The UMAP Journal

Vol. 17, No. 3

Subscription Rates

Membership Plus for Individual Subscribers

Individuals subscribe to *The UMAP Journal* through COMAP's Membership Plus. This subscription includes print copies of quarterly issues of *The UMAP Journal*, our annual collection *UMAP Modules: Tools for Teaching*, our organizational newsletter *Consortium*, on-line membership that allows members to search our on-line catalog, download COMAP print materials, and reproduce for use in their classes, and a 10% discount on all COMAP materials.

(Domestic)	**#2020**	**$69**
(Outside U.S.)	**#2021**	**$79**

Institutional Plus Membership Subscribers

Institutions can subscribe to the *Journal* through either Institutional Pus Membership, Regular Institutional Membership, or a Library Subscription. Institutional Plus Members receive two print copies of each of the quarterly issues of *The UMAP Journal*, our annual collection *UMAP Modules: Tools for Teaching*, our organizational newsletter *Consortium*, on-line membership that allows members to search our on-line catalog, download COMAP print materials, and reproduce for use in **any class taught in the institution, and** a 10% discount on all COMAP materials.

(Domestic)	**#2070**	**$395**
(Outside U.S.)	**#2071**	**$405**

Institutional Membership Subscribers

Regular Institutional members receive only print copies of *The UMAP Journal*, our annual collection *UMAP Modules: Tools for Teaching*, our organizational newsletter *Consortium*, **and** a 10% discount on all COMAP materials.

(Domestic)	**#2040**	**$165**
(Outside U.S.)	**#2041**	**$175**

Library Subscriptions

The Library Subscription includes quarterly issues of *The UMAP Journal* and our annual collection *UMAP Modules: Tools for Teaching* and our organizational newsletter *Consortium*.

(Domestic)	**#2030**	**$140**
(Outside U.S.)	**#2031**	**$160**

To order, send a check or money order to COMAP, or call toll-free
1-800-77-COMAP (1-800-772-6627).

The UMAP Journal is published quarterly by the Consortium for Mathematics and Its Applications (COMAP), Inc., Suite 210, 57 Bedford Street, Lexington, MA, 02420, in cooperation with the American Mathematical Association of Two-Year Colleges (AMATYC), the Mathematical Association of America (MAA), the National Council of Teachers of Mathematics (NCTM), the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and The Institute for Operations Research and the Management Sciences (INFORMS). The Journal acquaints readers with a wide variety of professional applications of the mathematical sciences and provides a forum for the discussion of new directions in mathematical education (ISSN 0197-3622).

Second-class postage paid at Boston, MA
and at additional mailing offices.
Send address changes to:
The UMAP Journal
COMAP, Inc.
57 Bedford Street, Suite 210, Lexington, MA 02420

COMAP

关注数学模型

获取更多资讯

Vol. 17 No. 3 1996

Table of Contents

Publisher's Editorial

New Directions

Solomon A. Garfunkel
Executive Director
COMAP, Inc.
57 Bedford St., Suite 210
Lexington, MA 02173
s.garfunkel@mail.comap.com

Sometimes it seems that things at COMAP never stay the same for very long. The second derivative is always positive. No sooner does one project come to fruition than another seems to heat up and still more project ideas emerge. Case in point: *Principles and Practice of Mathematics* is now out and in use. The text, COMAP's new first-year course for undergraduate mathematics majors, has been published by Springer-Verlag. Desk copies are available from the publisher, and we will have a major event at the joint MAA–AMS meetings in January to feature the book. Publication of *Principles and Practice* caps five years of effort by a distinguished author team and advisory board, not to mention the tireless efforts of the project editor, Walter Meyer.

Moreover, the fourth edition of *For All Practical Purposes* has just been published by W.H. Freeman and Company. The text has a new four-color design and, as with each new edition, a substantial amount of original material not present in earlier editions.

Work continues apace on the ARISE Project, our grades 9–11 comprehensive secondary-school mathematics curriculum. All of the project materials are currently undergoing revision for publication in the fall of 1997. In addition, we have recently received funding from the National Science Foundation (NSF) through the STREAM Project to prepare staff-development materials (print, video, and interactive) to support all high-school reform efforts.

Now, about that second derivative. We have just begun a major new undergraduate initiative: Project Intermath, also funded by NSF. This project is part of a new national effort to *institutionalize* reform. Our vision is to help establish the interdisciplinary cooperation necessary for designing integrated college-level experiences in mathematics, science, and technology. We and our partners at the U.S. Military Academy at West Point intend to foster continuous coordination among departments presenting mathematics-based curricula. The strategies for Project Intermath include the development and dissemination of Interdisciplinary Lively Applications (ILAP) modules as well as the testing of integrated curriculum models. The director of this project and new addition to the COMAP staff is Brig. Gen. (Ret.) Frank Giordano. Frank will continue to direct the Mathematical Contest in Modeling; and we are grateful

to have his energy, wisdom, and expertise.

On a different note, by the time you read this editorial, COMAP'S web site, `www.comap.com`, will be up and running. Initially, as with most organizations, the Web site will be informational——describing our organization, its projects, and its products. But our plans are more grandiose. We intend to put most, if not all, of COMAP's modular materials on the Web. Our intent is to make it possible for faculty to preview our supplemental materials in order to decide better which modules will fit their course structure. As with many other companies that act as publishers, we are still debating the economics of this form of distribution. Stay tuned.

As exciting as new technologies and new modes of delivery can be, nothing is more exciting than new programmatic ideas. COMAP, with the help and guidance of our president, Uri Treisman, is planning a major new initiative in the area of service. Tentatively titled "Volunteering Our Expertise," this program is designed to encourage and reward college mathematics departments for new projects that serve their community. Importantly, these will be projects in which faculty and students make use of their mathematical expertise to aid local schools, hospitals, churches, and community organizations. COMAP plans to publish reports on these programs and to give annual awards. It is our sincere hope that through these efforts we will foster increased service activities by undergraduate mathematics departments and, not coincidentally, demonstrate to our neighbors the centrality of our discipline.

This annual guest editorial on COMAP activities, summing up where we are and where we hope to go, is a pleasant task. I would like to thank our editor, Paul Campbell, for this opportunity and for all of his hard work in making *The UMAP Journal* the fine publication it has become.

About the Author

Sol Garfunkel received his Ph.D. in mathematical logic from the University of Wisconsin in 1967. He was at Cornell University and at the University of Connecticut at Storrs for eleven years and has dedicated the last 20 years to research and development efforts in mathematics education. He has been the Executive Director of COMAP since its inception in 1980.

He has directed a wide variety of projects, including UMAP (Undergraduate Mathematics and Its Applications Project), which led to the founding of this *Journal*, and HiMAP (High School Mathematics and Its Applications Project), both funded by the NSF. For Annenberg/CPB, he directed three telecourse projects: *For All Practical Purposes* (in which he appeared as the on-camera host), *Against All Odds: Inside Statistics*, and *In Simplest Terms: College Algebra*. He is currently co-director of the Applications Reform in Secondary Education (ARISE) project, a comprehensive curriculum development project for secondary-school mathematics.

Modeling Forum

Results of the 1996 Mathematical Contest in Modeling

Frank Giordano, MCM Director
COMAP, Inc.
57 Bedford St., Suite 210
Lexington, MA 02173
f.giordano@mail.comap.com
FRGiordano@aol.com

Introduction

A total of 393 teams of undergraduates (a 23% increase from 1995!), from 225 schools, spent the second weekend in February working on applied mathematics problems. They were part of the twelfth Mathematical Contest in Modeling (MCM). On Friday morning, the MCM faculty advisor opened a packet and presented each team of three students with a choice of one of two problems. After a weekend of hard work, typed solution papers were mailed to COMAP on Monday. Seven of the top papers appear in this issue of *The UMAP Journal*.

Results and winning papers from the first eleven contests were published in special issues of *Mathematical Modeling* (1985–1987) and *The UMAP Journal* (1985–1995). The 1994 volume of *Tools for Teaching*, commemorating the tenth anniversary of the contest, contains all of the 20 problems used in the first ten years of the contest and a winning paper for each. Limited quantities of that volume and of the special MCM issues of the *Journal* for the last few years are available from COMAP.

Problem A: The Submarine Detection Problem

The world's oceans contain an ambient noise field. Seismic disturbances, surface shipping, and marine mammals are sources that, in different frequency ranges, contribute to this field. We wish to consider how this ambient noise might be used to detect large moving objects, e.g., submarines located below the ocean surface. Assuming that a submarine makes no intrinsic noise, develop a method for detecting the presence of a moving submarine, its speed, its size, and

its direction of travel, using only information obtained by measuring changes to the ambient noise field. Begin with noise at one fixed frequency and amplitude.

Problem B: The Contest Judging Problem

When determining the winner of a competition like the Mathematical Contest in Modeling, there is generally a large number of papers to judge. Let's say that there are $P = 100$ papers. A group of J judges is collected to accomplish the judging. Funding for the contest constrains both the number of judges that can be obtained and the amount of time that they can judge. For example, if $P = 100$, then $J = 8$ is typical.

Ideally, each judge would read all papers and rank-order them, but there are too many papers for this. Instead, there are a number of screening rounds in which each judge reads some number of papers and gives them scores. Then some selection scheme is used to reduce the number of papers under consideration: If the papers are rank-ordered, then the bottom 30% that each judge rank-orders could be rejected. Alternatively, if the judges do not rank-order the papers, but instead give them numerical scores (say, from 1 to 100), then all papers falling below some cutoff level could be rejected.

The new pool of papers is then passed back to the judges, and the process is repeated. A concern is that the total number of papers that each judge reads must be substantially less than P. The process is stopped when there are only W papers left. These are the winners. Typically, for $P = 100$, we have $W = 3$.

Your task is to determine a selection scheme, using a combination of rank-ordering, numerical scoring, and other methods, by which the final W papers will include only papers from among the "best" $2W$ papers. (By "best" we assume that there is an absolute rank-ordering to which all judges would agree.) For example, the top three papers found by your method will consist entirely of papers from among the "best" six papers. Among all such methods, the one that requires each judge to read the least number of papers is desired.

Note the possibility of systematic bias in a numerical scoring scheme. For example, for a specific collection of papers, one judge could average 70 points, while another could average 80 points. How would you scale your scheme to accommodate for changes in the contest parameters (P, J, and W)?

The Results

The solution papers were coded at COMAP headquarters so that names and affiliations of the authors would be unknown to the judges. Each paper was then read preliminarily by two "triage" judges at Carroll College, Montana. At the triage stage, the summary and overall organization are the basis for judging a paper. If the judges' scores diverged for a paper, the judges conferred; if they still did not agree on a score, a third judge evaluated the paper.

Final judging took place at Harvey Mudd College, Claremont, California. The judges classified the papers as follows:

	Outstanding	Meritorious	Honorable Mention	Successful Participation	Total
Submarine Detection	4	16	37	69	126
Contest Judging	5	38	77	147	267
	9	54	114	216	393

The nine papers that the judges designated as Outstanding appear in this special issue of *The UMAP Journal*, together with commentaries by judges and practitioners. We list those teams and the Meritorious teams (and advisors) below; the list of all participating schools, advisors, and results is in the **Appendix**.

Outstanding Teams

Institution and Advisor **Team Members**

Submarine Detection Papers

"Gone Fishin' "
Pomona College
Claremont, CA
Ami Radunskaya

Douglas Martin
Robert A. Moody
Woon (Larry) Wong

"How to Locate a Submarine by Detecting Changes in the Ambient Noise"
University of North Carolina
Chapel Hill, NC
Ancel C. Mewborn

Carl Leitner
Akira Negi
Katherine Scott

"Detection of a Silent Submarine from Ambient Noise Field Fluctuations"
Wake Forest University
Winston-Salem, NC
Stephen B. Robinson

Andrew R. Frey
Joseph R. Gagnon
J. Hunter Tart

"Imaging Underwater Objects with Ambient Noise"
Worcester Polytechnic Institute
Worcester, MA
Arthur C. Heinricher

Aron C. Atkins
Henry A. Fink
Jeffrey D. Spaleta

Contest Judging Papers

"The Paper Selection Scheme Simulation Analysis"	
Fudan University	Zheng Yuan Zhu
Shanghai, China	Jian Liu
Yongji Tan	Haonan Tan
"Modeling Better Modeling Judges"	
Gettysburg College	Brian E. Ellis
Gettysburg, PA	Chad Hall
James P. Fink	Charles A. Ross
"Judging a Mathematics Contest"	Daniel A. Calderón Brennan
St. Bonaventure University	Philip J. Darcy
St. Bonaventure, NY	David T. Tascione
Albert G. White	
"Select the Winners Fast"	
University of Science and Technology of China	Haitao Wang
Hefei, Anhui, China	Chunfeng Huan
Qingjuan Yu	Hongling Rao
"The Inconsistent Judge"	
Washington University	Dan Scholz
St. Louis, MO	Jade Vinson
Hiro Mukai	Derek Oliver Zaba

Meritorious Teams

Submarine Detection Papers (16 teams)
Beijing Institute of Technology, Beijing, China (Liang Sun)
Chongqing University, Chongqing, Sichuan, China (Fu Li)
Duke University, Durham, NC (Richard A. Scoville)
Eastern Mennonite University, Harrisonburg, VA (John L. Horst)
Nankai University, Tianjin, China (Wu Qun Huang)
Ohio State University, Columbus, OH (Dijen Ray-Chaudhuri)
Rhodes College, Memphis, TN (David A. Feil)
Southeast University, Nanjing, China (Huang Jun)
Southeast University, Nanjing, China (Zhizhong Sun)
Trinity University, San Antonio, TX (Diane G. Saphire)
Tsinghua University, Beijing, China (Celi Gao)
University of North Florida, Jacksonville, FL (Peter A. Braza)
University of Northern Iowa, Cedar Falls, IA (Gregory M. Dotseth)
University of Science and Technology of China, Hefei, Anhui, China (Jixin Cheng)
Xiangtan University, Xiangtan, Hunan, China (Zhou Yong)
Zhongshan University, Guangzhou, China (Ren Shu Chen)

Contest Judging Papers (38 teams)
Abilene Christian University, Abilene, TX (Thomas D. Hendricks)
Bellarmine College, Louisville, KY (John A. Oppelt)
California Polytechnic State University, San Luis Obispo, CA (Thomas O'Neil)
Colorado College, Colorado Springs, CO (Deborah P. Levinson)
East China University of Science and Technology, Shanghai, China (Sanbao Xu)
Eastern Oregon State College, LaGrande, OR (Mark R. Parker)
Harvard University, Cambridge, MA (Howard Georgi)
Harvey Mudd College, Claremont, CA (David L. Bosley)
Information & Engineering Institute, Zhengzhou, Henan, China (Hongwei Duan)
Kenyon College, Gambier, OH (Dana N. Mackenzie)
Lewis & Clark College, Portland, OR (Harvey Schmidt, Jr.)
Luther College, Decorah, IA (Reginald D. Laursen)
Macalester College, St. Paul, MN (Karla V. Ballman)
Messiah College, Grantham, PA (Douglas C. Phillippy)
Mt. St. Mary's College, Emmitsburg, MD (Fred J. Portier)
Mt. St. Mary's College, Emmitsburg, MD (Theresa A. Francis)
North Carolina School of Science and Mathematics, Durham, NC (Dot Doyle)
National University of Defence Technology, Chang Sha, Hunan, China (MengDa Wu)
New Mexico State University, Las Cruces, NM (Caroline Sweezy)
Northern Arizona University, Flagstaff, AZ (Terence R. Blows)
Rose-Hulman Institute of Technology, Terre Haute, IN (Aaron D. Klebanoff)
South China University of Technology, Guangzhou, China (Lejun Xie)
Southern Connecticut State University, New Haven, CT (Ross B. Gingrich)
University College Cork, Cork Ireland (J.B. Twomey)
University College Galway, Galway Ireland (Patrick M. O'Leary)
University of Alaska Fairbanks, Fairbanks, AK (John P. Lambert)
University of Dayton, Dayton, OH (Thomas E. Gantner)
University of Latvia, Riga, Latvia (Andris B. Cibulis)
University of Massachusetts–Amherst, Amherst, MA (Edward A. Connors)
University of Missouri–Rolla, Rolla, MO (Michael G. Hilgers)
University of Toronto, Toronto, Ontario, Canada (James G.C. Templeton)
University of Utah, Salt Lake City, UT (Don H. Tucker)
University of Wisconsin–Madison, Madison, WI (Anatole Beck)
University of Wisconsin–Platteville, Platteville, WI (John A. Krogman)
Western Washington University, Bellingham, WA (Tjalling J. Ypma)
Xidian University, Xian, Shaanxi, China (Mao Yong Cai)
Youngstown State University, Youngstown, OH (J. Douglas Faires)
Zhejiang University, Hangzhou, China (Daoyuan Fang)

Awards and Contributions

Each participating MCM advisor and team member received a certificate signed by the Contest Director and the appropriate Head Judge.

INFORMS, the Institute for Operations Research and the Management Sci-

ences, awarded to each member of two Outstanding teams a cash award and a three-year membership. The teams were from Gettysburg College (Contest Judging Problem) and Washington University (Contest Judging Problem). Moreover, INFORMS gave free one-year memberships to all members of Meritorious and Honorable Mention teams.

The Society for Industrial and Applied Mathematics (SIAM) designated one Outstanding team from each problem as a SIAM Winner. Each team member received a cash prize. The teams were from Pomona College (Submarine Detection Problem) and from St. Bonaventure University (Contest Judging Problem). They gave presentations at a special session at the July SIAM Annual Meeting in Kansas City, MO.

The Mathematical Association of America designated one Outstanding team as an MAA Winner. The team was from Pomona College (Submarine Detection Problem).

Judging

Director
Frank R. Giordano, Dept. of Mathematics, Carroll College, Helena, MT

Associate Directors
Chris Arney, Dept. of Mathematical Sciences, U.S. Military Academy, West Point, NY
Robert L. Borrelli, Mathematics Dept., Harvey Mudd College, Claremont, CA

Submarine Detection Problem
Head Judge
Marvin S. Keener, Mathematics Dept., Oklahoma State University, Stillwater, OK
Associate Judges
Courtney Coleman, Mathematics Dept., Harvey Mudd College, Claremont, CA
Patrick Driscoll, Mathematics Dept., Virginia Polytechnic Institute and State University, Blacksburg, VA
Ben A. Fusaro, Dept. of Mathematical Sciences, Florida State University, Tallahassee, FL
Mario Juncosa, RAND Corporation, Santa Monica, CA
Daphne Liu, Dept. of Mathematics and Computer Science, California State University Los Angeles, Los Angeles, CA
Jack Robertson, Mathematics Dept., Georgia College, Milledgeville, GA
Lee Seitelman, Glastonbury, CT
John L. Scharf, Carroll College, Helena, MT
Theodore H. Sweetser III, Jet Propulsion Lab, Pasadena, CA
Daniel Zwillinger, Zwillinger & Associates, Arlington, MA

Contest Judging Problem
Head Judge
Maynard Thompson, Mathematics Dept., University of Indiana, Bloomington, IN
Associate Judges
Karen Bolinger, Mathematics Dept., Arkansas State University, State University, AR
James Case, Baltimore, Maryland
Alessandra Chiareli, Computational Science Center, 3M, St. Paul, MN
William Fox, Dept. of Mathematical Sciences, U.S. Military Academy, West Point, NY
Ted Forsman, Onward, Inc.
Jerry Griggs, University of South Carolina, Columbia, SC
John Kobza, Virginia Polytechnic Institute and State University, Blacksburg, VA
Veena Mendiratta, Lucent Technologies, Naperville, IL
Don Miller, Dept. of Mathematics, St. Mary's College, Notre Dame, IN
Keith Miller, National Security Agency, Fort Meade, MD
Peter Olsen, National Security Agency, Fort Meade, MD
Theresa Sandifer, Mathematics Dept., Southern Connecticut State University, New Haven, CT
Robert M. Tardiff, Dept. of Mathematical Sciences,
Michael Tortorella, Lucent Technologies, NJ
Marie Vanisko, Carroll College, Helena, MT

Triage Session (all judges from Carroll College, Helena, MT)
Director
Frank Giordano
Head Judge, Submarine Detection Problem
John L. Scharf
Head Judge, Contest Judging Problem
Marie Vanisko
Associate Judges
Peter Biskis
Philip LaRue
Terry Mullen
Jack Oberweiser
Philip Rose
Anthony M. Szpilka

Sources of the Problems

Both the Submarine Detection Problem and the Contest Judging Problem were contributed by Daniel Zwillinger, Zwillinger & Associates, Arlington, MA.

Acknowledgments

The MCM was funded this year by the National Security Agency, whose support we deeply appreciate. We thank Dr. Gene Berg of NSA for his coordinating efforts. The MCM is also indebted to INFORMS, SIAM, and the MAA, which provided judges and prizes.

I thank the MCM judges and MCM Board members for their valuable and unflagging efforts. Harvey Mudd College, its Mathematics Dept. staff, and Prof. Borrelli were gracious hosts to the judges.

Cautions

To the reader of research journals:

Usually a published paper has been presented to an audience, shown to colleagues, rewritten, checked by referees, revised, and edited by a journal editor. Each of the student papers here is the result of undergraduates working on a problem over a weekend; allowing substantial revision by the authors could give a false impression of accomplishment. So these papers are essentially au naturel. Light editing has taken place: minor errors have been corrected, wording has been altered for clarity or economy, and style has been adjusted to that of *The UMAP Journal*. Please peruse these student efforts in that context.

To the potential MCM Advisor:

It might be overpowering to encounter such output from a weekend of work by a small team of undergraduates, but these solution papers are highly atypical. A team that prepares and participates will have an enriching learning experience, independent of what any other team does.

Appendix: Successful Participants

KEY:

P = Successful Participation
H = Honorable Mention
M = Meritorious
O = Outstanding (published in this special issue)
A = Submarine Detection Problem
B = Contest Judging Problem

INSTITUTION	CITY	ADVISOR	A	B
ALABAMA				
Birmingham-Southern Coll.	Birmingham	Raju Sriram		P
University of Alabama	Huntsville	Claudio H. Morales		P
ALASKA				
Univ. of Alaska Fairbanks	Fairbanks	John P. Lambert		M,H
ARIZONA				
Northern Arizona Univ.	Flagstaff	Terence R. Blows		M,P
ARKANSAS				
Hendrix College	Conway	Ze'ev Barel		H,P
Williams Baptist College	Walnut Ridge	Lana S. Rhoads		P
		Michael Milligan		P
CALIFORNIA				
Calif. Inst. of Technology	Pasadena	Richard M. Wilson		P
Calif. Poly. State Univ.	San Luis Obispo	Thomas O'Neil		M,P
Calif. State University	Bakersfield	Maureen E. Rush	H	P
	Northridge	Gholam-Ali Zakeri		P
Harvey Mudd College	Claremont	David L. Bosley	P	M
Humboldt State Univ.	Arcata	Kathleen M. Crowe		P
Pomona College	Claremont	Ami Radunskaya	O	
Sonoma State University	Rohnert Park	Clement E. Falbo		P
COLORADO				
Colorado College	Colorado Springs	Deborah P. Levinson		M
Metro. State College	Denver	Thomas E. Kelley	H	
Trinidad State Jr. College	Trinidad	A. Philbin		H,P
U.S. Air Force Academy	USAF Academy	Scott G. Frickenstein	H	
		Jonathan D. Robinson		H

INSTITUTION	CITY	ADVISOR	A	B
U. of Northern Colorado	Greeley	Donald D. Elliott		H
U. of Southern Colorado	Pueblo	Bruce N. Lundberg		P
CONNECTICUT				
Sacred Heart University	Fairfield	Antonio A. Magliaro	P	
Southern Connecticut State Univ.	New Haven	Ross B. Gingrich		M
Western Connecticut State Univ.	Danbury	Judith A. Grandahl		H
		Edward Sandifer		H
DISTRICT OF COLUMBIA				
Georgetown University	Washington	Andrew Vogt	P	H
George Washington University	Washington	Daniel H. Ullman	P	
Trinity College	Washington	Suzanne E. Sands		H
FLORIDA				
Florida A&M University	Tallahassee	Bruno Guerrieri		P
Florida Inst. of Technology	Melbourne	Laurene V. Fausett	P	
Florida State University	Tallahassee	Hong Wen	H	
Jacksonville University	Jacksonville	Robert A. Hollister	P	
Stetson University	Deland	Lisa O. Coulter		H
University of North Florida	Jacksonville	Peter A. Braza	M	
Univ. of South Forida–Fort Myers	Fort Myers	Charles E. Lindsey	P	P
GEORGIA				
Georgia College	Milledgeville	Craig Turner		P
ILLINOIS				
Greenville College	Greenville	Galen R. Peters		P
Illinois Wesleyan University	Bloomington	Zahia Drici	P	
Morton College	Cicero	Elaine B. Pavelka		P
Northern Illinois University	Dekalb	Hamid Bellout		H
		Linda R. Sons	P	
Olivet Nazarene University	Bourbonnais	Dale K. Hathaway	H	P
Wheaton College	Wheaton	Paul Isihara	B	H
INDIANA				
Indiana University	Bloomington	James F. Davis	H	H
Rose-Hulman Inst. of Tech.	Terre Haute	Aaron D. Klebanoff		M
Saint Mary's College	Notre Dame	Peter D. Smith	P	H

INSTITUTION	CITY	ADVISOR	A	B
IOWA				
Clarke College	Dubuque	Carol A. Spiegel		P,P
Drake University	Des Moines	Alexander F. Kleiner	P	
Graceland College	Lamoni	Steve K. Murdock		H
Grinnell College	Grinnell	Thomas L. Moore		H,H
Luther College	Decorah	Reginald D. Laursen		M
Univ. of Northern Iowa	Cedar Falls	Gregory M. Dotseth	M	
		Timothy L. Hardy		P
KENTUCKY				
Asbury College	Wilmore	Kenneth P. Rietz	P,P	
Bellarmine College	Louisville	John A. Oppelt		M,P
Univ. of Louisville	Louisville	Adel S. Elmaghraby		P
Western Kentucky Univ.	Bowling Green	Douglas D. Mooney	P	
LOUISIANA				
Louisiana State Univ.	Shreveport	Robert J. Fraga	P	
McNeese State Univ.	Lake Charles	Sid L. Bradley		P
Northwestern St. Univ.	Natchitoches	Lisa R. Galminas		P,P
MAINE				
Colby College	Waterville	Amy H. Boyd		P,P
MARYLAND				
Frostburg State University	Frostburg	Kurtis H. Lemmert	H	P
Goucher College	Baltimore	Robert E. Lewand		P
Hood College	Frederick	John Boon		P
		Betty Mayfield	P	
Loyola College	Baltimore	George B. Mackiw		P
		William D. Reddy	H	
Mt. St. Mary's College	Emmitsburg	Theresa A. Francis		M
		Fred J. Portier		M
Salisbury State University	Salisbury	Steven M. Hetzler	P	
		Kathleen M. Shannon		P
MASSACHUSETTS				
Fitchburg State College	Fitchburg	Richard C. Bisk	P	
Harvard University	Cambridge	Howard Georgi		M
Smith College	Northampton	Ruth Haas		H,P
Univ. of Massachusetts	Amherst	Edward A. Connors		M,H
Worcester Poly. Inst.	Worcester	Arthur C. Heinricher	O	
		Bogdan Vernescu	P	

INSTITUTION	CITY	ADVISOR	A	B
MICHIGAN				
Calvin College	Grand Rapids	Thomas L. Jager	H	
Eastern Michigan Univ.	Ypsilanti	Christopher E. Hee	H	P
Lawrence Tech. Univ.	Southfield	Ruth G. Favro	P	
		Howard Whitston	H	
Siena Heights College	Adrian	Toni Carroll		P
University of Michigan	Dearborn	Jennifer Zhao		H
MINNESOTA				
Bethany Lutheran College	Mankato	Julie M. Kjeer	H	
Macalester College	St. Paul	Karla V. Ballman		M
		Daniel A. Schwalbe		H
University of Minnesota	Duluth	Zhuangyi Liu	P	
		Bruce B. Peckham		H
MISSISSIPPI				
Belhaven College	Jackson	Robert A. Jones		P
		Janie Smith		P
Jackson State University	Jackson	David C. Bramlett		P
MISSOURI				
College of the Ozarks	Point Lookout	Albert T. Dixon	P	P
Missouri Southern St. Coll.	Joplin	Patrick Cassens		H,H
Northeast Missouri St. U.	Kirksville	Steven J. Smith	H,P	
Northwest Missouri St. U.	Maryville	Russell N. Euler		P
Southeast Missouri St. U.	Cape Girardeau	Robert W. Sheets		H,H
University of Missouri	Rolla	Michael G. Hilgers		M
Washington University	St. Louis	Hiro Mukai	P	O
MONTANTA				
Carroll College	Helena	Terence J. Mullen		H
		Anthony M. Szpilka		P
NEBRASKA				
Hastings College	Hastings	David B. Cooke		H
Nebraska Wesleyan Univ.	Lincoln	P. Gavin LaRose		H
NEVADA				
Sierra Nevada College	Incline Village	Sue Welsch	P	
University of Nevada	Reno	Mark M. Meerschaert		P

INSTITUTION	CITY	ADVISOR	A	B
NEW JERSEY				
Camden County College	Blackwood	Allison Sutton		P
New Jersey Inst. of Tech.	Newark	Bruce G. Bukiet		P
Rutgers University	Newark	Lee Mosher		P
NEW MEXICO				
New Mexico State Univ.	Las Cruces	Marcus S. Cohen		H
		Caroline Sweezy		M
New Mexico Inst. of Mining and Tech.	Socorro	Brian T. Borchers		P
NEW YORK				
Canisius College	Buffalo	L. Christine Kinsey		P
City College of CUNY	New York	Izidor Gertner	P	
		George Wolberg		H
Colgate University	Hamilton	Thomas W. Tucker		H
Hofstra University	Hempstead	Raymond N. Greenwell	P,P	
Ithaca College	Ithaca	John C. Maceli		H
		Osman Yurekli		P
Manhattanville College	Purchase	Edward Schwartz		H
Nazareth College	Rochester	Ronald W. Jorgensen		H
Queens College / CUNY	Flushing	Ari Gross		H
Siena College	Loudonville	Thomas H. Rousseau		H,P
St. Bonaventure Univ.	St. Bonaventure	Francis C. Leary		P
		Albert G. White		O
U.S. Military Academy	West Point	Douglas L. Bentley, Jr.		H
		Timothy M. Petit		P
Westchester Comm. Coll.	Valhalla	Rowan Lindley		P,P,P
Yeshiva College	New York	Thomas H. Otway		P
NORTH CAROLINA				
Appalachian State Univ.	Boone	Terry G. Anderson		H
		Alan T. Arnholt	P	
		Holly P. Hirst	P	
Duke University	Durham	Richard A. Scoville	M	
N.C. School of Sci. & Math.	Durham	Dot Doyle		M
		John Kolena		H
		Daniel J. Teague		P
North Carolina St. Univ.	Raleigh	John Bishir		H
Salem College	Winston-Salem	Debbie L. Harrell		P
		Craig J. Richardson		P
		Paula G. Young		P

INSTITUTION	CITY	ADVISOR	A	B
Univ. of North Carolina	Chapel Hill	Ancel C. Mewborn	O	
	Wilmington	Russell L. Herman		P
Wake Forest University	Winston-Salem	Stephen B. Robinson	O	
Western Carolina Univ.	Cullowhee	Jeff A. Graham		H
NORTH DAKOTA				
Univ. of North Dakota	Williston	Wanda M. Meyer		P
OHIO				
College of Wooster	Wooster	Matt Brahm		P
Hiram College	Hiram	James R. Case		H
		Michael A. Grajek		P
Kenyon College	Gambier	Dana N. Mackenzie		M,H
Miami University	Oxford	Douglas E. Ward		P
Ohio State University	Columbus	Dijen Ray-Chaudhuri	M	P
University of Dayton	Dayton	Thomas E. Gantner		M
Xavier University	Cincinnati	Richard J. Pulskamp		H,P
Youngstown St. Univ.	Youngstown	J. Douglas Faires	H	M
OKLAHOMA				
Oklahoma State Univ.	Stillwater	John E. Wolfe		H
Southeastern Okla. St. U.	Durant	John M. McArthur		H
OREGON				
Eastern Oregon St. Coll.	LaGrande	Mark R. Parker		M
		Holly S. Zullo	P	
Lewis & Clark College	Portland	Harvey Schmidt, Jr.		M
Southern Oregon St. Coll.	Ashland	Kemble R. Yates		H
PENNSYLVANIA				
Bloomsburg University	Bloomsburg	Scott E. Inch		P
Chatham College	Pittsburgh	Angela A. Fishman		P
Franklin & Marshall Coll.	Lancaster	Tim C. Hesterberg	P	
Gannon University	Erie	Thomas M. McDonald	P	
Gettysburg College	Gettysburg	James P. Fink		O
Lafayette College	Easton	Thomas Hill		H
Messiah College	Grantham	Douglas C. Phillippy		M
Muhlenberg College	Allentown	David A. Nelson		P
Susquehanna University	Selinsgrove	Kenneth A. Brakke		P

INSTITUTION	CITY	ADVISOR	A	B
SOUTH CAROLINA				
Central Carolina Tech. Coll.	Sumter	Karen G. McLaurin	P	P
Coastal Carolina Univ.	Conway	Prashant S. Sansgiry	P	
Columbia College	Columbia	Scott A. Smith		P
SOUTH DAKOTA				
Northern State University	Aberdeen	A. S. Elkhader		H,P
TENNESSEE				
Austin Peay St. Univ.	Clarksville	Mark C. Ginn		H
David Lipscomb Univ.	Nashville	Gary C. Hall		P
		Mark A. Miller	P	
Rhodes College	Memphis	David A. Feil	M	
TEXAS				
Abilene Christian Univ.	Abilene	Thomas D. Hendricks		M
Baylor University	Waco	Frank H. Mathis		H
Rice University	Houston	Douglas W. Moore	H	
Southwestern University	Georgetown	Therese N. Shelton		P
Texas A & M University	College Station	Denise E. Kirschner	P	
Trinity University	San Antonio	Diane G. Saphire	M	P
University of Dallas	Irving	Charles A. Coppin		P
	Irving	Edward P. Wilson	P	
U. of Texas–Pan American	Edinburg	Roger A. Knobel	P	P
U. of Texas–Permian Basin	Odessa	Marcin Paprzycki		P
UTAH				
University of Utah	Salt Lake City	Don H. Tucker		M,H
VERMONT				
Johnson State College	Johnson	Glenn D. Sproul		H
Norwich University	Northfield	Leonard C. Gambler	P	
VIRGINIA				
College of William & Mary	Williamsburg	Larry M. Leemis		H
		Hugo J. Woerdeman	P	
Eastern Mennonite Univ.	Harrisonburg	John L. Horst	M	
James Madison Univ.	Harrisonburg	James S. Sochacki	P	
Roanoke College	Salem	Roland B. Minton	P	
University of Richmond	Richmond	Kathy W. Hoke	P	P

INSTITUTION	CITY	ADVISOR	A	B
WASHINGTON				
Pacific Lutheran Univ.	Tacoma	Rachid Benkhalti		H,H
Univ. of Puget Sound	Tacoma	Robert A. Beezer		H
		Martin Jackson		H
		Andrew F. Rex	H	
Western Washington U.	Bellingham	Tjalling J. Ypma	P	M
WISCONSIN				
Beloit College	Beloit	Philip D. Straffin		H,H
Northcentral Tech. Coll.	Wausau	Frank J. Fernandes	P	
		Robert J. Henning		P,P
Northland College	Ashland	Nicholas C. Bystrom	P	
St. Norbert College	De Pere	John A. Frohliger		P
Univ. of Wisconsin	Green Bay	Nikitas L. Petrakopoulos		P
	Madison	Anatole Beck		M
	Oshkosh	K.L.D. Gunawardena		P
	Platteville	Clement T. Jeske		H
		John A. Krogman		M
		Sherrie Nicol		P
	Stevens Point	Norm D. Curet	P	H
Wisc. Lutheran Coll.	Milwaukee	Marvin C. Papenfuss		P
AUSTRALIA				
U. of South Queensland	Toowoomba, Qld.	Christopher J. Harman		P
CANADA				
University of Calgary	Calgary, Alb.	David R. Westbrook		H,P
Univ. of Saskatchewan	Saskatoon, Sask.	James A. Brooke	H	
University of Toronto	Toronto, Ont.	James G. C. Templeton		M
York University	North York, Ont.	Jianhong Wu		H,H
CHINA				
Auto. Eng. Coll. of Beijing Union U.	Beijing	Wang Xin Feng	P	
		Ren Kai Long		H
Beijing Inst. of Tech.	Beijing	Liang Sun	M	P
Beijing Normal Univ.	Beijing	Laifu Liu		P
		Wenyi Zeng	P	
Beijing Univ. of Aero. & Astro.	Beijing	Weiguo Li	P	
		Ling Ma	H	P
		Wei Guo Li	P	

INSTITUTION	CITY	ADVISOR	A	B
Beijing Univ. of Post & Tel.	Beijing	Shou-shan Luo	P	P
Chongqing University	Chongqing, Sichuan	Fu Li	M	
		Gong Chu		P
		Qiongsen Liu		P
		Shanqiang Ren		P
Dalian Univ. of Technology	Dalian, Liaoning	Hong-Quan Yu	P	
		Li-Zhong Zhao		P
		Ming-Feng He		H,P
E. China Normal Univ.	Shanghai	Zhen-Dong Yuan		P
E. China U. of Sci. & Tech.	Shanghai	Sanbao Xu		M,P
		Xiwen Lu	H	
		Yuanhong Lu	H	
Fudan University	Shanghai	Yuan Cao	P	
		Yongji Tan		O
		Wei Liao You	H	
FuYang Teachers College	FuYang, Anhui	Yang Song	P	P
		Yunfei Yao		P,P
Harbin Engineering Univ.	Harbin, Heilongjiang	Zhenbin Gao		P
		Jihong Shen		P
		Xiaowei Zhang	H	
Harbin Inst. of Tech.	Harbin, Heilongjiang	Peilin Shi		P,P,P,P
Harbin U. of Sci. & Tech.	Harbin, Heilongjiang	Chen Dong Yan		P
Hefei Univ. of Tech.	Hefei, Anhui	Xueqiao Du		P
		Yongwu Zhou	P	
		Youdu Huang	P	
Hohai University	Nanjing	Gen-Hong Ding		P
		Ru-Yun Wang	H	
Huazhong U. of Sci. & Tech.	Wuhan, Hubei	Wuwen Guo	H	
		Huan Qi		P
		Nanzhong He	P	
		Xiaoyang Zhou	H	
Info. & Eng'ng. Institute	Zhengzhou, Henan	Hongwei Duan		M
		Zhonggeng Han		P
Inst. of Elec. & Auto. Eng. of Beijing Union Univ.	Beijing	Zifa Wang	P	
Jilin Inst. of Technology	Changchun, Jilin	Xiaogang Dong	H	
		Yunhui Xu		H
Jilin University	Changchun, Jilin	Zhenghua Lin	P	H
		Xianyui Lu	P	P
Jinan University	Guangzhou, Guangdong	Ye Shi Qi		P
		Yijun Zeng	P	P
Nanjing Normal Univ.	Nanjing, Jiangsu	Huai Ping Zhu	P	P

INSTITUTION	CITY	ADVISOR	A	B
Nanjing U. of Sci. & Tech.	Nanjing	Long Sheng Cheng	H	
		Xiong Ping Qian		P
		Xin Min Wu		P
		Chong Gao Zhao		P
Nankai University	Tianjin	Tian Ping Ye		H
		Qun Huang Wu	M	
		Xia Sheng Wang		H
		Xing Wei Zhou		P
National U. of Defence Tech.	Chang Sha, Hunan	MengDa Wu		M
		Yi Wu	H	
Peking University	Beijing	Chong-shi Wu	H	
		Qiao-jun Gao	H	
		Sheng Huang		P,P
Shandong University	Jinan	Baogang Xu	H	
		Guijuie Qi		P
Shanghai Jiaotong University	Shanghai	Wensong Chu		P
		Zhihe Chen	P	
		Longwan Xiang		P
		Gang Zhou		P
Shanghai Normal Univ.	Shanghai	Jiaxiang Xiang		P,P
South China Univ. of Tech.	Guangzhou, Guangdong	Fengfeng Zhu	H	
		Hongzuo Fu	H	
		Lejun Xie		M
		Zhihua Can		P
Southeast University	Nanjing, Jiangsu	Jun Huang	M	
		Jianming Deng		P
		Zhizhong Sun	M	
		Daoyuan Zhu		P
Tsinghua University	Beijing	Binheng Song	H	P
		Celi Gao	M	
		Jin-Xing Xie		P
U. of Sci. & Tech. Beijing	Beijing	Xiaoming Yang		P
U. of Sci. & Tech. of China	Hefei, Anhui	Jixin Cheng	M	
		Yu Feng		P
		Zhi Zhou		P
		Qingjuan Yu		O
Xian Mining Institute	Xian, Shaanxi	PanZhu Wei		P
Xiangtan University	Xiangtan, Hunan	Yangjin Cheng		H
		Zhou Yong	M	

INSTITUTION	CITY	ADVISOR	A	B
Xidian University	Xian, Shaanxi	Hu Yu Pu		P
		Ma Jian Feng	H	
		Mao Yong Cai		M
		Yu Ping Wang		P
Zhejiang University	Hangzhou	Daoyuan Fang	P	M
		Huixiang Gong	H	P
Zhongshan University	Guangzhou, Guangdong	Ren Shu Chen	M	
		Liu Jun Xu		P
HONG KONG				
Hong Kong Baptist Univ.	Kowloon Tong	Li Zhi Liao		H
		Wai Chee Shiu		P
IRELAND				
Trinity College Dublin	Dublin	Timothy G. Murphy		H
		James C. Sexton	P	
Univ. College Cork	Cork	Patrick Fitzpatrick		H
		J. B. Twomey		M
Univ. College Galway	Galway	Patrick M. O'Leary		M
		Michael P. Tuite	P	
LATVIA				
University of Latvia	Riga	Andris B. Cibulis		M
LITHUANIA				
Vilnius University	Vilnius	Ricardas Kudzma	H	
		Algirdas Zabulionis		P
MEXICO				
Univ. Autón. de Yucatán	Merida, Yucatán	María C. Fuente-Florencia	P	

The editor wishes to thank Lay May Yeap of Beloit College for her help with Chinese names.

Gone Fishin'

Douglas Martin
Robert A. Moody
Woon (Larry) Wong
Pomona College
Claremont, CA 91711

Advisor: Ami Radunskaya

Abstract

We develop a method to measure the ambient field, using directional transducers and programmable electronic delay circuitry, so that we can determine the ambient sound pressure for a given cell in three-dimensional space.

We discuss the expected observations from the interaction of the target object and the ambient field, and the associated limitations of our method.

We conducted a simple experiment that verified that our approach worked in an anechoic environment (outdoors), and we argue that the results should generalize to the underwater environment.

We show how our method produces data sufficient to reconstruct object size, position, and velocity, and we present the algorithms required to accomplish this reconstruction. We discuss the advantages and problems associated with various reference frequencies, and methods to optimize the technique. Finally, we present a model of the anticipated characteristic performance of our method.

Ambient Sounds

(All of the background information in this section comes from Naval Sea Systems Command [1984].)

Ambient noise in the ocean can be classified into background or continuous noise, which is present for extended durations of time, and intermittent sounds or noises generated at random intervals for essentially random durations.

The background is comprised primarily of wind-based noises, noises from commercial shipping, and other similar human-made sources. Some seismic noise is present, particularly at lower frequencies.

Intermittent noises can be further divided into biological and nonbiological. Multi-watt whale sounds range from low-frequency moans to high frequency clicks and have a frequency range of at least 2 Hz–60 kHz.

Intermittent nonbiological noise include the rain, earthquakes, explosions, and volcanoes. Although there is some additional noise generated by the surf, the poor propagation characteristics of the shallow water in which such noise

is generated cause a rapid decay; thus these sounds do not appear to contribute heavily to the ambient field in the deep sea. The contributions of rain, earthquakes, and volcanoes are continuous enough for us to group them under the general heading "background noise," since they can reasonably be expected to have periods vastly larger than our scanning period. Explosions are, from our perspective, an overwhelming problem.

Studies have confirmed both the directionality and spatial coherence characteristics of ambient sound, both background and intermittent [Urick 1983]. Shipping noise, which is primarily in the 100 Hz range, tends to travel horizontally and at low angles. Wind noise, dominant in the higher frequencies (1,000 Hz and above), travels over relatively direct paths and tends to arrive at angles between 45° and 80° relative to the surface.

Sonar

Active sonar involves generating sound and determining the bearing and distance of the target object by measuring the time that it takes the echo to return to elements of the sensor array. In passive sonar, the target is the source of the detected sound. In both cases, the detected sound is effectively radiated from a point source, the target. The parameters that relate to the efficacy of passive sonar are [Urick 1983]:

SL, source level;

DI, receiving directivity index;

DT, detection threshold;

TL, transmission loss attributable to the medium; and

NL, ambient noise level.

These are related through the passive sonar equation:

$$\text{SL} - \text{TL} \quad = \quad \text{NL} - \text{DI} + \text{DT}.$$

Since we are detecting the noise directly and attempting to measure the absence of noise along a particular directional axis, we reverse the SL and NL terms. Thus, we attempt to locate the loud fields and thereby detect the lower-power fields via elimination. Since we predict extremely small deviations in the source and noise levels, we depend quite heavily on the directivity index as the parameter that enables us to construct an effective system.

Acoustic Impact of Submarine

At low speeds or at rest, the submarine will look like a hole in the noise field. Since the ambient field is homogeneous (that is, same intensity from any direction), the submarine appears as an absorber. It reflects sound waves; sound coming toward the observer from the other side of the submarine is deflected away from the observer, while sound coming from the same side as the observer is reflected towards the observer. Both these sound waves have the same intensity, so the net effect is that the submarine will look the same as the ambient field but reduced by a reflection coefficient. Similar considerations apply to sound coming at off-axis angles. See **Figure 1**, in which A, B, C, and D are incident ambient sound waves of equal intensity, with aA, aB, aC, and aD being their reflections from the submarine, for reflection coefficient a. Thus, the submarine masks A and B from the detector but reflects aC and aD toward the detector. Since the each wave is of equal intensity, the effect of the submarine is to change the ambient field coming from behind it by a factor of a.

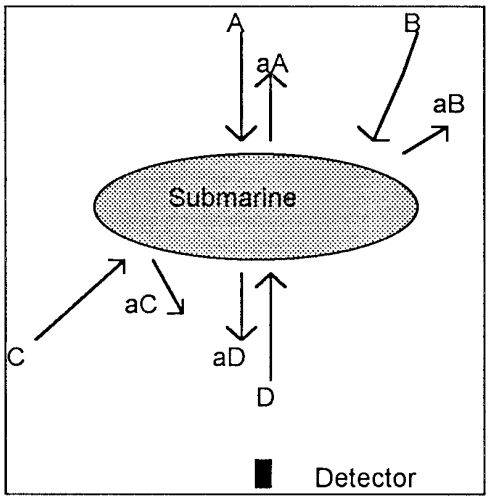

Figure 1. Effect of a submarine on ambient noise.

The formula for the reflection coefficient for a liquid-liquid boundary is

$$a = \left(\frac{R_2 - R_1}{R_2 + R_1}\right)^2,$$

where R_2 and R_1 are the acoustic impedances of the reflecting medium and the incident medium [Tucker and Gazey 1966, 91]. The liquid-solid reflection coefficient is far more complicated and, for our purposes, not significantly different. For the case of steel and water, we have $R_2 = 3{,}900{,}000$ and $R_1 = 154{,}000$ [Horton 1957, 26], which give $a = .854$. However, practical submarine designs minimize a to avert sonar detection; so a more likely value of a is probably about .1.

We can predict the difference in the ambient field from the submarine's presence. For ambient field intensity I_a, the reflected intensity is $I_r = aI_a$.

So the difference in decibels between the reflected intensity and the ambient intensity is

$$20 \log I_a - 20 \log(aI_a) = 20 \log I_a - 20(\log a + \log I_a) = -20 \log a.$$

For a steel submarine with no noise reflection damping, the difference is 1.37 dB; for a more realistic submarine, the difference is 20 dB; and for the perfectly noiseless submarine, the difference is infinite in dB, which corresponds to the entire ambient noise level (that is, for ambient noise level of 60 dB, the difference between the submarine and the ambient noise level is also 60 dB). Using a hydrophone (underwater microphone) array, we can detect the lower level of the ambient noise field and know that a submarine is there.

Microphones, Hydrophones, and Other Transducers

Most of the literature about hydrophones implies their use in linear arrays, generally either bottom-fixed or fixed to a vertical line that is stretched by a weight affixed to the sunken end of the line. Sonar receptors are mounted in either circular or cylindrical arrays with bearing determined by sound arrival times. In both of these general cases, the hydrophones used are omnidirectional, an undesirable feature for our purposes. Albers [1965] and Horton [1957] describe methods to determine source location using either linear or planar arrays of omnidirectional hydrophones or hydrophones with cardioid response patterns

There are three alternative techniques for limiting microphone pickup patterns that are vastly superior to reliance on a simple cardioid pattern.

- The "tuned port" microphone tube (see **Figure 2**) cancels off-axis sound by allowing it to arrive at the microphone element via several separate paths, slightly out of phase with each other, so that the off-axis sound interferes with itself, thus canceling the unwanted sound.

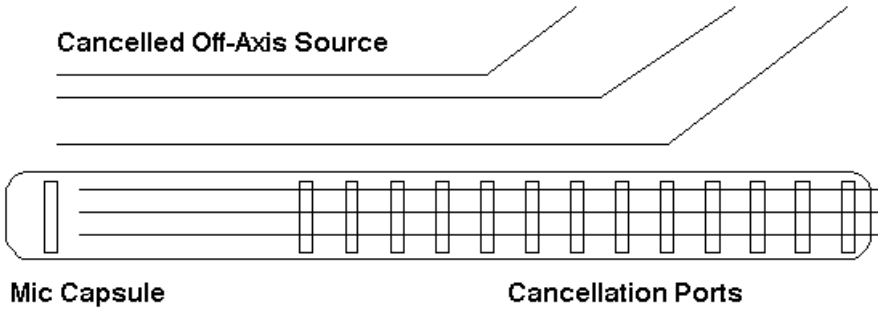

Figure 2. Shotgun tuned port microphone.

- A parabolic reflector selectively increases the impact of head-on waves with respect to the microphone element. The parabolic reflector involves several problems with water flow that we are not prepared to address.

The frequency limitations of both techniques present tradeoffs to various aspects of the model. Higher frequencies are absorbed at a greater rate in water (or any other medium), and as a result, the effective distance of the method will be adversely affected by choosing higher frequencies. Conversely, lower frequencies are less directional, have longer wavelengths (and so are less selective), and are much more difficult to deal with when attempting to decrease the beamwidth of the sound pickup pattern.

- "Pressure zone microphones" (PZM), also known as boundary microphones, are basically condenser microphone elements located close to a planar reflective boundary. Increasing the size of the boundary decreases the frequency at which the microphone becomes directional, in accordance with the formula $F = 188/D$, where F is the frequency at which the microphone becomes directional and D is the boundary dimension in feet (assuming a square boundary). Thus, a 2×2-ft boundary results in directionality at or above 94 Hz and a supercardioid response (rejection greater than 12 dB at or over 30° off-axis) at frequencies around 500 Hz [Bartlett 1991]. By using two panels with two microphones, an extremely directional response pattern can be achieved as a result of phase cancellation. Using 2×2-ft panels at an angle of 60° to each other, the on-axis sensitivity angle is approximately 7.5° (see **Figure 3**). Although this method is not mentioned in the underwater literature, we believe that we can achieve beam widths of arbitrarily narrow widths (with a 5° arc probably representing a practical minimum) by using this microphone design, as adapted to underwater use, in conjunction with Horton's phase-delay methods [1957, 247–248].

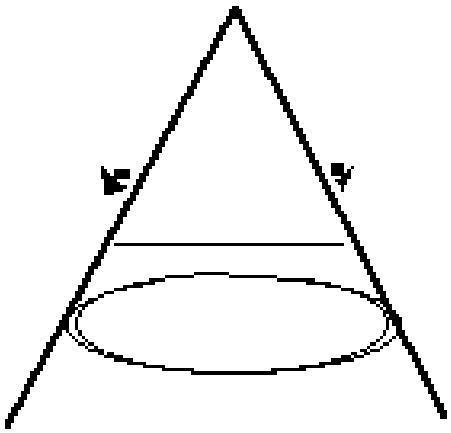

Figure 3. Pillon PZM shotgun, with 2×2-ft panels at an angle of 60° to each other. The PZMs are the dark objects outside the angle; the apex of the angle is pointed toward the sound source.

Details of Electronics

With two digital delay circuits between the microphone elements and a summing amplifier, we can bias the beam to the left or right of the center axis. The high accuracy and precision of available delay circuitry enable us to set the center point of the beam with high precision. Thus, we have a method for scanning along the plane parallel to the base of the Pillon structure that contains the two microphones, without the requirement of physical motion.

A second Pillon assembly located on a parallel plane vertically above the first assembly provides a second beam that we can move in conjunction with the first beam. By electronically delaying these summation voltages prior to a second summation, we can bias the combined beam, thereby defining a plane (actually a hemicylindrical surface) over which the beam traverses with discrete scanning intervals with respect to both the x- and y-coordinates.

Calibrating the beams is can be done by using a fixed reference source local to the PZM assemblies. A duplicate assembly located at a distance from the first provides a second plane over the area of observation. The intersections of target measurements between these two planes provides us with the third dimension through a change in basis; thus, we are able to detect the outline of an object in the area intersected by the cones bounded by the detection beams and the detection limits of the microphone elements. We use a notch filter with a very narrow bandwidth (a readily available device) and a computer controller in conjunction with the delay circuits to search selectively for any frequency within the limits of the microphone element range. Initially, we propose 1,500 Hz as a reasonable tradeoff between selectivity and sensitivity, but nothing in the mechanism or model requires this particular frequency in preference to any other. We suggest empirical analysis to determine which frequencies produce optimal results in real-world application.

The use of phase cancellation limits the arc angle of the scanning planes. This limit is more practical than theoretical, but realistically we cannot expect to achieve viable cancellations with delays in excess of a single wave period. Beyond this limit, we are faced with superposition problems, including accounting for the reflections from the boundary as they interact with incoming wavefronts. This means that the effective scanning area is limited by the choice of frequency, the sensitivity of the microphone at the selected frequency, and the quantization of the planar grid required for sufficient object resolution. Closer objects will be "fuzzier" but easier to range. Higher frequencies will improve image definition, but at the expense of range. If we are willing to complicate the detection apparatus, we can defer the selection of tradeoff criteria until measurement time. With this approach, we include motors that can vary the angles between the boundary planes and vary the orientation of the two scanning planes by rotating the vertical assemblies about the vertical axis. We can address the frequency characteristics by simultaneously processing several frequencies through a distribution amplifier and a set of narrow notch filters. The complete apparatus can be generalized as shown in **Figures 4** and **5**.

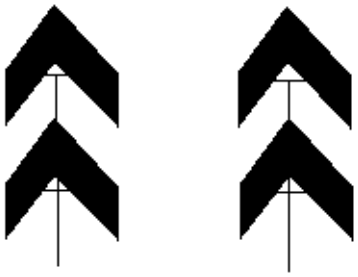

Figure 4. Detection array.

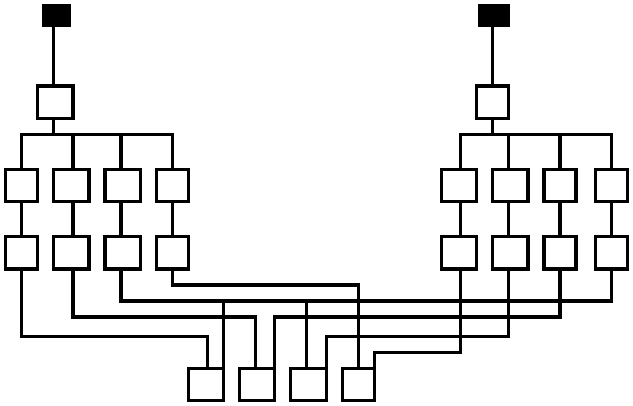

Figure 5. Electronic configuration.

Whether we exploit the characteristics of the boundary microphones or instead rely on conventional directional microphones, the scanning equations are identical. There are two factors of concern, the arrival time differential and the amplitude difference. We compute these as follows [Bartlett 1991, 44–45]:

$$\Delta T = \frac{\sqrt{D^2 + \left[\frac{S}{2} + D\tan\theta_s\right]^2} - \sqrt{D^2 + \left[\frac{S}{2} - D\tan\theta_s\right]^2}}{c},$$

where

T is the time differential between the microphones,

θ_s is is the desired source angle,

S is the spacing between the microphone elements,

c is the speed of sound, and

D is the distance from the microphones to the detection plane (the median of the depth plane).

Amplitude signals must be adjusted between the incident pairs in order to flatten the scanning plane. By weighting the amplitude values between the pair, we essentially eliminate the distance factor from the measured sector, which is

appropriate since we compute distance via triangulation rather than absolute amplitude. The amplitude difference in decibels between the two microphones of each pair is

$$\Delta = 20 \log \left[\frac{a + b \cos\left(\frac{\theta_m}{2} - \theta_s\right)}{a + b \cos\left(\frac{\theta_m}{2} + \theta_s\right)} \right],$$

where θ_m is the angle between the boundary planes (or between the microphone elements, in the case of directional elements), and a and b are constants that define the polar characteristics of the microphone elements.

There are several possible methods to image the three-dimensional target using our methods.

- We could try to use one additional convergence operation and scan both planes synchronously. This would yield sound pressure levels at points in three-space; in other words, it would result in a parallelepiped field of observation, demarcated by a grid on two of its faces, with pressure readings for each three-dimensional segment. This would yield the best image, but we are concerned that we are asking too much of the convergence techniques.

- We could determine the center of the target in each of the two observation planes and then find the distance to the center of the object using a simple triangulation (since we know the distance between the detection arrays and both angles of detection).

- We could apply a change of basis to one of the planes to map it onto the three-space coordinates of the other, weigh the measurements in the rotated plane by their z-coordinate values, and add the contents of the matrices. This would yield a two-dimensional matrix whose elements contain a value that should be proportional to the locations depth in the z-plane. A simpler interpretation of this is that we are intersecting the cylinder formed by extending the two-dimensional shape in each of the observation planes and interpreting this bounding shape as the shape of the object itself. This is more than sufficient image resolution given the limits of the arc formations. By summing the matrices derived from each frequency observed, we should get a reasonably detailed image with this method.

Deriving Object Characteristics from Data

Edge Detection

Through the array of hydrophones, we can construct a three-dimensional description of the target. By phase cancellation, we can construct precise narrow cones from the directional hydrophones and effectively orient them in an arbitrary direction. Each cone measures the sound intensity of a point in space, and we can represent this by a value in a two-dimensional matrix. By varying

the delay time, we can sweep the detector field across the plane representing the field of observation; and thus we construct a two-dimensional grid. Each element of the matrix contains the data for the sound intensity of a particular position in space. Since the submarine blocks out the ambient noise, we detect a decrease in sound intensity in the positions of the submarine relative to its surrounding space. This gives us an outline of the submarine in two dimensions relative to the perspective angle of the microphone array (**Figure 6**).

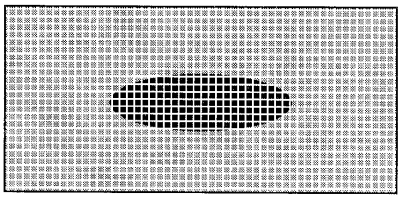

Figure 6. Plane scanned by an array of microphones. The dark region indicates the change in sound intensity level relative to the surrounding space, yielding the outline of the submarine.

From a single array of microphones, we cannot obtain the third dimension (the depth) of the submarine; therefore, the model proposes that another array of microphones be oriented at a slight angle to the observation plane and thereby obtain another two-dimensional view of the submarine relative to the perspective angle of this new array, which we call `array2`.

Knowing the distance between the two arrays of microphones and the angles of their respective planes of observation, we can determine the intersection set of any given set of values—that is, we can locate the same subfield in three-space by deriving the third spatial vector of the first plane from the two vectors (representing the x- and y-coordinates) of the second plane. Thus, we can determine the absolute location of the sound field depression in three-space.

Distance of scanned plane from the microphones

Referring to **Figure 7**, we observe that by the law of sines we have

$$\frac{\sin\beta}{b} = \frac{\sin(180^\circ - \alpha - \beta)}{a} = \frac{\sin\alpha}{c},$$

so that

$$b = \frac{a\sin\beta}{\sin(180^\circ - \alpha - \beta)}, \qquad c = \frac{a\sin\alpha}{\sin(180^\circ - \alpha - \beta)}.$$

Size of each grid box

From **Figure 8**, we see that the area of each grid point is simply the area of the circles traced out by the cone. The radius of the circle is $r = D\sin\theta$, so the area of each grid box is $A = \pi(D\sin\theta)^2$, where D is the distance from

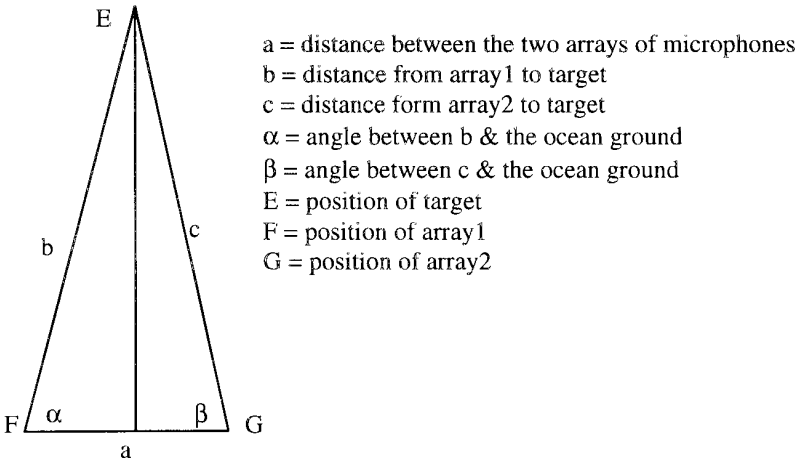

Figure 7. Geometry of microphone arrays.

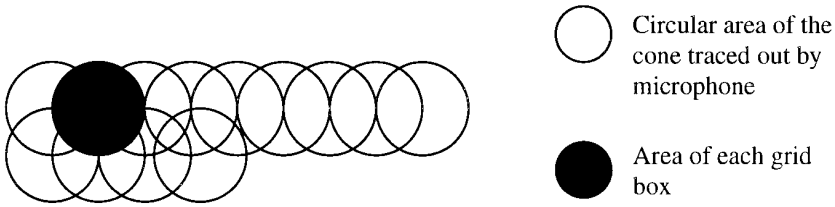

Figure 8. Geometry of grid boxes.

the microphone to the target and θ is the angle subtended by the cone of the microphone.

Since the two arrays of microphones are pointed in different directions in space, we have two different directional bases to work with. From `array1`, we have a basis with unit vectors X_1, Y_1, and Z_1; and each point of the plane represented by `array1` is expressed in terms of this basis with X_1 as the horizontal direction of the plane, Z_1 as the vertical direction of the plane, and Y_1 as the depth, perpendicular to the plane. Similarly, each point on plane 2 represented by `array2` will be expressed in termed of the basis with respect to `array2`. Looking at the two planes separately, we cannot determine the y-component of the picture in either case. Now, if we perform a change of basis for the second plane, we can obtain the depth of the submarine in terms of the first basis. This can be done simply by multiplying the coordinates of each of the points on plane 2 by the matrix

$$\begin{bmatrix} \cos\phi\cos\theta & -\sin\theta & -\sin\phi\cos\theta \\ \cos\phi\sin\theta & \cos\theta & -\sin\phi\sin\theta \\ \sin\phi & 0 & \cos\phi \end{bmatrix},$$

where θ is the angle between X_2 (unit vector in the x-direction of `array2`) and the X_1Z_1-plane, and ϕ is the angle between X_2 and the X_1Y_1 plane. Here we are assuming that the distance between the two microphone arrays is significantly small compared to the distance between the submarine and the microphone

arrays; therefore, we have the same origin for both bases and can use this single matrix to perform the change of basis. After the transformation of basis, we obtain the Y_1-components for each scanned point (X_{1i}, Z_{1j}) on plane 1. Thus we have constructed a three-dimensional "portrait" of the submarine; the location of each of its points is expressed with respect to the basis defined by `array1`.

Now we define a standard basis with respect to the ocean. That is, we use the plane of the bottom of the ocean as the xy plane and the altitude of the ocean as the z-axis. We now transform every point in the X_1, Y_1, Z_1 basis into the defined standard basis and obtain an "upright" three-dimensional figure of the submarine. By doing the transformation of the bases in real time on the computer, we can generate a three-dimensional continuous moving shadow of the submarine.

Size Computation

From the three-dimensional view, we can extract the positions of any given surface point. Hence we can compute the length and width of the submarine. Better yet, we can determine the volume of the ship by summing up all the grid space taken up by the submarine.

Velocity and Direction

To calculate the velocity of the submarine, we pick any point on the three-dimensional graph and locate its position at time t and then again at time $(t + \Delta t)$, then divide changes in each direction to find the components of the velocity.

Fun with Microphones

We experimented to test our model of how the submarine would affect the ambient field. Not having a large body of water, we did an air-based experiment. To simulate the ambient noise field, we used speakers connected to a noise generator. Since we did not have a large number of speakers to simulate the entire ambient noise field, we tried several representative orientations of speakers and detector. Our entire set of equipment was

- noise generator,
- amplifier,
- two speakers,
- sound-level meter,

- rolling trash can, and
- small table.

We set up the equipment in several different orientations. For each setup, we took several readings at each location to get an idea of the uncertainty of measurements. Our results are summarized at the end of this section.

In each setup, we pointed the detector directly toward the far noise source $c1$ (speaker 1) and took readings at three submarine positions A, B, and C. In every setup, the submarine in location A has a negligible effect on the ambient field.

Figure 9 shows the first setup, to simulate the noise field from the opposite side of the submarine, with the noise field from the detector side reflecting off the submarine back toward the detector.

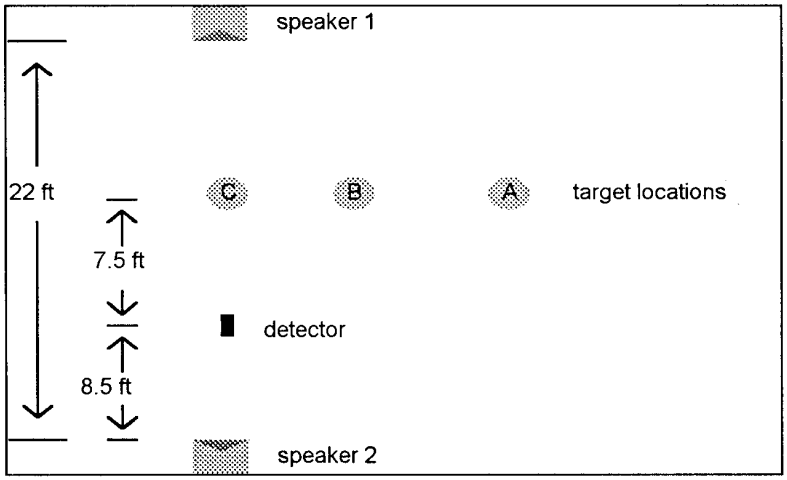

Figure 9. Setup 1.

Figure 10 shows the setup with the sources off to an angle behind the detector so as to measure the reflection of the ambient field off the submarine.

Figure 11 shows the sources at an angle in front of the detector, to get the opposite of the second setup. By combining the data from both the second and third setups, we hoped to get an idea of the difference between the magnitude of reflection from the submarine and absorption by the submarine.

Figure 12 is similar to **Figure 11** but with much greater distance to the detector (approximately two and a half times the distance for setup 3).

Table 1 summarizes our findings for the experiments. We decided that any measured change with an uncertainty greater than the change was not significant. In setup 1, the difference in the ambient field between the submarine in location A and in location B was negligible, whereas the difference between the submarine in location A and in location C was readily detectable. In the ocean, this would correspond to a measurable decrease in the ambient field when the array is directed toward the submarine, whereas when the array is pointed a little away from the submarine, the submarine would not be detected (i.e.,

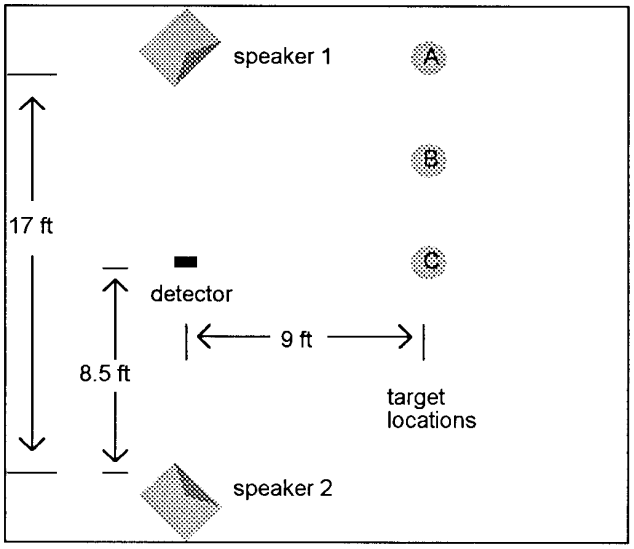

Figure 10. Setup 2.

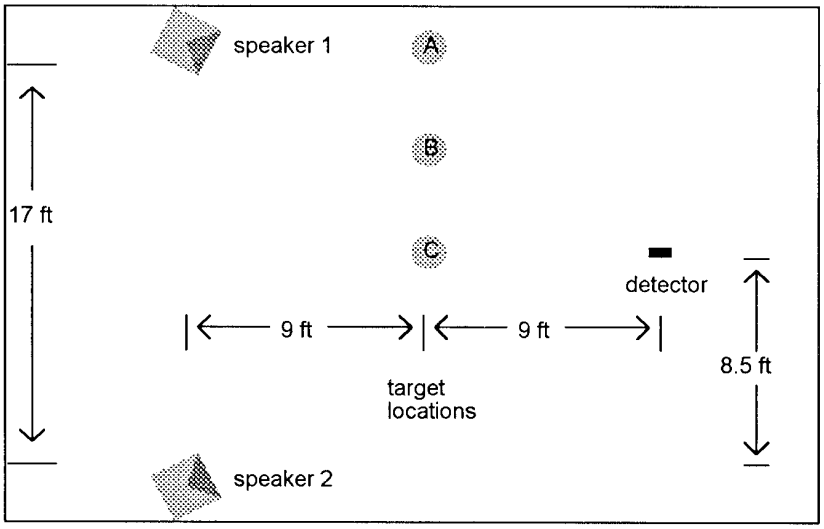

Figure 11. Setup 3.

Table 1.
Results of the experiment.

Setup	Change (in dB) between	
	A and B	A and C
1	not significant	-1.34 ± 0.16
2	not significant	$+0.54 \pm 0.27$
3	-2.20 ± 0.23	not significant
4	-1.04 ± 0.25	not significant

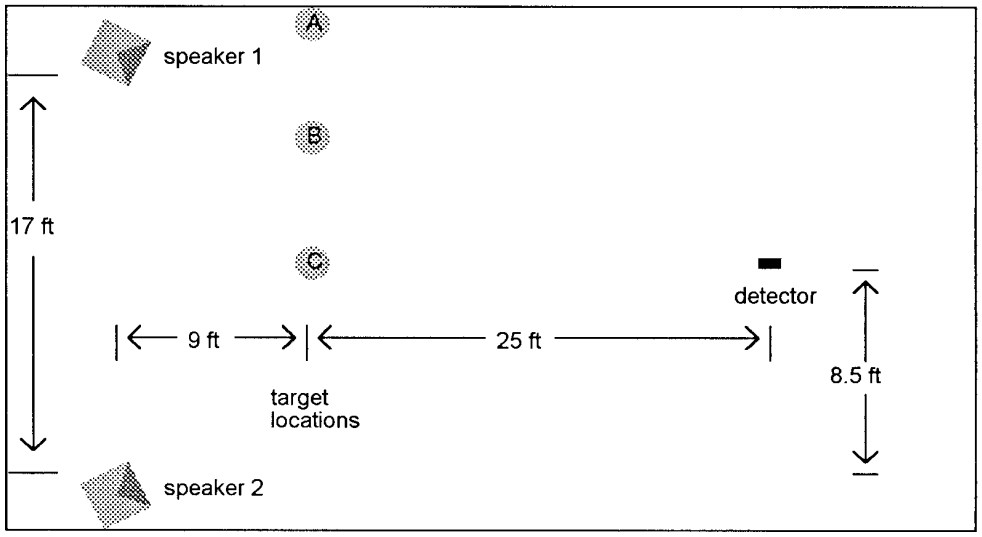

Figure 12. Setup 4.

ambient noise coming off-axis onto the submarine has a negligible effect). This conclusion is corroborated by the data in setups 3 and 4: When the submarine was not directly between the source and the detector, it had little effect on the ambient field; when it was directly between the source and the detector, the submarine had a large effect on the field. So, only noise that comes directly through the submarine to the detector is substantially affected.

The data from all four setups show that the reflected noise from the submarine is lower in intensity than the noise absorbed by the submarine. From setup 1 this is direct. From setup 2 in conjunction with 3, we see that sources of the same intensity reflect back to the detector at $+0.54$ dB, while the submarine interrupts sources at the same distance from the submarine by 2.20 dB. This difference becomes smaller in setup 4 because the detector can now register more of the ambient field, or the submarine takes up less of the detector's field of view. However, it is still detectable at 1.0 dB.

The data corroborate the fundamental idea that the submarine is detectable as an absence of noise in the ambient field.

Limitations

There are several problems that our model must deal with.

For a wave to "see" the target, that is, be reflected by the target, it must have a wavelength smaller than the the target. We assumed that the smallest dimension (length, width, height) of a submarine would be approximately 5 m. Smaller wavelengths might reflect off large fish; however, we are looking at an array of measurements, and from that we will know if the target is submarine size or fish size.

We also want the frequency to be as low as possible, since low frequencies

tend to attenuate in water far less than high frequencies. We can find the lowest usable frequency from the equation $c = \lambda f$, where c is the speed of sound in water (1,500 m/sec), f is the frequency of the wave, and λ is the wavelength (5 m). We find that 300 Hz is the lowest optimal frequency. There are other considerations, however. The directionality of the hydrophone array is highly frequency-dependent; in that case, the higher the frequency, the more directional the array can become. We decided that a frequency of 1,500 Hz would not attenuate too much for the additional directionality.

We also find that there is a maximum distance at which we can detect the submarine, even under conditions of ambient noise with fixed frequency and amplitude. The passive sonar equation,

$$\text{SL} - \text{TL} = \text{NL} - \text{DI} + \text{DT},$$

can be solved to find the maximum distance. Note that for detection of a submarine in an ambient noise field, $(\text{NL} - \text{SL})$ is the relevant noise relation term. This is the ratio of noise that the submarine will absorb (SL) to the ambient noise level (NL). Rearranging, we get

$$\text{TL} = \text{NL} - \text{SL} - \text{DT} + \text{DI}.$$

From Urick [1983, 385], we have

$$\text{DT} = 5 \log \left(\frac{dw}{t} \right),$$

where d is a parameter relating the probability of detection and the probability of false alarm, w is the frequency range, and t is the length of time listening for the submarine. From Urick [1983, 23], we have

$$\text{DI} = 10 \log \left(\frac{P_{\text{equiv}}}{P_{\text{actual}}} \right),$$

where P_{equiv} is the noise power generated by an equivalent nondirectional hydrophone and P_{actual} is the noise power generated by the actual hydrophone. Another way of looking at this is in terms of the relative area of detection for a nondirectional hydrophone compared to the actual array of hydrophones, so an equivalent equation is

$$\text{DI} = 10 \log \left(\frac{\text{surface area of } 360^\circ \text{ scan}}{\text{surface area of actual scan}} \right).$$

Since the arclength of a section on a sphere is given by $s = r\phi$ for angle ϕ measured in radians, and since for small ϕ the arclength s is approximately equal to the diameter of the circle inscribed on the sphere, we get the surface area of the scan as

$$A = \pi \left(\frac{r}{2\phi} \right)^2.$$

So we have

$$\text{DI} = 10\log\left(\frac{4\pi r^2}{\pi\cdot\frac{r^2}{4}\cdot\phi^2}\right) = 10\log\left(\frac{16}{\phi^2}\right) = 12 - 20\log\phi.$$

Finally,

$$\text{TL} = 20\log r + \left(\alpha\times 10^{-3}\right) r,$$

with α given by [Urick 1983, 108] as

$$\alpha = \frac{0.1f^2}{1+f^2} + \frac{40f^2}{4.100+f^2} + 2.75\times 10^{-4}\times f^2 + 0.003,$$

for f expressed in kHz [Urick 1983, 108]. For our purposes, we neglect the term involving α.[1]

Using these values, we can now solve for r as

$$\log r = .05\left(\text{NL} - \text{SL} - 5\log\left(\frac{dw}{t}\right) + 12 - 20\log\phi\right).$$

Included as an appendix are graphs for a perfectly absorbing submarine of range vs. time (**Figure 13**), and, given a sweep time of 1 second, range versus ϕ and $(\text{NL}-\text{SL})$ (**Figure 14**). Notice that if the submarine is not perfectly absorbing, it quickly becomes hard to detect. For example, recall that the difference between the ambient noise field alone and the field with the submarine present is $-20\log a$. For $a = .1$, we have a difference of 20 dB. If this is the case, the submarine is detectable only within 1 km or less, even with a low angle of observation. We can adjust for this somewhat with a slower scan; changing from 1 sec to 3 sec would almost double the range limit for detection. However, the better the absorption factor of the submarine, the easier for us to detect. This implies a tradeoff between absorption for active sonar and reflectivity for passive sonar.

The uncertainty in successive readings limits the smallest noise difference that we can detect, lowering the range limit by approximately the standard deviation of the noise. A reference deviation of a hydrophone is 0.5 dB [Wagstaff and Baggeroer 1985].

There is variation in the amplitude of the ambient field. If the variation is 7 dB, any difference of 7 dB may be insignificant random noise; again, this lowers the range limit by this deviation level. That is, it will lower the effective $(\text{NL}-\text{SL})$ difference. A common standard deviation in the ambient noise field over one-hour periods is about 6 dB, which results from changing wind speed

[1] AUTHORS' NOTE: The relative size of the α-term depends on the frequency used. Either we see the object or we don't, and some variation in TL is negligible with respect to both detection and imaging. At our suggested frequency of 1500 Hz, and at a range of 3 km, the α-term is about 38% of the total. But our proposed apparatus could use arbitrarily low frequencies. At 200 Hz (probably a practical minimum), the effect of the omitted term at a range of 10 km is only 5% of the total.

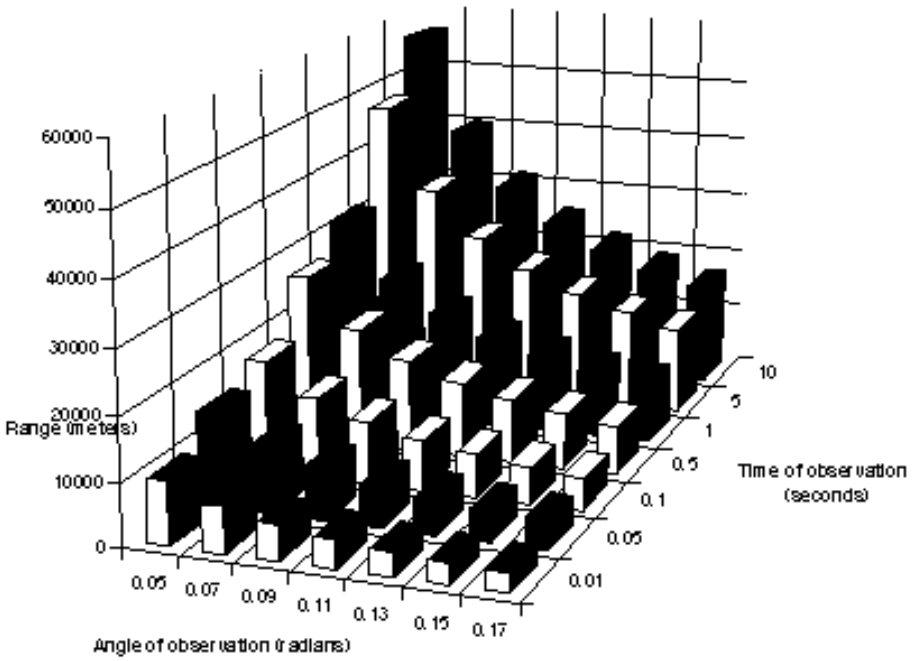

Figure 13. Range vs. time of observation and vs. angle of observation.

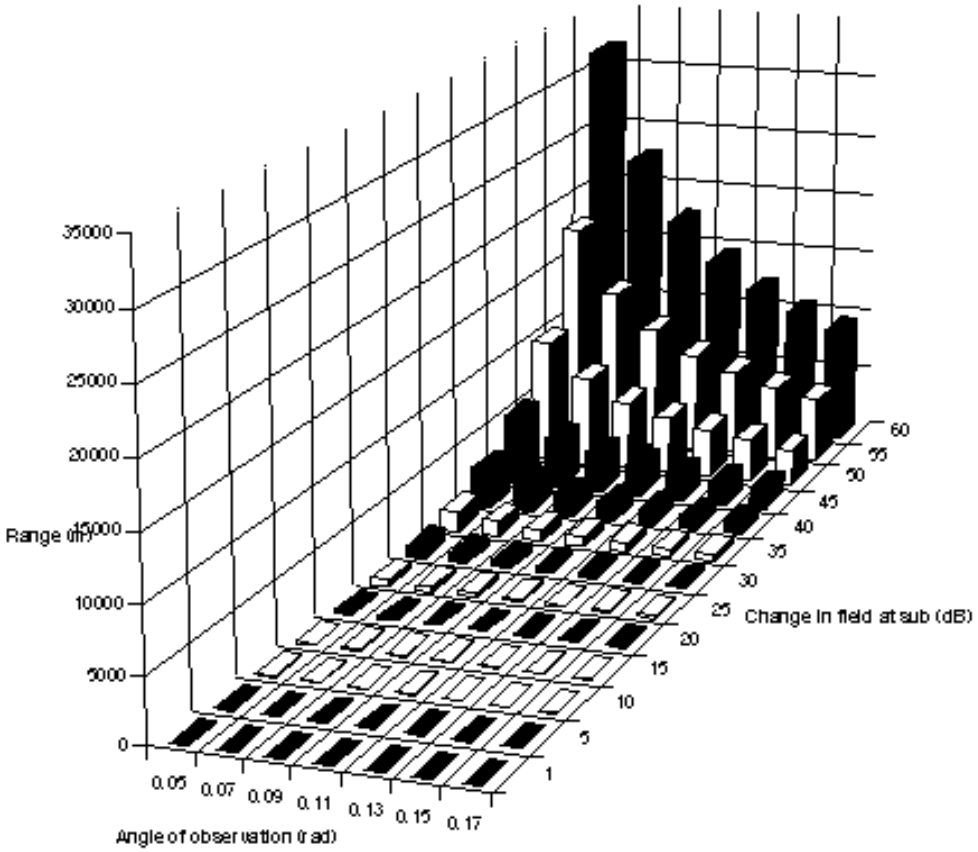

Figure 14. Range vs. change in field with reflection and vs. angle of observation.

during the observing period [Urick 1983]. However, we postulate that the deviation over one- or two-second intervals would be far lower, as the wind speed change would be far less. A reasonable assumption would probably be about 1–2 dB.

We can get around these deviations to some extent by taking a lot of data. That is, if on several successive scans we find a deviation of 1 dB in the same location, it is more likely to be caused by a submarine than by a statistical deviation. Thus, we can probably notice a deviation of less than half the deviation of the ambient noise field or the hydrophone.

A very large problem in positioning the submarine is not directly apparent from our model. The temperature difference at different levels in the ocean combined with pressure gives varying speeds of sound at different depths. While the maximum difference in the speeds of sound between depths of 0 to 1,500 m is only about 10 m/s, a velocity gradient as well as a speed gradient are forced upon the sound wave. This leads to rather strange characteristics in the path of travel of sound (called a *ray*) in the deep ocean (see Robinson and Lee [1994] and Tucker and Gazey [1966] for pictures of ray tracings).

An interesting note is the fact that a "shadow zone" occurs at some distance from the source. This impacts the placement of the hydrophone; we do not want it in the upper portion of Region II (200–1,500 m) or the lower part of Region I (0–200 m), as these areas are missed by all sound rays at some distance from the source. It also requires some data processing once the location of the submarine has been found. From the equation for the length of an arc, it is easy to verify that

$$\cos\phi_m = \cos\phi_n + \frac{d_n - d}{R_n},$$

where ϕ_m and ϕ_n are the angles of inclination of the incident wave at the detector and at the target, respectively, d_n is the depth of the detector; d is the depth of the source; and R_n is the radius of curvature of the ray path [Tucker and Gazey 1966, 105]. The horizontal distance to the target, s_n, is given by

$$s_n = R_n(\sin\phi_n - \sin\phi_m).$$

Combining these equations, we can solve for d_n and ϕ_n if we use the distance to the target computed by the array as s_n. Of course, this will not be exact, but iterations using ϕ_n to find a new s_n will give a fairly exact picture. The only problem is when the ray has been reflected by the surface of the ocean or the surface between two regions, or if the ray has gone through more than one region. If this is the case, the actual location of the target is far more difficult to ascertain, and is beyond the scope of this paper.

Discussion of Alternative Approaches

There are several other methods of detecting the submarine.

One very interesting and purely hypothetical method is the idea of a "sound camera." This incorporates a sound lens and a nonreflecting box. It works in much the same way as an ordinary camera: It takes the view, inverts it through the lens, and displays it inside the box. This display would probably be in the form of a pressure-sensitive liquid that would have some characteristic change stemming from very small pressure changes and be proportional to those pressure changes. Thus, one would be able to get a "picture" of the sound that the lens is "seeing" by observing the changes in the liquid in the box. Naturally, there are several problems, the foremost being the velocity of sound. Since the velocity of light is essentially infinite with respect to the camera, the entire plane of observation is "seen" at the same time. However, because the speed of sound is very finite, things with distances differing by kilometers or more would be offset in the picture by times of seconds. If they were relatively constant sources, this would not be of much consequence as the source 10 sec before the picture would probably be much the same as the source at the picture. This would be a hard device to design, but it would be very useful in searching the ocean for sound sources. Reconstructing a visual image from the sound field could probably be done using acoustic holography, a technique described below as an additional alternative approach.

A second alternative is a variant on a method described by Farrah et al. [1970] regarding sound holography. They describe the techniques for reconstructing a holographic image using sonar data. Just as a laser can be used to read the pits of a compact disk, or scan the surface of an LP recording, it would be possible for us to "read" the sound waves in water by measuring the refraction of a laser beam passing through them. We would be able to read specific waves from within the compendium of the ambient field, theoretically to any precision. Scanning the plane representing the observation threshold, we could reconstruct a plane parallel to it by subtracting the waves whose source can be determined to be off-axis to the observation plane. Time provides the third spatial dimension, so we should be able to locate and measure a target item within the observation space.

Conclusion

Our model allows for the prediction of a perfectly absorbing submarine to extreme distances with exceedingly good technology (see **Figure 13**). However, with a more realistic model of a submarine, the detection limit falls to less than 5 km, and possibly less than 1 km. Nevertheless, the array can also detect a submarine generating non-machinery-based noise (such as flow noise and cavitation), so we may hear a quickly-moving submarine at large distances.

References

Albers, Vernon. 1965. *Underwater Acoustics Handbook*, vol. 2. University Park, PA: Pennsylvania State University Press.

Bartlett, Bruce. 1991. *Stereo Microphone Techniques*. Stoneham, MA: Butterworth-Heineman.

Cox, Albert. 1982. *Sonar and Underwater Sound*. Lexington, MA: Lexington Books.

Farrah, H.R., E. Marom, and R.K. Mueller. 1970. An underwater viewing system using sound holography. In *Proceedings of the 2nd International Symposium on Acoustical Holography*, edited by A.F. Metherall and L. Larmore, 173–184. London: Plenum.

Hassab, Joseph. 1989. *Underwater Signal and Data Processing*. Boca Raton, FL: CRC Press.

Horton, Joseph. 1957. *Fundamentals of Sonar*. Annapolis: U.S. Naval Institute.

Metherell, A.F., and Lewis Larmore (eds.). 1970. *Acoustical Holography*, vol. 2. New York: Plenum Press.

Naval Sea Systems Command. 1984. *Ambient Noise in the Sea*. Washington: Dept. of the Navy.

Nisbett, Alec. 1974. *The Use of Microphones*. New York: Hastings House.

Robinson, Allan, and Ding Lee (eds.). 1994. *Oceanography and Acoustics*. Woodbury, NY: AIP Press.

Tucker, David, and B. Gazey. 1966. *Applied Underwater Acoustics*. London: Pergamon Press.

Urick, Robert J. 1967. *Principles of Underwater Sound for Engineers*. New York: McGraw-Hill.

_____. 1982. *Sound Propagation in the Sea*. Los Altos, CA: Peninsula Publishing.

_____. 1983. *Principles of Underwater Sound*. 3rd ed. New York: McGraw-Hill.

Wagstaff, Ronald, and Arthur Baggeroer (eds.). 1985. *High-Resolution Spatial Processing in Underwater Acoustics*. NSTL, MS: Naval Ocean Research and Development Activity.

Wilson, Oscar. 1985. *An Introduction to the Theory and Design of Sonar Transducers*. Washington: U.S. Government Printing Office.

How to Locate a Submarine by Detecting Changes in Ambient Noise

Carl Leitner
Akira Negi
Katherine Scott
University of North Carolina
Chapel Hill, NC 27599–3250

Advisor: Ancel C. Mewborn

Summary

We generate an artificial ambient noise field and give an algorithm for locating a submarine.

Our model for the ambient noise field allows any frequency range, but we use 30 kHz, where environmental sounds, such as surface activity and biologics, dominate. We assume a normal distribution of noise in the area of interest and that a submarine impedes measuring the noise beyond it.

Our recognition algorithm is simple: We look for contour lines on our ambient noise field, then for closed contours of the right size, and finally for the intensity patterns that match those of submarines. To help detect significant changes, we smooth the data. Since our algorithm is reasonably fast, it can also detect the changing locations of the center of the submarine as it moves and hence compute the speed and direction of the submarine. We calculate the size and the depth of the submarine by finding the maximum intensity of the dampening and the average dampening in the area around it.

Using the algorithm on our artificial data, we can spot a submarine within 25 m of the actual location, when we are working with an area of 7 km $\times$ 7 km, with the ratio of about 1.8 of maximum dampening effect to standard deviation of ambient noise. We were within a computer error scale for determining the speed of the submarine. Our model at the latest stage computed the depth of the submarine and the size of the submarine within a factor of 10^2.

Physics Facts

- Acoustic intensity is related to the distance r that sound travels and the attenuation coefficient α [Apel 1987, 368] via

$$I(r) = I_0 \left(\frac{r_0}{r}\right)^2 e^{-\alpha(r-r_0)},$$

where I_0 is the acoustic intensity at distance r_0.

  - "[B]elow 20 Hz, a frequency-independent attenuation coefficient, α_1, occurs and is approximated by: $\alpha_1 = 6.9 \times 10^{-7}\text{m}^{-1}$" [Apel 1987, 341, 369]. At higher frequencies, the attenuation constant is a function of frequency; at still higher frequencies, the dominant factor is water viscosity [Apel 1987, 368–371].
  - "[T]he [acoustic] energy radiates in all directions as a spherical wave," which causes an attenuation proportional to 1 / distance^2. [Dera 1992, 434]

- "[T]he velocity of sound in the ocean varies from 1430–1540 m / s near the surface to 1580 m / s at great depths" [Dera 1992, 436].
- Different underwater sounds have their own specific frequency ranges, such as shipping and machinery (less than 2 kHz), biologics (0.1 to 100 kHz), and ice (several Hz to kHz) [Hassab 1989, 3].

Assumptions and Justifications

- We monitor a small area of ocean, say 10 km $\times$ 10 km, with a uniform depth of D. In a small area, ocean depth tends to vary less, and we do not have to consider the curvature of the earth. If we need to locate a submarine in a larger area, we divide up the area into 10 $\times$ 10 squares..
- Effects of submarines on ambient noise:
  - When there is a submarine between a sound and the observation point, the ambient noise field is dampened at nearby observation points, the same effect as the submarine absorbing almost all sounds (about 98%). Most of the sound reflected off an ellipsoid is scattered far away from our sensors.
  - The dampening effects are independent of the speed of the submarine.
- Assumptions about the ambient noise field:
  - We have instruments to measure acoustic intensity as a function of the point on the plane at a particular depth d.
  - Sensors do not malfunction.
  - Atmospheric sound does not propagate to the ocean [Pain 1983, 158].
  - The dominant effect of sound hitting the bottom of the ocean is scattering off bumps.
  - We set d = 100 m, for realism and for simplicity ("sources located in shallow surface ducts can give complex ray arrival patterns" [Munk et al. 1995, 382]).

  - Just 14% of sound is transmitted through the steel-water boundary [Pain 1983]; so $(14\%)^2 = 2\%$ of the sound is transmitted completely through the submarine's two steel-water boundaries.
  - We measure a small interval of frequencies. "At the high-frequency end, sound absorption by seawater is very high. ... At the low end, below 1 cps, one has great difficulty in generating sound (except with earthquakes and very large explosions)" [Tolstoy and Clay 1966, 3].
  - We ignore the effects of ambient water velocity on sound propagation, the common practice [Keller 1977, 2].

- Other properties of submarines:

  - The submarine shape is an ellipsoid. The thickness of the submarine is negligible in comparison to the depth of the ocean and the distance from which we are measuring the noise field.
  - The speeds of submarines do not exceed 35 knots [Friedman 1984, 105].
  - The lengths of submarines vary from 50 m to 150 m; the width is always about 10 m Friedman [1984].
  - The submarine is parallel to the ocean surface; for operational reasons, submarines do not tilt by a very large angle.

Development of the Model

Choice of Frequency

We limit our noise detection to a convenient frequency range, near 30 kHz; our model can adapt to different choices.

Noise near 30 kHz is caused mostly by surface water movements, thermal activity, and biologics, which either affect relatively large areas uniformly, or are distributed randomly throughout the region. Thus, it is reasonable to assume that intensity of noise in the ambient noise field is distributed according to a smooth function with random fluctuations according to a normal or uniform distribution.

We could alternatively look at low frequencies, near a few Hz, for which we would need a different attenuation constant. The dominant noise in this range is seismic activity, and seismic information is readily available in real life.

Dampening Effects

We assume that reflections off the submarine and off the ocean bottom are dominated by scatterings, which means that the effects of reflection cannot be measured in the range in which we are working. This means that any sound that hits the submarine effectively disappears, i.e., it has the same effects as if the

sound were absorbed by the submarine. Thus, we assume that the submarine blocks all sounds that come toward the observation point from the other side.

An ellipsoidal submarine presents an elliptical profile, which at depth d can be represented by the equation

$$\frac{(x-x')^2}{a^2}+\frac{(y-y')^2}{b^2}=1,$$

where (x', y') is the center of the ellipse. In addition, we assume that the ocean depth is uniformly D (see **Figure 1**).

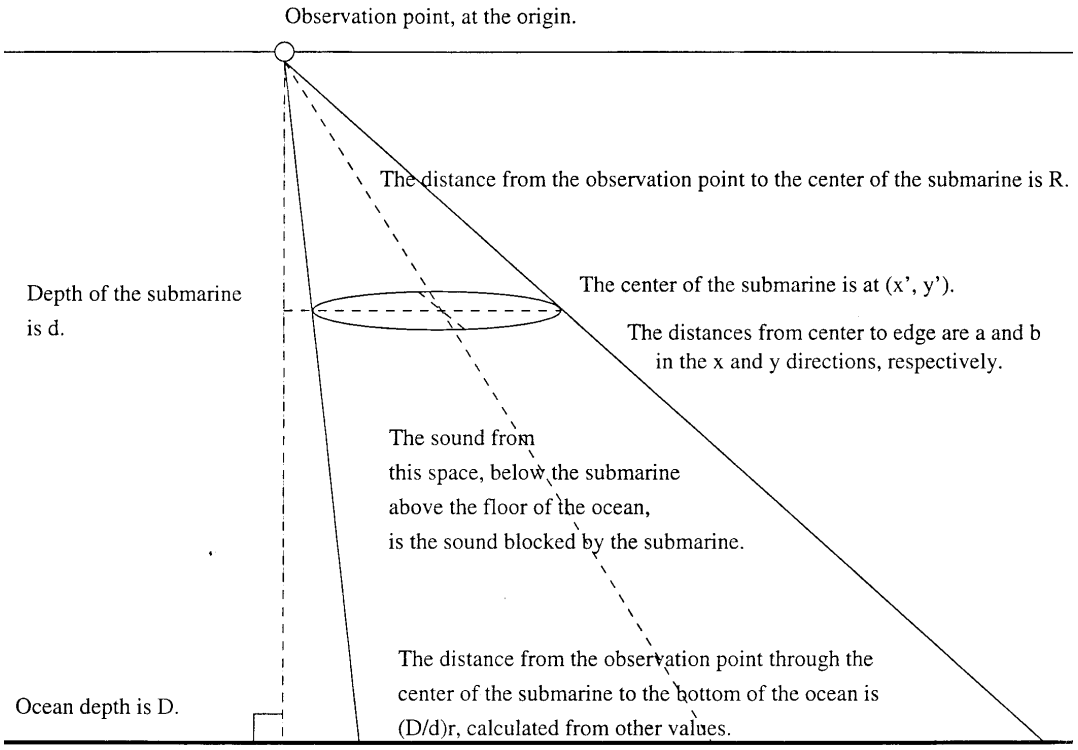

Figure 1. Setup for the integral to calculate the dampening caused by the submarine. The z-axis points downward, for convenience.

The sound intensity measured is inversely proportional to the square of the distance from the source. Ignoring seawater's absorption of sound, the total sound blocked is expressed by the integral

$$\iiint_V \frac{1}{(x^2+y^2+z^2)}\,dx\,dy\,dz \qquad \textbf{(1)}$$

where V is the volume of water blocked by the submarine. The evaluation of the integral proved exceedingly difficult, so we used the following approximation:

$$\int_R^{(D/d)R} \frac{ab\pi\left(\frac{r}{R}\right)^2\left(\frac{d}{R}\right)}{r^2}\,dr \qquad \textbf{(2)}$$

where $R = \sqrt{x'^2 + y'^2 + d^2}$ is the distance from the observation point to the center of the submarine. This integral reduces to

$$\frac{ab\pi}{R^2}(D-d). \tag{3}$$

Note that the dampening is *not* affected by the direction in which the submarine is pointing.

Water dampens the sound by a factor of $e^{-\alpha(r-r_0)}$, where r is the distance from the source of the sound, r_0 is a reference distance (usually taken to be 1), and α is a constant dependent on the frequency of the sound that we are measuring. For example, $\alpha = 3\times 10^{-2}$ dB / km at 30 kHz [Apel 1987]. Incorporating this factor into **(1)** makes the integral even harder to evaluate. Thus, we use **(2)** to get

$$\int_R^{(D/d)R} \frac{ab\pi\left(\frac{r}{R}\right)^2\left(\frac{d}{R}\right)}{r^2} e^{-\alpha(r-r_0)}\,dr,$$

which reduces to

$$\begin{aligned}\int_R^{(D/d)R} \frac{ab\pi d}{R^3} e^{-\alpha(r-r_0)}\,dr &= \left.\frac{abe^{-\alpha(r-r_0)}\pi d}{-\alpha R^3}\right|_R^{(D/d)R} \\ &= \frac{abde^{\alpha r_0}\pi}{\alpha R^3}\left(e^{-\alpha R} - e^{-\alpha RD/d}\right).\end{aligned}$$

Setting $r_0 = 1$, we get

$$\frac{abde^{\alpha}\pi}{\alpha R^3}\left(e^{-\alpha R} - e^{-\alpha RD/d}\right). \tag{4}$$

We see that the amount of noise measured at each point is approximately proportional to $2\pi e^{\alpha}/\alpha$. For derivation of the integrals and these values, see the **Appendix**.

Our computer model reveals that **(4)** produces similar maximal dampening effects to **(3)**, but the effects of **(4)** are registered in an area roughly one-third to one-half the radius of **(3)**.

We use **(3)** with the additional constraint of a smaller dampening radius. Dampening is more than some criterion constant c when

$$h < \sqrt{\frac{ab\pi(D-d)}{c} - d^2} < \sqrt{\frac{ab\pi D}{c}},$$

where h is the horizontal distance from the observation point. Note that the middle expression depends on the unknown depth of the submarine; the expression on the right contains known quantities, except for the length of the submarine, which we assume varies by a factor of only 3. We can assume c to be a relatively small constant, such as the standard deviation of the ambient noise, properly scaled. Doing so produces an area of radius approximately 100 m where the effects of the submarine are detectable.

We still need to consider the motion of the submarine. Since the typical speed of about 30 knots (15 m/s) is greatly less than the speed of the sound (1430–1540 m/s), the time delay due to movement is small enough to ignore.

Analysis of the Problem and Model Design

We developed a graphical simulation in MATLAB, with data on a grid (each square representing a sensor) and with color to indicate acoustic intensity.[1]

Generation of Simple Random Noise

We had MATLAB generate a data set containing only random noise. To start, we used 1 as the mean of the noise intensity. The adjustment values at each point came from a uniform distribution on the interval $(-0.00005, 0.00005)$. Later, we used a normal distribution mean 0 and standard deviation 0.00005 to create random noise.

Dampening Effect of the Submarine

We first guessed that the submarine would create an ellipse-shaped area of dampening, with greater dampening in the center than at the edges.

Later, we derived two dampening functions using **(3)** and **(4)**. The "attenuated" model included the effect of sound absorption by seawater, while the "unattenuated" model did not.

Smoothing Functions

Having simulated the data, we sought a way to locate the submarine.

We began by removing the random noise via a smoothing function. The first smoothing method, "consecutive," used least-squares to fit a line to the first 5 points in each row, and then to the next 5 points, etc. We specified an acceptable range of "noise" and then checked each point to see whether it was within that (vertical) distance from the least-squares line. If so, the value of the point was reassigned to the value on the regression line. We also wrote a routine that performed the same type of smoothing on each column. This method proved ineffective, but it inspired a second method.

The second method, "overlapping," was very similar to the consecutive method, except instead of doing the process on points 1 to 5, then 6 to 10, then 11 to 15, we did the process on points 1 to 5, then 2 to 6, then 3 to 7, etc. This was much more effective. We found that running the overlapping row and column

[1] AUTHORS' NOTE: The source code, together with a fuller version of this paper and additional figures, is available at `http://www.math.unc.edu/Undergrads/cleitner/ambient` .

smoothers several times each did a very good job of enhancing the contrast between average noise and dampening effect.

If we take the average of 5 points at each step, we theoretically expect the standard deviation of the mean to be reduced by a factor of $1/\sqrt{5}$ at each step. After 6 smoothings, the standard deviation was reduced by about $1/12$, which is greater than $(1/\sqrt{5})^6$. However, since there are many points that we are not affecting when they are outside of our noise tolerance (twice the standard deviation), this is consistent with the theoretical result.

Location of the Submarine

We first had MATLAB produce a contour map of the data. Then we located contours that might represent a submarine's detectable dampening radius. Finally, we checked these "suspicious" contours for whether they behaved like submarine-dampening contours.

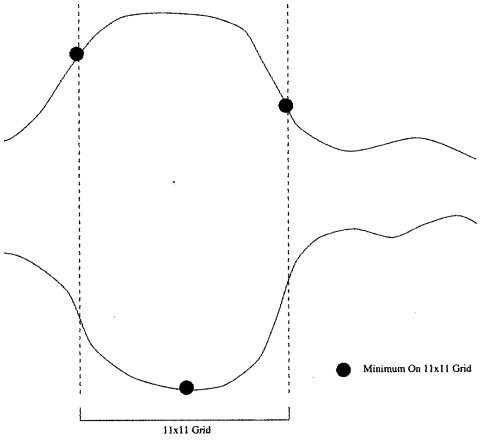

Figure 2. Cross-section view of the ambient field on an 11×11 grid. The marked points are minima.

We identified the contours that stayed within the grid. Then we eliminated all contours with diameters significantly greater than the detectable dampening radius. We found the "center of mass" of each contour by averaging all of the contour's vectors. Then we found clusters of contours by grouping those whose "centers of mass" were within one-fourth of the detectable dampening radius. Within each cluster, we averaged the contour centers to get a cluster center.

We eliminated contour clusters with fewer than 4 contours because submarines tend to have 4 or more contours associated with them.

For each cluster, we then searched the 11×11 square around the center and found the absolute minimum intensity in that square. We looked at the 4 grid squares adjacent to the minimum to see whether the absolute minimum was a local minimum (the 4 adjacent squares might not all be part of the 11×11 square). If the center was not a local minimum, then the cluster marked an area of high intensity rather than an area of dampening, so we eliminated such clusters from our set.

The remaining clusters should indicate the presence of submarines at their centers.

Speed and Direction

To approximate the speed of the submarine, we locate it at two different times and use the time-distance formula from algebra to find the speed. We would approximate the direction as a vector from the first location of the submarine to the second.

Size and Depth

We use the unattenuated model, **(3)**, to approximate depth and size. Recall that the dampening is

$$I = \frac{ab\pi}{R^2}(D-d).$$

The maximum of this function occurs when R is smallest, which is when $R = d$. Also, by rewriting the equation, we see that

$$\begin{aligned} ab\pi(D-d) &= I(h^2+d^2) \\ &= ab\pi(D-d) = m\big(I(h^2+d^2)\big) \\ &= ab\pi(D-d) = d^2 m(I) + m(Ih^2), \end{aligned}$$

where h is the horizontal distance and m is the averaging function over some region. By substitution, we see that

$$d^2 \max(I) = ab\pi(D-d) = d^2 m(I) + m(Ih^2),$$

or

$$d = \sqrt{\frac{m(Ih^2)}{\max(I) - m(I)}}.$$

Since we assume the width to be constant, we can assume a to be constant. So the length b of the submarine is

$$\frac{d^2 \max(I)}{a(D-d)\pi}.$$

Figures

Figures 3–6 are a clear visual representation of our algorithm. The mean of the measurements in these figures is scaled by $2\pi e^{\alpha}/\alpha \approx 2.095 \times 10^4$.

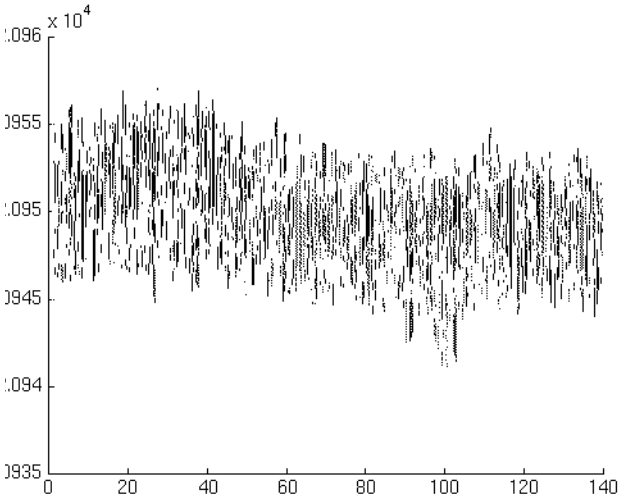

Figure 3. Plot of the sample data.

- **Figure 3** is a plot of our sample data. Where the graph dips down is where the submarine is. (The figure is a horizontal view of the 3-D graph.)
- **Figure 4** is a contour map of the 3-D data from which **Figure 3** was made.

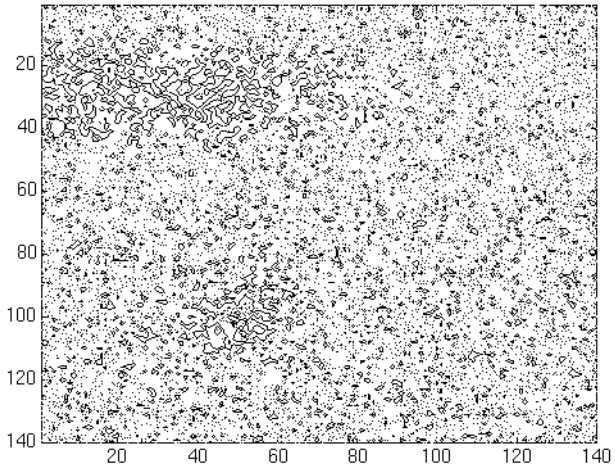

Figure 4. Contour map of the sample data.

- **Figure 5** shows the effect of "overlapping" smoothing on the data from **Figure 3**. Notice that the noise-to-dampening ratio has drastically improved. Unfortunately, the maximum dampening has also decreased, so it is necessary to return to the original data for information on submarine size.
- **Figure 6** is a contour map of the data in **Figure 5**. Notice how many fewer contours appear on **Figure 6** than on **Figure 4**.

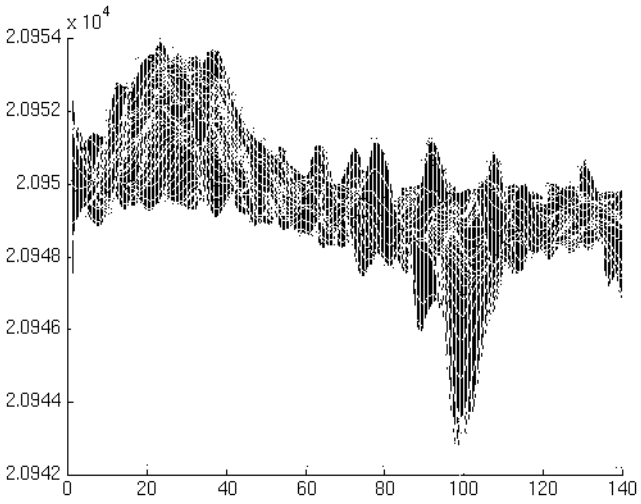

Figure 5. The effect of "overlapping" smoothing on the sample data from **Figure 3**.

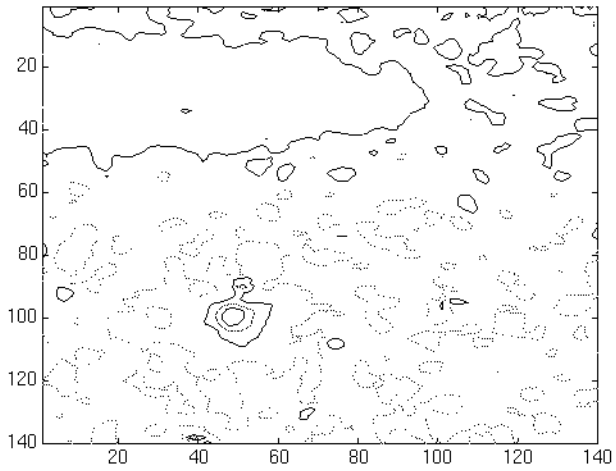

Figure 6. Contour map of the data in **Figure 5**.

Results and Model Testing

Using our algorithm on our modeled data, we located a submarine within 25 m of the actual location, working with an area of 7 km by 7 km, with a ratio of about 1.8 of maximum dampening effect to standard deviation of ambient noise. Considering that our scale is 25 m to a computer unit, this is the best that we can hope for. This remained true when we assumed that the mean noise level was not constant and replaced it with some smooth function with relatively small variation.

We were also within this same range for determining the speed of the submarine. Our model correctly located the center of the submarine at each subsequent step in the model, even when we assumed that the submarine is moving.

Since our model is functional with a ratio of about 1.8, we expect it to work just as well when the ratio is lower.

Unfortunately, our model at the latest stage computed the depth of the submarine and the size of the submarine only to within a factor of 10^2.

We need to test the model with various standard deviations in the random noise and with various means for the random noise as a function of grid location. We also need to do real-world experimentation.

Strengths and Weaknesses of the Model

Strengths

- Our computer program contains many constants that are easy to adjust. These include:
  - Scale: We considered one unit equal to 25 m of real scale.
  - Size of the field we observe: We chose a grid of 140 $\times$ 140, but MATLAB is capable of handling a much larger matrix.
  - Depth of ocean.
  - Attenuation constant: We used the 3×10^{-2} dB / m^2 as the default, since that is Apel's value for 30 kHz sound [1987].
  - Conversion factor from computer scale noise to real noise: We chose a value that was theoretically easy to work with (2.095×10^4), but it could be replaced with a real average amount of noise.
  - Absorption / reflection factor: In our current model, the amount of noise that is actually transmitted rather than blocked is 98%.
  - Radius of the area affected: This is a constant relative to several factors, including what noise level is expected and what level of statistical accuracy we want.
- The graphical interface is a very intuitive way to organize data and to see the effects of each stage of the modeling.
- Our computer model of ambient noise considers dampening caused by a submarine; the model is faithful to the sound wave propagation and absorption patterns in the ocean.
- The computer algorithm is fast; it ran in a few seconds on a SPARC 20 station. Hence, our model can be applied in real time or near real time.
- We considered many possible factors that could affect the noise field, including marine biology, surface activities, human-generated noise, geothermal activity, and seismic activity.

- Our calculations are independent of scale, as long as the submarine is more than 1 grid square in length.
- We could incorporate other smoothing functions, such as polynomial smoothing or spline approximation on 20 points or so.

Weaknesses

- Our model assumes a very large number of sensors (the square of the number of grid units on a side), which would mean a high cost of implementation. To get adequate data from fewer sensors would require more-sophisticated sensors and might pose more difficulty if a sensor failed. Having fewer sensors would also mean a more"bumpy" field of data points.
- We could not test our model on real data.
- Our assumptions may make our approximations inaccurate. Such assumptions include the regular shape of the submarine, uniform depth of the ocean within our area of interest, and sound reflection / refraction patterns.
- Our integral approximations may be inaccurate, which would cause our computer-generated noise field to be inaccurate.
- Our model could not calculate the size of the submarine or its depth very accurately. We believe, however, that we have the relative scale on these numbers right, and it is a matter of finding the right corrective factor.
- Our ambient noise field model is just for a single submarine, though incorporating more than one submarine should be a relatively easy project.

Appendix: Derivation of Integrals

To approximate the sound blocked by the submarine, we use two techniques:

- We draw a line l_1 from the observation point, through the center of the submarine, to the bottom of the ocean (see **Figure 7**). We attempt to integrate along that line, on a portion of a spherical surface. That is, let the distance from the point of observation to the point on l_1 be r, consider the sphere of radius r, and consider the portion of the sphere that would project onto the submarine. We take the limits of integration on r to be the center of the submarine and the point where l_1 hits the ocean floor.
- It is very hard to approximate the portion of the sphere that projects onto the submarine. When r is at the center of the submarine, we can approximate the area by projecting the ellipse of the submarine onto the plane perpendicular

to l_1. This reduces the length in some direction by a factor of d/R, where d is the depth of the submarine and $R = \sqrt{x^2 + y^2 + d^2}$ is the distance from the observation point to the center of the submarine. On the other hand, in the direction perpendicular to that direction, the length is not reduced and the area is reduced by d/r.

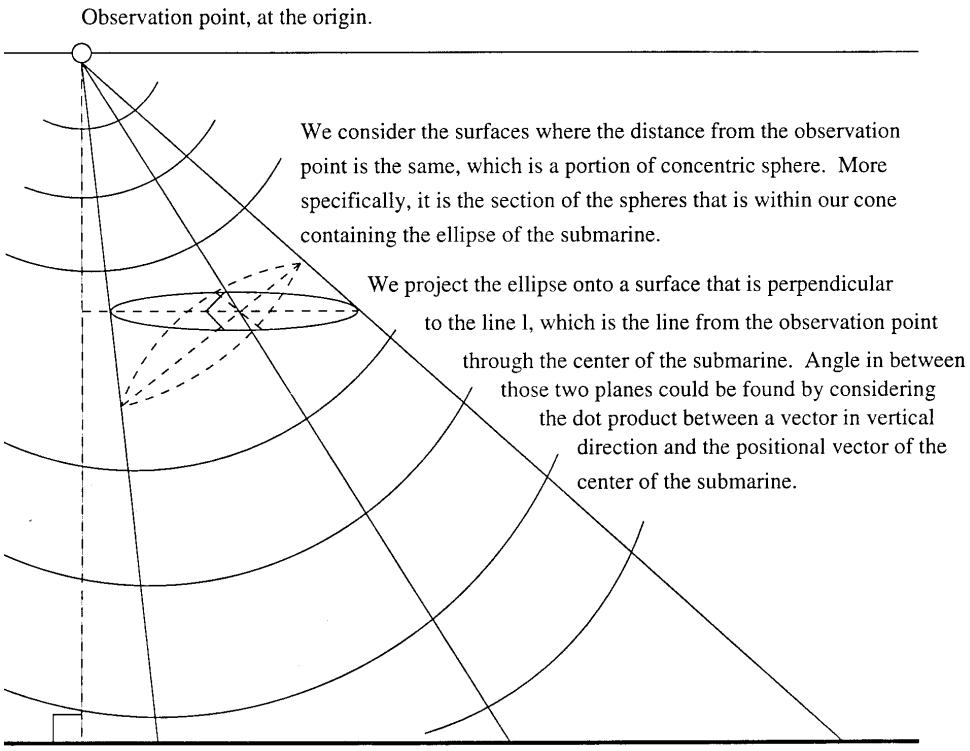

Figure 7. Approximation of the noise dampened by the submarine.

The area of the ellipse is $ab\pi$. We approximate the section of the sphere by a projection of this ellipse, which gives the area of $ab\pi(r/R)^2$, since the dimension in each direction changes by the factor of r/R as we move along l_1. Thus, we have the integral

$$\int_R^{R(D/d)} \frac{ab\pi \left(\frac{r}{R}\right)^2 \left(\frac{d}{R}\right) e^{-\alpha(r-r_0)}\, dr}{r^2} = \int_R^{R(D/d)} \frac{abd\pi}{R^3}\, e^{-\alpha(r-r_0)}\, dr.$$

To calculate the total ambient noise at a point in this scale, we set up a similar calculation. We ignore the noise that might be coming from above the sensor. By doing so, we have

$$\int_0^D 2\pi r^2 (1/r^2) e^{-\alpha(r-r_0)}\, dr + \int_D^\infty (A(r)/r^2) e^{-\alpha(r-r_0)}\, dr,$$

where $A(r)$ is the appropriate area function. We approximate $A(r) = 2\pi rD$, which results in

$$\int_0^D 2\pi e^{-\alpha(r-r_0)}\,dr + \int_D^\infty 2\pi r D(1/r^2)e^{-\alpha(r-r_0)}\,dr.$$

We again cannot integrate this immediately, so now we approximate $1/r$ by $1/D$, since this is the definite upper bound (also, we lost some noise in the approximation, so this will make up for the loss). This results in

$$\begin{aligned}\int_0^D 2\pi e^{-\alpha(r-r_0)}\,dr + \int_D^\infty 2\pi e^{-\alpha(r-r_0)}\,dr &= \int_0^\infty 2\pi e^{-\alpha(r-r_0)}\,dr\\ &= \frac{2\pi}{-\alpha}\left[e^{-\alpha(r-r_0)}\right]_0^\infty\\ &= 2\pi e^{\alpha}/\alpha.\end{aligned}$$

This integral treats the ocean as infinitely deep, which may make sense due to the constant reflections of sound off the bottom and surface.

References

Apel, John R. 1987. *Principles of Ocean Physics*, New York: Academic Press.

Centro Interdisciplinare di Bioacoustica-Universita degli Studi di Pavia. 1996. `http://www.unipv.it/~webcib/cib.html\#surf`

Dera, Jerzy. 1992. *Marine Physics*. Warsaw: Polish Scientific Publishers.

Friedman, Norman. 1984. *Submarine Design and Development*. Annapolis, Maryland: Naval Institute Press.

Hassab, Joseph C. 1989. *Underwater Signal and Data Processing*. Boca Raton, FL: CRC Press.

Keller, Joseph B. 1977. Survey of Wave Propagation and Underwater Acoustics. *Wave Propagation and Underwater Acoustics*, Springer-Verlag: Berlin, 1977.

Munk, Walter H., Peter Worcester, and Carl Wunsch. 1995. *Ocean Acoustic Tomography*. New York: Cambridge University Press.

Pain, H.J. 1983. *The Physics of Vibrations and Waves*. 3rd ed. Chichester, Great Britain: John Wiley & Sons.

Tolstoy, Ivan, and Clarence Clay. 1966. *Ocean Acoustics: Theory and Experiment in Underwater Sound*. New York: McGraw-Hill.

U.S. Department of Commerce. 1996. Marine mammal acoustics. `http://www.pmel.noaa.gov/vents/whales/whales.html`

Detection of a Silent Submarine from Ambient Noise Field Fluctuations

Andrew R. Frey
Joseph R. Gagnon
J. Hunter Tart
Wake Forest University
Winston-Salem, NC 27109

Advisor: Stephen B. Robinson

Summary

We developed a method for detecting intrinsically silent submarines in the ocean by measuring only the fluctuations in the ambient noise field. This method allows us to calculate the position, velocity, and approximate size of a submarine.

Our model relies on measuring the noise field at four different listening stations, with each station composed of four microphones a relatively small distance apart. We calculate an amplitude spectrum of the noise at each microphone using a Fourier analysis and compare this spectrum to the previously measured baseline spectrum for ambient noise. The difference between these spectra is the noise reflected from the submarine.

We use the four microphones at a particular station to measure the gradient of the peak amplitude from the submarine noise spectrum. Because amplitude varies inversely with distance from the submarine, we can compute the submarine's location from the amplitude and the gradient at each listening station. The approximate size, in terms of the radius of a similarly sized sphere, follows from the distance and peak amplitude.

Our comparison of the frequencies of the peak amplitudes of the submarine and ambient noise spectra provides a measure of the Doppler shift at each listening station caused by the submarine's motion. The Doppler shift gives us a component of the submarine's velocity in the direction of each station. We select a basis from among unit vectors in these four directions and convert the submarine's velocity into standard Cartesian components.

We wrote a Fortran program to implement our algorithm. Our simulations show that we can determine position with better than 8% accuracy in each dimension. Size calculations suggest a systematic error of roughly 20–30%. Error in the velocity computations varied for each component with changes in submarine position and in the dominant frequency of the ambient noise, but was within about 30% for a single frequency of 1,000 Hz.

The model could be modified to remove some of our assumptions, such as the absence of currents. Our model uses a minimum number of listening stations, but a larger number would significantly improve the results.

Assumptions

- The speed of sound in the ocean is constant. Though the speed of sound depends on temperature, the range of detection devices is small enough to render speed of sound fluctuations negligible.
- Ambient noise has the same frequency and amplitude everywhere, so reflections from the surface and bottom of the ocean need not be accounted for [Horton 1959].
- All submarines are approximately spherical, are made of steel, and reflect a fraction k of the sound energy incident upon them. Although the surface of the submarine, as a two-dimensional surface curved in three dimensions, is not a simple harmonic oscillator (SHO), an SHO is a reasonable analogy. In this case, the SHO is both forced (by ambient noise) and damped (by the water and by the flexibility of the metal). The steady-state solution for such a forced and damped SHO is vibration at the forcing frequency. The high damping coefficient likely with a submarine indicates that the response of the submarine should be independent of frequency (and Krasil'nikov [1963] lists a single reflectivity for all frequencies). Certainly, at distances large compared to the dimensions of the submarine, a sound wave reflected off it will approximate a spherical wave.
- Sound waves reflected from a submarine which then reflect off the surface or bottom of the ocean have negligible intensity. This is to say that non-ambient noises detected by our microphones can be considered to have reflected directly from a submarine, and not secondarily from the surface or bottom of the ocean.
- The ocean is of homogeneous consistency, so there are no large animals or objects, aside from the submarine, which significantly influence the transmission of sound waves. Furthermore, we assume that only one submarine is present at any given time.
- The submarine, in addition to creating no intrinsic noise, does not by its movement generate any turbulence that affects noise transmission.
- A typical submarine cannot move faster than 20 m/s.
- There is no appreciable current, and the detection stations are at rest with respect to the water.
- We begin by assuming that there is only one frequency (and corresponding amplitude) of ambient noise, and then we generalize to multiple frequencies.

Description of the Model

Our detection scheme consists of four clusters of microphones in a pyramidal orientation. We require four non-coplanar clusters in order to determine with certainty the submarine's direction of travel in case the submarine is located in the plane of three of the clusters. Four clusters is a minimum; see **Analysis of Error and Sensitivity** for advantages and disadvantages of having only four clusters. We define a Cartesian coordinate system with the xy-plane parallel to the surface of the ocean and the positive z-axis pointing up. We place one microphone cluster at the origin and at the points $(d, 0, 0)$, $(0, d, 0)$, and $(0, 0, d)$. (See **Figure 1**.)

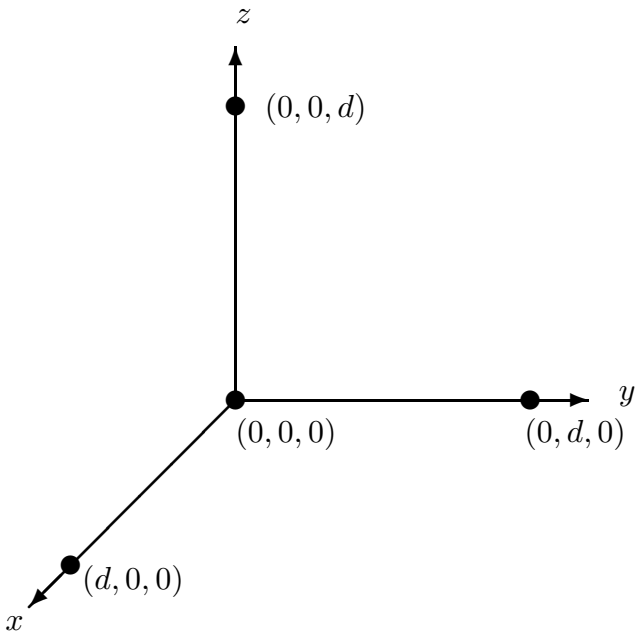

Figure 1. Array of microphone clusters.

Because submarines typically do not descend below 1,500 m, we place the origin of our system 1,000 m below the surface of the water, so that one microphone cluster is $(1{,}000 - d)$ m below the surface and the other three clusters are 1,000 m down. We chose to let d = 500 m so that the detection clusters are well-spaced throughout the potential depth-range of a submarine. We envision the detection clusters as either buoyant anchored rigs or as pods at the end of lines dropped from the surface (e.g., suspended from ships). However, the analysis is not affected by the location of the origin or the orientation of the coordinate axes, as long as the entire array is sufficiently submerged.

Each of the four clusters (listening stations) in turn consists of four microphones, one at the precise location we gave for the cluster and the other three a small distance δ away, one each in each of the coordinate directions. (See **Figure 2**.)

We first measure the ambient noise waveform when no submarines are present, so that we can determine the ambient noise's frequencies and associ-

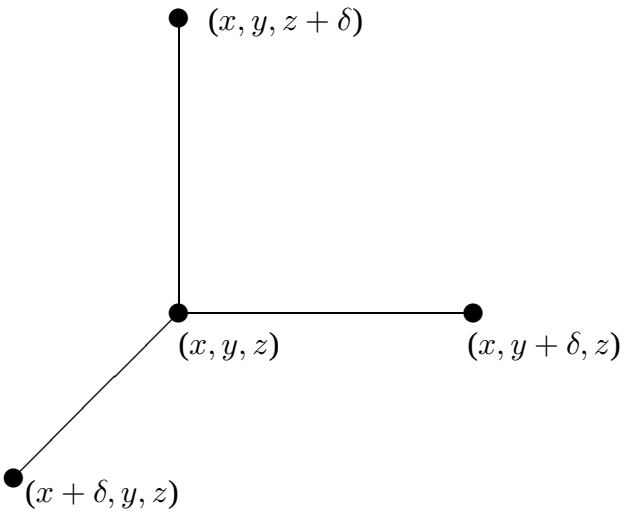

Figure 2. Microphone arrangement within each detection station.

ated amplitudes (at first, only one frequency was present). We then measure the sound at each microphone location for a short period of time, perform a Fourier analysis on the resulting wave pattern to determine the frequencies present and their respective amplitudes, and use these data to figure out the location, speed, direction of travel, and size of any submarines.

Data Required for Calculations

We seek the position of a submarine relative to our coordinate system, its velocity vector, and its size. Because we assume that a submarine can be treated as a sphere, its size is just its radius R, and its position can be described by its center. Hence, a complete solution to our detection problem is comprised of radius R, position coordinates (x, y, z), and velocity vector $\vec{v} = (v_x, v_y, v_z)$.

The data available to calculate these quantities consist of the frequencies and their respective amplitudes received by our array of microphones. We list the constants and variables that are required for our calculations:

f = the frequency of the ambient noise. We chose 1,000 Hz, a frequency well within the range of real oceanic ambient noise, for our single-frequency simulations.

I_0 = the intensity of the ambient noise. A reasonable intensity at the frequency 1,000 Hz is 5.4457×10^{-10} Pa^2 [Munk et al. 1995, 179].

A_0 = the amplitude of the ambient noise. Amplitude squared is proportional to intensity. Because the proportionality constant has already been taken into account in calculating I_0, we get $A_0 = \sqrt{I_0} = 2.3336 \times 10^{-5}$ Pa.

k = the percentage of the sound energy (or intensity) reflected by the surface of the submarine. We let $k = 0.86$ [Krasil'nikov 1963, 172]. Since amplitude is proportional to the square root of intensity, the amplitude of sound waves immediately after reflecting from the submarine surface will be $\sqrt{k}A_0 = 0.9274A_0$.

$A(s) =$ the amplitude of sound waves reflected from the submarine's surface at distance s from the center of the (spherical) submarine. Note that

$$A(s) = \frac{\sqrt{k}A_0R}{s}.$$

This formula is a well-known consequence of conservation of energy.

Step One: Detecting the Submarine

The raw data that we receive through each of the microphones consist of a waveform recorded over a short time interval (see **Figure 3** for a sample plot of multiple frequency noise).

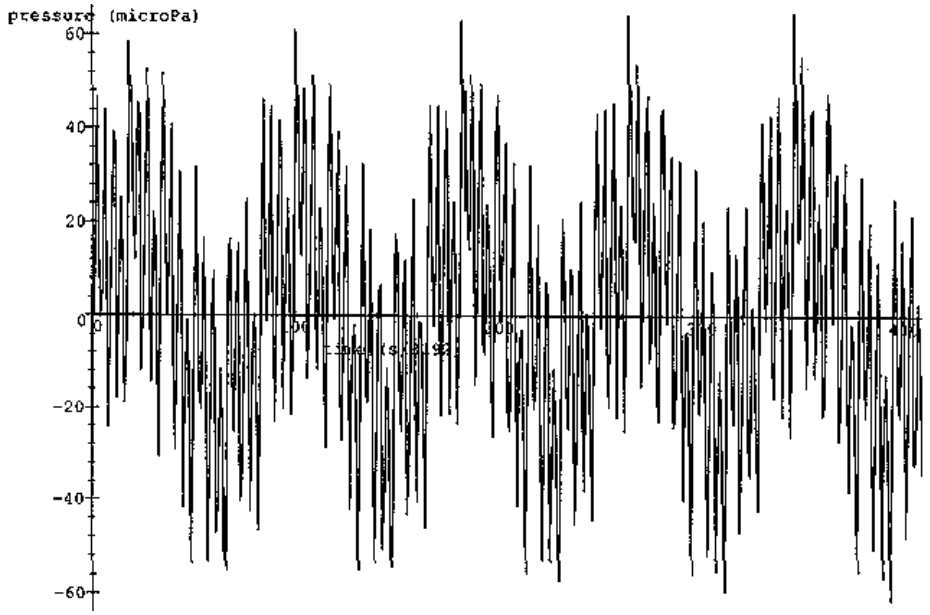

Figure 3. Ambient noise with five frequencies, showing pressure as a function of time.

To convert these data into useful frequency and amplitude figures, we use a fast Fourier transform, which isolates the particular frequencies in a given signal. The Fourier transforms that we used were the sine and cosine transforms provided with Press et al. [1986]. Once this computation is completed, we have amplitude values $A_{i,j}$ associated with the frequency f (with each frequency if there is more than one) of recorded noise for each microphone, where i is the station number and j is the microphone number within that station.

Our algorithm subtracts the ambient noise amplitude spectrum (determined as discussed previously) from the new amplitude spectrum recorded at each microphone. If all the differences are equal to zero, the current noise in the ocean is only the homogeneous ambient noise, so there must not be a submarine within a detectable distance. The microphones then collect a new set of data, and we begin the process again.

If any differences are nonzero, we must determine whether they are caused by a change in the ambient noise or by the presence of a submarine. If the differences in the spectra are due to a change in the ambient noise, all of the microphones should have recorded the same data (by the definition of ambient noise). However, any changes caused by a submarine should vary from station to station and from microphone to microphone because of varying positions of the microphones relative to the submarine. Therefore, we compare the difference spectrum (the amplitude spectrum from the microphone minus the ambient noise spectrum) of the first microphone of each station to the difference spectra of the first microphones of the other stations. We could compare all the difference spectra for all the microphones, but since the greatest variation is from station to station, we need to compare only among stations.

If there is no variation among the difference spectra among the stations, the algorithm must take the change in the ambient noise into account. It replaces the ambient noise spectrum by the new ambient noise spectrum.

If there is a difference from station to station, our algorithm has detected a submarine! In this case, the difference spectra give us the amplitude values $A_{i,j}$ of the sound reflected by the submarine for each microphone. Now the algorithm finds the frequency with the greatest amplitude for each microphone. Because all frequencies of noise reflect off the submarine with the same proportionality constant $\sqrt{k}$, this frequency and the corresponding peak amplitude must be the reflection of the frequency with the peak amplitude in the ambient noise. We consider only these peak amplitudes and their corresponding frequencies throughout the remainder of the algorithm, whether the ambient noise is composed of one frequency or many (see **Figure 4**).

Step Two: The Submarine's Position

Our strategy now is to compute $\vec{\nabla} A_i$ for each of the four stations by approximating the derivatives of $A_{i,1} = A_i(x, y, z)$, where (x, y, z) are the coordinates of microphone $(i, 1)$, in each of the x, y, and z directions. This explains our rationale for having four microphones at each of the four stations, since we can compute

$$\frac{A_i(x+\delta, y, z) - A_i(x, y, z)}{\delta} \approx \frac{\partial A_i(x, y, z)}{\partial x},$$

as well as the derivatives in the y and z directions, to find $\vec{\nabla} A_i$.

We note that $|\vec{\nabla} A_i|$ is the absolute value of the derivative of A_i with respect to s, the distance from the center of the submarine. Since we now have values

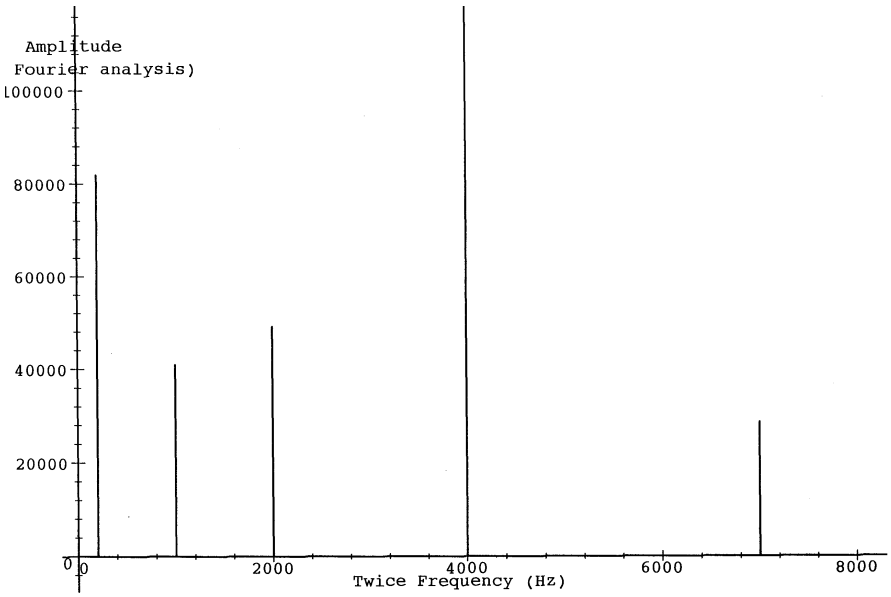

Figure 4. Amplitude spectrum of ambient noise, showing amplitude as a function of twice the frequency (in Hz).

for A_i and for $|\vec{\nabla} A_i|$, and since

$$A_i(s) = \frac{\sqrt{k} A_0 R}{s}, \qquad |\vec{\nabla} A_i| = \frac{\sqrt{k} A_0 R}{s^2}, \tag{1}$$

we see that $A_i/|\vec{\nabla} A_i| = s$. At this point, because we know the vector $\vec{\nabla} A_i$ that points to the submarine and the distance s to its center, we know the position of the submarine relative to detection station i. In fact, the coordinates of the submarine's center are given by

$$(a, b, c) = s \frac{\vec{\nabla} A_i}{|\vec{\nabla} A_i|} + (x, y, z).$$

While this calculation for one of the i stations is sufficient to get coordinates for the submarine's position, our four stations allow us to compute these coordinates four different ways. Since there will be some random error in each computation, averaging the four different points provides a better approximation of the submarine's position.

Step Three: The Submarine's Size

With values for the amplitude A_i at station i and the distance s of station i from the submarine, computing the radius of the submarine is easy. From **(1)**, we have

$$R = \frac{A_i s}{\sqrt{k} A_0}. \tag{2}$$

We obtain a better estimate by averaging the four values of R.

Step Four: The Submarine's Velocity

To calculate the velocity, we use the frequency f of sound reflected from the submarine's surface. Since we know the frequency of the ambient noise and thus the frequency shift between ambient and reflected sound, we can solve the equation describing the Doppler effect for the speed of the submarine in the direction of a particular detection station. The general Doppler effect for sound [Krane et al. 1992] is

$$f_o = f_s \left(\frac{c - v_o}{c + v_s} \right),$$

where f_o is the frequency received by the observer, f_s is the frequency of the source (the ambient frequency), v_o and v_s are the components of the velocities of the observer and the source along the line between them, and c is the speed of sound, in this case in water. Because our detection stations are stationary, we have $v_o = 0$. Also, we let v_s be positive if the submarine is moving away from station. Solving for v_s, we get

$$v_s = \left(\frac{f_s}{f_o} - 1 \right) c.$$

Once this v_s has been calculated for a particular listening station (let's rename it v_i), we can express this component of the submarine's velocity in terms of a vector. Since $\hat{u}_i = \vec{\nabla} A_i / |\vec{\nabla} A_i|$ is just the unit vector pointing from detection station i to the center of the submarine, the vector $\vec{v}_i = v_s \hat{u}_i$ is the component of the submarine's velocity in the direction of station i.

Note that we need velocity components in only three linearly independent directions to compute the velocity vector. However, we have four potential basis vectors, the four $\hat{u}_i$. To determine which set of three vectors is the most useful basis for our analysis, we consider the four matrices formed by taking combinations of the $\hat{u}_i$ as column vectors. First, we know that any coplanar set of three vectors will not form a basis at all, so the matrix composed of them will be singular. By perturbation, any set of three vectors that are almost coplanar will form an almost singular matrix; and, in the real world of measurement and computation errors, such a basis would not be useful. Therefore, we choose the set of vectors $\{\hat{u}_i, \hat{u}_j, \hat{u}_k\}$ whose matrix has the largest determinant as the basis likely to be most useful to our algorithm.

Now, the vector (v_i, v_j, v_k) is a coordinate vector in terms of our chosen basis. We want to change our basis to the standard $\{\vec{x}, \vec{y}, \vec{z}\}$ basis to determine the velocity vector of the submarine with respect to our coordinate system. This change of basis can be performed by a simple matrix multiplication,

$$\left[\begin{array}{c|c|c} & & \\ \hat{u}_i & \hat{u}_j & \hat{u}_k \\ & & \end{array}\right] \begin{bmatrix} v_i \\ v_j \\ v_k \end{bmatrix} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \vec{v}, \tag{3}$$

resulting in the velocity vector $\vec{v}$.

Extensions of the Model

Although most of our simulations were carried out with only one frequency in the ambient noise, we find a broad range of frequencies in the ambient noise of the ocean. Fortunately, the algorithm can process multiple frequencies because it uses only the peak amplitudes and corresponding frequencies. This technique reduces the scenario to a single-frequency problem.

The algorithm, as described above and encoded in the computer simulation, can adjust to fluctuations over time in the ambient noise field. However, if both the appearance of a submarine and a significant change in the ambient noise field coincide, the algorithm assumes that all of the difference spectra represent noise due to the presence of the submarine. This effect could cause some error.

We also can track a single submarine over a period of time, because the algorithm, in the form of a computer program, runs quickly enough to provide frequent data on the position and velocity of the submarine.

The presence of two or more submarines in the observation region presents more of a problem for the algorithm because it would not recognize the presence of the submarine with the smaller effect on the ambient noise field. However, the algorithm could be modified to detect the presence of a single submarine, calculate its effects on the ambient noise field, and compare the recorded data with those effects.

Our model could be extended to eliminate the assumptions of no current and stationary listening stations. The general Doppler equation already provides for a moving observer, so moving listening stations would be relatively easy to handle. Also, a constant surrounding current could be integrated into the Doppler equation, though this computation would be a bit more complicated.

Simulation Results

We wrote a Fortran program to simulate the implementation of our algorithm. The program uses only the parameter k, the positions of all the microphones, a waveform representing the ambient noise, and a waveform for each of the microphones representing the noise field with the presence of a submarine.

To run the simulation program, we needed to create sound data for the microphones to receive. We used another Fortran program to produce a discretized version of the soundwaves, with more than 8,000 samples per second. We first created a data set with only the ambient noise present, as a benchmark, and then added in the additional sound caused by a submarine of a particular size with specific position and velocity vectors. This data-generation scheme provided us with an easy check on the accuracy of our simulation program.

At first, we created an ambient noise data set with one fixed frequency and amplitude. We then created data files for three different selections of submarine radius, position, and velocity. We ran our simulation program for each of these

three data sets in the presence of a submarine. The results of these simulations are provided in **Table 1**. These results show that our simulation was relatively successful in picking out the position and the velocity of the submarine, though appreciable error was present. While our simulation apparently did not do a particularly good job of calculating the radius of the submarine, the percentage error for the three simulations was fairly consistent, suggesting that this error may be systematic and thus correctable.

Since Fourier analysis can result in an apparent smearing of a particular frequency over a frequency range, the amplitude we calculate for this frequency may be consistently smaller than it should be. This seems like the sort of problem that could be corrected through more sophisticated, more powerful Fourier analysis, or at least through the inclusion of an amplitude correction factor.

Table 1.

Computer program output of radius, position coordinates, and velocity coordinates for three different submarine data sets, with one ambient frequency.

Simulation		R (m)	x (m)	y (m)	z (m)	v_x (m/s)	v_y (m/s)	v_z (m/s)
1	Input	13.8	−3,000	−2,000	300	15	10	5
	Output	9.73	−2,765	−1,843	281	14.9	10.5	4.7
2	Input	5	200	1,000	−500	2	10	2
	Output	4.02	190	953	−464	1.3	9.5	1.6
3	Input	10	2,000	2,000	−500	5	5	5
	Output	7.95	1,871	1,871	−468	4.8	4.8	3.9

While the results of our simulations are encouraging, we point out that our computer program is merely one realization of our general mathematical algorithm for finding a submarine's radius, position, and velocity. The results would be improved by providing more complete and accurate sound data (corresponding to better microphones) or using a more accurate and perhaps more appropriate fast Fourier transform algorithm.

We also ran the same simulations using ambient noise with multiple frequencies. The results are in **Table 2**.

Analysis of Error and Sensitivity

Table 2 shows the differences between actual and calculated position, radius, and velocity of a submarine. The error in the position coordinates most likely arises due to the dependence of our algorithm on measured amplitude figures and the amplitude derivative that we compute using them, which appear not to be accurate. At least part of this error may be an artifact of our need to use discrete data. Furthermore, our numerical calculation of $\vec{\nabla} A_i$ for station i introduces more error, since it is based on a finite, and in fact quite large,

Table 2.
Radius, position coordinates, and velocity coordinates for three different simulated submarine data sets with ambient noise of five different frequencies.

Simulation		R (m)	x (m)	y (m)	z (m)	v_x (m/s)	v_y (m/s)	v_z (m/s)
1	Input	13.8	−3,000	−2,000	300	15	10	5
	Output	8.48	−2,826	−1,886	287	14.6	11.0	52.8
2	Input	5	200	1,000	−500	2	10	2
	Output	4.19	192	939	−454	−0.3	6.1	−4.2
3	Input	10	2,000	2,000	−500	5	5	5
	Output	7.72	1,839	1,839	−448	5.0	5.0	−1.5
Frequencies (Hz):				100	500	1,000	2,000	3,500
Amplitudes (μPa):				30	10	20	12	7

distance δ between microphones. This error is difficult to eliminate, since increasing δ creates a worse approximation of a derivative, while decreasing δ to very small distances requires unreasonably sensitive microphones to perceive tiny differences in amplitude. Because of this practical consideration, we are forced to accept some error in our calculation of a submarine's position.

We have a nice measure of position error, since our algorithm computes the position vector four times, each time using only data from the four microphones at one listening station. For the same reason, we also have a measure of the error in the radius measurements (neglecting the apparent systematic error). The relative errors (standard deviation divided by mean) listed in **Table 3** are all quite small.

Table 3.
Relative error for radius and position calculations.

Simulation	R	x-coordinate	y-coordinate	z-coordinate
1	.043	.034	.014	.023
2	.078	.009	.009	.023

Error in our radius calculation is largely inherited from the problem with amplitudes discussed above. From **(2)**, we see that the radius is determined by the measured values of A_i and s_i. Thus, if the A_i calculated by Fourier analysis is too small, the calculated radius value will similarly be too small. We suspect that this fact is the cause of the fairly systematic error in our values of R shown in **Table 1**.

Velocity error arises primarily because of two factors:

- error in our position calculation, since the velocity computation relies upon the submarine's location, and

Table 4.
Relative error of calculations for two different simulations for different values of δ, the distance separating microphones in each station lattice.

δ (m)	Simulation	R (m)	x (m)	y (m)	z (m)	$\lvert\vec{v}\rvert$ (m/s)
5	1	−.195	−.041	−.046	−.068	−.058
	2	−.295	−.078	−.080	−.054	.007
10	1	−.192	−.049	−.046	−.070	−.058
	2	−.294	−.078	−.079	−.061	.007
15	1	−.197	−.057	−.047	−.072	−.057
	2	−.278	−.078	−.077	−.068	.007
20	1	−.198	−.067	−.047	−.075	−.056
	2	−.272	−.077	−.076	−.075	.007

- error in the observed frequency f of noise reflected from the submarine as determined through the Fourier analysis.

The error in the observed frequency becomes particularly important when low frequencies predominate the ambient noise. The error is amplified because a small absolute error in the frequency measurement becomes a large relative error in the low frequency range, and the measurement error is a constant absolute error for all frequencies. Note the errors in the velocity components in **Table 4**, the results of a simulation in which a frequency of 100 Hz dominates the ambient noise.

We performed some additional simulations in which we permuted the parameter δ, the distance between microphones within stations, in order to determine the sensitivity of our calculations to this parameter. The relative errors between our calculated values and the actual values are shown in **Table 2**. From this table, it appears that our model is not particularly sensitive to fluctuations in δ.

Using the fast sine and cosine transforms limits our model; their use assumes that the data have an initial phase of zero, so we do not take into account phase shift in the reflected noise due to travel time from the submarine to the stations.

Finally, we note that some of the error in our results arose because of the small number of microphones we arranged in our stations. We designed the model with some frugality, using only four stations because four is the minimum number needed to guarantee that we can pinpoint a submarine's velocity (three fail if the submarine is in the same plane as all three stations). Our calculations would have benefitted from some redundancy in our measurements, for instance, using eight stations arranged as vertices of a cube. However, our minimal scheme is cheaper, requires less superstructure, and provides for simpler computer calculations than would a more redundant arrangement.

Conclusions

This model successfully detects a silent submarine using only distortions of the ambient noise field as data. It accomplishes this task with a small number of microphones (16) arranged in a lattice structure beneath the surface of the ocean, and it provides relatively accurate data for a range of submarine sizes, positions, and velocities. The model may lack realism in that it requires such assumptions as a homogeneous ocean, nearly spherical submarines, and isolated ambient noise frequencies. However, the first two of these assumptions have fairly solid physical bases in most normal circumstances, and our algorithm does provide a solid foundation that can be extended to take into account such complicating factors as a continuous frequency distribution.

Acknowledgments

The authors would like to thank Dr. Edward Allen and Dr. Stephen Robinson for their assistance in preparation for the MCM.

References

Horton, J.W. 1959. *Fundamentals of Sonar*. Annapolis, MD: United States Naval Institute.

Krane, K.S., D. Halliday, and R. Resnick. 1992. *Physics*. 5th ed. New York: John Wiley & Sons.

Koffman, Elliot B., and Frank L. Friedman. 1993. *Fortran with Engineering Applications*. 5th ed. New York: Addison-Wesley.

Krasil'nikov, V.A. 1963. *Sound and Ultrasound Waves in Air, Water and Solid Bodies*. 3rd ed. Jerusalem: Israel Program for Scientific Translations.

Munk, Walter H., Peter Worcester and Carl Wunsch. 1995. *Ocean Acoustic Tomography*. New York: Cambridge University Press.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1986. *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press.

Imaging Underwater Objects with Ambient Noise

Aron C. Atkins
Henry A. Fink
Jeffrey D. Spaleta
{ atkins, haf, spaletaj } @wpi.edu
Worcester Polytechnic Institute
Worcester, MA 01609

Advisor: Arthur C. Heinricher

Introduction

We present a method of locating and tracking undersea objects of various sizes. We list assumptions that help define our approach, together with requirements that must be met for an object to be detectable. We construct a detection scheme for two dimensions, which we generalize to three dimensions.

In constructing our models, we make several design decisions which include time limitations, range, and space and resource requirements. Many of these choices are desirable for confidently detecting objects.

Assumptions

We make several assumptions regarding the type of environment to which we will be listening.

- We consider an extreme case in which our object must be a near-perfect acoustic reflector; such an object resembles a submarine, which tends to absorb very little sound. (If submarines absorbed a predominant amount of sound, sonar systems would be ineffective in tracking their movement.)

  By requiring our target object to be a near-perfect acoustic reflector, an individual scanning frequency can be chosen from a range of transmittable frequencies in ocean water. This range is given as between 5 and 50 kHz [Buckingham et al. 1992]. We can limit our monitoring process to a single frequency because a near-perfect acoustic reflector will give the same characteristic response at all frequencies [Pedrotti and Pedrotti 1993]. Having information on more than one frequency would produce redundant information. It is better to pick a single frequency that is strongly represented in the ambient noise field than to monitor several different frequencies.

- We require directionality in the ambient field. Without an ambient noise bias in some direction, there will not be a detectable difference between reflected and background noise. Experimental research shows there is a directional bias in the ambient noise field near the ocean's surface and floor [Stephens 1970, 124–125; Urick 1983, 227]. For acoustic absorbers, ambient field directionality would be unnecessary; absorbers would always be detected as "holes" in the ambient noise field intensity regardless of field bias.
- We also require that the ambient noise field remain relatively stable while we are searching. Frequent changes in the field over short time intervals make it nearly impossible to find an object.
- We require that at least one of the following two conditions holds to detect confidently the presence of objects in our scan region:
  - The object being scanned is in motion relative to our scan location.
  - We perform a background scan prior to the target object entering into our scan region, reliable for the ocean conditions during the search time.
- We assume a minimum target size of approximate 10 m, the average width of submarines in the U.S. Navy. Submarines of other nations are of comparable size.
- Our sensing equipment consists of multiple parabolic reflectors, commonly used for focusing weak signals to one focal point, where we place a hydrophone, a device designed to measure acoustic intensity under water [Geil 1992]. This type of sensor has been used previously in related experiments [Buckingham et al. 1992] with promising results. Additionally, we assume that our equipment can detect angle differences to 0.1°.

We would like a system capable of detecting objects as far away as possible for relatively small sensor size. The relation between resolving distance and sensor diameter is given in **Appendix A**.

Object Detection

Prior Experimental Results

Our procedure relies on the results of an experiment performed by Buckingham et al. [1992] in which neoprene-coated boards submerged near a pier were detected using the ambient noise in the water. In their experiment, three targets each 0.9 m high and 0.77 m wide were placed 7 m from a reflector. The detector was a hydrophone located at the focal point of a parabolic reflector of diameter 1.22 m with a neoprene rubber surface. The back of the hydrophone was shielded to prevent detecting noise from that side. The boards first were

turned edge-on to the reflector and the noise level was recorded over a frequency range of 5–50 kHz. When the boards were rotated so that they were face-on to the parabolic dish, the noise spectrum was recorded again. By subtracting the two spectra, the authors found an average intensity difference of 4 dB. They reasoned that this difference was due to a directionality in the noise emanating from the nearby pier.

Our procedure for detecting objects in the ocean using the ambient noise field extends this experiment. We first consider locating and tracking an object in the ocean in two dimensions, using two of these reflectors.

Object Detection in Two Dimensions

We can uniquely determine any point in the plane if we know its distance from two known points. These known points can be our parabolic reflectors. (We assume that the object moving through the ocean is never collinear with our two reflectors; if there is reason to believe this may happen, we can add a noncollinear third reflector.)

To simplify visualizing the problem, we consider the special case when both of the parabolic reflectors are located at a fixed depth near the shore line; this does not make the problem any less interesting and is likely to be the case in many practical implementations (see **Figure 1**). The process of detecting a distant object begins with both reflectors sweeping through 180° while monitoring a fixed frequency. If the background noise were constant for all angles, then a plot of noise vs. angle would yield a horizontal line. However, due to directional biases in the ambient noise field, it is likely that this noise will be a function of the angle at which the reflector is directed. For this reason, it would be useful to have a background scan (a scan with no object present) for comparison with later scans.

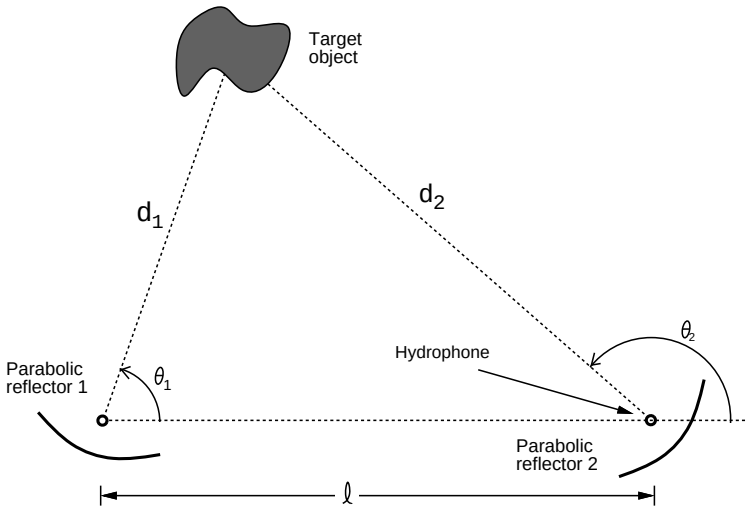

Figure 1. Point location in two dimensions.

Assuming that a background scan has been made, it is a simple matter for both reflectors to scan through angles and record the ambient noise level when searching for an object. By subtracting away the two plots of noise level in dB vs. the angle each reflector has swept from its initial position, a characteristic disturbance pattern should appear over a small angle range, indicating the presence of an object. The angle at the center of this disturbance represents the approximate viewing angle to the center of the object. Knowing this angle measurement from each reflector and the distance between the reflectors, it is a simple matter to triangulate the position of the object. A method to perform this triangulation is described in **Appendix C**.

The disturbance pattern in the intensity plot gives more information than just the viewing angle at which the center of the object is located. Since the angle is being swept out, the broader the disturbance pattern, the larger the object. By measuring the angular width of the disturbance region, we can calculate an approximate size of the object. Once the distance to the object has been found, the triangulation formulas can be applied using the extreme angular values of the signal disturbance to find the size of the object. The size of the object is approximated from the difference between the two extreme locations calculated from triangulation.

Once the reflectors have found the object initially, they can track it by scanning through a small interval centered around the initial position. This interval can come from an estimate of the velocity of the object.

One possibility that can occur while scanning is that the object could be out of range of one of the two reflectors. The position of the object can still be approximated, since the ranges of both of the reflectors are known. The object would lie in crescent region inside the circle of detection of one reflector and outside the circle of detection of the second. This gives a rough location of the object that may be good enough in practical applications. A potential intersection of two sensor regions is shown in **Figure 2**.

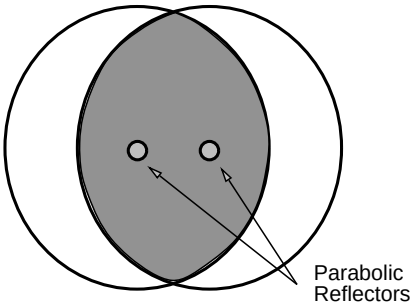

Figure 2. Intersection of two scan regions.

Detection without a Background Scan

If a background scan is not available, or if the background noise has changed sufficiently (due to a storm or other disturbances in the ocean), then an object

can still be detected and tracked. Again, for simplicity, we consider the case where the reflectors lie along a shore. The process begins as before, with each reflector sweeping through $180°$ and recording the ambient noise level. This scan will be a combination of the background noise with the influence of an object, if present. Although a variance may be present in this scan, it cannot be ruled out that its presence is merely an artifact of the normal ambient noise at that angle. To determine whether there is a true disturbance, a second full scan is needed a short time later.

Once a second scan has been made, the two can be subtracted as before to yield the difference in noise level at each angle. Assuming that the target object has moved during the two scan periods, there will be a shift in the disturbance on the second noise plot. The subtraction will yield a pattern. The midpoint in the subtracted disturbance region gives an approximation to the angle at which the object was viewed halfway during the time interval. If the two scans are taken close enough together to narrow down the time, yet far enough apart to detect the shift in the object's position, then the position of the object can be determined with nearly the same accuracy as when a background scan is used.

The drawback to this approach is that it requires each reflector, at least initially, to scan through a full $180°$ ($360°$ when not near the shore) at least twice to detect the object. Once the object is found, two passes over a much smaller angular range can be used to track it as before. A second shortcoming is that the method can detect only moving objects. To detect a stationary object, a background scan is required.

Object Detection in Three Dimensions

Three-dimensional object detection requires substantially more time than the two-dimensional case (see **Figure 3**). Each reflector has to sweep out the entire volume of space, a formidable task. The position of an object can be determined from the angles to it from two known points, just as in the two-dimensional case. The direction of each reflector is described by two angles, using their spherical coordinate representation, or alternatively by vectors. The case of both reflectors and the object being collinear can be resolved by using a third noncollinear reflector.

Ideal Detection of Objects

To cover a large region effectively, we use an array of reflectors positioned so that each point lies within the range of at least two reflectors. An ideal object-location scheme would use not two but three (noncollinear) parabolic reflectors to detect objects in the scan region. An object entering the region would be detected immediately by at least one reflector. Then the remaining reflectors would scan along the line of sight of the first reflector until they detect the object. Next, the three-dimensional point-location scheme discussed

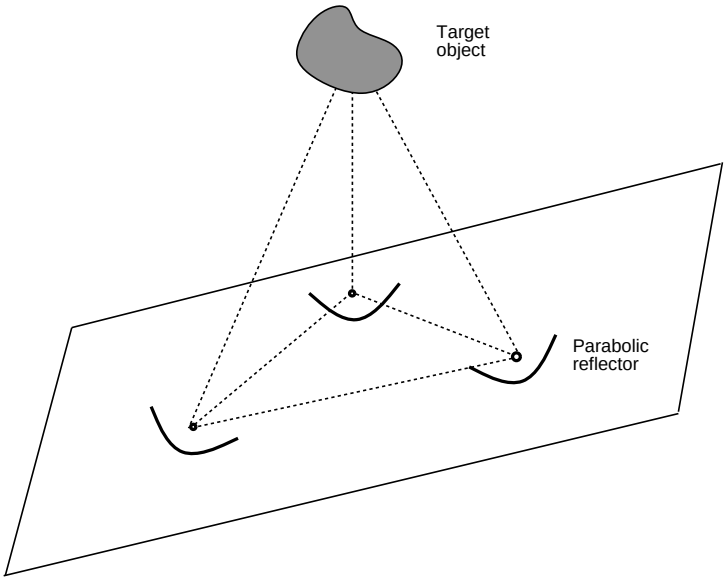

Figure 3. Point location in three dimensions.

in **Appendix D** determines the object's position. With a second scan a short time later, the two positions determine the heading and velocity.

Detection with Reflector Arrays

A major problem with our scheme is that it does not produce a complete sweep of the full area in a reasonable time. We can compensate for this by adding additional reflectors or by changing the shape of the reflectors. The array configurations we consider are a trough system (**Figure 4**), a linear system (**Figure 5**), and a parabolic torus configuration (**Figure 6**).

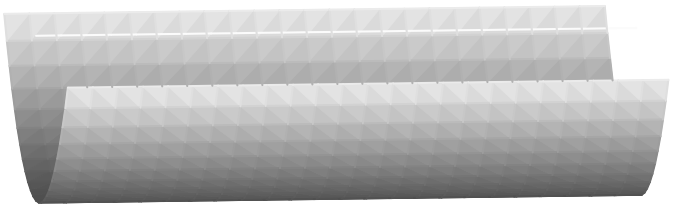

Figure 4. Parabolic trough.

The Parabolic Trough

Parabolic troughs were used as solar energy collectors as early as the late 1800s, to focus the sun's energy to heat water-filled pipes. Many solar energy plants still use parabolic troughs because of their advantages over flat mirrors or dishes [Smith 1996].

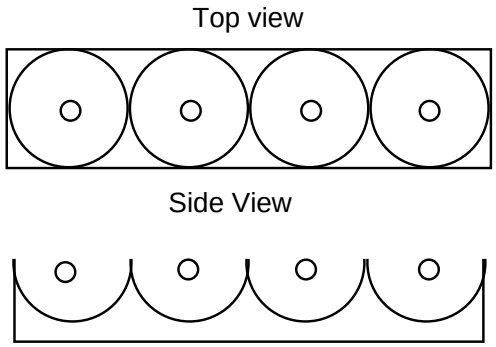

Figure 5. Linear array configuration.

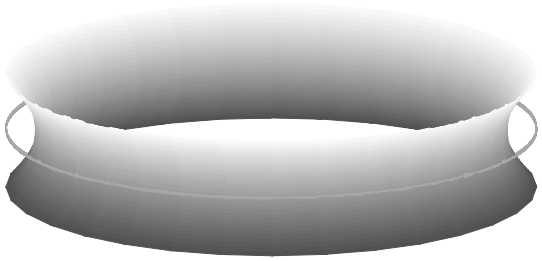

Figure 6. Parabolic torus.

A parabolic trough is shaped something like a soup can cut in half lengthwise. Unlike a parabolic dish, which only focuses rays to a specific focal point, a parabolic trough focuses rays into a line. A parabolic trough is simpler and cheaper to construct than a parabolic dish and can scan a much wider field. An array might be constructed of three of these parabolic troughs oriented at different angles. As each of them rotates, they sweep out a plane. The intersection of these three planes gives the location of the object. It is unlikely that an object would evade detection if the troughs are long enough and are able to rotate through some angle to compensate for their finite length.

A drawback to using a parabolic trough is its inefficiency, since it focuses rays to a line rather than a point. This would lead to lower detected intensities when an object is encountered.

The Linear Array

The same effect of using a trough can be produced using a linear string of dishes. If all the reflectors point in the same direction, then this would act as a trough with a finite number of focus points. The disadvantage is the cost of construction and the need for careful alignment of each dish.

The Parabolic Torus

A parabolic trough still does not sweep out an infinite plane through its rotations. One reflector shape that can focus the ambient noise in an entire

plane into a detectable area is a parabolic torus, which looks like a tire rim. Each cross section is a parabola, so the parabolic torus focuses energy a constant distance away from the vertex. For this torus, the focus becomes a ring around the reflecting surface. In theory, this provides a tremendous advantage over either troughs or dishes. A trough is limited to scanning areas only as wide as its length. The parabolic torus scans an entire plane during one instant and, when it is swept through $180°$ perpendicular to its axis, scans the entire volume around it. Thus, it is impossible to "hide" from this sensor, as is possible with the previous two.

A weakness of this reflector construction is that even less energy is focused from an object in the surroundings than from a parabolic trough. Unless the object creates a strong disturbance pattern in the ambient noise field, the object may not be seen with this reflector. Additionally, we can expect the range of this system to be less than for a system using only parabolic dishes.

A detection system could then employ three of these reflectors mounted at different angles. Each would sweep through their surroundings. After a full revolution, they would have two scans and corresponding angles for each scan.

A more accurate procedure might use two parabolic tori and a parabolic dish. The tori could be used to find the line along which the object lies and the dish could then sweep out along that line with greater accuracy to find the object. A drawback of the tori configuration is that we do not know a way of calculating the resolving power, so we do not know the range.

Strengths and Limitations

We discuss the necessity of some of our assumptions and what happens if they are relaxed or removed. We also reveal some intrinsic limitations due to equipment and environment.

- We can relax our assumption that the target object is a near-perfect reflector. Strong sound-absorbing objects produce intensity profiles with negative difference regions compared to the background, in contrast to the positive regions for near-perfect reflectors. The similarities suggest that our assumption is not needed. If we allow frequency-dependent reflectivity, we can determine the acoustic color of an object by scanning over a range of frequencies.

- If the assumption requiring directionality is removed, detection becomes difficult for all objects except very strong acoustic absorbers. Absorbers would always show up as "holes" in the background ambient noise field. Since interesting strong acoustic absorbers are rare, we recommend that this constraint remain in any implementation.

- There is no need for a background reference scan if we use a multiple scanning technique on moving objects. If a reliable background scan can be

obtained, it should be used for validation; for detection of stationary objects, a confident background is still necessary.

- Ambient noise near the surface of the ocean is unstable because of highly variant surface conditions, so stationary objects would be harder to detect. This is contrasted by the seismically directed field at the ocean floor, which is constant over periods as long as seasons [Stephens 1970].
- We are uncertain which reflector type is the best in practical situations. We suggest experimentation like that in Buckingham et al. [1992].
- The angle-measuring precision of our tracking equipment is crucial to our ability to precisely find objects. If we alter the precision of our equipment even slightly, we might incorrectly locate the object, especially when it is far away.
- We must limit the total size of our reflector. Large reflectors are impractical because moving them underwater poses a significant problem and because one placed on a ship must not restrict the ship's movement.
- One last limitation that dominates any design is the acoustic energy absorption of sea water. This attenuation of energy effectively cuts our viewing distance to 1 km if we are using a frequency of 40 kHz [Stephens 1970, 12], which is disappointing.

Conclusions

Our findings lead us to believe that this technology is a viable detection system for most circumstances. Theoretical results for distant objects, or for stationary objects in unstable ambient fields, are not so positive; and this type of system would be not advisable for these conditions. We are disappointed in the limited maximum viewing range in ocean water. The range of 1 km greatly limits the ability to detect approaching submarines.

This is, however, a great way to detect stationary submarines at deeper levels, which are hard to detect using traditional sonar techniques [Stephens 1970]. A scan depth of 1 km is a reasonable maximum for submarines.

Another potential application is the monitoring of otherwise undetectable disturbances in the ocean, including changes in seismic conditions and the behavior of marine life. This application presents a nonmilitary use for equipment that might otherwise lie unused.

Appendix A: Equipment Constraints and Resolving Capabilities

A parabolic reflector has constraints dependent on its diameter. Given the diameter w of our reflector, we can apply Rayleigh's criterion [Pedrotti and Pedrotti 1993, 335], which states that our object must be closer than a fixed distance for the object to be resolved:

$$\theta = \frac{1.22\lambda}{w}. \tag{1}$$

Here, θ identifies the minimum angle that can be used to resolve our object, and λ is the wavelength of the sound waves.

Pedrotti and Pedrotti [1993] state that λ is proportional to the speed of sound in water of a certain temperature divided by the frequency of the sound waves. This gives λ as

$$\lambda = \frac{c}{\mu}.$$

Substituting into **(1)** for λ, we have

$$\theta = \frac{1.22c}{w\mu}.$$

The speed of sound in sea water, c, is dependent on the temperature of the water. We assume a constant temperature of 25° C. For this temperature, $c = 1531$ m/s [Lide 1992, **14**–31]. The value of μ will be determined when we choose the sound wave frequencies to monitor.

We now derive a relation between smallest angle θ and the ratio

$$\frac{x = \text{object size}}{r = \text{maximum object distance}}.$$

Assuming that θ is a relatively small angle, we have

$$\frac{\theta}{2} \approx \tan\left(\frac{\theta}{2}\right) = \frac{x}{2r}, \qquad \frac{x}{r} \approx \frac{1.22c}{w\mu}.$$

Table 1 lists maximum object distances vs. required reflector diameter to view a $x = 10$ m-wide object at a frequency view of $\mu = 40$ kHz. To view an object the width of a submarine at a distance of 20 km, we would need a reflector approximately the length of a submarine, a truly impractical requirement.

Appendix B: 2-D Example

The following example demonstrates what an expected signal response will be for a specific geometrical situation. We consider a specific two-dimensional case with the following conditions (see **Figure 7**):

Table 1.
Range vs. required diameter of parabolic reflector.

Distance r (km)	1	2	5	10	20
Diameter w (m)	5	9	23	47	93

- We use a single parabolic reflector.
- The object to detect is a plane line segment, and our sensor lies below the outward normal of the object.
- We measure all angles relative to a line parallel to the line segment. The line segment can be viewed as defining the horizontal. The angle θ formed by the horizontal and the normal to the object directed towards the reflector is then $-90°$.
- We define several variables for use in our example:
  - The reflector is located directly below the midpoint of the object at a distance $r = 19.7$ m.
  - The object has a width of $x = 6.74$ m.
  - The diameter of the reflector is $w = 5$ m.

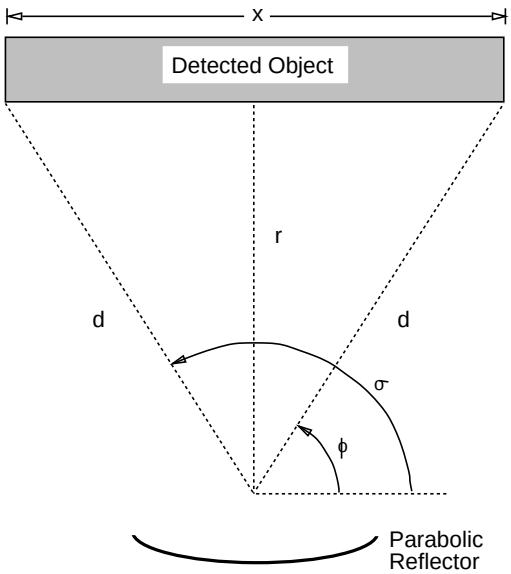

Figure 7. Relation between object and reflector.

The distance d from the corner of the object to the center of the reflector is

$$d = \sqrt{\left(\frac{x}{2}\right)^2 + r^2} = 20 \text{ m}.$$

The angle when the center of the reflector is pointed directly at the first corner of the object, measured from a line parallel to the object line segment, is

$$\phi = 90^\circ - \arcsin\left(\frac{x}{2d}\right) = 80^\circ.$$

Similarly, the angle when the center of the reflector is pointing at the last corner of the object is

$$\sigma = 90^\circ + \arcsin\left(\frac{x}{2d}\right) = 100^\circ.$$

To visualize the reflection signal from the object, we must introduce an ambient noise field. Let us assume that we are near the ocean bottom, where seismic disturbances produce a sound field that is more intense when scanning is directed at the sea floor than when scanning is directed upward toward the surface [Stephens 1970, 124–125]. The optical analogy is having the sun at your back while looking into a mirror. We consider a case in which the sound intensity I of the ambient noise field, with no object in the scan region, depends on the scanning angle λ as follows:

$$I(\lambda) = -2\sin\lambda + 90.$$

Figure 8 shows a plot of the function I.

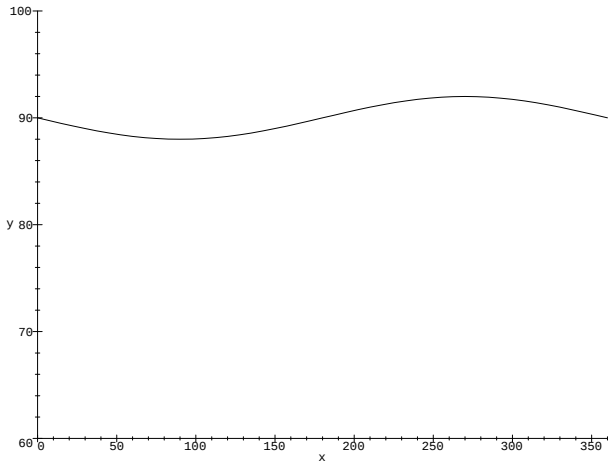

Figure 8. Plot of intensity I vs. scanning angle λ (in degrees).

With an object in the scan region, the reflector begins to pick up reflected noise from the object before being pointed directly at it. **Figure 9** shows the large region Ω (between the dotted rays) in which some noise reflected from the object is received. There is a small transition region at each end of the object where the object takes up only part of the view, including a part E in which the view is mostly of the object. Inside the region Γ, the object is in full view of the reflector and there are no boundary transition effects. Beyond region Ω, the signal received is just background noise.

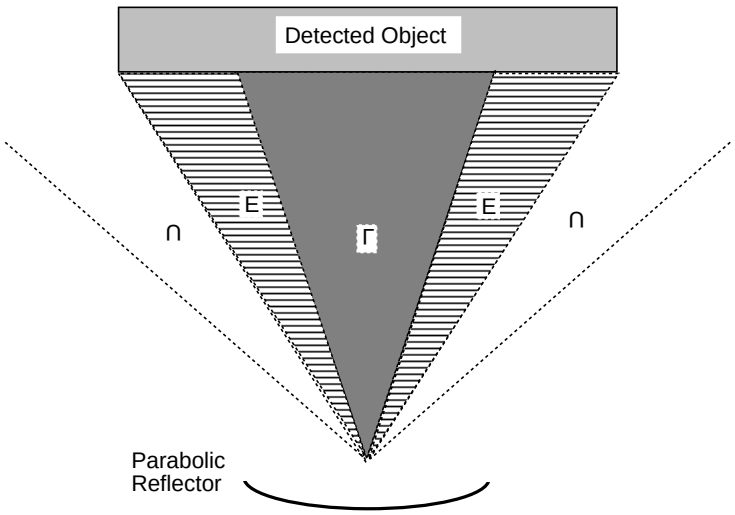

Figure 9. Reflected noise is received in region Γ but also in part in region E and in fact throughout region Ω (denoted here by ∩), which is bounded by the dotted rays.

The transition region can be described in terms of an angle offset ψ from the contact angles (σ and ϕ):

$$\psi = \arcsin\left(\frac{w}{2d}\right),$$

which is shown in **Figure 10**. The outer boundaries of region E are the angles ϕ and σ; the boundaries of region Ω are angles $\phi - \psi$ and $\sigma + \psi$; the boundaries of region Γ are the angles $\phi + \psi$ and $\sigma - \psi$. The right-hand transition region corresponds to the angle interval $(\phi - \psi, \phi + \psi)$, and the left-hand region to $(\sigma - \psi, \sigma + \psi)$, so that each region has angular width 2ψ.

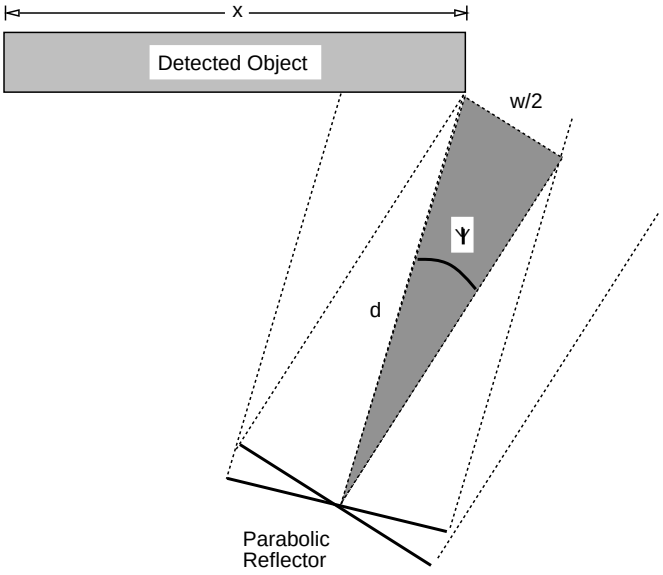

Figure 10. Definition of the angle ψ.

The intensity pattern from inside the region Ω depends on the angle between the direction at which the reflector is pointed and the normal from the target object (see **Figure 11**). All sound reaching the reflector must be parallel to the direction at which it is pointed, due to the characteristics of the parabolic shape of the reflector. Following sound from the reflector back to the object reveals from which direction sound is received when reflected.

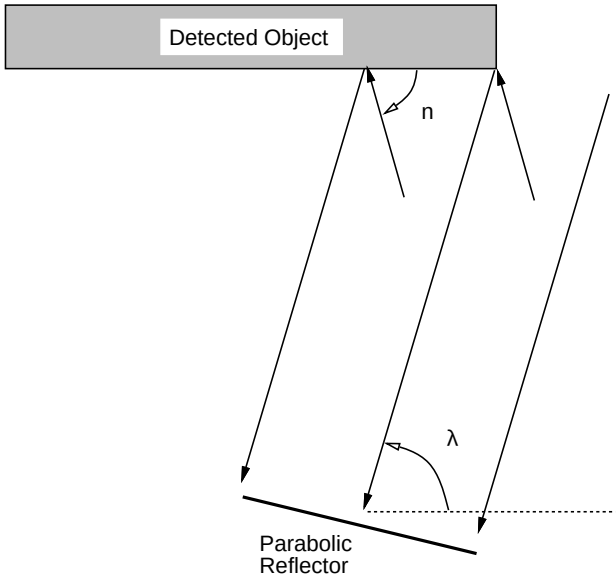

Figure 11. The scan angle λ and the view angle η.

The parallel rays from the reflector travel back to the object surface and reflect. The angle of incidence must equal the angle of reflection for a nearly perfect reflective surface. The angle the rays make from the horizontal after reflecting off the surface is

$$\eta(\lambda) = \pi - \lambda - 2\theta.$$

Using this reflected angle equation, we can produce a graph of the reflected intensity when the object takes up the full width of the sensor view, that is, when the object is in region E, where θ ranges from $\phi + \psi$ to $\sigma - \psi$. The intensity in this region is given by

$$I(\eta(\lambda)) = 2\sin\left(\frac{\lambda\pi - 2\pi^2}{180}\right)$$

and is plotted in **Figure 12**.

We approximate the boundary transitions of region Ω by fitting a Gaussian that matches the noise intensity of background at region Ω's outer edge and fits the noise reflected intensity at region E's outer edge.

The values of I in the two boundary regions are

$$I\big(\eta(\phi+\psi)\big)e^{\ln\left(\frac{I(\phi-\psi)}{I(\eta(\phi+\psi))}\right)(\lambda-\phi-\psi)^2/4\psi^2} = I\big(\eta(\phi+\psi)\big)\left[\frac{I(\phi-\psi)}{I(\eta(\phi+\psi))}\right]^{(\lambda-\phi-\psi)^2/4\psi^2},$$

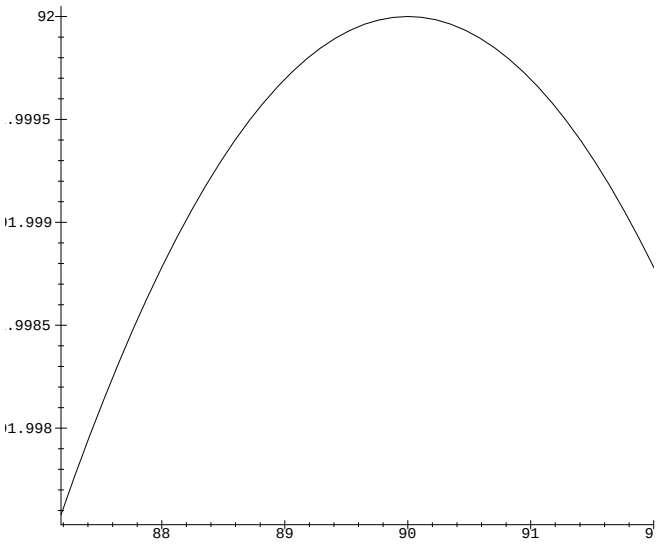

Figure 12. Object signal response vs. angle λ (in degrees).

$$I\big(\eta(\sigma-\psi)\big)e^{\ln\left(\frac{I(\sigma+\psi)}{I(\eta(\sigma-\psi))}\right)(\lambda-\sigma-\psi)^2/4\psi^2} = I\big(\eta(\sigma-\psi)\big)\left[\frac{I(\sigma+\psi)}{I(\eta(\sigma-\psi))}\right]^{(\lambda-\sigma+\psi)^2/4\psi^2}.$$

From these expressions, we can build a piecewise function that approximates the noise signal response from the object. Notice that this function is only an approximation, because it is not differentiable at all points. A better approximation can be obtained by fitting derivatives at boundaries as well. **Figure 13** gives a plot of the total intensity response over the entire domain of scan angle λ.

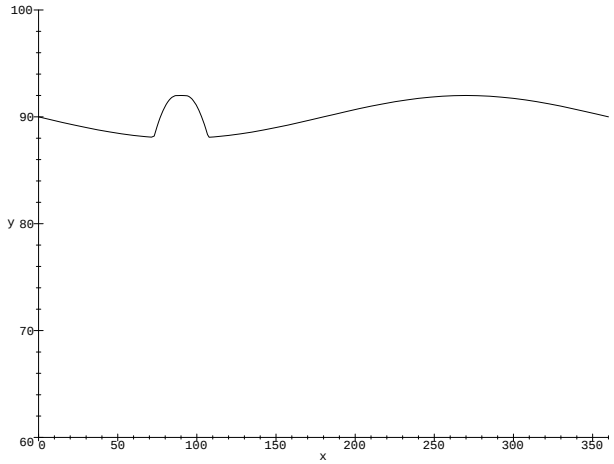

Figure 13. Total response vs. angle λ (in degrees).

To isolate the object image, we subtract the background noise field to obtain a difference plot (**Figure 14**). The distinct hump seen there will not always be the case. The intensity response depends on object surface geometry, reflectivity,

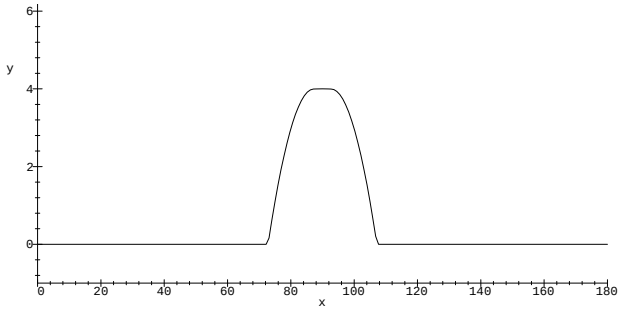

Figure 14. Difference plot.

and noise field directional characteristics. Certain parameter values will lead to difference plots that contain hills, valleys, and zero regions inside the overall object intensity response. Note that these complexities do not change the overall process of determining characteristics of the object, as long as a minimum and maximum angles of response can be found.

Appendix C: Point Location in Two Dimensions

We can determine the location of the object by triangulation from two reflectors, provided the object is not collinear with them (see **Figure 1**). Let d_i be the distance from reflector i to the target, θ_i the angle formed at reflector i, and ℓ the distance between the two reflectors.

Applying the law of sines to the triangle formed by the three objects yields

$$\begin{aligned}\frac{d_1}{\sin(\pi-\theta_2)} &= \frac{d_2}{\sin\theta_1} = \frac{\ell}{\sin(\pi-\theta_1-(\pi-\theta_2))},\\ \frac{d_1}{\sin\theta_2} &= \frac{d_2}{\sin\theta_1} = \frac{\ell}{\sin(\theta_2-\theta_1)}.\end{aligned}$$

Solving this for d_1 and d_2 gives

$$d_1 = \frac{\ell\sin\theta_2}{\sin(\theta_2-\theta_1)}, \qquad d_2 = \frac{\ell\sin\theta_1}{\sin(\theta_2-\theta_1)}. \tag{2}$$

Setting the origin of a coordinate system on reflector 1 and setting the x-axis to extend along toward reflector 2, the position of the object is then given by the ordered pair $(x, y) = (d_1\cos\theta_1,\ d_1\sin\theta_1)$.

Appendix D: Point Location in Three Dimensions

We assume that two reflectors have located a target and are aimed at it with direction vectors $\mathbf{v_1} = \langle a_1, b_1, c_1\rangle$ and $\mathbf{v_2} = \langle a_2, b_2, c_2\rangle$. We superimpose a three-dimensional coordinate system on our reflectors with the first reflector located at the origin. For simplicity, we build our coordinate system so that the second reflector lies on the x-axis a distance ℓ from the first one.

The position of the object is given by the intersection of the two lines with direction vectors $\mathbf{v_1}$ and $\mathbf{v_2}$. The parametric equations for them are

$$\text{line 1}: \begin{cases} x &= s\,a_1 \\ y &= s\,b_1 \\ z &= s\,c_1 \end{cases}, \qquad \text{line 2}: \begin{cases} x &= \ell + t\,a_2 \\ y &= t\,b_2 \\ z &= t\,c_2 \end{cases} \tag{3}$$

Their intersection is the solution of the system

$$\begin{aligned} s\,a_1 &= \ell + t a_2 \\ s\,b_1 &= t\,b_2 \\ s\,c_1 &= t\,c_2 \end{aligned} .$$

With only two unknowns, one equation is redundant. Solving the first two equations for s gives

$$s = \frac{\ell b_2}{a_1 b_2 - b_1 a_2}. \tag{4}$$

Substituting into **(3)** gives the location of the object as (sa_1, sb_1, sc_1), where s is given by **(4)**.

Object Velocity in Two Dimensions

We consider a worked-out example of a plausible trial if reflectors were set up and moved as outlined in this paper. In this scenario, two reflectors are placed on shore 100 m from each other (for this geometry, a third reflector is not needed to resolve the special case when the object and the two reflectors are collinear).

During the first sweep of both reflectors through a full $180°$, the reflectors find a peak in the ambient noise plot at angles of $\theta_1 = 84.3°$ and $\theta_2 = 91.9°$. Using **(2)** from **Appendix C** gives the location of the object:

$$\begin{aligned} x &= \frac{\ell \sin(\theta_2)\cos(\theta_1)}{\sin(\theta_2 - \theta_1)} = \frac{100\sin(91.9°)\cos(84.3)°}{\sin(91.9° - 84.3°)} = 75.1 \text{ m}, \\ y &= \frac{\ell \sin\theta_1 \sin\theta_2}{\sin(\theta_2 - \theta_1)} = \frac{100\sin(84.3)°\sin(91.9)°}{\sin(91.9° - 84.3°)} = 752.0 \text{ m}. \end{aligned}$$

Then, 5 sec later, the reflectors find the object at the angles $\theta_1 = 84.5°$ and $\theta_2 = 92.4°$. Using **(2)** again produces the new location of the object

$$x = 69.7 \text{ m}, \quad y = 723.6 \text{ m}.$$

Subtracting the coordinates gives the direction in which the object has moved:

$$\text{direction} = \langle 69.7 - 75.1, 723.6 - 752.0\rangle = \langle -5.4, -28.4\rangle.$$

So, the object is moving predominantly shoreward and slightly toward reflector 1. It speed is

$$\text{speed} = \frac{\text{distance}}{\text{time}} = \frac{\sqrt{(-5.4)^2 + (-28.4)^2}}{5} = 5.8 \text{ m/s}.$$

Therefore, the object is moving at 5.8 m / s (11 knots) at an angle of $79.2°$ SSW from the line through both reflectors. When first detected, it is 756 m from the first reflector and 752 m from the second reflector; at the second detection, it is 727 m and 724 m from the reflectors.

References

Buckingham, M.J., B.V. Berkhout, and S.A.L. Glegg. 1992. Imaging the ocean with ambient noise. *Nature* 356: 327–329.

French, A.P. 1971. *Vibrations and Waves.* New York: W.W. Norton.

Geil, F.G. 1992. Hydrophone techniques for underwater sound pickup. *Journal of the Audio Engineering Society* 40: 711–718.

Lide, D.R. (ed.). 1992. *CRC Handbook of Chemistry and Physics.* 73rd ed. London: CRC Press.

Pedrotti, F.L., and L.S. Pedrotti. 1993. *Introduction to Optics.* 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Smith, C. 1995. Revisiting solar power's past. *Technology Review* (July 1995). `http://web.mit.edu/afs/athena/org/t/techreview/www/articles/july95/Smith` (11 Feb 1996).

Stephens, R.W.B. (ed.). 1970. *Underwater Acoustics.* New York: Wiley-Interscience.

Thornton, S.T., and A. Rex. 1993. *Modern Physics for Scientists and Engineers.* New York: Saunders.

Toppan, A. 1996. Active USN ships—Submarines. `http://www.wpi.edu/~elmer/navy/current/usn_submarines.html` (11 Feb 1996).

Urick, R.J. 1983. *Principles of Underwater Sound.* 3rd ed. New York: McGraw-Hill.

Young, H.D. 1992. *University Physics.* 8th ed. New York: Addison-Wesley.

Judge's Commentary: The Outstanding Submarine Detection Papers

John S. Robertson
Dept. of Mathematics and Computer Science
Georgia College and State University
Milledgeville, GA 31061–0490
jroberts@mail.gac.peachnet.edu

Introduction

The problem of locating, classifying, and tracking objects under the ocean's surface is extremely important and has stimulated a great deal of significant oceanographic research. Despite the collapse of the Soviet Union and the end of the Cold War a few years ago, a number of countries possess submarine fleets that represent a very real strategic threat to other nations. Therefore, this kind of modeling problem will retain its importance for many decades.

Modeling

The fundamental approach to mathematical modeling can be summed in these three steps:

- Formulate a scientific problem in mathematical terms.
- Solve the underlying mathematical problem, perhaps inventing new mathematical methods in the process.
- Interpret the mathematical results in light of the original problem.

During the last step, the accuracy of the model's predictions are considered. If they are not good enough, then they can be used to highlight weaknesses in the model. Refinements are made and the three-step process is repeated as appropriate.

The Outstanding papers excelled in their application of both the first and third steps. For example, a critical factor noted by the judges was whether teams considered the environmental effect that the ocean has on sound propagation. Another factor weighed by the judges involved accounting for the properties of the ambient noise field itself. The literature contains extensive discussions

of both ideas, and too many teams did little or nothing in this area. A great number of papers made absolutely no attempt to do any true acoustic modeling. Instead, they looked like homework sets for a signal-processing course, rolling out page after page of theory without ever making a clear linkage to the problem posed. While the judges did not doubt the mathematical prowess present in some of those papers, those papers contained very little modeling—and modeling, after all, is what the contest was all about. Papers with simple models that were well conceived and whose shortcomings were clearly noted tended to fair much better that papers with extremely elaborate calculations and little connection with the real world.

Novel Approaches

Several papers stood out for the novel ideas that they incorporated into their problem analysis. This usually involved clever schemes for designing receivers so that they would work well under the conditions specified in the problem. Even though the mathematical analysis may have been a bit short, evidence of creative thinking generally gave teams that tried something new a substantial boost in the judges' eyes.

Literature Searching

Many teams made little or no effort to search the literature to discover relevant references. Between the time that the problem was chosen for the MCM and the contest date, *Scientific American* published an article that treated aspects of this subject shortly before the competition began [Buckingham et al. 1996]. Very few teams mentioned this paper among their references.

Conclusion

The very best papers displayed a healthy balance among the three modeling steps. Lots of powerhouse mathematics was certainly not sufficient for a paper to be competitive; teams in future years should bear this point in mind as they organize their write-ups. *Mathematical modeling is as much about modeling as it is about mathematical detail.*

Reference

Buckingham, M.L., John R. Potter, and Chad L. Epifanio. 1996. Seeing underwater with background noise. *Scientific American* 274 (2) (February 1996): 86–90.

Acknowledgment

The author is particularly grateful to Sedes Sapientiae for her help with this work.

About the Author

John S. Robertson is Chair of the Dept. of Mathematics and Computer Science at Georgia College and State University. He received his Ph.D. from Rensselaer Polytechnic Institute in 1986. He studied under Mel Jacobson and Bill Siegmann, two applied mathematicians who have made substantial contributions to the understanding of underwater sound propagation. Dr. Robertson subsequently became interested in problems related to atmospheric sound propagation and has written a number of research papers in both ocean and atmospheric acoustic propagation. He is passionately interested in applied mathematics and loves to teach students with all kinds of backgrounds. He enjoys living as a Yankee transplant in Georgia, where he no longer needs to shovel snow.

Practitioner's Commentary: The Outstanding Submarine Location Papers

Michael J. Buckingham
Marine Physical Laboratory
Scripps Institution of Oceanography
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093–0213
mjb@mpl.ucsd.edu
and
Institute of Sound and Vibration Research
The University
Southampton SO17 1BJ
England

Background

Underwater acoustics has a long history, dating back to ancient Greece, where scholars were interested in the hearing of fish; and to the ancient Chinese, who, taking a more pragmatic approach, listened at the end of a bamboo pole with the other end placed in the water, in an attempt to detect shoals of fish. This idea was further developed by Leonardo da Vinci, who, in 1490, described the use of a listening tube to detect distant shipping. However, it was not until September 1826 that the first quantitative investigation into underwater acoustics was performed, when the speed of sound in water was measured by two young scientists, Daniel Colladon and Charles Sturm, in a classic experiment conducted on Lake Geneva, Switzerland [Lasky 1977]. Surprisingly, in view of the rudimentary nature of their experiment, the result they obtained was within 3% of the currently accepted value of the speed of sound in water.

Most progress on methods of underwater acoustic detection has, of course, taken place in the twentieth century, with the two world wars providing the primary impetus to the development effort. Passive detection, in which a target is detected simply by listening for the sound that it makes, was the mainstay of submarine detection during World War I. As the war drew to a close, both Britain and the U.S. developed an accelerating interest in active detection techniques, in which a pulse of sound is projected into the water and the presence of a target inferred from the returning echo. However, active

systems came too late to play a significant role in influencing the outcome of World War I.

Development of sonar systems continued between the wars at a relatively leisurely pace; but, with the advent of World War II, underwater acoustic detection technology became a principal concern on both sides of the Atlantic. As a result, active echo-ranging systems were used extensively for submarine detection on surface ships and submarines. Other types of submarine detection device were also used, for example, magnetic anomaly detectors; but acoustics was and is the preferred approach to undersea detection, simply because the ocean is essentially transparent to sound and opaque to all other forms of radiation.

Today, half a century since the end of World War II, underwater acoustics is still performed using active and passive techniques. Although the technology has improved enormously over that period, the basic principles underlying these sonar techniques remain unchanged. Modern applications of underwater sound are not just military in nature but include bottom surveying and mapping by the offshore oil and gas industry, fish location and monitoring, and population studies of marine mammals. In some of these applications, neither active nor passive is ideal. Passive detection fails when the target is very quiet or silent, and active is undesirable in situations where the target, for example dolphins or whales, may be disturbed by the transmitted signal. Active detection has the further disadvantage of giving away the presence of the observer. In many military scenarios, this lack of covertness may preclude the use of active sonar altogether.

An alternative to passive and active sonar was introduced several years ago, based on the idea that ambient noise in the ocean acts as a form of acoustic illumination. The ambient noise is generated by a variety of sources, including breaking surface waves, precipitation, shipping, and biological sources, such as snapping shrimp in near-shore locations and numerous types of marine mammal throughout the oceans. Far from being the silent deep, the ocean is in fact a naturally noisy environment. The noise has much in common with daylight in the atmosphere in that both are random radiation fields, with components propagating in all directions.

By thinking of the noise as the acoustic daylight of the ocean, it is natural to pursue the opto-acoustic analogy, with a view to developing a new type of underwater acoustic detection system and ultimately perhaps an underwater imaging capability. The important question to be addressed is: *Can an object in the ocean be detected from the disturbance it introduces into the ambient noise field?* Any such object will scatter and reflect some of the incident noise, suggesting that by focusing the scattered component with a suitable acoustic lens, it should be possible to create an image of the object space. This, after all, is the essential process underlying conventional photography using daylight in the atmosphere.

Acoustic daylight imaging in the ocean has been the subject of an extensive research program at Scripps Institution of Oceanography for the past five years.

In the early days, a simple experiment was performed off the end of Scripps pier, the purpose of which was to establish whether detection or even imaging of an object is feasible through its effect on the noise field. The results of the experiment were encouraging [Buckingham et al. 1992], to the extent that a more ambitious program was launched with the aim of creating recognizable images of objects from the noise. This objective has now been achieved. Numerous real-time moving images of silent targets at ranges around 40 m have been created using ambient noise as the sole form of acoustic illumination [Buckingham et al. 1996]. The images consist of 126 pixels in an elliptical configuration; they show geometrical targets in the water column, as well as oil drums partially embedded in a sedimentary bottom.

Several theoretical analyses have been developed in support of the acoustic daylight imaging program. In one approach, Helmholtz-Kirchhoff scattering forms the basis of a numerical simulation of ambient noise imaging [Potter 1994]. The output of the numerical code is in the form of computer-generated images that closely resemble the acoustic daylight images obtained from the ocean experiments. On a different tack, a theoretical analysis of noise anisotropy and its effect on acoustic daylight images [Buckingham 1993] indicates that imaging should be possible under a wide range of illumination conditions. This three-pronged attack, comprising numerical simulation, theoretical analysis, and experimental observations, provides a substantial and growing body of evidence in support of the ambient noise imaging concept.

Ambient noise detection and imaging represents an ambitious technological challenge, hingeing on questions of fundamental physics, engineering design, extensive software development, theoretical analysis, numerical simulation, and delicate experimental procedures. Each of these aspects of the problem has to be integrated with the others in order to achieve a successful imaging capability. Although, as the culmination of a long-term effort, the most recent research has indeed reached a point where images can be created from the noise, the technique is still in its infancy. Perhaps a useful analogy is with the earliest days of television, when the fact that the pictures were of poor quality was beside the point: The critically important issue was that the pictures existed at all.

The Outstanding Papers

Continuing in the tradition of innovation in underwater acoustics, the four student teams addressed the question of detecting or imaging a submarine underwater with ambient noise. All the papers were well written and in each case gave careful consideration to the advantages and limitations of the technique. Since there is little in the literature on the problem, the students had to rely largely on their own imagination and creativity to make progress, the result of which is a number of interesting suggestions on the theory and practice of ambient noise imaging.

The Worcester Polytechnic Institute paper is interesting in that it considers various types of reflector as candidates for an acoustic lens. In fact, a similar approach was implemented in the original investigations of ambient noise detection, in the first instance using a parabolic dish [Buckingham et al. 1992] and subsequently a larger spheroidal dish [Buckingham et al. 1996]. This paper was the only one of the four to cite a reference to the research that has been performed on ambient noise detection and imaging. The authors propose an interesting triangulation technique for locating the position of targets based on the use of several parabolic reflectors, and then go on to consider two other detector geometries, a parabolic trough, and a parabolic torus. With regard to detection performance, their conclusions were cautious. They found that the technique would work, but only over limited ranges, out to about 1 km, due to the absorption of sound by seawater. Indeed, this is a limitation of ambient noise detection at frequencies in the tens of kHz range.

The Wake Forest University team took a different approach by developing a computer simulation of noise detection with four close-packed hydrophones at each of four locations. Based on several simplifying assumptions concerning the nature of the noise in the ocean, and by representing the submarine as a sphere, they proposed a detection technique to give the submarine's position and, through a Doppler shift in the noise spectrum, its velocity. The idea of determining target speed from the noise field is novel, although probably would be difficult to achieve in practice, since the noise is broadband, whereas the authors assumed it to be single frequency. Another of the assumptions adopted by the authors of this paper is that the intensity of the noise is statistically stationary everywhere, which in reality is definitely not the case. It is not clear whether their algorithm would be successful in detecting the noise fluctuations introduced by the target submarine against the fluctuations that are a natural feature of the ambient noise field. Still, this paper shows a good deal of imagination and gives a thorough discussion of the errors and sensitivity of the proposed technique.

An extensive discussion of ambient noise, its properties, and sources is given by the Pomona College team. The authors go on to propose the use of an array of directional hydrophones for detecting the presence of a submarine from the disturbance that it introduces into the noise field. In an extension of their technique, a second array is introduced to give a three-dimensional view of the submarine. This is the only paper that comes close to the idea that the ambient noise can be used not only for detection but also for imaging of objects. They also went beyond theoretical analysis by performing simple experiments in air, using loudspeakers to simulate an ambient noise field and a rolling trash can as the target. Silhouettes of the target were obtained when the trash can was between a source and the receiver. Toward the end of the paper, various schemes for producing images of a submarine with ambient noise are proposed. One of these is a sound camera, and a second is a variant of Schlieren photography, using a laser interference technique to observe the pressure fluctuations in the sound field. In the latter case, the authors appear to

be unfamiliar with Schlieren methods but worked out the essential principles for themselves.

The University of North Carolina team developed a computer-generated model of the ambient noise field and used it in conjunction with a recognition algorithm, in a simulation of submarine detection using ambient noise. The submarine was represented as an ellipsoidal object, which acted as an obstruction in the noise field, giving rise to a reduction in noise intensity (a silhouette) at the receiver. Contouring of the noise intensity forms the basis of the recognition algorithm. As in the Wake Forest paper, the contouring technique is claimed to provide position and velocity information on the target. In this case, however, the velocity is obtained, not by Doppler shifts, but by observing the submarine position at successive times. Apparently, a grid of sensors was assumed as the detector, with resolution limited to the inter-sensor spacing, although the details of the detection scheme are not clear from the paper. The contouring technique seems to work reasonably, although some of the arguments in the discussion are difficult to follow.

Overall, the quality of the papers is very high, and some of the ideas presented may well work in practice. It is interesting, though, that none of the candidates considered using a multi-beam phased array as an acoustic lens. Such a system is a strong contender for future ambient noise imaging systems. Perhaps we shall also see some of the novel ideas presented in the students' papers turned into real systems in the future.

References

Buckingham, M.J. 1993. Theory of acoustic imaging in the ocean with ambient noise. *Journal of Computational Acoustics* 1: 117–140.

_____, B.V. Berkhout, and S.A.L. Glegg. 1992. Imaging the ocean with ambient noise. *Nature* 356: 327–329.

_____, John R. Potter, and Chad L. Epifanio. 1996. Seeing underwater with background noise. *Scientific American* 274 (2) (February 1996): 86–90.

Lasky, M. 1977. Review of undersea acoustics to 1950. *Journal of the Acoustical Society of America* 61: 283–297.

Potter, J.R. 1994. Acoustic imaging using ambient noise: Some theory and simulation results. *Journal of the Acoustical Society of America* 95: 21–33.

About the Author

Michael Buckingham is Professor of Ocean Acoustics at Scripps Institution of Oceanography, La Jolla, California, U.S.A., and also a Visiting Professor at the Institute of Sound and Vibration Research, University of Southampton, U.K. He

is the originator of the acoustic daylight imaging concept, which his group at Scripps has developed over the past six years. He has been a Visiting Professor in the Department of Ocean Engineering, M.I.T., and an Exchange Scientist at the Naval Research Laboratory, Washington, D.C. Before joining Scripps in 1990, he was at the Royal Aerospace Establishment, Farnborough, U.K., where he developed his interests in underwater acoustics. In 1984, he received the A.B. Wood medal from the Institute of Acoustics, and he was the recipient of the Clerk Maxwell Premium from the I.E.R.E. in 1972. He is a Fellow of the Acoustical Society of America, the Institute of Acoustics, and the Institution of Electrical Engineers.

The Paper Selection Scheme Simulation Analysis

Zheng Yuan Zhu
Jian Liu
Haonan Tan
Fudan University
Shanghai, China

Advisor: Yongji Tan

Summary

We provide five models for selection schemes and, based on computer simulation, we propose optimal schemes for each model.

In analyzing the problem, we use a cost function to evaluate a scheme.

We quantify a judge's capability in terms of the variance d of accidental error and the magnitude e of systematic bias.

We enumerate our assumptions and give the algorithm for computer simulation. We then discuss the possible ranges of parameters d and e, finding that the judges' capability must reach a certain level to accomplish their task.

We discuss the five models. The Ideal Model can be used to explain the selection scheme under ideal conditions. The Round-Table Model and the Classic Model produce expensive solutions. To save money, we put forward the Cutoff Model and the Advanced Round-Table Model.

The Cutoff Model is based on numerical scoring and rejects papers under a certain cutoff level in each round. Its flexibility on changing the rejection proportion in each round, depending on the capability of the judge, leads to a more economical scheme. The Advanced Round-Table Model is a combination of rank-ordering and numerical scoring; it can generate a scheme that is both economical and easy to operate.

We compared all the models (see **Table 6**). The schemes produced by the Cutoff Model and the Advanced Round-Table Model reduce the expense drastically.

Later we generalize and find that all models except the Round-Table Model are suitable for different values of P, J, and W. For the purpose of classifying winners to various ranks, the Cutoff Model works best.

We find that the expense depends on the capabilities of the judges. A small decrease in the capability of the judges can lead to a great increase in expense in each model. So the best advice that we can give to the contest committee is

Do not hesitate to choose the best judges!

Assumptions

- There is an absolute rank-ordering and numerical scoring to which all judges would agree.
- The absolute numerical scores of all the papers are integers from 1 to 100, and are distributed $\mathrm{N}(70, 100)$ ($\mu = 70$, $\sigma^2 = 100$).
- A scheme is accepted if and only if it guarantees with 95% probability that the final W winners are among the "best" $2W$ papers.
- The judges score papers individually and do not influence one other.
- The judges' scoring has accidental errors that have the normal distribution; the magnitude of errors can be obtained from a judge's past records.
- Some judges have systematic bias for a specific kind of paper, hence they will mark higher or lower scores on such papers.
- In using the rank-ordering method, only the bottom 30% that each judge rank-orders are rejected.

Analysis of the Problem

Our main task is to provide a scheme that can reliably select the W winners and significantly reduce the number of papers for each judge to read.

The Evaluation Method

We can adopt either rank-ordering or numerical scoring to select the best papers. We elaborate the case of scoring, since a rank-ordering can be produced on the basis of the scores.

Expense

The purpose of reducing the number of papers read by each judge is to save on the expense of the contest. According to the theory of marginal utility, more papers to read means more money to be paid for each paper. So the reading number for different judges should be as equal as possible. The cost function depends on the actual situation, but we use the following function: Papers 1 to 20 cost \$$m$ each, 21 to 50 cost \$$2m$ each, and 51 to 100 cost \$$4m$ each. In mathematical form, the function is

$$C = m \cdot \sum_{i=1}^{J} \{a_i + (a_i - 20) \cdot u(a_i - 20) + 2 \cdot (a_i - 50) \cdot u(a_i - 50)\},$$

where a_i is the number of papers that judge i reads and

$$u(x-a) = \begin{cases} 0, & x < a; \\ 1, & x \geq a. \end{cases}$$

Small variation in the cost function has little effect on the scheme; with so little information, we might as well let $m = 10$.

The Judge

After preliminary simulation, we found that the capability of the judge is the most important factor in deciding the selection scheme. We use two parameters to describe the capability of a judge:

- The variance of accidental errors in scoring. The smaller the variance, the more abundant the judge's experience and the more accurate the judge's scores, and vice versa. The variance can be obtained from the judges' past performance.
- The magnitude of the systematic bias. This is hard to quantify, since it is difficult to determine the type and the magnitude of an individual's bias in real life. So we simplify the situation and classify both the judges and papers into three types: radical, neutral, and conservative. A radical judge gives higher scores to radical papers and lower scores to conservative papers and has no bias towards neutral papers. A conservative judge stands opposite to the radical one. A neutral judge has no bias at all.

Constructing the Model

Because of too many random factors in the judgment process, it is difficult to solve the problem theoretically. To solve the problem, we adopt computer simulation based on theoretical analysis.

The Algorithm of the Simulation

1. Generate 100 random integers from 1 to 100, as the papers' "real" scores, from the distribution $\mathrm{N}(70, 100)$. Put them in the array `paper_score[1,i]`.
2. Take a constant d as the upper bound of all the judges' accidental errors. Generate 8 random integers d_j as the standard deviations of the judges' accidental errors, using a discrete uniform distribution on the integers from 0 to d. Put these numbers in the array `judge[1,j]`.
3. Take a constant $e > 0$ as the systematic bias value. Let 1, 0, and -1 represent radical, neutral, and conservative, respectively. Give every paper a

number in $\{1, 0, -1\}$ and put them in array `paper_score[0,i]`. Give every judge a number in $\{1, 0, -1\}$ and put them in `judge[0,j]`. We calculate the systematic bias s from the expression

$$s = e * \texttt{paper\_score}[0, \texttt{i}] * \texttt{judge}[0, \texttt{j}].$$

For example, when a conservative judge meets a radical paper, $s = -e$.

4. The method that judge j scores paper i: Let

$$u = \texttt{paper\_score}[1, \texttt{i}] + e * \texttt{paper\_score}[0, \texttt{i}] * \texttt{judge}[0, \texttt{j}].$$

Generate random integers in [1,100] as scores that have the normal distribution $\mathrm{N}(u, d_j^2)$. Put them in array `judge_score[i,j]`. Thus we generate the judges' score matrix.

Determining the Parameters

We need to determine the values of d and e. First, we discuss how to determine the range of d.

From probability theory, we have

Lemma. *Let X_1, X_2,..., X_n be independent random variables with variances σ_j^2. Then $\overline{X} = \sum X_i/n$ has variance*

$$\sigma^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2.$$

So we have

Corollary 1.
$$\frac{1}{\sqrt{n}} \min_{1 \le i \le n} \{\sigma_i\} \le \sigma \le \frac{1}{\sqrt{n}} \max_{1 \le i \le n} \{\sigma_i\}.$$

From the corollary, we conclude that the accuracy of judgment can be improved if several judges work on each paper and average their scores.

Using the Cauchy inequality, we have

$$\sigma^2 \ge \frac{1}{n^3} \left(\sum_{i=1}^{n} \sigma_i \right)^2,$$

that is,

Corollary 2.
$$\sigma \ge \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} \sigma_i}{n}.$$

Since by our assumption σ is distributed discrete uniform on $[0, d]$, we have

$$\frac{\sum_{i=1}^{n} \sigma_i}{n} \approx \frac{d}{2}.$$

For $n \leq 8$, this becomes

$$\sigma \geq \frac{1}{2\sqrt{2}} \cdot \frac{\sum_{i=1}^{n} \sigma_i}{n} \approx \frac{\sqrt{2}}{8} d.$$

Now we have

Conclusion 1. *Generally speaking, accidental errors can be reduced when several judges work on the same paper. The more judges involved, the more accurate the result.*

Conclusion 2. *For the most part, in scoring a single paper, the standard deviation of the mean accidental error will not be lower than* $\frac{\sqrt{2}}{8} d$.

We have not proved the two conclusions; in fact, there are exceptions. But the probability of exception is too slim to have effect on the practical problem. So we grant these conclusions in our later discussion.

Based on lots of experiments with computer simulation, we find the experimental law listed below.

Law 1. $$d < 10,$$

where d is the upper bound on the standard deviation of the accidental judging error.

Verification: We need only explain that when $d = 10$, there is no selection scheme that guarantees with 95% probability that the final W winners are among the "best" $2W$ papers. We consider the ideal situation, under which each judge reads all the papers, for the case of 6 papers and 8 judges.

Because of **Conclusion 2**, the standard deviation of the 8 judges' mean accidental error is $\sigma \geq d\sqrt{2}/8$. We may suppose $\sigma = d\sqrt{2}/8$ as well, and here $d = 10$. We set the systematic bias to zero: $e = 0$. Using a simulation of the Round-Table Model and 10,000 iterations, 10,000 times, we observed 9,460 times when the judges chose the 3 winners correctly (i.e., they are among the "best" 6 papers). This probability of 94.6% is a little lower than our standard.

A calculation using Mathematica gave probability of failure (at least one winner not among the "best" 6 papers) as 5.6%. □

By using the simulation program for Round-Table again, with a changed value of the parameter d, we get:

Law 2. *When $d \leq 3$, Round-Table meets our 95% standard if each paper is read by only one judge.*

Comments on Laws 1 and 2

Law 1 points out that to succed in the task of selecting winners in a contest, the judges' capability should reach a certain level. If a judge's standard deviation d is more than 10, even if there is no systematic bias, a paper that deserves the score of 70 may be scored higher than 80 or lower than 60 with probability greater than 30%. The probability of a score higher than 90 or lower than 50 is no less than 5%. Such a person obviously cannot be qualified as a judge in any serious competition.

Law 2 points out that if all the judges are reliable enough—in other words, they are all experienced and have little systematic bias—a single judge's score is sufficient for determining the winners. When $d = 3$, running of Round-Table 5,000 times shows that the average failure rate is about 1.2%.

We conclude that if $e = 0$, we need to consider only d from 3 to 10; if $e \neq 0$, then d ranges from 0 to 10.

The range of e is quite difficult to determine. A reasonable supposition is that e has the same magnitude as d. We ran Round-Table with different values of d and e and used Mathematica to plot a three-dimensional bar chart (see **Figure 1**).

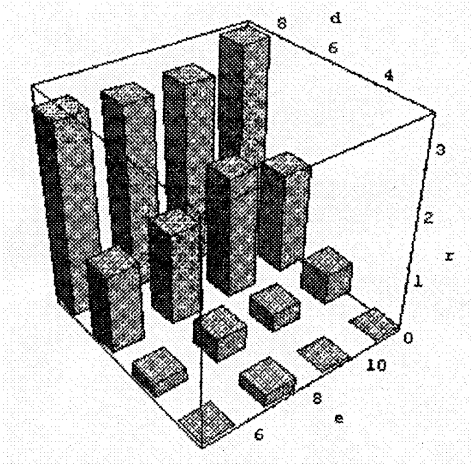

Figure 1. The failure rate $(r/100)$ is much more dependent on accidental error d than on bias e.

From the figure, we find that the standard deviation d has significant effect on the result, whereas the bias e has little effect.

Later we put forward several practical models for paper-judging and find some optimal schemes for different values of d and e, based on computer simulation.

We may suppose that $e \in \{0, 5, 10\}$ and $d \in \{1, 3, 5, 7, 9\}$. These limited ranges suffice to reveal the relationship between the scheme selection and the capability of judges.

The Ideal Model

When $d = e = 0$, every judge can rank-order or score all the papers as correctly as the absolute rank-ordering. This is the "ideal situation."

For 100 papers and 8 judges, 4 judges have 13 papers each and 4 have 12 each; a single reading of a paper, by any judge, suffices. The winners are the papers with the highest scores. The total cost is \$1,000, and the 3 winners must be the "best" 3 papers.

A good scheme to rank-order the papers is shown in **Figure 2**. This scheme guarantees that the winners are the "best" 3 papers. The cost is \$1,210.

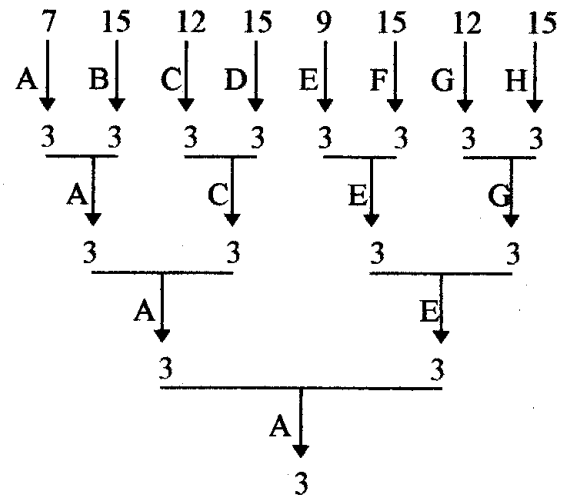

Figure 2. A scheme for rank-ordering the papers. The letters represent the 8 judges. The number at the tail of each arrow is the number of papers that a judge reads, and the number at the head of the arrow is the number that the judge selects.

But the most economical method is presented in **Figure 3**. Here, judges A, B, and C each rank-order 14 papers, and the other judges rank-order 13 papers each. We suppose that A is the Head Judge, who is responsible for picking out 3 winners from the 8 papers left after the first screening round. The cost of this method is \$1,070. Though it cannot ensure that winners are the top 3, it ensures that the "best 3" papers are among the 6 winners with probability of 99.3%.

We prove this assertion. If the "best" 6 papers distribute among at least 3 groups (i.e., are judged by at least 3 different judges at the initial screening), then the 3 winners are "qualified" (i.e., among the "best" 6). If the top 6 distribute among no more than 2 groups, there must be some unqualified winners. The number of the unfavorable events is

$$8 + \binom{6}{1} \cdot 8 \cdot 7 + \binom{6}{2} \cdot 8 \cdot 7 + \binom{6}{3} \cdot 8 \cdot 7 \cdot \frac{1}{2} = 1{,}744.$$

The number of ways to distribute 6 papers to groups arbitrarily is $8^6 = 262{,}144$, and $1{,}744/262{,}144 = 0.66\%$. In other words, all the winners are qualified in 99.3% of all the cases.

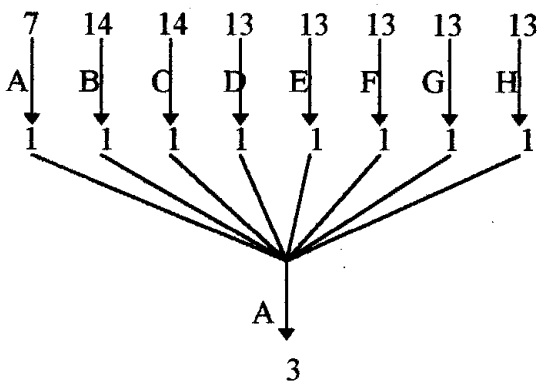

Figure 3. A more economical scheme for rank-ordering the papers.

The Ideal Model sets a lower bound for the cost function. When d and e are small (e.g., d=3, e=0), which means the judges are experienced and have almost no bias, the ideal model can work as the selection scheme.

The Round-Table Model

The distinguishing feature of this model is its simplicity.

1. Determine n, the number of rounds, according to the specific d and e.
2. Have all judges sit around a round table (hence the name of the model) and share the papers equally. At each round, after every judge has finished scoring, each judge passes the papers already read to the the right, where the neighboring judge scores these papers again in the next round.
3. After n rounds, each judge averages the n scores on a paper; the average score is the final score for the paper. The final score determines the rank-order of the paper.

The key question for this method is how to determine the value of n. Through numerical experiment, we found that systematic bias exerts only slight influence on n. So all our discussion later is based on the assumption that n is thoroughly determined by d.

When the distributions of the judges' accidental errors are all $\mathrm{N}(0, d^2)$, it is not hard to find that after n rounds the error of final score will be $d_n = d/\sqrt{n}$. For 8 judges, and with $d < 10$, we have $d_n < \sqrt{2}/8 \cdot 10 \approx 1.77$. Further simulation shows that the scheme is desirable if $d_n \leq 1.6$. We have

Law 3. *If all judges' error are distributed $N(0, d^2)$, and if $d_n = d/\sqrt{n} \leq 1.6$, i.e., $n \geq d^2/1.6^2$, then an n-round scheme is desirable; if $d_n \geq 1.77$, i.e., $n \leq d^2/1.77^2$, then an n-round scheme is not desirable.*

When all the conditions of **Law 3** are satisfied, we can easily find the optimal value of n. But the requirement on the error distribution is too harsh. When the variance of error has the uniform distribution on $[0, d]$, there is an empirical formula:

$$n = \min_{K \in N} \left\{ K \geq \left(\frac{d}{2 \cdot 1.6} \right)^2 \right\}.$$

The values of n obtained from the formula agree with the optimal n obtained from computer simulation (see **Table 1**).

Table 1.

Failure rate in 1,000 iterations, and expense, for various numbers of rounds n.

Bias e	Max variance d	Number of rounds n	% Failure rate	Expense
0	3	1	1.2	$1,000
	5	2	4.4	$2,400
	7	5	3.2	$10,400
	9	8	2.8	$22,400
5	3	1	3.6	$1,000
	5	2	4.7	$2,400
	7	5	4.8	$10,400
	9	8	4.4	$22,400

We see from the table that the expense increases rapidly as d rises. So we had better not use this scheme if the capability of the judges is quite ordinary.

The Classical Model

We put forward a classical model by combining rank-ordering and scoring methods.

1. Distribute the papers to each judge equally as far as possible. If a judge meets a paper already scored by that judge, the judge exchanges it with another judge. The judges score the papers.

2. Every judge rank-orders the papers that that judge has just scored and determines the bottom 30%.

3. Each judge's bottom 30% of papers are rejected.

4. If only 3 papers remain, they are winners. If there have been 8 rounds, then all the papers left have been scored by the 8 judges. So average each paper's scores and select the highest 3 as the winners. Otherwise go back to Step 1.

This model strictly limits the rejection of the papers during each round and refrains as much as possible from rejecting the good papers. The stability and precision of the model are very high, but the flexibility is low. Because of the

progressive rejection method, it costs less than the Round-Table Model when d is comparatively large; but in general, this scheme is comparatively expensive. See the result of the simulation of the Classical Model in **Table 2**.

Table 2.
Simulation results for the Classical Model.

Bias e	Max variance d	% Failure rate	Expense
0	0	0.0	$4,851
0	5	2.0	$5,022
0	9	4.1	$5,563
5	0	1.1	$5,462
5	5	1.9	$5,779
5	9	4.8	$6,250
10	0	5.5	$6,528
10	5	6.3	$6,653
10	9	10.7	$7,395

The Round-Table Model and the Classical Model can be used as selection schemes. However, the expenses are not encouraging. So we put forward two models based on them that will be more economical.

The Cutoff Model

This model is based on the Classical Model. However, it changes the cutoff levels in each round. Thus, it is not constrained to reject 30% in each round but can determine the rejection proportion by the circumstances. So it is more flexible.

1. Determine the failing percentage of each round. When there are n rounds, the falling percentage is $x = \sqrt[n]{0.03}$. Each paper is scored in each round, so we can complete the judging work in no more than 8 rounds. Hence $n = 8$.

2. Distribute the papers to each judge equally; each judge should get papers that that judge has not previously scored.

3. The judges score the papers. Average each paper's scores as its this-round-score. Determine this round's cutoff level using the failing percentage and the papers' this-round-score. Any paper below the cutoff level falls.

4. If only three papers remain, they are the winners. Otherwise, go back to Step 2.

For fixed variance d and systematic bias e, we experimented with different values of n to find the optimal scheme, that is, the scheme with the smallest number of rounds that satisfies the criterion of an failure rate less than 5%. **Table 3** gives some results.

Table 3.
Simulation results for the Cutoff Model, giving percentage failure rates for combinations of e and d with values of n.

Bias e	Max variance d	Number of rounds n								Expense
		1	2	3	4	5	6	7	8	
0	0	0								
	3	0								
	5	9	6	1						$1,189
	7	25	17	8	6	7	4			$1,661
	9	47	28	16	15	6	6	7	5	
5	0	16	10	7	3					
	3	11	5	4	6	1				
	5	29	7	3	7	2				$1,725
	7	65	45	18	15	8	6	8	7	
	9	62	33	23	18	18	11	10	3	

Figure 4 shows the rank of the third-best paper for two of the optimal schemes, and **Figure 5** shows that the simulation of these schemes is stable.

We see that the cost is significantly less than for the two previous models; the total reading times are reduced because many of the unqualified papers are rejected earlier. However, the method of distributing papers in each round is comparatively complicated and may cause trouble in practice.

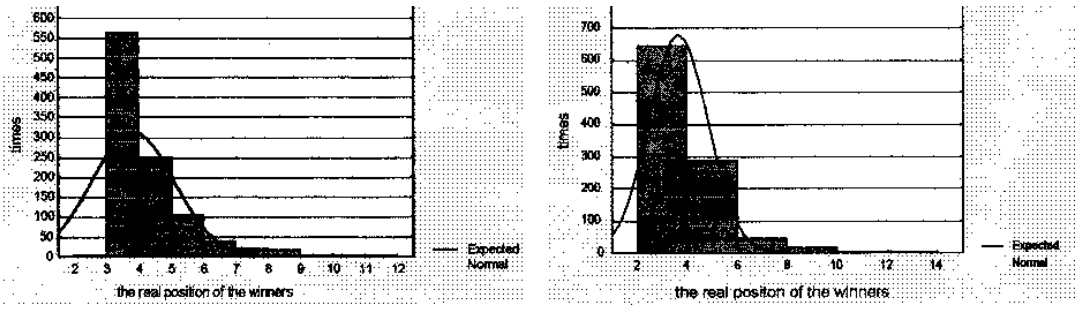

Figure 4. Rank of the third-best paper for two optimal schemes, over 1,000 iterations.

a. $e = 0$, $\max d = 5$, and 2 rounds. **b.** $e = 5$, $\max d = 5$, and 4 rounds.

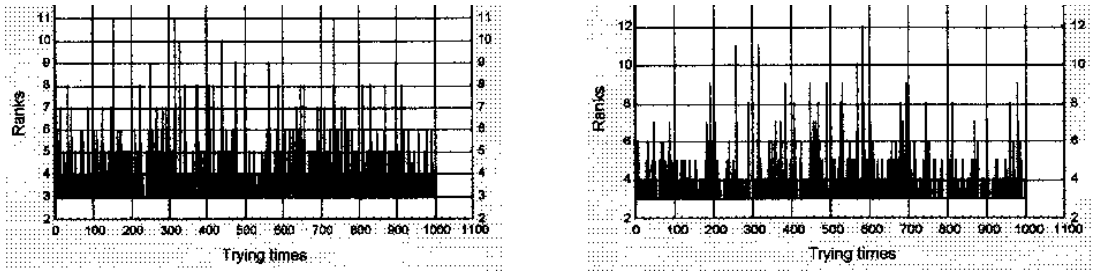

Figure 5. Rank of the third-best paper vs. iteration number, for two optimal schemes.

a. $e = 0$, $\max d = 5$, and 2 rounds. **b.** $e = 5$, $\max d = 5$, and 4 rounds.

The Advanced Round-Table Model

This model use a combination of rank-ordering and numerical scoring. In its early stage, we ingeniously use the method of group sequencing and partially exchanging to realize rejecting in definite proportion without violating the assumption that only the bottom 30% that each judge rank-orders are rejected. This selection scheme is stable in structure and easy to implement. Setting the times of paper-exchanging each round also makes it flexible. In the final stage, we determine the winners from average scores to speed the selection process.

1. We distribute the papers equally to the judges. The rejection proportion in each round is 30%. After n rounds of screening, each judge has only one paper left (when 30% is less than 1, we consider it to be 1). At the first round, we control the roundoff method to ensure that each judge has an equal number of papers left. At other rounds, we round down.) Given the capability of judges, we can determine the exchange times K_i in each round.

   For our problem with n=6, the numbers of papers each judge has after each round are 9, 6, 4, 3, 2, and 1. The numbers of papers rejected in each round are 4, 3, 2, 1, 1, and 1.

2. The judges sit around a round table. Let $K_i = i$ (later we thoroughly discuss the method of selecting the value of K_i). At the first round, $K_0 = 0$, judges do not exchange papers but rank-order them and eliminate the worst 30%.

3. At the second round, $K_1 = 1$. Each judge passes the worst 30% papers to the right. Then each judge scores the new papers received and re-rank-orders *all* current papers (including the ones not passed on) and cuts off the worst 30%.

4. For $K_i \geq 2$, the passing, scoring, re-rank-ordering, and rejecting the worst 30% takes place i times.

5. When each judge has only one paper, the paper is passed K_n times. If the paper has already been scored for K_n times, the other judges do not score it. From the mean value from the K_n scorings, we select the best three papers as winners.

The exchange method in this scheme is somewhat like that of the Round-Table Model, so we call it Advanced Round-Table.

Why Exchange the Bottom 30%?

We assumed that only the bottom 30% papers that each judge rank-orders could be rejected. Hence the number of papers left after each round will not be a fixed number, which makes the method more complicated, more unstable, and increases costs. But this undesirable feature can be avoided if we circulate

the bottom 30% papers. For example, consider a paper among the bottom 30% that J_1 has passed to J_2. If it is still among the bottom 30% of J_2's papers after J_2's re-rank-ordering, it is certain that it should be eliminated. If this bottom 30% also contains papers that are not from J_1, the paper in question can be eliminated without violating the 30% rejection rule (it has been rank-ordered only by J_2). Moreover, if J_2 believes that it is even worse than some of the bottom 30% from J_1, it is reasonable for J_2 to reject it.

How Many Times to Pass the Papers?

Since at each round we must determine how many times to pass the papers, the process of searching for the optimal scheme is much more complicated than the Round-Table and Cutoff Models. However, the flexibility of the model increases with the complexity, which makes it possible to find a selection scheme that is both efficient and economical.

First, we note two properties of $\{K_i\}$:

1. $\{K_i\}$ is bounded, i.e., $0 \leq K_i < J$.

2. $\{K_i\}$ is monotone increasing, i.e., $i < j \Rightarrow K_i \leq K_j$.

There are only J judges, and all the papers are divided into J groups. When $K_i > J$, a judge may find that a paper previously passed on comes back again! That is anything but efficient. So $K_i < J$.

At the last several rounds, it seems more likely that the qualified papers (the top 6 papers) may be eliminated, so there should be more passings at later rounds than at early ones. So $\{K_i\}$ is monotone increasing.

Second, there is a relationship between $\{K_i\}$ and the cost C. That is, C is a monotone increasing with the total number K of papers that all judges have rank-ordered, $K = 8\sum_{i=0}^{n} P_i K_i + 100$, where P_i is number of papers to be eliminated at round i. In this model, the numbers of papers read by each judge are almost equal (the difference is no more than one paper), so C is monotone increasing with K.

It's clear that the expense and the precision of selection scheme are at odds: The less the expense, the bigger the probability of producing an unqualified winner. We might begin as well with the $\{K_i\}$ whose corresponding C is the smallest. By testing the $\{K_i\}$ according to the order of the sequence one by one, we increase the expense step by step and at the same time increase the precision of the scheme. When the precision meets the request, the scheme that this $\{K_i\}$ corresponds to is the optimal scheme. We can use our simulation program to test the precision of the scheme.

We should say that this method of searching for the optimal scheme is time-consuming. To find an optimal scheme corresponding to a group of specific values of d and e, we may spend hours; but compared to the savings, several hours of machine time is nothing.

A more efficient approach is to use binary search to find the "best" $\{K_i\}$.

Table 4.
Failure rate and expense for various sets of $\{K_i\}$.

Bias e	Max variance d	K_n		Iterations	% Failure rate	Expense
0	5	1, 1, 1, 1, 1, 2, 4	*	20,000	1.8	$1,120
	7	1, 1, 1, 1, 1, 4, 5	*	10,000	3.9	$1,560
	9	1, 1, 1, 2, 2, 4, 5		5,000	4.8	$1,960
5	5	1, 1, 1, 2, 2, 2, 4		1,000	0.7	$1,480
	7	2, 2, 2, 2, 2, 4, 8		1,000	2.7	$3,760
	9	2, 2, 2, 2, 2, 4, 8		1,000	6.7	$3,760

Listed in **Table 4** are several $\{K_i\}$ (the two starred sets are optimal).

From the table we notice that under the same conditions this model's cost is less than that of all the other models, while its operating process is clear and definite, easy to understand, and easy to apply in practice.

The weakness is also obvious: finding the optimal $\{K_i\}$, which is time-consuming.

Comparison and Critique of the Models

We have discussed five models, all suited for practical use except for the Ideal Model, which can be used under ideal conditions only. **Table 5** compares the precision and expense of the different models under specific conditions.

Table 5.
Precision and expense of the different models.

Model	Bias e	Max variance d	Iterations	% Failure rate	Expense
Round-Table	0	5	1,000	4.7	$2,400
Classical			1,000	2.0	$5,022
Cutoff			1,000	1.8	$1,414
Advanced Round-Table			20,000	1.8	$1,120
Round-Table	0	7	1,000	4.8	$10,400
Classical			1,000	2.3	$5,389
Cutoff			1,000	4.4	$1,661
Advanced Round-Table			10,000	3.9	$1,560

From the table we see that when the judges's capability is high ($d = 5$), the Classical, Cutoff, and Advanced Round-Table Models' precisions are very sharp, while the Round-Table Model's is a little lower. As far as cost is concerned, the Classical Model is highest, followed by the Round-Table Model, the Cutoff Model, and the Advanced Round-Table Model. The costs of the latter two are extraordinarily low.

When the judge's capability is comparatively low ($d = 7$), the precision of every model is almost the same. The cost of the Round-Table Model is too

huge to assume and the Cutoff Model and the Advanced Round-Table Model are both again considerably lower.

To get a comprehensive idea of all the models, we summarize various criteria in **Table 6**. The contest committee can decide which model to use, based on this able and the concrete circumstances. We advise that the committee choose the best judges, even though they may cost more than others, for it would prove to be economical on the whole.

Table 6.
Comparative features of the models.

	Ideal	Round-Table	Classical	Cutoff	Advanced Round-Table
Adaptability	v. low	high	high	high	high
Precision	v. high	medium	high	high	high
Expense (large d)	—	v. high	high	low	low
Expense (small d)	low	low	v. high	low	low
Complexity of initializing	—	low	—	medium	v. high
Complexity of execution	low	low	high	high	low
Determinacy	aboslute	high	low	low	high
Flexibility	low	high	low	high	v. high

Generalization of the Model

Different Values for Parameters (P, J, and W)

The Classical Model can be applied directly with different parameter values. For the other models, all we need to do is to determine the parameters for the optimal scheme based on the new values of P, J, and W. For the Round-Table Model, the parameter is the number of times to pass the papers; for the Cutoff Model, it is the number of rejection rounds n; and for the Advanced Round-Table, it is the sequence $\{K_i\}$ of numbers of times to pass papers in each round.

All the empirical formula and laws in our paper are deduced for the particular given values of P, J, and W (100, 8, and 3), so they do not apply to a new problem automatically. However, using the method offered by our paper combined with computer simulation, we can find new empirical formulas and laws and determine the new parameters of optimal scheme easily and quickly.

For instance, let us take Problem B of the 1995 MCM, which had $P = 174$, $J = 12$, and $W = 4$. Let us assume that $d = e = 5$.

Using the algorithm of the Round-Table Model, we find that $n = 4$ is the best choice; the failure rate is 3.0%, and the expense is \$13,440. For the Classical Model, we get 2.8% and \$21,320. For the Cutoff Model, the optimal value is $n = 4$, the failure rate is 4.2%, and the cost is \$3,502. Finally, for the Advanced

Round-Table Model, the optimal set of parameters is $K_1 = K_2 = K_3 = K_4 = K_5 = K_6 = 0, K_7 = K_8 = 1$; the failure rate is 3.0% and the cost is \$1,700.

More Winners

Sometimes a few outstanding papers are not the only result of a contest. We may be asked to classify other papers as Meritorious, Honorable Mention, and Successful Participation, as the MCM does, in order to encourage the participants. Except for the Ideal Model, all the selection schemes discussed in our paper are suitable for this task. Compared with other schemes, the Cutoff Model is best, since it ranks all papers.

Strengths and Weaknesses

Strengths

The Cutoff Model and the Advanced Round-Table Model successfully generate selection schemes that can drastically reduce the cost of judges, and we have given practical methods for determining the optimal selection scheme of both models. Both models and their methods are easy not only to understand but to apply in practice, and they easily generalize to most situations.

Our simulation program requires little memory and runs very fast, so it is especially suitable for practical application.

Weaknesses

Since we largely rely on computer simulation to test our models, to select the optimal selection schemes, and to verify our laws, we cannot assure that our result is 100% definitive. However, we ran the simulation program more than 1,000 times before drawing any critical conclusion, and the result of the simulation are very stable. We believe that all our results are reliable enough for application.

Because of the absence of relevant information, our cost function may not conform to reality.

Modeling Better Modeling Judges

Brian E. Ellis
Chad Hall
Charles A. Ross
Gettysburg College
Gettysburg, PA 17325

Advisor: James P. Fink

Summary

We designed a system for judging a contest of papers with two main goals in mind: to minimize the number of reads by a judge and to ensure a fair contest. We first designed a model that would best predict the choices of human judges comparing just two papers. The basic premise is that the closer two papers are in an absolute ordering, the more likely the order of the papers is to be reversed by a judge, whereas the farther they are apart, the less likely a reversal.

Our model accommodates arbitrary numbers of judges, papers, and winners. The P papers are split into S stacks. To ensure fairness, two judges read each stack. From every pair of stacks, W papers advance to the next round. If two judges cannot agree which W should advance, the Head Judge decides. The rounds continue until $2W$ papers remain, when a balloting process among four judges and the Head Judge determines the W winners.

We can predict the total number of reads made in the judging process and the maximum number of reads by any judge. We calculate an optimal number of judges so that all judges have nearly the same number of reads.

Testing on a computer, we found that our model fails to pick W out of the top $2W$ no more than 0.1% of the time. These failures are attributable to the human factor in judging. For the given problem of 8 judges deciding on 3 winners from 100 papers, our model predicts and tested at 254 total reads, with 32 papers read by each judge; the model fails to select 3 of the top 6 only 0.08% of the time. For 32 judges deciding on 7 winners from 350 papers, our model predicts and tested at 1,162 total reads, and 36 papers read by each judge; the model fails to select 7 of the top 14 only 0.01% of the time.

Assumptions with Justifications

- Papers:
  - Ranking: There is an absolute ordering of the papers, so we can determine if the winning papers are within in the top $2W$ total papers.

  - Number: The number of papers is far greater than the number of winners.

- Judges:
  - Knowledge: All judges are knowledgeable about the question posed and can easily determine if a paper has merit; otherwise, a paper cannot be fairly evaluated.
  - Preferences: All judges will agree on the ranking of a particular paper within some margin of error. Each judge has personal preferences about what is desirable in a paper. Also, when a judge is asked to read a large number of papers, there must be some margin of error in the ranking process.
  - Ability: A judge can read up to 20 papers at a sitting and still pick out the top papers with a reasonable amount of accuracy. In speaking with a number of professors and contest evaluators, we found that 20 is the upper bound on the papers that the professors feel that they can evaluate fairly at one time.
  - Head Judge: The Head Judge only settles disputes and votes in the final round; the Head Judge is not counted in the number J of judges.
  - Number: The minimum number of judges is 5, including the Head Judge. There must be enough judges to evaluate all of the papers fairly; the more judges, the more accurate the process will be.

- *Fairness* is the ultimate variable. In any contest, judges must be willing to sacrifice time and energy to ensure that the best papers will win the contest. The credibility of the contest is based upon the fairness and correctness of the judging.

Definitions of Constants and Terms

P: total number of papers

J: total number of judges, not including the Head Judge

J_k: representation of judge k

W: total number of winners

read: one judge reading one paper one time

round: a process of elimination in which a set of papers is cut to W papers

R_a: the representation of round a

S_a: the number of stacks in round a. A *stack* is a set of papers of size $< P$.

N: the number of papers in a stack

S_{jk}: representation of stack j in round k

error: a judge's ordering that contradicts the absolute ordering

The Paper Contest Model

The model begins by dividing the P papers into S stacks. Judges then perform an elimination round in which two judges work together to combine two stacks into one stack of W papers. The comparisons are made by rank ordering, using no numerical scoring system. The process is repeated for the new stacks until two stacks are left. The final round then enacts a voting process on the last two stacks to declare the winners.

Preliminaries

We first determine the number of stacks, S_1, needed for the first round. To ensure a symmetric elimination, we need S_1 to be a power of 2. By our assumptions, each judge can read up to 20 papers, so the size of a stack cannot exceed 20. The number of papers in each stack is $N = P/2^n$, where n is the smallest value that satisfies

$$N = \frac{P}{2^n} \leq 20.$$

If 2^n does not divide P evenly, N is rounded up. The papers are distributed as evenly as possible about the S_1 stacks. We assign each judge one stack until we run out of either stacks or judges. If we run out of judges, then some judges will be asked to repeat the first round.

First Round

Judges J_1 and J_2 are assigned stacks S_{11} and S_{21}. Judge J_i chooses W papers from the stack S_{i1}; W, to ensure that all of the W best papers cannot be eliminated in round R_1. Once done, they swap stacks. Judge J_1 then chooses W from S_{21}, while J_2 chooses W from S_{11}. Together, they compare their lists and determine W from the union of S_{11} and S_{21}. If there is a dispute, the Head Judge determines which paper advances. Each pair of stacks is cut to W papers in the same manner. At the completion of the first round, there are $S_2 = 2^{n-1}$ stacks and $N = W$ papers.

Why Choose W Every Time?

The scenario could arise in which the top $2W$ papers fall into one stack in any round. If we return any fewer than W papers, the model would automatically

fail. To return more than W papers would increase the stability of the model, but not to a degree that would warrant the increased number of reads required.

Second and Subsequent Rounds

There will be $n-2$ "middle" rounds (see **Appendix A**). The procedure for these rounds can be generalized with the introduction of a variable r that holds the value of the round number. At the beginning of R_r, we have $S_r = 2n-r+1$ stacks and $N = W$ papers. The next two available judges are assigned the stacks S_{1r} and S_{2r}. Each chooses W papers from the union of stacks S_{1r} and S_{2r}, and then they agree upon a final W to advance, with the Head Judge settling any disputes. Every pair of stacks is cut to W papers in the same manner. This is repeated round by round up to and including round R_{n-1}, at the completion of which there will be $2W$ papers remaining.

Final Round

The final round, R_n, is a voting process. To ensure fairness and to account for the importance of the final decision, we choose five judges, including the Head Judge, to evaluate the papers. These judges read the remaining $2W$ papers and rank-order them. An official, possibly an extra judge, tallies the votes, giving W points to first place, $W-1$ points to second place, and so forth, down to 1 point for place W. The W papers receiving the most points are the winners. If there are any ties in the points, the ballot of the Head Judge breaks the tie.

Human Factor

The one variable that this or any model cannot control is the human factor. We simulate this human factor by using a probability distribution that models what an actual judge may do. If all the judges are exactly the same, paper 1 will always be ranked ahead of paper 2. However, individual judges will have preferences about what they would like to see in a paper. The most common example is one judge who would weight presentation over substance while another judge would rate substance over presentation. In this case, paper 2 could easily be rated above paper 1. To model this factor, we chose the following function as the probability that a judge's ranking of two papers differs from the absolute ordering

$$E(P,d) = \frac{1.46 + \arctan(1 - 60d/P)}{2.92 + \frac{\pi}{2}},$$

where there are P papers in the contest and d is the distance between the two compared papers on the absolute scale.

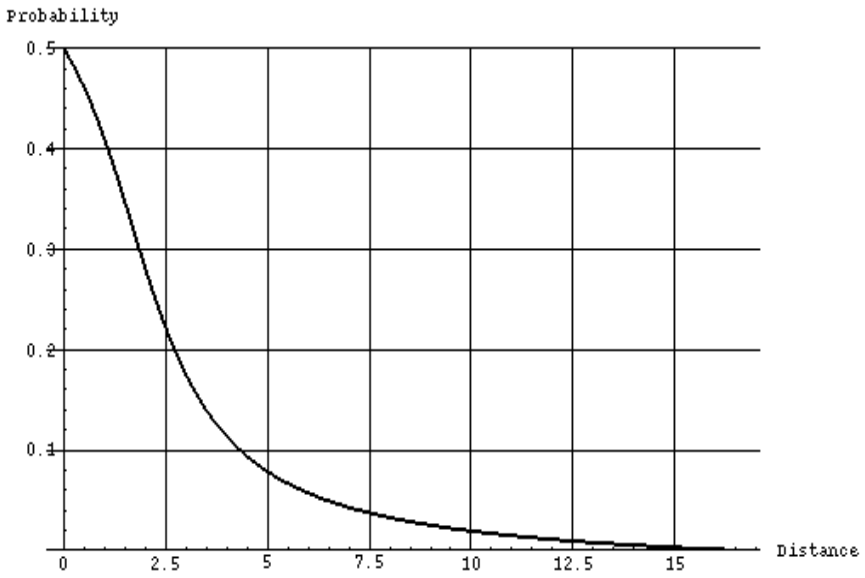

Figure 1. The operating characteristic curve for a judge's ranking of two papers. (Note that this is not a probability density function.)

The formula gives the probability of the judge making an error as a function of the true difference in ranks between two papers. As the distance between two ranks increases, the probability of reversing of the papers decreases quickly. The probability of error when there is a $0.01P$ difference between two papers is approximately 50%. So, the choice between papers 5 and 6 for $P = 100$ is completely random. The probability of error when there is more than a $0.17P$ difference is 0. In this case, the magnitude of the difference between the papers is too great; it would be impossible to err in the comparison. The values for the probability of error between $.01P$ and $.17P$ are representative of real life—the closer two papers are, the more likely a judge's personal preferences of style will influence the ordering of the papers. Similarly, the further two papers are apart, the less likely the judges preferences will be able to affect their comparison.

Results

Total Reads

The total number of reads, excluding arbitrations by the Head Judge, is given by

$$2P + \sum_{i=2}^{n-1} 2^{n-i+2}w + 5(2w).$$

The first term is to the number of reads in R_1, the second takes care of rounds R_2 through R_{n-1}, and the third is for R_n (see **Appendix A**).

Number of Judges

The model requires five judges, including the Head Judge. The model can accommodate any $J \geq 4$, but there is an optimal number of judges J_O that minimizes the maximum number of reads per judge. That optimal number of judges equals S_1, not counting the Head Judge. All J_O judges are needed in R_1, with one-half used in R_2, one-fourth in R_3, and so on. We use each judge in the first and in one subsequent round, leading to nearly the same number of reads (see **Appendix A**).

Maximum Reads per Judge

If $J \geq J_O$, the maximum number of reads is

$$2\left\lceil \frac{P}{2^n} \right\rceil + 2W.$$

If $J < J_O$, the maximum number of reads can be very large, even unreasonably large. In this situation, some of the J judges will be required to read two or more pairs of stacks in round one. Already these judges will have to read at least 40 papers and possibly more, because the second and subsequent rounds haven't even started. If a $J < J_O$ is chosen, J must be close to J_O or there will be many unhappy judges.

Testing the Model

We implemented the model in the C programming language, making some minor additional assumptions (see **Appendix B**).

We then ran tests for various combinations of P and W, always using the optimal number J_O of judges. We did 10,000 iterations each for the cases in **Table 1**. The test data returned an average failure rate of 0.023%.

Table 1.
Combinations of number of papers P and number of winning papers W, for each of which 10,000 iterations were run.

P	W	percentage failure	total reads	max reads per judge
50	2	0.02%	104	26
100	3	0.08%	254	32
200	5	0.01%	570	36
350	7	0.01%	1,162	36

The model is valid. It agrees with the formula for the maximum number of reads, total reads, and, most important, the final W are consistently among the top $2W$. The small failure rate is attributable to the human factor. Whenever the human element is involved, there are bound to be rare cases that occur.

Strengths and Weakness of the Model

Strengths

- The probability of the model failing is extremely low, usually less than 0.1%.
- The model takes into account the possibility of human error.
- All judging is done via direct comparison, and at least two judges must concur for a paper to advance. There is no numerical scoring, which can be biased by the grading scale of the judge; and an error by a judge has less of a chance of advancing an unworthy paper.
- Our model performs extremely well for the example posed in the original question ($P = 100$, $W = 3$, and $J = 8$) (see **Figure 2**). It fails only 0.08% of the time, while limiting the judges to 32 reads each (one-third of the total number of papers) and the total reads to 254.
- Most important, we would feel very comfortable using the model to judge our paper in the 1996 MCM. The model is fair. The top papers win virtually every time.

Weaknesses

- The model has definite bounds of effectiveness. We have set a bound on the number of papers that a judge can read at one sitting at 20. After the first round, judges read $2W$ papers each round. Thus, the number of winners must be less than or equal to 10. Allowing 2% of all papers to be winners, the total number of papers must be less than or equal to 500. A possible solution for large P is to break the contest into two halves of less than 500 each and run the model for each half.
- We had to model the human factor, which is difficult. The data on which we based the curve were our best estimate of human nature. We did not have data to consult to see how humans would actually perform under these circumstances. All of our testing and the validity of our results are based upon the assumption that our equation is actually representative of what occurs in the real world. If further research deems that equation to be inaccurate, it will be easy to adjust our model to a new equation.

Appendix A: Rationale and Proofs

There Are $n - 2$ Middle Rounds

In each round, exactly $W(S_r+1)$ papers advance into the next round, where $S_r + 1 = 2^{n-r}$. This is true because the number of stacks is halved with each

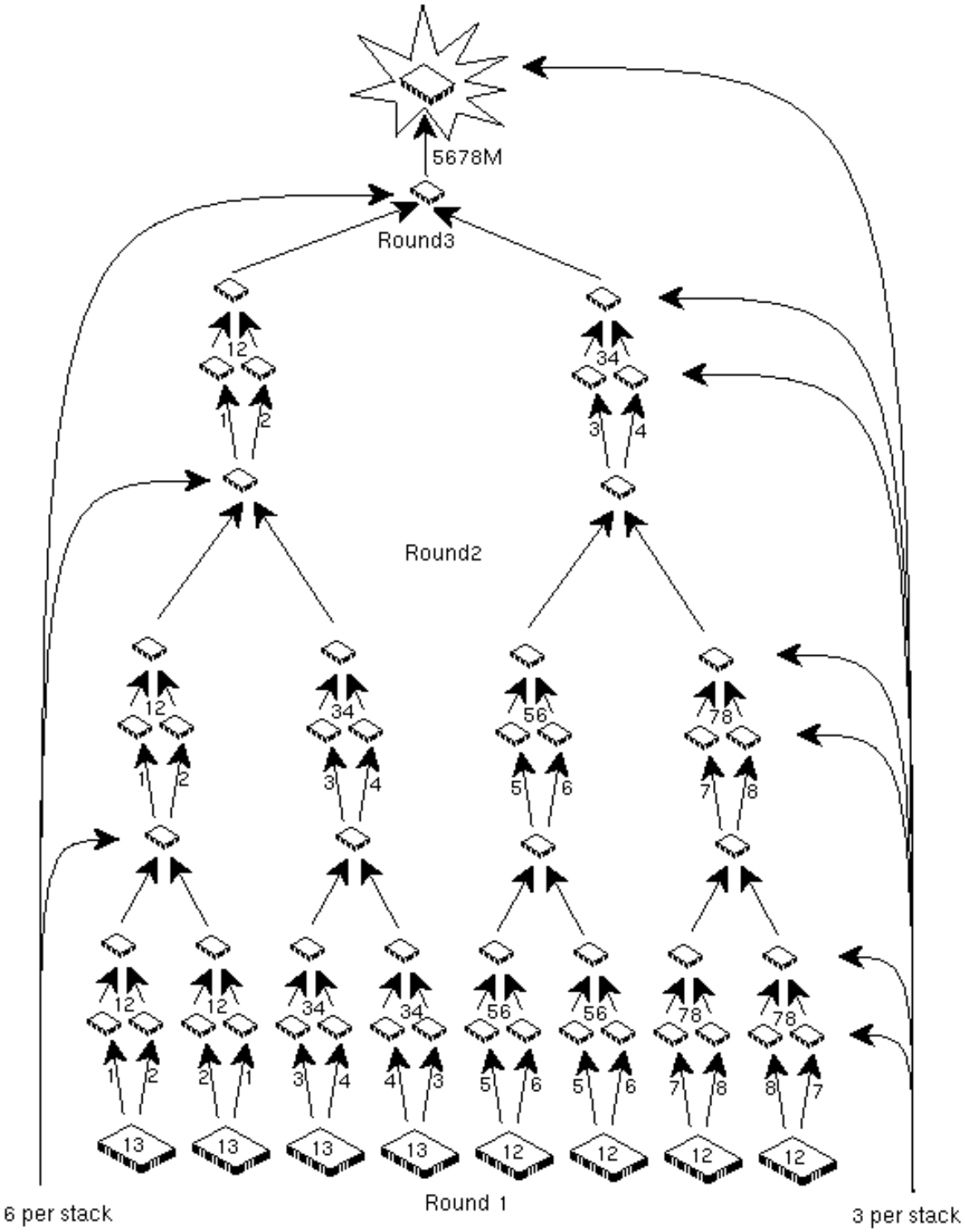

Figure 2. Diagram of the operation of the model for the original setting of $P = 100$, $W = 3$, and $J = 8$.

successive round, and there are W papers in each new stack. The final round begins when only $2W$ papers remain. We are in the final round when $2^{n-r} = 1$, or $2^n = 2^r$. So $n = r$ and the total number of rounds is n, including the first round and the final round. Hence there are $n - 2$ middle rounds.

Total Number of Reads

In R_1, every paper is read twice, yielding $2P$ reads. In the middle rounds, there are 2^{n-r+1} stacks of W papers each. Each stack is read twice, yielding $2^{n-r+2}W$ reads per round, for $n - 2$ rounds. Round R_n has 5 judges, each reading the final $2W$ papers. Thus, the sum of the reads for all rounds is

$$2P + \sum_{i=2}^{n-1} 2^{n-i+2}w + 5(2w).$$

Rationale for Optimal Number of Judges, J_O

We would like each judge to read approximately the same number of papers. This happens in the first round (some judges may read one additional paper if the stacks cannot be divided evenly). In each successive round, the number of papers read by a judge is $2W$. If the number of judges is $2n$, each judge is guaranteed exactly 2 rounds of judging. The number of judges needed is halved with each subsequent round. Thus, there will always be 4 judges for R_{n-1}, leaving 4 judges who have yet to judge a second time. These four judges plus the Head Judge make up the 5 who judge the final round. If J_O judges are used, each judge reads nearly the same number of papers, since every judge reads in the first round and one subsequent round.

Maximum Number of Reads for a Judge

If $J = J_O$, each judge reads in exactly 2 rounds: the first round, with approximately

$$2\left\lceil \frac{P}{2^n} \right\rceil$$

papers, and either a middle or a final round, with $2W$ papers. Thus, the maximum number of reads is

$$2\left\lceil \frac{P}{2^n} \right\rceil + 2W.$$

If $J > J_O$, some judges read in only one round. So long as $J < 2J_O$, at least one judge will read in two rounds, so the maximum number of reads is still be the same as above.

If $J < J_O$, some judges read more than one pair of stacks in the first round, possibly many pairs; the maximum number of reads could become very large.

Appendix B: The Computer Test Model

In using a computer to test a model, we must make some assumptions about human behavior that can be implemented by the computer. For the most part, our assumptions about human behavior are taken care of in the equation for the error factor. The other assumptions that we make are:

- Judges always require a third party to settle any disputes. It is very difficult to implement the power persuasion would have in a discussion between two judges.
- Judges are reasonable. If a judge has read two papers earlier and reads them again, they will again receive the same relative ranking, except possibly in the last round, which allows for more scrutiny to each paper.
- There is no Head Judge in the program. The person who settles any disputes is simply the next available judge who has never before compared the papers.
- The optimal number of judges is always used.

Judging a Mathematics Contest

Daniel A. Calderón Brennan
Philip J. Darcy
David T. Tascione
St. Bonaventure University
St. Bonaventure, NY 14778

Advisor: Albert G. White

Overview

Our model is based on breaking the problem down into four main areas and dealing with each: the distribution of papers among the judges, scoring methods, the number of papers to eliminate per round and the number of rounds, and the performance of the model with larger numbers of papers.

In each component, we focused on the goals of maintaining fairness and variety in all judging procedures, eliminating as many papers as possible in each round, minimizing the number of rounds, and, most important, seeing that no one goal was attained at the expense of any other.

The papers were coded and sent on from judge to judge without any prior knowledge of that particular participant or paper. There is no way to avoid having a judge read a paper twice, but a judge will not read a paper twice in a row until perhaps the final two rounds.

Assumptions

- Budget constraints affect only the number of judges.
- Time constraints affect only the number of papers that each judge can read.
- An approximate "absolute" ranking system exists among the judges; i.e., if every paper were to be scored or ranked by each judge, the results of each judge would generally agree with every other (allowing for a few places where consecutive papers may be "flip-flopped").
- All papers are eligible to win (none disqualified for cheating, missing sections, etc.).
- Judges need not be in the same location, but a copy of each paper (electronic or hard copy) is readily available to each judge.
- Judges remain ignorant of other judges' opinions on all papers.

- There is no way to avoid having a judge read a paper twice (nonconsecutively) during the reading process, but re-reading has no effect on judges' opinions of papers (i.e., a judge rates a paper as effectively on a second read as on the first).
- A minimum of 5 judges per 100 papers are needed.

Developing the Model

- Paper distribution:
  - Maintain judges' ignorance of other judges' opinions of papers.
  - Distribute papers so that the top $2W$ (6) are kept in competition throughout the contest.
  - Distribute papers such that no judge sees the same paper in consecutive rounds until the final two rounds (ensures efficiency in judging and fairness, as multiple opinions of papers are necessary).
  - Computer distribution both frees one human to judge instead of distributing papers and accomplishes the above tasks.
- Equating numerical judging systems:
  - Deals with the possibility of systematic bias in use of a numerical scoring scheme.
  - For each judge in the numerical scoring round, we adjust scores so that the highest score is equated to 100% and the others are adjusted proportionally; for example, scores of 92, 86, 89, 79 are adjusted to the same values divided by 92.
  - Given our assumption of an absolute ranking scheme and that a numerical scheme will closely follow, this method in essence allows us to put a numerical value on a judge's ranking so that it may be compared with ranks of other judges.
- Cuts at end of each round; number of rounds; total papers read per judge:
  - As few rounds as possible (time / budget constraints)
  - As few reads per judge as possible (time / budget constraints)
  - Keep the top $2W$ papers intact
- Cut approximate 44% of papers remaining in each round. When possible, leave a multiple of J intact, so that only in the first round do some judges read more than others. In the first round, do not cut to fewer than $2W + 1$ papers, in case the best $2W$ papers go to the same judge. After the first round, distribution of papers will prevent this problem from occurring again (fairness).

Methods for Distribution

The method for redistributing the papers after the first round is based upon a matrix created by the judges, each of whom enters in a column, from top to bottom, the numbers of the paper read, in decreasing order of rank. Then the first row consists of the highest-ranked papers of each judge.

Using falling diagonals of the matrix, it is possible to ensure that in the next round no judge receives any paper just read in the previous round. Further, it is possible to ensure that one judge does not receive all $2W$ top papers, which would force them to be cut the next round. The matrix below illustrates the method for redistributing the papers.

$$\begin{array}{c} \\ J_7 \\ J_6 \\ J_5 \\ J_4 \\ J_3 \\ J_2 \end{array} \begin{array}{c} \begin{array}{cccccccc} J_8 & J_1 & J_2 & J_3 & J_4 & J_5 & J_6 & J_7 \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & a_{67} & a_{68} \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & a_{77} & a_{78} \end{pmatrix} \end{array}$$

From the diagonals of the matrix that run downward from left to right, we get the assignments of papers to judges for the next round:

$$\begin{array}{lllllllll}
J_1: & a_{12}, & a_{23}, & a_{34}, & a_{45}, & a_{56}, & a_{67}, & a_{78} \\
J_2: & a_{13}, & a_{24}, & a_{35}, & a_{46}, & a_{57}, & a_{68}, & a_{71} \\
J_3: & a_{14}, & a_{25}, & a_{36}, & a_{47}, & a_{58}, & a_{61}, & a_{72} \\
J_4: & a_{15}, & a_{26}, & a_{37}, & a_{48}, & a_{52}, & a_{62}, & a_{73} \\
J_5: & a_{16}, & a_{27}, & a_{38}, & a_{41}, & a_{53}, & a_{63}, & a_{74} \\
J_6: & a_{17}, & a_{28}, & a_{31}, & a_{42}, & a_{54}, & a_{64}, & a_{75} \\
J_7: & a_{18}, & a_{21}, & a_{32}, & a_{43}, & a_{55}, & a_{65}, & a_{76} \\
J_8: & a_{11}, & a_{22}, & a_{33}, & a_{44}, & a_{56}, & a_{66}, & a_{77}
\end{array}$$

The Judges rank-order the newly distributed papers, and again a matrix is used to redistribute the papers that make the cut:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} & a_{38} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & a_{47} & a_{48} \end{pmatrix}$$

Again, diagonals are used to redistribute four papers to each judge.

$$
\begin{aligned}
J_1: &\quad a_{12}, \quad a_{23}, \quad a_{34}, \quad a_{45} \\
J_2: &\quad a_{13}, \quad a_{24}, \quad a_{35}, \quad a_{46} \\
J_3: &\quad a_{14}, \quad a_{25}, \quad a_{36}, \quad a_{47} \\
J_4: &\quad a_{15}, \quad a_{26}, \quad a_{37}, \quad a_{48} \\
J_5: &\quad a_{16}, \quad a_{27}, \quad a_{38}, \quad a_{41} \\
J_6: &\quad a_{17}, \quad a_{28}, \quad a_{31}, \quad a_{42} \\
J_7: &\quad a_{18}, \quad a_{21}, \quad a_{32}, \quad a_{43} \\
J_8: &\quad a_{11}, \quad a_{22}, \quad a_{33}, \quad a_{44}
\end{aligned}
$$

The next round of judging results in only two papers being passed by each judge. For this round, we convert the 2×2 matrix of the papers that pass this cut into a 4×4 matrix as follows:

$$
\left(\begin{array}{cccc|cccc} a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} & a_{18} & a_{11} \\ a_{23} & a_{24} & a_{25} & a_{26} & a_{27} & a_{28} & a_{21} & a_{22} \end{array}\right) \longrightarrow \begin{pmatrix} a_{12} & a_{13} & a_{14} & a_{15} \\ a_{16} & a_{17} & a_{18} & a_{11} \\ a_{23} & a_{24} & a_{25} & a_{26} \\ a_{27} & a_{28} & a_{21} & a_{22} \end{pmatrix}
$$

We split each row between columns 4 and 5 to create four rows of four elements. The papers are once again redistributed among the judges for numerical scoring. We must distribute each paper to two judges, according to our judging scheme, and we have devised a distribution method that ensures that no judge shares more than one paper to be read with any other (so that each paper may be numerically scored and therefore weighted against the widest possible range of papers allowable; in this case, each paper is scored against seven others). We do this first by adopting the diagonal distribution scheme that we have used throughout this problem and assigning the groups of four papers to four of the judges. Then, once we have distributed each paper once, we distribute the papers a second time by assigning the next four judges one column of the matrix as it stands above. In this case, we really cannot prevent a judge from reading a same paper as in the previous round, but the different scoring system calls for a more in-depth look at the papers anyway, and previous knowledge of a paper should not hinder the fairness of the distribution / scoring scheme.

In the final round, each judge reads the final eight papers and ranks them. The papers are then scored according to low-rank-sum method described in the **Judging Methods** section. The top three papers, the ones with the lowest rank sum, are the winners.

Number of Rounds, Papers Read per Judge, Paper Elimination

We must limit the number of rounds of judging and the number of papers read by each judge by eliminating the greatest possible number of papers per

round, while protecting papers that have an earnest chance of winning from being eliminated in a round in which too many papers are eliminated. This amounts to protecting the "best" $2W$ papers until they may compete against each other in the final round of judging.

In the first round, we call for elimination of only so many papers as to leave $2W + 1$ papers from each judge remaining in the competition. This ensures against the unlikely occurrence that a single judge receives the top $2W$ papers in the initial round. After this round, the methods of paper distribution prevent this from happening. A maximum of W papers may be given to a single judge in the second round, which calls for elimination of all but $W + 1$ papers; and in the third round, elimination of one-half of the remaining papers is allowable, since by this time the distribution scheme has spread out the $2W$ "best" papers enough to protect them from being eliminated.

Now that the papers have been thinned out, and the "cream" has been allowed to rise to the top, we may begin more forceful elimination measures. We make fourth round one in which the judges assign numerical grades to papers, with only the top $2W + 2$ papers surviving elimination The scoring procedure is discussed in the **Judging Methods** section; it suffices here to say here that the judging saves the top papers from being eliminated while allowing a drastic reduction in the number of papers remaining.

With only $2W + 2$ papers left in the running after the fourth round, we can have each judge read each of them and rank them in order, with the W papers ranked consistently high enough emerging as victors.

The judges have been subjected to the fewest number of rounds and the fewest possible number of papers to read.

Judging Methods

Our model uses two methods of judging, rank-ordering and numerical judging. Rank-ordering offers the advantage that there should not be any bias to interfere with the ranking. We base this on the assumption of an absolute ranking system, i.e., if the judges were to read all the papers, they would agree to an absolute ranking of the papers (with some reversals of consecutive papers). The ranking of any subset of papers must also conform to the absolute ranking, that is, the process of ranking must be order-preserving. The disadvantage with rank-ordering is that if a judge were to receive all the top papers, some of the top papers would get cut.

The other method, numerical grading, has the disadvantage of allowing for systematic bias. If a judge on average gives lower scores, our adjustment procedure compensates.

Since we use both methods of judging, we developed corrective measures for both to prevent these problems. The simpler of the two methods to correct is the rank-ordering method. We base the redistribution of papers that passed the cut on the rank that they had in the previous cut, ensuring that the best

$2W+1$ papers pass the first rank cut. After that cut, the redistribution prevents any judge from receiving a majority of the top $2W$ papers until the numerical grading round.

To prevent bias from removing one of the top $2W$ papers at the numerical round was a difficult task. We decided finally to base the numerical grading on a percentage curve based on the grades that the judges assign to all the papers. The top-scoring paper in a given judge's group determines the 100% paper, with the each of the remaining papers curved assigned its percentage of the top score. The distribution of the papers in the scoring round ensures that any two judges will have only one paper in common and that every paper is graded twice. Therefore every paper will have two "curved" percentage grades. These two grades are averaged together, yielding an overall weighted percentage ranking. The top $2W+2$ papers pass this cut in a further correction to ensure that the top $2W$ papers pass into the final round.

The eight papers in the final round are scored by using low-rank scoring. One of the eight judges assigns a number respectively to a paper, 1 (outstanding) through 8 (poor). At the end of the final round, the scores are added together and the paper with the lowest score wins.

Generalizing the Model

Since we have assumed that the ratio of judges to papers is greater than .05, we may test our model's growth rate, in terms of reads per judge and percentage of total papers read per judge, for larger ratios. We examined the changes as the number of papers doubles from 100 to 200, and then to 400, for ratios of .05 and .10. We assumed that each time P doubles, so does W.

Table 1 gives the details of our analysis, for a variety of different situations. For a ratio of .10, we see a tremendous increase in efficiency over the situation for a ratio of .05! Each judge reads about one-third of the total papers instead of more than half.

Demonstration of the Model

[EDITOR'S NOTE: At this point the authors illustrate in detail the application of their model to two sets of data devised by them, each with $P = 100$, $J = 8$, and $W = 3$. In the first example, the top three papers in fact win. The second example shows what happens when one judge receives all $2W$ top papers in the first round: The top three papers still win. For reasons of space, we omit these extended examples.]

Table 1.
Analysis of contest situations for ratios of judges to papers of .05 and .10.

Ratio	P	J	W	Round	Papers	Papers/judge	Method	Eliminated/judge
.05	100	5	3	1	100	20	rank	8
				2	60	12	rank	4
				3	40	8	rank	4
				4	20	8	score	bottom 14 averages
				5	8	8	rank	5 highest rank sums
						Total: 56 (56%)		
	200	10	6	1	200	20	rank	8
				2	120	12	rank	4
				3	80	8	rank	4
				4	40	8	score	bottom 26 averages
				5	14	14	rank	8 highest rank sums
						Total: 62 (31%)		
	400	5	12	1	400	20	rank	8
				2	240	12	rank	4
				3	160	8	rank	4
				4	80	8	score	bottom 54 averages
				5	26	26	rank	14 highest rank sums
						Total: 74 (18.5%)		
.10	100	10	3	1	100	10	rank	4
				2	60	6	rank	2
				3	40	4	rank	2
				4	20	4	score	bottom 12 averages
				5	8	8	rank	5 highest rank sums
						Total: 32 (32%)		
	200	20	6	1	200	10	rank	4
				2	120	6	rank	2
				3	80	4	rank	2
				4	40	4	score	bottom 26 averages
				5	14	14	rank	6 highest rank sums
						Total: 38 (19%)		
	400	40	12	1	400	10	rank	4
				2	240	6	rank	2
				3	160	4	rank	2
				4	80	4	score	bottom 54 averages
				5	26	14	rank	14 highest rank sums
						Total: 50 (12.5%)		

Strengths and Weaknesses

Strengths

- Paper distribution: The distribution method keeps the required $2W$ papers in circulation while eliminating the maximum number of papers at each round. It also allows for a wide variety of judges' opinions to be given on the top $2W$ papers.
- Judging methods: The numerical grades overcome systematic bias by the judges. The rank-ordering allows the maximum number of papers to be culled in early rounds while preserving the top $2W$ papers.
- Elimination methods: We minimize the number of rounds and the number of papers judges must read, while maintaining fairness.
- Growth rate: For a given ratio of papers to judges, as the number of papers increases, the percentage of papers read per judge diminishes rapidly.

Weaknesses

- Paper distribution: In the worst-case scenario, we cannot prevent a single judge from receiving all of the top P/J papers and eliminating the bottom portion of those.
- Judging methods: The numerical grading adjustments requires redundant readings, and extra time is consumed in applying the adjustments. Further, a less-qualified paper may receive a higher weighted average than a paper with a higher absolute ranking, if the lower-ranked paper is judged against papers with even lower ranks and the higher-ranked paper is judged against papers with even higher ranks.
- Elimination methods: Entering the final round, too few papers may be eliminated. This apparent inefficiency is necessary to carry the top $2W$ papers into the final round.
- Growth rate: With a low judge-to-total-papers ratio and a relatively low number of papers, the percentage of total papers read by each judge is quite high. The total number of papers read per judge increases even as the percentage drops.

References

Decker, Rich and Stuart Hirschfield. 1995. *The Object Concept: An Introduction to Computer Programming Using C++*. Boston, MA: PWS Publishing.

The selection sort that we used comes from this book.

Select the Winners Fast

Haitao Wang
Chunfeng Huang
Hongling Rao
Center for Astrophysics
University of Science and Technology of China
Hefei, China

Advisor: Qingjuan Yu

Summary

Assuming that judges are ideal, we provide a model to determine the top W papers in almost the shortest time. We use a matrix record the orderings that we get from judges, and we reject as many papers as possible after each round.

We then consider real-life judges and estimate the probability that the final W papers contain a paper not among the best $2W$ papers.

Furthermore, considering the possibility of systematic bias in a scoring scheme, we improve the model by using a Bayesian estimation method, which makes it possible to some extent to compare different judges' scores.

We performed many computer simulations to test the feasibility of our model. We find that our model would be improved by increasing the number of papers selected from the first round. We also made a stability analysis by altering P, J, W and got an empirical formula to predicts the total time of judging.

We used data from real life to test our model and got a perfect result: For $P = 50$, $J = 3$, and $W = 2$, we get the first- and third-best papers with our scheme; with $W = 3$, we got the top three papers.

We conclude by summarizing a practicable and flexible scheme and offering some suggestions, and estimating the budget with an empirical formula.

Assumptions

- The judges are equal. None is more authoritative than the others.
- When a judge is evaluating a paper, the judging result is not influenced by adventitious factors, such as taking bribes.
- The time that a judge takes is proportional to the number of papers to read.

- There exists an objective criterion with which we can tell which of two papers is "better." Therefore, we can use an absolute rank-order or absolute scores to describe the quality of the papers measured by the criterion.
- The absolute rank-order is transitive: If A is better than B and B is better than C, we can say A is better than C.

Analysis of Problem

Our primary goal is to include the top W papers among the "best" $2W$ papers.

A subsidiary goal is that each judge read the fewest possible number of papers. We interpret this goal into two points:

- It is the duration of the whole judging process, the total time for all rounds, that is constrained by funding. If the time for a round is how long it takes the judge who has the most papers to read, it is wise to distribute the papers to the judges as evenly as possible in each round.
- We want to get as much information as possible.

The two usual methods of judging are rank-ordering and numerical scoring. Systematic bias is possible in a scoring scheme, that is, each judge may have a subjective tendency in scoring, which results in incomparability among scores given by different judges. However, it is reasonable to believe that the scores that the same judge gives to different papers are comparable, even if they are obtained in different rounds. Therefore, compared with a rank-ordering method, scoring is a more meaningful way to record the results for papers judged in earlier rounds. We use a scoring scheme instead of a rank-ordering scheme in our later model, so a paper need not be read more than once by the same judge. Note that we do not compare the scores of different judges directly, that is, we mainly use scores to obtain a rank-ordering.

We first consider the simplified problem with the significant assumption that the ordering from each judge's evaluation coincides with the absolute ordering. In this event, we can definitely find the best W papers. Furthermore, we can optimally adjust the allocation of papers in every round to get an efficient scheme.

But judges in real life cannot rate the papers with perfect precision. For example, a paper with absolute rank 7 (we denote it $P_{(7)}$) may get a higher score than $P_{(6)}$ from a judge. We call that *misjudgment*. Misjudgments prevent us from getting the best W papers, so their effect must be taken into account.

There are also subjective differences among the scorings of different judges. For example, for the same two papers, one judge may give 80 and 83, while another gives 65 and 72. If we know the distribution of each judge's scores, we can to some extent compare scores given by different judges. The real distribution for each judge is unknown, so we have to use estimates.

Table 1.
Notation.

Symbol	Meaning
P	total number of papers
W	number of winners
J	number of judges
T	total judging time (or number of papers that can be judged in the time)
P_i	paper i
$P_{(i)}$	paper with the absolute rank of i
S_i	the absolute score for paper i
$P_i > P_j$	paper i is better than paper j in absolute rank-order
$P_i(A) > P_j(A)$	paper i is better than paper j in judge A's opinion
R_i	number of papers currently known to be better than P_i
ORD	matrix of currently known relations between pairs of papers
$\lceil x \rceil$	the smallest integer not less than x
$N(\mu_0, \sigma_0^2)$	normal distribution with mean μ_0 and standard deviation σ_0
σ_1	standard deviation of the judges' scoring
μ_j, σ_j	mean and standard error of judge j's scoring
$\hat{\mu}_j, \hat{\sigma}_j$	estimated values of μ_j, σ_j
P_{error}	probability of error occurring

Design of the Model

Top W in the Least Time

Ideally, the ordering in each judge's opinion coincides with the absolute ordering, expressed mathematically by

$$P_i(A) > P_j(A) \Leftrightarrow P_i > P_j.$$

So, based on the transitivity of the absolute score, if $P_i(A) > P_j(A)$ and $P_j(B) > Pk_(B)$, we can say that $P_i > P_k$.

To find the top W as soon as possible, as many papers as possible should be rejected after each round. So if there are W papers or more in the current paper pool that are better than P_i , reject P_i.

In the first round, P papers are dispatched to J judges evenly. After performing the above rejection rule, each judge selects W papers. In later rounds, how do we dispatch the remaining $W \cdot J$ papers to judges to obtain the greatest number of new orderings from each round?

Let us consider the simple case in which $W = 2$, with $P_1 > P_2$ and $P_3 > P_4$ known after the first round. If we compare P_2 with P_4 (or P_1 with P_3), no matter what the result is, we can always gain an extra relation (if, say, $P_2 > P_4$, the extra relation is $P_1 > P_4$). If we compare P_1 with P_4 (or P_2 with P_3), on some occasions no extra relations can be obtained. A similar result holds for $W = 3$.

This example indicates a fact: If we use *current rank* R_i to denote the number of papers known to be better than P_i from current known information, we should try to distribute the papers with close current rank to the same judge in order to get more relations in a round.

Use matrix ORD to describe known orders. Define ORD_{ij} by

$$\text{ORD}_{ij} = \begin{cases} 1, & \text{if } P_i > P_j; \\ -1, & \text{if } P_j > P_i; \\ 0, & \text{if } P_j = P_i; \\ \infty, & \text{if } P_i \text{ and } P_j \text{ have not been compared by any judge.} \end{cases}$$

At the beginning of a round, we dispatch papers to judges and judges give each paper a score. We then find every P_i and P_j that have scores from the same judge in the finished rounds and fill ORD_{ij} and ORD_{ji}. We replace ORD with its transitive closure [Wang 1986], which, put simply, adds all the indirect order gained from ORD_{ij} into the matrix ORD. At the end of each round, for each paper P_i, calculate R_i from ORD and reject P_i if $R_i \geq W$. Repeat the above process until the final W papers are left.

Consider Misjudgment

By *misjudgment*, we mean that the final W papers are not the best W. If the final W papers contain a paper not among the best $2W$, an *error* occurs.

Assume that for a paper with an absolute score of μ_1, the score given by a certain judge is a random number following a normal distribution $N(\mu_1, \sigma_1^2)$. The standard deviation σ_1 is the parameter that describes the degree of precision in a measurement. Misjudgment originates from the deviation of a judge's scoring from the absolute score, as **Figure 1** shows. The shaded area in **Figure 1** shows the misjudgment area.

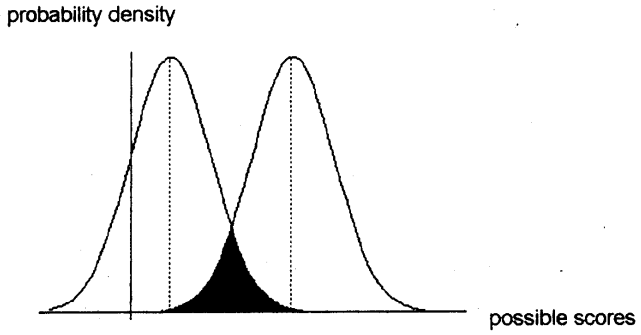

Figure 1. Possible score distributions for two papers.

There must be a distribution of the absolute scores of all the papers. We assume that it is a normal distribution $N(\mu_0, \sigma_0^2)$, so that the ratio σ_1/σ_0 reflects the judge's ability to distinguish the quality of these papers and also determines the probability of misjudgment. Using the basic model, given P, J, W, and σ_1/σ_0, we can estimate the probability of error (P_{error}). If the probability is small enough, we can expect the model to provide the desired result.

Taking the random feature of scoring into account, some conflict is likely to happen, such as $\text{ORD}_{ij} = 1$ but judge A's scores show $P_j(A) > P_i(A)$. One

way to solve the conflict is to find all judges who have read both P_i and P_j, sum up the scores given to P_i and P_j by these judges, and determine a new ORD_{ij} by comparing the two sums.

Systematic Bias Among Judges

Considering differences among the scoring tendencies of different judges (systematic biases), it is undesirable that each judge select out the same number of papers in the first round, for then it will be more likely that excellent papers will be rejected in the first round.

Instead, when the first round of judging is over, we input the scores of each group of papers into computer, which gives the estimate of each judge's parameters (mean score and standard deviation) and computes each group of papers' score threshold for rejecting papers corresponding to a certain absolute level. This way, excellent papers have less possibility of being rejected in the first round. Estimating the parameters of all the judges enables us to compare the scores from different judges to some extent.

We use Bayesian estimation [Box and Tiao 1973] to determine the estimate of judge j's parameters (μ_j, σ_j). Suppose that judge j gives scores $S_1, \ldots, S_n$ to papers $P_1, \ldots, P_n$. We use the method of maximum likelihood to estimate σ_j:

$$\hat{\sigma}_j = \frac{1}{n} \sum [S - E(S_i)]^2.$$

We then use Bayes's method to estimate μ_j. In reality, we may have a priori knowledge of each judge's scoring tendency. Even if not, we still have reason to assume an a priori distribution of each judge's score. If the prior parameters are (μ_0, σ_0^2), then the posterior parameter is

$$\hat{\mu}_j = \frac{n \cdot E(S)}{n + \left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2} + \frac{\left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2 \cdot \mu_0}{n + \left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2}.$$

Then we can use $\text{quantile}(N(\hat{\mu}_j, \hat{\sigma}_j^2), \text{LEVEL})$ as the score threshold, where $1-\text{LEVEL}$ is the expected proportion of papers should be retained. One suitable value is

$$\text{LEVEL} = 1 - \frac{W \cdot J}{P}.$$

Test of the Model

The most important test is to verify that the model makes sense. We do a computer simulation to see how our model behaves as the two practical factors are gradually taken into account. As detailed later, our results agree with our expectations (**Feasibility Test**). In addition, a finding in the course of testing

leads us to make some improvement to the model. A more thorough test is made by revising the parameters (**Stability Test**). Lastly, we try to apply our model to more complicated example in real world. The results meet reality very well (**A Real-Life Example**).

Feasibility Test

We fix $P = 100$, $J = 8$, and $W = 3$.

Test of Basic Model

We assign P papers absolute scores of $1, 2, \ldots, 100$. These values are used only to provide the relative order of the papers.

All papers are randomly allocated to 8 judges at the beginning of the simulation. We calculate the total judging time.

The results of 1,000 iterations of simulation (see **Figure 2**) show that the basic model can select the top three in quite a short time, as we analyzed before.

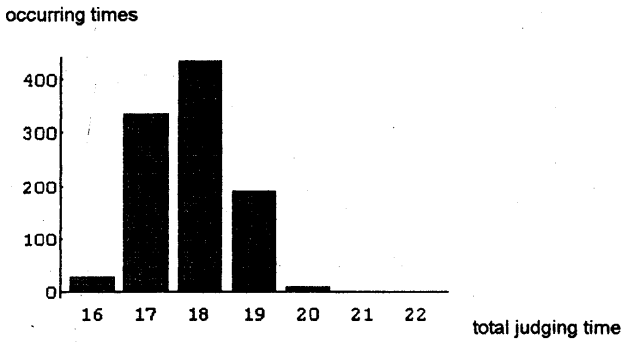

Figure 2. Frequency vs. total judging time.

Take Misjudgment into Account

These two simulations are vital:

- *Simulating the distribution of the absolute score.* Generally, we have reason to use a normal distribution. In order to assign the scores of 100 papers, we generate 100 random numbers following $N(60, 30^2)$, truncated at 0 and 100.
- *Simulating the score given to one paper.* We simulate a judge's scoring by adding a normal random number to the absolute score of the paper.

The quantity σ_1/σ_0 should be fairly small (say, ≤ 0.1), because a judge should have good competence in judgment. We take σ_1/σ_0 = $1/30$, $2/30$, and $3/30$ as cases in our simulation.

We also did a theoretical estimate for these cases of P_{error} under the worst of circumstances. Take $\sigma_1/\sigma_0 = 3/30$ for example. An error occurs when one or

more of $P_{(7)}, P_{(8)}, \ldots$ enter the final three. The probability of $P_{(7)}$ entering the final three contributes the most to P_{error}. We let $P_{\text{error}}(i,j)$ be the probability of misjudging papers i and j. We approximate P_{error} by the probability of $P_{(7)}$ entering the final three:

$$\begin{aligned} P_{\text{error}} &\approx \frac{1}{8}P_{\text{error}}(3,7) + \left(\frac{1}{8}\right)^2 \sum P_{\text{error}}(3,i)P_{\text{error}}(i,7) + \cdots \\ &\approx \frac{1}{8}P_{\text{error}}(3,7) + \left(\frac{1}{8}\right)^2 \sum_{i=4}^{6} P_{\text{error}}(3,i)P_{\text{error}}(i,7). \end{aligned}$$

We computed $P_{\text{error}}(i,j)$ using Mathematica by calculating the area of the shaded region in **Figure 1**. In this way, we get the estimate $P_{\text{error}} \approx 0.4\%$.

The results of the simulations accord with the theoretical estimate (see **Table 2**).

Table 2.

Results of 1,000 trials for each value of σ_1/σ_0 vs. theoretical estimates.

$P = 100$, $J = 8$, $W = 3$

σ_1/σ_0	Mean T	Max T	Errors	Observed P_{error}	Estimate of P_{error}
1/30	17.7	21	0	.000	10^{-7}
2/30	17.7	21	0	.000	.0006
3/30	17.8	21	4	.004	.004

An Extra Improvement to the Model

The simulation results demonstrate that the model behaves reasonably so far. Surprisingly, a slight modification improves the model remarkably. If in the first round we select more papers, say W_1 instead of W, and select W papers from the next round, we find that P_{error} declines greatly but the total judging time is scarcely affected. The chances of a excellent paper being rejected in the first round are much more than in the later round, because the papers rejected after the first round are read by only one judge, while those rejected later are read by more judges. **Table 3** gives simulation results for several values of W_1.

Table 3.

Results of 1,000 trials for each value of W_1.

$P = 100$, $J = 8$, $W = 3$

W_1	Mean T	Max T	Errors
6	17.9 ± 0.9	20	0
5	17.8 ± 0.8	20	0
3	17.8 ± 0.8	20	1

Take Judges' Systematic Biases into Account

We might as well simulate the scores from different judges by using the normal distribution with randomly generated mean value and variance.

Using the method offered in **Design of Model**, we get the results of **Table 4**. With increasing LEVEL, the total judging time declines but more errors occur; it is difficult to minimize both time and number of errors.

Table 4.

Results of 1,000 trials for each value of LEVEL.

$P = 100$, $J = 8$, $W = 3$

LEVEL	Mean T	Max T	Errors
50%	22.0	24	0
70%	19.5	23	3
75%	18.7	21	4
80%	18.0	21	4
85%	17.2	20	7
90%	16.3	19	11

Stability Test

We change the parameters P, J, W to test the model's stability. **Table 5** gives the results of 100 iterations for each of several groups of parameters.

Table 5.

Results of 100 trials, for $\sigma_1/\sigma_0 = 3/30$, for each combination of values of P, J, and W.

P	J	W	Mean T	Max T	Errors	$\lceil P/J \rceil + W + 2$
50	4	4	18.1 ± 1.2	22	0	19
80	8	3	14.9 ± 0.9	17	0	15
100	7	3	19.5 ± 0.7	21	0	20
	8	3	17.7 ± 0.8	20	0	18
	8	4	18.8 ± 0.8	21	0	19
		5	19.8 ± 0.9	22	1	20
	10	3	15.4 ± 0.9	18	1	15
120	8	3	19.8 ± 0.9	22	0	20
140	8	3	23.0 ± 1.0	27	1	23
	13	1	14.4 ± 0.7	16	3	14
		2	15.8 ± 0.8	18	1	15
		3	16.9 ± 0.8	19	0	16
		5	18.2 ± 0.9	21	0	18

Analyzing these data, we discover an empirical formula

$$\left\lceil \frac{P}{J} \right\rceil + W + 2,$$

which fits the data for the average value of T wonderfully. Another finding is that small W/P will cause considerable P_{error}. So when W/P is too small (say $\leq 1/100$), the model does not work well. But properly reducing the number of papers rejected in each round will reduce P_{error}.

A Real-Life Example

One idea for testing our model would be to use data from a tennis competition. The table of international standings can be treated as the absolute order, and the result of each formal match acts as a "judge." It is a pity that we have no data!

So we use a substitute for the data of a real competition. We obtained from our department real scores of 50 students for three semesters taking the same three courses. We consider that the sum of each student's three scores stands for the student's level in this major; we take it as an absolute score, and this gives the absolute ordering. In any one semester, the order of students' scores can differ from the absolute order. So we can use these data to simulate a contest, in which each semester acts as a judge assigning a score.

We use these data in our computer program and get the results of **Table 6**.

Table 6.
Results of analysis of departmental data ($P = 50$, $J = 3$).

W	T	papers selected
2	19	$P_{(1)}, P_{(3)}$
3	22	$P_{(1)}, P_{(2)}, P_{(3)}$

Generalization

How to Budget?

Funding for the contest constrains both the number of judges that can be obtained and the amount of time that they can judge.

Assume that each judge can mark n papers / day and the judge's salary is \$$s$ / day. We can hypothesize that funding f is a function of T, n, and J: $f = f(T, n, J)$, where T is total judging time. Obviously, $\partial f/\partial J > 0$, $\partial f/\partial T > 0$, $\partial f/\partial n < 0$, and $T = T(J)$. Fortunately, we have the an empirical formula

$$\left\lceil \frac{P}{J} \right\rceil + W + 2.$$

A reasonable functional form for f is

$$f = \left\lceil \frac{T}{n} \right\rceil \cdot J \cdot s = J \cdot s \cdot \left\lceil \frac{\left\lceil \frac{P}{J} \right\rceil + W + 2}{n} \right\rceil.$$

Since $k \le \lceil k \rceil \le k + 1$, we get

$$\frac{(P + (W + 2) \cdot J) \cdot s}{n} \le f \le \frac{(P + (W + 3 + n) \cdot J) \cdot s}{n},$$

which allows us to budget for the contest if the number of the judges has been given (see **Table 7**).

Table 7.

Cost of the contest, for various combinations of papers per day per judge and number of judges.

n	J	$\min f$	$\max f$
15	8	\$3,080	\$6,067
20	8	\$2,310	\$5,250
15	7	\$2,987	\$5,600
20	7	\$2,240	\$4,812

On the other hand, we can turn the equation around into the form

$$\frac{f \cdot \frac{n}{s} - P}{W + 3 + n} \le J \le \frac{f \cdot \frac{n}{s} - P}{W + 2}.$$

According to this, if funding is known, we can decide the number of judges (see **Table 8**). Of course, these are rough estimates.

Table 8.

Number of judges that can be hired, for various combinations of number of papers per day per judge and budget.

$P = 100$, $s = 350$, $W = 3$

n	f	$\min J$	$\max J$
15	\$5,000	6	22
10		3	8
15	\$7,000	10	40
10		7	20

Applying the Model to Different Kinds of Competition

For contests that give awards to just a few winners, our model is an effective and rational scheme. For contests that give various awards at different levels, we can modify a few parameters in our model. There are two methods.

- Method 1: Suppose that the contest committee expects to classify the participants into different levels in some given proportions, say four levels of 5%, 10%, 35%, and 50%, similar to the MCM. In the first round, we reject 50% as Successful Participation; reject 35% in the 2nd round as Honorable Mention; reject 10% in the third round as Meritorious; the remaining 5% are Outstanding.
- Method 2: We set the value of LEVEL as needed in each round to distinguish participants of different levels. This method is more flexible and fairer than Method 1.

Final Scheme

We summarize our final scheme:

- Divide the judging process into several screening rounds and follow the principles below in each round until W papers remain.
- Use a scoring scheme.
- Do not compare the scores from different judges.
- In the first round, distribute papers to all judges evenly. After scoring, select out the top $2W$ papers in each group to enter the next round.
- At the end of each round, for each paper, calculate the number of papers better than it (which we call the current rank of the paper), then reject every paper whose current rank is more than $W - 1$.
- At the beginning of each round, dispatch papers with a close current rank to the same judge, if possible.
- The number of papers distributed to each judge in each round should be as equal as possible.

Our Suggestions

- Properly reducing the number of rejected papers in the first round would decrease the error probability.
- Altering the number rejected in each round as needed is helpful in competitions that determine different levels of the participants.
- To be more practical and efficient, we suggest prejudging the papers at first, that is, rejecting the papers of distinctly poor quality.

- Between rounds, have some discussion among judges so that they gain some knowledge of the levels of papers as a whole. Such a feedback mechanism surely helps reduce the standard deviation of judgment.
- When there are about $2W$ papers left, all the judges gather to read the remaining papers together, if time permits, to select the top W papers.

Strengths and Weaknesses

Strengths

- We have shown how our model provides an efficient scheme in correct selecting winners. The model was not only tested in a computer simulation but also proved adaptable to real cases.
- Our model is also very stable. All parameters, which we set arbitrarily, can be changed without changing the quality of the model.
- We obtain from our model an empirical formula for T, based on P, J, and W.
- We use Bayesian estimation to take into account the differences among judges.
- Our model is flexible enough to be applied to different kinds of competitions.

Weaknesses

- We are unable to demonstrate that our model is optimal.
- We would like to be able to improve our estimation of the parameters for each judge.

References

Box, G.E.P. and Tiao, G.C. 1973. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Wang, Yihe. 1986. *Introduction to Discrete Mathematics*. Harbin, China: Harbin Institute of Technology Press.

The Inconsistent Judge

Dan Scholz
Jade Vinson
Derek Oliver Zaba
Dept. of Systems Science and Mathematics
Washington University
St. Louis, MO 63130

Advisor: Hiro Mukai

Summary

We provide a judging process that is robust enough to ensure that the intrinsically best papers are chosen in spite of randomness and subjectivity in the judging process. We increase the scope of the problem by introducing inconsistency into the judging process. We model this inconsistency by expressing the actual paper score as the sum of the intrinsic numerical score, the overall bias of the judge, and an error term:

$$S_{jp} = S_p + B_j + \epsilon_{jp}.$$

We use an iterative computer-guided process to determine the judging procedure. After each round of judging, the computer program uses bias estimates to calculate confidence intervals for the intrinsic score of each paper. These confidence intervals are used to reject as many papers as possible while guaranteeing, within a specified level of confidence, that the top W papers advance to the next round. Less cautious rejection criteria in each round adapt the method to select the winning papers from among the top $2W$ papers.

We did a computer simulation over a range of values for the parameters. Intrinsic scores were normally distributed with mean 50 and standard deviation 20; bias and consistency parameters were varied. We compare the method results to the intrinsic scores of the papers. For $P = 100$, $J = 8$, $W = 3$, the method proved correct 95% of the time with an average of 175 papers read.

Assumptions

Given that papers have an absolute intrinsic ranking, we assume that papers also have an associated intrinsic numerical score. The score that a judge gives a paper reflects not only the intrinsic score of the paper and the overall bias of the judge but also the inconsistency of the judge. This assumption is more realistic than assuming that all judges would agree to an absolute ranking and will produce a more robust judging procedure. We assume:

- Papers have intrinsic scores which follow a normal distribution.
- Judges have constant numerical bias.
- The range of biases for all judges follows a normal distribution.
- Judges' inconsistency follows a normal distribution.

The normal distribution is used for analytical convenience and is justified by historical precedent [DeGroot 1986, 263–264].

The Model

We express our assumptions mathematically by equating the score S_{jp} that judge j assigns paper P to the sum of the intrinsic score S_p of the paper, a bias term B_j for the judge, and an error term ϵ_{jp} for the score:

$$S_{jp} = S_p + B_j + \epsilon_{jp}.$$

Our model has parameters μ, σ, B, and Δ. The distribution of intrinsic scores is parameterized by μ and σ. That is, S_p is a random variable with distribution $N(\mu, \sigma^2)$. Parameter B is a measure for the bias of all the judges; the bias B_j for a particular judge comes from the distribution $N(0, \sigma^2)$. The parameter Δ is measure for the overall consistency of the judges. The error for an individual grading, ϵ_{jp}, comes from the distribution $N(0, \Delta^2)$. The terms B_j and ϵ_{jp} account for the subjective nature of the judging process.

The Method

Our model estimates the intrinsic scores of papers by producing estimates for the bias of the judges and adjusting their scores accordingly. Our confidence in the estimated intrinsic scores is used to reject as many papers as possible while maintaining that the probability of rejecting one of the top W papers is less than a predetermined α. The method proceeds as follows:

Distribution of Papers

Our model distributes papers according to the following prioritized criteria:

- No judge reads the same paper more than once.
- The numbers of papers read by each judge for a given round do not differ by more than one. This minimizes time spent reading for a round.
- Workload is distributed equally among the judges.

Estimation of the Intrinsic Score Distribution

Since the distribution and values of the actual intrinsic scores are not known, we attempt to estimate them. We estimate the mean and variance of the intrinsic scores after the first round. The mean and variance are estimated by the following (see **Appendix A**):

$$\hat{\mu} = \overline{S}_{jp}, \qquad \hat{\sigma^2} = \frac{1}{J}\sum_j \frac{1}{P(j)-1}\sum \left(X_j - \mu_j\right)^2,$$

where the $P(j)$ denotes the number of papers judge j has read.

Calculation of Bias

After round one, each judge will have scored approximately P/J papers. If the average of the scores for a given judge is significantly greater than the mean of the scores, either the judge is positively biased or the judge happened to receive a sample of unusually good papers, or both. If $X_1, \ldots, X_n$ are the scores of papers reas by a particular judge, then the conditional distribution for B_j after round one (see **Appendix B**) is normal with mean and variance given by

$$B_j^1 = \frac{\sum(X_i - \mu)}{n + \frac{\sigma^2+\Delta^2}{B^2}}, \qquad V_j^1 = \frac{1}{\frac{1}{B^2} + \frac{n}{\sigma^2+\Delta^2}}.$$

Note that in the special case when $B = 0$, the estimate for B_j is also zero, but if B is large, the distribution has mean approximately $\overline{X}$ and variance σ^2/n.

Recalculation of the Bias

If the judges are unusually consistent, i.e., if Δ is very small, we would like our judging procedure to recognize and take advantage of this fact. In the extreme case, when $\Delta = 0$, we can precisely rank all P papers with only $P + J - 1$ readings: Start by dividing the papers evenly in the first round; in the second round, judge A retires while each of the other judges reads one of judge A's papers; by subtracting out each judge's bias relative to judge A, we learn the precise ranking of the papers.

The method optimized for the trivial case above is successful because it uses the fact that $\Delta = 0$ to calculate the biases exactly after the second round. We could adapt this simple example to improve our judging procedure. In the first round, the biases are estimated according to the preceding section. These are used for the first cut. In subsequent rounds, first re-estimate Δ. Using the new value, re-estimate the biases B_j and their variances V_j of our estimates; if the new value of Δ is small, so is the uncertainty of our bias estimate. The combination of a small inconsistency Δ and accurate knowledge of the biases would allow us to calculate an estimated intrinsic score more

accurately. With sharpened values of the estimated intrinsic score, more papers could confidently be eliminated after each round. We derive and present the formulas for this bias re-estimation in **Appendix B**.

Estimation of Intrinsic Scores

The estimated bias for each judge is used to calculate for each paper p a net score that estimates the intrinsic score after taking into account the bias of the judges and the number of readings. The mean and variance for the net score are (see **Appendix B**)

$$\text{mean} = \frac{\frac{\mu}{\sigma^2} + \sum_{j \text{ judges } p} \frac{S_{jp} - B_j}{V+\Delta^2}}{\frac{1}{\sigma^2} + (\#\text{ readings})\frac{1}{V+\Delta^2}}, \qquad \text{variance} = \frac{1}{\frac{1}{\sigma^2} + (\#\text{ readings})\frac{1}{V+\Delta^2}}.$$

Here $V = \max V_j$ is used instead of V_j to simplify forthcoming calculations.

Rejection of Papers

At the end of each round, we seek to eliminate as many papers as possible while still ensuring that the best W papers are selected within a specified degree of confidence. **Appendix D** derives the inequality

$$\Pr(\text{mistaken rejection}) < W \cdot \sum_{p=1}^{R} \Phi\left(\frac{Sj,p - S_{j,p-w+1}}{\sqrt{2}\sigma_T}\right).$$

The variable S_{jp} reflects the computed score of paper p in ascending order, $S_{j,p-w+1}$ is the score of the paper whose score is ranked w, and σ_T^2 is the total variance of the score distributions. This inequality lends itself to an iterative process in which the lowest papers are rejected one by one until the inequality reaches a desired level of confidence. If there still remain more than W papers after the confidence level is reached, a new round is initiated. This iterative process involves repeating the earlier steps of this section.

Model Implementation

We simulated the model with a C++ program to demonstrate its validity and scope. The simulation compares the actual $2W$ intrinsic score winners to the W model-determined winners. Due to time constraints, the re-estimation of the bias was omitted from the simulation.

Initialization

The simulation assumes that there are 100 papers, 8 judges, and 3 winning papers. It generates absolute intrinsic scores from a normal distribution with mean 50 and standard deviation 20. The parameters B and Δ, which determine the generation of judges' parameters, are varied over a realistic range, empirically determined to be between 5 and 10. For all simulation calculations, the judging process assumes $B = 8$ and $\Delta = 8$, in order to demonstrate the validity of the model with no knowledge of the distributions of the biases and inconsistencies.

Simulation

We ran the program for 1,000 competitions with various levels of confidence per round and distributions of scores. The model was successful approximately 93–97% of the time with 160–190 readings. Due to the slack in the confidence inequalities, a strict lower bound of 0.7 confidence per round produced these encouraging results while significantly reducing the total number of readings.

Real-World Implementation

A fully implemented computer program would allow a judging team to input the number of papers to be evaluated, the number of winning papers, and the number of judges. The program would have the judges input the scores for each paper that they judged in round one. The program would then ask for a degree of confidence that the winning papers will be drawn from the top $2W$ papers. Output is the designation of the papers that are to be advanced to the next round. The judges now enter their scores for round two, and the process is repeated until W papers remain.

Stability

The formulas used thus far have relied upon exact values for parameters μ and Δ for the distribution of intrinsic scores to greatly simplify calculations. This information, however, would not be available in actual implementation of our judging process. Fortunately, small inaccuracies in the calculated values of $\hat{\mu}$ and $\hat{\sigma}$ do not undermine the validity of our judging process.

Strengths and Weaknesses

Our model provides a great deal of flexibility for variations in the judging procedure. We do not assume that the judges will agree in absolute ranking for a given competition. We are able to do this with a confidence of 95% for 3

winning papers with an average of 175 total readings. These numbers reflect simulation without re-estimating the biases. If re-estimation calculations are implemented, the numbers would improve.

Shortcomings of the model include its assumption that papers have an intrinsic score. This confines the validity of the model to more technical papers. Additionally, the model does not account for a situation in which the inconstancy of judges may vary in the distribution of the scores they report. This would happen if one judge used the full range of scores from 0 to 100 and another judge had a tighter range of reported scores. In this sense, the model does not fully reflect reality. The assumptions of a normal distribution and simulation over normally distributed data are also approximations of reality.

Appendix A: Estimation of Intrinsic Score Distribution

We seek to estimate the mean and variance of the intrinsic scores by observing the scores of the first round. The most reasonable estimate for the mean of the intrinsic scores is simply the mean of the scores observed in the first round, $\hat{\mu} = \overline{X}$. The scores assigned to various papers in the first round by a particular judge are of the form $S_{jp} = S_p + B_j + \epsilon_{jp}$ for the various values of p. The variance of these numbers for fixed j is the sum of the variance of the intrinsic score and the inconsistency Δ^2. Since the Δ^2 is insignificant compared to σ^2, we may approximate σ^2 by the variance Var S_{jp} for a fixed value of j. Averaging this variance over all judges yields a reasonable estimate for the variance σ^2:

$$\hat{\sigma}^2 = \frac{1}{J} \sum \text{Var } S_{jp}.$$

Appendix B: Estimation of Biases and Net Scores

Theorem. *Suppose that A is a random variable with distribution $N(\mu, \sigma^2)$ and A is hidden from observation, but the independent random variables $X_i = N(A, \sigma_i^2)$ are observed. Given observations X_i, the conditional distribution for A is normal with*

$$\textit{mean} = \frac{\frac{\mu}{\sigma^2} + \sum \frac{X_i}{\sigma_i^2}}{\frac{1}{\sigma^2} + \sum \frac{1}{\sigma_i^2}}, \qquad \textit{variance} = \frac{1}{\frac{1}{\sigma^2} + \sum \frac{1}{\sigma_i^2}}.$$

Proof: [EDITOR'S NOTE: This theorem was formulated by the authors, who could not find a reference for it. For reasons of space, we omit their proof, which is based on results in DeGroot [1986].]

Corollary 1. *If a judge's bias comes from the distribution $N(0, B^2)$ and the scores $X_1, \ldots, X_n$ reflect the bias, the variation σ^2 of intrinsic scores of papers with mean μ, and the inconsistency Δ of the judge, then our estimate for the bias of this judge and the variance of this estimate are*

$$B_j = \frac{\sum(X_i - \mu)}{n + \frac{\sigma^2+\Delta^2}{B^2}}, \qquad V_j = \frac{1}{\frac{1}{B^2} + \frac{n}{\sigma^2+\Delta^2}}.$$

Corollary 2. *Suppose that the intrinsic score of a paper comes from $N(\mu, \sigma^2)$. The scores S_{jp} reflect the intrinsic score of the paper, the biases of the judges, and the inconsistency Δ of the judging process. Our estimates B_j of the biases each have variance V. Then our estimate of the intrinsic score for paper p has mean and variance:*

$$\begin{aligned} \textit{mean} &= \frac{\frac{\mu}{\sigma^2} + \sum_{j \text{ judged } p} \frac{S_{jp}-B_j}{V+\Delta^2}}{\frac{1}{\sigma^2} + (\#\ \textit{readings})\frac{1}{V+\Delta^2}} \\ \textit{variance} &= \frac{1}{\frac{1}{\sigma^2} + (\#\ \textit{readings})\frac{1}{V+\Delta^2}}. \end{aligned}$$

Each paper has a normal score distribution with mean $\hat{S}_j$ and variance σ_T^2. The variances are the same for each paper. Then $S_{jp} - B_j$ (based on our best estimate of B_j, which may change from round to round), which is our best estimate of a judge's score, has variance $V_j + \Delta^2\sigma^2$. If we just up all bias variances to $V = \max V_j$, this becomes $V + \Delta^2\sigma^2$. So overall for this paper,

$$\text{mean} = \frac{\frac{\mu}{\sigma^2} + \sum \frac{S_{jp}-B_j}{V+\Delta^2\sigma^2}}{\frac{1}{\sigma^2} + n\left(\frac{1}{V+\Delta^2\sigma^2}\right)}, \qquad \text{variance} = \frac{1}{\frac{1}{\sigma^2} + n\left(\frac{1}{V+\Delta^2\sigma^2}\right)}.$$

Note that $\sigma_T = V + \Delta^2\sigma^2$.

Appendix C: Re-estimation of Parameters

First we seek to re-estimate the parameter Δ. If we consider all papers (at least two) read by both judge j and judge k, the differences are distributed according to

$$S_{jp} - S_{kp} = B_j - B_k + (\epsilon_{jp} - \epsilon_{kp}) = B_j - B_k + N\big(0, 2\Delta^2\big).$$

By computing the variance of the differences $S_{jp} - S_{kp}$ for a fixed pair of independent judges, we obtain an estimate of the variance $2\Delta^2$. The more papers the pair of judges has read in common, the more precise this estimate will be. We obtain a still more precise estimate of $2\Delta^2$ by averaging these variances

over each pair of judges, weighting each average according to the number of papers read by both judges:

$$\hat{\Delta}^2 = \frac{\frac{1}{2}\sum(P(j,k)-1)\text{Var}\left[S_{jp}-S_{kp}\right]}{\sum(P(j,k)-1)},$$

where $P(j,k)$ is the number of papers read by both judges j and k. Using the updated estimate of Δ, we may now re-estimate the biases B_j as well as their variances V_j. We use an iterative procedure and demonstrate that for $\Delta \neq 0$ the successive calculations for B_j and V_j converge. We cannot rigorously demonstrate the validity of this iterative procedure. Instead, we justify this procedure by intuitively motivating each step. [EDITOR'S NOTE: For reasons of space, we omit the details.]

Appendix D: The Confidence Inequality

Theorem. *Let $S_{j1},\ldots,S_{jP}$ denote the computed scores of the papers sorted in ascending order and let σ_T denote the standard deviation of the score estimates. If $R \leq P-W$, then the probability of accidentally rejecting one of the best W papers by rejecting the R lowest-ranked papers is bounded by*

$$\text{Pr(mistaken rejection)} < W \cdot \sum_{p=1}^{R} \Phi\left(\frac{Sj,p-S_{j,p-w+1}}{\sqrt{2}\sigma_T}\right).$$

Proof: To mistakenly eliminate one of the best W papers, it is necessary that one of the rejected papers have an intrinsic score greater than that of one of the top W papers. Thus

$$\begin{aligned}
\text{Pr(mistaken rejection)} &< \sum_{p=1}^{R}\sum_{q=P-W+1}^{P} \text{Pr}(S_p > S_q) \\
&< \sum_{p=1}^{R}\sum_{q=P-W+1}^{P} \text{Pr}(S_p > S_{P-W+1}) \\
&= W \cdot \sum_{p=1}^{R} \text{Pr}(S_p - S_{P-W+1}) \\
&= W \cdot \sum_{p=1}^{R} \Phi\left(\frac{S_{j,p}-S_{j,P-W+1}}{\sqrt{2}\sigma_T}\right).
\end{aligned}$$

References

DeGroot, Morris H. 1986. *Probability and Statistics.* Reading, MA: Addison-Wesley.

Judge's Commentary: The Outstanding Contest Judging Papers

Veena B. Mendiratta
Bell Labs
Lucent Technologies
2000 N. Naperville Road
Naperville, IL 60566
veena@lucent.com

The Contest Judging Problem provided the contestants with a challenging real-world problem that lent itself to a range of analysis and modeling methods. In coming up with a "best" selection scheme, the contestants used methods such as rank-ordering, numerical scoring, and bias estimation.

What made the problem interesting and challenging to model, and also to judge, were the less well-defined aspects of the problem. For example, how was bias of judges with respect to ranking and scoring papers handled? Also, how was the issue of ensuring that the best $2W$ papers were not screened out in the early rounds of reading addressed? Consequently, we could not select the winning papers for MCM 1996 based solely on the criterion of how many paper-readings a team's algorithm required; we also considered how the issues were addressed.

Almost all of the successful papers were able to develop a basic model for the ideal case to select the top W papers for the specific parameter values specified in the problem statement. A key assumption for the ideal case is that every judge rank-orders or scores all the the papers as the absolute rank-ordering, that is, there is no judge bias. The so-called ideal model, though unrealistic, sets a lower bound for the total number of reads. The stronger papers went significantly beyond the ideal case.

Characteristics of the best papers included the following:

- An explicit modeling of judges' bias, addressing the issues of systematic bias and the variance of accidental errors in scoring.
- Estimating statistical bounds on the probability of failure to pick the best W papers out of the top $2W$ papers.
- In addition, some of the papers realized the importance of more judges reading the papers remaining in the later screening rounds and included this factor in their models.

- The better papers provided a clear statement of results in terms of the total number of paper readings, the confidence level of the results, and sensitivity of the results to the model parameters.

Various different approaches were selected to address the above issues and some of these are summarized below.

The Gettysburg College team minimizes the probability of eliminating the W best papers in the first round by having two judges read each paper in that round. This same paper models judge error through a functional relationship between the probability of judge error in ranking (with respect to the absolute ranking) and the distance between two compared papers on the absolute scale. The team from the University of Science and Technology of China uses Bayesian estimation to address systematic bias in scoring. They also model the error probability as a function of the percentage of papers eliminated in each round. The Fudan University team shows statistically the conditions under which the "ideal" model can work. The Washington University team models the judge bias and error and, after each round of judging, uses new bias estimates to calculate confidence intervals that determine the number of papers rejected. The St. Bonaventure team implements a novel distribution scheme, which they illustrated very effectively with matrices, to ensure that judges do not receive the same paper more than once and also that the same judge does not receive the top $2W$ papers.

Lastly, the best papers were characterized by clear and logical presentations that brought forth the team's underlying analytical thought process. These papers were well organized and well written, with appropriate tables and graphics for presentation of results, and included a comprehensive summary, all of which made it easier to understand the material presented.

The Contest Judging problem was challenging and many excellent solutions were offered. Finally, however, five papers stood out from the others, and the members of those teams should feel proud of their accomplishments.

About the Author

Dr. Mendiratta (Ph.D., Northwestern University, 1981) has been at Bell Labs (now part of Lucent Technologies) since 1984, working on a wide range of systems. She currently works in the Architecture and Performance area. Her work at Bell Labs has focused on reliability modeling and performance analysis of switching systems, as well as mathematical programming models for switch configurations. Prior to 1984, she worked for three years as Manager of Operations Research for the Illinois Central Gulf Railroad, where she directed the implementation of an empty-freight-car distribution optimization model that was developed as part of her Ph.D. dissertation. Her professional activities include serving on the MCM Advisory Board as well as an MCM judge, being co-President of the INFORMS Chicago Chapter, and being a SIAM Visiting Lecturer.

Judge's Commentary: The Outstanding Contest Judging Papers

Donald E. Miller
Department of Mathematics
Saint Mary's College
Notre Dame, IN 46556
dmiller@saintmarys.edu

This year's problem, while framed in the terminology of a competition such as the MCM, finds application in several other areas of decision-making. One such area, itself a competition of sorts, occurs when a school makes its decisions on the award of scholarships. Another occurs in the screening of applicants for a specified position. In both of these situations, the potential exists for more nominally qualified applicants than positions. Thus, those making the decision, the judges in our problem, must either rank-order or otherwise quantify the applicants in an attempt to decide which ones are "best." Further, these decisions must be made under time constraints that make it impossible for each "judge" to evaluate every applicant; even if they were to do so, it is doubtful that the evaluations would be in complete agreement. It is this element that complicated evaluation of the contest papers.

The assumption of an absolute rank-ordering made the problem seem deceptively easy, resulting in a broad range of papers, from quite simple to very elaborate. At one end of the spectrum, we found papers that assumed absolute rank-ordering and simply developed a heuristic solution to the basic problem; some even recognized the assumption was unrealistic but chose to model the problem as stated, since that is what was requested. Others who used this assumption clearly didn't believe it, since they rejected a simple merge-sort in favor of more complicated algorithms that denied existence of the assumption. Still others attempted to refine the problem using theories ranging from topics in graph theory to concepts of fuzzy sets. Some of the better bias-elimination refinements included matrix reduction, regression with error terms, and scoring normalization with specified probability distributions.

The judges felt that the best models were those that solved the basic problem for 100 papers, 8 judges, and 3 winners, then produced a successful refinement with adequate complexity to accurately model the process but with enough simplicity for the model to be useful. Further, each model would be clearly stated and its use demonstrated with a simple example. Thus, the ideal paper would solve the basic problem and demonstrate that the solution was optimal, or at least close to optimal. It would then generalize the solution to accommodate different numbers of papers, judges, and winners. Having completed this,

it would address judge bias and measure the success rate of the algorithm as a function of some quantitative measure of judge bias. It might then examine alternative algorithms, finding a relation between levels of judges bias and the success rate of these algorithms. It would continue by addressing the strengths and weaknesses of these methods, while being sure to clearly address all the points requested in the contest rules.

While many papers showed much insight into the problem and its complexities, the Outstanding papers were distinguished by the way that they addressed the problem of judge bias effectively, and with appropriate documentation, to allow for implementation of the recommended model. The team from the University of Science and Technology of China used Bayesian statistics, with a normal prior, to adjust for judge bias, then ran simulations to test for sensitivity of their method. Normal distributions, a common assumption, were also used by the Washington University team for both the intrinsic score of the paper and judge bias. A distinguishing feature of the paper from Fudan University was its stability analysis for different levels of judge bias.

Finally, here's a note on good practice in mathematical modeling as related to this problem. Common in requests for models are unrealistic assumptions that would make any model created with these assumptions of minimal use. Thus, it is necessary for the modeler to evaluate critically all assumptions and, if necessary, refine the problem to one that is realistic. Judges viewed the statement, "there is an absolute rank-ordering to which all judges would agree," to be such an assumption. Under this assumption, with 100 papers and 8 judges, there are methods of finding the top three papers, in order, with each judge reading at most 14 papers and with at most 109 papers read. In another situation and under different rules, the modeler would ask the author for elaboration on the statement before proceeding with the model. But in the absence of such consultation, the modeler should answer the question as stated and then refine it with realistic assumptions.

About the Author

Donald Miller is Associate Professor and of Chair of Mathematics at Saint Mary's College. He has served as an associate judge of the MCM for four years and prior to that mentored two Meritorious teams. He has done considerable consulting and research in the areas of modeling and applied statistics. He is currently a member of SIAM's Education Committee and past president of the Indiana Section of the Mathematical Association of America.

Contest Director's Commentary: Judging the MCM

Frank Giordano
COMAP, Inc.
57 Bedford St., Suite 210
Lexington, MA 02173
f.giordano@mail.comap.com
FRGiordano@aol.com

Overview

Each paper submitted to the MCM is classified as Non-successful, Successful Participant, Honorable Mention, Meritorious, or Outstanding. Further, from among the Outstanding papers, judges from SIAM, INFORMS, and the MAA pick winners to receive awards from their societies. Typically, the top 2–3% are classified as Outstanding, the next 15% as Meritorious, and the next 25% as Honorable Mention. A paper classified as Successful Participant is a complete paper that satisfies the rules of the contest.

The process of determining the classification of each paper consists of distinct phases: screening, grading, and judging. As detailed below, each of these phases employs a mixture of normalized grading, ranking, classification against a judge's "absolute ideal," and qualitative judging. We now describe each of the phases.

Screening Rounds

Typically, there are three screening rounds. The primary purpose of screening rounds is to identify papers that are not going to be among the top 43% (Honorable Mention or above). After a preliminary reading of several papers (not graded), judges jointly design a 7-point grading scale. Discriminators used on the scale are designed to classify the papers into three categories. The top classification reflects exceptional quality; the second category are papers that the judge feels should be retained; and the third category are papers that the judge thinks should be eliminated. The grading scale emphasizes heavily the organization of the paper: What ideas have the contestants used in developing and analyzing their models? The summary is weighted heavily, as the contest rules require that it reflect the major ideas that the contestants used. Typically, a judge will devote 10–15 minutes to screen a paper.

The first two of the three screening rounds are accomplished by "Triage Judges." For the 1996 contest, the triage judging took place at Carroll College

in Helena, MT. Typically, 10–14 judges participate in the triage judging, and each paper is read by one judge in each round. After the second round, papers that were judged in the lowest category by both the judges who read the paper are eliminated. If both judges vote the paper weak, but their scores differ by more than 2 points, a third judge classifies the paper before it is eliminated. Typically, about 25% of the papers are eliminated after the first two screening rounds.

Grading Rounds

Grading takes place in Claremont, California by a different set of judges. Normalized scores from the first two screening rounds are used to rank-order the papers. The normalized scores are used to organize the papers into as many stacks as there are judges. Upon arrival at Claremont, judges are given a stratified packet of papers to read for "calibration" (not grade). After reading the papers, the judges jointly design a 7-point screening scale. The judges then conduct a third and final screening round.

After the screening round at Claremont, the judges design a 100-point scale to use in the grading rounds. Typically, there are 4 grading rounds. A judge typically spends 30–45 minutes to grade each paper. In addition to grading the paper, the judge is asked to rank it against all papers read in the round. For example, "2/7" would be the second best of 7 papers read in a grading rank. Ranking requires the judge to pick out the top papers each round. The judge is also required to assign an "absolute" classification: Successful Participant, Honorable Mention, Meritorious, or Outstanding. This classification permits the judges to render an opinion on the "absolute" quality of the paper independent of their grading and ranking procedures.

Final Judging

After the three screening rounds and four final grading rounds, typically 6–12 papers remain in the contest. Time is then provided for judges to read papers that they have not yet read. The judges then meet to judge the merits of each of the remaining papers. Judges who have studied the papers debate the strong points and weak points of each paper. The papers are then compared against one another. By consensus, the Outstanding papers are chosen.

Ranking the Papers

No attempt is made to rank the papers until after the first two screening rounds. Normalized scores are used for ranking the papers thereafter. Beginning with the third screening round, judges are required to rank the papers

against all papers read that round. Additionally, they are required to judge the paper as Successful Participant, Honorable Mention, Meritorious, or Outstanding.

Stratified Packets

Beginning with the third screening round, the papers are organized into as many stacks as there are judges. The papers are distributed modulo the number of judges, using the cumulative normalized scores. The cumulative scores are weighted, with the grading rounds counting more than the screening rounds.

Eliminating Papers

Typically, about 25% of the papers are eliminated after the first two screening rounds. Beginning with the third screening round, the Contest Director, the Associate Contest Director, and the two Head Judges meet to discuss the elimination of papers. The information that they use is

- the overall rank based on normalized scores,
- the round rank assigned each round, and
- the overall classification given to the paper by the judge.

Each item is quite useful in determining which papers should be eliminated. Since judges receive a stratified packet, the round ranks are especially useful. No "quota" is used to determine how many papers to cut each round. Typically, the top 43% (Honorable Mention, Meritorious, and Outstanding) remain after the first grading round; the top 18% (Meritorious and Outstanding) remain after the third grading round.

The Judges

Typically there are 10–14 judges accomplishing the first two screening rounds (Triage). These judges are led by experienced Head Judges who have graded several years at Claremont before becoming a Head Triage Judge. Generally, calibration sessions are held, where papers are read but not graded. After the calibration reading, 7-point scales are developed. For the grading and judging at Claremont, 25 judges are chosen. The professional societies choose their own judges from among volunteers with strong credentials. COMAP rounds out the judging team by choosing a field of judges with a wide range of expertise. The judges selected are respected in the mathematical science community for their integrity and dedication. If needed, the field of judges is augmented with subject-matter experts.

Conclusion

Judging a contest that receives as many creative student solutions as the MCM does is a difficult task. The judges deserve our gratitude for a difficult job that has been done well, and with great dedication and integrity.

About the Author

Frank Giordano has been the Director of the MCM since 1991.

Practitioner's Commentary: Computer Support for the MCM

Steve Harper
Mathematics Dept.
Caroll College
Helena, MT 59625
sharper@saints.carroll.edu

Judging the math modeling contest is mostly a human endeavor, with a proper place for a computer to help. Since human judges are, by definition, human, there need to be ways to watch for human bias in order to get fair contest results.

The current scheme that Contest Director Frank Giordano uses has evolved to address these concerns. To make a fair judgment, each judge needs to see excellent, good, and average papers. There are too many papers for each judge to read them all, and assigning papers by random draw will not always give a good balance. After being ranked using the scores from prior rounds, the papers are distributed to ensure that each judge gets a variety in quality.

Since different people do not give the same score to the same paper, the scores are weighted (to account for natural human tendency to "grade low" or "grade high"). A more subtle problem is that people tend to "root" for a paper that may, for instance, resonate with how the judge would approach the problem (the "right answer," so to speak). Making the scoring scale 40–100, rather than 1–100, can reduce the impact of rooters who give low scores to other papers. Having the judge rank all the papers for that round also gives the contest director more control to ensure that a paper is not arbitrarily eliminated. For instance, if a paper with a low score has a ranking of 2nd out of 15 papers, it may deserve to stay for another round and another opinion. Furthermore, the ranking provides a way for judges to give a numerical score based on the established criteria yet still note that there are unquantifiable factors that make one paper rank better than another paper with a higher score.

The judging for this contest previously used a spreadsheet. This gave some last minute flexibility, at the cost of data entry errors plus late-round bleary eyes trying to line up too much numerical data that would not fit on one page. In 1996, the contest used a custom FoxPro database program with three database tables (for scores, for judge information and weightings, and for round names and elimination scores).

Judges are assigned using letter codes rather than numbers since numerical scores abound on the printed score sheets. Each judge needs a random set of papers containing both good and average papers (that this judge has not read before). After the two triage rounds, the remaining contestant papers are split into 4 stratified layers based upon accumulated weighted scores. The

computer tries to give an equal number of papers to each level and assigns papers so that each judge will get a variety of quality in papers. It also ensures that no judge reads the same paper twice. The contest director resolves any problems generated (such as the computer not assigning any paper to a judge if, by luck, the only papers left in the contest have all been read by that judge).

The computer prints out the judge assignments for the contest director in score order, and for the individual judges in document number order (so the judge doesn't know how the paper has fared so far).

The computer checks scores as they are recorded and alerts the human input operator for typing or recording errors. (The score that a judge intends to put on a paper should not be rendered invalid by a typo or two.) It notes problems if the human tries to enter a score for a document that doesn't exist or is already out of the contest. It verifies that the actual judge is the same as was assigned (though this can be overridden). Other problems noted are scores out of range, changes to an existing score ("Is this really a change, or did you type the wrong document number?"), and ranking within a round (saying "This is the 5th best paper out of 4" will not pass inspection).

Some individual judges tend to grade harder or easier than others. A weighting formula (from Frank Giordano, who inherited it from Ben Fusaro) is designed to try to account for that with a minimum of effort. The formula is:

$$\text{Weighted Score} = \begin{cases} \dfrac{\text{Population Mean}}{\text{Judge's Mean}} \times \text{Raw Score}, & \text{if Population Mean} \le \text{Judge's Mean;} \\ \dfrac{100 - \text{Population Mean}}{100 - \text{Judge's Mean}} \times (\text{Raw Score} - 100) + 100, & \text{otherwise.} \end{cases}$$

(Note: no weighted score can be greater than 100, no matter how tough the judge is on the other papers.)

To allow for human modification, this weighting is applied in two steps. The program first calculates the judge's ratio, and the contest director can then individually adjust the judge's weighting. For example, a judge could read only a couple of papers before getting an emergency phone call and having to leave, and the weighting may be too far off (in the opinion of the contest director) to be useful.

Next the weighting ratio that goes with each judge is applied to each document that the judge scored. Again, the contest director can adjust the weighted score for any one paper, if there is good reason to do so.

To calculate the weighting, the only papers considered are the ones still left in the contest. All prior rounds for those papers are considered in calculating the weighting for each judge. Then the new weighted score is figured retroactively for the prior rounds to obtain a new weighted total. Thus, in each round, the weighting has to be recalculated. (Note: In each round, the low-scoring papers drop out. While the number of recorded scores is increasing, the number of papers left is decreasing, so the total number of scores included may go up or down. There always is a concern with weighting too small a sample.)

The length of time that a judge spends on the paper is also a factor, since a score for a Triage Round reading of 5 minutes does not deserve the same confidence as a 30-minute reading in the Finals.

Then, based on the weighted score, the computer suggests which papers to eliminate. (The elimination scheme is stored in the database and is easy to change.) It calculates what score will leave a (preselected) percentage of the original contestants after the current round, compares that to a (preselected) minimum score, and suggests the higher number.

After review, the contest director can decide to draw the elimination line in a different place, as well as check the set of scores for individual papers below the line to decide which papers deserve another chance. (If a paper got a 99 and a 40, it would probably deserve another read before being tossed out of the contest.)

The percentage and minimum scores for the 1996 contest are shown in **Table 1**.

Table 1.
Percentage and minimum scores for judging rounds of the 1996 contest.

Round	Score range	Minimum continuation score	Percentage remaining
Triage 1	0–7	1	99%
Triage 2	0–7	6	90%
Screening	0–7	12	43%
Final 1	20–50	25	30%
Final 2	40–100	85	18%
Final 3	40–100	150	10%
Final 4	40-100	220	6%

The computer also records the final classification based on when the document is eliminated:

- Survive Final 4: Outstanding
- Survive Final 3: Meritorious
- Survive Final 2: Meritorious
- Survive Final 1: Honorable Mention
- Survive Screening: Honorable Mention
- Survive Triage 1: Successful

The contest director can review these classifications and change them.

To help the contest director make elimination decisions, the computer prints out reports in weighted score order. At all times, a printout is available (in document number order) to answer the question, "Whatever happened to paper number such and such?"

Using this database program solves many problems. However, as one would expect, there is a cost—flexibility. The data are not readily accessible to noncomputer folks, so changing the structure of how the contest is judged is not too easy. It was not feasible in the time allotted to make a program generally usable for every judging situation possible. Given the stability of the contest over the last several years plus time constraints, trading future flexibility to solve present needs (for data entry, judge assignment, score weighting and elimination) seemed a fair trade.

Are there improvements? Are you kidding? The final round ends with judges arguing over the top 3% of papers about which ones are really Outstanding and which of those deserve special awards. For all judges to participate, each needs to have read several of the top papers—more than just luck would allocate. Next year's program already has a better judge assignment algorithm in place!

About the Author

Steve Harper teaches at Carroll College in Helena, MT, where students are encouraged to have a variety of background work in areas other than just computing. (Steve himself has a background in accounting, politics, wind energy, and consulting.) He designed and coded the database and program for the 1996 contest.

COMAP ANNOUNCES

THE MATHEMATICAL CONTEST IN MODELING

MCM 1997

FEBRUARY 7–10, 1997

The thirteenth annual international Mathematical Contest in Modeling will be held February 7–10, 1997. The contest will offer students the opportunity to compete in a team setting, using mathematics to solve real-world problems.

For registration information, contact:
MCM 1997, COMAP, Inc., Suite 210, 57 Bedford Street, Lexington, MA 02173
email: mcm@comap.com voice: 617-862-7878

Major funding provided by the National Security Agency.

Additional support for this project is provided by the Institute for Operations Reasearch and the Management Sciences, the Society for Industrial and Applied Mathematics, and the Mathematical Association of America.