UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,
Petitioner,

v.

BUFFALO PATENTS, LLC,
Patent Owner.

Case No. IPR2023-01386
U.S. Patent No. 6,904,405

Declaration of Shauna L. Wiest Regarding *Vo*

I, Shauna L. Wiest, state and declare as follows:

I. Introduction

1. I have prepared this Declaration in connection with Google LLC's ("Petitioner") Petition for *Inter Partes* Review of U.S. Patent No. 6,904,405, which I understand will be filed concurrently with this Declaration.

2. I am currently a senior research analyst with the Research & Information Services team at Finnegan, Henderson, Farabow, Garrett & Dunner, LLP located at 3300 Hillview Avenue, Palo Alto, CA 94304 ("Finnegan").

3. I am over eighteen years of age, and I am competent to make this Declaration. I make this Declaration based upon my own personal knowledge, and my professional knowledge of library science practices.

4. I earned a Master of Science in Library Science degree from the University of North Carolina at Chapel Hill in 1999, and a Bachelor of Arts in Political Science degree from the University of California at San Diego in 1989. I have worked as a law librarian for over eighteen years. I have been employed in the Research & Information Services Department at Finnegan since 2021. Before that, from 2000-2015, I was employed as a Law Librarian at Stoel Rives LLP, and from 2015-2016, I was employed as a Competitive Intelligence Specialist for Nossaman LLP.

II. Standard Library Practice for Receiving, Cataloging, Shelving, and Making Materials, including Conference Publications, Publicly Available

5. I have knowledge of and experience with standard library practices regarding receiving, cataloging, shelving, and making materials, including conference publications, available to the public. I am fully familiar with and have knowledge of and experience with the Machine-Readable Cataloging (MARC) system, an industry-wide standard that libraries use to catalog materials.

6. The MARC system was developed during the 1960s to standardize bibliographic catalog records so they could be read by computers and shared among libraries. By the mid-1970s, MARC had become the international standard for cataloging bibliographic materials and is still used today. Many libraries provide public access to their MARC records via the Internet and/or their electronic cataloging systems at the library. In a MARC record, each field provides specific information about the cataloged item, including how materials are held and made available to the public.

III. MARC Records

7. The MARC record system uses a specific three-digit numeric code ("field tags") (from 001-999) to identify each field in a catalog record. For example, MARC field tag 008 provides the six-digit date the item was received

and catalogued (*Date entered on file*). The first six characters of field tag 008 are always in the "**YYMMDD"** format. Descriptions and definitions of all of the character positions of field tag 008 are outlined here: https://www.loc.gov/marc/bibliographic/bd008a.html

8. As is relevant to this Declaration, MARC field tag 245 identifies the title and statement of responsibility for the work. MARC field tag 260 identifies the place of publication, name of publisher, and date of the publication. MARC field tag 310 sets forth the current stated publication frequency of either an item or an update to an item MARC field tag 362 sets for the dates of publication and/or sequential date(s) of publication designation for a work. MARC field tag 515 provides any numbering or publishing pattern peculiarities for the resource. And MARC field tag 991 provides local holdings information to assist the public in identifying the location of a desired resource.

9. Based on standard library practice, when a library receives an item, it generally stamps (and/or labels) the item with the library name, barcode, often with a date that is within a few days or weeks of receipt. Next, the library will catalog the item within a matter of a few days or weeks of receiving it. As a general practice, cataloguing is centralized and performed by a cataloguing department within a library or university setting. In certain circumstances the catalogued item may be subsequently sent to a library location within the

library or university setting where it may be stamped and/or labeled after it has been catalogued.

10. Generally, after an item is cataloged, the public may access the item by searching a library catalog, browsing the library shelves, and either requesting or electronically accessing the item from the library. Standard library practice is to make the item available to the public within a few days or weeks of cataloging it.

IV. Cataloguing of Conference Proceedings (Monograph v. Serial Treatment)

11. Conference proceedings often consist of collections of papers presented at a meeting, program, symposium or conference held during a specific time-period in a designated geographic place. Bibliographic control of conference proceedings to promote public discovery and access has long been a priority of catalogers. Although some ongoing conference publications are catalogued as serials (if a conference publication exhibits continuity, regularity, and evidence of seriality), cataloguing a conference publication as a monograph is preferred for public retrieval purposes. Cataloguing a conference publication as a monograph is an optimal way for the public to easily access the conference resource, and its papers, by searching for a unique title and other descriptive data elements added by the cataloger. Monographs are efficient for public searching and offer the public an easier way to discover

what an institution holds, while changes to the record (unlike serials) are not an issue. However, sometimes conference proceeding cataloging is done serially, and the MARC record will be updated after each serial publication is received.

V. Public Availability of *Vo*

12. This Declaration relates to the dates of receipt and public availability of the following reference: Minh Tue Vo and C. Wood, "Building an application framework for speech and pen input integration in multimodal learning interfaces," *("Vo")* found within the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Atlanta, GA, USA, May 7-10, 1996, pp. 3545-3548 Volume 6 ("1996 IEEE Publication"). I understand that *Vo* has been submitted as Exhibit 1009 in this proceeding. That same reference is appended to my Declaration as **Appendix A**.

13. As detailed below, I have reviewed the print reference, public holdings information, and Library of Congress MARC record for *Vo* to determine the date of public availability for this reference.

14. **Appendix A** to this Declaration is a true and accurate copy of the cover page (including binding), title page, front matter, table of contents for the copy of the 1996 IEEE Publication containing *Vo* held by the Library of Congress. **Appendix A** also includes pages 3545-3548, which comprise the publication,

"Building an application framework for speech and pen input integration in multimodal learning interfaces," *i.e., Vo*. **Appendix A** notes that the "Copyrights Manager" is the "IEEE Service Center," with an address of "445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 008855-1331" and has a Copyright date of 1996. It also explains that the conference was "Sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers, Inc."

15. **Appendix D** to this Declaration is a true and accurate copy of search results received from a search of the Library of Congress public catalog for records on International Conference on Acoustics, Speech, and Signal Processing (ICASSP) conferences, which was downloaded from https://catalog.loc.gov/vwebv/search?searchArg=IEEE+International+Conference+on+Acoustics%2C+Speech%2C+and+Signal+Processing&searchCode=GKEY%5E*&searchType=0&recCount=100&chkdTitles=0&sortBy=RELEVANCE&recPointer=0. The search results indicate that the Library of Congress public catalog holds 8 separate records for ICASSP conferences held since 1976. Some records are only associated with an ICASSP conference for one year, whereas other records are associated with ICASSP conferences from multiple years.

16. **Appendix B** to this Declaration is a true and accurate copy of the Library of Congress public catalog record for its copy of the 1996 IEEE Publication containing *Vo*, which was downloaded from

https://lccn.loc.gov/98655050 on September 5, 2023. I confirmed with librarians at the Library of Congress that the public catalog record is the correct and operative cataloguing record for the Library of Congress holdings for the 1996 IEEE Publication containing *Vo*.

17. The Library of Congress public catalog record (**Appendix B**) sets forth the public holdings and onsite location information for members of the public seeking a physical copy of the 1996 IEEE Publication containing *Vo*. The public catalog record indicates it is a "Conference proceedings" record and, specifically, identifies the meeting name as "ICASSP (Conference)," and the main title as "International Conference on Acoustics, Speech, and Signal Processing." The public catalog record notes that the publications associated with the record are (i) published "annual[ly]," (ii) published in 1995 and 1996, (iii) issued in five (5) or more parts, and (iv) should be requested in the Jefferson or Adams Building Reading Rooms at Call Number TK7882.S65 I37a. This information indicates serial cataloguing, as discussed above in paragraph 11. The public catalog record also notes a publication location in Piscataway, NJ, with copyright dates of 1995 and 1996, and IEEE as the publisher. It also explains that the conference was "sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers." Based on my experience as a librarian, the information provided in the public catalog record (**Appendix B**) indicates that the record

references multiple publications, including publications dated 1995 and 1996. Further, based on my experience as a librarian, the public catalog record (**Appendix B**) references the 1996 IEEE Publication containing *Vo* (**Appendix A**).

18. **Appendix C** to this Declaration is a true and accurate copy of the Library of Congress MARC record for its holdings of the 1996 IEEE Publication containing *Vo*, which was downloaded from https://catalog.loc.gov/vwebv/staffView?searchId=24554&recPointer=0&recCount=25&searchType=7&bibId=11501706 on September 5, 2023. I confirmed with librarians at the Library of Congress that the MARC record (**Appendix C**) is the correct and operative cataloguing record for the Library of Congress holdings for the 1996 IEEE Publication containing *Vo*. The Library of Congress MARC record sets forth information for members of the public seeking a physical copy of the 1996 IEEE Publication containing *Vo*. The Library of Congress MARC record field tag 245 sets forth the full title statement for the resource as: **a** Conference proceedings / |**c** the … International Conference on Acoustics, Speech, and Signal Processing; sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers. MARC field tag 260 sets forth the place of publication, publisher, and publication dates as: |**a** Piscataway, NJ : |**b** Institute of Electrical and Electronics Engineers, |**c** c1995-c1996. MARC field tag 310 indicates that the IEEE International Conference on Acoustics, Speech, and Signal

Processing Conference Proceedings is an annual publication. MARC field tag 362 indicates the sequential dates of publication for the resource as 1995-1996, indicating publication dates of both 1995 and 1996. MARC field tag 515 denotes numbering peculiarities for the resource as being issued in 5 or more parts. Finally, MARC field tag 991 sets forth the Library of Congress local holdings and call number information for the 1996 IEEE Publication containing *Vo* as General Collection, Call Number TK7882.S65 I37a (Serials). Based on my experience as a librarian, the MARC record (**Appendix C**) references the 1996 IEEE Publication containing *Vo* (**Appendix A**).

19. **Appendix C** confirms the fixed data elements of MARC field tag 008 as 950630d19951996njuar 1 0eng c. Data elements "d1995-1996" in field tag 008 indicates (i) new issues of the publication associated with this MARC record have published after 1996, or that a change in author or title has caused a successive MARC entry record to be created, and (ii) the first volume of the publication associated with the MARC record is dated 1995 and the last publication is dated 1996. As discussed above in paragraph 7, the first six characters "950630" are in typical "**YYMMDD**" format and indicate that the first volume of publication for the resource associated with the MARC record was received by The Library of Congress on June 30, 1995.

20. **Appendix E** to this Declaration is a true and accurate copy of the IEEE Xplore webpage associated with the 1995 IEEE ICASSP, which was downloaded from https://ieeexplore.ieee.org/xpl/conhome/3469/proceeding. This webpage indicates that the 1995 IEEE ICASSP was held on May 9-12, 1995. As explained above, the 1995 publication associated with the MARC record in **Appendix C** was received by The Library of Congress on June 30, 1995, a mere six weeks after the conclusion of the 1995 IEEE ICASSP.

21. **Appendix F** to this Declaration is a true and accurate copy of the cover, title and verso pages, and table of contents of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, held by the University of Wisconsin at Madison, Kurt F. Wendt Library. **Appendix G** to this Declaration is a true and accurate copy of the University of Wisconsin at Madison MARC record for its holdings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, held by the University of Wisconsin at Madison, Kurt F. Wendt Library, which was downloaded from https://search.library.wisc.edu/catalog/999858382602121 (*See:* Library Staff Details -staff view) on August 22, 2023. **Appendix F** indicates that the conference related to the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing was held on April 21-24, 1997. The MARC record in **Appendix G** confirms the fixed data elements of MARC field tag 008 as

970707d19971997cauar 1 0eng d, indicating that the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings was received on July 7, 1997, a mere ten weeks from the end of the 1997 conference.

22. Based on standard library practices and to assure the public can access desired resources, if the 1996 IEEE Publication containing *Vo* was received sometime in 1997, such information would be documented in the MARC record and/or the public catalog record in both the public holdings information and MARC field tags 008, 260, 362, or 991. Because no such information was included in the MARC record, it is my opinion that the Library of Congress received the 1996 IEEE Publication containing *Vo* before January 1, 1997.

23. Based on the information set forth in the appendices attached to this Declaration, standard library practices, and my experience as a librarian, the 1996 IEEE Publication containing *Vo* was received by the Library of Congress shortly after May 10, 1996 (the date of the 1996 conference), between June 1996 and July 1996, and at the latest by December 31, 1996.

24. Based on standard library practices and my experience, *Vo* would have been processed and catalogued by the Library of Congress within a matter of a few days or weeks after it was received in June or July of 1996 (or, at the latest by December 31, 1996). And, based on my experience as a librarian, **Appendix A** is the document the Library of Congress received, as recorded in the MARC record

(**Appendix C**), and was cataloged by the Library of Congress within a matter of a few days or weeks after receiving it.

25. Accordingly, *Vo* would have been made available to the public within a few days or weeks of being checked-in and catalogued. The public could have accessed *Vo* no later than a few days or weeks after December 31, 1996 by (i) searching the Library of Congress online catalog by, for example, searching by title, author, and/or subject keywords, or (ii) by asking a library staff member and being directed to General Collections at Call Number TK7882.S65 I37a.

VI. Conclusion

26. In signing this Declaration, I understand it will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I understand I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

27. I declare that all statements made herein of my knowledge are true, that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and

the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on September 11, 2023.

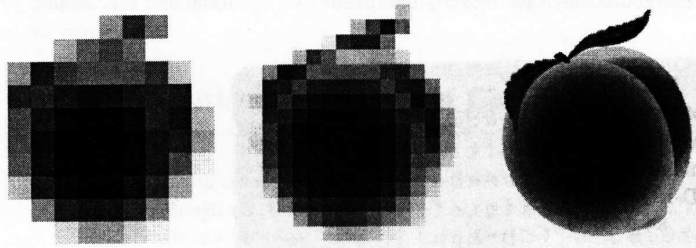


Shauna L. Wiest

APPENDIX A

1996 IEEE ICASSP VOL. 6





The 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings

Sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers

May 7-10, 1996
Marriott Marquis Hotel
Atlanta, Georgia, USA

The 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE

Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright 1996 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number 96CH35903
ISBN 0-7803-3192-3 (softbound)
ISBN 0-7803-3193-1 (casebound)
ISBN 0-7803-3194-X (microfiche)
ISBN 0-7803-3195-8 (CD-Rom)
Library of Congress: 84-645139

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
1-800-678-IEEE

Volume VI

UA1 Detection Classification and Localization

IER Clutter Reduction In Shallow Water VI-3041
F. Shin, D. Kil - Loral Defense Systems, USA
R. Wayland - Naval Air Warfare Center, USA

A Modified Eckart Post-Beamformer Filter for Improved Detection Using Broadband Features VI-3045
S. Mehta, J. Fay, P. Maciejewski - Naval Undersea Warfare Center Division Newport, USA

Feature Extraction and Sufficient Statistics in Detection and Classification VI-3049
E. Real - Sanders, A Lockheed-Martin Co., USA

Gerschgorin Radii Based Source Number Detection for Closely Spaced Signals VI-3053
J. Yang, H. Wu - National Cheng Kung University, ROC

A Wavelet Transform Frequency Classifier for Stochastic Transient Signals VI-3057
F. Garcia, I. Lourtie - ISR/IST, Portugal

Classification with a Combined Information Test VI-3061
R. Lynch - Naval Undersea Warfare Center, USA
P. Willett - University of Connecticut, USA

3-D MUSIC with Polynomial Rooting for Near-Field Source Localization VI-3065
H. Hung, S. Chang, C. Wu - National Taiwan Ocean University, Taiwan

Closed-Form Elliptic Location with an Arbitrary Array Topology VI-3069
U. Hanebeck, G. Schmidt - Technical University of Munich, Germany

Source Localisation with Recurrent Neural Networks VI-3073
B. Colnet, J. Di Martino - CRIN-INRIA CNRS, France

UA2 Matched Field Processing and Special Environments

A Test for Detection of Local Modeling Mismatches in Shallow Water VI-3077
G. Fostik, J. Tabrikian, H. Messer - Tel Aviv University, Israel

Broadband Matched Field Processing Using Robust Prewhitening and Multiple Window Techniques VI-3081
C. Mecklenbrauker, M. Geravanchizadeh, J. Bohme - Ruhr-University of Bochum, Germany

Focalization in the Gulf of Mexico VI-3085
Z. Michalopoulou, M. Porter - New Jersey Institute of Technology, USA

Robust ML Source Localization by Exploiting Predictable Acoustic Modes VI-3089
J. Tabrikian - Tel-Aviv University, Israel
J. Krolik - Duke University, USA
H. Messer - Tel-Aviv University, Israel

A Sub-Space Method for High Resolution Bathymetry VI-3093
A. Antoniou, R. Pillai - University of Victoria, Canada

Multiresolution Stochastic Models for the Efficient Solution of Large-Scale Space Time Estimation Problems VI-3097
T. Ho, P. Fieguth, A. Willsky - LIDS, MIT, USA

A Comparison of Two Techniques for Detecting Seizure in Newborn EEG Data VI-3101
M. Roessgen - Queensland University of Technology, Australia

Classification of Radar Returns Using Wigner-Ville Distribution VI-3105
P. Kumar, K. Prabhu - Indian Institue of Technology, India

On the Design of a Transversal Filter Bank for Parallel Processing Multiple Image Lines in Real-Time Acoustic Imaging VI-3109
J. Shen, E. Ebbini - The University of Michigan, USA

A Minimax Optimization Approach to Sidelobe Suppression Filter Design VI-3113
S. Park, J. Doherty - Iowa State University, USA

Volume VI

UA3 Time-Delay Estimation

An EM Algorithm for Multisensor TDOA/DD Estimation in a Multipath Propagation Environment VI-3117
S. Belanger - Loral Federal Systems, USA

Adaptive Realization of a Maximum Likelihood Time Delay Estimator VI-3121
P. Hahn, V. Mathews, T. Tsen - University of Utah, USA

Robust Time Delay Estimation Using Lower-Order Statistics VI-3125
X. Ma, C. Nikias - University of Southern California, USA

New Perspectives for Maximum-Likelihood Parameter and Time Delay Estimation Using Two Channels VI-3129
J. Stuller, N. Hubing - University of Missouri-Rolla, USA

Iterative Maximum-Likelihood/Cross-Correlation Algorithm for Reflection and Pulse T.O.A. Estimation VI-3133
A. Doser, P. Delaney, L. Schooley - University of Arizona, USA

The Issue of Phase in Specular Multipath Parameter Estimation VI-3137
Z. Mitrovski, E. Titlebaum - University of Rochester, USA

Short Time Correlator Delay Estimation in the Presence of Interfering Sources VI-3141
B. Daku - University of Saskatchewan, USA
C. McIntyre - Defense Research Establishment Atlantic, USA

Estimation of Delay and Doppler by Wavelet Transform VI-3145
K. Ho, Y. Chan, M. Johnson - Royal Military College of Canada, Canada

Blind Deconvolution of Echosounder Envelopes VI-3149
D. Caughey, R. Kirlin - University of Victoria, Canada

The Bootstrap Applied to Passive Acoustic Aircraft Parameter Estimation VI-3153
D. Reid, A. Zoubir, B. Boashash - Queensland University of Technology, Australia

UA4 Sonar Beamforming

Capability of Minimum Variance Beamformer for Direction Finding for Fast Perturbations of Array Sensors VI-3157
G. Serebryakov - State University, Russia

Linear Programming in Spectral Estimation: Application to Array Processing VI-3161
J. Fuchs - IRISA - Universite de Rennes I, France

Interpolated Array Minimum Variance Beamforming For Correlated Interference Rejection VI-3165
A. Zeira - Signal Processing Technology, USA
B. Friedlander - University of California at Davis, USA

Surface-Mounted Acoustic Vector-Sensor Array Processing VI-3169
M. Hawkes, A. Nehorai - Yale University, USA

Spatial Processing Technique Adaptive Beamforming (SPTABF) Via Compactly Supported Orthonormal Wavelets VI-3173
S. Chang, C. Chang, C.L. Chang, W. Fang - National Taiwan Ocean University, Taiwan

Beamforming Design with Partially Adaptive and Minimum Redundancy Arrays VI-3177
L. Lo Presti, G. Montalbano, G. Povero - Politecnico di Torino, Italy

Rank Enhancement for Eigenstructure Based Direction Finding Using Arrays with Non-Uniform Element Responses VI-3181
B. Jeffs - Brigham Young University, USA

Coarrray Analysis of Wideband Pulse-Echo Imaging Systems VI-3185
F. Ahmad, S. Kassam - University of Pennsylvania, USA

A Novel High-Resolution Algorithm for Complex-Direction Search VI-3189
R. Qiu, I. Lu - Polytechnic University, USA

Broadband Eigenvector Methods for Towed Array Shape Estimation with a Single Source VI-3193
J. Smith, Y. Leung, A. Cantoni - Australian Telecommunications Research Institute, Australia

VLSI1 Novel Approaches Towards DSP Application

A High Performance Microphone Array System for Hearing Aid Applications VI-3197
A. Wang, K. Yao, R. Hudson, D. Korompis - University of California at Los Angeles, USA

Control and Data Flow Aspects in the Design of a Wireless Data Radio Modem: A Case Study VI-3201
G. Post, O. Mauss - Aachen University of Technology, Germany

Variance Mismatch: Identifying Random Test Resistance in DSP Datapaths VI-3205
L. Goodby, A. Orailoglu - University of California at San Diego, USA

Digital Oscillators Over Finite Fields VI-3209
W. Chren - Grand Valley State University, USA
B. Wyman - Ohio State University, USA

Simultaneous Speech Segmentation and Phoneme Recognition Using Dynamic Programming VI-3213
R. Bajwa, R. Owens - Pennsylvania State University, USA
T. Kelliher - Westminster College, USA

Engineering Multirate Convolutions for Radar Imaging VI-3217
L. Bierens, E. Deprettere - Delft University of Technology, The Netherlands

A Fast Full-Search Motion Estimation Method for Programmable Processors with a Multiply-Accumulator VI-3221
Y. Naito, T. Miyazaki, I. Kuroda - NEC Corporation, Japan

Parallel Computation of Discrete Legendre Transforms VI-3225
M. Aburdene - Bucknell University, USA
J. Dorband - Goddard Space Flight Center, USA

New Look-Ahead Algorithm for Pipelined Implementation of Recursive Digital Filters VI-3229
A. Shaw, M. Imtiaz - Wright State University, USA

Telekom's MAGENTA Algorithm for En-/Decryption in the Gigabit/sec Range VI-3233
K. Huber, S. Wolter - German Telekom, Germany

VLSI2 Architectures for Video, Speech, and Filtering Applications

Optoelectronic Image Processor for Multiresolution Gabor Filtering VI-3236
H. Urey, W. Rhodes, S. DeWeerth, T. Drabik - Georgia Institute of Technology, USA

Low-Power Video Encoder-Decoder Using Wavelet/TSVQ with Conditional Replenishment VI-3240
W. Namgoong, N. Chaddha, T. Meng - Stanford University, USA

Low-Complexity Algorithm and Architecture for Block-Based Motion Estimation via One-Bit Transforms VI-3244
B. Natarajan, V. Bhaskaran, K. Konstantinides - Hewlett-Packard Laboratories, USA

A Binary Block Matching Architecture with Reduced Power Consumption and Silicon Area Requirement VI-3248
M. Mizuki, U. Desai, I. Masaki, A. Chandrakasan - MIT, USA

A Programmable Application-Specific CELP Processor with Parallel Architectures VI-3252
A. Suen, J. Wang, B. Liu - National Cheng Kung University, ROC

Efficient Realizations of Analysis and Synthesis Filters Based on the 2-D Discrete Wavelet Transform VI-3256
C. Chakrabarti, C. Mumford - Arizona State University, USA

Design Tradeoffs in CMOS FIR Filters VI-3260
C. Nagendra, M. Irwin - Pennsylvania State University, USA

M-Array VLSI Architecture for Order Statistic Filters VI-3264
C. Lin, C. Kuo - National Chung Cheng University, Taiwan

Time-Area Efficient Multiplier-free Recursive Filter Architectures for FPGA Implementations VI-3268
M. Shajaan, J. Sorensen - Technical University of Denmark, Denmark

VLSI3 DSP Architectures

Comparison of Two Architectures for Implementation of the Discrete Cosine Transform VI-3272
I. Sundsbo, G. Lokken Hansen, E. Aas - The Norwegian Institute of Technology, Norway

Multidimensional Systolic Arrays for Multidimensional DFTs VI-3276
H. Lim, E. Swartzlander - University of Texas at Austin, USA

ASAP-A 2-D DFT Processor and Architecture VI-3280
J. Mellott, M. Lewis, F. Taylor - University of Florida, USA
P. Coeffield - USAF Wright Labs, USA

A New Survivor Memory Management Method in Viterbi-Decoders: Trace-Delete Method and its Implementation VI-3284
S. Jung, H. Choi - Sung-Kyun-Kwon University, Korea
M. Lee - Samsung Electronics, Korea

Efficient Standard Basis Reed-Solomon Encoder VI-3287
S. Jain - Intel Corporation, USA
K. Parhi - University of Minnesota, USA

An Area-Efficient VLSI Architecture for Decoding of Reed-Solomon Codes VI-3291
J. Hsu, C. Wang - National Tsing Hua University, ROC

High-Speed Parallel Hard and Soft-Decision Golay Decoder: Algorithm and VLSI-Architecture VI-3295
W. Cao - Fraunhofer-Institute for Integrated Circuits, Germany

Granularly-Pipelined Cordic Processors for Sine and Cosine Generators VI-3298
S. Wang - University of Texas at Austin, USA
V. Piuri - Politecnico di Milano, Italy
E. Swartzlander - University of Texas at Austin, USA

RNIW: A Novel General-Purpose DSP Architecture VI-3302
H. Qing, H. Chao-Huan - Chinese Academy of Sciences, China

VLSI4 Design Aids for Signal Processing Systems

An Extension of Multi Dimensional Synchronous Dataflow to Handle Arbitrary Sampling Lattices VI-3306
P. Murthy, E. Lee - University of California at Berkeley, USA

On Retiming of Multirate DSP Algorithms VI-3310
V. Zivojnovic, R. Schoenen, H. Meyr - Aachen University of Technology, Germany

Real-Time Scheduling for Pipelined Execution of Data Flow Graphs on a Realistic Multiprocessor Architecture VI-3314
J. Jonsson, J. Vasell - Chalmers University of Technology, Sweden

Optimal Communication Scheduling Based on Collision Graph Model VI-3318
D. Surma, S. Tongsima, E. Sha - University of Notre Dame, USA

Hierarchization of a Priori Defined Dataflow Subgraphs VI-3322
M. Willems, V. Zivojnovic, S. Ritz - Aachen University of Technology, Germany

Instruction-Level Power Profiling VI-3326
H. Mehta, R. Owens, M. Irwin - Pennsylvania State University, USA

Two-Dimensional Retiming with Low Memory Requirements VI-3330
T. Denk, M. Majumdar, K. Parhi - University of Minnesota, USA

Knowledge-Based Transformation Ordering: Implementing a Linear Computation Using an Arbitrary Low Power at an Arbitrary High Throughput VI-3334
M. Srivastava - AT&T Bell Laboratories, USA
M. Potkonjak - NEC USA, USA

A Hierarchical Approach to Transistor-Level Power Estimation of Arithmetic Units VI-3338
J. Satyanarayana, K. Parhi - University of Minnesota, USA

Calypso: An Implementation of the Order Graph Method VI-3342
G. Cato, D. Reeves - North Carolina State University, USA

PRNN1 Neural Networks for Speech Processing

A Telephone-Based Directory Assistance System Adaptively Trained Using Minimum Classification Error / Generalized Probabilistic Descent VI-3346
E. McDermott - ATR HIP Research Laboratories, Japan
E. Woudenberg - ATR International, Japan
S. Katagiri - ATR IT Research Laboratories, Japan

REMAP - Experiments with Speech Recognition VI-3350
Y. Konig, H. Bourlard, N. Morgan - ICSI, USA

Strategies for Reducing the Complexity of a RNN Based Speech Recognizer VI-3354
K. Kasper, H. Reininger, H. Wust - Institut fur Agewandte Physik, Germany

Environment-Independent Continuous Speech Recognition Using Neural Networks and Hidden Markov Models VI-3358
D. Yuk, C. Che, L. Jin, Q. Lin - Rutgers University, USA

Comparison Between Two Hybrid HMM/MLP Approaches in Speech Recognition VI-3362
C. Ris, V. Fontaine, H. Leich - Faculte Polytechnique de Mons, Belgium

GPD Training of the State Weighting Functions in Hidden Control Neural Network VI-3366
K. Na, S. Chae - Seoul National University, ROK

Continuous Mandarin Speech Recognition Using Heirarchical Recurrent Neural Networks VI-3370
Y. Liao, W. Chen, S. Chen - National Chiao Tung University, ROC

Probabilistic Mapping Networks for Speaker Recognition VI-3374
H. Li, Y. Gong - CRIN-CNRS/INRIA - Lorraine, France

Shift-Tolerant K-Subspaces for Phoneme Recognition VI-3378
D. Wu, J. Gowdy - Clemson University, USA

Speaker-Adaptation in a Hybrid HMM-MLP Recognizer VI-3382
J. Neto, C. Martins, L. Almeida - INESC/IST, Portugal

PRNN2 Neural Network Techniques for Image Processing and Analysis

Automatic Target Recognition Using Modularly Cascaded Vector Quantizers and Multilayer Perceptrons VI-3386
L. Chan, N. Nasrabadi, V. Mirelli - SUNY at Buffalo, USA

Distortion Equalized Fuzzy Competitive Learning for Image Data Vector Quantization VI-3390
D. Butler - Bolton Institute, UK
J. Jiang - Loughborough University, UK

Revisiting Boltzmann Learning: Parameter Estimation in Markov Fields VI-3394
L. Hansen, L. Andersen, U. Kjems, J. Larsen - Technical University of Denmark, Denmark

Morphological/Rank Neural Networks and their Adaptive Optimal Design for Image Processing VI-3398
L. Pessoa, P. Maragos - Georgia Institute of Technology, USA

Nonlinear Color Image Restoration using Extended Redial Basis Function Networks VI-3402
I. Cha, S. Kassam - University of Pennsylvania, USA

A Unification of Relaxation Labeling and Associative Memory VI-3406
X. Zhuang - University of Missouri- Columbia, USA
Y. Zhao - University of Illinois- Urbana- Champaign, USA

Motion-Based Segmentation by Principal Singular Vector (PSV) Clustering Method VI-3410
S. Kung, Y. Lin, Y. Chen - Princeton University, USA

Multiplierless Neural Networks for Application to Digital Video VI-3414
M. Cavanagh, V. Jain - University of South Florida, USA

Neural Network Object Recognition for Inspection of Patient Setup in Radiation Therapy Using Portal Images VI-3418
S. Young, C. Sibata, A. Ho, K. Shin - Roswell Park Cancer Institute, USA

Efficient Learning of Standard Finite Normal Mixtures for Image Quantification VI-3422
Y. Wang - Georgetown University Medical Center, USA
T. Adali - University of Maryland at College Park, USA

Volume VI

PRNN3 Neural Networks and Pattern Recognition 1

A Chain Coding Approach for Real-Time Recognition of On-line Handwritten Characters VI-3426
H. Yuen - Hong Kong Polytechnic University, Hong Kong

A Unified Approach to Hierarchical Classification VI-3430
D. Kil, F. Shin - Loral Defense Systems, USA

A Soft-Decision Approach for Symbol Segmentation Within Handwritten Mathematical Expressions VI-3434
S. Lehmberg, H. Winkler, M. Lang - Munich University of Technology, Germany

HMM-Based Handwritten Symbol Recognition Using On-Line and Off-Line Features VI-3438
H. Winkler - Munich University of Technology, Germany

From Characters to Words: Dynamical Segmentation and Predictive Neural Networks VI-3442
S. Garcia-Salicetti - Institut National des Telecommunications, France
P. Gallinari - LAFORIA National des Telecommunications, France
B. Dorizzi, Z. Wimmer, J. Gentric - Institut National des Telecommunications, France

A New Data Reduction Algorithm for Pattern Classification VI-3446
H. Tahani, B. Plummer, N. Hamamalini - University of Missouri, USA

Fast Online Video Image Sequence Recognition with Statistical Methods VI-3450
M. Schuster, G. Rigoll - University of Duisburg, Germany

Handwritten Word Recognition Using Dynamic Matching with Variable Duration VI-3454
G. Kim, V. Govindaraju, S. Srihari - CEDAR/State University of New York at Buffalo, USA

Scale and Rotational Invariant Object Recognition Using Fractal Transformations VI-3458
G. Neil, K. Curtis - University of Nottingham, UK

Optimal Discriminant Feature-Based Waveform Recognition with Neural Networks VI-3462
J. Miao, M. Clements - Georgia Institute of Technology, USA

PRNN4 Neural Networks and Pattern Recognition 2

Parsing in Attributed Grammars: Application to Automatic Sleep Analysis VI-3466
A. Fred, J. Leitao - Instituto Superior Tecnico, Portugal

Affine Invariant Shape Representation and Recognition Using Gaussian Kernels and Multi-dimensional Indexing VI-3470
J. Ben-Arie - University of Illinois at Chicago, USA
Z. Wang, K. Rao - Illinois Institute of Technology, USA

Road Scene Analysis from Perspective Image VI-3474
J. Yang, S. Ozawa - Keio University, Japan

Writer Dependent Recognition of On-Line Unconstrained Handwriting VI-3478
J. Subrahmonia, K. Nathan, M. Perrone - IBM T.J. Watson Research Center, USA

Duration Modelling Results for an On-Line Handwriting Recognizer VI-3482
A. Senior, J. Subrahmonia, K. Nathan - IBM T.J. Watson Research Center, USA

Properties of a Hand-Printed Chinese Character Recognizer Based on Contextual Vector Quantization VI-3486
P. Chee, S. Leung, P. Wong, C. Chan - The University of Hong Kong, Hong Kong

Optimized Concensus Theory VI-3490
J. Benediktsson, J. Sveinsson, P. Swain - University of Iceland

Chinese All Syllables Recognition Using Combination of Multiple Classifiers VI-3494
L. Zhou, S. Imai - Tokyo Institute of Technology, Japan

Similarity Analysis of Cardiac Signals Using Temporal Features VI-3498
B. de Jesus, L. Lucke - University of Minnesota, USA

Initialization of Hidden Markov Models for Unconstrained On-Line Handwriting Recognition VI-3502
K. Nathan, A. Senior, J. Subrahmonia - IBM T.J. Watson Research Center, USA

PRNN5 Nonlinear System Identification

Diagnosing and Correcting System Anomalies with a Robust Classifier VI-3506
J. Hampshire - Carnegie Mellon University, USA
D. Watola - Jet Propulsion Laboratory, USA

Blind Separation of Delayed Sources Based on Information Maximization VI-3509
K. Torkkola - Motorola, USA

Modeling Nonlinear Systems With Cellular Neural Networks VI-3513
F. Puffer, R. Tetzlaff, D. Wolf - Institut fur Angewandte Physik, Germany

Energy Demodulation of Two-Component AM-FM Signals with Application to Speaker Separation VI-3517
B. Santhanam, P. Maragos - Georgia Institute of Technology, USA

Adaptive Non-linear System Identification Using Minimal Radial Basis Function Neural Network VI-3521
L. Yingwei, N. Sundararajan, P. Saratchandran - Nanyang Technological University, Singapore

Expanding Gaussian Kernels for Multivariate Conditional Density Estimation VI-3525
D. Davis, J. Hwang - University of Washington, USA

Fast Adaptive IIR-MLP Neural Networks for Signal Processing Applications VI-3529
P. Campolucci, A. Uncini, F. Piazza - University of Ancona, Italy

Information Geometry of Maximum Partial Likelihood Estimation for Channel Equalization VI-3533
J. Xuan, T. Adali, X. Liu - University of Maryland at College Park, USA

Statistical Analysis of the Single-Layer Backpropagation Algorithm with Bias Terms: Mean Weight Behavior VI-3537
N. Bershad, J. Shynk - University of California, USA

Fast Nonlinear Adaptive Filtering Using a Partial Window Conjugate Gradient Algorithm VI-3541
A. Birkett, R. Goubran - Carleton University, Canada

PRNN6 Modeling Theory and Architecture in Neural Networks

Building an Application Framework for Speech and Pen Input Integration in Multimodal Learning Interfaces VI-3545
M. Vo, C. Wood - Carnegie Mellon University, USA

TOM, A New Temporal Neural Net Architecture for Speech Signal Processing VI-3549
S. Durand, F. Alexandre - Crin-CNRS/ INRIA Lorraine, France

A Probabilistic Decision-Based Neural Network for Locating Deformable Objects and its Applications to Surveillance System and Video Browsing VI-3553
J. Lin, Y. Chan, S. Kung - Princeton University, USA

Self-Organizing Feature Maps with Perfect Organization VI-3557
L. Owsley, L. Atlas - University of Washington, USA
G. Bernard - Boeing Commercial Airplane Group, USA

Partial Likelihood for Real-Time Signal Processing VI-3561
T. Adali - University of Maryland at College Park, USA
K. Sonmez - University of Maryland Baltimore County, USA
X. Liu - University of Maryland at College Park, USA

Robust Learning Algorithms for Nonlinear Filtering VI-3565
D. Neumerkel, R. Shorten - Daimler-Benz AG, Germany
A. Hambrecht - CEGELEC AEG, Germany

Advances in Using Hierarchical Mixture of Experts for Signal Classification VI-3569
V. Ramamurti, J. Ghosh - The University of Texas at Austin, USA

Comparison of a Neural Network Detector vs Neyman-Pearson Optimal Detector VI-3573
D. Andina, J. Sanz-Gozanlez - Universidad Politecnica Madrid, Spain

An Algorithm to Determine Hidden Markov Model Topology VI-3577
R. Vasko, A. El-Jaroudi, J. Boston - University of Pittsburgh, USA

Wreath Products for Image Processing VI-3581
D. Healy, Jr., G. Mirchandani, T. Olson, D. Rockmore - Dartmouth College, USA

This material may be protected by Copyright law (Title 17 U.S. Code)

BUILDING AN APPLICATION FRAMEWORK FOR SPEECH AND PEN INPUT INTEGRATION IN MULTIMODAL LEARNING INTERFACES

Minh Tue Vo and Cindy Wood

Interactive Systems Laboratories
Carnegie Mellon University
Pittsburgh, PA 15213-3890, U.S.A.
Email: {tue, caw}@cs.cmu.edu

ABSTRACT

While significant advances have been made in recent years to improve speech recognition performance, and more recently, gesture and handwriting recognition as well, speech- and pen-based systems have still not found broad acceptance in everyday life. One reason for this is the inflexibility of each input modality when used alone. Human communication is very natural and flexible because we can take advantage of a multiplicity of communication signals working in concert to supply complementary information or increase robustness with redundancy. In this paper we present a multimodal interface capable of jointly interpreting speech, pen-based gestures, and handwriting in the context of an appointment scheduling application. The interpretation engine based on semantic frame merging correctly interprets 80% of a multimodal data set assuming perfect speech and gesture/handwriting recognition; in the presence of recognition errors the interpretation performance is in the range of 35-62%. A dialog processing scheme uses task domain knowledge to guide the user in supplying information and permits human-computer interactions to span several related multimodal input events.

1. INTRODUCTION

Speech recognition is coming of age and is being incorporated into modern computer systems. Although pen input processing is still in its infancy compared to speech, pen-based systems have begun to appear in the form of personal digital assistants (PDAs) and pen laptops. However, current speech and pen systems are still not very popular and did not live up to the promise of bringing the power of computing to the population at large. Their failure is partly due to users' frustrations generated by inadequate recognition performance, especially for handwriting recognition. However, this failure is also an indication that current systems are too inflexible for the important task of facilitating the use of computers for everyone, especially people who may not be computer-literate. PDA users frustrated by gesture and handwriting recognition errors would naturally wish they could talk to their machines to correct the errors or simply to avoid having to write everything. Likewise users of speech-enabled systems quickly find out that there are tasks that cannot be conveniently expressed by spoken commands but would be enormously simplified by the ability to point to or circle objects on the screen in addition to speaking commands. These intuitive assertions were confirmed by a user study [1] conducted at Carnegie Mellon University, in which people interacting with a computer much preferred a combination of both speech and gestures over speech or gestures alone. Our own experiments with a calendar interface also show that when given a choice, people tend to use the communication method (or methods) most suitable for the task at hand and freely switch among available methods.

Research efforts at our Interactive Systems Laboratories (Carnegie Mellon University and University of Karlsruhe) are focused on producing a sensible and useful user interface by integrating multiple input modalities, rather than building better speech and gesture recognizers alone. An interface supporting the kind of highly flexible interaction we envision must be capable of integrating information from both speech and non-verbal input sources to arrive at a correct understanding of complete multimodal events. Some of our initial works along this line have been reported in previous publications [2][3][4][5]. This paper presents an approach to integrating speech, pen-based gestures, and handwriting, in the context of an appointment scheduling application. A multimodal interpretation engine jointly interprets information from all input sources by merging semantic frames. A domain-independent dialog processor maintains context information across input events. We evaluated the system with data collected in a user study conducted using the Wizard-of-Oz paradigm [6].

2. JEANIE: A MULTIMODAL CALENDAR

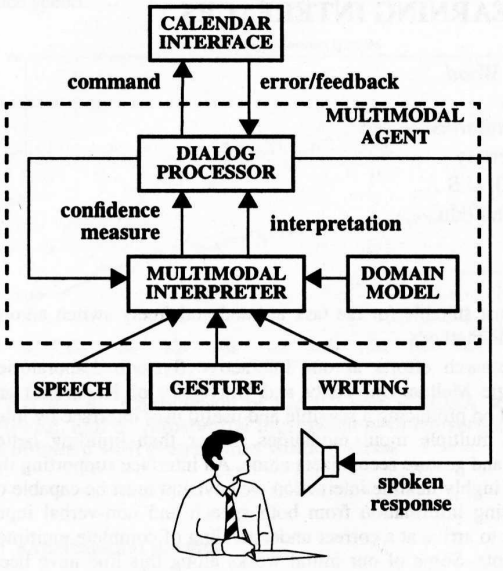
We have developed a prototype of a multimodal interface for an appointment scheduling program. A person using our Jeanie multimodal calendar can employ any combination of spoken input, gesturing with a pen on a touch-sensitive screen, or handwritten words to interact with the system. In typical scenarios, the user might say "Schedule a meeting on Monday" while at the same time drawing a box on the calendar to indicate where the new meeting should be inserted; write words on the newly scheduled meeting to annotate it; draw a cross on another meeting to cancel it; or point to a meeting and say "Reschedule this on Tuesday" or simply draw an arrow from that meeting to the new time slot on Tuesday.

An earlier version of this system was described in [4]. The interpretation engine in that system was based on an information-theoretic connectionist network [7] capable of incremental learning during use. Although that approach worked well for the preliminary system designed to explore multimodal interpretation, we found that the connectionist network was difficult to scale up when the task domain representation was significantly expanded in the new version. A semantic frame merging scheme was found to work much better on the larger and more complex action space, although we had to sacrifice the incremental learning capability.

Figure 1 shows a block diagram of the Jeanie system. The individual modality components (speech, gesture, and handwriting recognizers) are separate modules that can be replaced with ease. The calendar interface embodies domain knowledge and serves to isolate task-specific components. The heart of the system is the multimodal agent which interprets user input and performs

0-7803-3192-3/96 $5.00©1996 IEEE

FIGURE 1: The Jeanie multimodal calendar



requested operations through the calendar interface. This section describes the calendar interface and the modality processors. The next section presents the multimodal agent and the multimodal interpretation algorithm.

2.1. Calendar Interface

The object-oriented calendar interface is based on ICAL, a public domain calendar program developed by Sanjay Ghemawat at MIT. We extended ICAL to support a weekly view and client/server operations. The multimodal agent acting as a client of the calendar interface can query it for information such as gesture contexts (see Section 2.2) and ask the interface to carry out scheduling operations which include adding, removing, and changing appointments and notes.

2.2. Modality Processors

Inputs captured by a speech recorder and a touch-sensitive screen are processed by separate recognizers for each modality. Semantic frames extracted from recognizer outputs are merged by the multimodal interpreter to arrive at a unified interpretation as described in Section 3.

Speech. Our speech recognition subsystem is based on the recognition front-end of the JANUS speech translation system [8][9] which is capable of processing speaker-independent, spontaneous speech and was trained on human-human dialogs in the appointment scheduling domain. On a small data set of 128 utterances collected in user study experiments, the word recognition rate was 82%; if we take into account insertion, substitution, and deletion errors, the word accuracy is 76%.

Gesture. In earlier implementations we employed a gesture recognition module [2] based on a TDNN classifier [10]. The present implementation increases flexibility by decomposing gestures into sequences of strokes recognized as basic shapes such as lines, arcs, arrows, circles, crosses... Each gesture component is augmented by *gesture contexts* indicating spatial relationships between the gesture and nearby objects in the calendar interface. The current implementation of the gesture recognizer uses simple template matching and does not perform well on noisy input. A version based on the handwriting recognition algorithm described below is being developed and should greatly improve recognition performance.

Handwriting. Our handwriting recognizer developed by Stefan Manke at University of Karlsruhe based on the MS-TDNN [11] is capable of processing writer-independent, continuous (cursive) handwriting [12] at a recognition rate of over 90% on a 20,000-word vocabulary. Handwriting recognition is performed only when the gesture recognizer cannot identify the input strokes as basic shapes. This simple heuristic works fine for true cursive handwriting but can mistake writing strokes for gestures when it encounters printed block letters. This was unfortunately the case for some of the data collected in our user study; in addition most of the collected handwriting included uppercase letters for which the recognizer was not trained. These facts combined with gestures that were not adequately covered by templates led to a combined gesture/handwriting recognition rate of only 32%.

3. MULTIMODAL INTERPRETATION BY SEMANTIC FRAME MERGING

The multimodal interpreter is responsible for producing an interpretation of user intent (i.e., a command to send to the calendar interface) from the output of the modality processors. We represent this interpretation as a frame consisting of slots specifying pieces of information such as the action to carry out or the date and time of a meeting. Recognition output from the modality processors are parsed into partially filled frames that are merged together to produce the combined interpretation. The underlying algorithm is domain-independent although the encoding of information in frames must necessarily depend on the task domain.

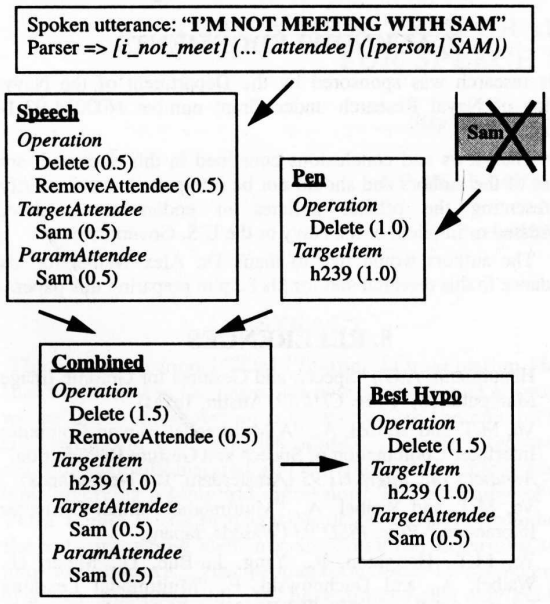
This frame merging technique represents a much extended and improved implementation of the approach sketched in [2]. It leads to uniform handling of high-level information from all input sources, which is very important for modularity and extensibility. To add another input modality we need only provide a module to convert low-level recognizer output to a partially filled frame to be merged with others. In addition, context information can be retained across input events by merging with previous interpretation frames as implemented in our dialog processor (Section 3.3).

3.1. Parsing Inputs From Individual Modalities

The text string output from the speech recognizer is processed by the Phoenix semantic parser developed by Ward [13]. The parser compiles a grammar specifying semantically meaningful fragments of text into an efficient recurrent transition network that identifies these fragments in the recognized utterance. It is capable of skipping unknown words and unmatched fragments and can therefore deal gracefully with ungrammatical sentences. The concepts identified by the parser serve to determine which slots of the speech frame should be filled and what scores should be assigned to the slot values.

The gesture/handwriting recognizers produce a sequence of gesture/handwriting components. The same gesture shape may mean different things depending on the calendar context surrounding it; for instance, a line across an appointment may indicate deletion but the same line falling on an empty time slot may visually describe a new meeting to be scheduled. In the present implementation of Jeanie, gestures and handwritten words are encoded into frames according to simple hard-coded heuristics. The development of a grammar to drive this process may become necessary for more complex gestures and operations.

FIGURE 2: Example of interpretation using frames



3.2. Merging Frames and Extracting Hypotheses

Frames are merged by taking the union of the sets of values filling each slots and adding corresponding scores. Embedded frames are merged recursively. This produces an aggregate frame encoding all alternative interpretation hypotheses. The multimodal interpreter extracts these hypotheses in descending order of aggregate scores and sends the best ones to the dialog processor. Figure 2 shows an example taken from our actual test data.

3.3. Dialog Processing

A dialog between a user and the system can span several related multimodal events, where each event may contain one or more spoken utterances and/or pen gestures. The dialog processor maintains a single frame per dialog to hold all information obtained since the start of the dialog; new information from each additional input event is merged into this frame and the scores are weighted appropriately so that frames produced by different input events in the same dialog contribute equally to the end result.

The domain-independent dialog processor also consults the domain interface (i.e., the calendar in this case) to filter out hypotheses that would cause domain errors such as missing or conflicting parameters. If it fails to find an acceptable hypothesis with a high enough score, it uses the error information returned by the domain interface to construct an appropriate feedback message to guide the user (only this process depends on the task domain). As soon as a good hypothesis is found, the dialog processor sends it to the domain interface as the operation to carry out and starts a new dialog. The user has the option of undoing the operation just performed if it is incorrect.

4. USER STUDY AND DATA COLLECTION

In order to create a useful and robust multimodal system, we need to find out how people would use a system with such capabilities. We have begun a series of experiments following the Wizard-of-Oz paradigm [6], in which the test subjects were presented with the Jeanie calendar interface and a hidden operator takes place of the real system in interpreting and carrying out multimodal commands. In parallel with the design of Jeanie, we conducted pilot experiments that contributed to the design of the frame representation for our task domain and yielded preliminary test data to evaluate the system.

Before the pilot experiment we conducted some small preliminary experiments to find out the types of gestures and speech people would use to make changes to a schedule, and to see if one modality was preferred over the other for certain tasks. That information was used to design pilot experiments that would illicit a mixture of speech and gestures from the test subjects. To get a variety of gestures that people might use the test subjects were first asked to make changes to a schedule using gestures only in any way they like. The same experiment was then repeated asking subjects to use speech only.

For the pilot experiments subjects were asked to make the changes by either giving voice commands or drawing on the touch screen or any combination of the two. Based on observations from the preliminary experiments we settled on instructing subjects by giving them a printed calendar with handwritten changes using as few words as possible and varying the wording to avoid speech contamination.

8 test subjects were used and a total of 13 trials were done for the preliminary and first set of pilot tests. The second pilot experiment consisted of 8 trials for 4 subjects using two completely different calendars to include a greater variety of tasks. In the pilot tests it was observed that some subjects preferred to use gestures more often, some preferred speech, others mixed the two. It was also observed that some tasks proved to be much easier to do by one method or the other; in those cases subjects chose the easier method regardless of general preference.

The pilot experiments showed, among other things, that test subjects often referred to a calendar event by using the name of a person or event, not just by a date and time; in addition, some subjects gave implied commands such as "On Tuesday I'm meeting with Bill, Tony, and Melanie" instead of "Add Tony and Melanie to the meeting with Bill on Tuesday". Test subjects with little computer experience tended to use implied commands most often, although we still do not have enough data to confirm such trends. The above observations significantly influenced the design of the frame representation.

5. PRELIMINARY PERFORMANCE EVALUATION

From out pilot Wizard-of-Oz experiments we obtained a small data set consisting of 185 multimodal events (77 speech-alone events, 57 pen-alone events, and 51 combination events). We transcribed each spoken utterance and pen gesture/handwriting as well as passing them through the modality recognizers. Results for all-transcribed inputs give us an idea of the performance of the multimodal interpreter, while results for recognized inputs offer some indication of performance degradation due to recognition errors.

The top-scored hypothesis produced by the interpreter for each input event is classified as *perfect*, *ok* (producing the intended result despite not being strictly correct), *ambiguous* (score too close to the next best hypothesis), *partially bad* (mostly correct except for one parameter), or *bad*. Table 1 shows the performance results for the pilot data. The pen data from the first pilot experiment could be transcribed but was unusable for recognition because of a flaw in the recording module which was cor-

TABLE 1: Multimodal interpretation performance

(TS/TP: transcribed speech/pen, RS/RP: recognized speech/pen)

	number of inputs	perfect (%)	ok (%)	ambi. (%)	partially bad (%)	bad (%)
TS/TP	185	78	2	7	3	10
RS/TP	185	57	5	16	6	16
TS/RP	100	49	2	26	9	14
RS/RP	100	28	7	29	12	24

rected for the second pilot experiment; because of this the bottom half of the table shows results for only the data set collected in this second experiment.

In the absence of recognition errors the multimodal interpreter will do the right thing 80% of the time (total of the *perfect* and *ok* columns), which is adequate for a usable system. At the other end of the spectrum the interpretation rate for all-recognized inputs is only 35%, a direct result of the poor 32% gesture/handwriting recognition rate.

The second row of the table shows that with a 76% word accuracy in speech recognition the interpretation rate drops to 62%. A closer analysis reveals that of the 18% that went from good to bad, about 15% occurred on speech-only events and only 3% occurred on combined speech/pen events. To put it another way, 27 of 77 (=35%) speech-only events caused interpretation errors due to speech recognition errors, compared to only 6 of 51 (=12%) combined events. On the pen-input side (only for the second data set of 100 input events), 11 of 18 (=61%) pen-only events caused interpretation errors due to gesture/handwriting recognition errors, compared to 19 of 39 (=49%) combined events. The sample is too small to assert any statistical significance, but this may be an indication of cross-modal redundancy at work.

We also observe that one significant effect of recognition errors is to increase the number of ambiguous outputs, more than the number of incorrect outputs. Recognition errors obviously reduce the amount of information available for interpretation, thereby causing more confusion. In interactive mode, the dialog processor can prompt the user with judicious guidance messages to supply clarifying information that could resolve ambiguities.

6. CONCLUSION

We have presented an architecture capable of joint interpretation of multimodal inputs and an example of a speech- and pen-enabled application using that architecture. The interpretation engine combines multiple input sources in a uniform and flexible way. The dialog processor maintains context information across input events and produces context-sensitive feedback to the user. The interpretation error rate is not unduly high but recognition errors may degrade performance significantly; however, we have seen indications that cross-modal redundancy can partly compensate for this degradation.

The pilot experiments have shown that the type of tasks we used in our Wizard setup will provide data for gesture, speech and a mixture of the two and will therefore be useful for the development and training of the Jeanie system. We are working on collecting more data to evaluate the system more fully, improve recognition rates, and refine parsing grammars. Experiments with the actual system in the loop rather than a Wizard operator are also being planned. Future studies should include experiments to investigate further the effect of cross-modal synergy in the presence of recognition errors to prove conclusively our intuitive assertions about the benefits of multimodal integration.

7. ACKNOWLEDGEMENTS

This research was sponsored by the Department of the Navy, Office of Naval Research under Grant number N00014-93-1-0806.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Navy or the U.S. Government.

The authors would like to thank Dr. Alex Waibel for his guidance in this research and for his help in preparing this paper.

8. REFERENCES

[1] Hauptmann, A.G., "Speech and Gestures for Graphic Image Manipulation," *Proc. CHI'89* (Austin, Texas).

[2] Vo, M.T. and Waibel, A., "A Multimodal Human-Computer Interface: Combination of Speech and Gesture Recognition," *Adjunct Proc. InterCHI'93* (Amsterdam, The Netherlands).

[3] Vo, M.T. and Waibel, A., "Multimodal Human-Computer Interaction;" *Proc. ISSD'93* (Waseda, Japan).

[4] Vo, M.T., Houghton, R., Yang, J., Bub, U., Meier, U., Waibel, A., and Duchnowski, P., "Multimodal Learning Interfaces," *Proc. ARPA SLT Workshop 95* (Austin, Texas).

[5] Waibel, A., Vo, M.T., Duchnowski, P., and Manke, S., "Multimodal Interfaces," *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, McKevitt, P. (Ed.), Vol. 10, Nos. 3-4, 1995.

[6] Salber, D. and Coutaz, J., "Applying the Wizard of Oz Technique to the Study of Multimodal Systems," *Proc. EWHCI'93* (Moscow, Russia).

[7] Gorin, A.L., Levinson, S., Gertner, A., and Goldman, E., "Adaptive Acquisition of Language," *Computer, Speech and Language*, Vol. 5, No. 2, pp. 101-132, April 1991.

[8] Woszczyna, M., Aoki-Waibel, N., Buø, F.D., Coccaro, N., Horiguchi, K., Kemp, T., Lavie, A., McNair, A., Polzin, T., Rogina, I., Rose, C.P., Schultz, T., Suhm, B., Tomita, M., and Waibel, A., "JANUS 93: Towards Spontaneous Speech Translation," *Proc. ICASSP'94* (Adelaide, Australia).

[9] Suhm, B., Geutner, P., Kemp, T., Lavie, A., Mayfield, L., McNair, A., Rogina, I., Schultz, T., Sloboda, T., Ward, W., Woszczyna, M., and Waibel, A., "JANUS: Towards Multilingual Spoken Language Translation," *Proc. ARPA SLT Workshop 95* (Austin, Texas).

[10] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K., "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 3, 1989, pp. 328-339.

[11] Haffner, P. and Waibel, A., "Multi-State Time Delay Neural Networks for Continuous Speech Recognition," *Advances in Neural Network Information Processing Systems 4*, Morgan Kaufmann Publishers, 1992, pp. 135-142.

[12] Manke, S., Finke, M., and Waibel, A., "The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System," *Advances in Neural Information Processing Systems 6*, Morgan Kaufmann, 1994.

[13] Ward, W., "Understanding Spontaneous Speech: the Phoenix System," *Proc. ICASSP'91* (Toronto, Canada).

APPENDIX B

LIBRARY OF CONGRESS | CATALOG

1. Conference proceedings

LCCN	98655050
Type of material	Periodical or Newspaper
Meeting name	ICASSP (Conference)
Main title	Conference proceedings / the ... International Conference on Acoustics, Speech, and Signal Processing ; sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers.
Published/Created	Piscataway, NJ : Institute of Electrical and Electronics Engineers, c1995-c1996.
Publication history	1995-1996.
Description	2 v. : ill. ; 28 cm. Issued in 5 or more pts.
Current frequency	Annual
LC classification	TK7882.S65 I37a
Continues	ICASSP (Conference). ICASSP (DLC) 94655093 (OCoLC)31179684
Continued by	ICASSP (Conference). IEEE International Conference on Acoustics, Speech, and Signal Processing (DLC) 98655049 (OCoLC)37249153
Variant title	International Conference on Acoustics, Speech, and Signal Processing ... proceedings
Other title	ICASSP ... proceedings
Related names	Institute of Electrical and Electronics Engineers. IEEE Signal Processing Society.
LC Subjects	Speech processing systems--Congresses. Signal processing--Congresses. Electro-acoustics--Congresses. Underwater acoustics--Congresses.
Other Subjects	Electro-acoustics. Signal processing. Speech processing systems. Underwater acoustics.
Form/Genre	Conference papers and proceedings.
Additional formats	Also issued online by subscription, in PDF.
Dewey class no.	621.38/043/05
Other system no.	(OCoLC)ocm32748056
Invalid LCCN	sn 95037131
Additional Links	http://ieeexplore.ieee.org/xpl/conferences.jsp Search: Acoustics, speech, and signal processing ...

CALL NUMBER	TK7882.S65 I37a
Request in	Jefferson or Adams Building Reading Rooms

LC CATALOG
Library of Congress
101 Independence Ave., SE
Washington, DC 20540

Questions? Ask a Librarian:
https://ask.loc.gov/

APPENDIX C

LIBRARY CATALOG

Search Navigation 1 of 8

PERIODICAL OR NEWSPAPER

Conference proceedings

Full Record MARC Tags

000		02397cas a2200517 a 4500
001		11501706
005		20210221092333.0
008		950630d19951996njuar 1 0eng c
010	__	|**a** 98655050 |**z** sn 95037131
035	__	|**a** (OCoLC)ocm32748056
040	__	|**a** IND |**b** eng |**c** IND |**d** MYG |**d** DLC |**d** MYG |**d** IUL |**d** CUV |**d** OCLCQ |**d** OCLCF |**d** OCLCA |**d** OCL |**d** OCLCO |**d** COO |**d** OCLCA |**d** IUL
042	__	|**a** pcc
050	00	|**a** TK7882.S65 |**b** I37a
082	00	|**a** 621.38/043/05 |**2** 19
111	2_	|**a** ICASSP (Conference)
245	10	|**a** Conference proceedings / |**c** the ... International Conference on Acoustics, Speech, and Signal Processing ; sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers.
246	13	|**a** ICASSP ... proceedings
246	3_	|**a** International Conference on Acoustics, Speech, and Signal Processing ... proceedings
260	__	|**a** Piscataway, NJ : |**b** Institute of Electrical and Electronics Engineers, |**c** c1995-c1996.
300	__	|**a** 2 v. : |**b** ill. ; |**c** 28 cm.
310	__	|**a** Annual
362	0_	|**a** 1995-1996.
515	__	|**a** Issued in 5 or more pts.
530	__	|**a** Also issued online by subscription, in PDF.
590	__	|**a** SERBIB/SERLOC merged record

650	_0	|**a** Speech processing systems |**v** Congresses.
650	_0	|**a** Signal processing |**v** Congresses.
650	_0	|**a** Electro-acoustics |**v** Congresses.
650	_0	|**a** Underwater acoustics |**v** Congresses.
650	_7	|**a** Electro-acoustics. |**2** fast |**0** (OCoLC)fst00906339
650	_7	|**a** Signal processing. |**2** fast |**0** (OCoLC)fst01118281
650	_7	|**a** Speech processing systems. |**2** fast |**0** (OCoLC)fst01129243
650	_7	|**a** Underwater acoustics. |**2** fast |**0** (OCoLC)fst01161058
655	_7	|**a** Conference papers and proceedings. |**2** fast |**0** (OCoLC)fst01423772
710	2_	|**a** Institute of Electrical and Electronics Engineers.
710	2_	|**a** IEEE Signal Processing Society.
780	00	|**a** ICASSP (Conference). |**t** ICASSP |**w** (DLC) 94655093 |**w** (OCoLC)31179684
785	00	|**a** ICASSP (Conference). |**t** IEEE International Conference on Acoustics, Speech, and Signal Processing |**w** (DLC) 98655049 |**w** (OCoLC)37249153
850	__	|**a** DLC |**a** MCM
859	41	|**u** http://ieeexplore.ieee.org/xpl/conferences.jsp |**z** Search: Acoustics, speech, and signal processing ...
890	__	|**a** Conference proceedings (ICASSP) |**i** 98-655050
906	__	|**a** 7 |**b** cbc |**c** serials |**d** u |**e** ncip |**f** 19 |**g** n-oclcserc
920	__	|**a** Keep 1
985	__	|**e** eserial 200406
991	__	|**b** c-GenColl |**h** TK7882.S65 |**i** I37a |**w** SERIALS
992	__	|**b** B/L |**w** SERLOC

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK7882.S65 I37a
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

APPENDIX D

LIBRARY | CATALOG

You Searched: ALL: IEEE International Conference on Acoustics, Speech, and Signal Processing

Titles List: **1-8** of 8

Relevance

Select All

1
Periodical or Newspaper
ICASSP (Conference),
Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing.
[Piscataway, N.J.] : [IEEE Service Center], [1998-]
Electronic Resource
Onsite Access Only

2
Periodical or Newspaper
ICASSP (Conference)
Conference proceedings / the ... International Conference on Acoustics, Speech, and Signal Processing ; sponsored by the Signal Processing Society of the Institute of Electrical and Electronics Engineers.
Piscataway, NJ : Institute of Electrical and Electronics Engineers, c1995-c1996.
TK7882.S65 I37a
Request in Jefferson or Adams Building Reading Rooms

3
Periodical or Newspaper
ICASSP (Conference)
IEEE International Conference on Acoustics, Speech, and Signal Processing : [proceedings] / sponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society.

Los Alamitos, CA : IEEE Computer Society Press, ©1997.
SELECT TITLE FOR HOLDINGS INFORMATION

☐ 4
Periodical or Newspaper
ICASSP (Conference)
Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing / sponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society.
Piscataway, NJ : IEEE Service Center, ©1998-
TK7882.S65 I37a
Request in Jefferson or Adams Building Reading Rooms

☐ 5
Periodical or Newspaper
ICASSP (Conference)
ICASSP : [proceedings] / IEEE International Conference on Acoustics, Speech, and Signal Processing.
New York, NY : Institute of Electrical and Electronics Engineers ; Piscataway, NJ : IEEE Service Center, ©1992-©1994.
SELECT TITLE FOR HOLDINGS INFORMATION

☐ 6
Periodical or Newspaper
ICASSP (Conference),
Proceedings / ICASSP.
New York, NY : Institute of Electrical and Electronics Engineers ; Piscataway, NJ : Available from IEEE Service Center, ©1980-
SELECT TITLE FOR HOLDINGS INFORMATION

☐ 7
Periodical or Newspaper
ICASSP (Conference)
Record / IEEE International Conference on Acoustics, Speech & Signal Processing.
New York, NY : Institute of Electrical and Electronics Engineers ; Piscataway, NJ : Available from IEEE Service Center, ©1977-©1979.
TK7882.S65 I37a
Request in Jefferson or Adams Building Reading Rooms

☐ 8

Periodical or Newspaper

ICASSP (Conference)

Conference record / IEEE International Conference on Acoustics, Speech, and Signal Processing.

New York, N.Y. : Institute of Electrical and Electronics Engineers, ©1976.

TK7882.S65 I37a

Request in Jefferson or Adams Building Reading Rooms

☐ Select All

APPENDIX E

PageVault

Document title:	IEEE Xplore - Conference Table of Contents
Capture URL:	https://ieeexplore.ieee.org/xpl/conhome/3469/proceeding?isnumber=10219&sortType=vol-only-seq&rowsPerPage=100&pageNumber=1
Page loaded at (UTC):	Mon, 11 Sep 2023 16:44:08 GMT
Capture timestamp (UTC):	Mon, 11 Sep 2023 16:48:16 GMT
Capture tool:	10.29.1
Collection server IP:	34.230.137.168
Browser engine:	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/108.0.5359.215 Safari/537.36
Operating system:	Windows_NT (Node 16.17.1)
PDF length:	19
Capture ID:	kiQUfj5tyHbsG9xiymXydR
User:	finnegan-admin

PDF REFERENCE #: tDhWvwYxiYXgUurkLWQo39

International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

1995 International Conference on Acoustics, Speech, and Signal Processing
9-12 May 1995
DOI: 10.1109/ICASSP.1995

Volume : 5

Cramer-Rao bounds for matched field tomography and ocean acoustic tomography
A.B. Baggeroer; H. Schmidt
Publication Year: 1995 , Page(s): 2763 - 2766 vol.5
Cited by: Papers (13) | Patents (2)
Abstract

Signals and signal processing for acoustic monitoring of ocean processes
T.G. Birdsall
Publication Year: 1995 , Page(s): 2767 - 2770 vol.5
Cited by: Papers (3)
Abstract

A Cramer-Rao bound on acoustic measurement of ocean climate change in the presence of mesoscale sound-speed variability
J.L. Krolik; S. Narasimhan
Publication Year: 1995 , Page(s): 2771 - 2774 vol.5
Cited by: Papers (1)

Need Full-Text access to IEEE Xplore for your organization?
CONTACT IEEE TO SUBSCRIBE ›



IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS
9–13 June 2024
Denver, CO, USA

ocean climate change in the presence of mesoscale sound-speed variability

J.L. Krolik; S. Narasimhan

Publication Year: 1995 , Page(s): 2771 - 2774 vol.5

Cited by: Papers (1)

Abstract

Environmental limits to source localization

J.M.F. Moura; M.J. Rendas; G. Bienvenu

Publication Year: 1995 , Page(s): 2775 - 2778 vol.5

Cited by: Papers (1)

Abstract

Ocean acoustic tomography experiment using a vertical array

C. Noel; C. Viala; F. Evennou; Y. Stephan; F.R. Martin-Lauzer

Publication Year: 1995 , Page: 2779 vol.5

Abstract

Multiresolution statistical analysis and assimilation of large ocean data sets

P.W. Fieguth; W.C. Karl; A.S. Willsky

Publication Year: 1995 , Page(s): 2781 - 2784 vol.5

Abstract

Kalman filtering of large-scale geophysical flows by approximations based on Markov random field and wavelet

T.M. Chin; A.J. Mariano

Publication Year: 1995 , Page(s): 2785 - 2788 vol.5

Cited by: Papers (1)

Abstract

Assimilation of satellite data in beta-plane ocean circulation models

A. Asif; J.M.F. Moura

Publication Year: 1995 , Page(s): 2789 - 2792 vol.5

Cited by: Papers (2)

Abstract

High frequency acoustic imaging in the ocean

J.S. Jaffe; G. Chandran; E. Reuss

Publication Year: 1995 , Page(s): 2793 - 2796 vol.5

Abstract

Improved analysis of remote sensed data using

IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS
9–13 June 2024
Denver, CO, USA
— CALL FOR PAPERS —
Submit A Paper
IEEE ICC IEEE ComSoc

J.S. Jaffe; G. Chandran; E. Reuss
Publication Year: 1995 , Page(s): 2793 - 2796 vol.5

Abstract

Improved analysis of remote sensed data using combined physical and engineering approaches
J.J. Simpson
Publication Year: 1995 , Page(s): 2797 - 2800 vol.5

Abstract

Interferometric synthetic aperture radar for terrain mapping
J.C. Curlander
Publication Year: 1995 , Page(s): 2801 - 2804 vol.5
Cited by: Papers (4) | Patents (2)

Abstract

Robust material identification in hyperspectral data via multiresolution wavelet techniques
J.D. Gorman; N.S. Subotic; B.J. Thelen
Publication Year: 1995 , Page(s): 2805 - 2808 vol.5
Cited by: Papers (2)

Abstract

Space-time adaptive processing for airborne radar
J. Ward
Publication Year: 1995 , Page(s): 2809 - 2812 vol.5
Cited by: Papers (61) | Patents (4)

Abstract

Ground penetrating radar tomography
J.E. Molyneux
Publication Year: 1995 , Page(s): 2813 - 2816 vol.5
Cited by: Papers (2)

Abstract

Seismic methods for imaging physical properties of the Earth
G. Mavko; N. Lucet; T. Mukerji
Publication Year: 1995 , Page(s): 2817 - 2820 vol.5

Abstract

Signal processing challenges in climate data; global temperature and CO/sub 2/
D.J. Thomson

Abstract

Signal processing challenges in climate data; global temperature and CO/sub 2/

D.J. Thomson

Publication Year: 1995 , Page(s): 2821 - 2824 vol.5

Abstract

An approach for random media parameter estimation using seismic reflection data

M.G. Imhof; M.N. Toksoz

Publication Year: 1995 , Page(s): 2825 - 2828 vol.5

Abstract

Managing complexity in heterogeneous system specification, simulation, and synthesis

A. Kalavade; J.L. Pino; E.A. Lee

Publication Year: 1995 , Page(s): 2833 - 2836 vol.5

Cited by: Papers (3)

Abstract

Design guidance in the power dimension

J. Rabaey; L. Guerra; R. Mehra

Publication Year: 1995 , Page(s): 2837 - 2840 vol.5

Cited by: Papers (19)

Abstract

MAT2DSP-A tool for evaluating implementation complexity of signal processing algorithms

B. Friedlander

Publication Year: 1995 , Page(s): 2841 - 2844 vol.5

Cited by: Papers (1)

Abstract

System and algorithm implementation techniques on the TMS320 family

P. Papamichalis; J. Reimer; J. Rowlands

Publication Year: 1995 , Page(s): 2845 - 2848 vol.5

Abstract

A survey of architectures for the discrete and continuous wavelet transforms

Chaitali Chakrabarti; Mohan Vishwanath; R.M. Owens

Publication Year: 1995 , Page(s): 2849 - 2852 vol.5

Cited by: Papers (9)

Abstract

A survey of architectures for the discrete and continuous wavelet transforms

Chaitali Chakrabarti; Mohan Vishwanath; R.M. Owens

Publication Year: 1995 , Page(s): 2849 - 2852 vol.5

Cited by: Papers (9)

Abstract

Optimal VLSI architecture for vector quantization

Yu Hen Hu

Publication Year: 1995 , Page(s): 2853 - 2856 vol.5

Cited by: Papers (5)

Abstract

Approximate signal processing using incremental refinement and deadline-based algorithms

S.H. Nawab; J.M. Winograd

Publication Year: 1995 , Page(s): 2857 - 2860 vol.5

Cited by: Papers (11)

Abstract

VLSI implementation of high throughput DSP using finite ring arithmetic

G.A. Jullien

Publication Year: 1995 , Page(s): 2861 - 2864 vol.5

Abstract

Efficient approaches to testing VHDL DSP models

J.R. Armstrong

Publication Year: 1995 , Page(s): 2865 - 2867 vol.5

Cited by: Patents (2)

Abstract

DSP design education at Georgia Tech

V.K. Madisetti; J.H. McClellan; T.P. Barnwell

Publication Year: 1995 , Page(s): 2869 - 2872 vol.5

Cited by: Papers (3)

Abstract

Integrating analysis, simulation, and implementation tools in electronic courseware for teaching signal processing

R.H. Bamberger; B.L. Evans; E.A. Lee; J.H. McClellan; M.A. Yoder

Publication Year: 1995 , Page(s): 2873 - 2876 vol.5

Cited by: Papers (1)

Abstract

R.H. Bamberger; B.L. Evans; E.A. Lee; J.H. McClellan; M.A. Yoder
Publication Year: 1995 , Page(s): 2873 - 2876 vol.5
Cited by: Papers (1)

Abstract

Improving DSP design expertise by stepwise educational intensification
P. Pype; H. De Man
Publication Year: 1995 , Page(s): 2877 - 2880 vol.5

Abstract

DSP-based mobile and satellite receivers, from algorithm to implementation: a design course at Aachen University of Technology
O.C. Mauss; M. Pankert; F. Classen; H. Meyr
Publication Year: 1995 , Page(s): 2881 - 2884 vol.5

Abstract

A distance learning laboratory design experiment in undergraduate digital signal processing
D.M. Etter; G.C. Orsak; D.H. Johnson
Publication Year: 1995 , Page(s): 2885 - 2887 vol.5
Cited by: Papers (6)

Abstract

An overview of signal processing in astronomy
S.R. Kulkarni
Publication Year: 1995 , Page: 2889 vol.5
Cited by: Papers (3)

Abstract

Cosmology with gravitational lenses: time delay and microlensing
R. Schild; D.J. Thomson
Publication Year: 1995 , Page(s): 2891 - 2894 vol.5

Abstract

Time series data in geophysics/space physics
L.J. Lanzerotti; D.J. Thomson
Publication Year: 1995 , Page(s): 2895 - 2898 vol.5

Abstract

High resolution astronomical imaging by post-detection image processing
G.J.M. Aitken
Publication Year: 1995 , Page(s): 2899 - 2902 vol.5

Abstract

High resolution astronomical imaging by post-detection image processing

G.J.M. Aitken

Publication Year: 1995 , Page(s): 2899 - 2902 vol.5

Abstract

Some recent results in tomography

Y. Bresler

Publication Year: 1995 , Page(s): 2903 - 2906 vol.5

Abstract

Tractable models and efficient algorithms for Bayesian tomography

C.A. Bouman; K. Sauer; S.S. Saquib

Publication Year: 1995 , Page(s): 2907 - 2910 vol.5

Cited by: Papers (3)

Abstract

Neuromagnetic source reconstruction

P.S. Lewis; J.C. Mosher; R.M. Leahy

Publication Year: 1995 , Page(s): 2911 - 2914 vol.5

Cited by: Papers (2)

Abstract

Geometric signal processing and applications to brain mapping

O. Tretiak; A. Goldszal

Publication Year: 1995 , Page(s): 2915 - 2918 vol.5

Cited by: Patents (5)

Abstract

Efficient geometric transformations and 3-D image registration

P. Thevenaz; M. Unser

Publication Year: 1995 , Page(s): 2919 - 2922 vol.5

Cited by: Papers (3)

Abstract

Joint best bases for fast encoding in magnetic resonance imaging

D. Healy; J. Weaver

Publication Year: 1995 , Page(s): 2923 - 2926 vol.5

Cited by: Papers (1)

Abstract

resonance imaging

D. Healy; J. Weaver

Publication Year: 1995 , Page(s): 2923 - 2926 vol.5

Cited by: Papers (1)

Abstract

Adaptive matched filtering of steady-state visual evoked potentials

C.E. Davila; R. Srebro; H. Azmoodeh; I. Ghaleb

Publication Year: 1995 , Page(s): 2927 - 2930 vol.5

Cited by: Papers (2)

Abstract

Crosstalk resistant adaptive noise cancellation applied to somatosensory evoked potential enhancement

V. Parsa; P. Parker; R. Scott

Publication Year: 1995 , Page(s): 2931 - 2934 vol.5

Cited by: Papers (2)

Abstract

Hidden process modeling

B.R. Kosanovic; L.F. Chaparro; R.J. Sclabassi

Publication Year: 1995 , Page(s): 2935 - 2938 vol.5

Cited by: Papers (8)

Abstract

Narrowband delay estimation for thalamocortical epileptic seizure pathways

D.L. Sherman; Yien Che Tsai; L.A. Rossell; M.A. Mirski; N.V. Thakor

Publication Year: 1995 , Page(s): 2939 - 2942 vol.5

Cited by: Papers (6)

Abstract

Matrix kernels for MEG and EEG source localization and imaging

J.C. Mosher; R.M. Leahy; P.S. Lewis

Publication Year: 1995 , Page(s): 2943 - 2946 vol.5

Cited by: Papers (16)

Abstract

Quantification of injury-related EEG signal changes using Itakura distance measure

Xuan Kong; V. Goel; N. Thakor

Publication Year: 1995 , Page(s): 2947 - 2950 vol.5

Cited by: Papers (6)

Abstract

using Itakura distance measure

Xuan Kong; V. Goel; N. Thakor

Publication Year: 1995 , Page(s): 2947 - 2950 vol.5

Cited by: Papers (6)

Abstract

A comparative study of multiple tests based techniques for optimal sensor location for knock detection

A.M. Zoubir

Publication Year: 1995 , Page(s): 2951 - 2954 vol.5

Abstract

Lane detection for automotive sensors

S. Lakshmanan; K.C. Kluge

Publication Year: 1995 , Page(s): 2955 - 2958 vol.5

Cited by: Papers (14)

Abstract

Multi-scale signal feature processing for automatic, objective vehicle noise and vibration quality analysis

D.A. Whitney; B. Broder

Publication Year: 1995 , Page(s): 2959 - 2962 vol.5

Cited by: Patents (3)

Abstract

Advanced signal processing for misfire detection in automotive engines

W.B. Ribbens; S. Bieser

Publication Year: 1995 , Page(s): 2963 - 2966 vol.5

Cited by: Papers (3) | Patents (8)

Abstract

Active sensory tuning of windnoise using a genetic algorithm

A. Sterian; P. Runkle; G.H. Wakefield

Publication Year: 1995 , Page(s): 2967 - 2970 vol.5

Cited by: Papers (1)

Abstract

Analysis and processing of shaft angular velocity signals in rotating machinery for diagnostic applications

Y.W. Kim; G. Rizzoni; B. Samimy; Y.Y. Wang

Publication Year: 1995 , Page(s): 2971 - 2974 vol.5

Cited by: Papers (12) | Patents (8)

Abstract

signals in rotating machinery for diagnostic applications

Y.W. Kim; G. Rizzoni; B. Samimy; Y.Y. Wang

Publication Year: 1995 , Page(s): 2971 - 2974 vol.5

Cited by: Papers (12) | Patents (8)

Abstract

Signal processing for automotive applications

T. Gioutsos

Publication Year: 1995 , Page(s): 2975 - 2978 vol.5

Cited by: Patents (2)

Abstract

A case study in IVHS implementation: an image processing application for I-15 HOV lanes

H. Abut; A. Bilgin; R.M. Bernardi; L.A. Wherry

Publication Year: 1995 , Page(s): 2979 - 2982 vol.5

Abstract

Adaptive filtering of stable processes for active attenuation of impulsive noise

R. Leahy; Zhenyu Zhou; Yung-Chih Hsu

Publication Year: 1995 , Page(s): 2983 - 2986 vol.5

Cited by: Papers (54)

Abstract

High quality and low complexity pitch modification of acoustic signals

Gang-Janp Lin; Sau-Gee Chen; T. Wu

Publication Year: 1995 , Page(s): 2987 - 2990 vol.5

Cited by: Papers (1) | Patents (1)

Abstract

A digital hearing aid that compensates loudness for sensorineural hearing impairments

J.C. Tejero; S. Bernal; J.A. Hidaldo; J. Fernandez; R. Urquiza; A. Gago

Publication Year: 1995 , Page(s): 2991 - 2994 vol.5

Cited by: Papers (4)

Abstract

Frequency-independent beamformer with low response error

T. Chou

Publication Year: 1995 , Page(s): 2995 - 2998 vol.5

Cited by: Papers (59) | Patents (18)

Abstract

error

T. Chou

Publication Year: 1995 , Page(s): 2995 - 2998 vol.5

Cited by: Papers (59) | Patents (18)

Abstract

Desktop mic array for teleconferencing

P.L. Chu

Publication Year: 1995 , Page(s): 2999 - 3002 vol.5

Cited by: Papers (2) | Patents (5)

Abstract

Wavelet-based noise reduction

N.A. Whitmal; J.C. Rutledge; J. Cohen

Publication Year: 1995 , Page(s): 3003 - 3006 vol.5

Cited by: Papers (6) | Patents (1)

Abstract

Multi-channel deconvolution using Pade approximation

Hong Wang

Publication Year: 1995 , Page(s): 3007 - 3010 vol.5

Cited by: Papers (3)

Abstract

High resolution analysis of bird sounds

C. Rogers

Publication Year: 1995 , Page(s): 3011 - 3014 vol.5

Cited by: Papers (5) | Patents (2)

Abstract

A user-friendly system for microphone array research

S.E. Kirtman; H.F. Silverman

Publication Year: 1995 , Page(s): 3015 - 3018 vol.5

Cited by: Papers (6)

Abstract

A closed-form method for finding source locations from microphone-array time-decay estimates

M.S. Brandstein; J.E. Adcock; H.F. Silverman

Publication Year: 1995 , Page(s): 3019 - 3022 vol.5

Cited by: Papers (17) | Patents (2)

Abstract

The fast affine projection algorithm

S.L. Gay; S. Tavathia

Publication Year: 1995 , Page(s): 3023 - 3026 vol.5

Abstract

The fast affine projection algorithm

S.L. Gay; S. Tavathia

Publication Year: 1995 , Page(s): 3023 - 3026 vol.5

Cited by: Papers (204) | Patents (11)

Abstract

A stereophonic echo canceler using single adaptive filter

S. Minami

Publication Year: 1995 , Page(s): 3027 - 3030 vol.5

Cited by: Papers (5)

Abstract

Simulation of time variant room impulse responses

C. Antweiler; H.-G. Symanzik

Publication Year: 1995 , Page(s): 3031 - 3034 vol.5

Cited by: Papers (6) | Patents (2)

Abstract

Acoustic echo cancellation using NLMS-neural network structures

A.N. Birkett; R.A. Goubran

Publication Year: 1995 , Page(s): 3035 - 3038 vol.5

Cited by: Papers (29) | Patents (1)

Abstract

Finite-precision design and implementation of all-pass polyphase networks for echo cancellation in sub-bands

O. Tanrikulu; B. Baykal; A.G. Constantinides; J.A. Chambers; P.A. Naylor

Publication Year: 1995 , Page(s): 3039 - 3042 vol.5

Cited by: Papers (3) | Patents (2)

Abstract

Coupled adaptive filters for acoustic echo control and noise reduction

R. Martin; J. Altenhoner

Publication Year: 1995 , Page(s): 3043 - 3046 vol.5

Cited by: Papers (21) | Patents (14)

Abstract

Optimization of an acoustic echo canceller combined with adaptive gain control

Cited by: Papers (21) | Patents (14)

Abstract

Optimization of an acoustic echo canceller combined with adaptive gain control

P. Heitkamper

Publication Year: 1995 , Page(s): 3047 - 3050 vol.5

Cited by: Papers (6) | Patents (2)

Abstract

A new adaptive intersubband tap-assignment algorithm for subband adaptive filters

A. Sugiyama; F. Landais

Publication Year: 1995 , Page(s): 3051 - 3054 vol.5

Cited by: Papers (5) | Patents (3)

Abstract

Cepstral prefiltering for time delay estimation in reverberant environments

A. Stephenne; B. Champagne

Publication Year: 1995 , Page(s): 3055 - 3058 vol.5

Cited by: Papers (18) | Patents (1)

Abstract

Stereo projection echo canceller with true echo path estimation

S. Shimauchi; S. Makino

Publication Year: 1995 , Page(s): 3059 - 3062 vol.5

Cited by: Papers (50) | Patents (6)

Abstract

Optimizing high quality audio coding: advantages of full system observability

A.J.S. Ferreira

Publication Year: 1995 , Page(s): 3063 - 3066 vol.5

Cited by: Papers (1)

Abstract

High quality audio coding using multipulse LPC and wavelet decomposition

S. Boland; M. Deriche

Publication Year: 1995 , Page(s): 3067 - 3069 vol.5

Cited by: Papers (4) | Patents (3)

Abstract

Audio coding with signal adaptive filterbanks

Publication Year: 1995 , Page(s): 3067 - 3069 vol.5
Cited by: Papers (4) | Patents (3)

Abstract

Audio coding with signal adaptive filterbanks
J. Princen; J.D. Johnston
Publication Year: 1995 , Page(s): 3071 - 3074 vol.5
Cited by: Papers (19) | Patents (14)

Abstract

Computationally efficient wavelet packet coding of wide-band stereo audio signals
M. Black; M. Zeytinoglu
Publication Year: 1995 , Page(s): 3075 - 3078 vol.5
Cited by: Papers (21) | Patents (5)

Abstract

Incorporation of biorthogonality into lapped transforms for audio compression
Shiufun Cheung; J.S. Lim
Publication Year: 1995 , Page(s): 3079 - 3082 vol.5
Cited by: Papers (3) | Patents (6)

Abstract

Comparison of the wavelet decomposition and the Fourier transform in TCX encoding of wideband speech and audio
J.-M. LeRoux; R. Lefebvre; J.-P. Adoul
Publication Year: 1995 , Page(s): 3083 - 3086 vol.5
Cited by: Papers (3)

Abstract

On the performance of wavelets for low bit rate coding of audio signals
P.E. Kudumakis; M.B. Sandler
Publication Year: 1995 , Page(s): 3087 - 3090 vol.5
Cited by: Papers (15)

Abstract

Lossless compression of high fidelity audio and adaptive linear prediction
A.L. Adams; S.W. McLaughlin
Publication Year: 1995 , Page(s): 3091 - 3094 vol.5

Abstract

High-quality audio-coding at less than 64 kbit/s by using

A.L. Adams; S.W. McLaughlin
Publication Year: 1995 , Page(s): 3091 - 3094 vol.5

Abstract

High-quality audio-coding at less than 64 kbit/s by using transform-domain weighted interleave vector quantization (TwinVQ)
N. Iwakami; T. Moriya; S. Miki
Publication Year: 1995 , Page(s): 3095 - 3098 vol.5
Cited by: Papers (48) | Patents (16)

Abstract

Adaptive filtering algorithms for stereophonic acoustic echo cancellation
J. Benesty; F. Amand; A. Gilloire; Y. Grenier
Publication Year: 1995 , Page(s): 3099 - 3102 vol.5
Cited by: Papers (64) | Patents (9)

Abstract

The partitioned eigenvector method for towed array shape estimation
J.J. Smith; Y.H. Leung; A. Cantoni
Publication Year: 1995 , Page(s): 3103 - 3106 vol.5
Cited by: Papers (2)

Abstract

Monochromatic single-mode excitation in shallow water using feedback control
J.R. Buck; J.C. Preisig; M. Johnson; J. Catipovic
Publication Year: 1995 , Page(s): 3107 - 3110 vol.5

Abstract

Source localization in shallow water using polynomial rooting
J. Tabrikian; H. Messer
Publication Year: 1995 , Page(s): 3111 - 3114 vol.5

Abstract

Source localization in an acoustic waveguide with uncertain sound speed
B.F. Harrison; R.J. Vaccaro; D.W. Tufts
Publication Year: 1995 , Page(s): 3115 - 3118 vol.5
Cited by: Papers (1)

Abstract

Maximum a posteriori probability estimation of seafloor

D.F. Harrison; R.J. Vaccaro; D.W. Tufts

Publication Year: 1995 , Page(s): 3115 - 3118 vol.5

Cited by: Papers (1)

Abstract

Maximum a posteriori probability estimation of seafloor microroughness parameters from backscatter spatial coherence

V. Premus; D. Alexandrou

Publication Year: 1995 , Page(s): 3119 - 3122 vol.5

Abstract

F-test in matched field processing: identifying multimode propagation

C.F. Mecklenbrauker; D. Maiwald; J.F. Bohme

Publication Year: 1995 , Page(s): 3123 - 3126 vol.5

Cited by: Papers (2)

Abstract

GLRT detection of signals in reverberation

R.N. Carpenter; S. Kay

Publication Year: 1995 , Page(s): 3127 - 3130 vol.5

Cited by: Papers (2)

Abstract

Robust detection of signal classes

O.P. Kenny; L.B. White

Publication Year: 1995 , Page(s): 3131 - 3134 vol.5

Cited by: Papers (1)

Abstract

Group delay inversion for ocean properties

M.A. Dzieciuch

Publication Year: 1995 , Page(s): 3135 - 3137 vol.5

Abstract

Passive target motion analysis using multipath differential time-delay and differential Doppler shifts

P. Blanc-Benon; G. Bienvenu

Publication Year: 1995 , Page(s): 3139 - 3142 vol.5

Cited by: Papers (5) | Patents (1)

Abstract

Multipath time-delay estimation for long data records

T.G. Manickam; R.J. Vaccaro

Publication Year: 1995 , Page(s): 3143 - 3146 vol.5

Abstract

Multipath time-delay estimation for long data records

T.G. Manickam; R.J. Vaccaro

Publication Year: 1995 , Page(s): 3143 - 3146 vol.5

Cited by: Papers (1)

Abstract

Performance analysis of integrated polyspectrum based time delay estimators

Yisong Ye; J.K. Tugnait

Publication Year: 1995 , Page(s): 3147 - 3150 vol.5

Cited by: Papers (1)

Abstract

An improvement to the explicit time delay estimator

H.C. So; P.C. Ching; Y.T. Chan

Publication Year: 1995 , Page(s): 3151 - 3154 vol.5

Cited by: Papers (12)

Abstract

Digital estimation of frequencies of sinusoids from wide-band under-sampled data

D.W. Tufts; Hongya Ge

Publication Year: 1995 , Page(s): 3155 - 3158 vol.5

Cited by: Papers (7)

Abstract

An efficient self-recovering adaptive algorithm for BPSK signals transmitted through underwater acoustic channels

E. Petit; G. Jourdain

Publication Year: 1995 , Page(s): 3159 - 3162 vol.5

Cited by: Papers (1)

Abstract

IEEE Personal Account

CHANGE USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS

VIEW PURCHASED DOCUMENTS

Profile Information

COMMUNICATIONS PREFERENCES

PROFESSION AND EDUCATION

TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800 678 4333

WORLDWIDE: +1 732 981 0060

CONTACT & SUPPORT

Follow

Abstract

Performance analysis of integrated polyspectrum based time delay estimators

Yisong Ye; J.K. Tugnait

Publication Year: 1995 , Page(s): 3147 - 3150 vol.5

Cited by: Papers (1)

Abstract

An improvement to the explicit time delay estimator

H.C. So; P.C. Ching; Y.T. Chan

Publication Year: 1995 , Page(s): 3151 - 3154 vol.5

Cited by: Papers (12)

Abstract

Digital estimation of frequencies of sinusoids from wide-band under-sampled data

D.W. Tufts; Hongya Ge

Publication Year: 1995 , Page(s): 3155 - 3158 vol.5

Cited by: Papers (7)

Abstract

An efficient self-recovering adaptive algorithm for BPSK signals transmitted through underwater acoustic channels

E. Petit; G. Jourdain

Publication Year: 1995 , Page(s): 3159 - 3162 vol.5

Cited by: Papers (1)

Abstract

IEEE Personal Account

CHANGE USERNAME/PASSWORD

Purchase Details

PAYMENT OPTIONS

VIEW PURCHASED DOCUMENTS

Profile Information

COMMUNICATIONS PREFERENCES

PROFESSION AND EDUCATION

TECHNICAL INTERESTS

Need Help?

US & CANADA: +1 800 678 4333

WORLDWIDE: +1 732 981 0060

CONTACT & SUPPORT

Follow

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting | Sitemap | IEEE Privacy Policy

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

Copyright 2023 IEEE - All rights reserved.

APPENDIX F

1997 IEEE ICASSP VOL. 1



89059970699
b89059970699a

KURT F. WENDT LIBRARY
COLLEGE OF ENGINEERING

UW-MADISON, WI 53706

1997 IEEE

International Conference on Acoustics, Speech, and Signal Processing

Volume I

KURT F. WENDT LIBRARY
COLLEGE OF ENGINEERING

UW-MADISON, WI 53706

1997 IEEE International Conference on Acoustics, Speech, and Signal Processing

Volume I of V

Plenary, Expert Summaries, Special, Audio, Underwater Acoustics, VLSI

April 21-24, 1997
Munich, Germany

Sponsored by

The Institute of Electrical and Electronics Engineers
Signal Processing Society

Co-Sponsored by

DPG, GI, ITG, and TUM



IEEE Computer Society Press
Los Alamitos, California

Washington • Brussels • Tokyo

Kurt F. Wendt Library
University of Wisconsin - Madison
215 N. Randall Avenue
Madison, WI 53706-1688

IEEE Computer Society Press
10662 Los Vaqueros Circle
P.O.Box 3014
Los Alamitos, CA 90720-1264

Copyright ©1997 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society Press, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Press Order Number PR07919
ISBN 0-8186-7919-0
ISBN 0-8186-7920-4 (case)
ISBN 0-8186-7921-2 (microfiche)
IEEE Order Plan Catalog Number 97CB36052
Library of Congress Number 96-80328

Additional copies may be ordered from:

IEEE Computer Society Press
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: +1-714-821-8380
Fax: +1-714-821-4641
Email: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: +1-908-981-1393
Fax: +1-908-981-9667
misc.custserv@computer.org

IEEE Computer Society
13, Avenue de l'Aquilon
B-1200 Brussels
BELGIUM
Tel: +32-2-770-2198
Fax: +32-2-770-8505
euro.ofc@computer.org

IEEE Computer Society
Ooshima Building
2-19-1 Minami-Aoyama
Minato-ku, Tokyo 107
JAPAN
Tel: +81-3-3408-3118
Fax: +81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Bob Werner
Printed in the United States of America by Sony Electronic Publishing Services

The Institute of Electrical and Electronics Engineers, Inc.

Wanot
TA
365
I102
V.1997
1

AWC 9763

Table of Contents

Volume I
Plenary, Expert Summaries, Special, Audio, Underwater Acoustics, VLSI

Conference Committee xvi
IEEE Signal Processing Society xvii
ICASSP 1998 in Seattle xviii
Call for Papers xix
ICASSP-98 Paper Cover Sheet xx
IEEE Copyright Form xxi

Plenary Lectures and Expert Summaries

Space-Time Processing for Wireless Communications 1
A. Paulraj

Variability of Performance in Video Coding 5
D. Pearson

Expert Summaries 9
R. Chellapa, R. de Mori, G. Giannakis, H. Musmann, H. Ney, M. Smith, J. Treichler and M. Zoltowski

Signal Processing for Education

Expanding Team Experiences in DSP Education 11
D. Etter and G. Orsak

Interactive Classroom for DSP/Communication Courses 15
H. Abut and Y. Öztürk

Experiences in Teaching DSP First in the ECE Curriculum 19
J. McClellan

Analog Signal Processing: A Replacement for the Sophomore-Level Circuit Analysis Course 23
D. Munson

Re-Engineering the Electrical Engineering Curriculum 27
S. Mitra

New Methods for Design of SP-Algorithms

Structural Subband Decomposition: A New Concept in Digital Signal Processing 31
S. Mitra

A New Algorithm for the Generalized Eigenvalue Problem 35
K. Hüper and U. Helmke

A Lattice Structure for Perfect Reconstruction Linear Time Varying Filter Banks with All Pass Analysis Banks 39
S. Dasgupta, C. Schwarz and M. Fu

Algorithm Design for Structured Systems: Application to Pole Placement 43
S. Paul and J. Nossek

Actions of Noncompact Groups and Algorithm Design: A Case Study 47
K. Diepold and R. Pauli

Discretization Issues for the Design of Optimal Blind Algorithms 51
R. Kennedy, D. Borah and Z. Ding

Continuous-Time Envelope-Constrained Filter Design via Laguerre Filters and H(Inf) Optimization Methods 55
Z. Zang, A. Cantoni and K. Teo

Local Adaptive Algorithms for Information Maximization in Neural Networks, and Application to Source Separation 59
J. Dehaene and N. Twum-Danso

Quick Aggregation of Markov Chain Functionals via Stochastic Complementation 63
K. Dogançay and V. Krishnamurthy

A Rank Preserving Flow Algorithm for Quadratic Optimization Problems Subject to Quadratic Equality Constraints 67
J. Moore and D. Jiang

Speech-to-Speech Translation

Verbmobil: The Combination of Deep and Shallow Processing for Spontaneous Speech Translation 71
T. Bub, W. Wahlster and A. Waibel

Prosodic Processing and its Use in Verbmobil 75
H. Niemann, E. Nöth, A. Kießling, R. Kompe and A. Batliner

The Language Components in Verbmobil 79
H. Block

The Karlsruhe-Verbmobil Speech Recognition Engine 83
M. Finke, P. Geutner, H. Hild, T. Kemp, K. Ries and M. Westphal

An Experiment on Korean-to-English and Korean-to-Japanese Spoken Language Translation 87
J. Yang and J. Park

Multilingual Person to Person Communication at IRST 91
B. Angelini, M. Cettolo, A. Corazza, D. Falavigna and G. Lazzari

Fast Word-Graph Generation for Spontaneous Conversational Speech Translation 95
T. Shimizu, H. Singer and Y. Sagisaka

JANUS-III: Speech-to-Speech Translation in Multiple Languages 99
A. Lavie, A. Waibel, L. Levin, M. Finke, D. Gates, M. Gavalda, T. Zeppenfeld and P. Zhan

State-Transition Cost Functions and an Application to Language Translation 103
H. Alshawi and A. Buchsbaum

Hybrid Language Processing in the Spoken Language Translator 107
M. Rayner and D. Carter

Finite-State Speech-to-Speech Translation 111
E. Vidal

An Experimental Bidirectional Japanese/English Interpreting Video Phone System Using Internet 115
S. Hiraoka, M. Hoshimi, K. Matsui and J. Junqua

Advanced Neural Applications

From Neural Networks to Neural Strategies 119
C. Goerick, B. Sendhoff and W. von Seelen

Neural and Traditional Techniques in Diagnostic ECG Classification 123
R. Silipo and G. Bortolan

Unsupervised Learning for Blind Source Separation: An Information-Theoretic Approach 127
D. Obradovic and G. Deco

Applications of Neural Blind Separation to Signal and Image Processing 131
J. Karhunen, A. Hyvärinen, R. Vigário, J. Hurri and E. Oja

Communications and Neural Networks: Theory and Practice 135
M. Plumbley

Robust Vector Quantization by Competitive Learning 139
J. Buhmann and T. Hofmann

Recognizing Faces from a New Viewpoint 143
T. Vetter

Hybrid Optimization of Feedforward Neural Networks for Handwritten Character Recognition 147
W. Utschick and J. Nossek

Reading Checks with Multilayer Graph Transformer Networks 151
Y. LeCun, L. Bottou and Y. Bengio

Neural Networks for Process Control in Steel Manufacturing 155
M. Schlang, E. Broese, B. Feldkeller, O. Gramckow, M. Jansen, T. Poppe, C. Schaeffner and G. Soergel

A Neuro-Dynamic Programming Approach to Admission Control in ATM Networks: The Single Link Case 159
P. Marbach and J. Tsitsiklis

Signal Processing Technology for Multi-Media Human-Machine Interaction

Issues in Measuring the Benefits of Multimodal Interfaces 163
J. Flanagan and I. Marsic

Multimodal Interfaces for Multimedia Information Agents 167
A. Waibel, B. Suhm, M. Vo and J. Yang

Smart Rooms, Desks, and Clothes 171
A. Pentland

Human Machine Interaction by Voice and Gesture 175
N. Jayant

Audio-Visual Interaction in Multimedia Communication 179
T. Chen and R. Rao

LIP Motion Modeling and Speech Driven Estimation 183
F. Lavagetto, S. Lepsoy, C. Braccini and S. Curinga

Voice Source Localization for Automatic Camera Pointing System in Videoconferencing 187
H. Wang and P. Chu

Video Interface for Spatiotemporal Interactions Based on Multi-Dimensional Video Computing 191
A. Akutsu, Y. Tonomura and H. Hamada

Indexing and Search of Multimodal Information 195
A. Hauptmann and H. Wactlar

Acoustic Indexing for Multimedia Retrieval and Browsing 199
S. Young, J. Foote, G. Jones, K. Sparck-Jones and M. Brown

Broadcast News Transcription 203
F. Kubala, H. Jin, L. Nguyen, R. Schwartz and S. Matsoukas

Image/Speech Processing that Adopts an Artistic Approach Toward Integration of Art and Technology 207
R. Nakatsu

Microphone Array Signal Processing

Noise Cancelling for Microphone Arrays 211
J. Meyer and C. Sydow

A Microphone Array System for Speech Recognition 215
K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura and J. Kojima

Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays 219
W. Kellermann

A Steerable and Variable First-Order Differential Microphone Array 223
G. Elko and A. Pong

Microphone Array Based Speech Recognition with Different Talker-Array Positions 227
M. Omologo, M. Matassoni, P. Svaizer and D. Giuliani

Acoustic Source Location in a Three-Dimensional Space Using Crosspower Spectrum Phase 231
P. Svaizer, M. Matassoni and M. Omologo

Superdirective Microphone Array for a Set-Top Videoconferencing System 235
P. Chu

Simultaneous Echo Cancellation and Car Noise Suppression Employing a Microphone Array 239
M. Dahl, I. Claesson and S. Nordebo

Analytical Evaluation of a Self-Calibrating Microphone Array 243
S. Nordholm and I. Claesson

Microphone Array Response to Speaker Movements 247
Y. Grenier and S. Affes

A Digital Processing System for Source Location and Sound Capture by Large Microphone Arrays 251
H. Silverman, W. Patterson, J. Flanagan and D. Rabinkin

DSP for Mobile Communication

3-D Unitary ESPRIT for Joint 2-D Angle and Carrier Estimation 255
M. Haardt and J. Nossek

Quality Enhancement of Coded and Corrupted Speeches in GSM Mobile Systems Using Residual Redundancy 259
T. Hindelang, W. Xu and C. Erben

Pilot Assisted Coherent DS-CDMA Reverse-Link Communications with Optimal Robust Channel Estimation 263
F. Ling

A New Frequency Estimator Applied to Burst Transmission 267
C. Bergogne, M. Bousquet and P. Sehier

Unified Specification of Control and Data Flow 271
T. Groetker, R. Schoenen and H. Meyr

Reconfigurable Processing: The Solution to Low-Power Programmable DSP 275
J. Rabaey

DSP Cores for Mobile Communications: Where Are We Going? 279
G. Fettweis

DSPs in Mobile Communication in the United States 283
S. Kasturia and C. Warwick

FRIDGE: An Interactive Code Generation Environment for HW/SW CoDesign 287
M. Willems, V. Bürsgens, T. Grötker and H. Meyr

Staying Ahead of the Game in Silicon for Digital Mobile Communications 291
R. Subramanian, M. Barberis and H. Dawid

Echo Cancellation

Approximation of Optimal Step Size Control for Acoustic Echo Cancellation 295
C. Antweiler, J. Grunwald and H. Quack

Subband Stereo Echo Canceller Using the Projection Algorithm with Fast Convergence to the True Echo Path 299
S. Makino, K. Strauss, S. Shimauchi, Y. Haneda and A. Nakagawa

A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation 303
J. Benesty, D. Morgan and M. Sondhi

Comparison of Three Post-Filtering Algorithms for Residual Acoustic Echo Reduction 307
V. Turbin, A. Gilloire and P Scalart

Audio Coding & Transducer

Audio Coding Using Sinusoidal Excitation Representation 311
W. Chang, D. Wang and L. Wang

Optimum Bit Allocation and Decomposition for High Quality Audio Coding 315
X. Wei, M. Shaw and M. Varley

The D_5 Lattice Quantization for a 64 Kbit/S Low-Delay Subband Audio Coder with a 15 Khz Bandwidth 319
K. Hay, S. Saoudi and L. Mainard

An Experimental Audio Codec Based on Warped Linear Prediction of Complex Valued Signals 323
A. Härmä, U. Laine and M. Karjalainen

High Quality Low Complexity Scalable Wavelet Audio Coding 327
W. Dobson, J. Yang, K. Smart and F. Guo

An Efficient Tonal Component Coding Algorithm for MPEG-2 Audio NBC 331
Y. Takamizawa, M. Iwadare and A. Sugiyama

Spectral Amplitude Warping (SAW) for Noise Spectrum Shaping in Audio Coding 335
R. Lefebvre and C. Laflamme

A Fast Noise-Scaling Algorithm for Uniform Quantization in Audio Coding Schemes 339
C. Serantes, A. Pena and N. Prelcic

Pyramid Vector Coding for High Quality Audio Compression 343
D. Cadel and G. Parladori

Subband Audio Coding with Synthesis Filters Minimizing a Perceptual Distortion 347
K. Gosse, F. Moreau de Saint-Martin, X. Durot, P. Duhamel and J. Rault

New Results in Low Bitrate Audio Coding Using a Combined Harmonic-Wavelet Representation 351
S. Boland and M. Deriche

Adaptive Inverse Control of Weakly Nonlinear Systems 355
W. Klippel

Microphone Array & Active Noise Control

Broadband Beamforming with Adaptive Postfiltering for Speech Acquisition in Noisy Environments 359
S. Fischer and K. Kammeyer

Near-Field Beamforming for Microphone Arrays 363
J. Ryan and R. Goubran

A Robust Adaptive Microphone Array with Improved Spatial Selectivity and its Evaluation in a Real Environment 367
O. Hoshuyama, A. Sugiyama and A. Hirano

Tracking Multiple Talkers Using Microphone-Array Measurements 371
D. Sturim, H. Silverman and M. Brandstein

A Robust Method for Speech Signal Time-Delay Estimation in Reverberant Rooms 375
M. Brandstein and H. Silverman

A Model-Based Approach to Active Noise Cancellation Using Loudspeaker Array 379
J. Gu and S. Yau

Reverberant Sound Field Analysis Using a Microphone Array 383
W. Täger and Y. Mahieux

Minimisation of the Maximum Error Signal in Active Control 387
A. Gonzalez, A. Albiol and S. Elliott

Subband Active Noise Control Algorithm Based on a Delayless Subband Adaptive Filter Architecture 391
J. Yun, D. Youn and Y. Park

Nonlinear Active Noise Control in a Linear Duct 395
P. Strauch and B. Mulgrew

Fast Exact Filtered-X LMS and LMS Algorithms for Multichannel Active Noise Control 399
S. Douglas

A Novel Frequency Domain Filtered-X LMS Algorithm for Active Noise Reduction 403
T. Kosakat, S. Elliott and C. Boucher

Hearing Aids and Computer Music

Practical Supergain Head Sized Arrays 407
D. Masmoudi, D. Dallet and J. Dom

A Multichannel Compression Strategy for a Digital Hearing Aid 411
T. Schneider and R. Brennan

Multi-Microphone Sub-Band Adaptive Signal Processing for Improvement of Hearing Aid Performance: Preliminary Results Using Normal Hearing Volunteers 415
P. Shields and D. Campbell

Environmental Noise Reduction Based on Speech/Non-Speech Identification for Hearing Aids 419
K. Itoh and M. Mizushima

Blind Separation of Multiple Speakers in a Multipath Environment 423
R. Lambert and A. Bell

A Single-Chip 1,200 Sinusoid Real-Time Generator for Additive Synthesis of Musical Signals 427
F. De Bernardinis, R. Roncella, R. Saletti, P. Terreni and G. Bertini

A Generalized Musical-Tone Generator with Application to Sound Compression and Synthesis 431
C. Drioli and D. Rocchesso

A Singing Voice Synthesis System Based on Sinusoidal Modeling 435
M. Macon, L. Jensen-Link, J. Oliverio, M. Clements and E. George

Time-Scale Modification of Audio Signals with Combined Harmonic and Wavelet Representations 439
K. Hamdy, A. Tewfik, S. Takagi and T. Chen

A Waveguide Model for Slapbass Synthesis 443
E. Rank and G. Kubin

Minimum Perceptual Spectral Distance FIR Filter Design 447
S. Wu and W. Putnam

A Phase Interpolation Algorithm for Sinusoidal Model Based Music Synthesis 451
X. Qian and Y. Ding

Analytical Approximations of Fractional Delays: Lagrange Interpolators and Allpass Filters 455
S. Tassart and P. Depalle

Improved Discrete-Time Modeling of Multi-Dimensional Wave Propagation Using the Interpolated Digital Waveguide Mesh 459
L. Savioja and V. Välimäki

Matched Field Processing

Generalized Likelihood Ratio Test for Selecting a Geo-Acoustic Environmental Model 463
C. Mecklenbräuker, P. Gerstoft, P. Chung and J. Böhme

Tuning Genetic Algorithms for Underwater Acoustics Using a priori Statistical Information 467
M. Rendas and G. Bienvenu

Robust Beamformer Weight Design for Broadband Matched-Field Processing 471
K. Harmanci and J. Krolik

FASTMAP: A Fast, Approximate Maximum a posteriori Probability Parameter Estimator with Application to Robust Matched-Field Processing 475
B. Harrison, R. Vaccaro and D. Tufts

Electromagnetic Matched Field Processing for Source Localization 479
D. Gingras, P. Gerstoft, N. Gerr and C. Mecklenbräuker

Detection, Classification and Localisation

Power-Law Processors for Detecting Unknown in Signals in Colored Noise 483
I. Kirsteins, S. Mehta and J. Fay

Multitarget Detection/tracking of Echoes with Known Waveform: Algorithm and Applications 487
V. Rampa and U. Spagnolini

Detection of Gaussian Bandpass Transients Under Impulsive Noise: A Wavelet Transform Approach 491
F. Garcia and I. Lourtie

Maximum Likelihood Estimator for Magneto-Acoustic Localisation 495
G. Dassot, R. Blanpain and C. Jauffret

Barankin Bound for Source Localization in Shallow Water 499
J. Tabrikian and J. Krolik

Underwater Transient Signal Processing: Marine Mammal Identification, Localization, and Source Signal Deconvolution 503
Z. Michalopoulou

Numerical Optimization of Non-Adaptive Microphone Arrays 507
A. Goldin

Joint Direction-of-Arrival and Array Shape Tracking for Multiple Moving Targets 511
J. Goldberg, A. Perez-Neira and M. Lagunas

Comparison of Probabilistic Least Squares and Probabilistic Multi-Hypothesis Tracking Algorithms for Multi-Sensor Tracking 515
M. Krieg and D. Gray

Direction Finding with Imperfect Wavefront Coherence: A Matrix Fitting Approach Using Genetic Algorithm ..... 51[illegible]
A. Gershman, C. Mecklenbräuker and J. Böhme

Design of an Optimum Wideband Active Sonar Array with Robustness 523
S. Abeysekera and Y. Leung

Ocean Signal Processing

Multipath Time-Delay Estimation 527
J. Fuchs

Fast Maximum Likelihood Estimation with Multiple Signal Initialization 531
R. MacLeod and R. Vaccaro

An Algorithm for Detecting Closely Spaced Delay/Doppler Components 535
A. Habboosh, R. Vaccaro and S. Kay

Improvement of TDOA Measurement Using Wavelet Denoising with a Novel Thresholding Technique 539
S. Wu, H. So and P. Ching

A Short-Time Wiener Filter for Noise Removal in Underwater Acoustic Data 543
C. Therrien, K. Frack and N. Fontes

Fast Approximate Subspace Tracking (FAST) 547
D. Tufts, E. Real and J. Cooley

A Subspace Framework for Fast Parameter Estimation with Known Waveforms 551
B. Freburger, D. Tufts and T. Palka

Terrain Classification in Polarimetric SAR Using Wavelet Packets 555
N. Keshava and J. Moura

Electromagnetic Matched-Field Processing for Target Height Finding with Over-the-Horizon Radar 559
M. Papazoglou and J. Krolik

Time-Frequency Classification Using a Multiple Hypotheses Test: An Application to the Classification of Humpback Whale Signals 563
G. Roberts, A. Zoubir and B. Boashash

Source Classification Using Pole Method of AR Model 567
J. Huang, J. Zhao and Y. Xie

The LMMSE Estimate-Based Multiuser Detector: Performance Analyses and Adaptive Implementation 571
H. Ge

Improved Doppler Tracking and Correction for Underwater Acoustic Communications 575
M. Johnson, L. Freitag and M. Stojanovic

A Blind Multichannel Combiner for Long Range Underwater Communications 579
B. Sharif, J. Neasham, D. Thompson, O. Hinton and A. Adams

DSP Processors

Very Long Instruction Word Architectures for Digital Signal Processing 583
J. Mellott and F. Taylor

A Novel 32 Bit RISC Architecture Unifying RISC and DSP 587
C. Baumhof, F. Müller, O. Müller and M. Schlett

A Dual-Issue RISC Processor for Multimedia Signal Processing 591
H. Sato, E. Holmann, T. Yoshida, M. Matsuo and T. Kengaku

A Processor-Coprocessor Architecture for High End Video Applications 595
E. Maas, D. Herrmann, R. Ernst, P. Rüffer, S. Hasenzahl and M. Seitz

An MPEG-2 Encoder Architecture Based on a Single-Chip Dedicated LSI with a Control MPU 599
Y. Ooi, O. Ohnishi, Y. Yokoyama, Y. Katayama, M. Mizuno, M. Yamashina, H. Takano, N. Hayashi and I. Tamitani

An Efficient and Reconfigurable VLSI Architecture for Different Block Matching Motion Estimation Algorithms 603
X. Zhang and C. Tsui

An Operation-Saving VLSI Geometry Engine Core 607
K. Karagianni, G. Diamantakos, V. Paliouras and T. Stouraitis

The FFT Butterfly Operation in 4 Processor Cycles on a 24 Bit Fixed-Point DSP with a Pipelined Multiplier 611
M. Grajcar and B. Sick

New Unified VLSI Architectures for Computing DFT and Other Transforms 615
S. Hsiao and C. Yen

Half-Rate GSM Vocoder Implementation on a Dual Mac Digital Signal Processor 619
M. Prasad, P. D'Arcy, A. Gupta, M. Diamondstein and H. Srinivas

VLSI Implementation of an Area-Efficient Architecture for the Viterbi Algorithm 623
C. Cabrera, M. Boo and J. Bruguera

DSP Building Blocks & Arithmetic

Low-Area Dual Basis Divider Over GF(2^M) 627
L. Song and K. Parhi

VLSI Architecture for Datapath Integration of Arithmetic Over GF(2^M) on Digital Signal Processors 631
W. Drescher, K. Bachmann and G. Fettweis

A Fast Direction Sequence Generation Method for CORDIC Processors 635
S. Nahm and W. Sung

A Radix-4 Redundant Cordic Algorithm with Fast On-Line Variable Scale Factor Compensation 639
C. Li and S. Chen

Pipelining of Cordic Based IIR Digital Filters 643
J. Ma, K. Parhi and E. Deprettere

An Asynchronous Implementation of the MAXLIST Algorithm 647
C. Myers and H. Zheng

A Novel Systematic Mapping Approach for Highly Efficient Multiplexed FIR-Filter Architectures 651
W. Wilhelm and T. Noll

Design Automation for DSP

An Upper Bound of the Throughput of Multirate Multiprocessor Schedules 655
R. Schoenen, V. Zivojnovic and H. Meyr

Minimizing the Number of Operations in DSP Computations 659
I. Hong and M. Potkonjak

BEEHIVE: An Adaptive, Distributed, Embedded Signal Processing Environment 663
S. Famorzadeh, V. Madisetti, T. Egolf and T. Nguyen

On Objective Function Selection in List Scheduling Algorithms for Digital Signal Processing Applications 667
J. Jonsson and J. Vasell

VLSI High Level Synthesis of Fast Exact Least Mean Square Algorithms Based on Fast FIR Filters 671
J. Diguet, O. Sentieys, D. Chillet and J. Philippe

Hierarchical VHDL Libraries for DSP ASIC Design 675
J. McCanny, D. Ridge, Y. Hu and J. Hunter

DSP Quant: Design, Validation, and Applications of DSP Hard Real-Time Benchmark 679
C. Lee, D. Kirovski, I. Hong and M. Potkonjak

Constructing Memory Layouts for Address Generation Units Supporting Offset 2 Access 683
B. Wess and M. Gotschlich

Modulo-Addressing Utilization in Automatic Software Synthesis for Digital Signal Processors 687
M. Willems, H. Keding, V. Zivojnovic and H. Meyr

Cooperative Register Assignment and Code Compaction for Digital Signal Processors with Irregular Datapaths 691
W. Kreuzer and B. Wess

Optimization of Embedded DSP Programs Using Post-Pass Data-Flow Analysis 695
A. Sudarsanam, S. Malik, S. Tjiang and S. Liao

Code Positioning to Reduce Instruction Cache Misses in Signal Processing Applications on Multimedia RISC Processors 699
H. Stolberg, M. Ikekawa and I. Kuroda

Code Generation by Using Integer-Controlled Dataflow Graph 703
T. Miyazaki and E. Lee

Fixed-Point C Compiler for TMS320C50 Digital Signal Processor 707
J. Kang and W. Sung

Author Index **A1**

This material may be protected by Copyright law (Title 17 U.S. Code)

MULTIMODAL INTERFACES FOR MULTIMEDIA INFORMATION AGENTS

Alex Waibel, Bernhard Suhm, Minh Tue Vo and Jie Yang

Interactive Systems Laboratories
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA 15217 (USA)
{waibel,bsuhm,tue,yang+}@cs.cmu.edu

ABSTRACT

When humans communicate they take advantage of a rich spectrum of cues. Some are verbal and acoustic. Some are non-verbal and non-acoustic. Signal processing technology has devoted much attention to the recognition of speech, as a single human communication signal. Most other complementary communication cues, however, remain unexplored and unused in human-computer interaction. In this paper we show that the addition of non-acoustic or non-verbal cues can significantly enhance robustness, flexibility, naturalness and performance of human-computer interaction. We demonstrate computer agents that use speech, gesture, handwriting, pointing, spelling *jointly* for more robust, natural and flexible human-computer interaction in the various tasks of an information worker: information creation, access, manipulation or dissemination.

1. INTRODUCTION

Human-computer interfaces today are limited and inflexible and do not take advantage of the many communication channels humans use to communicate verbal and non-verbal ways. Humans speak, point, gesture, write, fixate, use facial expressions, head-motion, eye-contact, etc. to express ideas, intentions and feelings. To build computer systems that operate more flexibly and robustly, and that are more intuitive to use by anyone, computer interfaces must be able to understand and process this multiplicity of human communication cues. Recently many research projects have been conducted to address this problem (e.g. [1], [2], [3], [4]). The need for more intuitive interfaces becomes all the more pressing as multimedia presentation, that is output, is becoming commonly available.

In this paper we describe information agents that use multimodal interfaces to access, manipulate, create and disseminate information. In particular, we seek effective methods by which human users can easily interact with a computer by speaking, pointing, drawing, spelling, etc. We show how cross-modal cues can be used effectively to recover from errors or miscommunications, since confusions in one modality usually are unambiguous in another. We describe how multiple modalities can cooperate and complement each other in various phases of the preparation of multimedia documents. Moreover, the added flexibility and robustness significantly increases reliability and naturalness, and with it the usability and acceptance of information systems. We present QuickDoc as a prototypical application that combines our multimodal subsystems in a simple, but powerful way.

2. INTERPRETATION OF MULTIMODAL INPUT

2.1. Multimodal Components

Our labs have implemented recognizers and processing modules for various input modalities, most notably speech, pen input, and face tracking. These modality processors constitute the basic components of all our multimodal systems.

Speech

Our speech recognition subsystem is based on the recognition front-end of the JANUS speech translation system [5] which is capable of processing speaker-independent, spontaneous speech. The recognizer can be adapted to any task domain by retraining the language models and possibly tuning the acoustic models if the domain involves special vocabulary. We also have a high-performance, real-time continuous spelling recognizer for large lists of 100,000 or more names.

Gesture

Our approach to pen-based gesture recognition is to decompose pen strokes into sequences of basic shapes such as line, arc, arrow, circle, cross...[6] The same gesture shape may mean different things depending on the surrounding context, hence each gesture component is augmented by *gesture contexts* indicating spatial relationships between the gesture and nearby objects in the user interface.

Handwriting

Our MS-TDNN-based handwriting recognizer [7] is capable of processing writer-independent, continuous (cursive) handwriting at a recognition rate of 94% on a 20,000-word vocabulary. We employ simple heuristics to decide when to invoke handwriting recognition on pen input, e.g., when the gesture recognizer cannot identify the input strokes as basic shapes.

Face Tracking

We base our face-tracking subsystem on face color clustering and motion detection [8]. The face tracker can control a pan-tilt-zoom camera to follow a freely moving person, producing a constant-sized image of the face area in real time. Another software layer built on top of this face tracker can identify the eyes and other face features, build a model of the face, and estimate the gaze direction to track the head pose [9].

2.2. Joint Interpretation

In order to make sense of input from all available sources, we need a multimodal interpreter capable of producing an interpreta-

0-8186-7919-0/97 $10.00 © 1997 IEEE

tion of user intent (e.g., a command to execute in the application interface) from the output of the modality processors.

In our joint interpretation scheme, the user intent is represented by a frame consisting of slots specifying pieces of information such as the action to carry out or the parameters for that action. Recognition output from the modality processors are parsed into partially filled frames that are merged together to produce the combined interpretation as described in [6]. This technique leads to uniform handling of high-level information from all input sources, which is very important for modularity and extensibility. To add another input modality we need only provide a module to convert low-level recognizer output to a partially filled frame to be merged with others. In addition, context information can be retained across input events by merging with previous interpretation frames.

3. MULTIMODAL ERROR CORRECTION

Intensive research over recent years has boosted the performance of speech recognition technology significantly. Nevertheless it is widely believed that recognition performance will remain limited, at least for the foreseeable future. The potential for error with any speech interface will be balanced by the redundant and alternate ability to enter information or queries by spelled, handwritten or typed input. Rather than fighting a particular recognition error the user can therefore quickly circumvent the problem by choosing a different modality. Such alternates not only eliminate user frustration with potentially recurring errors, but also provide an avenue for background learning and adapting.

3.1. Multimodal Interactive Error Repair

Our approach is to have the user collaborate with the system in recovering from interpretation errors [10]. First, errors have to be identified, either initiated by the system, for instance based on some confidence measure, or by the user. If a graphical user interface (GUI) is available, the user can simply highlight erroneous words (the reparandum) in the recognition hypothesis displayed. For error correction, the user provides additional input, choosing among different correction methods: repair by repeating the reparandum by respeaking, spelling out loud, or handwriting; repair by paraphrasing the reparandum; repair by pen gestures (e.g. to delete or insert), in addition to the standard repair by typing or selecting among N-best alternatives.

The rationale for this multimodal approach to error recovery is twofold. First, it exploits the fact that different input modalities are orthogonal: words which are confusable in one modality, can be disambiguated in a different modality (e.g. "road" and "rote" spoken versus spelled). Secondly, recent studies [11] show that switching modality after system misinterpretation alleviates user frustration.

3.2. Error Repair for Multimedia Information Agents

The design of speech user interfaces is constrained by the context of the application, and so is the design of error repair. The application context varies along several dimensions: available modalities (speech only versus GUI), what is repaired (isolated words, phrases, sentences), goal of repair is (get verbatim every word versus get the intended action), dialogue metaphor (command controls versus conversational). For multimedia information agents, we see the following two application contexts as particularly relevant:

- (Mixed initiative) Spoken dialogue, for example in user queries and command control of the application. In this context, a GUI may be available, repair will typically be performed at the level of phrases, the goal of repair is to initiate the intended action (semantic repair), and a conversational dialogue is desirable.
- Dictation, for example to entry in fields, or to dictate reports. In this context, a GUI is very appropriate, repair will occur from isolated word to sentence level, the goal of repair is to get each word correct (verbatim repair).

We have built prototypical speech user interfaces with error repair capabilities for dictation and form filling tasks. In this application context, requiring the user to identify errors by highlighting them is appropriate, since the user naturally focuses on the actual recognition hypothesis. For error correction, we offer to repeat the input (potentially switching to another modality), and pen gestures to indicate where to insert words, or which word to delete. In form filling tasks, knowledge of the current field can provide powerful constraints for recognition by restricting vocabulary and language model accordingly.

3.3. Evaluating Interactive Error Repair

Given a set of error recovery methods feasible with current technology in a given application context, a crucial issue is to predict which methods a user will prefer. Based on the assumption of a rational user [13], we use the time to complete some input, including the time needed to correct errors, as an objective and easily quantifiable measure for the effectiveness of different error recovery methods, and main predictor of user preference. By assuming stochastic independence of the various repair attempts, the expected accuracy after a certain number of repair attempts can be estimated as a geometric series. Given a high level of accuracy sufficient for the application (e.g. 99%), the number of repair attempts can be estimated, and thus the total time required to input including repair [12].

A pilot evaluation on a form filling task [12] suggests that given current technology, repair by spelling and handwriting can be very effective, and significantly better than standard choice from N-best alternatives. Using the context of a repair, e.g. in rescoring N-best lists obtained by decoding repair input, can substantially improve the accuracy of repair. Although results are still preliminary, they show that our multimodal approach to interactive error recovery is very promising.

4. MULTIMODAL INFORMATION AGENTS

The objective for multmodal information agents is to quickly access, create, manipulate, and disseminate multimedia information. A multimodal agent can label multimedia information with appropriate classifications by voice or/and gesture annotations. All annotations are incorporated with original information. The agent will automatically attach hotlinks on the world-wide-web to broaden information sources. The agent then will generate an HTML format report. The agent can search all the information in the report including voice and gesture annotations and voice mails. The agent can repair errors and add more information. Finally, the agent can disseminate the report in electronic form or printed form.

4.1. Information Access

Multimodal information agents can speed up navigation of Information space and queries of a database. A city map, for example, can be queried by a combination of gesture, handwriting and speech. A query such as: "Show me all hospitals in this <circling-gesture> area," or "How far is from here to there <tracing-gesture>?" will access information that would otherwise have to be assembled and combined painstakingly in a series of steps. Multimodal agents will recognize gesture and speech concepts and map them onto a semantic representation. The semantic representation can be filled interchangeably by both gestures and verbal phrases using pattern recognition and chart parsing techniques. The parsing result is then sent to query the database. If interpretation errors occur, the user corrects them using multimodal correction techniques. The whole process is transparent to the user.

Searching multimedia data can be viewed as natural extension of multimodal database queries. For instance, voice annotations or video clips added during user interaction can be interpreted and indexed using multimodal interpretation components, and then accessed using database queries techniques.

4.2. Information Creation

A user often needs to add information to a database or modify existed information. Multimodal agents are also useful in such a process. For example, a doctor needs to mark his/her diagnoses on X-ray, CT, MR images and make comments on the case. He/she can use speech, gesture, and handwriting to do this. Similarly iconic gestures may be used on a touch sensitive display or writing pad, to quickly enter resources or objects in a database, such as a map, in non-verbal ways. Iconic gestures can be defined to represent objects that would be entered in the database. In the same way, handwriting symbols can be used to mark or tag objects on a screen. Multimodal input also need not always be interpreted, but can often be stored for later review. Thus a voice annotation, or a circling gesture can be simply stored to go along with an image for later review and retrieval. The efficiency of multimodal agents for information creation can illustrated by the following example. The scenario is an image analyst sitting in the front of a computer and classifying various targets. Images can be popped up quickly and in rapid succession and the analyst provides a quick classification. The image and its classification are entered into a database along with the analyst's spoken and gestured annotation. With the human in the loop to make decisions on intelligence relevance, but with the availability of a better (multimodal) interface, to do so quickly, the job of data creation and manipulation can be carried out with greater reliability and efficiency compared to a fully automated system.

4.3. Information Manipulation

Information will not only be requested to viewed, but will also need to be updated. A multimodal multimedia information system can offer new ways to maintain the underlying databases: the user can use speech, gesture, writing, spelling to update database information efficiently. manipulating visual representations on the display rather than having to learn and use a separate database manipulation system. Gestures play an important role in specifying object parameters (e.g. different iconic shapes to represent different object types) and spatial constraints (e.g., location and extent of objects), while speech is useful in specifying parameters not easily expressed visually. For example he/she might circle a bridge on a map and say "this bridge has been destroyed", or might cross out an object without words, etc.

In some situations, the answer for a query cannot be directly obtained from the database and the information from the database has to be future processed. For example, the query "Show me the nearest restaurant" requires not only database access but also other computations.

4.4. Information Dissemination

In addition to information access, creation and manipulation, the product has to be packaged and readied for later viewing. The world wide web can serve as a natural medium to disseminate multimedia information. The multimodal input signals are interpreted and compiled, and an HTML format report is automatically generated on the fly. It is then automatically available for viewing together with the original aligned speech waveform attached to it as hotlinks. Multimodal interaction can also be used to manipulate objects in a report, such as positioning objects on a page appropriately, or attaching appropriate hotlinks.

To quickly generate a multimedia report, speech dictation can be combined with point and click actions to arrange reports on the fly. Such a multimodal report generation would be faster than typing and can generate a hierarchy of facts and notions. The generation thus would consist of some dictation or typing, some multimodal object manipulation ("move this <point> here <point> and link it as an explanation for that <click>") and some handwriting or drawings that are to be attached. The result of these communicative acts will compile into HTML code, that is ready to be transferred. Time savings are possible in several ways: the report generation can be done more efficiently by speech, gesture, or handwriting, and by using references to already existing multimedia stickies, factoids, video footage or explanations, and the recipient of the report can probe and browse a hierarchical multimedia report in a non-linear fashion more efficiently, calling up aspects of the assessment as needed.

4.5. Controlling the Interface

Although the goal of a multimodal interface is to provide as transparent an access to functionality as possible, interaction to control various aspects of the system can't be eliminated.

For example for navigation within visually presented image data, the user typically modifies the view by operations such as zooming and panning. The integration of speech and gesture increases flexibility by allowing different attributes of an operation to be specified in whatever modality is more appropriate for each attribute; for instance, the effect of a spoken "zoom in to this <circle> area" accompanied by a circle drawn around the desired area is much more difficult to achieve using a single input modality

Additionally, multiple modalities can be used to enhance the reliability. For example, a gaze tracker can be used to detect the user's focus of attention. When the user is looking at a window on a screen, the window will be highlighted. No action is taken unless the user uses a voice command to confirm the selection. The voice commands could be ``select this window" or ``close window", etc. This can reduce unintended actions caused by the user randomly looking around.

5. THE QUICKDOC APPLICATION

QuickDoc is an example of an application that combines our multimodal subsystems in a simple but powerful way, letting the user perform a repetitive task with speed and convenience. The task is for a doctor to go through a series of images such as X-rays or computer-aided tomography scans, quickly identify an anomalous

area, label the area with the name of a disease or condition, and attach relevant comments. The end product is an HTML report that summarizes the doctor's findings in a compact table listing the annotated images, the corresponding preliminary diagnoses, and automatically generated hotlinks to relevant sites based on the diagnoses.

The QuickDoc user labels each presented image by circling an area on the image (drawing directly on a touch-sensitive screen) and speaking a disease name as well as a percentage representing the confidence level of the diagnosis. The gesture recognizer identifies the circle and generates an area marker on the image, attaching other parts of the gesture as written annotations (a future version may run the handwriting recognizer to turn those into text). The user can issue a spoken command to initiate a voice annotation attachment; the recorded audio file is also run through the speech recognizer to produce an automatic transcription. When all the images have been processed, another spoken command causes the HTML report to be generated. Optionally the collected data can also be edited in a multimodal dictation and repair facility before being processed into HTML.

The generated report contains thumbnails of the annotated images accompanied by disease label, confidence level, location in the image (automatically extracted from the circling gesture), and voice annotation playback icon if applicable. The thumbnail is linked to a more detailed page containing the full-size image and the automatic voice transcription. The name of the disease is looked up in a database and turned into a hotlink to a page listing Web sites relevant to that disease. A keyword search mode allows to search spoken keywords in the text of the report and the automatic voice transcriptions; entries containing hits are then highlighted in the report.

QuickDoc thus embodies multimodal creation, manipulation, dissemination, and access of multimedia information in a simple yet surprisingly useful application.

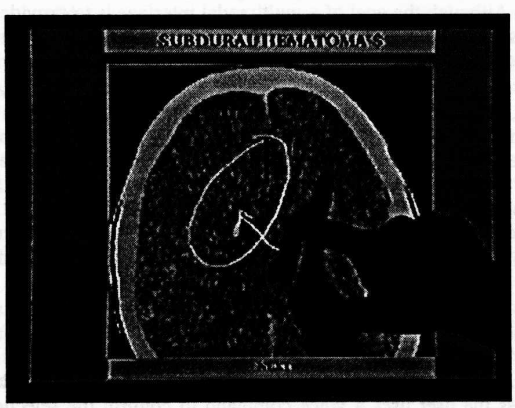


FIGURE 1: Multimodal information creation in QuickDoc: labeling some area by a circle gesture and speech ("subdural hematoma")

6. CONCLUSIONS

In this paper we have described systems that combine speech, gesture, handwriting, pointing, spelling, and other communication modalities into interfaces that are robust, flexible, and intuitive to use. We described how multimodal error recovery can ease the problem of unreliable automatic interpretation of communication modalities, in particular speech, thus removing a major obstacle in making multimodal interfaces truly usable. We demonstrated how these concepts can be combined for a more user friendly multimedia document production technology.

7. ACKNOWLEDGEMENTS

This research was sponsored by the DARPA under the Department of the Navy, Naval Research Office under grant number N00014-93-1-0806. Views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Navy or the U.S. Government.

8. REFERENCES

[1] H. Ando, Y. Kitahara and N. Hataoka: "Evaluation of Multimodal Interface Using Spoken Language and Pointing Gesture on Interior Design System", *Proc. ICSLP*, 1994, Vol. 2, pp. 567-570.

[2] T. Nishimoto, N. Shida, T. Kobayashi, K. Shirai: "Multimodal Drawing Tool Using Speech, Mouse and Keyboard", *Proc. ICSLP*, 1994, Vol. 3, pp. 1287-1290.

[3] J. Wang: "Integration of Eye-gaze, Voice and Manual Response in Multimodal User Interface", *Proc. ICSMC*, 1995, Vol. 5, pp. 3938-42.

[4] A. Waibel, M.T. Vo, P. Duchnowski, and S. Manke: "Multimodal Interfaces," *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, McKevitt, P. (Ed.), Vol. 10, Nos. 3-4, 1995.

[5] B. Suhm, P. Geutner, T. Kemp, A. Lavie, L.J. Mayfield, A. McNair, I. Rogina, T. Schultz, T. Sloboda, W. Ward, M. Woszczyna and A. Waibel: "JANUS: Towards Multilingual Spoken Language Translation," *Proc. ARPA SLT Workshop 95* (Austin, Texas).

[6] M.T. Vo and C. Wood: "Building and application framework for speech and pen input integration in multimodal learning interfaces," *Proc. ICASSP'96* (Atlanta, GA).

[7] S. Manke, M. Finke and A. Waibel: "The Use of Dynamic Writing Information in a Connectionist On-Line Cursive Handwriting Recognition System," *Advances in Neural Information Processing Systems 6*, Morgan Kaufmann, 1994.

[8] J. Yang and A. Waibel: "A real-time face tracker," *Proc. WACV'96*, pp. 142-147.

[9] R. Stiefelhagen, J. Yang and A. Waibel: "A Modelbased Gaze Tracking System," *Proc. IEEE International Joint Symposia on Intelligence and Systems - Image, Speech & Natural Language Systems*, pp. 304-310, 1996.

[10] A.E. McNair and A. Waibel: "Improving Recognizer Acceptance through Robust, Natural Speech Repair", *Proc. ICSLP*, 1994, Vol. 3, pp. 1299-1302.

[11] S.L. Oviatt and R. VanGent: "Error Resolution during Multimodal Human-Computer Interaction", *Proc. ICSLP*, 1996.

[12] B. Suhm, B. Myers and A. Waibel: "Interactive Recovery from Speech Recognition Errors in Speech User Interfaces", *Proceedings of the International Conference on Spoken Language Processing - ICSLP*, 1996.

[13] A. Newell, S. Card and P. Moran: "The Psychology of Human Computer Interaction", Lawrence Earlbaum Associates, 1983.

APPENDIX G

Journals, Magazines, Newspapers

IEEE International Conference on Acoustics, Speech, and Signal Processing : [proceedings]

AUTHOR / CREATOR
ICASSP (Conference)

CONFERENCES
ICASSP (Conference)

AVAILABLE AS
Online
Physical

View Online

Available Issues / Request a Copy

Save to Lists

Citation

Physical Locations

Multiple items exist. Use Available Issues / Request a Copy to see what's on the shelf.

Verona Shelving Facility	**by request only** TA365 I102

This Location Has
1997

Publication Details

CREATOR
sponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society

INCLUDED TITLE
IEEE Xplore conferences

FORMAT
Journals, Magazines, Newspapers

LANGUAGE
English

CONTRIBUTORS
IEEE Signal Processing Society

PUBLICATION
Los Alamitos, CA : IEEE Computer Society Press, c1997

PUBLICATION DATES
1997

PHYSICAL DETAILS
1 volume : illustrations ; 28 cm

OCLC
ocm37249153

PRECEDING TITLES
Conference proceedings

SUCCEEDING TITLES

Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing

Notes

Also available in electronic format on the World Wide Web.

Subjects

Speech processing systems -- Congresses.

Signal processing -- Congresses.

Electro-acoustics -- Congresses.

Underwater acoustics -- Congresses.

Content Types

Congresses

Items Related By Call Number

Shelf Browse

Journals, Magazines, Newspape...	**Journals, Magazines, Newspape...**	**Journals, Magazines, Newspape...**	**Journals, Mag**
ICASSP : [proceedings]	Record	Conference record	Proceedings
ICASSP (Conference)	ICASSP (Conference)	ICASSP (Conference)	ICASSP (Confe

Physical
TA365 I102

Physical
TA365 I102

Online
Physical
TA365 I102

Physical
TA365 I102

Information from the Web

Library Staff Details

MARC Bibliographic Record

LEADER	01800mas a2200445 a 4500
001	9943740843602122
005	20180323022142.0
007	cr |||||||||||
008	970707d19971997cauar 1 0eng d
010	$a 98655049 $zsn 97038149
012	$a2$i9707
035	$a(OCoLC)ocm37249153
035	$9AYE1159UW

035		$a(WU)4374084-uwmadisondb
035		$a(EXLNZ-01UWI_NETWORK)999858382602121
040		$aMYG$beng$cMYG$dDLC$dMYG$dGZM
042		$alc
049		$aGZMA
050	00	$aTK7882.S65$bI37a
082	00	$a621.38/043/05$219
111	2_	$aICASSP (Conference)
245	10	$aIEEE International Conference on Acoustics, Speech, and Signal Processing :$b[proceedings] /$csponsored by the Institute of Electrical and Electronics Engineers Signal Processing Society.
260		$aLos Alamitos, CA :$bIEEE Computer Society Press,$cc1997.
300		$a1 volume :$billustrations ;$c28 cm
310		$aAnnual
336		$atext$btxt$2rdacontent
337		$acomputer$bc$2rdamedia
338		$aonline resource$bcr$2rdacarrier
362	0_	$a1997.

530		$aAlso available in electronic format on the World Wide Web.
515		$aIssued in 5 or more pts.
650	_0	$aSpeech processing systems$vCongresses.
650	_0	$aSignal processing$vCongresses.
650	_0	$aElectro-acoustics$vCongresses.
650	_0	$aUnderwater acoustics$vCongresses.
710	2_	$aIEEE Signal Processing Society.
730	0_	$aIEEE Xplore conferences.
780	00	$aICASSP.$tConference proceedings$w(DLC) 98655050$w(OCoLC)32748056
785	00	$aICASSP.$tProceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing$x1520-6149$w(DLC)sn 98038280$w(OCoLC)40157209
856	41	$zAvailable through IEEE Xplore$uhttp://ieeexplore.ieee.org/servlet/opac?punumber=1000002
997		$aMARCIVE

MMS IDs
Document ID: 999858382602121
Network Electronic IDs: 999858382602121
Network Physical IDs: 999858382602121
mms_mad_ids: 9943740843602122