

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

GOOGLE LLC,

Petitioner,

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner.

Case IPR2022-01027

Patent No. 8,467,543

Declaration of Shauna L. Wiest Regarding *Hussain*

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

I. Introduction

1. I have prepared this Declaration in connection with the Petition for *Inter Partes* Review of U.S. Patent No. 8,467,543, 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 901 New York Avenue, NW, Washington, DC 20001-4413.

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

4. I earned a Master of Science in Library Science from the University of North Carolina at Chapel Hill in 1999, and a Bachelor of Arts in Political Science from the University of California at San Diego in 1989. I have worked as a law librarian for over twenty years. I have been employed in the Research & Information Services Department at Finnegan, Henderson, Farabow, Garrett & Dunner, LLP 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, and Making Materials Publicly Available

5. I have knowledge of and experience with standard library practices regarding the receipt, cataloging, and making materials available to the public. I have knowledge of and experience with the Machine-Readable Cataloging (MARC) system, an industry-wide standard that libraries use to catalog materials. I also have knowledge of and experience with OCLC (previously called Online Computer Library Center, Inc.) and OCLC Control Numbers, which are unique accession numbers assigned by the OCLC system when a bibliographic record is cataloged and made available to the public through WorldCat. OCLC control numbers are often used by the public to search for specific records within WorldCat, the latter being an online database of 520,593,877 bibliographic records from approximately 72,000 libraries as of December 2021 (Source: <https://www.oclc.org/en/worldcat/inside-worldcat.html>).

6. By the mid-1970s, standard library practice involved cataloging items utilizing the MARC system. The MARC system was developed during the 1960s to standardize bibliographic records so they could be read by computers and shared among libraries. By the mid-1970s, MARC had become the international standard for bibliographic cataloging and is still used today.

7. Many libraries provide public access to their MARC records via the Internet and/or their public catalog systems at the library. The MARC record system uses a specific three-digit numeric code (from 001-999) to identify each field in a catalog record. For example, field tag 008 provides the six-digit date the item was cataloged (Date entered on file). The first six characters of field tag 008 are always in the “YYMMDD” format. Field tag 245 identifies the full title statement of the work; field tags 260 and 264 denotes the place of publication, including date of publication, distribution, and copyright for the item; and field tag 300 identifies the physical description, often including holdings information, for the work.

8. Based on standard library practice, when a library receives an item, it stamps the item with the library name and 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.

9. Generally, after an item is cataloged, the public may access the item by searching a catalog, browsing the 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.

III. MARC Records for *Hussain*

10. This Declaration relates to the dates of receipt and public availability of the following: Amir Hussain, Douglas R. Campbell, and

Thomas J. Moir, *A New Metric For Selecting Sub-Band Processing in Adaptive Speech Enhancement Systems*, presented at Eurospeech '97: 5th European Conference on Speech Communication and Technology, Rhodes, Greece, 22-25 September 1997 (“*Hussain*”). The *Hussain* reference appears at page 2611, identified by paper number ThAA.15, in Volume V of the Eurospeech '97 Conference Proceedings. I understand that *Hussain* has been submitted as Exhibit 1006 in this proceeding.

11. **Appendix A** to this declaration is a true and correct copy of *Hussain* obtained from The Ohio State University Library. Appendix A includes the front matter and table of contents for Volume V containing *Hussain*. Appendix A also contains handwritten notes identifying the call number for *Hussain* as P95 E88 1997 V. 5.

12. **Appendix B** to this declaration is a true and accurate copy of the MARC catalog record from The Ohio State University for its copy of Eurospeech '97 containing *Hussain*, downloaded from <https://library.ohio-state.edu/search~S7?/.b5440967/.b5440967/1%2C1%2C1%2CB/marc~b5440967> on May 17, 2022.

13. In The Ohio State University MARC catalog record, the fixed data elements of MARC field 008 are listed as 000630s1997. This indicates that Eurospeech '97 containing *Hussain* was cataloged by The Ohio State University on June 30, 2000. The subfield “code “s” appearing in MARC field 008 refers to the

“single known date” for the item, which in this catalog record is 1997. Based on standard library practices, *Hussain* would have been shelved and made available to the public a few days or weeks after June 30, 2000.

14. In The Ohio State University MARC catalog record, the Title Statement field 245 denotes the title and statement of responsibility for the work as Eurospeech 97 :|b5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September, 1997. The subfield code “b” refers to the remainder of the title, which in this record is 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September, 1997.

15. Field 264 of The Ohio State University MARC catalog record denotes the place of publication, distribution, and date/copyright information for Eurospeech '97 containing *Hussain*. Field 264 is identified as Grenoble, France :|bESCA,|c[1997]. The subfield code “b” refers to the name of the producer, publisher, distributor, or manufacturer, which in this record is ESCA, or European Speech Communication Association. The subfield code “c” refers to the copyright date, which for this record is 1997.

16. Field 300 of The Ohio State University MARC catalog record identifies the physical description and holdings of the item. The MARC catalog record denotes the physical description of Eurospeech '97 containing *Hussain* as containing 5 volumes.

17. **Appendix C** to this declaration is a true and correct copy of The Rutgers University Libraries MARC record for its copy of *Hussain*, which was downloaded from <https://bit.ly/3sdtXkt> on May 17, 2022

18. In The Rutgers University Libraries MARC catalog record, the fixed data elements of MARC field 008 are listed as 990730s1997. This indicates that Eurospeech '97 was cataloged by The Rutgers University Libraries on July 30, 1999. The subfield “code “s” appearing in MARC field 008 refers to the “single known date” for the item, which for this catalog record is 1997. Based on standard library practices, Eurospeech '97 would have been shelved and made available to the public a few days or weeks after July 30, 1999.

19. The Rutgers University Libraries MARC catalog record field tag 245 denotes the title and statement of responsibility for the work as “Eurospeech '97 : \$b5th European Conference on Speech Communication and Technology : Rhodes, Greece, 24-25 September 1997.”

20. The Rutgers University Libraries MARC catalog record field tag 260 denotes that Eurospeech '97 was published in Patras, Greece by the European Speech Communication Association in 1997.

21. Field 300 of the Rutgers University Libraries MARC catalog record identifies the physical description and holdings of the item. The MARC catalog record denotes the physical description of Eurospeech '97 as 2 volumes (containing volumes 1 and 2). Based on the public availability of volumes 1 and 2,

other volumes, including volume 5, would have been made publicly available if requested within a few days or weeks after July 30, 1999.

IV. OCLC Records for *Hussain*

22. A public search of WorldCat lists nine individual OCLC catalog entries for Eurospeech '97: 5th European Conference on Speech Communication and Technology: Rhodes, Greece, 22-25 September 1997, as of January 31, 2022.

Appendix D to this declaration is a true and accurate copy of the list of the nine individual OCLC catalog entries from WorldCat containing Eurospeech '97.

Appendix D was downloaded from:

https://www.worldcat.org/search?q=au%3Aeurospeech&dblist=638&fq=yr%3A1997&qt=facet_yr%3A on May 17, 2022. Each of the nine catalog entries can be accessed to view the libraries who have cataloged and made publicly available, Eurospeech '97 containing *Hussain*.

23. **Appendix E** to this declaration is a true and accurate copy of OCLC Control Number 105130238, which is among the nine individual OCLC catalog entries listed in Appendix D. Appendix E was downloaded from:

<http://www.worldcat.org/oclc/1051302387> on May 17, 2022 and lists the cataloged publicly available library holdings for OCLC Control Number 1051302387. OCLC Control Number 1051302387 denotes the date (1997) of the publication, the physical description (5 volumes), and title (EuroSpeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25

September, 1997 : proceedings) for the publication. Each of the eight other entries in Appendix D list similar cataloging dates, physical description, title information, and publicly available library holdings information for *Hussain*.

V. Conclusion

24. 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.

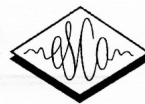
25. I declare that all statements made herein of my knowledge are true, that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on May 17, 2022, in Washington, DC.

A handwritten signature in black ink, appearing to read "Shauna L. Wiest", written over a horizontal blue line.

Shauna L. Wiest

APPENDIX A



EuroSpeech '97

5th EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY

RHODES - GREECE

22-25 September 1997

UNDER THE AUSPICES OF

THE MINISTRY OF CULTURE

THE MINISTRY OF THE AEGEAN

THE GENERAL SECRETARIAT OF SCIENCE AND RESEARCH

PROCEEDINGS VOLUME 5



ORGANIZER:

UNIVERSITY OF PATRAS

WIRE COMMUNICATIONS LABORATORY

261 10 Rion - Patras - Greece

Meta Platforms, Inc. - Exhibit 1010

Page 11 of 38

EUROPEAN SPEECH COMMUNICATION ASSOCIATION (ESCA)

**5th European Conference on Speech Communication and Technology:
EUROSPEECH '97**

Title: Eurospeech '97 Proceedings
Subtitle: ESCA 5th European Conference on Speech Communication and Technology
Volumes: I, II, III, IV, V
Editors: G. Kokkinakis, N. Fakotakis, E. Dermatas
ISSN: 1018-4074

P95
E88
1997
V.5

Additional copies may be ordered from:

ESCA - European Speech Communication Association
ICP, Universite Stendhal,
BP 25X, 38040 GRENOBLE Cedex 9, FRANCE

Tel: (+33).4.76.82.43.36

Fax: (+33).4.76.82.43.35

Cover Design: Artemis Kokkinaki-Papapetrou
Edited at: WCL, University of Patras, Greece
Printed by: TYPOFFSET, K. Manouri-Stanidi, Patras, Greece

VOLUME 5

SESSION: ThMD
Speaker Recognition and Language Identification
Chair: Douglas Reynolds, MIT, USA

ThMD.1 Gaussian Mixture Models with Common Principal Axes and Their Application in Text-Independent Speaker Identification 2279
 Kuo-Hwei Yuo, Hsiao-Chuan Wang
National Tsing Hua Univ., ROC China

ThMD.2 Speaker Models Designed from Complete Data Sets: A New Approach to Text-Independent Speaker Verification 2283
 Dominik R. Dersch, *Robin W. King
Univ. of Sydney, Australia
 *Univ. of South Australia, Australia

ThMD.3 A Double Gaussian Mixture Modeling Approach To Speaker Recognition 2287
 Vergin Rivarol, Douglas O'Shaughnessy
INRS Telecommunications, Canada

ThMD.4 An Acoustic Subword Unit Approach to Non-Linguistic Speech Feature Identification 2291
 Mohamed Afify, Yifan Gong, Jean-Paul Haton
CRIN-CNRS, France

ThMD.5 N-Best GMM's for Speaker Identification 2295
 Chakib Tadj, *Pierre Dumouchel, †Yu Fang
Ecole de Technologie Supérieure, Canada
 *Centre de Recherche Informatique, Canada
 †Institut Universitaire de Technologie, Canada

ThMD.6 Model Dependent Spectral Representations for Speaker Recognition 2299
 Guillaume Gravier, *Chafic Mokbel, Gerard Chollet
ENST/SIG, France
 *CNET-DIH/RCP, France

ThMD.7 Equalizing Sub-Band Error Rates in Speaker Recognition 2303
 Roland Auckenthaler, *John S. Mason
Technical Univ. Graz, Austria
 *Univ. of Wales Swansea, UK

ThMD.8 Automatic Gender Identification Under Adverse Conditions 2307
 Stefan Slomka, Sridha Sridharan
Queensland Univ. of Technology, Australia

ThMD.9 Acoustic Features and Perceptive Processes in the Identification of Familiar Voices 2311
 Yizhar Lavner, Isak Gath, Judith Rosenhouse
Israel Institute of Technology, Israel

ThMD.10 On the Use Acoustic Segmentation in Speaker Identification 2315
 Leandro Rodriguez-Linares, Carmen Garcia-Mateo
Univ. of Vigo, Spain

ThMD.11 Speaker Recognition by Humans and Machines
 Herman J.M. Steeneken, David A. Van Leeuwen
TNO-HFRI, The Netherlands 2319

ThMD.12 Foreign Speaker Accent Classification Using Phoneme-Dependent Accent Discrimination Models and Comparisons with Human Perception Benchmarks
 Karsten Kumpf, *Robin W. King
Univ. of Sydney, Australia
 *Univ. of South Australia, Australia 2323

ThMD.13 A Comparison of Human and Machine In Speaker Recognition 2327
 Li Liu, Jialong He, Günther Palm
Univ. of Ulm, Germany

ThMD.14 Evaluation of Second Language Learners' Pronunciation Using Hidden Markov Models 2331
 Simo M.A. Goddijn, *Gaus de Krom
Forensic Science Laboratory, Rijswijk
 *Univ. of Utrecht, The Netherlands

ThMD.15 Delta Vector Taylor Series Environment Compensation for Speaker Recognition 2335
 Brian Eberman, Pedro J. Moreno
Digital Equipment Corp., USA

ThMD.16 Wavelet-Like Regression Features in the Cepstral Domain for Speaker Recognition 2339
 Jonathan Hume
Univ. of Wales Swansea, UK

ThMD.17 Minimum Classification Error Linear Regression (MCELR) for Speaker Adaptation Using HMM with Trend Functions 2343
 Rathinavelu Chengalvarayan
Bell Labs-Lucent Technologies, USA

ThMD.18 A Continuous HMM Text Independent Speaker Recognition System Based on Vowel Spitting 2347
 Nikos Fakotakis, *Anastasios Tsopanoglou, Kallirroi Georgila
Univ. of Patras, Greece
 *KNOWLEDGE SA, Greece

ThMD.19 On the Independence of Digits in Connected Digit Strings 2351
 Johan W. Koolwaaij, Lou Boves
Nijmegen University, The Netherlands

ThMD.20 A New Procedure for Classifying Speakers in Speaker Verification Systems 2355
 Johan W. Koolwaaij, Lou Boves
Nijmegen University, The Netherlands

ThMD.21 Sound Channel Video Indexing 2359
 Claude Montacie, Marie-Jose Caraty
Univ. Pierre et Marie Curie - CNRS, France

ThMD.22 CDHMM Speaker Recognition By Means of Frequency Filtering of Filter-Bank Energies 2363
 Javier Hernando, Climent Nadeu
Universitat Politècnica de Catalunya, Spain

SESSION: Th3A
Style and Accent Recognition
Chair: Gerard Chollet, ENST/SIG, Switzerland

Th3A.1 Using Accent-Specific Pronunciation Modelling for Improved Large Vocabulary Continuous Speech Recognition 2367
J. J. Humphries, P. C. Woodland
Cambridge Univ., UK

Th3A.2 Automatic Speech Recognition for Children
Alexandros Potamianos, Shrikanth Narayanan, Sungbok Lee
AT&T Labs-Research, USA 2371

Th3A.3 Recognition of Non-Native Accents 2375
Carlos Teixeira, Isabel Trancoso, Antonio Serralheiro
INESC, Portugal

Th3A.4 Speaking Mode Dependent Pronunciation Modeling in Large Vocabulary Conversational Speech Recognition 2379
Michael Finke, Alex Waibel
Carnegie Mellon Univ., USA

Th3A.5 A Prosody-Only Decision-Tree Model for Disfluency Detection 2383
Elizabeth Shriberg, *Rebecca Bates, Andreas Stolcke
SRI International, USA
**Boston Univ., USA*

Th3A.6 A Novel Training Approach for Improving Speech Recognition Under Adverse Stressful Conditions 2387
Sahar E. Bou-Ghazale, John H.L. Hansen
Duke Univ., USA

SESSION: Th3B
Phonetics
Chair: Joaquim Llosterri, Univ. of Barcelona, Spain

Th3B.1 From Phone Identification to Phone Clustering Using Mutual Information 2391
Peter O'Boyle, Ji Ming, Marie Owens, F. Jack Smith
Queen's Univ. of Belfast, N. Ireland

Th3B.2 Phonetic Code Emergence in a Society of Speech Robots: Explaining Vowel Systems and the MUAUF Principle 2395
Ahmed-Reda Berrah, Rafael Laboissiere
Institut de la Communication Parleé, France

Th3B.3 Effects of Voicing on /t,d/ Tongue/Palate Contact in English and Norwegian 2399
Inger Moen, Hanne Gram Simonsen
Univ. of Oslo, Norway

Th3B.4 Fieldwork Techniques for Relating Formant Frequency, Amplitude and Bandwidth 2403
Peter Ladefoged, *Gunnar Fant
UCLA, USA
**KTH, Sweden*

Th3B.5 Word Juncture Modelling Based on the TIMIT Database 2407
Xue Wang, Louis C.W. Pols
Univ. of Amsterdam, The Netherlands

Th3B.6 The Phonology and Phonetics of Second Language Intonation: The Case of "Japanese English" 2411
Motoko Ueyama
UCLA, USA

SESSION: Th3C (SPECIAL SESSION)
Towards Robust ASR for Car and Telephone Applications
Chair: Jean-Claud Junqua, Panasonic Technologies Inc., California, USA

Th3C.1 Methods for Microphone Equalization in Speech Recognition 2415
L. Fissore, Giorgio Micca, C. Vair
Centro Studi e Laboratori Telecomunicazioni (CSELT), Italy

Th3C.2 Room Acoustics and Reverberation: Impact on Hands-Free Recognition 2419
Satoshi Nakamura, Kiyohiro Shikano
Nara Institute of Science and Technology, Japan

Th3C.3 Echo and Noise Reduction for Hands-Free Terminals -State of the Art- 2423
Gerard Faucon, Regine Le Rouquin-Jeannes
Univ. de Rennes I, France

Th3C.4 Robust Speech Recognition for Wireless Networks and Mobile Telephony 2427
Reinhold Haeb-Umbach
Philips GmbH, Germany

Th3C.5 Robust ASR for the Cellular Environment
Jay Naik
Nynex, USA
(Not arrived in time to be included in the Proceedings)

Th3C.6 Speech Recognition in the Car From Phone Dialing to Car Navigation 2431
Dirk Van Compernelle
Lernout & Hauspie Speech Products NV, Belgium

SESSION: Th3D
Language Specific Systems
Chair: Christel Sorin, CNET, Lannion, France

Th3D.1 A Keyvowel Approach to the Synthesis of Regional Accents of English 2435
Briony Williams, Stephen Isard
Univ. of Edinburgh, UK

Th3D.2 Experimental Implementation of Pitch-Synchronous Synthesis Methods for the ROMVOX Text-to-Speech System 2439
Attila Ferencz, Radu Arsinte, *Istvan Nagy, Teodora Ratiu, Maria Ferencz, †Gavril Todorean, †Diana Zaiu, Tunde-Csilla Kovacs, Lajos Simon
Software ITC SA, Romania
**Music Academy Gh.Dima, Romania*
†Technical Univ. of Cluj-Napoca, Romania

Th3D.3 The Bell Labs German Text-to-Speech System: An Overview 2443
Bernd Mobius, Richard Sproat, Jan P.H. van Santen, Joseph P. Olive
Bell Labs-Lucent Technologies, USA

Th3D.4 The Generation of Regional Pronunciations of English for Speech Synthesis 2447
Susan Fitt
Univ. of Edinburgh, UK

Th3D.5 Bell Laboratories Russian Text-To-Speech System
Elena Pavlova, Yuri Pavlov, Richard Sproat, Chilin Shih, Jan P.H. van Santen
Bell Labs-Lucent Technologies, USA 2451

Th3D.6 A Bilingual Text-To-Speech System in Spanish and Catalan 2455
Antonio Bonafonte, Ignasi Esquerra, Albert Febrer, Francesc Vallverdu
Universitat Politècnica de Catalunya, Spain

SESSION: Th4A
Pronunciation Models
Chair: Jean-Paul Haton, CRIN/CNRS-INRIA, France

Th4A.1 Automatic Rule-based Generation of Word Pronunciation Networks 2459
Nick Cremelie, Jean-Pierre Martens
Univ. of Gent, Belgium

Th4A.2 Creating User Defined New Vocabularies for Voice Dialing 2463
Jose Maria Elvira, Juan Carlos Torrecilla, Javier Caminero
Telefonica I+D, Spain

Th4A.3 Automatic Generation of Context-Dependent Pronunciations 2467
Mosur Ravishankar, Maxine Eskenazi
Carnegie Mellon Univ., USA

Th4A.4 Automatic Generation of a Pronunciation Dictionary Based on a Pronunciation Network 2471
Toshiaki Fukada, Yoshinori Sagisaka
ATR ITL, Japan

Th4A.5 What is Wrong with the Lexicon-An Attempt to Model Pronunciations Probabilistically 2475
Uwe Jost, Henrik Heine, Gunnar Evermann
Hamburg Univ., Germany

Th4A.6 Lexical Tuning Based on Triphone Confidence Estimation 2479
Kevin L. Markey, *Wayne Ward
Berdy Medical Systems, USA
*Carnegie Mellon Univ., USA

SESSION: Th4B
Auditory Modelling and Psychoacoustics
Chair: William Ainsworth, Keele Univ., UK

Th4B.1 Improving of Amplitude Modulation Maps for FO-Dependent Segregation of Harmonic Sounds 2483
Frederic Berthommier, *Georg Meyer
ICP, INPG, France
*Univ. of Keele, UK

Th4B.2 Psychophysical Evaluation of PSOLA: Natural Versus Synthetic Speech 2487
Reinier Kortekaas, Armin Kohlrausch
IPO, The Netherlands

Th4B.3 Perception of Noised Words by Normal Children and Children with Speech and Language Impairments
Valentina V. Lyblinskaja, *Inna V. Korolova, A.N. Kornev, Elena V. Lagounova 2491
Pavlov Institute of Physiology, Russia
*Institute of Ear, Throat, Nose and Speech Pathology, Russia

Th4B.4 Modeling the Perception of Simultaneous Semi-Vowels 2495
Georg F. Meyer, William A. Ainsworth
Keele Univ., UK

Th4B.5 Properties of Auditory Model Representations
Fernando Santos Perdigao, Luis V. Sá 2499
Universidade de Coimbra, Portugal

Th4B.6 Impact of "Ascending Sequence" AI (Auditory Primary Cortex) Cells on Stop Consonant Perception
Marta Eduardo Sa, de Sa Luis Vieira 2503
Universidade de Coimbra, Portugal

SESSION: Th4C
Voice Conversion and Data Driven F0-Models
Chair: Yoshinori Sagisaka, ATR Interpret. Telecom. Res. Lab., Japan

Th4C.1 Application-Dependent Prosodic Models for Text-To-Speech Synthesis and Automatic Design of Learning Database Corpus Using Genetic Algorithm 2507
Olivier Boeffard, Emerard F.
France Telecom-CNET, France

Th4C.2 Combinatorial Issues in Text-To-Speech Synthesis
Jan P.H. van Santen 2511
Bell Labs-Lucent Technologies, USA

Th4C.3 Automatic Corpus-Based Training of Rules for Prosodic Generation in Text-To-Speech 2515
Eduardo Lopez-Gonzalo, Jose M. Rodriguez-Garcia, Luis Hernandez-Gomez, Juan M. Villar
ETSIT-UPM, Spain

Th4C.4 Hidden Markov Model Based Voice Conversion Using Dynamic Characteristics of Speaker 2519
Eun-Kyoung Kim, Sangho Lee, Yung-Hwan Oh
KAIST, Korea

Th4C.5 Speaker Interpolation in HMM-Based Speech Synthesis System 2523
Takayoshi Yoshimura, *Takashi Masuko, Keiichi Tokuda, *Takao Kobayashi, Tadashi Kitamura
Nagoya Institute of Technology, Japan
*Tokyo Institute of Technology, Japan

Th4C.6 Designing a Speaker Adaptable Formant-Based Text-To-Speech System 2527
Vassilios Darsinos, Dimitrios Galanis, George Kokkinakis
Univ. of Patras, Greece

SESSION: Th4D

Vocal Tract Analysis

Chair: Andreas Paoloni, Fondazione Ugo Bordoni, Italy

Th4D.1 On Using Fractal Features of Speech Sounds in Automatic Speech Recognition.....2531

Petros Maragos, *Alexandros Potamianos

ILSP & Georgia Tech, Greece & USA

**AT&T Labs-Research, USA*

Th4D.2 Dynamic Constraint Weighting in the Context of Articulatory Parameter Estimation.....2535

Hywel B. Richards, John S. Bridle, Melvyn J. Hunt, *John S. Mason

Dragon Systems, UK

**Univ. of Wales Swansea, UK*

Th4D.3 Estimation of Vocal Tract Front Cavity Resonance in Unvoiced Fricative Speech.....2539

Minkyu Lee, Donald G. Childers

Univ. of Florida, USA

Th4D.4 A Software Tool to Study Portuguese Vowels

António Teixeira, Francisco Vaz, *Jose Carlos Principe

INESC, Portugal

**Univ. of Florida, USA.....2543*

Th4D.5 Post-Synchronization Via Formant-to-Area Mapping of Asynchronously Recorded Speech Signals and Area Functions.....2547

Jean Schoentgen, Sorin Ciocea

Univ. Libre de Bruxelles, Belgium

Th4D.6 Geometrically and Acoustically Optimized Codebook for Unique Mapping from Formants to Vocal-Tract Shape.....2551

Zhenli L. Yu, P.C. Ching

The Chinese Univ. of Hong Kong, Hong Kong

SESSION: ThAA

Noise Mitigation, Speech Enhancement II

Chair: Bayya Yegnanarayana, IIT MADRAS, India

ThAA.1 Noisy Speech Enhancement by Fusion of Auditory and Visual Information: A Study of Vowel Transitions.....2555

Laurent Girin, Gang Feng, Jean-Luc Schwartz

Univ. of Stendhal, France

ThAA.2 Spectral Subtraction Using a Non-Critically Decimated Discrete Wavelet Transform.....2559

Andreas Engelberg, Thomas Gulzow

Univ. of Kiel, Germany

ThAA.3 Bayesian Affine Transformation of HMM Parameters for Instantaneous and Supervised Adaptation in Telephone Speech Recognition.....2563

Jen-Tzung Chien, Hsiao-Chuan Wang, *Chin-Hui Lee

National Tsing Hua Univ., ROCina

**Bell Labs, USA*

ThAA.4 Integrated Bias Removal Techniques for Robust Speech Recognition.....2567

Craig Lawrence, Mazin Rahim

Univ. of Maryland, USA

ThAA.5 Acoustic Front Ends for Speaker-Independent Digit Recognition in Car Environments.....2571

Detlev Langmann, Alexander Fischer, Friedhelm Wuppermann,

Reinhold Haeb-Umbach, Thomas Eisele

Philips GmbH, Germany

ThAA.6 Signal Bias Removal Using the Multi-Path Stochastic Equalization Technique.....2575

Lionel Delphin-Poulat, Chafic Mokbel

France Telecom, France

ThAA.7 Subband Echo Cancellation in Automatic Speech Dialog Systems.....2579

Andrej Miksic, Bogomir Horvat

Univ. of Maribor, Slovenia

ThAA.8 Speech Enhancement Via Energy Separation

Hesham Tolba, Douglas O'Shaughnessy

Univ. du Quebec, Canada.....2583

ThAA.9 A Method of Signal Extraction from Noisy Signal

Masashi Unoki, Masato Akagi

Japan Advanced Institute of Science and Technology, Japan.....2587

ThAA.10 Multi-Channel Noise Reduction Using Wavelet Filter Bank.....2591

Jiri Sika, Vratislav Davidek

Czech Technical Univ., Czech Republic

ThAA.11 Speech Signal Detection in Noisy Environment Using a Local Entropic Criterion.....2595

Imad Abdallah, Silvio Montresor, Marc Baudry

Laboratoire d'Informatique de l'Univ. du Maine, France

ThAA.12 A New Algorithm for Robust Speech Recognition: The Delta Vector Taylor Series Approach

Pedro J. Moreno, Brian Eberman

Digital Equipment Corp., USA.....2599

ThAA.13 Robust Enhancement of Reverberant Speech Using Iterative Noise Removal.....2603

David Cole, Miles Moody, Sridha Sridharan

Queensland Univ. of Technology, Australia

ThAA.14 A Network Speech Echo Canceller with Comfort Noise.....2607

David J. Jones, Scott D. Watson, Kenneth G. Evans, Barry

M.G. Cheetham, *Robert A. Reeves

Univ. of Liverpool, UK

**BT Laboratories, UK*

ThAA.15 A Metric for Selecting Sub-Band Processing in Adaptive Speech Enhancement Systems.....2611

Amir Hussain, Douglas R. Campbell, Thomas J. Moir

Univ. of Paisley, UK

ThAA.16 Estimation of LPC Cepstrum Vector of Speech Contaminated by Additive Noise and its Application to Speech Enhancement.....2615

Hidefumi Kobatake, Hideta Suzuki

Tokyo Univ. of Agriculture & Technology, Japan

ThAA.17 Multi-Band and Adaptation Approaches to Robust Speech Recognition.....2619

Sangita Tibrewala, Hynek Hermansky

Oregon Graduate Institute of Science and Technology, USA

ThAA.18 Non-Quadratic Criterion Algorithms for Speech Enhancement 2623
Enrique Masgrau, Eduardo Lleida, Luis Vicente
Universidad de Zaragoza, Spain

SESSION: ThAB

F0 and Duration Modelling, Spoken language processing

Chair: Richard Schwartz, BBN Systems and Techs, USA

ThAB.1 Modeling Segmental Duration with Multivariate Adaptive Regression Splines 2627
Marcel Riedi
ETH Zentrum TIK, Switzerland

ThAB.2 High Quality Speech Synthesis for Phonetic Speech Segmentation 2631
Fabrice Malfrere, Thierry Dutoit
Circuits Theory and Signal Processing Lab, Belgium

ThAB.3 Factors Affecting Perceived Quality and Intelligibility in the CHATR Concatenative Speech Synthesiser 2635
Nick Campbell, Itoh Yoshiharu, Wen Ding, Norio Higuchi
ATR Interpreting Telecommunications Res. Labs., Japan

ThAB.4 Reduced Lexicon Trees for Decoding in a MMI-Connectionist/HMM Speech Recognition System 2639
Christoph Neukirchen, Daniel Willett, Gerhard Rigoll
Gerhard-Mercator-Univ. Duisburg, Germany

ThAB.5 A Stochastic Model of Intonation for French Text-to-Speech Synthesis 2643
Jean Veronis, Philippe Di Cristo, Fabienne Courtois, Benoit Lagrue
Univ. de Provence & CNRS, France

ThAB.6 Phonetic Rules for a Phonetic-to-Speech System
Angelien A. Sanderma, *René Collier 2647
KPN Research, The Netherlands
**Institute for Perception Research, The Netherlands*

ThAB.7 Multi-Lingual Duration Modeling 2651
Jan P.H. van Santen, Chilin Shih, Bernd Mobius, Evelyne Tzoukermann, Michael Tanenblatt
Bell Labs-Lucent Technologies, USA

ThAB.8 A Model of Segment (and Pause) Duration Generation for Brazilian Portuguese Text-to-Speech Synthesis 2655
Plinio A. Barbosa
State Univ. of Campinas, Brazil

ThAB.9 Parsing Strategy for Spoken Language Interfaces with a Lexicalized Tree Grammar 2659
Ariane Halber, David Roussel
Thomson-CSF, France

ThAB.10 What's in a Word Graph -- Evaluation and Enhancement of Word Lattices 2663
Jan W. Amtrup, Henrik Heine, Uwe Jost
Hamburg Univ., Germany

ThAB.11 Accelerated DP Based Search for Statistical Translation 2667
Christoph Tillmann, Stefan Vogel, Hermann Ney, A. Zubiaga, H. Sawaf
RWTH Aachen, Germany

ThAB.12 Use of Pitch Pattern Improvement in the CHATR Speech Synthesis System 2671
Ken Fujisawa, Toshio Hirai, Norio Higuchi
ATR ITL, Japan

ThAB.13 Generating Segment Durations in a Text-to-Speech System: A Hybrid Rule-Based/Neural Network Approach 2675
Gerald Corrigan, Noel Massey, Orhan Karaali
Motorola, USA

ThAB.14 On the Global F0 Shape Model Using a Transition Network for Japanese Text-To-Speech Systems
Yasushi Ishikawa, Takashi Ebihara 2679
Mitsubishi Electric Corporation, Japan

ThAB.15 An Alternative and Flexible Approach in Robust Information Retrieval Systems 2683
Jose Colas, Juan M. Montero, Javier Ferreiros, Jose M. Pardo
Universidad Politecnica de Madrid, Spain

ThAB.16 A Probabilistic Approach to Analogical Speech Translation 2687
Keiko Horiguchi, Alexander Franz
Sony, Japan

ThAB.17 Dynamic Lexicon for a Very Large Vocabulary Vocal Dictation 2691
Marie-José Caraty, Claude Montacie, Fabrice Lefèvre
Univ. Pierre et Marie Curie - CNRS, France

SESSION: ThAC

Language Modelling

Chair: Ronald Rosenfeld, Carnegie Mellon Univ., USA

ThAC.1 Construction of Language Models Using the Morphic Generator Grammatical Inference (MGGI) Methodology 2695
Encarna Segarra, Luis Hurtado
Universidad Politecnica de Valencia, Spain

ThAC.2 An Integrated Language Modeling with N-Gram Model and WA Model for Speech Recognition 2699
Shuwu Zhang, Taiyi Huang
Chinese Academy of Sciences, China

ThAC.3 Statistical Analysis of Dialogue Structure 2703
Ye-Yi Wang, Alex Waibel
Carnegie Mellon Univ., USA

ThAC.4 Statistical Language Modeling Using the CMU-Cambridge Toolkit 2707
Philip Clarkson, *Ronald Rosenfeld
Cambridge Univ., UK
**Carnegie Mellon Univ., USA*

ThAC.5 Text Normalization and Speech Recognition in French 2711
Gilles Adda, Martine Adda-Decker, Jean-Luc Gauvain, Lori Lamel
LIMSI, France

ThAC.6 A Novel Tree-Based Clustering Algorithm for Statistical Language Modeling.....2715
 Geraldine Damnati, Jacques Simonin
France Telecom, France

ThAC.7 Variable-Length Language Modeling Integrating Global Constraints.....2719
 Shoichi Matsunaga, Shigeki Sagayama
NTT, Japan

ThAC.8 An Hybrid Language Model for Continuous Dictation Prototype.....2723
 Kamel Smaili, Imed Zitouni, Francois Charpillet, Jean-Paul Haton
CRIN-CNRS & INRIA, Lorraine, France

ThAC.9 Dealing with Pronunciation Variants at the Language Model Level for the Continuous Automatic Speech Recognition of French.....2727
 Laure Pousse, Guy Perennou
IRIT - Equipe IHMPT, France

ThAC.10 Rational Interpolation of Maximum Likelihood Predictors in Stochastic Language Modeling 2731
 Ernst Günter Schukat-Talamazzini, *Florian Gallwitz, *Stefan Harbeck, *Volker Warnke
Univ. of Jena, Germany
**Univ. of Erlangen, Germany*

ThAC.11 N-Gram Language Model Adaptation Using Small Corpus for Spoken Dialog Recognition 2735
 Akinori Ito, Hideyuki Saitoh, Masaharu Katoh, Masaki Kohda
Yamagata Univ., Japan

ThAC.12 Variable N-Gram Language Modeling and Extensions for Conversational Speech 2739
 Man-Hung Siu, *Mari Ostendorf
BBN Inc, USA
**Boston Univ., USA*

ThAC.13 Fuzzy Class Rescoring: A Part-of-Speech Language Model.....2743
 Petra Geutner
Univ. of Karlsruhe, Germany

ThAC.14 Speech Understanding Based on Integrating Concepts By Conceptual Dependency 2747
 Akito Nagai, Yasushi Ishikawa
Mitsubishi Electric Corporation, Japan

ThAC.15 Dynamic Language Models for Interactive Speech Applications.....2751
 Fabio Brugnara, Marcello Federico
Istituto per la Ricerca Scientifica e Tecnologica (IRST), Italy

ThAC.16 Large-Scale Lexical Semantics for Speech Recognition Support 2755
 George Demetriou, Eric Atwell, Clive Souter
Univ. of Leeds, UK

ThAC.17 Integration of Grammar and Statistical Language Constraints for Partial Word-Sequence Recognition....2759
 Hajime Tsukada, Hirofumi Yamamoto, Yoshinori Sagisaka
ATR Interpreting Telecommunications Res. Labs., Japan

ThAC.18 Using Intonation to Constrain Language Models in Speech Recognition 2763
 Paul Taylor, Simon King, Stephen Isard, Helen Wright, Jaqueline Kowtko
Univ. of Edinburgh, UK

ThAC.19 Incorporating POS Tagging into Language Modeling 2767
 Peter A. Heeman, *James F. Allen
France Telecom, France
**Univ. of Rochester, USA*

ThAC.20 Confidence Metrics Based on N-Gram Language Model Backoff Behaviors.....2771
 Carl Uhrík, *Wayne Ward
Berdy Medical Systems, USA
**Carnegie Mellon Univ., USA*

ThAC.21 Structure and Performance of a Dependency Language Model.....2775
 Ciprian Chelba, *David Engle, Frederick Jelinek, †Victor M. Jimenez, Sanjeev Khudanpur, Lidia Mangu, ‡Harry Printz, ***Eric Ristad, ††Ronald Rosenfeld, ‡‡Andreas Stolcke, ‡‡Dekai Wu
John Hopkins Univ., USA
**Dept. of Defense Fort Meade, MD, USA*
†Universitat Politècnica de Valencia, Spain
‡IBM, USA
***Princeton Univ., USA*
††Carnegie Mellon Pittsburgh, PA, USA
‡‡SRI International, USA
‡‡Hong Kong Tech University, Hong Kong

ThAC.22 Modeling Linguistic Segment and Turn Boundaries for N-Best Rescoring of Spontaneous Speech
 Andreas Stolcke
SRI International, USA.....2779

ThAC.23 Hybrid Language Models: Is Simpler better?
 Peter E. Kenne, Mary O'Kane
Univ. of Adelaide, Australia.....2783

ThAC.24 Internal and External Tagsets in Part-of-Speech Tagging.....2787
 Thorsten Brants
Univ. of the Saarland, Germany

SESSION: ThAD
Auditory Modelling and Psychoacoustics, Neural Networks for Speech Processing and Recognition
Chair: Phil D. Green, Univ. of Sheffield, UK

ThAD.1 A Probabilistic Model of Double-Vowel Segregation 2791
 Laurent Varin, Frederic Berthommier
ICP, INPG, France

ThAD.2 Stimulus Signal Estimation From Auditory-Neural Transduction Inverse Processing 2795
 Houshang Habibzadeh Vaneghi, Shigeyoshi Kitazawa
Shizuoka Univ., Japan

ThAD.3 FDVQ Based Keyword Spotter Which Incorporates A Semi-Supervised Learning for Primary Processing..... 2799
 Chakib Tadj, Pierre Dumouchel, *Franck Poirier
Ecole de Technologie Supérieure, Canada
**Institut Universitaire Professionnalise, France*

ThAD.4 The Initial Time Span of Auditory Processing Used for Speaker Attribution of the Speech Signal..... 2803
 Valentina V. Lublinskaja, *Christian Sappok
Pavlov Institute of Physiology, Russia
**Ruhr Universität, Germany*

ThAD.5 Sparse Connection and Pruning in Large Dynamic Artificial Neural Networks 2807
 Nikko Strom
KTH, Sweden

ThAD.6 A Modular Initialization Scheme for Better Speech Recognition Performance Using Hybrid Systems of MLPs/HMMs 2811
 Roxana Teodorescu, Dirk Van Compernelle, Ioannis Dologlou
K.U Leuven-ESAT, Belgium

ThAD.7 Lateralization for Auditory Perception of Foreign Words 2815
 Tatiana Chernigovskaya
Russian Academy of Sciences, Russia

ThAD.8 The Structural Weighted Sets Method for Continuous Speech and Text Recognition 2819
 Yuri Kosarev, Pavel Jarov, Alexander Osipov
Russian Academy of Sciences, Russia

ThAD.9 Lateral Inhibitory Networks for Auditory Processing..... 2823
 Christian J. Sumner, Duncan F. Gillies
Imperial College, UK

ThAD.10 Missing Fundamentals: A Problem of Auditory or Mental Processing?..... 2827
 Henning Reetz
Univ. of Konstanz, Germany

ThAD.11 Predictive Neural Networks Applied to Phoneme Recognition..... 2831
 Felix Freitag, Enric Monte, Josep M. Salavedra
Polytechnic University of Catalunya, Spain

ThAD.12 Empirical Comparison of Two Multilayer Perceptron-Based Keyword Speech Recognition Algorithms 2835
 Suhardi, *Klaus Fellbaum
Technical Univ. of Berlin, Germany
**Brandenburg Technical Univ. of Cottbus, Germany*

ThAD.13 Segment Boundary Estimation Using Recurrent Neural Networks 2839
 Toshiaki Fukada, *Sophie Aveline, Mike Schuster, Yoshinori Sagisaka
ATR Interpreting Telecommunications Res. Labs.,
**ENST, France*

ThAD.14 Incorporation of HMM Output Constraints in Hybrid NN/HMM Systems During Training 2843
 Mike Schuster
ATR ITL, Japan

ThAD.15 Principles of the Hearing Periphery Functioning in New Methods of Pitch Detection and Speech Enhancement..... 2847
 Ludmila Babkina, *Sergey Koval, Alexander Molchanov
Research Institute of Ear, Nose, Throat and Speech Disorders, Russia
**Speech Technology Centre, Russia*

ThAD.16 The Locus of the Syllable Effect: Prelexical or Lexical?..... 2851
 Christine Meunier, *Alain Content, Uli H. Frauenfelder, †Ruth Kearns
Univ. of Geneva, Switzerland
**Univ. Libre de Bruxelles, Belgium*
†Medical Research Council, UK

ThAD.17 On Not Remembering Disfluencies 2855
 Ellen Gurman Bard, Robin J. Lickley
Univ. of Edinburgh, UK

ThAD.18 Using an Auditory Model and Leaky Autocorrelators to Tune In to Speech 2859
 Tjeerd Andringa
Univ. of Groningen, The Netherlands

A NEW METRIC FOR SELECTING SUB-BAND PROCESSING IN ADAPTIVE SPEECH ENHANCEMENT SYSTEMS

Amir Hussain, Douglas R. Campbell and Thomas J. Moir

Department of Electronic Engineering and Physics,

University of Paisley, High St., Paisley PA1 2BE, Scotland U.K.

Corresponding author's email: huss_ee0@paisley.ac.uk

ABSTRACT

A multi-microphone adaptive speech enhancement system employing diverse sub-band processing is presented. A new robust metric is developed, which is capable of real-time implementation, in order to automatically select the best form of processing within each sub-band. It is based on an adaptively estimated inter-channel Magnitude Squared Coherence (MSC) relationship, which is used to detect the level of correlation between in-band signals from multiple sensors during noise-alone periods in intermittent speech. This paper reports recent results of comparative experiments with simulated anechoic data extended to include simulated reverberant data. The results demonstrate that the method is capable of significantly outperforming conventional noise cancellation schemes.

1. INTRODUCTION

Background noise contamination of speech signals reduces the Signal to Noise Ratio (SNR) of hands-free telephones, portable phones, and security screens. Speech recognition systems in particular, are known to experience problems due to levels of background noise that are in fact, considered quite acceptable to human listeners [7]. In addition to the noise level, the presence of multiple noise sources, reverberant environments, moving noise sources and statistically non-stationary noise sources considerably complicates the situation.

Classical speech enhancement methods based on full-band multi-microphone noise cancellation implementations which attempt to model acoustic path transfer functions can produce excellent results in anechoic environments with localized sound radiators [6], however performance deteriorates in reverberant environments. Multi-band processing has been found to be important in combating reverberation effects [10] [11]. Adaption is necessary to compensate for changing noise fields [2] [10] due to for example, non-Gaussian sources, source/sensor motion, or time-varying acoustic paths. Multi-sensor methods are necessary to compensate for reverberation [13], and speech/noise spectral overlap [2].

Studies on noise in automobiles by various researchers [9] [10] [11] have found that the long-term noise correlation between two microphone locations was high for frequencies below 500Hz, and decreased gradually with virtually no correlation above 2kHz. Wallace and Goubran [12] applied the classical full-band linear FIR based correlated noise canceller, adapted using the Least Mean Square (LMS) algorithm, to recorded automobile noise and obtained significant noise

reduction in the low frequency range. but at high frequencies, where both the correlation and noise energy were low, the noise increased. In some situations, as reflected in the measurements from office and automobile environments recently reported in [2], significant short term correlation may also occur at high frequencies depending upon the relative locations of the microphones and the nature of the noise sources. The above evidence implies that processing appropriate in one sub-band, may not be so in another. This therefore supports a general approach involving the use of diverse processing in frequency bands dependent on the correlation between the in-band signals from multiple sensors.

Van Compernelle et al [5] and Dabis et al [6] used closely spaced microphones in a full-band adaptive noise cancellation scheme involving the identification of a differential acoustic path transfer function during a noise only period in intermittent speech. A Multi-Microphone Sub-Band Adaptive (MMSBA) speech enhancement system has been described which extends this method by applying it within a set of sub-bands provided by a filterbank [2] [8] [11]. Even non-optimised MMSBA processing has shown the potential to yield more than 6dB SNR improvements over conventional full-band methods in real reverberant environments.

The MMSBA system has been further developed by employing diverse Sub-Band Processing (SBP) in order to allow inter-channel features within the sub-bands to influence the subsequent processing [4]. In order to realize this, a robust practical metric has been developed, which is capable of real-time implementation, based on the inter-channel Magnitude Squared Coherence (MSC) relationship in order to automatically select the best SBP option. The choice of SBP options include: (i) no processing; (ii) intermittent coherent noise canceller; and (iii) continuous incoherent noise canceller.

2. THE PROPOSED MMSBA SYSTEM INCORPORATING THE NEW METRIC

Two or more relatively closely spaced microphones may be used in an adaptive noise cancellation scheme [4] [5] [6] [11] to identify a differential acoustic path transfer function during a noise only period in intermittent speech. The extension of this work, termed the Multi-Microphone sub-band Adaptive (MMSBA) speech enhancement system, applies the method within a set of sub-bands provided by a filter bank as shown in Figure 1. The filter bank can be implemented using various orthogonal transforms or by a parallel filter bank approach. The sub-bands can be distributed either in a linear or a non-linear fashion.

It is assumed in this work that the speaker is close enough to the microphones so that room acoustic effects on the speech are insignificant, that the noise signal at the microphones may be represented as a point source modified by two different acoustic path transfer functions H_1 and H_2 , and that an effective voice activity detector (VAD) is available.

The sub-band processing (SBP) can be accomplished in a number of ways, for example:

1. No Processing: Examine the noise power in a sub-band and if below (or the SNR above) some arbitrary threshold, then set the processing transfer function to one, that is, the signal in that band need not be modified.

2. Intermittent coherent noise canceller: If the noise power is significant and the noise between the two channels is significantly correlated in a sub-band, then perform adaptive intermittent noise cancellation, wherein an adaptive filter may be determined which models the differential acoustic-path transfer function between the microphones during the noise alone period. This can then be used in a noise cancellation format during the speech plus noise period to process the noisy speech signal. This scheme illustrated in Figure 1 can be described mathematically as follows. Assuming N , S , P , R represent the z-transforms of the noise signal, speech signal, primary signal and reference signal, respectively. The primary and reference signals in each sub-band are thus

$$P = B(S + H_1 N) \quad ; \quad R = B(S + H_2 N)$$

The transformed error signal is thus,

$$E = B[(1 - H_3)S + (H_1 - H_3 H_2)N]$$

which is a frequency domain error, weighted by the band-limiting transfer function B , and H_3 represents the sub-band adaptive filter. The Mean Squared Error (MSE) function is,

$$J_E = (2\pi j)^{-1} \oint_{|z|=1} E E^* z^{-1} dz$$

The sub-band noise cancellation problem is thus, to find an H_3 such that within the sub-band defined by B , the variance of J_E is minimised. During a noise only period $S = 0$, defining the noise spectral density Φ_{nn} , then

$$J_E = (2\pi j)^{-1} \oint_{|z|=1} B(H_1 - H_3 H_2) \Phi_{nn} (H_1 - H_3 H_2)^* B^* z^{-1} dz$$

which is minimised in the least squares sense when

$$H_3 = (B H_1)(B H_2)^{-1}$$

That is, H_3 is a band-limited transfer function that minimises the noise power in E .

Now using H_3 as a fixed processing filter when speech and noise are present ideally gives:

$$E = B(1 - H_3)S$$

where the output E is a noise reduced, filtered version of the sub-band speech signal. This approach will fail if $H_1 = H_2$, however in practical situations it is often possible to arrange the sensor placement to avoid this acoustic path balancing.

3. Incoherent noise canceller: If the noise power is significant but not highly correlated between the two channels in a sub-band, then the incoherent noise cancellation approach of Ferrara and Widrow [15] or Zelinski [17] may be applied during the noisy speech period. Since in this case, the primary signal noise component $B H_1 N$ is uncorrelated with the reference signal noise component $B H_2 N$, the filtered reference is an estimate of the speech signal S .

In this paper, we examine the above three SBP options and implement the processing using the Least Mean Squares

(LMS) algorithm [16] to perform the adaption. A metric for selecting the appropriate SBP option is now derived next.

2.1 The adaptively estimated Magnitude Squared Coherence (MSC) Metric

The coherence function is a complex function of frequency defined as [14]:

$$\rho(f) = \frac{E[Y(f)Z^*(f)]}{\sqrt{E[|Y(f)|^2]E[|Z(f)|^2]}}$$

where $Y(f)$ and $Z(f)$ are the Fourier transforms of the signals $y(n)$ and $z(n)$ picked up at the same time by two microphones. The magnitude of the complex coherence function varies between 0 and 1 and gives, for each frequency, the percentage of signal energy coming from correlated sources; with a value close to 1 implying strong correlation between the two signals and a value close to 0 indicating uncorrelated (or weakly correlated) signals.

In the context of dereverberation, the Magnitude Squared Coherence (MSC) has been previously used by Allen et al [14] to correct the magnitude of the reflected signal. Recently, Bouquin and Faucon [1] have applied the MSC on noisy speech signals for noise reduction and successfully employed it as a VAD for the case of spatially uncorrelated noises. In this work, we propose the use of a modified MSC as a part of a system for selecting the best SBP option in a MMSBA speech enhancement system.

Since the speech signal uttered by the speaker is submitted to modifications due to its propagation, the observations received by the two microphones, mic 1 and mic 2, as shown in Figure 1, may be written as (assuming that the speech and noise signals are independent):

$$\text{At mic 1:} \quad x_1 = s_1 + n_1$$

$$\text{and, at mic 2:} \quad x_2 = s_2 + n_2$$

where s_i and n_i ($i=1,2$) represent the clean speech signal and the disturbing additive noise, respectively.

For each block l and frequency bin f_k , the coherence function is given by [13]:

$$\rho(f_k, l) = \frac{S_{x_1 x_2}(f_k, l)}{\sqrt{S_{x_1 x_1}(f_k, l) S_{x_2 x_2}(f_k, l)}}$$

where $S_{x_1 x_2}(f_k, l)$ is the cross-spectral density,

$S_{x_1 x_1}(f_k, l)$ and $S_{x_2 x_2}(f_k, l)$ are the auto-spectral densities; which can be estimated using a simple recursive calculation on a block by block basis [14]:

$$S_{x_i x_j}(f_k, l) = \beta S_{x_i x_j}(f_k, l-1) + (1-\beta) X_i(f_k, l) X_j^*(f_k, l),$$

$$i, j = 1, 2$$

where β is the forgetting factor. During the noise alone or speech free period, for each overlapped and Hann windowed block l we compute the Magnitude Squared Coherence (MSC) at each of the frequency bins f_k , $k=0, \dots, L/2$, (where L corresponds to the length of the short term FFT and is set to $L=256$) as:

$$\text{MSC}(f_k, l) = \frac{|S_{x_1 x_2}(f_k, l)|^2}{S_{x_1 x_1}(f_k, l) S_{x_2 x_2}(f_k, l)}$$

which is then averaged over all the previous overlapped blocks to give (at each frequency bin):

$$\overline{\text{MSC}}(f_k) = \frac{1}{I} \sum_{i=1}^I \text{MSC}(f_k, i)$$

The above *adaptively averaged* MSC criterion can thus be used as an effective means for determining the level of correlation between the disturbing noises at various frequencies bands (by averaging the above MSC over each respective sub-band), during the noise alone period in intermittent speech. The subsequent form of processing in each respective frequency band can therefore be selected on the basis of the inter-channel correlation.

On initial trials, a threshold value of 0.6 for the adaptive MSC has been found to be suitable for distinguishing between highly correlated and weakly correlated sub-band noise signals. For 50% block overlap, a forgetting factor of $\beta = 0.8$ has been found to be adequate, which compares well with the figures reported by Bouquin and Faucon [1].

3. SIMULATION RESULTS

3.1 Simulated Anechoic Data:

Two simulated independent white noise sequences with an averaged Mean Squared Coherence relationship depicted in Figure 2 were used to corrupt a real anechoic speech signal sampled at 10kHz. The initial SNR was fixed at -2dB, and a noise alone period was manually labelled comprising the first 1024 samples. Three noise cancellation systems were compared namely:

1. The conventional full-band noise canceller intermittently adapted using the LMS algorithm (FBLMS). The order of the adaptive filter was arbitrarily chosen to be 256.
2. A two sensor MMSBA system with four sub-bands employing intermittent adaptive LMS update in each of the 4 sub-bands (MBLMS). The sub-band filter order was set to 256/4.
3. And the proposed MMSBA system also with four sub-bands, employing diverse sub-band processing (MBDLMS) with the intermittent adaptive LMS update employed in each of the first two bands (in order to more effectively cancel the correlated noises in those sub-bands cf. Figure 2) and the continuous Ferrara-Widrow LMS update was employed in each of the last two bands (in order to more effectively cancel the uncorrelated noises in those sub-bands cf. Figure 2).

System	FBLMS	MBLMS	MBDLMS
SNR improv.	1.8dB	5.7dB	10.8dB

Table 1: Performance Comparison of various adaptive noise cancellers for synthetic data.

As can be seen from Table 1, for this test case the use of a MBLMS system gives better performance in cancelling the simulated interference compared to the conventional FBLMS noise canceller. However, the use of the proposed MBDLMS system employing diverse sub-band processing, dependent on the inter-channel coherence information, can be seen to produce a much greater performance increment over the MBLMS system. Informal listening tests also showed the MBDLMS processed speech to be both enhanced in SNR and of better perceived quality than that obtained by the other methods.

3.2 Simulated Reverberant Data:

For this case, the impulse responses between the noise source and two microphones were calculated by an image program

which simulates room acoustics using room dimensions, reflection coefficients and source/receiver locations as parameters. At a sampling rate of 10kHz, realistic room responses would be of length > 1024, but for testing purposes a length of 256 was selected. The room was approximated to a (6x5x4)m rectangular enclosure. The walls, floor and ceiling were given the same reflection coefficient value of 0.6 to generate a medium room reverberation level, and the noise-to-microphones (NTM) and microphone-to-microphone (MTM) spacing were set to 1m and 15cm respectively. Two microphone signals were then generated by convolving a white noise sequence with each of the simulated impulse responses to yield the primary and reference noise signals, which were then added to the previous anechoic speech signal. The initial SNR was fixed at -3dB, and a noise alone period was manually labelled comprising the first 1024 samples. For this particular test case with the selected MTM, NTM, and noise orientation angle values, the MSC relationship between the two microphone signals during the noise-alone period was found to exhibit a significant level of correlation across all the frequency bands. Thus for this case, the MBDLMS noise canceller is identical to the MBLMS canceller. The performance of the FBLMS and the 4-band MBDLMS intermittent noise cancellers is compared in Table 2.

System	FBLMS	MBDLMS
SNR improv.	1.9dB	10.1dB

Table 2: Performance Comparison of adaptive noise cancellers for "realistic" room data.

As can be seen from Table 2, the use of a MBDLMS system gives significantly better performance in cancelling the simulated room reverberated noise compared to the conventional FBLMS noise canceller. Informal listening tests again confirmed the performance improvements.

4. CONCLUSIONS

A multi-microphone sub-band adaptive (MMSBA) speech enhancement system employing diverse sub-band processing has been presented. An adaptively estimated inter-channel MSC measure has been proposed for selecting the best form of processing within each sub-band. Comparative results achieved in simulation experiments demonstrate that the method is capable of outperforming conventional noise cancellation schemes. Current experiments are extending the work reported here by using anechoic speech signals corrupted with reverberated noises recorded in real room and automobile environments.

5. ACKNOWLEDGEMENTS

This work is supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) Project Reference Number GR/K48907.

6. REFERENCES

- [1] R. Le Bouquin and G. Faucon, *Study of a voice activity detector and its influence on a noise reduction system*, Speech Communication, Vol.16, pp.245-254, 1995.

[2] E.Toner, *Speech Enhancement using Digital Signal Processing*, PhD thesis, Department of Electronic Engineering and Physics, University of Paisley, 1993.

[3] D.R.Campbell and E.Toner, *Speech enhancement with sub-band processing in an automobile environment*, 26th International Symposium on Automotive Technology and Automation, Dedicated Conference on Mechatronics, Aachen, Germany, 13th-17th September 1993.

[4] A.Hussain, D.R.Campbell and T.J.Moir, *A Multi-Microphone Sub-band Adaptive Speech Enhancement System employing diverse sub-band processing*, Proceedings ESCA-NATO Workshop on *Robust Speech Recognition for Unknown Communication Channels*, pp.123-126, Pont-a-Mousson, France, 17-18 April 1997

[5] D.Van Compernelle, W. Ma and M.M. Van Diest, *Speech recognition in noisy environment with the aid of microphone arrays*, European Conf. On Speech Technology, Vol.2, pp.657-660, Paris, France, 1989.

[6] H.S.Dabis, T.J.Moir and D.R.Campbell, *Speech enhancement by recursive estimation of differential transfer functions*, Proceedings of International Conference on Signal Processing (ICSP), pp.345-348, Beijing, China, 1990.

[7] H.S.Dabis and A.Wrench, *An evaluation of adaptive noise cancelling for speech recognition*, Eurospeech, pp.1301-1304, 1991.

[8] D.Darlington and D.R.Campbell, *Sub-band Adaptive Filtering Applied to Speech Enhancement*, ESCA Research Workshop, The Auditory Basis of Speech Perception, Keele University, 15-19 July, 1996.

[9] R.A.Goubran, R. Herbert and H.M.Hafez, *Acoustic noise suppression using regressive adaptive filtering*, 40th Vehicular Technology Conference, Florida U.S.A, pp.48-53, 1990.

[10] M.M.Goulding and J.S.Bird, *Speech enhancement for mobile telephony*, IEEE Trans. On Vehicular Technology, Vol.39 no.4, pp.316-326, 1990.

[11] E.Toner and D.R.Campbell, *Speech Enhancement using sub-band intermittent adaption*, International Journal of Speech Communication, Vol.12, pp.253-259, 1993.

[12] R.B.Wallace and R.A.Goubran, *Improved tracking adaptive noise canceller for non-stationary environments*, IEEE Trans. on Signal Processing, Vol.40, no.3, pp.700-703, 1992.

[13] H.Yamada, H.Wang and F.Itakura, *Recovering of broad band reverberant speech signal by sub-band MINT method*, ICASSP, Toronto, Canada, pp.969-972, 1991.

[14] J.B.Allen, D.A.Berkley and J.Blauert, *Multi-microphone signal processing technique to remove room reverberation from speech signals*, J. Acoustic Soc. Amer., Vol.62, No.4, pp.912-915, 1977.

[15] E. R. Ferrara, B. Widrow, *Multi-channel Adaptive Filtering for signal enhancement*, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.29, no.3, pp.766-770, 1981.

[16] B.Widrow, S.D.Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

[17] Z.R. Zelinski, *Noise reduction based on microphone array with LMS adaptive post filtering*, Electronic Letters, Vol.26, No.24, pp.2036-2037.

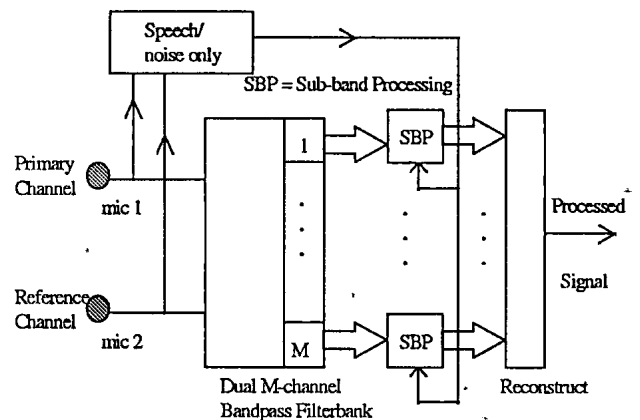


Figure 1: The MMSBA system

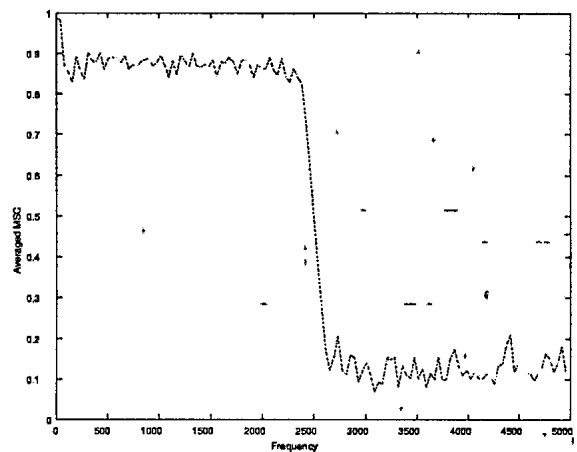


Figure 2: The adaptively estimated MSC between the disturbing synthetic noises

APPENDIX B

The Ohio State University Libraries Library Catalog

Catalog home

Save for later

Regular display

Another search

LEADER 0000cam a2200325Ia
000 44512506
000DC
20010623043656.0 100 0 eng
00030s1997
f035 .b54409676 d
040 ATT|cATT|dOCL|dOSU|dwaOLN|
049 dUtOrBLW OSUU
090 P95|b.E88 1997
090 P95|b.E88 1997
092 302.2242|bE89e, 1997 :|cRhodes,
11 Eurospeech 97|d(1997 :|cRhodes,
245 10 Eurospeech 97 :|b5th European Conference on
Speech Communication and Technology : Rhodes, Greece,
22-25 September, 1997
24 3 ESCA : 5th European Conference on Speech Communication
6 and Technology
26 1 Grenoble, France :|bESCA,|
40 c[1997]umes ;|c30 cm
03 text|btxt|2rdacontent
03 unmediated|bn|
33 2rdamedia volume|bnc|
50 2rdacarrier Cover
05 0 Speech|vCongresses.|0http://id.loc.gov/
0 audio|vCongresses|
65 0 B000011441vCongresses.|0http://id.loc.gov/
0 audio|vCongresses|
71 2 B000011441vCongresses|vSpeech Communication Association.|
0 0http:// id.loc.gov/authorities/names/n94005950
91 RDA ENRICHED|aMARS
01 MARS
01 xxS
0

LOCATION	CALL	YEA	STATU	NOT
Book	P95 .E88 1997	R	AVAILABLE	E
Depository	v.5		E	

Book Depository	P95 .E88 1997 v.4		AVAILABLE	
Book Depository	P95 .E88 1997 v.3		AVAILABLE	
Book Depository	P95 .E88 1997 v.2		AVAILABLE	
Book Depository	P95 .E88 1997 v.1		AVAILABLE	

[The Ohio State University Libraries](#)

© The Ohio State University - University Libraries

1858 Neil Avenue Mall | Columbus, Ohio 43210 | 614-292-OSUL (6785)

[Request an alternate format of this page](#) | [Accessibility](#) | [Privacy Policy](#) | [Contact Us](#)



[Copyright Information](#) | [Details and Exceptions](#)



APPENDIX C



CONFERENCE PROCEEDING

Eurospeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 242-25 September 1997 : proceedings.

Eurospeech 97 (5th : 1997 : Rhodes, Greece); European Speech Communication Association. 1997?

📖 Available at Math & Physics Library Stacks (P95.E88 1997) >

TOP

SHARE

FIND IN LIBRARY

DETAILS

VIRTUAL BROWSE

LINKS

Share



CITATION



EMAIL



PRINT



PERMALINK



REFWORKS



ENDNOTE WEB



EXPORT RIS

See something that doesn't look right?



Report a Problem

Find In Library

Sign in for request options



Sign in

LOCATION ITEMS

Math & Physics Library
Available , Stacks P95.E88 1997



Item in place (0 requests) v.2
Loanable



Item in place (0 requests) v.1
Loanable



Details

Title	Eurospeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 242-25 September 1997 : proceedings.
Creator	Eurospeech 97 (5th : 1997 : Rhodes, Greece) European Speech Communication Association. Speech Phonetics Congresses.
Subject	
Genre	
Other title	Eurospeech ninety seven Proceedings, Eurospeech '97
Publisher	Patras : European Speech Communication Association : University of Patras, Greece
Year	1997?
Format	2 v. (lvii + lvii, 1074 p.) : ill. ; 29 cm.
Language	English
General note	"ISSN 1018-4074"--T.p. verso, v. 1-2.
Bibliography	Includes bibliographical references and index.
Identifier	OCLC : (OCoLC)ocm55169276
MMS ID	991015673189704646
Source	Library Catalog

Virtual Browse

ve got to
elieved to
eard
...

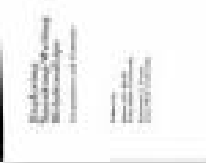
...

Directory of
graduate
programs in
the
communicatio ...
19??

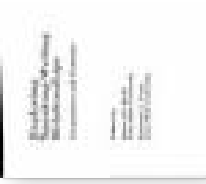


Talk and social
theory :
ecologies of
speaking and
listening in ...
2004

Eurospeech
'97 : 5th
European
Conference on
Speech
1997?



Exploring
speaking-
writing
relationships :
connections ...
c1981



Exploring
speaking-
writing
relationships :
connections
c1981

>

Links

[Display Source Record](#) >

APPENDIX D

Search WorldCat

au:euromspeech

Search

Advanced Search Find a Library

Search results for 'au:euromspeech' > '1997'

Open Content

Results 1-9 of about 9 (.04 seconds)

Open Access

Select All

Clear All

Save to:

[New List]

Save

Sort by:

Relevance

Save Search

<< First

< Prev

1

Next >

>>

Format

☒ All Formats (9)

☐ Print book (7)

☐ Computer file (2)

Refine Your Search

Author

1997, Rodos> Euro... (5)

European Speech C... (2)

1997, Rodos) Euro... (1)

G Kokkinakis (1)

Year

<< All Years

1997

Language

English

1.

174594294

1

[Proceedings / Eurospeech '97, 5th European Conference on Speech Communication and Technology / 1.](#)

by Eurospeech <5, 1997, Rodos>

Print book [View all formats and languages »](#)

Language: English

Publisher: Patras, Greece : Univ. of Patras, Wire Communications Lab., (1997)

[View all editions »](#)

2.

247058441

2

[Conference proceedings on CD-ROM : Rhodes, Greece, 22-25 September 1997](#)

by European Speech Communication Association; Wire Communications Laboratory;

Computer file : CD for computer

Language: English

Publisher: [Grenoble] ESCA 1997

3.

833245408

3

[Proceedings Eurospeech '97, 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September 1997. Abstracts](#)

by G Kokkinakis; European Speech Communication Association; Wire Communications Laboratory;

Print book : Conference publication

Language: English

Publisher: Grenoble ESCA 1997

☐ 4. 174594296

4

[Proceedings / Eurospeech '97, 5th European Conference on Speech Communication and Technology / 2.](#)

by Eurospeech <5, 1997, Rodos>

 Print book [View all formats and languages »](#)

Language: English

Publisher: Patras, Greece : Univ. of Patras, Wire Communications Lab., (1997)

[View all editions »](#)

☐ 5. 174594329

5

[Proceedings / Eurospeech '97, 5th European Conference on Speech Communication and Technology / 5.](#)

by Eurospeech <5, 1997, Rodos>

 Print book [View all formats and languages »](#)

Language: English

Publisher: Patras, Greece : Univ. of Patras, Wire Communications Lab., (1997)

[View all editions »](#)

☐ 6. 174594312

6

[Proceedings / Eurospeech '97, 5th European Conference on Speech Communication and Technology / 4.](#)

by Eurospeech <5, 1997, Rodos>

 Print book [View all formats and languages »](#)

Language: English

Publisher: Patras, Greece : Univ. of Patras, Wire Communications Lab., (1997)

[View all editions »](#)

☐ 7. 174594308

7

[Proceedings / Eurospeech '97, 5th European Conference on Speech Communication and Technology / 3.](#)

by Eurospeech <5, 1997, Rodos>

 Print book [View all formats and languages »](#)

Language: English

Publisher: Patras, Greece : Univ. of Patras, Wire Communications Lab., (1997)

[View all editions »](#)

☐ 8. 180047965

8

Conference proceedings on CD-ROM / Eurospeech '97, 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September 1997

by Eurospeech (5, 1997, Rodos); European Speech Communication Association.; Wire Communications Laboratory (Patrai)

Computer file : 5.25 in. disc : CD for computer

Language: English

Publisher: [Grenoble] : ESCA, 1997.

9.

1051302387

9

EuroSpeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September, 1997 : proceedings

by European Speech Communication Association.; Panepistēmio Patrōn. Wire Communications Laboratory.;

Print book : Conference publication

View all formats and languages »

Language: English

Publisher: Grenoble : European Speech Communication Association, [1997]

View all editions »

Results 1-9 of about 9 (.04 seconds)

<<

First

<

Prev

1

Next

>

Meta Platforms, Inc. - Exhibit 1010

Page 34 of 38

APPENDIX E

Search

Search

Advanced Search Find a Library

<< Return to Search Results

Add to list

Add tags

Write a review

Rate this item:

1 2 3 4 5

Cite/Export

Print

E-mail

Share

Permalink

EuroSpeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September, 1997 : proceedings

Author : [European Speech Communication Association.](#); [Panepistémio Patrôn.](#)
[Wire Communications Laboratory.](#)

Publisher: Grenoble : European Speech Communication Association,

140071
Edition/Format: Print book : Conference publication : English [View all editions and formats](#)

Rating: (not yet rated) [0 with reviews - Be the first.](#)

Subjects More like this
[Spoken language processing](#)
[Similar items](#)

Get a Copy
Find a copy in the library

Find a copy in the library

Enter your location

complete postal address for best results.

2011 Find libraries

Displaying libraries 1-6 out of 6 for all 3 editions (Manassas, VA 20112, USA)

Show libraries holding [just this edition](#)

Library	Held formats	Distance
1. Rutgers University Libraries Piscataway, NJ 08854 United States	Book	204 MAP IT
	Book	209 miles
		Library info

 Tags

[Add tags](#) for "EuroSpeech '97 : 5th European Conference on Speech Communication and Technology : Rhodes, Greece, 22-25 September, 1997 : proceedings". Be the first.

 Similar Items

Related Subjects: (1)

[Spoken language processing](#)