11th Annual Conference of the International Speech Communication Association 2010

(INTERSPEECH 2010)

Makuhari, Chiba, Japan
26-30 September 2010

Volume 1 of 4

ISBN: 978-1-61782-123-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Printed by Curran Associates, Inc. (2011)

International Speech Communications Association
c/o Emmanuelle Foxonet
Lous Tourils
F-66390 Baixas, France

Phone: 33 468 385 827
Fax: 49 228 735 639

secretariat@isca-speech.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

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
Chiu-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
Hosung Nam, Vikramjit Mitra, Mark Tiede, Elliot Saltzman, Louis Goldstein, Carol Espy-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, Masanori Morise, Toru Takahashi, Hideki Banno, Ryuichi Nisimura, Toshio Irino

Phase Equalization-Based Autoregressive Model of Speech Signals 42
Sadao Hiroya, Takemi Mochida

Articulatory-Functional Modeling of Speech Prosody: A Review 46
Yi Xu, Santitham Prom-On

Two New Estimation Methods for a Superpositional Intonation Model 50
Humberto M. Torres, Hansjorg Mixdorff, 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
Rajesh Balchandran, Leonid Rachevsky, Bhuvana Ramabhadran, Miroslav Novak

Optimizing Spoken Dialogue Management with Fitted Value Iteration 86
Senthilkumar Chandramohan, Matthieu Geist, Olivier Pietquin

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 Suendermann

Minimally Invasive Surgery for Spoken Dialog Systems ... 98
David Suendermann, Jackson Liscombe, Roberto Pieraccini

MON-SES2-O3: SPEECH PERCEPTION I: FACTORS INFLUENCING PERCEPTION

Detecting Categorical Perception in Continuous Discrimination Data ... 102
Paul Boersma, Katerina Chladkova
The Interrelation Between the Stimulus Range and the Number of Response Categories in Vowel Categorization ... 106
Titia Benders, Paola Escudero
The Relation Between Pitch Perception Preference and Emotion Identification ... 110
Marie Nilsenova, Martijn Goudbeek, Luuk Kempen
Competition in the Perception of Spoken Japanese Words ... 114
Takashi Otake, James M. McQueen, Anne Cutler
Influence of Musical Training on Perception of L2 Speech ... 118
Makiko Sadakata, Lotte Van Der Zanden, Kaoru Sekiyama
Full Body Aero-Tactile Integration in Speech Perception ... 122
Donald Derrick, Bryan Gick

MON-SES2-O4: PROSODY: MODELS

Nucleus Position Within the Intonation Phrase: A Typological Study of English, Czech and Hungarian ... 126
Tomas Dubeda, Katalin Mady
Focus-Sensitive Operator or Focus Inducer: Always and Only ... 130
Yong-Cheol Lee, Satoshi Nambu
F_0 Declination in English and Mandarin Broadcast News Speech ... 134
Jiahong Yuan, Mark Liberman
Frequency of Occurrence Effects on Pitch Accent Realisation ... 138
Katrin Schweitzer, Michael Walsh, Bernd Mobius, Hinrich Schutze
On the Automatic ToBI Accent Type Identification from Data ... 142
Cesar Gonzalez-Ferreras, Carlos Vivaracho-Pascual, David Escudero-Mancebo, Valentin Cardenoso-Payo
AuToBI --- A Tool for Automatic ToBI Annotation ... 146
Andrew Rosenberg

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

A Classifier-Based Target Cost for Unit Selection Speech Synthesis Trained on Perceptual Data ... 150
Volker Strom, Simon King
Applying Scalable Phonetic Context Similarity in Unit Selection of Concatenative Text-to-Speech ... 154
Wei Zhang, Xiaodong Cui
Speech Database Reduction Method for Corpus-Based TTS System ... 158
Mitsuaki Isogai, Hideyuki Mizuno
Automatic Error Detection for Unit Selection Speech Synthesis Using Log Likelihood Ratio Based SVM Classifier ... 162
Heng Lu, Zhen-Hua Ling, Si Wei, Lirong Dai, Ren-Hua Wang
Using Robust Viterbi Algorithm and HMM-Modeling in Unit Selection TTS to Replace Units of Poor Quality ... 166
Hanna Silen, Elina Helander, Jani Nurminen, Konsta Koppinen, Moncef Gabbouj
Automatic Detection of Abnormal Stress Patterns in Unit Selection Synthesis ... 170
Yeon-Jun Kim, Mark C. Beutnagel
Enhancements of Viterbi Search for Fast Unit Selection Synthesis ... 174
Daniel Tihelka, Jiri Kala, Jindrich Matousek
Accurate Pitch Marking for Prosodic Modification of Speech Segments ... 178
Thomas Ewender, Beat Pfister
A Novel Hybrid Approach for Mandarin Speech Synthesis ... 182
Shifeng Pan, Meng Zhang, Jianhua Tao
Modeling Liaison in French by Using Decision Trees ... 186
Josafa De Jesus Aguiar Pontes, Sadaoki Furui
Improvement on Plural Unit Selection and Fusion ... 190
Jian Luan, Jian Li
Improving Speech Synthesis of Machine Translation Output ... 194
Alok Parlikar, Alan W. Black, Stephan Vogel
Paraphrase Generation to Improve Text-to-Speech Synthesis ... 198
Ghislain Putois, Jonathan Chevelu, Cedric Boidin

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

Phone Mismatch Penalty Matrices for Two-Stage Keyword Spotting via Multi-Pass Phone Recognizer ... 202
Chang Woo Han, Shin Jae Kang, Chul Min Lee, Nam Soo Kim

English Spoken Term Detection in Multilingual Recordings 206
Petr Motlicek, Fabio Valente, Philip N. Garner
A Hybrid Approach to Robust Word Lattice Generation via Acoustic-Based Word Detection 210
Icksang Han, Chiyoun Park, Jeongmi Cho, Jeongsu Kim
Direct Observation of Pruning Errors (DOPE): A Search Analysis Tool 214
V. Steinbiss, Martin Sundermeyer, Hermann Ney
Direct Construction of Compact Context-Dependency Transducers from Data 218
David Rybach, Michael Riley
Incremental Composition of Static Decoding Graphs with Label Pushing 222
Miroslav Novak
A Novel Path Extension Framework Using Steady Segment Detection for Mandarin Speech Recognition 226
Zhanlei Yang, Wenju Liu
On the Relation of Bayes Risk, Word Error, and Word Posteriors in ASR 230
Ralf Schluter, Markus Nussbaum-Thom, Hermann Ney
Time Conditioned Search in Automatic Speech Recognition Reconsidered 234
D. Nolden, Hermann Ney, Ralf Schluter
Efficient Data Selection for Speech Recognition Based on Prior Confidence Estimation Using Speech and Context Independent Models 238
Satoshi Kobashikawa, Taichi Asami, Yoshikazu Yamaguchi, Hirokazu Masataki, Satoshi Takahashi
A Novel Confidence Measure Based on Marginalization of Jointly Estimated Error Cause Probabilities 242
Atsunori Ogawa, Atsushi Nakamura

MON-SES2-P3: SPECIAL-PURPOSE SPEECH APPLICATIONS

Evaluation of a Silent Speech Interface Based on Magnetic Sensing 246
Robin Hofe, Stephen R. Ell, Michael J. Fagan, James M. Gilbert, Phil D. Green, Roger K. Moore, Sergey I. Rybchenko
Advanced Speech Communication System for Deaf People 250
R. San-Segundo, V. Lopez, R. Martin, S. Lufti, J. Ferreiros, R. Cordoba, J. M. Pardo
Unsupervised Acoustic Model Adaptation for Multi-Origin Non Native ASR 254
Sethserey Sam, Eric Castelli, Laurent Besacier
Speech-Based Automated Cognitive Status Assessment 258
Dilek Hakkani-Tur, Dimitra Vergyri, Gokhan Tur
Speech Recognition with a Seamlessly Updated Language Model for Real-Time Closed-Captioning 262
Toru Imai, Shinichi Homma, Akio Kobayashi, Takahiro Oku, Shoei Sato
The Comparison Between the Deletion-Based Methods and the Mixing-Based Methods for Audio CAPTCHA Systems 266
Takuya Nishimoto, Takayuki Watanabe
Comparing Mono- & Multilingual Acoustic Seed Models for a Low e-Resourced Language: A Case-Study of Luxembourgish 270
Martine Adda-Decker, Lori Lamel, Natalie D. Snoeren
Manipulating Treacheoesophageal Speech 274
R. J. J. H. Van Son, Irene Jacobi, Frans Hilgers
Towards Mixed Language Speech Recognition Systems 278
David Imseng, Herve Bourlard, Mathew Magimai Doss
Voice Search for Development 282
Etienne Barnard, Johan Schalkwyk, Charl Van Heerden, Pedro J. Moreno
Cross-Cultural Investigation of Prosody in Verbal Feedback in Interactional Rapport 286
Gina-Anne Levow, Susan Duncan, Edward T. King
Multimodal Speaker Diarization Using Oriented Optical Flow Histograms 290
Mary Tai Knox, Gerald Friedland
Towards an ASR-Free Objective Analysis of Pathological Speech 294
Catherine Middag, Yvan Saeys, Jean-Pierre Martens

MON-SES2-P4: SPEECH ANALYSIS

Session Variability Contrasts in the MARP Corpus 298
Keith W. Godin, John H. L. Hansen
Estimation of Two-to-One Forced Selection Intelligibility Scores by Speech Recognizers Using Noise-Adapted Models 302
Kazuhiro Kondo, Yusuke Takano
Analysis of Gender Normalization Using MLP and VTLN Features 306
Thomas Schaaf, Florian Metze
Discovering an Optimal Set of Minimally Contrasting Acoustic Speech Units: A Point of Focus for Whole-Word Pattern Matching 310
Guillaume Aimetti, Roger K. Moore, L. Ten Bosch
Improvements to the Equal-Parameter BIC for Speaker Diarization 314
Themos Stafylakis, Xavier Anguera
A Multistream Multiresolution Framework for Phoneme Recognition 318
Nima Mesgarani, Samuel Thomas, Hynek Hermansky

Cluster Analysis of Differential Spectral Envelopes on Emotional Speech 322
Giampiero Salvi, Fabio Tesser, Enrico Zovato, Piero Cosi
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
Yuya Akita, 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
Jiulong Shan, Genqing Wu, Zhihong Hu, Xiliu Tang, Martin Jansche, 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 IIR NIST SRE 2008 and 2010 Summed Channel Speaker Recognition Systems 366
Hanwu Sun, Bin Ma, Chien-Lin Huang, Trung Hieu Nguyen, Haizhou Li
Speaker Characterization Using Long-Term and Temporal Information 370
Chien-Lin Huang, Hanwu 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 Convolutive Mixtures 390
Yasmina Benabderrahmane, Sid Ahmed Selouani, Douglas O'Shaughnessy
Online Gaussian Process for Nonstationary Speech Separation 394
Hsin-Lung Hsieh, Jen-Tzung Chien
Convexity and Fast Speech Extraction by Split Bregman Method 398
Meng Yu, Wenye Ma, Jack Xin, Stanley Osher
Reducing Musical Noise in Blind Source Separation by Time-Domain Sparse Filters and Split Bregman Method 402
Wenye Ma, Meng Yu, Jack Xin, Stanley Osher
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, Francois Mairesse, 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
Shuji 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
Qingju 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. Florescu, Lise Crevier-Buchman, Bruce Denby, Thomas Hueber, Antonia Colazo-Simon, Claire Pillot-Loiseau, Pierre Roussel, Cedric Gendrot, Sophie Quattrocchi
Comparison of HMM and TMDN Methods for Lip Synchronisation 454
Gregor Hofer, 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 Yanushevskaya, Christer Gobl, John Kane, Ailbhe Ni Chasaide
Modeling Perceived Vocal Age in American English 466
James D. Harnsberger, Rahul Shrivastav, W. S. Brown Jr.
Multivariate Analysis of Vocal Fatigue in Continuous Reading 470
Marie-Jose Caraty, Claude Montacie
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 Polzehl, Bjorn Schuller, Stefan Steidl
A Novel Feature Extraction Strategy for Multi-Stream Robust Emotion Identification 482
Gang Liu, Yun Lei, John H. L. Hansen
Setup for Acoustic-Visual Speech Synthesis by Concatenating Bimodal Units 486
Asterios Toutios, Utpala Musti, Slim Ouni, Vincent Colotte, Brigitte Wrobel-Dautcourt, Marie-Odile Berger
Towards Affective State Modeling in Narrative and Conversational Settings 490
Bart Jochems, Martha Larson, Roeland Ordelman, Ronald Poppe, Khiet P. Truong
Detection of Anger Emotion in Dialog Speech Using Prosody Feature and Temporal Relation of Utterances 494
Narichika Nomoto, Hirokazu Masataki, Osamu Yoshioka, Satoshi Takahashi
Gesture and Speech Coordination: The Influence of the Relationship Between Manual Gesture and Speech 498
Benjamin Roustan, Marion Dohen
Analysis and Detection of Cognitive Load and Frustration in Drivers' Speech 502
Hynek Boril, Seyed Omid Sadjadi, Tristan Kleinschmidt, John H. L. Hansen
Acoustic-Based Recognition of Head Gestures Accompanying Speech 506
Akira Sasou, Yasuharu Hashimoto, Katsuhiko Sakaue
Multimodal Dialog in the Car: Combining Speech and Turn-and-Push Dial to Control Comfort Functions 510
Sandro Castronovo, Angela Mahr, Margarita Pentcheva, Christian Muller
Hands Free Audio Analysis from Home Entertainment 514
Danil Korchagin, Philip N. Garner, Petr Motlicek
Affective Story Teller: A TTS System for Emotional Expressivity 518
Mostafa Al Masum Shaikh, Antonio Rui Ferreira Rebordao, 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 Ghai, 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
Ravichander Vipperla, Steve Renals, Joe Frankel
Discriminative Adaptation Based on Fast Combination of DMAP and DfMLLR 534
Lukas Machlica, Zbynek Zajic, Ludek Muller
Revisiting VTLN Using Linear Transformation on Conventional MFCC 538
D. R. Sanand, Ralf Schluter, Hermann Ney

Speaker Adaptation Based on Nonlinear Spectral Transform for Speech Recognition 542
Toyohiro Hayashi, Yoshihiko Nankaku, Akinobu Lee, Keiichi Tokuda
Speaker Adaptation Based on System Combination Using Speaker-Class Models 546
Tetsuo Kosaka, Takashi Ito, Masaharu Kato, Masaki Kohda
Speaker Adaptation in Transformation Space Using Two-Dimensional PCA 550
Yongwon Jeong, Young Rok Song, Hyung Soon Kim
On Speaker Adaptive Training of Artificial Neural Networks 554
Jan Trmal, Jan Zelinka, Ludek Muller
Model Synthesis for Band-Limited Speech Recognition 558
Yongjun He, Jiqing Han
Performance Estimation of Reverberant Speech Recognition Based on Reverberant Criteria RSR-D_n with Acoustic Parameters 562
Takahiro Fukumori, Masanori Morise, Takanobu Nishiura
A Novel Approach for Matched Reverberant Training of HMMs Using Data Pairs 566
Armin Sehr, Christian Hofmann, Roland Maas, Walter Kellermann
An Auditory Based Modulation Spectral Feature for Reverberant Speech Recognition 570
Harikrishna Maganti, Marco Matassoni
On the Potential of Channel Selection for Recognition of Reverberated Speech with Multiple Microphones 574
Martin Wolf, Climent Nadeu
An Improved Wavelet-Based Dereverberation for Robust Automatic Speech Recognition 578
Randy Gomez, Tatsuya Kawahara
Methods for Robust Speech Recognition in Reverberant Environments: A Comparison 582
Rico Petrick, Thomas Feher, Masashi Unoki, Rudiger Hoffmann

MON-SES3-P3: LANGUAGE LEARNING, TTS, AND OTHER APPLICATIONS

Integration of Multilayer Regression Analysis with Structure-Based Pronunciation Assessment 586
Masayuki Suzuki, Yu Qiao, Nobuaki Minematsu, Keikichi Hirose
Using Non-Native Error Patterns to Improve Pronunciation Verification 590
Joost Van Doremalen, Catia Cucchiarini, Helmer Strik
Regularized-MLLR Speaker Adaptation for Computer-Assisted Language Learning System 594
Dean Luo, Yu Qiao, Nobuaki Minematsu, Yutaka Yamauchi, Keikichi Hirose
Automatic Evaluation of English Pronunciation by Japanese Speakers Using Various Acoustic Features and Pattern Recognition Techniques 598
Kuniaki Hirabayashi, Seiichi Nakagawa
Decision Tree Based Tone Modeling with Corrective Feedbacks for Automatic Mandarin Tone Assessment 602
Hsien-Cheng Liao, Jiang-Chun Chen, Sen-Chia Chang, Ying-Hua Guan, Chin-Hui Lee
CASTLE: A Computer-Assisted Stress Teaching and Learning Environment for Learners of English as a Second Language 606
Jingli Lu, Ruili Wang, Liyanage C. De Silva, Yang Gao, Jia Liu
Automatic Reference Independent Evaluation of Prosody Quality Using Multiple Knowledge Fusions 610
Shen Huang, Hongyan Li, Shijin Wang, Jiaen Liang, Bo Xu
Landmark-Based Automated Pronunciation Error Detection 614
Su-Youn Yoon, Mark Hasegawa-Johnson, Richard Sproat
HMM Based TTS for Mixed Language Text 618
Zhiwei Shuang, Shiyin Kang, Yong Qin, Lirong Dai, Lianhong Cai
An Analysis of Language Mismatch in HMM State Mapping-Based Cross-Lingual Speaker Adaptation 622
Hui Liang, John Dines
Classroom Note-Taking System for Hearing Impaired Students Using Automatic Speech Recognition Adapted to Lectures 626
Tatsuya Kawahara, Norihiro Katsumaru, Yuya Akita, Shinsuke Mori
Exploring Web-Browser Based Runtimes Engines for Creating Ubiquitous Speech Interfaces 630
Paul R. Dixon, Sadaoki Furui

MON-SES3-P4: PITCH AND GLOTTAL-WAVEFORM ESTIMATION AND MODELING I

Efficient Three-Stage Pitch Estimation for Packet Loss Concealment 633
Xuejing Sun, Sameer Gadre
On Evaluation of the F_0 Estimation Based on Time-Varying Complex Speech Analysis 637
Keiichi Funaki
Pitch Estimation in Noisy Speech Based on Temporal Accumulation of Spectrum Peaks 641
Feng Huang, Tan Lee
Multi-Pitch Estimation by a Joint 2-D Representation of Pitch and Pitch Dynamics 645
Tianyu T. Wang, Thomas F. Quatieri
On the Effect of Fundamental Frequency on Amplitude and Frequency Modulation Patterns in Speech Resonances 649
Pirros Tsiakoulis, Alexandros Potamianos
Pitch Determination Using Autocorrelation Function in Spectral Domain 653
M. Shahidur Rahman, Tetsuya Shimamura

Chirp 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 Cinneide, David Dorran, Mikel Gainza, Eugene Coyle
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 Harish Reddy M., Kishore Prahallad, Suryakanth V. Gangashetty, B. Yegnanarayana
On Using Voice Source Measures in Automatic Gender Classification of Children's Speech 673
Gang Chen, Xue Feng, Yen-Liang Shue, Abeer Alwan

TUE-SES1-S1: SPECIAL SESSION: OPEN VOCABULARY SPOKEN DOCUMENT RETRIEVAL

Constructing Japanese Test Collections for Spoken Term Detection 677
Yoshiaki Itoh, Hiromitsu Nishizaki, Xinhui Hu, Hiroaki Nanjo, Tomoyosi Akiba, Tatsuya Kawahara, Seiichi Nakagawa, Tomoko Matsui, Yoichi Yamashita, Kiyoaki Aikawa
Japanese Spoken Term Detection Using Syllable Transition Network Derived from Multiple Speech Recognizers' Outputs 681
Satoshi Natori, Hiromitsu Nishizaki, Yoshihiro Sekiguchi
Combining Chinese Spoken Term Detection Systems via Side-Information Conditioned Linear Logistic Regression 685
Sha Meng, Wei-Qiang Zhang, Jia Liu
Metric Subspace Indexing for Fast Spoken Term Detection 689
Taisuke Kaneko, Tomoyosi Akiba
Unsupervised Spoken-Term Detection with Spoken Queries Using Segment-Based Dynamic Time Warping 693
Chun-An Chan, Lin-Shan Lee
Contextual Verification for Open Vocabulary Spoken Term Detection 697
Daniel Schneider, Timo Mertens, Martha Larson, Joachim Kohler
Augmented Set of Features for Confidence Estimation in Spoken Term Detection 701
Javier Tejedor, Doroteo T. Toledano, Miguel Bautista, Simon King, Dong Wang, Jose Colas
Cluster-Based Language Model for Spoken Document Retrieval Using NMF-Based Document Clustering 705
Xinhui Hu, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura

TUE-SES1-O1: ROBUST ASR

Asymptotically Exact Noise-Corrupted Speech Likelihoods 709
R. C. Van Dalen, M. J. F. Gales
A MMSE Estimator in Mel-Cepstral Domain for Robust Large Vocabulary Automatic Speech Recognition Using Uncertainty Propagation 713
Ramon Fernandez Astudillo, Reinhold Orglmeister
Non-Negative Matrix Factorization Based Compensation of Music for Automatic Speech Recognition 717
Bhiksha Raj, Tuomas Virtanen, Sourish Chaudhuri, Rita Singh
Feature versus Model Based Noise Robustness 721
Kris Demuynck, Xueru 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, Sung Joo Lee, Yunkeun 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, Driss Matrouf, Jean-Francois Bonastre, Jean Hennebert
Selecting Phonotactic Features for Language Recognition 737
Rong Tong, Bin Ma, Haizhou Li, Eng Siong Chng
Improved Language Recognition Using Mixture Components Statistics 741
Abualsoud Hanani, Michael Carey, Martin J. Russell
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 Abad, Isabel Trancoso, Ceu Viana
Dialect Recognition Using a Phone-GMM-Supervector-Based SVM Kernel 753
Fadi Biadsy, 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
Minh Duong, Jack Mostow
Predicting Word Accuracy for the Automatic Speech Recognition of Non-Native Speech 773
Su-Youn Yoon, Lei Chen, Klaus Zechner
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
Dongrui Wu, Thomas D. Parsons, Shrikanth S. Narayanan
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, Shrikanth S. Narayanan
A Cluster-Profile Representation of Emotion Using Agglomerative Hierarchical Clustering 797
Emily Mower, Kyu J. Han, Sungbok Lee, Shrikanth 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 Sorin
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 Sig 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. Russell
Global Variance Modeling on the Log Power Spectrum of LSPs for HMM-Based Speech Synthesis 825
Zhen-Hua Ling, Yu Hu, Lirong 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 Pilkington, Heiga Zen
Quantized HMMs for Low Footprint Text-to-Speech Synthesis 837
Alexander Gutkin, 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
Ayami Mase, Keiichiro Oura, Yoshihiko Nankaku, Keiichi Tokuda
An Unsupervised Approach to Creating Web Audio Contents-Based HMM Voices 849
Jinfu 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, Makoto Shozakai

Brazilian Portuguese Acoustic Model Training Based on Data Borrowing from Other Language ... 861
Kazuhiko Abe, Sakriani Sakti, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura

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

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

Cross-Lingual Acoustic Modeling for Dialectal Arabic Speech Recognition ... 873
Mohamed Elmahdy, Rainer Gruhn, Wolfgang Minker, Slim Abdennadher

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

Latent Perceptual Mapping: A New Acoustic Modeling Framework for Speech Recognition ... 881
Shiva Sundaram, Jerome R. Bellegarda

Unsupervised Model Adaptation on Targeted Speech Segments for LVCSR System Combination ... 885
Richard Dufour, Fethi Bougares, Yannick Esteve, Paul Deleglise

Incremental Word Learning Using Large-Margin Discriminative Training and Variance Floor Estimation ... 889
Irene Ayllon 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, Jort F. Gemmeke, Antti Hurmalainen

Similarity Scoring for Recognizing Repeated Out-of-Vocabulary Words ... 897
Mirko Hannemann, Stefan Kombrink, Martin Karafiat, Lukas Burget

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
Akiko 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 Electroglottograph and Inverse Filtered Glottal Waveforms ... 913
Daniel Bone, Samuel Kim, Sungbok Lee, Shrikanth S. Narayanan

Modal Analysis of Vocal Fold Vibrations Using Laryngotopography ... 917
Ken-Ichi Sakakibara, Hiroshi Imagawa, Miwako Kimura, Hisayuki Yokonishi, Niro Tayama

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

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

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

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

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

TUE-SES1-P4: SPEECH ENHANCEMENT

Multichannel Noise Reduction Using Low Order RTF Estimate ... 953
Subhojit Chakladar, Nam Soo Kim, Yu Gwang Jin, Tae Gyoon Kang

Reinforced Blocking Matrix with Cross Channel Projection for Speech Enhancement ... 957
Inho Lee, Jongsung 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
Yusuke Sato, Tetsuya Hoya, Hovagim Bakardjian, Andrzej Cichocki

Novel Probabilistic Control of Noise Reduction for Improved Microphone Array Beamforming 969
Jungpyo Hong, Seungho Han, Sangbae Jeong, Minsoo Hahn

Speech Enhancement Using Improved Generalized Sidelobe Canceller in Frequency Domain with Multi-Channel Postfiltering 973
Kai Li, Qiang Fu, Yonghong Yan

Close Speaker Cancellation for Suppression of Non-Stationary Background Noise for Hands-Free Speech Interface 977
Jani Even, Carlos Ishi, Hiroshi Saruwatari, Norihiro Hagita

Multi-Channel Iterative Dereverberation Based on Codebook Constrained Iterative Multi-Channel Wiener Filter 981
Ajay S., T. V. Sreenivas

Speaker-Dependent Mapping of Source and System Features for Enhancement of Throat Microphone Speech 985
J. M. Anand, R. M. Sri Harish, B. Yegnanarayana

An Analytic Modeling Approach to Enhancing Throat Microphone Speech Commands for Keyword Spotting 989
Jun Cai, Stefano Marini, Pierre Malarme, Francis Grenez, Jean Schoentgen

Single-Channel Speech Enhancement Using Kalman Filtering in the Modulation Domain 993
Stephen So, Kamil K. Wojcicki, Kuldip K. Paliwal

Integrated Feedback and Noise Reduction Algorithm in Digital Hearing Aids via Oscillation Detection 997
Miao Yao, Weiqian Liang

A Blind Signal-to-Noise Ratio Estimator for High Noise Speech Recordings 1001
Charles Mercier, Roch Lefebvre

TUE-SES2-S1: SPECIAL SESSION: FACT AND REPLICA OF SPEECH PRODUCTION

Estimation of Glottal Area Function Using Stereo-Endoscopic High-Speed Digital Imaging 1005
Hiroshi Imagawa, Ken-Ichi Sakakibara, Isao T. Tokuda, Mamiko Otsuka, Niro Tayama

Toward Aero-Acoustical Analysis of the Sibilant /s/: An Oral Cavity Modeling 1009
Kazunori Nozaki, Youhei Ohnishi, Takeshi Suda, Shigeo Wada, Shinji Shimojo

Effects of Wall Impedance on Transmission and Attenuation of Higher-Order Modes in Vocal-Tract Model 1013
Kunitoshi Motoki

Articulatory Synthesis and Perception of Plosive-Vowel Syllables with Virtual Consonant Targets 1017
Peter Birkholz, Bernd J. Kroger, Christiane Neuschaefer-Rube

Speech Robot Mimicking Human Articulatory Motion 1021
Kotaro Fukui, Toshihiro Kusano, Yoshikazu Mukaeda, Yuto Suzuki, Atsuo Takanishi, Masaaki Honda

Mechanical Vocal-Tract Models for Speech Dynamics 1025
Takayuki Arai

Prosodic Timing Analysis for Articulatory Re-Synthesis Using a Bank of Resonators with an Adaptive Oscillator 1029
Michael C. Brady

TUE-SES2-O1: ASR: LANGUAGE MODELING

Decoding with Shrinkage-Based Language Models 1033
Ahmad Emami, Stanley F. Chen, Abraham Ittycheriah, Hagen Soltau, Bing Zhao

Enhanced Word Classing for Model M 1037
Stanley F. Chen, Stephen M. Chu

Improved Neural Network Based Language Modelling and Adaptation 1041
J. Park, X. Liu, M. J. F. Gales, P. C. Woodland

Recurrent Neural Network Based Language Model 1045
Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cernocky, Sanjeev Khudanpur

Discriminative Language Modeling Using Simulated ASR Errors 1049
Preethi Jyothi, Eric Fosler-Lussier

Learning a Language Model from Continuous Speech 1053
Graham Neubig, Masato Mimura, Shinsuke Mori, Tatsuya Kawahara

TUE-SES2-O2: SPEAKER CHARACTERIZATION AND RECOGNITION II

Looking for Relevant Features for Speaker Role Recognition 1057
Benjamin Bigot, Julien Pinquier, Isabelle Ferrane, Regine Andre-Obrecht

Prosodic Speaker Verification Using Subspace Multinomial Models with Intersession Compensation 1061
Marcel Kockmann, Lukas Burget, Ondrej Glembek, Luciana Ferrer, Jan Cernocky

The Estimation and Kernel Metric of Spectral Correlation for Text-Independent Speaker Verification 1065
Eryu Wang, Kong Aik Lee, Bin Ma, Haizhou Li, Wu Guo, Lirong Dai

Improving Monaural Speaker Identification by Double-Talk Detection 1069
Rahim Saeidi, Pejman Mowlaee, Tomi Kinnunen, Zheng-Hua Tan, Mads Graesboll Christensen, Soren Holdt Jensen, Pasi Franti

Exploring Subsegmental and Suprasegmental Features for a Text-Dependent Speaker Verification in Distant Speech Signals 1073
B. Avinash, S. Guruprasad, B. Yegnanarayana

A Fast Implementation of Factor Analysis for Speaker Verification 1077
Qingsong Liu, Wei Huang, Dongxing Xu, Hongbin Cai, Beiqian Dai

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, Kees Janse

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. Ogbureke, 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 Mattheyses, Lukas Latacz, Werner Verhelst

Latent Affective Mapping: A Novel Framework for the Data-Driven Analysis of Emotion in Text 1117
Jerome R. Bellegarda

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 Espinosa, Michael White, Eric Fosler-Lussier, Chris Brew

TUE-SES2-P1: PROSODY: BASICS & APPLICATIONS

Acoustic Correlates of Meaning Structure in Conversational Speech 1129
Alexei V. Ivanov, Giuseppe Riccardi, S. Ghosh, 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
Margaret Zellers, 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
Mickael Rouvier, Richard Dufour, Georges Linares, Yannick Esteve

Speech Dominoes and Phonetic Convergence 1153
Gerard Bailly, Amelie Lelong

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
Hussein Hussein, 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
Suman V. Ravuri, 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, Eiichi Sueyoshi, Seiichi Nakagawa
Low-Dimensional Space Transforms of Posteriors in Speech Recognition 1193
Jan Zelinka, Jan Trmal, Ludek Muller
Hierarchical Bottle Neck Features for LVCSR 1197
Christian Plahl, Ralf Schluter, Hermann Ney
Hierarchical Neural Net Architectures for Feature Extraction in ASR 1201
Frantisek Grezl, Martin Karafiat
Mutual Information Analysis for Feature and Sensor Subset Selection in Surface Electromyography Based Speech Recognition 1205
Vivek Kumar Rangarajan Sridhar, Rohit Prasad, Prem Natarajan
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, Yosuke 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
Hsin-Yi Lin, Janice Fon
Effect of Spatial Separation on Speech-in-Noise Comprehension in Dyslexic Adults 1229
Marjorie Dole, Michel Hoen, Fanny Meunier
Speech Categorization Context Effects in Seven- to Nine-Month-Old Infants 1233
Ellen Marklund, Francisco Lacerda, 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, Jian 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 Otake

TUE-SES2-P4: SLP SYSTEMS

Comparison of Methods for Topic Classification in a Speech-Oriented Guidance System 1261
Rafael Torres, Shota Takeuchi, Hiromichi Kawanami, Tomoko Matsui, Hiroshi Saruwatari, Kiyohiro 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 Sethy, 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, Seiichi Nakagawa
Spoken Document Retrieval for Oral Presentations Integrating Global Document Similarities into Local Document Similarities 1285
Hiroaki Nanjo, Yusuke Iyonaga, Takehiko Yoshimi
Combining 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 Rosset, Herve Glotin
Prominence Based Scoring of Speech Segments for Automatic Speech-to-Speech Summarization 1297
Sree Harsha Yella, Vasudeva Varma, Kishore Prahallad
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 Assessment of Nonlinearly Processed Speech 1309
Anton Schlesinger, Marinus M. Boone
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-Turbin, Alexander Raake, Sebastian Moller
It Takes Two to Tango --- Assessing the Impact of Delay on Conversational Interactivity on Perceived Speech Quality 1321
Sebastian Egger, Raimund Schatz, Stefan Scherer
Comparison of Approaches for Instrumentally Predicting the Quality of Text-to-Speech Systems 1325
Sebastian Moller, Florian Hinterleitner, 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-O1: ASR: ACOUSTIC MODELS II

Boosting Systems for LVCSR 1341
George Saon, Hagen Soltau
Incorporating Sparse Representation Phone Identification Features in Automatic Speech Recognition Using Exponential Families 1345
Vaibhava Goel, Tara N. Sainath, Bhuvana Ramabhadran, Peder Olsen, David Nahamoo, Dimitri Kanevsky
Integrating MLP Features and Discriminative Training in Data Sampling Based Ensemble Acoustic Modeling 1349
Xin Chen, Yunxin Zhao
Semi-Supervised Training of Gaussian Mixture Models by Conditional Entropy Minimization 1353
Jui-Ting Huang, Mark Hasegawa-Johnson
A Study of Irrelevant Variability Normalization Based Training and Unsupervised Online Adaptation for LVCSR 1357
Guangchuan Shi, Yu Shi, Qiang Huo
Improvements to Generalized Discriminative Feature Transformation for Speech Recognition 1361
Roger Hsiao, Florian Metze, Tanja Schultz

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
Frederic Tantini, Christophe Cerisara, 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
Vahid Khanagha, Khalid Daoudi, Oriol Pont, Hussein Yahia
Phone Boundary Detection Using Sample-Based Acoustic Parameters 1397
You-Yu Lin, Yih-Ru Wang, Yuan-Fu Liao
HMM-Based Automatic Visual Speech Segmentation Using Facial Data 1401
Utpala Musti, Asterios Toutios, Slim Ouni, Vincent Colotte, Brigitte Wrobel-Dautcourt, Marie-Odile Berger
Bayes Factor Based Speaker Segmentation for Speaker Diarization 1405
D. Wang, Robert Vogt, Sridha 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, Tan Lee
Discriminative Training and Unsupervised Adaptation for Labeling Prosodic Events with Limited Training Data 1429
Raul Fernandez, Bhuvana Ramabhadran
Prosody for the Eyes: Quantifying Visual Prosody Using Guided Principal Component Analysis 1433
Erin Cvejic, Jeesun 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 Jourani, Khalid Daoudi, Regine Andre-Obrecht, Driss Aboutajdine
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 GMMSVM-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
Donglai Zhu, Bin Ma, Kong Aik Lee, Cheung-Chi Leung, Haizhou Li
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
Jouni Pohjalainen, Rahim Saeidi, Tomi Kinnunen, Paavo Alku
The Use of Subvector Quantization and Discrete Densities for Fast GMM Computation for Speaker Verification 1481
Guoli Ye, Brian Mak

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

Parallel Lexical-Tree Based LVCSR on Multi-Core Processors 1485
Naveen Parihar, Ralf Schluter, David Rybach, Eric A. Hansen
Exploring Recognition Network Representations for Efficient Speech Inference on Highly Parallel Platforms 1489
Jike Chong, Ekaterina Gonina, Kisun You, Kurt Keutzer
WFST Compression for Automatic Speech Recognition 1493
Diamantino Caseiro
Speech Recognizer Optimization Under Speed Constraints 1497
Ivan Bulyko
The 2010 CMU GALE Speech-to-Text System 1501
Florian Metze, Roger Hsiao, Qin Jin, Udhyakumar Nallasamy, Tanja Schultz
Speaker Diarization in Meeting Audio for Single Distant Microphone 1505
Tin Lay Nwe, Hanwu Sun, Bin Ma, Haizhou Li
Extending the Punctuation Module for European Portuguese 1509
Fernando Batista, Helena Moniz, Isabel Trancoso, Hugo Meinedo, Ana Isabel Mata, Nuno Mamede
Utilizing a Noisy-Channel Approach for Korean LVCSR 1513
Sakriani Sakti, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura
The RWTH 2009 Quaero ASR Evaluation System for English and German 1517
Markus Nussbaum-Thom, Simon Wiesler, Martin Sundermeyer, Christian Plahl, Stefan Hahn, Ralf Schluter, Hermann Ney

TUE-SES3-P3: PHONETICS

When is Indexical Information About Speech Activated? Evidence from a Cross-Modal Priming Experiment 1521
Benjamin Munson, Renata Solum

The Influence of Actual and Perceived Sexual Orientation on Diadochokinetic Rate in Women and Men 1525
Benjamin Munson

Laryngealization and Features for Chinese Tonal Recognition 1529
Kristine M. Yu

Production and Perception of Vietnamese Short Vowels in V1V2 Context 1533
Viet Son Nguyen, Eric Castelli, Rene Carre

Measuring Basic Tempo Across Languages and Some Implications for Speech Rhythm 1537
Gertraud Fenk-Oczlon, August Fenk

Durational Structure of Japanese Single/Geminate Stops in Three- and Four-Mora Words Spoken at Varied Rates 1541
Yukari Hirata, Shigeaki Amano

Distribution and Trichotomic Realization of Voiced Velars in Japanese --- An Experimental Study 1545
Shin-Ichiro Sano, Tomohiko Ooigawa

Specification in Context --- Devoicing Processes in Polish, French, American English and German Sonorants 1549
Jagoda Sieczkowska, Bernd Mobius, Grzegorz Dogil

Phonetic Imitation of Japanese Vowel Devoicing 1553
Kuniko Nielsen

VOLUME 3

Post-Aspiration in Standard Italian: Some First Cross-Regional Acoustic Evidence 1557
Mary Stevens, John Hajek

Articulatory Grounding of Southern Salentino Harmony Processes 1561
Mirko Grimaldi, Andrea Calabrese, Francesco Sigona, Luigina Garrapa, Bianca Sisinni

Effects of Accent Typicality and Phonotactic Frequency on Nonword Immediate Serial Recall Performance in Japanese 1565
Yuuki Tanida, Taiji Ueno, Satoru Saito, Matthew A. Lambon Ralph

How Abstract Is Phonetics? 1568
Osamu Fujimura

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
Michael I. Proctor, Daniel Bone, Athanasios Katsamanis, Shrikanth S. Narayanan

Improved Real-Time MRI of Oral-Velar Coordination Using a Golden-Ratio Spiral View Order 1580
Yoon-Chul Kim, Shrikanth S. Narayanan, Krishna S. Nayak

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

Adaptation of a Tongue Shape Model by Local Feature Transformations 1596
Chao Qin, Miguel A. Carreira-Perpinan, Mohsen Farhadloo

Vocal Tract Contour Analysis of Emotional Speech by the Functional Data Curve Representation 1600
Sungbok 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 Iskarous

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

Estimation of Speech Lip Features from Discrete Cosinus Transform 1612
Zuheng Ming, Denis Beautemps, Gang Feng, Sebastien Schmerber

Autoregressive Modelling for Linear Prediction of Ultrasonic Speech 1616
Farzaneh Ahmadi, Ian V. McLoughlin, Hamid R. Sharifzadeh

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 Arai, 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-O1: ASR: ACOUSTIC MODEL ADAPTATION

Prior Information for Rapid Speaker Adaptation 1644
C. Breslin, K. K. Chin, M. J. F. Gales, Kate Knill, H. Xu

Discriminative Adaptation for Log-Linear Acoustic Models 1648
Jonas Loof, Ralf Schluter, Hermann Ney

Automatic Speech Recognition of Multiple Accented English Data 1652
Dimitra Vergyri, Lori Lamel, Jean-Luc Gauvain

Shrinkage Model Adaptation in Automatic Speech Recognition 1656
Jinyu Li, Yu Tsao, Chin-Hui Lee

Unscented Transform with Online Distortion Estimation for HMM Adaptation 1660
Jinyu Li, Dong Yu, Yifan Gong, L. Deng

HMM Adaptation Using Linear Spline Interpolation with Integrated Spline Parameter Training for Robust Speech Recognition 1664
Michael L. Seltzer, Alex Acero

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 Troncy

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
Juhan Nam, Gautham J. Mysore, Joachim Ganseman, Kyogu Lee, Jonathan S. Abel

Fast Least-Squares Solution for Sinusoidal, Harmonic and Quasi-Harmonic Models 1700
Georgios Tzedakis, 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 Svendsen

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

WED-SES1-O4: VOICE CONVERSION

Maximum a posteriori Voice Conversion Using Sequential Monte Carlo Methods 1716
Elina Helander, Hanna Silen, Joaquin 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, Takao Kobayashi

Probabilistic Integration of Joint Density Model and Speaker Model for Voice Conversion 1728
Daisuke Saito, Shinji Watanabe, Atsushi Nakamura, Nobuaki Minematsu
Text-Independent F0 Transformation with Non-Parallel Data for Voice Conversion 1732
Zhi-Zheng Wu, Tomi Kinnunen, Eng Siong Chng, 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 Svantesson, 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
Jianjing Kuang
Text-Based Unstressed Syllable Prediction in Mandarin 1752
Ya Li, Jianhua 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, Tulika 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
Sankalan 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 Standard Chinese (Beijing) 1788
Na Zhi, Daniel Hirst, 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 Fujisaki 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 Momtazi, Friedrich Faubel, Dietrich Klakow
Impact of Word Classing on Shrinkage-Based Language Models 1804
Ruhi Sarikaya, Stanley F. Chen, Abhinav Sethy, Bhuvana Ramabhadran
Combination of Probabilistic and Possibilistic Language Models 1808
Stanislas Oger, Vladimir Popescu, Georges Linares
On-Demand Language Model Interpolation for Mobile Speech Input 1812
Brandon Ballinger, Cyril Allauzen, 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 Kurimo
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
Markpong Jongtaveesataporn, Sadaoki Furui
Augmented Context Features for Arabic Speech Recognition 1832
Ahmad Emami, Hong-Kwang J. Kuo, Imed Zitouni, Lidia Mangu
A Statistical Segment-Based Approach for Spoken Language Understanding 1836
Lucia Ortega, Isabel Galiano, Lluis-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, Janise Shanley
Linguistic Rhythm in Foreign Accent 1848
Jiahong 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, Yoshinori Sagisaka
Foreign Accent Matters Most When Timing is Wrong 1854
Chiharu Tsurutani
Effects of Korean Learners' Consonant Cluster Reduction Strategies on English Speech Recognition Performance 1858
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 Masuda, Takayuki Arai
Perception of Estonian Vowel Categories by Native and Non-Native Speakers 1870
Lya Meister, Einar Meister
Spoken English Assessment System for Non-Native Speakers Using Acoustic and Prosodic Features 1874
Qin Shi, Kun Li, Shilei Zhang, Stephen M. Chu, Ji Xiao, Zhijian Ou
Russian Infants and Children's Sounds and Speech Corpuses for Language Acquisition Studies 1878
Elena E. Lyakso, Olga V. Frolova, Anna V. Kurazhova, Julia S. Gaikova
Language-Specific Influence on Phoneme Development: French and Drehu Data 1882
Julia Monnin, Helene Loevenbruck
Did you Say 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 T^3, Juicer, HDecode and Sphinx3 Decoders 1890
Josef R. Novak, Paul R. Dixon, Sadaoki Furui
Tracter: A Lightweight Dataflow Framework 1894
Philip N. Garner, John Dines
Verifying Pronunciation Dictionaries Using Conflict Analysis 1898
Marelie H. Davel, Febe De Wet
Automatic Estimation of Transcription Accuracy and Difficulty 1902
Brandon C. Roy, Soroush 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, Kaisuke Nakajima, Linne 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, Dongning Wang, Jinsong Zhang, Ziyu Xiong
How Children Acquire Situation Understanding Skills?: A Developmental Analysis Utilizing Multimodal Speech Behavior Corpus 1926
Shogo Ishikawa, Shinya Kiriyama, Yoichi Takebayashi, Shigeyoshi Kitazawa
The Influence of Expertise and Efficiency on Modality Selection Strategies and Perceived Mental Effort 1930
Ina Wechsung, Stefan Schaffer, Robert Schleicher, Anja Naumann, Sebastian Moller
Parameters Describing Multimodal Interaction --- Definitions and Three Usage Scenarios 1934
Christine Kuhnel, Benjamin Weiss, Sebastian Moller
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 Gros
Recognition of Spontaneous Conversational Speech Using Long Short-Term Memory Phoneme Predictions 1946
Martin Wollmer, Florian Eyben, Bjorn Schuller, 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
Ladan Golipour, Douglas O'Shaughnessy
Online Adaptive Learning for Speech Recognition Decoding 1958
Jeff Bilmes, 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 I. Watson
On Generating Combilex 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 Godde, Sebastian Moller
A New Multichannel Multi Modal Dyadic Interaction Database 1982
Viktor Rozgic, 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, Tien-Ping Tan, Eng Siong 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-Osuna
Investigating Articulatory Setting --- Pauses, Ready Position, and Rest --- Using Real-Time MRI 1994
Vikram Ramanarayanan, Dani Byrd, Louis Goldstein, Shrikanth S. Narayanan
Articulatory Inversion of American English /ɹ/ by Conditional Density Modes 1998
Chao Qin, Miguel A. Carreira-Perpinan
Can Tongue Be Recovered from Face? The Answer of Data-Driven Statistical Models 2002
Atef Ben Youssef, Pierre Badin, Gerard Bailly
Phrase-Medial Vowel Devoicing in Spontaneous French 2006
Francisco Torreira, Mirjam Ernestus
Exploring the Mechanism of Tonal Contraction in Taiwan Mandarin 2010
Chierh Cheng, Yi Xu, Michele Gubian

WED-SES2-O4: PARALANGUAGE & COGNITION

Voice Attributes Affecting Likability Perception 2014
Benjamin Weiss, Felix Burkhardt
Turn-Alignment Using Eye-Gaze and Speech in Conversational Interaction 2018
Kristiina Jokinen, Kazuaki Harada, Masafumi Nishida, Seiichi 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 Goudbeek, 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-SES2-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
Takeshi Yamada, Tomohiro Nakajima, Nobuhiko Kitawaki, Shoji Makino
Estimating Noise from Noisy Speech Features with a Monte Carlo Variant of the Expectation Maximization Algorithm 2046
Friedrich Faubel, 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 Mushtaq, Yu Tsao, Chin Hui-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 Badiezadegan, 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
Wooil Kim, Jun-Won Suh, 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 Saito, Yoshihiko Nankaku, Akinobu Lee, Keiichi Tokuda
A Comparative Study of Noise Estimation Algorithms for VTS-Based Robust Speech Recognition 2090
Yong Zhao, Biing-Hwang Juang
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 Drugman, Thierry Dutoit
Acoustic Feature Diversity and Speaker Verification 2110
R. Padmanabhan, Hema A. Murthy
A Discriminative Performance Metric for GMM-UBM Speaker Identification 2114
Omid Dehzangi, Bin Ma, Eng Siong 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, Driss 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
Axel Robel
A Phonetic Alternative to Cross-Language Voice Conversion in a Text-Dependent Context: Evaluation of Speaker Identity 2150
Kayoko Yanagisawa, Mark Huckvale
Evaluation of Speaker Mimic Technology for Personalizing SGD Voices 2154
Esther Klabbers, 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 Saruwatari, 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, Yijian 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, Miaomiao Wang, Keikichi Hirose, 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 Krstulovic, Kate Knill
HMM-Based Text-to-Articulatory-Movement Prediction and Analysis of Critical Articulators 2194
Zhen-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, Taufiq Hasan, John H. L. Hansen
Detecting Novel Objects in Acoustic Scenes Through Classifier Incongruence 2206
Jorg-Hendrik Bach, Jorn Anemuller
A Multidomain Approach for Automatic Home Environmental Sound Classification 2210
Stavros Ntalampiras, Ilyas 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, Ilyas Potamitis, Nikos Fakotakis
Lightly Supervised Recognition for Automatic Alignment of Large Coherent Speech Recordings 2222
Norbert Braunschweiler, M. J. F. Gales, Sabine Buchholz
Incremental Diarization of Telephone Conversations 2226
Oshry Ben-Harush, Itshak Lapidot, Hugo Guterman
Audio Analytics by Template Modeling and 1-Pass DP Based Decoding 2230
Srikanth Cherla, V. Ramasubramanian
Perceptual Wavelet Decomposition for Speech Segmentation 2234
Mariusz Ziolko, Jakub Galka, Bartosz Ziolko, Tomasz Drwiega
A Comparative Study of Constrained and Unconstrained Approaches for Segmentation of Speech Signal 2238
Venkatesh Keri, Kishore Prahallad
Automatic Discriminative Measurement of Voice Onset Time 2242
Morgan Sonderegger, Joseph Keshet
Selective Gammatone Filterbank Feature for Robust Sound Event Recognition 2246
Yi Ren Leng, Huy Dat Tran, Norihide Kitaoka, 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
Tara N. Sainath, Bhuvana Ramabhadran, David Nahamoo, Dimitri Kanevsky, Abhinav Sethy
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, Ulpu 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. V. S. Sivaram, Sriram Ganapathy, Hynek Hermansky

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

FSM-Based Pronunciation Modeling Using Articulatory Phonological Code 2274
Chi Hu, Xiaodan Zhuang, Mark Hasegawa-Johnson
Detailed Pronunciation Variant Modeling for Speech Transcription 2278
Denis Jouvet, Dominique Fohr, Irina Illina
A Minimum Classification Error Approach to Pronunciation Variation Modeling of Non-Native Proper Names 2282
Line Adde, Bert Reveil, Jean-Pierre Martens, Torbjorn Svendsen
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

Learning New Word Pronunciations from Spoken Examples 2294
Ibrahim Badr, Ian McGraw, James Glass

WED-SES3-O2: SPEAKER RECOGNITION AND DIARIZATION

Phonetic Subspace Mixture Model for Speaker Diarization 2298
I-Fan Chen, Shih-Sian Cheng, Hsin-Min Wang
Overlap Detection for Speaker Diarization by Fusing Spectral and Spatial Features 2302
Martin Zelenak, Carlos Segura, Javier Hernando
Floor Holder Detection and End of Speaker Turn Prediction in Meetings 2306
Alfred Dielmann, Giulia Garau, Herve Bourlard
Confidence Measures for Speaker Segmentation and Their Relation to Speaker Verification 2310
Carlos Vaquero, Alfonso Ortega, Jesus Villalba, Antonio Miguel, Eduardo Lleida
Decoupling Session Variability Modelling and Speaker Characterisation 2314
Anthony Larcher, Christophe Levy, Driss Matrouf, Jean-Francois Bonastre
Incorporating MAP Estimation and Covariance Transform for SVM Based Speaker Recognition 2318
Cheung-Chi Leung, Donglai Zhu, Kong Aik Lee, Bin Ma, Haizhou Li

WED-SES3-O3: SPEECH AND AUDIO CLASSIFICATION

Single-Speaker/Multi-Speaker Co-Channel Speech Classification 2322
Stephane Rossignol, Olivier Pietquin
Discriminative Training for Hierarchical Clustering in Speaker Diarization 2326
Oriol Vinyals, Gerald Friedland, Nelson Morgan
GMM-UBM Based Open-Set Online Speaker Diarization 2330
Jurgen Geiger, Frank Wallhoff, Gerhard Rigoll
A Segment-Based Non-Parametric Approach for Monophone Recognition 2334
Ladan Golipour, Douglas O'Shaughnessy
A Fast One-Pass-Training Feature Selection Technique for GMM-Based Acoustic Event Detection with Audio-Visual Data 2338
Taras Butko, Climent Nadeu
Effects of Modelling Within- and Between-Frame Temporal Variations in Power Spectra on Non-Verbal Sound Recognition 2342
Nobuhide Yamakawa, Tetsuro Kitahara, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno

VOLUME 4

WED-SES3-O4: EMOTION RECOGNITION

On the Importance of Glottal Flow Spectral Energy for the Recognition of Emotions in Speech 2346
Ling He, Margaret Lech, Nicholas Allen
Real-Life Emotion-Related States Detection in Call Centers: A Cross-Corpora Study 2350
Laurence Devillers, Christophe Vaudable, Clement Chastagnol
Multi-Class and Hierarchical SVMs for Emotion Recognition 2354
Ali Hassan, Robert I. Damper
Determining Optimal Features for Emotion Recognition from Speech by Applying an Evolutionary Algorithm 2358
David Hubner, Bogdan Vlasenko, Tobias Grosser, Andreas Wendemuth
Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression Using Bidirectional LSTM Modeling 2362
Martin Wollmer, Angeliki Metallinou, Florian Eyben, Bjorn Schuller, Shrikanth S. Narayanan
Data-Dependent Evaluator Modeling and Its Application to Emotional Valence Classification from Speech 2366
Kartik Audhkhasi, Shrikanth S. Narayanan

WED-SES3-P1: SPEECH CODING, MODELING, AND TRANSMISSION

Modelling Speech Line Spectral Frequencies with Dirichlet Mixture Models 2370
Zhanyu Ma, Arne Leijon
PDF-Optimized LSF Vector Quantization Based on Beta Mixture Models 2374
Zhanyu Ma, Arne Leijon
Non-Linear Predictive Vector Quantization of Feature Vectors for Distributed Speech Recognition 2378
Jose Enrique Garcia, Alfonso Ortega, Antonio Miguel, Eduardo Lleida
Superwideband Extension of G.718 and G.729.1 Speech Codecs 2382
Lasse Laaksonen, Mikko Tammi, Vladimir Malenovsky, Tommy Vaillancourt, Mi Suk Lee, Tomofumi Yamanashi, Masahiro Oshikiri, Claude Lamblin, Balazs Kovesi, Lei Miao, Deming Zhang, Jon Gibbs, Holly Francois
A Multipulse FEC Scheme Based on Amplitude Estimation for CELP Codecs Over Packet Networks 2386
Jose L. Carmona, Angel M. Gomez, Antonio M. Peinado, Jose L. Perez-Cordoba, Jose A. Gonzalez

Voice Quality Evaluation of Recent Open Source Codecs ... 2390
Anssi Ramo, Henri Toukomaa
Efficient HMM-Based Estimation of Missing Features, with Applications to Packet Loss Concealment ... 2394
Bengt J. Borgstrom, Per H. Borgstrom, Abeer Alwan
Speech Inventory Based Discriminative Training for Joint Speech Enhancement and Low-Rate Speech Coding ... 2398
Xiaoqiang Xiao, Robert M. Nickel
Quality-Based Playout 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
Heikki Rasilo, Unto K. Laine, Okko Rasanen

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

Improving Back-Off Models with Bag of Words and Hollow-Grams ... 2418
Benjamin Lecouteux, Raphael Rubino, Georges Linares
Study on Interaction Between Entropy Pruning and Kneser-Ney Smoothing ... 2422
Ciprian Chelba, Thorsten Brants, Will Neveitt, Peng Xu
Dynamic Language Model Adaptation Using Keyword Category Classification ... 2426
Hitoshi Yamamoto, Ken Hanazawa, Kiyokazu Miki, Koichi Shinoda
Integration of Cache-Based Model and Topic Dependent Class Model with Soft Clustering and Soft Voting ... 2430
Welly Naptali, Masatoshi Tsuchiya, Seiichi Nakagawa
Conditional Models for Detecting Lambda-Functions in a Spoken Language Understanding System ... 2434
Frederic Duvert, Renato De Mori
Novel Weighting Scheme for Unsupervised Language Model Adaptation Using Latent Dirichlet Allocation ... 2438
Akmal Haidar, Douglas O'Shaughnessy
Automatic Speech Recognition System Channel Modeling ... 2442
Qun Feng Tan, Kartik Audhkhasi, Panayiotis G. Georgiou, Emil Ettelaie, Shrikanth S. Narayanan
Round-Robin Discrimination Model for Reranking ASR Hypotheses ... 2446
Takanobu Oba, Takaaki Hori, Atsushi Nakamura
On-the-Fly Lattice Rescoring for Real-Time Automatic Speech Recognition ... 2450
Hasim Sak, Murat Saraclar, Tunga 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, Kailash Patil, Sriram 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. Beeston, Guy J. Brown
Predicting Human Perception and ASR Classification of Word-Final [t] by Its Acoustic Sub-Segmental Properties ... 2466
Barbara Schuppler, Mirjam Ernestus, Wim 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 Yoshioka, 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 Kagomiya, Seiji Nakagawa
Challenging the Speech Intelligibility Index: Macroscopic vs. Microscopic Prediction of Sentence Recognition in Normal and Hearing-Impaired Listeners ... 2482
Tim Jurgens, Stefan Fredelake, 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 Ketabdar, Alexander Raake
An Effect of Formant Amplitude in Vowel Perception ... 2494
Masashi Ito, Keiji Ohara, Akinori Ito, Masafumi Yano

WED-SES3-P4: SPOKEN LANGUAGE UNDERSTANDING AND SPOKEN LANGUAGE TRANSLATION I

Functional Imaging of Brain Regions Sensitive to Communication Sounds in Primates 2498
Christopher I. Petkov, 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 Jabaian, Laurent Besacier, Fabrice Lefevre

Learning Naturally Spoken Commands for a Robot 2510
Anja Austermann, Seiji Yamada, Kotaro Funakoshi, Mikio Nakano

A Semi-Supervised Cluster-and-Label Approach for Utterance Classification 2514
Amparo Albalate, Aparna Suchindranath, David Suendermann, 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, Hui 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 Ettelaie, Panayiotis G. Georgiou, Shrikanth S. Narayanan

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

THU-SES1-S1: SPECIAL SESSION: SOCIAL SIGNALS IN SPEECH

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
Emina Kurtic, Guy J. Brown, Bill Wells

Disambiguating the Functions of Conversational Sounds with Prosody: The Case of 'Yeah' 2558
Khiet P. Truong, Dirk Heylen

Prosody and Voice Quality of Vocal Social Signals: The Case of Dominance in Scenario Meetings 2562
Marcela Charfuelan, Marc Schroder, Ingmar Steiner

THU-SES1-O2: PHYSIOLOGY AND PATHOLOGY OF SPOKEN LANGUAGE

Online SLU Model Adaptation with a Partial Oracle 2566
Pierre Gotab, Geraldine Damnati, Frederic Bechet, Lionel Delphin-Poulat

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

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

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

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

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

THU-SES1-O3: PITCH AND GLOTTAL-WAVEFORM ESTIMATION AND MODELING II

Pre- and Short-Term Posttreatment Vocal Functioning in Patients with Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy 2590
Irene Jacobi, 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
Wei Chu, Abeer Alwan

Robust and Efficient Pitch Estimation Using an Iterative ARMA Technique 2602
Jung Ook Hong, Patrick J. Wolfe

Statistical Modeling of F0 Dynamics in Singing Voices Based on Gaussian Processes with Multiple Oscillation Bases 2606
Yasunori Ohishi, Hirokazu Kameoka, Daichi Mochihashi, Hidehisa Nagano, Kunio Kashino
Applying Geometric Source Separation for Improved Pitch Extraction in Human-Robot Interaction 2610
Martin Heckmann, Claudius Glaser, Frank Joublin, Kazuhiro Nakadai

THU-SES1-O4: ASR: FEATURE EXTRACTION II

A Spectral LF Model Based Approach to Voice Source Parameterisation 2614
John Kane, Mark Kane, Christer Gobl
Glottal-Based Analysis of the Lombard Effect 2618
Thomas Drugman, Thierry Dutoit
Hidden Logistic Linear Regression for Support Vector Machine Based Phone Verification 2622
Bo Li, Khe Chai Sim
Jointly Optimized Discriminative Features for Speech Recognition 2626
Tim Ng, Bing Zhang, Long Nguyen
Invariant Integration Features Combined with Speaker-Adaptation Methods 2630
Florian Muller, Alfred Mertins
Multi Resolution Discriminative Models for Subvocalic Speech Recognition 2634
Mark Raugas, Vivek Kumar Rangarajan Sridhar, Rohit Prasad, Prem Natarajan

THU-SES1-P1: SPEAKER DIARIZATION

A Comparative Large Scale Study of MLP Features for Mandarin ASR 2638
Fabio Valente, Mathew Magimai Doss, Christian Plahl, Suman V. Ravuri, Wen Wang
Recognizing Cochlear Implant-Like Spectrally Reduced Speech with HMM-Based ASR: Experiments with MFCCs and PLP Coefficients 2642
Cong-Thanh Do, Dominique Pastor, Gael Le Lan, Andre Goalic
A Hybrid Approach to Online Speaker Diarization 2646
Carlos Vaquero, Oriol Vinyals, Gerald Friedland
System Output Combination for Improved Speaker Diarization 2650
Simon Bozonnet, Nicholas Evans, Xavier Anguera, Oriol Vinyals, Gerald Friedland, Corinne Fredouille
An Integrated Top-Down/Bottom-Up Approach to Speaker Diarization 2654
Simon Bozonnet, Nicholas Evans, Corinne Fredouille, Dong Wang, Raphael Troncy
Advances in Fast Multistream Diarization Based on the Information Bottleneck Framework 2658
Deepu Vijayasenan, Fabio Valente, Herve Bourlard
Audio-Visual Synchronisation for Speaker Diarisation 2662
Giulia Garau, Alfred Dielmann, Herve Bourlard
An Improved Cluster Model Selection Method for Agglomerative Hierarchical Speaker Clustering Using Incremental Gaussian Mixture Models 2666
Kyu J. Han, Shrikanth S. Narayanan
Dialog Prediction for a General Model of Turn-Taking 2670
Nigel G. Ward, Olac Fuentes, Alejandro Vega

THU-SES1-P2: MULTI-MODAL ASR, INCLUDING AUDIO-VISUAL ASR

Speaker Tracking in an Unsupervised Speech Controlled System 2674
Tobias Herbig, Franz Gerl, Wolfgang Minker
MultiBIC: An Improved Speaker Segmentation Technique for TV Shows 2678
Paula Lopez-Otero, Laura Docio-Fernandez, Carmen Garcia-Mateo
Automatic Speech Recognition for Assistive Writing in Speech Supplemented Word Prediction 2682
John-Paul Hosom, Tom Jakobs, Allen Baker, Susan Fager
Viseme-Dependent Weight Optimization for CHMM-Based Audio-Visual Speech Recognition 2686
Alexey Karpov, Andrey Ronzhin, Konstantin Markov, Milos Zelezny
Audio-Visual Anticipatory Coarticulation Modeling by Human and Machine 2690
Louis H. Terry, Karen Livescu, Janet B. Pierrehumbert, Aggelos K. Katsaggelos
Impact of Lack of Acoustic Feedback in EMG-Based Silent Speech Recognition 2694
Matthias Janke, Michael Wand, Tanja Schultz
Using Prosody to Improve Mandarin Automatic Speech Recognition 2698
Chong-Jia Ni, Wenju Liu, Bo Xu
A Robust Audio-Visual Speech Recognition Using Audio-Visual Voice Activity Detection 2702
Satoshi Tamura, Masato Ishikawa, Takashi Hashiba, Shin'Ichi Takeuchi, Satoru Hayamizu
Efficient Manycore CHMM Speech Recognition for Audiovisual and Multistream Data 2706
Dorothea Kolossa, Jike Chong, Steffen Zeiler, Kurt Keutzer

THU-SES1-P3: SPEAKER AND LANGUAGE RECOGNITION

Two-Layered Audio-Visual Integration in Voice Activity Detection and Automatic Speech Recognition for Robots 2710
Takami Yoshida, Kazuhiro Nakadai

Non-Audible Murmur Recognition Based on Fusion of Audio and Visual Streams 2714
Panikos Heracleous, Norihiro Hagita

Improved N-Gram Phonotactic Models for Language Recognition 2718
Mohamed Faouzi Benzeghiba, Jean-Luc Gauvain, Lori Lamel

A Study of Term Weighting in Phonotactic Approach to Spoken Language Recognition 2722
Sirinoot Boonsuk, Donglai Zhu, Bin Ma, Atiwong Suchato, Proadpran Punyabukkana, Nattanun Thatphithakkul, Chai Wutiwiwatchai

Exploiting Context-Dependency and Acoustic Resolution of Universal Speech Attribute Models in Spoken Language Recognition 2726
Sabato Marco Siniscalchi, Jeremy Reed, Torbjorn Svendsen, Chin-Hui Lee

Hierarchical Multilayer Perceptron Based Language Identification 2730
David Imseng, Mathew Magimai Doss, Herve Bourlard

The NIST 2010 Speaker Recognition Evaluation 2734
Alvin F. Martin, Craig S. Greenberg

Bayesian Speaker Recognition Using Gaussian Mixture Model and Laplace Approximation 2738
Shih-Sian Cheng, I-Fan Chen, Hsin-Min Wang

What Else is New Than the Hamming Window? Robust MFCCs for Speaker Recognition via Multitapering 2742
Tomi Kinnunen, Rahim Saeidi, Johan Sandberg, Maria Hansson-Sandsten

Fast Computation of Speaker Characterization Vector Using MLLR and Sufficient Statistics in Anchor Model Framework 2746
A. K. Sarkar, S. Umesh

Graph-Embedding for Speaker Recognition 2750
Zahi N. Karam, W. M. Campbell

A Hybrid Modeling Strategy for GMM-SVM Speaker Recognition with Adaptive Relevance Factor 2754
Chang Huai You, Haizhou Li, Kong Aik Lee

THU-SES1-P4: SOURCE LOCALIZATION AND SEPARATION

Robust Mixture Modeling Using T-Distribution: Application to Speaker ID 2758
Harshavardhan S., T. V. Sreenivas

A Variable Frame Length and Rate Algorithm Based on the Spectral Kurtosis Measure for Speaker Verification 2762
Chi-Sang Jung, Kyu J. Han, Hyunson Seo, Shrikanth S. Narayanan, Hong-Goo Kang

Near Field Sound Source Localization Based on Cross-Power Spectrum Phase Analysis with Multiple Microphones 2766
Kohei Hayashida, Masanori Morise, Takanobu Nishiura

A Maximum a Posteriori Sound Source Localization in Reverberant and Noisy Conditions 2770
Jinho Choi, Chang D. Yoo

A DOA Estimation Algorithm Based on Equalization-Cancellation Theory 2774
Duc Thanh Chau, Junfeng Li, Masato Akagi

Concurrent Speaker Localization Using Multi-Band Position-Pitch (M-PoPi) Algorithm with Spectro-Temporal Pre-Processing 2778
Tania Habib, Harald Romsdorfer

On Using Gaussian Mixture Model for Double-Talk Detection in Acoustic Echo Suppression 2782
Ji-Hyun Song, Kyu-Ho Lee, Yun-Sik Park, Sang-Ick Kang, Joon-Hyuk Chang

Catalog-Based Single-Channel Speech-Music Separation 2786
Cemil Demir, A. Taylan Cemgil, Murat Saraclar

Unvoiced Speech Segregation Based on CASA and Spectral Subtraction 2790
Ke Hu, Deliang Wang

THU-SES2-S1: SPECIAL SESSION: INTERSPEECH 2010 PARALINGUISTIC CHALLENGE

Unsupervised Sequential Organization for Cochannel Speech Separation 2794
Ke Hu, Deliang Wang

The INTERSPEECH 2010 Paralinguistic Challenge 2798
Bjorn Schuller, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Muller, Shrikanth S. Narayanan

Age and Gender Classification from Speech Using Decision Level Fusion and Ensemble Based Techniques 2802
Florian Lingenfelser, Johannes Wagner, Thurid Vogt, Jonghwa Kim, Elisabeth Andre

Level of Interest Sensing in Spoken Dialog Using Multi-Level Fusion of Acoustic and Lexical Evidence 2806
Je Hun Jeon, Rui Xia, Yang Liu

Fuzzy Support Vector Machines for Age and Gender Classification 2810
Phuoc Nguyen, Trung Le, Dat Tran, Xu Huang, Dharmendra Sharma

Gender and Affect Recognition Based on GMM and GMM-UBM Modeling with Relevance MAP Estimation 2814
Rok Gajsek, Janez Zibert, Tadej Justin, Vitomir Struc, Bostjan Vesnicer, France Mihelic

Age Recognition Based on Speech Signals Using Weights Supervector 2818
Royi Porat, Dan Lange, Yaniv Zigel

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 Burget, Jan Cernocky

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

Age and Gender Recognition Based on Multiple Systems --- Early vs. Late Fusion 2834
Tobias Bocklet, Georg Stemmer, Viktor Zeissler, Elmar Noth

THU-SES2-O1: NEW PARADIGMS IN ASR II

Automatic Speaker Age and Gender Recognition in the Car for Tailoring Dialog and Mobile Services 2838
Michael Feld, Felix Burkhardt, Christian Muller

Improved Topic Classification and Keyword Discovery Using an HMM-Based Speech Recognizer Trained Without Supervision 2842
Man-Hung Siu, Herbert Gish, Arthur Chan, William Belfield

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

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

Mandarin Tone Recognition Using Affine-Invariant Prosodic Features and Tone Posteriorgram 2854
Yow-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
Om D. Deshmukh, Harish Doddala, Ashish Verma, Karthik Visweswariah

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

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

Phrase Alignment Confidence for Statistical Machine Translation 2878
Sankaranarayanan 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
Kiyoaki Aikawa, Junko Uenuma, Tomoko Akitake

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

A Singing Style Modeling System for Singing Voice Synthesizers 2894
Keijiro Saino, Makoto Tachibana, Hideki Kenmochi

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, Sihyun Joo, Chang D. Yoo

Modified Spatial Audio Object Coding Scheme with Harmonic Extraction and Elimination Structure for Interactive Audio Service 2906
Jihoon Park, Kwangki Kim, Jeongil 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 Gubian, Lou Boves

Unsupervised Learning of Vowels from Continuous Speech Based on Self-Organized Phoneme Acquisition Model 2914
Kouki Miyazawa, Hideaki Kikuchi, Reiko Mazuka

Learning Speaker Normalization Using Semisupervised Manifold Alignment 2918
Andrew R. Plummer, Mary E. Beckman, Mikhail Belkin, Eric Fosler-Lussier, Benjamin Munson

Fully Unsupervised Word Learning from Continuous Speech Using Transitional Probabilities of Atomic Acoustic Events 2922
Okko Johannes Rasanen
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 Vesely, Lukas Burget, Frantisek Grezl
The Use of Sense in Unsupervised Training of Acoustic Models for ASR Systems 2938
Rita Singh, Benjamin Lambert, Bhiksha Raj
Boosted Mixture Learning of Gaussian Mixture HMMs for Speech Recognition 2942
Jun Du, Yu Hu, Hui Jiang
On the Exploitation of Hidden Markov Models and Linear Dynamic Models in a Hybrid Decoder Architecture for Continuous Speech Recognition 2946
Volker Leutnant, Reinhold Haeb-Umbach
Context Dependent Modelling Approaches for Hybrid Speech Recognizers 2950
Alberto Abad, Thomas Pellegrini, Isabel Trancoso, Joao Neto
A Regularized Discriminative Training Method of Acoustic Models Derived by Minimum Relative Entropy Discrimination 2954
Yotaro Kubo, Shinji Watanabe, Atsushi Nakamura, Tetsunori Kobayashi
Decision Tree State Clustering with Word and Syllable Features 2958
Hank Liao, Chris Alberti, Michiel 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 Wollmer, Yang Sun, Florian Eyben, Bjorn Schuller
One-Model Speech Recognition and Synthesis Based on Articulatory Movement HMMs 2970
Tsuneo Nitta, Takayuki Onoda, Masashi Kimura, Yurie Iribe, Kouichi Katsurada
Acoustic Modeling with Bootstrap and Restructuring for Low-Resourced Languages 2974
Xiaodong Cui, Jian Xue, Pierre L. Dognin, Upendra V. Chaudhari, Bowen Zhou
Lecture Speech Recognition by Combining Word Graphs of Various Acoustic Models 2978
Tetsuo Kosaka, Keisuke Goto, Takashi Ito, Masaharu Kato
Semi-Parametric Trajectory Modelling Using Temporally Varying Feature Mapping for Speech Recognition 2982
Khe Chai Sim, Shilin 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 Malkin, 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-Cozar, David Griol
A Stochastic Finite-State Transducer Approach to Spoken Dialog Management 3002
Lluis-F. Hurtado, Joaquin Planells, Encarna Segarra, Emilio Sanchis, David Griol
Enhanced Monitoring Tools and Online Dialogue Optimisation Merged into a New Spoken Dialogue System Design Experience 3006
Romain Laroche, Philippe Bretier, Ghislain Putois
Optimising a Handcrafted Dialogue System Design 3010
Romain Laroche, Ghislain Putois, Philippe Bretier
Utterance Selection for Speech Acts in a Cognitive Tourguide Scenario 3014
Felix Putze, Tanja Schultz
Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System 3018
Gabriel Parent, Maxine Eskenazi
Combining User Intention and Error Modeling for Statistical Dialog Simulators 3022
Silvia Quarteroni, Meritxell Gonzalez, Giuseppe Riccardi, Sebastian Varges
Parallel Processing of Interruptions and Feedback in Companions Affective Dialogue System 3026
Jaakko Hakulinen, Markku Turunen, Raul Santos De La Camara, Nigel Crook
Dynamic Language Modeling Using Bayesian Networks for Spoken Dialog Systems 3030
Antoine Raux, Neville Mehta, Deepak Ramachandran, Rakesh Gupta
Automatic Detection of Task-Incompleted Dialog for Spoken Dialog System Based on Dialog Act N-Gram 3034
Sunao Hara, Norihide Kitaoka, Kazuya Takeda

Dialogue Act Detection in Error-Prone Spoken Dialogue Systems Using Partial Sentence Tree and Latent Dialogue 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, Kouhei Sumi, Zhi-Qiang Chang, Katsuya Takanashi
Psychological Evaluation of a Group Communication Activation Robot in a Party Game ... 3046
Yoichi Matsuyama, Shinya Fujie, Hikaru Taniyama, Tetsunori Kobayashi
Analyzing User Utterances in Barge-in-Able Spoken Dialogue System for Improving Identification Accuracy ... 3050
Kyoko Matsuyama, Kazunori Komatani, Ryu Takeda, Toru Takahashi, Tetsuya Ogata, Hiroshi G. Okuno
Pitch Similarity in the Vicinity of Backchannels ... 3054
Mattias Heldner, Jens Edlund, Julia Hirschberg
A Rule-Based Backchannel Prediction Model Using Pitch and Pause Information ... 3058
Khiet P. Truong, Ronald Poppe, Dirk Heylen

THU-SES2-P3: DISCOURSE AND DIALOGUE

Combining Text Categorization and Dialog Modeling for Speaker Role Identification on Call Center Conversations ... 3062
Remi Lavalley, 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 Benedi
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 Isotani, Hisashi 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 Watanabe, Tomohiro Nakatani
Spectral Entropy-Based Voice Activity Detector for Videoconferencing Systems ... 3106
Bowon Lee, Debargha Muhkerjee
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 Oonishi, Koji Iwano, Sadaoki 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
Yohei Kawaguchi, Masahito Togami, Yasunari Obuchi
Author Index