructed for ergonomic testing, performance understanding, training, simulation modeling, and refinement.

Once a manufacturing system is installed, it can be more effectively changed using VM. A manufacturer can determine the value of a new or nonexistent process, machine, or manufacturing system before buying it. New processes can be introduced into the enterprise in a virtual world to assess their impacts. New products can be infused into the virtual factory allowing the effects of new product introductions to be known and dealt with before actual product introduction. The impacts of new processing capabilities on new and existing products can be determined. These evaluations can be performed without disrupting the current operations. VM can provide an environment for continuous experimentation to improve the actual system. Risky ideas can be tested by any one in a risk-free virtual world. A case for change can be quickly substantiated. Changes can then be managed so that there are no surprises. A manufacturing system can be implemented in stages and modifications can be introduced gradually. The effects of change can be localized. Continuous improvement can be supported.

VM can improve the operations of manufacturing systems. Processes can be better characterized and validated. Productivity can be better measured and verified. The plants productivity can be optimized. Downtimes can be reduced by better planning. VM can allow the manufacturing facilities and processes to be continually improved. The manufacturing infrastructure can be improved. VM can accelerate improvement in the manufacturing system, and new technologies can be transitioned more rapidly. The manufacturing system can be programmed off-line and programs can be verified before risking or using expensive manufacturing resources. Factory control and scheduling can be improved. The on-line shop information and models required by VM can provide the necessary knowledge to effectively respond to shop-floor problems.

A better understanding of manufacturing enterprises can occur due to VM. In order to simulate a manufacturing enterprise, the enterprise will have to be modeled, and through the modeling process the manufacturing enterprise can be better understood. VM models can standardize the format, the content, and the medium for unambiguously communicating the expected activities of an enterprise. The interdependencies of components can be studied. The VM environment can provide a catalyst for human comprehension of the manufacturing enterprise. It can enable human beings to understand the complexity, the kinetics, and the chaos of modern manufacturing. VM can elevate and strengthen the collective problem solving

abilities of decision makers. Decision making can be improved and people can be forced to rethink past manufacturing paradigms.

Product service and repair can be improved using VM. VM can provide the design interface to service and support by testing reliability, maintainability, and supportability in a virtual world. Special packaging requirements (for shipping, handling, or transportation) can be identified. VM can determine necessary support equipment, data, computers, facilities, personnel, and supplies. Training needs and equipment can be identified. Repair technicians can see how a product was manufactured in order to properly re-manufacture it years later, after the original facility and processes are gone. VM can provide a better understanding of the product so it can better be repaired. Information about past products are readily available. VM can provide enhanced rapid prototyping techniques for repair.

VM can also provide a vehicle for manufacturing education and research. Manufacturing facilities are very expensive and it is often difficult to idle their production for training or research. VM can provide a fully dedicated virtual manufacturing facility. VM can allow movement around the synthetic manufacturing system and provide views from any of a number of perspective without actually having to be in far away, dangerous or difficult to access facilities. Manufacturing scenarios can be modified without the expense of actually performing the modifications. Inexperienced operators can be given the responsibility of managing, controlling, and overseeing a factory in a risk-free training environment. The confusion, complexity, and emergencies of real manufacturing situations can be simulated to train manufacturing operators to make the right decisions without having to risk that they might make the wrong decisions. Simulations can be played over again so trainees responses can be discussed and improved. Operators can be trained in manufacturing systems which are not yet available, to speed the system's start-up, and provide operators with a base-line of expectation for the new manufacturing system's performance. People can be trained to make weapons without actually expending the resources to make them, or risk having them around after production. Group training can be facilitated. New, high-risk even eccentric manufacturing research can be performed without risking or using costly manufacturing capital.

In conclusion, if fully realized, Virtual Manufacturing can provide many benefits throughout a manufacturing enterprise, from product concept to disposal.

3. TECHNICAL ASPECTS OF VIRTUAL MANUFACTURING

The benefits from fully implementing VM are very broad and so are the technical aspects which must be further developed in order to obtain VM's full potential. VM has been defined as an integrated, synthetic manufacturing environment exercised to enhance all levels of decision and control in a manufacturing enterprise. This definition has two major dimensions: the virtual environment, and synthetic manufacturing.

Synthetic manufacturing mixes distributed, real and simulated objects and processes. It addresses all levels of manufacturing including resources and facilities, products and processes. Resources and facilities at all levels encompass equipment, factories, enterprises, and the

markets or battlefields they support. VM concerns product representation at all levels of the product-life-cycle from concept to disposal. Manufacturing processes at all levels include material, data, information, and knowledge transformation.

The virtual environment is an integrated environment. It is used to enhance decision and control by facilitating visualization, providing understanding, and identifying alternatives. The virtual environment must be constructed for a given manufacturing system. It is built on technologies of virtual reality, modeling, analysis, simulation, implementation, and control. The virtual environment is integrated through standards, unified methods, consistent practices, and overarching theories. Corporate business practices must support use of the virtual environment.

The two dimensions of VM are shown in Table 1. Here, virtual reality technologies encompass computer graphics, computer animation, and technologies for providing virtual interaction with manufacturing resources, products, and processes. These technologies include such things as the animation of equipment, the simulation of a control room for a non-existent factory, or the interaction of a human with a non-existent product.

Environment construction technologies are those tools for constructing a virtual environment for a given manufacturing system. These include tools to extract information about the current manufacturing system, and software tools to construct a new virtual environment. For example, data collection systems can be used to gather information about the manufacturing system. Object oriented technologies can be used to build a virtual manufacturing system from standard building blocks. CAD tools can allow product information construction.

Modeling technologies encompass the tools used to analytically describe a manufacturing system. These include IDEF models, product models, and differential equations, depending on the level of manufacturing being modeled.

The integrating infrastructure allows all aspects of the VM environment to work together in a seamless fashion. The integrating infrastructure is built on standards, methodologies, practices and theories. It includes such things as the PDES standard, NC programming standards such as BCL, enterprise integration methodologies, and hybrid systems theory.

Analysis and simulation technologies provide understanding to decision makers, and can be used to suggest alternatives. They include closed-form mathematical analysis tools, and they include computer simulations.

Implementation and control technologies assure that a manufacturing system in the virtual environment has predicted behavior which matches its real behavior when the manufacturing system is implemented. These technologies assure that the models used for simulation are the ones which are implemented. Emulation technologies, code generators, rapid prototyping, and IPPD (integrated product process development) technologies all can be used to assure that predicted behavior matches actuality.

The virtual environment depends on a business environment which will use its capabilities to their fullest. Such a business must be highly technically adept. It must be proactive in deploying new technology.

SYNTHETIC MANUFACTURING AT ALL LEVELS mixes distributed, real and imaginary: resources, products, and processes		INTEGRATED ENVIRONMENT FOR ENHANCED DECISION AND CONTROL facilitate visualization, organize, provide understanding, identify alternatives, and effect implementation and control						
		A. virtual reality technologies	B. environment construction technologies	C. modeling technologies	D. integrating infrastructure	E. analysis and simulation technologies	F. implementation and control technologies	G. business practices
1. resource/facility levels								
	1.1. equipment	equipment animation(9), machine-tool simulators(1)	equipment sensing and monitoring tools(8)	continuous- and discrete-time systems models(9)	continuous- and discrete-time systems theory, practice, and stds.(9)	continuous- and discrete-time system analysis and simulation(9)	continuous- and discrete-time system code generation(7), next generation controller(5)	acceptance of automatic systems
	1.2. factory	factory animation(8), control-room simulators(9)	factory data collection tools(8)	discrete-event dynamic system (DEDS) models(2), chaos models(1), feature-based resource models(2)	DEDS theory(2)	DEDS analysis(4), factory Monte-Carlo simulation (9), distributed discrete-event simu-lation(5), virtual activity cost-ing (0), ergonomic analysis (8), chaos analysis(1)	Petri-net-based control tools(7)	acceptance in new DEDS technologies
	1.3. enterprise	other(0)		enterprise modeling(3)	enterprise integration methodologies(4)	enterprise metrics(4), work flow tools(8), benchmarking(8)	education, reorganization, executive edict	active human resource management
	1.4. marketplace / battlefield	DIS(7)	EDI environment construction(6)	micro-economic models(8)	EDI stds.(6), micro-economic theory(8)	econometrics(8)	laws, education	
	1.5. equipment -to-factory	factory and equipment animation(4)		hybrid systems models(2)	hybrid systems theory(2)	hybrid systems simulation(6)		
	1.6. factory-to-enterprise	other(0)		other(0)		other(0)		
	1.7. enterprise- to-marketplace			marketing models(8), other(0)		market analysis tools(8), battlefield needs analysis(8), requirements analysis(8), COEA(10), other(0)		
2. product levels								
	2.1. pre-concept / concept		causal diagrams(8)	other(0)		design for manufacturability (6), design for serviceability (5), design for retirement(4)		market-driven, understand market, create new markets
	2.2. prototype	virtual prototyping(4), product animation(9), interactive product manipulation(6), IPPD(5)	CAD(8)	feature-based product models(5), proprietary CAD/CAM product models(8), IPPD(5), VHDL(8), other(0)	IGES std.(10), PDES std.(6), IPPD(5), VHDL(8)	static and dynamic models(8), IPPD(5), VHDL(7), design for manufacturability (6), design for serviceability(5), design	rapid prototyping(8), IPPD(5)	quick-to-market attitude
	2.3. eng./mfg. design	product animation(9), interactive product manipulation(6), IPPD(5)	rapid prototyping(8), CAD(8)	feature-based product models(5), proprietary CAD/CAM product models(8), IPPD(5), VHDL(8)	IGES std.(10), PDES std.(6), IPPD(5), VHDL(8)	static and dynamic models(8), IPPD(5), VHDL(7), design for manufacturability (6), design for serviceability(5), design for retirement(4), failure mode analysis FMECA(8), repair	process planning tools (esp. distributed & generative)(5), IPPD(5)	embrace concurrent engineering philosophies
	2.4. operation / support	on-line product documentation(8), virtual disassembly & repair (5)			IGES std.(10), PDES std.(6)			customer-oriented / loyalty to customer
	2.5. retirement	disassembly animation(8)	in-situ microsensor data collection systems(5)	effluent distribution models(7), materials interaction models(5)		knowledge-based materials selection systems(3), design for environmentally conscien-tious mfg. and recycling(4)	automatic control for emissions reduction(5)	environmentally responsible
3. process levels								
	3.1. material transformation	material transformation animation(8), IPPD(5)	CAD(8)	material transformation models(5), IPPD(5)	NC code stds.(10), process plan documentation practices(6), IPPD(5)	material transformation simulation(8), IPPD(5)	CAD/CAM code generation(10), IPPD(5)	interest in manufacturing process
	3.2. information transformation			structured analysis and design models(8), abstract data types(10)		—too numerous to list—	programming languages(10), applications programs(10)	proactive in implementing information technology
	3.3. knowledge transformation			knowledge-based models(5)			inference engines(9)	
4. technologies common to most rows		database (real-time, multi-media, distributed) (2), distributed animation(5), factory computer-human interaction tools (6)	object-oriented technologies (8)	functional, informational, & behavior modeling(6)	communications stds (4), layered implementation methods(6), autonomous agent architectures (4), hypermedia(7), coopera-tive work environments(6)	mathematics(8), AI problem-solving approaches (5), simulation with intelligent agents(4), computer-human interaction analysis(5), logistics support analysis(8)	executable specification technology(3)	pro-technology culture, attitude of continuous improvement, willingness to change, acceptance of electronic signoff and policy changes

Table 1: Virtual Manufacturing Technologies

4. ASSESSMENT OF KEY VIRTUAL MANUFACTURING TECHNOLOGIES

In Table 1 are listed some of the essential technologies for the realization of each technical aspect of VM. Each technology is ranked on a scale from zero to ten according to an initial assessment of its maturity (see Figure 1). This paper only provides an initial assessment of the state of these technologies. Some table entries are missing. Some information about the state of technologies has not been collected. Some technologies may be more or less mature than presented here. This section discusses only some of these table entries.†

4.1. Continuous- and Discrete-Time Systems Models, Theory, Analysis, Simulation, Practices, Standards, and Code Generators (1.1. C-F)

There are three basic types of systems from a systems theory perspective: continuous-time systems, discrete-time systems, and discrete-event systems.

Continuous time systems are parameterized by the passage of continuous time. Most natural systems are continuous-time systems: a falling ball, a spring-mass-damper system, water flowing out of a vessel, etc. Continuous-time systems are often modeled and analyzed using differential equations, Fourier transforms, or Laplace transforms.

Discrete-time systems are parameterized by discrete or integral time. Discrete-time systems often occur when using digital systems which are clocked at a specific rate. Computer systems, digital filters, digital controllers are some of the more common discrete-time systems. Discrete-time systems are often modeled and analyzed using difference equations and Z transforms.

Discrete-event systems are systems which are parameterized by the occurrence of events. Time in the usual sense is not a factor in a pure discrete-event system. Elevators, discrete manufacturing systems, traffic systems, communications networks, and large-scale computer systems respond more to the occurrence of events than the passage of time. The best discrete-event system models are not yet agreed upon. Discrete-event systems are discussed in the next section. Continuous- and discrete-time systems are discussed here.

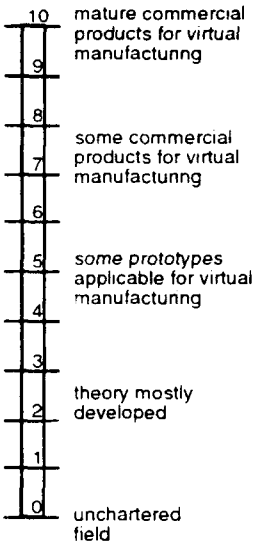


FIGURE 1: Technology Maturity Scale

† Each technology discussed in this section spans one or more technical aspects given in Table 1. These aspects are listed in parentheses after the subsection heading. For example, the "Continuous- and Discrete-Time Systems" heading has in parenthesis (1.1. C-F). Thus, these technologies cover row 1.1 columns C through F.

4.1.1. Importance of Continuous- and Discrete-Time Systems to Virtual Manufacturing

Continuous-time system models are commonly used to model the dynamics and kinematics of mechanical systems such as machine tools, robots, and conveyor systems. Discrete-time system models are commonly used to model digital controllers for these systems such as machine-tool controllers, chemical process controllers, etc. Continuous- or discrete-time models may be used for analog or digital signal processing in process identification or feedback. These models, are fundamental to understanding and controlling manufacturing equipment.

4.1.2. State-of-the-Art in Continuous- and Discrete-Time Systems

Continuous-time systems theory is centuries old. Discrete-time systems theory has been developed over the last half of this century. Systems theory is a standard component of many engineering curriculums, often as a part of a controls, communications and signal processing study. There are numerous textbooks on the subject [3-8]. Some are quite theoretical, some are quite practical. A number of vendors have developed products for simulating both continuous- and discrete-time systems. Continuous-time simulations are usually based on numerical methods for solving differential equations. Discrete-time simulators can use analytic solution techniques or can implement difference-equation models directly. Some of the simulation vendors have also developed code generation packages which implement discrete-time controllers and filters according to the simulation design. Code is generated in a standard programming language, usually C.

Continuous- and discrete-time systems theory, models, analysis, simulation, practices, and standards are very mature technologies (ranked 9). Code generators are relatively new and are given a maturity ranking of 7.

4.1.3. Key Players in Continuous- and Discrete-Time Systems

4.1.3.1. Conferences

American Control Conference (technical)

IEEE Conference on Decision and Control (technical)

International Federation for Automatic Control Triennial World Congress (technical)

Instrument Society of America Annual Conference (trade show)

numerous regional and special interest conferences

4.1.3.2. Vendors

Integrated Systems, Inc., 3260 Jay Street, Santa Clara, CA 95054-9799 (408) 980-1500
Products: $MATRIX_x$, SystemBuild, HyperBuild
Product Features: continuous- and discrete-time system simulation, and code generation

The MathWorks, Inc., 24 Prime Park Way, Natick, MA 01760-1520 (508) 653-1415, info@mathworks.com
Products: MATLAB, SIMULINK
Product Features: continuous- and discrete-time system simulation, and code generation

Mitchell and Gauthier Associates (MGA), Inc., 200 Baker Avenue, Concord, MA 01742-2100 (508) 369-5115, software@mga.com

Products: ACSL
Product Features: continuous-time system simulation

4.2. Discrete Event Dynamic System (DEDS) Models, Theory, and Analysis (1.2. C-E), Factory Monte-Carlo Simulation (1.2. E), and Petri-net-based Control Tools (1.2. F)

Discrete-event dynamic systems (DEDS) are those systems which are primarily parameterized by the occurrence of events. Time is a secondary issue in a pure discrete-event system. For example, an elevator controller is a discrete-event system because it only needs know the events of different floor buttons being pushed and different floors being visited. Similarly, a manufacturing system controller may not be as concerned with how much time it took for a box to reach the end of a conveyor, as it is concerned with the fact that the box is now at the end of the conveyor.

4.2.1. Importance of Discrete-Event Systems to Virtual Manufacturing

The factory level of manufacturing resources are best thought of as discrete-event systems. Thus, discrete-event systems are very important to VM at the factory level. Discrete-event simulators can be used to predict such things as a factory's performance, capacity utilization, bottlenecks, and work-in-progress inventories.

4.2.2. State-of-the-Art in Discrete-Event Systems

Discrete-event systems have been modeled by queuing theory, finite-state machines, Petri nets, formal languages, Markov chains, min-max algebra, and partial difference equations [9]. The best models for discrete-event systems are still a subject of intense study and debate. Thus, we have analytic models of only limited classes of discrete-event systems. General analytic models are not yet developed. Discrete-event system models and theory are given a ranking of 2 in maturity because of the lack of general underlying theories. Analysis techniques are given a ranking of 4, though they lack generality, these analysis techniques can be quite powerful in their limited domains.

Discrete-event systems which can be modeled as queuing networks can be simulated using Monte-Carlo simulation techniques. Monte-Carlo simulators simulate the activities of discrete-event driven factories where the events have a probability associated with when they will happen. A random number generator selects a time for the next event which matches the probability distribution for that event. The use of a random number generator is akin to rolling the dice in Monte Carlo, thus the term "Monte-Carlo" simulation. This field is well developed. There are numerous textbooks [10-14] and conferences on the subject. A variety of vendors sell discrete-event simulators. This field is given a maturity ranking of 9.

Discrete-event systems which can be modeled using Petri-nets lend themselves to implementation using Petri-net based controllers. Some Programmable Logic Controllers (PLCs) can be programmed using Petri-net based languages such as GRAFCET a French standard [15], or Sequential Function Charts from the IEC 1131-3 standard [16]. Numerous researchers have developed prototype Petri-net analysis and control implementation tools. Implementation of Petri-net based control is given a maturity ranking of 7.

4.2.3. Key Players in Discrete-Event Systems

4.2.3.1. Key Conferences in Discrete-Event Systems

Biannual WODES/CODES Conference on Discrete-Event Systems (technical,

Winter Simulation Conference (technical)

European Workshop on Application and Theory of Petri Nets (technical)

4.2.3.2. Key Researchers in Discrete-Event Systems

Bruce Krogh, CMU, Petri-net models of manufacturing systems, discrete-event systems, relationships between Petri-nets and PLC programming languages

Larry Ho, Harvard, discrete-event system models, simulation performance improvement using perturbation analysis

W.M. Wonham, U. Toronto, relations between formal language and automata models of discrete-event systems

Thomas J. Schriber, U of Michigan, simulation modeling of discrete-event systems

Stephen D. Roberts, NC State, simulation language design, application of simulation to medical-care problems

Joseph M. Sussman, MIT, simulation of transportation systems

4.2.3.3. Vendors in Discrete-Event Systems Software

AT&T Istel Visual Interactive Systems, Inc., 25800 Science Park Dr., Ste. 100, Cleveland, OH 44122 (216) 292-2668
Products: WITNES
Product Features: discrete-event system simulation

AutoSimulations, 655 East Medical Dr., P.O. Box 307, Bountiful, UT 84011-0307 (801) 298-1398
Products: AutoMod
Product Features: discrete-event system simulation

Pritsker Corp., 8910 Purdue Rd., Ste. 500, Indianapolis, IN 46268 (317) 879-1011
Products: SLAMSYSTEM
Product Features: discrete-event system simulation

Systems Modeling Corp., The Park Building, 504 Beaver St., Sewickley, PA 15143 (412) 741-3727
Products: SIMAN V
Product Features: discrete-event system simulation

Wolverine Software Corp., 4115 Annandale Rd., Ste. 200, Annandale, VA 22003-2500 (703) 750-3910
Products: GPSS/H
Product Features: discrete-event system simulation

4.3. Distributed Discrete-Event Simulation (1.2. E)

Distributed discrete-event simulation technology performs Monte-Carlo simulation using parallel processors. By having multiple processors work on the same simulation at once, the end simulation result can be obtained much faster.

4.3.1. Importance of Distributed Discrete-Event Simulation to Virtual Manufacturing

The Virtual Manufacturing vision will require substantial computer resources for its full implementation. If computation tasks used to create the virtual environment can be distributed across processors, then information provided to decision makers can be provided at an increased rate, in greater detail. Distributed discrete-event simulation technology will provide the highest level of computational performance in a VM system.

4.3.2. State-of-the-Art in Distributed Discrete-Event Simulation

Distributed discrete-event simulation depends on technologies to partition the tasks (usually called logical processes), assign the tasks to processes, and synchronize the parallel tasks. Task partitioning and task assignment are general problems in the parallel processing community, which are not discussed here. The synchronization issues discussed here are peculiar to distributed discrete-event simulation [17, 18].

There are three basic techniques for synchronizing parallel simulations: synchronous, conservative, and optimistic.

In synchronous simulation, all logical processes advance according to a common clock, and no logical process advances beyond the currently serviced event time. There are several techniques to perform synchronous discrete-event simulation.

In conservative simulation, all logical processes operate concurrently. No processes advances past time *t* until it is guaranteed that no events (distributed as messages) will arrive before time *t*. Logical processes can guarantee that no events will arrive before time *t* by querying source processes about the minimum time for the next event [19] .

In optimistic simulation, the processes execute autonomously and errors are allowed to occur. Errors are caused by one process receiving an event that happened in its past. It has to invoke some mechanism to incorporate the past event into its current state. The three most common techniques for implementing optimistic synchronization are: time warp [20], space-time [21], and moving time windows. Time warp is the most commonly used technique. Space-time is difficult to understand and to implement. The moving-time-windows technique does not repair damage caused by timing errors. Optimistic simulation can become the highest performance parallel, discrete-event simulation method. But, it requires the computational overhead of additional messages, and reprocessing when timing errors occur. These problems are compounded by the fact that the slowest processors are slowed even more by this overhead.

This field is given a maturity ranking of 5.

4.3.3. Key Players in Distributed Discrete-Event Simulation

4.3.3.1. Conferences on Distributed Discrete-Event Simulation

Workshop on Parallel and Distributed Simulation (PADS) (technical)

4.3.3.2. Researchers in Distributed Discrete-Event Simulation

David Jefferson, UCLA, Virtual time, time warp, time warp operating system (TWOS).

J. Lin, Bellcore, Conservative and optimistic simulation, analytical models, optimizations to time warp.

Philip Wilsey, U. of Cincinnati, Optimistic synchronization, digital system simulation, dynamic parameter adjustment, analytical modeling of time warp.

R. Fujimoto, Georgia Tech, Time warp, analytical models of time warp, hardware support for time warp.

4.3.3.3. Vendors in Distributed Discrete-Event Simulation

Jade Simulations International Corporation, 1833 Crowchild Trail N.W. Calgary, Alberta Canada, T2M 4S7 (403) 282-5711
Product Features: C and C++ support libraries for programming time warp simulations

University of Georgia, cosline.cosmic.uga.edu (internet linkup, telnet to this address)
Products: TWOS
Product Feature: Operating system simulator based on time warp.

4.4. Autonomous Agent Architectures (4. D)

The concept of autonomous agents in manufacturing is to use off-the-shelf modules which have the intelligence to act together to form an automated manufacturing system [22]. Autonomous agent research has also been referred to as the study of intelligent agents, multi-agent systems, autonomous systems, and coordination theory. The field is usually thought of as fitting under the general category of Distributed Artificial Intelligence (DAI) [23-28]. Autonomous agents can be considered a higher level abstraction than an object, which is a higher level abstraction than an abstract data type.

4.4.1. Importance of Autonomous Agents to Virtual Manufacturing

Autonomous agent technology seems essential in the ability of Virtual Manufacturing systems to mix real and imaginary objects in a synthetic manufacturing environment. Some form of well structured and intelligent modularity is required to be able to mix and match the real and the simulated. Only if the manufacturing system is modular can parts of it be real and parts of it be imaginary. These modules will require some form of intelligence to operate in a number of possible groupings. These modular concepts are also required to form virtual agile enterprises. Modularity will also help isolate design changes in one part of the manufacturing system from the rest of the manufacturing system. An agent-based architecture would enable the VM promise of manufacturing system implementation in stages, and it would enable the VM concept of allowing modifications to be introduced gradually.

4.4.2. State-of-the-Art in Autonomous Agents

Smith effectively started the Distributed Artificial Intelligence field with his introduction of the Contract Net in 1979 [29]. Autonomous agents have since been proposed and used for factory control and scheduling in a number of different ways. They were first proposed for dispatching jobs in a factory by Shaw in 1984 [30]. In 1986, Parunak et. al. were the first to dispatch jobs in a factory testbed using autonomous agents [31]. A great deal of research and testbed installations of autonomous-agent-based dispatching systems have since been developed. In 1991 Flavors Technology was the first company to install a commercial product for using au-

tonomous agents to dispatch jobs in a factory [32]. Baker also showed the feasibility of advance scheduling using autonomous agents in 1991 [33]. And in 1992 a United Technologies team installed an autonomous agent based advance scheduler in a Pratt-Witney shop [22]. Autonomous agents have also been used to simulate pull-signal control in an automotive factory [34].

There are currently limited products using autonomous agents for factory control. The only commercial product, offered by Flavors Technology, uses autonomous agents for dispatching, but has no advance scheduling capabilities. This product has also been used to simulate jobs being pulled in a factory, but agents have not been used to actually implement pull-signal control of a factory. Flavors has installed this product in only a limited number of factories. It is not clear now what other manufacturing functions can be performed by autonomous agents. Cutkosky et. al. have proposed that agents could be used for concurrent engineering [35]. Further study is required to tie agent-based advanced scheduling with agent-based process planning.

Autonomous agent technology for manufacturing systems is given a maturity rating of 4. Some theoretical work still needs to be done in some areas. In other areas, commercial products need to obtain greater acceptance.

4.4.3. Key Players in Autonomous Agents

Listed here are key players relative to the use of autonomous agents in manufacturing. The autonomous agents field outside manufacturing is quite broad with numerous significant researchers not listed here.

4.4.3.1. Conferences on Autonomous Agents

NCMS SIG on Autonomous Agents' Workshops

IMS Holonic Manufacturing Workshops

4.4.3.2. Researchers in Autonomous Agents

H. V. D. Parunak, Industrial Technology Institute, distributed artificial intelligence, contract net dispatching, the use of autonomous agents in manufacturing

Neil Duffie, University of Wisconsin, agent-based dispatching using parts as agents, heterarchical control architectures

Albert D. Baker, U. of Cincinnati, agent-based advance scheduling

Mike Shaw, University of Illinois, autonomous agent architectures for manufacturing

4.4.3.3. Vendors of Autonomous Agent Software

Flavors Technology, 3 Northern Boulevard, Amherst, NH 03031 (603) 672-3340.
Products: Parallel Inference Machine (PIM) and Paracell programming environment
Product Features: dispatching with autonomous agents, factory

5. ASSESSMENT OF VIRTUAL MANUFACTURING'S TECHNICAL ASPECTS

Each cell in (technical aspect of) Table 1 is classified along three dimensions in Table 2. The technology identified in the cell was ranked for how fully it covers the required technology to implement VM within the scope of that cell. This is termed coverage, and each cell is ranked as having high or low coverage. The identified cells are also ranked according to maturity. Cells

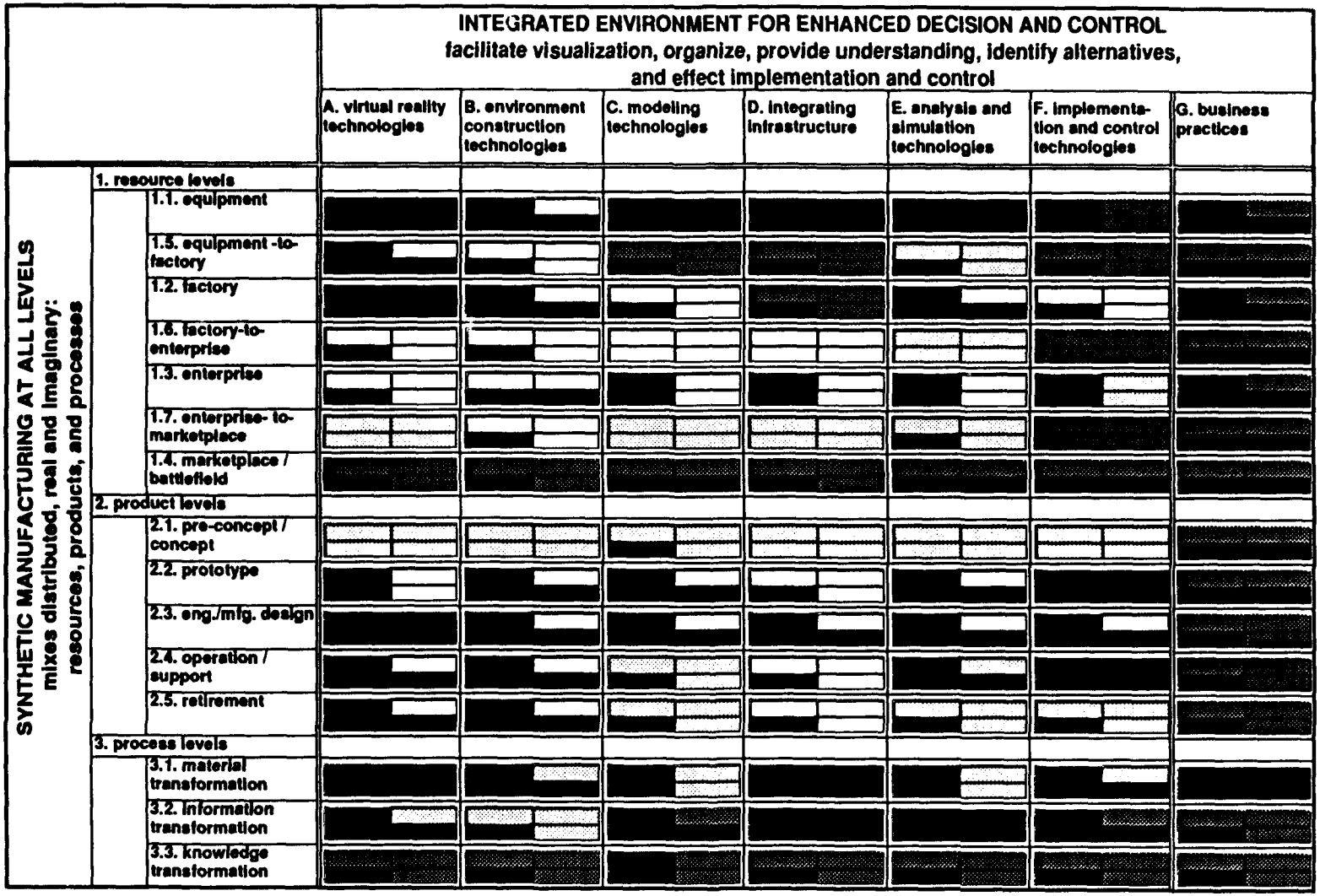


Table 2: Virtual Manufacturing's Technical Aspects - Coverage, Maturity, and Potential Impact

whose most important technologies are given high maturity ratings on the 0-10 scale of Figure 1 are considered high maturity, others are given a ranking of low maturity. The third dimension of each cell is potential impact, whether the uncovered or immature technical aspects can be highly impacted by the VM Initiative. Some cells have low potential impact because other organizations are already vigorously pursuing these areas. Other cells have low potential impact because underlying theories have been elusive to develop.

Coverage and maturity are ranked using matrices like the one in Figure 2. A square is filled in to represent placement along each axis. The third dimension of potential impact is represented with the grayscales shown in Figure 3. These combine to form shaded matrices in each cell such as the low coverage, low maturity, and medium potential impact of cell 2.4.C, operation and support modeling technologies.

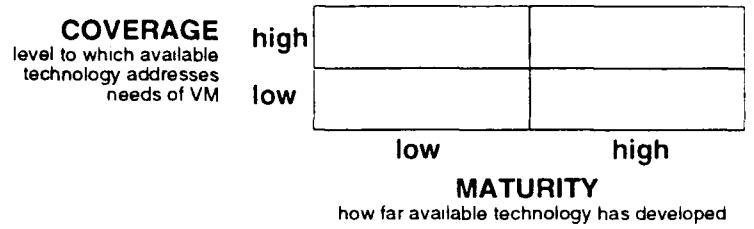


FIGURE 2: Coverage vs. Maturity Matrix



Figure 3: Potential Impact Grayscale

These ratings are not scientifically assigned. But are done based on discussions and research about what is expected from the technology of each cell, and the state of these technologies. These cells could be more accurately shaded by performing more in-depth discussions with leading technologists covering each cell and with those who best understand the VM vision.

Further studies and discussions should also be performed to determine which aspects of VM to develop. This not only depends on the potential impact which could be made in that aspect, but also the desirability of developing that aspect of VM. For example, enterprise level virtual reality technologies are given low coverage, low maturity, and high potential impact; but they do not seem as important as for example, enterprise modeling technologies.

6. CONCLUSIONS

This paper has presented and initial technical assessment of the limitations in Virtual Manufacturing Technology. It has provided a working definition of Virtual Manufacturing. Based on that definition, Virtual Manufacturing is decomposed into two dimensions. As many technologies as could be identified were put in the cells of those two dimensions. The maturity of each of these technologies is rated on a 0-10 scale. Each cell is then ranked according to how well the identified technologies cover the Virtual Manufacturing vision, the maturity of those technologies, and the potential impact the Virtual Manufacturing Initiative could have on those immature uncovered technical aspects. Further work is required to prioritize which high impact items are most desirable to pursue.

7. REFERENCES

[1] S. L. Goldman, "Agile Manufacturing: A New Production Paradigm for Society," AMEF Technical Report, Iacocca Institute, Lehigh University, Bethlehem, PA, 1993.

[2] J. P. Womack, D. T. Jones and D. Roos, *The Machine that Changed the World: The Story of Lean Production*, New York, NY: Harper Perennial, 1990.

[3] R. E. Ziemer, W. H. Tranter and D. R. Fannin, *Signals and Systems: Continuous and Discrete*, New York, NY: Macmillan, 1983.

[4] S. W. Director and R. A. Rohrer, *Introduction to Systems Theory*, New York, NY: McGraw-Hill, 1972.

[5] J. Herbert P. Neff, *Continuous and Discrete Linear Systems*, New York, NY: Harper & Row, 1984.

[6] C. D. McGillem and G. R. Cooper, *Continuous and Discrete Signal and System Analysis*, Philadelphia, PA: Saunders College Pub., 1991.

[7] A. D. Poularikas and S. Seely, *Signals and Systems*, Boston, MA: PWS Engineering, 1984.

[8] S. S. Soliman and M. D. Srinath, *Continuous and Discrete Signals and Systems*, Englewood Cliffs, NJ: Prentice Hall, 1990.

[9] Y.-C. Ho, Ed., *Discrete Event Dynamic Systems: Analyzing Complexity and Performance in the Modern World*, New York, NY: Institute of Electrical and Electronics Engineers, 1992.

[10] G. S. Fishman, *Principles of Discrete Event Simulation*, New York, NY: John Wiley, 1978.

[11] C. D. Pegden, R. E. Shannon and R. P. Sadowski, *Introduction to SIMAN with Version 3.0 Enhancements*, New York, NY: McGraw-Hill, 1990.

[12] A. A. B. Pritsker, *Introduction to Simulation and SLAM*, New York, NY: John Wiley, 1986.

[13] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*, New York, NY: John Wiley, 1981.

[14] T. J. Schriber, *An Introduction to Simulation Using GPSS/H*, New York, NY: John Wiley, 1991.

[15] Diagramme Fonctionnel 'GRAFCET' Pour La Description Des Systems Logiques de Commande, French Standard NFC03-190, Union Technique de l'Electricite, 12, Place des Etats-Unis, 75783 Paris CEDEX 16, France, June 1982.

[16] Programmable Controllers - Part 3: Programming Languages, International Standard IEC 1131-3, International Electrotechnical Commission, Geneva, Switzerland, 1992.

[17] P. A. Wilsey, "Distributed Simulation," Personal Communication, August 18, 1993.

[18] R. M. Fujimoto, "Parallel Discrete Event Simulation," *Communications of the ACM*, vol. 33, no. 10, pp. 30-53, October 1990.

[19] J. Misra, "Distributed Discrete-Event Simulation," *Computing Surveys*, vol. 18, no. 1, pp. 39-65, March 1986.

[20] D. R. Jefferson, "Virtual Time," *ACM Transactions on Programing Languages and Systems*, vol. 7, no. 3, pp. 404-425, July 1985.

[21] K. M. Chandy and R. Sherman, "Space-Time and Simulation," in *Distributed Simulation, 1989*, B. Unger and R. Fujimoto, Eds. San Diego, CA: Society for Computer Simulation International, pp. 53-57.

[22] H. V. D. Parunak, "Briefing 1: What is an Agent?," in *Principles of Autonomous Systems*, Ann Arbor, MI, September 21-22, 1993, National Center for Manufacturing Sciences.

[23] A. H. Bond and L. Gasser, "An Analysis of Problems and Research in DAI," in *Readings in Distributed Artificial Intelligence*, A. H. Bond and L. Gasser, Eds. Morgan Kaufmann, pp. 3-36, 1988.

[24] K. S. Decker, E. H. Durfee and V. R. Lesser, "Evaluating Research in Cooperative Distributed Problem Solving," in *Distributed Artificial Intelligence*, vol. 2, L. Gasser and M. Huhns, Eds. London, England: Pitman, pp. 487-519, 1989.

[25] Y. Demazeau and J. P. Mueller, "Decentralized Artificial Intelligence," in *Decentralized AI*, Y. Demazeau and J. P. Mueller, Eds. Amsterdam, The Netherlands: North-Holland, pp. 3-13, 1990.

[26] E. H. Durfee, et. al., "Cooperative Distributed Problem Solving," in *The Handbook of Artificial Intelligence*, vol. 4, A. Barr, P. R. Cohen and E. A. Feigenbaum, Eds. Addison-Wesley, pp. 83-147, 1989.

[27] M. R. Genesereth and N. J. Nilsson, *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann, 1987.

[28] M. N. Huhns, "Foreward," in *Distributed Artificial Intelligence*, M. N. Huhns, Ed. London, England: Pitmman, pp. v-ix, 1988.

[29] R. G. Smith, "A Framework for Distributed Problem Solving," Ph.D. Dissertation, Stanford University, 1979.

[30] J. P. Shaw, "The Design of a Distributed Knowledge-Based System for the Intelligent Manufacturing Information System," Ph.D. Dissertation, Krannert Graduate School, Purdue University, 1984.

[31] H. V. D. Parunak, J. F. White, P. W. Lozo, R. Judd, B. W. Irish and J. Kindrick, "An Architecture for Heuristic Factory Control," in *Proceedings of the American Control Conference*, Seattle, WA, 1986, pp. 548-558.

[32] D. Morley and C. Schelberg, "An Analysis of a Plant-Specific Dynamic Scheduler," in *Conference on Complex Systems in Manufacturing*, Ann Arbor, MI, April 13-15, 1993, National Center for Manufacturing Sciences.

[33] A. D. Baker, "Case Study Results with the Market-Driven Contract Net Manufacturing Computer-Control Architecture," in *AUTOFACT '92*, Detroit, MI, November, SME Technical Report MS92-338, pp. 31:17-31:35.

[34] C. Schelberg, "Work Flow and Demand Thru a Simple Workstation," in *Principles of Autonomous Systems*, Ann Arbor, MI, September 21-22, 1993, National Center for Manufacturing Sciences.

[35] M. Cutkosky, R. Englemore, R. Fikes, T. Gruber, M. Genesereth, W. Mark, J. Tenenbaum and J. Weber, "PACT: An Experiment in Integrating Concurrent Engineering Systems," *IEEE Computer*, pp. 28-37, Jan 1993.

VISUALIZATION OF RAT BRAIN

Robert V. Blystone
Professor
Department of Biology

Trinity University
715 Stadium Drive
San Antonio, Texas 78212

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

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

August 1993

VISUALIZATION OF RAT BRAIN

Robert V. Blystone
Professor
Department of Biology
Trinity University

Abstract

The anatomy presented by the brain is quite complex, especially at the histological level. Computer visualization of brain anatomy could facilitate various facets of brain research being performed at Armstrong Laboratory. Three areas of computer-based brain visualization were developed during the summer research period at Armstrong Labs: 1) methodology for building a digital brain atlas, 2) techniques for 3-D mapping of brain grafts, and 3) an MRI-based volume reconstruction of a rat head. These visualization approaches utilized microscopy and computer resources both at Trinity University and Brooks AFB. Computer networking capabilities were expanded to expedite transfer of binary digital images between Brooks and Trinity. Visualization software at both sites were explored to maximize the software's utility for AL/OER research efforts. A joint research proposal to the NIH Human Brain Project was prepared and submitted. Collaboration was initiated with the RIC group at the University of Texas Health Science Center San Antonio for the purpose of acquiring MRI data of rat. Numerous computer visualizations were constructed from AL/OER data which in turn have suggested new research approaches and possibilities.

VISUALIZATION OF RAT BRAIN

Robert V. Blystone

Introduction

Members of the Directed Effects group at Armstrong Laboratory are engaged in several projects that involve examination of rat brains at a histological level. Their work often requires interpretation of serially sectioned brain tissue in reconstructed three dimensional space. Additionally, work in progress requires the careful implantation of electrodes and canulae into precise areas of the rat brain. Current brain atlas reference diagrams and photographs are two dimensional as are the serial histological sections from the experimental tissue. Dynamic three dimensional, computer-based visualizations of rat brain offer a significant aid in the manipulation of living brain tissue, in its subsequent preparation for microscopic examination, and in its interpretation as sectioned material. The purpose of the Summer Faculty Research period was to develop the computer visualization procedures.

Methods

1. Networking

The three principals requiring computer graphics support envisioned by this project have quite different computer operating systems and machines. Dr. Mason's lab has a UNIX-based Silicon Graphics entry-level Indigo and a PC 386 DOS system for microscope image capture. Dr. Ziriax has a VMS-based DEC VAX computer system. Dr. Blystone has an Apple-based Macintosh lab operating

under Mac OS and AUX-3. The computers are scattered between three different buildings at two physically separated sites. It was obligatory that these sites be networked to facilitate digital image data exchange.

During May, 1993, we installed Gopher and NCSA Mosaic software on the computers at Brooks. Coupled with FTP capabilities already in place, a better system for the transfer of data was now possible. Other software utilities were installed to provide a broader base for public domain software use within the UNIX environment. Subsequently numerous digital images were transferred between Trinity and Brooks using TELNET, Fetch, and FTP.

2. Visualizing Brain Grafts

Lt. Col. Mickley has established an extensive library of sectioned rat brain tissue. The rats are part of a research protocol that is exploring taste aversion transferal from fetal brain tissue of behaviorally modified rat embryos to untrained adult animals. Suspensions of embryonic tissue are carefully placed in the amygdala or gustatory neocortex of adult rat brains. After five months post transfer, the adult rats are sacrificed and the brains prepared and examined by light microscopy. The grafts are identified and marked on a two dimensional map based on the serial sections of the rat brain.

The adult rats were formalin perfused and 6μm serial coronal sections were taken through Bregma +1.2 or - 2.3 brain regions. Every tenth section was stained with Luxol fast blue and cresyl violet acetate. This histology work was performed by personnel under the direction of Dr. Mickley. I selected six sectioned animal brains for detailed examination.

The stained sections were viewed with an Olympus Vanox microscope and the resultant images digitized using the Image-1 system. (Universal Imaging Corporation, West Chester, Penn.) It is critical that the serial images be properly

aligned. Initially, fiduciary points were created by placing tape on the monitor screen and aligning landmarks by this method. Later, three different digital methods were used. Image-1, Photoshop 2.5, and NIH Image 1.50 all have overlay features that allow two digital images to be viewed over one another. The preferable method was to use the Image-1 feature as it saved time by aligning images at the time of image capture.

The aligned TIFF images were enhanced using Adobe Photoshop 2.5 software (Adobe Systems Incorporated, Mountain View, Calf.). The enhancement procedure involved gray scale balancing each image to provide consistent tonal quality between all images because the Luxol staining process does not yield uniform color balance between sections. The images were edge contrasted to improve detail. The enhancement process was completed by digitally cutting away all non-graft tissue. The resultant image could take on several forms depending on the visual effect desired. One approach would be to leave edge landmarks and threshold the graft as a strict binary image. Upon completing the enhancement process, the images were converted into the PICT file format.

The PICT images were then converted to HDF format using the Spyglass Data Transform Utility. (Spyglass Incorporated, Champaign, Ill.) Spyglass Dicer 2.0 was used to produce three-dimensional rendered volumes of the HDF enhanced data sets. The Dicer software allows the rotation and reslicing of the volume data set for different graft perspectives. Care was taken to bring the x, y, and z pixel planes into proper dimensional agreement. Although the x and y image planes may be several hundred pixels across, the z plane is only one pixel deep. This plane must be adjusted to reflect the proper distance. To minimize the effect of increasing the pixel depth of the z plane, an image doubling macro from NIH Image 1.50 was used to smooth the z plane transitions. (NIH Image is public

domain software obtained from Zippy.NIMH.NIH.gov) NIH Image may also be used to play consecutive images as a movie. NIH Image software has an image projection feature which will allow the rotation of a series of images in either the x, y, or z plane. The total time needed to build each graft reconstruction beginning with image capture was twenty hours.

3. Interactive Computer Brain Atlas

The DEBL research facility makes extensive use of a printed stereotaxic atlas of rat brain. (Praxinos and Watson, 1986). The atlas is used during surgical procedures, during histological preparation, and during interpretation of specimens undergoing microscopic evaluation. Earlier, Dr. Mason had initiated the development of a computer-based version of the stereotaxic atlas. Each page of the atlas had been reduced to a line drawing and the drawing optically scanned into digital format. Using UNIX based software, a three dimensional composite was made using VoxBlast. (VayTek, Incorporated, Fairfield, Iowa) Blystone suggested a more flexible model could be produced using Spyglass software on the Macintosh computer. Additionally, the reconstructed brain grafts from the taste aversion project could be projected on the envisioned three dimensional stereotaxic model.

The image files were converted to PICT format and opened in Photoshop 2.5. The line drawings were edge enhanced and all background "noise" removed. The images required hand alignment and scaling as the original data source (Praxinos and Watson, 1986) did not use consistent scaling features. The reregistered files were then taken into Spyglass Dicer and rendered into a three dimensional volume. Again care was taken to properly express the z pixel plane in its proper depth.

It was possible to computer cut the brain volume map at specific Bregma distances. By taking properly scaled graft reconstructions, the visualized brain graft could be superimposed onto the stereotaxic brain reconstruction. By doing this, the graft could be viewed on an idealized rat brain map.

4. Computer Reconstruction of a Rat Head MRI

Three dimensional reconstruction based on sectioned histological material visualizes only limited volumes of a specimen. An MRI data set can visualize considerably larger volumes of a specimen. Developing dimensional visualizations of the whole head of a rat based on MRI data was the next logical step. The histological data and the MRI data could be combined in the computer for even greater effect when applied to research information. Arrangements were made to use a small animal, research grade, GE MRI chamber at the Univ. of Texas Health Science Center, San Antonio. This chamber operates at 2 Tesla with a typical resolution of from 2 to 3 mm. Images resemble in many ways those made with the light microscope. The x and y axes yield a 256 by 256 pixel image and the z plane is once again 1 pixel deep and represents a depth of approximately 2 mm. The MRI unit available takes eight interleaved images at a time. Once a set of images is made it must be shuffled with another set of eight adjacent images to produce a contiguous set of 16 images. A set of images typically takes 15 to 20 minutes to produce. The first specimen used was a whole rat that required 74 images along its long axis to resolve in its totality.

The raw MRI images are in a proprietary file format that were converted to a PICT format using software made available at the Medical School. Once in the PICT format, the digital images were moved to computers at Trinity and density equalized. The tissue was Sobel edge contrasted and the background gray removed by thresholding by using Photoshop. The enhanced images were then converted to

HDF format and processed in Spyglass Dicer. After calibrating the z depth, a rat was three dimensionally reconstructed from its MRI data set. Printing instructions posed by this report process make it impossible to display the results of the MRI reconstruction.

A week later a rat head was imaged at higher resolution in the MRI. The resultant 24 images were treated in a manner similar to the whole body images as reported above. In both cases a dynamic volumetric database of images was now available for further data visualization. It takes about two days to "scrub" and prepare the data set. A final set of images requires about 1.5 Megabytes to store. These images were converted to film, 35 mm slides, and videotape. A preliminary report of the MRI work was given at the Scientific and Engineering Applications of Macintosh meeting held this year in Boston (August, 1993).

Results

With networking utilities in place, Dr. Mason realized that vast amounts of RF research data could be transferred between laboratories. At his suggestion a grant proposal to the NIH Human Brain Project was initiated. The proposal outlined how NCSA Mosaic's hypertext feature could be used to develop interactive approaches to stored RF data involving brain response to different fields. Another concept pursued in the proposal was to provide an on-line site for the retrieval of computer animation data related to RF research. Computer animated research data is highly difficult to share by conventional means. Mason and Blystone suggested a system in which publication data that included computer graphics could be shared over the Internet by reviewers and later by the public. The proposed network model also outlined how RF dosimetry data could be shared and modeled over the Internet. This type of network sharing would be extremely

valuable to the RF research community. Concluding the ideas developed in the proposal, the Internet was envisioned as a training location for converting brain research data into computer-based visualizations. Complementing this last idea, the networked, interactive RF brain database could be used also by educational facilities. The proposal is still in review at NIH at this time.

The three-dimensional reconstruction of the brain grafts provided a volume view of the grafts within the brain tissue. It was possible to see how the injection of fetal brain cells followed the canula path into the adult brain and how those cells sought their position within the brain tissue. Too few grafts were reconstructed to generalize on the overall anatomy of the graft within the brain. The reconstructions provided a dramatic view of the grafts within the brain tissue. Unfortunately, the limitations of this report do not allow for the appreciation of the detail or animation made possible by the computer graphics. However, an abstract describing the brain graft visualization process was submitted to the American Society for Cell Biology for presentation at its annual meeting in December, 1993.

The interactive computer brain atlas was only partially completed. The diagrammatic brain visualizations created were quite useful; however, there was not enough time to organize the images into an interactive program such as Hypercard. (Claris, Cupertino, Calf.) The concept was that a researcher or student could use a three dimensional image to locate two dimensional views of histology that were embedded in the volume reference image. Initial results of this approach suggested that it could be successful.

A limitation was encountered with the brain atlas project; brains reconstructed from serial sections often compromise gross anatomical features. An MRI data set of the brain and surrounding area could better serve as landmarks for embedded serial sections. Embedded means that the serial section is mapped onto

a volume region of the MRI. When the MRI is computer sectioned to reveal an inner two dimensional view, that view corresponds to an appropriate serial section. If this visualization can be performed, other types of information can be mapped onto the MRI volume; for example, temperature data corresponding to RF treated tissue. These approaches to viewing complex data sets were predicated on having an MRI data set. No MRI images were available for the whole rat or for the rat brain. Dr. Mason and I undertook the acquisition of such a data set and were successful by developing original images with the help of Dr. Peter Fox's group at UTHSC, SA.

To summarize, the following digital data sets were produced during the summer research period:

- four brain graft reconstructions,
- two different rat atlas brain reconstructions, and
- two MRI-based rat visualizations.

These visualizations can be rotated in three space, resectioned within the computer, and animated through time. These images can serve as reference standards for mapping other research data on their visual framework.

Discussion

Computer-based scientific visualization provides a methodology by which the full value of collected research data can be extracted. Often computer visualization can reveal information about an experiment that is not immediately evident from the data. In some ways scientific visualization mirrors the use of statistics in research. Just as statistics is used to find meaning within experimental data, scientific visualization can place large volumes of data into understandable and digestible formats. To a large extent the just concluded summer research

period was directed to such an end: to place large volumes of data into understandable and digestible formats.

We were able to visualize brain grafts in their volume context and suggest how additional information could be extracted from the histology through graft volume estimation. Two dimensional atlas section data could be better understand by seeing each sections position within a three dimensional framework. The MRI volume data sets of rats can serve as templates for the representation of complex numerical dosimetry data. To this end the summer was very successful. The Directed Effects group at Armstrong Lab have new tools to better extract the meaning from their experimental data.

Bibliography

Praxinos, G. and C. Watson, 1986. The Rat Brain in Stereotaxic Coordinates. Academic Press, New York, NY.

Acknowledgments

Appreciation is extended to Dr. Patrick Mason for opening his lab to me. Our many discussions were quite valuable and his encouragement during the preparation of the NIH proposal was the pivotal element leading to the completion and submission of the proposal. Thanks are extended to Dr. John Ziriax for the useful repartee and computer help. Lt. Col. Mickley made his personnel and research materials available to me. The staff at the DEBL provided me many courtesies and help.

MODULATION AND CELLULAR RESPONSES OF NITRIC OXIDE

B. DeSales Lawless

Associate Professor

Department of Science and Mathematics

Fordham University College at Lincoln Center

New York, N.Y., 10023

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

Sponsored by:

Air Force Office of Scientific Research

Bolling Air Force Base, Washington, D.C.

August, 1993

INTRODUCTION

Nitric Oxide has recently become the focus of research in many areas. This short-lived free radical functions as a neurotransmitter, and is involved in vasodilation, inflammatory responses, autoimmunity, platelet aggregation, etc. High levels of nitric oxide are associated with hypotension, endotoxic shock, and in the formation of carcinogenic N-nitrosamines. Its cytotoxicity role in tumor and pathogen distruction is well known. Evidence suggests that in some instances it is responsible for apoptosis (programmed cell death).

Nitric oxide is produced from L-arginine with the enzyme Nitric Oxide Synthase (NOS, EC 1.14.13.39). This enzyme has a constitutive isoform that has been isolated from bovine endothelium. There is also an inducible isoform found in mouse macrophages. Both forms require NADPH and oxygen. Plants produce NO using Nitrate Reductase (NR).

The inducible Nitric oxide synthase (NOS) is produced in the murine macrophage RAW 264.7 by stimulation with gamma interferon. This reaction has been used in the research reporter below. RAW 264.7 cell line is available from American Type Culture Collection, Rockville, MD.

Flow cytometry is useful in cell phenotyping, in delineating the stages of the cell cycle, and in detecting autofluorescence and aneuploidy. These procedures were used in this work. 2N and 4N stages can readily be distinguished. 4N gives the percentage of dividing cells.

METHODOLOGY

Nitric Oxide Assay. In these investigations, nitric oxide content in solutions is determined with the Griess Reagents (0.8% sulfanilic acid and 0.5% N,N-dimethyl-alpha-napththylamine, both in 5N acetic acid). Baxter Medical Corp, Sacramento, CA. The assay colorimetrically measures nitrite concentration which correlates with NO.

In a typical experiment, 10,000 cells/100ul are placed in a well of 96 well microtiter plate with 100 ul of each Griess reagent. The plate is incubated 10 min at 37 deg C. Absorbance is read at 546 nm on a BioRad model 450 microplate reader.

When mouse macrophage RAW 264.7 cells were stimulated overnight with 5 U/ml of gamma interferon and 10 ng/ml lipopolysaccharide they produced increased amounts of nitric oxide. Using the nitric oxide assay, we observed that 3-

aminotyrosine was an inhibitor of this NO synthesis. Aminoguanidine produced some stimulation.

Investigators in the Armstrong Laboratory have transfected RAW 264.7 cells with genomic fractions of the nitrogen reductase enzyme which was found in cells of the barley plant. The fragments used nitrate as substrate to produce nitric oxide. This was similarly detected with Griess reagents. Transfected RAW 264.7 cell lysates were passed over an ion-exchange column and also yielded active fractions as detected with the assay.

Flow Cytometry

Cell Cycle analyses were performed with FACScan Flow Cytometer (Becton-Dickinson, San Jose, CA.) Cells were excited with a single 488 nm argon laser and 5,000-10,000 events were collected. Data were analyzed with CellFit software (Becton-Dickinson).

In cell cycle analyses, the cells were fixed for twenty-four hours in 70% ethyl alcohol at 4 deg C. according to procedure of Coligan, et al in Current Protocols in Immunology. The DNA of the cells was stained with 1 ml solution of propidium iodide (50 ug/ml) containing RNase A (100 U/ml). For cell cycle analysis, the instrument was

calibrated using chicken erythrocytes and calf thymus cells which contined 33% and 80% of the 2n DNA of human cells respectively.

Data graphics were presented with dot plots and histograms including appropriate statistics. Autofluorescence was estimated using dot plots with forward scatter vs side scatter.

Phenotyping for CD4+(T-helper)/CD8+(T-suppressor) cells was effected with the Becton-Dickinson Simultest fluorescent antibody mixture. Anti CD4 was fluoroisothiocyanate conjugated; Anti CD8 phycoerythrin conjugated. Cells were incubated on ice for 30 minutes with 1 ul of the fluorescent antibody mixture. Statistics were obtained with quadrant analysis of dot plots.

INVESTIGATIONS

The following are a selected list of investigations undertaken as a faculty research associate at Brooks Air Force Base during the summer of 1993.

#1. Several investigators from the Hyperbaric Medicine Division of Brooks Air Force Base asked us to devise a protocol whereby they could evaluate the effects of hyperbaric conditions on blood chemistry and immunology. They were advised to submit two ml samples of heparinized bloo taken from each participant prior to having entered the hyperbaric chamber, at the end of end of the third 30 minute oxygen breathing period prior to being removed from oxygen. We suggested a sample at the end upon surfacing and another control sample 24 hours after treatment to study time of rebound.

These investigators came to our laboratory. We instructed them in preparing peripheral blood mononuclear cells by density gradient centrifugation from whole blood. They were also instructed in making cell counts and in determining viabilities by trypan blue dye exclusion on each sample. Samples were assayed for functionality by estimating its T-lymphocyte response to Concanavalin-A stimulation. Cell cycle analysis was performed by flow cytometry on stimulated

and a control population. Production of nitric oxide measurements were taken from the supernatant culture medium (RPMI without phenol red) plus 10% fetal calf serum supplemented with penicillin and streptomycin. Nitric oxide was assayed using Griess reagents. The CD4+/CD8+ (T-inducer/T-suppressor) cell ratios were determined using flow cytometry.

The data indicated that there was immunosuppression under hyperbaric conditions but that the subjects quickly rebounded after the test. This data is being submitted for publication by the investigators in hyperbaric medicine. Their report list us as coauthors.

#2. We repeated some of the previously published studies performed in this laboratory. Our data confirmed the findings. RAW 264.7 cells were grown in overnight at 5.7 x 10 cells/ml in RPMI-10% heat inactivated Fetal Calf Serum, and supplemented with penicillin-streptomycin. The RPMI contained no phenol red indicator so that there would be no interference in our colorometric nitrogen oxide assay. After a four-hour incubation of 1 x 10 fresh RAW 264.7 cells in the supernatant medium from the over-night incubation of the high-density population, the morphology of the cells changed

from round to large spindle shaped adherent cells. They were mitotically and morphologically altered. Members of our laboratory are engaged in isolating and characterizing the putative autocrine that appears in the medium as result of

present in the medium when other stresses were applied to the RAW cells.

#3. We substituted PBMC (peripheral blood mononuclear leukocytes) for the RAW 264.7 cells using the "autocrine" enriched supernatant from RAW cells grown in high concentrations. The autocrine produced in response to stress by the mouse macrophage line is effective against human leukocytes. We are determining if the effect is on lymphocytes or monocytes. We are also observing the effect of the autocrine medium on the T-helper/T-suppressor cells.

#4. We confirmed that nitric oxide produced by RAW 264.7 cells is toxic to the macrophages that produced it.

#5. When human PBMC are stimulated with gamma Interferon/Lipopolysaccharide, they respond as the RAW cells do -- they produce nitric oxide. Furthermore, the nitric oxide in this medium is cytotoxic to the leukocytes. It is the plastic adherent monocyte-macrophage population that is affected in each situation.

#6. Human dendritic cells were perpared from whole human blood at Rockefeller University and sent overnight to our laboratory. These cells produced nitric oxide when stimulated with gamma-Interferon/LPS. The nitric oxide was cytotoxic to the dendritic population. The Dendritic Cell is a potent antigen-presenting cell. It also transports the HIV virus to the T-helper cells in human circulation.

#7. To test the hypothesis that cell-cell contact is required for the production of the putative autocrine that is mitogenic, we sent RAW 264.7 cells at the high density to Huntsville, Alabama. Colleagues there rotated the cells overnight in their bioractors which simulate microgravity -- conditions encountered on NASA Shuttle flights. If the supernatant fluid from the reactor fails to exhibit properties of the autocrine enriched medium, we would conclude that the cell-cell contact is required for its production. Cell-cell contacts are reduced in microgravity conditions. If the expected results are found, we would arrange to repeat the project on a future NASA Shuttle mission. This could be arrange with our colleague, Dr. Marian Lewis in Huntsville, Alabama.

#8. We verified the hypothesis "NO is needed for lymphocyte proliferation." We incubated human PBMC in medium containing exogenous nitroprusside. This compound effectively added nitric oxide to the medium. In control studies, we added 3 aminotyrosine, a nitric oxide inhibitor. We stimulated PBMC with Con A in medium with and without exogenous NO. These studies were combined with determination of the CD4/CD8 ratios to see if NO selectively affected either of these populations.

#9. We completed preliminary studies to to learn if NO production via the nitrogen oxide synthetase in RAW cells resulted in DNA fragmentation. This was done with flow cytometry. In indiscriminate necrosis, DNA fragmentation would appear as multiple fractions below the 2N DNA; in apoptosis or programed death, there should be a few breaks expected. These can be detected on the histograms of Propidium Iodide stained nuclei.

#10. We found results in preliminary studies on AGE (advanced glycosylated end products; and leukocytes cultured in medium containing high glucose concentrations. AGE is fluorescent; leukocytes show no autofluorescence. AGE also combine with NO and should reduce its concentration in the medium.

11. We investigated autofluorescence in three immortalized lines, HL-60 (human lymphocyte), RAW adherent and RAW non adherent populations, and human PBMC. There was no detectable autofluorescence in PBMC and amounts varied in the immortalized cell lines. We would like to test the hypothesis that reduced pyridine nucleotides are responsible for autofluorescence in the immortalized cell lines. We expect elevated metabolic activity in such cells.

#12. The clone RAW 264.7 when taken from liquid nitrogen and put into culture develops into an adherent population and a suspended population of cells. These behave differently when examined for autofluorescence. We want to see if they produce NO equally well when stimulated with INF/LPS and if they are affected equally by nitric oxide. Aminoguanidine is known to block NO production. We would like to repeat the our studies on PBMCs after separating the adherent monocytes from the non-adherent lymphocytes.

#13. We have completed several studies on apoptosis but have no definitive answers. We plan to continue apoptosis study as it is affected by nitric oxide.

#14. We have made preliminary studies of the action of the free radical on nucleic acids. We would like to continue this research.

Hydrothermal Effects on the Structural Integrity of Graphite Fiber-Cyanate Ester Resin Composites

B. L. ("Les") Lee
Associate Professor
Department of Engineering Science and Mechanics

The Pennsylvania State University
227 Hammond Building
University Park, PA 16802

Final Report for:
Summer Faculty Research Program
Air Force Wright Laboratory
Materials Directorate, Non-Metallic Materials Division
(USAF Researcher: Capt. Michael W. Holl)

Sponsored by:
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

September, 1993

Hydrothermal Effects on the Structural Integrity of Graphite Fiber-Cyanate Ester Resin Composites

by

B. L. ("Les") Lee
Associate Professor
Department of Engineering Science and Mechanics

The Pennsylvania State University
227 Hammond Building
University Park, PA 16802

Abstract

The weight change and the retention of in-plane shear (+/-45°) strength of graphite fiber-reinforced cyanate ester resin matrix composites have been measured on the exposure to *high humidity* and *thermal cycling* respectively. Cyanate ester resin matrix composites absorbed a remarkably small amount of moisture on the exposure to 95% RH condition at 60°C up to 36 days. However, the degree of moisture absorption underwent a rather sudden increase to an equilibrium level of 1% after the prolonged exposure. The morphology study showed the occurrence of extensive cracking of matrix/interface region in the form of the delamination between the plies as well as translaminar cracking within the ply. The phenomenon is believed to be caused by weakening of the fiber-matrix *interface* which was confirmed by microscopy analysis of fracture surface. A sudden moisture gain associated with extensive matrix/interface cracking was found to reduce in-plane shear strength and fatigue lifetime at a given stress amplitude. From the assumed relationship between the slope of S-N curve and 'm' factor of Paris law, it was hypothesized that the rate of crack growth is higher for wet specimens already with extensive cracks initiated. The rate of in-plane shear strength degradation was also measured on the static exposure to dry heat as well as thermal cycling to a peak temperature of 150 or 204°C. At a frequency of 10 min/cycle and for a relatively short duration, the effect of thermal cycling seems to be represented by the cumulative sum of thermal degradation effect at the peak temperature.

INTRODUCTION

Fiber-reinforced polymer matrix composites which form various structural components of aircraft are subjected to a complex history of temperature, moisture and other environmental conditions. Their effects, particularly the effects of moisture absorption and thermal cycling, have been extensively studied since the early 1970's [1-9]. Excellent review articles have been available on these subjects. Among them, worthwhile to note are the reviews done by Springer [4], Adams [5], Ashbee [6], and Wolff [7] on the moisture effects. Reflecting the state of technology, the past investigations were concentrated in the epoxy resin matrix composites typically reinforced by graphite, glass or aramid fibers. As a result, hydrophilic nature of epoxy resins as composite matrices has been well-documented. Since many applications cannot accept moisture-sensitive nature of epoxy resin matrix composites, intensive research works were performed to introduce new thermosetting resins with a reduced tendency of moisture absorption without penalizing the processability. One important example of the achievement in this direction was the synthesis of polycyanurate, so-called *cyanate ester*, resins for composites [7,10,11]. Neat resin casting of cyanate ester was reported to exhibit an equilibrium level of water absorption as low as 0.7% at 100°C [7].

Cyanate ester resin matrix composites reinforced by glass or aramid fibers established their presence in the application for multi-layer electric circuit board as early as the late 1970's, because of the following considerations [10]: (a) their glass transition temperatures (Tg) exceeding those of epoxy resins and matching molten solder temperature (220-270°C); (b) low dielectric loss properties; and (c) excellent processability (epoxy-like). Recently, successful efforts for toughening of the resins by the inclusion of elastomer or thermoplastic secondary-phase have led to a serious consideration of cyanate ester resin matrix composites for their use in primary structure applications [10, 12-17]. Currently available cyanate ester resins for structural composites offer a desirable combination of high resistance to moisture absorption, reasonably good fracture toughness, and acceptable processability. In addition, a relatively low cure temperature (177°C) and a large amount of free volume after gelation were reported to allow the cure shrinkage less than 1% creating more residual stress-free composite parts [10].

In the case of epoxy resin matrix composites, moisture absorption was shown to relieve the residual stresses by matrix swelling and temporarily increase the resistance to matrix/interface cracking [18,19]. The occurrence of this type of interaction is expected to

be negligible in the cyanate ester resin matrix composites which absorb a very small amount of water. However, the question of whether this small amount of water absorbed will eventually influence the fiber-matrix *interface* of cyanate ester resin matrix composites as in the case of epoxy resin composites has not been resolved [6,7]. Aside from the issue of moisture effect on the interfacial adhesion, the structural integrity of cyanate ester resin matrix composites should be assured under the condition of thermal spike or cycling. Here a systematic data base is lacking, despite the claims by the resin producers of better microcrack resistance of cyanate ester resin composites over the epoxy resin composites upon thermal cycling. One of the first questions to resolve in this area appears to be whether the effect of thermal cycling can be represented by the cumulative sum of thermal degradation effect at the peak temperature. An equally important question is what factors control a critical frequency of thermal cycling above which the deviation from cumulative effects occurs. These questions were not fully examined even in the case of conventional thermoset resin matrix composites where extensive data base exists for thermal cycling.

Our study plans to assess the long-term structural integrity of graphite fiber-reinforced cyanate ester resin matrix composites with a special attention to the issues raised and discussed so far. As a first step toward this goal, the weight change and the retention of matrix/interface dependent mechanical properties of the composite systems have been measured on the exposure to *high humidity* and *thermal cycling* respectively. In consideration of the long-term structural integrity, the study has included the evaluation of not only static strength but also *fatigue* lifetime profile of composites. A study has been initiated to resolve the question of whether the effect of thermal cycling can be represented by the cumulative sum of thermal degradation effect. Continuing study will eventually assess cumulative effects of temperature cycling or spike in conjunction with moisture exposure on the mechanical performance of composites. When completed, the efforts will to clarify the mechanisms of hydrothermal environment-induced change of the composite performance, and thereby provide a technical basis for the future work on the environmental fatigue of composite structures. Preliminary findings of our efforts are reported in this paper which will be the first one of the series.

OBJECTIVES

(a) To examine the effects of moisture absorption, temperature cycling, and their combination on the long-term structural integrity, particularly matrix/interface dependent mechanical properties, of graphite fiber-reinforced cyanate ester resin matrix composites; (b) to assess the mechanisms of hydrothermal environment-induced change of the composite performance; and (c) to provide a technical basis for the future work on the environmental fatigue of composite structures.

EXPERIMENTS

Specimen Preparation Unidirectional composite prepreg was obtained from ICI Fiberite Composite Materials Inc. (Tempe, AZ). The reinforcement was Hercules IM7 graphite fiber. The resin matrix for the prepreg was ICI 954-2 cyanate ester resin which reportedly contains thermoplastic toughener of a proprietary nature. *Off-axis angle-plied* laminates were prepared by laying up four prepreg plies with a symmetric configuration of $(+45/-45^{\circ})_s$ and curing them in autoclave. The following cure conditions recommended by the materials supplier were used: heat-up rate of 3°C/min; 1 hour at 121°C (250°F) followed by 2 hours at 177°C (350°F); pressure of 0.69 MPa (100 psi). Some of the laminates were subjected to an optional postcure cycle consisting of 2 hours at 232°C (450°F) under contact pressure. The thickness and the resin content of cured composite panels was approximately 0.57 mm and ??% by weight respectively. Two types of straight strip specimens were cut with water coolant from the panels: 12.5 mm X 62.5 mm (Type I), and 18.75 mm X 125 mm (Type II). The specimens were immediately placed in a desiccator after cutting. Their weights were monitored until they reached an asymptote due to drying-out of moisture.

Environmental Exposure Both Type I and Type II specimens of graphite fiber/cyanate ester resin composite were placed for various periods of time in a humidity chamber which were maintained at 95% RH and 60°C (140°F). The weight gain of each specimen due to moisture absorption was measured in a chemical balance after the moisture was removed from the specimen surface by soft paper tissue. The results of the specimens exposed to high humidity continuously were compared with those of the specimens which were taken out of the humidity chamber for weighing, re-exposed to the humidity, and re-

weighed. Both techniques produced the same consistent results within the range of data scattering.

A separate lot of composite specimens were also exposed to dry heat at 150°C (302°F) or 204°C (400°F) in either cyclic or static mode. Static exposure of the composite specimens to dry heat was performed using an oven without air-circulation. Thermal cycling between elevated temperature and ambient temperature was accomplished by shuttling the specimen between the preheated oven and an open area with cooling fan. Shuttling action was performed using a hydraulic test machine operated at a frequency of 1.67×10^{-3} Hz (10 min/cycle) in a stroke-controlled mode. The above cyclic frequency allowed 5 minutes' exposure at elevated temperature per each cycle. After the exposure to various environments described so far, the Type I specimens were used for the measurement of weight change. The Type II specimens were used for tensile testing in static as well as dynamic mode.

Mechanical Testing After the exposure to moist heat (95% RH, 60°C), dry heat (150 or 204°C), or thermal cycling (to 150 or 204°C from ambient temperature) for various periods of time, the Type II specimens of graphite fiber/cyanate ester resin composite were subjected to static uniaxial tension until failure occurs. The rate of deflection and the initial gage length of the specimen were 1.25 mm/min and 75mm respectively. The in-plane shear strength and modulus were defined according to the procedure of ASTM D3518. Monitoring of strain along the transverse as well as axial direction, which is needed for the calculation of in-plane shear modulus, was limited to the case of dry specimens. In the case of axial strain measurement, bonded strain gage technique was used along with an extensometer. A reasonably close agreement between the two types of measurement was observed. For wet specimens, only the axial strain was measured by an extensometer.

In the case of dry specimens and the specimens exposed to moist heat (95% RH, 60°C) for 50 days, tension-tension fatigue experiments were performed at ambient condition. The number of cycles to failure were measured at various levels of stress range to define S-N (stress range vs fatigue life) relationship. Cyclic frequency and the minimum-to-maximum stress ratio were fixed at 10 Hz and 0.1 respectively. No preventive measure was taken to reduce moisture desorption of wet specimens during the fatigue experiment.

Failure Analysis Along with mechanical testing, the morphology and the extent of damage propagation of graphite fiber/cyanate ester resin composites were examined by

optical microscopy. Scanning electron microscopy (SEM) was used to observe the mode of failure before and after the exposure to moist heat, dry heat, or thermal cycling. Small pieces of composite laminates were mounted on specimen studs. The samples were coated in a sputter coater with a thin layer of gold palladium. The fracture patterns of the coated samples were examined using the secondary electron mode.

RESULTS AND DISCUSSIONS

Moisture Absorption As discussed previously, the weight change and the retention of matrix/interface dependent mechanical properties of composites are our immediate and foremost concern in assessing the long-term structural integrity of graphite fiber-reinforced cyanate ester resin matrix composites. Our test results clearly showed that IM7 graphite fiber-reinforced 954-2 cyanate ester resin composites absorb a remarkably small amount of moisture on the exposure to 95% RH condition at 60°C up to 36 days (Figure 1). The moisture content remained almost constant at a level of around 0.3% by weight after the initial pick-up within 1 day. Postcured specimens appeared to follow the same trend with a slightly higher moisture content. These values are comparable to the assumed equilibrium level of moisture content reported by the materials supplier [10-12]. However, a possible error in our estimation of moisture content cannot be ruled out, mainly because of extremely low degree of weight change involved and uncertain effectiveness of removing the moisture absorbed during cutting of the specimen.

One interesting finding was that the degree of moisture absorption undergoes a rather sudden increase to the level of 1% after the exposure for around 50 days. After this increase, the moisture content remained constant around 1% up to 100 days' exposure forming a second level of equilibrium. Even at this level, the moisture content of cyanate ester resin composite is still lower than that of epoxy resin composites which typically exhibit an equilibrium moisture content of 1.5 to 1.8% [4]. The observed pattern of moisture gain, a sudden increase of moisture content after 50 days and the existence of dual equilibrium levels, clearly indicates that moisture absorption behavior of IM7 graphite fiber-reinforced 954-2 cyanate ester resin composites cannot be described by a concentration-dependent form of Fick's law. Moisture diffusion behavior of fiber-reinforced composites can deviate from Fick's law for the following reasons [4]: (a) the development of cracks or delamination altering the structure of the material, (b) moisture

propagation along the fiber-matrix interface, (c) voids in the matrix, and (d) non-Fickian behavior of matrix itself without defects.

Our microscopy study show : that the composite specimens in dry state are virtually void-free and well-compacted. vever, on the exposure to moist heat (95% RH, 60°C) for 50 days, the composite under ent extensive cracking of matrix/interface region (Figure 2). The damage took the form of not only the delamination between the plies but also translaminar cracking within the ply. Some matrix cracks remained in the subsurface region but lying parallel to the plane of surface. Less cracks were observed in the matrix-rich pockets of composites. The observed damage of matrix/interface region is believed to be a primary source of a sudden moisture gain after 50 days' exposure. Other potential sources for anomalous diffusion behavior of cyanate ester resin matrix composites will be checked in our future work. The question of moisture propagation along the fiber-n rix interface will be examined by comparing the moisture absorption behavior of the comp te specimens with their edges exposed and that of the specimens with the sealed edges.

Although it is almost certain that the damage of matrix/interface region is responsible for a sudden moisture gain after 50 days' exposure, the cause of damage formation in the composites still remains to be answered. As discussed, matrix swelling effect should not be a factor in the cyanate ester resin matrix composites which initially absorb a very small amount of water. At the temperature of 60°C, the possibility of having the cracks due to the volume expansion of absorbed moisture is rather slim. Therefore, the only explanation of the phenomenon is that, on a prolonged exposure to humidity, this small amount of absorbed water gradually weakens the fiber-matrix *interface* of cyanate ester resin composites as in the case of epoxy resin composites. In the case of graphite fiber/epoxy resin composites, Browning et a. [20] showed that matrix/interface dependent 90° strength of unidirectional lamina decreases by almost 40% at the moisture content of around 1%. As will be discussed later, in-plane shear strength of our graphite fiber/cyanate ester resin composites decreased by 20% at the same level of moisture content. Although weakening of fiber-matrix bonding has been confirmed by subsequent failure analysis, a rather unusual occurrence of moisture-induced cracks in the absence of external loading is still difficult to explain. Moisture-induced cracks might be related to the fact that cyanate ester resins exhibit a large amount of free volume after gelation which led to very low level of cure shrinkage and residual stresses [10]. More study including chemical analyses of cured vs postcured composite systems is planned to probe the cause of matrix cracking.

Moisture Effects on the Mechanical Properties A sudden moisture gain after 50 days' humidity exposure associated with extensive matrix/interface cracking was found to have direct influences on the static strength and fatigue lifetime profile of IM7 graphite fiber-reinforced 954-2 cyanate ester resin composites. Static in-plane shear strength of composites decreased from 70 MPa in dry state to 58 MPa after the humidity exposure. As shown in Figure 3, S-N (stress range vs the number of cycles to failure) curve of wet specimens clearly shifted to the left of that of dry specimens, although the data tended to be more scattered for the wet specimens. At a given stress amplitude, the composite specimens which contained extensive cracks as a result of humidity exposure exhibited almost exponentially shorter fatigue lifetime. When the same data are plotted in log-log scale (Figure 4), the S-N relationship took the following form of $S = a N^{-b}$ where a and b are constants. The slope of S-N curve, represented by (-b), were found to be lower for wet specimens with extensive matrix cracking (-0.045) than dry specimens (-0.059).

Past research work [21,22] showed that, when the fatigue lifetime is dominated by the damage propagation phase, (-b) is equal to (-1/m) where m is a power-law factor of Paris' law. Fatigue crack growth data for most polymers and composites were reported to follow Paris' law which takes the following form: $da/dN = A(\Delta K)^m$ where a is crack length, N is cycles, ΔK is the range of stress intensity factor on each cycle, A and m are constants. If the above-discussed relationship is valid for our composite system at least in a qualitative sense, the rate of crack growth at a given range of stress intensity factor should be higher for the wet specimens (already with extensive matrix/interface cracks initiated) than dry specimens. This hypothetical conclusion seems to be consistent with the previous assumption of fiber-matrix interface weakened by absorbed moisture. Separate experiments involving the measurement of delamination growth under cyclic loading are planned to verify the relationship between the slope of S-N curve and m factor of Paris law.

Moisture Effects on Failure Modes Weakening of fiber-matrix interface due to moisture absorption was confirmed by extensive failure analysis of the composite specimens based on scanning electron microscopy (SEM). One difficulty in the failure analysis was the lack of any quantitative means of measuring the extent to which matrix remnants adhere to the fibers in the crack path region. However, the fracture surface of dry composite specimens displayed a clear tendency of having the remnant of matrix resin clinging to the fibers even in a highly-fibrillated region with many bare fibers, as shown in Figure 5-a. Compared with dry specimens, wet specimens showed a much more reduced

tendency of having the deposit of matrix resin on the fibers. The frequency of observing the resin deposit over the given area was found to be distinctly lower for wet specimens. When observed, the remnants of matrix resin often showed an appearance of being easily peeled (Figure 5-b).

Thermal Cycling vs Static Exposure In addition to the study of moisture effects, our research program examined how thermal cycling influences the static strength of cyanate ester resin matrix composites. The study is intended to be a preliminary phase of the future investigation which will assess cumulative effects of temperature cycling or spike in conjunction with moisture exposure on the mechanical performance of composites. As discussed, one of the first questions to resolve was whether the effect of thermal cycling can be represented by cumulative sum of thermal degradation effect at peak temperature. Within this context, the in-plane shear strength of unpostcured laminates of graphite fiber/cyanate ester resin composites was measured after static exposure at 150°C as well as temperature cycling to 150°C for various durations up to 5 days. The in-plane shear strength of postcured laminates was measured after static exposure as well as cycling to the heat of 204°C. Both test results were plotted as a function of exposure time at 150 or 204°C. As shown in Figures 6 and 7, the in-plane shear strength of IM7 graphite fiber/954-2 cyanate ester resin composites decreased steadily on the exposure to dry heat. The rate of strength degradation on static exposure to heat (about 4% per day) was found to be close to that of thermal cycling expressed as a function of time at peak temperature. In the case of postcured specimens, a slight deviation occurred within the range of data scattering.

From the above test results, it is reasonable to conclude that, at a frequency of 10 min/cycle and for a relatively short duration, the effect of thermal cycling can be represented by the cumulative sum of thermal degradation effect at peak temperature. This conclusion was further supported by the results of ensuing morphology study. The study based on fluorescent dye-enhanced optical microscopy revealed that any of the specimens subjected to static heat or thermal cycling do not contain microcracks (Figure 8). Obviously the degradation of materials occurred at a molecular level. The same mechanisms of strength degradation can be assumed under static heat as well as thermal cycling. However, the situation may be altered under thermal cycling condition at higher frequency range and/or longer durations. Our future program will examine this possibility and, if necessary, define a critical frequency above which thermal cycling effects deviates from the cumulative sum of thermal degradation effect at peak temperature.

CONCLUDING REMARKS

Our study plans to assess the long-term structural integrity of graphite fiber-reinforced cyanate ester resin matrix composites. As a first step toward this goal, the weight change and the retention of in-plane shear strength of the composite systems have been measured on the exposure to *high humidity* and *thermal cycling* respectively. Initial test results clearly showed that cyanate ester resin matrix composites absorb a remarkably small amount of moisture on the exposure to 95% RH condition at 60°C up to 36 days. Postcured specimens appeared to follow the same trend with a slightly higher moisture content. One interesting finding was that the degree of moisture absorption undergoes a rather sudden increase to an equilibrium level of 1% after the exposure for around 50 days. The observed pattern of moisture gain clearly indicates that moisture absorption behavior of graphite fiber/cyanate ester resin composites cannot be described by a concentration-dependent form of Fick's law.

The morphology study showed that, on the exposure to moist heat for 50 days, the composite undergoes extensive cracking of matrix/interface region in the form of the delamination between the plies as well as translaminar cracking within the ply. The observed damage of matrix/interface region is believed to be responsible for a sudden moisture gain. Since matrix swelling effect is negligible in the cyanate ester resin matrix composites which initially absorb a very small amount of water, the phenomenon is presumed to be caused by weakening of the fiber-matrix *interface* as in the case of epoxy resin composites. Weakening of fiber-matrix bonding was confirmed by the failure analysis. Rather unusual occurrence of moisture-induced cracks in the absence of external loading might be related to the fact that cyanate ester resins exhibit a large amount of free volume after gelation which led to very low level of cure shrinkage and residual stresses.

A sudden moisture gain associated with extensive matrix/interface cracking was found to reduce in-plane shear strength of composites by 20%. At a given stress amplitude, the composite specimens exposed to the humidity for 50 days suffered an exponential decrease of fatigue lifetime. From the assumed relationship between the slope of S-N curve and the m factor of Paris law, it was hypothesized that the rate of crack growth is higher for wet specimens (already with extensive matrix/interface cracks initiated) than dry specimens. This hypothetical conclusion seems to be consistent with the results of the failure analysis which indicates weakening of fiber-matrix interface due to the absorbed moisture. The fracture surface of wet specimens showed a reduced tendency of having the deposit of

matrix resin on the fibers. When observed, the remnants of matrix resin often showed an appearance of being easily peeled in wet specimens.

In addition to the effects of moisture absorption, the study examined the question of whether the effect of thermal cycling can be represented by the cumulative result of thermal degradation effect. The retention of in-plane shear strength of graphite fiber/cyanate ester resin composites was monitored on the exposure to dry heat at 150 or 204°C. The rate of strength degradation on the static exposure to heat was found to be close to that of thermal cycling expressed as a function of time at peak temperature. At a frequency of 10 min/cycle and for a relatively short duration, the effect of thermal cycling seems to be represented by the cumulative sum of thermal degradation effect at peak temperature. None of the specimens subjected to static heat or thermal cycling exhibited microcracking, suggesting that the degradation of materials occurred at molecular level.

Acknowledgements

I wish to thank the Materials Directorate of Air Force Wright Laboratory and the Air Force Office of Scientific Research for sponsorship of this research. The Research and Laboratories, Inc. helped us in all administrative aspects of this program.

I sincerely appreciate continuing support and encouragement for this research work from Capt. Michael Holl, Mr. Kenneth Johnson, and Mr. Roger Griswold at the Non-Metallic Materials Division of the Materials Directorate. Mr. Tom Witman, a summer student from the Wright State Univ., provided a valuable help throughout the research program.

I am also very grateful to Dr. Ran Y. Kim at the Univ. of Dayton Research Institute, and Messrs. Brett Bolan and Terry Christiansen at the Materials Directorate for their advices for the research program.

Finally my sincere thanks should go to Messrs. Ron Cornwell, Ron Esterline and Ken Lindsay at the Univ. of Dayton Research Institute who played an essential role in this program by preparing the specimens and performing various physical testing.

References

[1] Herz, J., "Moisture Effects on the High Temperature Strength of Fiber-Reinforced Resin Composites", Proc. of 4th SAMPE Nat'l Tech. Conf., 1 (1972).

[2] Browning, C. E., "The Effects of Moisture on the Properties of High Performance Structural Resins and Composites", Proc. of 28th SPI Reinforced Plastics/Composite Inst. Annual Tech. Conf., Sect 15-A (1973).

[3] Plueddemann, E. P., ed., Interfaces in Polymer Matrix Composites (Composite Materials, Vol.6), Academic Press, New York, NY (1974).

[4] Springer, G. S., ed., Environmental Effects on Composite Materials, Vol. 1-3, Technomic Publishing Co., Westport , CT (1981, 1984, 1988).

[5] Adams, D. F., Environmental Effects on Composite Materials (Seminar Notes), Technomic Publishing Co., Lancaster, PA (1993).

[6] Ashbee, K. H. G., Fundamental Principles of Fiber-Reinforced Composites, Chapter 10, Technomic Publishing Co., Lancaster, PA (1989).

[7] Wolff, E. G., "Moisture Effects on Polymer Matrix Composites", SAMPE J, Vol. 29, No. 3, 11 (1993).

[8] Herakovich, C. T., J. G. Gavis, and J. S. Mills, "Thermal Microcracking of Celion 6000/PMR15 Graphite/Polyimde", in Thermal Stresses in Severe Environments edited by D. P. H. Hasselman and R. A. Heller, Plenum Press, New York, NY (1980).

[9] S. S. Tompkins and S. L. Williams, "Effects of Thermal Cycling on Mechanical Properties of Graphite/Polyimide Composites", *J. Spacecraft*, Vol. 21, No. 3, 274 (1984).

[10] McConnell, V. P., "Tough Promises from Cyanate Ester", *Advanced Composites*, Vol. 7, No. 3, 28 (1992).

[11] Shimp, D. A., J. R. Christenson, and S. L. Ising, "Cyanate Esters - An Emerging Family of Versatile Composite Resins", Proc. of 34th Int'l SAMPE Symposium, 222 (1989).

[12] Almen, G., P. Mackenzie, V. Malhotra, and R. Maskell, "Toughened Cyanates for Aerospace Applications", Proc. of 35th Int'l SAMPE Symposium (1990).

[13] Lee, F. W., M. A. Boyle, and P. Lefebvre, "High Service Temperature, Damage Tolerant Prepreg Systems Based on Cyanate Chemistry", Proc. of 35th Int'l SAMPE Symposium (1990).

[14] Yang, P. C., D. M. Pickelman, and E. P. Woo, "A New Cyanate Matrix Resin with Improved Toughness: Toughening Mechanism and Composite Properties", Proc. of 35th Int'l SAMPE Symposium, 1131 (1990).

[15] Yang, P. C., E. P. Woo, S. A. Laman, J. J. Jakubowski, D. M. Pickelman, and H. J. Sue, "Rubber-Toughened Cyanate Composites: Properties and Toughening Mechanism", Proc. of 36th Int'l SAMPE Symposium, 437 (1991).

[16] Speak, S. C., H. Sitt, and R. H. Fuse, "Novel Cyanate Ester Based Products for High Performance Radome Applications", Proc. of 36th Int'l SAMPE Symposium, 336 (1991).

[17] Cinquin, J., and P. Abjean, "Correlation between Wet Ageing, Humidity Absorption and Properties of Composite Materials Based on Different Resins Family", Proc. of 38th Int'l SAMPE Symposium, 1539 (1993).

[18] Hahn, H. T., "Residual Stresses in Polymer Matrix Composite Laminates", J. of Composite Materials, Vol. 10, 266 (1976).

[19] Lee, B. L., R. W. Lewis, and R. E. Sacher, "Effects of Static Immersion in Water on the Tensile Strength of Crossply Laminates", Proc. of the 2nd Int'l Conf. on Composite Materials, 1560 (1978).

[20] Browning, C. E., G. E. Husman, and J. M. Whitney, "Moisture Effects in Epoxy Matrix Composites", in Composite Materials: Testing and Design (4th Conf.), ASTM STP #617, 481, ASTM, Philadelphia, PA (1976).

[21] Tetelman, A. S., and A. J. McEvily, Fracture of Structural Materials, p.371, John Wiley, New York (1967).

[22] Mandell, J. F., J. P. F. Chevaillier, K. L. Smith, and D. D. Huang, "Fatigue of PVC and Polysulfone", M. I. T. Research Report R84-1, M. I. T., Cambridge, MA (1984).

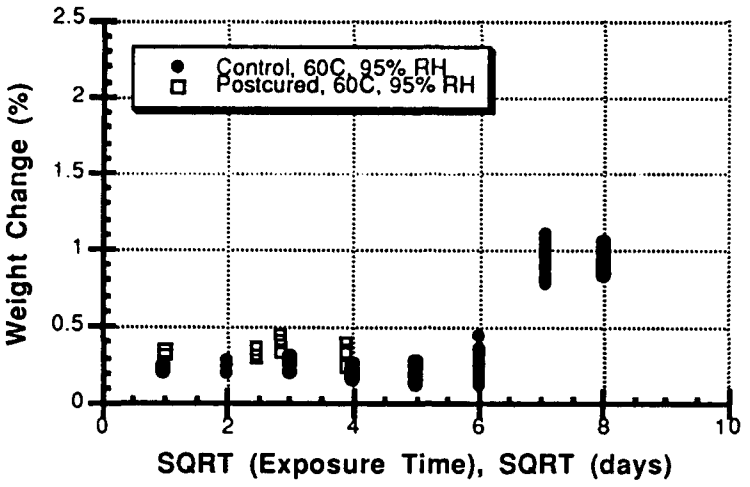


Figure 1. Moisture absorption of graphite fiber-cyanate ester resin composite on the exposure to 95% RH condition at 60°C



Figure 2. Cracking of matrix/interface region of graphite fiber-cyanate ester resin composite on the exposure to 95% RH condition at 60°C for 50 days

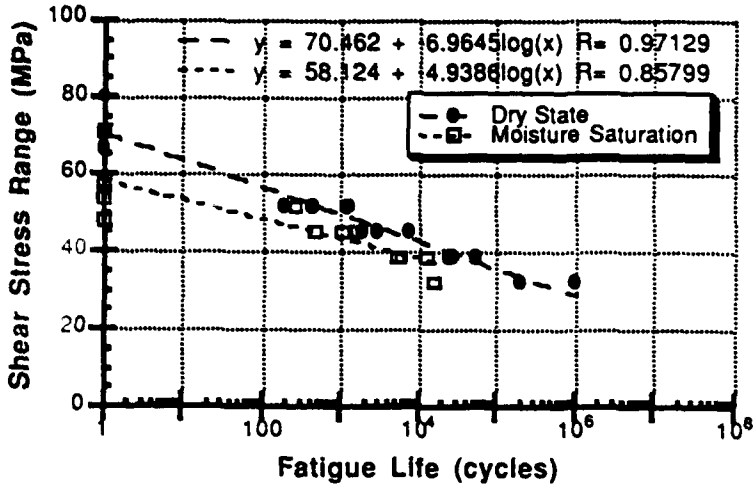


Figure 3. S-N curves of dry and wet (50 days, 95% RH, 60°C) specimens of graphite fiber-cyanate ester resin composite; Semi-log scale.

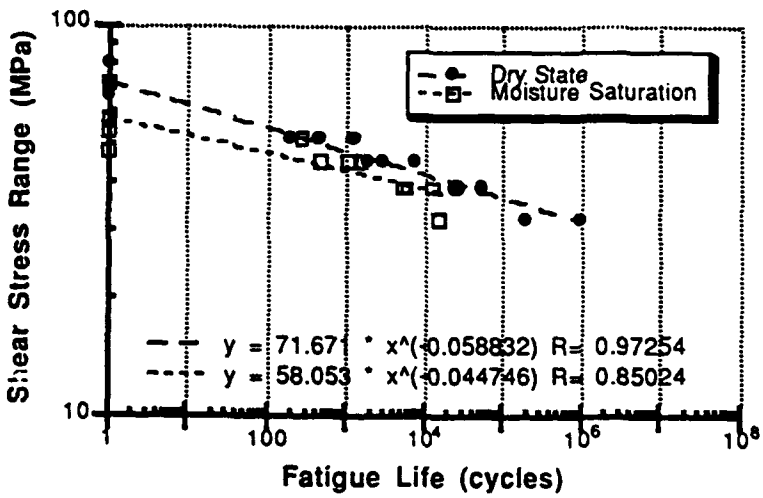
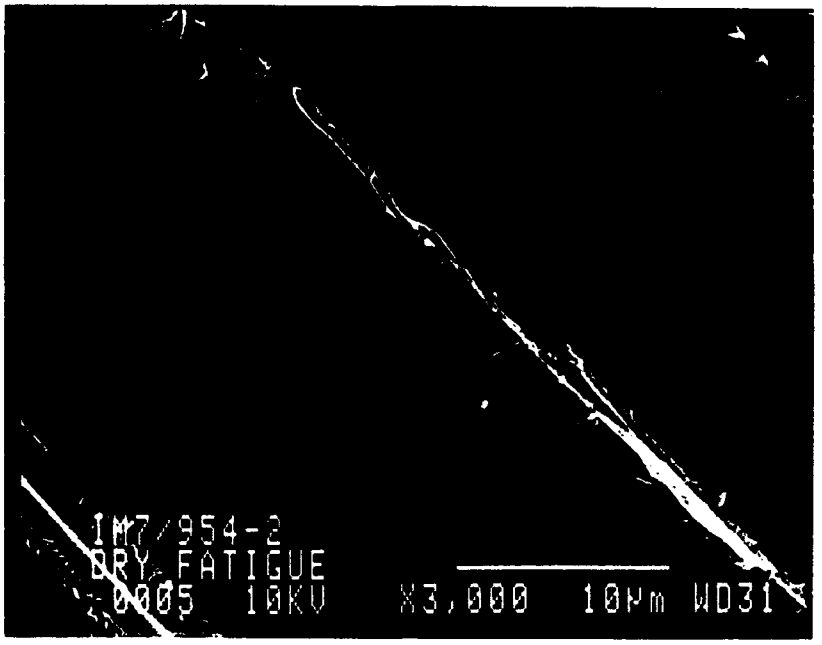


Figure 4. S-N curves of dry and wet (50 days, 95% RH, 60°C) specimens of graphite fiber-cyanate ester resin composite; Log-log scale.

(a)



(b)

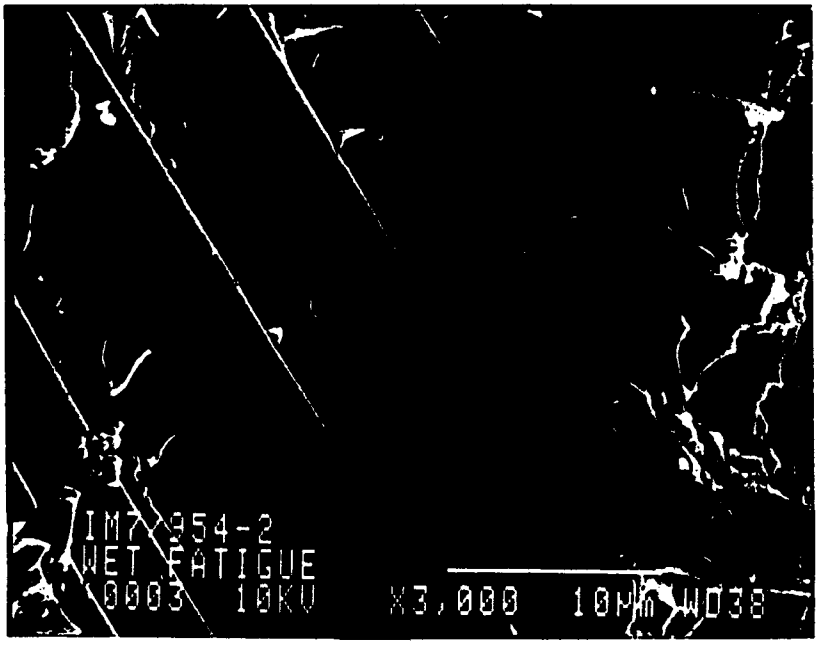


Figure 5. The fracture surfaces of graphite fiber-cyanate ester resin composite specimens: (a) the remnant of matrix resin clinging to the fibers even in a highly-fibrillated region (dry specimen); (b) the remnants of matrix resin with an appearance of being easily peeled (wet specimen).

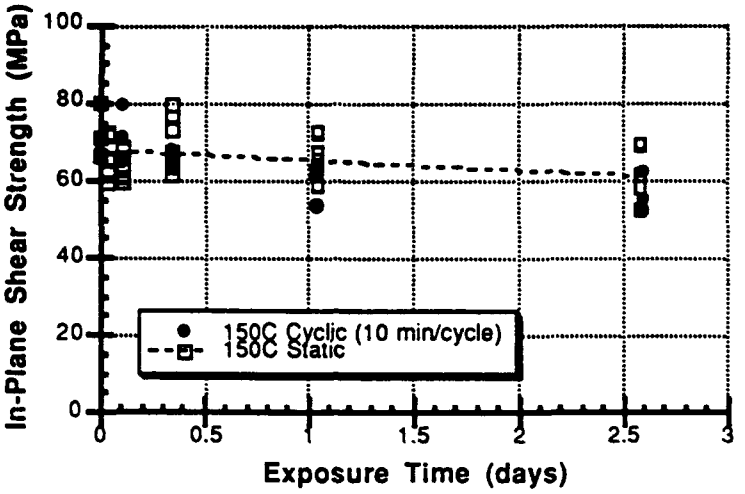


Figure 6. The changes of in-plane shear strength of graphite fiber-cyanate ester resin composites after static exposure vs cycling to 150°C (Not postcured)

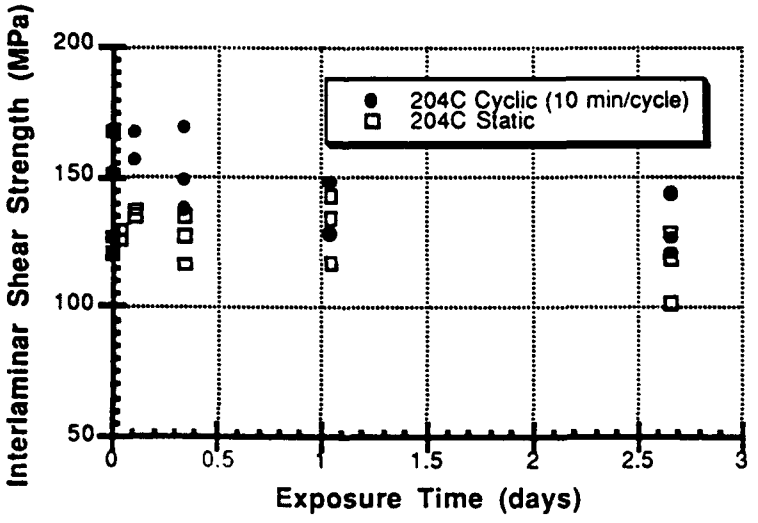


Figure 7. The changes of in-plane shear strength of graphite fiber-cyanate ester resin composites after static exposure vs cycling to 204°C (Postcured)

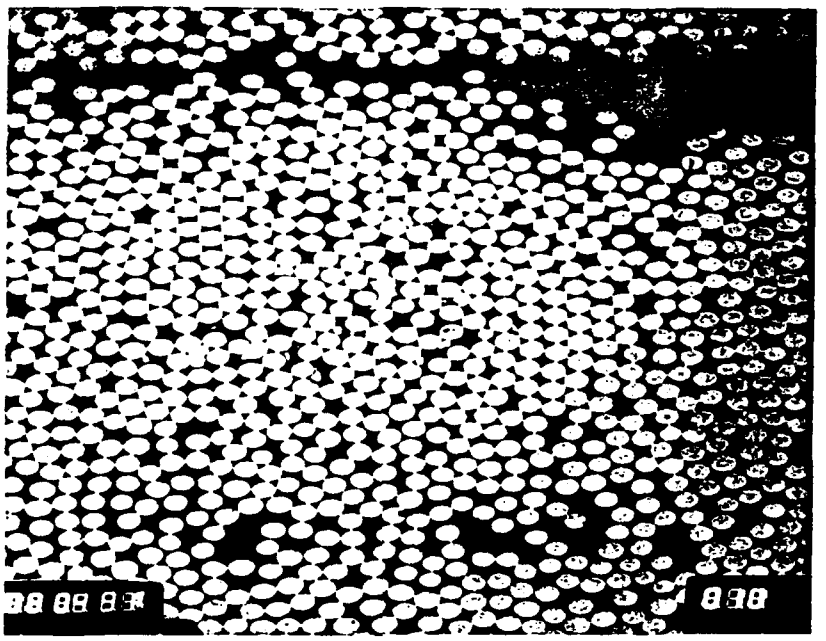


Figure 8. Fluorescent dye-enhanced optical micrograph for the graphite fiber-cyanate ester resin composite specimen subjected to thermal cycling to 150°C

AN ERGONOMIC STUDY OF AIRCRAFT SHEETMETAL WORK

Paul S. Ray
Assistant Professor
Department of Industrial Engineering
The University of Alabama
Tuscaloosa, AL 35487

and

Gina Masterson
Graduate Student
Department of Industrial Engineering
Auburn University
Auburn, AL 36830

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory/OEMO
Brooks Air Force Base
San Antonio, TX

August 1993

An Ergonomic Study of Aircraft Sheetmetal Work

Paul S. Ray
Assistant Professor
Department of Industrial Engineering
The University of Alabama
Tuscaloosa, Alabama 35487

and

Gina Masterson
Graduate Student
Department of Industrial Engineering
Auburn University
Auburn, Alabama 36830

ABSTRACT

Ergonomic risk factors of aircraft sheetmetal tasks were studied at Kelly Air Force Base in San Antonio, Texas. It was observed that most of the tasks performed by the workers involved fairly high risk of developing cumulative trauma disorders (CTDs) of the upper limbs. An ergonomic screening questionnaire filled out by workers indicated a prevalence of mild forms of CTDs among 77% of the workers. Control measures were recommended to reduce progress and/or development of CTDs; a training program in basic Ergonomics was also suggested as a means to make workers aware of CTD and increase adherence to recommended improvements in work methods. When implemented, the recommended control measures will have a positive effect on sheetmetal worker health and help to reduce future compensation claims associated with CTDs.

An Ergonomic Study of Aircraft Sheetmetal Work

Paul S. Ray
and
Gina Masterson

INTRODUCTION:

Currently forty percent (40%) of all occupational illnesses reported to Armstrong Laboratory is related to CTDs. A large number of these CTDs occur in sheetmetal shops throughout the Air Force. In recent years, the Air Force has seen an increase in incidence of cumulative trauma disorders (CTDs) in the sheetmetal shop at Kelly Air Force Base, reported through the Air Force occupational illness data registry. This shop involves extensive manual operations using powered and non-powered hand tools and presents a great potential for causing cumulative trauma disorders.

The objectives of this study were to identify the ergonomic risk factors and determine ways to minimize the incidence of musculoskeletal injuries resulting from the sheetmetal tasks required for aircraft maintenance.

EXPERIMENTAL SETTING:

The sheetmetal workshop is located in building #375 of the Aircraft Directorate at Kelly Air Force Base. It is a large hanger accommodating the maintenance, inspection, and repair of B-52 and C-5 aircraft. The hanger is divided into two main sections: the "big hanger" houses a number of aircraft in various stages of maintenance, and the "backshop" section houses more than twenty-five different work areas. About one thousand military and civilian

employees work in this hanger. The day shift starts at 6:00 A.M. and ends at 4:45 P.M. There is a lunch break of forty-five minutes and two ten minute breaks, resulting in a net work time of five hundred and eighty minutes. A second shift, comprised mostly of volunteer workers, begins at 4:00 P.M. and ends at 2:00 A.M. The schedule has a four-day work week, starting on Monday and ending on Thursday. The work at the shop is self-paced.

DEMOGRAPHIC COMPOSITION OF SHEETMETAL WORKERS:

There are one hundred and forty-four civilian workers in the shop. Out of these, twenty-two workers are assigned on loan to other departments. Seventeen workers are assigned to the second shift. The allocation of workers varies each week to accommodate the variation in workload at the base. Fifteen percent of the workers are female. Eighty-three percent of the workers belong to the age group of thirty to forty nine years and sixty-five percent of them have over ten years of experience at the shop. The workers belong to four ethnic groups, as shown in Table 1.

Table 1: Ethnic Origin Of Workers

ETHNIC GROUP	PERCENTAGE
a. American Indian	1.5
b. Black	4.5
c. Caucasian	18.2
d. Hispanic	75.8
Total	100

METHODOLOGY:

Evaluation of The Baseline Status

The current level of occurrence of cumulative trauma disorders

among the sheetmetal workers was assessed from:
1) historical records, and 2) a survey using an ergonomic disorder screening questionnaire.

Historical Records

A) Safety Office Record: This record is generated from worker complaints evaluated by a military physician and classified as injury cases. The mishap records from October 1991 through April 1993, were reviewed. This report presents cumulative trauma injury cases for six body parts: neck, back, arm, hand including wrist, fingers, and trunk. The incidence rate was found to be comparatively high in recent years as given in Table 2. Most of these reported cases involved back and arm injuries due to overexertion.

Table 2: Incidence Rate of Injury

YEAR	INCIDENCE RATE/100 WORKERS
1991	7.7
1992	7.7
1993	6.9

B) Occupational Illness/Injury Report (AF Form 190): This record is generated from worker complaints evaluated by a military physician, and classified by Military Public Health as occupational illness cases. A review of the last five years' records of AF Form 190 revealed the following facts:

1) The total number of CTD cases reported for all Air Force bases during the last five years (1987-92) is 1,119.

2) The number of sheetmetal workers represent 13% of the total

CTD cases.

3) The numbers of CTD cases reported during 1987-92 by selected (former) logistic bases are as shown in Table 3.

Table 3: Number of CTD Cases Reported During 1987-1992					
BASE	Kelly	Tinker	Hill	McClellan	Robins
NO. OF CASES	12	44	44	27	13

Reporting of CTD cases from other (former) logistics bases is significantly larger than that from Kelly. This fact might be an indication of lack of ergonomics awareness at Kelly.

4) The average number of days on limited duty per CTD case for sheetmetal workers during 1987-92 is given in Table 4.

Table 4: Average Number of Days On Limited Duty 1987-1992				
BASE	Kelly	Tinker	Hill	McClellan
NO. OF DAYS	119	8	15	42

The number of days of limited duty at Kelly might be an indication of much worse disease conditions at this base. Increased awareness and worker training could facilitate identification of the disease and reduce the need of long periods on limited duty.

5) CTD cases classified by type for selected (former) logistics bases during 1987-92 are given in Table 5.

Table 5: CTD Cases by Types for Kelly and Hill AFB				
CTD TYPE	BASE		TOTAL	PERCENTAGE
	Kelly	Hill		
CTS	8	26	34	60.7
Tendinitis	4	18	22	39.3
Totals	12	44	56	100

The incidence of carpel tunnel syndrome (CTS) was found to be the most frequent disease, followed by tendinitis of the upper limbs. This fact indicated the worst problem to be frequent deviation of the wrist while performing the sheetmetal tasks, followed by hyperextension of the arms and shoulder.

Disorder Screening Questionnaire

A survey was conducted to assess the existing cases of CTD among the sheetmetal workers using a disorder screening questionnaire developed by the Air Force. A total of sixty-six people responded to the survey and returned the completed questionnaires. Seventy-seven percent (77%) of workers reported suffering from some kind of cumulative trauma disorder (CTD). Fifty-six percent (56%)of them reported suffering from multiple disorders. The most frequently cited CTD was for the wrists (44%), closely followed by cases of hand disorders (42%), and low back pain cases (36%). Fifty-seven percent (57%) of the workers had introductory training in ergonomics and seventy-four percent (74%) in using back belts.

The survey indicates that the prevalence of CTDs among the sheetmetal workers was very high (77%). However, it appeared that severity of the disease was still within bearable limits and as such had not been reported to the authorities.

The details of CTD cases at Kelly reported during the last five years (AF Form 190) are given in Table 6. The following inferences can be drawn from this table:

1) Twenty-five percent of the total CTD cases reported are females. This level is high in comparison to the percentage of female workers (15%) in the group.
2) There is no incidence of lost days indicating the severity of the cases is within bearable limits.
3) The workers in the age group above 35 years account for most (75%) of the CTD cases. This fact probably indicates that these employees had adequate time to develop CTDs.

Table 6: CTD Cases Among The Sheetmetal Workers at Kelly

Case No.	Gender	Age	Limited Duty Days
1	F	56	2
2	M	49	0
3	M	44	7
4	M	51	30
5	M	37	0
6	M	52	56
7	M	42	273
8	M	34	0
9	M	27	21
10	M	40	62
11	F	34	973
12	F	33	0

IDENTIFICATION OF ERGONOMIC RISK FACTORS:

The approach to identify the ergonomic hazards consisted of taking observations on the shop floor to record the details of typical sheetmetal tasks, and analyzing the details with a focus on ergonomic risk factors.

Most of the sheetmetal tasks were found to consist of drilling, riveting, and countersinking operations using pneumatically operated hand guns. The major sheetmetal tasks observed consisted of repairing the following: 1) B-52 nose cowls, 2) remove skin - wrap panel B-52, 3) install skin - lower wrap panel B-52, 4) install skin - upper wrap panel B-52, 5) align fittings - cowl panel B-52, 6) flight controls C-5, 7) replace web - side cowls C-5, 8) align fittings - side cowls C-5, 9) underside pylon C-5, and 10) inside pylon C-5. Based on this study, six groups of risk factors identified were as follows:

1) Posture: The workers in this shop were found to work in awkward postures during most of their work time, resulting in bent neck, hyperextension of shoulder, hyperextension of arms, and extreme deviations of hand and wrists. These are illustrated in Figure 1.

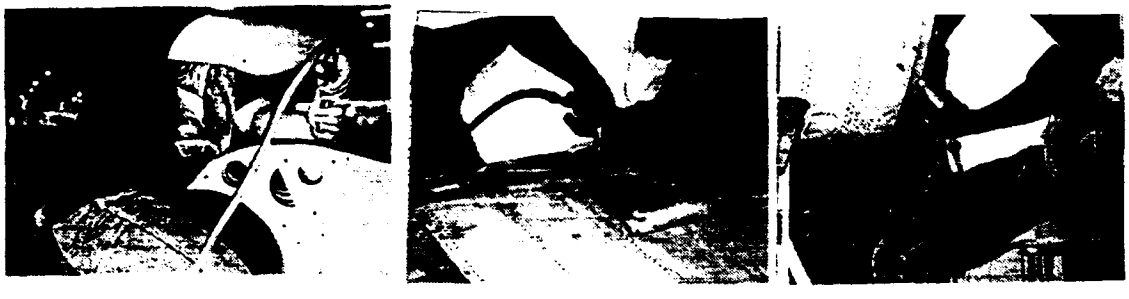


Figure 1: Tasks Performed With Deviated Limbs

Repair work of C-5 pylons required workers to perform overhead work for long periods of time (Figure 2), while other tasks required them to lie down inside the pylon and work in confined space with extremely deviated hands.

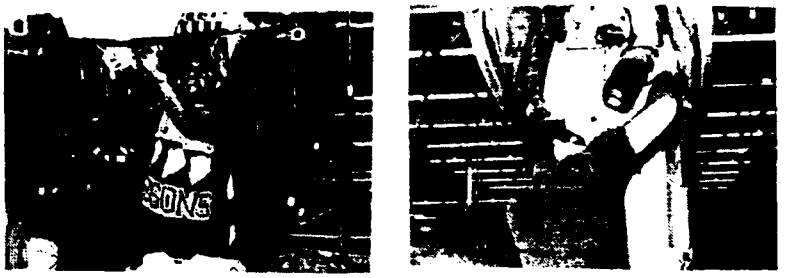


Figure 2: Overhead Tasks at Pylon

2) Force: The sheetmetal tasks required extensive use of pneumatically operated hand guns. These operations required considerable manual force to keep them at specific locations on metal surfaces. The situation was aggravated due to the fact that the hand guns, weighing from two to fourteen pounds, had to be carried manually during operations. Several areas used rivets made of alloy steel (MONEL) which were very hard. Replacing these required drilling through the rivets using significantly heavy force. Most of the pneumatic guns were operated by one finger, either the index finger or the middle finger.

3) Vibration: Another frequent operation, driving rivets, required the use of pneumatic rivet guns. These guns made too much vibration. Only a few workers were found to use hand gloves. Many of the workers had discontinued their use because the gloves supplied did not fit well, or were not useful, while others did not get replacements of worn out ones. There appeared to be no concerted program for selection, usage, and replacement of hand gloves. However, hand gloves alone were not adequate to eliminate the injurious effects of vibration. The guns that we observed in the shop were not provided antivibration type handles and therefore offered no protection from tool vibration. Ear plugs and ear muffs were provided to the workers for protecting their ears from noise exposure.

4) Tool Design: The repair work in the sheetmetal shop required an extensive use of pneumatic tools. The pneumatic tools which are currently in use are operated by only one finger, either the index finger or the middle finger. In addition to this, most of the pneumatically operated tools were not balanced. The center of gravity of the tools should be near their handles. Otherwise, the unbalanced weight of the tools tires the hand and arm muscles quickly. This is particularly true when the arms are extended outward as illustrated in Figure 3. The handles of the existing pneumatic tools consisted of bare metal and exposed hands to skin compression and abrasion from contact with tool surfaces.

5) Mechanical Stress: Manual trimming of metal sheet to size using a large scissor required significant force to be applied by hand and fingers. The scissor handles were not covered with pliant rubber and thus caused significant stress in fingers (Figure 4).



Figure 3: Tasks with Extended Arms

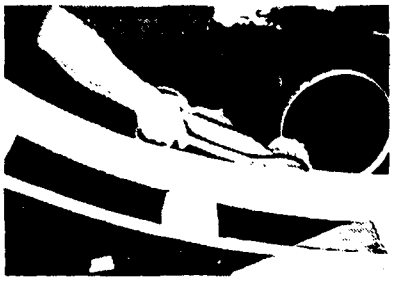


Figure 4: Using Scissor

6) Repetitiveness: This shop practiced self-paced work, but it was found that the workers stopped only when they felt pain, when the task was complete, or when they had to make a tool change. They were not taking micro-rest pauses consciously to relieve the musculoskeletal stress. Drilling or riveting , at a rate of 10-15 seconds per cycle, continuously for half of an hour or more, if practiced daily is likely to induce CTD.

RISK FACTOR SCORES FOR TASKS OBSERVED

Evaluation of the ergonomic severity of the ten representative sheetmetal tasks was done using the a checklist similar to one developed by Armstrong and Lifshitz (1987). The summary of the analysis is given in Table 7.

Table 7: CTD Risk Factor Scores

CTD Risk Factors	replace o/skin B-52 nose cowl	remove skin B-52 wrap panel	install skin B-52 lower wrap panel	install skin B-52 upper wrap panel	align fitting B-52 cowl panel	flight control c-5	replace web c-5 side cowl	align fitting c-5 side cowl	under-side c-5 pylon	inside c-5 pylon
Physical Stress	0	0	0	0	67	0	0	67	0	0
Force	50	50	50	50	0	50	50	0	50	50
Posture	25	25	0	25	25	25	0	25	25	0
Work Station	33	67	33	33	33	33	67	33	33	33
Repeti-tiveness	0	0	0	0	0	100	0	0	0	0
Tool Design	60	60	60	60	20	60	60	20	60	60
overall score	33	39	28	33	31	39	33	28	33	28

Legend Max Job Stress= 0; Min. Job Stress= 100

The overall score of the risk factors of sheetmetal tasks carried a higher than average risk of developing CTDs. Repetitiveness, work station design, and posture of the workers were judged to carry high risk. The nature of jobs resulted in repetitive motions. The quality of work station design affects the posture of workers during the work. As such, these two factors are

related. Other factors were judged to have an average level of CTD risk. The repair of the C-5 pylon underside involving overhead work was judged to carry very high risk of developing a CTD of all the upper body parts including the neck. While working inside a pylon, the workers were subjected to considerable thermal stress. During the summer months, the temperature inside the pylon was found to be greater than 90 degrees fahrenheit. The workers were required to use coveralls as protection from contaminants (cleaning chemicals, grease, etc.) while working inside the pylons. As such, the workers were subjected to temperature ranges exceeding 100 degrees fahrenheit. In addition, no provision of ventilation was found to exist inside the pylons.

CONTROL MEASURES:

The overall control measures consist of: a) training the workers in the basics of ergonomics, b) improving the design of work stations to facilitate less stressful postures, c) using ergonomically designed hand tools, and d) implementing administrative procedures that support the prevention of cumulative trauma disease among the workers.

Control measures appropriate for a number of operations observed on the shop floor are detailed below.

REPETITIVE MOTIONS

1) Rest Pauses: Train the workers to take micro-rest pauses and not to continue drilling or riveting for more than five minutes at a stretch. A rest pause of a few seconds will help to reduce musculoskeletal stress significantly. The job is self-paced and small rest pauses are not expected to reduce productivity at all.

2) Mix Work Locations: Revise work method by shifting the work area during drilling and riveting from lower to upper level and left hand to right hand side, and vice-versa of the work piece every five minutes. This step will shift work stress from one group of muscles to another and thus further limit the musculoskel-

etal stress.

WORK STATION AND POSTURE

In general, provision of adjustable work tables and chairs is required to improve postures during work. In addition, some specific measures are suggested for the ten components observed.

3) B-52 Nose Cowl: Provide step-stool (Figure 5) with height adjustable from twenty to thirty-two inches for sheetmetal repair work of the B-52 nose cowls. Currently, the nose cowl is positioned on a thirty inch high wooden stand. The repair work is to be performed sitting on the step-stool for the lower areas and standing on the step-stool for the higher areas.

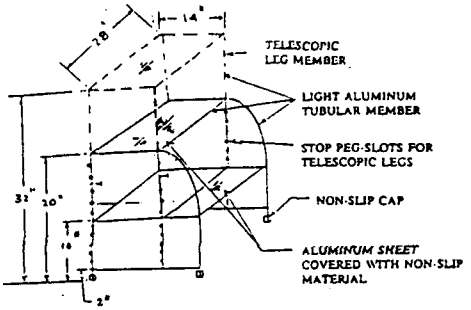


Figure 5: Step-stool

4) B-52 Nose Cowl: Mechanize the operation of cleaning the nose cowl. The suggested fixture should have the ability to rotate the cowl in any direction and jolt it to shake out any loose material left inside the cowl from the repair work.

5) B-52 Lower Wrap Panel: Modify the assembly fixture (F/S 56000) so that the wrap panel on the fixture can be rotated to bring all work areas within easy reach of the workers.

6) B-52 Upper Wrap Panel: Provide step-stools (Figure 5) with height adjustable from twenty to thirty-three inches for sheetmetal repair work of wrap panels on the fixture (#MBPSB/46511). For low level work areas, perform the work sitting on the stool and for high level work areas, perform the repair standing on the step-stool.

7) B-52 Upper Wrap Panel: Use special riveting tool bit with inclined stem for low and high level work area, if required, to bring the wrist into a neutral position (Figure 6).

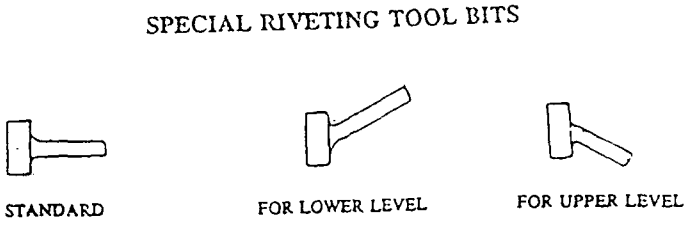


Figure 6: Special Riveting Tool Bits

8) C-5 Flight Controls: Provide adjustable height dollies for the flight control component repair work as shown in Table 8.

Table 8: Suggested Dolly Heights

COMPONENT	WORK-AREA HEIGHT	DOLLY HEIGHT	WORK POSITION
outer flap	21-19 in	21* in	sit
lower rudder	36-46 in	18-28 in	sit/stand
elevator	36-46 in	18-28 in	sit/stand
aileron	36-46 in	18-28 in	sit/stand
outboard elevator	21-31 in	9-19 in	sit

9) C-5 Side Cowl: Modify the assembly fixture (C/N 02688A) so that the cowl mounted on the fixture can be rotated through 180 degrees to bring all of the work areas within easy reach of workers.

10) B-52 Side Cowl: Provide an ergonomic chair (Figure 7) for

the overhead repair work done under the side cowl positioned on the sawhorse (work stand) with the convex side up (Figure 8).

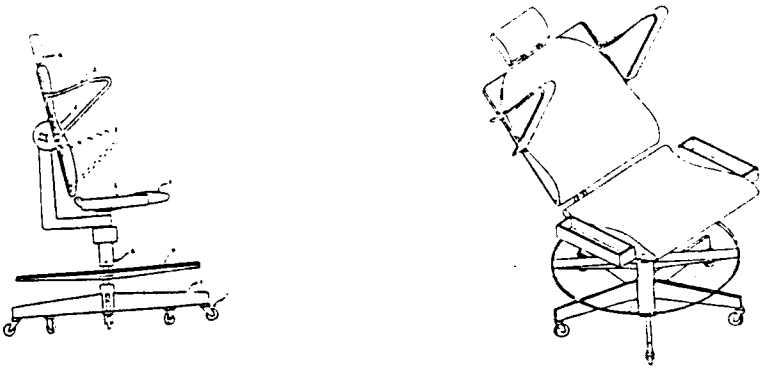


Figure 7: Ergonomic Chair

11) C-5 Pylon: Provide ergonomic work chair (Figure 7) for maintenance and sheetmetal work done under the C-5 pylon. This device will provide the needed support for head, neck, shoulder, and arms at appropriate heights and thus reduce the high risk of developing CTD among the workers.

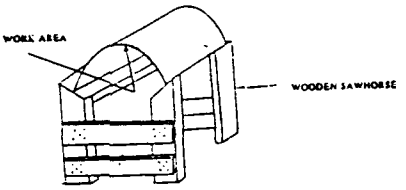


Figure 8: Side Cowl on Sawhorse

12) C-5 Pylon: Provide piped cool air flow to the workers inside the pylon. A good arrangement consisting of an air conditioner unit near the pylon work area supplying piped cool air inside each pylon will alleviate the thermal stress of workers

inside the pylon.

FORCE

13) All Departments: Provide overhead balancing devices to carry the weight of pneumatic tools. These devices will eliminate the static load of carrying the tools during sixty to seventy percent of the task time. See Figure 9.

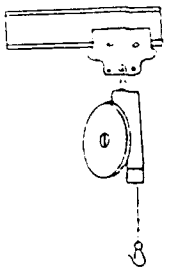


Figure 9: Balancing Device

14) All Departments: Mechanize the manual trimming of sheets to size. A pneumatically operated saw will be adequate. The present method of trimming with hand scissors is stressful to fingers, wrist, and hand.

TOOL DESIGN AND VIBRATION HAZARD

15) All Departments: Use heavier tools to reduce vibration, provided they are suspended on overhead balancing devices.

16) All Departments: Provide antivibration type hand tools and handles.

17) All Departments: Select new pneumatic tools having the following features:

a) Larger triggers suitable for operation by two or more fingers instead of only one finger.
b) Tool handles should be located under the tool's

center of gravity as shown in Figure 10.

c) The tool handles should be of antivibration type and covered with non-slip pliant material.

TYPES OF RIVET GUNS



UNBALANCED GUN BALANCED GUN

Figure 10: Handle Location of Balanced Guns

MECHANICAL STRESS

18) All Departments: Use hand gloves when handling sheetmetal having sharp edges. Bare hands must not be used during these operations.

19) Riveting with the bucking operation is one of the most hazardous tasks in the sheetmetal shop. An effort should be made to design a special riveting gun to eliminate the need for bucking. It will be worthwhile to investigate the possibility of developing special blind rivets that can be used in all areas of aircraft without need of bucking.

PROTECTIVE CLOTHING

20) All Departments: Provide hand gloves to reduce the effect of vibration during riveting, impact strain during bucking, and contact strain with hard surfaces. The hand glove should be suitable for the vibration frequencies of the tools used in the shop and have the following characteristics:

a) padding at palm side and all areas of hand coming into contact with the tool,
b) wrist support band, and
c) covered fingers (finger tips may be exposed for some tasks).

(ER CONTROL STEPS

21) Organize a training program on Ergonomics for the workers and shop supervisors. Video tapes are available from The National Institute of Occupational Safety and Health and are a good training method.

22) Install safety posters at suitable shop locations to keep the workers aware of the ergonomic hazards in the shop.

23) Install periodical review for the tool maintenance program to ensure availability of tools with proper sharpness. This step will help to prevent development of excessive tool vibration.

24) Plan to rotate workers from high risk tasks to low risk tasks periodically.

25) Include screening for ergonomic disorders in occupational clinical exam program for workers of this shop to identify the onset of CTDs, so that corrective steps may be taken at a much lower cost and avoid compensation claims in the future.

DISCUSSION:

The awareness of ergonomic risk factors at the workplace has increased significantly during recent years. The manifold increase in the number of CTD cases reported reflects this awareness among workers. This trend is likely to continue in the near future and is likely to result in significant compensation claims. As such, it will be economically advantageous to implement an ergonomic intervention program to prevent progression of CTD among workers and thus reduce the risk of exposure to compensation claims in the future.

As sheetmetal work for all aircraft maintenance work is very similar in nature, the current study may be applicable to other Air Force bases as well. As such, the potential benefits of this study are much larger and need not be limited to Kelly only. The ergonomic chair proposed for the C-5 pylon maintenance can be beneficially used in similar maintenance jobs of other aircraft and is expected to have universal applicability.

CONCLUSION:

Most of the sheetmetal repair tasks at Kelly carry an average to high level risk of developing CTDs. Twenty-five control steps have been suggested to minimize the ergonomic risk factors from the current work methods followed on the shop floor. A large percentage of the workers at present are suffering from some sort of cumulative trauma disorder, affecting various body parts. The reporting of such incidence has been low, so far, probably due to very little awareness of these job-related diseases. Implementation of a concerted ergonomic intervention program now will be timely, to prevent the progression of CTDs among the workers and help in reducing the potential for significant compensation claims in the near future.

REFERENCES

Armstrong, Thomas. J., and Litshitz. Y.,1987. Evaluation and Design Of Jobs for Control of Cumulative Trauma Disorders. In Ergonomic Interventions To Prevent Musculoskeletal Injuries in Industry (American Conference of Industrial Hygienists), Lewis Publishers, Michigan. pp 73-85.

Taboun, S. M., and Dutta, S. P., 1993. Ergonomic Analysis and Evaluation of Electronic Assembly Operations. In: Advances in Industrial Ergonomics and Safety V. Edited by Nielsen, R. and Jorgensen, K., Taylor & Francis, London. pp 29-36.

EVALUATION OF HEXAMETHYLENE DIISOCYANATE SAMPLING AND ANALYSIS IN SPRAY-PAINTING OPERATIONS

Walter E. Rudzinski
Professor
Department of Chemistry
Southwest Texas State University
San Marcos TX, 78666

Final Report for
Summer Faculty Research Program
Armstrong Laboratory
Occupational and Environmental Health Directorate
Brooks Air Force Base
San Antonio Tx, 78235-5501

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

September, 1993

EVALUATION OF HEXAMETHYLENE DIISOCYANATE SAMPLING AND ANALYSIS IN SPRAY-PAINTING OPERATIONS

Walter E. Rudzinski
Professor
Department of Chemistry
Southwest Texas State University
San Marcos TX, 78666

Abstract

Several methods were used for the analysis of 1,6-hexamethylene diisocyanate, HDI, monomer and oligomer. Samples were collected during several spray-painting operations, and two impinger collection methods were compared with fiber filter, sorbent and total particulate collection methods.

The results demonstrate that the impinger collection methods, NIOSH Method 5521 and a new NIOSH method for Isocyanates, give higher results than those which employ a 1-(2-pyridyl)-piperazine-coated filter (OSHA Method 42) or a tryptamine-coated XAD-2 sorbent tube.

The results also indicate that capillary zone electrophoresis may be used as an analytical tool for the separation of 1-(2-pyridyl)-piperazine-derivatized HDI, and may offer better resolution than HPLC when HDI is in a complicated polyurethane paint matrix.

Finally, the results from total particulate sampling indicate that the concentration of polyisocyanate in air may be estimated (within a factor of two) from the relative amount of hardener in the paint formulation.

EVALUATION OF HEXAMETHYLENE DIISOCYANATE SAMPLING AND ANALYSIS IN SPRAY-PAINTING OPERATIONS

Walter E. Rudzinski

Introduction

Workers who are involved in polyurethane spray-painting may experience adverse health effects which have been ascribed to the activator in the paint formulation. Hexamethylene diisocyanate or HDI is the most-commonly employed activator used in U.S. Air Force polyurethane spray paint operations and may be found either in the form of a volatile monomer or as a polyisocyanate in the aerosol form.

HDI monomer because of its volatility and the reactivity of its NCO groups is an occupational hazard for which the American Council of Governmental Industrial Hygienists, ACGIH, has recommended a threshhold limit value (TLV) of 34 $\mu g/m^3$ (5 ppb) as an 8-hr time-weighted average, TWA[1]. A TLV for the polyisocyanates of HDI, the biuret and the isocyanurate trimer, has not been established even though the polyisocyanates contain reactive isocyanate groups, and in an aerosol form may be inhaled during spray-painting operations[2,3].

Some early reports suggest that polyisocyanates can cause occupational asthma during spray-painting[4-6]. The studies often do not differentiate between the adverse effects caused by diisocyanate monomer and polyisocyanate because of the presence of both in paint formulations. However, a recent report has demonstrated that prepolymers of HDI are unequivocally a cause of occupational asthma[7]. For this reason, a permissable exposure limit (PEL) should be established not only for isocyanate monomer but for all isocyanates regardless of their chemical form, as has already been done in some countries[3,8] and in the state of Oregon[9].

The levels of HDI monomer and polyisocyanate oligomer during spray painting are significant. A ten year survey in the state of Oregon indicated that 6 % of all the air samples were higher than 0.02 ppm HDI monomer which is higher than the PEL established in the state of Oregon; while an even higher, 42%, of the samples exceeded the Oregon PEL of

1 mg/m^3 for HDI polyisocyanates[9]. There is a need therefore to ensure that methods used for the evaluation of isocyanate monomer and polyisocyanates in the workplace give accurate and reproducible results.

Currently most methods for the analysis of isocyanates have been developed for the monomer. In these methods the sample is collected with either an impinger or filter, derivatized, then separated using high performance liquid chromatography, HPLC, with either ultraviolet absorbance (uv), fluorescent(fluor) or electro- chemical (ec) detection. The entire area has been the subject of various reviews[10-12].

In an effort to devise simpler sampling and analysis procedures for both the monomer and the polyisocyanates we have compared and in some cases modified four different methods for the determination of HDI:

1. NIOSH Method 5521[13] which uses an impinger filled with 1-(2-methoxy phenyl)-piperazine in toluene for collection and derivitization of the sample. The analyte is then separated using HPLC with simultaneous uv and ec detection. This method is the best available method for HDI polyisocyanates and has been used as a "benchmark" for developing new methodologies.
2. A new,draft NIOSH Method: Isocyanates[14] which uses an impinger filled with tryptamine in DMSO for the collection and derivitization of diisocyanate monomer and oligomers, HPLC separation, and simultaneous fluorescence and ec detection. The method has been extended to include sampling with a tryptamine-coated XAD-2 resin, and the results have been compared with those obtained from sampling with an impinger filled with tryptamine in a DMSO solution[15].
3. OSHA Method 42[16] which involves the collection of sample by drawing air through a fiber filter coated with 1-(2-pyridyl)-piperazine. The derivatized isocyanate is then analyzed by using HPLC. The method has been compared with NIOSH 5521 in order to see whether filter sampling can provide results comparable to those obtained using an impinger. The analytical procedure has also been modified to se whether capillary zone electrophoresis offers any advantages for the separation and analysis of HDI monomer.

4. NIOSH Method 0500[17] which involves the collection of total particulates by drawing air through preweighed filters. The method has been used to determine polyisocyanate based on the total mass of particulates and the relative weight percent of polyisocyanate in the original paint formulation.

Experimental

Reagents

1-(2-methoxyphenyl) piperazine (MOPIP) from Fluka; 3-(2-aminoethyl) indole (tryptamine) from Sigma.1-(2-pyridyl)piperazine (PPIP) and 98% pure HDI monomer were obtained from Aldrich Chemical Co. (Milwaukee, Wis.). Desmodur N-75 which is 35-40% biuret trimer of HDI and 35% polyisocyanate in xylene (MSDS data sheet) was obtained from Miles Chemical Co. (Pittsburgh, PA).The hardener used for preparing tryptamine-derivatized polyisocyanate standards was obtained from Deft (Irvine,CA) and contained 60% aliphatic diisocyanate and 40% organic solvents. XAD-2 was purchased from Supelco (Bellefontane, PA). All other chemicals and solvents were reagent grade or better.

PPIP, MOPIP and tryptamine were used as derivatizing agents in preparing HDI urea standards as described in OSHA Method 42[16], NIOSH Method 5521[13] and NIOSH Method: Isocyanates[14] respectively. MOPIP-derivatized polyisocyanate standards were prepared from Desmodur N-75 according to NIOSH Method 5521. Tryptamine-derivatized polyisocyanate standards were prepared according to NIOSH Method: Isocyanates from the Deft hardener.

Sampling Apparatus

Constant-flow personal air-sampling pumps capable of drawing up to 2.0 L/min., and standard midget impingers filled with MOPIP or tryptamine solutions were used to collect samples according to NIOSH Method 5521 and NIOSH Method:Isocyanates protocols. Solid sorbent tubes (5 mm i.d.) were filled up to a height of 3 cm with tryptamine coated on XAD-2 resin (0.1% w/w)[15,18] Glass fiber filters were coated with 2.0 mg of 1-(2-pyridyl) piperazine) and then mounted in 37 mm cassettes.Air samples were collected open-faced at a flow rate of 1.0 L/min. Total particulates were collected open-faced according to NIOSH Method 0500[17] at a flow rate of 2.0 L/min.

Instrumentation

The HPLC system consisted of a Hewlett-Packard Series 1090 chromatograph with autosampler, and diode array uv-vis absorbance detector. A Hewlett-Packard 1049A electrochemical detector operated in the amperometric mode at +0.8 V was used to detect the MOPIP and tryptamine derivatives. An ABI 90 fluorescence detector operating at an excitation wavelength of 275 nm was used to detect the tryptamine derivatives. The fluorescence detector was not equipped with a monochromator to limit the bandpass of the emission and no filters were used, and therefore the emission intensity was detected over the entire range,190-700 nm, of the instrument.

The columns used were either Hypersil ODS 5µ (NIOSH Method 5521) or LiChrospher 100 RP 18 5µ (OSHA Method 42 and NIOSH Method:Isocyanates). The mobile phases were those specified in NIOSH Method 5521, NIOSH Method Isocyanates or OSHA Method 42. In the analysis of polyisocyanate, the mobile phase was sometimes adjusted to a higher concentration of organic modifier in order to shorten the retention time.

Hewlett-Packard 3396 series II integrators were used to quantitate the chromatographic peaks. The concentration of HDI monomer was determined by comparing the integrated area with that of authentic standards. The concentration of polyisocyanate was determined by comparing either the area of the largest peak (attributable to biuret trimer) or the total area under the three largest peaks in the chromatogram with that of a series of standards.

A Waters Quanta 4000 CZE system equipped with a Waters 730 data module was used to analyze for HDI. All samples were injected using hydrostatic injection for 8 sec. The height differential of the reservoirs was 9.8 cm. The capillary column had an effective separation length of 50.3 cm and an inner diameter of 75 µm. The total column length was 57.8 cm. The operating voltage was set at 20 kV. Either 0.1 M sodium acetate or 0.01 M sodium phosphate buffers adjusted to a pH of 3.0 were employed. The background current through the capillary was 75 µA when using 0.01 M phosphate buffer. The detector wavelegth was set at 254 nm. The column was reconditioned daily with 0.5 M NaOH, and the capillary was purged with buffer for 2 minutes between each analysis.

Field Operation, Keesler AFB

Six different spray paint operations were evaluated. Operations 1 and 4 involved spray-painting of a truck and trencher respectively in a paint booth. Operations 2, 3, 5 and 6 involved spray-painting wheels, signs, a generator and aircraft wing parts respectively. High volume low pressure (HVLP) spray guns were used throughout. Area samples and personal samples were collected in the area of overspray and in the breathing zone respectively. Area samplers were 3-4 ft above the floor and about 2-5 feet downdraft from the equipment being painted.

Two different Deft polyurethane paint formulation were used. Both formulations contained pigment: hardener in a 1:1 ratio.

Field Operation, Tinker AFB

The spray painting operation involved the painting of a B-52 bomber. Six hours were spent painting the aircraft, and the remaining two, cleaning the Kraco Pro AA 4000 electrostatic paint guns. Four personal samples were collected. Two of these were in the breathing zone of the manlift operator, while the other two were in the breathing zone of two spray painters. The four area samplers were positioned on the east and west sides of the hanger dock approximately 15 feet from the wing tips of the aircraft.

All sampling was performed using tryptamine-coated XAD-2 sorbent tubes which were positioned next to impingers filled with tryptamine in DMSO.

A Deft polyurethane paint formulation was used which contained pigment: hardener in a 3:1 ratio.

Results and Discussion

Table I. (on the next page) shows the results of spray-painting operations conducted at Keesler AFB. Detectable HDI concentrations range from below the limit of detection to a high of 0.14 mg/m^3. For all spray paint operations, the HDI concentration obtained by using impinger collection (NIOSH Method 5521) is either the same (within experimental error) or higher than that obtained by collection on a filter (OSHA Method 42). The results indicate that 44% of the spray-painting operations exceeded the TLV of 0.034 mg/m^3[1] so that personal protective equipment was needed.

TABLE I. Keesler AFB, Comparison of Diisocyanate Concentrations Obtained Using Impinger, Filter and Total Particulate Methods.

op	type	HDI (mg/m^3)		N-75 (mg/m^3)		T.P.
		NIOSH	OSHA	NIOSH	Ratio	(mg/m^3)
1	personal	0.01	-----	2.4		
	personal	0.003	-----	1.2		
	area	0.027	-----	6.5		
2	personal	0.017	<0.013	3.0		
	area	0.015	0.014	1.5		
	personal	0.022	<0.028	2.6		
	area	0.020	<0.059	3.0		
3	personal	<0.016	-----	1.0	0.8	4.0
	personal	0.007	-----			
	area	0.14	<0.026	1.6	2.9	15.2
	area	0.004	<0.016			
4	personal	0.11	-----	5.6		
	personal	0.12	-----	5.4		
	area	0.12	<0.012			
	area	0.023	<0.015	4.1	3.2	12.0
5	personal	0.14	-----	1.9	2.2	8.0
	area	0.059	0.025	2.8	4.3	15.8
6	personal	0.11	-----	2.3		
	area	0.043	<0.037	2.5	2.1	7.8
	blank	<0.0001	<0.001			

op = operation; ratio = concentration of N-75 polyisocyanate based on weight percent hardener and non-volatiles in paint

T.P. = total particulates

The N-75 polyisocyanate concentration in air during spray-painting operations as determined by NIOSH Method 5521 ranged from 1.0 - 6.5 mg/m^3. At present there are no OSHA permissible exposure limits (PELs) or short-term exposure limits (STELs), no NIOSH recommended exposure limits (RELs), and no American Conference of Governmental Industrial Hygienists (ACGIH) threshold limit values (TLVs) for polyisocyanates. The manufacturer (Miles) recommends a 1 mg/m^3 STEL and a 0.5 mg/m^3 8 hour TWA exposure limit, while the Swedish standard lists a 5-min STEL of 0.2 mg/m^3 and a TLV-TWA of 0.09 mg/m^3[8]. The Oregon OSHA standard has a ceiling standard of 1 mg/m^3 and a 8-hr TWA of 0.5 mg/m^3[9]. The spray-painting operations at Keesler AFB exceed the ceiling standard of 1 mg/m^3 for polyisocyanates in every case. These results underline the importance of analyzing for polyisocyanates especially since a definite link between occupational asthma and isocyanates has been established[4-7,19,20].

In an effort to find a simple method to analyze for polyisocyanate, total particulate concentrations were determined.These fell within a relatively small range, 4.0 - 15.8 mg/m^3. The ratio of polyisocyanate to total non-volatiles in the original paint formulation was 0.27 and 0.19. These ratios were used to calculate the mass ratio of polyisocyanate in the paint formulation.The values (se Table I, column 6) were within a factor of two of the NIOSH value.

If a linear regression analysis is attempted between the values of polyisocyanate obtained from NIOSH Method 5521 and the values from the ratio method, then a slope of 0.87 and a correlation coefficient, r, of 0.74 is obtained. This value exceeds the critical value of r = 0.46 for N = 14 samples and indicates a positive correlation[21].

In an effort to confirm the results obtained using OSHA Method 42, several of the samples were re analyzed using a novel capillary zone electrophoresis approach[22]. Table II (on the next page) shows the results obtained for six of the samples. The results from three samples were not used: the area sample from operation 1 exhibited poor precision for the HPLC analysis with up to 100% variability in the integrated area, while area samples from operations 5 and 6 had HDI concentrations which were below the limit of quantitation.

TABLE II. Keesler AFB, CZE and HPLC Results for HDI

Operation	Results (µg HDI)		Air Conc'n (mg/m^3)
	HPLC	CZE	
2	1.00	1.08	0.014
	0.67	0.64	0.059
3	0.57	0.57	0.026
4	0.39	0.47	0.012
	0.39	0.47	0.015
6	0.25	0.30	0.037

The pooled standard deviation for the HPLC data is 0.02 µg (N = 15), while for the CZE data it's 0.04 µg (N = 10). In all cases the results were comparable as measured by a plot of CZE response as a function of HPLC response. The r value was equal to 0.998 and the ratio of responses (slope) was 0.98.

In a comparison of CZE and HPLC, the limit of detection is the same, and the time reproducibility is comparable. Overall, the ability of CZE to analyze for diisocyanate in spray-paint operations is comparable to HPLC except that the CZE method offers 250,000 theoretical plates (calculated using the equation of Lukacs and Jorgenso[22]) as compared with the HPLC approach which provides 2500.

For all field samples, analyzed by CZE, the HDI peak was baseline-resolved from other polyurethane paint components, but for two samples, analyzed by HPLC, components in the paint formulation interfered in the analysis.

The results are promising because the CZE approach provides the same results as can be obtained with HPLC, but CZE consumes little solvent, requires as little as 3 µL of sample and offers baseline-resolution of the derivitized HDI within a complicated polyurethane paint matrix; this is useful whenever the HDI component is difficult to resolve and quantitate.

TABLE III . Tinker AFB, Comparison of Sorbent and Impinger Collection Methods.

sample	type	HDI		polyisocyanate	
		XAD-2	DMSO	XAD-2	DMSO
E1-1	painting	0.014	0.026	0.014	0.045
E2-1	operating manlift	0.005	0.083	0.074	0.160
A1-1	east side	0.021	0.017	0.070	0.085
A2-1	west side	0.014	0.014	0.094	0.164
E1-2	operating manlift	0.009	0.024	0.023	0.039
E2-2	painting	0.014	0.039	0.125	0.105
A1-2	east side	0.014	0.028	0.086	0.120
A2-2	west side	0.026	0.331	0.220	0.039

sample: E prefix denotes personal, A prefix denotes area sampling. The ventilation system was in the exhaust mode for the first four samples, while in the recirculation mode for the last four.
type denotes the type of labor being performed by the worker for personal and the location of the sampler for area sampling.
polyisocyanate concentrations are expressed in terms of mg of HDI/m^3 which would have the same number of isocyanate groups available for derivatization as the polyisocyanate.
XAD-2 and **DMSO** denote the sorbent and the solvent used to dissolve the tryptamine.

Table III. shows the results of spray-painting operations conducted at Tinker AFB. For the analysis of the tryptamine derivative of HDI, desorbed from an XAD-2 resin, the HDI concentration ranges between 0.005 and 0.026 mg/m^3. For the analysis of HDI collected in the tryptamine DMSO solution, the concentration ranges between 0.014 and 0.083 mg/m^3 (excluding the results of sample A2-2 which seem too high). While the concentration of HDI in air as determined by collection on a solid sorbent,

never exceeds the threshhold limit value (TLV) of 34 $\mu g/m^3$ [1], the results based on impinger collection show that two of seven samples exceed the TLV. 87% of the polyisocyanate concentrations exceed the TLV when measured in terms of HDI equivalents.

The relative collection efficiency of the tryptamine-coated XAD-2 sorbent(when compared to collection in a tryptamine DMSO solution) ranges from 6.3% to 126.6% with an average of 52.2%. For the polyisocyanate analysis, the average collection efficiency of XAD-2 is 66.6% when compared to that of the DMSO solution[15].In most cases, the results obtained with a sorbent tube are lower than those obtained by collection with an impinger. These results are in contrast to those obtained by Wu and Graind who found 90% recovery of phenylisocyanate when using tryptamine-coated XAD-2 as a sorbent[18]. The difference in the results can probably be ascribed to the reactivity of HDI which is higher than that of phenylisocyanate.

Conclusions

NIOSH Method 5521 and NIOSH Method:Isocyanates which both use an impinger for sample collection give higher results for the analysis of HDI monomer than either a 1-(2-pyridyl)-piperazine-coated filter (OSHA Method 42) or a tryptamine-coated XAD-2 sorbent tube.

2. When evaluating field data, polyisocyanate concentrations need to be reported as well as monomer concentrations, even though there are no current, regulatory requirements for reporting polyisocyanate concentration. Field studies show that concentrations of polyisocyanate in excess of 1 mg/m^3 are common and therefore may pose a health hazard to spray paint workers.

3. Capillary zone electrophoresis may be used as an analytical tool for the separation of 1-(2-pyridyl)-piperazine-derivatized HDI. CZE offers better resolution of HDI than the validated HPLC approach especially when the HDI is in a complicated polyurethane paint formulation.

4. The concentration of polyisocyanate in air may be estimated (within a factor of two) from the concentration of particulates in air, and the percent hardener and non-volatiles in the paint formulation.

References

1. American Council of Governmental Industrial Hygienists: *TLV's Threshhold Limit Values and Biological Exposure Indices for 1986-1987* . Cincinatti, Ohio: ACGIH,1987.

2. Hardy, H., J. Devine: Use of Organic Isocyanates in Industry. Some Industrial Hygiene Aspects. *Ann. Occup. Hyg. 22* : 421-427 (1979).

3. Silk, S. , H. Hardy: Control Limits for Diisocyanates. *Ann. Occup. Hyg. 27* : 333-339 (1983).

4. Belin L., U. Hjortsberg, U. Wass: Life-Threatening Pulmonary Reaction to Car Paint Containing a Prepolymerized Isocyanate. *Scand. J. Work Environ. Health 7* :310-312 (1981).

5. Malo J.L., G. Ouimet, A. Cartier, D. Levitz, R. Zeiss: Combined Alveolitis and Asthma Due to Hexamethylene Diisocyanate (HDI), with Demonstration of Crossed Respiratory and Immunologic Reactivities to Diphenylmethane Diisocyanate (MDI). *J. Allergy Clin. Immunol. 72* : 413-419 (1983).

6. Nielsen J., C. Sango, G. Winroth, T. Hallberg, S. Skerfving: Systemic Reactions Associated with Polyisocyanate Exposure. *Scand. J. Work Environ. Health.11* : 51-54 (1985).

7. Vandenplas, O., A. Cartier, J. LeSage,Y. Cloutier, G. Perreault, L.C. Grammer, M.A. Shaughnessy, and J-L Malo: Prepolymers of Hexamethylene Diisocyanate as a Cause of Occupational Asthma. *J. Allergy Clin. Immunol. 91(4)* : 850-861 (1993).

8. Tornling, G., R. Alexandersson, G. Hedenstierna, and N. Plato: Decreased Lung Function and Exposure to Diisocyanate (HDI) and (HDI-BT) in Car Repair Painters; Observations on Re-examination 6 Years after Initial Study. *Am. J. Ind. Med. 17* : 299-300 (1990).

9. Janko, M., K. McCarthy, M. Fajer, J. van Raalte: Occupational Exposure to 1,6-Hexamethylene Diisocyanate-Based Polyisocyanates in the State of Oregon, 1980-1990. *Am. Ind. Hyg. Assoc. J. 53(5)* : 331-338 (1992).

10. Melcher, R.G.: Industrial Hygiene. *Anal. Chem. 55* : 40R-56R (1983).

11. Purnell C.J. , R. Walker: Methods for the Determination of Atmospheric Organic Isocyanates. A Review. *Analyst. 110* : 893 (1985).

12. Research and Development Laboratories: Analysis of Isocyanates in Spray Paint Operations by W. E. Rudzinski (RDL 34). Culver City, CA : Research and Development Laboratories, 1992.

13. National Institute for Occupational Safety and Health: Isocyanates (Method 5521) by M.J. Seymour and A.W. Teass. In *Manual of Analytical Methods*. 3d ed. Vol. 2. (Publication No. 84-100). Washington D.C. : U.S. Department of Health and Human Services,1989.

14. National Institute for Occupational Safety and Health: Isocyanates (Method: Isocyanates) by R.J. Key-Schwartz and S.P Tucker. In *NIOSH Manual of Analytical Methods*. 4th ed., 8/15/93 (draft). Washington D.C.: U.S. Department of Health and Human Services,1993.

15. Research and Development Laboratories: Tryptamine as a Derivatizing Agent for the Analysis of Isocyanates in Spray-Painting Operations by R. Sutcliffe (RDL 30). Culver City, CA : Research and Development Laboratories, 1993.

16. Occupational Safety and Health Administration: Diisocyanates. (Method 42). In *OSHA Methods Manual* . 1983. pp 42-1: 42-39.

17. National Institute for Occupational Safety and Health: Total Nuisance Dust.(Method 0500) by K. Morring,J. Clere, F. Hearl.In *NIOSH Manual of Analytical Methods*. 3d ed.
Vol.2.(Publication No. 84-100). Washington D.C.: U.S. Department of Health and Human Services, 1989.

18. Wu W.S., V.S. Gaind: Application of Tryptamine as a Derivatizing Agent for the Determination of Airborne Isocyanates. Part 5. Investigation of Tryptamine-coated XAD-2 Personal Sampler for Airborne Isocyanates in Workplaces. *Analyst. 117* : 9-12 (1992).

19. Tyrer F.: Hazards of Spraying with Two-Pack Paints Containing Isocyanates. *J. Soc. Occup. Med. 29* : 22-24 (1979).

20. Cockcroft D., J. Mink: Isocyanate-Induced Asthma in an Automobile Spray Painter. *Can. Med. Assoc. J. 121* : 602-604 (1979).

21. Hoel P.: *Elementary Statistics*. 4th ed. New York : John Wiley and Sons, 1976. pp 213-235.

22. Lukacs, K.D., J.W. Jorgenson: *J. High Res. Chromatogr. Chromatogr. Commun. 8* : 407 (1985).

EFFECT OF TOLUENE INHALATION ON THE HYPOTHALAMIC-PITUITARY-OVARIAN ENDOCRINE AXIS IN THE CYCLING RAT

Syed Saiduddin, D.V.M., Ph.D.
Professor
Department of Veterinary Physiology and Pharmacology

The Ohio State University
1900 Coffey Road
Columbus, OH 43210-1092

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

The Ohio State University
October 1993

EFFECT OF TOLUENE INHALATION ON THE HYPOTHALAMIC-PITUITARY OVARIAN ENDOCRINE AXIS IN THE CYCLING RAT

Syed Saiduddin D.V.M., Ph.D.
Professor
Department of Veterinary Physiology and Pharmacology
The Ohio State University

Abstract

Menstrual disorders and miscarriages in women exposed to toluene in workplaces have been reported. The male rat has been extensively used to study the effects of toluene on reproductive hormones. The present study was designed to elucidate the effect of toluene inhalation on the estrous cycle and the associated neural and endocrine responses in female F_{344} rats. Three concentrations (100 ppm, 500 ppm and 1000 ppm) of toluene vapor was used. Control group (0 ppm) was exposed to air only. Rats were treated for 4 hours daily for 3 weeks. Daily vaginal cytology was examined to monitor the estrous cycles. Inhalation of toluene for 4 hours daily caused shortening of estrous cycles, elevation of serum enzymes and enlargement of liver. Analysis of various tissues will be performed to determine the effect of toluene on the neural, endocrine, hepatic and reproductive system.

EFFECT OF TOLUENE INHALATION ON THE HYPOTHALAMIC-PITUITARY-OVARIAN ENDOCRINE AXIS IN THE CYCLING RAT

Syed Saiduddin D.V.M., Ph.D.

INTRODUCTION:

Toluene, an aromatic hydrocarbon, is extensively used in the petrochemical, paint and printing industries and is a component of jet fuels and gasoline (1). Such widespread use of toluene, increases the potential risk of exposure of humans to toluene by inhalation of toluene vapors (2). The adverse health effects associated with toluene vapor inhalation are well documented (2,3). To date, the majority of the reports of toluene toxicity have dealt with the neuromuscular, cardiovascular, teratogenic and developmental behavioral aspects of the compound (4-6).

Toluene has also been shown to adversely affect the reproductive system (3). Some clinical reports and epidemiologic studies conducted on industrial workers have revealed evidence of menstrual disorders and miscarriages in women exposed to toluene vapors in the work place (7-11). In addition to occupationally related toluene exposure, toluene abuse has been shown to be a problem in both the teenage and adult female (12). Toluene abuse during pregnancy has been reported to result in a significantly higher incidence of preterm delivery, perinatal death and fetal growth retardation (13-15). The above studies would suggest that toluene may be exerting at least some of its effects on reproduction by altering the neuroendocrine balance in the cycling and pregnant female. To our knowledge, however, there are no citations in the literature which would confirm or refute this suspicion.

Most of the research to date involving the effects of toluene exposure on the endocrine system have been conducted using male rats and the results suggest an alteration of brain catecholamine levels and blood hormone concentrations

following toluene exposure (16-20). The data from the foregoing studies in male rats suggest that exposure to toluene has the potential to cause alteration in the reproductive function by affecting the hypothalamic-pituitary axis resulting from abnormal concentrations of neurotransmitters (norepinephrine and dopamine) and pituitary hormones (prolactin, luteinizing hormone and follicle stimulating hormone). These same neurotransmitters and hormones are also important regulators of female reproduction. It is therefore reasonable to propose that the synthesis and release of pituitary gonadotropins and possibly prolactin may be altered by toluene in the female. These hormones regulate the synthesis and secretion of sex steroid hormones by the ovary and significantly affect ovarian function and the reproductive cycle.

The following study was designed to elucidate the effect of toluene inhalation on the estrous cycle and its controlling neuroendocrine factors in the female rat.

MATERIALS AND METHODS:

The use of animals in this study was approved by a protocol File No. SGO R93-020a of the office of the Air Force Surgeon General and protocol No. 93A0061 of the Institutional Laboratory Animal Care and Use Committee, The Ohio State University.

Female rats (60-70 days of age) of Fischer 344 (F_{344}) strain were purchased from Charles River, Raleigh, NC. After quarantine, the rats were group housed (3 per cage) in clear plastic cages with wood chip bedding. Tap water and feed (Purina Formulab #5008, St. Louis, MO) were available *ad libitum*. Ambient temperature was maintained at 21° to 27°C and lights were on 0600 h for a 14 h light/10 h dark cycle. The animals used in this study were handled in accordance with the principles stated in the *Guide for the Care and Use of Laboratory Animals*, prepared by the U.S. Department of Health and Human Services (1985).

Daily vaginal cytology was examined (between 0600 and 0900 h) to determine the stage of estrous cycle. During the pretreatment period, rats were moved into the 490 L inhalation exposure chambers and kept there for four hours daily to acclimate the rats. Rats showing 2-3 consecutive normal cycles were assigned randomly to one of four treatment groups. Each treatment group contained 20 rats for a total of 80 rats in this study (Table 1). Control (O ppm) rats were placed in an exposure chamber with air only for four hours daily (1000 to 1400 h) for three weeks. The second, third and fourth group of rats were exposed to 100 ppm, 500 ppm and 1000 ppm of toluene vapors, respectively. Concentration of toluene in the chambers was monitored and recorded continuously during the exposure period.

Table 1 - Summary of study design and number of rats

Treatment[a]	Number of Rats
Control (air only)	20
Toluene 100 ppm	20
Toluene 500 ppm	20
Toluene 1000 ppm	20

[a]Rats were killed on the first day of diestrus following three weeks of treatment. Therefore, for some rats, treatment continued beyond the 21-day period until they were sacrificed.

The rats were sacrificed on the first day of diestrous (leukocytic) smear following the three-week treatment. Rats were anesthetized with halothane, blood collected from the posterior vena cava and sacrificed by exsanguination under anesthesia. Blood was allowed to clot at 4°C and centrifuged to separate the serum. Serum alkaline phosphatase (ALKP) and alanine aminotranferase (ALT) were measured in fresh serum samples. The remaining serum was aliquoted and frozen at -80°C for later determination of luteinizing hormone (LH), prolactin, estradiol, progesterone and hippuric acid (principle matabolite of toluene).

Hypothalamus (n = 4/group) was dissected and frozen immediately for measurement of norepinephrine and dopamine concentration. Pituitary (n = 5/group) was fixed *in situ* in Histochoice tissue fixative for immunocytochemistry (LH, FSH and prolactin). Staining index and/or number of LH, FSH and prolactin-positive cells will be quantitated using computer based image analysis (Quantimet 570c).

Livers (n = 4 per group) were perfused with buffer, homogenized and microsomes were prepared and frozen for measurement of UDP glucuronyltransferase activity. The remaining livers were weighed and representative samples were preserved in 10% buffered formalin. Uteri, ovaries and vagina were weighed and also preserved in 10% buffered formalin. Corpora lutea were counted in the ovaries before fixation.

All data presented in this report are means ± SEM. One-way analysis of variance and Fisher's least significant differences test was used to test the differences between control and treatment groups. Probability level of $P<0.05$ was considered significant.

RESULTS:

The average estrous cycle length (4.836 days) in the toluene (1000 ppm) treated rats was significantly shorter when compared to control rats (5.104 days, Table 2). The typical control cycle consisted of three days of leukocytic vaginal smear followed by one day of nucleated and one day of cornified cells. The shortening of the cycle in the toluene treated rats is mainly due to the presence of both nucleated and cornified cells on the same day instead of on two separate days. Also, this condition occurred predominantly in the cycles immediately after the treatment began. The

foregoing observation, i.e., shortened cycles due to a compressed proestrous/estrous periods, may suggest altered endogenous hormones that regulate estrous cycles.

Table 2 - Effect of toluene inhalation on the estrous cycle length in F_{344} rats

	Concentration of toluene			
	0 ppm (control)	100 ppm	500 ppm	1000 ppm
Mean length of estrous cycle Days ± S.E.M.	5.107 ±0.061	4.961 ± 0.067	4.967 ±0.044	4.836* ±0.043

*Significantly different from control value ($P<0.05$).
Vaginal cytology was examined daily between 6 and 9 AM. Rats were exposed to different concentrations of toluene (air only for control group) for 4 hours (10 AM - 2 PM) daily for 3 weeks. There were 20 rats per group.

Body and organ weights are presented in Table 3. With the exception of the ovaries, all other weights were not statistically different from controls. The toluene treated rats had significantly ($P<0.05$) heavier ovaries compared to controls. However, the number of corpora lutea (data not presented) were not different due to treatment. The size of corpora lutea was not assessed. The interpretation of this finding awaits the completion of histopathology of ovaries and the measurement of serum progesterone.

Serum alkaline phosphatase and ALT enzymes were measured as indicators of liver damage (21). Alkaline phosphatase level was not different due to treatment, however ALT level was significantly ($P<0.05$, Table 4) elevated in the 100 ppm and 500 ppm groups when compared to controls. Although the absolute weight of the liver was not different due to treatment (Table 3), when the liver weights expressed as a percent of body weight there was a significant ($P<0.05$, Table 4) increase in liver weight of rats in 500 ppm and 100 ppm groups when compared to controls. Further

interpretation of the elevation in serum ALT and enlargement of liver due to toluene inhalation await the completion of histopathology and measurement of liver microsomal enzymes and the blood steroid hormone determinations.

Table 3 - *Effect of toluene inhalation on body and organ weights of F_{344} female rats*

	Concentration of toluene			
	0 ppm (control)	100 ppm	500 ppm	1000 ppm
Body wt	182.9[a] ±1.0	180.3 ± 1.6	181.8 ±2.4	180.7 ±1.5
Liver wt	5.38 (15) ±0.05	5.22 (16) ± 0.14	5.65 (15) ±0.13	5.61 (16) ±0.09
Brain	1.68 (18) ±0.01	1.69 (19) ± 0.01	1.69 ±0.01	1.68 ±0.01
Adrenal	0.048 ±0.001	0.047 ± 0.001	0.047 ±0.001	0.047 ±0.001
Uterus	0.2.86 ±0.008	0.277 ± 0.006	0.285 ±0.008	0.290 ±0.008
Ovary	0.064 ±0.003	0.076* ± 0.003	0.077* ±0.003	0.076* ±0.003

*Significantly different from controls ($P<0.05$).
Rats were exposed to different concentrations of toluene (air only for control group) for 4 hours daily for 3 weeks. Rats were sacrificed on the first day of diestrus following 3 weeks of treatment.
[a]Values are mean weight (g) ± SEM, n = 20 except where indicated in parentheses.

The following is a list of tissues that need to be analyzed:

1. Blood hormone measurements: Serum samples are frozen and will be used to measure (by radioimmunoassay) the following hormones:

   Luteinizing hormone

   Prolactin

   Estradiol

   Progesterone

Table 4 - Effect of toluene inhalation on serum enzymes and liver weights as a % of body weight in F_{344} rats

	Concentration of toluene			
	0 ppm (control)	100 ppm	500 ppm	1000 ppm
Alkaline phosphatase IU/L of serum	232[a] ± 5.3 (17)	233 ± 8.1 (15)	238 ± 5.8 (19)	232 ± 5.7 (19)
Alanine transaminase IU/L of serum	62.9 ± 2.9 (17)	77.7* ± 4.7 (15)	82.7* ± 6.0 (19)	69.6 ± 2.5 (19)
Liver as a % of body weight	2.943 ± 0.036 (15)	2.915 ± 0.083 (16)	3.111* ± 0.043 (15)	3.127* ± 0.043 (16)

*Significantly different from control value ($P<0.05$).
[a]Values are mean ± SEM, numbers in ( ) are numbers of observations.

2. Immunocytochemistry of the pituitary to quantitate the luteinizing hormone, follicle stimulating hormone and prolactin.
3. Hypothalamic catecholamine measurement (serotonin and dopamine) by HPLC method.
4. Liver histopathology and measurements of microsomal enzyme activity.
5. Histopathologic examinations of the ovary, uterus, and vagina.

It is anticipated that the above studies will be completed in the near future and will permit the interpretation of the overall effects of toluene on the female reproductive system.

CONCLUSION:

Toluene caused significant shortening of the estrous cycle, elevation of serum enzymes and enlargement of liver. These results are considered important findings and emphasize the need to complete the remaining parts of the study (as listed

above) to make proper interpretations about the toxicology of toluene on the female reproductive system. Liver enlargement at 1000 ppm of toluene and the serum ALT increase in lower toluene doses suggest differential dose effects and may involve time dependent hepatic degeneration and hepatic compensatory growth. Considering the short term nature of the study this finding is of significant importance to further investigate the reproductive toxicity of toluene in the female rat by further assessment of endocrine, neural and liver end points.

ACKNOWLEDGMENTS:

The contributions of the following individuals to this project is greatfully acknowledged.

Mr. Richard Godfrey, Mr. Jerry Nicholson and Ms. Peggy Parish for their technical help.

Mr. Edwin Kinkead for his technical and professional advise and guidance during the design and experimental phase of the project.

Mr. Harry Lahey for operation of animal exposure chambers.

Dr. John R. Latendress (LTC) for participation in every phase of this study.

Dr. John C. Lipscomb (Capt) for neurochemical measurements.

Thanks is also extended to many personnel at the Toxicology Division of the Armstrong Laboratory, Wright-Patterson AFB for their encouragement and support.

REFERENCES:

1. Donald, J.M., Hooper, K. and Hopenhayn-Rich, C.; Reproductive and developmental toxicity of toluene: A review environmental health perspectives, 94:237-244, 1991.
2. Low, L.K., Meeks, J. R. and Mackerer, C.R.; Health effects of the alkylbenzenes. I. Toluene Toxicology and Industrial Health, 4:49-75, 1988.

3. Roychowdhury, M.; Reproductive hazards in the work environment; professional safety, 35:17-22, 1990.
4. Baelum, J.; Human solvent exposure, factors influencing the pharmacokinetics and acute toxicity, Pharmacology & Toxicology, 68:7-36, 1991.
5. Baelum, J., Lundqvist, G.R., Molhave, L. and Andersen, N.T.; Human response to varying concentrations of toluene, Int Arch Occup Environ Health, 62:65-71, 1990.
6. Lorenzana-Jimenez, M. and Salas, M.; Behavioral effects of chronic toluene exposure in the developing rat neurotoxicology and tetratology, 12, 353-357, 1990.
7. Matsushita, T., Arimatsu, Y., Ueda, A., Satoh, K. and Nomura, S.; Hematological and neuro-muscular response of workers exposed to low concentration of toluene vapor, Ind Health, 13:115-121, 1975.
8. Linbohm, M.-L., Taskinen, H., Sallmen, M. and Hemminki, K.; Spontaneous abortions among women exposed to organic solvents, American of Industrial Medicine, 17:449- 463, 1990.
9. Axelsson, G. and Rylander, R.; Outcome of pregnancy in women engaged in laboratory work at a petrochemical plant, American of Industrial Medicine, 16:539-545, 1989.
10. Axelsson, G. and Rylander, R.; Exposure to solvents and outcome of pregnancy in university laboratory employees, British of Industrial Medicine, 41:305-312, 1984.
11. Taskinen, H., Anttila, A., Lindbohm, M.L., Sallmen, M. and Hemminki, K.; Spontaneous abortions and congenital malformations among the wives of men occupationally exposed to organic solvents, Scand J Work Environ Health, 15:345-352, 1989.
12. Streicher, H.Z., Gabow, P.A., Moss, A.H., Kono, D. and Kaehny, W.D.; Syndromes of toluene sniffing in adults," Annals of Internal Medicine, 94:758-762, 1981.
13. Wilkins-Haug, L. and Gabow, P.A.; Toluene abuse during pregnancy: obstetric complications and perinatal outcomes, Obstet Gynecol, 77:504-509, 1991.
14. Hersh, J.H., Podruch, P.E., Rogers, G. and Weisskopf, B.; Toluene embryopathy, J. Pediatrics, 106:922-927, 1985.

15. Ron, M.A., Volatile substance abuse: a review of possible long-termneurological, intellectual and psychiatric sequelae, British J. of Psychiatry, 148:235-246, 1986.
16. Andersson, K., Fuxe, K., Toftgard, R., Nilsen, O.G., Eneroth, P. and Gustafsson, J.; Toluene-induced activation of certain hypothalamic and median eminence catecholamine nerve terminal systems of the male rate and its effects on anterior pituitary hormone secretion, Toxicology Letters, 5:393-398, 1980.
17. Andersson, K., Nilsen, O.G., Toftgard, R., Eneroth, P., Gustafsson, J.A., Battistini, N. and Agnati, L.F.; Increased amine turnover in several hypothalamic noradrenaline nerve terminal systems and changes in prolactin secretion in the male rat by exposure to various concentrations of toluene, NeuroToxicology, 4:43-55, 1983.
18. Von Euler, G., Fuxe, K., Hansson, T., Ogren, S.-O., Agnatti, L.F., Eneroth, P., Harfstrand, A. and Gustafsson, J.-A.; Effects of chronic toluene exposure on central monoamine and peptide receptors and their interactions in the adult male rat, Toxicology, 52:103-126, 1988.
19. Von Euler, G., Fuxe, K., Hansson, T., Eneroth, P. and Gustafsson, J.-A.; Persistent effects of neonatal toluene exposure on regional brain catecholamine levels and turnover in the adult male rat, Toxicology, 54:1-16, 1989.
20. Von Euler, G., Ogren, S.-O., Bondy, S.C., McKee, M., Warner, M., Gustafsson, J.-A., Eneroth, P. and Fuxe, K.; Subacute exposure to low concentrations of toluene affects dopamine-mediated locomotor activity in the rat, Toxicology, 67:333-349, 1991.
21. Textbook of Clinical Chemistry. Tietz, N.W. (ed) W.B. Saunders Co., N.Y. 1986, pp 671.

DEVELOPING THE SOFTWARE FOR

SEM-EDXA ANALYSES OF AIRBORNE INORGANIC FIBERS

Larry R. Sherman
Department of Chemistry
University of Scranton
Scranton, Pennsylvania 18520-4626

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory
Brooks Air Force Base, San Antonio, Texas 78235

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

August 20, 1993

Larry R. Sherman

Thomas C. Thomas
Focal Coordinator
Armstrong Laboratories

DEVELOPING THE SOFTWARE FOR SEM-EDXA ANALYSES OF AIRBORNE INORGANIC FIBERS

LARRY R. SHERMAN
Department of Chemistry
University of Scranton
Scranton, Pennsylvania 18510-4626

ABSTRACT

The health hazard posed by airborne asbestos fibers is well documented. Numerous analytical methods for their identification and characterization have also been published but Analytical procedures for other fibers are not as well defined and this project involved using the Amray 1820 Scanning Electron Microscope equipped with an Electron Dispersed X-ray Analyzer to overcome this deficiency in the methods used at the Armstrong Laboratory.

During a previous tenure at AL/OEA, the author and his colleagues used a number of fiber standards to establish a basic fiber identification library. However, the procedures were cumbersome for characterized the fibers and no one except the author was able to determine many fibers.

During the current fellowship the library was expanded and four user friendly computer programs were written in the C++ format for fiber/mineral identification. In preliminary trials, approximately 90% of the material were identified on samples containing more than 30 fibers/100 fields (NIOSH 7400). A base report program has also been prepared so that fiber analyses are a routine procedure for the Al/OEAO asbestos function.

DEVELOPING THE SOFTWARE FOR SEM-EDXA ANALYSES OF AIRBORNE INORGANIC FIBERS

Larry R. Sherman

INTRODUCTION

Although asbestos is currently believed to pose the greatest microscopic fiber health hazard to workers in either the fabricating or users' industries, other inorganic fibers also appear to cause lung diseases (1,2,3,4). When exposed to high temperature, man-made fibers can be converted into cristobalite, a crystalline silica, which is believed to be a carcinogen (2,5) and OSHA has established an exposure limit for fibers with a length to aspect ratio of 3:1. Workers are exposed at this level if they wear no breathing protection equipment when removing and repairing high temperature equipment. The health hazards posed by other inorganic fibers, like fiberglass or ceramics, is much less known and methods for their identification, characterization or lung abatement has not been well studied (5,6). Even though there are no exposure limits for most inorganic fibers, many industrial hygienists have adopted asbestos levels of 0.2 f/cc for an eight hour exposure and 0.0 1 f/cc for a 30 minute exposure for most fibers until more data is available (6).

The normal asbestos procedure at AL/OEAO uses Phase Contrast Microscopy, NIOSH Method 7400, but this procedure can not distinguish non-asbestos inorganic fibers from asbestos fibers or innoxious organic fibers. One of the major problems has been separating acicular minerals, fi-

brous minerals and asbestiform minerals. Furthermore, on the sub-micrometer scale, cleavage fragments can appear as fibrous material which add further ambiguity when TEM is used. On the other hand, assuming that all fibers are inorganic or even upgrading the fiber count to asbestos fibers can be an uneconomical evaluation and require much more effort than justified (7).

During two previous tenures at Brooks AFB, the author had shown that SEM-EDXA is a fast economic method for identifying asbestos fibers with a greater accuracy for amphiboles than a TEM method. He and his coworkers demonstrated that SEM-EDXA can also be used for identifying other airborne materials with a high degree of certainty (8,9); however, the methods developed during the previous tenures were cumbersome at best and required a sophisticated knowledge of computer programing and analytical skills. Thus the methods never became a routinely used one. During the current summer fellowship, the author and his graduate student added over sixty new minerals to the library and reduced the analyses procedures to a routine level.

BACKGROUND

Since the majority of airborne fibers collected and counted by NIOSH 7400 are organic, a lack of identification distorts the true picture of a health hazard. The majority of inorganic minerals in air samples are naturally occurring "dust particles" which depends upon geographic location and environment, i.e. semi-desert and dusty atmosphere contain amosite is a natural asbestos derived from rock weathering.

On the other hand, samples obtained in office building may contain talc from cosmetics, face powders, body powders and some prescription medicines.

Morphology can be used to distinguish a fibrous from a non-fibrous material, but without chemical analysis, gives little insight into the type of mineral in the samples. Furthermore, although morphology can be distinguish between chrysotile and amphiboles, it cannot be used to distinguish individual amphiboles or amphiboles and non-asbestos fibrous materials.

Asbestos contains silicon, magnesium, iron, calcium, manganese, and sodium with aluminum replacing some silicon atoms and manganese replacing some of the iron atoms in the crystal structure. Most natural airborne inorganic materials are calcium, magnesium or aluminum silicates but may contain sulfur, phosphorus or potassium. In rare situations they also contain chlorine (one sample submitted to the laboratory was potassium chloride). Most of these mineral can easily be identified by SEM-EDXA. Man-made fibers, which are now believed to be as dangerous as asbestos (1,3,6,9), are more difficult to identify because of the near identical compositions, i.e. $CaSiO_3$ or $AlSiO_3$.

METHODOLOGY

SEM-EDXA

Standard inorganic fiber samples and minerals were mounted on carbon coated aluminum SEM studs and overlaid with a Au/Pd coating using the Anatech Ltd. Hummer VI as

previously described in the literature (8,9,10,11,12,13). They are analyzed by standard methods and the weight percent used for further chemical analyses (9,10,11,13).

STANDARDS

All seven asbestos types and over ninety non-asbestos materials were analyzed by the SEM-EDXA (9,13). The EDXA data was processed to determine the mean composition and standard deviation to establish a library of mineral analyses. Since EDXA unit can only decipher elements with an atomic number greater than 11, the stoichiometic composition and the library values do not match. Furthermore, SEM-EDXA is a surface analyses procedure and common elemental substitution will also distort the results from the stoichiometic values. Nevertheless, the substitutions are quite reproducible and windows (ranges of composition) have been established from the mole element ratios to quickly identify most materials.

COMPUTER PROGRAMS

Identification of minerals as either fibrous or non-fibrous is extremely tedious with composition tables. To ease the work and use technicians with less chemistry expertise, several user friendly computer programs have been prepared for identifying the materials. To be portable the analytical computer programs were written in C++. They require a 286 computer with DOS 3.3 or higher. Since the programs are long, they are not reproduced in this report but are available from either the laboratory focal coordinator or the faculty fellow.

All programs contain extensive documentation inside the programs. The first 20 -100 lines are devoted to defining variables and special functions. Next a matrix of mineral data appears for the purpose of printing tables of standard minerals, followed by keyboard entries of the X-ray data. The latter are initially processed (molar quantities and molar ratios are determined) to assure that the data fit into the program, they are further manipulated to be compared with the windows established from the table of standards. Finally the data is organized and printed in a readable table so that the chemist can review the information before preparing an analysis final report.

ASBT1.CPP

This program analysis asbestos and is similar to those previously reported (9,12). It compares molar element/silicon ratios and cation/anion ratios based upon the library material and "real" samples. When compiled and mounted in the file manager, the chemist uses the arrow and numerical keys along with the directions printed on the screen to determine if a fiber is asbestos. The program prepares a laboratory report which is stored as ASBFIBER.TXT and can be printed from the file manager.

CALCIUM.TXT

In the previous study (9), calcium was discovered to be a key element in determining real samples. During this project all minerals containing more than 5% calcium, including the calcium asbestos, tremolite, actinolite and ferroactinolite, were put into a single computer analysis

program, CALCIUM.CPP program. A range of compositions based on the molar element/calcium ratio was established for each mineral. Windows were created for each calcium bearing minerals but by the end of this summer project, we had received insufficient base samples to field test all aspects of the program.

To perform an analysis, the operator uses the arrow and numerical keys to enter the data into the program. The laboratory report is stored as CALCIUM.TXT and is accessible from the file manager.

HISIL.CPP

HISIL.CPP program is designed for samples containing more than 10% silicon and less than 5% calcium. Some overlap exists of minerals found within other files. The file contains over 40 minerals including four asbestos, chrysotile, amosite, crocidolite and anthophylite. All weight percent mineral data are stored in a matrix from which the standard tables are prepared. Unknown molar data is manipulated to check against standard molar element/silicon ratios.

The program is accessible with arrow and numerical keys; the laboratory report is stored as HISIL.TXT in the file manager.

MISCEL.CPP

The fourth program MISCEL.CPP will process materials with less than 5% calcium, less than 10% silicon or which contain unusual elements, e.g. copper, tin, zinc or titanium. The analytical computer problems are inherent; each mineral has a different element in highest concentration,

making it very difficult to establish molar element ratios which are meaningful. Although this program was never completely finished, it is accessible through the file manager and the laboratory report is stored as MISCEL.TXT. The program can be used in a limited sense but will have little use, since it contains minerals that are rarely seen in airborne samples. This program required further data manipulation to be fully usable. The author will request a mini-grant to finish this work.

RPT.93

The fifth program was put together for sending the results of a SEM-EDXA analysis to the base submitting a sample. Unfortunately, the usable program is written in basic and is less user friendly than the number crunching programs above. RPT.93 will produce a very satisfactory base report (Appendix A).

A program has been written in C++ for a base report; however, it is too cumbersome and not user friendly. As part of the mini-grant, the author wishes to prepare a report program which is portable and only requires arrow keys for printing. In the mean time, the asbestos function chemists can access the program through the file manager, then press F3 followed by typing B.BAS then press the F2 key. Other information needed to print the report is requested on the screen.

RESULTS

To save time in identifying a fiber, the analyst should inspect the Mg/Si, Na/Si and Ca/Si ratios. If the Mg/Si

ratio is less than 0.15 or greater than 1.0, the sample can not be asbestos. Furthermore, if the Na/Si ratio is greater than 0.45 or the Ca/Si ratio is greater than 0.6, the sample cannot be asbestos and either the Calcium or the Hisil pro grams should be used. If the above mentioned ratios indicate that the sample could be asbestos, the ASBT1 program should first be executed to determine if the fiber is asbestos. If the latter is correct, the asbestos type is printed on the library report.

All programs, as developed in this study, are being used on a routine basis by the asbestos function group chemists. They are processing all samples with more than 30 fibers/100 fields and attaching a "fiber/mineral" analysis report to the NIOSH 7400 results. As the chemists use the programs, they are finding inconsistencies and small errors, The faculty fellow is correcting these as they are brought to his attention and will request a mini-grant to "fine tune" the programs.

INSTRUMENTATION

AL-OEAO possesses an Amray 1820 Scanning Electron Microscope with a Tracor Northern X-ray Analyzer. The instrument is capable of viewing and analyzing fibers with a diameter or length of 0.3 μm or greater. The use of the instrument in fiber analyses has been previously documented (5,7,8). All data processing was performed on an upgraded Zenith 287 microcomputer.

FUTURE WORK

Although the fiber/mineral analysis programs have

reached the level that they can be used in routine analyses of samples which have "action level" concentration of fibers, a great deal of fine tuning is necessary to help the chemists make rational decisions, e.i. in a real sample the mineral was cyanite but the computer printed topaz. The error is due to the very close compositions of the two minerals. The program needs to be reworked so that it prints both minerals and leaves the final decision for the chemist.

ACKNOWLEDGEMENTS.

The P.I. wishes to thank Messieurs T. C. Thomas, A. Richardson III and K. T. Robertson for their assistance and encouragements in this project; especially to my graduate students Mr. Gary Monka and Miss Carolyn Mermon without whose assistance this project could not have been completed. Last of all, he wishes to thank to the Air Force Office of Scientific Research for the opportunity to perform this work.

REFERENCES

1. N.A. Esmen, Y.Y. Hammad "Recent Studies of the Environment in Ceramic Fiber Production. In: Biological Effects of Man-made Mineral Fibers, Vol. 1, pp 222-231. WHO/IARC Publication Lyon France (1984).
2. Int. Agency Res. Cancer "Silica and Some Silicates. IARC Monograph, Vol. 42 Lyon France (1987).
3. K.H. Kilburn, D. Powers & R.H. Warshaw, "Pulmonary Effects of Exposure to Fine Fibreglass: Irregular Opacities and Small Airways Obstruction," Br. J. Ind. Med. **49**:714-720 (1992).
4. A. Dufresne, P.l Loosereewanich, M. Harrigan, P. Sebastien, G. Perrault & R. Begin "Pulmonary Dust Retention in a Silicon Carbide Worker," Am. Ind. Hyg. Assoc. J. **54**:327-330 (1993).
5. J.R. Kramer "Fibrous and Asbestiform Minerals" in Proc. Workshop on Asbestos: Definitions and Measurement Methods" ed. by C.C. Gravat, P.D. LaFleur & K.F.J. Heinrich, Nat. Bureau of Stand. U.S. Govt Printing Office, Washington, D.C. (1978) p.19-33.
6. R.T. Cheng, H.J. McDermott, G.M. Gia, T.L.G. Cover & M.M. Dude "Exposures to Refractory Ceramic Fibers in Refineries and Chemical Plants" Appl. Occup. Environ. Hyg. **7**, 361-7 (1992).
7. T.R. Jacop, J.G. Hadley, J.R. Bender & W. Eastes "Airborne Glass Fiber Concentrations During Manufacturing Operations Involving Glass Wool Insulation," Am. Ind. Hyg. Assoc. J. **54**:32-326 (1993).

8. R. Diskin "Final Report Summer Fellowship," Inorganic Fiber Analysis by SEM-EDXA. US Air Force Office of Scientific Research (1992).

9. L.R. Sherman "Final Report Summer Fellowship," US Air Force Office of Scientific Research (1992).

10. L.R. Sherman, K.T. Roberson & T.C. Thomas "Qualitative Analysis of Asbestos Fibers in Air, Water and Bulk Samples by SEM-EDXA" J.PA Acad. of Sci. **63**, 28-33 (1989).

11. K.T. Roberson, T.C. Thomas & L.R. Sherman, "Comparison of Asbestos Air Samples by SEM-EDXA and TEM-EDXA" Ann. Occup. Hyg. **36**, 265-9 (1992).

12. L.R. Sherman "Final Report Summer Fellowship," US Air Force Office of Scientific Research (1988).

13. G. Monka "Final Report Summer Fellowship," Using the Scanning Electron Microscopy for Fiber Analysis, US Air Force Office of Scientific Research (1993).

APPENDIX A

AIR FORCE
OCCUPATIONAL AND ENVIRONMENTAL HEALTH DIRECTORATE
AL/OEA 2402 E. Drive
Brooks Air Force Base, Texas 78235-5114
Report of Fiber Analysis as Determined by
Scanning Electron Microscopy

ANALYSIS DATE: 08-17-1993

The results of this report are based on the composition and mass ratios of the elements characterized as asbestos, ceramic fibers, fiberglass and calcium minerals. These mass ratios allow this laboratory to make an estimation of the type of fibers found on filters with fiber counts greater than 30 fibers/100 fields The results of this report are based upon the best analytical techniques and fiber data libraries now available in AL/OEA.

The number of fibers/100 field: 50
Calculated number of fibers/cc: .09
The number of fibers analyzed on
the electron microscope: 2

OEAO SAMPLE NO: 930678 BASE SAMPLE NO: SX0678

FIBER NO.	IDENTIFICATION	COMMENT
1	CHRYSOTILE	Serpentine asbestos This fiber is an asbestos
2	CHRYSOTILE	Serpentine asbestos This fiber is an asbestos

Analyzed by: Kenneth T. Roberson GS-12
Chemist, Asbestos Section

APPENDIX B

AUTOEXE FILE ON 286 COMPUTER

Main menu:

A: DOS 3.3	H: CALCIUM FIBERS
B: File Manager	I: CALCIUM TEXT
C: SPC from WP	J: HIGH SILICA
D: A: 5 1/4 720KB	K: HIGH SILICA TEXT
E: A: 3.5 720KB	L: MISCELLANEOUS
F: ASBESTOS ANALYSIS	M: BASE REPORT
G: ASBESTOS FIBER TEXT	N: Shut Down, Park Head

```
Description: Asbestos Analysis
Commands:    CD FIBER
             ASBT1
             CD..
             c:\AUTO
Description: Calcium fibers
Commands:    CD FIBER
             CALCIUM
             CD..
             c:\AUTO
Description: Calcium Text
Commands:    CD FIBER
             Copy Calcium.TXT LPT1
             CD..
             c:\AUTO
Description: High-Silica
Commands:    CD FIBER
             Copy HISIL
             c:\AUTO
Description: HIGH SILICA TEXT
Commands:    CD FIBER
             Copy HISIL.TXT LPT1
             CD..
             c:\AUTO
Description: Miscellaneous
Commands:    CD FIBER
             Copy MISCELer.TXT LPT1
             CD..
             c:\AUTO
Description: BASE REPORT
Commands:    CD BASIC
             BASICA
             CD..
             c:\AUTO
```


END

6-01-94