# TABLE OF CONTENTS

**Volume 1**

**KEYNOTE SPEAKER I**

The Multiple-Channel Cochlear Implant: Interfacing Electronic Technology to Human Consciousness ................................................................. 1
  Graeme M. Clark

**SESSION: SPEECH RECOGNITION – LANGUAGE MODELING I**

Dynamic Language Model Adaptation Using Variational Bayes Inference ......................................................... 4
  Yik-Cheung Tam, Tanja Schultz

The Hidden Vector State Language Model .................................................................................................................. 8
  Vidura Seneviratne, Steve Young

Class-Based Variable Memory Length Markov Model ................................................................................................. 12
  Shinsuke Mori, Gakuto Kurata

Context-Sensitive Statistical Language Modeling ..................................................................................................... 16
  Alexander Gruenstein, Chao Wang, Stephanie Seneff

Language Model Data Filtering via User Simulation and Dialogue Resynthesis ......................................................... 20
  Chao Wang, Stephanie Seneff, Grace Chung

Bayesian Learning for Latent Semantic Analysis ....................................................................................................... 24
  Jen-Tzung Chien, Meng-Sung Wu, Chia-Sheng Wu

The Effect of Stress and Boundaries on Segmental Duration in a Corpus of Authentic Speech (British English) ...................................................................................... 28
  Daniel Hirst, Caroline Bouzon

Investigation of the Relationship Between Turn-taking and Prosodic Features in Spontaneous Dialogue ......................................................................................................................... 32
  Tomoko Ohsuga, Masafumi Nishida, Yasuo Horiuchi, Akira Ichikawa

Filled Pauses as Cues to the Complexity of Following Phrases ................................................................................. 36
  Michiko Watanabe, Keikichi Hirose, Yasuharu Den, Nobuaki Minematsu

Perceptual Magnet Effect in German Boundary Tones .................................................................................................... 40
  Katrin Schneider, Bernd Mobius

Constraints on the Acquisition of Simplex and Complex Words in German .......................................................... 44
  Angela Grimm, Jochen Trommer

Whistled Speech: A Natural Phonetic Description of Languages Adapted to Human Perception and to the Acoustical Environment .............................................................................. 48
  Julien Meyer

**SESSION: SPOKEN LANGUAGE EXTRACTION / RETRIEVAL I**

Fast Vocabulary-Independent Audio Search Using Path-Based Graph Indexing ......................................................... 52
  Olivier Siohan, Michiel Bacchiani
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Effects of Speech Recognition and Punctuation on Information Extraction Performance</td>
<td>56</td>
</tr>
<tr>
<td>John Makhouli, Alex Baron, Ivan Bulyko, Long Nguyen, Lance Ramshaw, David Stallard, Richard Schwartz, Bing Xiang</td>
<td></td>
</tr>
<tr>
<td>Indexing Uncertainty for Spoken Document Search</td>
<td>60</td>
</tr>
<tr>
<td>Ciprian Chelba, Alex Acero</td>
<td></td>
</tr>
<tr>
<td>Exploiting Passage Retrieval for N-Best Rescoring of Spoken Questions</td>
<td>64</td>
</tr>
<tr>
<td>Tomoyosi Akiba, Hiroyuki Abe</td>
<td></td>
</tr>
<tr>
<td>Multi-Stage Compaction Approach to Broadcast News Summarisation</td>
<td>68</td>
</tr>
<tr>
<td>BalaKrishna Kolluru, Heidi Christensen, Yoshihiko Gotoh</td>
<td></td>
</tr>
<tr>
<td>Audio-Video Summarization of TV News Using Speech Recognition and Shot Change Detection</td>
<td>72</td>
</tr>
<tr>
<td>Chien-Lin Huang, Chia-Hsin Hsieh, Chung-Hsien Wu</td>
<td></td>
</tr>
<tr>
<td>The Blizzard Challenge --- 2005: Evaluating Corpus-Based Speech Synthesis on Common Datasets</td>
<td>76</td>
</tr>
<tr>
<td>Alan W. Black, Keiichi Tokuda</td>
<td></td>
</tr>
<tr>
<td>A Probabilistic Approach to Unit Selection for Corpus-Based Speech Synthesis Systems</td>
<td>80</td>
</tr>
<tr>
<td>Shinsuke Sakai, Han Shu</td>
<td></td>
</tr>
<tr>
<td>The Blizzard Challenge 2005 CMU Entry --- A Method for Improving Speech Synthesis Systems</td>
<td>84</td>
</tr>
<tr>
<td>John Kominek, Christina L. Bennett, Brian Langner, Arthur R. Toth</td>
<td></td>
</tr>
<tr>
<td>Automatic Personal Synthetic Voice Construction</td>
<td>88</td>
</tr>
<tr>
<td>H. Timothy Bunnell, Chris Pennington, Debra Yarrington, John Gray</td>
<td></td>
</tr>
<tr>
<td>An Overview of Nitech HMM-Based Speech Synthesis System for Blizzard Challenge 2005</td>
<td>92</td>
</tr>
<tr>
<td>Heiga Zen, Tomoki Toda</td>
<td></td>
</tr>
<tr>
<td>On Building a Concatenative Speech Synthesis System from the Blizzard Challenge Speech Databases</td>
<td>96</td>
</tr>
<tr>
<td>Wael Hamza, Raimo Bakis, Zhi Wei Shuang, Heiga Zen</td>
<td></td>
</tr>
<tr>
<td>Multisyn Voices from ARCTIC Data for the Blizzard Challenge</td>
<td>100</td>
</tr>
<tr>
<td>Robert A.J. Clark, Korin Richmond, Simon King</td>
<td></td>
</tr>
<tr>
<td>Large Scale Evaluation of Corpus-Based Synthesizers: Results and Lessons from the Blizzard Challenge 2005</td>
<td>104</td>
</tr>
<tr>
<td>Christina L. Bennett</td>
<td></td>
</tr>
<tr>
<td>SESSION: NEW APPLICATIONS</td>
<td></td>
</tr>
<tr>
<td>Speech Retrieval of Mandarin Broadcast News via Mobile Devices</td>
<td>108</td>
</tr>
<tr>
<td>Berlin Chen, Yi-Ting Chen, Chih-Hao Chang, Hung-Bin Chen</td>
<td></td>
</tr>
<tr>
<td>State Estimation of Meetings by Information Fusion Using Bayesian Network</td>
<td>112</td>
</tr>
<tr>
<td>Michiaki Katoh, Kiyoshi Yamamoto, Jun Ogata, Takashi Yoshimura, Futoshi Asano, Hideki Asoh, Nobuhiko Kitawaki</td>
<td></td>
</tr>
<tr>
<td>Results from a Survey of Attendees at ASRU 1997 and 2003</td>
<td>116</td>
</tr>
<tr>
<td>Roger K. Moore</td>
<td></td>
</tr>
<tr>
<td>Speech Processing in the Networked Home Environment --- A View on the Amigo Project</td>
<td>120</td>
</tr>
<tr>
<td>Reinhold Haeb-Umbach, Basilis Kladis, Joerg Schmalenstroeer</td>
<td></td>
</tr>
<tr>
<td>Fixed Distortion Segmentation in Efficient Sound Segment Searching</td>
<td>124</td>
</tr>
<tr>
<td>Masahide Sugiyama</td>
<td></td>
</tr>
<tr>
<td>Identifying Singers of Popular Songs</td>
<td>128</td>
</tr>
<tr>
<td>Tin Lay Nwe, Haizhou Li</td>
<td></td>
</tr>
</tbody>
</table>
Speech Repair: Quick Error Correction Just by Using Selection Operation for Speech Input Interfaces  
Jun Ogata, Masataka Goto

Steerable Highly Directional Audio Beam Loudspeaker  
Dirk Olszewski, Fransiskus Prasetyo, Klaus Linhard

Automatic Music Genre Classification Using Second-Order Statistical Measures for the Prescriptive Approach  
Hassan Ezzaidi, Jean Rouat

Effect of Head Orientation on the Speaker Localization Performance in Smart-Room Environment  
Alberto Abad, Dusan Macho, Carlos Segura, Javier Hernando, Climent Nadeu

Application of Automatic Speaker Recognition Techniques to Pathological Voice Assessment (Dysphonia)  
Corinne Fredouille, G. Pouchoulin, Jean-Francois Bonastre, M. Azzarello, A. Giovanni, A. Ghio

Adaptive Speech Analytics: System, Infrastructure, and Behavior  
Upendra V. Chaudhari, Ganesh N. Ramaswamy, Eddie Epstein, Sasha P. Caskey, Mohamed Kamal Omar

SESSION: E-LEARNING AND SPOKEN LANGUAGE PROCESSING

Correlating Student Acoustic-Prosodic Profiles with Student Learning in Spoken Tutoring Dialogues  
Katherine Forbes-Riley, Diane J. Litman

Speech Recognition Performance and Learning in Spoken Dialogue Tutoring  
Diane J. Litman, Katherine Forbes-Riley

Structural Representation of the Non-Native Pronunciations  
Satoshi Asakawa, Nobuaki Minematsu, Toshiko Isei-Jaakkola, Keikichi Hirose

Ya-Ya Language Box --- A Portable Device for English Pronunciation Training with Speech Recognition Technologies  
Fu-chiang Chou

Pronunciation Error Detection Method Based on Error Rule Clustering Using a Decision Tree  
Akinori Ito, Yen-Ling Lim, Motoyuki Suzuki, Shozo Makino

Modeling and Automating Detection of Errors in Arabic Language Learner Speech  
Abhinav Sethy, Shrikanth Narayanan, Nicolaus Mote, W. Lewis Johnson

Effects of F0 Feedback on the Learning of Chinese Tones by Native Speakers of English  
Felicia Zhang, Michael Wagner

SESSION: E-INCLUSION AND SPOKEN LANGUAGE PROCESSING I

Voice-Controlled Internet Browsing for Motor-Handicapped Users. Design and Implementation Issues  
Tom Brondsted, Erik Aaskoven

Creating an Ongoing Research Capability in Speech Technology for Two Minority Languages: Experiences from the WISPR Project  
Briony Williams, Delyth Prys, Ailbhe Ni Chasaide

Speech Operated Smart-Home Control System for Users with Special Needs  
A. Vovos, Basilis Kladis, Nikolaos Fakotakis
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoken Dialog System and its Evaluation of Geographic Information System for Elderly Persons’ Mobility Support</td>
<td>196</td>
</tr>
<tr>
<td>Takatoshi Jitsuhiro, Shigeki Matsuda, Yutaka Ashikari, Satoshi Nakamura, Ikuko Eguchi Yairi, Seiji Igi</td>
<td></td>
</tr>
<tr>
<td>A Frame Based Spoken Dialog System for Home Care</td>
<td>200</td>
</tr>
<tr>
<td>Daniele Falavigna, Toni Giorgino, Roberto Gretter</td>
<td></td>
</tr>
<tr>
<td>Frame Based Model Order Selection of Spectral Envelopes</td>
<td>204</td>
</tr>
<tr>
<td>Matthias Wolfel</td>
<td></td>
</tr>
<tr>
<td>On Variable-Scale Piecewise Stationary Spectral Analysis of Speech Signals for ASR</td>
<td>208</td>
</tr>
<tr>
<td>Vivek Tyagi, Christian Wellekens, Herve Bourlard</td>
<td></td>
</tr>
<tr>
<td>Efficient Pitch-Based Estimation of VTLN Warp Factors</td>
<td>212</td>
</tr>
<tr>
<td>Arlo Faria, David Gelbart</td>
<td></td>
</tr>
<tr>
<td>Accent Detection and Speech Recognition for Shanghai-Accented Mandarin</td>
<td>216</td>
</tr>
<tr>
<td>Variability of Automatic Speech Recognition Systems Using Different Features</td>
<td>220</td>
</tr>
<tr>
<td>Loic Barrault, Renato de Mori, Roberto Gemello, Franco Mana, Driss Matrouf</td>
<td></td>
</tr>
<tr>
<td>Crosslingual and Bilingual Speech Recognition with Slovak and Czech SpeechDat-E Databases</td>
<td>224</td>
</tr>
<tr>
<td>Slavomir Lihan, Jozef Juhar, Anton Cizmar</td>
<td></td>
</tr>
<tr>
<td>Automatic Data Selection for MLP-Based Feature Extraction for ASR</td>
<td>228</td>
</tr>
<tr>
<td>Carmen Pelaez-Moreno, Qifeng Zhu, Barry Y. Chen, Nelson Morgan</td>
<td></td>
</tr>
<tr>
<td>Rapid Porting of ASR-Systems to Mobile Devices</td>
<td>232</td>
</tr>
<tr>
<td>Thilo W. Kohler, Christian Fugen, Sebastian Stuker, Alex Waibel</td>
<td></td>
</tr>
<tr>
<td>A Stream-Based Audio Segmentation, Classification and Clustering Pre-Processing System for Broadcast News Using ANN Models</td>
<td>236</td>
</tr>
<tr>
<td>Hugo Meinedo, Joao Neto</td>
<td></td>
</tr>
<tr>
<td>Speech Activity Detection Fusing Acoustic Phonetic and Energy Features</td>
<td>240</td>
</tr>
<tr>
<td>Etienne Marcheret, Karthik Visweswaria, Gerasimos Potamianos</td>
<td></td>
</tr>
<tr>
<td>Robust Voice Activity Detection Based on the Entropy of Noise-Suppressed Spectrum</td>
<td>244</td>
</tr>
<tr>
<td>Zoltan Tuske, Peter Mihajlik, Zoltan Tobler, Tibor Fegyo</td>
<td></td>
</tr>
<tr>
<td>Multiple Moving Speaker Tracking by Microphone Array on Mobile Robot</td>
<td>248</td>
</tr>
<tr>
<td>Masamitsu Murase, Shunichi Yamamoto, Jean-Marc Valin, Kazuhiro Nakadai, Kentaro Yamada, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno</td>
<td></td>
</tr>
<tr>
<td>A Speaker Biased SI Recognizer for Embedded Mobile Applications</td>
<td>252</td>
</tr>
<tr>
<td>Yaxin Zhang, Bian Wu, Xiaolin Ren, Xin He</td>
<td></td>
</tr>
<tr>
<td>Fast Unsupervised Speaker Adaptation Through a Discriminative Eigen-MLLR Algorithm</td>
<td>256</td>
</tr>
<tr>
<td>Bart Bakker, Carsten Meyer, Xavier Aubert</td>
<td></td>
</tr>
<tr>
<td>Incremental Largest Margin Linear Regression and MAP Adaptation for Speech Separation in Telemedicine Applications</td>
<td>260</td>
</tr>
<tr>
<td>Rusheng Hu, Jian Xue, Yunxin Zhao</td>
<td></td>
</tr>
<tr>
<td>Applying Vocal Tract Length Normalization to Meeting Recordings</td>
<td>264</td>
</tr>
<tr>
<td>Giulia Garau, Steve Renals, Thomas Hain</td>
<td></td>
</tr>
<tr>
<td>Implementing Frequency-Warping and VTLN Through Linear Transformation of Conventional MFCC</td>
<td>268</td>
</tr>
<tr>
<td>S. Umesh, Andras Zolnay, Hermann Ney</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>MLLR-Like Speaker Adaptation Based on Linearization of VTLN with MFCC Features</td>
<td>272</td>
</tr>
<tr>
<td>Xiaodong Cui, Abeer Alwan</td>
<td></td>
</tr>
<tr>
<td>Model Adaptation by State Splitting of HMM for Long Reverberation</td>
<td>276</td>
</tr>
<tr>
<td>Chandra Kant Raut, Takuya Nishimoto, Shigeki Sagayama</td>
<td></td>
</tr>
<tr>
<td>Online Speaker Adaptation and Tracking for Real-Time Speech Recognition</td>
<td>280</td>
</tr>
<tr>
<td>Daben Liu, Daniel Kiecza, Amit Srivastava, Francis Kubala</td>
<td></td>
</tr>
<tr>
<td>Automatic Speech Recognition Based on Adaptation and Clustering Using Temporal-Difference Learning</td>
<td>284</td>
</tr>
<tr>
<td>Masafumi Nishida, Yasuo Horiuchi, Akira Ichikawa</td>
<td></td>
</tr>
<tr>
<td>Improving the Speech Recognition Performance of Beginners in Spoken Conversational Interaction for Language Learning</td>
<td>288</td>
</tr>
<tr>
<td>Hui Ye, Steve Young</td>
<td></td>
</tr>
<tr>
<td>Rapid Unsupervised Speaker Adaptation Based on Multi-Template HMM</td>
<td>292</td>
</tr>
<tr>
<td>Sufficient Statistics in Noisy Environments</td>
<td></td>
</tr>
<tr>
<td>Randy Gomez, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano</td>
<td></td>
</tr>
<tr>
<td>Rapid Speaker Adaptation for Continuous Speech Recognition Using Merging Eigenvoices</td>
<td>296</td>
</tr>
<tr>
<td>Dong-jin Choi, Yung-Hwan Oh</td>
<td></td>
</tr>
<tr>
<td>Real-Time Pitch Tracking Based on Combined SMDSF</td>
<td>300</td>
</tr>
<tr>
<td>Jian Liu, Thomas Fang Zheng, Jing Deng, Wenhu Wu</td>
<td></td>
</tr>
<tr>
<td>Fundamental Frequency Estimation by Least-Squares Harmonic Model Fitting</td>
<td>304</td>
</tr>
<tr>
<td>Andras Banhalmi, Kornel Kovacs, Andras Kocsor, Laszlo Toth</td>
<td></td>
</tr>
<tr>
<td>Harmonic Filtering for Joint Estimation of Pitch and Voiced Source with Single-Microphone Input</td>
<td>308</td>
</tr>
<tr>
<td>S.W. Lee, Frank K. Soong, P.C. Ching</td>
<td></td>
</tr>
<tr>
<td>High-Resolution Noise-Robust Spectral-Based Pitch Estimation</td>
<td>312</td>
</tr>
<tr>
<td>Marian Kepesi, Luis Weruaga</td>
<td></td>
</tr>
<tr>
<td>F0 Estimation for Adult and Children’s Speech</td>
<td>316</td>
</tr>
<tr>
<td>John-Paul Hosom</td>
<td></td>
</tr>
<tr>
<td>Fundamental Frequency and Voicing Prediction from MFCCs for Speech Reconstruction from Unconstrained Speech</td>
<td>320</td>
</tr>
<tr>
<td>Ben Milner, Xu Shao, Jonathan Darch</td>
<td></td>
</tr>
<tr>
<td>F_0 Stylistation with a Free-Knot B-Spline Model and Simulated-Annealing Optimization</td>
<td>324</td>
</tr>
<tr>
<td>N. Barbot, Olivier Boeffard, D. Lolive</td>
<td></td>
</tr>
<tr>
<td>Voiced Excitation as Entrained Primary Response of a Reconstructed Glottal Master Oscillator</td>
<td>328</td>
</tr>
<tr>
<td>F.R. Drepper</td>
<td></td>
</tr>
<tr>
<td>Estimation of LF Glottal Source Parameters Based on an ARX Model</td>
<td>332</td>
</tr>
<tr>
<td>Damien Vincent, Olivier Rosec, Thierry Chonavel</td>
<td></td>
</tr>
<tr>
<td>Some Experiments on Iterative Reconstruction of Speech from STFT Phase and Magnitude Spectra</td>
<td>336</td>
</tr>
<tr>
<td>Leigh D. Alsteris, Kuldip K. Paliwal</td>
<td></td>
</tr>
<tr>
<td>Statistical Properties of the Warped Discrete Cosine Transform Cepstrum Compared with MFCC</td>
<td>340</td>
</tr>
<tr>
<td>R. Muralishankar, Abhijeet Sangwan, Douglas O’Shaughnessy</td>
<td></td>
</tr>
<tr>
<td>New Signal Features for Robust Identification of Isolated Vowels</td>
<td>344</td>
</tr>
<tr>
<td>Anibal J.S. Ferreira</td>
<td></td>
</tr>
<tr>
<td>Amplitude Modulation of Frication Noise by Voicing Saturates</td>
<td>348</td>
</tr>
<tr>
<td>Jonathan Pincas, Philip J.B. Jackson</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Extraction of Relevant Speech Features Using the Information Bottleneck Method</td>
<td>352</td>
</tr>
<tr>
<td>Comparing Several Models for Perceptual Long-Term Modeling of Amplitude and Phase Trajectories of Sinusoidal Speech</td>
<td>356</td>
</tr>
<tr>
<td>Multi-Resolution RASTA Filtering for TANDEM-Based ASR</td>
<td>360</td>
</tr>
<tr>
<td>A Category-Dependent Feature Selection Method for Speech Signals</td>
<td>364</td>
</tr>
<tr>
<td>Voicing Features for Robust Speech Detection</td>
<td>368</td>
</tr>
<tr>
<td>Joint Bayesian Predictive Classification and Parallel Model Combination for Robust Speech Recognition</td>
<td>372</td>
</tr>
<tr>
<td>Gaussian Elimination Algorithm for HMM Complexity Reduction in Continuous Speech Recognition Systems</td>
<td>376</td>
</tr>
<tr>
<td>Robust Speech Recognition in Cars Using Phoneme Dependent Multi-Environment Linear Normalization</td>
<td>380</td>
</tr>
<tr>
<td>Energy-Based Frame Selection for Reliable Feature Normalization and Transformation in Robust Speech Recognition</td>
<td>384</td>
</tr>
<tr>
<td>Remodeling of the Sensor for Non-Audible Murmur (NAM)</td>
<td>388</td>
</tr>
<tr>
<td>Focused Word Segmentation for ASR</td>
<td>392</td>
</tr>
<tr>
<td>Lexical Tone Perception in Musicians and Non-Musicians</td>
<td>396</td>
</tr>
<tr>
<td>Contextual Effect on Perception of Lexical Tones in Cantonese</td>
<td>400</td>
</tr>
<tr>
<td>Visual Cues in Mandarin Tone Perception</td>
<td>404</td>
</tr>
<tr>
<td>Cross-Language Perception of Word Stress</td>
<td>408</td>
</tr>
<tr>
<td>The Lexical Statistics of Word Recognition Problems Caused by L2 Phonetic Confusion</td>
<td>412</td>
</tr>
<tr>
<td>A Multi-Layer Fuzzy Logical Model for Emotional Speech Perception</td>
<td>416</td>
</tr>
<tr>
<td>Utterance Verification Incorporating In-Domain Confidence and Discourse Coherence Measures</td>
<td>420</td>
</tr>
<tr>
<td>Using Symbolic Prominence to Help Design Feature Subsets for Topic Classification and Clustering of Natural Human-Human Conversations</td>
<td>424</td>
</tr>
<tr>
<td>Tightly Integrated Spoken Language Understanding using Word-to-Concept Translation</td>
<td>428</td>
</tr>
</tbody>
</table>
Exploiting Unlabeled Data Using Multiple Classifiers for Improved Natural Language Call-Routing
Ruhi Sarikaya, Hong-Kwang Jeff Kuo, Vaibhava Goel, Yuqing Gao

Active Learning with Minimum Expected Error for Spoken Language Understanding
Hong-Kwang Jeff Kuo, Vaibhava Goel

Lexical Out-of-Vocabulary Models for One-Stage Speech Interpretation
Matthias Thomae, Tibor Fabian, Robert Lieb, Gunther Ruske

Speech Technology for e-Inclusion of People with Physical Disabilities and Disordered Speech
Mark S. Hawley, Phil Green, Pam Enderby, Stuart Cunningham, Roger K. Moore

Speech Technology for Language Training and e-Inclusion
Bjorn Granstrom

Supporting the Creation of TTS for Local Language Voice Information Systems
Roger Tucker, Ksenia Shalonova

Access for All --- A Talking Internet Service
Ove Andersen, Christian Hjulmand

A Speech Centric Mobile Multimodal Service Useful for Dyslectics and Aphasics
Knut Kvale, Narada Warakagoda

SESSION: PARALINGUISTIC AND NONLINGUISTIC INFORMATION IN SPEECH

No Laughing Matter
Nick Campbell, Hideki Kashioka, Ryo Ohara

A Study on the Automatic Detection and Characterization of Emotion in a Voice Service Context
C. Blouin, V. Maffiolo

Classical and Novel Discriminant Features for Affect Recognition from Speech
Raul Fernandez, Rosalind W. Picard

SESSION: LOW-DIMENSIONAL FEATURE SPACE DERIVATION FOR EMOTION RECOGNITION

Low-Dimensional Feature Space Derivation for Emotion Recognition
Jaroslaw Cichosz, Krzysztof Slot

Proposal of Acoustic Measures for Automatic Detection of Vocal Fry
Carlos Toshinori Ishi, Hiroshi Ishiguro, Norihiro Hagita

Automatic Detection of Laughter
Khiet P. Truong, David A. van Leeuwen

Tales of Tuning --- Prototyping for Automatic Classification of Emotional User States
Anton Battiner, Stefan Steidl, Christian Hacker, Elmar Noth, Heinrich Niemann

Automatic Emotion Recognition Using Prosodic Parameters
Iker Luengo, Eva Navas, Inmaculada Hernaez, Jon Sanchez

An Articulatory Study of Emotional Speech Production
Sungbok Lee, Serdar Yildirim, Abe Kazemzadeh, Shrikanth Narayanan

Informed Blending of Databases for Emotional Speech Synthesis
Gregor O. Hofer, Korin Richmond, Robert A.J. Clark
Emotional FESTIVAL-MBROLA TTS Synthesis ......................................................... 504
Fabio Tesser, Piero Cosi, Carlo Drioli, Graziano Tisato

Emofilt: The Simulation of Emotional Speech by Prosody-Transformation........................ 508
Felix Burkhardt

Acoustic/Prosodic and Lexical Correlates of Charismatic Speech ........................................ 512
Andrew Rosenberg, Julia Hirschberg

Communicative Speech Synthesis Using Constituent Word Attributes .................................. 516
Yoko Greenberg, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka

Emotions in Dubbed Speech: An Intercultural Approach with Respect to F0 .......................... 520
Angelika Braun, Matthias Katerbow

The Prosodic Dimensions of Emotion in Speech: the Relative Weights of Parameters ............... 524
Nicolas Audibert, Veronique Auberge, Albert Rilliard

Stimulus Duration and Type in Perception of Female and Male Speaker Age .......................... 528
Susanne Schotz

Perceptions of Emotions in Expressive Storytelling ........................................................... 532
Cecilia Ovesdotter Alm, Richard Sproat

Nearly Defect-Free F0 Trajectory Extraction for Expressive Speech Modifications Based on STRAIGHT ........................................................................................................................ 536
Hideki Kawahara, Alain de Cheveigne, Hideki Banno, Toru Takahashi, Toshio Irino

Gradually Changing Expression of Singing Voice Based on Morphing ................................ 540
Tomoko Yonezawa, Noriko Suzuki, Kenji Mase, Kiyoshi Kogure

SESSION: ISSUES IN LARGE VOCABULARY DECODING

A Multi-Pass, Dynamic-Vocabulary Approach to Real-Time, Large-Vocabulary Speech Recognition ................................................................................................................................. 544
I. Lee Hetherington

Anatomy of an Extremely Fast LVCSR Decoder ...................................................................... 548
George Saon, Daniel Povey, Geoffrey Zweig

Evaluation of a Long-Contextual-Span Hidden Trajectory Model and Phonetic Recognizer Using A** Lattice Search ................................................................. 552
Dong Yu, Li Deng, Alex Acero

Generalized Fast On-the-Fly Composition Algorithm for WFST-Based Speech Recognition ................................................................................................................................. 556
Takaaki Hori, Atsushi Nakamura

Minimum Bayes-Risk Decoding Considering Word Significance for Information Retrieval System ................................................................................................................................. 560
Hiroaki Nanjo, Teruhisa Misu, Tatsuya Kawahara

On Improvements to CI-Based GMM Selection ...................................................................... 564
Arthur Chan, Mosur Ravishankar, Alexander I. Rudnicky

Scalable Language Model Look-Ahead for LVCSR ............................................................... 568
Dominique Massonie, Pascal Nocera, Georges Linares

Memory Efficient Approximative Lattice Generation for Grammar Based Decoding ............... 572
Miroslav Novak

Improved Semi-Dynamic Network Decoding Using WFSTs .................................................. 576
Dong-Hoon Ahn, Su-Byeong Oh, Minhwa Chung

New Pruning Criteria for Efficient Decoding ................................................................. 580
Janne Pylkkonen
A Confidence-Guided Dynamic Pruning Approach --- Utilization of Confidence Measurement in Speech Recognition ................................................................................................................. 584
Tibor Fabian, Robert Lieb, Gunther Ruske, Matthias Thomae

SESSION: SPOKEN LANGUAGE EXTRACTION / RETRIEVAL II

Discrimination of Speech, Musical Instruments and Singing Voices Using the Temporal Patterns of Sinusoidal Segments in Audio Signals ................................................................. 588
Toru Taniguchi, Akishige Adachi, Shigeki Okawa, Masaaki Honda, Katsuhiko Shirai

Extractive Summarization of Meeting Recordings ................................................................................................................................. 592
Gabriel Murray, Steve Renals, Jean Carletta

IR-Based Classification of Customer-Agent Phone Calls ...................................................................................................................... 596
Arjan van Hessen, Jaap Hinke

Mining Broadcast News Data: Robust Information Extraction from Word Lattices .................................................................................. 600
Benoit Favre, Frederic Bechet, Pascal Nocera

To Recover from Speech Recognition Errors in Spoken Document Retrieval ......................................................................................... 604
Mikko Kurimo, Ville Turunen

Unsupervised Clustering of Spontaneous Speech Documents ...................................................................................................................... 608
Edgar Gonzalez, Jordi Turmo

Spectral Cross-Correlation Features for Audio Indexing of Broadcast News and Meetings ........................................................................... 612
Masahide Yamaguchi, Masaru Yamashita, Shoichi Matsunaga

Spontaneous Speech Consolidation for Spoken Language Applications ........................................................................................................ 616
Chiori Hori, Alex Waibel

Comparing Lexical, Acoustic/Prosodic, Structural and Discourse Features for Speech Summarization ................................................................. 620
Sameer Maskey, Julia Hirschberg

Hierarchical Topic Organization and Visual Presentation of Spoken Documents Using Probabilistic Latent Semantic Analysis (PLSA) for Efficient Retrieval/Browsing Applications ........................................................................................................... 624
Te-Hsuan Li, Ming-Han Lee, Berlin Chen, Lin-Shan Lee

The COST278 Broadcast News Segmentation and Speaker Clustering Evaluation --- Overview, Methodology, Systems, Results .......................................................................................... 628
Janez Zibert, France Mihelic, Jean-Pierre Martens, Hugo Meinedo, Joao Neto, Laura Docio, Carmen Garcia Mateo, Petr David, Jindrich Zd'ansky, Matus Pleva, Anton Cizmar, Andrej Zgank, Zdravko Kacic, Csaba Teleki, Klara Vicsi

Comparison of Keyword Spotting Approaches for Informal Continuous Speech .......................................................................................... 632
Igor Szoke, Petr Schwarz, Pavel Matejka, Lukas Burget, Martin Karafiat, Michal Fapso, Jan Cernocky

Dialogue Strategy to Clarify User’s Queries for Document Retrieval System with Speech Interface .......................................................................................... 636
Teruhisa Misu, Tatsuya Kawahara

Comparison of Different Phone-Based Spoken Document Retrieval Methods with Text and Spoken Queries .................................................................................. 640
Nicolas Moreau, Shan Jin, Thomas Sikora

SESSION: SIGNAL ANALYSIS, PROCESSING AND FEATURE ESTIMATION II

PCA of Perturbation Parameters in Voice Pathology Detection ........................................................................................................................ 644
Pedro Gomez, Francisco Diaz, Agustin Alvarez, Rafael Martinez, Victoria Rodellar, Roberto Fernandez-Baillo, Alberto Nieto, Francisco J. Fernandez
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic Programming Based Segmentation Approach to LSF Matrix</td>
<td>648</td>
</tr>
<tr>
<td>Explicit Segmentation of Speech Based on Frequency-Domain AR Modeling</td>
<td>652</td>
</tr>
<tr>
<td>Non-Parametric Speaker Turn Segmentation of Meeting Data</td>
<td>656</td>
</tr>
<tr>
<td>Unsupervised Segmentation of Continuous Speech Using Vector Autoregressive Time-Frequency Modeling Errors</td>
<td>660</td>
</tr>
<tr>
<td>The Analysis on Band-Limited Hypernasal Speech Using Group Delay Based Formant Extraction Technique</td>
<td>664</td>
</tr>
<tr>
<td>Detection of Acoustic Change-Points in Audio Records via Global BIC Maximization and Dynamic Programming</td>
<td>668</td>
</tr>
<tr>
<td>Multi-Band Approach of Audio Source Discrimination with Empirical Mode Decomposition</td>
<td>672</td>
</tr>
<tr>
<td>Application of Auditory Image Model for Speech Event Detection</td>
<td>676</td>
</tr>
<tr>
<td>Unsupervised Identification of Speech Segments Using Kernel Methods for Clustering</td>
<td>680</td>
</tr>
<tr>
<td>Speech Event Detection Using Multiband Modulation Energy</td>
<td>684</td>
</tr>
<tr>
<td>Measuring Unsupervised Acoustic Clustering Through Phoneme Pair Merge-and-Split Tests</td>
<td>688</td>
</tr>
</tbody>
</table>

Volume 2

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variational Bayesian Speaker Change Detection</td>
<td>692</td>
</tr>
<tr>
<td>Distinctive Feature Based SVM Discriminant Features for Improvements to Phone Recognition on Telephone Band Speech</td>
<td>696</td>
</tr>
<tr>
<td>Detection of Hypernasality Using Statistical Pattern Classifiers</td>
<td>700</td>
</tr>
<tr>
<td>Self-Organizing Chirp-Sensitive Artificial Auditory Cortical Model</td>
<td>704</td>
</tr>
<tr>
<td>On the Use of a Decimative Spectral Estimation Method Based on Eigenanalysis and SVD for Formant and Bandwidth Tracking of Speech Signals</td>
<td>708</td>
</tr>
<tr>
<td>Frequency-Domain Auditory Suppression Modelling (FASM) --- A WDFT-Based Anthropomorphic Noise-Robust Feature Extraction Algorithm for Speech Recognition</td>
<td>712</td>
</tr>
</tbody>
</table>
SESSION: KEYNOTE SPEAKER II

Linear Models for Structure Prediction ................................................................. 716
Fernando C.N. Pereira

Discriminative Maximum Entropy Language Model for Speech Recognition .......... 720
Chuang-Hua Chueh, To-Chang Chien, Jen-Tzung Chien

Open Vocabulary Speech Recognition with Flat Hybrid Models .......................... 724
Maximilian Bisani, Hermann Ney

An Error-Corrective Language-Model Adaptation for Automatic Speech Recognition ................................................................. 728
Minwoo Jeong, Jihyun Eun, Sangkeun Jung, Gary Geunbae Lee

Discriminative Training of Finite State Decoding Graphs ....................................... 732
Shiuan-Sung Lin, Francois Yvon

Building Continuous Space Language Models for Transcribing European Languages ........................................................................................................ 736
Holger Schwenk, Jean-Luc Gauvain

Using Random Forest Language Models in the IBM RT-04 CTS System............... 740
Peng Xu, Lidia Mangu

Perceptual Development of the Duration Cue in Dutch /textscripta/-lengthmark / ....... 744
Willemijn Heeren

Pronunciation Variations of Spanish-Accented English spoken by Young Children ................................................................. 748
Hong You, Abeer Alwan, Abe Kazemzadeh, Shrikanth Narayanan

L2 Development of Quantity Perception: Dutch Listeners Learning Finnish /t/-lengthmark / ................................................................. 752
Willemijn Heeren

Phonetic Inventories in Italian Children Aged 18--27 Months: A Longitudinal Study ........................................................................................................ 756
Claudio Zmarich, Serena Bonifacio

Pitch Patterns of Intonational Phrases and Intonational Phrase Groups in Native and Non-Native Speech ........................................................................... 760
Hiroko Hirano, Goh Kawai

Measuring Liveliness in Presentation Speech .......................................................... 764
Rebecca Hincks

Non-Verbal Speech Processing for a Communicative Agent ..................................... 768
Nick Campbell

Physiologically Motivated Audio-Visual Localisation and Tracking ........................ 772
Stuart N. Wrigley, Guy J. Brown

Discriminatively Trained Features Using fMPE for Multi-Stream Audio-Visual Speech Recognition ................................................................. 776
Jing Huang, Daniel Povey

INTERFACE: A New Tool for Building Emotive/Expressive Talking Heads .......... 780
Graziano Tisato, Piero Cosi, Carlo Drioli, Fabio Tesser

Variance Reduction by Using Separate Genuine-Impostor Statistics in Multimodal Biometrics ................................................................. 784
P. Ejarque, Javier Hernando

The Dialog Application Metalanguage GDialogXML ............................................ 788
Volker Schubert, Stefan W. Hamerich
SESSION: EMOTIONAL SPEECH ANALYSIS AND SYNTHESIS

Data-Driven Synthesis of Expressive Visual Speech Using an MPEG-4 Talking Head
Jonas Beskow, Mikael Nordenberg ................................................................. 792

Voice Quality Interpolation for Emotional Text-to-Speech Synthesis
Oytun Turk, Marc Schroder, Baris Bozkurt, Levent M. Arslan ................. 796

Investigating the Role of Phoneme-Level Modifications in Emotional Speech Resynthesis
Murtaza Bulut, Carlos Busso, Serdar Yildirim, Abe Kazemzadeh, Chul Min Lee, Sungbok Lee, Shrikanth Narayanan ................................................. 800

Speaker Independent Emotion Recognition by Early Fusion of Acoustic and Linguistic Features Within Ensembles
Bjorn Schuller, Ronald Muller, Manfred Lang, Gerhard Rigoll .................... 804

Integrating Information from Speech and Physiological Signals to Achieve Emotional Sensitivity
Jonghwa Kim, Elisabeth Andre, Matthias Rehm, Thurid Vogt, Johannes Wagner ................................................................. 808

Multimodal Databases of Everyday Emotion: Facing up to Complexity
Ellen Douglas-Cowie, Laurence Devillers, Jean-Claude Martin, Roddy Cowie, Suzie Savvidou, Sarkis Abrilian, Cate Cox .............................. 812

SESSION: SPOKEN / MULTI-MODAL DIALOGUE SYSTEMS I

Learning of Stochastic Dialog Models Through a Dialog Simulation Technique
Francisco Torres, Emilio Sanchis, Encarna Segarra ........................................ 816

Evaluating the Di@L-log System on a Cohort of Elderly, Diabetic Patients: Results from a Preliminary Study
Lesley-Ann Black, Michael McTear, Norman Black, Roy Harper, Michelle Lemon ................................................................. 820

Combination of Classifiers for Automatic Recognition of Dialog Acts
Pavel Kral, Christophe Cerisara, Jana Kleckova ............................................ 824

Rapidly Developing Spoken Chinese Dialogue Systems with the d-Ear SDS SDK
Xiaojun Wu, Thomas Fang Zheng, Michael Brasser, Zhanjiang Song ........ 828

Robust Algorithms and Interaction Strategies for Voice Spelling
Daniela Oria, Akos Vetek ............................................................................. 832

Modality Integration and Dialog Management for a Robotic Assistant
Ioannis Toptsis, Axel Haasch, Sonja Huwel, Jannik Fritsch, Gernot A. Fink ................................................................. 836

An Integration Framework for a Mobile Multimodal Dialogue System Accessing the Semantic Web
Norbert Reithinger, Daniel Sonntag ................................................................ 840

Operating a Public Spoken Guidance System in Real Environment
Ryuichi Nisimura, Akinobu Lee, Masashi Yamada, Kiyohiro Shikano ....... 844

Distributed Dialogue Management for Smart Terminal Devices
Esa-Pekka Salonen, Markku Turunen, Jaakko Hakulinen, Leena Helin, Perttu Prusi, Anssi Kainulainen ......................................................... 848

Visualization of Spoken Dialogue Systems for Demonstration, Debugging and Tutoring
Jaakko Hakulinen, Markku Turunen, Esa-Pekka Salonen ............................. 852

Development and Evaluation of a Spoken Dialog System to Access a Newspaper Web Site
Cesar Gonzalez-Ferreras, Valentin Cardenoso-Payo ....................................... 856
Comparing ASR Modeling Methods for Spoken Dialogue Simulation and Optimal Strategy Learning ................................................................. 860
  Olivier Pietquin, Richard Beaufort

An Approach to Multi-Strategy Dialogue Management .................................................. 864
  Shiu-Wah Chu, Ian O’Neill, Philip Hanna, Michael McTear

  Anna Hjalmarsson

Reducing the Description Amount in Authoring MMI Applications ............................... 872
  Kouichi Katsurada, Kazumine Aoki, Hirobumi Yamada, Tsuneo Nitta

Contextual Constraints Based on Dialogue Models in Database Search Task for Spoken Dialogue Systems .......................................................... 876
  Kazunori Komatani, Naoyuki Kanda, Tetsuya Ogata, Hiroshi G. Okuno

Using Word-Level Pitch Features to Better Predict Student Emotions During Spoken Tutoring Dialogues ........................................................... 880
  Mihai Rotaru, Diane J. Litman

Let's Go Public! Taking a Spoken Dialog System to the Real World ................................ 884
  Antoine Raux, Brian Langner, Dan Bohus, Alan W. Black, Maxine Eskenazi

Back-Channel Feedback Generation Using Linguistic and Nonlinguistic Information and its Application to Spoken Dialogue System .......................... 888
  Shinya Fujie, Kenta Fukushima, Tetsunori Kobayashi

Learning User Simulations for Information State Update Dialogue Systems .................. 892
  Kallirroi Georgila, James Henderson, Oliver Lemon

Design of a Voice-Enabled Interface for Real-Time Access to Stock Exchange from a PDA Through GPRS ......................................................... 896
  Dario Martin-Iglesias, Yago Pereiro-Estevan, Ana I. Garcia-Moral, Ascension Gallardo-Antolin, Fernando Diaz-de-Maria

Integrating Denotational Meaning into a DBN Language Model ................................... 900
  William Schuler, Tim Miller

Improving Out-of-Coverage Language Modelling in a Multimodal Dialogue System Using Small Training Sets ..................................................... 904
  Louis ten Bosch

Ritel: An Open-Domain, Human-Computer Dialog System ........................................ 908
  Olivier Galibert, Gabriel Illouz, Sophie Rosset

A Comparison of Particle Filtering Variants for Speech Feature Enhancement ................. 912
  Reinhold Haeb-Umbach, Joerg Schmalenstroer

Enhancement of Mel Log-Power Spectrum of Speech Using Particle Filtering ................ 916
  Ilyas Potamitis, Nikolaos Fakotakis

Improving Robustness of Speech Recognition Performance to Aggregate of Noises by Two-Dimensional Visualization .............................................. 920
  Makoto Shozakai, Goshu Nagino

Feature Compensation Based on Switching Linear Dynamic Model and Soft Decision ................................................................. 924
  Woohyung Lim, Bong Kyoung Kim, Nam Soo Kim

Using Output Probability Distribution for Improving Speech Recognition in Adverse Environment ................................................................. 928
  Shilei Huang, Xiang Xie, Jingming Kuang

A Generalized Framework for Compensation of Mel-Filterbank Outputs in Feature Extraction for Robust ASR ..................................................... 932
  Eric H.C. Choi
Robust Automatic Speech Recognition Using a Perceptually-Based Optimal Spectral Amplitude Estimator Speech Enhancement Algorithm in Various Low-SNR Environments.......................................................................................................................... 936

Hesham Tolba, Zili Li, Douglas O'Shaughnessy

Improved Noise-Robustness in Distributed Speech Recognition via Perceptually-Weighted Vector Quantisation of Filterbank Energies.......................................................................................................................... 940

Stephen So, Kuldeep K. Paliwal

Sub-Band Weighted Projection Measure for Robust Sub-Band Speech Recognition .......................................................................................................................................................................................................................................................... 944

Babak Nasersharif, Ahmad Akbari

Noise Compensation Using Interacting Multiple Kalman Filters .......................................................................................................................................................................................................................................................... 948

Jianping Deng, Martin Bouchard, Tet Hin Yeap

Kalman and Unscented Kalman Filter Feature Enhancement for Noise Robust ASR .......................................................................................................................................................................................................................................................... 952

Veronique Stouten, Hugo Van hamme, Patrick Wambacq

Histogram-Based Quantization (HQ) for Robust and Scalable Distributed Speech Recognition .......................................................................................................................................................................................................................................................... 956

Chia-yu Wan, Lin-Shan Lee

A Data-Driven Approach for the Model Parameter Compensation in Noisy Speech Recognition .......................................................................................................................................................................................................................................................... 960

Yong-Joo Chung

Rapid Response and Robust Speech Recognition by Preliminary Model Adaptation for Additive and Convolutional Noise .......................................................................................................................................................................................................................................................... 964

Satoshi Kobashikawa, Satoshi Takahashi, Yoshikazu Yamaguchi, Atsunori Ogawa

Nonlinear and Linear Transformations of Speech Features to Compensate for Channel and Noise Effects .......................................................................................................................................................................................................................................................... 968

Saurabh Prasad, Stephen A. Zahorian

Construction Method of Acoustic Models Dealing with Various Background Noises Based on Combination of HMMs .......................................................................................................................................................................................................................................................... 972

Motoyuki Suzuki, Yusuke Kato, Akinori Ito, Shozi Makino

Robust Speech Recognition Based on Noise and SNR Classification --- A Multiple-Model Framework .......................................................................................................................................................................................................................................................... 976

Haitian Xu, Zheng-Hua Tan, Paul Dalsgaard, Borge Lindberg

Eigen-Environment Based Noise Compensation Method for Robust Speech Recognition .......................................................................................................................................................................................................................................................... 980

Hwa Jeon Song, Hyung Soon Kim

Robust Feature Compensation in Nonstationary and Multiple Noise Environments .......................................................................................................................................................................................................................................................... 984

Martin Graciarena, Horacio Franco, Greg Myers, Victor Abrash

Maximum Mutual Information SPLICE Transform for Seen and Unseen Conditions .......................................................................................................................................................................................................................................................... 988

Jasha Droppo, Alex Acero

Speech Recognition with Support Vector Machines in a Hybrid System .......................................................................................................................................................................................................................................................... 992

Sven E. Kruger, Martin Schaffoner, Marcel Katz, Edin Andelic, Andreas Wendemuth

Experiments on Speaker Profile Portability .......................................................................................................................................................................................................................................................... 996

Vincent Barreaud, Douglas O'Shaughnessy, Jean-Guy Dahan

A Confidence Measure Invariant to Language and Grammar .......................................................................................................................................................................................................................................................... 1000

Daniele Colibro, Luciano Fissore, Claudio Vair, Emanuele Dalmasso, Pietro Laface

Robust Detection of Sonorant Landmarks .......................................................................................................................................................................................................................................................... 1004

Ken Schutte, James Glass

The Labial-Coronal Effect and CVCV Stability During Reiterant Speech Production: An Acoustic Analysis .......................................................................................................................................................................................................................................................... 1008

Amelie Rochet-Capellan, Jean-Luc Schwartz
The Labial-Coronal Effect and CVCV Stability During Reiterant Speech Production: An articulatory Analysis
Amelie Rochet-Capellan, Jean-Luc Schwartz

Articulatory Constraints and Coronal Stops: An EPG Study
Mitsuhiro Nakamura

Strategies of Labial Coarticulation
Vincent Robert, Brigitte Wrobel-Dautcourt, Yves Laprie, Anne Bonneau

Investigation and Modeling of Coarticulation During Speech
Jianwu Dang, Jianguo Wei, Takeharu Suzuki, Pascal Perrier

Tongue Kinematics in Diphthong Production in Ningbo Chinese
Fang Hu

Comparing Tongue Positions of Vowels in Oral and Nasal Contexts
Takayuki Arai

Can We Retrieve Vocal Tract Dynamics That Produced Speech? Toward a Speaker Articulatory Strategy Model
Slim Ouni

Modeling the Production of VCV Sequences via the Inversion of a Biomechanical Model of the Tongue
Pascal Perrier, Liang Ma, Yohan Payan

Estimation of the Acoustic Properties of the Nasal Tract During the Production of Nasalized Vowels
Xiaoqian Niu, Alexander Kain, Jan P.H. van Santen

A Web-Based Articulatory Speech Synthesis System for Distance Education
Kohichi Ogata

Group Delay Function as a Means to Assess Quality of Glottal Inverse Filtering
Paavo Alku, Matti Airas, Tom Backstrom, Hannu Pulakka

Subglottal Pressure and NAQ Variation in Voice Production of Classically Trained Baritone Singers
Eva Bjorkner, Johan Sundberg, Paavo Alku

Covariation of Subglottal Pressure, F0 and Intensity
Gunnar Fant, Anita Kruckenber

Automatic Voice-Source Parameterization of Natural Speech
Javier Perez, Antonio Bonafonte

Physiological Study of Whispered Speech in Moroccan Arabic
Chakir Zeroual, John H. Esling, Lise Crevier-Buchman

Voice Quality in Down Syndrome Children Treated with Rapid Maxillary Expansion

Synthesis of Disordered Speech
Julien Hanquinet, Francis Grenez, Jean Schoentgen

Quasi-Automatic Extraction of Tongue Movement from a Large Existing Speech Cineradiographic Database
Julie Fontecave, Frederic Berthommier

The Working Memory Token Test (WMTT): Preliminary Findings in Young Adults With and Without Dyslexia
Shimon Sapir, Ravit Cohen Mimran

Reducing the Corpus-Based TTS Signal Degradation Due to Speaker’s Word Pronunciations
Sergio Paulo, Luis C. Oliveira
A Phonetic Study of the “er-hua” Rimes in Beijing Mandarin
Wai-Sum Lee ................................................................. 1092

Learning Statistically Characterized Resonance Targets in a Hidden Trajectory
Model of Speech Coarticulation and Reduction
Li Deng, Dong Yu, Alex Acero ........................................ 1096

Articulatory Motivated Acoustic Features for Speech Recognition
Danil Kocharov, Andras Zolnay, Ralf Schluter, Hermann Ney
................................................................. 1100

Effects of Bayesian Predictive Classification Using Variational Bayesian
Posteriors for Sparse Training Data in Speech Recognition
Shinji Watanabe, Atsushi Nakamura ................................... 1104

A Study on Separation Between Acoustic Models and its Applications
Yu Tsao, Jinyu Li, Chin-Hui Lee ........................................ 1108

Extended Baum-Welch Reestimation of Gaussian Mixture Models Based on
Reverse Jensen Inequality
Mohamed Afify ............................................................. 1112

Hidden Conditional Random Fields for Phone Classification
Asela Gunawardana, Milind Mahajan, Alex Acero, John C. Platt
........................................................................ 1116

Asymptotically Exact AM-FM Decomposition Based on Iterated Hilbert Transform
Francesco Gianfelici, Giorgio Biagetti, Paolo Crippa, Claudio Turchetti
........................................................................ 1120

Advances in Statistical Estimation and Tracking of AM-FM Speech Components
Athanassios Catsamanis, Petros Maragos .......................... 1124

Formant Frequency Prediction from MFCC Vectors in Noisy Environments
Jonathan Darch, Ben Milner, Saeed Vaseghi.................. 1128

Detection of Vowel Onset Point Events Using Excitation Information
S.R. Mahadeva Prasanna, B. Yegnanarayana .......................... 1132

Pitch-Synchronous Time-Scaling for Prosodic and Voice Quality Transformations
Joao P. Cabral, Luis C. Oliveira ................................. 1136

Discrimination Between Singing and Speaking Voices
Yasunori Ohishi, Masataka Goto, Katunobu Itou, Kazuya Takeda
........................................................................ 1140

Two Experiments Comparing Reading with Listening for Human Processing of
Conversational Telephone Speech
Douglas Jones, Wade Shen, Elizabeth Shriberg, Andreas Stolcke, Teresa Kamm, Douglas
Reynolds .................................................................. 1144

The ESTER Phase II Evaluation Campaign for the Rich Transcription of French
Broadcast News
Sylvain Galliano, Edouard Geoffrois, Djamel Mostefa, Khalid Choukri, Jean-Francois
Bonastre, Guillaume Gravier ............................................ 1148

A Method of Multi-Layered Speech Segmentation Tailored for Speech Synthesis
Takashi Saito .............................................................. 1152

Generation of Word Alternative Pronunciations Using Weighted Finite State
Transducers
Sergio Paulo, Luis C. Oliveira ........................................ 1156

Multword Expressions in Spontaneous Speech: Do We Really Speak Like That?
Helmur Strik, Diana Binnenpoorte, Catia Cucchiarini ............ 1160

Czech Spontaneous Speech Corpus with Structural Metadata
Jachym Kolar, Jan Svec, Stephanie Strassel, Christopher Walker, Dagmar Kozlikova, Josef
Psutka .................................................................... 1164
SESSION: EARLY LANGUAGE ACQUISITION

A Longitudinal Analysis of the Spectral Peaks of Vowels for a Japanese Infant .......................... 1168
Kentaro Ishizuka, Ryoko Mugitani, Hiroko Kato, Shigeaki Amano

Cross-Linguistic Comparison of Two-Year-Old Children's Acoustic Vowel Spaces: Contrasting Hungarian with Dutch................................................. 1172
Krisztina Zajdo, Jeannette M. van der Stelt, Ton G. Wempe, Louis C.W. Pols

Acoustic Correlates of Contrastive Stress in German Children ................................................... 1176
Britta Lintfert, Katrin Schneider

Ecological Language Acquisition via Incremental Model-Based Clustering.................................... 1180
Giampiero Salvi

Perceptual and Linguistic Category Formation in Infants ............................................................. 1184
Tamami Sudo, Ken Mogi

SESSION: MULTI-MODAL / MULTI-MEDIA PROCESSING II

Myoelectric Signals for Multimodal Speech Recognition .......................................................... 1188
Raghunandan S. Kumaran, Karthik Narayanan, John N. Gowdy

Is Color Information Really Useful for Lip-Reading? (or What is Lost When Color is not Used) ........................................................................................................ 1192
Philippe Daubias

A System for Audio-Visual Speech Recognition ......................................................................... 1196
I. Shdaifat, R.-R. Grigat

Multimodal Interface for Organization Name Input Based on Combination of Isolated Word Recognition and Continuous Base-Word Recognition ........................................................................ 1200
Norihide Kitaoka, Hironori Oshikawa, Seichi Nakagawa

Recognition of 3 Party Conversation Using Prosody and Gaze .................................................. 1204
Yosuke Matsusaka

Combining Voiceprint and Face Biometrics for Speaker Identification Using SDWS .................... 1208
Dongdong Li, Yingchun Yang, Zhaohui Wu

Using the Focus of Visual Attention to Improve Spontaneous Speech Recognition ................. 1212
Neil Cooke, Martin Russell

Real-Time Outer Lip Contour Tracking for HCI Applications ..................................................... 1216
Sabri Gurbuz

Improving Lip-Reading with Feature Space Transforms for Multi-Stream Audio-Visual Speech Recognition .................................................................................. 1220
Jing Huang, Karthik Visweswaranah

Are There Facial Correlates of Thai Syllabic Tones? .................................................................... 1224
Hansjorg Mixdorff, Denis Burnham, Guillaume Vignali, Patavee Charnvivit

A New Posterior Based Audio-Visual Integration Method for Robust Speech Recognition ........ 1228
Rowan Seymour, Ji Ming, Darryl Stewart

On Integrating Insights from Human Speech Perception into Automatic Speech Recognition ....... 1232
Sorin Dusan, Larry R. Rabiner

Parallels Between HSR and ASR: How ASR can Contribute to HSR ........................................ 1236
Odette Scharenborg

ASR Decoding in a Computational Model of Human Word Recognition .................................... 1240
Louis ten Bosch, Odette Scharenborg
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>An investigation into a Simulation of Episodic Memory for Automatic Speech Recognition</td>
<td>124</td>
</tr>
<tr>
<td>Viktoria Maier, Roger K. Moore</td>
<td></td>
</tr>
<tr>
<td>Phonetic Ignorance is Bliss: Investigating the Effects of Phonetic Information Reduction on ASR Performance</td>
<td>124</td>
</tr>
<tr>
<td>Eric Fosler-Lussier, C. Anton Rytting, Soundararajan Srinivasan</td>
<td></td>
</tr>
<tr>
<td>Automatic Speech Recognition with Neural Spike Trains</td>
<td>125</td>
</tr>
<tr>
<td>Marcus Holmberg, David Gelbart, Ulrich Ramacher, Werner Hemmert</td>
<td></td>
</tr>
<tr>
<td>A Speech Similarity Distance Weighting for Robust Recognition</td>
<td>125</td>
</tr>
<tr>
<td>Michael J. Carey, Tuan P. Quang</td>
<td></td>
</tr>
<tr>
<td>Japanese Vowel Recognition Based on Structural Representation of Speech</td>
<td>126</td>
</tr>
<tr>
<td>Takao Murakami, Kazutaka Maruyama, Nobuaki Minematsu, Keikichi Hirose</td>
<td></td>
</tr>
<tr>
<td>Modeling the Perception of Multitalker Speech</td>
<td>126</td>
</tr>
<tr>
<td>Soundararajan Srinivasan, DeLiang Wang</td>
<td></td>
</tr>
<tr>
<td>Binaural Feature Selection for Missing Data Speech Recognition</td>
<td>126</td>
</tr>
<tr>
<td>Sue Harding, Jon Barker, Guy J. Brown</td>
<td></td>
</tr>
<tr>
<td>Oldenburg Logatome Speech Corpus (OLLO) for Speech Recognition Experiments with Humans and Machines</td>
<td>127</td>
</tr>
<tr>
<td>Thorsten Wesker, Bernd Meyer, Kirsten Wagener, Jorn Anemuller, Alfred Mertins, Birger Kollmeier</td>
<td></td>
</tr>
</tbody>
</table>

**SESSION: SPEECH RECOGNITION – LANGUAGE MODELING III**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Word Error Based Discriminative Training of Language Models</td>
<td>1276</td>
</tr>
<tr>
<td>Jen-Wei Kuo, Berlin Chen</td>
<td></td>
</tr>
<tr>
<td>On the Use of Morphological Constraints in N-gram Statistical Language Model</td>
<td>1280</td>
</tr>
<tr>
<td>A. Ghaoui, Francois Yvon, C. Mokbel, G. Chollet</td>
<td></td>
</tr>
<tr>
<td>A posteriori Multiple Word-Domain Language Model</td>
<td>1284</td>
</tr>
<tr>
<td>Elvira I. Sicilia-Garcia, Ji Ming, F. Jack Smith</td>
<td></td>
</tr>
<tr>
<td>Effective Topic-Tree Based Language Model Adaptation</td>
<td>1288</td>
</tr>
<tr>
<td>Javier Dieguez-Tirado, Carmen Garcia Mateo, Antonio Cardenal-Lopez</td>
<td></td>
</tr>
<tr>
<td>Building Topic Specific Language Models from Webdata Using Competitive Models</td>
<td>1292</td>
</tr>
<tr>
<td>Abhinav Sethy, Panayiotis G. Georgiou, Shrikanth Narayanan</td>
<td></td>
</tr>
<tr>
<td>Trigger-Based Language Model Adaptation for Automatic Meeting Transcription</td>
<td>1296</td>
</tr>
<tr>
<td>Carlos Troncoso, Tatsuuya Kawahara</td>
<td></td>
</tr>
<tr>
<td>Statistical Language Models for Large Vocabulary Spontaneous Speech Recognition in Dutch</td>
<td>1300</td>
</tr>
<tr>
<td>Jacques Duchateau, Dong Hoon Van Uytsel, Hugo Van hamme, Patrick Wambacq</td>
<td></td>
</tr>
<tr>
<td>Diachronic Vocabulary Adaptation for Broadcast News Transcription</td>
<td>1304</td>
</tr>
<tr>
<td>Alexandre Allauzen, Jean-Luc Gauvain</td>
<td></td>
</tr>
<tr>
<td>Growing an n-gram Language Model</td>
<td>1308</td>
</tr>
<tr>
<td>Vesa Siivola, Bryan L. Pellom</td>
<td></td>
</tr>
<tr>
<td>Embedding Grammars into Statistical Language Models</td>
<td>1312</td>
</tr>
<tr>
<td>Harald Huning, Manuel Kirschner, Fritz Class, Andre Berton, Udo Haiber</td>
<td></td>
</tr>
<tr>
<td>Methods for Combining Language Models in Speech Recognition</td>
<td>1316</td>
</tr>
<tr>
<td>Simo Broman, Mikko Kurimo</td>
<td></td>
</tr>
<tr>
<td>Review of Statistical Modeling of Highly Inflected Lithuanian Using Very Large Vocabulary</td>
<td>1320</td>
</tr>
<tr>
<td>Airenas Vaiciinas, Gailius Raskinis</td>
<td></td>
</tr>
</tbody>
</table>
SESSION: SPEECH RECOGNITION – PRONUNCIATION MODELING

Automatic Generation of Domain-Dependent Pronunciation Lexicon with Data-Driven Rules and Rule Adaptation .................................................. 1336
  Je Hun Jeon, Minhwa Chung

Pronunciation Variation Modelling Using Accent Features ........................................... 1340
  Michael Tjalve, Mark Huckvale

Automatic Detection of Frequent Pronunciation Errors Made by L2-Learners ............... 1344
  Khiet P. Truong, Ambra Neri, Febe de Wet, Catia Cucchiarini, Helmer Strik

Automatic Transcription of Czech, Russian, and Slovak Spontaneous Speech in the MALACH Project ................................................................. 1348
  Josef Psutka, Pavel Ircing, J.V. Psutka, Jan Hajic, William J. Byrne, Jiri Mirovsky

A Study of Implicit and Explicit Modeling of Coarticulation and Pronunciation Variation ................................................................. 1352
  Stephane Dupont, Christophe Ris, Laurent Couvreur, Jean-Marc Boite

Detection of Coughs from User Utterances Using Imitated Phoneme Model .................. 1356
  Shin-ya Takahashi, Tsuyoshi Morimoto, Sakashi Maeda, Naoyuki Tsuruta

Stochastic Pronunciation Modeling by Ergodic-HMM of Acoustic Sub-Word Units .......... 1360
  V. Ramasubramanian, P. Srinivas, T.V. Sreenivas

An Automated Linguistic Knowledge-Based Cross-Language Transfer Method for Building Acoustic Models for a Language Without Native Training Data .................. 1364
  Chen Liu, Lynette Melnar

Fully Automated Non-Native Speech Recognition Using Confusion-Based Acoustic Model Integration ............................................................ 1368
  Ghazi Bouselmi, Dominique Fohr, Irina Illina, Jean-Paul Haton

The Focus Prosody: More Than a Simple Binary Function .......................................... 1372
  Veronique Auberge, Albert Rilliard

Volume 3

Peak Timing in Two Dialects of Connaught Irish .......................................................... 1376
  Martha Dalton, Ailbhe Ni Chasaide

Compound Rises and “Uptalk” in Spoken English ......................................................... 1380
  Janet Fletcher

Duration and the Temporal Structure of Mandarin Discourse .................................... 1384
  Li-chiung Yang

Prosodic Realization of Split Noun Phrases in Mandarin Chinese Compared in Topic and Focus Contexts ............................................................ 1388
  Bei Wang

Downstep Effect on Disyllabic Words of Citation Forms in Standard Chinese ............... 1392
  Ziyu Xiong
Estimation of Intonation Variation with Constrained Tone Transformations
Jinfu Ni, Hisashi Kawai, Keikichi Hirose

Voice Quality of Falling Tones in Taiwan Min
Ho-hsien Pan

Duration, Intensity and Pause Predictions in Relation to Prosody Organization
Chiu-yu Tseng, Bau-Ling Fu

Pitch Accent Prediction: Effects of Genre and Speaker
Jiahong Yuan, Jason M. Brenier, Daniel Jurafsky

Analysis and Modeling of Fundamental Frequency Contours of Hindi Utterances
Hiroya Fujisaki, Sumio Ohno

Fundamental Frequency and Tone in isiZulu: Initial Experiments
Natasha Govender, Etienne Barnard, Marelie Davel

Intonational Sequences in Tuscan Italian
Judith Bishop, Marc Peake, Dmitry Sityaev

Effects of Raddoppiamento Sintattico on Tonal Alignment in Italian
Caterina Petrone

Acoustic Analysis of Czech Stress: Intonation, Duration and Intensity Revisited
Tomas Dubeda, Jan Votrubec

Variability of F0 Peak Alignment in Moroccan Arabic Accentual Focus
Mohamed Yeou

Phonological Analysis of Schwa and Liaison within the PFC Project (Phonologie du Francais Contemporain): How Determinant are the Prosodic Factors?
A. Lacheret, Ch. Lyche, Michel Morel

Abstractness in Speech-Metronome Synchronisation: P-Centres as Cyclic Attractors
Plinio A. Barbosa, Pablo Arantes, Alexsandro R. Meireles, Jussara M. Vieira

SESSION: APPLICATIONS OF CONFIDENCE RELATED MEASURES TO ASR

Improvement of Rejection Performance of Keyword Spotting Using Anti-Keywrods Derived from Large Vocabulary Considering Acoustical Similarity to Keywords
Makoto Yamada, Tsuneo Kato, Masaki Naito, Hisashi Kawai

Bayes Risk Minimization Using Metric Loss Functions
Ralf Schluter, T. Scharrenbach, V. Steinbiss, Hermann Ney

Word Error Rate Minimization Using an Integrated Confidence Measure
Akio Kobayashi, Kazuo Onoe, Shoei Sato, Toru Imai

Fast Confidence Measure Algorithm for Continuous Speech Recognition
Bin Dong, Qingwei Zhao, Yonghong Yan

Developing and Enhancing Posterior Based Speech Recognition Systems
Hamed Ketabdar, Jithendra Vepa, Samy Bengio, Herve Bourlard

Background Model Based Posterior Probability for Measuring Confidence
Peng Liu, Ye Tian, Jian-Lai Zhou, Frank K. Soong

SESSION: MULTILINGUAL TTS

Foreign Accents in Synthetic Speech: Development and Evaluation
Laura Mayfield Tomokiyo, Alan W. Black, Kevin A. Lenzo
Toward Multiple-Language TTS: Experiments in English and Mandarin ........................................... 1472
Raul Fernandez, Wei Zhang, Ellen Eide, Raimo Bakis, Wael Hamza, Yi Liu, Michael Picheny, John F. Pitrelli, Yong Qing, Zhi Wei Shuang, Li Qin Shen

Cross-Language Synthesis with a Polyglot Synthesizer .................................................................. 1476
Javier Latorre, Koji Iwano, Sadaoki Furui

Development of a Kiswahili Text to Speech System ........................................................................ 1480
Mucemi Gakuru, Frederick K. Iraki, Roger Tucker, Ksenia Shalonova, Kamanda Ngugi

Multilingual Models in the IBM Bilingual Text-to-Speech Systems .............................................. 1484
J. Botella Ordinas, V. Fischer, C. Waast-Richard

Reconstruction of Polish Diacritics in a Text-to-Speech System ..................................................... 1488
Artur Janicki, Piotr Herman

SESSION: SPEECH BANDWIDTH EXTENSION

Design of Bandwidth Scalable LSF Quantization Using Interframe and Intraframe Prediction .............. 1492
Hiroyuki Ehara, Toshiyuki Morii, Masahiro Oshikiri, Koji Yoshida, Kouichi Honma

Artificial Bandwidth Extension of Speech Supported by Watermark-Transmitted Side Information ....... 1496
Bernd Geiser, Peter Jax, Peter Vary

Speech Bandwidth Extension by Improved Codebook Mapping Towards Increased Phonetic Classification ................................................................. 1500
Rongqiang Hu, Venkatesh Krishnan, David V. Anderson

Bandwidth Expansion of Narrowband Speech Using Non-Negative Matrix Factorization .................. 1504
Dhananjay Bansal, Bhiksha Raj, Paris Smaragdis

Robust Bandwidth Extension of Noise-Corrupted Narrowband Speech ........................................... 1508
Michael L. Seltzer, Alex Acero, Jasha Droppo

Pitch-Synchronous Time-Scaling for High-Frequency Excitation Regeneration ................................ 1512
Joao P. Cabrál, Luis C. Oliveira

SESSION: SPOKEN LANGUAGE RESOURCES AND TECHNOLOGY EVALUATION II

A Database of German Emotional Speech ....................................................................................... 1516
Felix Burkhardt, A. Paeschke, M. Roffes, W. Sendlmeier, B. Weiss

Evaluating the Pronunciation of Proper Names by Four French Grapheme-to-Phoneme Converters .......... 1520
Philippe Boula de Mareuil, Christophe d'Alessandro, Gerard Baily, Frederic Bechet, Marie-Neige Garcia, Michel Morel, Romain Prudon, Jean Veronis

A Human-Human Train Timetable Dialogue Corpus ...................................................................... 1524
Filip Jurcicek, Jiri Zahradiil, Libor Jelinek

A Portuguese Spoken and Multi-Modal Dialog Corpora .................................................................. 1528
Gloria Branco, Luis Almeida, Rui Gomes, Nuno Beires

Development of a Cantonese-English Code-Mixing Speech Corpus ............................................... 1532
Joyce Y.C. Chan, P.C. Ching, Tan Lee

BNSI Slovenian Broadcast News database --- Speech and Text Corpus ........................................... 1536
Andrej Zgank, Darinka Verdonik, Aleksandra Zogling Markus, Zdravko Kacic

Confronting HMM-Based Phone Labelling with Human Evaluation of Speech Production .................. 1540
Jan Volin, Radek Skarnitzl, Petr Pollak
Structural Metadata Annotation: Moving Beyond English .......................................................... 154
Stephanie Strassel, Jachym Kolar, Zhiyi Song, Leila Barclay, Meghan Glenn

Neologos: An Optimized Database for the Development of New Speech Processing Algorithms ................................................................................................................. 154
Delphine Charlet, Sacha Krstulovic, Frederic Bimbot, Olivier Boeffard, Dominique Fohr, Odile Mella, Filip Korkmazsky, Djamel Mostefa, Khalid Choukri, Arnaud Vallee

A Hybrid Approach to Automatic Segmentation and Labeling for Mandarin Chinese Speech Corpus .......................................................................................................................... 155
Cheng-Yuan Lin, Kuan-Ting Chen, J.-S. Roger Jang

The Multiple Pronunciations in Taiwanese and the Automatic Transcription of Buddhist Sutra with Augmented Read Speech .................................................................................. 155
Yuang-Chin Chiang, Min-Siong Liang, Hong-Yi Lin, Ren-Yuan Lyu

Bootstrapping Pronunciation Dictionaries: Practical Issues ......................................................... 156
Marelie Davel, Etienne Barnard

Root Causes of Lost Time and User Stress in a Simple Dialog System ........................................ 156
Nigel G. Ward, Anais G. Rivera, Karen Ward, David G. Novick

Evaluating Communication Effectiveness in Team Collaboration ................................................. 156
Julie A. Parisi, Douglas S. Brungart

Bilingual Aligned Corpora for Speech to Speech Translation for Spanish, English and Catalan ......................................................................................................................... 157
David Conejero, Alan Lounds, Carmen Garcia Mateo, Leandra Rodriguez-Linares, Raquel Mochales, Asuncion Moreno

Design and Collection of Czech Lombard Speech Database .......................................................... 158
Hynek Boril, Petr Pollak

TBALL Data Collection: The Making of a Young Children's Speech Corpus ................................ 158
Abe Kazemzadeh, Hong You, Markus Iseli, Barbara Jones, Xiaodong Cui, Margaret Heritage, Patti Price, Elaine Anderson, Shrikanth Narayanan, Abeer Alwan

Construction and Utilization of Bilingual Speech Corpus for Simultaneous Machine Interpretation Research .................................................................................................................. 159
Hitomi Tohyama, Shigeki Matsubara, Nobuo Kawaguchi, Yasuyoshi Inagaki

Meeting Acts: A Labeling System for Group Interaction in Meetings ........................................ 159
Rebecca Bates, Patrick Menning, Elizabeth Willingham, Chad Kuyper

A New Evaluation Criteria for Keyword Spotting Techniques and a New Algorithm .................... 160
Marius C. Silaghi, Rachna Vargiya

PhattSessionz: Recording 1000 Adolescent Speakers in Schools in Germany ............................... 160
Chr. Draxler, Alexander Steffen

An Amharic Speech Corpus for Large Vocabulary Continuous Speech Recognition .................... 161
Solomon Teferra Abate, Wolfgang Menzel, Bairu Tafila

The FASIL Speech and Multimodal Corpora ............................................................................... 161
Hans Dolfin, David Reitter, Luis Almeida, Nuno Beires, Michael Cody, Rui Gomes, Kerry Robinson, Roman Zielinski

Revealing Phonological Similarities Between German and Dutch .................................................. 162
Karin Muller

SESSION: LARGE VOCABULARY SPEECH RECOGNITION SYSTEMS

Development of a Conversational Telephone Speech Recognizer for Levantine Arabic .................. 162
Dimitra Vergyri, Katrin Kirchhoff, R. Gadde, Andreas Stolcke, Jing Zheng
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploiting Large Quantities of Spontaneous Speech for Unsupervised Training of Acoustic Models</td>
<td>1616</td>
</tr>
<tr>
<td>Bhuvana Ramabhadran</td>
<td></td>
</tr>
<tr>
<td>Improved Spontaneous Mandarin Speech Recognition by Disfluency Interruption Point (IP) Detection Using Prosodic Features</td>
<td>1620</td>
</tr>
<tr>
<td>Che-Kuang Lin, Lin-Shan Lee</td>
<td></td>
</tr>
<tr>
<td>Improvements to the BBN RT04 Mandarin Conversational Telephone Speech Recognition System</td>
<td>1624</td>
</tr>
<tr>
<td>Jeff Z. Ma, Spyros Matsoukas</td>
<td></td>
</tr>
<tr>
<td>Incorporating a Bayesian Wide Phonetic Context Model for Acoustic Rescoring</td>
<td>1628</td>
</tr>
<tr>
<td>Sakriani Sakti, Satoshi Nakamura, Konstantin Markov</td>
<td></td>
</tr>
<tr>
<td>Modeling Vowels for Arabic BN Transcription</td>
<td>1632</td>
</tr>
<tr>
<td>Abdel Messaoudi, Lori Lamel, Jean-Luc Gauvain</td>
<td></td>
</tr>
<tr>
<td>Recent Progress in Arabic Broadcast News Transcription at BBN</td>
<td>1636</td>
</tr>
<tr>
<td>Mohamed Afify, Long Nguyen, Bing Xiang, Sherif Abdou, John Makhoul</td>
<td></td>
</tr>
<tr>
<td>The 2004 BBN 1xRT Recognition Systems for English Broadcast News and Conversational Telephone Speech</td>
<td>1640</td>
</tr>
<tr>
<td>Spyros Matsoukas, Rohit Prasad, Srinivas Laxminarayan, Bing Xiang, Long Nguyen, Richard Schwartz</td>
<td></td>
</tr>
<tr>
<td>The 2004 BBN/LIMSI 20xRT English Conversational Telephone Speech Recognition System</td>
<td>1644</td>
</tr>
<tr>
<td>The BBN Mandarin Broadcast News Transcription System</td>
<td>1648</td>
</tr>
<tr>
<td>Bing Xiang, Long Nguyen, Xuefeng Guo, Dongxin Xu</td>
<td></td>
</tr>
<tr>
<td>The LIUM Speech Transcription System: A CMU Sphinx III-Based System for French Broadcast News</td>
<td>1652</td>
</tr>
<tr>
<td>Paul Deleglise, Yannick Esteve, Sylvain Meignier, Teva Merlin</td>
<td></td>
</tr>
<tr>
<td>Transcribing Lectures and Seminars</td>
<td>1656</td>
</tr>
<tr>
<td>Lori Lamel, G. Adda, E. Bilinski, Jean-Luc Gauvain</td>
<td></td>
</tr>
<tr>
<td>Transcription of Conference Room Meetings: An Investigation</td>
<td>1660</td>
</tr>
<tr>
<td>Thomas Hain, John Dines, Giulia Garau, Martin Karafiat, Darren Moore, Vincent Wan, Roeland Ordelman, Steve Renals</td>
<td></td>
</tr>
<tr>
<td>Where Are We in Transcribing French Broadcast News?</td>
<td>1664</td>
</tr>
<tr>
<td>Jean-Luc Gauvain, G. Adda, Martine Adda-Decker, Alexandre Allauzen, V. Gendner, Lori Lamel, Holger Schwenk</td>
<td></td>
</tr>
<tr>
<td>Two-Pass Strategy for Handling OOVs in a Large Vocabulary Recognition Task</td>
<td>1668</td>
</tr>
<tr>
<td>Odette Scharenborg, Stephanie Seneff</td>
<td></td>
</tr>
<tr>
<td>The BBN RT04 English Broadcast News Transcription System</td>
<td>1672</td>
</tr>
<tr>
<td>Long Nguyen, Bing Xiang, Mohamed Afify, Sherif Abdou, Spyros Matsoukas, Richard Schwartz, John Makhoul</td>
<td></td>
</tr>
<tr>
<td>Investigations on Ensemble Based Semi-Supervised Acoustic Model Training</td>
<td>1676</td>
</tr>
<tr>
<td>Rong Zhang, Ziad Al Bawab, Arthur Chan, Ananlada Chotimongkol, David Huggins-Daines, Alexander I. Rudnicky</td>
<td></td>
</tr>
<tr>
<td>Fully Automated System for Czech Spoken Broadcast Transcription with Very Large (300K+) Lexicon</td>
<td>1680</td>
</tr>
<tr>
<td>Jan Nouza, Jindrich Zdanský, Petr David, Petr Cerva, Jan Kolorenc, Dana Nejedlova</td>
<td></td>
</tr>
<tr>
<td>Experiments with Probabilistic Principal Component Analysis in LVCSR</td>
<td>1684</td>
</tr>
<tr>
<td>Mike Schuster, Takaaki Hori, Atsushi Nakamura</td>
<td></td>
</tr>
<tr>
<td>Vietnamese Large Vocabulary Continuous Speech Recognition</td>
<td>1688</td>
</tr>
<tr>
<td>Thang Tat Vu, Dung Tien Nguyen, Mai Chi Luong, John-Paul Hosom</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Data Sampling for Improved Speech Recognizer Training</td>
<td>1692</td>
</tr>
<tr>
<td>Takahiro Shinozaki, Mari Ostendorf, Les Atlas</td>
<td></td>
</tr>
<tr>
<td>Influence of F0 on Vietnamese Syllable Perception</td>
<td>1696</td>
</tr>
<tr>
<td>Do Dat Tran, Eric Castelli, Jean-Francois Serignat, Van Loan Trinh, Xuan Hung Le</td>
<td></td>
</tr>
<tr>
<td>Lexical Tone and Pitch Perception in Tone and Non-Tone Language Speakers</td>
<td>1700</td>
</tr>
<tr>
<td>Barbara Schwanhausser, Denis Burnham</td>
<td></td>
</tr>
<tr>
<td>Intonational Contrasts in EP: A Categorical Perception Approach</td>
<td>1704</td>
</tr>
<tr>
<td>Isabel Fale, Isabel Hub Faria</td>
<td></td>
</tr>
<tr>
<td>Does Narrow Focus Activate Alternative Referents?</td>
<td>1708</td>
</tr>
<tr>
<td>Bettina Braun, Andrea Weber, Matthew Crocker</td>
<td></td>
</tr>
<tr>
<td>Audiovisual Interaction on the Perception of Frequency Glide of Linear Sweep Tones</td>
<td>1712</td>
</tr>
<tr>
<td>Kiyoaki Aikawa, Hayato Hashimoto</td>
<td></td>
</tr>
<tr>
<td>Audiovisual Integration in Dichotic Listening</td>
<td>1716</td>
</tr>
<tr>
<td>Kei Omata, Ken Mogi</td>
<td></td>
</tr>
<tr>
<td>Perception Experiment Combining a Parametric Loudspeaker and a Synthetic Talking Head</td>
<td>1720</td>
</tr>
<tr>
<td>Gunilla Svanfeldt, Dirk Olszewski</td>
<td></td>
</tr>
<tr>
<td>Multidimensional Scaling of Listener Responses to Synthetic Speech</td>
<td>1724</td>
</tr>
<tr>
<td>Catherine Mayo, Robert A.J. Clark, Simon King</td>
<td></td>
</tr>
<tr>
<td>A Timbre Space for Speech</td>
<td>1728</td>
</tr>
<tr>
<td>Hiroko Terasawa, Malcolm Slaney, Jonathan Berger</td>
<td></td>
</tr>
<tr>
<td>Voice Quality Assessment by Means of Comparative Judgments of Speech Tokens</td>
<td>1732</td>
</tr>
<tr>
<td>A. Kacha, Francis Grenez, Jean Schoentgen</td>
<td></td>
</tr>
<tr>
<td>Speech Intelligibility Derived from Time-Frequency and Source Smearing</td>
<td>1736</td>
</tr>
<tr>
<td>Toshio Irino, Satoru Satou, Shunsuke Nomura, Hideki Banno, Hideki Kawahara</td>
<td></td>
</tr>
<tr>
<td>Steady-State Pre-Processing for Improving Speech Intelligibility in Reverberant Environments: Evaluation in a Hall with an Electrical Reverberator</td>
<td>1740</td>
</tr>
<tr>
<td>Nahoko Hayashi, Takayuki Arai, Nao Hodashima, Yusuke Miyauchi, Kiyohiro Kurisu</td>
<td></td>
</tr>
<tr>
<td>Neural Bases of Listening to Speech in Noise</td>
<td>1744</td>
</tr>
<tr>
<td>Patrick C.M. Wong, Kiara M. Lee, Todd B. Parrish</td>
<td></td>
</tr>
<tr>
<td>The Intelligibility of Tracheoesophageal Speech: First Results</td>
<td>1748</td>
</tr>
<tr>
<td>P. Jongmans, F.J.M. Hilgers, Louis C.W. Pols, C.J. van As-Brooks</td>
<td></td>
</tr>
<tr>
<td>A Computational Model of the Speech Reception Threshold for Laterally Separated Speech and Noise</td>
<td>1752</td>
</tr>
<tr>
<td>Guy J. Brown, Kalle J. Palomaki</td>
<td></td>
</tr>
<tr>
<td>Lexical Inhibition Effects in Time-Compressed Speech</td>
<td>1756</td>
</tr>
<tr>
<td>Esther Janse</td>
<td></td>
</tr>
<tr>
<td>Perception of Time-Compressed Rapid Acoustic Cues in French CV Syllables</td>
<td>1760</td>
</tr>
<tr>
<td>Caroline Jacquier, Fanny Meunier</td>
<td></td>
</tr>
<tr>
<td>Reversed Speech Comprehension Depends on the Auditory Efferent System Functionality</td>
<td>1764</td>
</tr>
<tr>
<td>C. Grataloup, M. Hoen, F. Pellegrino, E. Veuillet, L. Collet, Fanny Meunier</td>
<td></td>
</tr>
<tr>
<td>Perceptual Space of English Fricatives for Japanese Learners</td>
<td>1768</td>
</tr>
<tr>
<td>Won Tokuma, Shinichi Tokuma</td>
<td></td>
</tr>
<tr>
<td>Perceptual Salience of Language-Specific Acoustic Differences in Autonomous Fillers Across Eight Languages</td>
<td>1772</td>
</tr>
<tr>
<td>Ioana Vasilescu, Maria Candea, Martine Adda-Decker</td>
<td></td>
</tr>
</tbody>
</table>
Effects of Cortical and Subcortical Brain Damage on the Processing of Emotional Prosody
Marc D. Pell

Spontaneous Speech: How People Really Talk and Why Engineers Should Care
Elizabeth Shriberg

SESSION: SPEECH RECOGNITION – ADAPTATION II

Feature Adaptation Using Projection of Gaussian Posteriors
Karthik Visweswariah, Peder Olsen

Maximum Margin Learning and Adaptation of MLP Classifiers
Xiao Li, Jeff Bilmes, Jonathan Malkin

Leveraging Speaker-Dependent Variation of Adaptation
Arindam Mandal, Mari Ostendorf, Andreas Stolcke

A Comparative Study of Two Kernel Eigenspace-Based Speaker Adaptation Methods on Large Vocabulary Continuous Speech Recognition
Roger Hsiao, Brian Mak

Environmental Compensation Using ASR Model Adaptation by a Bayesian Parametric Representation Method
Xuechuan Wang, Douglas O'Shaughnessy

Discriminative Speaker Adaptation with Eigenvoices
Jun Luo, Zhijian Ou, Zuoying Wang

Context in Multi-Lingual Tone and Pitch Accent Recognition
Gina-Anne Levow

Automatic Prominence Identification and Prosodic Typology
Fabio Tamburini

Influence of Syntax on Prosodic Boundary Prediction
Tommy Ingulfsen, Tina Burrows, Sabine Buchholz

Using Prosodic Information for Disambiguation Purposes
Roberto Gretter, Dino Seppi

Analysis of the Effects of Word Emphasis and Echo Question on F_0 Contours of Cantonese Utterances
Wentao Gu, Keikichi Hirose, Hiroya Fujisaki

Combining Models of Prosodic Phrasing and Pausing
Tina Burrows, Peter Jackson, Katherine Knill, Dmitry Sityaev

Distinguishing Deceptive from Non-Deceptive Speech
Julia Hirschberg, Stefan Benus, Jason M. Brenier, Frank Enos, Sarah Friedman, Sarah Gilman, Cynthia Girand, Martin Graciarena, Andreas Kathol, Laura Michaelis, Bryan L. Pellom, Elizabeth Shriberg, Andreas Stolcke

Detecting Certainty in Spoken Tutorial Dialogues
Jackson Liscombe, Julia Hirschberg, Jennifer J. Venditti

Detection of Real-Life Emotions in Call Centers
Laurence Vidrascu, Laurence Devillers

Using Context to Improve Emotion Detection in Spoken Dialog Systems
Jackson Liscombe, Giuseppe Riccardi, Dilek Hakkani-Tur

Voice Quality and f_0 Cues for Affect Expression: Implications for Synthesis
Irena Yanushevskaya, Christer Gobl, Ailbhe Ni Chasaide

Voice and Emotional Expression Transformation Based on Statistics of Vowel Parameters in an Emotional Speech Database
Toru Takahashi, Takeshi Fujii, Masashi Nishi, Hideki Banno, Toshio Irino, Hideki Kawahara
SESSION: RAPID DEVELOPMENT OF SPOKEN DIALOGUE SYSTEMS

Automated Wizard-of-Oz for Spoken Dialogue Systems .......................................................... 1856
  Giuseppe Di Fabbrizio, Gokhan Tur, Dilek Hakkani-Tur

A Rapid Prototyping Tool for Constructing Web-Based MMI Applications ...................... 1860
  Kouichi Katsurada, Kunitoshi Sato, Hiroaki Adachi, Hirobumi Yamada, Tsuneo Nitta

Developing Extensible and Reusable Spoken Dialogue Components: An Examination of the Queen's Communicator .......................................................... 1864
  Philip Hanna, Ian O'Neill, Xingkun Liu, Michael McTear

SGStudio: Rapid Semantic Grammar Development for Spoken Language Understanding ............................................................................................................................... 1868
  Ye-Yi Wang, Alex Acero

Rapid Transition to New Spoken Dialogue Domains: Language Model Training Using Knowledge from Previous Domain Applications and Web Text Resources 1872
  Murat Akbacak, Yuqing Gao, Liang Gu, Hong-Kwang Jeff Kuo

A Methodology for Comparing Grammar-Based and Robust Approaches to Speech Understanding ................................................................................................................. 1876
  Manny Rayner, Pierrette Bouillon, Nikos Chatzichristofis, Beth Ann Hockey, Marianne Santaholma, Marianne Starlander, Hitoshi Isahara, Kyoko Kanzaki, Yukie Nakao

SESSION: TEXT-TO-SPEECH I

Learning to Personalize Spoken Generation for Dialogue Systems .................................... 1880
  Francois Mairesse, Marilyn Walker

Optimization of Text-to-Speech Phonetic Transcriptions Using a posteriori Signal Comparison .......................................................... 1884
  S. Revelin, D. Cadic, C. Waast-Richard

Voice Transformation Using Principle Component Analysis Based LSF Quantization and Dynamic Programming Approach .......................................................... 1888
  Ozgul Salor, Mubeccel Demirekler

Adapt Mandarin TTS System to Chinese Dialect TTS Systems ........................................... 1892
  Hai Ping Li, Wei Zhang

Grapheme-to-Phoneme Conversion Based on TBL Algorithm in Mandarin TTS System .............................................................................................................................. 1896
  Min Zheng, Qin Shi, Wei Zhang, Lianhong Cai

An Automaton-Based Machine Learning Technique for Automatic Phonetic Transcription ................................................................................................................................. 1900
  Paolo Massimino, Alberto Pacchiotti

Comparative Objective and Subjective Evaluation of Three Data-Driven Techniques for Proper Name Pronunciation .......................................................... 1904
  Tanawan Soonklang, Robert I. Damper, Yannick Marchand

Articulatory Synthesis Using Corpus-Based Estimation of Line Spectrum Pairs ............... 1908
  Olov Engwall

Effects of Pitch Accent Type on Interpreting Information Status in Synthetic Speech .............................................................................................................................. 1912
  Aoju Chen, Els den Os

Towards Generic Spatial Object Model and Route Guidance Grammar for Speech-Based Systems .............................................................................................................................. 1916
  Perttu Prusi, Anssi Kainulainen, Jaakko Hakulinen, Markku Turunen, Esa-Pekka Salonen, Leena Helin
Duration-Embedded Bi-HMM for Expressive Voice Conversion
Chi-Chun Hsia, Chung-Hsien Wu, Te-Hsien Liu

Analysis of Major Factors of Naturalness Degradation in Concatenative Synthesis
Toshio Hirai, Hisashi Kawai, Minoru Tsuzaki, Nobuyuki Nishizawa

Duration Modeling and Memory Optimization in a Mandarin TTS System
Jilei Tian, Jani Nurminen, Imre Kiss

A Bi-Lingual Mandarin-to-Taiwanese Text-to-Speech System
Min-Siong Liang, Ke-Chun Chuang, Rhuei-Cheng Yang, Yuang-Chin Chiang, Ren-Yuan Lyu

Using Morphology and Phoneme History to Improve Grapheme-to-Phoneme Conversion
Uwe D. Reichel, Florian Schiel

Predicting Consonant Duration with Bayesian Belief Networks
Olga Goubanova, Simon King

Inducing Decision Tree Pronunciation Variation Models from Annotated Speech Data
Per-Anders Jande

Phonetic Transcription Verification with Generalized Posterior Probability
Lijuan Wang, Yong Zhao, Min Chu, Frank K. Soong, Zhigang Cao

Training a Maximum Entropy Model for Surface Realization
Hua Cheng, Fuliang Weng, Niti Hantawee, Lawrence Cavedon, Stanley Peters

NAM-to-Speech Conversion with Gaussian Mixture Models
Tomoki Toda, Kiyohiro Shikano

Which Italian do Current Systems Speak? A First Step Towards Pronunciation Modelling of Italian Varieties
Michelina Savino, Mario Refice, Massimo Mitaritonna

Modelling Pitch Accent Types for Polish Speech Synthesis
Dominika Oliver, Robert A.J. Clark

Learning Methods and Features for Corpus-Based Phrase Break Prediction on Thai
Chatchawarn Hansakunbuntheung, Ausdang Thangthai, Chai Wutiwiwatchai, Rungkarn Sincharoenschai

Hidden Markov Models for Grapheme to Phoneme Conversion
Paul Taylor

Robust Distant Speaker Recognition Based on Position Dependent Cepstral Mean Normalization
Longbiao Wang, Norhide Kitaoka, Seiichi Nakagawa

Speaker Adaptation in the NIST Speaker Recognition Evaluation 2004
David A. van Leeuwen

A Distance Measure Between GMMs Based on the Unscented Transform and its Application to Speaker Recognition
Jacob Goldberger, Hagai Aronowitz

Estimation of Speaker's Height and Vocal Tract Length from Speech Signal
Sorin Dusan

On the Relationship Between Phonetic Modeling Precision and Phonetic Speaker Recognition Accuracy
Doroteo Torre Toledano, Carlos Fombella, Joaquin Gonzalez Rodriguez, Luis Hernandez Gomez

Open-Set Speaker Identification Using Adapted Gaussian Mixture Models
J. Fortuna, P. Sivakumaran, A. Anyaeinia, A. Malegaonkar
Speaker Verification in Noisy Conditions Using Correlated Subband Features
James McAuley, Ji Ming, Pat Corr

Probabilistic Anchor Models Approach for Speaker Verification
Mikaël Collet, Yassine Mami, Delphine Charlet, Frédéric Bimbot

A Bayesian Network Approach Combining Pitch and Spectral Envelope Features to Reduce Channel Mismatch in Speaker Verification and Forensic Speaker Recognition
Mijail Arciengá, Aníl Alexander, Philipp Zimmermann, Andrzej Drygajlo

Channel Robust Speaker Verification via Bayesian Blind Stochastic Feature Transformation
Kwok-Kwong Yiu, Man-Wai Mak, Sun-Yuan Kung

dPLRM-Based Speaker Identification with Log Power Spectrum
Tomoko Matsui, Kunio Tanabe

Speaker Verification Using Gaussian Mixture Models Within Changing Real Car Environments
Xianxian Zhang, John H.L. Hansen, Pongtep Angkititrakul, Kazuya Takeda

The Correspondences Between the Perception of the Speaker Individualities Contained in Speech Sounds and Their Acoustic Properties
Kanae Amino, Tsutomu Sugawara, Takayuki Arai

A Noise-Robust Pitch Synchronous Feature Extraction Algorithm for Speaker Recognition Systems
Samuel Kim, Sungwan Yoon, Thomas Eriksson, Hong-Goo Kang, Dae Hee Youn

Modeling High-Level Information by Using Gaussian Mixture Correlation for GMM-UBM Based Speaker Recognition
Jing Deng, Thomas Fang Zheng, Zhanjiang Song, Jian Liu

In-Set/Out-of-Set Speaker Identification Based on Discriminative Speech Frame Selection
Xianxian Zhang, John H.L. Hansen

Mixture of Support Vector Machines for Text-Independent Speaker Recognition
Zhenchun Lei, Yingchun Yang, Zhaohui Wu

Optimal Model Order Selection Based on Regression Tree in Speaker Identification
Shilei Zhang, Junmei Bai, Shuwu Zhang, Bo Xu

Speaker Verification Improvement Using Blind Inversion of Distortions
Marcos Faundez-Zanuy, Jordi Sole-Casals

Supergaussian GARCH Models for Speech Signals
Israel Cohen

A Spectral Conversion Approach to Feature Denoising and Speech Enhancement
A. Mouchtaris, J. Van der Spiegel, P. Mueller, P. Tsakalides

Acoustic Feedback Cancellation in Speech Reinforcement Systems for Vehicles
Alfonso Ortega, Eduardo Lleida, Enrique Masgrau, Luis Buera, Antonio Miguel

Implicit Control of Noise Canceller for Speech Enhancement
Julien Bourgeois, Jurgen Freudenberg, Guillaume Lathoud

Speech Enhancement Using Markov Model of Speech Segments
T.M. Sunil Kumar, T.V. Sreenivas

Volume 4

A Wavelet Based Noise Reduction Algorithm for Speech Signal Corrupted by Coloured Noise
Vladimir Braquet, Takao Kobayashi
Speech Enhancement in Temporal DFT Trajectories Using Kalman Filters
Esfandiar Zavarehei, Saeed Vaseghi

Formant-Tracking Linear Prediction Models for Speech Processing in Noisy Environments
Qin Yan, Saeed Vaseghi, Esfandiar Zavarehei, Ben Milner

Statistical Noise Compensation for Cochlear Implant Processing
Hui Jiang, Qian-Jie Fu

WPD-Based Noise Suppression Using Nonlinearly Weighted Threshold Quantile Estimation and Optimal Wavelet Shrinking
Tuan Van Pham, Gernot Kubin

Subjective and Objective Quality Assessment of Regression-Enhanced Speech in Real Car Environments
Weifeng Li, Katunobu Itou, Kazuya Takeda, Fumitada Itakura

A Model for Selective Segregation of a Target Instrument Sound from the Mixed Sound of Various Instruments
Masashi Unoki, Masaaki Kubo, Atsushi Haniu, Masato Akagi

Improved Decision Directed Approach for Speech Enhancement Using an Adaptive Time Segmentation
Richard C. Hendriks, Richard Heusdens, Jesper Jensen

Generalized Filter-Bank Equalizer for Noise Reduction with Reduced Signal Delay
Heinrich W. Lollmann, Peter Vary

A Pitch-Based Model for Separation of Reverberant Speech
Nicoleta Roman, DeLiang Wang

On Noise Gain Estimation for HMM-Based Speech Enhancement
David Y. Zhao, W. Bastiaan Kleijn

Speech Enhancement Using Auditory Phase Opponency Model
Om Deshmukh, Carol Espy-Wilson

High-Density Discrete HMM with the Use of Scalar Quantization Indexing
Brian Mak, Siu-Kei Au Yeung, Yiu-Pong Lai, Manhung Siu

Improved Discriminative Training Using Phone Lattices
Jing Zheng, Andreas Stolcke

Improved MLP Structures for Data-Driven Feature Extraction for ASR
Qifeng Zhu, Barry Y. Chen, Frantisek Grezl, Nelson Morgan

Investigations on Error Minimizing Training Criteria for Discriminative Training in Automatic Speech Recognition
Wolfgang Macherey, Lars Haferkamp, Ralf Schluter, Hermann Ney

Temporally Varying Model Parameters for Large Vocabulary Continuous Speech Recognition
K.C. Sim, M.J.F. Gales

Using MLP Features in SRI's Conversational Speech Recognition System
Qifeng Zhu, Andreas Stolcke, Barry Y. Chen, Nelson Morgan

SESSION: SPEECH PRODUCTION II

A Toolkit for Voice Inverse Filtering and Parametrisation
Matti Airas, Hannu Pulakka, Tom Backstrom, Paavo Alku

Stylization of Glottal-Flow Spectra Produced by a Mechanical Vocal-Fold Model
Denisse Sciamarella, Christophe d'Alessandro
Numerical Glottal Sound Source Model as Coupled Problem Between Vocal Cord Vibration and Glottal Flow ................................................................. 2152
Hideyuki Nomura, Tetsuo Funada

A Tagged-Cine MRI Investigation of German Vowels ........................................... 2156
Marianne Pouplier, Maureen Stone

A Three-Dimensional Linear Articulatory Model of Velum Based on MRI Data .................. 2160
Antoine Serrurier, Pierre Badin

On the Relationship Between Intra-Oral Pressure and Speech Sonority .................... 2164
Anne Cros, Didier Demolin, Ana Georgina Flesia, Antonio Galves

SESSION: SPEAKER CHARACTERIZATION AND RECOGNITION II

Maximum Conditional Mutual Information Modeling for Speaker Verification.................. 2168
Mohamed Kamal Omar, Jiri Navratil, Ganesh N. Ramaswamy

Class-Dependent Score Combination for Speaker Recognition .................................... 2172
Luciana Ferrer, Kemal Sonmez, Sachin Kajarekar

Modeling Intra-Speaker Variability for Speaker Recognition ....................................... 2176
Hagai Aronowitz, Dror Irony, David Burshtein

Liveness Detection Using Cross-Modal Correlations in Face-Voice Person Authentication ................................................................. 2180
Girija Chetty, Michael Wagner

Stream-Weight Optimization by LDA and Adaboost for Multi-Stream Speaker Verification ........................................................................ 2184
Taichi Asami, Koji Iwano, Sadaoki Furui

Considering Speech Quality in Speaker Verification Fusion ....................................... 2188
Yosef A. Solewicz, Moshe Koppel

SESSION: GENDER AND AGE ISSUES IN SPEECH AND LANGUAGE RESEARCH I

Speaker Adaptive Acoustic Modeling with Mixture of Adult and Children’s Speech ........ 2192
Matteo Gerosa, Diego Giuliani, Fabio Brugnara

A Comparison of Human and Computer Recognition Accuracy for Children’s Speech ................................................................. 2196
Shona D’Arcy, Martin Russell

Italian Children’s Speech Recognition for Advanced Interactive Literacy Tutors ............ 2200
Piero Cosi, Bryan L. Pellom

Do Speech Recognizers Prefer Female Speakers? ........................................................... 2204
Martine Adda-Decker, Lori Lamel

Detecting Politeness and Frustration State of a Child in a Conversational Computer Game ........................................................................ 2208
Serdar Yildirim, Chul Min Lee, Sungbok Lee, Alexandros Potamianos, Shrikanth Narayanan

Gender in Everyday Speech and Language: A Corpus-Based Study ............................... 2212
Diana Binnenpoorte, Christophe Van Bael, Els den Os, Louis Boves

SESSION: SPOKEN LANGUAGE ACQUISITION, DEVELOPMENT AND LEARNING II

Developmental Change of Phoneme Duration in a Japanese Infant and Mother ............... 2216
Shigeaki Amano
Mora Timing Organization in Producing Contrastive Geminate/Single Consonants and Long/Short Vowels by Native and Non-Native Speakers of Japanese: Effects of Speaking Rate

Haiping Jia, Hiroki Mori, Hideki Kasuya

Mutual Intelligibility of American, Chinese and Dutch-Accented Speakers of English

Hongyan Wang, Vincent J. van Heuven

Deriving a Bi-Lingual Dictionary from Raw Transcription Data

Peter Juel Henrichsen

A Statistical Method of Evaluating Pronunciation Proficiency for Japanese Words

Kei Ohta, Seiichi Nakagawa

SESSION: LANGUAGE AND DIALECT IDENTIFICATION I

Phonotactic Language Identification Using High Quality Phoneme Recognition

Pavel Matejka, Petr Schwarz, Jan Cernocky, Pavel Chylil

Advances in Word Based Dialect/Accent Classification

Rongqing Huang, John H.L. Hansen

Syllable Structure in Spoken Arabic: A Comparative Investigation

Rym Hamdi, Salem Ghazali, Melissa Barkat-Defradas

A Transformation-Based Learning Approach to Language Identification for Mixed-Lingual Text-to-Speech Synthesis

J.C. Marcadet, V. Fischer, C. Waast-Richard

Constructing Family Trees of Multilingual Speech Using Gaussian Mixture Models

Shuichi Itahashi, Shiwai Zhu, Mikio Yamamoto

Modeling Long and Short-Term Prosody for Language Identification

Jean-Luc Rouas

SESSION: SPOKEN LANGUAGE TRANSLATION I

Document Driven Machine Translation Enhanced ASR

M. Paulik, Christian Fugen, Sebastian Stuker, Tanja Schultz, Thomas Schaal, Alex Waibel

Automatic Text Dictation in Computer-Assisted Translation

Shahram Khadivi, Andras Zolnay, Hermann Ney

On the Use of Speech Recognition in Computer Assisted Translation

L. Rodriguez, J. Civera, E. Vidal, Francisco Casacuberta, C. Martinez

Speech Translation for Low-Resource Languages: The Case of Pashto

Andreas Kathol, Kristin Precoda, Dimitra Vergyri, Wen Wang, Susanne Riehemann

Finite-State Transducer Inference for a Speech-Input Portuguese-to-English Machine Translation System

David Pico, Jorge Gonzalez, Francisco Casacuberta, Diamantino Caseiro, Isabel Trancoso

Quantitative Evaluation of Effects of Speech Recognition Errors on Speech Translation Quality

Kenko Ohta, Keiji Yasuda, Genichiro Kikui, Masuzo Yanagida

SESSION: MULTI-CHANNEL SPEECH ENHANCEMENT

A Stereo Input-Output Superdirective Beamformer for Dual Channel Noise Reduction

Thomas Lotter, Bastian Sauert, Peter Vary
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalman Filters for Time Delay of Arrival-Based Source Localization</td>
<td>2288</td>
</tr>
<tr>
<td>Ulrich Klee, Tobias Gehrig, John McDonough</td>
<td></td>
</tr>
<tr>
<td>Simultaneous Adaptation of Echo Cancellation and Spectral Subtraction for In-Car Speech Recognition</td>
<td>2292</td>
</tr>
<tr>
<td>Osamu Ichikawa, Masafumi Nishimura</td>
<td></td>
</tr>
<tr>
<td>Variable Step Size Adaptive Decorrelation Filtering for Competing Speech Separation</td>
<td>2296</td>
</tr>
<tr>
<td>Rong Hu, Yunxin Zhao</td>
<td></td>
</tr>
<tr>
<td>Speech Extraction in a Car Interior Using Frequency-Domain ICA with Rapid Filter Adaptations</td>
<td>2300</td>
</tr>
<tr>
<td>Daisuke Saitoh, Atsunobu Kaminuma, Hiroshi Saruwatari, Tsuyoki Nishikawa, Akinobu Lee</td>
<td></td>
</tr>
<tr>
<td>Speech Enhancement Using Non-Acoustic Sensors</td>
<td>2304</td>
</tr>
<tr>
<td>Rongqiang Hu, Sunil D. Kamath, David V. Anderson</td>
<td></td>
</tr>
<tr>
<td>Improved Blind Dereverberation Performance by Using Spatial Information</td>
<td>2308</td>
</tr>
<tr>
<td>Marc Delcroix, Takafumi Hikichi, Masato Miyoshi</td>
<td></td>
</tr>
<tr>
<td>A Hybrid Microphone Array Post-Filter in a Diffuse Noise Field</td>
<td>2312</td>
</tr>
<tr>
<td>Junfeng Li, Masato Akagi</td>
<td></td>
</tr>
<tr>
<td>A Framework for Estimation of Clean Speech by Fusion of Outputs from Multiple Speech Enhancement Systems</td>
<td>2316</td>
</tr>
<tr>
<td>Venkatesh Krishnan, Phil S. Whitehead, David V. Anderson, Mark A. Clements</td>
<td></td>
</tr>
<tr>
<td>A Study of Weighted CSP Analysis with Average Speech Spectrum for Noise Robust Talker Localization</td>
<td>2320</td>
</tr>
<tr>
<td>Yuki Denda, Takanobu Nishiura, Yoichi Yamashita</td>
<td></td>
</tr>
<tr>
<td>Sound Segregation Based on Binaural Zero-Crossings</td>
<td>2324</td>
</tr>
<tr>
<td>Young-Ik Kim, Sung Jun An, Rhee Man Kil, Hyung-Min Park</td>
<td></td>
</tr>
<tr>
<td>A Two-Microphone Diversity System and its Application for Hands-Free Car Kits</td>
<td>2328</td>
</tr>
<tr>
<td>Jurgen Freudenberger, Klaus Linhard</td>
<td></td>
</tr>
<tr>
<td>Directionally Constrained Minimization of Power Algorithm for Speech Signals</td>
<td>2332</td>
</tr>
<tr>
<td>Takahiro Murakami, Kiyoshi Kurihara, Yoshihisa Ishida</td>
<td></td>
</tr>
<tr>
<td>Oriented Global Coherence Field for the Estimation of the Head Orientation in Smart Rooms Equipped with Distributed Microphone Arrays</td>
<td>2336</td>
</tr>
<tr>
<td>Alessio Brulti, Maurizio Omologo, Piergiorgio Svaizer</td>
<td></td>
</tr>
<tr>
<td>Robust Speaker Localization Through Adaptive Weighted Pair TDOA (AWEPAT) Estimation</td>
<td>2340</td>
</tr>
<tr>
<td>Nilesh Madhu, Rainer Martin</td>
<td></td>
</tr>
<tr>
<td>A Spectrogram Model for Enhanced Source Localization and Noise-Robust ASR</td>
<td>2344</td>
</tr>
<tr>
<td>Guillaume Lathoud, Mathew Magimai.-Doss, Bertrand Mesot</td>
<td></td>
</tr>
<tr>
<td>Denoising Through Source Separation and Minimum Tracking</td>
<td>2348</td>
</tr>
<tr>
<td>Srinam Srinivasan, Mattias Nilsson, W. Bastiaan Kleijn</td>
<td></td>
</tr>
<tr>
<td>Collaborative Voice Activity Detection for Hearing Aids</td>
<td>2352</td>
</tr>
<tr>
<td>Louisa Busca Grisoni, John H.L. Hansen</td>
<td></td>
</tr>
<tr>
<td>Using Inter-Frequency Decorrelation to Reduce the Permutation Inconsistency Problem in Blind Source Separation</td>
<td>2356</td>
</tr>
<tr>
<td>Enrique Robledo-Amuncio, Biing-Hwang Juang</td>
<td></td>
</tr>
<tr>
<td>A Graphical Model for Multi-Sensory Speech Processing in Air-and-Bone Conductive Microphones</td>
<td>2360</td>
</tr>
<tr>
<td>Amarnag Subramanya, Zhengyou Zhang, Zicheng Liu, Jasha Droppo, Alex Acero</td>
<td></td>
</tr>
</tbody>
</table>
## SESSION: PROSODY IN LANGUAGE PERFORMANCE II

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase</td>
<td>2364</td>
</tr>
<tr>
<td>Heejin Kim, Jennifer Cole</td>
<td></td>
</tr>
<tr>
<td>Segmental &quot;Anchor&quot; and the French Late Rise</td>
<td>2368</td>
</tr>
<tr>
<td>Pauline Welby, Helene Loevenbruck</td>
<td></td>
</tr>
<tr>
<td>Prosodic Cues for Syntactically-Motivated Junctures</td>
<td>2372</td>
</tr>
<tr>
<td>Ivan Chow</td>
<td></td>
</tr>
<tr>
<td>A Glimpse of the Time-Course of Intonation Processing in European Portuguese</td>
<td>2376</td>
</tr>
<tr>
<td>Isabel Fale, Isabel Hub Faria</td>
<td></td>
</tr>
<tr>
<td>Great Expectations --- Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates</td>
<td>2380</td>
</tr>
<tr>
<td>Petra Wagner</td>
<td></td>
</tr>
<tr>
<td>Choosing a Scale for Measuring Perceived Prominence</td>
<td>2384</td>
</tr>
<tr>
<td>Christian Jensen, John Tondering</td>
<td></td>
</tr>
<tr>
<td>The Effects of Prosodic Features on the Interpretation of Clarification Ellipses</td>
<td>2388</td>
</tr>
<tr>
<td>Jens Edlund, David House, Gabriel Skantze</td>
<td></td>
</tr>
<tr>
<td>Exploration of Different Types of Intonational Deviations in Foreign-Accented and Synthesized Speech</td>
<td>2392</td>
</tr>
<tr>
<td>Matthias Jilka</td>
<td></td>
</tr>
<tr>
<td>A Rhythmic-Prosodic Model of Poetic Speech</td>
<td>2396</td>
</tr>
<tr>
<td>Jorg Broggelwirth</td>
<td></td>
</tr>
<tr>
<td>Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech</td>
<td>2400</td>
</tr>
<tr>
<td>Sonja Biersack, Vera Kempe, Lorna Knapton</td>
<td></td>
</tr>
<tr>
<td>An Analysis of the Intonational Structure of Stuttered Speech</td>
<td>2404</td>
</tr>
<tr>
<td>Timothy Arbis-Kelm</td>
<td></td>
</tr>
<tr>
<td>Voice Quality Dimensions of Pitch Accents</td>
<td>2408</td>
</tr>
<tr>
<td>Britta Lintfert, Wolfgang Wokurek</td>
<td></td>
</tr>
<tr>
<td>Audiovisual Production and Perception of Contrastive Focus in French: A Multispeaker Study</td>
<td>2412</td>
</tr>
<tr>
<td>Marion Dohen, Helene Loevenbruck</td>
<td></td>
</tr>
<tr>
<td>Predicting End of Utterance in Multimodal and Unimodal Conditions</td>
<td>2416</td>
</tr>
<tr>
<td>Pashiera Barkhuysen, Emiel Krahmer, Marc Swerts</td>
<td></td>
</tr>
<tr>
<td>Production of Prominence in Japanese Sign Language</td>
<td>2420</td>
</tr>
<tr>
<td>Sao ň Tanaka, Masafumi Nishida, Yasuo Horluchi, Akira Ichikawa</td>
<td></td>
</tr>
<tr>
<td>MLLR Transforms as Features in Speaker Recognition</td>
<td>2424</td>
</tr>
<tr>
<td>Andreas Stolcke, Luciana Ferrer, Sachin Kajarekar, Elizabeth Shriberg, Anand Venkataram</td>
<td></td>
</tr>
<tr>
<td>Gaussian Mixture Modelling of Broad Phonetic and Syllabic Events for Text-Independent Speaker Verification</td>
<td>2428</td>
</tr>
<tr>
<td>Brendan Baker, Robbie Vogt, Sridha Sridharan</td>
<td></td>
</tr>
<tr>
<td>Efficient Speaker Identification and Retrieval</td>
<td>2432</td>
</tr>
<tr>
<td>Hagai Aronowitz, David Burshtein</td>
<td></td>
</tr>
<tr>
<td>The Cambridge University March 2005 Speaker Diarisation System</td>
<td>2436</td>
</tr>
<tr>
<td>R. Sinha, S.E. Tranter, M.J.F. Gales, P.C. Woodland</td>
<td></td>
</tr>
<tr>
<td>Combining Speaker Identification and BIC for Speaker Diarization</td>
<td>2440</td>
</tr>
<tr>
<td>Xuan Zhu, Claude Barras, Sylvain Meignier, Jean-Luc Gauvain</td>
<td></td>
</tr>
</tbody>
</table>
Broadcast News Speaker Tracking for ESTER 2005 Campaign ..................................................... 2444
  Dan Istrate, Nicolas Scheffer, Corinne Fredouille, Jean-Francois Bonastre

On the Nature of Acoustic Information in Identification of Coarticulated Vowels ......................... 2448
  Sorin Dusan

Impact of Duration on F1/F2 Formant Values of Oral Vowels: An Automatic Analysis of Large Broadcast News Corpora in French and German ................................................. 2452
  Cedric Gendrot, Martine Adda-Decker

Modeling of Between-Speaker and Within-Speaker Variation in Spontaneous Speech Tempo ................................................................................................................................. 2456
  Hugo Quene

Vowel Devoicing vs. Mora-Timed Rhythm in Spontaneous Japanese --- Inspection of Phonetic Labels of OGI_TS ........................................................................................................... 2460
  Masahiko Komatsu, Makiko Aoyagi

Does Vowel Space Size Depend on Language Vowel Inventories? Evidence from Two Arabic Dialects and French ................................................................................. 2464
  Jalal-Eddin Al-Tamimi, Emmanuel Ferragne

Understanding Phonology by Phonetic Implementation .................................................................. 2468
  Chilin Shih

User Evaluation of Conversational Agent H. C. Andersen ................................................................. 2472
  Niels Ole Bernsen, Laila Dybkjaer

Integrated Development and on-the-Fly Simulation of Multimodal Dialogs .................................. 2476
  Silke Goronzy, Nicole Beringer

Interactions Between Speech Recognition Problems and User Emotions ...................................... 2480
  Mihai Rotaru, Diane J. Litman, Katherine Forbes-Riley

WebTalk: Mining Websites for Interactively Answering Questions ................................................. 2484
  Junlan Feng, Srinath Reddy, Murat Saraclar

Towards Generic Quality Prediction Models for Spoken Dialogue Systems --- A Case Study ........ 2488
  Sebastian Moller

Robust Access to Large Structured Data Using Voice Form-Filling .............................................. 2492
  S. Parthasarathy, Cyril Allauzen, R. Munkong

SESSION: HUMAN FACTORS, USER EXPERIENCE AND NATURAL LANGUAGE APPLICATION DESIGN

Spoken Dialog System for Real-Time Data Capture ........................................................................ 2496
  Esther Levin, Alex Levin

A User Study on the Influence of Mobile Device Class, Synthesis Method, Data Rate and Lexicon on Speech Synthesis Quality ................................................................. 2500
  Michael Pucher, Peter Frohlich

User's Experience of a Commercial Speech Dialogue System ...................................................... 2504
  Fang Chen, Yael Katzenellenbogen

Voice User Interface Design for Automated Directory Assistance ................................................. 2508
  Esther Levin, Amir M. Mane

Optimizing User Experience Through Design of the Spoken Language Understanding (SLU) Module ..................................................................................................................... 2512
  Maria Gabriela Alvarez-Ryan, Narendra Gupta, Barbara Hollister, Tirso Alonso

Interactive Visualization of Human-Machine Dialogs ..................................................................... 2516
  Jeremy Wright, David Kapilow, Alicia Abella
Synthesising Hyperarticulation in Unit Selection TTS .......................................................... 2520
Matthew P. Aylett

Symbolic Prosody Driven Unit Selection for Highly Natural Synthetic Speech ......................... 2524
Daniel Tihelka

Hybrid Syllable/Triphone Speech Synthesis ........................................................................... 2528
Jindrich Matousek, Zdenek Hanzlicek, Daniel Tihelka

A Neural Network Approach for the Design of the Target Cost Function in Unit-Selection Speech Synthesis ........................................................................................................ 2532
Francisco Campillo Diaz, Jose Luis Alba, Eduardo Rodriguez Banga

FSM and K-Nearest-Neighbor for Corpus Based Video-Realistic Audio-Visual Synthesis .................. 2536
Christian Weiss

An Embedded and Concatenative Approach to TTS of Multiple Languages ............................... 2540
Gui-Lin Chen, Ke-Song Han, Zhen-Li Yu, Dong-Jian Yue, Yi-Qing Zu

Morphing Spectral Envelopes Using Audio Flow .................................................................. 2544
Tony Ezzat, Ethan Meyers, James Glass, Tomaso Poggio

Linguistic Features Weighting for a Text-to-Speech System Without Prosody Model .......................................................... 2548
Vincent Colotte, Richard Beaufort

Unit Selection Synthesis Database Development Using Utterance Verification ........................ 2552
Ingunn Amdal, Torbjorn Svendsen

Refining Phoneme Segmentations Using Speaker-Adaptive Context Dependent Boundary Models .......................................................................................................................... 2556
Yong Zhao, Lijuan Wang, Min Chu, Frank K. Soong, Zhigang Cao

Customizing Base Unit Set with Speech Database in TTS Systems ............................................. 2560
Yining Chen, Yong Zhao, Min Chu

Unit Selection for Speech Synthesis Based on a New Acoustic Target Cost .................................. 2564
Soufiane Rouibia, Olivier Rosec

Small Footprint Concatenative Text-to-Speech Synthesis System Using Complex Spectral Envelope Modeling ........................................................................................................... 2568
Dan Chazan, Ron Hoory, Zvi Kons, Ariel Sagi, Slava Shechtman, Alexander Sorin

High Quality Spanish Restricted-Domain TTS Oriented to a Weather Forecast Application ............ 2572
Francesc Alias, Ignasi Iriondo, Lluis Formiga, Xavier Gonzalvo, Carlos Monzo, Xavier Sevillano

Comparing Spectral Distance Measures for Join Cost Optimization in Concatenative Speech Synthesis .......................................................................................................................... 2576
Ingmund Bjorkan, Torbjorn Svendsen, Snorre Farner

HMM-Based European Portuguese TTS System ...................................................................... 2580
Maria Joao Barros, Rannier Maia, Keichi Tokuda, Fernando Gil Resende, Diamantino Freitas

Combining the Flexibility of Speech Synthesis with the Naturalness of Pre-Recorded Audio: A Comparison of Two Approaches to Phrase-Splicing TTS .................................................. 2584
Wael Hamza, John F. Pitrelli

Codec Integrated Voice Conversion for Embedded Speech Synthesis ........................................ 2588
Guntram Strecha, Oliver Jokisch, Matthias Eichner, Rudiger Hoffmann

Evaluation of VTLN-Based Voice Conversion for Embedded Speech Synthesis ....................... 2592
David Sundermann, Guntram Strecha, Antonio Bonafonte, Harald Hoge, Hermann Ney

Model Adaptation and Adaptive Training Using ESAT Algorithm for HMM-Based Speech Synthesis .......................................................................................................................... 2596
Juri Isogai, Junichi Yamagishi, Takao Kobayashi
Embedded Cantonese TTS for Multi-Device Access to Web Content .............................................. 2600
Tien-Ying Fung, Yuk-Chi Li, Eddie Sio, Icarus Lee, Helen Meng, P.C. Ching

Model Based Analysis of a Diphone Database for Improved Unit Concatenation ................... 2604
Karl Schnell, Arild Lacroix

SESSION: ROBUST SPEECH RECOGNITION III

Context-Dependent Word Duration Modelling for Robust Speech Recognition ......................... 2608
Ning Ma, Phil Green

An Energy Search Approach to Variable Frame Rate Front-End Processing for Robust ASR ................................................................. 2612
Julien Epps, Eric H.C. Choi

Non-Linear Estimation of Voice Activity to Improve Automatic Recognition of Noisy Speech ...................................................................................... 2616
Roberto Gemello, Franco Mana, Renato de Mori

Voice Activity Detection Based on Optimally Weighted Combination of Multiple Features .............................................................................................................. 2620
Yusuke Kida, Tatsuya Kawahara

Soft Decision Strategy and Adaptive Compensation for Robust Speech Recognition Against Impulsive Noise ............................................................. 2624
Pei Ding

Statistical Class-Based MFCC Enhancement of Filtered and Band-Limited Speech for Robust ASR .................................................................................................................. 2628
Nicolas Morales, Doroteo Torre Toledano, John H.L. Hansen, Jose Colas, Javier Garrido

Spectral Entropy Feature in Full-Combination Multi-Stream for Robust ASR ............................ 2632
Hemant Misra, Herve Bourlard

Environment-Independent Mask Estimation for Missing-Feature Reconstruction ....................... 2636
Wooil Kim, Richard M. Stern, Hanseok Ko

Soft Harmonic Masks for Recognising Speech in the Presence of a Competing Speaker ............ 2640
Andre Coy, Jon Barker

Comb Filter Decomposition for Robust ASR .................................................................................. 2644
Lech Szymanski, Martin Bouchard

Investigating the Role of the Lombard Reflex in Non-Audible Murmur (NAM) Recognition ...................................................................................................................... 2648
Panikos Heracleous, Tomomi Kaino, Hiroshi Saruwatari, Kiyohiro Shikano

Improved "TEO" Feature-Based Automatic Stress Detection Using Physiological and Acoustic Speech Sensors .................................................................................................. 2652
Evan Ruzanski, John H.L. Hansen, Don Finan, James Meyerhoff, William Norris, Terry Wollert

Spectral Subtraction Using Elliptic Integral for Multiplication Factor ........................................ 2656
Takeshi S. Kobayakawa

Robust Distant Speech Recognition Based on Position Dependent CMN Using a Novel Multiple Microphone Processing Technique ...................................................... 2660
Longbiao Wang, Northide Kitaoka, Seiichi Nakagawa

Data Collection and Evaluation of Speech Recognition for Motorbike Riders ........................ 2664
H. Tanaka, H. Fujimura, C. Miyajima, T. Nishino, Katunobu Itou, Kazuya Takeda

Application of a First-Order Differential Microphone for Efficient Voice Activity Detection in a Car Platform ................................................................................................. 2668
Agustin Alvarez, Pedro Gomez, V. Nieto, Rafael Martinez, Victoria Rodellar
Robust Speech Recognition for Mobile Devices in Car Noise ................................................................. 2672
Panji Setiawan, Suhadi Suhadi, Tim Fingscheidt, Sorel Stan

Evaluation and Optimization of Noise Robust Front-End Technologies for the
Automatic Recognition of Hungarian Telephone Speech ........................................................................... 2676
Peter Mihajlik, Zoltan Tobler, Zoltan Tuske, Geza Gordos

A Performance Investigation of Noisy Voice Recognition Over IP Telephony
Networks ...................................................................................................................................................... 2680
Gang Chen, Douglas O'Shaughnessy, Hesham Tolba

Internal Noise Suppression for Speech Recognition by Small Robots ...................................................... 2684
Akinori Ito, Takashi Kanayama, Motonori Suzuki, Shozo Makino

Temporal ICA for Classification of Acoustic Events in a Kitchen Environment ........................................ 2688
Florian Kraft, Robert Malkin, Thomas Schaaf, Alex Waibel

"Hello -- Is Anybody at Home?" --- About the Minimum Word Accuracy of a Smart
Home Spoken Dialogue System ............................................................................................................. 2692
Jan Felix Krebber

The Simulation of Realistic Acoustic Input Scenarios for Speech Recognition Systems ................................ 2696
H. Gunter Hirsch, Harald Finster

An Agent-Based Framework for Speech Investigation .............................................................................. 2700
Michael Walsh, Gregory M.P. O'Hare, Julie Carson-Berndsen

Switched Split Vector Quantisation of Line Spectral Frequencies for Wideband Speech Coding ................ 2704
Stephen So, Kuldip K. Paliwal

A Novel Voicing Cut-Off Determination for Low Bit-Rate Harmonic Speech Coding ............................ 2708
Changchun Bao, Jason Lukasiak, Christian Ritz

A Partial Decorrelation Scheme for Improved Predictive Open Loop Quantization with Noise Shaping ............................................................... 2712
Hauke Kruger, Peter Vary

Using Dynamic Codebook Re-Ordering to Exploit Inter-Frame Correlation in MELP Coders .................. 2716
Venkatesh Krishnan, Thomas P. Barnwell III, David V. Anderson

Enhanced Speech Coding Based on Phonetic Class Segmentation ............................................................ 2720
Adriane Swalm Durey, Venkatesh Krishnan, Thomas P. Barnwell III

A Pitch-Synchronous Pitch-Cycle Modification Method for Designing a Hybrid I-MELP/Waveform-Matching Speech Coder ............................................................... 2724
Ali Erdem Ertan, Thomas P. Barnwell III

A New Structural Preprocessor for Low-Bit Rate Speech Coding ............................................................ 2728
Joon-Hyuk Chang, Jong-Won Shin, Seung Yeol Lee, Nam Soo Kim

An Improved GMM-Based Voice Quality Predictor ................................................................................. 2732
Tiago H. Falk, Wai-Yip Chan, Peter Kabal

High-Quality Memoryless Subband Coding of Impulse Responses at 22 Bits per Frame .......................... 2736
Jan Erkelens

A Study of Variable Pulse Allocation for MPE and CELP Coders Based on PESQ Analysis ...................... 2740
Shi-Han Chen, Kuo-Guan Wu, Chih-Chung Kuo

Joint Source-Channel Coding of LSP Parameters for Bursty Channels .......................................................... 2744
Jose L. Perez-Cordoba, Antonio M. Peinado, Angel M. Gomez, Antonio J. Rubio
SESSION: GENDER AND AGE ISSUES IN SPEECH AND LANGUAGE RESEARCH II

Adaptation and Normalization Experiments in Speech Recognition for 4 to 8 Year old Children ................................................................................................................................... 274
Daniel Elenius, Mats Blomberg

PROSPECT Features and their Application to Missing Data Techniques for Vocal Tract Length Normalization .................................................................................................................. 272
Wim Jansen, Hugo Van hamme

Data Driven Subword Unit Modeling for Speech Recognition and its Application to Interactive Reading Tutors ............................................................................................................ 275
Andreas Hagen, Bryan L. Pellom

The PF_STAR Children's Speech Corpus .................................................................................. 276
Anton Batliner, Mats Blomberg, Shona D'Arcy, Daniel Elenius, Diego Giuliani, Matteo Gerossi, Christian Hacker, Martin Russell, Stefan Steidl, Michael Wong

The Swedish NICE Corpus --- Spoken Dialogues Between Children and Embodied Characters in a Computer Game Scenario ................................................................................. 276
Linda Bell, Johan Boye, Joakim Gustafson, Mattias Heldner, Anders Lindstrom, Mats Wiren

A Preprocessing Technique for Improving Speech Intelligibility in Reverberant Environments: The Effect of Steady-State Suppression on Elderly People.................................................. 276
Yusuke Miyauchi, Nao Hodoshima, Keiichi Yasu, Nahoko Hayashi, Takayuki Arai, Mitsuko Shindo

SESSION: DISCOURSE AND DIALOGUE I

Synchronizing Dialogue Contributions of Human Users and Virtual Characters in a Virtual Reality Environment ...................................................................................................... 273
Norbert Pfleger, Markus Lockelt

Volume 5

Does Active Learning Help Automatic Dialog Act Tagging in Meeting Data?......................... 2776
Anand Venkataraman, Yang Liu, Elizabeth Shriberg, Andreas Stolcke

A Principled Approach for Rejection Threshold Optimization in Spoken Dialog Systems ................................................................................................................................. 2780
Dan Bohus, Alexander I. Rudnicky

Application of Confidence Measures for Dialogue Systems Through the Use of Parallel Speech Recognizers .............................................................................................................. 2784
David Perez-Pinar Lopez, Carmen Garcia Mateo

Multi-Level Information and Automatic Dialog Acts Detection in Human-Human Spoken Dialogs .................................................................................................................. 2788
Sophie Rosset, Delphine Tribout

From Question Answering to Spoken Dialogue: Towards an Information Search Assistant for Interactive Multimodal Information Extraction ................................................. 2792
Rieks op den Akker, Harry Bunt, Simon Keizer, Boris van Schooten

SESSION: TEXT-TO-SPEECH II

Pitch-Effects in Diphone Recording: Are Logatomes inappropriate? ........................................ 2796
Ulrich Reubold, Alexander Steffen
Speech Parameter Generation Algorithm Considering Global Variance for HMM-Based Speech Synthesis  .......................................................... 2800
Tomoki Toda, Keiichi Tokuda

Performance Evaluation of Style Adaptation for Hidden Semi-Markov Model Based Speech Synthesis .......................................................... 2804
Makoto Tachibana, Junichi Yamagishi, Takashi Masuko, Takao Kobayashi

A Comparison of Methods for Speaker-Dependent Pronunciation Tuning for Text-to-Speech Synthesis .......................................................... 2808
Gabriel Webster, Tina Burrows, Katherine Knill

Perceptually-Based Data-Driven Join Costs: Comparing Join Types .......................................................... 2812
Ann K. Syrdal, Alistair D. Conkie

Discontinuity Detection in Concatenated Speech Synthesis Based on Nonlinear Speech Analysis .......................................................... 2816
Yannis Pantazis, Yannis Stylianou, Esther Klabbers

I

SESSION: LANGUAGE AND DIALECT IDENTIFICATION II

Improving the Discrimination Between Native Accents When Recorded Over Different Channels .......................................................... 2820
Tingyao Wu, Dirk Van Compernolle, Jacques Duchateau, Qian Yang, Jean-Pierre Martens

Aligning and Recognizing Spoken Books in Different Varieties of Portuguese .......................................................... 2824
Isabel Trancoso, Antonio Serralheiro, Ceu Viana, Diamantino Caseiro

An Acoustic Segment Modeling Approach to Automatic Language Identification .......................................................... 2828
Bin Ma, Haizhou Li, Chin-Hui Lee

Different Size Multilingual Phone Inventories and Context-Dependent Acoustic Models for Language Identification .......................................................... 2832
Dong Zhu, Martine Adda-Decker, Fabien Antoine

A Text Categorization Approach to Automatic Language Identification .......................................................... 2836
Sheng Gao, Bin Ma, Haizhou Li, Chin-Hui Lee

Advances in Regional Accent Clustering in Swedish .......................................................... 2840
Giampiero Salvi

SESSION: SPEECH RECOGNITION IN UBQUITOUS NETWORKING AND CONTEXT-AWARE COMPUTING

An Architecture for Seamless Access to Distributed Multimodal Services .......................................................... 2844
David Pearce, Jonathan Engelsma, James Ferrans, John Johnson

Robust Speech Recognition in Ubiquitous Networking and Context-Aware Computing .......................................................... 2848
Zheng-Hua Tan, Paul Dalsgaard, Borge Lindberg, Haitian Xu

Unified Probabilistic Approach to Error Concealment for Distributed Speech Recognition .......................................................... 2852
Valentin Ion, Reinhold Haeb-Umbach

Combining Packet Loss Compensation Methods for Robust Distributed Speech Recognition .......................................................... 2856
Alastair James, Ben Milner

Distributed ASR Using Speech Coder Data for Efficient Feature Vector Representation .......................................................... 2860
Trond Skogstad, Torbjorn Svendsen

Cluster-based Modeling for Ubiquitous Speech Recognition .......................................................... 2864
Sadaoki Furui, Tomohisa Ichiba, Takahiro Shinozaki, Edward W.D. Whittaker, Koji Iwano
SESSION: PHONETICS AND PHONOLOGY II

The Feature [sonorant] in Lexical Access ................................................................. 2868
Danny R. Moates, Z.S. Bond, Russell Fox, Verna Stockmal

Voice and Aspiration in German and East Bengali Stops: A Cross-Language Study ......................................................... 2872
Simone Mikuteit

Polder Dutch: Aspects of the /t[textesil]i/-Lowering in Standard Dutch .................. 2876
Irene Jacobi, Louis C.W. Pols, Jan Stroop

Production and Perception of Vietnamese Vowels .............................................. 2880
Eric Castelli, Rene Carre

Using Open Quotient for the Characterisation of Vietnamese Glottalised Tones .......... 2884
Tuan Vu Ngoc, Christophe d'Alessandro, Alexis Michaud

On the Acoustic Characterization of Ejective Stops in Waima’a ................................ 2888
John Hajek, Mary Stevens

Spirantization of /p t k/ in Sienese Italian and So-Called Semi-Fricatives ................. 2892
Mary Stevens, John Hajek

Italian Geminates Under Speech Rate and Focalization Changes: Kinematic, Acoustic, and Perception Data .................................................. 2896
Barbara Gili Fivela, Claudio Zmarich

Durational Characteristics of Korean Lombard Speech ........................................ 2900
Sunhee Kim

A Cross-Linguistic Study of Vowel Quantity in Different Word Structures:
Japanese, Finnish and Czech ............................................................................. 2904
Toshiko Isei-Jaakkola, Satoshi Asakawa

Acoustic Properties of Foreign Accent: VOT Variations in Moroccan-Accented Italian .......................................................... 2908
Laura Mori, Melissa Barkat-Defradas

The Interrelation Between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese ........................................... 2912
Andreia S. Rauber, Paola Escudero, Ricardo A.H. Bion, Barbara O. Baptista

Recognition of German Obstruents .................................................................. 2916
Julia Hoelterhoff

Czech Voiced Labiodental Continuant Discrimination from Basic Acoustic Data ........ 2920
Radek Skarnitzl, Jan Volin

An Elitist Approach for Extracting Automatically Well-Realized Speech Sounds with High Confidence ........................................................ 2924
Jean-Baptiste Maj, Anne Bonneau, Dominique Fohr, Yves Laprie

Applying Multiple Regression Models for Predicting Word Duration in a Corpus of Spontaneous Speech .............................................................. 2928
Na'im R. Tyson

On European Portuguese Automatic Syllabification .......................................... 2932
Catarina Oliveira, Lurdes Castro Moutinho, Antonio J.S. Teixeira

Rule-Based Grapheme-to-Phoneme Method for the Greek .................................... 2936
A. Chalamandaris, S. Raptis, Pirros Tsiakoulis

Assimilation and Deletion Phenomena Involving Word-Final /n/ and Word-Initial /p, t, k/ in Modern Greek: a Codification of the Observed Variation Intended for Use in TTS Synthesis ................................................................. 2940
Constandinos Kalimeris, George Mikros, Stelios Bakamidis
A German Viseme-Set for Automatic Transcription of Input Text Used for Audio-Visual-Speech-Synthesis .......................................................... 294
Christian Weiss, Bianca Aschenberner

Visual Perception of Anticipatory Rounding Gestures in French .................................................. 2948
Johanna-Pascale Roy

SESSION: ACOUSTIC PROCESSING FOR ASR III

Hierarchical Clustering of Mixture Tying Using a Partially Observable Markov Decision Process .............................................................. 2952
Michael Jonas, James G. Schmolze

Flavors of Gaussian Warping ........................................................................................................ 2956
Pierre Ouellet, Gilles Boulianne, Patrick Kenny

Phoneme Alignment Based on Discriminative Learning .............................................................. 2960
Joseph Keshet, Shai Shalev-Shwartz, Yoram Singer, Dan Chazan

Comparison of Low Footprint Acoustic Modeling Techniques for Embedded ASR Systems .......................................................... 2964
Jussi Leppanen, Imre Kiss

Factors in Classification of Stop Consonant Place of Articulation ........................................ 2968
Atiwong Suchato, Proadpran Punyabukkana

Cross-Speaker Articulatory Position Data for Phonetic Feature Prediction .................................. 2972
Arthur R. Toth, Alan W. Black

Improvements to fMPE for Discriminative Training of Features ............................................. 2976
Daniel Povey

Incorporating Tone-Related MLP Posteriors in the Feature Representation for Mandarin ASR .................................................................. 2980
Xin Lei, Mei-Yuh Hwang, Mari Ostendorf

Speech Trajectory Clustering for Improved Speech Recognition ................................................. 2984
Yan Han, Johan de Veth, Louis Boves

Selection of Features and Combination of Classifiers Using a Fuzzy Approach for Acoustic Event Classification ........................................ 2988
Andrey Temko, Dusan Macho, Climent Nadeu

Multi-Task Learning Strategies for a Recurrent Neural Net in a Hybrid Tied-Posteriors Acoustic Model .................................................................. 2992
Jan Stadermann, Wolfram Koska, Gerhard Rigoll

Revising Perceptual Linear Prediction (PLP) .............................................................................. 2996
Florian Honig, Georg Stemmer, Christian Hacker, Fabio Brugnara

Confidence Measures in Speech Recognition Based on Probability Distribution of Likelihoods ................................................................. 3000
Joel Pinto, R.N.V. Sitaram

Continuous Local Codebook Features for Multi- and Cross-Lingual Acoustic Phonetic Modelling .................................................................. 3004
Frank Diehl, Asuncion Moreno, Enric Monte

Augmented State Space Acoustic Decoding for Modeling Local Variability in Speech ................................................................. 3008
Antonio Miguel, Eduardo Lleida, Richard Rose, Luis Buera, Alfonso Ortega

Auditory Teager Energy Cepstrum Coefficients for Robust Speech Recognition .................. 3012
Dimitris Dimitriadis, Petros Maragos, Alexandros Potamianos

A Hybrid MaxEnt/HMM Based ASR System ........................................................................... 3016
Yasser Hifny, Steve Renals, Neil D. Lawrence
Regularizing Linear Discriminant Analysis for Speech Recognition
Hakan Erdogan

Comprehensive Modulation Representation for Automatic Speech Recognition
Yadong Wang, Steven Greenberg, Jayaganesh Swaminathan, Ramdas Kumaresan, David Poeppel

Segment-Based Phonetic Class Detection Using Minimum Verification Error (MVE) Training
Qiang Fu, Biing-Hwang Juang

Acoustic and Phonetic Confusions in Accented Speech Recognition
Yi Liu(1), Pascale Fung

Auditory Image Model Features for Automatic Speech Recognition
Mario E. Munich, Qiguang Lin

Applications of NAM Microphones in Speech Recognition for Privacy in Human-Machine Communication
Panikos Heracleous, Tomomi Kaino, Hiroshi Saruwatari, Kiyohiro Shikano

A Hybrid ANN/DBN Approach to Articulatory Feature Recognition
Joe Frankel, Simon King

SESSION: SPEAKER CHARACTERIZATION AND RECOGNITION IV

Experiments on Speaker Tracking and Segmentation in Radio Broadcast News
Daniel Moraru, Mathieu Ben, Guillaume Gravier

Unsupervised Segmentation and Verification of Multi-Speaker Conversational Speech
Emanuele Dalmasso, Pietro Laface, Daniele Colibro, Claudio Vair

Focal Speakers: A Speaker Selection Method Able to Deal with Heterogeneous Similarity Criteria
Sacha Krstulovic, Frederic Bimbot, Delphine Charlet, Olivier Boeffard

A Model Space Framework for Efficient Speaker Detection
Mathieu Ben, Guillaume Gravier, Frederic Bimbot

Speaker Detection Using Acoustic Event Sequences
Nicolas Scheffer, Jean-Francois Bonastre

Speaker Clustering of Unknown Utterances Based on Maximum Purity Estimation
Wei-Ho Tsai, Hsin-Min Wang

Modified DISTBIC Algorithm for Speaker Change Detection
Petra Zochova, Vlasta Radova

Decision Trees with Improved Efficiency for Fast Speaker Verification
Gilles Gonon, Remi Gribonval, Frederic Bimbot

A Speaker Independent "Liveness" Test for Audio-Visual Biometrics
Nicolas Eveno, Laurent Besacier

Distributed Speaker Recognition Using Speaker-Dependent VQ Codebook and Earth Mover's Distance
Shingo Kuroiwa, Yoshiyuki Umeda, Satoru Tsuge, Fuji Ren

Speaker Verification via Articulatory Feature-Based Conditional Pronunciation Modeling with Vowel and Consonant Mixture Models
Ka-Yee Leung, Man-Wai Mak, Manhung Siu, Sun-Yuan Kung

Prosodic Features Based on Wavelet Analysis for Speaker Verification
Jixu Chen, Beiqian Dai, Jun Sun
Relevant Information Extraction for Discriminative Training Applied to Speaker Identification.................................................................................................................................. 3096
M. Mihoubi, Douglas O'Shaughnessy, P. Dumouchel

Conceiving a New Sequence Kernel and Applying it to SVM Speaker Verification......................... 3100
Jerome Louradour, Khalid Daoudi

The Predictive Differential Amplitude Spectrum for Robust Speaker Recognition in Stationary Noises...................................................................................................................... 3104
Jing Deng, Thomas Fang Zheng, Jian Liu, Wenhu Wu

Data-Driven Clustering for Blind Feature Mapping in Speaker Verification........................................ 3108
Michael Mason, Robbie Vogt, Brendan Baker, Sridha Sridharan

Improved Covariance Modeling for GMM in Speaker Identification.................................................. 3112
Xi Zhou, Zhi-qiang Yao, Beiqian Dai

Modelling Session Variability in Text-Independent Speaker Verification........................................... 3116
Robbie Vogt, Brendan Baker, Sridha Sridharan

Overlapping Wavelet Packet Features for Speaker Verification.......................................................... 3120
Mihalis Siafarikas, Todor Ganchev, Nikolaos Fakotakis, George Kokkinakis

Using Hadamard ECOC in Multi-Class problems Based on SVM....................................................... 3124
An-rong Yin, Xiang Xie, Jingming Kuang

SESSION: ROBUST SPEECH RECOGNITION IV

Joint Uncertainty Decoding for Noise Robust Speech Recognition..................................................... 3128
H. Liao, M.J.F. Gales

Confidence Scoring and Rejection Using Multi-Pass Speech Recognition.......................................... 3132
Vincent Vanhoucke

Memory-Enhanced MMSE-Based Channel Error Mitigation for Distributed Speech Recognition.......................................................... 3136
Cheng-Lung Lee, Wen-Whei Chang

Designing Multiple Distinctive Phonetic Feature Extractors for Canonicalization by Using Clustering Technique.............................................................................................................. 3140
Takashi Fukuda, Muhammad Ghulam, Tsuneo Nitta

Efficient Blind Dereverberation Framework for Automatic Speech Recognition.................................. 3144
Keisuke Kinoshita, Tomohiro Nakatani, Masato Miyoshi

Combining Multi-Source Far Distance Speech Recognition Strategies: Beamforming, Blind Channel and Confusion Network Combination......................................................... 3148
Matthias Wolfel, John McDonough

SESSION: SPEECH CODING AND QUALITY ASSESSMENT

Objective Quality Assessment of Wideband Speech by an Extension of ITU-T Recommendation P.862.................................................................................................................. 3152
Akira Takahashi, Atsuko Kurashima, Chiharu Morioka, Hideaki Yoshino

Quality Control for UMTS-AMR Speech Channels.............................................................................. 3156
Marc Werner, Peter Vary

Perceptual Postfilter Estimation for Low Bit Rate Speech Coders Using Gaussian Mixture Models................................................................................................................................. 3160
Wei Chen, Peter Kabal, Turaj Z. Shabestary

SNR-dependent Background Noise Compensation of PESQ Values for Cellular Phone Speech......................................................................................................................... 3164
Kengo Fujita, Tsuneo Kato, Hideaki Yamada, Hisashi Kawai
# A MFCC-Based CELP Speech Coder for Server-Based Speech Recognition in Network Environments

Gil Ho Lee, Jae Sam Yoon, Hong Kook Kim

---

# Distortion Measures for Vector Quantization of Noisy Spectrum

Volodya Grancharov, Jonas Samuelsson, W. Bastiaan Kleijn

---

## SESSION: SPOKEN LANGUAGE TRANSLATION II

### On the Integration of Speech Recognition and Statistical Machine Translation

E. Matusov, S. Kanthak, Hermann Ney

### Integrated N-Best Re-Ranking for Spoken Language Translation

V.H. Quan, M. Federico, M. Cettolo

### An Ngram-Based Statistical Machine Translation Decoder

Josep M. Crego, Jose B. Marino, Adria de Gispert

### Use of Maximum Entropy in Natural Word Generation for Statistical Concept-Based Speech-to-Speech Translation

Liang Gu, Yuqing Gao

### Improving Statistical Machine Translation by Classifying and Generalizing Inflected Verb Forms

Adria de Gispert, Jose B. Marino, Josep M. Crego

### Improved Speech Recognition Word Lattice Translation by Confidence Measure

Abdulvohid Bozarov, Yoshinori Sagisaka, Ruiqiang Zhang, Genichiro Kikui

### Vocal Tract Area Function Inversion by Linear Regression of Cepstrum

Parham Mokhtari, Tatsuya Kitamura, Hironori Takemoto, Kiyoshi Honda

### Introducing Visual Cues in Acoustic-to-Articulatory Inversion

Olov Engwall

### Speech Inversion and Re-Synthesis

V.N. Sorokin, A.S. Leonov, I.S. Makarov, A.I. Tsyplikhin

### Teaching a Vocal Tract Simulation to Imitate Stop Consonants

Mark Huckvale, Ian Howard

### Using Phonetic Constraints in Acoustic-to-Articulatory Inversion

Blaise Potard, Yves Laprie

### A Support Vector Approach to the Acoustic-to-Articulatory Mapping

Asterios Toutios, Konstantinos Margaritis

---

## SESSION: PROSODY MODELING AND SPEECH TECHNOLOGY II

### Analysis by Synthesis of Speech Prosody: The ProZed Environment

Daniel Hirst, Cyril Auran

### A Discriminative Approach to Phrase Break Modelling

Stephen Cox

### Stochastic and Syntactic Techniques for Predicting Phrase Breaks

Ian Read, Stephen Cox

### Tree-Based Prediction of Prosodic Phrase Breaks on Top of Shallow Textual Features

Gerasimos Xydas, Panagiotis Zervas, Georgios Kouropetrouoglou, Nikolaos Fakotakis, George Kokkinakis

### Chinese Prosodic Phrasing with a Constraint-Based Approach

Honghui Dong, Jianhua Tao, Bo Xu
A Probabilistic Approach to Prosodic Word Prediction for Mandarin Chinese TTS .......... 3244
Minghui Dong, Kim-Teng Lua, Haizhou Li

Evaluation of a System for F_0 Contour Prediction for European Portuguese .......... 3248
Joao Paulo Teixeira, Diamantino Freitas, Hiroya Fujisaki

Analysis on Command Sequences of a F0 Generation Model for Mandarin Speech and its Application to their Automatic Extraction ............................................................... 3252
Ke Li, Yoshinori Sagisaka

Corpus-Based Extraction of F_0 Contour Generation Process Model Parameters .......... 3256
Keikichi Hirose, Yusuke Furuyama, Nobuaki Minematsu

Optimized Selection of Intonation Dictionaries in Corpus Based Intonation Modelling .......................................................... 3260
David Escudero, Valentin Cardenoso

Generation of Fundamental Frequency Contours for Mandarin Speech Synthesis Based on Tone Nucleus Model ................................................................. 3264
Qinghua Sun, Keikichi Hirose, Wentao Gu, Nobuaki Minematsu

On the Inter-Syllable Coarticulation Effect of Pitch Modeling for Mandarin Speech .... 3268
Chen-Yu Chiang, Yih-Ru Wang, Sin-Hong Chen

Training the Tilt Intonation Model Using the JEMA Methodology .......................... 3272
Matej Rojc, Pablo Daniel Aguero, Antonio Bonafonte, Zdravko Kacic

Piecewise Linear Stylization of Pitch Via Wavelet Analysis ................................... 3276
Dagen Wang, Shrikanth Narayanan

Phonetic Labeling and Segmentation of Mixed-Lingual Prosody Databases .............. 3280
Harald Romsdorfer, Beat Pfister

Exploratory Analysis of Linguistic Data Based on Genetic Algorithm for Robust Modeling of the Segmental Duration of Speech ......................................................... 3284
Edmilson Morais, Fabio Violaro

Annotation-Mining for Rhythm Model Comparison in Brazilian Portuguese ............. 3288
Dafydd Gibbon, Flaviane Romani Fernandes

A Stochastic Approach to Phoneme and Accent Estimation .................................... 3292
Tohru Nagano, Shinsuke Mori, Masafumi Nishimura

The Detection of Emphatic Words Using Acoustic and Lexical Features .................. 3296
Jason M. Brenier, Daniel M. Cer, Daniel Jurafsky

Tone Recognition in Mandarin Using Focus ......................................................... 3300
Dinoj Surendran, Gina-Anne Levow, Yi Xu

An Automatic Intonation Recognizer for the Polish Language Based on Machine Learning and Expert Knowledge ................................................................. 3304
Mikolaj Wypych

Generalized Envelope Matching Technique for Time-Scale Modification of Speech (GEM-TSM) ................................................................. 3308
Atsuhiro Sakurai

SESSION: TOPICS IN SPEECH RECOGNITION

Comparing HMM, Maximum Entropy, and Conditional Random Fields for Disfluency Detection ................................................................. 3312
Yang Liu, Elizabeth Shriberg, Andreas Stolcke, Mary Harper

Recognizing Speech from Simultaneous Speakers ............................................... 3316
Bhiksha Raj, Rita Singh, Paris Smaragdis
Polynomial Dynamic Time Warping Kernel Support Vector Machines for Dysarthric Speech Recognition with Sparse Training Data .......................................................... 3320
  Vincent Wan, James Carmichael

Flavoured Acoustic Model and Combined Spelling to Sound for Asymmetrical Bilingual Environment ................................................................. 3324
  R. Lejeune, J. Baude, C. Tchong, H. Crepy, C. Waast-Richard

Genetic Triangulation of Graphical Models for Speech and Language Processing .................................................. 3328
  Chris Bartels, Kevin Duh, Jeff Bilmes, Katrin Kirchhoff, Simon King

Improving Speech Recognition Using a Data-Driven Approach ........................................................... 3332
  Guillermo Aradilla, Jithendra Vepa, Herve Bourlard

Outlier Detection for Acoustic Model Training Using Robust Statistics .......................................................... 3336
  Shigeki Matsuda, Wolfgang Herbordt, Satoshi Nakamura

Optimization Methods for Discriminative Training ........................................................................ 3340
  Jonathan Le Roux, Erik McDermott

Segmentation of Recordings Based on Partial Transcriptions .......................................................... 3344
  Patrick Cardinal, Gilles Boulianne, Michel Comeau

A Speaker Independent Continuous Speech Recognizer for Amharic .................................................. 3348
  Hussien Seid, Bjorn Gamback

Optimizing the Structure of Partly-Hidden Markov Models Using Weighted Likelihood-Ratio Maximization Criterion .................................................. 3352
  Tetsuji Ogawa, Tetsunori Kobayashi

Multilingual Speech Recognition: A Unified Approach .......................................................... 3356
  C. Santhosh Kumar, V.P. Mohandas, Haizhou Li

Detection of Recognition Errors Based on Classifiers Trained on Artificially Created Data .......................................................... 3360
  Tomas Bartos, Ludek Muller

On Designing and Evaluating Speech Event Detectors ........................................................................ 3364
  Jinyu Li, Chin-Hui Lee

Local Word Confidence Measure Using Word Graph and N-Best List .................................................. 3368
  Joseph Razik, Odile Mella, Dominique Fohr, Jean-Paul Haton

Mandarin/English Mixed-Lingual Name Recognition for Mobile Phone .................................................. 3372
  Xiaolin Ren, Xin He, Yaxin Zhang

New Word-Level and Sentence-Level Confidence Scoring Using Graph Theory Calculus and its Evaluation on Speech Understanding .................................................. 3376
  Javier Ferreiros, Ruben San Segundo, Fernando Fernandez, Luis-Fernando D'Haro, Valentin Sama, Roberto Barra, Pedro Mellen

Analysis of Spectral Space Reduction in Spontaneous Speech and its Effects on Speech Recognition Performances .................................................. 3380
  Masanobu Nakamura, Koji Iwano, Sadaoki Furui

SVitchboard 1: Small Vocabulary Tasks from Switchboard ........................................................................ 3384
  Simon King, Chris Bartels, Jeff Bilmes

SESSION: DISCOURSE AND DIALOGUE II

Timing of Experimentally Elicited Minimal Responses as Quantitative Evidence for the Use of Intonation in Projecting TRPs .................................................. 3388
  Wieneke Wesseling, R.J.J.H. van Son

Linguistic and Acoustic Features Depending on Different Situations --- The Experiments Considering Speech Recognition Rate .................................................. 3392
  Shinya Yamada, Toshihiko Itoh, Kenji Araki
Towards VoiceXML Compilation for Portable Embedded Applications in Ubiquitous Environments .......................................................... 3396
Dirk Buhler, Stefan W. Hamerich

Prosody in Public Speech: Analyses of a News Announcement and a Political Interview ................................................................. 3400
Eva Strangert

Characterising Dialogue Call-Flows for Pervasive Environments ................................................................................................... 3404
Amit Anil Nanavati, Nitendra Rajput

An Architecture for Pluggable Disambiguation Mechanism for RDC Based Voice Applications ......................................................... 3408
Tanveer Faruquie, Pankaj Kankar, Nitendra Rajput, Abhishek Verma

Adapting Dialog Call-Flows for Pervasive Devices ............................................................................................................................ 3412
Nitendra Rajput, Amit Anil Nanavati, Abhishek Kumar, Neeraj Chaudhary

Clarification Questions to Improve Dialogue Flow and Speech Recognition in Spoken Dialogue Systems .............................................. 3416
Ulf Krum, Hartwig Holzapfel, Alex Waibel

Speech Interface for Controlling an Hi-Fi Audio System Based on a Bayesian Belief Networks Approach for Dialog Modeling ........... 3420
Fernando Fernandez, Javier Ferreiros, Valentin Sama, Juan Manuel Montero, Ruben San Segundo, Javier Macias-Guarasa, Rafael Garcia

SESSION: SPOKEN LANGUAGE UNDERSTANDING II

Hierarchical Language Models for One-Stage Speech Interpretation ................................................................................................. 3424
Matthias Thomae, Tibor Fabian, Robert Lieb, Gunther Ruske

Spoken Language Understanding Using Layered N-gram Modeling .................................................................................................. 3428
Nick J.C. Wang

Named Entity Recognition from Spontaneous Open-Domain Speech ................................................................................................. 3432
Mihai Surdeanu, Jordi Turmo, Eli Comelles

Discriminative Training and Support Vector Machine for Natural Language Call Routing ......................................................................... 3436
Imed Zitouni, Hui Jiang, Qiru Zhou

A Multiple Classifier-Based Concept-Spotting Approach for Robust Spoken Language Understanding ........................................ 3440
Jihyun Eun, Minwoo Jeong, Gary Geunbae Lee

A Flexible and Integrated Interface Between Speech Recognition, Speech Interpretation and Dialog Management ...................... 3444
Robert Lieb, Matthias Thomae, Gunther Ruske, Daniel Bobbert, Frank Althoff

Incremental Dependency Parsing of Japanese Spoken Monologue Based on Clause Boundaries ......................................................... 3448
Tomohiro Ohno, Shigeki Matsubara, Hideki Kashioka, Naoto Kato, Yasuyoshi Inagaki

Situation Based Speech Recognition for Structuring Baseball Live Games .................................................................................. 3452
Atsushi Sako, Tetsuya Takiguchi, Yasuo Ariki

Semantic Annotation of the French Media Dialog Corpus .............................................................................................................. 3456
H. Bonneau-Maynard, Sophie Rosset, C. Ayache, A. Kuhn, Djamel Mostefa

Robust and Efficient Semantic Parsing of Free Word Order Languages in Spoken Dialogue Systems ................................................... 3460
Ralf Engel

Conceptual Language Model Design for Spoken Language Understanding .................................................................................. 3464
Catherine Kobus, Geraldine Damnati, Lionel Delphine-Poulat, Renato de Mori
From Robust Spoken Language Understanding to Knowledge Acquisition and Management .......................................................... 3468
  Luis Seabra Lopes, Antonio J.S. Teixeira, Marcelo Quindere, Mario Rodrigues

Improving End-to-End Performance of Call Classification Through Data Confusion Reduction and Model Tolerance Enhancement .......................................................... 3472
  Cheng Wu, Xiang Li, Hong-Kwang Jeff Kuo, E.E. Jan, Vaibhava Goel, David Lubensky

Author Index