



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
265 10 Pion - Patras - Greece

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 <i>National Tsing Hua Univ., ROCina</i>	
ThMD.2 Speaker Models Designed from Complete Data Sets: A New Approach to Text-Independent Speaker Verification	2283
Dominik R. Dersch, *Robin W. King <i>Univ. of Sydney, Australia</i> *Univ. of South Australia, Australia	
ThMD.3 A Double Gaussian Mixture Modeling Approach To Speaker Recognition	2287
Vergin Rivarol, Douglas O'Shaughnessy <i>INRS Telecommunications, Canada</i>	
ThMD.4 An Acoustic Subword Unit Approach to Non-Linguistic Speech Feature Identification	2291
Mohamed Afify, Yifan Gong, Jean-Paul Haton <i>CRIN-CNRS, France</i>	
ThMD.5 N-Best GMM's for Speaker Identification	2295
Chakib Tadj, *Pierre Dumouchel, †Yu Fang <i>Ecole de Technologie Supérieure, Canada</i> *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 <i>ENST/SIG, France</i> *CNET-DIH/RCP, France	
ThMD.7 Equalizing Sub-Band Error Rates in Speaker Recognition	2303
Roland Auckenthaler, *John S. Mason <i>Technical Univ. Graz, Austria</i> *Univ. of Wales Swansea, UK	
ThMD.8 Automatic Gender Identification Under Adverse Conditions	2307
Stefan Slomka, Sridha Sridharan <i>Queensland Univ. of Technology, Australia</i>	
ThMD.9 Acoustic Features and Perceptive Processes in the Identification of Familiar Voices	2311
Yizhar Lavner, Isak Gath, Judith Rosenhouse <i>Israel Institute of Technology, Israel</i>	
ThMD.10 On the Use Acoustic Segmentation in Speaker Identification	2315
Leandro Rodriguez-Linares, Carmen Garcia-Mateo <i>Univ. of Vigo, Spain</i>	
ThMD.11 Speaker Recognition by Humans and Machines	
Herman J.M. Steeneken, David A. Van Leeuwen <i>TNO-HFRI, The Netherlands</i> 2319	

ThMD.12 Foreign Speaker Accent Classification Using Phoneme-Dependent Accent Discrimination Models and Comparisons with Human Perception Benchmarks	
Karsten Kumpf, *Robin W. King <i>Univ. of Sydney, Australia</i> *Univ. of South Australia, Australia 2323	
ThMD.13 A Comparison of Human and Machine In Speaker Recognition	2327
Li Liu, Jialong He, Günther Palm <i>Univ. of Ulm, Germany</i>	
ThMD.14 Evaluation of Second Language Learners' Pronunciation Using Hidden Markov Models	2331
Simo M.A. Goddijn, *Guus de Krom <i>Forensic Science Laboratory, Rijswijk</i> *Univ. of Utrecht, The Netherlands	
ThMD.15 Delta Vector Taylor Series Environment Compensation for Speaker Recognition	2335
Brian Eberman, Pedro J. Moreno <i>Digital Equipment Corp., USA</i>	
ThMD.16 Wavelet-Like Regression Features in the Cepstral Domain for Speaker Recognition	2339
Jonathan Hume <i>Univ. of Wales Swansea, UK</i>	
ThMD.17 Minimum Classification Error Linear Regression (MCELR) for Speaker Adaptation Using HMM with Trend Functions	2343
Rathinavelu Chengalvarayan <i>Bell Labs-Lucent Technologies, USA</i>	
ThMD.18 A Continuous HMM Text Independent Speaker Recognition System Based on Vowel Spotting	2347
Nikos Fakotakis, *Anastasios Tsopanoglou, Kallirroi Georgila <i>Univ. of Patras, Greece</i> *KNOWLEDGE SA, Greece	
ThMD.19 On the Independence of Digits in Connected Digit Strings	2351
Johan W. Koolwaaij, Lou Boves <i>Nijmegen University, The Netherlands</i>	
ThMD.20 A New Procedure for Classifying Speakers in Speaker Verification Systems	2355
Johan W. Koolwaaij, Lou Boves <i>Nijmegen University, The Netherlands</i>	
ThMD.21 Sound Channel Video Indexing	2359
Claude Montacie, Marie-Jose Caraty <i>Univ. Pierre et Marie Curie - CNRS, France</i>	
ThMD.22 CDHMM Speaker Recognition By Means of Frequency Filtering of Filter-Bank Energies	2363
Javier Hernando, Climent Nadeu <i>Universitat Politècnica de Catalunya, Spain</i>	

SESSION: Th3A**Style and Accent Recognition***Chair: Gerard Chollet, ENSP/SIG, Switzerland***Th3A.1 Using Accent-Specific Pronunciation Modelling for Improved Large Vocabulary Continuous Speech Recognition**.....2367J. 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 Llisterrí, 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 MUAF Principle**.....2395
Ahmed-Reda Berrah, Rafael Laboissiere
*Institut de la Communication Parléé, 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 Bouquin-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
Atila 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-IRIA, 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. Lublinskaja, *Inna V. Koroleva, A.N. Kornev, Elena V. Iagounova 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
Universidade de Coimbra, Portugal 2499

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

SESSION: Th4C
Voice Conversion and Data Driven FO-Models
Chair: Yoshinori Sagisaka, ATR Interpret. Telecom. Res. Labs., 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
Bell Labs-Lucent Technologies, USA 2511

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: Antreas 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**

Antonio 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., ROC China***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.....2583***Univ. du Quebec, Canada.....2583***ThAA.9 A Method of Signal Extraction from Noisy Signal Masashi Unoki, Masato Akagi.....2587***Japan Advanced Institute of Science and Technology, Japan***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**

Hidetomi 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 Malfre, 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 MMF-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 Lague
Univ. de Provence & CNRS, France

ThAB.6 Phonetic Rules for a Phonetic-to-Speech System
 Angelien A. Sanderman, *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 ITR, 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	ThAC.18 Using Intonation to Constrain Language Models in Speech Recognition 2763
Geraldine Damnati, Jacques Simonin	Paul Taylor, Simon King, Stephen Isard, Helen Wright,
<i>France Telecom, France</i>	Jaqueline Kowtko
	<i>Univ. of Edinburgh, UK</i>
ThAC.7 Variable-Length Language Modeling Integrating Global Constraints 2719	ThAC.19 Incorporating POS Tagging into Language Modeling 2767
Shoichi Matsunaga, Shigeki Sagayama	Peter A. Heeman, *James F. Allen
<i>NTT, Japan</i>	<i>France Telecom, France</i>
	<i>*Univ. of Rochester, USA</i>
ThAC.8 An Hybrid Language Model for Continuous Dictation Prototype 2723	ThAC.20 Confidence Metrics Based on N-Gram Language Model Backoff Behaviors 2771
Kamel Smaili, Imed Zitouni, Francois Charpillat, Jean-Paul Haton	Carl Uhrig, *Wayne Ward
<i>CRIN-CNRS & INRIA, Lorraine, France</i>	<i>Berdy Medical Systems, USA</i>
	<i>*Carnegie Mellon Univ., USA</i>
ThAC.9 Dealing with Pronunciation Variants at the Language Model Level for the Continuous Automatic Speech Recognition of French 2727	ThAC.21 Structure and Performance of a Dependency Language Model 2775
Laure Pousse, Guy Perennou	Ciprian Chelba, *David Engle, Frederick Jelinek, †Victor M. Jimenez, Sanjeev Khudanpur, Lidia Mangu, †Harry Printz, **Eric Ristad, ††Ronald Rosenfeld, ‡‡Andreas Stolcke, ‡‡Dekai Wu
<i>IRIT - Equipe IHMPT, France</i>	<i>John Hopkins Univ., USA</i>
ThAC.10 Rational Interpolation of Maximum Likelihood Predictors in Stochastic Language Modeling 2731	<i>*Dept. of Defense Fort Meade, MD, USA</i>
Ernst Günter Schukat-Talamazzini, *Florian Gallwitz, *Stefan Harbeck, *Volker Warnke	<i>†Universitat Politecnica de Valencia, Spain</i>
<i>Univ. of Jena, Germany</i>	<i>‡IBM, USA</i>
<i>*Univ. of Erlangen, Germany</i>	<i>**Princeton Univ., USA</i>
ThAC.11 N-Gram Language Model Adaptation Using Small Corpus for Spoken Dialog Recognition 2735	<i>††Carnegie Mellon Pittsburgh, PA, USA</i>
Akinori Ito, Hideyuki Saitoh, Masaharu Katoh, Masaki Kohda	<i>‡‡SRI International, USA</i>
<i>Yamagata Univ., Japan</i>	<i>‡‡Hong Kong Tech University, Hong Kong</i>
ThAC.12 Variable N-Gram Language Modeling and Extensions for Conversational Speech 2739	ThAC.22 Modeling Linguistic Segment and Turn Boundaries for N-Best Rescoring of Spontaneous Speech
Man-Hung Siu, *Mari Ostendorf	Andreas Stolcke
<i>BBN Inc, USA</i>	<i>SRI International, USA</i> 2779
<i>*Boston Univ., USA</i>	ThAC.23 Hybrid Language Models: Is Simpler better?
ThAC.13 Fuzzy Class Rescoring: A Part-of-Speech Language Model 2743	Peter E. Kenne, Mary O'Kane
Petra Geutner	<i>Univ. of Adelaide, Australia</i> 2783
<i>Univ. of Karlsruhe, Germany</i>	ThAC.24 Internal and External Tagsets in Part-of-Speech Tagging 2787
ThAC.14 Speech Understanding Based on Integrating Concepts By Conceptual Dependency 2747	Thorsten Brants
Akito Nagai, Yasushi Ishikawa	<i>Univ. of the Saarland, Germany</i>
<i>Mitsubishi Electric Corporation, Japan</i>	
ThAC.15 Dynamic Language Models for Interactive Speech Applications 2751	
Fabio Brugnara, Marcello Federico	
<i>Istituto per la Ricerca Scientifica e Tecnologica (IRST), Italy</i>	
ThAC.16 Large-Scale Lexical Semantics for Speech Recognition Support 2755	
George Demetriou, Eric Atwell, Clive Souter	
<i>Univ. of Leeds, UK</i>	
ThAC.17 Integration of Grammar and Statistical Language Constraints for Partial Word-Sequence Recognition 2759	
Hajime Tsukada, Hirofumi Yamamoto, Yoshinori Sagisaka	
<i>ATR Interpreting Telecommunications Res. Labs., Japan</i>	
	ThAD.1 A Probabilistic Model of Double-Vowel Segregation 2791
	Laurent Varin, Frederic Berthommier
	<i>ICP, INPG, France</i>
	ThAD.2 Stimulus Signal Estimation From Auditory-Neural Transduction Inverse Processing 2795
	Houshang Habibzadeh Vaneghi, Shigeyoshi Kitazawa
	<i>Shizuoka Univ., Japan</i>

SESSION: THAD

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

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 Professionnalisé, 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,

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

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 $BH_1 N$ is uncorrelated with the reference signal noise component $BH_2 N$, 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:

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

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

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

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.

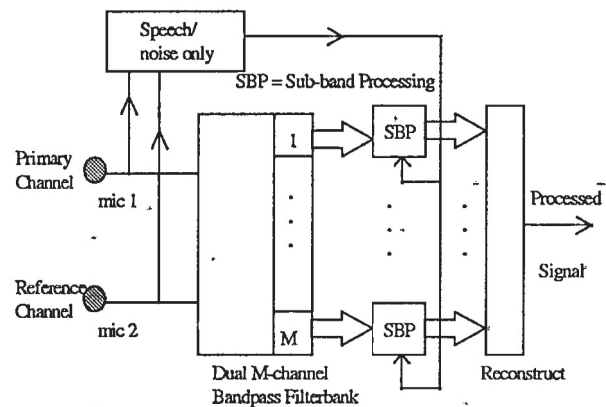


Figure 1: The MMSBA system

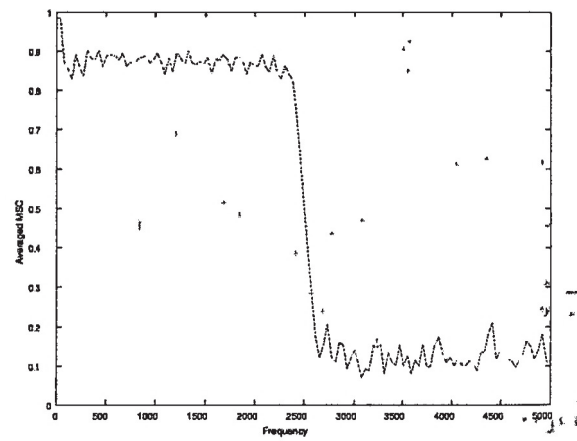


Figure 2: The adaptively estimated MSC between the disturbing synthetic noises