

2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays

(HSCMA 2011)

**Edinburgh, United Kingdom
30 May – 1 June 2011**



**IEEE Catalog Number: CFP11HSM-PRT
ISBN: 978-1-4577-0997-5**

Program

Monday, May 30

W: Welcome and Opening Remarks

K1: Keynote 1

A Versatile Speech Front-End for Telecommunication and Speech Recognition - The Last One Mile: Implementation Issues for a Better Product

Akihiko K. Sugiyama (NEC Corporation, Japan)

MB1: Break

T1: Talks 1 - Distant Speech Recognition

Channel Selection based on Multichannel Cross-Correlation Coefficients for Distant Speech Recognition

Kenichi Kumatani (Disney Research, Pittsburgh, USA); John McDonough (Carnegie Mellon University, USA); Jill Lehman (Carnegie Mellon University, USA); Bhiksha Raj (Carnegie Mellon University, USA)

pp. 1-6

Discriminative approach to dynamic variance adaptation for noisy speech recognition

Marc Delcroix (NTT Communication Science Laboratories, Japan); Shinji Watanabe (NTT Communication Science Laboratories, Japan); Tomohiro Nakatani (NTT Corporation, Japan); Atsushi Nakamura (NTT Communication Science Laboratories, Japan)

pp. 7-12

Extension of the REMOS Concept to Frequency-Filtering-Based Features for Reverberation-Robust Speech Recognition

Roland Maas (University of Erlangen-Nuremberg, Germany); Martin Wolf (Universitat Politècnica de Catalunya, Spain); Armin Sehr (University of Erlangen-Nuremberg, Germany); Climent Nadeu (UPC, Spain); Walter Kellermann (University Erlangen-Nuremberg, Germany)

pp. 13-18

Theoretical analysis of parametric blind spatial subtraction array and its application to speech recognition performance prediction

Ryoichi Miyazaki (Nara Institute of Science and Technology, Japan); Hiroshi Saruwatari (Graduate School of Information Science, Nara Institute of Science and Technology, Japan); Ryo Wakisaka (Nara Institute of Science and Technology, Japan); Kiyohiro Shikano (Graduate School of Information Science,

Nara Institute of Science and Technology, Japan); Tomoya Takatani (Toyota Motor Corporation, Japan)
pp. 19-24

ML: Lunch

P1: Posters 1

Audio Spatio-Temporal Fingerprints for Cloudless Real-Time Hands-Free Diarization on Mobile Devices

Danil Korchagin (Idiap Research Institute, Switzerland)
pp. 25-30

Worst-case performance optimization for spherical microphone array modal beamformers

Haohai Sun (Norwegian University of Science and Technology, Norway); Shefeng Yan (Chinese Academy of Science, P.R. China); Ulf Peter Svensson (Norwegian University of Science and Technology (NTNU), Norway)
pp. 31-35

Blind source separation of mixed speech in a high reverberation environment

Keiju Iso (University of Tsukuba, Japan); Shoko Araki (NTT Communication Science Laboratories, Japan); Shoji Makino (University of Tsukuba, Japan); Tomohiro Nakatani (NTT Corporation, Japan); Hiroshi Sawada (NTT communication Science Laboratories, Japan); Takeshi Yamada (University of Tsukuba, Japan); Atsushi Nakamura (NTT Communication Science Laboratories, Japan)
pp. 36-39

Diffuseness estimation via surface arrays for directional audio coding

Michael Meier (Fraunhofer IIS, Germany); Giovanni Del Galdo (Fraunhofer Institute for Integrated Circuits IIS, Germany)
pp. 40-45

Median Tracking in Noise Subspace for Noise Floor Estimation

Mahdi Triki (Philips Research, The Netherlands)
pp. 46-51

Self-clustering non-Euclidean kernels for improving the estimation of multidimensional TDOA of multiple sources

Francesco Nesta (Fondazione Bruno Kessler - irst, Italy); Alessio Brutti (Fondazione Bruno Kessler, Italy)
pp. 52-57

An Improved Combination of Directional BSS and a Source Localizer for Robust Source Separation in Rapidly Time-varying Acoustic Scenarios

Yuanhang Zheng (University of Erlangen-Nuremberg, Germany); Anthony Lombard (University of Erlangen-Nuremberg, Germany); Walter Kellermann (University Erlangen-Nuremberg, Germany)
pp. 58-63

Group Delay based methods for Speech Source Localization over Circular Arrays

Ardhendu Tripathy (Indian Institute of Technology, Kanpur, India); Lalan Kumar (Indian Institute of Technology Kanpur, India); Rajesh M Hegde (IIT Kanpur, India)
pp. 64-69

Improving Hands-Free Speech Recognition in a Car Through Audio-Visual Voice Activity Detection

Friedrich Faubel (Saarland University, Germany); Munir Georges (Saarland University, Germany); Kenichi Kumatani (Disney Research, Pittsburgh, USA); Andrés Bruhn (Saarland University, Germany); Dietrich Klakow (Saarland University, Germany)
pp. 70-75

Sub-Nyquist Spatial Sampling Using Arrays of Directional Microphones

Vladimir Tourbabin (Ben-Gurion University of the Negev & Ben-Gurion University of the Negev, Israel); Boaz Rafaely (Ben-Gurion University, Israel)
pp. 76-80

A Time-domain Implementation of Data-Independent Robust Broadband Beamformers with Low Filter Order

Edwin Mabande (University of Erlangen-Nuremberg, Germany); Adrian Schad (Technology University of Darmstadt, Germany); Walter Kellermann (University Erlangen-Nuremberg, Germany)
pp. 81-85

Linearly constrained minimum variance method for spherical microphone arrays in a coherent environment

Yotam Peled (Ben-Gurion University of the Negev, Israel); Boaz Rafaely (Ben-Gurion University, Israel)
pp. 86-91

An Integrated Framework for Multi-Channel Multi-Source Localization and Voice Activity Detection

Mohammad J. Taghizadeh (Idiap Research Institute & EPFL, Switzerland); Philip Garner (Idiap Research Institute, Switzerland); Herve Boursard (IDIAP, Switzerland); Hamid Reza Abutalebi (Yazd University, Iran); Afsaneh Asaei (Idiap Research Institute & EPFL, Switzerland)
pp. 92-97

Combined Echo and Noise Reduction for Distributed Microphones

Eric Böhmler (University of Applied Sciences Constance, Germany); Juergen Freudenberger (University of Applied Sciences Constance & Institute for System Dynamics (ISD), Germany); Sebastian Stenzel (University of Applied Sciences Constance & Institute for System Dynamics (ISD), Germany)
pp. 98-103

Time-frequency masking for convolutive and noisy mixtures

Juergen Freudenberger (University of Applied Sciences Constance & Institute for System Dynamics (ISD), Germany); Sebastian Stenzel (University of Applied Sciences Constance & Institute for System Dynamics (ISD), Germany)
pp. 104-108

***Speech Measurements Using a Laser Doppler Vibrometer Sensor:
Application to Speech Enhancement***

Yekutiel Avargel (AudioZoom Ltd, Israel); Israel Cohen (Technion, Israel)
pp. 109-114

***Non-linear Spectro-temporal Modulations for Reverberant Speech
Recognition***

Marco Matassoni (Fondazione Bruno Kessler, Italy); Hari Krishna Maganti
(Fondazione Bruno Kessler - Center for Information Technology -IRST, Italy);
Maurizio Omologo (Fondazione Bruno Kessler - irst, Italy)
pp. 115-120

WR: Welcome Reception

Tuesday, May 31

K2: Keynote 2

***Model-based approaches to handling additive noise in reverberant
environments***

Mark Gales (University of Cambridge, United Kingdom); Y Wang (University of
Cambridge, United Kingdom)
pp. 121-126

TB1: Break

T2: Talks 2 - Microphone arrays

Towards Acoustic Self-Localization of Ad Hoc Smartphone Arrays

Marius Hennecke (TU Dortmund University, Germany); Gernot Fink (TU Dortmund
University, Germany)
pp. 127-132

***A Wave-Domain Model for Acoustic MIMO Systems with Reduced
Complexity***

Martin Schneider (University for Erlangen-Nuremberg, Germany); Walter
Kellermann (University Erlangen-Nuremberg, Germany)
pp. 133-138

Closed-Form Self-Localization of Asynchronous Microphone Arrays

Pasi Perilä (Tampere University of Technology, Finland); Mikael Mieskolainen
(Tampere University of Technology, Finland); Matti S Hämäläinen (Nokia
Research Center, Finland)
pp. 139-144

***Dereverberation Performance of Rigid and Open Spherical Microphone
Arrays: Theory & Simulation***

Daniel P Jarrett (Imperial College London, United Kingdom); Emanuel Habets (International Audio Laboratories Erlangen, Germany); Mark Thomas (Imperial College London, United Kingdom); Nikolay D Gaubitch (Imperial College London, United Kingdom); Patrick A Naylor (Imperial College London, United Kingdom)
pp. 145-150

TL: Lunch

D1: Demo session

Low-latency meeting recognition and understanding using distant microphones

Shoko Araki (NTT Communication Science Laboratories, Japan); Takaaki Hori (NTT Communication Science Laboratories, Japan); Takuya Yoshioka (NTT Corporation, Japan); Masakiyo Fujimoto (NTT Communication Science Laboratories, Japan); Shinji Watanabe (NTT Communication Science Laboratories, Japan); Takanobu Oba (NTT Communication Science Laboratories, Japan); Atsunori Ogawa (NTT Corporation, Japan); Kazuhiro Otsuka (NTT, Japan); Dan Mikami (NTT, Japan); Marc Delcroix (NTT Communication Science Laboratories, Japan); Keisuke Kinoshita (NTT Communication Science Laboratories, Japan); Tomohiro Nakatani (NTT Corporation, Japan); Atsushi Nakamura (NTT Communication Science Laboratories, Japan); Junji Yamato (NTT, Japan)
pp. 151-152

A Speech-Based Conversation System for Accessing Agriculture Commodity Prices in Indian Languages

Gautam Varma Mantena (IIIT Hyderabad, India); Rajendran S. (IIIT Hyderabad, India); Rambabu B. (IIIT Hyderabad, India); Suryakanth Gangashetty (IIIT Hyderabad, India); Yegnanarayana B. (International Institute of Information Technology, Hyderabad, India); Kishore Prahallad (IIIT Hyderabad, India)
pp. 153-154

Positioning System for Mobile Terminals Using a Microphone Array Network as an Intuitive Interface

Shimpei Soda (Kobe University, Japan); Koji Kugata (Kobe University, Japan); Tomoya Takagi (Kobe University, Japan); Hiroki Noguchi (Kobe University, Japan); Shintaro Izumi (Kobe University, Japan); Masahiko Yoshimoto (Kobe University, Japan); Hiroshi Kawaguchi (Kobe University, Japan)
pp. 155-156

IllumiSense: Context Sensitive Illumination

Marius Hennecke (TU Dortmund University, Germany); Christian Kleine-Cosack (TU Dortmund University, Germany); Gemot Fink (TU Dortmund University, Germany)
pp. 157-158

A Practical Beamformer-Postfilter System for Adaptive Speech Enhancement in Non-Stationary Noise Environments

Tobias Wolff (Nuance Communications, Germany); Markus Buck (Nuance Communications, Germany)

pp. 159-160

Real-time prototype for multiple source tracking through Generalized State Coherence Transform and Particle Filtering

Francesco Nesta (Fondazione Bruno Kessler - irst, Italy); Alessio Brutti (Fondazione Bruno Kessler, Italy); Luca Cristoforetti (Fondazione Bruno Kessler - irst, Italy)

pp. 161-162

The Ambient Spotlight: Personal meeting capture with a microphone array

Jonathan Kilgour (University of Edinburgh, United Kingdom); Carletta (University of Edinburgh, United Kingdom); Steve Renals (University of Edinburgh, United Kingdom)

pp. 163-164

P2: Posters 2

Use of reflected wavefronts for acoustic source localization with a line array

Piergiorgio Svaizer (Fondazione Bruno Kessler, Italy); Alessio Brutti (Fondazione Bruno Kessler, Italy); Maurizio Omologo (Fondazione Bruno Kessler - irst, Italy)

pp. 165-169

A Double Talk Control Method Improving Estimation Speed by Adjusting Required Error Level

Kensaku Fujii (University of Hyogo, Japan); Takuto Yoshioka (University of Hyogo, Japan); Kana Yamasaki (University of Hyogo, Japan); Mitsuji Muneyasu (Kansai University, Japan); Masakazu Morimoto (University of Hyogo, Japan)

pp. 170-175

First-order Superdirectional Acoustic Zooming in the Presence of Directional Interferences

Rene Derkx (Philips Research, Eindhoven, The Netherlands)

pp. 176-179

Functional Link Based Architectures For Nonlinear Acoustic Echo Cancellation

Danilo Communiello (Sapienza University of Rome, Italy); Luis Azpicueta-Ruiz (Universidad Carlos III de Madrid, Spain); Michele Scarpiniti (University of Rome "La Sapienza", Italy); Aurelio Uncini (University of Rome "La Sapienza", Italy); Jerónimo Arenas-García (Universidad Carlos III de Madrid, Spain)

pp. 180-184

Generating Virtual Microphone Signals Using Geometrical Information Gathered by Distributed Arrays

Giovanni Del Galdo (Fraunhofer Institute for Integrated Circuits IIS, Germany); Oliver Thiergart (Friedrich-Alexander University Erlangen-Nürnberg, Germany); Tobias Weller (Fraunhofer Institute for Integrated Circuits IIS, Germany); Emanuel Habets (International Audio Laboratories Erlangen, Germany)

pp. 185-190

Joint Dereverberation and Noise Reduction Using a Two-Stage Beamforming

Approach

Emanuel Habets (International Audio Laboratories Erlangen, Germany); Jacob Benesty (INRS-EMT, University of Quebec, Canada)
pp. 191-195

Multi-Style Training of HMMs with Stereo Data for Reverberation-Robust Speech Recognition

Armin Sehr (University of Erlangen-Nuremberg, Germany); Christian Hofmann (University of Erlangen-Nuremberg, Germany); Roland Maas (University of Erlangen-Nuremberg, Germany); Walter Kellermann (University Erlangen-Nuremberg, Germany)
pp. 196-200

An Analysis of Nonstationary Variance Estimates in the Maximum Negentropy Beamformer

Barbara Rauch (Saarland University, Germany); Friedrich Faubel (Saarland University, Germany); Dietrich Klakow (Saarland University, Germany)
pp. 201-206

Incorporating localisation cues in a fragment decoding framework for distant binaural speech recognition

Ning Ma (University of Sheffield, United Kingdom); Jon Barker (University of Sheffield, United Kingdom); Heidi Christensen (University of Sheffield, United Kingdom); Phil Green (University of Sheffield, United Kingdom)
pp. 207-212

CD: Conference Dinner (The Caves)

Wednesday, June 1

K3: Keynote 3

Environmental sound recognition and classification

Daniel P W Ellis (Columbia University, USA)

WB: Break

T3: Talks 3 - Microphone arrays and applications

Statistical Method to Identify Key Anthropometric Parameters in HRTF Individualization

Mengqiu Zhang (Australian National University, Australia); Rodney Andrew Kennedy (Australian National University, Australia); Thushara D. Abhayapala (Australian National University, Australia); Wen Zhang (Process Science and Engineering, CSIRO, Australia)
pp. 213-218

A microphone array system integrating beamforming, feature enhancement, and spectral mask-based noise estimation

Takuya Yoshioka (NTT Corporation, Japan); Tomohiro Nakatani (NTT Corporation, Japan)

pp. 219-224

Single and Multichannel Enhancement of Distant Speech using characteristics of Speech Production

Yegnanarayana B. (International Institute of Information Technology, Hyderabad, India); Guruprasad S (IIIT Hyderabad, India); S. R. Mahadeva Prasanna (Indian Institute of Technology Guwahati, India); Suryakanth Gangashetty (IIIT Hyderabad, India)

pp. 225-230

CT: Closing Talk

HSCMA: Hands-free Sound Capture and Microphone Array in Kinect

Ivan Tashev (Microsoft Research, USA)

CR: Closing remarks

WL: Lunch