Interspeech 2013: Program

Welcome Session
Monday August 26, 2013
9:00 - 9:30 Amphithéâtre

Keynote 1: Hynek Hermansky, ISCA Medalist
Monday August 26, 2013
9:30 - 10:30 Amphithéâtre
My Adventures with Speech
Chair: Jean-François Bonastre, ISCA President

Mon-Oral 1
Monday August 26, 2013
11:00 - 13:00 Room Gratte-Ciel 1
Oral 1: Systems for Search/Retrieval of Speech Documents
Chair: Martha Larson, Delft Univ. of Technology, Stavros Tsakalidis, BBN Technologies, Cambridge

230 Information Retrieval-based Dynamic Time Warping
Xavier Anguera

747 On the Computation of Document Frequency Statistics from Spoken Corpora using Factor Automata
Dogan Can and Shrikanth Narayanan

581 Acceleration of Spoken Term Detection Using a Suffix Array by Assigning Optimal Threshold Values to Sub-Keywords
Kouichi Katsurada, Seiichi Miura, Kheang Seng, Yurie Iribe, Tsuneo Nitta

1488 Strategies for High Accuracy Keyword Detection in Noisy Channels
Arindam Mandal, Julien van Hout, Yi-K Cheung Tam, Vikramjit Mitra, Yun Lei, jing zheng, Dimitra Vergyri, Luciana Ferrer, Martin Graciarena, Andreas Kathol, Horacio Franco

562 On the Calibration and Fusion of Heterogeneous Spoken Term Detection Systems
Alberto Abad, Luis Javier Rodriguez-Fuentes, Mikel Penagarikano, Amparo Varona, German Bordel

Mon-Oral 2
Monday August 26, 2013
11:00 - 13:00 Room Gratte-Ciel 3
Oral 2: Speech Analysis I
Chair: Masami Akamine, Toshiba R+D Center, Kawasaki, Nathalie Henrich, GIPSA-lab, Grenoble
Using Phonetic Feature Extraction to Determine Optimal Speech Regions for Maximising the Effectiveness of Glottal Source Analysis
John Kane, Irena Yanushevskaya, John Dalton, Christer Gobl, Alibhe Ni Chasaide

Beyond bandlimited sampling of speech spectral envelope imposed by the harmonic structure of voiced sounds
Hideki Kawahara, Masanori Morise, Tomoki Toda, Ryuichi Nisimura, Toshio Irino

A Source-Filter based Adaptive Harmonic Model and Its Application to Speech Prosody Modification
JeeSok Lee, Frank K. Soong, Hong-Goo Kang

Detection of Glottal Opening Instants using Hilbert Envelope
Ramesh K, S R Mahadeva Prasanna, D Govind

Robust formant detection using group delay function and stabilized weighted linear prediction
Dhananjaya Gowda, Jouni Pohjalainen, Mikko Kurimo, Paavo Alku

A source-filter separation algorithm for voiced sounds based on an exact anticausal/causal pole decomposition for the class of periodic signals.
Thomas Hézard, Thomas Hélie, Boris Doval

---

**Mon-Oral 3**

Monday August 26, 2013
11:00 - 13:00 Room Prestige Gratte-Ciel

**Oral 3: Language and Dialect Recognition**

Chairs: Kay Berkling, Cooperative State Univ. of Baden-Württemberg, Karlsruhe, David van Leeuwen, Radboud Univ., Nijmegen

Parallel Absolute-Relative Feature Based Phonotactic Language Recognition
Weiwei Liu, Weiqiang Zhang, Zhiyi Li, Jia Liu

Dimensionality Reduction of Phone Log-Likelihood Ratio Features for Spoken Language Recognition
Mireia Diez, Amparo Varona, Mikel Penagarikano, Luis Javier Rodriguez-Fuentes, Germán Bordel

Improvements in Language Identification on the RATS Noisy Speech Corpus
Jeff Ma, Bing Zhang, Spyros Matsoukas, Sri Harish Mallidi, Feipeng Li, Hynek Hermansky

Regularized Subspace n-Gram Model for Phonotactic iVector Extraction
Mehdi Soufifar, Lukas Burget, Oldrich Plchot, Sandro Cumani, Jan Cernocky

Foreign Accent Detection from Spoken Finnish Using i-Vectors
Hamid Behravan, Ville Hautamaki, Tomi Kinnunen

Adaptive Gaussian Backend for Robust Language Identification
Mitchell McLaren, Aaron Lawson, Yun Lei, Nicolas Scheffer
Mon-Oral 4
Monday August 26, 2013
11:00 - 13:00 Room Tête d'Or 2

Oral 4: ASR – Neural Networks


Lattice-Based Training of Bottleneck Feature Extraction Neural Networks
Matthias Paulik

Modular Combination of Deep Neural Networks for Acoustic Modeling
Jonas Gehring, Wonkyun Lee, Kevin Kilgour, Ian Lane, Yajie Miao, Alexander Waibel

Informative Spectro-Temporal Bottleneck Features for Noise-Robust Speech Recognition
Shuo-Yiin Chang and Nelson Morgan

A Scalable Approach to Using DNN-Derived Features in GMM-HMM Based Acoustic Modeling For LVCSR
Zhi-Jie Yan, Qiang Huo, Jian Xu

Improved Feature Processing for Deep Neural Networks
Shakti P. Rath, Daniel Povey, Karel Vesely, Jan Cernocky

Deep vs. Wide: Depth on a Budget for Robust Speech Recognition
Oriol Vinyals and Nelson Morgan

Mon-Oral 5
Monday August 26, 2013
11:00 - 13:00 Room Bellecour 2

Oral 5: Speech Acoustics

Chairs: Kunitoshi Motoki, Hokkai-Gakuen Univ., Jacqueline Vaissière, Univ. Sorbonne Nouvelle, Paris

An early case of “VOT”
Angelika Braun

Pitch pattern variations in three regional varieties of American English
Robert A. Fox, Ewa Jacewicz, Jessica Hart

Fine-grain voice strength estimation from vowel spectral cues
Jean-Sylvain Lienard and Claude Barras

Linking Loudness Increases in Normal and Lombard Speech to Decreasing Vowel Formant Separation
Elizabeth Godoy, Catherine Mayo, Yannis Stylianou
Three-dimensional rectangular vocal-tract model with asymmetric wall impedances  
Kunitoshi Motoki

Quasi closed phase analysis for glottal inverse filtering  
Manu Airaksinen, Brad Story, Paavo Alku

---

**Mon-SP1**

**Monday August 26, 2013**

**11:00 - 13:00  Room Bellecour 1**

**Special Session 1 (Part a): Paralinguistic Challenge**

**Chairs:** Anton Batliner, Friedrich-Alexander Univ., Erlangen, Björn Schuller, Technische Univ., München

The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism (Part 1)

<table>
<thead>
<tr>
<th>Session</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social</td>
<td>Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed</td>
</tr>
<tr>
<td>317</td>
<td>Non-linguistic Vocalisation Recognition Based on Hybrid GMM-SVM</td>
<td>Artur Janicki</td>
</tr>
<tr>
<td>374</td>
<td>Characteristic Contours of Syllabic-Level Units in Laughter</td>
<td>Jieun Oh, Eunjoon Cho, Malcolm Slaney</td>
</tr>
<tr>
<td>429</td>
<td>Detection of nonverbal vocalizations using Gaussian Mixture Models:</td>
<td>Teun Krikke and Khiet Truong</td>
</tr>
<tr>
<td></td>
<td>looking for fillers and laughter in conversational speech</td>
<td></td>
</tr>
<tr>
<td>485</td>
<td>Using Phonetic Patterns for Detecting Social Cues in Natural</td>
<td>Johannes Wagner, Florian Lingenfelser, Elisabeth André</td>
</tr>
<tr>
<td>534</td>
<td>Speech Paralinguistic Event Detection Using Probabilistic Time-Series</td>
<td>Rahul Gupta, Kartik Audhkhasi, Sungbok Lee, Shrikanth Narayanan</td>
</tr>
<tr>
<td></td>
<td>Smoothing and Masking</td>
<td></td>
</tr>
<tr>
<td>1403</td>
<td>Detecting Laughter and Filled Pauses Using Syllable-based Features</td>
<td>Guozhen An, David Guy Brizan, Andrew Rosenberg</td>
</tr>
<tr>
<td>234</td>
<td>Classifying Language-Related Developmental Disorders from Speech</td>
<td>Daniel Bone, Theodora Chaspari, Kartik Audhkhasi, James Gibson, Andreas Tsiartas, Sungbok Lee, Shrikanth Narayanan</td>
</tr>
<tr>
<td></td>
<td>Cues: the Promise and the Potential Confounds</td>
<td></td>
</tr>
<tr>
<td>1254</td>
<td>Classification of Developmental Disorders from Speech Signals Using</td>
<td>Katrin Kirchhoff, Yuzong Liu, Jeff Bilmes</td>
</tr>
<tr>
<td></td>
<td>Submodular Feature Selection</td>
<td></td>
</tr>
<tr>
<td>1453</td>
<td>Robust and Accurate Features for Detecting and Diagnosing Autism</td>
<td>Meyesam Asgari, Alireza Bayestehtashk, Izhak Shafran</td>
</tr>
<tr>
<td></td>
<td>Spectrum</td>
<td></td>
</tr>
</tbody>
</table>
Mon-Poster 1

Monday August 26, 2013

11:00 - 13:00 Room Forum 6

Poster 1: Perception of Prosody

Chair: Carlos Gussenhoven, Radboud Univ., Nijmegen

Resistance is futile – The intonation between continuation rise and calling contour in German
Oliver Niebuhr

The Influence of F0 Contour Continuity on Prominence
Perception
Hansjörg Mixdorff and Oliver Niebuhr

Native English listeners' perceptions of prosody in L1 and L2 reading
Caroline Smith and Paul Edmunds

Naturalness Judgement of L2 Mandarin Chinese - Does timing matter?
Chiharu Tsurutani and Dean Luo

Language background affects the strength of the pitch bias in a duration discrimination task
daniel aalto, Juraj Simko, Martti Vainio

Pitch and Lengthening as Cues to Turn Transition in Swedish
Margaret Zellers

Perception of Glottalization in Varying Pitch Contexts Across Languages
Maria Paola Bissiri and Margaret Zellers

Exemplar-based pitch accent categorisation using the Generalized Context Model
Michael Walsh, Katrin Schweitzer, Nadja Schaufler

Double contrast is signalled by prenuclear and nuclear accent types alone, not by f0-plateaux
Bettina Braun and Yuki Asano

Word stress perception in European Portuguese
Susana Correia, Sónia Prota, Joseph Butler, Marina Vigario

Using generalized additive models and random forests to model German prosodic prominence
Denis Arnold, Petra Wagner, R. Harald Baayen
Mon-Poster 2

Monday August 26, 2013
11:00 - 13:00 Room Forum 6
Poster 2: Prosody, Phonetics of Language Varieties

Chair: Lya Meister, Univ. of Tallin

On the Robustness of some Acoustic Parameters for Signalling Word Stress across Styles in Brazilian Portuguese
Plinio Barbosa, Anders Eriksson, Joel Åkesson

Reexamine the Sandhi Rules and the Merging Tones in Hakka Language
Shao-ren Lyu and Ho-hsien Pan

A Preliminary Spectral Analysis of Palatal and Velar Stop Bursts in Pitjantjatjara
Marija Tabain, Richard Beare, Andrew Butcher

Presentational focus realisation in Nalbaria Variety of Assamese
Shakuntala Mahanta and Asim Twaha

On the relation between intonational phrasing and pitch accent distribution. Evidence from European Portuguese varieties
Marisa Cruz and Sónia Frota

How are word-final schwas different in the North and South of France?
Rena Nemoto and Martine Adda-Decker

Modeling Postcolonial Language Varieties: Challenges and Lessons Learned from Mozambican Portuguese
Simone Ashby, Silvia Barbosa, Catarina Silva, Paulino Fumo, José Pedro Ferreira

Prosody of Contrastive Focus in Estonian
Heete Sahkai, Mari-Liis Kalvik, Meelis Mihkla

Exploring the connection of acoustic and distinctive features
Thomas Kisler and Uwe Reichel

A physiological analysis of the tense/lax vowel contrast in two varieties of German
Conceição Cunha, Jonathan Harrington, Phil Hoole

Production of Estonian quantity contrasts by native speakers of Finnish
Einar Meister and Lya Meister
Aerodynamic and durational cues of phonological voicing in whisper
Yohann Meynadier and Yulia Gaydina

Information theoretic syllable structure and its relation to the c-center effect
Uwe Reichel

The Bulgarian Stressed and Unstressed Vowel System. A Corpus Study.
Bistra Andreeva, William Barry, Jacques Koreman

---

**Mon-Poster 3**

**Monday August 26, 2013**

11:00 - 13:00  Room Forum 6

**Poster 3: Speech Synthesis I**

Chair: Marcela Charfuelan, DFKI GmbH, Kaiserslautern

**Training an articulatory synthesizer with continuous acoustic data**
Santitham Prom-on, Peter Birkholz, Yi Xu

**Estimating Speaker-Specific Intonation Patterns using the Linear Alignment Model**
Geza Kiss and Jan P. H. van Santen

**Factored Maximum Likelihood Kernelized Regression for HMM-based Singing Voice Synthesis**
June Sig Sung, Doo Hwa Hong, Hyun Woo Koo, Nam Soo Kim

**Improvements to HMM-Based Speech Synthesis Based on Parameter Generation with Rich Context Models**
Shinnosuke Takamichi, Tomoki Toda, Yoshihori Shiga, Sakriani Sakti, Graham neubig, Satoshi Nakamura

**Voice Conversion in High-order Eigen Space Using Deep Belief Nets**
Torus Nakashika, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Ariki

**Voice Conversion for Non-Parallel Datasets Using Dynamic Kernel Partial Least Squares Regression**
Hanna Silen, Jani Nurminen, Elina Helander, Moncef Gabbouj

**A Style Control Technique for Singing Voice Synthesis Based on Multiple-Regression HMM**
Takashi Nose, Misa Kanemoto, Tomoki Koriyama, Takao Kobayashi

**Predicting the Quality of Text-To-Speech Systems from a Large-Scale Feature Set**
Florian Hinterleitner, Christoph Norrenbrock, Sebastian Möller, Ulrich Heute

**Speaker-specific retraining for enhanced compression of unit selection text-to-speech databases**
Jani Nurminen, Hanna Silén, Moncef Gabbouj
Avatar Therapy: an audio-visual dialogue system for treating auditory hallucinations
Mark Huckvale, Julian Leff, Geoff Williams

Optimizations and Fitting Procedures for the Liljen crants-Fant model for Statistical Parametric Speech Synthesis
Prasannakumar Muthukumar, Alan W Black, H Timothy Bunnell

Analysis and Modeling of "Focus" in Context
Dirk Hovy, Gopala Krishna Anumanchipalli, Alok Parlikar, Caroline Vaughn, Adam Lammert, Eduard Hovy, Alan W Black

Keynote 2: Benjamin Munson
Monday August 26, 2013
14:30 - 15:30 Amphithéâtre

On the Interaction of Social and Linguistic Factors in Phonetic Variation in Typical and Atypical Speakers
Chair: Jennifer Pardo, Montclair Univ.

Mon-Oral 6
Monday August 26, 2013
16:00 - 18:00 Room Gratte-Ciel 1

Oral 6: Perception, Dialectal Differences
Chairs: Catherine Best, MARCS Inst., Sydney, Benjamin Munson, Univ. of Minnesota, Minneapolis

Production and perception of pseudo-V1CV2 outside the vowel triangle: Speech illusion effects
Thi Anh Xuan Tran, Viet Son Nguyen, Eric Castelli, René Carré

Recent Evolution of Non Standard Consonantal Variants in French Broadcast
Maria Candea, Martine Adda-Decker, Lori Lamel

Architekt or Archtekt? Perception of devoiced vowels produced by Japanese speakers of German
Frank Zimmerer, Rei Yasuda, Henning Reetz

Comparing Vowel Category Response Surfaces over Age-varying Maximal Vowel Spaces within and across Language Communities
Andrew Plummer, Lucie Menard, Benjamin Munson, Mary Beckman

Perceived vocal attractiveness across dialects is similar but not uniform
Molly Babel and Grant McGuire

Mutual intelligibility of American, Chinese and Dutch-accented speakers of English tested by SUS and SPIN sentences
Hongyan Wang and Vincent J. van Heuven
Mon-Oral 7
Monday August 26, 2013
16:00 - 18:00 Room Gratte-Ciel 3

Oral 7: Speech Enhancement: Single Channel

Chairs: Yifan Gong, Microsoft, Hiroshi Saruwatari, NAIST, Ikoma

Speech Enhancement Based on Deep Denoising

592  Autoencoder
Xugang Lu, Yu Tsao, Shigeki Matsuda, Chiori Hori

Musical Noise Analysis for Bayesian Minimum Mean-Square Error Speech Amplitude Estimators Based on Higher-Order Statistics

961  Non-negative Matrix Factorization with Linear Constraints for Single-Channel Speech Enhancement
Nikolay Lyubimov and Mikhail Kotov

A Single Channel Speech Enhancement Approach by Combining Statistical Criterion and Multi-Frame Sparse Dictionary Learning

1235  Speech Enhancement
Hung-Wei Tseng, Srikanth Vishnubhotla, Mingyi Hong, Xiangfeng Wang, Jinjun Xiao, Zhi-Quan Luo, Tao Zhang

Speech Enhancement Using Convolutive Nonnegative Matrix Factorization with Cosparsity Regularization

1483  Speech
Majid Mirbagheri, Yanbo Xu, Sahar Akram, Shihab Shamma

Joint Stochastic-Deterministic Wiener Filtering with Recursive Bayesian Estimation of Deterministic Speech

1558
Matthew McCallum and Bernard Guillemin

Mon-Oral 8
Monday August 26, 2013
16:00 - 18:00 Room Prestige Gratte-Ciel

Oral 8: Dialog Modeling

Chairs: Olivier Pietquin, Supelec, Metz, Jérôme Bellagarda, Apple, Cupertino

Automatic self-supervised learning of associations between speech and text

217  Particle Swarm Optimisation of Spoken Dialogue System Strategies
Juho Knuuttila, Okko Räsänen, Unto K. Laine

Model-based Bayesian Reinforcement Learning for Dialogue Management

457  Evaluating Spoken Dialogue Models under the Interactive Pattern Recognition Framework
Fabrizio Ghigi, María Inés Torres Barañano, Raquel Justo, Jose Miguel Benedi

Multi-Layer Mutually Reinforced Random Walk with Hidden Parameters for Improved Multi-Party Meeting Summarization

992
Yun-Nung Chen and Florian Metze
Pei-hao Su, Yow-Bang Wang, Tsung-Hsien Wen, Tien-han Yu, Lin-shan Lee

Mon-Oral 9
Monday August 26, 2013
16:00 - 18:00 Room Tête d'Or 2
Oral 9: ASR – Lexical, Prosodic and Cross/Multi-Lingual
Chairs: Kate Knill, Cambridge Univ., Torbjorn Svendsen, Univ. of Trondheim

Improving LVCSR with Hidden Conditional Random Fields for Grapheme-to-Phoneme
Stefan Hahn, Patrick Lehnen, Simon Wiesler, Ralf Schlüter, Hermann Ney

Context-dependent phone mapping for LVCSR of under-resourced languages
Van Hai Do, Xiong Xiao, Eng Siong Chng, Haizhou Li

Improving Grapheme-based ASR by Probabilistic Lexical Modeling
Ramya Rasipuram and Mathew Magimai Doss

Crosslingual Tandem-SGMM: Exploiting Out-Of-Language Data for Acoustic Model and Feature Level Adaptation
Petr Motlicek, David Imseng, Philip N. Garner

Multilingual Multilayer Perceptron For Rapid Language Adaptation Between and Across Language Families
Ngoc Thang Vu and Tanja Schultz

Modeling Prosodic Sequences with K-Means and Dirichlet Process
Andrew Rosenberg

Mon-Oral 10
Monday August 26, 2013
16:00 - 18:00 Room Bellecour 2
Oral 10: Phonetic Convergence
Chairs: Véronique Delvaux, Univ. of Mons, Jason Shaw, MARCS Inst., Sydney

Convergence of Articulation Rate in Spontaneous Speech
Antje Schweitzer and Natalie Lewandowski

Phonetic convergence in shadowed speech: A comparison of perceptual and acoustic measures.
Jennifer Pardo

Pitch and duration as a basis for entrainment of overlapped speech onsets
Marcin Wlodarczak, Juraj Simko, Petra Wagner
Investigating fine temporal dynamics of prosodic and lexical accommodation
Francesca Bonn, Celine De Looze, Sucheta Ghosh, Emer Gilmartin, Carl Vogel, Anna Polychroniou, Hugues Salamin, Alessandro Vinciarelli, Nick Campbell

Spontaneous and explicit speech imitation
Jeesun Kim, Reuben Demirdjian, Chris Davis

Imitation interacts with one’s second-language phonology but it does not operate cross-linguistically
Václav Jonáš Podlipský, Šárka Šimáčková, Kateřina Chládková

Mon-SPI
Monday August 26, 2013
16:00 – 18:00 Room Bellecour 1

Special Session 1 (Part b): Paralinguistic Challenge

Chairs: Anton Batliner, Friedrich-Alexander Univ., Erlangen, Björn Schuller, Technische Univ., München

The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism (Part 2)
Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed

(b) Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, Marcello Mortillaro, Hugues Salaman, Anna Polychroniou, Fabio Valente, Samuel Kim

Let Me Finish: Automatic Conflict Detection Using Speaker Overlap
Felix Grezes, Justin Richards, Andrew Rosenberg

GMM based Speaker Variability Compensated System for Interspeech 2013 ComParE Emotion Challenge
Vidhyasaharan Sethu, Julien Epps, Eliathamby Ambikairajah, Haizhou Li

Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech
Okko Räsänen and Jouni Pohjalainen

Ensemble of Machine Learning and Acoustic Segment Model Techniques for Speech Emotion and Autism Spectrum Disorders Recognition

Detecting Autism, Emotions and Social Signals Using AdaBoost
Gábor Gosztolya, Róbert Busa-Fekete, László Tóth

Overview
Overview of the 2013 Challenge, comparison of results.
Stefan Steidl

Repository
Presentation of Repository
Björn Schuller

Past & Future
The Challenges: Past and Present
Anton Batliner

The Future
The Future: Ideas, Questionnaire, Open Discussion
Björn Schuller, Stefan Steidl, Anton Batliner
Mon-Poster 4
Monday August 26, 2013
16:00 - 18:00 Room Forum 6

Poster 4: Speech Production, Acquisition and Development I

Chair: Takayuki Arai, Sophia Univ., Yotsuya

Prosodic Markings of Semantic Predictability in Taiwan
Mandarin
Po-jen Hsieh

How Did it Work? Historic Phonetic Devices Explained by Coeval Photographs
Ruediger Hoffmann, Dieter Mehnert, Rolf Dietzel

Eliciting speech with sentence lists – a critical evaluation with special emphasis on segmental anchoring
Lea Susan Kohzt and Oliver Niebuhr

An MRI-based Acoustic Study of Mandarin Vowels
Yuguang Wang, Jianwu Dang, Xi Chen, Jianguo Wei, Kiyoshi Honda, Hongcui Wang

Daniel Hirst

Velic Coordination in French Nasals: a Real-time Magnetic Resonance Imaging Study
Michael Proctor, Louis Goldstein, Adam Lammert, Dani Byrd, Asterios Toutios, Shrikanth Narayanan

Learning to imitate adult speech with the KLAIR virtual infant
Mark Huckvale and Anmita Sharma

Physics-based synthesis of disordered voices
Jorge Lucero, Jean Schoentgen, Mara Behlau

Place assimilation and articulatory strategies: the case of sibilant sequences in French as L1 and L2
Sonia d’Apolito and Barbara Gili Fivela

Effects of Lexical Class and Lemma Frequency on German Homographs
Barbara Samlowski, Petra Wagner, Bernd Möbius

Measuring laryngealization in running speech: interaction with contrastive tones in Yalálag Zapotec
Leonardo Lancia, Heriberto Avelino, Daniel Voigt

A Neural Oscillator Model of Speech Timing and Rhythm
Erin Rusaw
Observations of Perseverative Coarticulation in Lateral Approximants Using MRI
Nicole Wong, Maojing Fu, Zhi-Pei Liang, Ryan Shoosted, Bradley Sutton

Mon-Poster 5
Monday August 26, 2013
16:00 - 18:00 Room Forum 6

Poster 5: General Topics in ASR

Chair: Tanel Alumäe, Univ. of Tallinn

Comparing Computation in Gaussian Mixture and Neural Network Based Large-Vocabulary Speech Recognition
vishwa Gupta and Gilles Boulianne

Improved Speed and Quality for Automatic Speech Recognition using Simultaneous Perturbation Stochastic Approximation
Daniel Stein, Jochen Schwenninger, Michael Stadschnitzer

Hardware/Software Codesign for Mobile Speech Recognition
David Sheffield, Michael Anderson, Yunsup Lee, Kurt Keutzer

Exploiting the Succeeding Words in Recurrent Neural Network Language Models
Yangyang Shi, Martha A Larson, Pascal Wiggers, Catholijn M Jonker

Speech Acoustic Unit Segmentation Using Hierarchical Dirichlet Processes
Amir Hossein Harati Nejad Torbati, Joseph Picone, Marc Sobel

Transducer-based Speech Recognition with Dynamic Language Models
Munir Georges, Stephan Kanthak, Dietrich Klakow

A Method for Structure Estimation of Weighted Finite-State Transducers and Its Application To Grapheme-to-Phoneme Conversion
Yotaro Kubo, Takaaki Hori, Atsushi Nakamura

Combining Forward-based and Backward-based Decoders for Improved Speech Recognition Performance
Denis Jouvet and Dominique Fohr

iVector-based Acoustic Data Selection
Olivier Siohan and Michiel Bacchiani

Accurate and Compact Large Vocabulary Speech Recognition on Mobile Devices
Xin Lei, Andrew Senior, Alexander Gruenstein, Jeffrey Sorensen

Pre-Initialized Composition For Large-Vocabulary Speech Recognition
Cyril Allauzen and Michael Riley
Speaker Dependent Activation Keyword Detector Based On GMM-UBM
Evelyn Kurniawati and Sapna George

Written-Domain Language Modeling for Automatic Speech Recognition
Hasim Sak, Yun-hsuan Sung, Francoise Beaufays, Cyril Allauzen

---

Mon-Poster 6
Monday August 26, 2013
16:00 - 18:00  Room Forum 6

Poster 6: Voice Activity Detection and Speech Segmentation

Chair: Ascensión Gallardo Antolín, Univ. Carlos III de Madrid
Detecting words in speech using linear separability in a bag-of-events vector
Maarten Versteegh and Louis ten Bosch

On the Improvement of Multimodal Voice Activity Detection
Matt Burlick, Dimitrios Dimitriadis, Eric Zavesky

Using Linguistic Information to Detect Overlapping Speech
Jürgen T. Geiger, Florian Eyben, Nicholas Evans, Björn Schuller, Gerhard Rigoll

Incremental Acoustic Subspace Learning for Voice Activity Detection using Harmonicity-based Features
Jiaxing Ye, Takumi Kobayashi, Masahiro Murakawa, Tetsuya Higuchi

Endpoint Detection using Weighted Finite State Transducer
Hoon Chung, SungJoo Lee, YunKeun Lee

A robust frontend for VAD: exploiting contextual, discriminative and spectral cues of human voice
Maarten Van Segbroeck, Andreas Tsiartas, Shrikanth Narayanan

All for One: Feature Combination for Highly Channel-Degraded Speech Activity Detection
Martin Graciarena, Abeer Alwan, Dan Ellis, Horacio Franco, Luciana Ferrer, John H.L. Hansen, Adam Janin, Byung-Suk Lee, Yun Lei, Vikramjit Mitra, Nelson Morgan, Seyed Omid Sadjadi, TJ Tsai, Nicolas Scheffer, Lee Ngee Tan, Benjamin Williams

Superposed Speech Localisation using Frequency Tracking
Maxime Le Coz, Julien Pinquier, Régine Andre-Obrecht

Multi-band long-term signal variability features for robust voice activity detection
Andreas Tsiartas, Theodora Chaspari, Athanasios Katsamanis, Prasanta Ghosh, Ming Li, Maarten Van Segbroeck, Alexandros Potamianos, Shrikanth Narayanan

A Low-Complexity Voice Activity Detector for Smart Hearing Protection of Hyperacusic Persons
Narimene Lezzoum, Ghyslain Gagnon, Jérémie Voix
Speech Activity Detection on YouTube Using Deep Neural Networks
Neville Ryant, Mark Liberman, Jiahong Yuan

Speaker and Noise Independent Voice Activity Detection
François Germain, Dennis Sun, Gautham Mysore

Confidence-Based Scoring: A Useful Diagnostic Tool for Detection Tasks
TJ Tsai and Adam Janin

Concurrent processing of voice activity detection and noise reduction using empirical mode decomposition and modulation spectrum analysis
Yasuaki Kanai, Shota Morita, Masashi Unoki

Show & Tell 1
Monday August 26, 2013
16:00 - 18:00 Salon Tête d'Or

Chairs: Laurence Devillers, LIMSI-CNRS/Paris-Sorbonne 4, Fabrice Lefèvre, LIA, Avignon

The Furhat Social Companion Talking Head
Samer Al Mouhayed, Jonas Beskow, Gabriel Skantze

Ultraspeech-Player: Intuitive Visualization of Ultrasound Articulatory Data for Speech Therapy and Pronunciation Training
Thomas Hueber

Laughter Modulation: from Speech to Speech-Laugh
Jieun Oh, Ge Wang

ReFr: An Open-Source Reranker Framework
Daniel M. Bikel, Keith B. Hall

Embedding Speech Recognition to Control Lights
Alessandro Sosi, Fabio Brugnara, Luca Cristoforetti, Marco Matassoni, Mirco Ravanelli, Maurizio Omologo

The MUTE Silent Speech Recognition System
Geoffrey S. Meltzer, James T. Heaton, Yunbin Den

The Edinburgh Speech Production Facility Corpus DoubleTalk CorpusDoubleTalk
James M Scobbie, Alice Turk, Christian Geng, Simon King, Robin Lickley, Korin Richmond

Lexee: Cloud-Based Platform for Building and Deploying Voice-Enabled Mobile Applications
Dmitry Sityaev, Jonathan Hotz, Vadim Snitkovsky

A Tool to Elicit and Collect Multicultural and Multimodal Laughter
Mariette Soury, Clément Gossart, Martine Adda, Laurence Devillers

Design of a Mobile App for Interspeech Conferences: Towards an Open Tool for the Spoken Language Community
Robert Schleicher, Tilo Westermann, Jinjin Li, Moritz Lavitschka, Benjamin Mateev, Ralf Reichmuth, Sebastian Möller

Visualizing Articulatory Data with VisArtico
Slim Ouni

Audition: the Most Important Sense for Humanoid Robots?
Rodolphe Gelin, G. Barbieri
Welcome Event and Aperitif
Monday August 26, 2013
18:00 - 20:00 Amphithéâtre

Keynote 3: Anne-Lise Giraud
Tuesday August 27, 2013
9:00 - 10:00 Amphithéâtre

Are Cortical Oscillations a Useful Ingredient of Speech Perception?
Chair: Jean-Luc Schwartz, GIPSA-lab, Grenoble

Tue - Oral 11
Tuesday August 27, 2013
10:30 - 12:30 Room Gratte-Ciel 1

Oral 11: Discourse, Intonation, Prosody
Chairs: Plinio Barbosa, State Univ. of Campinas, Marija Tabain, La Trobe Univ., Melbourne

The Acoustics of Word Stress in Swedish: A Function of Stress Level, Speaking Style and Word Accent
Anders Eriksson, Plinio Barbosa, Joel Åkesson

Intonational contrasts encode speaker's certainty in neutral vs. incredulity declarative questions in French
Amandine Michelas, Cristel Portes, Maud Champagne-Lavau

Prosodic Changes Pre-announcing a Syntactic Completion Point in Japanese Utterance
Yuichi Ishimoto, Mika Enomoto, Hitoshi Iida

Prosodic encoding of declarative, interrogative and imperative sentences in Jaminjung, a language of Australia
Candide Simard

Crosslinguistic priming in interactive reference: Evidence for conceptual alignment in speech production
Anne Vullinghs, Martijn Goudbeek, Emiel Krahmer

A cross-linguistic study on turn-taking and temporal alignment in verbal interaction
Spyros Kousidis, David Schlangen, Stavros Skopeteas

Tue - Oral 12
Tuesday August 27, 2013
10:30 - 12:30 Room Gratte-Ciel 3

Oral 12: Source Separation
Chairs: Shoji Makino, Univ. of Tsukuba, Emmanuel Vincent, LORIA, Nancy

Discriminative Nonnegative Dictionary Learning using Cross-Coherence Penalties for Single Channel Source Separation
Emad M. Grais and Hakan Erdogan
Monaural Speech Segregation Based on Pitch Track Correction Using An Ensemble Kalman Filter
Han-Gyu Kim, Gil-Jin Jang, Jeong-Sik Park, Yung-Hwan Oh

Voice Activity Classification for Automatic Bi-Speaker Adaptive Beamforming in Speech Separation
Thuy Ngoc Tran, William Cowley, Andre Pollok

Blind source separation using spatially distributed microphones based on microphone-location dependent source activities
Keisuke Kinoshita, Mehrez Souden, Tomohiro Nakatani

Non-negative Tensor Factorisation of Modulation Spectrograms for Monaural Sound Source Separation
Tom Barker and Tuomas Virtanen

Iterative Sinusoidal-Based Partial Phase Reconstruction in Single-Channel Source Separation
Pejman Mowlaee and Mario Kaoru Watanabe

**Tue - Oral 13**

**Tuesday August 27, 2013**
10:30 - 12:30 Room Prestige Gratte-Ciel

**Oral 13: Paralinguistic Information**

Chairs: Elizabeth Shriberg, Microsoft Speech Labs, Mountain View, Michael Wagner, Univ. of Camberra

Classification of speech under stress by modeling the aerodynamics of the laryngeal ventricle
Xiao Yao, Takatoshi Jitsuhiro, Chiyomi Miyajima, Norihide Kitaoka, Kazuya Takada

“Sure, I Did The Right Thing”: A System for Sarcasm Detection in Speech
Rachel Rakov and Andrew Rosenberg

Investigating Voice Quality as a Speaker-Independent Indicator of Depression and PTSD
Stefan Scherer, Jonathan Gratch, Louis-Philippe Morency

A corpus-based study between elderly and young speakers of European Portuguese: acoustic correlates and their impact on speech recognition performance
Thomas Pellegrini, Annika Hämäläinen, Philippe Boula de Mareuil, Michael Tjalve, Isabel Trancoso, Sara Candeias, Miguel Sales Dias, Daniela Braga

Modeling Spectral Variability for the Classification of Depressed Speech
Nicholas Cummins, Julien Epps, Vidhyasaharan Sethu, Michael Breakspear, Roland Goecke

Sentiment Analysis of Online Spoken Reviews
Veronica Perez-Rosas and Rada Mihalcea
Tue - Oral 14

Tuesday August 27, 2013
10:30 - 12:30  Room Tête d’Or 2

Oral 14: ASR – Robustness Against Noise I

Chairs: John Hansen, Univ. of Texas, Denis Jouvet, Loria/Inria, Nancy

Using Twin-HMM-Based Audio-Visual Speech Enhancement as a front-end for robust audio-visual speech recognition
Ahmed Hussen Abdelaziz, Steffen Zeiler, Dorothea Kolossa

Spectro-Temporal Directional Derivative Features for Automatic Speech Recognition
James Gibson, Maarten Van Segbroeck, Antonio Ortega, Panayiotis Georgiou, Shrikanth Narayanan

Attribute-based histogram equalization (HEQ) and its adaptation for robust speech recognition
Xiong Xiao, Eng Siong Chng, Haizhou Li

Modified Cepstral Mean Normalization - Transforming to utterance specific non-zero mean
Vikas Joshi, Vishnu Prasad, Umesh S.

Damped Oscillator Cepstral Coefficients for Robust Speech Recognition
Vikramjit Mitra, Horacio Franco, Martin Graciarena

Regularized MVDR Spectrum Estimation-based Robust Feature Extractors for Speech Recognition
Md Jahangir Alam, Patrick Kenny, Douglas O’Shaughnessy


Tue - Oral 15

Tuesday August 27, 2013
10:30 - 12:30  Room Bellecour 2

Oral 15: Neural Basis of Speech Perception

Chairs: Anne-Lise Giraud, Univ. of Geneva, Michel Hoen, CRNL, Lyon

Optimization of sigmoidal rate-level function based on acoustic features
Víctor Poblete, Nestor Becerra Yoma, Richard M. Stern

Composing auditory ERPs: Cross-linguistic comparison of auditory change complex for Japanese fricative consonants
Makiko Sadakata, Loukianos Spyrou, Mizuki Shingai, Kaoru Sekiyama

How voicing, place and manner of articulation differently modulate event-related potentials associated with response inhibition
Nathalie Bedoin, Jennifer Krzonowski, Emmanuel Ferragne

Categorization of speech in early auditory evoked responses
Ludovic Bellier, Michel Mazzuca, Hung Thai-Van, Anne Caclin, Rafael Laboissière

Perception and production of Italian vowels: an ERP study
Anna Dora Manca and Mirko Grimaldi
Implicit learning leads to familiarity effects for intonation but not for voice
Ann-Kathrin Grohe and Bettina Braun

---

**Tue - SP2**

**Tuesday August 27, 2013**

10:30 - 12:30 Room Bellecour 1

**Special Session 2: Spoofing and Countermeasures for Automatic Speaker Verification**

Chairs: Nicholas Evans, EURECOM, Sophia Antipolis, Sebastien Marcel, IDIAP, Martigny

Spoofing and countermeasures for automatic speaker verification

Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi

I-vectors meet imitators: on vulnerability of speaker verification systems against voice mimicry

Rosa González Hautamäki, Tomi Kinnunen, Ville Hautamaki, Timo Leino, Anne-Maria Laukkanen

Security Evaluation of I-Vector Based Speaker Verification Systems Against Hill-Climbing Attacks

Marta Gomez-Barrero, Javier Gonzalez-Domínguez, Javier Galbally, Joaquin Gonzalez-Rodriguez

A new speaker verification spoofing countermeasure based on local binary patterns

Federico Alegre, Ravichander Vipperla, Asmaa Amehraye, Nicholas Evans

Voice Transformation-based Spoofing of Text-Dependent Speaker Verification Systems

Zvi Kons and Hagai Aronowitz

Vulnerability evaluation of speaker verification under voice conversion spoofing: the effect of text constraints

Zhizheng Wu, Anthony Larcher, Kong Aik Lee, Eng Siong Chng, Tomi Kinnunen, Haizhou Li

---

**Tue - Poster 7**

**Tuesday August 27, 2013**

10:30 - 12:30 Room Forum 6

**Poster 7: Speech Production, Acquisition and Development II**

Chair: Qiang Fang, Institute of Linguistics, Beijing

Timing differences in articulation between voiced and voiceless stop consonants: An analysis of cine-MRI data

Masako Fujimoto, Tatsuya Kitamura, Hiroaki Hatano, Ichiro Fujimoto

Vocal Tract Cross-Distance Estimation from Real-Time MRI using Region-of-Interest Analysis

Adam Lammert, Vikram Ramanarayan, Michael Proctor, Shrikanth Narayanan

Syllable Nuclei Detection Using Perceptually Significant Features

A. Apoorv Reddy, Nivedita Chennupati, Bayya Yegnanarayana
Truncation of Pharyngeal Gesture in English Diphthong [au]
Fang-Ying Hsieh, Louis Goldstein, Dani Byrd, Shrikanth Narayanan

The Effect of Word Frequency and Lexical Class on Articulatory-Acoustic Coupling
Zhaojun Yang, Vikram Ramanarayanan, Dani Byrd, Shrikanth Narayanan

Discrimination between fricative and affricate in Japanese using time and spectral domain variables.
Kimiko Yamakawa and Shigeaki Amano

L2 syntax acquisition: the effect of oral and written computer assisted practice
Polina Drozdova, Catia Cucchiarini, Helmer Strik

The physiological use of the charismatic voice in political speech
Rosario Signorello and Didier Demolin

Crosslinguistic Corpus of Hesitation Phenomena: A corpus for investigating first and second language speech performance
Ralph Rose

Real-time control of a 2D animation model of the vocal tract using optopalatography
Simon Preuß, Christiane Neuschaefer-Rube, Peter Birkholz

The influence of accentuation and polysyllabicity on compensatory shortening in German
Jessica Siddins, Jonathan Harrington, Felicitas Kleber, Ulrich Reubold

An Investigation of Vowel Epenthesis in Chinese Learners’ Production of German Consonants
Hongwei Ding and Rüdiger Hoffmann

On the Evaluation of Inversion Mapping Performance in the Acoustic Domain
Korin Richmond, Zhen-Hua Ling, Junichi Yamagishi, Benigno Uria

**Tue - Poster 8**
Tuesday August 27, 2013
10:30 - 12:30 Room Forum 6

**Poster 8: Speech Synthesis II**

Chair: Olivier Rosec, VOXYGEN, Pleumeur Bodou

Probabilistic speech F0 contour model incorporating statistical vocabulary model of phrase-accent command sequence
Tatsuma Ishihara, Hirokazu Kameoka, Kota Yoshizato, Daisuke Saito, Shigeki Sagayama

Reconstruction of continuous voiced speech from whispers
Ian McLoughlin, Jingjie Li, Yan Song
Efficient Speech Transcription Through Respeaking
Matthias Sperber, Graham Neubig, Christian Fügen, Satoshi Nakamura, Alex Waibel

Annotation and Classification of Political Advertisements
Samuel Kim, Panayiotis Georgiou, Shrikanth Narayanan

Using Role Play for Collecting Question-Answer Pairs for Dialogue Agents
Ryuichiro Higashinaka, Kohji Dohsaka, Hideki Isozaki

Individual differences of emotional expression in speaker’s behavioral and autonomic responses
Yoshiko Arimoto and Kazuo Okanoya

Development and Validation of the Conversational Agents Scale (CAS)
Ina Wechsung, Benjamin Weiss, Christine Kühnel, Patrick Ehrenbrink, Sebastian Möller

Motivational Feedback in Crowdsourcing: a Case Study in Speech Transcription
Giuseppe Riccardi, Arindam Ghosh, Shammar Absar Chowdhury, Ali Orkan Bayer

Sheffield Wargames Corpus
Charles Fox, Yulan Liu, Erich Zwyssig, Thomas Hain

Formalizing Expert Knowledge for Developing Accurate Speech Recognizers
Anuj Kumar, Florian Metze, Wenyi Wang, Matthew Kam

Analysis of gaze and speech patterns in three-party quiz game interaction
Samer Al Moubayed, Jens Edlund, Joakim Gustafson

Methodologies for the evaluation of Speaker Diarization and Automatic Speech Recognition in the presence of overlapping speech
Olivier Galibert

‘Houston, We have a solution’ : Speech and Language Processing Advancements using NASA Apollo Missions
Abhijeet Sangwan

---

Tue-RoundTable
Tuesday August 27, 2013
13:00 - 14:30 Amphithéâtre

Round Table: Innovative Products and Services Based on Speech Technologies

Chair: Roberto Pieraccini, ICSI, Berkeley
Co-chairs: Laurence Devillers, LIMSI-CNRS/Paris-Sorbonne 4, Fabrice Lefèvre, LIA, Avignon
Invited speakers:

- Jérôme Bellagarda, Apple, Cupertino,
- Rodolphe Gelin, Aldebaran Robotics, Paris,
- Dilek Hakkani-Tür, Microsoft Research, Mountain View,
- Alex Waibel, Karlsruhe Institute of Technology/Carnegie Mellon Univ./Jibbigo LLC.

---

**Tue -Oral 16**

Tuesday August 27, 2013

14:00 - 16:00  Room Gratte-Ciel 1

**Oral 16: Speech Technology for Speech Disorders**

Chairs: Corinne Fredouille, LIA, Avignon, Elmar Nöth, Friedrich-Alexander Univ., Erlangen

*Performance of the MVOCA Silent Speech Interface Across Multiple Speakers*

Robin Hofe, Jie Bai, Lam A. Cheah, Stephen R. Ell, James M. Gilbert, Roger K. Moore, Phil D. Green

*Automatic Glottal Tracking From High-Speed Digital Images Using a Continuous Normalized Cross Correlation*

Gustavo Andrade-Miranda and Juan Ignacio Godino

*Automatic Evaluation of Parkinson's Speech - Acoustic, Prosodic and Voice Related Cues*

Tobias Bocklet, Stefan Steidl, Elmar Nöth, Sabine Skodda

*Comparison of approaches for an efficient phonetic decoding*

Luiza Orosanu and Denis Jouvet

*Learning speaker-specific pronunciations of disordered speech*

Heidi Christensen, Phil Green, Thomas Hain

*Adapting a Speech into Sign Language Translation System to a new domain*

Verónica López-Ludeña, Rubén San Segundo Hernández, Carlos González-Morcillo, Juan Carlos López, Emilio Ferreiro

---

**Tue -Oral 17**

Tuesday August 27, 2013

14:00 - 16:00  Room Gratte-Ciel 3

**Oral 17: Speech Analysis II**

Chairs: Paavo Alku, Aalto Univ., Helsinki, Abeer Alwan, USC, Los Angeles

*Assessing the Intelligibility Impact of Vowel Space Expansion via Clear Speech-Inspired Frequency Warping*

Elizabeth Godoy, Maria Koutsogiannaki, Yannis Stylianou

*Prediction of Intelligibility of Noisy and Time-Frequency Weighted Speech based on Mutual Information Between Amplitude Envelopes*
Frequency-adaptive post-filtering for intelligibility enhancement of narrowband telephone speech
Emma Jokinen, Marko Takanen, Paavo Alku

Comparative investigation of objective speech intelligibility prediction measures for noise-reduced signals in Mandarin and Japanese
Junfeng Li, Fei Chen, Masato Akagi, Yonghong Yan

Monitoring the Effects of Temporal Clipping on VoIP Speech Quality
Andrew Hines, Jan Skoglund, Anil Kokaram, Naomi Harte

The spectral dynamics of vowels in Mandarin Chinese
Jiahong Yuan

Tue -Oral 18
Tuesday August 27, 2013
14:00 - 16:00 Room Prestige Gratte-Ciel

Oral 18: Discriminative Training Methods for Language Modeling
Chairs: Hermann Ney, RWTH Aachen Univ., Murat Saraclar, Boğaziçi Univ., Istanbul

CSLM - A modular Open-Source Continuous Space Language Modeling Toolkit

Speed Up of Recurrent Neural Networks Language Models With Sentence Independent Subsampling Stochastic Gradient Decent
Yangyang shi, Mei-Yuh Hwang, Kaisheng Yao, Martha A Larson

Improving Unsupervised Language Model Adaptation with Discriminative Data Filtering
Shuangyu Chang, Michael Levit, Partha Parthasarathy, Benoit Dumoulin

Lightly Supervised Training for Risk-Based Discriminative Language Models
Akio Kobayashi, Takahiro Oku, Yuya Fujita, Shoeyi Sato

Investigation of MT-based ASR Confusion Models for Semi-Supervised Discriminative Language Modeling
Erinç Dikici, Emily Prud'hommeaux, Brian Roark, Murat Saraclar

Unsupervised Discriminative Language Modeling Using Error Rate Estimator
Takanobu Oba, Atsunori Otagawa, Takaaki Hori, Hirokazu Masataki, Atsushi Nakamura

Tue -Oral 19
Tuesday August 27, 2013
14:00 - 16:00 Room Tête d'Or 2

Oral 19: ASR – Adaptive Training
Chairs: Jen-Tzung Chien, NCKU,Taiwan, Jean-Luc Gauvain, LIMSI/CNRS
A region-specific feature-space transformation for speaker adaptation and singularity analysis of Jacobian matrix
Shakti P. Rath, Lukas Burget, Martin Karafiát, Ondrej Glembek, Jan Cernocky

An Explicit Independence Constraint for Factorised Adaptation in Speech Recognition
Yongqiang Wang and Mark Gales

Asynchronous Factorisation of Speaker and Background with Feature Transforms in Speech Recognition
Oscar Saz and Thomas Hain

Cluster Adaptive Training With Factorized Decision Trees For Speech Recognition
Kai Yu and Hainan Xu

Rapid and Effective Speaker Adaptation of Convolutional Neural Network Based Models for Speech Recognition
Ossama Abdel-Hamid and Hui Jiang

Text-to-Speech Inspired Duration Modeling for Improved Whole-Word Acoustic Models
Keith Kintzley, Aren Jansen, Hynek Hermansky

---

**Tue - Oral 20**

Tuesday August 27, 2013
14:00 - 16:00 Room Bellecour 2

Oral 20: Speech Acquisition and Development

Chairs: Fangfang Li, Dept of Psychology, Univ. of Lethbridge, Mark Huckvale, UCL, London

Duration of early vocalisations
Adele Gregory, Marija Tabain, Michael Robb

Acoustic development of vowel production in American English children
Jing Yang and Robert Fox

The role of intrinsic motivations in learning sensorimotor vocal mappings: a developmental robotics study
Clément Moulin-Frier and Pierre-Yves Oudeyer

Children’s timing and repair strategies for communication in adverse listening conditions
Valerie Hazan and Michèle Pettinato

Speech planning as an index of speech motor control maturity
Guillaume Barbier, Pascal Perrier, Lucie Ménard, Yohan Payan, Mark Tiede, Joseph Perkell

The Relationship between Gender-differentiated Productions of /s/ and Gender Role Behaviour in Young Children
Melissa Kinsman and Fangfang Li
**Tue -SP3**

**Tuesday August 27, 2013**

14:00 - 16:00 ; 16:30 - 18:30  Room Bellecour 1

**Special Session 3 (a & b): Articulatory Data Acquisition and Processing**

Chairs: **Slim Ouni**, LORIA, Nancy, **Korin Richmond**, CSTR, Edinburgh

**Data-driven Design of a Sentence List for an Articulatory Speech Corpus**

Jeffrey Berry and Luciano Fadiga

**Faster 3D Vocal Tract Real-time MRI Using Constrained Reconstruction**

Yinghua Zhu, Asterios Toutios, Shrikanth Narayanan, Krishna Nayak

**Relevance-Weighted-Reconstruction of Articulatory Features in Deep-Neural-Network-Based Acoustic-to-Articulatory Mapping**

Claudia Canevari, Leonardo Badino, Luciano Fadiga, Giorgio Metta

**Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience**

Fabian Tomaschek, Martijn Wieling, Denis Arnold, Harald Baayen

**Towards a Systematic and Quantitative Analysis of Vocal Tract Data**

Samuel Silva, António Teixeira, Catarina Oliveira, Paula Martins

**A two-step technique for MRI audio enhancement using dictionary learning and wavelet packet analysis**

Colin Vaz, Vikram Ramanarayanan, Shrikanth Narayanan

**Electromagnetic Articulography with AG500 and AG501**

Massimo Stella, Antonio Stella, Francesco Sigona, Paolo Bernardini, Mirko Grimaldi, Barbara Gili Fivela

**Development and implementation of fiducial markers for vocal tract MRI imaging and speech articulatory modelling**

Pierre Badin, Julían Andrés Valdés Vargas, Arielle Koncki, Laurent Lamalle, Christophe Savariaux

**Functional Data Analysis of Tongue Articulation in Palatal Vowels: Gothenburg and Malmohus Swedish /i/, y/, u-/**

Susanne Schötz, Johan Frid, Lars Gustafsson, Anders Löfqvist

**SMASH: A Tool for Articulatory Data Processing and Analysis**

Jordan R. Green, Jun Wang, David L. Wilson

---

**Tue-Poster 10**

**Tuesday August 27, 2013**

14:00 - 16:00 Room Forum 6

**Poster 10: Topics in Speech Perception and Emotion**

Chair: **Angelika Braun**, Univ. of Trier

**Emotion Recognition of Conversational Affective Speech Using Temporal Course Modeling**

Jen-Chun Lin, Chung-Hsien Wu, Wen-Li Wei
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>The Role of Empathy in the Recognition of Vocal Emotions</td>
<td>Rene Altrov, Hille Pajupuu, Jaan Pajupuu</td>
</tr>
<tr>
<td>55</td>
<td>Electrophysiological evidence for benefits of imitation during the processing of spoken words embedded in sentential contexts</td>
<td>Angèle Brunellière and Sophie Dufour</td>
</tr>
<tr>
<td>152</td>
<td>Compensatory Speech Response to Time-Scale Altered Auditory Feedback</td>
<td>Rintaro Ogane and Masaaki Honda</td>
</tr>
<tr>
<td>161</td>
<td>Bhattacharyya Distance based Emotional Dissimilarity Measure in Multi-Dimensional Space for Emotion Classification</td>
<td>Tin Lay Nwe, Nguyen Trung Hieu, Dilip Kumar Limbu</td>
</tr>
<tr>
<td>245</td>
<td>On the Enhancement of Dereverberation Algorithms Based on a Perceptual Evaluation Criterion</td>
<td>Thiago Prego, Amaro Lima, Sergio Netto</td>
</tr>
<tr>
<td>640</td>
<td>Revisiting pitch slope and height effects on perceived duration</td>
<td>Carlos Gussenhoven and Wencui Zhou</td>
</tr>
<tr>
<td>887</td>
<td>Adaptation to natural fast speech and time-compressed speech in children</td>
<td>Hélène Guiraud, Emmanuel Ferragne, Nathalie Bedoin, Véronique Boulenger</td>
</tr>
<tr>
<td>944</td>
<td>Modeling Durational Incompressibility</td>
<td>Andreas Windmann, Juraj Simko, Britta Wrede, Petra Wagner</td>
</tr>
<tr>
<td>1136</td>
<td>Perceived prosodic correlates of smiled speech in spontaneous data</td>
<td>Caroline Emond, Lucie Ménard, Marty Laforest</td>
</tr>
<tr>
<td>1192</td>
<td>Predicting speech quality based on interactivity and delay</td>
<td>Alexander Raake, Katrin Schoenenberg, Janto Skowronek, Sebastian Egger</td>
</tr>
<tr>
<td>1467</td>
<td>Perceptual, acoustic and electroglottographic correlates of 3 aggressive attitudes in French: a pilot study</td>
<td>Charlotte Kouklia and Nicolas Audibert</td>
</tr>
</tbody>
</table>

**Tue-Poster 11**

Tuesday August 27, 2013

14:00 - 16:00 Room Forum 6

**Poster 11: Discourse and Machine Learning, Paralinguistic and Nonlinguistic Cues**

Chair: Martijn Goudbeek, Tilburg Univ.

**Theme Identification in Telephone Service Conversations using Quaternions of Speech**

Mohamed Morchid, Georges Linares, Marc El-Beze, Renato de Mori
Detection of Laughter in Children’s Speech Using Spectral and Prosodic Acoustic Features
Hrishikesh Rao, Jonathan Kim, Agata Rozga, Mark Clements

Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee
Khiet Truong

Annotation and Detection of conflict escalation in political debates
Samuel Kim, Fabio Valente, Alessandro Vinciarelli

Machine Learning of Probabilistic Phonological Pronunciation Rules from the Italian CLIPS Corpus
Florian Schiel, Mary Stevens, Uwe Reichel, Francesco Cutugno

Human perception of alcoholic intoxication in speech
Barbara Baumeister and Florian Schiel

Phonetic Manifestation and Influence of Zero Anaphora in Chinese Reading Texts
Luying Hou and Yuan Jia

Diacritics restoration for Arabic dialect texts
Salima Harrat, Mourad Abbas, Karima Meftouh, Kamel Smaili

Effects of talk-spurt silence boundary thresholds on distribution of gaps and overlaps
Marcin Wlodarczak and Petra Wagner

Final Lengthening in Russian: a Corpus-based Study
Tatiana Kachkovskaia, Nina Volskaya, Pavel Skrelin

From segmentation bootstrapping to transcription-to-word conversion
Uwe Reichel

Geneviève Caelen-Haumont and Katarina Bartkova

Non-canonical Syntactic Structures in Discourse: Tonality, Tonicity and Tones in English (Semi-)spontaneous Speech
Laetitia Leonarduzzi and Sophie Herment

Prediction of Negotiation Strategy and Outcome by Using Basic Verbal and Behavioral Features
Elnaz Nouri, Sunghyun Park, Stefan Scherer, Jonathan Gratch, Peter Carnevale, Louis-Philippe Morency, David Traum
Tue-Poster 12

Tuesday August 27, 2013

14:00 - 16:00 Room Forum 6

Poster 12: Language Identification, Speaker Diarization

Chair: Pietro Laface, Politecnico di Torino

Unsupervised Naming of Speakers in Broadcast TV: Using Written Names, Pronounced Names or Both?
Johann Poignant, Laurent Besacier, Viet Bac Le, Sophie Rosset, Georges quénot

Integer Linear Programming for Speaker Diarization and Cross-Modal Identification in TV Broadcast
Hervé Bredin and Johann Poignant

Native Accent Classification via I-Vectors and Speaker Compensation Fusion
Andrea DeMarco and Stephen J. Cox

A Free State-of-the-art Toolbox for Broadcast News Diarization
Mickael Rouvier, Grégor Dupuy, Elie Khoury, Paul Gay, Téva Merlin, Sylvain Meignier

Audio Event Classification Using Deep Neural Networks
Zvi Kons and Orith Toledo-Ronen

Code-Switching Event Detection Based on Delta-BIC Using Phonetic Eigenvoice Models
Wei-Bin Liang, Chung-Hsien Wu, Chun-Shan Hsu

Automatic Estimation of Dialect Mixing Ratio for Dialect Speech Recognition
Naoki Hirayama, Koichiro Yoshino, Katsutoshi Itoyama, Shinsuke Mori, Hiroshi G. Okuno

The Albayzin 2012 Language Recognition Evaluation
Luis Javier Rodriguez-Fuentes, Niko Brümmer, Mikel Penagarikano, Amparo Varona, Germán Bordel, Mireia Diez

TRAP Language Identification System for RATS Phase II Evaluation
Kyu Han, Sriram Ganapathy, Ming Li, Mohamed Omar, Shrikanth Narayanan

Improving Language Identification Robustness to Highly Channel-Degraded Speech through Multiple System Fusion
Aaron Lawson, Mitchell McLaren, Yun Lei, Vikramjit Mitra, Nicolas Schieffer, Luciana Ferrer, Martin Graciarena

Tue-Oral 21

Tuesday August 27, 2013

16:30 - 18:30 Room Gratte-Ciel 1

Oral 21: Metadata, Evaluation and Resources


Annotation Errors Detection in TTS Corpora
Jindrich Matousek and Daniel Tihelka
Technique for Automatic Sentence Level Alignment of Long Speech and Transcripts
Imran Ahmed and Sunil Kumar Kopparapu

Text-to-Speech Alignment of Long Recordings Using Universal Phone Models
Sarah Hoffmann and Beat Pfister

Lightly Supervised Discriminative Training of Grapheme Models for Improved Sentence-level Alignment of Speech and Text Data
Adriana Stan, Peter Bell, Junichi Yamagishi, Simon King

Automatic Social Role Recognition In Professional Meetings Using Conditional Random Fields
Ashtosh Sapru and Herve Bourlard

Same same but different - an acoustical comparison of the automatic segmentation of high quality and mobile telephone speech
Christoph Draxler and Hanna S. Feiser

---

Tue-Oral 22
Tuesday August 27, 2013
16:30 - 18:30 Room Gratte-Ciel 3

Oral 22 – Speech Synthesis: Prosody and Emotion

Chairs: Nick Campbell, Trinity College, Dublin, Emily Mower Provost, Univ. of Michigan, Ann Harbor

Multi-Centroidal Duration Generation Algorithm for HMM-Based TTS
Yongguo Kang, Yan Deng, Jian Li

Analysis and Synthesis of Shouted Speech
Tuomo Raitio, Antti Suni, Jouni Pohjalainen, Manu Airaksinen, Martti Vainio, Paavo Alku

Robust Estimation of Multiple-Regression HMM Parameters for Dimension-Based Expressive Dialogue Speech Synthesis
Tomohiro Nagata, Hiroki Mori, Takashi Nose

A New Prosody Annotation Protocol for Live Sports Commentaries
Sandrine Brognaux, Benjamin Picart, Thomas Dragman

Unsupervised Prominence Prediction for Speech Synthesis
Mahnoosh Mehrabani, Taniya Mishra, Alistair Conkie

Expressive speech synthesis in MARY TTS using audiobook data and EmotionML
Marcela Charfuelan and Ingmar Steiner
**Tue-Oral 23**

Tuesday August 27, 2013

16:30 - 18:30 Room Prestige Gratte-Ciel

**Oral 23: Spoken Language Information Retrieval**

Chairs: Giuseppe Di Fabbrizio, Amazon, Cambridge, Haizhou Li, Institute for Infocomm Res., Singapore

**Using Dialog-Activity Similarity for Spoken Information Retrieval**

Nigel Ward and Steven Werner

**A Hybrid HMM/DNN Approach to Keyword Spotting of Short Words**

I-Fan Chen and Chin-Hui Lee

**Leveraging Locality for Topic Identification of Conversational Speech**

Jonathan Wintrode

**Person name spotting by combining acoustic matching and LDA topic models**

Grégory Senay, Benjamin Bigot, Richard Dufour, Georges Linares, Corinne Fredouille

**Using Phonological Phrase Segmentation to Improve Automatic Keyword Spotting for the Highly Agglutinating Hungarian Language**

Gyorgy Szaszak and Andras Beke

**Leveraging Knowledge Graphs for Web-Scale Unsupervised Semantic Parsing**

Larry Heck, Dilek Hakkani-Tur, Gokhan Tur

---

**Tue-Oral 24**

Tuesday August 27, 2013

16:30 - 18:30 Room Tête d’Or 2

**Oral 24: Speaker Recognition**

Chairs: Tomi Kinnunen, Univ. of Eastern Finland, Joensuu, Nicolas Scheffer, SRI, Menlo Park

**Fast and Memory Effective I-Vector Extraction using a Factorized Subspace**

Sandro Cumani and Pietro Laface

**Effective estimation of a multi-speaker model using information on signal parameters**

Konstantin Simonchik, Andrey Shulipa, Timur Pekhovskiy

**Automatic regularization of cross-entropy cost for speaker recognition fusion**

Ville Hautamaki, Kong Aik Lee, David van Leeuwen, Rahim Saeidi, Anthony Larcher, Tomi Kinnunen, Taufiq Hasan, Seyed Omid Sadjadi, Gang Liu, Hynek Boril, John H.L. Hansen, Benoit Fauve

**Speaker verification based on fusion of acoustic and articulatory information**

Ming Li, Jangwon Kim, Prasanta Kumar Ghosh, Vikram Ramanarayanan, Shrikanth Narayanan

**The distribution of calibrated likelihood-ratios in speaker recognition**

David van Leeuwen and Niko Brummer
Tue-Oral 25
Tuesday August 27, 2013
16:30 - 18:30 Room Bellecour 2

Oral 25: Multimodal Speech Perception

Chairs: Chris Davis, MARCS Inst., Sydney, Lucie Ménard, UQAM, Montréal

1295 Effects of Mouth-Only and Whole-Face Displays on Audio-Visual Speech Perception in Noise: Is the Vision of a Talker’s Full Face Truly the Most Efficient Solution?
Grozdana Erjavec and Denis Legros

Acoustic and visual phonetic features in the McGurk effect – an audiovisual speech illusion
Kaisa Trippana, Mikko Tiainen, Lari Vainio, Martti Vainio

The effect of visual speech timing and form cues on the processing of speech and nonspeech
Chris Davis and Jeesun Kim

Effect of context, rebinding and noise on audiovisual speech fusion
Ganesh Attigodu Chandrashekara, Frédéric Berthommier, Olha Nahorna, Jean Luc Schwartz

Social face to face communication – American English attitudinal prosody
Albert Rilliard, Donna Erickson, Takaaki Shochi, João Antônio de Moraes

Inter-personnal adaptation of respiratory patterns in read speech
Gerard Bailly, Amélie Rochet-Capellan, Coriandre Vilain

Tue-Poster 13
Tuesday August 27, 2013
16:30 - 18:30 Room Forum 6

Poster 13: Speech Analysis

Chair: Tom Quatieri, MIT, Cambridge

68 A comparative study of glottal open quotient estimation techniques
John Kane, Stefan Scherer, Louis-Philippe Morency, Christer Gobl

Estimation of multiple-branch vocal tract models: the influence of prior assumptions
Christian H. Kasess and Wolfgang Kreuzer

Detecting Overlapping Speech with Long Short-Term Memory Recurrent Neural Networks
Jürgen T. Geiger, Florian Eyben, Björn Schuller, Gerhard Rigoll
Evaluation of Fundamental Validity in Applying AR-HMM with Automatic Topology Generation to Pathology-Voice Analysis
Akira Sasou

Significance of Instants of Significant Excitation for Source Modeling
Nagaraj Adiga and S R Mahadeva Prasanna

Significance of variable height-bandwidth group delay filters in the spectral reconstruction of speech
Devanshu Arya, Anant Raj, Rajesh Hegde

Nonlinear Prediction of Speech Signal using Volterra-Wiener Series
Hemant Arjun Patil and Tanvina Patel

Evaluation of Speech-Based Protocol for Detection of Early-Stage Dementia
Aharon Satt, Alexander Sorin, Orith Toledo-Ronen, Ioannis Kompatsiaris, Athina Kokonozi, Magda Tsolaki, Oren Barkan

Instantaneous Harmonic Representation of Speech Using Multicomponent Sinusoidal Excitation
Elias Azarov, Maxim Vashkevich, Alexander Petrovsky

A Quantitative Comparison of Glottal Closure Instant Estimation Algorithms on a Large Variety of Singing Sounds
Onur Babacan, Thomas Drugman, Nicolas d'Alessandro, Nathalie Henrich, Thierry Dutoit

Automatic gender recognition in normal and pathological speech
Jorge Andrés Gómez García, Juan Ignacio Godino Ilorente, Germán Castellanos-Domínguez

Unsupervised Vocal-Tract Length Estimation Through Model-based Acoustic-to-Articulatory Inversion
Shanqing Cai, H Timothy Bunnell, Rupal Patel

Model Order Estimation Using Bayesian NMF for Discovering Phone Patterns in Spoken Utterances
Sayeh Mirzaei, Hugo Van hamme, Yaser Norouzi

---

**Tue-Poster 14**

**Tuesday August 27, 2013**
16:30 - 18:30 Room Forum 6

**Poster 14: ASR – Feature Extraction**

Chair: Long Nguyen, BBN Technologies, Cambridge

Convolutional Deep Rectifier Neural Nets for Phone Recognition
Laszlo Toth

Pitch Synchronous Spectral Analysis for a Pitch Dependent Recognition of Voiced Phonemes - PISAR
Hans-Guenter Hirsch
New parameters for Automatic Speech Recognition based on the mammalian cochlea Model using resonance analysis
José Luis Oropeza Rodríguez

Using an autoencoder with deformable templates to discover features for automatic speech recognition
Navdeep Jaitly and Geoffrey Hinton

Speaking Rate Normalization with Lattice-based Context-dependent Phoneme Duration Modeling for Personalized Speech Recognizers on Mobile Devices
Ching-Feng Yeh, Hung-Yi Lee, Lin-Shan Lee

Subspace Models for Bottleneck Features
Jun Qi, Dong Wang, Javier Tejedor

Bottleneck Features based on Gammatone Frequency Cepstral Coefficients
Jun Qi, Dong Wang, Ji Xu, Javier Tejedor

Cross-Entropy vs. Squared Error Training: a Theoretical and Experimental Comparison
Pavel Golik, Patrick Doetsch, Hermann Ney

Acoustic Features for Detection of Phonemic Aspiration in Voiced Plosives
Vaishali Patil and Preeti Rao

Estimating Phoneme Class Conditional Probabilities from Raw Speech Signal using Convolutional Neural Networks
Dimitri Palaz, Ronan Collobert, Mathew Magimai-Doss

Hierarchical models based on a continuous acoustic space to identify phonological features
Javier Mikel Olaso and Maria Inès Torres

Locality Sensitive Hashing for Fast Computation of Correlational Manifold Learning based Feature space Transformations
Vikrant Tomar and Richard Rose

Evaluating speech features with the Minimal-Pair ABX task: Analysis of the classical MFC/PLP pipeline
Thomas Schatz, Vijayaditya Peddinti, Francis Bach, Aren Jansen, Hynek Hermansky, Emmanuel Dupoux

---

**Tue-Poster 15**

**Tuesday August 27, 2013**

**Poster 15: ASR – Pronunciation, Prosodic and New Paradigms**

Chair: Tanja Schultz, Karlsruhe Institute of Technology

Knowledge Integration for Improving Performance in LVCSR
Chen-Yu Chiang, Sabato Marco Siniscalchi, Sin-Horng Chen, Chin-Hui Lee
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>138</td>
<td>Inter-Speaker Variability in Audio-Visual Classification of Word Prominence</td>
<td>Martin Heckmann</td>
</tr>
<tr>
<td>167</td>
<td>Parameter Clustering for Temporally Varying Weight Regression for Automatic Speech Recognition</td>
<td>Khe Chai Sim</td>
</tr>
<tr>
<td>214</td>
<td>Phone Duration Modeling Using Clustering of Rich Contexts</td>
<td>Tanel Alumäe and Rena Nemoto</td>
</tr>
<tr>
<td>526</td>
<td>Human Mouth State Detection Using Low-Frequency Ultrasound</td>
<td>Farzaneh Ahmadi, Mousa Ahmadi, Ian McLoughlin</td>
</tr>
<tr>
<td>539</td>
<td>Lexical Stress Detection for L2 English Speech Using Deep Belief Networks</td>
<td>Kun Li, Xiaojun Qian, Shiyan Kang, Helen Meng</td>
</tr>
<tr>
<td>763</td>
<td>MLP-HMM Two-Stage Unsupervised Training for Low-Resource Languages on Conversational Telephone Speech Recognition</td>
<td>Yanmin Qian and Jia Liu</td>
</tr>
<tr>
<td>812</td>
<td>Failure transitions for Joint n-gram Models and G2P Conversion</td>
<td>Josef Novak, Nobuaki Minematsu, Keikichi Hirose</td>
</tr>
<tr>
<td>1043</td>
<td>Generative modeling of speech F0 contours</td>
<td>Hirokazu Kameoka, Kota Yoshizato, Tatsuma Ishihara, Yasunori Ohishi, Kunio Kashino, Shigeki Sagayama</td>
</tr>
<tr>
<td>1211</td>
<td>G2P variant prediction techniques for ASR and STD</td>
<td>Marelie Davel, Charl van Heerden, Etienne Barnard</td>
</tr>
<tr>
<td>1298</td>
<td>Rhythm Analysis of Second-Language Speech through Low-Frequency Auditory</td>
<td>Jing Zheng and Joseph Tepperman</td>
</tr>
<tr>
<td>1342</td>
<td>Graph-Based Semi-supervised Learning for Phone and Segment Classification</td>
<td>Yuzong Liu and Katrin Kirchhoff</td>
</tr>
<tr>
<td>1472</td>
<td>Selective use of gaze information to improve ASR performance in noisy environments by cache-based class language model adaptation</td>
<td>Ao Shen, Neil Cooke, Martin Russell</td>
</tr>
<tr>
<td>1539</td>
<td>Deep Segmental Neural Network for Automatic Speech Recognition</td>
<td>Ossama Abdel-Hamid, Li Deng, Dong Yu, Hui Jiang</td>
</tr>
<tr>
<td>1269</td>
<td>Quantifying cross-linguistic variation in grapheme-to-phoneme mapping</td>
<td>Martine Coene, Annemiek Hammer, Wojtek Kowalczyk, Louis ten Bosch, Bart Vaerenberg, Paul Govaerts</td>
</tr>
</tbody>
</table>
Show & Tell 2
Tuesday August 27, 2013
16:30 - 18:30 Salon Tête d'Or

Chairs: Laurence Devillers, LIMSI-CNRS/Paris-Sorbonne 4, Fabrice Lefèvre, LIA, Avignon

The Speech Recognition Virtual Kitchen
Florian Metze, Eric Fosler-Lussier, Rebecca Bates

Multilingual Web Conferencing Using Speech-to-Speech Translation
John Chen, Shufei Wen, Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore

ROCme! Software for the Recording and Management of Speech Corpora
Emmanuel Ferragne, Sébastien Flavier, Christian Fressard

Voice Search in Mobile Applications with the Rootvoole Framework
Felix Burkhardt

On-Line Audio Dilation for Human Interaction
John S. Novak III, Jason Archer, Valeriy Shafiro, Robert Kenyon, Jason Leigh

Phase-Aware Single-Channel Speech Enhancement
Pejman Mowlaee, Mario Kaoru Watanabe, Rahim Saeidi

A Free Online Accent and Intonation Dictionary for Teachers and Learners of Japanese
Hiroko Hirano, Ikuki Nakamura, Nobuaki Minematsu, Masayuki Suzuki, Cheiko Nakagawa, Noriko Nakamura, Yukinori Tagawa, Keikichi Hirose, Hiroya Hashimoto

Reactive Accent Interpolation Through an Interactive Map Application
Maria Astrinaki, Junichi Yamagishi, Simon King, Nicolas d’Alessandro, Thierry Dutoit

A Tool to Elicit and Collect Multicultural and Multimodal Laughter
Mariette Soury, Clément Gossart, Martine Adda, Laurence Devillers

Design of a Mobile App for Interspeech Conferences: Towards an Open Tool for the Spoken Language Community
Robert Schleicher, Tilo Westermann, Jinjin Li, Moritz Lawitschka, Benjamin Mateev, Ralf Reichmuth, Sebastian Müller

A Non-Experts User Interface for Obtaining Automatic Diagnostic Spelling Evaluations for Learners of the German Writing System
Kay Berkling

ISCA GENERAL ASSEMBLY
Tuesday August 27, 2013
18:30 - 19:45 Amphithéâtre

Keynote 4: Maureen Clerc
Wednesday, August 28, 2013
9:00 - 10:00 Amphithéâtre

Verbal Communication Through Brain Computer interfaces
Chair: Roger Moore, Univ. of Sheffield

Wed-Oral 26
Wednesday, August 28, 2013
10:30 - 12:30 Room Gratte-Ciel 1

Oral 26: Dialog Systems
Chairs: Gary Geunbae Lee, Postech, Pohang, Sophie Rosset, LIMSI/CNRS, Orsay
Estimation of Interest and Comprehension Level of Audience through Multi-modal Behaviors in Poster Conversations
Tatsuya Kawahara, Soichiro Hayashi, Katsuya Takanashi

A New DNN-based High Quality Pronunciation Evaluation for Computer-Aided Language Learning (CALL)
Wenping Hu, Yao Qian, Frank K. Soong

A Multi-domain Dialog System to integrate heterogeneous Spoken Dialog Systems
Joaquin Planells Lerma, Lluis-F. Hurtado, Encarna Segarra, Emilio Sanchis

Development and Evaluation of Spoken Dialog Systems with One Agent and Two Agents
Yuki Todo, Ryota Nishimura, Kazumasa Yamamoto, Seiichi Nakagawa

User feedback in human-robot interaction: Prosody, gaze and timing
Gabriel Skantze, Catharine Oertel, Anna Hjalmarsson

KPCatcher – A Keyphrase Extraction System for Enterprise Videos
Yongxin Xi, Matthias Paulik, Venkata Gadde, Ananth Sankar

---

**Wed-Oral 27**

Wednesday, August 28, 2013
10:30 - 12:30  Room Gratte-Ciel 3

**Oral 27: Speech Analysis III**

Chairs: Hideki Kawahara, Wakayama Univ., Yannis Stylianou, Univ. of Crete, Heraklion

**Pitch-Gesture Modeling Using Subband Autocorrelation Change**
Malcolm Slaney, Elizabeth Shriberg, Jui-Ting Huang

**Analysis of emotional speech at subsegmental level**
Gangamohan Paidi, Sudarsanareddy Kadiri, Yegnanarayana Bayya

**Periodicity extraction for voiced sounds with multiple periodicity**
Masanori Morise, Hideki Kawahara, Kenji Ozawa

**Modelling and estimation of the fundamental frequency of speech using a hidden Markov model**
John Taylor and Ben Milner

**Extended Weighted Linear Prediction Using the Autocorrelation Snapshot - A Robust Speech Analysis Method and its Application to Recognition of Vocal Emotions**
Jouni Pohjalainen and Paavo Alku

**Improving the Accuracy and the Robustness of Harmonic Model for Pitch Estimation**
Meysam Asgari and Izhak Shafran
**Wed-Oral 28**

**Wednesday, August 28, 2013**

10:30 - 12:30  Room Prestige Gratte-Ciel

**Oral 28: ASR – Pronunciation Variants and Modeling**

Chairs: Frederic Bechet, Univ. Aix Marseille, Eric Fosler-Lussier, Ohio State Univ.

**Discriminative pronunciation modeling based on minimum phone error training**

364

Meixu Song, Qingqing Zhang, Jielin Pan, Yonghong Yan

**Grapheme-to-phone conversion based on adaptive regularization of weight vectors**

983

Keigo Kubo, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura

**An efficient method to estimate pronunciation from multiple utterances**

991

Tofigh Naghibi, Sarah Hoffmann, Beat Pfister

**Category-based phoneme-to-grapheme transliteration**

1111

Willem Basson and Marelie Davel

**Discriminative training of WFST factors with application to pronunciation modeling**

1114

Preethi Jyothi, Eric Fosler-Lussier, Karen Livescu

**Discriminative training of a phoneme confusion model for a dynamic lexicon in ASR**

1264

Penny Karanasou, François Yvon, Thomas Lavergne, Lori Lamel

---

**Wed-Oral 29**

**Wednesday, August 28, 2013**

10:30 - 12:30  Room Tête d'Or 2

**Oral 29: Speaker Recognition Evaluation**

Chairs: Niko Brummer, Agnitio Research, Somerset West, Patrick Kenny, CRIM, Montréal

**The 2012 NIST Speaker Recognition Evaluation**

1544

Craig Greenberg, Vincent Stanford, Alvin Martin, Meghana Yadagiri, George Doddington, John Godfrey, Jaime Hernandez-Cordero

**Likelihood-ratio calibration using prior-weighted proper scoring rules**

81

Niko Brummer and George Doddington

**A Noise-Robust System for NIST 2012 Speaker Recognition Evaluation**

298

Luciana Ferrer, Mitchell McLaren, Nicolas Scheffer, Yun Lei, Martin Graciarena, Vikramjit Mitra

**IU submission to NIST SRE 2012: A large-scale collaborative effort for noise-robust speaker verification**

693

**Improved Unsupervised NAP Training Dataset Design for Speaker Recognition**

Hanwu Sun and Bin Ma

---

**Nuance - Politecnico di Torino's 2012 NIST Speaker Recognition Evaluation System**

Daniele Colibro, Claudio Vair, Kevin Farrell, Gennady Karvitsky, Nir Krause, Sandro Cumani, Pietro Laface

---

**Wednesday, August 28, 2013**

**10:30 - 12:30 Room Bellecour 2**

**Oral 30: Physiology and Models of Speech Production**

**Chairs:** Masaaki Honda, Waseda Univ., Tokyo, Amélie Rochet Capellan, GIPSA-lab, Grenoble

**A perceptually and physiologically motivated voice source model**

Gang Chen, Marc Garellek, Jody Kreiman, Bruce Gerratt, Abeer Alwan

** Stable articulatory tasks and their variable formation: Tamil retroflex consonants**

Caitlin Smith, Michael Proctor, Khalil Iskarous, Louis Goldstein, Shrikanth Narayanan

**Articulatory settings facilitate mechanically advantageous motor control of vocal tract articulators**

Vikram Ramanarayanan, Adam Lammert, Louis Goldstein, Shrikanth Narayanan

**The interplay of linguistic structure and breathing in German spontaneous speech**

Amelie Rochet-Capellan and Susanne Fuchs

**Physical Models of the Vocal Tract with a Flapping Tongue for Flap and Liquid Sounds**

Takayuki Arai

**Articulatory copy synthesis from cine X-ray films**

Yves Laprie, Matthieu Loosvelt, Shinji Maeda, Rudolph Sock, Fabrice Hirsch

---

**Wednesday, August 28, 2013**

**10:30 - 12:30 Room Bellecour 1**

**Special Session 4: Speech Science in End-User Applications**

**Chairs:** Felix Burkhardt, Telekom Innovation Lab., Berlin, Juergen Schroeter, ATT, Florham Park

**Accepted Papers**

**Large-Scale Personal Assistant Technology Deployment: the Siri Experience**

Jérôme Bellagarda

**Evaluating an adaptive dialog system for the public**

Benjamin Weiss, Simon Willkomm, Sebastian Möller
Self-taught assistive vocal interfaces: An overview of the ALADIN project
Jort F. Gemmeke, Bart Ons, Netsanet Tessema, Hugo Van hamme, Janneke van de Loo, Guy De Pauw, Walter Daelemans, Jonathan Huyghe, Jan Derboven, Lode Vuegen, Bert Van Den Broeck, Peter Karsmakers, Bart Vanrumste

Affect recognition in real-life acoustic conditions -- A new perspective on feature selection
Florian Eyben, Felix Weninger, Björn Schuller

A distributed system for recognizing home automation commands and distress calls in the Italian language
Emanuele Principi, Stefano Squartini, Francesco Piazza, Danilo Fuselli, Maurizio Bonifazi

Probabilistic Trainable Segmenter for Call Center Audio Using Multiple Features
Nina Zinovieva, Xiaodan Zhuang, Pat Peterson, Joe Alwan, Rohit Prasad

ADDITIONAL CONTRIBUTIONS
Voice Search in Mobile Applications: Strategies, Evaluations and the Use of Linked Open Data
Felix Burkhardt and Hans Ulrich Nägeli

Evaluation of a Real-Time Voice Order Recognition System from Multiple Audio Channels in a Home
Michel Vacher, Benjamin Lecouteux, Dan Istrate, Thierry Joubert, François Portet, Mohamed Sehili, Pedro Chahuara

In-home detection of distress calls: the case of aged users
Frédéric Aman, Michel Vacher, Solange Rossato, François Portet

Data Driven Methods for Utterance Semantic Tagging
Ding Liu, Anthea Cheung, Anna Margolis, Patrick Redmond, Jun-Won Suh, Chao Wang

The AT&T Speech API A Study on Practical Challenges for Customized Speech to Text Service
Evandro Gouvea, Antonio Moreno-Daniel, Aarthi Reddy, Rathinavelu Chengalvarayan, David Thomson, Andrej Ljolje

In-Vehicle Destination Entry by Voice: Practical Aspects
Bart D'hoore and Alfred Wiesen

Wed-Poster 16
Wednesday, August 28, 2013
10:30 - 12:30 Room Forum 6
Poster 16: Perception of Non Native Sounds

Chair: Ricardo Gutierrez-Osuna, Texas A & M Univ., College Station
Intelligibility at a multilingual cocktail party: Effect of concurrent language knowledge
Aurore Gautreau, Michel Hoen, Fanny Meunier

Regional accents affect speech intelligibility in a multitalker environment
Ewa Jacewicz and Robert Fox
Perception of English minimal pairs in noise by Japanese listeners: does clear speech for L2 listeners help?
Shinichi Tokuma and Won Tokuma

Salento Italian listeners’ perception of American English vowels
Bianca Sisinni, Paola Escudero, Mirko Grimaldi

TP 3.1 Software: A Tool for Designing Audio, Visual, and Audiovisual Perceptual Training Tasks and Perception Tests
Andreia Rauber, Anabela Rato, Denise Kluge, Giane Santos

Effect of linguistic masker on the intelligibility of Mandarin sentences
Fei Chen, Junfeng Li, Lena Wong, Yonghong Yan

The Learning and Generalization of Contrasts Consistent or Inconsistent with Native Biases
Kyuwon Moon and Meghan Sumner

L2 English Learners’ Recognition of Words Spoken in Familiar versus Unfamiliar English Accents
Jia Ying, Jason Shaw, Catherine Best

The Effects of Perceptual and/or Productive Training on the Perception and Production of English Vowels /ɪ/ and /iː/ by Cantonese ESL learners
Janice Wing Sze Wong

On the role of L1 speech production in L2 perception: Evidence from Spanish learners of French
Natalia Kartushina and Ulrich Hans Frauenfelder

Looking for lexical feedback effects in /tl/ > /kl/ repairs
Pierre Halle, Natalia Kartushina, Juan Segui, Ulricht Frauenfelder

Recognizing words across regional accents: The role of perceptual assimilation in lexical competition
Catherine Best, Jason Shaw, Elizabeth Clancy

---

Wed-Poster 17

Wednesday, August 28, 2013
10:30 - 12:30 Room Forum 6

Poster 17: Speech Disorders – Data and Methodology

Chair: Heidi Christensen, Univ. of Sheffield

Dysarthria Intelligibility Assessment in a Factor Analysis Total Variability
David Martínez González, Phil Green, Heidi Christensen

Perceptual interference between regional accent and voice/speech disorders
Alain Ghio, Médéric Gasquet-Cyrus, Juliette Roquel, Antoine Giovanni
Linguistic disfluency in narrative speech: Evidence from story-telling in 6-year-olds
Ingrida Balciuniene

Assessing the Utility of Judgments of Children’s Speech Production Made by Untrained Listeners in Uncontrolled Listening Environments
Benjamin Munson

Consonant distortions in dysarthria due to Parkinson’s disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia
Tanja Kocjančič Antolík and cécile fougeron

Study of coarticulation and F2 transitions in French and Italian adult stutterers
Marine Verdurand-Pen deliau, Solange Rossato, Lionel Granjon, Daria Balbo, Claudio Zmarich

Automatic Tracheoesophageal Voice Typing Using Acoustic Parameters
Renee Clapham, Corina Van As-Brooks, Michiel Van den Brekel, Frans Hilgers, Rob van Son

Burst-based Features for the Classification of Pathological Voices
Julie Mauclair, Lionel Koenig, Marina Robert, Peggy Gatignol

Classification of depression state based on articulatory precision
Brian Helfer, Thomas Quatieri, James Williamson, Daryush Mehta, Rachelle Horwitz, Beaux Yu

Using text and acoustic features to diagnose progressive aphasia and its subtypes
Kathleen C. Fraser, Frank Rudzicz, Elizabeth Rochon

---

**Wed-Poster 18**

**Wednesday, August 28, 2013**

10:30 - 12:30 Room Forum 6

**Poster 18: Search and Computational Issues in LVCSR**

Chair: Thomas Hain, Univ. of Sheffield

Multi-domain Neural Network Language
Tanel Alumäe

Improving Lightly Supervised Training for Broadcast Transcription
Yanhua Long, Gales Mark, Pierre Lanchantin, Xunying Liu, Matthew Seigel, Phil Woodland

Weakly supervised parsing with rules
Christophe Cerisara, Alejandra Lorenzo, Pavel Kral

Relative Error Bounds for Statistical Classifiers based on the f-Divergence
Markus Nussbaum-Thom, Eugen Beck, Tamer Alkhouli, Ralf Schlüter, Hermann Ney
Experiments towards a better LVCSR System for Tamil
Melvin Jose Johnson Premkumar, Ngoc Thang Vu, Tanja Schultz

A Hybrid Language Model for Open-Vocabulary Thai LVCSR
Kwanchiva Thangthai, Ananlada Chotimongkol, Chai Wuttiwiwatchai

Hierarchical Pitman-Yor and Dirichlet Process for Language Model
Jen-Tzung Chien and Ying-Lan Chang

Unsupervised Confidence Calibration Using Examples of Recognized Words and Their Contexts
Taichi Asami, Satoshi Kobashikawa, Hirokazu Masataki, Osamu Yoshioka, Satoshi Takahashi

Multilingual Hierarchical MRASTA Features for ASR
Zoltán Tüske, Ralf Schlüter, Hermann Ney

Heuristic Selection of Training Sentences from Historical TV Guide for Semi-supervised LM Adaptation
Harry Chang

Combination of Random Indexing based Language Model and N-gram Language Model for Speech Recognition
Dominique Fohr and Odile Mella

Improving Low-Resource CD-DNN-HMM using Dropout and Multilingual DNN Training
Yajie Miao and Florian Metze

Finding Recurrent Out-of-Vocabulary Words
Long Qin and Alexander Rudnicky

Using Conversational Word Bursts in Spoken Term Detection
Justin Chiu and Alexander Rudnicky

Wed-Oral 31
Wednesday, August 28, 2013
14:00 - 16:00 Room Gratte-Ciel 1

Oral 31: Speech and Hearing Disorders

Chairs: Suzanne Boyce, Univ. of Cincinnati, Valerie Hazan, UCL, London

Brain activations in speech recovery process after intra-oral surgery: an fMRI study
Audrey Acher, Marc Sato, Laurent Lamalle, Coriandre Vilain, Arnaud Attye, Alexandre Krainik, Georges Bettega, Christian Adrien Righini, Brice Carlot, Muriel Brix, Pascal Perrier

Acoustic and perceptual analysis of vocal tremor
Christophe Mertens, Jean Schoentgen, Francis Grenze, Sabine Skodda
<table>
<thead>
<tr>
<th>Session</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral 32</td>
<td>Lexical Tone Perception in Thai Normal-Hearing Adults and Those Using Hearing Aids: A Case Study</td>
<td>Charturong Tantibundhut, Chutamanee Onsawan, Nittayapa Klangpornkun, Phongphhan Phunphanich, Tanawan Saimai, Nuntaporn Saimai, Pitthapon Pitthawahatchai</td>
</tr>
<tr>
<td></td>
<td>Evaluation of a Bone-Conducted Ultrasonic Hearing Aid in Vocal Emotion</td>
<td>Takayuki Kagomiya and Seiji Nakagawa</td>
</tr>
<tr>
<td></td>
<td>Processing of /i/ and /u/ in Italian cochlear-implant children: a behavioral and neurophysiologic study</td>
<td>Luigia Garrapa, Davide Bottari, Mirko Grimaldi, Francesco Pavani, Andrea Calabrese, Michele De Benedetto, Silvano Vitale</td>
</tr>
<tr>
<td></td>
<td>Predicting the Bilateral Advantage in Cochlear Implantees using a Non-Intrusive Speech Intelligibility</td>
<td>Stefano Cosentino, Tiago Falk, David McAlpine</td>
</tr>
</tbody>
</table>

**Wed-Oral 32**  
Wednesday, August 28, 2013
14:00 - 16:00 Room Gratte-Ciel 3

**Oral 32: Speech and Audio Segmentation**


**143**  
A Blind Segmentation Approach to Acoustic Event Detection Based on I-Vector  
Zhen Huang, You-Chi Cheng, Kehuang Li, Ville Hautamäki, Chin-Hui Lee

**168**  
A dynamic programming framework for neural network-based automatic speech segmentation  
Van Zyl van Vuuren, Louis ten Bosch, Thomas Niesler

**440**  
Acoustic segmentation of speech using zero time liftering (ZTL)  
Ravi Shankar Prasad and Yegnanarayana Bayya

**839**  
Unsupervised Mining of Acoustic Subword Units With Segment-level Gaussian Posteriorgrams  
Haipeng Wang, Tan Lee, Cheung-Chi Leung, Bin Ma, Haizhou Li

**1305**  
Combination of Auditory Attention Features with Phone Posteriors for Better Automatic Phoneme Segmentation  
Ozlem Kalinli

**1430**  
Automatic Phonetic Segmentation using Boundary Models  
Jiahong Yuan, Neville Ryan, Mark Liberman, Andreas Stolcke, Vikramjit Mitra, Wen Wang

**Wed-Oral 33**  
Wednesday, August 28, 2013
14:00 - 16:00 Room Prestige Gratte-Ciel

**Oral 33: Speech Synthesis – Various Topics**

Chairs: Keiichi Tokuda, Nagoya Inst. of Tech, Petra Wagner, Univ. of Bielefeld
HMM-based TTS for Hanoi Vietnamese: Issues in Design and Evaluation
Thu Thu Trang NGUYEN, Christophe D’ALESSANDRO, Albert RILLIARD, Do Dat TRAN

HMM-based synthesis of creaky voice
Tuomo Raitio, John Kane, Thomas Drugman, Christer Gobl

Integrating Conditional Random Fields and Joint Multi-gram Model with Syllabic features for Grapheme-to-Phone Conversion
Xiaoxuan Wang and Khe Chai Sim

Structure learning in Hidden Conditional Random Fields for Grapheme-to-Phoneme conversion
Patrick Lehnen, Alexander Allauzen, Thomas Lavergne, Francois Yvon, Stefan Hahn, Hermann Ney

TUNDRA: A Multilingual Corpus of Found Data for TTS Research Created with Light Supervision
Adriana Stan, Oliver Watts, Yoshitaka Mamiya, Mircea Giurgiu, Rob Clark, Junichi Yamagishi, Simon King

Minimum mean squared error based warped complex cepstrum analysis for statistical parametric speech synthesis
Ranniery Maia, Mark Gales, Yannis Stylianou, Masami Akamine

---

**Wed-Oral 34**

**Wednesday, August 28, 2013**

14:00 - 16:00  Room Tête d'Or 2

**Oral 34: ASR – Discriminative Training**

Chairs: Karen Livescu, TTI-Chicago, Phil Woodland, Cambridge Univ.

Augmented Conditional Random Fields Modeling Based on Discriminatively Trained Features
Yasser Hifny

Sequence-discriminative training of deep neural networks
Karel Vesely, Arnab Ghoshal, Lukas Burget, Dan Povey

Discriminatively Trained sparse Inverse Covariance Matrices for Low Resource Speech Acoustic Modeling
Weibin Zhang

Discriminative training of acoustic models for system combination
Yuuki Tachioka and Shinji Watanabe

Semi-Supervised GMM and DNN Acoustic Model Training with Multi-system Combination and Confidence Re-calibration
Yan Huang, Dong Yu, Yifan Gong, Chaojun Liu

Restructuring of Deep Neural Network Acoustic Models with Singular Value Decomposition
Jian Xue, Jinyu Li, Yifan Gong
**Wed-Oral 35**

**Wednesday, August 28, 2013**

14:00 - 16:00 Room Bellecour 2

**Oral 35: L2 Acquisition, Multilingualism**

Chairs: Kathy Huet, Univ. of Mons, Caroline Smith, Univ. of New Mexico

**Languages**

93  
Large-Scale Characterization of Mandarin Pronunciation Errors Made by Native Speakers of European Languages  
Nancy Chen, Vivaek Shivakumar, Mahesh Harikumar, Bin Ma, Haizhou Li

111  
Production training in Second Language Acquisition: A comparison between objective measures and subjective judgments  
Veronique Delvaux, Kathy Huet, Myriam Piccaluga, Bernard Harmegnies

**Program**

712  
The Production and Perception of Voice Onset Time in English-Speaking Children Enrolled in a French immersion  
Nicole Netelenbos and Fangfang Li

**Pronunciation errors by Spanish learners of Dutch: A data-driven study for ASR-based pronunciation training**

802  
Pepi Burgos, Catia Cucchiarini, Roeland van Hout, Helmer Strik

**Realisation of tonal alignment in the English of Japanese-English late bilinguals**

1603  
Calbert Graham and Brechtje Post

**The Influence of language and speech task upon creaky voice use among six young American women learning French**

1386  
Agathe Benoist-lucy and Claire Pillot-loiseau

---

**Wed-SP5**

**Wednesday, August 28, 2013**

14:00 - 16:00 & 16:30 - 18:30 Room Bellecour 1

**Special Session 5 (a & b): Child Computer Interaction**


**Acoustic-Prosodic, Turn-taking, and Language Cues in Child-Psychologist Interactions for Varying Social Demand**

233  
Daniel Bone, Chi-Chun Lee, Theodora Chaspari, Matthew Black, Marian Williams, Sungbok Lee, Pat Levitt, Shrikanth Narayanan

**A Preliminary Study of Child Vocalization on a Parallel Corpus of US and Shanghainese Toddlers**

1445  
Hynek Boril, Qian Zhang, Pongtep Angkititrakul, John H.L. Hansen, Dongxin Xu, Gill Gilkerson, Jeffrey A. Richards

**A Survey about Databases of Children's Speech**

908  
Felix Claus, Hamurabi Gamboa Rosales, Rico Petrick, Horst-Udo Hain, Rüdiger Hoffmann

**Affective Evaluation of Multimodal Dialogue Games for Preschoolers Using Physiological Signals**

1188  
Vassiliki Kouloumenta, Manolis Perakakis, Alexandros Potamianos

**Amplitude Modulation Features for Emotion Recognition from Speech**

1616  
Md Jahangir Alam, Yazid Attabi, Pierre Dumouchel, Patrick Kenny, Douglas O'Shaughnessy
Analyzing Eye-Voice Coordination in Rapid Automatized Naming
Daniel Bone, Chi-Chun Lee, Vikram Ramanarayanan, Shrikanth Narayanan, Renske Hoedemaker, Peter Gordon

Analyzing the structure of parent-moderated narratives from children with ASD using an entity-based approach
Theodora Chaspari, Emily Provost, Shrikanth Narayanan

Automated speech scoring for non-native middle school students with multiple task types
Keelan Evanini and Xinhao Wang

Identification of Gender from Children’s Speech by Computers and Humans
Saeid Safavi, Peter Jancovic, Martin Russell, Michael Carey

On Why Japanese /r/ Sounds are Difficult for Children to Acquire
Takayuki Arai

Wed-Poster 19
Wednesday, August 28, 2013
14:00 - 16:00  Room Forum 6

Poster 19: Speaker Recognition

Chair: Craig Greenberg, NIST, Gaithersburg

Anchor and UBM-based Multi-Class MLLR M-Vector System for Speaker Verification
Achintya Sarkar and Claude Barras

Ensemble approach in Speaker Verification
Leibny Paola Garcia Perera, Bhiksha Raj, Juan Arturo Nolazco Flores

Sequential Model Adaptation for Speaker Verification
Jun Wang, Dong Wang, Xiaojun Wu, Thomas Fang Zheng, Javier Tejedor

Improving Short Utterance based i-vector Speaker Recognition using Source and Utterance-Duration Normalization Techniques
Ahilam Kanagasundaram, David Dean, Javier Gonzalez-Dominguez, Sridha Sridharan, Daniel Ramos, Joaquin Gonzalez-Rodriguez

On Leveraging Conversational Data for Building a Text Dependent Speaker Verification System
Hagai Aronowitz and Oren Barkan

THU-EE System Fusion for the NIST 2012 Speaker Recognition Evaluation
Wei-Qiang Zhang, Zhi-Yi Li, Weiwei Liu, Jia Liu

Subspace-constrained Supervector PLDA for Speaker Verification
Daniel Garcia-Romero and Alan McCree
Augmenting Short-term Cepstral Features with Long-term Discriminative Features for Speaker Verification of Telephone Data
Cong-Thanh Do, Claude Barras, Viet-Bac Le, Achintya Sarkar

Using group delay functions from all-pole models for speaker recognition
Padmanabhan Rajan, Tomi Kinnunen, Cemal Hanilci, Jouni Pohjalainen, Paavo Alku

Secure Binary Embeddings of Front-end Factor Analysis for Privacy Preserving Speaker Verification
José Portêlo, Alberto Abad, Bhiksha Raj, Isabel Trancoso

On Von-Mises Fisher Mixture Model in Text-Independent Speaker Identification
Jalil Taghia, Zhanyu Ma, Arne Leijon

Using Phone Log-Likelihood Ratios as Features for Speaker Recognition
Mireia Diez, Amparo Varona, Mikel Penagarikano, Luis Javier Rodríguez-Fuentes, Germán Bordel

Handling Recordings Acquired Simultaneously over Multiple Channels with PLDA
Jesus Villalba, Mireia Diez, Amparo Varona, Eduardo Lleida Solano

Bayesian distance metric learning on i-vector for speaker verification
Xiao Fang, Najim Dehak, James Glass

Merging human and automatic system decisions to improve speaker recognition performance
Rosa González Hautamäki, Ville Hautamäki, Padmanabhan Rajan, Tomi Kinnunen

---

Wed-Poster 20

Wednesday, August 28, 2013
14:00 - 16:00 Room Forum 6

Poster 20: Dialog Systems and Applications I

Chair: Martine Garnier-Rizet, IMMI/ANR, Paris

Recurrent Neural Networks for Language Understanding
Kaisheng Yao, Geoffrey Zweig, Mei-Yuh Hwang, Yangyang Shi, Dong Yu

A Study on LVCSR and Keyword Search for Tagalog
Korbinian Riedhammer, Hai Do Van, Jim Hieronymus

Characterising Depressed Speech for Classification
Sharifa Alghowinem, Roland Goecke, Michael Wagner, Julien Epps, Gordon Parker, Michael Breakspear

Combining Acoustic Name Spotting and Continuous Context Models to improve Spoken Person Name Recognition in Speech
Benjamin Bigot, Grégory Senay, Georges Linarès, Corinne Fredouille, Richard Dufour
A Resource-Dependent Approach to Word Modeling for Keyword Spotting
I-Fan Chen and Chin-Hui Lee

Markers of confidence and correctness in spoken medical narratives
Kathryn Womack, Cecilia Ovesdotter Alm, Cara Calvelli, Jeff B. Pelz, Pengcheng Shi, Anne Haake

Ibuki Nakamura, Nobuaki Minematsu, Masayuki Suzuki, Hiroko Hirano, Chieko Nakagawa, Noriko Nakamura, Yukinori Tagawa, Keikichi Hirose, Hiroya Hashimoto

Addressee Detection for Dialog Systems Using Temporal and Spectral Dimensions of Speaking
Style
Elizabeth Shriberg, Andreas Stolcke, Suman Ravuri

Analysis of factors involved in the choice of rising or non-rising intonation in question utterances appearing in conversational speech
Hiroaki Hatano, Miyako Kiso, Carlos Ishi

IsNL? A Discriminative Approach to Detect Natural Language Like Queries for Conversational Understanding
Asli Celikyilmaz, Gokhan Tur, Dilek Hakkani-Tür

Automatic Accent Quantification of Indian Speakers of English
Jian Cheng, Nikhil Bojja, Xin Chen

Semantic Parsing Using Word Confusion Networks With Conditional Random Fields
Gokhan Tur, Anoop Deoras, Dilek Hakkani-Tür

Timing responses to questions
Sofia Strömbergsson, Anna Hjalmarsson, Jens Edlund, David House

BUT BABEL system for spontaneous Cantonese
Martin Karafiat, Frantisek Grezl, Mirko Hannemann, Karel Vesely, Honza Cernocky

Semi-Supervised Manifold Learning Approaches for Spoken Term Verification
Atta Norouzian, Richard Rose, Aren Jansen

Language Modeling for Mixed Language Speech Recognition using Weighted Phrase Extraction
Ying LI and Pascale Fung

---

Wed-Poster 21

Wednesday, August 28, 2013

14:00 - 16:00  Room Forum 6

Poster 21: Spoken Machine Translation and Speech Natural Language Processing

Chair: Chiori Hori, NICT, Kyoto
Quality Assessment of Asymmetric Multiparty Telephone Conferences: a Systematic Method from Technical Degradations to Perceived Impairments
Janto Skowronek, Julian Herlinghaus, Alexander Raake

User Activity Estimation Method Based on Probabilistic Generative Model of Acoustic Event Sequence with User Activity and Its Subordinate Categories
Keisuke Imoto, Suehiro Shimauchi, Hisashi Uematsu, Hitoshi Ohnuro

Generalizing Continuous-space Translation of Paralinguistic Information
Takamori Kano, Shinichiro Takamichi, Sakri Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura

An Empirical Comparison of Joint Optimization Techniques for Speech Translation
Masaya Ohgushi, Graham Neubig, Sakri Sakti, Tomoki Toda, Satoshi Nakamura

A sequential repetition model for improved disfluency detection
Mari Ostendorf and Sangyun Hahn

Disfluency Detection Based on Prosodic Features for Univ. Lectures
Henrique Medeiros, Helena Moniz, Fernando Batista, Isabel Trancoso, Luis Nunes

What's the difference? Comparing humans and machines on the Aurora 2 speech recognition task
Bernd T. Meyer

Calibration of distance measures for unsupervised query-by-example
Michele Gubian, Lou Boves, Maarten Versteegh

Indexing multimedia documents with acoustic concept recognition lattices
Diego Castan and Murat Akbacak

MINT.tools: Tools and Adaptors Supporting Acquisition, Annotation and Analysis of Multimodal Corpora
Spyros Kousidis, Thies Pfeiffer, David Schlangen

Show & Tell 3
Wednesday, August 28, 2013
14:00 - 16:00 Salon Tête d'Or

Chairs: Laurence Devillers, LIMSI-CNRS/Paris-Sorbonne 4, Fabrice Lefèvre, LIA, Avignon

ST3-1 Presentation of the Simple4All Project
*The Simple4All project consortium*

ST3-2 On-Line Learning of Lexical Items and Grammatical Constructions via Speech, Gaze and Action-Based Human-Robot Interaction
Grégoire Pointeau, Maxime Petit, Xavier Hinaut, Guillaume Gibert, Peter Ford Dominey

ST3-3 Development of a Pronunciation Training System Based on Auditory-Visual Elements
Harako Miyakoda

ST3-4 Real-time and Non-real-time Voice Conversion Systems with Web Interfaces
Elias Azarov, Maxim Vashkevich, Denis Likhachov, Alexander Petrovsky
Wed-Oral 36

Wednesday, August 28, 2013
16:30 - 18:30 Room Gratte-Ciel 1

Oral 36: Language Model Adaptation


128
Fitting Long-range Information Using Interpolated Distanced N-grams and Cache Models into a Latent Dirichlet Language Model for Speech Recognition
Md. Akmal Haidar and Douglas O'Shaughnessy

349
Incorporating Proximity Information for Relevance Language Modeling in Speech Recognition
Yi-Wen Chen, Bo-Han Hao, Kuan-Yu Chen, Berlin Chen

1014
Instance-Based On-line Language Model Adaptation
Ali Orkan Bayer and Giuseppe Riccardi

886
Unsupervised Topic Adaptation for Morph-Based Speech Recognition
André Mansikkanemi and Mikko Kurimo

1033.2.0
Unsupervised Language Model Adaptation for Automatic Speech Recognition of Broadcast News Using Web Crowdsourcing
Tim Schlippe, Łukasz Gren, Ngoc Thang Vu, Tanja Schultz

292
Recurrent Neural Network Based Language Model Personalization by Social Network Crowdsourcing
Tsung-Hsien Wen, Aaron Heidel, Hung-yi Lee, Yu Tsao, Lin-Shan Lee
**Wed-Oral 37**

**Wednesday, August 28, 2013**

16:30 - 18:30  Room Gratte-Ciel 3

**Oral 37: Spoken Language Summarization and Understanding**

Chairs: **Dilek Hakkani-Tür**, Microsoft Research, Mountain View, **Jingjing Liu**, MIT, Cambridge

*Language-Independent Call Routing using the Large Margin Estimation Principle*

Moataz El Ayadi and Mohamed Afify

*Deep Belief Network based Semantic Taggers for Spoken Language Understanding*

Anoop Deoras and Ruhi Sarikaya

*Error-Corrective Discriminative Joint Decoding of Automatic Spoken Language Transcription and Understanding*

Bassam Jabaian and Fabrice Lefèvre

*Detecting Summarization Hot Spots in Meetings Using Group Level Involvement and Turn-taking Features*

Catherine Lai, Jean Carletta, Steve Renals

*Supervised Spoken Document Summarization Based on Structured Support Vector Machine with Utterance Clusters as Hidden Variables*

Sz-Rung Shiang, Hung-yi Lee, Lin-shan Lee

*Web Data Harvesting for Speech Understanding Grammar Induction*

Ioannis Klasinas, Alexandros Potamianos, Elias Iosif, Spiros Georgiladakis, Gianluca Mameli

---

**Wed-Oral 38**

**Wednesday, August 28, 2013**

16:30 - 18:30  Room Prestige Gratte-Ciel

**Oral 38: Speech Synthesis – Multimodal and Articulatory Synthesis**

Chairs: **Jonas Beskow**, KTH, Stockholm, **Jianwu Dang**, Tianjin Univ /JAIST

*Articulatory Synthesis of French Connected Speech from EMA Data*

Asterios Toutios and Shrikanth Narayanan

*A New Language Independent, Photo-realistic Talking Head Driven by Voice Only*

Xinjins Zhang, Lijuan Wang, Gang Li, Frank Seide, Frank Soong

*Binocular Photometric Stereo Acquisition and Reconstruction for 3D Talking Head Applications*

Chaoyang Wang, Lijuan Wang, Yasuyuki Matsushita, Bojun Huang, Magnetro Chen, Frank Soong

*Speaker Adaptation of an Acoustic-Articulatory Inversion Model using Cascaded Gaussian Mixture Regressions*

Thomas Hueber, Gérard Bailly, Pierre Badin, Frédéric Elisei

*
Articulatory features for speech-driven head motion synthesis
Atef Ben Youssef, Hiroshi Shimodaira, David A. Braude

Template-Warping Based Speech Driven Head Motion Synthesis
David Braude, Hiroshi Shimodaira, Atef Ben Youssef

Wed-Oral 39
Wednesday, August 28, 2013
16:30 - 18:30 Room Tête d’Or 2
Oral 39: Speaker Diarization and Recognition
Chairs: Sylvain Meignier, LIUM, Jason Pelecanos, IBM, Yorktown Heights

ALIZE 3.0 - Open Source Toolkit for State-of-the-Art Speaker Recognition
Anthony Larcher, Jean-Francois Bonastre, Benoit Fauve, Kong Aik Lee, Christophe Levy, Haizhou Li, John S.D. Mason, Jean-Yves Parfait

New Cosine Similarity Scorings to Implement Gender-independent Speaker Verification
Mohammed Senoussaoui, Patrick Kenny, Pierre Dumouchel, Najim Dehak

Improving Speaker Identification in TV-shows Using Person Name Detection in Overlaid Text and Speech
Delphine Charlet, Corinne Fredouille, Geraldine Damnati, Gregory Senay

Exploring methods of improving speaker accuracy for speaker diarization
Mary Knox, Nikki Mirghafori, Gerald Friedland

Combining Deep Speaker Specific Representations with GMM-SVM for Speaker Verification
Ryan Price, Sangeeta Biswas, Koichi Shinoda

Using Spectral Moments as a Speaker Specific Feature in Nasals and Fricatives
Carola Schindler and Christoph Draxler

Wed-Oral 40
Wednesday, August 28, 2013
16:30 - 18:30 Room Bellecour 2
Oral 40: Models of Speech Perception
Chairs: Louis Ten Bosch, Radboud Univ., Nijmegen, Odile Bagou, Univ. of Geneva

A computational model of perceptuo-motor processing in speech perception: learning to imitate and categorize synthetic CV syllables
Raphael Laurent, Jean-Luc Schwartz, Pierre Bessière, Julien Diard

Talker-specific perceptual processing: Influences on internal category structure
Rachel Theodore
Elicitation and analysis of a corpus of robust noise-induced word misperceptions in Spanish
Maria Luisa Garcia Lecumberri, Attila Mate Toth, Yan Tang, Martin Cooke

Vocabulary structure and spoken-word recognition: Evidence from French reveals the source of embedding asymmetry
Anne Cutler and Laurence Bruggeman

How do multiple sublexical cues converge in lexical segmentation? An artificial language learning study.
Odile Bagou and Ulrich Hans Frauenfelder

Towards an end-to-end computational model of speech comprehension: simulating a lexical decision task
Louis ten Bosch, Lou Boves, Mirjam Ernestus

Wed-Poster 22
Wednesday, August 28, 2013
16:30 - 18:30 Room Forum 6
Poster 22: Paralinguistic Information

Chair: Julia Hirschberg, Columbia Univ., New York

Demographic Recommendation by means of Group Profile Elicitation Using Speaker Age and Gender Recognition
Sven Shepstone, Zheng-Hua Tan, Søren Holdt Jensen

Affective Classification of Generic Audio Clips using Regression Models
Nikolaos Malandrakis, Shiva Sundaram, Alexandros Potamianos

A Preliminary Study of Cross-lingual Emotion Recognition from Speech: Automatic Classification versus Human Perception
Je Hun Jeon, Duc Le, Rui Xia, Yang Liu

Active learning for dimensional speech emotion recognition
Wenqing Han, HaiFeng Li, Huabin Ruan, Lin Ma, Jiayin Sun, Björn Schuller

Auditory detectability of vocal ageing and its effect on forensic automatic speaker recognition
Finnian Kelly and Naomi Harte

Comparative Study of Speaker Personality Traits Recognition in Conversational and Broadcast News Speech
Firoj Alam and Giuseppe Riccardi

Active Learning by Label Uncertainty for Acoustic Emotion Recognition
Zixing Zhang, Jun Deng, Erik Marchi, Björn Schuller

Modeling Therapist Empathy and Vocal Entrainment in Drug Addiction Counseling
Bo Xiao, Panayiotis Georgiou, Zac Imel, David Atkins, Shrikanth Narayanan
Estimating Callers’ Levels of Knowledge in Call Center Dialogues
Chiaki Miyazaki, Ryuichiro Higashinaka, Toshiko Makino, Yoshihiro Matsuo

Energy and F0 contour modeling with Functional Data Analysis for Emotional Speech Detection
Juan Arias, Carlos Busso, Nestor Becerra Yoma

Incremental Emotion Recognition
Taniya Mishra and Dimitrios Dimitriadis

Comparison of Spectrum Estimators in Speaker Verification: Mismatch Conditions Induced by Vocal Effort
Cemal Hanilci, Tomi Kinnunen, Pudmanabhan Rajan, Jouni Pohjalainen, Paavo Alku, Figen Ertas

Using Denoising Autoencoder For Emotion Recognition
ruixia and Yang Liu

Wed-Poster 23
Wednesday, August 28, 2013
16:30 - 18:30 Room Forum 6

Poster 23: Speech and Audio Signal Processing
Chair: Piero Cosi, ISTC-CNR, Padua

A Phase-Modified Approach for TDE-Based Acoustic Localization
Georgios Athanasopoulos and Werner Verhelst

Interference Robust DOA Estimation of Human Speech by Exploiting Historical Information and Temporal Correlation
Wei Xue, Shan Liang, Wenju Liu

Identifying new bird species from differences in birdsong
Naomi Harte, Sadhbh Murphy, David Kelly, Nicola Marples

Controlling “shout” expression in a Japanese POP singing performance: analysis and suppression study
Yuri Nishigaki, Ken-Ichi Sakakibara, Masanori Morise, Ryuichi Nisimura, Toshio Irino, Hideki Kawahara

Dimensionality Analysis of Singing Speech Based on Locality Preserving Projections
Mahnoosh Mehrabani and John H. L. Hansen

Audio Classification Using Dominant Spatial Patterns in Time-Frequency Space
Md. Khademul Islam Molla and Keikichi Hirose

Spectro-temporal Modulation Based Singing Detection Combined with Pitch-based Grouping for Singing Voice Separation
Tse-En Lin, Chung-Chien Hsu, Yi-Cheng Chen, Jian-Huang Chen, Tai-Shih Chi
NMF-based Temporal Feature Integration for Acoustic Event Classification
Jimmy Ludeña-Choez and Ascension Gallardo-Antolin

Robust Audio Codebooks for Large Scale Event Detection in Consumer Videos
Shourabh Rawat, Peter Schulam, Susanne Burger, Duo Ding, Yipei Wang, Florian Metze

Person identification using biometric markers from footsteps sound
Muhammad Umair Bin Altaf, Taras Butko, Biing-Hwang(Fred) Juang

Learning Binaural Spectrogram Features for Azimuthal Speaker Localization
Wiktor Mlynarski

An Unsupervised Bayesian Classifier for Multiple Speaker Detection and Localization
Youssef Oualil, Friedrich Faubel, Dietrich Klakow

Joint recognition and direction-of-arrival estimation of simultaneous meeting-room acoustic events
Rupayan Chakraborty and Climent Nadeu

Audio Self Organized Units for High-level Event Detection
Xiaodan Zhuang, Shuang Wu, Pradeep Natarajan, Rohit Prasad, Prem Natarajan

---

Wed-Poster 24
Wednesday, August 28, 2013
16:30 - 18:30  Room Forum 6

Poster 24: ASR – Robustness Against Noise III

Chair: Nelson Morgan, ICSI, Berkeley

Distribution-based Feature Normalization for Robust Speech Recognition Leveraging Context and Dynamic Cues
Yu-Chen Kao and Berlin Chen

An Investigation of Temporally Varying Weight Regression for Noise Robust Speech Recognition
Khe Chai Sim

Feature Space Generalized Variable Parameter HMMs for Noise Robust Recognition
Yang Li, Xunying Liu, Lan Wang

Bidirectional Truncated Recurrent Neural Networks for Efficient Speech Denoising
Philemon Brakel, Dirk Stroobandt, Benjamin Schrauwen

Multi-stream recognition of noisy speech with performance monitoring
Ehsan Variani, Feipeng Li, Hynek Hermansky
Model-based noise suppression using unsupervised estimation of hidden Markov model for non-stationary noise
Masakiyo Fujimoto and Tomohiro Nakatani

Joint Noise Cancellation and Dereverberation using Multi-Channel Linearly Constrained Minimum Variance Filter
Karan Nathwani and Rajesh M. Hegde

Is speech enhancement pre-processing still relevant when using deep neural networks for acoustic modeling?
Marc Delcroix, Yotaro Kubo, Tomohiro Nakatani, Atsushi Nakamura

Histogram Equalization of Real and Imaginary Modulation Spectra for Noise-Robust Speech Recognition
Hsin-Ju Hsieh, Berlin Chen, Jeih-weih Hung

An Investigation of Spectral Restoration Algorithms for Deep Neural Networks based Noise Robust Speech Recognition
Bo Li, Yu Tsao, Khe Chai Sim

Bounded Conditional Mean Imputation with an Approximate Posterior
Ulpur Remes

Mixtures of Bayesian Joint Factor Analyzers for Noise Robust Automatic Speech Recognition
Xiaodong Cui, Vaibhava Goel, Brian Kingsbury

Robust Speech Enhancement Techniques for ASR in Non-Stationary Noise and Dynamic Environments
Gang Liu, Dimitrios Dimitriadis, Enrico Bocchieri

---

**Cultural and Scientific Entertainment**

**Wednesday, August 28, 2013**

18:30 - 19:15

---

**Conference Banquet**

**Wednesday, August 28, 2013**

19:30 - 23:30

---

**Thu-Oral 41**

**Thursday August 29, 2013**

9:00 - 11:00 Room Gratte-Ciel 1

**Oral 41: Linguistic Systems, Phonetics-Phonology Interface**

Chairs: Ioana Chitoran, Univ. Paris-Diderot, Henning Reetz, Institute for Phonetics, Frankfurt

LAPSyD: Lyon-Albuquerque Phonological Systems Database

Ian Maddieson, Sébastien Flavier, Egidio Marsico, Christophe Coupé, François Pellegrino
The Duration Compensation Issue Revisited
Plinio Barbosa

Cross-language Comparison of Functional Load for Vowels, Consonants, and Tones
Yoon Mi Oh, François Pellegrino, Christophe Coupé, Egidio Marsico

Notes on So-called Inter-speaker Difference in Spontaneous Speech: The Case of Japanese Voiced Obstruent
Kikuo Maekawa

The role of the pharynx and tongue in enhancement of vowel nasalization: A real-time MRI investigation of French nasal vowels
Christopher Carignan, Ryan Shosted, Maojing Fu, Zhi-Pei Liang, Bradley Sutton

Assimilation of word-final nasals to following word-initial place of articulation in UK English
Margaret Renwick, Ladan Baghai-Ravary, Rosalind Temple, John Coleman

---

**Thu-Oral 42**

*Thursday August 29, 2013*

*Oral 42: Speech Synthesis – Voice Conversion*

**Chairs:** Yao Qian, Microsoft Research Center., Beijing; Tomoki Toda, NAIST, Ikoma

**Joint Spectral Distribution Modeling Using Restricted Boltzmann Machines for Voice Conversion**
Ling-Hui Chen, Zhen-Hua Ling, Yan Song, Li-Rong Dai

Exemplar-based unit selection for voice conversion utilizing temporal information
Zhizheng Wu, Tuomas Virtanen, Tomi Kinnunen, Eng Siong Chng, Haizhou Li

Alleviating the Over-Smoothing Problem in GMM-Based Voice Conversion with Discriminative Training
Hsin-Te Hwang, Yu Tsao, Hsin-Min Wang, Yih-Ru Wang, Sin-Horng Chen

A Hybrid Approach to Electrolaryngeal Speech Enhancement Based on Spectral Subtraction and Statistical Voice Conversion
Ko Tanaka, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura

A Digital Signal Processor Implementation of Silent/Electrolaryngeal Speech Enhancement based on Real-Time Statistical Voice Conversion
Takuto Moriguchi, Tomoki Toda, Motoaki Sano, Hiroshi Sato, Graham Neubig, Sakriani Sakti, Satoshi Nakamura

Foreign Accent Conversion Through Voice Morphing
Sandesh Aryal, Daniel Felps, Ricardo Gutierrez-Osuna

---

**Thu-Oral 43**

*Thursday August 29, 2013*

*Oral 43: Large Vocabulary Continuous Speech Recognition Systems*

**Chairs:** Steve Renals, Univ. of Edinburgh; Dirk Van Compernolle, K.U. Leuven – ESAT, Leuven
Empirical Link Between Hypothesis Diversity and Fusion Performance in an Ensemble of Automatic Speech Recognition
Kartik Audhkhasi, Andreas Zavou, Panayiotis Georgiou, Shrikanth Narayanan

A lecture transcription system combining neural network acoustic and language models
Peter Bell, Hitoshi Yamamoto, Pawel Swietojanski, Youzheng Wu, Fergus McInnes, Chiori Hori, Steve Renals

Neural Network Acoustic Models for the DARPA RATS Program
Hagen Soltau, Hong-Kwang Kuo, Lidia Mangu, George Saon, Tomas Beran

Improved Models for Automatic Punctuation Prediction for Spoken and Written Text
Nicola Ueffing, Maximilian Bisani, Paul Voza

Some Issues affecting the Transcription of Hungarian Broadcast Audio
Anindya Roy, Lori Lamel, Thiago Fraga da Silva, Jean-Luc Gauvain, Ilya Oparin

Development of the RWTH Transcription System for Slovenian
Pavel Golik, Zoltán Tüske, Ralf Schütz, Hermann Ney

Thu-Oral 44
Thursday August 29, 2013
9:00 - 11:00 Room Tête d’Or 2
Oral 44: Robust Speaker Recognition

Chairs: Najim Dehak, MIT, Cambridge, Lukas Burget, Brno Univ. of Technology
Noise Robust Speaker Verification with Delta Cepstrum
Naoyuki Kanda, Ryu Takeda, Yasunari Obuchi

R-Norm: Improving Inter-Speaker Variability Modelling at the Score Level via Regression Score Normalisation
David Vandyke, Michael Wagner, Roland Goecke

Frequency Warping and Robust Speaker Verification: A Comparison of Alternative Mel-Scale Representations
Tomi Kinnunen, Md. Jahangir Alam, Pavel Matejka, Patrick Kenny, Jan Cernocky, Douglas O'Shaughnessy

Acoustic Factor Analysis based Universal Background Model for Robust Speaker Verification in Noise
Taufiq Hasan and John Hansen

A New Bayesian Network to Assess the Reliability of Speaker Verification Decisions
Jesus Villalba, Eduardo Lleida Solano, Alfonso Ortega, Antonio Miguel

The IBM RATS Phase II Speaker Recognition System: Overview and Analysis
Weizhong Zhu, Sibel Yaman, Jason Pelecanos
Thu-Oral 45
Thursday August 29, 2013
9:00 - 11:00 Room Bellecour 2

Oral 45: Acoustic and Articulatory Cues in Speech Perception

Chairs: Mary Beckman, OSU, Columbus, David House, KTH, Stockholm

Vowel identity conditions the time course of tone recognition
Jason Shaw, Michael Tyler, Benjawan Kassisopa, Yuan Ma, Michael Proctor, Chong Han, Donald Derrick, Denis Burnham

Changes in the Role of Intensity as a Cue for Fricative Categorisation
Odette Scharenborg and Esther Janse

Weighting of acoustic cues shifts to frication duration in identification of fricatives/affricates when auditory properties are degraded due to aging
Keiichi Yasu, Takayuki Arai, Kei Kobayashi, Mitsuko Shindo

Duration as a secondary cue for perception of voicing and tone in Shanghai
Chinese
Jiayin Gao and Pierre Halle

Development of Central Auditory Processes and their links with language skills in typically developing children
Marie Dekerle, Fanny Meunier, Sophie Donnadieu

Show me what you listen to! Auditory classification images reveal the processing of fine acoustic cues during speech categorization.
Léo Varnet, Kenneth Knoblauch, Fanny Meunier, Michel Hoen

Thu-Poster 25
Thursday August 29, 2013
9:00 - 11:00 Room Forum 6

Poster 25: Speech Production – Data and Models

Chair: Claire Pillot-Loiseau, Univ. Sorbonne Nouvelle

The organ stop “vox humana” as a model for a vowel synthesiser
Fabian Brackhane and Juergen Trouvain

Information theoretic acoustic feature selection for acoustic-to-articulatory inversion
Prasanta Ghosh and Shrikanth Narayanan

Formant Contours in Czech Vowels: Speaker-discriminating Potential
Dita Fejlová, David Lukeš, Radek Skarnitzl

An Anisotropic Diffusion Filter Based on Multidirectional Separability
Shen Liu, Jianguo Wei, Xin Wang, Wenhuan Lu, Quang Fang, Jianwu Dang

The phonological voicing contrast in Czech: An EPG study of phonated and whispered fricatives
Radek Skarnitzl, Pavel Šturm, Pavel Machač
Vowel and prosodic factor dependent variations of vocal-tract length
Shinji Maeda and Yves Laprie

Word Identification using Phonetic Features: towards a method to support multivariate fMRI speech decoding
Tijl Grootswagers, Karen Dijkstra, Louis ten Bosch, Alex Brandmeyer, Makiko Sadakata

Analysis of breathy, modal and pressed phonation based on low frequency spectral density
Dhananjaya Gowda and Mikko Kurimo

Is the Vowel Length Contrast in Japanese Exaggerated in Infant-Directed Speech?
Keiichi Tajima, Kuniyoshi Tanaka, Andrew Martin, Reiko Mazuka

Investigating the relationship between glottal area waveform shape and harmonic magnitudes through computational modeling and laryngeal high-speed videendoscopy
Gang Chen, Robin Samlan, Jody Kreiman, Abeer Alwan

Formant Frequency Tracking Using Gaussian Mixtures with Maximum A Posteriori Adaptation
Jonathan Kim, Hrishikesh Rao, Mark Clements

Devoicing of vowels in German, a comparison of Japanese and German speakers
Rei Yasuda and Frank Zimmerer

Identifying consonantal tasks via measures of tongue shaping: a real-time MRI investigation of the production of vocalized syllabic /l/ in American English
Caitlin Smith and Adam Lammert

---

**Thu-Poster 26**

Thursday August 29, 2013
9:00 - 11:00 Room Forum 6

**Poster 26: Speech Enhancement**

Chairs: Maurizio Omologo, FBK, Trento

A Speech Enhancement Method by Coupling Speech Detection and Spectral Amplitude Estimation
Feng Deng, Chang-chun Bao, Feng Bao

Late Reverberation Suppression Using MMSE Modulation Spectral Estimation
Chenxi Zheng and Wai-Yip Chan

A New Statistical Excitation Mapping for Enhancement of Throat Microphone Recordings
Mehmet Ali Tugtekin Turan and Engin Erzin

Classification based binaural dereverberation
Nicoleta Roman and Michael Mandel
Target-to-Non-target Directional Ratio Estimation Based on Dual-Microphone Phase Differences for Target-Directional Speech Enhancement
Seon Man Kim and Hong Kook Kim

Speech Spectrum Restoration Based on Conditional Restricted Boltzmann Machine
Xugang Lu, Shigeki Matsuda, Chiori Hori

Speaker Separation Using Visual Speech Features and Single-channel Audio
Faheem Khan and Ben Milner

Spectral Modulation Sensitivity Based Perceptual Acoustic Echo Cancellation
Wei-Lun Chuang, Kah-Meng Cheong, Chung-Chien Hsu, Tai-Shih Chi

Speech Enhancement Using Compressed Sensing
Vinayak Abrol, Pulkit Sharma, Anil Kumar Sao

Spectro-temporal post-enhancement using MMSE estimation in NMF based single-channel source separation
Emad M. Grais and Hakan Erdogan

A Pitch-Based Spectral Enhancement Technique for Robust Speech Processing
Kantapon Kaewtip, Lee Ngee Tan, Abeer Alwan

Stochastic-Deterministic Signal Modelling for the Tracking of Pitch in Noise and Speech Mixtures using Factorial HMMs
Matthew McCallum and Bernard Guillemin

Restoration of clipped signals with application to speech recognition
Shay Maymon, Etienne Marcheret, Vaibhava Goel

On the robustness of distributed EM based BSS in asynchronous distributed microphone array scenarios
Yasufumi Uezu, Keisuke Kinoshita, Mehrez Souden, Tomohiro Nakatani

Thu-Poster 27
Thursday August 29, 2013
9:00 - 11:00 Room Forum 6
Poster 27: ASR – Acoustic Modeling

Chair: Arnab Ghoshal, Univ. of Edinburgh

Infinite Support Vector Machines in Speech Recognition
Jingzhou Yang, Rogier van Dalen, Mark Gales

An On-line Incremental Speaker Adaptation Technique for Audio Stream Transcription
Diego Giuliani and Fabio Brugnara
Accent- and Speaker-Specific Polyphone Decision Trees for Non-Native Speech Recognition
Dominic Telaar and Mark C. Fuhs

Investigations on Hessian-Free Optimization for Cross-Entropy Training of Deep Neural Networks
Simon Wiesler, Jinyu Li, Jian Xue

Cross-lingual Acoustic Model Adaptation based on Transfer Vector Field Smoothing with MAP
Masahiro Saiko, Shigeki Matsuda, Ken Hanazawa, Ryosuke Isotani, Chiori Hori

N-Best Rescoring by Phoneme Classifiers using Subclass AdaBoost Algorithm
Hiroshi Fujimura, Yusuke Shinhoara, Takashi Masuko

Stream Selection and Integration in Multistream ASR Using GMM-Based Performance Monitoring
Tetsuji Ogawa, Feipeng Li, Hynek Hermansky

VTLN based on the linear interpolation of contiguous Mel filter-bank energies
Nestor Becerra Yoma, Claudio Garreton, Fernando Huemupan, Ignacio Catalan

Context-dependent modeling and speaker normalization applied to reservoir-based phone recognition
Fabian Triefenbach, Azarakhsh Jalalvand, Kris Demuynck, Jean-Pierre Martens

Interpolation of Acoustic Models for Speech Recognition
Thiago Fraga da Silva, Jean-Luc Gauvain, Lori Lamel

Training Log-Linear Acoustic Models in Higher-Order Polynomial Feature Space for Speech Recognition
Muhammad Ali Tahir, Heyun Huang, Ralf Schlüter, Hermann Ney, Louis ten Bosch, Bert Cranen, Lou Boves

Comparison of Spectral Analysis Methods for Automatic Speech Recognition
Neelima Parinam, Chandra Sekhar Vootkuri, Stephen Zahorian

Synthetic Speaker Models Using VTLN to Improve the Performance of Children in Mismatched Speaker Conditions for ASR
Rama Sanand Doddipatla and Torbjørn Svendsen

Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition
Ossama Abdel-Hamid, Li Deng, Dong Yu

---

Special Event: ESCA/ISCA Anniversary
Thursday August 29, 2013
11:30 - 12:30 Amphithéâtre

Chair: Joseph Mariani, IMMI&LIMSI CNRS, Orsay
Thu-Oral 46
Thursday August 29, 2013
14:00 - 16:00  Room Gratte-Ciel 1

Oral 46: Language Modeling for Conversational Speech

Chairs: Katrin Kirchhoff, Univ. of Washington, Seattle, Giuseppe Riccardi, Univ. of Trento

Feature-rich sub-lexical language models using a maximum entropy approach for German LVCSR

Mahaboob Ali Basha Shaik, Amr El-Desoky Mousa, Ralf Schlüeter, Hermann Ney

Morpheme Level Hierarchical Pitman-Yor Class-based Language Models for LVCSR of Morphologically Rich Languages

Amr El-Desoky Mousa, Mahaboob Ali Basha Shaik, Ralf Schlüeter, Hermann Ney

Discriminatively Trained Dependency Language Modeling for Conversational Speech Recognition

Benjamin Lambert, Bhiksha Raj, Rita Singh

Prefix Tree based N-best list Re-scoring for Recurrent Neural Network Language Model used in Speech Recognition System

yujing si, qingqing zhang, ta li, jielin pan, yonghong yan

Cross-domain Paraphrasing For Improving Language Modelling Using Out-of-domain Data

Xunying (Andrew) Liu, Mark Gales, Phil Woodland

Viterbi decoding for latent words language models using Gibbs sampling

Ryo Masumura, Hirokazu Masataki, Takanobu Oba, Osamu Yoshioka, Satoshi Takahashi

---

Thu-Oral 47
Thursday August 29, 2013
14:00 - 16:00  Room Gratte-Ciel 3

Oral 47: Speech Enhancement and Coding

Chairs: Gernot Kubin, Univ. of Graz, Isabel Trancoso, INESC-ID/IST, Lisbon

Computationally efficient objective function for algebraic codebook optimization in ACELP

Tom Bäckström

Speech Quality Prediction for Artificial Bandwidth Extension Algorithms

Sebastian Möller, Emilia Kelaidi, Friedemann Köster, Nicolas Côté, Patrick Bauer, Tim Fingscheidt, Thomas Schlien, Hannu Pulakka, Paavo Alku

Speech Enhancement with Weighted Denoising Auto-Encoder

Bing-yin Xia and Changchun Bao

Syllable-based Pitch Encoding for Low Bit Rate Speech Coding with Recognition/Synthesis Architecture

Milos Cernak, Xingyu Na, Philip N. Gamer

Artificial bandwidth extension based on regularized piecewise linear mapping with discriminative region weighting and long-span features

Duy Nguyen, Masayuki Suzuki, Nobuaki Minematsu, Keikichi Hirose
Bong-Ki Lee, Chungsoo Lim, Jihwan Park, Joon-Hyuk Chang

Thu-Oral 48
Thursday August 29, 2013
14:00 - 16:00 Room Prestige Gratte-Ciel

Oral 48: Spoken Machine Translation and Speech Natural Language Processing
Chairs: Satoshi Nakamura, NAIST, Ikoma, Monika Woszczyna, Multimodal Technologies, Pittsburgh

Thu-Oral 48
Thursday August 29, 2013
14:00 - 16:00 Room Prestige Gratte-Ciel

Oral 49: ASR – Robustness Against Noise II
Chairs: Li Deng, Microsoft Research, Redmond, George Saon, IBM, Yorktown Heights

Noise adaptive training for subspace Gaussian mixture
Thu-Oral 49
Thursday August 29, 2013
14:00 - 16:00 Room Tète d'Or 2

The IBM Speech Activity Detection System for the DARPA RATS Program
Thu-Oral 49
Thursday August 29, 2013
14:00 - 16:00 Room Tète d'Or 2

Conditional Emission Densities for Combining Speech Enhancement and Recognition Systems
Thu-Oral 50

Thursday August 29, 2013
14:00 - 16:00 Room Bellecour 2

Oral 50: Articulatory and Acoustic Cues of Speech Prosody

Chairs: Aijun Li, Institute of Linguistics, Beijing, Oliver Niebuhr, Univ. of Kiel

The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particle

Lan Nguyen Thi, Alexis Michaud, Dat Tran Do, Khoa Mac Dang

The Voice Prominence Hypothesis: the Interplay of F0 and Voice Source Features in Accentuation

Ailbhe Ní Chasaide, Irena Yanushevskaya, John Kane, Christer Gobl

Mora-based pre-low raising in Japanese pitch

Albert Lee, Yi Xu, Santitham Prom-on

Prosodic cues of sarcastic speech in French: slower, higher, wider

Hélène Loevenbruck, Mohamed Ameur Ben Jannet, Mariapaola D’Imperio, Mathilde Spini, Maud Champagne-Lavau

Correlates of contrastive focus in congenitally blind adults and sighted adults

Lucie Menard, Anne Leclerc, Mark Tiede, Amélie Prémont, Christine Turgeon, Pamèla Trudeau-Fisette, Dominique Côté

Is protrusion of French rounded vowels affected by prosodic positions?

Laurianne Georgeton and Nicolas Audibert

Thu-SP6

Thursday August 29, 2013
14:00 – 16:00 Room Bellecour 1

Special Session 6: Intelligibility-Enhancing Speech Modifications

Chairs: Martin Cooke, Ikerbasque, Vitoria, Catherine Mayo, CSTR, Edinburgh

ACCEPTED PAPERS

Intelligibility-enhancing speech modifications: the Hurricane Challenge

Martin Cooke, Catherine Mayo, Cassia Valentini-Botinhao
Statistical Synthesizer with Embedded Prosodic and Spectral Modifications to Generate Highly Intelligible Speech in Noise
Daniel Ero, Tudor-Catalin Zorila, Yannis Stylianou, Eva Navas, Inma Hernaez

Lombard Modified Text-to-Speech Synthesis for Improved Intelligibility: Submission for the Hurricane Challenge 2013
Antti Suni, Reima Karhila, Tuomo Raitio, Mikko Kurimo, Martti Vainio, Paavo Alku

Combining perceptually-motivated spectral shaping with loudness and duration modification for intelligibility enhancement of HMM-based synthetic speech in noise
Cassia Valentini-Botinhao, Junichi Yamagishi, Simon King, Yannis Stylianou

Increasing Speech Intelligibility via Spectral Shaping with Frequency Warping and Dynamic Range Compression plus Transient Enhancement
Elizabeth Godoy and Yannis Stylianou

Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression
Henning Schepker, Jan Rennies, Simon Doclo

SII-based Speech Preprocessing for Intelligibility Improvement in Noise
Cees Taal and Jesper Jensen

Rephrasing-Based Speech Intelligibility Enhancement
Mengqiu Zhang, Petko Petkov, Bastiaan Kleijn

Information-preserving temporal reallocation of speech in the presence of fluctuating maskers
Vincent Aubanel and Martin Cooke

Preservation of Speech Spectral Dynamics Enhances Intelligibility
Petko Nikolov Petkov and W. Bastiaan Kleijn

ADDITIONAL CONTRIBUTIONS
An Overview of the VUB Entry for the 2012 Hurricane Challenge
Henk Brouckxon and Werner Verhelst

Improvement of speech intelligibility by optimization of spectral energy
Reiko Takou, Nobumasa Seiyama, Atsushi Imai
Balancing word lists in speech audiometry through large spoken language corpora
Annemiek Hammer, Bart Vaerenberg, Wojtek Kowalczyk, Louis ten Bosch, Martine Coene, Paul Govaerts

Developing an Information System for Deaf
Verónica López-Ludeña, Rubén San Segundo Hernández, Javier Ferreiros, José Manuel Pardo, Emilio Ferreiro

Dysarthric Speech Recognition Using Dysarthria-Severity-Dependent and Speaker-Adaptive Models
Myung Jong Kim, Joohong Yoo, Hoirin Kim

Voice Pathology Detection and Classification Using MPEG-7 Audio Low-Level Features
Ghulam Muhammad and Moutasem Melhem

Empirical Mode Decomposition-Based Spectral Acoustic Cues for Disordered Voices Analysis
Abdellah Kacha, Francis Grenez, Jean Schoentgen

Exemplar-based Individuality-preserving Voice Conversion for Articulation Disorders in Noisy Environments
Ryo Aihara, Ryoichi Takashima, Tetsuya Takiguchi, Yasuo Arik

Improving recognition of disordered speech with out-of-domain knowledge
Heidi Christensen, Magdalena Aniol, Peter Bell, Phil Green, Thomas Hain, Simon King, Pawel Swietojanski

Effects of envelope filter cutoff frequency on the intelligibility of Mandarin noise-vocoded speech in babble noise: Implications for cochlear implants
Guangting Mai, James William Minett, William Shi-Yuan Wang

---

**Thu-Poster 29**

**Thursday August 29, 2013**
14:00 - 16:00 Room Forum 6

**Poster 29: Robust Speaker Recognition**

Chair: Douglas O'Shaughnessy, INRS, Montréal

Multi-session PLDA Scoring of I-vector for Partially Open-Set Speaker Detection
Kong Aik Lee, Anthony Larcher, Chang Huai You, Bin Ma, Haizhou Li

Impact of Noise Reduction and Spectrum Estimation on Noise Robust Speaker Identification
Keith Godin, Seyed Omid Sadjadi, John H. L. Hansen

Improvement of distant-talking speaker identification using bottleneck features of DNN
Takanori Yamada, Longbiao Wang, Atsuhiko Kai

Geometric contamination for GMM/UBM speaker verification in reverberant environments
Alessio Brutti and Maurizio Omologo
Towards a More Efficient SVM Supervector Speaker Verification System using Gaussian Reduction and a Tree-Structured Hash
Richard McClanahan and Phillip DeLeon

Improving the PLDA based Speaker Verification in Limited Microphone Data Conditions
Ahilan Kanagasundaram, David Dean, Javier Gonzalez-Dominguez, Sridha Sridharan, Daniel Ramos, Joaquin Gonzalez-Rodriguez

The I3A Speaker Recognition System for NIST SRE12: Post-evaluation Analysis
Jesus Villalba, Eduardo Lleida Solano, Alfonso Ortega, Antonio Miguel

Text-dependent speaker recognition using PLDA with uncertainty propagation
Themos Stafylakis, Patrick Kenny, Pierre Ouellet, Javier Perez, Marcel Kockmann, Pierre Dumouchel

Robust Speaker Recognition Using Spectro-Temporal Autoregressive Models
Sri Harish Mallidi, Sriram Ganapathy, Hynek Hermansky

Effect of multicondition training on i-vector PLDA configurations for speaker recognition
Padmanabhan Rajan, Tomi Kinnunen, Ville Hautamaki

Improving Robustness to Compressed Speech in Speaker Recognition
Mitchell McLaren, Victor Abrash, Martin Graciarena, Yun Lei, Jan Pesan

Modulation Features for Noise Robust Speaker Identification
Vikramjit Mitra, Mitchell McLaren, Horacio Franco, Martin Graciarena, Nicolas Scheffer

Minimax ivector extractor for short duration speaker verification
Ville Hautamaki, You-Chi Cheng, Padmanabhan Rajan, Chin-Hui Lee

Standoff Speaker Recognition
Mike Fowler, Mark McCurry, Jonathan Bramsen, Kehinde Dunsin, Jeremiah Remus

Thu-Poster 30
Thursday August 29, 2013
14:00 - 16:00 Room Forum 6

Poster 30: Dialog Systems and Applications II

Chair: Helen Meng, CUHK, Hong Kong

Correlates to intelligibility in deviant child speech – comparing clinical evaluations to audience response system-based evaluations by untrained listeners
Sofia Strombergsson and Christina Tännander

Using linguistic analysis to characterize conceptual units of thought in spoken medical narratives
Kathryn Womack, Cecilia Ovesdotter Alm, Cara Calvelli, Jeff B. Pelz, Pengcheng Shi, Anne Haake
Interacting with robots via speech and gestures, an integrated architecture
Francesco Cutugno, Alberto Finzi, Michelangelo Fiore, Enrico Leone, Silvia Rossi

Incorporating Named Entity Recognition into the Speech Transcription Process
Mohamed Hatmi, Christine Jacquin, Emmanuel Morin, Sylvain Meignier

DTW-Distance-Ordered Spoken Term Detection
Teppei Ohno and Tomoyosi Akiba

Refining Sentence Similarity with Discourse Information in Dialog System
Sangkeun Jung and Seung-Hoon Na

Two-step Correction of Speech Recognition Errors Based on N-gram and Long Contextual Information
Ryohei Nakatani, Tetsuya Takiguchi, Yasuo Ariki

Inferring Actor Communities from Videos
Sumit Negi, Ramanth Balasubramanyan, Santanu Chaudhury

Multiple topic identification in telephone conversations
Xavier Bost, Marc El-Beze, Renato De Mori

Variable-Span Out-of-Vocabulary Named Entity Detection
Wei Chen, Sankaranarayanan Ananthakrishnan, Rohit Prasad, Prem Natarajan

On the Feasibility of Using Pupil Diameter to Estimate Cognitive Load Changes for In-Vehicle Spoken Dialogues
Andrew Kun, Oskar Palinko, Zeljko Medenica, Peter Heeman

Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding
Grégoire Mesnil, Xiaodong He, Li Deng, Yoshua Bengio

Paraphrase Features to Improve Natural Language Understanding
Xiaohu Liu, Ruhi Sarikaya, Chris Brockett, Chris Quirk, William Dolan

Discovering New User Intents from Search Query Logs for Spoken Language Understanding
Dilek Hakkani-Tur, Asli Celikyilmaz, Larry Heck, Gokhan Tur

Exploiting Shared Information for Multi-intent Natural Language Sentence Classification
Puyang Xu and Ruhi Sarikaya

Closing Ceremony
Thursday August 29, 2013
16:00 - 17:00 Amphithéâtre