# TABLE OF CONTENTS

## VOLUME 1

### KEYNOTE I

Still Talking to Machines (Cognitively Speaking) ................................................................. 1  
Steve Young

### KEYNOTE II

Sound-Based Assistive Technology Supporting "Seeing", "Hearing" and "Speaking" for the Disabled and the Elderly ................................................................. 11  
Tohru Ifukube

### KEYNOTE III

Beyond Sentence Prosody ........................................................................................................ 20  
Chia-Yu Tseng

### MON-SES2-S1: SPECIAL SESSION: MODELS OF SPEECH – IN SEARCH OF BETTER REPRESENTATIONS

A Procedure for Estimating Gestural Scores from Natural Speech ............................................ 30  
Hosang Nam, Vikramjit Mitra, Mark Tiede, Elliot Saltzman, Louis Goldstein, Carol Eddy-Wilson, Mark Hasegawa-Johnson

On the Interdependencies Between Voice Quality, Glottal Gaps, and Voice-Source Related Acoustic Measures ................................................................. 34  
Yen-Liang Shue, Gang Chen, Abeer Alwan

Simplification and Extension of Non-Periodic Excitation Source Representations for High-Quality Speech Manipulation Systems ......................................................... 38  
Hideki Kawahara, Munanori Morese, Toru Takahashi, Hideki Banno, Ryuichi Nisimura, Toshio Iri

Phase Equalization-Based Autoregressive Model of Speech Signals ....................................... 42  
Sadakou Hiyoyo, Takemi Mochida

Articulatory-Functional Modeling of Speech Prosody: A Review ........................................... 46  
Yu Xa, Santitham Prom-On

Two New Estimation Methods for a Superpositional Intonation Model .................................... 50  
Humberto M. Torrez, Hansjorg Misdorff, Jorge A. Gurlekian, Hartmut R. Pfitzinger

### MON-SES2-O1: ASR: ACOUSTIC MODELS I

A Discriminative Splitting Criterion for Phonetic Decision Trees ........................................ 54  
Simon Wiesler, Georg Heigold, Markus Nussbaum-Thom, Ralf Schluter, Hermann Ney

Canonical State Models for Automatic Speech Recognition .................................................... 58  
M. J. F. Gales, Kai Yu

Restructuring Exponential Family Mixture Models ................................................................. 62  
Pierre L. Dognin, John R. Hershey, Vaibhava Goel, Peder Olsen

Unsupervised Discovery and Training of Maximally Dissimilar Cluster Models ....................... 66  
Francoise Beaufays, Vincent Vanhoucke, Brian Strope

Probabilistic State Clustering Using Conditional Random Field for Context-Dependent Acoustic Modelling .......................................................... 70  
Khe Chai Sim

Integrate Template Matching and Statistical Modeling for Speech Recognition ........................ 74  
Xie Sun, Yunxin Zhao

### MON-SES2-O2: SPOKEN DIALOGUE SYSTEMS I

Cross-Lingual Spoken Language Understanding from Unaligned Data Using Discriminative Classification Models and Machine Translation ................................................................. 78  
Fabrice Lefevre, Francois Mairesse, Steve Young

Techniques for Topic Detection Based Processing in Spoken Dialog Systems .......................... 82  
Brajesh Balchandran, Leonid Bachevry, Bhuvana Ramabhadran, Miroslav Novak

Optimizing Spoken Dialogue Management with Fitted Value Iteration ................................. 86  
Senthilkumar Chandramohan, Matthias Geiss, Olivier Pitequin

Natural Belief-Critic: A Reinforcement Algorithm for Parameter Estimation in Statistical Spoken Dialogue Systems .............................................................. 90  
F. Jurcicek, B. Thomson, S. Keizer, Francois Mairesse, M. Gasic, Kai Yu, Steve Young

Is It Possible to Predict Task Completion in Automated Troubleshooters? ............................. 94  
Alexander Schmitt, Michael Scholz, Wolfgang Minker, Jackson Liscombe, David Suerdernmann
Minimally Invasive Surgery for Spoken Dialog Systems

David Suendermann, Jackson Liscombe, Roberto Pieraccini

MON-SES2-P3: SPEECH PERCEPTION I: FACTORS INFLUENCING PERCEPTION

Detecting Categorical Perception in Continuous Discrimination Data

Paul Boersma, Katerina Chladkova

The Interrelation Between the Stimulus Range and the Number of Response Categories in Vowel Categorization

Tita Benders, Paola Escudero

The Relation Between Pitch Perception Preference and Emotion Identification

Marie Nilsenova, Martijn Goudbeeck, Luuk Kempen

Competition in the Perception of Spoken Japanese Words

Takashi Onake, James M. McQueen, Anne Collier

Influence of Musical Training on Perception of L2 Speech

Makiko Sadakata, Lotte Van Der Zanden, Kaoru Sekiyama

Full Body Aero-Tactile Integration in Speech Perception

Donald Derrick, Bryan Gick

MON-SES2-P3-04: PROSODY: MODELS

Nucleus Position Within the Intonation Phrase: A Typological Study of English, Czech and Hungarian

Tomas Dubedat, Katalin Mady

Focus-Sensitive Operator or Focus Inducer: Always and Only

Yong-Chool Lee, Satoshi Nambu

F0 Declination in English and Mandarin Broadcast News Speech

Jiahong Yuan, Mark Liberman

Frequency of Occurrence Effects on Pitch Accent Realisation

Katrin Schweitzer, Michael Walsh, Bernd Mohias, Hinrich Schutze

On the Automatic ToBI Accent Type Identification from Data

Cesar Gonzalez-Ferreras, Carlos Vivaracho-Pascual, David Escudero-Mancebo, Valentin Cardenoso-Payo

AuToBI --- A Tool for Automatic ToBI Annotation

Andrew Rosenberg

MON-SES2-P3-P1: SPEECH SYNTHESIS I: UNIT SELECTION AND OTHERS

A Classifier-Based Target Cost for Unit Selection Speech Synthesis Trained on Perceptual Data

Volker Strom, Simon King

Applying Scalable Phonetic Context Similarity in Unit Selection of Concatenative Text-to-Speech

Wei Zhang, Xiaodong Cui

Speech Database Reduction Method for Corpus-Based TTS System

Mitsuaki Igai, Hileyi Miyato

Automatic Error Detection for Unit Selection Speech Synthesis Using Log Likelihood Ratio Based SVM Classifier

Heng Lu, Zhen-Hua Ling, Si Wei, Lizong Dai, Ren-Hua Wang

Using Robust Viterbi Algorithm and HMM-Modeling in Unit Selection TTS to Replace Units of Poor Quality

Hanna Silen, Elina Helander, Jami Nurminen, Konsta Koppinen, Moncef Gabbouj

Automatic Detection of Abnormal Stress Patterns in Unit Selection Synthesis

Yeom-Jun Kim, Mark C. Beutnagel

Enhancements of Viterbi Search for Fast Unit Selection Synthesis

Daniel Tirolka, Jiri Kala, Jindrich Matousek

Accurate Pitch Marking for Prosodic Modification of Speech Segments

Thomas Esvender, Beat Pfister

A Novel Hybrid Approach for Mandarin Speech Synthesis

Shifeng Pan, Meng Zhang, Jianhua Tao

Modeling Liaison in French by Using Decision Trees

Josefa De Jesus Aguiar Pontes, Sadaoki Furui

Improvement on Plural Unit Selection and Fusion

Jian Luan, Jian Li

Improving Speech Synthesis of Machine Translation Output

Alok Parikh, Alan W. Black, Stephan Vogel

Paraphrase Generation to Improve Text-to-Speech Synthesis

Ghislain Putois, Jonathan Chevelu, Cedric Boidin

MON-SES2-P3-P2: ASR: SEARCH, DECODING AND CONFIDENCE MEASURES I

Phone Mismatch Penalty Matrices for Two-Stage Keyword Spotting via Multi-Pass Phone Recognizer

Chang Woo Han, Shin Jae Kang, Chul Min Lee, Nam Soo Kim
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Spoken Term Detection in Multilingual Recordings</td>
<td>206</td>
</tr>
<tr>
<td>A Hybrid Approach to Robust Word Lattice Generation via Acoustic-Based Word Detection</td>
<td>210</td>
</tr>
<tr>
<td>Direct Observation of Pruning Errors (DOPE): A Search Analysis Tool</td>
<td>214</td>
</tr>
<tr>
<td>Direct Construction of Compact Context-Dependency Transducers from Data</td>
<td>218</td>
</tr>
<tr>
<td>Incremental Composition of Static Graphs with Label Pushing</td>
<td>222</td>
</tr>
<tr>
<td>A Novel Path Extension Framework Using Steady Segment Detection for Mandarin Speech Recognition</td>
<td>226</td>
</tr>
<tr>
<td>On the Relation of Bayes Risk, Word Error, and Word Posteriors in ASR</td>
<td>230</td>
</tr>
<tr>
<td>Time Conditioned Search in Automatic Speech Recognition Reconsidered</td>
<td>234</td>
</tr>
<tr>
<td>Efficient Data Selection for Speech Recognition Based on Prior Confidence Estimation Using Speech and Context Independent Models</td>
<td>238</td>
</tr>
<tr>
<td>A Novel Confidence Measure Based on Marginalization of Jointly Estimated Error Cause Probabilities</td>
<td>242</td>
</tr>
<tr>
<td>Evaluation of a Silent Speech Interface Based on Magnetic Sensing</td>
<td>246</td>
</tr>
<tr>
<td>Advanced Speech Communication System for Deaf People</td>
<td>250</td>
</tr>
<tr>
<td>Unsupervised Acoustic Model Adaptation for Multi-Origin Non Native ASR</td>
<td>254</td>
</tr>
<tr>
<td>Speech-Based Automated Cognitive Status Assessment</td>
<td>258</td>
</tr>
<tr>
<td>Speech Recognition with a Seamlessly Updated Language Model for Real-Time Closed-Captioning</td>
<td>262</td>
</tr>
<tr>
<td>The Comparison Between the Deletion-Based Methods and the Mixing-Based Methods for Audio CAPTCHA Systems</td>
<td>266</td>
</tr>
<tr>
<td>Comparing Mono- &amp; Multilingual Acoustic Seed Models for a Low e-Resourced Language: A Case-Study of Luxembourgish</td>
<td>270</td>
</tr>
<tr>
<td>Manipulating TreacleophagelSpeech</td>
<td>274</td>
</tr>
<tr>
<td>Towards Mixed Language Speech Recognition Systems</td>
<td>278</td>
</tr>
<tr>
<td>Voice Search for Development</td>
<td>282</td>
</tr>
<tr>
<td>Cross-Cultural Investigation of Prosody in Verbal Feedback in Interactional Rapport</td>
<td>286</td>
</tr>
<tr>
<td>Multimodal Speaker Diarization Using Oriented Optical Flow Histograms</td>
<td>290</td>
</tr>
<tr>
<td>Towards an ASR-Free Objective Analysis of Pathological Speech</td>
<td>294</td>
</tr>
<tr>
<td>Session Variability Contrasts in the MARP Corpus</td>
<td>298</td>
</tr>
<tr>
<td>Estimation of Two-to-One Forced Selection Intelligibility Scores by Speech Recognizers Using Noise-Adapted Models</td>
<td>302</td>
</tr>
<tr>
<td>Analysis of Gender Normalization Using MLP and VTLN Features</td>
<td>306</td>
</tr>
<tr>
<td>Discovering an Optimal Set of Minimally Contrasting Acoustic Speech Units: A Point of Focus for Whole-Word Pattern Matching</td>
<td>310</td>
</tr>
<tr>
<td>Improvements to the Equal-Parameter BIC for Speaker Diarization</td>
<td>314</td>
</tr>
<tr>
<td>A Multistream Multiresolution Framework for Phoneme Recognition</td>
<td>318</td>
</tr>
</tbody>
</table>
Cluster Analysis of Differential Spectral Envelopes on Emotional Speech .......................................................... 322
Gianpietro Salvi, Fabio Tessier, Derico Zuvato, Piero Cossi
Modeling Pronunciation Variation with Context-Dependent Articulatory Feature Decision Trees .......................... 326
Sam Bowman, Karen Livescu
Ungrounded Independent Non-Negative Factor Analysis ............................................................................... 330
Bhiksha Raj, Kevin W. Wilson, Alexander Krueger, Reinhold Haeb-Umbach
Signal Interaction and the Devil Function ......................................................................................... 334
John R. Hershey, Peder Olsen, Steven J. Rennie

MON-SES3-O1: SYSTEMS FOR LVCSR

Semi-Automated Update of Automatic Transcription System for the Japanese National Congress .................. 338
Yu Yu, Akio, Masato Mimura, Graham Neubig, Tatsuya Kawahara
Language Model Cross Adaptation for LVCSR System Combination .................................................. 342
X. Liu, M. J. F. Gales, P. C. Woodland
Large Vocabulary Continuous Speech Recognition Using WFST-Based Linear Classifier for Structured Data ..... 346
Shinji Watanabe, Takaaki Hori, Atsushi Nakamura
Accelerating Hierarchical Acoustic Likelihood Computation on Graphics Processors .............................. 350
Pavel Kveton, Miroslav Novak
Search by Voice in Mandarin Chinese ........................................................................................................ 354
Jiadong Shan, Genqiong Wu, Zhilong Hu, Xifu Tang, Minlin Jancsche, Pedro J. Moreno
The AMIDA 2009 Meeting Transcription System ................................................................................ 358
Thomas Hain, Lukas Burget, John Dines, Philip N. Garner, Asmaa El Hannani, Marijn Huijbregts, Martin Karafiat, Mike Lincoln, Vincent Wan

MON-SES3-O2: SPEAKER CHARACTERIZATION AND RECOGNITION I

Simple and Efficient Speaker Comparison Using Approximate KL Divergence ................................................. 362
W. M. Campbell, Zahi N. Karam
The HR NIST SRE 2008 and 2010 Summed Channel Speaker Recognition Systems ...................................... 366
Hanyu Sun, Bin Ma, Chien-Lin Huang, Trung Hieu Nguyen, Haizhou Li
Speaker Characterization Using Long-Term and Temporal Information .................................................. 370
Chien-Lin Huang, Hanyu Sun, Bin Ma, Haizhou Li
Score-Level Compensation of Extreme Speech Duration Variability in Speaker Verification .......................... 374
Sergio Perez-Gomez, Daniel Ramos, Javier Gonzalez-Dominguez, Joaquin Gonzalez-Rodriguez
Speaker Recognition Experiments Using Connectionist Transformation Network Features .................. 378
Alberto Abad, Isabel Trancoso
Speaker Recognition Using Supervised Probabilistic Principal Component Analysis .................................. 382
Yun Lei, John H. L. Hansen

MON-SES3-O3: SOURCE SEPARATION

A Factorial Sparse Coder Model for Single Channel Source Separation ....................................................... 386
Robert Peharz, Michael Stark, Franz Pernkopf, Yannis Stylianou
Oriented PCA Method for Blind Speech Separation of Convolutative Mixtures ............................................. 390
Yasmina Benabderrahmane, Sid Ahmed Selouani, Douglas O'Shaughnessy
Online Gaussian Process for Nonstationary Speech Separation ........................................................................ 394
Hsin-Lung Hsieh, Jen-Tseng Chien
Convexity and Fast Speech Extraction by Split Bregman Method ............................................................. 398
Meng Yu, Wenyue Yu, Jack Xin, Stanley Ohler
Reducing Musical Noise in Blind Source Separation by Time-Domain Sparse Filters and Split Bregman Method 402
Wenyue Ma, Meng Yu, Jack Xin, Stanley Ohler
Combining Monaural and Binaural Evidence for Reverberant Speech Segregation ...................................... 406
John Woodruff, Rohit Prabhavalkar, Eric Fosler-Lussier, Deliang Wang

MON-SES3-O4: SPEECH SYNTHESIS II: HMM-BASED SPEECH SYNTHESIS

Speaker and Language Adaptive Training for HMM-Based Polyglot Speech Synthesis .................................. 410
Heiga Zen
Context Adaptive Training with Factorized Decision Trees for HMM-Based Speech Synthesis ...................... 414
Kai Yu, Heiga Zen, François Mairese, Steve Young
Roles of the Average Voice in Speaker-Adaptive HMM-Based Speech Synthesis ........................................ 418
Junichi Yamagishi, Oliver Watts, Simon King, Bela Usabaev
An HMM Trajectory Tiling (HTT) Approach to High Quality TTS .............................................................. 422
Yao Qian, Zhi-Jie Yan, Yijian Wu, Frank K. Soong, Xin Zhuang, Shengyi Kong
A Perceptual Study of Acceleration Parameters in HMM-Based TTS ......................................................... 426
Yi-Ning Chen, Zhi-Jie Yan, Frank K. Soong
Evaluation of Prosodic Contextual Factors for HMM-Based Speech Synthesis ................................................................. 430
Shoji Yokomizo, Takashi Nose, Takao Kobayashi

MON-SES3-O5: MULTI-MODAL SIGNAL PROCESSING

Learning Words and Speech Units Through Natural Interactions .................................................................................. 434
Jonas Hornstein, Jose Santos-Victor

Bimodal Coherence Based Scale Ambiguity Cancellation for Target Speech Extraction and Enhancement ................. 438
Qingia Liu, Wenwu Wang, Philip Jackson

Speech Estimation in Non-Stationary Noise Environments Using Timing Structures Between Mouth Movements and Sound Signals .................................................................................................................. 442
Hiroaki Kawashima, Yu Horii, Takashi Matsuyama

Synthesizing Photo-Real Talking Head via Trajectory-Guided Sample Selection ......................................................... 446
Lijuan Wang, Xiaojun Qian, Wei Han, Frank K. Soong

Silent vs Vocalized Articulation for a Portable Ultrasound-Based Silent Speech Interface ........................................... 450
Victoria M. Floreescu, Lise Crever-Buchman, Bruce Denby, Thomas Hueber, Antonia Colazo-Simon, Claire Pilloy-Loiseau, Pierre Roussel, Cedric Gendrot, Sophie Quattrocchi

Comparison of HMM and TMDN Methods for Lip Synchronisation ............................................................................. 454
Gregor Hoyer, Korin Richmond

MON-SES3-P1: PARALANGUAGE

Rhythm and Formant Features for Automatic Alcohol Detection ................................................................................ 458
Florian Schiel, Christian Heinrich, Veronika Neumeyer

An Exploration of Voice Source Correlates of Focus ................................................................................................. 462
Irena Yanashkevskaya, Chiister Gobl, John Kane, Aihhhe Ni Chasaide

Modeling Perceived Vocal Age in American English .............................................................................................. 466
James D. Harnsberger, Rahul Shivasthan, W. S. Brown Jr.

Multivariate Analysis of Vocal Fatigue in Continuous Reading .................................................................................. 470
Marie Jose Caraty, Claude Montacut

Frequency-Domain Delexicalization Using Surrogate Vowels .................................................................................... 474
Alexander Kain, Jan P. H. Van Santen

Emotion Recognition Using Imperfect Speech Recognition ...................................................................................... 478
Florian Metze, Anton Batliner, Florian Eyben, Tim Polzelh, Bjorn Schuller, Stefan Steinld

A Novel Feature Extraction Strategy for Multi-Stream Robust Emotion Identification ............................................. 482
Gang Liu, Yan Lei, John H. L. Hansen

Setup for Acoustic-Visual Speech Synthesis by Concatenating Bimodal Units ............................................................ 486
Aserton Tautut, Utpala Mastra, Simon Oost, Vincent Colotto, Brigitte Wrobel-Dautcourt, Marie-Odile Berger

Towards Affective State Modeling in Narrative and Conversational Settings ............................................................ 490
Bart Jochems, Martha Larson, Roeland Ordelman, Ronald Pappe, Khiet P. Truong

Detection of Anger Emotion in Dialog Speech Using Prosody Feature and Temporal Relation of Utterances ............ 494
Narichika Nomoto, Hirokazu Masatani, Osamu Yoshioka, Satoshi Takahashi

Gesture and Speech Coordination: The Influence of the Relationship Between Manual Gesture and Speech ............. 498
Benjamin Rostan, Marvin Dohen

Analysis and Detection of Cognitive Load and Frustration in Drivers' Speech .............................................................. 502
Hynek Boril, Seved Omud Sadjadi, Tristan Kleinschmidt, John H. L. Hansen

Acoustic-Based Recognition of Head Gestures Accompanying Speech ................................................................... 506
Aki Sasaou, Yashu Bhamasho, Katsukih Okaue

Multimodal Dialog in the Car: Combining Speech and Turn-and-Push Dial to Control Comfort Functions ............... 510
Sandro Castronovo, Angela Mahr, Margarita Penchueva, Christian Muller

Hands Free Audio Analysis from Home Entertainment ............................................................................................... 514
Dami Korchugan, Philip N. Garner, Petr Modicck

Affective Story Teller: A TTS System for Emotional Expressivity ............................................................................. 518
Mostafa Al Masum Shaikh, Antonio Rui Ferreira Rebdarao, Keikichi Hirose

MON-SES3-P2: ASR: SPEAKER ADAPTATION, ROBUSTNESS AGAINST REVERBERATION

Enhancing Children's Speech Recognition Under Mismatched Condition by Explicit Acoustic Normalization ............ 522
Shweta Gai, Rohit Sinha

Comparison of Discriminative Input and Output Transformations for Speaker Adaptation in the Hybrid NN/HMM Systems .................................................................................................................. 526
Bo Li, Khe Chai Sim

Augmentation of Adaptation Data .............................................................................................................................. 530
Bavishander Uppurlda, Steve Renals, Joe Frankel

Discriminative Adaptation Based on Fast Combination of DMAP and DMLLR .............................................................. 534
Lukas Matclich, Zhyn Lac, Jude Muller

Revisiting VTLN Using Linear Transformation on Conventional MFCC ...................................................................... 538
D. R. Sanand, Ralf Schuster, Hermann Ney
Speaker Adaptation Based on Nonlinear Spectral Transform for Speech Recognition .......................................................... 542
Speaker Adaptation Based on System Combination Using Speaker-Classifier Models .............................................................. 546
Speaker Adaptation in Transformation Space Using Two-Dimensional PCA ........................................................................ 550
On Speaker Adaptive Training of Artificial Neural Networks .................................................................................................. 554
Model Synthesis for Band-Limited Speech Recognition .............................................................................................................. 558
Performance Estimation of Reverberant Speech Recognition Based on Reverberant Criteria RSR-Ds with Acoustic Parameters ........................................................................................................... 562
A Novel Approach for Matched Reverberant Training of HMMs Using Data Pairs ................................................................. 566
An Auditory Based Modulation Spectral Feature for Reverberant Speech Recognition .............................................................. 570
On the Potential of Channel Selection for Recognition of Reverberated Speech with Multiple Microphones .............................. 574
Methods for Robust Speech Recognition in Reverberant Environments: A Comparison ............................................................ 582
MON-SES3-P3: LANGUAGE LEARNING, TTS, AND OTHER APPLICATIONS

Integration of Multilayer Regression Analysis with Structure-Based Pronunciation Assessment .................................................... 586
Using Non-Native Error Patterns to Improve Pronunciation Verification .................................................................................... 590
Regularized-MLLR Speaker Adaptation for Computer-Assisted Language Learning System ...................................................... 594
Decision Tree Based Tone Modeling with Corrective Feedbacks for Automatic Mandarin Tone Assessment .................................. 602
CASTLE: A Computer-Assisted Stress Teaching and Learning Environment for Learners of English as a Second Language .................................................................................................................... 606
Automatic Reference Independent Evaluation of Prosody Quality Using Multiple Knowledge Fusions ..................................... 610
Landmark-Based Automated Pronunciation Error Detection .................................................................................................... 614
HMM Based TTS for Mixed Language Text ............................................................................................................................... 618
An Analysis of Language Mismatch in HMM State Mapping-Based Cross-Lingual Speaker Adaptation ................................. 622
Classroom Note-Taking System for Hearing Impaired Students Using Automatic Speech Recognition Adapted to Lectures ............................................................................................................................ 626
Exploring Web-Browser Based Runtimes Engines for Creating Ubiquitous Speech Interfaces .................................................. 630
MON-SES3-P4: PITCH AND GLOTTAL-WAVEFORM ESTIMATION AND MODELING I

Efficient Three-Stage Pitch Estimation for Packet Loss Concealment ...................................................................................... 633
On Evaluation of the F0 Estimation Based on Time-Varying Complex Speech Analysis ............................................................... 637
Pitch Estimation in Noisy Speech Based on Temporal Accumulation of Spectrum Peaks .............................................................. 641
Multi-Pitch Estimation by a Joint 2-D Representation of Pitch and Pitch Dynamics .................................................................... 645
On the Effect of Fundamental Frequency on Amplitude and Frequency Modulation Patterns in Speech Resonances ................................................... 649
Pitch Determination Using Autocorrelation Function in Spectral Domain .................................................................................. 653
Christ Complex Cepstrum-Based Decomposition for Asynchronous Glottal Analysis .......................................................... 657
Thomas Drugman, Thierry Dutoit

Exploiting Glottal Formant Parameters for Glottal Inverse Filtering and Parameterization .......................................................... 661
Alan O’Connelde, David Dorrnan, Mikkel Gainza, Eugene Coyde

Glottal Parameters Estimation on Speech Using the Zeros of the Z-Transform .......................................................... 665
Nicolas Sturmel, Christophe D’Alessandro, Boris Doval

Significance of Pitch Synchronous Analysis for Speaker Recognition Using AANN Models .......................................................... 669
Sri Harth Reddy M., Kishore Pradhahlay, Suryakanth V. Ganagshetty, B. Tegnanarayana

On Using Voice Source Measures in Automatic Gender Classification of Children’s Speech .......................................................... 673
Gong Chen, Xue Feng, Yen-Liang Shue, Abder Elwan

TUE-SES1-O1: ROBUST ASR

Asymptotically Exact Noise-Corrupted Speech Likelihoods .................................................................................. 709
R. C. Van Dallen, M. J. F. Gates

A MMSE Estimator in Mel-Cepstral Domain for Robust Large Vocabulary Automatic Speech Recognition Using Uncertainty Propagation .................................................................................. 713
Ramón Fernández Astudillo, Reinhold Orglmeister

Non-Negative Matrix Factorization Based Compensation of Music for Automatic Speech Recognition .................................................................................. 717
Rakesh Astudillo, Tuomas Virtanen, Sourish Chaudhuri, Rita Singh

Feature versus Model Based Noise Robustness .......................................................................................... 721
Kris Demeyney, Xuern Zhang, Dirk Van Compernolle, Hugo Van Hamme

SNR-Based Mask Compensation for Computational Auditory Scene Analysis Applied to Speech Recognition in a Car Environment .......................................................................................... 725
Ji Hun Park, Seon Man Kim, Jae Sam Yoon, Hong Kook Kim, Sang Joo Lee, Yankun Lee

Automatic Selection of Thresholds for Signal Separation Algorithms Based on Interaural Delay .................................................................................. 729
Chanwoo Kim, Richard M. Stern, Kiwan Eom, Jaewon Lee

TUE-SES1-O2: LANGUAGE AND DIALECT IDENTIFICATION

Channel Detectors for System Fusion in the Context of NIST LRE 2009 .................................................................................. 733
Florian Verdet, Diiss Matrouf, Jean-Francois Bonastre, Jean Hennebert

Selecting Phonotactic Features for Language Recognition .................................................................................. 737
Rong Tong, Bin Ma, Haizhou Li, Eng Song Chng

Improved Language Recognition Using Mixture Components Statistics .................................................................................. 741
Abuabdou Hanani, Michael Carey, Martin J. Ruxell

Using Cross-Decoder Co-Occurrences of Phone N-Grams in SVM-Based Phonotactic Language Recognition .................................................................................. 745
Mikel Penagarikano, Amparo Varona, Luis Javier Rodriguez-Fuentes, German Bordel

Exploiting Variety-Dependent Phones in Portuguese Variety Identification Applied to Broadcast News Transcription .................................................................................. 749
Oscar Koller, Alberto Aham, Isabel Tereanoses, Cel Viana

Dialect Recognition Using a Phone-GMM-Supervector-Based SVM Kernel .................................................................................. 753
Fadi Badrta, Julia Hirschberg, Michael Collins
**TUE-SES1-O3: TECHNOLOGIES FOR LEARNING AND EDUCATION**

- Discriminative Acoustic Model for Improving Mispronunciation Detection and Diagnosis in Computer-Aided Pronunciation Training (CAPT) ................................................................. 757
  Xiaojun Qian, Frank K. Soong, Helen Meng
- Automatic Pronunciation Scoring Using Learning to Rank and DP-Based Score Segmentation ....................................................... 761
  Liang-Yu Chen, Jyh-Shing Roger Jang
- Automatic Derivation of Phonological Rules for Mispronunciation Detection in a Computer-Assisted Pronunciation Training System ........................................................................................................... 765
  Wai-Kit Lo, Shuang Zhang, Helen Meng

**VOLUME 2**

- Adapting a Duration Synthesis Model to Rate Children’s Oral Reading Prosody ............................................................................... 769
  Monh Dong, Jack Mostow
- Predicting Word Accuracy for the Automatic Speech Recognition of Non-Native Speech ................................................................. 773
  Su-Youn Yoon, Lei Chen, Klaus Zeichner
- A New Approach for Automatic Tone Error Detection in Strong Accented Mandarin Based on Dominant Set ................................. 777
  Taotao Zhu, Dengfeng Ke, Zhenbiao Chen, Bo Xu

**TUE-SES1-O4: EMOTIONAL SPEECH**

- Analysis of Excitation Source Information in Emotional Speech ................................................................. 781
  S. R. M. Prasanna, D. Govind
- Acoustic Feature Analysis in Speech Emotion Primitives Estimation ........................................................................................................ 785
  Yoshinori Shiga, Tomoki Toda, Shinsuke Sakai, Hisashi Kawai
- Spectro-Temporal Modulations for Robust Speech Emotion Recognition ................................................................................................. 789
  Lan-Ying Yeh, Tai-Shih Chi
- Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples ................................. 793
  Chi-Chun Lee, Matthew Black, Athanasios Katsamanis, Adam C. Lammert, Brian R. Baucom, Andrew Christensen, Panayiotis G. Georgiou, Shravan Narayanan
- A Cluster-Profile Representation of Emotion Using Agglomerative Hierarchical Clustering .................................................................. 797
  Emily Mower, Kya J. Han, Sanghok Lee, Shravan S. Narayanan
- Incremental Acoustic Valence Recognition: An Inter-Corpus Perspective on Features, Matching, and Performance in a Gating Paradigm ................................................................................. 801
  Bjorn Schuller, Laurence Devillers

**TUE-SES1-P1: SPEECH SYNTHESIS III: HMM-BASED SPEECH SYNTHESIS**

- Sinusoidal Model Parameterization for HMM-Based TTS System ........................................................................................................ 805
  Slava Shechtman, Alex Sarin
- Improved Training of Excitation for HMM-Based Parametric Speech Synthesis ................................................................. 809
  Yoshinori Shiga, Tomoki Toda, Shinsuke Sakai, Hisashi Kawai
- Excitation Modeling Based on Waveform Interpolation for HMM-Based Speech Synthesis ................................................................. 813
  June Sik Sung, Doo Hwa Hong, Kyung Hwan Oh, Nam Soo Kim
- Formant-Based Frequency Warping for Improving Speaker Adaptation in HMM TTS ................................................................. 817
  Xin Zhuang, Yao Qian, Frank K. Soong, Yijian Wu, Bo Zhang
- Improved Modelling of Speech Dynamics Using Non-Linear Formant Trajectories for HMM-Based Speech Synthesis .................. 821
  Hongwei Hu, Martin J. Rassell
- Global Variance Modeling on the Log Power Spectrum of LSPs for HMM-Based Speech Synthesis ........................................ 825
  Zhen-Hua Ling, Yu Hu, Livong Dai
- Autoregressive Clustering for HMM Speech Synthesis ........................................................................................................ 829
  Matt Shannon, William Byrne
- An Implementation of Decision Tree-Based Context Clustering on Graphics Processing Units ......................................................... 833
  Nicholas Plunkington, Heiga Zen
- Quantized HMMs for Low Footprint Text-to-Speech Synthesis ........................................................................................................ 837
  Alexander Gulin, Xavi Gonzalvo, Stefan Breuer, Paul Taylor
- The Role of Higher-Level Linguistic Features in HMM-Based Speech Synthesis ......................................................................................... 841
  Oliver Watts, Junichi Yamagishi, Simon King
- HMM-Based Singing Voice Synthesis System Using Pitch-Shifted Pseudo Training Data ................................................................. 845
  Ayumi Mae, Keichiro Oura, Yoshihiko Nankaku, Keichiro Tokuda
- An Unsupervised Approach to Creating Web Audio Contents-Based HMM Voices ................................................................. 849
  Jieju Ni, Hisashi Kawai
- Conversational Spontaneous Speech Synthesis Using Average Voice Model ......................................................................................... 853
  Tomoki Koriyama, Takashi Nose, Takao Kobayashi
TUE-SES1-P2: NEW PARADIGMS IN ASR I

Mandarin Digit Recognition Assisted by Selective Tone Distinction ......................................................... 857
Xiao-Dong Wang, Kunihiko Owa, Mokoto Shozakai

Brazilian Portuguese Acoustic Model Training Based on Data Borrowing from Other Language .................. 861
Kazuhiko Abe, Sakriami Sakhi, Royuske Isotani, Hitashi Kawai, Satoshi Nakamura

Rapid Bootstrapping of Five Eastern European Languages Using the Rapid Language Adaptation Toolkit ........ 865
Ngoc Thang Yu, Tim Schlippe, Franziska Krauss, Tanja Schultz

Cross-Lingual Speaker Adaptation via Gaussian Component Mapping ......................................................... 869
Huenwei Cuo, Tan Lee, P. C. Ching

Cross-Lingual Acoustic Modeling for Dialectal Arabic Speech Recognition ................................................. 873
Mohamed Elmhady, Rainer Grahn, Wolfgang Misker, Slim Abdenaddher

Cross-Lingual and Multi-Stream Posterior Features for Low Resource LVCSR Systems .............................. 877
Samuel Thomas, Sreeram Ganapathy, Hynek Hermansky

Shiva Sundaram, Jerome R. Bellogarda

Unsupervised Model Adaptation on Targeted Speech Segments for LVCSR System Combination .................. 885

Incremental Word Learning Using Large-Margin Discriminative Training and Variance Floor Estimation ........ 889
Irene Aylin Clemente, Martin Heckmann, Alexander Denecke, Britta Wrede, Christian Goerick

State-Based Labelling for a Sparse Representation of Speech and Its Application to Robust Speech Recognition 893
Tuomas Virtanen, Jori F. Gemmeke, Antti Harmalainen

Similarity Scoring for Recognizing Repeated Out-of-Vocabulary Words ..................................................... 897
Mirko Hannemann, Stefan Kombrink, Martin Karafiát, Lukás Burger

Data Pruning for Template-Based Automatic Speech Recognition ................................................................. 901
Dino Seppi, Dirk Van Compernolle

TUE-SES1-P3: SPEECH PRODUCTION I: VARIOUS APPROACHES

Speaking Style Dependency of Formant Targets ............................................................................................ 905
Atiko Amano-Kusumoto, John-Paul Hosom, Alexander Kain

Similarity of Effects of Emotions on the Speech Organ Configuration with and without Speaking .................. 909
Tatsuya Kitamura

A Study of Intra-Speaker and Inter-Speaker Affective Variability Using Electroglossograph and Inverse Filtered 913
Glottal Waveforms ........................................................................................................................................

Modal Analysis of Vocal Fold Vibrations Using Laryngotopography ............................................................ 917
Ken-Ichi Sakahara, Hiroshi Imagawa, Misako Kimura, Hisayuki Tokonishi, Niro Tayama

Laryngeal Voice Quality in the Expression of Focus ....................................................................................... 921
Martti Vainio, Matti Airus, Juhani Jarvihovi, Paavo Alku

Laryngeal Characteristics During the Production of Geminate Consonants .................................................. 925
Masako Fujimoto, Kiko Maekawa, Seiya Funatsu

Numerical Study of Turbulent Flow-Induced Sound Production in Presence of a Tooth-Shaped Obstacle: Towards 929
Sibilant [s] Physical Modeling ..............................................................................................................

Julien Cicioni, Kazunori Nozaki, Annemie Van Hirtum, Shigeo Wada

Morphological and Predictability Effects on Schwa Reduction: The Case of Dutch Word-Initial Syllables .......... 933
Iris Hanique, Barbara Schuppeler, Mirjam Ernestus

Acoustic-to-Articulatory Inversion Based on Local Regression ..................................................................... 937
Samer Al Moubayed, G. Ananthakrishnan

Korean Lenis, Fortis, and Aspirated Stops: Effect of Place of Articulation on Acoustic Realization .................. 941
Mirjam Brouwer

Speech Synthesis by Modeling Harmonics Structure with Multiple Function .............................................. 945
Toru Nakashika, Ryuichi Tachibana, Masahumi Nishimura, Tetsuya Takiguchi, Yasuo Ariki

Physics of Body-Conducted Silent Speech --- Production, Propagation and Representation of Non-Audible 949
Murmur ....................................................................................................................................................
Madoko Otani, Tatsuya Hirahara

TUE-SES1-P4: SPEECH ENHANCEMENT

Multichannel Noise Reduction Using Low Order RTF Estimate ................................................................. 953
Subhojit Chaklader, Nam Soo Kim, Yu Guang Jin, Tae Gyeon Kung

Reinforced Blocking Matrix with Cross Channel Projection for Speech Enhancement .................................. 957
Inho Lee, Jongsup Yoon, Yoonjae Lee, Hanseok Ko

Masking Property Based Microphone Array Post-Filter Design ................................................................. 961
Ning Cheng, Wenju Liu, Lan Wang

Reduction of Broadband Noise in Speech Signals by Multilinear Subspace Analysis ..................................... 965
Yasujiro Sato, Tetsuya Hoya, Hovagim Bakardjian, Andrzej Cichecki
TUE-SES2-O3: SINGLE-CHANNEL SPEECH ENHANCEMENT

Fast Converging Iterative Kalman Filtering for Speech Enhancement Using Long and Overlapped Tapered Windows with Large Side Lobe Attenuation .......................................................... 1081

Stephen So, Kuldip K. Paliwal

Robust Noise Estimation Using Minimum Correction with Harmonicity Control .................................................. 1085

Xuejing Sun, Kuan-Chieh Yen, Rogerio Alves

New Insights into Subspace Noise Tracking ........................................................................................................ 1089

Mahdi Triki

Bias Considerations for Minimum Subspace Noise Tracking ............................................................................. 1093

Mahdi Triki, Koos Jasz

A Corpus-Based Approach to Speech Enhancement from Nonstationary Noise .............................................. 1097

Ji Ming, Ramji Srinivasan, Danny Crookes

Bandwidth Expansion of Speech Based on Wavelet Transform Modulus Maxima Vector Mapping ................. 1101

Zhe Chen, You-Chi Cheng, Fuliang Yin, Chin-Hui Lee

TUE-SES2-O4: SPEECH SYNTHESIS IV: MISCELLANEOUS TOPICS

Hidden Markov Models with Context-Sensitive Observations for Grapheme-to-Phoneme Conversion ....................... 1105

Kalu U. Ogbaru, Peter Cahill, Julie Carson-Berndsen

Evaluating a Dialog Language Generation System: Comparing the Mountain System to Other NLG Approaches .... 1109

Brian Langner, Stephan Vogel, Alan W. Black

Active Appearance Models for Photorealistic Visual Speech Synthesis ............................................................. 1113

Wesley Mattheyes, Lukas Latacz, Werner Verhelst


Jerome R. Bellgardada

Native and Non-Native Speaker Judgements on the Quality of Synthesized Speech ........................................... 1121

Anna C. Janska, Robert A. J. Clark

Machine Learning for Text Selection with Expressive Unit-Selection Voices .................................................. 1125

Dominic Espinoza, Michael White, Eric Foster-Lussier, Chris Brew

TUE-SES2-P1: PROSODY: BASICS & APPLICATIONS

Acoustic Correlates of Meaning Structure in Conversational Speech .............................................................. 1129

Alexei V. Ivanov, Giuseppe Riccardi, S. Ghoosh, S. Tonelli, E. A. Stepanov

HMM-Based Prosodic Structure Model Using Rich Linguistic Context ............................................................ 1133

Nicolas Obin, Xavier Rodet, Anne Lacheret

Audiovisual Congruence and Pragmatic Focus Marking ..................................................................................... 1137

Charlotte Wollermann, Bernhard Schroder, Ulrich Schade

Redescribing Intonational Categories with Functional Data Analysis ............................................................. 1141

Michele Gubian, Brechtje Post

Exploring Goodness of Prosody by Diverse Matching Templates ....................................................................... 1145

Shen Huang, Hongyan Li, Shijin Wang, Jiaen Liang, Bo Xu

A Language-Identification Inspired Method for Spontaneous Speech Detection ............................................ 1149

Michaël Rouvier, Richard Dufour, Georges Linares, Yannick Esteve

Speech Dominoves and Phonetic Convergence .................................................................................................. 1153

A Quick Sequential Forward Floating Feature Selection Algorithm for Emotion Detection from Speech ........ 1157

Matyas Brendel, Riccardo Zaccarelli, Laurence Devillers

Automated Vocal Emotion Recognition Using Phoneme Class Specific Features ............................................. 1161

Geza Kiss, Jan P. H. Van Santen

Feature Selection for Pose Invariant Lip Biometrics .......................................................................................... 1165

Adrian Pass, Jianguo Zhang, Darryl Stewart

Signal-Based Accent and Phrase Marking Using the Fujisaki Model............................................................... 1169

Hasein Hassain, Rudiger Hoffmann

A Study of Interplay Between Articulatory Movement and Prosodic Characteristics in Emotional Speech Production .......................................................................................................................... 1173

Jangwon Kim, Sungbok Lee, Shrikanth S. Narayanan

TUE-SES2-P2: ASR: FEATURE EXTRACTION I

Improved Phoneme Recognition by Integrating Evidence from Spectro-Temporal and Cepstral Features .............. 1177

Shang-Wen Li, Liang-Che Sun, Lin-Shan Lee

Using Spectro-Temporal Features to Improve AFE Feature Extraction for ASR ............................................. 1181

Samuel V. Ravari, Nelson Morgan

Using Harmonic Phase Information to Improve ASR Rate .................................................................................. 1185

Ibon Saratxaga, Inma Hernaez, Igor Odriozola, Eva Navas, Iker Luengo, Daniel Erro
Speech Recognition Using Long-Term Phase Information .......................... 1189
Kazumasa Yamamoto, Etsushi Sueyoshi, Seisichi Nakagawa

Low-Dimensional Space Transforms of Posteriors in Speech Recognition ......................................................... 1193
Jan Zelinka, Jan Trmal, Laděk Muller

Hierarchical Bottle Neck Features for LVCSR ......................................................................................................................... 1197
Christian Plahl, Ralf Schluter, Hermann Ney

Hierarchical Neural Net Architectures for Feature Extraction in ASR .................................................................................. 1201
Franzisk Grezl, Martin Karafiat

Mutual Information Analysis for Feature and Sensor Subset Selection in Surface Electromyography Based Speech Recognition ......................................................................................................................... 1205
Vivek Kumar Bangarajan Sridhar, Rohit Prasad, Prem Natrajan

Learning from Human Errors: Prediction of Phoneme Confusions Based on Modified ASR Training ......................... 1209
Bernd T. Meyer, Birger Kollmeier

TUE-SES2-P3: SPEECH PERCEPTION II: CROSS LANGUAGE AND AGE

Speech Intelligibility of Diagonally Localized Speech with Competing Noise Using Bone-Conduction Headphones .......... 1213
Kazuhiro Kondo, Takayuki Kanda, Yusuke Kobayashi, Hiroyuki Yagyu

Masking of Vowel-Analog Transitions by Vowel-Analog Distracters .................................................................................. 1217
Pierre L. Divenyi

2010, a Speech Oddity: Phonetic Transcription of Reversed Speech ......................................................................................... 1221
Francois Pellegrino, Emmanuel Ferragne, Fanny Meunier

Perception on Pitch Reset at Discourse Boundaries ................................................................................................................. 1225
Hsii-Yi Lin, Janice Fon

Effect of Spatial Separation on Speech-in-Noise Comprehension in Dyslexic Adults ................................................................. 1229
Marjorie Dole, Michel Hoyn, Fanny Meunier

Speech Categorization Context Effects in Seven- to Nine-Month-Old Infants ........................................................................ 1233
Ellen Marklund, Francisco Luzcera, Anna Ericsson

Changes in Temporal Processing of Speech Across the Adult Lifespan .................................................................................. 1237
Diane Kewley-Port, Larry E. Humes, Daniel Fogerty

Fluency and Structural Complexity as Predictors of L2 Oral Proficiency .................................................................................. 1241
Jared Bernstein, Jun Cheng, Masanori Suzuki

Semantic Facilitation in Bilingual Everyday Speech Comprehension ......................................................................................... 1245
Marco Van De Ven, Benjamin V. Tucker, Mirjam Ernestus

L2 Experience and Non-Native Vowel Categorization of L1-Mandarin Speakers ................................................................. 1249
Bo-Ren Hsieh, Ho-Hsien Pan

Cross-Lingual Talker Discrimination ................................................................................................................................. 1253
Mirjam Wester

Dajare is Not the Lowest Form of Wit ................................................................................................................................. 1257
Takashi Otsuka

TUE-SES2-P4: SLP SYSTEMS

Comparison of Methods for Topic Classification in a Speech-Oriented Guidance System .................................................. 1261
Rafael Torres, Shunt Takachi, Hiroshi Kanwami, Tomoko Matsui, Hiroshi Saruwatari, Kiyoushi Shikano

Using Dependency Parsing and Machine Learning for Factoid Question Answering on Spoken Documents ......................... 1265
Pere R. Comas, Jordi Turmo, Lluis Marquez

A Spoken Term Detection Framework for Recovering Out-of-Vocabulary Words Using the Web ........................................ 1269
Carolina Parada, Abhinav Seby, Mark Dredze, Frederick Jelinek

Improved Spoken Term Detection by Discriminative Training of Acoustic Models Based on User Relevance Feedback .................................................................................................................................................. 1273
Hung-Yi Lee, Chia-Ping Chen, Ching-Feng Yeh, Lin-Shan Lee

A Lightweight Keyword and Tag-Cloud Retrieval Algorithm for Automatic Speech Recognition Transcripts ................. 1277
Sebastian Tschopel, Daniel Schneider

Lecture Subtopic Retrieval by Retrieval Keyword Expansion Using Subordinate Concept ......................................................... 1281
Noboru Kanedera, Tetsuo Funada, Seisichi Nakagawa

Spoken Document Retrieval for Oral Presentations Integrating Global Document Similarities into Local Document Similarities ......................................................................................................................................................... 1285
Hiroshi Nanjo, Yusuake Iyonaga, Takeshiho Yoshimi

Comparing Word-Based Features, Statistical Language Models, and Parsing for Named Entity Recognition ......................... 1289
Joseph Polifroni, Stephanie Seneff

Efficient Combined Approach for Named Entity Recognition in Spoken Language ......................................................................................................................................................... 1293
Azeddine Zidouni, Sophie Rossert, Herve Glotin

Prominence Based Scoring of Speech Segments for Automatic Speech-to-Speech Summarization ................................ 1297
Sree Harsha Yella, Vasudeva Varma, Kishore Prabhatad

Maximum Lexical Cohesion for Fine-Grained News Story Segmentation .................................................................................. 1301
Zihan Liu, Lei Xie, Wei Feng

Phoneme Lattice Based TextTiling Towards Multilingual Story Segmentation ............................................................... 1305
Xiaoxuan Wang, Lei Xie, Bin Ma, Eng Siong Chng, Haizhou Li
TUE-SES3-S1: SPECIAL SESSION: QUALITY OF EXPERIENCING SPEECH SERVICES

The Characterization of the Relative Information Content by Spectral Features for the Objective Intelligibility
Anton Schlesinger, Marinus M. Boorne

Assessment of Nonlinearly Processed Speech
..........................................................1309
Anton Schlesinger, Marinus M. Boorne

Analytical Assessment and Distance Modeling of Speech Transmission Quality
..........................................................................................................................................................................................1313
Marcel Waltermann, Alexander Raake, Sebastian Moller

An Intrusive Super-Wideband Speech Quality Model: DIAL
..........................................................................................................................................................................................................................................................1317
Nicolas Cote, Vincent Koehl, Valerie Gautier-Turnin, Alexander Raake, Sebastian Moller

It Takes Two to Tango --- Assessing the Impact of Delay on Conversational Interactivity on Perceived Speech Quality
..........................................................................................................................................................................................................................................................1321
Sebastian Eigger, Raimund Schatz, Stefan Scherer

Comparison of Approaches for Instrumentally Predicting the Quality of Text-to-Speech Systems
..........................................................................................................................................................................................................................................................1325
Sebastian Möller, Florian Hintereiter, Tiago H. Falk, Tim Polzehl

A Hybrid Architecture for Mobile Voice User Interfaces
..........................................................................................................................................................................................................................................................1329
Imre Kiss, Joseph Polifroni, Chao Wang, Ghinwa Choueiter, Mike Phillips

Assessment of Spoken and Multimodal Applications: Lessons Learned from Laboratory and Field Studies
..........................................................................................................................................................................................................................................................1333
Markku Turunen, Jaakko Hakulinen, Tomi Heimonen

Improving Cross Database Prediction of Dialogue Quality Using Mixture of Experts
..........................................................................................................................................................................................................................................................1337
Klaus-Peter Engelbrecht, Hamed Ketabdar, Sebastian Moller

TUE-SES3-O2: LANGUAGE PROCESSING

Improving ASR-Based Topic Segmentation of TV Programs with Confidence Measures and Semantic Relations
..........................................................................................................................................................................................................................................................1365
Camille Guinaudeau, Guillaume Gravier, Pascale Sebillot

The Relevance of Timing, Pauses and Overlaps in Dialogues: Detecting Topic Changes in Scenario Based Meetings
..........................................................................................................................................................................................................................................................1369
Saturnino Luz, Jing Su

Semi-Supervised Part-of-Speech Tagging in Speech Applications
..........................................................................................................................................................................................................................................................1373
Richard Dufour, Benoit Favre

Memory-Based Active Learning for French Broadcast News
..........................................................................................................................................................................................................................................................1377
Frédéric Tantini, Christophe Certessa, Claire Gardent

Can Conversational Word Usage Be Used to Predict Speaker Demographics?
..........................................................................................................................................................................................................................................................1381
Dan Gillick

Prosodic Word-Based Error Correction in Speech Recognition Using Prosodic Word Expansion and Contextual Information
..........................................................................................................................................................................................................................................................1385
Chao-Hong Liu, Chung-Hsien Wu

TUE-SES3-O3: SPEECH AND AUDIO SEGMENTATION

Fully Automatic Segmentation for Prosodic Speech Corpora
..........................................................................................................................................................................................................................................................1389
Sarah Hoffmann, Beat Pfister

A Novel Text-Independent Phonetic Segmentation Algorithm Based on the Microcanonical Multiscale Formalism
..........................................................................................................................................................................................................................................................1393
Yahid Khanapha, Khalid Daoudi, Oriol Pont, Hussein Yahia

Phone Boundary Detection Using Sample-Based Acoustic Parameters
..........................................................................................................................................................................................................................................................1397
You-Yu Lin, Yih-Bu Wang, Yuan-Fu Liao

HMM-Based Automatic Visual Speech Segmentation Using Facial Data
..........................................................................................................................................................................................................................................................1401
Utpala Musti, Asterios Tsiotis, Slim Omni, Vincent Colotte, Brigitte Wrobel-Daucourt, Marie-Odile Berger

Bayes Factor Based Speaker Segmentation for Speaker Diarization
..........................................................................................................................................................................................................................................................1405
D. Wang, Robert Vogt, Srídha Sridharan

Using High-Level Information to Detect Key Audio Events in a Tennis Game
..........................................................................................................................................................................................................................................................1409
Qiang Huang, Stephen Cox
**TUE-SES3-O4: PROSODY: ANALYSIS**

What Do You Mean, You're Uncertain?: The Interpretation of Cue Words and Rising Intonation in Dialogue .................................................. 1413
Catherine Lai

Coping Imbalanced Prosodic Unit Boundary Detection with Linguistically-Motivated Prosodic Features ......................................................... 1417
Yi-Fen Liu, Shu-Chuan Tseng, Jyh-Shing Roger Jang, C.-H. Alvin Chen

Improving Prosodic Phrase Prediction by Unsupervised Adaptation and Syntactic Features Extraction ......................................................... 1421
Zhigang Chen, Guoping Hu, Wei Jiang

Perception-Based Automatic Approximation of F0 Contours in Cantonese Speech ............................................................................. 1425
Yujia Li, Yun Lee

Discriminative Training and Unsupervised Adaptation for Labeling Prosodic Events with Limited Training Data ................................. 1429
Raul Fernandez, BhuvanaRamabhadran

Prosody for the Eyes: Quantifying Visual Prosody Using Guided Principal Component Analysis ............................................................. 1433
Erin Cvejic, Jeewan Kim, Chris Davis, Guillaume Gibert

**TUE-SES3-P1: SPEAKER CHARACTERIZATION AND RECOGNITION III**

An Investigation into Direct Scoring Methods Without SVM Training in Speaker Verification .......................................................... 1437
Ce Zhang, Rong Zheng, Bo Xu

Large Margin Gaussian Mixture Models for Speaker Identification .............................................................................................................. 1441
Reda Souami, Khalid Daoudi, Regine Andre-Obrecht, Dross Abou时代的

On the Use of Gaussian Component Information in the Generative Likelihood Ratio Estimation for Speaker Verification ........................................ 1445
Rong Zheng, Bo Xu

Acoustic Vector Resampling for GMM-SVM-Based Speaker Verification ........................................................................................................ 1449
Man-Wai Mak, Wei Rao

A Fast Speaker Indexing Using Vector Quantization and Second Order Statistics with Adaptive Threshold Computation ................................ 1453
Konstantin Biatov

Using Phoneme Recognition and Text-Dependent Speaker Verification to Improve Speaker Segmentation for Chinese Speech ...................... 1457
Gang Wang, Xiaojun Wu, Thomas Fang Zheng

On Enhancing Feature Sequence Filtering with Filter-Bank Energy Transformation in Speaker Verification with Telephone Speech ........................................... 1461
Claudio Garreton, Nestor Becerra Yoma

MAP Estimation of Subspace Transform for Speaker Recognition ................................................................................................................. 1465

A Longest Matching Segment Approach for Text-Independent Speaker Recognition ...................................................................................... 1469
Ayeh Jafari, Ramji Srinivasan, Danny Crookes, Ji Ming

Approaching Human Listener Accuracy with Modern Speaker Verification ..................................................................................................... 1473
Ville Hautamaki, Tomi Kinnunen, Mohaddeseh Nosratighods, Kong Aik Lee, Bin Ma, Haizhou Li

Extended Weighted Linear Prediction (XLP) Analysis of Speech and Its Application to Speaker Verification in Adverse Conditions .................................................. 1477
Guoli Ye, Brian Mak

The Use of Subvector Quantization and Discrete Densities for Fast GMM Computation for Speaker Verification .......................................................... 1481

**TUE-SES3-P2: SYSTEMS FOR LVCSR AND RICH TRANSCRIPTION**

Parallel Lexical-Tree Based LVCSR on Multi-Core Processors .............................................................................................................................. 1485
Navneet Parihar, Ralf Schluter, David Rybach, Eric A. Hansen

Exploring Recognition Network Representations for Efficient Speech Inference on Highly Parallel Platforms .................................................. 1489
Jieke Chong, Ekaterina Gornina, Kisan You, Kurt Keutzer

WFST Compression for Automatic Speech Recognition ............................................................................................................................. 1493
Diamantino Caceres

Speech Recognizer Optimization Under Speed Constraints ............................................................................................................................. 1497
Ivan Bulyko

The 2010 CMU GALE Speech-to-Text System .............................................................................................................................................. 1501
Florian Metze, Roger Hsiao, Qin Jin, Udhayakumar Nallasamy, Tanja Schultz

Speaker Diarization in Meeting Audio for Single Distant Microphone .............................................................................................................. 1505
Tin Lai Yue, Han Cai Sun, Bin Ma, Haizhou Li

Extending the Punctuation Module for European Portuguese .............................................................................................................................. 1509
Fernando Battista, Helena Moniz, Isabel Trancoso, Hugo Meinedo, Ana Isabel Mata, Nuno Mamede

Utilizing a Noisy-Channel Approach for Korean LVCSR ............................................................................................................................. 1513
Saikriri Sakti, Ryosuke Itoi, Hidashi Kawai, Satoshi Nakamura

The RWTH 2009 Quaero ASR Evaluation System for English and German ........................................................................................................ 1517
Markus Naussbaum-Thom, Simon Wiesler, Martin Sundermeyer, Christian Pohl, Stefan Hahn, Ralf Schluter, Hermann Ney
TUE-SES3-P4: SPEECH PRODUCTION II: VOCAL TRACT MODELING AND IMAGING

Data-Driven Analysis of Realtime Vocal Tract MRI Using Correlated Image Regions ................................................. 1572
Adam C. Lammert, Michael I. Proctor, Shrikanth S. Narayanan

Rapid Semi-Automatic Segmentation of Real-Time Magnetic Resonance Images for Parametric Vocal Tract
Analysis......................................................................................................................................................................... 1576
Yoon-Chul Kim, Shrikanth S. Narayanan, Krishna S. Nayak

Improved Real-Time MRI of Oral-Velar Coordination Using a Golden-Ratio Spiral View Order ........................................................................................................................................................................... 1580
Yuuki Tanida, Taiti Ueno, Satoru Saito, Matthew A. Lambon Ralph

Statistical Multi-Stream Modeling of Real-Time MRI Articulatory Speech Data............................................................... 1584
Erik Bresch, Athanasios Katsamanis, Louis Goldstein, Shrikanth S. Narayanan

Predicting Unseen Articulations from Multi-Speaker Articulatory Models ................................................................. 1588
G. Ananthakrishnan, Pierre Badin, Julian Andres Valdes Vargas, Olov Engwall

Estimating Missing Data Sequences in X-Ray Microbeam Recordings ........................................................................ 1592
Chao Qin, Miguel A. Carreira-Perpinan, Mohsen Farhadlloo

Adaptation of a Tongue Shape Model by Local Feature Transformations .................................................................. 1596
Zhuang Meng, Denis Beaudemps, Gang Feng, Sebastien Schmerber

Vocal Tract Contour Analysis of Emotional Speech by the Functional Data Curve Representation .............................. 1600
Sangbok Lee, Shrikanth S. Narayanan

Locally-Weighted Regression for Estimating the Forward Kinematics of a Geometric Vocal Tract Model .................. 1604
Adam C. Lammert, Louis Goldstein, Khalil Iskareus

Identifying Articulatory Goals from Kinematic Data Using Principal Differential Analysis ..................................... 1608
Michaël Reimer, Frank Rudzicz

Estimation of Speech Lip Features from Discrete Cosinus Transform ......................................................................... 1612
Farzaneh Ahmadi, Ian V. McLoughlin, Hamid R. Sharifzadeh

Autoregressive Modelling for Linear Prediction of Ultrasonic Speech........................................................................ 1616

WED-SES1-S1: SPECIAL SESSION: SPEECH INTELLIGIBILITY ENHANCEMENT FOR ALL AGES,
HEALTH CONDITIONS AND ENVIRONMENTS

Enhanced Speech Yielding Higher Intelligibility for All Listeners and Environments .................................................. 1620
Takayuki Arat, Nao Hodoshima
Quality Conversion of Non-Acoustic Signals for Facilitating Human-to-Human Speech Communication Under
Harsh Acoustic Conditions .................................................................................................................. 1624
Seyed Omid Sadjadi, Sanjay A. Patil, John H. L. Hansen

The Use of Air-Pressure Sensor in Electrolaryngeal Speech Enhancement Based on Statistical Voice Conversion .......................................................... 1628
Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, Kiyohiro Shikano

A New Binary Mask Based on Noise Constraints for Improved Speech Intelligibility ................................................................. 1632
Gibak Kim, Philipos C. Loizou

Energy Reallocation Strategies for Speech Enhancement in Known Noise Conditions .......................................................... 1636
Yan Tang, Martin Cooke

Effects of Enhancement of Spectral Changes on Speech Quality and Subjective Speech Intelligibility ................................................................. 1640
Jing Chen, Thomas Baer, Brian C. J. Moore

WED-SES1-O2: SLP SYSTEMS FOR INFORMATION EXTRACTION/RETRIEVAL

CRF-Based Stochastic Pronunciation Modeling for Out-of-Vocabulary Spoken Term Detection .......................................................... 1668
Dong Wang, Simon King, Nicholas Evans, Raphael Tromey

Improved Spoken Term Detection by Feature Space Pseudo-Relevance Feedback .......................................................... 1672
Chia-Ping Chen, Hung-Yi Lee, Ching-Feng Yeh, Lin-Shan Lee

Towards Spoken Term Discovery at Scale with Zero Resources .................................................................................. 1676
Aren Jansen, Kenneth Church, Hynek Hermansky

Vocabulary Independent Spoken Query: A Case for Subword Units .................................................................................. 1680
Evandro Gouvea, Tony Ezzat

Extractive Speech Summarization — From the View of Decision Theory .................................................................................. 1684
Shih-Hsiang Lin, Yao-Ming Yeh, Berlin Chen

The Impact of ASR on Abstractive vs. Extractive Meeting Summaries .................................................................................. 1688
Gabriel Murray, Giuseppe Carenini, Raymond Ng

WED-SES1-O3: SPEECH REPRESENTATION

Binary Coding of Speech Spectrograms Using a Deep Auto-Encoder .................................................................................. 1692
L. Deng, Michael L. Seltzer, Dong Yu, Alex Acero, Abdel-Rahman Mohamed, G. Hinton

A Super-Resolution Spectrogram Using Coupled PLCA .................................................................................. 1696
Juhun Nam, Gautham J. Mysore, Joachim Ganseran, Kyung Lee, Jonathan S. Abel

Fast Least-Squares Solution for Sinusoidal, Harmonic and Quasi-Harmonic Models .................................................................................. 1700
Georgios Tsidakis, Yannis Pantazis, Olivier Rosec, Yannis Stylianou

Sparse Component Analysis for Speech Recognition in Multi-Speaker Environment .................................................................................. 1704
Afsaneh Asaei, Herve Bourlard, Philip N. Garner

Intra-Frame Variability as a Predictor of Frame Classifiability .................................................................................. 1708
Trond Skogstad, Torbjorn Svensdalen

Autocorrelation and Double Autocorrelation Based Spectral Representations for a Noisy Word Recognition System .................................................................................. 1712
Tetuya Shimamura, Ngoc Dinh Nguyen

WED-SES1-O4: VOICE CONVERSION

Maximum a posteriori Voice Conversion Using Sequential Monte Carlo Methods .................................................................................. 1716
Elna Helander, Hanna Silen, Joaquim Miguez, Moncef Gabbouj

Dynamic Model Selection for Spectral Voice Conversion .................................................................................. 1720
Pierre Lanchantin, Xavier Rodet

Speaker-Independent HMM-Based Voice Conversion Using Quantized Fundamental Frequency .................................................................................. 1724
Takashi Nose, Taku Kobayashi
Probabilistic Integration of Joint Density Model and Speaker Model for Voice Conversion .......................................................... 1728

Text-Independent F0 Transformation with Non-Parallel Data for Voice Conversion ............................................................................. 1732

Zhi-Zheng Wu, Tori Kinnunen, Eng Siong Chong, Haizhou Li

A Minimum Converted Trajectory Error (MCTE) Approach to High Quality Speech-to-Lips Conversion ................................................. 1736

Xiaodan Zhuang, Lijuan Wang, Frank K. Soong, Mark Hasegawa-Johnson

WED-SES1-P1: PROSODY: LANGUAGE-SPECIFIC MODELS

Influence of Lexical Tones on Intonation in Kammu .................................................. 1740
Anastasia Karlsson, David House, Jan-Olof Stentesson, Damrong Tayanin

Phonetic Realization of Second Occurrence Focus in Japanese ...................................... 1744
Satoshi Nambu, Yong-Cheol Lee

Prosodic Grouping and Relative Clause Disambiguation in Mandarin .............................. 1748
Jianqiang Kuang

Text-Based Unstressed Syllable Prediction in Mandarin .................................................. 1752
Ya Li, Jianhui Tao, Meng Zhang, Shifeng Pan, Xiaoying Xu

"Flat Pitch Accents" in Czech ...................................................................................... 1756
Tomas Dubeda

Positional Variability of Pitch Accents in Czech ............................................................ 1760
Tomas Dubeda

Modeling of Sentence-Medial Pauses in Bangla Readout Speech: Occurrence and Duration ...................................................................... 1764
Shyamal Das Mandal, Arup Saha, Talika Basu, Keikichi Hirose, Hiroya Fujisaki

Declarative Sentence Intonation Patterns in 8 Swiss German Dialects .................................................. 1768
Adrian Leemann, Lucy Zuberbuhler

Syllable-Level Prominence Detection with Acoustic Evidence ........................................... 1772
Je Hun Jeon, Yang Liu

Prosody Cues for Classification of the Discourse Particle "ha" in Hindi ...................................... 1776
Sanjalan Prasad, Kalika Bali

Interaction of Syntax-Marked Focus and Wh-Question Induced Focus in Standard Chinese .................................................................. 1780
Yuan Jia, Aijun Li

Prominence Detection in Swedish Using Syllable Correlates ............................................. 1784
Samer Al Moubayed, Jonas Beskow

Automatic Analysis of the Intonation of a Tone Language. Applying the Momel Algorithm to Spontaneous Transactions in Chinese (Beijing) ........................................................................... 1788
Na Zhi, Daniel Hirtl, Pier Marco Bertinetto

Towards Long-Range Prosodic Attribute Modeling for Language Recognition .................. 1792
Raymond W. M. Ng, Cheung-Chi Leung, Ville Hautamaki, Tan Lee, Bin Ma, Haizhou Li

A Modified Parameterization of the Fujiwaki Model ....................................................... 1796
Robert Schubert, Oliver Jokisch, Diane Hirschfeld

WED-SES1-P2: ASR: LANGUAGE MODELING AND SPEECH UNDERSTANDING I

Within and Across Sentence Boundary Language Model .......................................... 1800
Saeedeh Montazer, Friedrich Faul, Dietrich Klukow

Impact of Word Classing on Shrinkage-Based Language Models .................................. 1804
Bahi Sarikaya, Stanley F. Chen, Abhinav Sethy, Bhuvana Ramabhadran

Combination of Probabilistic and Possibilistic Language Models ....................................... 1808
Stanislav Oger, Vladimir Popescu, Georges Linares

On-Demand Language Model Interpolation for Mobile Speech Input ............................. 1812
Brandon Ballinger, Cyril Aliazen, Alexander Gruenstein, Johan Schalkwyk

Text Normalization Based on Statistical Machine Translation and Internet User Support ........................................................................... 1816
Tim Schlippe, Chenfei Zhu, Jan Gebhardt, Tanja Schultz

Efficient Estimation of Maximum Entropy Language Models with N-Gram Features: An SRILM Extension ................................................. 1820
Tanel Alumae, Mikko Karimo

Similar N-Gram Language Model ................................................................................. 1824
Christian Gillot, Christophe Cerisara, David Langlois, Jean-Paul Haton

Topic and Style-Adapted Language Modeling for Thai Broadcast News ASR ......................... 1828
Markphong Jongmanussaporn, Sadaoki Furui

Augmented Context Features for Arabic Speech Recognition ........................................... 1832
Ahmad Emami, Hong-Kwoang J. Kao, Imad Ztouni, Lidia Mangal

A Statistical Segment-Based Approach for Spoken Language Understanding ..................... 1836
Lucia Ortega, Isabel Galisano, Luis-F. Hurtado, Emilio Sanchis, Encarna Segarra

WED-SES1-P3: FIRST AND SECOND LANGUAGE ACQUISITION

Cantonese Tone Word Learning by Tone and Non-Tone Language Speakers .................................................. 1840
Angela Cooper, Yue Wang
Validation of a Training Method for L2 Continuous-Speech Segmentation ......................................................................................................................................................................................1844
Anne Cutler, Janine Shandel

Linguistic Rhythm in Foreign Accent ......................................................................................................................................................................................1848
Jiaohong Yuan

The Effect of a Word Embedded in a Sentence and Speaking Rate Variation on the Perceptual Training of Geminate and Singleton Consonant Distinction ......................................................................................................................................................................................1850
Mee Sonu, Keiichi Tajima, Hiroaki Kato, Yoshimori Sagenaka

Foreign Accent Matters Most When Timing is Wrong .....................................................................................................................................................1854
Hyejin Hong, Jina Kim, Minhwa Chung

The Effects of EMA-Based Augmented Visual Feedback on the English Speakers' Acquisition of the Japanese Flap: A Perceptual Study ......................................................................................................................................................................................1862
June S. Levitt, William F. Katz

Perception of Voiceless Fricatives by Japanese Listeners of Advanced and Intermediate Level English Proficiency ......................................................................................................................................................................................1866
Hinako Matsuda, Takayuki Arai

Perception of Estonian Vowel Categories by Native and Non-Native Speakers ......................................................................................................................................................................................1870
Lyta Meister, Einar Meister

Spoken English Assessment System for Non-Native Speakers Using Acoustic and Prosodic Features ......................................................................................................................................................................................1874
Qin Shi, Kun Li, Shilici Zhang, Stephen M. Chu, Ji Xiao, Zhijian Ou

Russian Infants and Children's Sounds and Speech Corpora for Language Acquisition Studies ......................................................................................................................................................................................1878
Elena E. Lyakso, Olga V. Frolova, Anna V. Kuzhzhova, Julia S. Gautova

Language-Specific Influence on Phoneme Development: French and Drehu Data ......................................................................................................................................................................................1882
Julia Molvin, Helene Loevenbruck

Did you Susi or Shushi? Measuring the Emergence of Robust Fricative Contrasts in English- and Japanese-Acquiring Children ......................................................................................................................................................................................1886
Jeffrey J. Holliday, Mary E. Beckman, Chanelle Mays

WED-SES1-P4: SPOKEN LANGUAGE RESOURCES, SYSTEMS AND EVALUATION I

An Empirical Comparison of the T3, Juicer, HDecode and Sphinx3 Decoders ......................................................................................................................................................................................1890
Josef R. Novak, Paul R. Dixon, Sadao3 Furui

Tracter: A Lightweight Dataflow Framework ......................................................................................................................................................................................1894
Philip N. Garner, John Dines

Verifying Pronunciation Dictionaries Using Conflict Analysis ......................................................................................................................................................................................1898
Marjorie H. Devel, Febe De Wet

Automatic Estimation of Transcription Accuracy and Difficulty ......................................................................................................................................................................................1902
Brandon C. Roy, Sorouch Vosoughi, Deb Roy

Creating a Linguistic Plausibility Dataset with Non-Expert Annotators ......................................................................................................................................................................................1906
Benjamin Lambert, Rita Singh, Bhiksha Raj

Construction and Evaluations of an Annotated Chinese Conversational Corpus in Travel Domain for the Language Model of Speech Recognition ......................................................................................................................................................................................1910
Xinhui Hu, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura

Building Transcribed Speech Corpora Quickly and Cheaply for Many Languages ......................................................................................................................................................................................1914
Thad Hughes, Katsuke Nakajima, Lonne Ha, Atul Vasu, Pedro J. Moreno, Mike Lebeau

The CHIME Corpus: A Resource and a Challenge for Computational Hearing in Multisource Environments ......................................................................................................................................................................................1918
Heidi Christensen, Jon Barker, Ning Ma, Phil D. Green

Developing a Chinese L2 Speech Database of Japanese Learners with Narrow-Phonetic Labels for Computer Assisted Pronunciation Training ......................................................................................................................................................................................1922
Wen Cao, Dongqiang Wang, Jinsong Zhang, Ziya Xiong

How Children Acquire Situation Understanding Skills?: A Developmental Analysis Utilizing Multimodal Speech Behavior Corpus ......................................................................................................................................................................................1926
Shogo Ishikawa, Shinya Kuriyama, Yoichi Takebayashi, Shigeoishi Kitazawa

The Influence of Expertise and Efficiency on Modality Selection Strategies and Perceived Mental Effort ......................................................................................................................................................................................1930
Ina Wechuan, Stefan Schauffer, Robert Schleicher, Anja Sanneman, Sebastian Muller

Parameters Describing Multimodal Interaction — Definitions and Three Usage Scenarios ......................................................................................................................................................................................1934
Christine Kuhnle, Benjamin Weisz, Sebastian Muller

Repair Strategies on Trial: Which Error Recovery Do Users Like Best? ......................................................................................................................................................................................1938
Alexander Zgorzelski, Alexander Schmitt, Tobias Heinroth, Wolfgang Minker

WED-SES2-O1: ASR: SEARCH, DECODING AND CONFIDENCE MEASURES II

CRF-Based Combination of Contextual Features to Improve a posteriori Word-Level Confidence Measures ......................................................................................................................................................................................1942
Julien Fayolle, Fabienne Moreau, Christian Raymond, Guillaume Gravier, Patrick Gross

Recognition of Spontaneous Conversational Speech Using Long Short-Term Memory Phoneme Predictions ......................................................................................................................................................................................1946
Martin Wolnmer, Florian Eyben, Bjorn Schuler, Gerhard Rigoll

Improving ASR Error Detection with Non-Decoder Based Features ......................................................................................................................................................................................1950
Thomas Pellegrini, Isabel Trancoso
Phoneme Classification and Lattice Rescoring Based on a k-NN Approach ................................................................. 1954
Ludan Golipour, Douglas O'Shaughnessy

Online Adaptive Learning for Speech Recognition Decoding ................................................................. 1958
Jeff Böhm, Hui Lin

Improvements of Search Error Risk Minimization in Viterbi Beam Search for Speech Recognition ......................... 1962
Takaaki Hori, Shinji Watanabe, Atsushi Nakamura

WED-SES2-O2: SPOKEN LANGUAGE RESOURCES, SYSTEMS AND EVALUATION II

Say What? Why Users Choose to Speak Their Web Queries ................................................................. 1966
Maryam Kamvar, Doug Beeferman

The Effect of Audience Familiarity on the Perception of Modified Accent ........................................................ 1970
Jonathan Teutenberg, Catherine J. Watson

On Generating Comibllex Pronunciations via Morphological Analysis ........................................................ 1974
Korin Richmond, Robert A. J. Clark, Sue Fitt

Say It As You Mean It — Analyzing Free User Comments in the VOICE Awards Corpus ....................................... 1978
Florian Gode, Sebastian Moller

A New Multichannel Multi Modal Dyadic Interaction Database ................................................................. 1982
Viktor Rojec, Bo Xiao, Athanasios Katsamanis, Brian R. Baucom, Panayiotis G. Georgiou, Shrikanth S. Narayanan

SEAME: A Mandarin-English Code-Switching Speech Corpus in South-East Asia ............................................. 1986
Dau-Cheng Lyu, Tian-Ping Tao, Eng Shiong Chng, Haizhou Li

WED-SES2-O3: SPEECH PRODUCTION III: ANALYSIS

Relying on Critical Articulators to Estimate Vocal Tract Spectra in an Articulatory-Acoustic Database .................... 1990
Daniel Felps, Christian Geng, Michael Berger, Korin Richmond, Ricardo Gutierrez-Orra

Vikram Ramanarayanan, Dani Byrd, Louis Goldstein, Shrikanth S. Narayanan

Articulatory Inversion of American English /a/ by Conditional Density Modes .................................................. 1998
Chao Qin, Miguel A. Carreira-Perepon

Can Tongue Be Recovered from Face? The Answer of Data-Driven Statistical Models ............................................ 2002
Atif Ben Yousef, Pierre Budin, Gerard Baillif

Phrase-Medial Vowel Devoicing in Spontaneous French ............................................................................. 2006
Francisco Torreira, Mirjam Ernestus

Exploring the Mechanism of Tonal Contraction in Taiwanese Mandarin ......................................................... 2010
Chierh Cheng, Yi Xu, Michele Gabian

WED-SES2-O4: PARALANGUAGE & COGNITION

Voice Attributes Affecting Likability Perception .......................................................................................... 2014
Benjamin Weiss, Felix Barkhardt

Turn-Alignment Using Eye-Gaze and Speech in Conversational Interaction ....................................................... 2018
Kristina Jokinen, Kazuki Harada, Masafumi Nishida, Seichi Yamamoto

An Investigation of Formant Frequencies for Cognitive Load Classification ...................................................... 2022
Tet Fei Yap, Julien Epps, Eliathamby Ambikairajah, Eric H. C. Choi

Language Specific Effects of Emotion on Phoneme Duration ........................................................................ 2026
Martijn Goudbeeck, Mirjam Broersma

Automatic Classification of Married Couples’ Behavior Using Audio Features ..................................................... 2030
Matthew Black, Athanasios Katsamanis, Chi-Chun Lee, Adam C. Lammert, Brian R. Baucom, Andrew Christensen, Panayiotis G. Georgiou, Shrikanth S. Narayanan

Influence of Gestural Salience on the Interpretation of Spoken Requests ......................................................... 2034
Gideon Kowadlo, Patrick Ye, Ingrid Zukerman

WED-SE02-P1: ROBUST ASR AGAINST NOISE

Robust Word Recognition Using Articulatory-Trajectories and Gestures ......................................................... 2038
Vikramjit Mitra, Hosung Nam, Carol Espy-Wilson, Elliot Saltzman, Louis Goldstein

Performance Estimation of Noisy Speech Recognition Considering Recognition Task Complexity ......................... 2042
Takashi Yamada, Takehiro Nakajima, Nobuhiko Kitazaki, Shoji Makino

Estimating Noise from Noisy Speech Features with a Monte Carlo Variant of the Expectation Maximization Algorithm ................................................................. 2046
Friedrich Faulst, Dietrich Klakow

Template-Based Spectral Estimation Using Microphone Array for Speech Recognition ........................................ 2050
Satoshi Tamura, Eriko Hishikawa, Wataru Taguchi, Satoru Hayamizu

A Particle Filter Feature Compensation Approach to Robust Speech Recognition ................................................ 2054
Aleem Moshokq, Yu Tao, Chun-Hsi-Lee

Nonlinear Enhancement of Onset for Robust Speech Recognition .................................................................... 2058
Chanwoo Kim, Richard M. Stern
Mask Estimation in Non-Stationary Noise Environments for Missing Feature Based Robust Speech Recognition ................................................................. 2062  
Shirin Badieaeddajog, Richard C. Rose  
Robust Automatic Speech Recognition with Decoder Oriented Ideal Binary Mask Estimation ................................................................. 2066  
Lae-Hoon Kim, Kyung-Tae Kim, Mark Hasegawa-Johnson  
A Robust Speech Recognition System Against the Ego Noise of a Robot ................................................................. 2070  
Gokhan Ince, Kazuhiro Nakadai, Tobias Rodemann, Hiroshi Tsujino, Jun-Ichi Imura  
Empirical Mode Decomposition for Noise-Robust Automatic Speech Recognition ................................................................. 2074  
Kuo-Hao Wu, Chia-Ping Chen  
An Effective Feature Compensation Scheme Tightly Matched with Speech Recognizer Employing SVM-Based GMM Generation ................................................................. 2078  
Woel Kim, Jun-Won Sah, John H. L. Hansen  
Artificial and Online Acquired Noise Dictionaries for Noise Robust ASR ................................................................. 2082  
Jort F. Gemmeke, Tuomas Virtanen  
Voice Activity Detection Based on Conditional Random Fields Using Multiple Features ................................................................. 2086  
Akira Satno, Yoshitomo Nankaku, Akiho Lee, Kotaichi Toida  
A Comparative Study of Noise Estimation Algorithms for VTS-Based Robust Speech Recognition ................................................................. 2090  
Yang Zhao, Bing-Huang Jiang  
On Using Missing-Feature Theory with Cepstral Features — Approximations to the Multivariate Integral ................................................................. 2094  
Frank Seide, Pei Zhao  
Using a DBN to Integrate Sparse Classification and GMM-Based ASR ................................................................. 2098  
Yang Sun, Jort F. Gemmeke, Bert Cranen, L. Ten Bosch, Lou Boves  

WED-SES2-P2: SPEAKER CHARACTERIZATION AND RECOGNITION IV

Transcript-Dependent Speaker Recognition Using Mixer 1 and 2 ................................................................. 2102  
Fred S. Richardson, Joseph P. Campbell  
On the Potential of Glottal Signatures for Speaker Recognition ................................................................. 2106  
Thomas Dragman, Thierry Dutoit  
Acoustic Feature Diversity and Speaker Verification ................................................................. 2110  
R. Padmanabhan, Hema A. Murthy  
A Discriminative Performance Metric for GMM-UBM Speaker Identification ................................................................. 2114  
Amin Dehazangi, Bin Ma, Eng Song Chng, Haizhou Li  
A Novel Speaker Binary Key Derived from Anchor Models ................................................................. 2118  
Xavier Anguera, Jean-Francois Bonastre  
Variant Time-Frequency Cepstral Features for Speaker Recognition ................................................................. 2122  
Wei-Qiang Zhang, Yan Deng, Liang He, Jia Liu  
Exploitation of Phase Information for Speaker Recognition ................................................................. 2126  
Ning Wang, P. C. Ching, Tan Lee  
Effects of the Phonological Relevance in Speaker Verification ................................................................. 2130  
Yanhua Long, Lirong Dai, Bin Ma, Wu Guo  
Topological Representation of Speech for Speaker Recognition ................................................................. 2134  
Gabriel H. Sierra, Jean-Francois Bonastre, Dries Matrouf, Jose R. Calvo  
Assessment of Single-Channel Speech Enhancement Techniques for Speaker Identification Under Mismatched Conditions ................................................................. 2138  
Seyed Omid Sadjadi, John H. L. Hansen  
Speaker Recognition Using the Resynthesized Speech via Spectrum Modeling ................................................................. 2142  
Xiang Zhang, Chuan Cao, Lin Yang, Hongbin Suo, Jianping Zhang, Yonghong Yan  

WED-SES2-P3: VOICE CONVERSION AND SPEECH SYNTHESIS

Shape-Invariant Speech Transformation with the Phase Vocoder ................................................................. 2146  
Jiel Robel  
A Phonetic Alternative to Cross-Language Voice Conversion in a Text-Dependent Context: Evaluation of Speaker Identity ................................................................. 2150  
Kayoko Yamagisawa, Mark Huckvale  
Evaluation of Speaker Mimic Technology for Personalizing SGD Voices ................................................................. 2154  
Esther Klabbbers, Alexander Kain, Jan P. H. Van Santen  
Adaptive Voice-Quality Control Based on One-to-Many Eigenvoice Conversion ................................................................. 2158  
Kumi Ohta, Tomoki Toda, Yamato Ohtani, Hiroshi Sarawatari, Kiyohiro Shikano  
Applying Voice Conversion to Concatenative Singing-Voice Synthesis ................................................................. 2162  
Fernando Villavicencio, Jordi Bonada  
Improved Generation of Fundamental Frequency in HMM-Based Speech Synthesis Using Generation Process Model ................................................................. 2166  
Miaomiao Wang, Miaomiao Wen, Keikichi Hirose, Nobuaki Minematsu  
A Hierarchical F0 Modeling Method for HMM-Based Speech Synthesis ................................................................. 2170  
Ming Lei, Yiqun Wu, Frank K. Soong, Zhen-Hua Ling, Lirong Dai  
Training a Parametric-Based LogF0 Model with the Minimum Generation Error Criterion ................................................................. 2174  
Javier Latorre, M. J. F. Gales, Heiga Zen
Improving Mandarin Segmental Duration Prediction with Automatically Extracted Syntax Features ......................................................... 2178
Miaomiao Wen, Xiaomin Wang, Keshi Hirone, Nobuaki Minematsu

An Intonation Model for TTS in Sepedi .............................................................................................................................................. 2182
Daniel R. Van Niekerk, Etienne Barnard

Synthesis of Fast Speech with Interpolation of Adapted HSMMs and Its Evaluation by Blind and Sighted Listeners ........................................ 2186
Michael Pucher, Dietmar Schabus, Junichi Yamagishi

A Comparison of Pronunciation Modeling Approaches for HMM-TTS .............................................................................................. 2190
Gabriel Webster, Sacha Krsansavice, Kate Karvı

HMM-Based Text-to-Articulatory-Movement Prediction and Analysis of Critical Articulators ........................................................... 2194
Zheng-Hua Ling, Korin Richmond, Junichi Yamagishi

WED-SES2-P4: DETECTION, CLASSIFICATION, AND SEGMENTATION

Audio-Based Sports Highlight Detection by Fourier Local Auto-Correlations ............................................................................... 2198
Jiaxing Ye, Takumi Kobayashi, Tetsuya Higuchi

Automatic Excitement-Level Detection for Sports Highlights Generation ................................................................................. 2202
Hynek Boril, Abhijeet Sangwan, Tauqiq Hasan, John H. L. Hansen

Detecting Novel Objects in Acoustic Scenes Through Classifier Incongruence ........................................................................ 2206
Jörg-Hendrik Bach, Jann Anemüller

A Multidomain Approach for Automatic Home Environmental Sound Classification ................................................................. 2210
Stavros Ntalampiras, Ilias Potamitis, Nikos Fakotakis

Content-Based Advertisement Detection ............................................................................................................................................. 2214
Patrick Cardinal, Vishwa Gupta, Gilles Boulianne

Identification of Abnormal Audio Events Based on Probabilistic Novelty Detection ........................................................................ 2218
Stavros Ntalampiras, Ilias Potamitis, Nikos Fakotakis

Lightly Supervised Recognition for Automatic Alignment of Large Coherent Speech Recordings ................................................... 2222
Norbert Braunschweiler, M. J. F. Gales, Sabine Bachholz

Incremental Diarization of Telephone Conversations .................................................................................................................. 2226
Oshry Ben-Harush, Ithshak Lapadat, Hugo Guerman

Audio Analytics by Template Modeling and 1-Pass DP Based Decoding ..................................................................................... 2230
Srikant Cherla, V. Ramanabramanjan

Perceptual Wavelet Decomposition for Speech Segmentation ........................................................................................................ 2234
Mariusz Ziolkó, Jakub Gała, Bartosz Ziolkó, Tomasz Dwwiega

A Comparative Study of Constrained and Unconstrained Approaches for Segmentation of Speech Signal .................................... 2238
Venkatesh Keri, Rithore Prakshallad

Automatic Discriminative Measurement of Voice Onset Time ....................................................................................................... 2242
Morgan Sondereregger, Joseph Keshet

Selective Gammatone Filterbank Feature for Robust Sound Event Recognition ........................................................................ 2246
Yi Ren Leng, Huy Dat Tran, Norhide Kitaoaka, Haizhou Li

WED-SES3-S1: SPECIAL SESSION: COMPRESSIVE SENSING FOR SPEECH AND LANGUAGE PROCESSING

Towards a Robust Face Recognition System Using Compressive Sensing ................................................................................... 2250
Allen Y. Yang, Zihan Zhou, Yi Ma, S. Shankar Sastry

Sparse Representation Features for Speech Recognition ................................................................................................................. 2254
Akhil Sethy, Tara N. Sainath, Bhuvana Ramabhadran, Dimitri Kanevsky

Data Selection for Language Modeling Using Sparse Representations ............................................................................................. 2258
Abhinav Sethy, Tara N. Sainath, Bhuvana Ramabhadran, Dimitri Kanevsky

Observation Uncertainty Measures for Sparse Imputation ................................................................................................................ 2262
Jort F. Gemmeke, Ulju Remes, Kalle J. Palomaki

Sparse Representations for Text Categorization ............................................................................................................................. 2266
Tara N. Sainath, Sameer R. Maskey, Dimitri Kanevsky, Bhuvana Ramabhadran, David Nahamoo, Julia Hirschberg

Sparse Auto-Associative Neural Networks: Theory and Application to Speech Recognition ....................................................... 2270
G. S. F. S. Sivaram, Srima Canappathy, Hynek Hermansky

WED-SES3-01: ASR: LEXICAL AND PRONUNCIATION MODELING

FSM-Based Pronunciation Modeling Using Articulatory Phonological Code .................................................................................. 2274
Chi Hu, Xiaodan Zhaung, Mark Hasegawa-Johnson

Detailed Pronunciation Variant Modeling for Speech Transcription ............................................................................................... 2278
Denis Javet, Dominique Fohr, Irina Illina

A Minimum Classification Error Approach to Pronunciation Variation Modeling of Non-Native Proper Names ................................ 2282
Miaomiao Wen, Xiaomin Wang, Keshi Hirone, Nobuaki Minematsu

Acoustics-Based Phonetic Transcription Method for Proper Nouns ................................................................................................. 2286
Antoine Laurent, Sylvain Meignier, Teva Merlin, Paul Deleglise

Wiktionary as a Source for Automatic Pronunciation Extraction .................................................................................................. 2290
Tim Schlippe, Sebastian Ochs, Tanja Schultz
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning New Word Pronunciations from Spoken Examples</td>
<td>2294</td>
</tr>
<tr>
<td>Ibrahim Badr, Ian McGraw, James Glass</td>
<td></td>
</tr>
<tr>
<td>WED-SES3-O2: SPEAKER RECOGNITION AND DIARIZATION</td>
<td></td>
</tr>
<tr>
<td>Phonetic Subspace Mixture Model for Speaker Diarization</td>
<td>2298</td>
</tr>
<tr>
<td>J-Fan Chen, Shih-Sian Cheng, Hsin-Min Wang</td>
<td></td>
</tr>
<tr>
<td>Overlap Detection for Speaker Diarization by Fusing Spectral and Spatial Features</td>
<td>2302</td>
</tr>
<tr>
<td>Martin Zelenak, Carlos Segura, Javier Hernando</td>
<td></td>
</tr>
<tr>
<td>Floor Holder Detection and End of Speaker Turn Prediction in Meetings</td>
<td>2306</td>
</tr>
<tr>
<td>Alfred Dietmann, Giulia Garau, Herve Bourlard</td>
<td></td>
</tr>
<tr>
<td>Confidence Measures for Speaker Segmentation and Their Relation to Speaker Verification</td>
<td>2310</td>
</tr>
<tr>
<td>Carlos Vaquero, Affonso Ortega, Jeser Villaiba, Antonio Miguel, Eduardo Lleida</td>
<td></td>
</tr>
<tr>
<td>Decoupling Session Variability Modelling and Speaker Characterisation</td>
<td>2314</td>
</tr>
<tr>
<td>Anthony Larcher, Christophe Levy, Driss Matrouf, Jean-Francois Bonastre</td>
<td></td>
</tr>
<tr>
<td>Incorporating MAP Estimation and Covariance Transform for SVM Based Speaker Recognition</td>
<td>2318</td>
</tr>
<tr>
<td>Cheung-Chi Leung, Donglai Zhu, Kong Aik Lee, Bin Ma, Haizhou Li</td>
<td></td>
</tr>
<tr>
<td>WED-SES3-O3: SPEECH AND AUDIO CLASSIFICATION</td>
<td></td>
</tr>
<tr>
<td>Single-Speaker/Multi-Speaker Co-Channel Speech Classification</td>
<td>2322</td>
</tr>
<tr>
<td>Stephane Rossignol, Olivier Pietquin</td>
<td></td>
</tr>
<tr>
<td>Discriminative Training for Hierarchical Clustering in Speaker Diarization</td>
<td>2326</td>
</tr>
<tr>
<td>Oriol Vinyals, Gerald Friedland, Nelson Morgan</td>
<td></td>
</tr>
<tr>
<td>GMM-UBM Based Open-Set Online Speaker Diarization</td>
<td>2330</td>
</tr>
<tr>
<td>Jurgen Geiger, Frank Wallhoff, Gerhard Rigoll</td>
<td></td>
</tr>
<tr>
<td>A Segment-Based Non-Parametric Approach for Monophone Recognition</td>
<td>2334</td>
</tr>
<tr>
<td>Ladan Golipour, Douglas O'Shaughnessy</td>
<td></td>
</tr>
<tr>
<td>A Fast One-Pass-Training Feature Selection Technique for GMM-Based Acoustic Event Detection with Audio-Visual Data</td>
<td>2338</td>
</tr>
<tr>
<td>Tomas Balke, Clement Nadeau</td>
<td></td>
</tr>
<tr>
<td>Effects of Modelling Within- and Between-Frame Temporal Variations in Power Spectra on Non-Verbal Sound Recognition</td>
<td>2342</td>
</tr>
<tr>
<td>Nobuhide Yamakawa, Tetsuro Kitahara, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno</td>
<td></td>
</tr>
<tr>
<td>VOLUME 4</td>
<td></td>
</tr>
<tr>
<td>WED-SES3-O4: EMOTION RECOGNITION</td>
<td></td>
</tr>
<tr>
<td>On the Importance of Glottal Flow Spectral Energy for the Recognition of Emotions in Speech</td>
<td>2346</td>
</tr>
<tr>
<td>Ling He, Margaret Lech, Nicholas Allen</td>
<td></td>
</tr>
<tr>
<td>Real-Life Emotion-Related States Detection in Call Centers: A Cross-Corpora Study</td>
<td>2350</td>
</tr>
<tr>
<td>Laurence Devillers, Christophe Vaudable, Clement Chastagnol</td>
<td></td>
</tr>
<tr>
<td>Multi-Class and Hierarchical SVMs for Emotion Recognition</td>
<td>2354</td>
</tr>
<tr>
<td>Ali Hassan, Robert J. Damper</td>
<td></td>
</tr>
<tr>
<td>Determining Optimal Features for Emotion Recognition from Speech by Applying an Evolutionary Algorithm</td>
<td>2358</td>
</tr>
<tr>
<td>David Hubner, Bogdan Vlasenko, Tobias Grosse, Andreas Wendemuth</td>
<td></td>
</tr>
<tr>
<td>Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression Using Bidirectional LSTM Modeling</td>
<td>2362</td>
</tr>
<tr>
<td>Martin Wallner, Angeliki Metallinou, Florian Eyben, Bjorn Schuller, Shrikanth S. Narayanan</td>
<td></td>
</tr>
<tr>
<td>Data-Dependent Evaluator Modeling and Its Application to Emotional Valence Classification from Speech</td>
<td>2366</td>
</tr>
<tr>
<td>Karthik Audhkhasi, Shrikanth S. Narayanan</td>
<td></td>
</tr>
<tr>
<td>WED-SES3-P1: SPEECH CODING, MODELING, AND TRANSMISSION</td>
<td></td>
</tr>
<tr>
<td>Modelling Speech Line Spectral Frequencies with Dirichlet Mixture Models</td>
<td>2370</td>
</tr>
<tr>
<td>Zhanyu Ma, Arne Leijon</td>
<td></td>
</tr>
<tr>
<td>PDF-Optimized LSF Vector Quantization Based on Beta Mixture Models</td>
<td>2374</td>
</tr>
<tr>
<td>Zhanyu Ma, Arne Leijon</td>
<td></td>
</tr>
<tr>
<td>Non-Linear Predictive Vector Quantization of Feature Vectors for Distributed Speech Recognition</td>
<td>2378</td>
</tr>
<tr>
<td>Jose Enrique Garcia, Affonso Ortega, Antonio Miguel, Eduardo Lleida</td>
<td></td>
</tr>
<tr>
<td>Supercospectral Extension of G.718 and G.729.1 Speech Codes</td>
<td>2382</td>
</tr>
<tr>
<td>Lasse Laubkosten, Milko Tammi, Vladimir Malenovskiy, Tommy Vaillancourt, Mi Suk Lee, Tomofumi Yamanishi, Masahiro Oshikiri, Claude Lamblin, Balazs Kovese, Lei Miao, Deming Zhang, Jon Gibbs, Holly Francois</td>
<td></td>
</tr>
<tr>
<td>A Multipulse FEC Scheme Based on Amplitude Estimation for CELP Codes Over Packet Networks</td>
<td>2386</td>
</tr>
<tr>
<td>Jose L. Carmona, Angel M. Gomez, Antonio M. Peinado, Jose L. Perez-Cordoba, Jose A. Gonzalez</td>
<td></td>
</tr>
</tbody>
</table>
Voice Quality Evaluation of Recent Open Source Codecs ................................................................. 2390
Amasi Bamo, Henri Toukonmaa

Efficient HMM-Based Estimation of Missing Features, with Applications to Packet Loss Concealment .............................. 2394
Bengt J. Borgström, Per H. Borgström, Abeer Alwan

Speech Inventory Based Discriminative Training for Joint Speech Enhancement and Low-Rate Speech Coding .................. 2398
Xiaoqiang Xiao, Robert M. Nickel

Quality-Based Payout Buffering with FEC for Conversational VoIP ...................................................... 2402
Qipeng Gong, Peter Kabal

Sub-Band Basis Spectrum Model for Pitch-Synchronous Log-Spectrum and Phase Based on Approximation of Sparse Coding .................................................................................. 2406
Masatsune Tamura, Takehiko Kagoshima, Masami Akamine

A Multimodal Density Function Estimation Approach to Formant Tracking ......................................................... 2410
Harshavardhan S., Chandra Sekhar Seelamantula, T. V. Sreenivas

Estimation Studies of Vocal Tract Shape Trajectory Using a Variable Length and Lossy Kelly-Lochbaum Model ................. 2414
Hekki Rusilo, Unto K. Launis, Okko Rissanen

WED-SES3-P2: ASR: LANGUAGE MODELING AND SPEECH UNDERSTANDING II

Improving Back-Off Models with Bag of Words and Hollow-Grams .............................................................................. 2418
Benjamin Lecomte, Raphaël Rahina, Georges Linarès

Study on Interaction Between Entropy Pruning and Kneser-Ney Smoothing ................................................................. 2422
Ciprian Chelbsa, Thorsten Brants, Will Neveitl, Peng Xu

Dynamic Language Model Adaptation Using Keyword Category Classification .......................................................... 2426
Hitoshi Yamamoto, Ken Hanazawa, Kiyoju Miki, Koichi Shinoda

Integration of Cache-Based Model and Topic Dependent Class Model with Soft Clustering and Soft Voting ...................... 2430
Welly Naptali, Masatsugu Tsuchiya, Seiichi Nakagawa

Conditional Models for Detecting Lambda-Functions in a Spoken Language Understanding System .................................. 2434
Frederic Duverri, Renato De Mott

Novel Weighting Scheme for Unsupervised Language Model Adaptation Using Latent Dirichlet Allocation ....................... 2438
Atmal Haidar, Douglass O’Shaughnessy

Automatic Speech Recognition System Channel Modeling ......................................................................................... 2442
Qun Feng Tan, Karik Audhkhali, Panayotis G. Georgiou, Emil Etelsale, Shrkanth S. Narayanan

Round-Robin Discrimination Model for Reranking ASR Hypotheses ........................................................................... 2446
Takanobu Oba, Takashi Mori, Atsushi Nakamura

On-the-Fly Lattice Rescoring for Real-Time Automatic Speech Recognition ............................................................... 2450
Hastin Sah, Muru Saradlar, Pungo Gungor

WED-SES3-P3: SPEECH PERCEPTION III: PROCESSING AND INTELLIGIBILITY

A Feature Extraction Method for Automatic Speech Recognition Based on the Cochlear Nucleus ........................................ 2454
Serajul Haque, Roberto Togneri

A Phoneme Recognition Framework Based on Auditory Spectro-Temporal Receptive Fields ........................................ 2458
Samuel Thomas, Kaijash Patil, Srim Ganapathy, Nima Mesgarani, Hynek Hermansky

Perceptual Compensation for Effects of Reverberation in Speech Identification: A Computer Model Based on Auditory Efferent Processing .................................................................... 2462
Amy V. Boeston, Guy J. Brown

Predicting Human Perception and ASR Classification of Word-Final [t] by Its Acoustic Sub-Segmental Properties .................. 2466
Barbara Schuppier, Mirjam Ernestus, Wil Van Dommelen, Jacques Koreman

Multichannel Source Separation Based on Source Location Cue with Log-Spectral Shaping by Hidden Markov Source Model ........................................................................................................ 2470
Tomohiro Nakatani, Shoko Araki, Takuya Toshihoka, Masakiyo Fujimoto

A Speech-in-Noise Test Based on Spoken Digits: Comparison of Normal and Impaired Listeners Using a Computer Model .................................................................................................................................. 2474
Matthew Robertson, Guy J. Brown, Wendy Lecluyse, Manasa Panda, Christine M. Tan

Evaluation of Bone-Conducted Ultrasonic Hearing-Aid Regarding Transmission of Paralinguistic Information: A Comparison with Cochlear Implant Simulator ........................................................................................................ 2478
Takayuki Kogomiya, Seiji Nakagawa

Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-Impaired Listeners ........................................................................ 2482
Tim Jurgens, Stefan Freidelau, Ralf M. Meyer, Birger Kollmeier, Thomas Brand

Does Sentence Complexity Interfere with Intelligibility in Noise? Evaluation of the Oldenburg Linguistically and Audiologically Controlled Sentence Test (OLACS) ................................................. 2486
Verena N. Uslar, Thomas Brand, Mirko Hanke, Rebecca Carroll, Esther Ruigendijk, Cornelia Hamann, Birger Kollmeier

Intelligibility Predictions for Speech Against Fluctuating Masker ..................................................................................... 2490
Juan-Pablo Ramirez, Hamed Kebabdar, Alexander Raske

An Effect of Formant Amplitude in Vowel Perception ........................................................................................................ 2494
Masashi Ito, Keiji Ohara, Akinori Ito, Masafumi Yano
Functional Imaging of Brain Regions Sensitive to Communication Sounds in Primates .......................................................... 2498
Christopher I. Parkin, Benjamin Wilson

Strategies for Statistical Spoken Language Understanding with Small Amount of Data — An Empirical Study ........................................ 2502
Ye-Yi Wang

Investigating Multiple Approaches for SLU Portability to a New Language ............................................................................. 2506
Bassam Jabbar, Laurent Besacier, Fabrice Lefevre

Learning Naturally Spoken Commands for a Robot ..................................................................................................................... 2510
Anja Austermann, Seiji Yamada, Kotaro Funakoshi, Makio Nishino

A Semi-Supervised Cluster-and-Label Approach for Utterance Classification ............................................................................ 2514
Amguro Albalate, Aparna Sachindranath, David Suen dermann, Wolfgang Minker

Classifying Dialog Acts in Human-Human and Human-Machine Spoken Conversations ................................................................. 2518
Silvia Quarteroni, Giuseppe Riccardi

Exploring Speaker Characteristics for Meeting Summarization .................................................................................................... 2522
Fei Liu, Yang Liu

Semi-Supervised Extractive Speech Summarization via Co-Training Algorithm .............................................................................. 2526
Shasha Xie, Hai Lin, Yang Liu

Extractive Summarization Using a Latent Variable Model ........................................................................................................... 2530
Asli Celikyilmaz, Dilek Hakkani-Tur

Hierarchical Classification for Speech-to-Speech Translation ..................................................................................................... 2534
Emil Eteläe, Panayiotis G. Georgiou, Shrikanth S. Narayanan

Rapid Development of Speech Translation Using Consecutive Interpretation ................................................................................. 2538
Matthias Paulik, Alex Waibel

Combining Many Alignments for Speech to Speech Translation .................................................................................................... 2542
Sameer R. Maskey, Steven J. Rennie, Bowen Zhou

Detecting Politeness and Efficiency in a Cooperative Social Interaction ......................................................................................... 2546
Paul M. Brunet, Marcela Charfuelan, Roderick Cowie, Marc Schroder, Hastings Donnan, Ellen Douglas-Cowie

Comparing Measures of Synchrony and Alignment in Dialogue Speech Timing with Respect to Turn-Taking Activity ......................... 2550
Nick Campbell, Stefan Scherer

Resources for Turn Competition in Overlap in Multi-Party Conversations: Speech Rate, Pausing and Duration ............................. 2554
Khiet P. Truong, Dirk Heylen

Disambiguating the Functions of Conversational Sounds with Prosody: The Case of 'Yeah' .............................................................. 2558
Prosody and Voice Quality of Vocal Social Signals: The Case of Dominance in Scenario Meetings ...................................................... 2562

Online SLU Model Adaptation with a Partial Oracle ..................................................................................................................... 2566
Pierre Gouab, Geraldine Dennault, Frederic Bechet, Lionel Delph in-Poulat

The Prosody of Swedish Conversational Grunts ......................................................................................................................... 2570
D. Norberg, J. Gustafson

Reliable Tracking Based on Speech Sample Salience of Vocal Cycle Length Perturbations ........................................................... 2574
C. Mertens, Francis Grenze, Lise Crevier-Buchman, Jean Schoentgen

Longitudinal Changes of Selected Voice Source Parameters ......................................................................................................... 2578
Hideki Kasuya, Hajime Yoshida, Satoshi Eiburu, Hiroki Mori

Automatic Perceptual Categorization of Disordered Connected Speech ............................................................................................ 2582
A. Alpan, Jean Schoentgen, Y. Maryn, Francis Grenze

Kinematic Analysis of Tongue Movement Control in Spastic Dysarthria ......................................................................................... 2586
Heejin Kim, Panying Rong, Torrey M. Loucks, Mark Hasegawa-Johnson

Pre- and Short-Term Posttreatment Vocal Functioning in Patients with Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy ........................................................................................................... 2590
Irene Jacoby, Lisette Van Der Molen, Maya Van Rossum, Frans Hilgers

Acoustic Analysis of Intonation in Parkinson's Disease .................................................................................................................... 2594
Joan K. Y. Ma, Rudiger Hoffmann

SAFE: A Statistical Algorithm for F0 Estimation for Both Clean and Noisy Speech ........................................................................ 2598
Jung Ok Hong, Patrick J. Wolfe

Robust and Efficient Pitch Estimation Using an Iterative ARMA Technique ..................................................................................... 2602
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Modeling of F0 Dynamics in Singing Voices Based on Gaussian Processes with Multiple Oscillation Bases</td>
<td>2606</td>
</tr>
<tr>
<td>Applying Geometric Source Separation for Improved Pitch Extraction in Human-Robot Interaction</td>
<td>2610</td>
</tr>
<tr>
<td>THU-SES1-O4: ASR: FEATURE EXTRACTION II</td>
<td></td>
</tr>
<tr>
<td>A Spectral LF Model Based Approach to Voice Source Parameterisation</td>
<td>2614</td>
</tr>
<tr>
<td>John Kane, Mark Kane, Christer Gobl</td>
<td></td>
</tr>
<tr>
<td>Glottal-Based Analysis of the Lombard Effect</td>
<td>2618</td>
</tr>
<tr>
<td>Thomas Dragman, Thierry Dutot</td>
<td></td>
</tr>
<tr>
<td>Hidden Logistic Linear Regression for Support Vector Machine Based Phone Verification</td>
<td>2622</td>
</tr>
<tr>
<td>Be Li, Khe Chai Sim</td>
<td></td>
</tr>
<tr>
<td>Jointly Optimized Discriminative Features for Speech Recognition</td>
<td>2626</td>
</tr>
<tr>
<td>Tim Ng, Bing Zhang, Long Nguyen</td>
<td></td>
</tr>
<tr>
<td>Invariant Integration Features Combined with Speaker-Adaptation Methods</td>
<td>2630</td>
</tr>
<tr>
<td>Florian Muller, Alfred Mertins</td>
<td></td>
</tr>
<tr>
<td>Multi Resolution Discriminative Models for Subvocalic Speech Recognition</td>
<td>2634</td>
</tr>
<tr>
<td>Mark Raugas, Vivek Kumar Kangarajan Sridhar, Rohit Prasad, Prem Natarajan</td>
<td></td>
</tr>
<tr>
<td>THU-SES1-P1: SPEAKER DIARIZATION</td>
<td></td>
</tr>
<tr>
<td>A Comparative Large Scale Study of MLP Features for Mandarin ASR</td>
<td>2638</td>
</tr>
<tr>
<td>Fabio Valente, Mathew Magajin Doo, Christian Plath, Suman V. Ravari, Wen Wang</td>
<td></td>
</tr>
<tr>
<td>Recognizing Cochlear Implant-Like Spectrally Reduced Speech with HMM-Based ASR: Experiments with MFCCs and PLP Coefficients</td>
<td>2642</td>
</tr>
<tr>
<td>Cong-Thanh Do, Dominique Pastor, Gael Le Lan, Andre Goalic</td>
<td></td>
</tr>
<tr>
<td>A Hybrid Approach to Online Speaker Diarization</td>
<td>2646</td>
</tr>
<tr>
<td>Carlos Faquero, Oriol Vinyals, Gerald Friedland</td>
<td></td>
</tr>
<tr>
<td>System Output Combination for Improved Speaker Diarization</td>
<td>2650</td>
</tr>
<tr>
<td>Simon Bozomet, Nicholas Evans, Xavier Anguera, Oriol Vinyals, Gerald Friedland, Corinne Fredouille</td>
<td></td>
</tr>
<tr>
<td>An Integrated Top-Down/Bottom-Up Approach to Speaker Diarization</td>
<td>2654</td>
</tr>
<tr>
<td>Simon Bozomet, Nicholas Evans, Corinne Fredouille, Dong Wang, Raphael Troncy</td>
<td></td>
</tr>
<tr>
<td>Advances in Fast Multistream Diarization Based on the Information Bottleneck Framework</td>
<td>2658</td>
</tr>
<tr>
<td>Deepu Vijayasenan, Fabio Valente, Herve Bourlard</td>
<td></td>
</tr>
<tr>
<td>Audio-Visual Synchronisation for Speaker Diarisation</td>
<td>2662</td>
</tr>
<tr>
<td>Gualia Garaz, Alfred Diekmann, Herve Bourlard</td>
<td></td>
</tr>
<tr>
<td>An Improved Cluster Model Selection Method for Agglomerative Hierarchical Speaker Clustering Using Incremental Gaussian Mixture Models</td>
<td>2666</td>
</tr>
<tr>
<td>Kyu J. Han, Shrikanth S. Narayanan</td>
<td></td>
</tr>
<tr>
<td>Dialog Prediction for a General Model of Turn-Taking</td>
<td>2670</td>
</tr>
<tr>
<td>Nigel G. Ward, Olic Fuentes, Alejandro Vega</td>
<td></td>
</tr>
<tr>
<td>THU-SES1-P2: MULTI-MODAL ASR, INCLUDING AUDIO-VISUAL ASR</td>
<td></td>
</tr>
<tr>
<td>Speaker Tracking in an Unsupervised Speech Controlled System</td>
<td>2674</td>
</tr>
<tr>
<td>Tobias Herbig, Franz Gerl, Wolfgang Minker</td>
<td></td>
</tr>
<tr>
<td>MultiBIC: An Improved Speaker Segmentation Technique for TV Shows</td>
<td>2678</td>
</tr>
<tr>
<td>Paula Lopez-Dero, Laura Docio-Fernandez, Carmen Garcia-Mateo</td>
<td></td>
</tr>
<tr>
<td>Automatic Speech Recognition for Assistive Writing in Speech Supplemented Word Prediction</td>
<td>2682</td>
</tr>
<tr>
<td>John-Paul Hossom, Tom Jakobs, Allen Baker, Susan Fager</td>
<td></td>
</tr>
<tr>
<td>Viseme-Dependent Weight Optimization for CHMM-Based Audio-Visual Speech Recognition</td>
<td>2686</td>
</tr>
<tr>
<td>Alexey Karpov, Andrey Rzohnin, Konstantin Markov, Milos Zelezny</td>
<td></td>
</tr>
<tr>
<td>Audio-Visual Anticipatory Coarticulation Modeling by Human and Machine</td>
<td>2690</td>
</tr>
<tr>
<td>Louis H. Terry, Karen Lwiescu, Janet B. Pierrchumbert, Aggelos K. Katsaggelos</td>
<td></td>
</tr>
<tr>
<td>Impact of Lack of Acoustic Feedback in EMG-Based Silent Speech Recognition</td>
<td>2694</td>
</tr>
<tr>
<td>Matthias Janke, Michael Wand, Tanja Schultz</td>
<td></td>
</tr>
<tr>
<td>Using Prosody to Improve Mandarin Automatic Speech Recognition</td>
<td>2698</td>
</tr>
<tr>
<td>Chong-Jia Ni, Wenyu Liu, Bo Xu</td>
<td></td>
</tr>
<tr>
<td>A Robust Audio-Visual Speech Recognition Using Audio-Visual Voice Activity Detection</td>
<td>2702</td>
</tr>
<tr>
<td>Satoshi Tamura, Masato Ishikawa, Takashi Hashiba, Shiri’echi Takeuchi, Satoru Hayamizu</td>
<td></td>
</tr>
<tr>
<td>Efficient Manycore CHMM Speech Recognition for Audiovisual and Multistream Data</td>
<td>2706</td>
</tr>
<tr>
<td>Dorothese Kolossa, Jie Chong, Steffen Zeiler, Kurt Keutzer</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Non-Audible Murmur Recognition Based on Fusion of Audio and Visual Streams</td>
<td>2714</td>
</tr>
<tr>
<td>Improved N-Gram Phonotactic Models for Language Recognition</td>
<td>2718</td>
</tr>
<tr>
<td>A Study of Term Weighting in Phonotactic Approach to Spoken Language Recognition</td>
<td>2720</td>
</tr>
<tr>
<td>Fuzzy Support Vector Machines for Age and Gender Classification</td>
<td>2722</td>
</tr>
<tr>
<td>Level of Interest Sensing in Spoken Dialog Using Multi-Level Fusion of Acoustic and Lexical Evidence</td>
<td>2724</td>
</tr>
<tr>
<td>Age and Gender Classification from Speech Using Decision Level Fusion and Ensemble Based Techniques</td>
<td>2726</td>
</tr>
<tr>
<td>A Hybrid Modeling Strategy for GMM-SVM Speaker Recognition with Adaptive Relevance Factor</td>
<td>2728</td>
</tr>
<tr>
<td>Exploiting Context-Dependency and Acoustic Resolution of Universal Speech Attribute Models in Spoken Language Recognition</td>
<td>2730</td>
</tr>
<tr>
<td>Hierarchical Multilayer Perceptron Based Language Identification</td>
<td>2732</td>
</tr>
<tr>
<td>The NIST 2010 Speaker Recognition Evaluation</td>
<td>2734</td>
</tr>
<tr>
<td>Bayesian Speaker Recognition Using Gaussian Mixture Model and Laplace Approximation</td>
<td>2736</td>
</tr>
<tr>
<td>What Else is New Than the Hammering Window? Robust MFCCs for Speaker Recognition via Multitapering</td>
<td>2738</td>
</tr>
<tr>
<td>Fast Computation of Speaker Characterization Vector Using MLLR and Sufficient Statistics in Anchor Model Framework</td>
<td>2740</td>
</tr>
<tr>
<td>Graph-Embedding for Speaker Recognition</td>
<td>2742</td>
</tr>
<tr>
<td>A Hybrid Modeling Strategy for GMM-SVM Speaker Recognition with Adaptive Relevance Factor</td>
<td>2744</td>
</tr>
<tr>
<td>THU-SES1-P4: SOURCE LOCALIZATION AND SEPARATION</td>
<td>2746</td>
</tr>
<tr>
<td>Robust Mixture Modeling Using T-Distribution: Application to Speaker ID</td>
<td>2748</td>
</tr>
<tr>
<td>A Variable Frame Length and Rate Algorithm Based on the Spectral Kurtosis Measure for Speaker Verification</td>
<td>2750</td>
</tr>
<tr>
<td>Near Field Sound Source Localization Based on Cross-Power Spectrum Phase Analysis with Multiple Microphones</td>
<td>2752</td>
</tr>
<tr>
<td>A Maximum a Posteriori Sound Source Localization in Reverberant and Noisy Conditions</td>
<td>2754</td>
</tr>
<tr>
<td>A DOA Estimation Algorithm Based on Equalization-Cancellation Theory</td>
<td>2756</td>
</tr>
<tr>
<td>Concurrent Speaker Localization Using Multi-Band Position-Pitch (M-PoPi) Algorithm with Spectro-Temporal Pre-Processing</td>
<td>2758</td>
</tr>
<tr>
<td>On Using Gaussian Mixture Model for Double-Talk Detection in Acoustic Echo Suppression</td>
<td>2760</td>
</tr>
<tr>
<td>Catalog-Based Single-Channel Speech-Music Separation</td>
<td>2762</td>
</tr>
<tr>
<td>Unvoiced Speech Segregation Based on CASA and Spectral Subtraction</td>
<td>2764</td>
</tr>
<tr>
<td>THU-SES2-S1: SPECIAL SESSION: INTERSPEECH 2010 PARALINGUISTIC CHALLENGE</td>
<td>2766</td>
</tr>
<tr>
<td>Unsupervised Sequential Organization for Cochannel Speech Separation</td>
<td>2768</td>
</tr>
<tr>
<td>The INTERSPEECH 2010 Paralinguistic Challenge</td>
<td>2770</td>
</tr>
<tr>
<td>Age and Gender Classification from Speech Using Decision Level Fusion and Ensemble Based Techniques</td>
<td>2772</td>
</tr>
<tr>
<td>Level of Interest Sensing in Spoken Dialog Using Multi-Level Fusion of Acoustic and Lexical Evidence</td>
<td>2774</td>
</tr>
<tr>
<td>Fuzzy Support Vector Machines for Age and Gender Classification</td>
<td>2776</td>
</tr>
<tr>
<td>Gender and Affect Recognition Based on GMM and GMM-UBM Modeling with Relevance MAP Estimation</td>
<td>2778</td>
</tr>
<tr>
<td>Age Recognition Based on Speech Signals Using Weights Supervector</td>
<td>2780</td>
</tr>
<tr>
<td>A Variable Frame Length and Rate Algorithm Based on the Spectral Kurtosis Measure for Speaker Verification</td>
<td>2782</td>
</tr>
<tr>
<td>On Using Gaussian Mixture Model for Double-Talk Detection in Acoustic Echo Suppression</td>
<td>2784</td>
</tr>
<tr>
<td>Catalog-Based Single-Channel Speech-Music Separation</td>
<td>2786</td>
</tr>
<tr>
<td>Unvoiced Speech Segregation Based on CASA and Spectral Subtraction</td>
<td>2788</td>
</tr>
<tr>
<td>THU-SES1-P3: SPEAKER AND LANGUAGE RECOGNITION</td>
<td>2790</td>
</tr>
<tr>
<td>Hierarchical Multilayer Perceptron Based Language Identification</td>
<td>2792</td>
</tr>
<tr>
<td>The NIST 2010 Speaker Recognition Evaluation</td>
<td>2794</td>
</tr>
<tr>
<td>Bayesian Speaker Recognition Using Gaussian Mixture Model and Laplace Approximation</td>
<td>2796</td>
</tr>
<tr>
<td>What Else is New Than the Hammering Window? Robust MFCCs for Speaker Recognition via Multitapering</td>
<td>2798</td>
</tr>
<tr>
<td>Fast Computation of Speaker Characterization Vector Using MLLR and Sufficient Statistics in Anchor Model Framework</td>
<td>2800</td>
</tr>
<tr>
<td>Graph-Embedding for Speaker Recognition</td>
<td>2802</td>
</tr>
<tr>
<td>A Hybrid Modeling Strategy for GMM-SVM Speaker Recognition with Adaptive Relevance Factor</td>
<td>2804</td>
</tr>
<tr>
<td>THU-SES1-P4: SOURCE LOCALIZATION AND SEPARATION</td>
<td>2806</td>
</tr>
<tr>
<td>Robust Mixture Modeling Using T-Distribution: Application to Speaker ID</td>
<td>2808</td>
</tr>
<tr>
<td>A Variable Frame Length and Rate Algorithm Based on the Spectral Kurtosis Measure for Speaker Verification</td>
<td>2810</td>
</tr>
<tr>
<td>On Using Gaussian Mixture Model for Double-Talk Detection in Acoustic Echo Suppression</td>
<td>2812</td>
</tr>
<tr>
<td>Catalog-Based Single-Channel Speech-Music Separation</td>
<td>2814</td>
</tr>
<tr>
<td>Unvoiced Speech Segregation Based on CASA and Spectral Subtraction</td>
<td>2816</td>
</tr>
</tbody>
</table>
Age and Gender Classification Using Fusion of Acoustic and Prosodic Features ................................................................. 2822
Hugo Meinedo, Isabel Trancoso

Brno University of Technology System for Interspeech 2010 Paralinguistic Challenge .......................................................... 2826
Marcel Kockmann, Lukas Burger, Jan Cernocky

Combining Five Acoustic Level Modeling Methods for Automatic Speaker Age and Gender Recognition .......................... 2830
Ming Li, Chi-Sang Jang, Kyu J. Han

Age and Gender Recognition Based on Multiple Systems — Early vs. Late Fusion ................................................................. 2834
Tobias Bocklet, Georg Stemmer, Viktor Zeissier, Elmar Soeh

THU-SES2-O1: NEW PARADIGMS IN ASR II

Automatic Speaker Age and Gender Recognition in the Car for Tailoring Dialog and Mobile Services ........................... 2838
Man-Ing Siu, Herbert Gish, Arthur Chan, William Belfield

Improved Topic Classification and Keyword Discovery Using an HMM-Based Speech Recognizer Trained Without Supervision ................................................................. 2842

An Analysis of Sparseness and Regularization in Exemplar-Based Methods for Speech Classification ................................. 2846
Dimitri Kanovsky, Tara N. Sainath, Bhuvana Ramabhadran, David Nahamoo

Investigation of Full-Sequence Training of Deep Belief Networks for Speech Recognition ................................................. 2850
Abdel-Bahman Mohamed, Dong Yu, L. Deng

Mandarin Tone Recognition Using Affine-Invariant Prosodic Features and Tone Posteriorgram ........................................... 2854
Yen-Bang Wang, Lin-Shan Lee

Continuous Speech Recognition with a TF-IDF Acoustic Model .......................................................................................... 2858
Geoffrey Zweig, Patrick Nguyen, Jasha Droppo, Alex Acero

THU-SES2-O2: SPOKEN LANGUAGE UNDERSTANDING AND SPOKEN LANGUAGE TRANSLATION II

SCARF: A Segmental Conditional Random Field Toolkit for Speech Recognition ................................................................. 2862
Geoffrey Zweig, Patrick Nguyen

Role of Language Models in Spoken Fluency Evaluation .................................................................................................. 2866
Ose D. Dossmah, Harish Doddada, Ashish Verma, Kavitha Vireswarab

Social Role Discovery from Spoken Language Using Dynamic Bayesian Networks ........................................................... 2870
Sibel Yaman, Dilek Hakimani-Tur, Gokhan Tur

Domain Adaptation and Compensation for Emotion Detection ............................................................................................ 2874
Michelle Hewlett Sanchez, Gokhan Tur, Luciana Ferrer, Dilek Hakimani-Tur

Phrase Alignment Confidence for Statistical Machine Translation ....................................................................................... 2878
Sanakaranarayanan Ananthakrishnan, Rohit Prasad, Prem Natarajan

Named-Entity Projection and Data-Driven Morphological Decomposition for Field Maintainable Speech-to-Speech Translation Systems ................................................................................... 2882
Ian R. Lane, Alex Waibel

THU-SES2-O3: SIGNAL PROCESSING FOR MUSIC AND SONG

Acoustic Correlates of Voice Quality Improvement by Voice Training .................................................................................. 2886
Kiyoko Atsaka, Junko Uemura, Tomoko Akiyuki

Phonetic Segmentation of Singing Voice Using MIDI and Parallel Speech .............................................................................. 2890
Minghui Dong, Paul Chan, Ling Cen, Haihouli Li, Jason Yao, Ping Jen Kua

A Singing Style Modeling System for Singing Voice Synthesizers ......................................................................................... 2894

A Fast Query by Humming System Based on Notes .............................................................................................................. 2898
Jingzhou Yang, Jia Liu, Wei-Qiang Zhang

Melody Pitch Estimation Based on Range Estimation and Candidate Extraction Using Harmonic Structure Model ................. 2902
Seokhwan Jo, Siyun Joo, Chang D. You

Modified Spatial Audio Object Coding Scheme with Harmonic Extraction and Elimination Structure for Interactive Audio Service ......................................................................................................... 2906
Jihoon Park, Kwangki Kim, Jeonggi Seo, Minsoo Hahn

THU-SES2-O4: MODELING FIRST LANGUAGE ACQUISITION

Modelling the Effect of Speaker Familiarity and Noise on Infant Word Recognition ................................................................. 2910
Christina Bergmann, Michele Gabian, Lou Boam

Unsupervised Learning of Vowels from Continuous Speech Based on Self-Organized Phoneme Acquisition Model .............. 2914
Kouki Miyazawa, Hideo Kituchi, Reiko Matsuoka

Learning Speaker Normalization Using Semisupervised Manifold Alignment ........................................................................ 2918
Andrew R. Plummer, Mary E. Beckman, Mikhail Belkin, Eric Fosler-Lussier, Benjamin Mannon
Fully Unsupervised Word Learning from Continuous Speech Using Transitional Probabilities of Atomic Acoustic Events ................................................................. 2922
Oiko, Johannes Rasinen
Language Acquisition and Cross-Modal Associations: Computational Simulation of the Result of Infant Studies ................................................................. 2926
L. Ten Bosch, Lou Boves
Active Word Learning Under Uncertain Input Conditions ................................................................. 2930
Maarten Versteegh, L. Ten Bosch, Lou Boves

THU-SES2-P1: ASR: ACOUSTIC MODELS III

Parallel Training of Neural Networks for Speech Recognition ................................................................. 2934
Karel Velely, Lukas Burget, Frantisek Grezl
The Use of Sense in Unsupervised Training of Acoustic Models for ASR Systems ................................. 2938
Rita Singh, Benjamin Lamberti, Bhiksha Raj
Boosted Mixture Learning of Gaussian Mixture HMMs for Speech Recognition ........................................ 2942
Jun Du, Yu Hui, Jia Jiang
On the Exploitation of Hidden Markov Models and Linear Dynamic Models in a Hybrid Decoder Architecture for Continuous Speech Recognition ................................................................. 2946
Volker Leunathan, Reinhold Haeb-Umbach
Context Dependent Modelling Approaches for Hybrid Speech Recognizers ................................................................. 2950
A Regularized Discriminative Training Method of Acoustic Models Derived by Minimum Relative Entropy Discrimination ................................................................. 2954
Yotaro Kabo, Shinji Watanabe, Atsushi Nakamura, Tetsunori Kobayashi
Decision Tree State Clustering with Word and Syllable Features ................................................................. 2958
Hank Liao, Chris Alberg, Michel Bacchiani, Olivier Siohan
A Duration Modeling Technique with Incremental Speech Rate Normalization ................................................................. 2962
Hiroshi Fujimura, Takashi Masuko, Mitsuyoshi Tachimori
Long Short-Term Memory Networks for Noise Robust Speech Recognition ................................................................. 2966
Martin Wolmier, Yang Sun, Florian Eyben, Bjorn Schuller
One-Model Speech Recognition and Synthesis Based on Articulatory Movement HMMs ................................................................. 2970
Tsunes Nitta, Takayuki Onoda, Masashi Kinura, Yuri Iribe, Kouichi Katsurada
Acoustic Modeling with Bootstrap and Restructuring for Low-Resource Languages ................................................................. 2974
Xiaodong Cui, Jian Xue, Pierre L. Daghn, Upendra V. Chandhuri, Bowen Zhou
Lecture Speech Recognition by Combining Word Graphs of Various Acoustic Models ................................................................. 2978
Tetsuo Kusaka, Kenkake Goto, Takashi Ito, Masaharu Kato
Semi-Parametric Trajectory Modelling Using Temporally Varying Feature Mapping for Speech Recognition ................................................................. 2982
Khe Chai Sim, Shi Lin Liu
Deep-Structured Hidden Conditional Random Fields for Phonetic Recognition ................................................................. 2986
Dong Yu, L. Deng
Semi-Supervised Learning for Improved Expression of Uncertainty in Discriminative Classifiers ................................................................. 2990
Jonathan Makkin, Jeff Bilmes
Modeling Posterior Probabilities Using the Linear Exponential Family ................................................................. 2994
Peder Olsen, Vaibhava Goel, Charles Micchelli, John R. Hershey

THU-SES2-P2: SPOKEN DIALOGUE SYSTEMS II

New Technique to Enhance the Performance of Spoken Dialogue Systems Based on Dialogue States-Dependent Language Models and Grammatical Rules ................................................................. 2998
Ramon Lopez-Cazorla, David Gred
A Stochastic Finite-State Transducer Approach to Spoken Dialog Management ................................................................. 3002
Lluis-F. Hartado, Joaquim Planells, Encarna Serrada, Emilio Sanchez, David Gred
Enhanced Monitoring Tools and Online Dialogue Optimization Merged into a New Spoken Dialogue System Design Experience ................................................................. 3006
Romain Laporche, Philippe Herrier, Ghislain Patois
Optimising a Handcrafted Dialogue System Design ................................................................. 3010
Romain Laporche, Ghislain Patois, Philippe Herrier
Utterance Selection for Speech Acts in a Cognitive Tourguide Scenario ................................................................. 3014
Felix Putze, Tanja Schulte
Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System ................................................................. 3018
Gabriel Parent, Maxine Eckenazi
Combining User Intention and Error Modeling for Statistical Dialog Simulators ................................................................. 3022
Silvia Quarteroni, Meritsell Gonzalez, Giuseppe Riccardi, Sebastian Vargas
Parallel Processing of Interruptions and Feedback in Companions Affective Dialogue System ................................................................. 3026
Jaaakko Hakulinen, Markku Turunen, Raul Santos De La Camara, Nigel Crook
Dynamic Language Modeling Using Bayesian Networks for Spoken Dialog Systems ................................................................. 3030
Antoine Roos, Neville Mehla, Deepak Ramchandran, Rakesh Gupta
Automatic Detection of Task-Incompletely Dialog for Spoken Dialog System Based on Dialog Act N-Gram ................................................................. 3034
Sunao Hara, Northide Kitatsu, Kazuya Takada
Conversations

Act Matrix .....................................................................................................................................................................................................3038
Wei-Bin Liang, Chung-Hsien Wu, Yu-Cheng Hsiao

Detection of Hot Spots in Poster Conversations Based on Reactive Tokens of Audience ..............................................................3042
Tatsuya Kawahara, Koubei Beito, Zhi-Qiang Chang, Katsumi Takanashi

Psychological Evaluation of a Group Communication Activation Robot in a Party Game ..............................................................3046
Yoichi Matsuyma, Shinya Fujie, Hikaru Taniyama, Tetsunori Kobayashi

Analyzing User Utterances in Barge-in-Able Spoken Dialogue System for Improving Identification Accuracy ..............................................................3050
Kyoko Matsuyma, Kazunori Komatani, Ryu Takeda, Toru Takahashi, Tetsuya Ogata, Hiroshi G. Okuno

Pitch Similarity in the Vicinity of Backchannels ..................................................................................................................3054
Mattias Heldner, Jens Edlund, Julius Hirschberg

A Rule-Based Backchannel Prediction Model Using Pitch and Pause Information ..............................................................................................3058
Khiet P. Truong, Ronald Poppe, Dirk Heylen

THU-SES2-P3: DISCUSSION AND DIALOGUE

Combining Text Categorization and Dialog Modeling for Speaker Role Identification on Call Center Conversations.................................................................................................................................................................................................................................................3062
Remi Lavallée, Chloe Clavel, Patrice Bellot, Marc El-Beze

Topic-Dependent N-Gram Models Based on Optimization of Context Lengths in LDA ..................................................................................................................3066
Akira Nakamura, Satoru Hayamizu

Expectations for Discourse Genre Identification: A Prosodic Study ..................................................................................................................3070
Nicolas Obin, Volker Dellwo, Anne Lacheret, Xavier Rodet

Dialogue Act Tagging and Segmentation with a Single Perceptron ..................................................................................................................3074
Ramon Granell, Stephen Pulman, Carlos-D. Martinez-Hinarejos, Jose Miguel Benedit

Improving the Readability of Class Lecture ASR Results Using a Confusion Network ..................................................................................3078
Yasuhisa Fujii, Kazumasa Yamamoto, Seiichi Nakagawa

THU-SES2-P4: VOICE ACTIVITY AND TURN DETECTION

Toward Detecting Voice Activity Employing Soft Decision in Second-Order Conditional MAP ..................................................................................................................3082
Sang-Kyun Kim, Jae-Hun Choi, Sang-Ick Kang, Ji-Hyun Song, Joon-Hyuk Chang

Voice Activity Detection in a Regularized Reproducing Kernel Hilbert Space ..................................................................................................................3086
Xugang Lu, Masashi Unoki, Ryosuke Iosimaki, Hitoshi Kawai, Satoshi Nakamura

A New VAD Framework Using Statistical Model and Human Knowledge Based Empirical Rule ..................................................................................................................3090
Ji Wu, Xiao-Lei Zhang, Wei Li

Adaptive High Accuracy Approaches to Speech Activity Detection in Noisy and Hostile Audio Environments ..................................................................................................................3094
Mark Huggins, Brett Smolenski, Aaron Lawson

Robust Voice Activity Detection in Stereo Recording with Crosstalk ..................................................................................................................3098
Prasanta Kumar Ghosh, Andreas Tsiartas, Panayiotis G. Georgiou, Shrikanth S. Narayanan

Voice Activity Detection Using Frame-Wise Model Re-Estimation Method Based on Gaussian Pruning with Weight Normalization ..................................................................................................................3102
Masakiyo Fujimoto, Shinji Watambe, Tomohiro Nakatani

Spectral Entropy-Based Voice Activity Detector for Videoconferencing Systems ..................................................................................................................3106
Bowon Lee, Debargha Mukherjee

The QUT-NOISE-TIMIT Corpus for the Evaluation of Voice Activity Detection Algorithms ..................................................................................................................3110
David Dean, Sridha Sridharan, Robert Vogt, Michael Mason

A Bayesian Approach to Voice Activity Detection Using Multiple Statistical Models and Discriminative Training ..................................................................................................................3114
Tao Yu, John H. L. Hansen

Noise Robust Voice Activity Detection Using Features Extracted from the Time-Domain Autocorrelation Function ..................................................................................................................3118
Houman Ghaemmaghami, Brendan Baker, Robert Vogt, Sridha Sridharan

VAD-Measure-Embedded Decoder with Online Model Adaptation ..................................................................................................................3122
Tasuku Oomish, Koji Iwano, Sadashki Furui

Robust Statistical Voice Activity Detection Using a Likelihood Ratio Sign Test ..................................................................................................................3126
Shiwen Deng, Jiqing Han

Automatic Turn Segmentation in Spoken Conversations ..................................................................................................................3130
Alexei V. Ivanov, Giuseppe Riccardi

Turn Taking-Based Conversation Detection by Using DOA Estimation ..................................................................................................................3134
Yoshi Kawaguchi, Masahito Togami, Yasunari Obuchi

Author Index