# 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics

**Mohonk Mountain House**
**New Paltz, New York**
**October 21-24, 2007**

**WASPAA2007**

**http://www.kecl.ntt.co.jp/icl/signal/waspaa2007/**

IEEE Signal Processing Society

IEEE Catalog Number: CFP07AUDC
ISBN: 978-1-4244-1619-6
Library of Congress: 2007905904

## ORGANIZING COMMITTEE

***General Chair***
Shoji Makino
NTT CS Labs, Japan

***Technical Program Chairs***
Masato Miyoshi
NTT CS Labs, Japan

Tomohiro Nakatani
NTT CS Labs, Japan

***Finance Chair***
Michael Brandstein
MIT Lincoln Lab, USA

***Publications Chairs***
Hiroshi Saruwatari
NAIST, Japan

Hiroshi Sawada
NTT CS Labs, Japan

***Publicity Chair***
Steven Grant
University of Missouri
Rolla, USA

***Local Arrangements***
Yiteng (Arden) Huang
Bell Labs
Lucent Technologies, USA

***Registration Chairs***
Kunio Kashino
NTT CS Labs, Japan

Shoko Araki
NTT CS Labs, Japan

***Web Administrators***
Yu Takahashi
Yoshimitsu Mori
Shigeki Miyabe
NAIST, Japan

***Workshop Committee***
Takuya Yoshioka
Satoko Kaida
NTT CS Labs, Japan

## LIST OF REVIEWERS

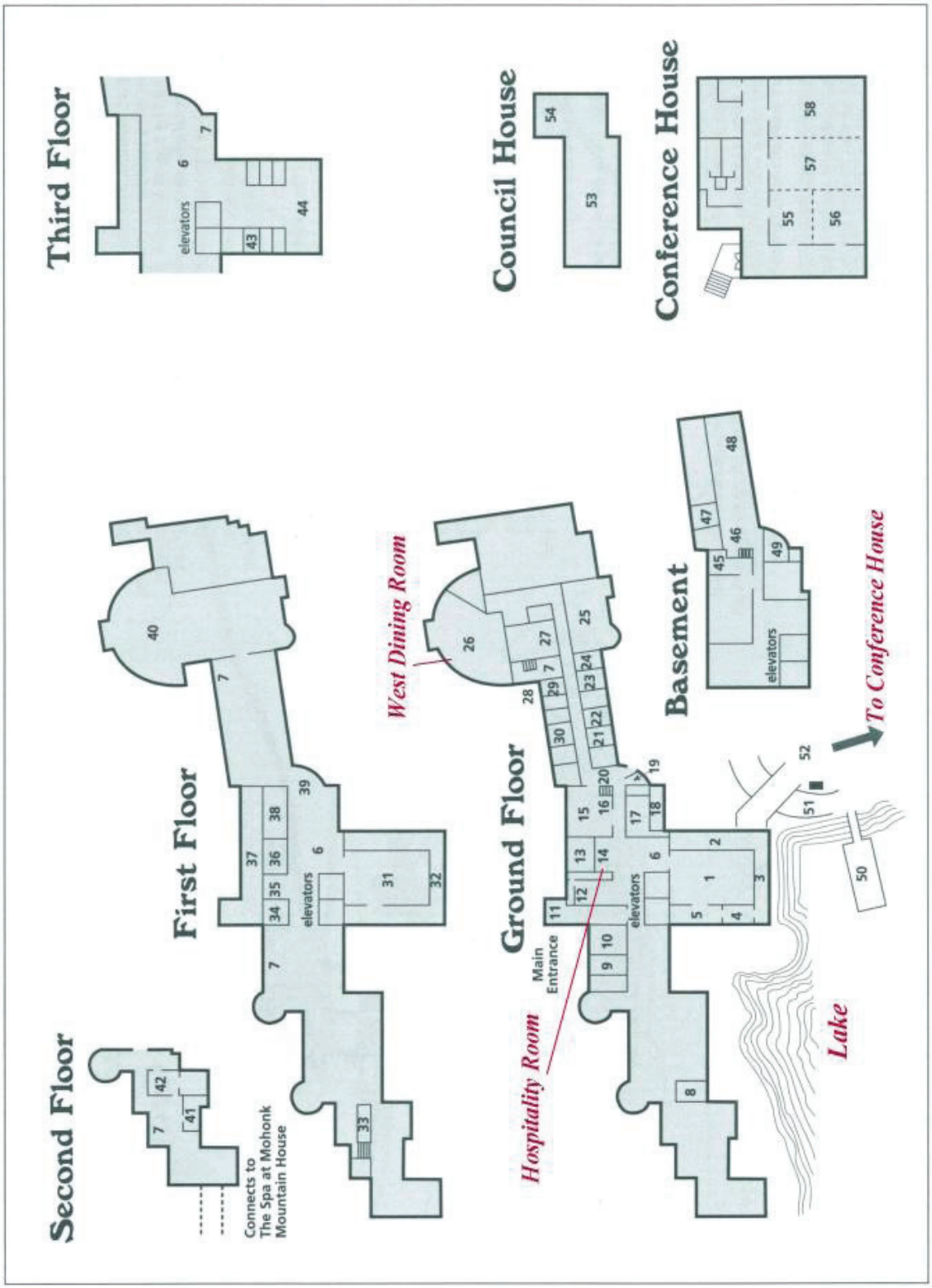
We would like to thank the following people for taking time from their busy schedules to review papers for this workshop.

Thushara D. Abhayapala
Shoko Araki
Michael Brandstein
Michael Casey
Simon Doclo
Daniel P. W. Ellis
Gianpaolo Evangelista
Nikolay D. Gaubitch
Simon Godsill
Michael Goodwin
Wolfgang Herbordt
Takafumi Hikichi
James R. Hopgood
Jyri Huopaniemi
Matti Karjalainen
Walter Kellermann
Tetsuro Kitahara
Bastiaan Kleijn
Eric Lehmann
Rainer Martin
Masahide Mizushima
Takehiro Moriya
Tomohiro Nakatani
Philip Nelson
Ryouich Nishimura
Maurizio Omologo
Naotoshi Osaka
Renato S. Pellegrini
Tomas F. Quatieri
Boaz Rafaely
Robert Aichner
Joerg Bitzer
Herbert Buchner
Jingdong Chen
Gary Elko
Cumhur Erkut
Christof Faller
Ralf Geiger
Alberto Gonzalez
Masataka Goto
Juergen Herre
Robert Hoeldrich
Yiteng (Arden) Huang
Yutaka Kaneda
Kunio Kashino
Youngmoo E. Kim
Anssi Klapuri
Dorothea Kolossa
Shoji Makino
Masato Miyoshi
Marc Moonen
Meinard Mueller
Patrick A. Naylor
Nikos Nikolaidis
Sven Nordholm
Nobutaka Ono
Lucas C Parra
Mark D. Plumbley
Rudolf Rabenstein
Mark Sandler

Augusto Sarti
Hiroshi Sawada
Gerald Schuller
Paris Smaragdis
Peter Steffen
Piergiorgio Svaizer
Heinz Teutsch
Vesa Valimaki
Wieslaw Woszczyk
Dmitry Zotkin

Hiroshi Saruwatari
Gerhard Schmidt
Malcolm Slaney
Sriram Srinivasan
Akihiko Sugiyama
Ivan Tashev
Masashi Unoki
Peter Vary
Udo Zoelzer

# FLOOR PLAN


Third Floor
elevators
6
7
43
44
Council House
54
53
Conference House
55
56
57
58
Basement
47
48
46
45
49
elevators
First Floor
40
7
39
37
38
36
35
34
6
elevators
31
32
7
33
Second Floor
7
42
41
Connects to
The Spa at Mohonk
Mountain House
West Dining Room
Ground Floor
26
28
27
7
29
25
24
23
22
21
30
20
19
15
16
17
18
13
14
12
11
6
elevators
5
1
2
4
3
Main
Entrance
9
10
8
Hospitality Room
52
51
50
To Conference House
Lake

# WORKSHOP SCHEDULE OVERVIEW

| | Sunday, October 21 | Monday, October 22 | Tuesday, October 23 | Wednesday, October 24 |
|---|---|---|---|---|
| 7:00<br>8:00 | | **Breakfast**<br>West Dining Room | **Breakfast**<br>West Dining Room | **Breakfast**<br>West Dining Room |
| 8:00<br>9:00 | | **Keynote Address 1**<br>*Simon Haykin*<br>Conference House | **Keynote Address 2**<br>*Albert S. Bregman*<br>Conference House | **Lecture WL1**<br>*Music and Signal Analysis and Synthesis*<br>Conference House |
| 9:00<br>10:00 | | **Lecture ML1**<br>*Microphone Array Signal Processing*<br>Conference House | **Lecture TL1**<br>*Signal Enhancement*<br>Conference House | |
| 10:00<br>10:20 | | **Break**<br>West Dining Room | **Break**<br>West Dining Room | **Break**<br>West Dining Room |
| 10:20<br>12:20 | | **Poster MP1**<br>West Dining Room | **Poster TP1**<br>West Dining Room | **Poster WP1**<br>West Dining Room |
| 12:20<br>14:00 | | **Lunch**<br>West Dining Room | **Lunch**<br>West Dining Room | **Lunch**<br>West Dining Room |
| 14:00<br>15:40 | | | | |
| 15:40<br>16:00 | | **Break**<br>Conference House | **Break**<br>Conference House | |
| 16:00<br>18:00 | **Registration**<br>Hospitality Room | **Lecture ML2**<br>*Source Localization and Blind Source Separation*<br>Conference House | **Lecture TL2**<br>*Speech and Audio Coding and Hearing Aid*<br>Conference House | |
| 18:00<br>20:00 | **Dinner**<br>West Dining Room | **Dinner**<br>West Dining Room | **Dinner**<br>West Dining Room | |
| 20:00<br>22:00 | **Open Bar**<br>West Dining Room | **Demo Session 1**<br>West Dining Room | **Demo Session 2**<br>West Dining Room | |

# Sunday, October 21

Registration (October 21, 16:00-18:00, Hospitality Room)

Dinner (October 21, 18:00-20:00, West Dining Room)

Open Bar (October 21, 20:00-22:00, West Dining Room)

# Monday, October 22

## Keynote Address1 **Simon Haykin (McMaster University)**

(October 22, 8:00-9:00, Conference House)
Chair: *Shoji Makino*

8:00-9:00

[Keynote1] **Coherent ICA: Implications for Auditory Signal Processing** . . . . . 1
*Simon Haykin, Kevin Kan*

## Lecture ML1 **Microphone Array Signal Processing**

(October 22, 9:00-10:00, Conference House)
Chair: *Gary W. Elko*

9:00-9:20

[ML1-1] **Enhanced Microphone-Array Beamforming Based on Frequency-Domain Spatial Analysis-Synthesis** . . . . . 2
*Michael M. Goodwin*

9:20-9:40

[ML1-2] **Real Time Capture of Audio Images and Their Use with Video** . . . . . 2
*Adam O'Donovan, Ramani Duraiswami, Nail A. Gumerov*

9:40-10:00

[ML1-3] **Subband Method for Multichannel Least Squares Equalization of Room Transfer Functions** . . . . . 2
*Nikolay D. Gaubitch, Mark R. P. Thomas, Patrick A. Naylor*

## Break (October 22, 10:00-10:20, West Dining Room)

## Poster MP1

(October 22, 10:20-12:20, West Dining Room)
Chair: *W. Bastiaan Kleijn*

[MP1-01] **Broadband Music: Opportunities and Challenges for Multiple Source Localization** 3
*Jacek P. Dmochowski, Jacob Benesty, Sofiène Affes*

[MP1-02] **Energy-Based Position Estimation of Microphones and Speakers for ad hoc Microphone Arrays** . . . . . 3
*Minghua Chen, Zicheng Liu, Li-Wei He, Phil Chou, Zhengyou Zhang*

[MP1-03] **Linear Regression on Sparse Features for Single-Channel Speech Separation** . . 3
*Mikkel N. Schmidt, Rasmus K. Olsson*

[MP1-04] **Sound Source Separation using Null-Beamforming and Spectral Subtraction for Mobile Devices** . . . . . 3
*Shintaro Takada, Satoshi Kanba, Tetsuji Ogawa, Kenzo Akagiri, Tetsunori Kobayashi*

[MP1-05] **On Dealing with Sampling Rate Mismatches in Blind Source Separation and Acoustic Echo Cancellation** . . . . . 4
*Enrique Robledo-Arnuncio, Ted S. Wada, Biing-Hwang (Fred) Juang*

[MP1-06] **Signal Deflation and Paraunitary Constraints in Spatio-Temporal FastICA-Based Convolutive Blind Source Separation of Speech Mixtures** . . . . . 4
*Malay Gupta, Scott C. Douglas*

[MP1-07] **Fast Convergence Blind Source Separation Based on Frequency Subband Interpolation by Null Beamforming** . . . . . 4
*Keiichi Osako, Yoshimitsu Mori, Yu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano*

[MP1-08] **Electronic Pop Protection for Microphones** . . . . . 5
*Gary W. Elko, Jens Meyer, Steven Backer, Jürgen Peissig*

[MP1-09] **A Practical Multichannel Dereverberation Algorithm using Multichannel DYPSA and Spatiotemporal Averaging** . . . . . 5
*Mark R. P. Thomas, Nikolay D. Gaubitch, Jon Gudnason, Patrick A. Naylor*

[MP1-10] **Isotropic Noise Suppression in the Power Spectrum Domain by Symmetric Microphone Arrays** . . . . . 5
*Hikaru Shimizu, Nobutaka Ono, Kyosuke Matsumoto, Shigeki Sagayama*

[MP1-11] **Acoustic Echo Cancelation for Dynamically Steered Microphone Array Systems** 6
*Matti Hämäläinen, Ville Myllylä*

[MP1-12] **A New Approach to Digital Audio Equalization** . . . . . 6
*S. Cecchi, L. Palestini, E. Moretti, F. Piazza*

[MP1-13] **Implementation of Directional Sources in Wave Field Synthesis** . . . . . 6
*Jens Ahrens, Sascha Spors*

[MP1-14] **A Comparsion of Acoustic and Psychoacoustic Measurements of Pass-Through Hearing Protection Devices** . . . . . 6
*Douglas S. Brungart, Brian W. Hobbs, James T. Hamil*

[MP1-15] **Improvement in Detectability of Alarm Signals in Noisy Environments by Utilizing Spatial Cues** . . . . . 7
*Hideaki Uchiyama, Masashi Unoki, Masato Akagi*

[MP1-16] **Estimation Model for the Speech-Quality Dimension "Directness / Frequency Content"** . . . . . 7
*Lu Huo, Marcel Wältermann, Kirstin Scholz, Alexander Raake, Ulrich Heute, Sebastian Möller*

[MP1-17] **Probabilistic Model Based Similarity Measures for Audio Query-By-Example** . 7
*Tuomas Virtanen, Marko Helén*

[MP1-18] **Improving Generalization for Classification-Based Polyphonic Piano Transcription** 8
*Graham E. Poliner, Daniel P. W. Ellis*

[MP1-19] **Acoustic Signal Processing for Degradation Analysis of Rotating Machinery to Determine the Remaining Useful Life** . . . . . 8
*Patricia Scanlon, Alan M. Lyons, Alan O'Loughlin*

[MP1-20] **Single-Frame Discrimination of Non-Stationary Sinusoids** . . . . . 8
*Jeremy J. Wells, Damian T. Murphy*

## Lunch (October 22, 12:20-15:40, West Dining Room)

## Break (October 22, 15:40-16:00, Conference House)

## Lecture ML2 Source Localization and Blind Source Separation

(October 22, 16:00-18:00, Conference House)
Chair: *Scott Douglas*

16:00-16:20

[ML2-1] **Modeling of Motion Dynamics and its Influence on the Performance of a Particle Filter for Acoustic Speaker Tracking** . . . . . 9
*Eric A. Lehmann, Anders M. Johansson, Sven Nordholm*

16:20-16:40

[ML2-2] **Multi Target Acoustic Source Tracking using Track Before Detect** . . . . . 9
*Maurice Fallon, Simon Godsill*

16:40-17:00

[ML2-3] **Blind Sparse-Nonnegative (BSN) Channel Identification for Acoustic Time-Difference-Of-Arrival Estimation** . . . . . 9
*Yuanqing Lin, Jingdong Chen, Youngmoo Kim, Daniel D. Lee*

17:00-17:20

[ML2-4] **Blind Criterion and Oracle Bound for Instantaneous Audio Source Separation using Adaptive Time-Frequency Representations** . . . . . 10
*Emmanuel Vincent, Rémi Gribonval*
17:20-17:40
[ML2-5] **Monaural Speech Separation using Source-Adapted Models** . . . . . 10
*Ron J. Weiss, Daniel P. W. Ellis*
17:40-18:00
[ML2-6] **A Soft Masking Strategy Based on Multichannel Speech Probability Estimation for Source Separation and Robust Speech Recognition** . . . . . 10
*Eugen Hoffmann, Dorothea Kolossa, Reinhold Orglmeister*

Dinner (October 22, 18:00-20:00, West Dining Room)

Demo Session 1 (October 22, 20:00-22:00, West Dining Room)

# Tuesday, October 23

## Keynote Address2 **Albert S. Bregman (McGill University)**

(October 23, 8:00-9:00, Conference House)
Chair: *Daniel P.W. Ellis*

8:00-9:00

[Keynote2] **Progress in the Study of Auditory Scene Analysis** . . . . . 11
*Albert S. Bregman*

## Lecture TL1 **Signal Enhancement**

(October 23, 9:00-10:00, Conference House)
Chair: *Patric A. Naylor*

9:00-9:20

[TL1-1] **Single-Channel Impact Noise Suppression with No Auxiliary Information for its Detection** . . . . . 12
*Akihiko Sugiyama*

9:20-9:40

[TL1-2] **Aliasing Reduction for Modified Discrete Cosine Transform Domain Filtering and its Application to Speech Enhancement** . . . . . 12
*Fabian Kuech, Bernd Edler*

9:40-10:00

[TL1-3] **Example-Driven Bandwidth Expansion** . . . . . 12
*Paris Smaragdis, Bhiksha Raj*

## Break (October 23, 10:00-10:20, West Dining Room)

## Poster TP1

(October 23, 10:20-12:20, West Dining Room)
Chair: *Paris Smaragdis*

[TP1-01] **A Two-Stage Frequency-Domain Blind Source Separation Method for Underdetermined Convolutive Mixtures** . . . . . 13
*Hiroshi Sawada, Shoko Araki, Shoji Makino*

[TP1-02] **Long-Term Gain Estimation in Model-Based Single Channel Speech Separation** 13
*M. H. Radfar, R. M. Dansereau*

[TP1-03] **Sparseness-Based 2ch BSS using the EM Algorithm in Reverberant Environment** 13
*Yosuke Izumi, Nobutaka Ono, Shigeki Sagayama*

[TP1-04] **Prior Structures for Time-Frequency Energy Distributions** . . . . . 14
*Ali Taylan Cemgil, Paul Peeling, Onur Dikmen, Simon Godsill*

[TP1-05] **Fast Time-Domain Spherical Microphone Array Beamforming** . . . . . 14
*Zhiyun Li, Ramani Duraiswami*

[TP1-06] **Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model** . . . . . 14
*Eric A. Lehmann, Anders M. Johansson, Sven Nordholm*

[TP1-07] **Overfitting-Resistant Speech Dereverberation** . . . . . 14
*Takuya Yoshioka, Tomohiro Nakatani, Takafumi Hikichi, Masato Miyoshi*

[TP1-08] **Novel and Efficient Download Test for Two Path Echo Canceller** . . . . . 15
*Mohammad Asif Iqbal, Steven L. Grant*

[TP1-09] **An Approach to Massive Multichannel Broadband Feedforward Active Noise Control using Wave-Domain Adaptive Filtering** . . . . . 15
*Sascha Spors, Herbert Buchner*

[TP1-10] **Enhancement of Residual Echo for Improved Frequency-Domain Acoustic Echo Cancellation** . . . 15
*Ted S. Wada, Biing-Hwang (Fred) Juang*

[TP1-11] **Effects of Pre-Processing Filters on a Wavelet Packet-Based Algorithm to Identify Speech Transients** . . . 16
*Daniel M. Rasetshwane, J. Robert Boston, Ching-Chung Li, John D. Durrant*

[TP1-12] **Modeling Spot Microphone Signals using the Sinusoidal Plus Noise Approach** . . 16
*Christos Tzagkarakis, Athanasios Mouchtaris, Panagiotis Tsakalides*

[TP1-13] **A Modified Spatio-Temporal Orthogonal Iteration Method for Multichannel Audio Signal Representation** . . . 16
*Scott C. Douglas, Malay Gupta*

[TP1-14] **A Low-Delay Audio Coder with Constrained-Entropy Quantization** . . . 17
*Minyue Li, W. Bastiaan Kleijn*

[TP1-15] **Extending Fine-Grain Scalable Audio Coding to Very Low Bitrates using Overcomplete Dictionaries** . . . 17
*Emmanuel Ravelli, Gaël Richard, Laurent Daudet*

[TP1-16] **Spectral Band Replication Tool for Very Low Delay Audio Coding Applications** . 17
*Tobias Friedrich, Gerald Schuller*

[TP1-17] **Methods for 2nd Order Spherical Harmonic Spatial Encoding in Digital Waveguide Mesh Virtual Acoustic Simulations** . . . 17
*Alex Southern, Damian Murphy*

[TP1-18] **Solo Voice Detection via Optimal Cancellation** . . . 18
*Christine Smit, Daniel P. W. Ellis*

[TP1-19] **Fast Sequential LS Estimation for Sinusoidal Modeling and Decomposition of Audio Signals** . . . 18
*Bertrand David, Roland Badeau*

[TP1-20] **Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices** . . . 18
*Takeshi Saitou, Masataka Goto, Masashi Unoki, Masato Akagi*

[TP1-21] **Convolutional Synthesis of Wind Instruments** . . . 19
*Tamara Smyth, Jonathan S. Abel*

## Lunch (October 23, 12:20-15:40, West Dining Room)

## Break (October 23, 15:40-16:00, Conference House)

## Lecture TL2 Speech and Audio Coding and Hearing Aid

(October 23, 16:00-18:00, Conference House)

Chair: *Thomas F. Quatieri*

16:00-16:20

[TL2-1] **Comparison of Reduced-Bandwidth MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids** . . . 20
*Simon Doclo, Tim van den Bogaert, Jan Wouters, Marc Moonen*

16:20-16:40

[TL2-2] **Distributed Spatial Audio Coding in Wireless Hearing Aids** . . . 20
*Olivier Roy, Martin Vetterli*

16:40-17:00

[TL2-3] **A Time-Frequency Modulation Model of Speech Quality** . . . 20
*James M. Kates, Kathryn H. Arehart*

17:00-17:20

[TL2-4] **Low Delay Filterbanks for Enhanced Low Delay Audio Coding** . . . . . 21
*Markus Schnell, Ralf Geiger, Markus Schmidt, Markus Multrus, Michael Mellar, Jürgen Herre, Gerald Schuller*
17:20-17:40
[TL2-5] **Lossless Audio Coding with Bandwidth Extension Layers** . . . . . 21
*Stephen Voran*
17:40-18:00
[TL2-6] **Rate Distribution between Model and Signal** . . . . . 21
*W. Bastiaan Kleijn, Alexey Ozerov*

## Dinner (October 23, 18:00-20:00, West Dining Room)

## Demo Session 2 (October 23, 20:00-22:00, West Dining Room)

# Wednesday, October 24

## Lecture WL1 Music and Signal Analysis and Synthesis

(October 24, 8:00-10:00, Conference House)
Chair: *Simon Godsill*


8:00-8:20

[WL1-1] **Sinewave Analysis/Synthesis Based on the Fan-Chirp Transform** . . . . . 22
*Robert Dunn, Thomas F. Quatieri*

8:20-8:40

[WL1-2] **Spectral Refinement and its Application to Fundamental Frequency Estimation** . 22
*Mohamed Krini, Gerhard Schmidt*

8:40-9:00

[WL1-3] **A Novel Method for Decomposition of Multicomponent Nonstationary Signals** . . 22
*A. Goli, D. M. McNamara, A. K. Ziarani*

9:00-9:20

[WL1-4] **Using Stereo Information for Instrument Identification in Polyphonic Mixtures** . 23
*David Sodoyer, Pierre Leveau, Laurent Daudet*

9:20-9:40

[WL1-5] **Bauer Method of MVDR Spectral Factorization for Pitch Modification in the Source Domain** . . . . . 23
*M. Ravi Shanker, R. Muralishankar, A. G. Ramakrishnan*

9:40-10:00

[WL1-6] **Waveguide Modeling of Lossy Flared Acoustic Pipes: Derivation of a Kelly-Lochbaum Structure for Real-Time Simulations** . . . . . 23
*Thomas Hélie, Rémi Mignot, Denis Matignon*


## Break (October 24, 10:00-10:20, West Dining Room)

## Poster WP1

(October 24, 10:20-12:20, West Dining Room)
Chair: *Jingdong Chen*


[WP1-01] **Sound Source Distance Learning Based on Binaural Signals** . . . . . 24
*Sampo Vesa*

[WP1-02] **EM Localization and Separation using Interaural Level and Phase Cues** . . . . 24
*Michael I. Mandel, Daniel P. W. Ellis*

[WP1-03] **Single Channel Speech and Background Segregation Through Harmonic-Temporal Clustering** . . . . . 24
*Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, Alain de Cheveigné, Shigeki Sagayama*

[WP1-04] **Joint Iterative Multi-Speaker Identification and Source Separation using Expectation Propagation** . . . . . 25
*John MacLaren Walsh, Youngmoo E. Kim, Travis M. Doll*

[WP1-05] **Audio Source Separation with Matching Pursuit and Content-Adaptive Dictionaries (MP-CAD)** . . . . . 25
*Namgook Cho, Yu Shiu, C.-C. Jay Kuo*

[WP1-06] **Post-Filter Design for Superdirective Beamformers with Closely Spaced Microphones** . . . . . 25
*Heinrich W. Löllmann, Peter Vary*

[WP1-07] **A Fast Microphone Array SRP-PHAT Source Location Implementation using Coarse-To-Fine Region Contraction (CFRC)** . . . . . 25
*Hoang Do, Harvey F. Silverman*

[WP1-08] **Importance of Energy and Spectral Features in Gaussian Source Model for Speech Dereverberation** . . . 26
*Tomohiro Nakatani, Biing-Hwang Juang, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi*

[WP1-09] **A Variable Step-Size for Frequency-Domain Acoustic Echo Cancellation** . . . 26
*Yin Zhou, Xiaodong Li*

[WP1-10] **A Novel Approach to Active Noise Control Based on Wave Domain Adaptive Filtering** . . . 26
*P. Peretti, S. Cecchi, L. Palestini, F. Piazza*

[WP1-11] **Semantic Colouration Space Investigation: Controlled Colouration in the Bark-Sone Domain** . . . 27
*Jimi Y. C. Wen, Patrick A. Naylor*

[WP1-12] **Robustness Analysis of Binaural Hearing Aid Beamformer Algorithms by Means of Objective Perceptual Quality Measures** . . . 27
*Thomas Rohdenburg, Volker Hohmann, Birger Kollmeier*

[WP1-13] **Privacy-Preserving Musical Database Matching** . . . 27
*Madhusudana Shashanka, Paris Smaragdis*

[WP1-14] **A Multichannel Linear Prediction Method for the MPEG-4 ALS Compliant Encoder** . . . 27
*Yutaka Kamamoto, Noboru Harada, Takehiro Moriya*

[WP1-15] **Enhanced Resampling for Sinusoidal Modeling Parameters** . . . 28
*Martin Raspaud, Sylvain Marchand*

[WP1-16] **Compressive Coding of Stereo Audio Signals Extracting Sparseness Among Sound Sources with Independent Component Analysis** . . . 28
*Shigeki Miyabe, Tadashi Mihashi, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Toshiyuki Nomura*

[WP1-17] **Distortion-Aware Query-By-Example for Environmental Sounds** . . . 28
*Gordon Wichern, Jiachen Xue, Harvey Thornburg, Andreas Spanias*

[WP1-18] **Multi-Object Tracking of Sinusoidal Components in Audio with the Gaussian Mixture Probability Hypothesis Density Filter** . . . 29
*Daniel Clark, Ali-Taylan Cemgil, Paul Peeling, Simon Godsill*

[WP1-19] **Separation of Harmonic and Speech Signals using Sinusoidal Modeling** . . . 29
*Peter Jančovič, Münevver Köküer*

[WP1-20] **An Instrument Timbre Model for Computer Aided Orchestration** . . . 29
*Damien Tardieu, Xavier Rodet*

Lunch (October 24, 12:20-14:00, West Dining Room)

# Keynote Address1 **Simon Haykin (McMaster University)**

## Monday, October 22 8:00-9:00, Conference House

Chair: *Shoji Makino*

[Keynote1] 8:00-9:00

**Coherent ICA: Implications for Auditory Signal Processing**

*Simon Haykin, Kevin Kan (McMaster University)*

In this paper, we describe a novel algorithm, called Coherent Independent Components Analysis, and referred to as Coherent ICA for short. The algorithm, rooted in information-theoretic learning, exploits the combined use of the Infomax and Imax principles. Experimental results, based on the auditory coding of natural sounds, are presented that demonstrate the ability of coherent ICA to extract the envelope of amplitude-modulated sounds in a manner similar to the behaviour of neurons in the cochlear nucleus and inferior colliculus.

**Biography** Simon Haykin received his B.Sc (First Class Honours), Ph.D. and D.Sc, all in Electrical Engineering at the University of Birmingham in England.

Presently he is a Distinguished Professor in the Department of Electrical and Computer Engineering at McMaster University, Canada. He is a Fellow of the IEEE and a Fellow of the Royal Society of Canada. He is the recipient of an Honourary Doctorate of Technical Sciences from ETH, Zurich, Switzerland, and the first recipient of the Henry Booker Gold Medal from URSI, as well as many other prizes and awards. He is the author/coauthor of many books, including the classic books: Adaptive Filter Theory (Prentice Hall), Neural Networks (Prentice Hall), and Communication Systems (Wiley)

His current research interests are focused on Cognitive Dynamic Systems with particular emphasis on the following : 1. The design of a new generation of adaptive hearing system for the hearing impaired ( encompassing a cocktail part processor and neorocompesator), and the modeling of human communication in a noisy background, 2. Nonlinear filtering for state estimation. 3. Cognitive radar networks involving the use of inexpensive radar sensors.. 4. Robust algorithms for transmit power control and spectrum management in cognitive radio.

# Lecture ML1 **Microphone Array Signal Processing**

## Monday, October 22 9:00-10:00, Conference House

Chair: *Gary W. Elko*

[ML1-1] 9:00-9:20

**Enhanced Microphone-Array Beamforming Based on Frequency-Domain Spatial Analysis-Synthesis**

*Michael M. Goodwin (Creative ATC)*

Distant-talking hands-free communication is hindered by reverberation and interference from unwanted sound sources. Microphone arrays have been used to improve speech reception in adverse environments, but small arrays based on linear processing such as delay-sum beamforming allow for only limited improvement due to low directionality and high-level sidelobes. In this paper, we propose a beamforming system that improves the spatial selectivity by forming multiple steered beams and carrying out a spatial analysis of the acoustic scene. The analysis derives a time-frequency mask that, when applied to a reference look-direction beam, enhances target sources and improves rejection of interferers that are outside of the specified target region. The performance of the system is demonstrated by simulations and audio examples.

[ML1-2] 9:20-9:40

**Real Time Capture of Audio Images and Their Use with Video**

*Adam O'Donovan, Ramani Duraiswami, Nail A. Gumerov (Perceptual Interfaces and Reality Laboratory, Computer Science & Umiacs, University of Maryland)*

Spherical microphone arrays provide an ability to compute the acoustical intensity corresponding to different spatial directions in a given frame of audio-data. These intensities may be exhibited as an image and these images updated at a high frame rate to achieve a video image if the data capture and intensity computations can be performed sufficiently quickly, there by creating a frame-rate audio camera. We describe how such a camera can be built and the processing done sufficiently quickly using graphics processors. The joint processing of and captured frame-rate audio and video images enables applications such as visual identification of noise sources, beamforming and noise-suppression in video conferenceing and others, provided it is possible to account for the spatial differences in the location of the audio and the video cameras. Based on the recognition that the spherical array can be viewed as a central projection camera it is possible to perform such joint analysis. We provide several examples of real-time applications.

[ML1-3] 9:40-10:00

**Subband Method for Multichannel Least Squares Equalization of Room Transfer Functions**

*Nikolay D. Gaubitch, Mark R. P. Thomas, Patrick A. Naylor (Imperial College London)*

Equalization of room transfer functions (RTFs) is important in many speech and audio processing applications. It is a challenging problem because RTFs are several thousand taps long and non-minimum phase and in practice only approximate measurements of the RTFs are available. In this paper, we present a subband multichannel least squares method for equalization of RTFs which is computationally efficient and less sensitive to inaccuracies in the measured RTFs compared to its fullband counterpart. Experimental results using simulations and real measurements demonstrate the performance of the algorithm.

# Poster MP1

## Monday, October 22 10:20-12:20, West Dining Room

Chair: *W. Bastiaan Kleijn*

[MP1-01]

**Broadband Music: Opportunities and Challenges for Multiple Source Localization**

*Jacek P. Dmochowski, Jacob Benesty, Sofiène Affes (INRS-EMT, Universite du Quebec)*

It is well-known that the subspace MUltiple SIgnal Classification (MUSIC) method provides high-resolution spatial spectral estimates in narrowband signal environments. However, for broadband signals, such high-resolution methods still elude researchers. This paper proposes a broadband version of the MUSIC method using a parameterized version of the spatial correlation matrix. The proposed algorithm utilizes both inter-microphone amplitude and phase differences; as a result, simulation results show that the proposed method allows two broadband sources which are not resolvable by conventional steered beamforming to be accurately resolved.

[MP1-02]

**Energy-Based Position Estimation of Microphones and Speakers for ad hoc Microphone Arrays**

*Minghua Chen, Zicheng Liu, Li-Wei He, Phil Chou, Zhengyou Zhang (Microsoft Research)*

We present a new algorithm, named $\delta$-majorization, to estimate the positions of microphones and speakers in an ad hoc microphone array setting. This new algorithm is an extension to our previous energy-based position estimation algorithm.

[MP1-03]

**Linear Regression on Sparse Features for Single-Channel Speech Separation**

*Mikkel N. Schmidt, Rasmus K. Olsson (Technical University of Denmark)*

In this work we address the problem of separating multiple speakers from a single microphone recording. We formulate a linear regression model for estimating each speaker based on features derived from the mixture. The employed feature representation is a sparse, non-negative encoding of the speech mixture in terms of pre-learned speaker-dependent dictionaries. Previous work has shown that this feature representation by itself provides some degree of separation. We show that the performance is significantly proved when regression analysis is performed on the sparse, non-negative features, both compared to linear regression on spectral features and compared to separation based directly on the non-negative sparse features.

[MP1-04]

**Sound Source Separation using Null-Beamforming and Spectral Subtraction for Mobile Devices**

*Shintaro Takada, Satoshi Kanba, Tetsuji Ogawa, Kenzo Akagiri, Tetsunori Kobayashi (Department of Computer Science, Waseda University)*

This paper presents a new type of speech segregation method for mobile devices in noisy sound situation, where two or more speakers are talking simultaneously. The proposed method consists of multiple null-beamformers, their minimum power channel selection and spectral subtraction. The proposed method is performed with space-saving and coplanar microphone arrangements and lowcost calculations, which are the very important requirements for the mobile application. Effectiveness of the proposed method is clarified in the segregation and the recognition experiments of two simultaneous

continuous speeches: the method improved the PESQbased MOS value by about one point and reduced 70% of word recognition errors compared with non-processing.

[MP1-05]

**On Dealing with Sampling Rate Mismatches in Blind Source Separation and Acoustic Echo Cancellation**

*Enrique Robledo-Arnuncio, Ted S. Wada, Biing-Hwang (Fred) Juang (Georgia Institute of Technology)*

The lack of a common clock reference is a fundamental problem when dealing with audio streams originating from or heading to different distributed sound capture or playback devices. When implementing multichannel signal processing algorithms for such kind of audio streams it is necessary to account for the unavoidable mismatches between the actual sampling rates. There are some approaches that can help to correct these mismatches, but important problems remain to be solved, among them the accurate estimation of the mismatch factors, and achieving both accuracy and computational efficiency in their correction. In this paper we present an empirical study on the performance of blind source separation and acoustic echo cancellation algorithms in this scenario. We also analyze the degradation in performance when using an approximate but efficient method to correct the rate mismatches.

[MP1-06]

**Signal Deflation and Paraunitary Constraints in Spatio-Temporal FastICA-Based Convolutive Blind Source Separation of Speech Mixtures**

*Malay Gupta, Scott C. Douglas (Southern Methodist University)*

The FastICA algorithm of Hyvarinen and Oja is a popular procedure for blind source separation of non-convolutive signal mixtures. Recently, two different extensions of this procedure have been proposed for convolutive blind source separation of speech and other signal mixtures. In this paper, we describe the major differences and compare the performances of these approaches, illustrating how signal deflation or coefficient orthogonalization is employed to maintain uniqueness of the separated system outputs for both synthetic convolutive mixtures and speech mixtures as recorded in real-room environments. Our numerical evaluations indicate that (a) coefficient orthogonality through paraunitary constraints provide more robust estimation behavior that least-squares signal deflation with unit-norm constraints, and (b) all-pass constraints can be used to improve performance when only signal deflation is employed.

[MP1-07]

**Fast Convergence Blind Source Separation Based on Frequency Subband Interpolation by Null Beamforming**

*Keiichi Osako, Yoshimitsu Mori, Yu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano (Graduate School of Information Science, Nara Institute of Science and Technology)*

We propose a new algorithm for blind source separation (BSS) approach that combines independent component analysis (ICA) and frequency subband beamforming interpolation. The slow convergence of the optimization of the separation filters is a problem in ICA. Our approach to resolve this problem is based on the relationship between ICA and null beamforming (NBF). The proposed method consists of the following three parts: step I a frequency subband selector part for learning ICA, step II a frequency domain ICA part with direction-of-arrivals (DOA) estimation of sound sources, and step III an interpolation part using null beamforming constructed with the estimated DOA. The results of the signal separation experiments under reverberant condition reveal that the convergence speed is superior to that of the conventional ICA based BSS methods.

[MP1-08]

**Electronic Pop Protection for Microphones**

*Gary W. Elko, Jens Meyer (mh acoustics LLC), Steven Backer, Jürgen Peissig (Sennheiser Research Lab)*

Pop noise caused by plosives generated by talkers, singers and other vocalists has long been a topic of interest to microphone manufacturers. The articulation of speech plosives (oral-stop consonants such as p, t, and k) can severely degrade the quality of a recording or performance. This is especially true for pressure-differential microphones, where not only is the unwanted artifact heard, but also there is the high potential for either microphone or associated electronics overload. Traditional wind/pop shields made of foam or a stretched fabric have been in use for many years, and are adequate for most common applications. However, it may be desired or advantageous to use an electronic version of pop protection under conditions where a pop or wind screen is not practical or sufficient, such as when using a lapel or podium microphone that is designed to be visually unobtrusive. The results presented in this paper describe a relatively simple procedure to reduce microphone sensitivity to speech plosives via physically-informed signal processing methods, or Electronic Pop Protection (EPP). The EPP system itself comprises two parts: 1) detection of the presence of pop (turbulent air-jet flow) in the signal, and 2) suppression of this transient noise. A DSP-based demonstrator was created to illustrate and measure the operation of the algorithm on actual hardware. Finally, an objective evaluation of the algorithm is presented.

[MP1-09]

**A Practical Multichannel Dereverberation Algorithm using Multichannel DYPSA and Spatiotemporal Averaging**

*Mark R. P. Thomas, Nikolay D. Gaubitch, Jon Gudnason, Patrick A. Naylor (Imperial College London)*

Speech signals for hands-free telecommunication applications are received by one or more microphones placed at some distance from the talker. In an office environment, for example, unwanted signals such as reverberation and background noise from computers and other talkers will degrade the quality of the received signal. These unwanted components have an adverse effect upon speech processing algorithms and impair intelligibility. This paper demonstrates the use of the Multichannel DYPSA algorithm to identify glottal closure instants (GCIs) from noisy, reverberant speech. Using the estimated GCIs, a spatiotemporal averaging technique is applied to attenuate the unwanted components. Experiments with a microphone array demonstrate the dereverberation and noise suppression of the spatiotemporal averaging method, showing up to a 5 dB improvement in segmental SNR and 0.33 in normalized Bark spectral distortion score.

[MP1-10]

**Isotropic Noise Suppression in the Power Spectrum Domain by Symmetric Microphone Arrays**

*Hikaru Shimizu, Nobutaka Ono, Kyosuke Matsumoto, Shigeki Sagayama (Graduate School of Information Science and Technology, The University of Tokyo)*

In this paper, we propose a new framework of array processing for suppressing isotropic noise on the power spectrum domain. As a theoretical basis, we discuss the characteristics of the isotropic noise covariance matrix and show that it can be diagonalized by a definite unitary matrix when the microphone array has a kind of symmetry. Based on it, our method estimates the power spectrum of the target source at a looking direction from the non-diagonal components of the transformed covariance matrix by ML (Maximum Likelihood) method. We also show some experimental results in both of stationary noise field and nonstationary noise field by simulation.

[MP1-11]

**Acoustic Echo Cancelation for Dynamically Steered Microphone Array Systems**

*Matti Hämäläinen (Nokia Research Center), Ville Myllylä (Nokia)*

A new method for integrating dynamically steered beamforming filters and acoustic echo cancellation is presented. The proposed method enables beam steering independent AEC processing without allocation of parallel AEC filters for each microphone input (a.k.a. AEC first configuration). Especially for larger microphone arrays the proposed method can provide significant computational savings with comparable performance to AEC first configuration.

[MP1-12]

**A New Approach to Digital Audio Equalization**

*S. Cecchi, L. Palestini, E. Moretti, F. Piazza (Universita Politecnica Delle Marche)*

The design and implementation of a M-band linear phase digital audio equalizer system is presented. Beginning from analysis/ synthesis filterbank, an innovative uniform and non uniform bands audio equalizer is derived using multirate properties. In literature fixed frequency response equalization has well-known problems due to algorithms implementation. The idea of this work derives from different techniques employed in filter banks to avoid aliasing in the case of adaptive filtering in each band. The effectiveness of the algorithm is shown comparing it with a simple filterbank and with an octave band equalizer based on frequency domain technique. The solution presented here has several advantages: low computational complexity, low delay and uniform frequency response avoiding ripple between adjacent bands.

[MP1-13]

**Implementation of Directional Sources in Wave Field Synthesis**

*Jens Ahrens, Sascha Spors (Deutsche Telekom Laboratories, Berlin University of Technology)*

Wave field synthesis (WFS) is a spatial audio reproduction technique aiming at physically synthesizing a desired sound field. Typically, virtual sound sources are rendered as emitting spherical or plane waves. In this paper we present an approach to the implementation of sources with arbitrary directivity. The approach is based on the description of the directional properties of a source by a set of circular harmonics. A time domain expression of the loudspeaker driving signals is derived allowing an efficient implementation. Consequences of sampling and truncation of the secondary source distribution as occurring in typical installations of WFS systems are discussed and simulated reproduction results are shown.

[MP1-14]

**A Comparsion of Acoustic and Psychoacoustic Measurements of Pass-Through Hearing Protection Devices**

*Douglas S. Brungart, Brian W. Hobbs (Air Force Research Laboratory), James T. Hamil (General Dynamics)*

In environments where listeners need to detect low-level sounds while being protected from high-level noises, electronic pass-through hearing protectors (EPHPs) offer an appealing alternative to traditional passive earplugs or earmuffs. In this paper, we compare acoustic measurements of the Head-Related Transfer Functions associated with eight different EPHPs to localization results measured on human listeners with the same devices. The results are discussed in terms of the insights they can provide for the design of improved EPHP systems.

[MP1-15]

**Improvement in Detectability of Alarm Signals in Noisy Environments by Utilizing Spatial Cues**

*Hideaki Uchiyama, Masashi Unoki, Masato Akagi (School of Information Science, JAIST)*

It is important to present alarm signals for peoples to be accurately perceived in real environments to avoid many dangerous situations. Because noises in real environments mask alarm signals so that nobody can detect it. In this paper, detection ability of alarm signals in a car noise were measured as a function of interaural time difference (ITD) and interaural phase difference (IPD). Pulse train signals and five alarm signals were used to confirm whether SRM occurs for these in the presence of realistic noise. Results showed that SRM occurs for all signals and detection ability of alarm signals can be improved not only by ITD but also by IPD of the signal. This influence was depended upon the relation between ITD and IPD related to the component frequency. In addition, ITD and IPD of the arrival direction difference of the alarm signal in the masker greatly influences occurrence of SRM by interpreting binaural masking level difference (BMLD). These suggest that spatial cues (ITD and IPD) of the arrival direction of alarm signal in comparison to masker direction have to be considered to convey warnings accurately and efficiently without loss of information.

[MP1-16]

**Estimation Model for the Speech-Quality Dimension ”Directness / Frequency Content”**

*Lu Huo (Institute for Circuit and System Theory, Faculty of Engineering, University of Kiel.), Marcel Wältermann (Deutsche Telekom Laboratories, TU Berlin, Berlin, Germany), Kirstin Scholz (Institute for Circuit and System Theory, Faculty of Engineering, University of Kiel.), Alexander Raake (Deutsche Telekom Laboratories, TU Berlin, Berlin, Germany), Ulrich Heute (Institute for Circuit and System Theory, Faculty of Engineering, University of Kiel.), Sebastian Möller (Deutsche Telekom Laboratories, TU Berlin, Berlin, Germany)*

In this paper, an instrumental method for estimating the quality-relevant dimension ”directness / frequency content” (DF) is presented. Apart from the perceptual dimensions ”continuity” and ”noisiness”, DF has been identified to be of crucial importance for the listening-quality of today’s telecommunication networks, especially if it comes to linear distortions. The presented work is part of a framework for a signal-based approach to measure the quality of transmitted speech on the basis of perceptual dimensions[2].

[MP1-17]

**Probabilistic Model Based Similarity Measures for Audio Query-By-Example**

*Tuomas Virtanen, Marko Helén (Tampere University of Technology)*

This paper proposes measures for estimating the similarity of two audio signals, the objective being in query-by-example. Both signals are first represented using a set of features calculated in short intervals, and then probabilistic models are estimated for the feature distributions. Gaussian mixture models and hidden Markov models are tested in this study. The similarity of the signals is measured by the congruence between the feature distributions or by a cross-likelihood ratio test. We calculate the Kullback-Leibler divergence between the distributions by sampling the distributions at the points of the obsevations vectors. The cross-likelihood ratio test is evaluated using the likelihood of the first signal being generated by the model of the second signal, and vice versa. Simulations were conducted to test the accuracy of the proposed methods on query-by-example of audio. On a database consisting of of speech, music, and environmental sounds the proposed methods enable better retrieval accuracy than the existing methods.

[MP1-18]

**Improving Generalization for Classification-Based Polyphonic Piano Transcription**

*Graham E. Poliner, Daniel P. W. Ellis (LabROSA, Columbia University)*

In this paper, we present methods to improve the generalization capabilities of a classification-based approach to polyphonic piano transcription. Support vector machines trained on spectral features are used to classify frame-level note instances, and the independent classifications are temporally constrained via hidden Markov model post-processing. Semi-supervised learning and multiconditioning are investigated, and transcription results are reported for a compiled set of piano recordings. A reduction in frame-level transcription error score of 10% was achieved by combining multiconditioning and semi-supervised classification.

[MP1-19]

**Acoustic Signal Processing for Degradation Analysis of Rotating Machinery to Determine the Remaining Useful Life**

*Patricia Scanlon, Alan M. Lyons, Alan O'Loughlin (Alcatel-Lucent - Bell Laboratories)*

Automated Condition Monitoring of machines typically involves the detection and diagnosis of developing defects. However increased demand for reliability requires these systems to also predict the Remaining Useful life (RUL) of the machine in order to schedule timely maintenance and reduce machine downtime by preventing catastrophic faults. In this paper, an automated approach to degradation analysis is proposed that uses the acoustic noise signal from a rotating machine to determine the RUL.We incorporate a novel approach to Feature Subset Selection to extract relevant features for classification. This method is applied to the high dimensionality multivariate time series data extracted from the acoustic data acquired over the lifetime of the fan to reduce confounds from redundant features as well as improve computational efficiency. Our approach requires no a-priori information regarding the spectral location of defects or class label boundaries of the training data. Using such an approach, the RUL of the machine was determined with an accuracy of 98.7%.

[MP1-20]

**Single-Frame Discrimination of Non-Stationary Sinusoids**

*Jeremy J. Wells, Damian T. Murphy (University of York)*

This paper introduces two new methods for discrimination of non-stationary sinusoids within a single analysis frame. Both methods use data from frequency and time reassignment of Fourier transform data. These methods are then compared with an adapted version of an existing discrimination method, both in terms of their effectiveness and their computational cost.

# Lecture ML2 Source Localization and Blind Source Separation

## Monday, October 22 16:00-18:00, Conference House

Chair: *Scott Douglas*

[ML2-1] 16:00-16:20

**Modeling of Motion Dynamics and its Influence on the Performance of a Particle Filter for Acoustic Speaker Tracking**

*Eric A. Lehmann, Anders M. Johansson, Sven Nordholm (Western Australian Telecommunications Research Institute)*

Methods for acoustic speaker tracking attempt to localize and track the position of a sound source in a reverberant environment using the data received at an array of microphones. This problem has received significant attention over the last few years, with methods based on a particle filtering principle perhaps representing one of the most promising approaches. As a Bayesian filtering technique, a particle filter relies on the definition of two main concepts, namely the measurement process and the transition equation (target dynamics). Whereas a significant research effort has been devoted to the development of improved measurement processes, the influence of the dynamics formulation on the resulting tracking accuracy has received little attention so far. This paper provides an insight into the dynamics modeling aspect of particle filter design. Several types of motion models are considered, and the performance of the resulting particle filters is then assessed with extensive experimental simulations using real audio data recorded in a reverberant environment. This paper demonstrates that the ability to achieve a reduced tracking error relies on both the chosen model as well as the specific optimization of its parameters.

[ML2-2] 16:20-16:40

**Multi Target Acoustic Source Tracking using Track Before Detect**

*Maurice Fallon, Simon Godsill (University of Cambridge)*

Particle Filter-based Source Localisation algorithms attempt to track the position of a sound source - a person speaking in a room - based on the current data from a distributed microphone array as well as all previous data up to that point. This paper introduces a multitarget methodology for acoustic source tracking. The methodology is based upon the Track Before Detect (TBD) framework. The algorithm also implicitly evaluates the source activity using a variable appended to the state vector. Examples of typical tracking performance are given using a set of real speech recordings with two sources active simultaneously.

[ML2-3] 16:40-17:00

**Blind Sparse-Nonnegative (BSN) Channel Identification for Acoustic Time-Difference-Of-Arrival Estimation**

*Yuanqing Lin (GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania), Jingdong Chen (Bell Laboratories, Lucent Technologies), Youngmoo Kim (Department of Electrical and Computer Engineering, Drexel University), Daniel D. Lee (GRASP Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania)*

Estimating time-difference-of-arrival (TDOA) remains as a challenging task when acoustic environments are reverberant and noisy. Blind channel identification approaches for TDOA estimation explicitly model multipath reflections and have been demonstrated to be effective in dealing with reverberation. Unfortunately, the existing blind channel identification algorithms are sensitive to ambient noise. This paper proposes to resolve the noise sensitivity issue by exploiting the prior knowledge about an acoustic room impulse response (RIR), that is, an acoustic RIR can be modeled

by a sparse-nonnegative FIR filter. This paper shows how to formulate a single-input two-output blind channel identification into a least square (LS) convex optimization, and how to incorporate the sparsity and nonnegativity priors so that the resulting optimization is still convex and can be solved efficiently. The proposed blind sparse-nonnegative (BSN) channel identification approach for TDOA estimation is not only robust to reverberation, but also robust to ambient noise, as demonstrated by simulations and experiments in real acoustic environments.

[ML2-4] 17:00-17:20

**Blind Criterion and Oracle Bound for Instantaneous Audio Source Separation using Adaptive Time-Frequency Representations**

*Emmanuel Vincent, Rémi Gribonval (IRISA-INRIA)*

The separation of multichannel audio mixtures is often addressed by the masking approach, which consists of representing the mixture signal in the time-frequency domain and associating each time-frequency bin with a small number of active sources. Adaptive time-frequency representations appear promising compared to usual fixed representations since they can increase the disjointness of the sources. However their use has not been conclusive so far. In this paper, we propose a new criterion for the blind estimation of an adapted representation and explain how to compute the oracle representation leading to the best possible performance given reference source signals. Experimental results suggest that a small separation performance improvement can indeed be achieved using adaptive representations, but that complementary approaches must be investigated to obtain larger improvements.

[ML2-5] 17:20-17:40

**Monaural Speech Separation using Source-Adapted Models**

*Ron J. Weiss, Daniel P. W. Ellis (LabROSA, Columbia University)*

We propose a model-based source separation system for use on single channel speech mixtures where the precise source characteristics are not known *a priori*. We do this by representing the space of source variation with a parametric signal model based on the eigenvoice technique for rapid speaker adaptation. We present an algorithm to infer the characteristics of the sources present in a mixture, allowing for significantly improved separation performance over that obtained using unadapted source models. The algorithm is evaluated on the task defined in the 2006 Speech Separation Challenge and compared with separation using source-dependent models.

[ML2-6] 17:40-18:00

**A Soft Masking Strategy Based on Multichannel Speech Probability Estimation for Source Separation and Robust Speech Recognition**

*Eugen Hoffmann, Dorothea Kolossa, Reinhold Orglmeister (TU Berlin)*

In this paper, we present a post processing algorithm that improves the quality of the solution of ICA-algorithms by applying a modified speech enhancement technique. The proposed method is based on estimating speech probabilities from the ICA outputs by means of two dimensional correlations. With these probabilities, a soft masking function can be applied on the ICA outputs, which results in significantly increased interferer suppression. In order to avoid negative influences on subsequent speech recognition, missing feature recognition has been applied to robustly recognize the nonlinearly processed speech signal. The algorithm has been tested on real-room speech mixtures with a reverberation time of 300ms, where an SIR-improvement of up to 32dB has been obtained, which was 10dB above ICA performance for the same dataset.

# Keynote Address2 **Albert S. Bregman (McGill University)**

## Tuesday, October 23 8:00-9:00, Conference House

Chair: *Daniel P.W. Ellis*

[Keynote2] 8:00-9:00

**Progress in the Study of Auditory Scene Analysis**

*Albert S. Bregman (Department of Psichology, McGill University)*

The early research on auditory scene analysis (ASA) - the subject of my talk at the corresponding IEEE workshop in 1995 - has been followed by many exciting studies that have opened up new directions of research. A number of them will be discussed under the following headings: (1) What is the role of attention in ASA? (2) What have we learned by using evoked potentials to study ASA? (3) To what extent has research on human babies and on non-human animals supported the idea that primitive ASA is ”wired into” the brain? (4) What is the physiological basis of ASA? (5) How is ”binding” carried out in the brain?

**Biography** Al Bregman was born in Toronto. He studied at the universities of Toronto (philosophy) and Yale (psychology), followed by a 3-year postdoctoral period doing research at the Center for Cognitive Studies and teaching in the Psychology Department at Harvard. He came to McGill in 1965 as its first professor in the newly developing area of cognitive psychology and taught this subject to a generation of students, many of whom subsequently became well-known in this field. He also taught auditory perception, experimental methods and conceptual issues in psychology. Although he studied memory for his Ph.D. (1963) at Yale and for a few years afterward, his main research has been on the perceptual organization of sound, and he has 75 publications in this area. These have described the cues for, and the perceptual results of auditory organization, and its role in the perception of speech, music and other sounds. His book, Auditory Scene Analysis (1990), defined and named this field and integrated a large number of phenomena under a set of principles that owe much to Gestalt psychology and to artificial intelligence. A later audio CD gave many examples of sound patterns that illustrate the operation of these principles. Currently he is an Emeritus Professor of Psychology at McGill University. He was appointed Fellow of the Canadian and American Psychological Associations and of the Royal Society of Canada, and held a 2-year Killam Research Fellow. He was also awarded the Jacques Rousseau medal for interdisciplinary contributions by the French Canadian Association for the Advancement of Science.

Further information can be obtained at http://www.psych.mcgill.ca/labs/auditory/Home.html

# Lecture TL1 **Signal Enhancement**

## Tuesday, October 23 9:00-10:00, Conference House

Chair: *Patric A. Naylor*

[TL1-1] 9:00-9:20

**Single-Channel Impact Noise Suppression with No Auxiliary Information for its Detection**

*Akihiko Sugiyama (NEC Corporation)*

This paper proposes impact-noise suppression with no auxiliary information for keystroke/thrum detection. Impact noise is successfully detected by the shape and the change of the power spectrum. The impact-noise components are either replaced by an ambient-noise estimate in nonspeech sections or substracted using an estimated impact-noise estimate in speech sections. Subjective evaluation results demonstrate that the proposed impact-noise suppression provides as much as 1.1 higher score in the 5-grade MOS than an ordinary noise suppressor and noisy speech with statistically significant difference. Because it is designed as a postprocessor, it is easily combined with a conventional noise suppressor.

[TL1-2] 9:20-9:40

**Aliasing Reduction for Modified Discrete Cosine Transform Domain Filtering and its Application to Speech Enhancement**

*Fabian Kuech (Fraunhofer Institute for Integrated Circuits IIS), Bernd Edler (Laboratorium fuer Informationstechnologie)*

Efficient combinations of coding and manipulation of audio signals in the spectral domain are often desirable in communication systems. The modified discrete cosine transform (MDCT) represents a popular spectral transform in audio coding as it leads to compact signal representations. However, as the MDCT corresponds to a critically sampled filter bank, it is in general not appropriate to directly apply it to filtering tasks. In this paper we present a method to compensate for aliasing terms that arise from such direct MDCT domain filtering. The discussion is thereby based on a rigorous matrix representation of critically sampled filter banks which also leads to corresponding efficient realizations. As an application showcase, noise reduction for MDCT based speech coding is considered in simulations.

[TL1-3] 9:40-10:00

**Example-Driven Bandwidth Expansion**

*Paris Smaragdis, Bhiksha Raj (Mitsubishi Electric Research Labs)*

In this paper we present an example-driven algorithm that allows the recovery of lost spectral components of band-limited signals. We present a generative spectral model which allows the extraction of salient information from audio snippets, and then apply this information to enhance the bandwidth of band-limited signals. We demonstrate the methodology that we use and present various examples and comparisons with other approaches.

# Poster TP1

## Tuesday, October 23 10:20-12:20, West Dining Room

Chair: *Paris Smaragdis*

[TP1-01]

**A Two-Stage Frequency-Domain Blind Source Separation Method for Underdetermined Convolutive Mixtures**

*Hiroshi Sawada, Shoko Araki, Shoji Makino (NTT Corporation)*

This paper proposes a two-stage method for the blind separation of convolutively mixed sources. We employ time-frequency masking, which can be applied even to an underdetermined case where the number of sensors is insufficient for the number of sources. In the first stage of the method, frequency bin-wise mixtures are classified based on Gaussian mixture model fitting. In the second stage, the permutation ambiguities of the bin-wise classified signals are aligned by clustering the posterior probability sequences calculated in the first stage. Experimental results for separating four speeches with three microphones in a reverberant condition show the superiority of the proposed method over existing methods based on time-difference-of-arrival estimations or signal envelope clustering.

[TP1-02]

**Long-Term Gain Estimation in Model-Based Single Channel Speech Separation**

*M. H. Radfar, R. M. Dansereau (The Department of Systems and Computer Engineering, Carleton University)*

Model-based single channel speech separation techniques commonly use trained patterns of the individual speakers to separate the speech signals. In most recent proposed techniques, it is assumed that data used in the train and test phase have the same level of energy, a prerequisite which is hardly met in the real situations. Considering this limitation, we propose a technique which estimates the gain associated with the individual speakers from the mixture and thus obviate the need for this assumption. The basic idea is to express the probability density function (PDF) of the mixture in terms of the individual speakersPDFs and corresponding gains. Then, those patterns and gains which maximize the mixtures PDF are selected and used to recover the speech signals. Experimental results conducted on a wide variety of mixtures with signal-to-signal ratios ranging from 0 to 18 dB show that the proposed technique estimates the speakersgain with 95% accuracy within the range of the actual gain +/- 20%. Comparing the separated speech signals with the original ones in terms of SNR criterion with/without including the gain estimation stage, we observe a significant SNR improvement (on average 5.73 dB) for the gain included scenario.

[TP1-03]

**Sparseness-Based 2ch BSS using the EM Algorithm in Reverberant Environment**

*Yosuke Izumi, Nobutaka Ono, Shigeki Sagayama (Graduate School of Information Science and Technology the Univ. of Tokyo)*

In this paper, we propose a new approach to sparseness-based BSS based on the EM algorithm, which iteratively estimates the DOA and the time-frequency mask for each source through the EM algorithm under the sparseness assumption. Our method has the following characteristics: 1) it enables the introduction of physical observation models such as the diffuse sound field, because the likelihood is defined in the original signal domain and not in the feature domain, 2) one does not necessarily have to know in advance the power of the background noise since they are also parameters which can be estimated from the observed signal, 3) it takes short computational time for instance at most 1 minute to separate three sources from two mixtures with a 2.8GHz CPU machine, 4) a common objective

function is iteratively increased in localization and separation steps, which correspond to the E-step and M-step, respectively.

[TP1-04]

**Prior Structures for Time-Frequency Energy Distributions**

*Ali Taylan Cemgil, Paul Peeling, Onur Dikmen, Simon Godsill (Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge)*

We introduce a framework for probabilistic modelling of time-frequency energy distributions based on correlated Gamma and inverse Gamma random variables. One advantage of the approach is that the resulting class of models are conjugate which makes inference easier. Moreover, both positivity and additivity follow naturally in this framework. We illustrate how generic models (applicable to a broad class of signals) and more specialised models can be designed to model harmonicity, spectral continuity and/or changepoints. We show simulation results that illustrate the potential of the approach on a large spectrum of audio processing applications such as denoising, source separation and transcription.

[TP1-05]

**Fast Time-Domain Spherical Microphone Array Beamforming**

*Zhiyun Li, Ramani Duraiswami (University of Maryland at College Park)*

Capturing and reproducing 3D audio has been finding increasing potentials in a wide range of applications. One of the main technologies for that is spherical beamforming, which, however, is a very expensive algorithm, especially for interactive applications. In this paper, we propose a fast spherical beamforming algorithm in time domain. It uses pre-computed data and can be steered to arbitrary 3D directions. It is ideal for interactive audio applications such as 3D games, etc. Simulation and experimental results are demonstrated to verify our algorithm.

[TP1-06]

**Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model**

*Eric A. Lehmann, Anders M. Johansson, Sven Nordholm (Western Australian Telecommunications Research Institute)*

The image-source method has become a ubiquitous tool in many fields of acoustics and signal processing. A technique was recently proposed to predict the energy decay (energy–time curve) in room impulse responses simulated using the image-source model. The present paper demonstrates how this technique can be efficiently used to determine the enclosure's absorption coefficients in order to achieve a desired reverberation level, even with a non-uniform distribution of the sound absorption in the room. As shown in this work, classical expressions for the prediction of an enclosure's reverberation time, such as Sabine and Eyring's formulae, do not provide accurate results when used in conjunction with the image method. The proposed approach hence ensures that the image-source model effectively generates impulse responses with a proper reverberation time, which is of particular importance, for instance, for the purpose of assessing the performance of audio signal processing algorithms operating in reverberant conditions.

[TP1-07]

**Overfitting-Resistant Speech Dereverberation**

*Takuya Yoshioka, Tomohiro Nakatani, Takafumi Hikichi, Masato Miyoshi (NTT CS Labs)*

This paper proposes a method that prevents the overfitting problem inherent in the joint source-channel estimation approach to speech dereverberation. The approach has several desirable attributes such as high estimation accuracy. However, the channel-related parameters estimated with the conventional implementation of this approach often overfit the source characteristics present in observed signals. This overfitting results in unstable behavior of the dereverberation process. The problem stems from the fact that the conventional implementation employs a point estimation scheme to obtain the parameters describing the source characteristics. The proposed method marginalizes the source parameters to mitigate the overfitting problem. Two kinds of experimental results are reported, one of which was obtained in a single source situation and shows the ability to prevent the overfitting; the other is obtained in a multi-source scenario indicating the applicability of the proposed method to multi-source situations.

[TP1-08]

**Novel and Efficient Download Test for Two Path Echo Canceller**

*Mohammad Asif Iqbal, Steven L. Grant (University of Missouri-Rolla)*

The two-path echo cancellation technique is a popular method for handling the double-talk problem in acoustic and line echo cancellation applications. The method uses two filters. A so-called background adaptive filter adapts its coefficients to predict echo all or most of the time regardless of signal activity on the near-end. A second foreground filter, that also predicts the echo, receives it coefficients from the background filter, but only when the background is performing better than the foreground. Only the foreground residual echo is sent to the far-end, so any background divergence due to double-talk is not observed by the user. The key to good two-path performance is in the definition of the background-to-foreground coefficient download tests. These typically contain a suite of various measures that attempt to ascertain the convergence state of the two filters. In this paper we present a novel, simple statistic that directly and accurately estimates a filter's convergence state. With the aid of this new statistic we show significant improvement in the overall performance of the two path echo canceller.

[TP1-09]

**An Approach to Massive Multichannel Broadband Feedforward Active Noise Control using Wave-Domain Adaptive Filtering**

*Sascha Spors, Herbert Buchner (Deutsche Telekom Laboratories)*

Multichannel active noise control (ANC) systems are increasingly being applied in scenarios where an enlarged quiet zone is desired. For few channels numerous solutions to this problem have been developed in the past. However, algorithms for multichannel ANC with a high number of channels (massive multichannel ANC), in order to achieve a large quiet zone, still remain a challenge. The fundamental limitations of current adaptation algorithms in the context of massive multichannel ANC are outlined in this contribution. As a solution to these limitations, the application of the generic concept of wave-domain adaptive filtering (WDAF) is proposed for ANC. Simulation results from a 60-channel ANC system illustrate the successful application of the proposed concepts.

[TP1-10]

**Enhancement of Residual Echo for Improved Frequency-Domain Acoustic Echo Cancellation**

*Ted S. Wada, Biing-Hwang (Fred) Juang (Center for Signal and Image Processing)*

This paper explores the technique of integrating a noise suppressing nonlinearity to the adaptive filter error feedback loop of a frequency-domain acoustic echo canceler (AEC) when there is an additive noise at the near-end. It was shown in a previous study that in the time-domain case, both the echo return loss enhancement (ERLE) and the misalignment from using normalized least-mean-squares (NLMS)

are improved significantly. By applying the same technique to the frequency-domain AEC, namely the generalized multidelay filter (GMDF), the misalignment can be decreased by as much as 5 dB in a numerical simulation, whereas the ERLE can be increased by over 5 dB in a real acoustic environment. New results further support the idea that removing the effects of distortion to the cancellation error helps to improve the performance of an adaptive filter.

[TP1-11]

**Effects of Pre-Processing Filters on a Wavelet Packet-Based Algorithm to Identify Speech Transients**

*Daniel M. Rasetshwane, J. Robert Boston, Ching-Chung Li, John D. Durrant (University of Pittsburgh)*

Speech transients have been shown to be important cues for identifying speech sounds and amplification of transients can improve the intelligibility of speech in noise. We have developed a time-frequency approach to identify transients that uses a pre-processing filter, but optimal filter parameters are difficult to determine due to the large number of possibilities. This paper describes the use of the Articulation Index (AI) to evaluate the effects of different high-pass filters on our algorithm. The best filter was found to depend on signal-to-noise ratio (SNR). AI results should be interpreted with caution, but they appear to provide a reasonable approach to selecting a pre-processing filter.

[TP1-12]

**Modeling Spot Microphone Signals using the Sinusoidal Plus Noise Approach**

*Christos Tzagkarakis, Athanasios Mouchtaris, Panagiotis Tsakalides (Department of Computer Science, University of Crete and Institute of Computer Science (FORTH-ICS))*

This paper focuses on high-fidelity multichannel audio coding based on an enhanced adaptation of the well-known sinusoidal plus noise model (SNM). Sinusoids cannot be used per se for high-quality audio modeling because they do not represent all the audible information of a recording. The noise part has also to be treated to avoid an artificial sounding resynthesis of the audio signal. Generally, the encoding process needs much higher bitrates for the noise part than the sinusoidal one. Our objective is to encode spot microphone signals using the SNM, by taking advantage of the interchannel similarities to achieve low bitrates. We demonstrate that for a given multichannel audio recording, the noise part for each spot microphone signal (before the mixing stage) can be obtained by using its noise envelope to transform the noise part of just one of the signals (the so-called ”reference signal”, which is fully encoded).

[TP1-13]

**A Modified Spatio-Temporal Orthogonal Iteration Method for Multichannel Audio Signal Representation**

*Scott C. Douglas, Malay Gupta (Southern Methodist University)*

In this paper, we present a novel algorithm for a spatio-temporal extension of the well-known method of orthogonal iterations in linear algebra. This algorithm estimates an n-input, m-output ($m<n$) paraunitary filter bank from a multichannel data autocorrelation sequence to maximize the total output power of the filter bank when applied to an n-dimensional input signal. We then show how this procedure can be used to generate reduced rank signal representations of recordings of m audio sources in a room as collected by an n-channel microphone array. The importance of the method for determining the number of active sound sources in a room for convolutive blind source separation is also discussed.

[TP1-14]

**A Low-Delay Audio Coder with Constrained-Entropy Quantization**

*Minyue Li, W. Bastiaan Kleijn (Sound and Image Processing Laboratory, KTH School of Electrical Engineering)*

We present a low-delay, constrained-entropy, backward adaptive, linear-predictive audio coder with low computational complexity. In contrast to most practical linear-predictive coders, the coder facilitates the exploitation of reverse waterfilling. The coder uses time-invariant quantization step sizes and constrained-entropy coding, thus eliminating convergence problems of backward adaptation near signal transitions. Yet rate variations are kept small by the usage of a mixture model density for the signal. The mixture model has the backward adapted model and a second model as components and the component probability is transmitted. Experimental results confirm the advantages of the coder structure and show that the coder provides good overall performance.

[TP1-15]

**Extending Fine-Grain Scalable Audio Coding to Very Low Bitrates using Overcomplete Dictionaries**

*Emmanuel Ravelli (University Pierre and Marie Curie - Paris 6), Gaël Richard (GET-ENST (Telecom Paris)), Laurent Daudet (University Pierre and Marie Curie - Paris 6)*

Signal representations in overcomplete dictionaries are considered here as an alternative to the traditional parametric/transform representations for audio coding. Such representations produce sparser decompositions and thus allow better coding efficiency than transform coding at very low rates. Moreover, the decomposition algorithms are intrinsically progressive, and flexible enough to allow an efficient transient modeling. We propose in this paper a fine-grain scalable audio coder which works on a large range of bitrates (2kbs to 128kbs). Objective measures as well as informal subjective evaluation show that this coder outperforms a comparable transform-based coder at very low bitrates.

[TP1-16]

**Spectral Band Replication Tool for Very Low Delay Audio Coding Applications**

*Tobias Friedrich, Gerald Schuller (Fraunhofer IDMT)*

In this paper a Spectral Band Replication (SBR) tool for low delay audio applications is presented. One goal of this enhancement tool is to reduce the needed bit rate for the representation of audio data using an arbitrary audio codec. Another goal is to keep the algorithmic delay as low as possible. A low coding delay is essential for instance for real time applications like distributed music production under live conditions or telephone conferencing. The low delay SBR approach proposed in this paper uses techniques developed for speech coding purposes and is associated with artificial bandwidth extension methods, particularly spectral folding. Further, the tool exclusively operates in the time domain using prediction methods and adaptive filters in order to avoid additional delay which can be caused by using a filter bank.

[TP1-17]

**Methods for 2nd Order Spherical Harmonic Spatial Encoding in Digital Waveguide Mesh Virtual Acoustic Simulations**

*Alex Southern, Damian Murphy (Department of Electronics, University of York)*

The Digital Waveguide Mesh (DWM) is a numerical simulation technique that has been shown to be suitable for modelling the acoustics of enclosed spaces. Previous work considered an approach using an array of spatially distributed receivers based on sound intensity probe theory to capture spatial room impulse responses (RIRs) from the DWM. A suitable process to facilitate spatial encoding of

the DWMinto second-order spherical harmonics has also been explored. The purpose of this paper is to introduce a new alternative processing scheme based on the Blumlein Difference Technique. Both the previous and currently presented techniques are newly formulated with the main processing in the frequency domain. In addition the ability of the newly proposed technique for capturing the 2nd order components is confirmed and further processing of the receiver array is considered to extend the usable frequency range.

[TP1-18]

**Solo Voice Detection via Optimal Cancellation**

*Christine Smit, Daniel P. W. Ellis (LabROSA, Columbia University)*

Automatically identifying sections of solo voices or instruments within a large corpus of music recordings would be useful, for example, to construct a library of isolated instruments to train signal models. We consider several ways to identify these sections, including a baseline classifier trained on conventional speech features. Our best results, achieving frame level precision and recall of around 70%, come from an approach that attempts to track the local periodicity of an assumed solo musical voice, then classifies the segment as a genuine solo or not on the basis of what proportion of the energy can be canceled by a comb filter constructed to remove just that periodicity.

[TP1-19]

**Fast Sequential LS Estimation for Sinusoidal Modeling and Decomposition of Audio Signals**

*Bertrand David, Roland Badeau (Ecole Nationale Superieure des Telecommunications (GET/Telecom Paris))*

This work demonstrates a sequential Least Squares algorithm applied to the decomposition of sounds into multiple sines on one side and a residual on the other. For a given basis of r distinct frequency components, the algorithm derives recursively the Least Squares estimates of the associated amplitudes and phases. While a direct calculation achieves a O(nr2) complexity the main cost of our implementation is only of 4r multiplications per sample, whatever the length n of the analysis window. The technique is extended to bases of exponentially increasing or decreasing frequency components, which provides a fast and enhanced decomposition of rapidly varying segments of the sound. Finally, the proposed method is successfully applied to a real piano note.

[TP1-20]

**Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices**

*Takeshi Saitou, Masataka Goto (National Institute of Advanced Industrial Science and Technology (AIST)), Masashi Unoki, Masato Akagi (School of Information Science, Japan Advanced Institute of Science and Technology)*

This paper describes a speech-to-singing synthesis system that can synthesize a singing voice given a speaking voice reading the lyrics of a song and its musical score. The system is based on the speech manipulation system STRAIGHT and comprises three models controlling three acoustic features unique to singing voices: the F0, duration, and spectral envelope. Given the musical score and its tempo, the F0 control model generates the F0 contour of the singing voice by controlling four F0 fluctuations: overshoot, vibrato, preparation, and fine fluctuation. The duration control model lengthens the duration of each phoneme in the speaking voice by considering the duration of its musical note. The spectral control model converts the spectral envelope of the speaking voice into that of the singing voice by controlling both the singing formant and the amplitude modulation of formants in synchronization with vibrato. Experimental results showed that the proposed system was able to convert speaking voices into singing voices whose naturalness is almost same with actual singing voices.

[TP1-21]

**Convolutional Synthesis of Wind Instruments**

*Tamara Smyth (School of Computing Science, Simon Fraser University), Jonathan S. Abel (CCRMA, Stanford University)*

In this work we propose a new physical modeling technique whereby a waveguide structure is replaced by a low latency convolution operation with an impulse response that is either measured, modified, and/or constructed, optionally parametrically. By doing so, there is not longer the constraint that succesive arrivals be uniformely spaced, nor need they decay exponentially as they must in a waveguide structure. Measured impulse responses allow for the estimation of filter transfer functions normally seen in a waveguide model of an acoustic tube, some of which are difficult to obtain theoretically, which may then be used to synthesize new impulse responses corresponding to wind instrument bores both existing and imagined. The technique is presented in the context of reed-based wind instruments, but may be extended to other musical instruments.

# Lecture TL2 **Speech and Audio Coding and Hearing Aid**

## Tuesday, October 23 16:00-18:00, Conference House

Chair: *Thomas F. Quatieri*

[TL2-1] 16:00-16:20

**Comparison of Reduced-Bandwidth MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids**

*Simon Doclo (Katholieke Universiteit Leuven, Dept. of Electrical Engineering), Tim van den Bogaert, Jan Wouters (Katholieke Universiteit Leuven, ExpORL), Marc Moonen (Katholieke Universiteit Leuven, Dept. of Electrical Engineering)*

In a binaural hearing aid noise reduction system, binaural output signals are generated by sharing information between the two hearing aids. When each hearing aid has multiple microphones and all microphone signals are transmitted between the hearing aids, a significant noise reduction can be achieved using the binaural multi-channel Wiener filter (MWF). To limit the number of signals being transmitted between the hearing aids, in order to comply with bandwidth constraints of the binaural link, this paper presents reduced-bandwidth MWF-based algorithms, where each hearing aid uses only a filtered combination of the contralateral microphone signals. One algorithm uses the output of a monaural MWF on the contralateral microphone signals, whereas a second algorithm involves a distributed binaural MWF scheme. Experimental results compare the performance of the presented algorithms.

[TL2-2] 16:20-16:40

**Distributed Spatial Audio Coding in Wireless Hearing Aids**

*Olivier Roy, Martin Vetterli (EPFL-I&C-LCAV)*

The information content of binaural signals can be beneficial to many algorithms deployed in current digital hearing aids. However, the exchange of such signals over a wireless communication link requires transmission schemes that must fulfill demanding technical constraints. We present a distributed coding algorithm that builds on psychoacoustic principles in order to achieve this goal with low bitrates, while still preserving affordable delays and complexity. The key steps of the proposed algorithm are detailed and the accuracy of the signal exchange mechanism is evaluated through simulations.

[TL2-3] 16:40-17:00

**A Time-Frequency Modulation Model of Speech Quality**

*James M. Kates (GN ReSound), Kathryn H. Arehart (University of Colorado)*

A new speech-quality metric, based on time-frequency modulation, is introduced in this paper. The metric uses a cochlear model, with the signal envelope in each frequency band converted to dB above threshold. Envelopes sampled across the frequency bands give short-time spectra that are approximated using a set of of mel cepstrum coefficients. The correlation between the cepstral coefficient sequences for the clean and degraded signals is used to compute the quality metric. The metric accurately models quality judgments made by normal-hearing and hearing-impaired listeners for speech degraded by additive noise, nonlinear distortion, and dynamic-range compression.

[TL2-4] 17:00-17:20

**Low Delay Filterbanks for Enhanced Low Delay Audio Coding**

*Markus Schnell, Ralf Geiger, Markus Schmidt, Markus Multrus, Michael Mellar, Jürgen Herre (Fraunhofer IIS), Gerald Schuller (Fraunhofer IDMT)*

Low delay perceptual audio coding has recently gained wide acceptance for high quality communications. While common schemes are based on the well-known Modified Discrete Cosine Transform (MDCT) filterbank, this paper describes novel coding algorithms that for the first time make use of dedicated low delay filterbanks, thus achieving improved coding efficiency while maintaining or even reducing the low codec delay. The MPEG-4 Enhanced Low Delay AAC (AAC-ELD) coder currently under development within ISO/MPEG combines a traditional perceptual audio coding scheme with spectral band replication (SBR), both running in a delay-optimized fashion by using low delay filterbanks.

[TL2-5] 17:20-17:40

**Lossless Audio Coding with Bandwidth Extension Layers**

*Stephen Voran (Institute for Telecommunication Sciences)*

Layered audio coding typically offers reduced distortion as bit rate is increased, but that distortion is spread across the entire band until the lossless coding bit rate is reached and distortion is eliminated. We proposed a layered audio coding paradigm of bandwidth extension, rather than distortion reduction. For example, a core layer can provide lossless coding of a 24 kHz bandwidth signal (fs=48 kHz), then first and second bandwidth extension lossless layers can extend that signal to losslessly coded 48 and then 96 kHz bandwidths (fs=96 and 192 kHz).

[TL2-6] 17:40-18:00

**Rate Distribution between Model and Signal**

*W. Bastiaan Kleijn, Alexey Ozerov (School of Electrical Engineering, KTH (Royal Institute of Technology))*

Knowledge of a statistical model of the signal can be used to increase coding efficiency. A common approach is to use a fixed model structure with parameters that adapt to the signal. The model parameters and a signal representation that depends on the model are encoded. We show that, if the signal is divided into segments of a particular duration, and the model structure is fixed, then the optimal bit allocation for the model parameters does not vary with the overall rate. We discuss in detail the parameter rate for the autoregressive (AR) model. Our approach shows that the square error criterion in the signal domain is consistent with the commonly used root mean square log spectral error for the model parameters. We show that without usage of perceptual knowledge we obtain a rate allocation for the model that is consistent with what is commonly used and which is independent of overall coding rate. We provide experimental results for the application of the autoregressive model to speech that confirm the theory.

# Lecture WL1 **Music and Signal Analysis and Synthesis**

## Wednesday, October 24 8:00-10:00, Conference House

Chair: *Simon Godsill*

[WL1-1] 8:00-8:20

**Sinewave Analysis/Synthesis Based on the Fan-Chirp Transform**

*Robert Dunn, Thomas F. Quatieri (MIT Lincoln Laboratory)*

There have been numerous recent strides at making sinewave analysis consistent with time-varying sinewave models. This is particularly important in high-frequency speech regions where harmonic frequency modulation (FM) can be significant. One notable approach is through the Fan Chirp transform that provides a set of FM-sinewave basis functions consistent with harmonic FM. In this paper, we develop a complete sinewave analysis/synthesis system using the Fan Chirp transform. With this system we are able to obtain more accurate sinewave frequencies and phases, thus creating more accurate frequency tracks, in contrast to a system derived from the short-time Fourier transform, particularly for high-frequency regions of large-bandwidth analysis. With synthesis, we show an improvement in segmental signal-to-noise ratio with respect to waveform matching with the largest gains during rapid pitch dynamics.

[WL1-2] 8:20-8:40

**Spectral Refinement and its Application to Fundamental Frequency Estimation**

*Mohamed Krini, Gerhard Schmidt (HarmanBecker Automotive Systems)*

In this paper a method for spectral refinement of speech and audio signals and its application to fundamental frequency estimation is presented. The SR procedure is applied as a post-processor on the output of a standard short-term frequency analysis. The algorithm is based on a linear combination of weighted subband signal vectors and thus has only low computational complexity. The new scheme can be applied either as a refinement of only a subset of the frequency bands or as a refinement of the entire frequency range including the computation of additional frequency supporting points. Several algorithmic parts, e.g., noise suppression or fundamental frequency estimation, can achieve better results if a better resolution - at least in the lower frequency range - can be provided. In this contribution an enhanced fundamental frequency estimation method is proposed, that allows reliable operation at low signal-to-noise scenarios even for very low fundamental frequencies. Evaluations have shown that a significant improvement can be accomplished when utilizing the SR method as a pre-processor for fundamental frequency estimation.

[WL1-3] 8:40-9:00

**A Novel Method for Decomposition of Multicomponent Nonstationary Signals**

*A. Goli, D. M. McNamara, A. K. Ziarani (Clarkson University)*

A method for decomposition of a multicomponent nonstationary signal is presented. In this method a nonlinear adaptive structure, which was first introduced as a sinusoidal tracking algorithm, is used. This structure provides excellent (instantaneous) frequency-adaptivity and (instantaneous) amplitude-adaptivity features; These features make it an ideal IF-IA estimator for a monocomponent nonstationary signal. We show that a parallel-scheme of this structure decomposites a multicomponent nonstationary signal into its composite monocomponent nonstationary signals and, moreover, estimates IF and IA of each component accurately. Unlike most other approaches in the literature, which use the complex representation for the given signal, this method employs a real representation for it. High resolution in time-frequency plane, cross terms free, simple structure, and the capability of real-time

implementation are among significant features of this method. Results on both simulated and real data, i.e. a bat echolocation signal, are given to confirm the above-mentioned performances.

[WL1-4] 9:00-9:20

**Using Stereo Information for Instrument Identification in Polyphonic Mixtures**

*David Sodoyer, Pierre Leveau, Laurent Daudet (University Pierre et Marie Curie - Paris 6)*

This paper addresses the localization of music instruments in the stereo space. The signal, composed of two channels, is decomposed into a linear combination of Stereo Instrument-Specific Harmonic atoms, that model the harmonic structure of instrument notes as a whole and whose individual angles give clues about the real angle of the sources. To get such decompositions, a Stereo Matching Pursuit algorithm has been implemented, with a phase adaptation for each signal channel. This decomposition give neat source localizations for instantaneous mixes, and the switch to realistic convolutive mixes seems to be possible with adequate post-processing.

[WL1-5] 9:20-9:40

**Bauer Method of MVDR Spectral Factorization for Pitch Modification in the Source Domain**

*M. Ravi Shanker (Indian Institute of Science), R. Muralishankar (PES Institute of Technology), A. G. Ramakrishnan (Indian Institute of Science)*

In our earlier work [1], we employed MVDR (minimum variance distortionless response) based spectral estimation instead of modified linear prediction method [2] in pitch modification. Here, we use the Bauer method of MVDR spectral factorization, leading to a causal inverse filter rather than a noncausal filter setup with MVDR spectral estimation [1]. Further, this is employed to obtain source (or residual) signal from pitch synchronous speech frames. The residual signal is resampled using DCT/IDCT depending on the target pitch scale factor. Finally, forward filters realized from the above factorization are used to get pitch modified speech. The modified speech is evaluated subjectively by 10 listeners and mean opinion scores (MOS) are tabulated. Further, modified bark spectral distortion measure is also computed for objective evaluation of performance. We find that the proposed algorithm performs better compared to time domain pitch synchronous overlapping and [2]. A good MOS score is achieved with the proposed algorithm compared to [1] with a causal inverse and forward filter setup.

[WL1-6] 9:40-10:00

**Waveguide Modeling of Lossy Flared Acoustic Pipes: Derivation of a Kelly-Lochbaum Structure for Real-Time Simulations**

*Thomas Hélie (IRCAM - CNRS UMR 9912), Rémi Mignot (IRCAM - ENST), Denis Matignon (ENST TSI-CNRS UMR 5141)*

This paper deals with the real-time simulation of flared acoustic pipes thanks to digital waveguides. The novelty relies on a refined 1D-acoustic model: the Webster-Lokshin equation. This model describes the propagation of longitudinal waves in axisymmetric acoustic pipes with a varying cross section, visco-thermal losses at the walls, and without assuming plane waves or spherical waves. Solving this model for a piece of pipe leads to a quadripole made of four transfer functions which imitate the global acoustic effects. Moreover, defining progressive waves and introducing some relevant physical interpretations allow to isolate elementary transfer functions associated to elementary acoustic effects. Thanks to this decomposition, a standard Kelly-Lochbaum structure is recovered, from which efficient and low-cost digital simulations are obtained. Thus, this work allows to improve the realism of the sound synthesis of wind instruments, while it preserves standard waveguide techniques which only involve delay lines and digital filters.

# Poster WP1

## Wednesday, October 24 10:20-12:20, West Dining Room

Chair: *Jingdong Chen*

[WP1-01]

**Sound Source Distance Learning Based on Binaural Signals**

*Sampo Vesa (Helsinki University of Technology)*

A learning approach for estimating sound source distance from binaural signals is presented. The frequency-dependent coherence between the left and right ear signals is used as the distance cue. The distance estimation is based on pre-calculated coherence profiles and an energy-weighted likelihood function. The system is evaluated with different speech samples. The accuracy is best in the frontal direction and poorer in the sides, due to shadowing of the head. However, the system shows promise for scenarios where the sound source location is restricted and could be integrated with a two-channel azimuth localization system.

[WP1-02]

**EM Localization and Separation using Interaural Level and Phase Cues**

*Michael I. Mandel, Daniel P. W. Ellis (Columbia University)*

We describe a system for localizing and separating multiple sound sources from a reverberant stereo (two-channel) recording. It consists of a probabilistic model of interaural level and phase differences and an EM algorithm for finding the maximum likelihood parameters of this model. By assigning points in the interaural spectrogram to the source with the best-fitting parameters and then estimating the parameters of the sources from the points assigned to them, the system is able to separate and localize more sound sources than available channels. It is also able to estimate frequency-dependent level differences from a synthetic mixture that correspond well to the synthesis parameters. In experiments in simulated anechoic and reverberant environments, the proposed system was better able to enhance the signal-to-noise ratio of target sources than two comparable algorithms.

[WP1-03]

**Single Channel Speech and Background Segregation Through Harmonic-Temporal Clustering**

*Jonathan Le Roux (Graduate School of Information Science and Technology, The University of Tokyo), Hirokazu Kameoka (NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation), Nobutaka Ono (Graduate School of Information Science and Technology, The University of Tokyo), Alain de Cheveigné (CNR, Universite Paris 5, Ecole Normale Superieure), Shigeki Sagayama (Graduate School of Information Science and Technology, The University of Tokyo)*

The design of effective algorithms for single-channel analysis of complex and varied acoustical scenes is a very important and challenging problem. We present here the application of the recently introduced Harmonic-Temporal Clustering (HTC) framework to single channel speech enhancement, background retrieval and speaker separation. HTC processing relies on a precise parametric description of the voiced parts of speech derived from the power spectrum. We explain the positioning of the algorithm inside the Computational Acoustic Scene Analysis (CASA) area, describe the theoretical background of the method, show through preliminary experiments its basic feasibility, and discuss potential improvements.

[WP1-04]

**Joint Iterative Multi-Speaker Identification and Source Separation using Expectation Propagation**

*John MacLaren Walsh, Youngmoo E. Kim, Travis M. Doll (Drexel University, Electrical and Computer Engineering)*

The identification of individuals through the sound of their voice, a seemingly effortless task for humans, has proven to be quite difficult to achieve in a robust way computationally. The majority of past work in speaker (talker) identification has focused on the single-speaker case. These types of systems are easily confounded by settings where multiple talkers may be overlapping or speaking simultaneously, such as a conference room. We propose a system that jointly identifies and separates the acoustic features of multiple talkers that fall within a library of known individuals. This system uses the probabilistic framework of expectation propagation (EP) to iteratively determine model-based statistics of both speaker identity and feature separation. This research has applications in the areas of human-computer interaction and audio forensics, as well as the analysis of real-world audio surveillance data that contains multiple simultaneous talkers.

[WP1-05]

**Audio Source Separation with Matching Pursuit and Content-Adaptive Dictionaries (MP-CAD)**

*Namgook Cho, Yu Shiu, C.-C. Jay Kuo (University of Southern California)*

A single-channel audio source separation algorithm based on the matching pursuit (MP) technique with content-adaptive dictionaries (CAD) is proposed in this work. The proposed MP-CAD algorithm uses content-dependent atoms that capture inherent characteristics of audio signals effectively. As compared with previous methods based on spectral decomposition and clustering in the time-frequency domain, the MP-CAD algorithm projects the time-domain audio signals onto a subspace spanned by content-adaptive atoms efficiently for their concise representation and separation. The effectiveness of the MP-CAD algorithm in audio signal approximation and single-channel source separation is demonstrated by computer simulation.

[WP1-06]

**Post-Filter Design for Superdirective Beamformers with Closely Spaced Microphones**

*Heinrich W. Löllmann, Peter Vary (RWTH Aachen University)*

In this paper, the post-filter design for superdirective beamformers with small microphone arrays is investigated, which can be used for speech enhancement systems in mobile communication devices or hearing aids. It is shown that coherent noise sources can be well suppressed by a multi-channel controlled post-filter. However, a sufficient suppression of diffuse noise sources can not be achieved by this. Such noise can be further suppressed by a single-channel controlled post-filter. This combined post-filter design leads to a significantly better speech quality compared to the related approach of Le Bouquin et al.

[WP1-07]

**A Fast Microphone Array SRP-PHAT Source Location Implementation using Coarse-To-Fine Region Contraction (CFRC)**

*Hoang Do, Harvey F. Silverman (LEMS, Division of Engineering, Brown University)*

In most microphone array applications, it is essential to localize sources in a noisy, reverberant environment. In such conditions, the steered response power using the phase transform (SRP-PHAT) has been shown to be more robust than faster, two-stage, direct time-difference of arrival methods. The complication is that the SRP-PHAT space has many local maxima which has required computationally

intensive grid-search methods. In this paper, we introduce the use of coarse-to-fine region contraction (CFRC) to make computing the SRP practical. We compare CFRC cost and performance to that of using stochastic region contraction (SRC), a method we presented recently at ICASSP 2007. Results from real data from human talkers show that CFRC costs about the same as SRC overall, but it saves about 40% over SRC under very noisy conditions.

[WP1-08]

**Importance of Energy and Spectral Features in Gaussian Source Model for Speech Dereverberation**

*Tomohiro Nakatani (NTT Corporation), Biing-Hwang Juang (Georgia Institute of Technology), Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi (NTT Corporation)*

In this paper, behavior of speech dereverberation based on a time-varying Gaussian source model (GSM) is investigated to provide a better perspective on solving the dereverberation problem. GSM is a generalization of the autocorrelation codebook (ACC) that has been recently shown to enable us to achieve high quality speech dereverberation with only a few second observation. Based on GSM, the speech dereverberation is formulated as a likelihood maximization problem with multi-channel linear prediction, where the reverberant speech signal is transformed into one that is probabilistically more like clean speech. For the investigation purpose, the autocorrelation matrix of GSM is first decomposed into energy, vocal tract filter, and excitation signal features by adopting an autoregressive GSM (ARGSM), and then analyzed based on experiments. They reveal that the energy feature in the models has a principal effect on reducing the reverberation components. This effect has not been well understood in the context of speech dereverberation. It is also shown that the other spectral features in the models further contribute to recover the short-time characteristics of the dereverberated signals.

[WP1-09]

**A Variable Step-Size for Frequency-Domain Acoustic Echo Cancellation**

*Yin Zhou, Xiaodong Li (Institute of Acoustics, Chinese Academy of Sciences)*

The presence of near-end speech and ambient noise in acoustic echo cancellation makes it necessary for the adaptive filter to introduce a variable step-size to achieve low residual error and high robustness against local disturbances. In this paper, such an optimum bin-wise variable step-size for the frequency-domain adaptive filter algorithm is derived and its calculation based on an estimated magnitude-squared coherence is proposed. The inherent estimation bias and variance and methods to mitigate their adverse effects are discussed. Simulation results confirm that this estimated optimum step-size well control the filter adaptation in various conditions.

[WP1-10]

**A Novel Approach to Active Noise Control Based on Wave Domain Adaptive Filtering**

*P. Peretti, S. Cecchi, L. Palestini, F. Piazza (Universita Politecnica Delle Marche)*

Wave Field Analisys and Synthesis are methods which permit 3D sound field recording and reproduction. They are based on the precise extrapolation and reconstruction of the desired wave field by using arrays of microphones and loudspeakers. In order to utilize these techniques in real world applications (e.g. cinema, home theatre, teleconferencing) it is necessary to apply multi-channel Digital Signal Processing algorithms, already developed for traditional systems. In this paper we approach the problem of Active Noise Control (ANC) from the mentioned point of view: we will present two cancellation schemes depending on whether the noise sources are localized in the recording room or in the reproduction environment. The two algorithms make use of the spatio-temporal transforms, on which Wave Domain Adaptive Filtering (WDAF) is based, in order to reduce the computational complexity. It will be shown through numerical simulations that it is possible to suppress noise sources localized either inside or outside the recording/ reproduction area.

[WP1-11]

**Semantic Colouration Space Investigation: Controlled Colouration in the Bark-Sone Domain**

*Jimi Y. C. Wen, Patrick A. Naylor (Imperial College London)*

This paper investigates the multidimensionality of colouration perception. Here spectral variation, a dominating physical acoustic property for colouration, and its relationship to the colouration perception was studied. A method for converting colouration specified in the bark-sone domain back to the operational magnitude-frequency domain was next formulated. A colouration space was then derived from common factor analysis using both verbal attributes from listening tests and colouration generation control parameters. The results show that one of the two dimensions of the colouration space was intuitively explainable and matches well with the control parameter.

[WP1-12]

**Robustness Analysis of Binaural Hearing Aid Beamformer Algorithms by Means of Objective Perceptual Quality Measures**

*Thomas Rohdenburg, Volker Hohmann, Birger Kollmeier (Universitaet Oldenburg, Medizinische Physik)*

In this contribution different microphone array-based noise reduction schemes for hearing aids are suggested and compared in terms of their performance, signal quality and robustness against model errors. The algorithms all have binaural output and are evaluated using objective perceptual quality measures [1, 2, 3]. It has been shown earlier that these measures are able to predict subjective data that is relevant for the assessment of noise reduction algorithms. The quality measures showed clearly that fixed beamformers designed with head models were relatively robust against steering errors whereas for adaptive beamformers the robustness was limited and the benefit due to higher noise reduction depended on the noise scenario and the reliability of a direction of arrival estimation. Furthermore, binaural cue distortions introduced by the different binaural output strategies could be identified by the binaural speech intelligibility measure [3] even in case monaural quality values were similar. Thus, this perceptual quality measure seems to be suitable to discover the benefit that the listener might have from the effect of spatial unmasking.

[WP1-13]

**Privacy-Preserving Musical Database Matching**

*Madhusudana Shashanka (Boston University), Paris Smaragdis (Mitsubishi Electric Research Labs)*

In this paper we present an illustratory process which allows privacy-preserving transactions in the context of musical databases. In particular we address the problem of matching a piece of music audio to a service database in such a way such that the database provider will not directly observe the query, nor its result, thereby preserving the privacy of the inquirer. We formulate this process within the field of secure multiparty computation and show how such a transaction can be achieved once we derive secure versions of basic signal processing operations.

[WP1-14]

**A Multichannel Linear Prediction Method for the MPEG-4 ALS Compliant Encoder**

*Yutaka Kamamoto, Noboru Harada, Takehiro Moriya (NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation.)*

A new linear prediction analysis method for multichannel signals was devised, with the goal of enhancing the compression performance of the MPEG-4 Audio Lossless coding (ALS) compliant encoder. The multichannel coding tool for this standard carries out an adaptively weighted subtraction of the residual signals of the coding channel from those of the reference channel, both of which are

produced by independent linear prediction. Our linear prediction method tries to directly minimize the amplitude of the predicted residual signal after subtraction of the signals of the coding channel. The results of a comprehensive evaluation show that this method reduces the size of a compressed file, the maximum improvement of compression ratio achieves 14.6%, at the cost of a small increase in computational complexity at the encoder and without increase in decoding time. This is a practical method because the compressed bit stream remains compliant with the MPEG-4 ALS standard.

[WP1-15]

**Enhanced Resampling for Sinusoidal Modeling Parameters**

*Martin Raspaud, Sylvain Marchand (LaBRI, University of Bordeaux 1)*

The sinusoidal modeling parameters can be regarded as control signals, and resampling can be used for synthesis or time-scaling purposes. However, these signals are not zero-centered, and consist of a slow time varying envelope together with modulations (vibrato, tremolo). Using directly the classic resampling method could have disastrous effects. We present an addition to classic resampling aimed at achieving better results on such non zero-centered signals. Applied to the control signals of sinusoidal modeling, the method locally removes a polynomial envelope to the signal to perform better resampling on the residual modulations.

[WP1-16]

**Compressive Coding of Stereo Audio Signals Extracting Sparseness Among Sound Sources with Independent Component Analysis**

*Shigeki Miyabe, Tadashi Mihashi, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano (Nara Institute of Science and Technology), Toshiyuki Nomura (Media and Information Research Laboratories, NEC Corporation)*

In this paper we propose a new compressive coding method of stereophonic audio signal using sparse coding based on independent component analysis. Some researchers have proposed a compressive coding method called binaural cue coding (BCC), and the ISO/MPEG standardization group discusses standard of the next generation audio based on BCC. BCC is based on an assumption that only single sound source exists in a filterbank of the multichannel audio signal. However, BCC degrades quality of complicated musical signal like performances of orchestra because the assumption hardly satisfies. To clarify the sparseness among the sources, the proposed method clarifies the assumption of sparseness using independent component analysis (ICA), a single dominant source is chosen efficiently in each of frequency bin but filterbank. In addition, transfer functions to restore stereo signal is also extracted by ICA. Experiments based on both objective and subjective evaluations assesses efficiency of the proposed method.

[WP1-17]

**Distortion-Aware Query-By-Example for Environmental Sounds**

*Gordon Wichern, Jiachen Xue, Harvey Thornburg, Andreas Spanias (Arizona State University)*

There has been much recent progress in the technical infrastructure necessary to continuously characterize and archive all sounds that occur within a given space or human life. Efficient and intuitive access, however, remains a considerable challenge. In other domains, i.e., melody retrieval, query-by-example (QBE) has found considerable success in accessing music that matches a specific query. We propose an extension of the QBE paradigm to the broad class of natural and environmental sounds. These sounds occur frequently in continuous recordings, and are often difficult for humans to imitate. We utilize a probabilistic QBE scheme that is flexible in the presence of time, level, and scale distortions along with a clustering approach to efficiently organize and retrieve the archived audio. Experiments on a test database demonstrate accurate retrieval of archived sounds, whose relevance to example queries is determined by human users.

[WP1-18]

**Multi-Object Tracking of Sinusoidal Components in Audio with the Gaussian Mixture Probability Hypothesis Density Filter**

*Daniel Clark, Ali-Taylan Cemgil, Paul Peeling, Simon Godsill (University of Cambridge)*

We address the problem of identifying individual sinusoidal tracks from audio signals using multi-object stochastic filtering techniques. Attractive properties for audio analysis include that it is conceptually straightforward to distinguish between measurements that are generated by actual targets and those which are false alarms. Moreover, we can estimate target states when observations are missing and can maintain the identity of these targets between time-frames. We illustrate a particularly useful variant, the Probability Hypothesis Density (PHD) filter, on measurements of musical harmonics determined by high resolution subspace methods which provide very accurate estimates of amplitudes, frequencies and damping coefficients of individual sinusoidal components. We demonstrate this approach in a musical audio signal processing application for extracting frequency tracks of harmonics of notes played on a piano.

[WP1-19]

**Separation of Harmonic and Speech Signals using Sinusoidal Modeling**

*Peter Jančovič, Münevver Köküer (University of Birmingham)*

This paper studies the problem of separation of two harmonic-based source signals from a single-channel mixture signal based on employment of a sinusoidal model. The sinusoidal model represents the signal as a sum of sine-waves, whose parameters (i.e., frequencies, amplitudes, and phases) are estimated by a least-square method that minimizes the reconstruction error between the model and the mixture signal. A comprehensive evaluation of the performance of the sinusoidal model for separation of simulated harmonic signals with various fundamental frequencies is presented. Very good performance, in terms of signal-to-distortion ratio, is observed without any a-priori knowledge about F0s of individual signals. The studied model is then demonstrated for separation of a mixture of two speech signals.

[WP1-20]

**An Instrument Timbre Model for Computer Aided Orchestration**

*Damien Tardieu, Xavier Rodet (IRCAM-CNRS-STMS)*

In this paper we propose a generative probabilistic model for instrument timbre dedicated to computer aided orchestration. We define the orchestration problem as the search of instruments sounds combinations that sound close to a given target. A system that addresses this problem must know a large variety of instruments sounds in order to be able to explore the timbre space of an orchestra. The proposed method is based on gaussian mixture modeling of signal descriptors and on a division of the learning problems that allows to learn many different instrument sounds with few trainig data, and to deduce the models of sounds that are not in the trainig set but that are known to be possible.

# Author Index


**A**

Abel, J. S. 19
Affes, S. 3
Ahrens, J. **6**
Akagi, M. 7, 18
Akagiri, K. 3
Araki, S. 13
Arehart, K. H. 20

**B**

Backer, S. 5
Badeau, R. 18
Benesty, J. 3
Boston, J. R. 16
Bregman, A. S. **11**
Brungart, D. S. **6**
Buchner, H. 15

**C**

Cecchi, S. **6**, 26
Cemgil, A. T. **14**, 29
Chen, J. 9
Chen, M. **3**
Cho, N. **25**
Chou, P. 3
Clark, D. **29**

**D**

Dansereau, R. M. 13
Daudet, L. 17, 23
David, B. **18**
de Cheveigné, A. 24
Dikmen, O. 14
Dmochowski, J. P. **3**
Do, H. **25**
Doclo, S. **20**
Doll, T. M. 25
Douglas, S. C. 4, **16**
Dunn, R. **22**
Duraiswami, R. 2, 14
Durrant, J. D. 16

**E**

Edler, B. 12
Elko, G. W. **5**
Ellis, D. P. W. 8, 10, 18, 24

**F**

Fallon, M. **9**
Friedrich, T. **17**

**G**

Gaubitch, N. D. **2**, 5
Geiger, R. 21
Godsill, S. 9, 14, 29
Goli, A. **22**
Goodwin, M. M. **2**
Goto, M. 18
Grant, S. L. 15
Gribonval, R. 10
Gudnason, J. 5
Gumerov, N. A. 2
Gupta, M. **4**, 16

**H**

Hämäläinen, M. **6**
Hamil, J. T. 6
Harada, N. 27
Haykin, S. **1**
He, L.-W. 3
Helén, M. 7
Hélie, T. **23**
Herre, J. 21
Heute, U. 7
Hikichi, T. 14
Hobbs, B. W. 6
Hoffmann, E. **10**
Hohmann, V. 27
Huo, L. **7**

**I**

Iqbal, M. A. **15**
Izumi, Y. **13**

**J**

Jančovič, P. **29**
Johansson, A. M. 9, 14
Juang, B.-H. 4, 15, 26

**K**

Kamamoto, Y. **27**
Kameoka, H. 24
Kan, K. 1
Kanba, S. 3
Kates, J. M. **20**
Kim, Y. E. 9, 25
Kinoshita, K. 26
Kleijn, W. B. 17, **21**
Kobayashi, T. 3
Köküer, M. 29
Kollmeier, B. 27
Kolossa, D. 10
Krini, M. **22**
Kuech, F. **12**
Kuo, C.-C. J. 25

**L**

Le Roux, J. **24**
Lee, D. D. 9
Lehmann, E. A. **9**, **14**
Leveau, P. 23
Li, C.-C. 16
Li, M. **17**
Li, X. 26
Li, Z. **14**
Lin, Y. **9**
Liu, Z. 3
Löllmann, H. W. **25**
Lyons, A. M. 8

**M**

Makino, S. 13
Mandel, M. I. **24**
Marchand, S. 28
Matignon, D. 23
Matsumoto, K. 5
McNamara, D. M. 22
Mellar, M. 21
Meyer, J. 5
Mignot, R. 23
Mihashi, T. 28
Miyabe, S. **28**
Miyoshi, M. 14, 26
Möller, S. 7
Moonen, M. 20
Moretti, E. 6
Mori, Y. 4
Moriya, T. 27
Mouchtaris, A. 16
Multrus, M. 21
Muralishankar, R. 23
Murphy, D. T. 8, 17
Myllylä, V. 6

**N**

Nakatani, T. 14, **26**
Naylor, P. A. 2, 5, 27
Nomura, T. 28
Nordholm, S. 9, 14

**O**

O'Donovan, A. **2**
O'Loughlin, A. 8
Ogawa, T. 3
Olsson, R. K. 3
Ono, N. 5, 13, 24
Orglmeister, R. 10
Osako, K. **4**
Ozerov, A. 21

**P**

Palestini, L. 6, 26
Peeling, P. 14, 29
Peissig, J. 5
Peretti, P. **26**
Piazza, F. 6, 26
Poliner, G. E. **8**

**Q**

Quatieri, T. F. 22

**R**

Raake, A. 7
Radfar, M. H. **13**
Raj, B. 12
Ramakrishnan, A. G. 23
Rasetshwane, D. M. **16**
Raspaud, M. **28**
Ravelli, E. **17**
Richard, G. 17
Robledo-Arnuncio, E. **4**
Rodet, X. 29
Rohdenburg, T. **27**
Roy, O. **20**

**S**

Sagayama, S. 5, 13, 24
Saitou, T. **18**
Saruwatari, H. 4, 28
Sawada, H. **13**
Scanlon, P. **8**
Schmidt, G. 22
Schmidt, M. 21
Schmidt, M. N. **3**
Schnell, M. **21**
Scholz, K. 7
Schuller, G. 17, 21
Shanker, M. R. **23**
Shashanka, M. **27**
Shikano, K. 4, 28
Shimizu, H. **5**
Shiu, Y. 25
Silverman, H. F. 25
Smaragdis, P. **12**, 27
Smit, C. **18**
Smyth, T. **19**
Sodoyer, D. **23**
Southern, A. **17**
Spanias, A. 28
Spors, S. 6, **15**
Sugiyama, A. **12**

**T**

Takada, S. **3**
Takahashi, Y. 4
Takatani, T. 28
Tardieu, D. **29**

Thomas, M. R. P. 2, **5**
Thornburg, H. 28
Tsakalides, P. 16
Tzagkarakis, C. **16**

**U**

Uchiyama, H. **7**
Unoki, M. 7, 18

**V**

van den Bogaert, T. 20
Vary, P. 25
Vesa, S. **24**
Vetterli, M. 20
Vincent, E. **10**
Virtanen, T. **7**
Voran, S. **21**

**W**

Wältermann, M. 7
Wada, T. S. 4, **15**
Walsh, J. M. **25**
Weiss, R. J. **10**
Wells, J. J. **8**
Wen, J. Y. C. **27**
Wichern, G. **28**
Wouters, J. 20

**X**

Xue, J. 28

**Y**

Yoshioka, T. **14**, 26

**Z**

Zhang, Z. 3
Zhou, Y. **26**
Ziarani, A. K. 22