# TABLE OF CONTENTS

## VOLUME 1

### ISCA MEDAL 2017 CEREMONY

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISCA Medal for Scientific Achievement</td>
<td>1</td>
</tr>
</tbody>
</table>

### MON-SS-1-8: SPECIAL SESSION: INTERSPEECH 2017 AUTOMATIC SPEAKER VERIFICATION SPOOFING AND COUNTERMEASURES CHALLENGE 1

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection</td>
<td>2</td>
</tr>
<tr>
<td>Experimental Analysis of Features for Replay Attack Detection --- Results on the ASVspoof 2017 Challenge</td>
<td>7</td>
</tr>
<tr>
<td>Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection</td>
<td>12</td>
</tr>
<tr>
<td>Countermeasures for Automatic Speaker Verification Replay Spoofing Attack : On Data Augmentation, Feature Representation, Classification and Fusion</td>
<td>17</td>
</tr>
<tr>
<td>Spoof Detection Using Source, Instantaneous Frequency and Cepstral Features</td>
<td>22</td>
</tr>
<tr>
<td>Audio Replay Attack Detection Using High-Frequency Features</td>
<td>27</td>
</tr>
<tr>
<td>Feature Selection Based on CQCCs for Automatic Speaker Verification Spoofing</td>
<td>32</td>
</tr>
</tbody>
</table>

### MON-SS-1-11: SPECIAL SESSION: SPEECH TECHNOLOGY FOR CODE-SWITCHING IN MULTILINGUAL COMMUNITIES

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal Speaker Clustering and Verification Corpus with Code-Switching Frisian-Dutch Speech</td>
<td>37</td>
</tr>
<tr>
<td>Exploiting Untranscribed Broadcast Data for Improved Code-Switching Detection</td>
<td>42</td>
</tr>
<tr>
<td>Jee haan, I'd like both, por favor: Elicitation of a Code-Switched Corpus of Hindi—English and Spanish—English Human—Machine Dialog</td>
<td>47</td>
</tr>
<tr>
<td>On Building Mixed Lingual Speech Synthesis Systems</td>
<td>52</td>
</tr>
<tr>
<td>Speech Synthesis for Mixed-Language Navigation Instructions</td>
<td>57</td>
</tr>
<tr>
<td>Addressing Code-Switching in French/Algerian Arabic Speech</td>
<td>62</td>
</tr>
<tr>
<td>Metrics for Modeling Code-Switching Across Corpora</td>
<td>67</td>
</tr>
<tr>
<td>Synthesising isiZulu-English Code-Switch Bigrams Using Word Embeddings</td>
<td>72</td>
</tr>
<tr>
<td>Crowdsourcing Universal Part-of-Speech Tags for Code-Switching</td>
<td>77</td>
</tr>
</tbody>
</table>

### MON-SS-2-8: SPECIAL SESSION: INTERSPEECH 2017 AUTOMATIC SPEAKER VERIFICATION SPOOFING AND COUNTERMEASURES CHALLENGE 2

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Replay Attack Detection with Deep Learning Frameworks</td>
<td>82</td>
</tr>
<tr>
<td>Ensemble Learning for Countermeasures of Audio Replay Spoofing Attack in ASVspoof2017</td>
<td>87</td>
</tr>
<tr>
<td>A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification</td>
<td>92</td>
</tr>
<tr>
<td>Replay Attack Detection Using DNN for Channel Discrimination</td>
<td>97</td>
</tr>
<tr>
<td>ResNet and Model Fusion for Automatic Spoofing Detection</td>
<td>102</td>
</tr>
</tbody>
</table>
SFF Anti-Spoofer: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures
Challenge 2017 ........................................................................................................................................................................107
K. N. R. R. Raju Alluri, Stivanand Achanta, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, Anil Kumar Vuppala

MON-O-1-1: CONVERSATIONAL TELEPHONE SPEECH RECOGNITION

Improved Single System Conversational Telephone Speech Recognition with VGG Bottleneck Features ..........................................................112
William Hartmann, Roger Hsiao, Tom Ng, Jeff Ma, FrancisKeith, Man-Hung Siu

Student-Teacher Training with Diverse Decision Tree Ensembles ........................................................................................................117
Jeremy H. M. Wong, Mark J. F. Gales

Embedding-Based Speaker Adaptive Training of Deep Neural Networks ................................................................................................122
Xiaodong Cui, Yathibha Goel, George Saon

Improving Deliverable Speech-to-Text Systems with Multilingual Knowledge Transfer ..............................................................................127
Jeff Ma, Francis Keith, Tim Ng, Man-Hung Siu, Owen Kimball

English Conversational Telephone Speech Recognition by Humans and Machines .............................................................................132
George Saon, Gakuto Karata, Tom Servu, Karitk Audhkhasi, Samuel Thomas, Dimitrios Dimitriadis, Xiaodong Cui, Bhuvana Ramabadran, Michael Picheny, Lynn-Li Lim, Bergul Roomi, Phil Hall

Comparing Human and Machine Errors in Conversational Speech Transcription ........................................................................................137
Andreas Stolcke, Jasha Droppo

MON-O-1-2: MULTIMODAL PARALINGUISTICS

Multimodal Makers of Persuasive Speech: Designing a Virtual Debate Coach ..................................................................................................142
Vedha Petukhava, Manoj Raju, Harry Bunt

Acoustic-Prosodic and Physiological Response to Stressful Interactions in Children with Autism Spectrum Disorder ........................................147
Daniel Bone, Julia Mertens, Emily Zane, Sungbok Lee, Shrikanth S. Narayanan, Ruth Grossman

A Stepwise Analysis of Aggregated Crowdsourced Labels Describing Multimodal Emotional Behaviors ......................................................152
Alec Barmania, Carlos Busso

An Information Theoretic Analysis of the Temporal Synchrony Between Head Gestures and Prosodic Patterns in Spontaneous Speech ........................................................................................................157
Gaurav Fotekar, Prasanta Kumar Ghosh

Multimodal Prediction of Affective Dimensions via Fusing Multiple Regression Techniques ........................................................................162
D.-Y. Huang, Wang Ding, Mingyu Xu, Huaping Ming, Minghui Dong, Xinguo Yu, Haizhou Li

Co-Production of Speech and Pointing Gestures in Clear and Perturbed Interactive Tasks: Multimodal Designation Strategies ..................................................166
Marion Dohem, Benjamin Roustan

MON-O-1-4: DEREVERBERATION, ECHO CANCELLATION AND SPEECH

Improving Speaker Verification for Reverberant Conditions with Deep Neural Network Dereverberation Processing ........................................................................................................171
Peter Gazewich, Stephen A. Zahorian

Stepszie Control for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods and the Estimated System Distance .................................................................176
Philipp Bulling, Klaus Linhard, Arthur Wolf, Gerhard Schmidt

A Delay-Flexible Stereo Acoustic Echo Cancellation for DFT-Based In-Car Communication (ICC) Systems ........................................................................181
Jan Francen, Tim Fingscheidel

Speech Enhancement Based on Harmonic Estimation Combined with MMSE to Improve Speech Intelligibility for Cochlear Implant Recipients ..................................................................................186
Dongmei Wang, John H. L. Hansen

Improving Speech Intelligibility in Binaural Hearing Aids by Estimating a Time-Frequency Mask with a Weighted Least Squares Classifier .................................................................................191
David Aylom, Roberto Gil-Pita, Manuel Roso-Zurera

Simulations of High-Frequency Vocoder on Mandarin Speech Recognition for Acoustic Hearing Preserved ................................................196
Trung-Chen Wu, Tai-Shih Chi, Chia-Fone Lee

MON-O-1-6: ACOUSTIC AND ARTICULATORY PHONETICS

Phonetic Correlates of Pharyngeal and Pharyngealized Consonants in Saudi, Lebanese, and Jordanian Arabic: An rt-MRI Study .........................................................................................................................201
Zainab Hermes, Marissa Barlaz, Ryan Shoosted, Zhi-Pei Liang, Brad Sutton

Glottal Opening and Strategies of Production of Fricatives ..................................................................................................................206
Benjamin Elie, Yves Laperie

Acoustics and Articulation of Medial versus Final Coronal Stop Geminination Contrasts in Moroccan Arabic ...............................................210
Mohamed Youness Frej, Christopher Cariguan, Catherine T. Best

How are Four-Level Length Distinctions Produced? Evidence from Moroccan Arabic ................................................................................215
Giuseppina Turco, Karim Shoul, Rachid Ridouane
Vowels in the Barunga Variety of North Australian Kriol ......................................................... 219
Nature of Contrast and Coarticulation: Evidence from Mizo Tones and Assamese Vowel Harmony ................................................................. 224

MON-O-1-10: MULTIMODEL AND ARTICULATORY SYNTHESIS

The Influence of Synthetic Voice on the Evaluation of a Virtual Character ......................................................... 229
Articulatory Text-to-Speech Synthesis Using the Digital Waveguide Mesh Driven by a Deep Neural Network ................................................................. 234
An HMM/DNN Comparison for Synchronized Text-to-Speech and Tongue Motion Synthesis ................................................................. 239
VCV Synthesis Using Task Dynamics to Animate a Factor-Based Articulatory Model ................................................................. 244
Beyond the Listening Test: An Interactive Approach to TTS Evaluation ................................................................. 249
Integrating Articulatory Information in Deep Learning-Based Text-to-Speech Synthesis ................................................................. 254

MON-O-2-1: NEURAL NETWORKS FOR LANGUAGE MODELING

Approaches for Neural-Network Language Model Adaptation ................................................................. 259
A Batch Noise Contrastive Estimation Approach for Training Large Vocabulary Language Models ................................................................. 264
Investigating Bidirectional Recurrent Neural Network Language Models for Speech Recognition ................................................................. 269
Fast Neural Network Language Model Lookups at N-Gram Speeds ................................................................. 274
Empirical Exploration of Novel Architectures and Objectives for Language Models ................................................................. 279
Residual Memory Networks in Language Modeling: Improving the Reputation of Feed-Forward Networks ................................................................. 284

MON-O-2-2: PATHOLOGICAL SPEECH AND LANGUAGE

Dominant Distortion Classification for Pre-Processing of Vowels in Remote Biomedical Voice Analysis ................................................................. 289
Automatic Paraphasia Detection from Aphasic Speech: A Preliminary Study ................................................................. 294
Evaluation of the Neurological State of People with Parkinson's Disease Using i-Vectors ................................................................. 299
Objective Severity Assessment from Disordered Voice Using Estimated Glottal Airflow ................................................................. 304
Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach ................................................................. 309
Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease ................................................................. 314

MON-O-2-4: SPEECH ANALYSIS AND REPRESENTATION 1

Phone Classification Using a Non-Linear Manifold with Broad Phone Class Dependent DNNs ................................................................. 319
An Investigation of Crowd Speech for Room Occupancy Estimation ................................................................. 324
Time-Frequency Coherence for Periodic-Aperiodic Decomposition of Speech Signals ................................................................. 329
Musical Speech: A New Methodology for Transcribing Speech Prosody ................................................................. 334
Estimation of Place of Articulation of Fricatives from Spectral Characteristics for Speech Training ................................................................. 339
Estimation of the Probability Distribution of Spectral Fine Structure in the Speech Source ................................................................. 344
MON-O-2-6: PERCEPTION OF DIALECTS AND L1

End-to-End Acoustic Feedback in Language Learning for Correcting Devoiced French Final-Fricatives ......................................................... 349
Sicheon Ghosh, Camille Fauch, Yves Laprise, Aphilas Sini

Dialect Perception by Older Children .................................................................................................................................................. 354
Eva Jacewicz, Robert A. Fox

Perception of Non-Contrastive Variations in American English by Japanese Learners: Flaps are Less Favored Than Stops ........................................... 359
Kiyoko Yoneyama, Mafuyu Kitahara, Keiichi Tajima

L1 Perceptions of L2 Prosody: The Interplay Between Intonation, Rhythm, and Speech Rate and Their Contribution to Accentedness and Comprehensibility .......................................................... 364
Changwoo Kim, Anun Narayanan, Tara N. Sainath, Michiel Bacchiani

Effects of Pitch Fall and L1 on Vowel Length Identification in L2 Japanese .............................................................................................................................. 369
Izumi Takiguchi

A Preliminary Study of Prosodic Disambiguation by Chinese EFL Learners ................................................................................................. 374
Yuanyuan Zhang, Hongwei Ding

MON-O-2-10: FAR-FIELD SPEECH RECOGNITION

Generation of Large-Scale Simulated Utterances in Virtual Rooms to Train Deep-Neural Networks for Far-Field Speech Recognition in Google Home ........................................................................................................ 379
Banriskhem K. Khonglah, K. T. Deepak, S. R. Mahadeva Prasanna

Neural Network-Based Spectrum Estimation for Online WPE Dereverberation ......................................................................................... 384
Keisuke Kinoshita, Marc Delcroix, Hae Yong Kwon, Takuma Mori, Tomohiro Nakatani

Factorial Modeling for Effective Suppression of Directional Noise ......................................................................................................................... 389
Osamu Ichikawa, Takashi Fukuuda, Gakuto Kurata, Steven J. Rennie

On Design of Robust Deep Models for CHiME-4 Multi-Channel Speech Recognition with Multiple Configurations of Array Microphones ........................................................................................................... 394
Yen-Huei Ts, Jun Du, Lei Sun, Feng Ma, Chin-Hui Lee

Acoustic Modeling for Google Home ...................................................................................................................................................... 399
Bo Li, Tara N. Sainath, Anun Narayanan, Joe Caroselli, Michiel Bacchiani, Ananya Misra, Izak Shafrun, Harman Sak, Golan Pandak, Kean Chin, Khe Chai Sim, Ron J. Weiss, Kevin W. Wilson, Ehsan Varniani, Chanwoo Kim, Olivier Siohan, Michiel Weintraub, Erik McDermott, Richard Rose, Matt Shumon

On Multi-Domain Training and Adaptation of End-to-End RNN Acoustic Models for Distinctive Vowel Recognition ......................................................................... 404
Seyedmahmoud Mirraezi, John R. L. Hansen

MON-P-1-1: SPEECH ANALYSIS AND REPRESENTATION 2

Low-Dimensional Representation of Spectral Envelope Without Deterioration for Full-Band Speech Analysis/Synthesis System .................................................................................................................................................. 409
Masanori Morise, Genta Miyashita, Kenji Ozawa

Robust Source-Filter Separation of Speech Signal in the Phase Domain ................................................................................................. 414
Erfan Loseimini, Jon Barker, Oscar Saez Torralba, Thomas Hain

A Time-Warping Pitch Tracking Algorithm Considering Fast F0 Changes..................................................................................................................... 419
Simon Stone, Peter Steiner, Peter Birkholz

A Modulation Property of Time-Frequency Derivatives of Filtered Phase and its Application to Aperiodicity and f0 Estimation .................................................................................................................................................. 424
Hideki Kawahara, Ken-Ichi Sakahibara, Masanori Morise, Hideki Banno, Tomoki Toda

Non-Local Estimation of Speech Signal for Vowel Onset Point Detection in Varied Environments ................................................................. 429
Avinash Kumar, S. Shahnawazuddin, Gayadhur Pradhan

Time-Domain Envelope Modulating the Noise Component of Excitation in a Continuous Residual-Based Vocoder for Statistical Parametric Speech Synthesis ........................................................................................................... 434
Mohammed Salah Al-Rudhi, Tomas Gabor Csapo, Geza Nemeth

Wavelet Speech Enhancement Based on Robust Principal Component Analysis .......................................................................................................... 439
Chia-Lung Wu, Hsiang-Ping Hsu, SY-Siang Wang, Jih-Hweih Hung, Ying-Hui Lai, Hsin-Min Wang, Yu Tsao

Vowel Onset Point Detection Using Sonority Information .......................................................................................................................... 444
Bishuo Sharma, S. R. Mahadeva Prasanna

Analytic Filter Bank for Speech Analysis, Feature Extraction and Perceptual Studies ................................................................................................. 449
Unto K. Laine

Learning the Mapping Function from Voltage Amplitudes to Sensor Positions in 3D-EMA Using Deep Neural Networks .................................................................................................................................................. 454
Christian Kroos, Mark D. Plumley

MON-P-1-2: SPEECH AND AUDIO SEGMENTATION AND CLASSIFICATION 2

Multilingual i-Vector Based Statistical Modeling for Music Genre Classification ................................................................................................. 459
Jia Dai, Wei Xue, Wenju Liu

Indoor/Outdoor Audio Classification Using Foreground Speech Segmentation ................................................................................................. 464
Banrikkhem K. Khonglah, K. T. Deepak, S. R. Mahadeva Prasanna
Attention Based CLDNNs for Short-Duration Acoustic Scene Classification ......................................................... 469
Jinzi Guo, Ning Xu, Li-Jia Li, Abhijit Abhay

Frame-Wise Dynamic Threshold Based Polyphonic Acoustic Event Detection ............................................... 474
Xianjun Xia, Roberto Togneri, Ferdous Sohel, David Huang

Enhanced Feature Extraction for Speech Detection in Media Audio ................................................................. 479
Inseon Jang, Chunjunnu Anh, Jeongseu Youn, Yuenseun Jang

Audio Classification Using Class-Specific Learned Descriptors ...................................................................... 484
Sukanya Sonowal, Tushar Sandhan, Indu Choi, Nam Soo Kim

Hidden Markov Model Variational Autoencoder for Acoustic Unit Discovery ............................................... 488
Janek Ellen, Jaden Heymann, Lukas Drude, Thomas Glarner, Reinhold Haehl-Umbach, Bhiksha Raj

Virtual Adversarial Training and Data Augmentation for Acoustic Event Detection with Gated Recurrent Neural Networks .................................................................................................................. 493
Matthias Zohrer, Franz Perknoof

Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi ...................................................... 498
Michael McGettige, Michaela Socolof, Sarah Milne, Michael Wagner, Morgan Soderberg

A Robust Voiced/Unvoiced Phoneme Classification from Whispered Speech Using the ’Color’ of Whispered Phonemes and Deep Neural Networks ............................................................................. 503
G. Nisha Meenakshi, Prasanta Kumar Ghosh

MON-P.1-4: SEARCH, COMPUTATIONAL STRATEGIES AND LANGUAGE MODELING

Rescoring-Aware Beam Search for Reduced Search Errors in Contextual Automatic Speech Recognition ................................................. 508
Ian Williams, Petar Alekovic

Comparison of Decoding Strategies for CTC Acoustic Models ......................................................................... 513
Thomas Zenkel, Ramon Sanabal, Florian Metse, Jan Niewi, Matthias Sperber, Sebastian Stuker, Alex Waibel

Phone Duration Modeling for LVCSR Using Neural Networks ........................................................................... 518
Hossein Hadian, Daniel Povey, Hossein Sameti, Sanjeev Khudanbaur

Towards Better Decoding and Language Model Integration in Sequence to Sequence Models .......................... 523
Jan Chorowski, Navdeep Jaitly

Empirical Evaluation of Parallel Training Algorithms on Acoustic Modeling .................................................. 528
Wenpeng Li, Binbin Zhang, Lei Xie, Dong Yu

Binary Deep Neural Networks for Speech Recognition ...................................................................................... 533
Xu Xiang, Yanmin Qian, Kai Yu

Hierarchical Constrained Bayesian Optimization for Feature, Acoustic Model and Decoder Parameter Optimization ........................................................................................................................................................................ 538
Akshay Chandrashekar, Ian Lane

Use of Global and Acoustic Features Associated with Contextual Factors to Adapt Language Models for Spontaneous Speech Recognition ................................................................................... 543
Shohei Toyama, Daisuke Saito, Nobuaki Minematsu

Joint Learning of Correlated Sequence Labeling Tasks Using Bidirectional Recurrent Neural Networks ............. 548
Yardanu Fasuja, Anirban Lah, Shachar Mirkin, Visk Raykar, Lili Koiferman, Guy Lev

Estimation of Gap Between Current Language Models and Human Performance .............................................. 553
Xiaoyu Shen, Yousef Oual, Clayton Greenberg, Mittal Singh, Dietrich Klakow

A Phonological Phrase Sequence Modelling Approach for Resource Efficient and Robust Real-Time Punctuation Recovery ........................................................................................................................... 558
Anna Moro, Gyorgy Szaszak

MON-P.2-1: SPEECH PERCEPTION

Factors Affecting the Intelligibility of Low-Pass Filtered Speech ........................................................................ 563
Lei Wang, Fei Chen

Phonetic Restoration of Temporally Reversed Speech ......................................................................................... 567
Shi Yu Wang, Fei Chen

Simultaneous Articulatory and Acoustic Distortion in L1 and L2 Listening: Locally Time-Reversed “Fast’’ Speech .............................................................................................................................................. 571
Mako Ishida

Lexically Guided Perceptual Learning in Mandarin Chinese ................................................................................. 576
L. Ann Burchfield, San-Hei Knei Luk, Mark Antoniou, Anne Cutler

The Effect of Spectral Profile on the Intelligibility of Emotional Speech in Noise .................................................. 581
JiHong Su, Ning Xu, Jie Song Chong, Jae Sun Kim

Whether Long-term Tracking of Speech Rate Affects Perception Depends on Who is Talking ........................ 586
Merel Maslovski, Antje S. Meyer, Hand Rutger Bosker

Emotional Thin-Slicing: A Proposal for a Short- and Long-Term Division of Emotional Speech .......................... 591
Daniel Oliveira Peres, Dominic Watt, Waldemar Ferreira Netto

Predicting Epenthetic Vowel Quality from Acoustics ......................................................................................... 596
Adriana Quevedo-Ruzco, Erik Parla-Olivera, Shi Yu, Yuko Hirose, Sharon Papercamp, Emmanuel Dupoux

The Effect of Spectral Tilt on Size Discrimination of Voiced Speech Sounds ..................................................... 601
Toshio Matusi, Toshiro Iri, Kodai Yamamoto, Hideki Kawahara, Roy D. Patterson

Misperceptions of the Emotional Content of Natural and Vocoder speech in a Car ........................................... 606
Jaime Lorenzo-Trueba, Cassia Valentini Botinhao, Gustav Eje Henter, Junichi Yamagishi
The Relative Cuing Power of F0 and Duration in German Prominence Perception .................................................. 611
Oliver Niebuhr, Jana Winkler
Perception and Acoustics of Vowel Nasality in Brazilian Portuguese ................................................................. 616
Luciana Marques, Rebecca Scarborough
Sociophonetic Realizations Guide Subsequent Lexical Access .............................................................................. 621
Jonny Kim, Katie Drager

MON-P-2-2: SPEECH PRODUCTION AND PERCEPTION

Critical Articulators Identification from RT-MRI of the Vocal Tract ........................................................................ 626
Samuel Silva, Antonio Teixeira
Semantic Edge Detection for Tracking Vocal Tract Air-Tissue Boundaries in Real-Time Magnetic Resonance
Images .............................................................................................................................................. 631
Krishna Somanandapalli, Asterios Toutios, Shrikant S. Narayanan
Vocal Tract Airway Tissue Boundary Tracking for rMRI Using Shape and Appearance Priors ................................. 636
Sasan Asadiadabi, Engin Erzin
An Objective Critical Distance Measure Based on the Relative Level of Spectral Valley ........................................ 641
T. V. Ananthapadmanabha, A. G. Ramakrishnan, Shubham Sharma
Database of Volumetric and Real-time Vocal Tract MRI for Speech Science ......................................................... 645
Tanner Sorensen, Zoris Shordilles, Asterios Toutios, Yoon-Chul Kim, Yinghua Zhu, Sangyeon Kim, Adam Lammert, Vikram
Ramamaruyan, Louis Goldstein, Dani Byrd, Krishna Nayak, Shrikant Narayanan

VOLUME 2

The Influence on Realization and Perception of Lexical Tones from Affricate's Aspiration ........................................ 650
Chong Cao, Yanlin Xie, Qi Zhang, Jinsong Zhang
Audiovisual Recalibration of Vowel Categories .................................................................................................... 655
Matthias K. Franken, Frank Esner, Jan-Mathis Schoffelen, Daniel J. Achexon, Peter Hagoort, James M. McQueen
The Effect of Gesture on Persuasive Speech ........................................................................................................... 659
Judith Peters, Marielle Hoeftjes
Auditory-Visual Integration of Talker Gender in Cantonese Tone Perception .......................................................... 664
Wei Lai
Event-Related Potentials Associated with Somatosensory Effect in Audio-Visual Speech Perception ........................ 669
Takayuki Ito, Hiroki Ohashi, Eva Montas, Vincent L. Gracco
When a Dog is a Cat and How it Changes Your Pupil Size: Pupil Dilation in Response to Information Mismatch ...... 674
Lena F. Renner, Marcin Wodarczak
Cross-Modal Analysis Between Phonation Differences and Texture Images Based on Sentiment Correlations ....... 679
Win Thuzar Kyaw, Yoshinori Sagisaka
Wireless Neck-Surface Accelerometer and Microphone on Flex Circuit with Application to Noise-Robust Monitoring
of Lombard Speech ........................................................................................................................................ 684
Daryush D. Mehta, Patrick C. Chwalek, Thomas F. Quatieri, Laura J. Brattain
Video-Based Tracking of Jaw Movements During Speech: Preliminary Results and Future Directions .................. 689
Andrea Bandini, Arvind Namasiyavam, Tanya Yunusova
Accurate Synchronization of Speech and EGG Signal Using Phase Information ..................................................... 694
S. B. Sanil Kumar, K. Sreenivasa Rao, Tanmay Mandal
The Acquisition of Focal Lengthening in Swedish .................................................................................................. 699
Anna Sara H. Romoren, Aoju Chen

MON-P-2-3: MULTI-LINGUAL MODELS AND ADAPTATION FOR ASR

Multilingual Recurrent Neural Networks with Residual Learning for Low-Resource Speech Recognition .................. 704
Shiyu Zhou, Yuan Yuan Zhao, Shuang Xu, Bo Xu
CTC Training of Multi-Phone Acoustic Models for Speech Recognition ............................................................... 709
Olivier Siohan
An Investigation of Deep Speech Networks for Multilingual Speech Recognition Training and Adaptation ........... 714
Sibo Tong, Philip N. Garner, Herve Bourlard
2016 BUT Babel System: Multilingual BLSTM Acoustic Model with i-Vector Based Adaptation .......................... 719
Martin Karafiast, Murali Karthick Baskar, Pavel Matjeja, Karel Veleky, Frantisek Grezl, Lukas Burget, Jan Cernocky
Optimizing DNN Adaptation for Recognition of Enhanced Speech ................................................................. 724
Marco Matassoni, Alessio Bratti, Daniele Falavigna
Deep Least Squares Regression for Speaker Adaptation .............................................................................................. 729
Younggwan Kim, Hyunjung Lim, Jaehoon Goo, Heesoo Kim
Multi-Task Learning Using Mismatched Transcription for Under-Resourced Speech Recognition ......................... 734
Yun Hai Do, Nancy F. Chen, Boon Pung Lim, Mark Hasegawa-Johnson
Generalized Distillation Framework for Speaker Normalization ............................................................................... 739
Neetha Mariam Joy, Sandeep Reddy Kothinti, S. Umesh, Basil Abraham
Learning Factorized Transform for Unsupervised Adaptation of LSTM-RNN Acoustic Models ............................ 744
Lahiru Samararakoon, Brian Mak, Khe Chai Sim
Factorised Representations for Neural Network Adaptation to Diverse Acoustic Environments

Joachim Fainberg, Steve Renals, Peter Bell

MON-P-2-4: PROSODY AND TEXT PROCESSING

An RNN Model of Text Normalization

Richard Sproat, Navdeep Jaitly

Weakly-Supervised Phrase Assignment from Text in a Speech-Synthesis System Using Noisy Labels

Asaf Rendel, Raul Fernandez, Zvi Kons, Andrew Rosenberger, Ron Hoory, Bhavana Ramabhadran

Prosody Aware Word-Level Encoder Based on BLSTM-RNNs for DNN-Based Speech Synthesis

Yasuake Iijima, Nobukatsu Hojo, Ryo Masumura, Taichi Asami

Global Syllable Vectors for Building TTS Front-End with Deep Learning

Jiefei Ni, Yoshiinori Shiga, Hisashi Kawat

Prosody Control of Utterance Sequence for Information Delivering

Ishin Fukuoka, Kazuhiro Iwata, Tetsunori Kobayashi

Multi-Task Learning for Prosodic Structure Generation Using BLSTM RNN with Structured Output Layer

Yuchen Huang, Zhiyong Wu, Runnan Li, Helen Meng, Lianhong Cai

Investigating Efficient Feature Representation Methods and Training Objective for BLSTM-Based Phone

Duration Prediction

Bo Chen, Tianling Bian, Kai Yu

Discrete Duration Model for Speech Synthesis

Discrete Duration Model for Speech Synthesis

Bo Chen, Tianling Bian, Kai Yu

Comparison of Modeling Target in LSTM-RNN Duration Model

Bo Chen, Jiahao Lai, Kai Yu

Learning Word Vector Representations Based on Acoustic Counts

M. Sam Rebeiro, Oliver Watts, Junichi Yamagishi

Synthesising Uncertainty: The Interplay of Vocal Effort and Hesitation Disfluencies

Eva Szekely, Joseph Mendelson, Joakim Gustafson

MON-S&T-1/2-A: SHOW & TELL 1

Prosographe: A Tool for Prosody Visualisation of Large Speech Corpora

Alp Oktem, Mireia Farrar, Lee Wanner

ChunkitApp: Investigating the Relevant Units of Online Speech Processing

Svetlana Vetчинникова, Anna Mauranen, Nina Mikusova

Extending the EMU Speech Database Management System: Cloud Hosting, Team Collaboration, Automatic

Revision Control

Markus Jochim

HomeBank: A Repository for Long-Form Real-World Audio Recordings of Children

Anne S. Warlaumont, Mark Vandam, Erika Bergelson, Alejandrina Cristia

A System for Real Time Collaborative Transcription Correction

Peter Bell, Joachim Fainberg, Catherine Lai, Mark Sinclair

MoPAREST — Mobile Phone Assisted Remote Speech Therapy Platform

Chitralekha Bhat, Anjali Kant, Bhavik Vachhani, Sarita Raotara, Ashok Kumar Sinha, Sunil Kumar Kopparapu

MON-S&T-1/2-B: SHOW & TELL 2

An Apparatus to Investigate Western Opera Singing Skill Learning Using Performance and Result Biofeedback, and Measuring its Neural Correlates

PercyConfigurator — Perception Experiments as a Service

Christoph Draxler

System for Speech Transcription and Post-Editing in Microsoft Word

Askars Salimbajev, Indra Bannicce

Emoji!? Collecting Emotion Data from Speech and Facial Expression Using Mobile Game App

Myll — The Mill: A New Platform for Processing Speech and Text Corpora Easily and Efficiently

Visual Learning 2: Pronunciation App Using Ultrasound, Video, and MRI

Kyori Suzuki, Ian Wilson, Hayato Watanabe

KEYNOTE 1: JAMES ALLEN

Dialogue as Collaborative Problem Solving

James Allen
TUE-SS-3-11: SPECIAL SESSION: SPEECH AND HUMAN-ROBOT INTERACTION

Elicitation Design for Acoustic Depression Classification: An Investigation of Articulation Effort, Linguistic Complexity, and Word Affect ................................................................. 834
Brian Staszak, Juliet Apgo, Roland Goecke

Robustness Over Time-Varying Channels in DNN-HMM ASR Based Human-Robot Interaction ................................................................. 839
Jose Novoa, Jorge Wuth, Juan Pablo Escudero, Josue Fredeus, Rodrigo Mahu, Richard M. Stern, Nestor Becerra Yoma

Analysis of Engagement and User Experience with a Laughter Responsive Social Robot ............................................................................. 844
Bekir Berker Turker, Zana Bucinca, Eneng Erzin, Yucef Temez, Meitin Segzin

Automatic Classification of Autistic Child Vocalisations: A Novel Database and Results ................................................................. 849
Alice Baird, Shahin Amiriparvar, Nicholas Cummins, Alyssa M. Alcorn, Anton Batliner, Sergey Pugachevskiy, Michael Freitag, Maurice Gerczuk, Bjorn Schuller

Crowd-Sourced Design of Artificial Attentive Listeners ............................................................................................................. 854
Catharine Oertel, Patrik Jonell, Dimosthenis Kontogiorgos, Joseph Mendelson, Jonas Beskow, Joakim Gustafson

Studying the Link Between Inter-Speaker Coordination and Speech Imitation Through Human-Machine Interactions............................................................................. 859
Leonardo Lancia, Thierry Chaminade, Noel Nguyen, Laurent Prevot

TUE-SS-4-11: SPECIAL SESSION: INCREMENTAL PROCESSING AND RESPONSIVE BEHAVIOUR

Adjusting the Frame: Biphasic Performative Control of Speech Rhythm ................................................................................................. 864
Samuel Delalez, Christophe D’Alessandro

Attentional Factors in Listeners’ Uptake of Gesture Cues During Speech Processing ............................................................................. 869
Bauleh Saryuzdi, Craig G. Chambers

Motion Analysis in Vocalized Surprise Expressions ...................................................................................................................... 874
Carlos Isbi, Takashi Minato, Hiroshi Ishiguro

Enhancing Backchannel Prediction Using Word Embeddings .................................................................................................................. 879
Robin Ruege, Markus Muller, Sebastian Stuker, Alex Wachel

A Computational Model for Phonetically Responsive Spoken Dialogue Systems ..................................................................................... 884
Eran Raveh, Ingmar Steiner, Bernd Mobius

Incremental Dialogue Act Recognition: Token- vs Chunk-Based Classification ......................................................................................... 889
Eustace Ebhotemhen, Voila Petukhova, Dietrich Klakow

TUE-SS-5-11: SPECIAL SESSION: ACOUSTIC MANIFESTATIONS OF SOCIAL CHARACTERISTICS

Clear Speech --- Mere Speech? How Segmental and Prosodic Speech Reduction Shape the Impression That Speakers Create on Listeners ............................................................................. 894
Oliver Niederhut

Relationships Between Speech Timing and Perceived Hostility in a French Corpus of Political Debates ............................................................................. 899
Charlotte Kocklis, Nicolas Audibert

Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution ......................................................................... 904
Laura Fernandez Gallardo, Benjamin Weiss

Prosodic Analysis of Attention-Drawing Speech ................................................................................................................................. 909
Carlos Isbi, Jun Arau, Norhiro Hagita

Perceptual and Acoustic Correlates of Gender in the Perpurbertual Voice .................................................................................................................. 914
Adriana Funk, Frederik Palmer

To See or Not to See: Interlocutor Visibility and Likeability Influence Convergence in Intonation ............................................................................. 919
Katrin Schweitzer, Michael Walsh, Antje Schweitzer

Acoustic Correlates of Parental Role and Gender Identity in the Speech of Expecting Parents ..................................................................................... 924
Melanie Weirich, Adrian P. Simpson

A Semi-Supervised Learning Approach for Acoustic-Prosodic Personality Perception in Under-Resourced Domains .................................................................................................................. 929
Rahen Solera-Urena, Helena Moniz, Fernando Batista, Vera Caburrao, Anna Pompili, Ramon Fernandez Astudillo, Joana Campan, Ana Patra, Isabel Trancoso

Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions ............................................................................. 934
Rachel Taftman, Connor Kasten

TUE-O-3-1: NEURAL NETWORK ACOUSTIC MODELS FOR ASR 1

A Comparison of Sequence-to-sequence Models for Speech Recognition .................................................................................................................. 939
Rohit Prabhavalkar, Kanishka Rao, Tara N. Sainath, Bo Li, Lei Johnson, Navdeep Jaitly

CTC in the Context of Generalized Full-Sum HMM Training ................................................................................................................................. 944
Albert Zeyer, Eugene Beck, Ralf Schluter, Hermann Ney

Advances in Joint CTC-Attention Based End-to-End Speech Recognition with a Deep CNN Encoder and RNN-LM ............................................................................. 949
Takahito Hori, Shinji Watanabe, Yu Zhang, William Chan

Multitask Learning with CTC and Segmental CRF for Speech Recognition ..................................................................................................... 954
Liang Lu, Lingpeng Kong, Chris Dyer, Noah A. Smith

Direct Acoustics-to-Word Models for English Conversational Speech Recognition ..................................................................................... 959
Kartik Audhkhasi, Bhuvana Ramabhadran, George Saon, Michael Picheny, David Nahamoo
Reducing the Computational Complexity of Two-Dimensional LSTMs
Bo Li, Tara N. Sainath

TUE-O-3-2: MODELS OF SPEECH PRODUCTION

Functional Principal Component Analysis of Vocal Tract Area Functions
Jorge C. Lucero

Analysis of Acoustic-to-Articulatory Speech Inversion Across Different Accents and Languages
Ganesh Sivaraman, Carol Espy-Wilson, Martijn Wieling

Integrated Mechanical Model of [r]--[l] and [b]--[m]--[w] Producing Consonant Cluster [br]
Takayuki Aru

A Speaker Adaptive DNN Training Approach for Speaker-Independent Acoustic Inversion
Leonardo Badino, Luca Franceschini, Raman Arora, Michele Donini, Massimiliano Pontil

Acoustic-to-Articulatory Mapping Based on Mixture of Probabilistic Canonical Correlation Analysis
Hideitsu Uchida, Daisuke Saito, Nobuki Minematsu

Test-retest Repeatability of Articulatory Strategies Using Real-time Magnetic Resonance Imaging
Tanner Sorensen, Asterios Toutios, Johannes Teger, Louis Goldstein, Shrikant Narayanan

TUE-O-3-4: SPEAKER RECOGNITION

Deep Neural Network Embeddings for Text-Independent Speaker Verification
David Snyder, Daniel Garcia-Romero, Daniel Povey, Sanjeev Khudanpur

Tied Variational Autoencoder Backends for i-Vector Speaker Recognition
Jesús Villalba, Niko Brunner, Nojim Dehak

Improved Gender Independent Speaker Recognition Using Convolutional Neural Network Based Bottleneck Features
Shivesh Ranjan, John H. L. Hansen

Autoencoder Based Domain Adaptation for Speaker Recognition Under Insufficient Channel Information
Sawan Shin, Seongkyu Mun, Woosil Kim, Hansuk Ko

Nonparametrically Trained Probabilistic Linear Discriminant Analysis for i-Vector Speaker Verification
Abbas Khorasvani, Mohammad Mehdi Homayounpour

DNN Bottleneck Features for Speaker Clustering
Jesus Jorrin, Paula Garcia, Luis Buera

TUE-O-3-6: PHONATION AND VOICE QUALITY

Creak as a Feature of Lexical Stress in Estonian
Katiin Aare, Partel Lippus, Juraj Simko

Cross-Speaker Variation in Voice Source Correlates of Focus and Deaccentuation
Irena Yanushevskaya, Abbebe Ni Chasaide, Christer Gobl

Acoustic Characterization of Word-final Glottal Stops in Mizo and Assam Sora

Iterative Optimal Preemphasis for Improved Glottal-Flow Estimation by Iterative Adaptive Inverse Filtering
Parham Mokhtari, Hiroshi Ando

Automatic Measurement of Pre-aspiration
Yaniv Sheena, Misa Hejna, Yossi Adi, Joseph Keshet

Acoustic and Electroglotographic Study of Breathy and Modal Vowels as Produced by Heritage and Native Gujarati Speakers
Kirupreets Narra

TUE-O-3-8: SPEECH SYNTHESIS PROSODY

An RNN-Based Quantized F0 Model with Multi-Tier Feedback Links for Text-to-Speech Synthesis
Xin Wang, Shinji Takaki, Junichi Yamagishi

Phrase Break Prediction for Long-Form Reading TTS: Exploiting Text Structure Information
Viacheslav Klimkov, Adam Nadolski, Aleciss Moinet, Bartosz Putrycz, Roberto Barra-Chicote, Thomas Merritt, Thomas Drugman

Physically Constrained Statistical F0 Prediction for Electrolaryngeal Speech Enhancement
Kou Tanaka, Hirokazu Kameoka, Tomoki Toda, Satoshi Nakamura

DNN-SPACE: DNN-HMM-Based Generative Model of Voice F0 Contours for Statistical Phrase/Accent Command Estimation
Nobukatsu Hata, Yutakito Ohashi, Yuseke Itjma, Hirokazu Kameoka

Controlling Prominence Realisation in Parametric DNN-Based Speech Synthesis
Zofia Malisz, Harald Berthelsen, Jonas Beskow, Joakim Gustafson

Increasing Recall of Lengthening Detection via Semi-Automatic Classification
Simon Betz, Jana Yosse, Sina Zarrirees, Petra Wagner
Comparing Languages Using Hierarchical Prosodic Analysis.......................................................................................................................... 1213
Jurej Simko, Antti Sari, Katja Hirsch, Martin Faini

Intonation Facilitates Prediction of Focus Even in the Presence of Lexical Tones .......................................................................................................................... 1218
Martin Ha Kwan Ip, Anne Cutler

Mind the Peak: When Museum is Temporarily Understood as Musical in Australian English ................................................................. 1223
Katharina Zahn, Heather Kember, Bettina Braun

Pashto Intonation Patterns .......................................................................................................................... 1228
Luca Rognoni, Judith Bishop, Miriam Corris

A New Model of Final Lowering in Spontaneous Monologue ................................................................................................. 1233
Kikuo Maekawa

TUE-O-4-8: EMOTION MODELING

Speech Emotion Recognition with Emotion-Pair Based Framework Considering Emotion Distribution Information in Dimensional Emotion Space ........................................................................................................................................................................................................................................ 1238
Xi Ma, Zhiyong Wu, Jia Jia, Mingxing Xu, Helen Meng, Lianhong Cai

Adversarial Auto-Encoders for Speech Based Emotion Recognition .......................................................................................................................... 1243
Saurabh Sahu, Rahul Gupta, Ganesh Sivaraman, Wael Abdalmageed, Carol Espy-Wilson

An Investigation of Emotion Prediction Uncertainty Using Gaussian Mixture Regression .......................................................................................................................... 1248
Tong Dang, Vidyasahasaran Sethu, Julien Epps, Eliazbamby Ambikairajah

Capturing Long-Term Temporal Dependencies with Convolutional Networks for Continuous Emotion Recognition .......................................................................................................................... 1253
Soheil Khorram, Zakaria Alidenah, Dimitrios Dimitriadis, Melvin McInnis, Emily Mower Provost

Voice-to-Affect Mapping: Inferences on Language Voice Baseline Settings .......................................................................................................................... 1258
Ailbhe Ni Chasaide, Irena Yanushkevich, Christer Gold

Attentive Convolutional Neural Network Based Speech Emotion Recognition: A Study on the Impact of Input Features, Signal Length, and Acted Speech .......................................................................................................................... 1263
Michael Neumann, Ngoc Thang Vu

TUE-O-4-10: VOICE CONVERSION 1

Voice Conversion Using Sequence-to-Sequence Learning of Context Posterior Probabilities .......................................................................................................................... 1268
Hiroki Miyoshi, Yuki Saito, Shinnosuke Takamichi, Hiroshi Saruwatari

Learning Latent Representations for Speech Generation and Transformation .......................................................................................................................... 1273
Wei-Ning Hsu, Yu Zhang, James Glass

Parallel-Data-Free Many-to-Many Voice Conversion Based on DNN Integrated with Eigenspace Using a Non-Parallel Speech Corpus .......................................................................................................................... 1278
Tetsuya Hashimoto, Hidetomo Uchida, Daitsuke Saito, Nhu Ha Minh

Sequence-to-Sequence Voice Conversion with Similarity Metric Learned Using Generative Adversarial Networks .......................................................................................................................... 1283
Takashiro Kameko, Hirokazu Kameoka, Kaoru Hiramatsu, Kanso Kashino

A Mouth Opening Effect Based on Pole Modification for Expressive Singing Voice Transformation .......................................................................................................................... 1288
Luc Ardaillon, Axel Roebel

Siamese Autoencoders for Speech Style Extraction and Switching Applied to Voice Identification and Conversion .......................................................................................................................... 1293
Sayed Hamidreza Mohammad, Alexander Kain

TUE-O-5-1: NEURAL NETWORK ACOUSTIC MODELS FOR ASR 2

Recurrent Neural Aligner: An Encoder-Decoder Neural Network Model for Sequence to Sequence Mapping .......................................................................................................................... 1298
Hasim Sak, Matt Shannon, Kanishka Rao, Francoise Beaufays

Highway-LSTM and Recurrent Highway Networks for Speech Recognition .......................................................................................................................... 1303
Golan Panduk, Tara N. Sainath

Improving Speech Recognition by Revising Gated Recurrent Units .......................................................................................................................... 1308
Mirco Ravanelli, Philemon Brukel, Maurizio Omologo, Yoshua Bengio

Stochastic Recurrent Neural Network for Speech Recognition .......................................................................................................................... 1313
Jen-Tzung Chien, Chen Shen

Frame and Segment Level Recurrent Neural Networks for Phone Classification .......................................................................................................................... 1318
Jurej Simko, Sebastian Tschatschek, Franz Pernkopf

Deep Learning-Based Telephony Speech Recognition in the Wild .......................................................................................................................... 1323
Kyu J. Han, Seongjung Hahn, Byung-Hak Kim, Jungsuk Kim, Ian Lane
VOLUME 3

TUE-O-5-2: SPEAKER RECOGNITION EVALUATION

The I4U Mega Fusion and Collaboration for NIST Speaker Recognition Evaluation 2016 ......................................................... 1328

The MIT-LI, JHU and LRDE NIST 2016 Speaker Recognition Evaluation System ................................................................. 1333
Pedro A. Torres-Carrascal, Fred Richardson, Shaham Nercissian, Douglas Sturim, William Campbell, Youngjune Gwon, Swaroop Fattam, Najim Dehak, Harish Mallidi, Pham Sankar Nidadavolu, Ruschi Li, Reda Dehak

Nuance - Politecnico di Torino's 2016 NIST Speaker Recognition Evaluation System ............................................................... 1338
Daniele Costabile, Claudio Vair, Emanuele Dalmasso, Kevin Farrell, Grzegorz Karczyski, Sandro Cumani, Pietro Lafface

UTD-CRSS Systems for 2016 NIST Speaker Recognition Evaluation .......................................................... 1343
Chunlei Zhang, Fahimeh Bahmaninezhad, Shivesh Ranjan, Chengzhui Yu, Navid Shokouhi, John H. L. Hansen

Analysis and Description of ABC Submission to NIST SRE 2016 .............................................................. 1348
Oldrich Pichtot, Pavel Matejka, Anna Silnova, Ondrej Novotny, Mireia Diz Sanchez, Johan Rohdin, Ondrej Glembek, Niko Brammer, Albert Swart, Jesus Jorrin-Prieto, Paola Garcia, Luis Buera, Patrick Kenny, Jahanger Alam, Gautam Bhattacharya

The 2016 NIST Speaker Recognition Evaluation .................................................................................................................. 1353
Seyed Omid Sadjadi, Timothee Kheyrkhah, Audrey Tong, Craig Greenberg, Douglas Reynolds, Elliot Singer, Lisa Mason, Jaime Hernandez-Cordero

TUE-O-5-4: GLOTTAL SOURCE MODELING

A New Cosine Series Anti-aliasing Function and its Application to Aliasing-Free Glottal Source Models for Speech and Singing Synthesis .......................................................................................................................... 1358
Hsieh-Kai Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, Tomoki Toda, Toshio Irih

Speaking Style Conversion from Normal to Lombard Speech Using a Glottal Vocoder and Bayesian GMMs .......................... 1363
Rodrigo Martinez, Sean D. Peterson, Pavel Prado, Patricio Orio, Marius Zanatau

Reducing Mismatch in Training of DNN-Based Glottal Excitation Models in a Statistical Parametric Text-to-Speech System ................................................................. 1368
Lauri Juvola, Bjaybou Bollepalli, Junichi Yamagishi, Puavo Aliku

Semi Parametric Concatenative TTS with Instant Voice Modification Capabilities ................................................................. 1373
Alexander Sorin, Sava Shechman, Asaf Rendel

Modeling Laryngeal Muscle Activation Noise for Low-Order Physiological Based Speech Synthesis ........................................ 1378
Rodrigo Martinez, Sean D. Peterson, Pavel Prado, Patricio Orio, Marius Zanatau

Direct Modelling of Magnitude and Phase Spectra for Statistical Parametric Speech Synthesis ............................................. 1383
Felipe Espic, Caxia Valentini Botinhao, Simon King

TUE-O-5-6: PROSODY: RHYTHM, STRESS, QUANTITY AND PHRASING

Similar Prosodic Structure Perceived Differently in German and English ............................................................................. 1388
Heather Kember, Ann-Kathrin Grobe, Katharina Zahnner, Bettina Braun, Andrea Weber, Anne Culer

Disambiguate or not? --- The Role of Prosody in Unambiguous and Potentially Ambiguous Anaphora Production in Strictly Mandarin Parallel Structures ................................................................. 1393
Laying Hou, Bert Le Brun, Rene Kager

Acoustic Properties of Canonical and Non-Canonical Stress in French, Turkish, Armenian and Brazilian Portuguese ................................ .................................................................................................................. 1398
Angeliki Athanasopoulou, Irene Vogel, Hossein Doltlan

Phonological Complexity, Segment Rate and Speech Tempo Perception ...................................................................................... 1403
Leendert Plag, Rachel Smith

On the Duration of Mandarin Tones ............................................................................................................................................. 1407
Jing Tong, Yu Zhang, Ajun Li, Li Xu

The Formant Dynamics of Long Close Vowels in Three Varieties of Swedish ........................................................................... 1412
Otto Ewald, Eva Lima Ann, Susanne Schoetz

TUE-O-5-8: SPEECH RECOGNITION FOR LANGUAGE LEARNING

Bidirectional LSTM-RNN for Improving Automated Assessment of Non-Native Children's Speech ........................................... 1417
Yao Qian, Keelan Evamini, Xinhuo Wang, Chong Min Lee, Matthew Mulholland

Automatic Scoring of Shadowing Speech Based on DNN Posteriors and Their DTW .......................................................... 1422
Jianwei Yue, Fumiya Shiozawa, Shohi Toyama, Yutaka Tamauchi, Kayoko Ito, Daisuke Saito, Nobuaki Minematsu

Off-Topic Spoken Response Detection Using Siamese Convolutional Neural Networks .......................................................... 1427
Chong Min Lee, So-Youn Yoon, Xinhuo Wang, Matthew Mulholland, Ikkyu Choi, Keelan Evamini

Phonological Feature Based Mispronunciation Detection and Diagnosis Using Multi-Task DNNs and Active Learning ............. 1432
Vipul Arora, Aditi Lahiri, Henning Reetz
Detection of Mispronunciations and Disfluencies in Children Reading Aloud ........................................... 1437
Automatic Assessment of Non-Native Prosody by Measuring Distances on Prosodic Label Sequences .......... 1442

TUE-P-3-2: SPEAKER CHARACTERIZATION AND RECOGNITION

Speaker Verification via Estimating Total Variability Space Using Probabilistic Partial Least Squares ................. 1537
Chen Chen, Jiqing Han, Yilin Pan
Deep Speaker Feature Learning for Text-Independent Speaker Verification .................................................. 1542
Lantian Li, Yiziang Chen, Ying Shi, Zhiyuan Tang, Dong Wang
Duration Mismatch Compensation Using Four-Covariance Model and Deep Neural Network for Speaker Verification ................................................................. 1547
Pierre-Michel Bousquet, Michael Rouvier
Extended Variability Modeling and Unsupervised Adaptation for PLDA Speaker Recognition .......................... 1552
Alan McCree, Gregory Sell, Daniel Garcia-Romero
Improving the Effectiveness of Speaker Verification Domain Adaptation with Inadequate In-Domain Data ......... 1557
Bent J. Borgstrom, Elliot Singer, Douglas Reynolds, Seyed Omid Sadjadi
i-Vector DNN Scoring and Calibration for Noise Robust Speaker Verification .............................................. 1562
Zhili Tan, Man-Wai Mak
Analysis of Score Normalization in Multilingual Speaker Recognition ...................................................... 1567
Pavel Matejka, Ondrej Novotny, Oldrich Plichta, Lukas Burget, Mireia Diez Sanchez, Jan Cernocky
Alternative Approaches to Neural Network Based Speaker Verification ................................................... 1572
Anna Silnova, Lukas Burget, Jan Cernocky
A Distribution Free Formulation of the Total Variability Model ...................................................................... 1576
Ruchir Travadi, Shrikanth S. Narayanan

TUE-O-5-10: STANCE, CREDIBILITY, AND DECEPTION

Inferring Stance from Prosody ........................................................................................................ 1447
Nigel G. Ward, Jason C. Carlson, Olac Fuentes, Diego Castan, Elizabeth E. Shriberg, Andreas Tsiartas
Exploring Dynamic Measures of Stance in Spoken Interaction ................................................................. 1452
Gina-Anne Levow, Richard A. Wright
Opinion Dynamics Modeling for Movie Review Transcripts Classification with Hidden Conditional Random Fields ............................................................... 1457
Valentin Barriere, Chloe Clavel, Slim Essid
Transfer Learning Between Concepts for Human Behavior Modeling: An Application to Sincerity and Deception Prediction ............................................................... 1462
Qinyi Luo, Rahul Gupta, Shrikanth S. Narayanan
The Sound of Deception - What Makes a Speaker Credible? .................................................................... 1467
Anne Schroder, Simon Stone, Peter Birkholz
Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection ........................................... 1472
Gideon Mendels, Sarah Ha Lestian, Kai-Zhan Lee, Julia Herschberg

TUE-P-3-4: SHORT UTTERANCES SPEAKER RECOGNITION

A Generative Model for Score Normalization in Speaker Recognition .......................................................... 1477
Albert Swart, Yoko Brammer
Content Normalization for Text-Dependent Speaker Verification ................................................................. 1482
Subhadeep Dey, Srikant Madikeri, Petr Motlicek, Marc Ferras
End-to-End Text-Independent Speaker Verification with Triplet Loss on Short Utterances ............................. 1487
Chunlei Zhang, Kazushito Kaishida
Adversarial Network Bottleneck Features for Noise Robust Speaker Verification ........................................ 1492
Hong Yu, Zheng-Hua Tan, Zhanyu Ma, Jun Guo
What Does the Speaker Embedding Encode? ............................................................................................... 1497
Shuai Wang, Yanmin Quan, Kai Yu
Incorporating Local Acoustic Variability Information into Short Duration Speaker Verification .................... 1502
Jianbo Ma, Vidhyasaharan Sethu, Eliahatham Ambikairajah, Kong Aik Lee
DNN i-Vector Speaker Verification with Short, Text-Constrained Test Utterances ......................................... 1507
Jinghua Zhong, Wenping Hu, Frank K. Soong, Helen Meng
Time-Varying Autoregressions for Speaker Verification in Reverberant Conditions .................................... 1512
Ville Vestman, Dhananjaya Gowda, Md. Sadiul Souad, Pauvo Alka, Tomi Kinnunen
Deep Speaker Embeddings for Short-Duration Speaker Verification ......................................................... 1517
Gautam Bhattacharya, Jahangir Alam, Patrick Kenny
Using Voice Quality Features to Improve Short-Utterance, Text-Independent Speaker Verification Systems ................................................................. 1522
Soo Jin Park, Gary Yeung, Jody Kreiman, Patricia A. Keating, Abeer Alwan
Gain Compensation for Fast i-Vector Extraction Over Short Duration ....................................................... 1527
Kong Aik Lee, Haizhou Li
Joint Training of Expanded End-to-End DNN for Text-Dependent Speaker Verification .................................. 1532
Hee-Soo Heo, Jee-Weon Jung, Bi-Ho Yang, Sung-Hyun Yoon, Ha-Jin Yu
Domain Mismatch Modeling of Out-Domain i-Vectors for PLDA Speaker Verification ................................................................. 1581
Md. Hafizur Rahman, Ivan Himawan, David Dean, Sridha Sridharan

TUE-P-4-1: ACOUSTIC MODELS FOR ASR 1

An Exploration of Dropout with LSTMs ........................................................................................................................................ 1586
Gangfeng Cheng, Vijayaditya Peddinti, Daniel Povey, Vimal Manohar, Sanjeev Khudanpur, Yonghong Yan
Residual LSTM: Design of a Deep Recurrent Architecture for Distant Speech Recognition .......................................................... 1591
Jaeyoung Kim, Mostafa El-Khamy, Jungwon Lee
Unfolded Deep Recurrent Convolutional Neural Network with Jump Ahead Connections for Acoustic Modeling ....................... 1596
Dung T. Tran, Marc Delcroix, Shigeki Karita, Michael Hentschel, Atsunori Ogaswa, Tomohiro Nakatani
Forward-Backward Convolutional LSTM for Acoustic Modeling ............................................................................................................. 1601
Shigeki Karita, Atsunori Ogaswa, Marc Delcroix, Tomohiro Nakatani
Convolutional Recurrent Neural Networks for Small-Footprint Keyword Spotting ................................................................. 1606
Serkan O. Arik, Markus Kriegel, Rewon Child, Joel Hestness, Andrew Gibiansky, Chris Fougnier, Ryan Prenger, Adam Coates
Deep Activation Mixture Model for Speech Recognition ..................................................................................................................... 1611
Chunyang Wu, Mark J. F. Gales
Ensembles of Multi-Scale VGG Acoustic Models ................................................................................................................................. 1616
Michael Heck, Masayuki Suzuki, Takashi Fukuda, Gaohao Kurata, Satoshi Nakamura
Training Context-Dependent DNN Acoustic Models Using Probabilistic Sampling ...................................................... 1621
Tamas Grosz, Gabor Gosztolya, Laszlo Toth
A Comparative Evaluation of GMM-Free State Tying Methods for ASR ............................................................................................ 1626
Tamas Grosz, Gabor Gosztolya, Laszlo Toth

TUE-P-4-2: ACOUSTIC MODELS FOR ASR 2

Backstitch: Counteracting Finite-Sample Bias via Negative Steps .................................................................................................... 1631
Yiming Wang, Vijayaditya Peddinti, Hainan Xu, Xiaohui Zhang, Daniel Povey, Sanjeev Khudanpur
Node Pruning Based on Entropy of Weights and Node Activity for Small-Footprint Acoustic Model Based on Deep Neural Networks .......................................................................................................................................................... 1636
Ryo Takeda, Kazehiro Nakadai, Kazunori Komatani
End-to-End Training of Acoustic Models for Large Vocabulary Continuous Speech Recognition with TensorFlow .......................... 1641
Ehsan Variani, Tom Bagby, Erik McDermott, Michel Bacchiani
An Efficient Phone N-Gram Forward-Backward Computation Using Dense Matrix Multiplication ............................................. 1646
Khe Chai Sim, Arun Narayanan
Parallel Neural Network Features for Improved Tandem Acoustic Modeling ........................................................................ 1651
Zohran Tuske, Wilfried Michel, Ralf Schluter, Hermann Ney
Acoustic Feature Learning via Deep Variational Canonical Correlation Analysis ................................................................. 1656
Qingming Tang, Weiran Wang, Karen Livescu

TUE-P-4-3: DIALOG MODELING

Online End-of-Turn Detection from Speech Based on Stacked Time-Asynchronous Sequential Networks ........................................ 1661
Ryo Masumura, Yuichi Asami, Hirokazu Masataki, Ryo Ishii, Ryouichi Higashimaka
Improving Prediction of Speech Activity Using Multi-Participant Respiratory State ............................................................... 1666
Marcin Wodarczak, Kornel Laskowski, Mattias Heldner, Katrin Aure
Turn-Taking Offsets and Dialogue Context ................................................................................................................................. 1671
Peter A. Heeman, Rebecca Lumsford
Towards Deep End-of-Turn Prediction for Situated Spoken Dialogue Systems ............................................................. 1676
Angelika Mair, Juliane Hough, David Schlangen
End-of-Utterance Prediction by Prosodic Features and Phrase-Dependence Structure in Spontaneous Japanese Speech ........................................................................................................................................................................... 1681
Yuichi Ishimoto, Takehiro Teraoka, Mika Enomoto

Turn-Taking Estimation Model Based on Joint Embedding of Lexical and Prosodic Contents ...................................................... 1686
Chaoran Liu, Carlos Ishi, Hiroshi Ishiguro
Social Signal Detection in Spontaneous Dialogue Using Bidirectional LSTM-CTC .............................................................................. 1691
Hirofumi Inaguma, Koji Inoue, Masato Mimura, Tatuya Kawahara
Entrainment in Multi-Party Spoken Dialogues at Multiple Linguistic Levels .................................................................................. 1696
Zahra Rahimi, Anish Kumar, Diane Litman, Susannah Paletz, Mingzhi Yu
Measuring Synchrony in Task-Based Dialogues ................................................................................................................................. 1701
Justine Reverdy, Carl Vogel

Sequence to Sequence Modeling for User Simulation in Dialog Systems ......................................................................................... 1706
Paul Crook, Alex Marin
Human and Automated Scoring of Fluency, Pronunciation and Intonation During Human--Machine Spoken Dialog Interactions ......................................................................................................................................................... 1711
Vikram Ramanarayanan, Patrick L. Lange, Keelan Evansini, Hillary R. Molloy, David Suendermann-Oeft
Hierarchical LSTMs with Joint Learning for Estimating Customer Satisfaction from Contact Center Calls ........................................... 1716
Atsushi Ando, Ryo Masumura, Hosana Kamiyama, Satoshi Kobashikawa, Yuichi Aono
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain-Independent User Satisfaction Reward Estimation for Dialogue Policy Learning</td>
<td>1721</td>
</tr>
<tr>
<td>Stefan Utes, Pavel Budzianowski, Inigo Casabuena, Nikola Mrkic, Lina Rojas-Barahona, Pei-Hao Su, Tsung-Hsien Wen, Milica Gasic, Steve Young</td>
<td></td>
</tr>
<tr>
<td>Analysis of the Relationship Between Prosodic Features of Fillers and its Forms or Occurrence Positions</td>
<td>1726</td>
</tr>
<tr>
<td>Shizuka Nakamura, Ryosuke Nakanishi, Katsuya Takanashi, Tatsuya Kawahara</td>
<td></td>
</tr>
<tr>
<td>Cross-Subject Continuous Emotion Recognition Using Speech and Body Motion in Dyadic Interactions</td>
<td>1731</td>
</tr>
<tr>
<td>Syeda Narijis Fatima, Engin Erzin</td>
<td></td>
</tr>
<tr>
<td><strong>TUE-P-5-1: L1 AND L2 ACQUISITION</strong></td>
<td></td>
</tr>
<tr>
<td>An Automatically Aligned Corpus of Child-Directed Speech</td>
<td>1736</td>
</tr>
<tr>
<td>Micha Elsner, Kiwako Ito</td>
<td></td>
</tr>
<tr>
<td>A Comparison of Danish Listeners’ Processing Cost in Judging the Truth Value of Norwegian, Swedish, and English Sentences</td>
<td>1741</td>
</tr>
<tr>
<td>Patrick F. Reidy, Mary E. Beckman, Jan Edwards, Benjamin Munson</td>
<td></td>
</tr>
<tr>
<td>On the Role of Temporal Variability in the Acquisition of the German Vowel Length Contrast</td>
<td>1745</td>
</tr>
<tr>
<td>Felicitas Kleber</td>
<td></td>
</tr>
<tr>
<td>A Data-driven Approach for Perceptually Validated Acoustic Features for Children's Sibilant Fricative Productions</td>
<td>1750</td>
</tr>
<tr>
<td>Yujia Xiao, Frank K. Soong</td>
<td></td>
</tr>
<tr>
<td>Proficiency Assessment of ESL Learner’s Sentence Prosody with TTS Synthesized Voice as Reference</td>
<td>1755</td>
</tr>
<tr>
<td>Si Chen, Yunqian He, Chun Wang, Yuen, Bei Li, Yihe Yang</td>
<td></td>
</tr>
<tr>
<td>Changes in Early L2 Cue-Weighting of Non-Native Speech: Evidence from Learners of Mandarin Chinese</td>
<td>1765</td>
</tr>
<tr>
<td>Sech Warner</td>
<td></td>
</tr>
<tr>
<td>Directing Attention During Perceptual Training: A Preliminary Study of Phonetic Learning in Southern Min by Mandarin Speakers</td>
<td>1770</td>
</tr>
<tr>
<td>Ying Chen, Eric Pederson</td>
<td></td>
</tr>
<tr>
<td>Prosody Analysis of L2 English for Naturalness Evaluation Through Speech Modification</td>
<td>1775</td>
</tr>
<tr>
<td>Dean Luo, Ruxin Luo, Lixin Wang</td>
<td></td>
</tr>
<tr>
<td>Measuring Encoding Efficiency in Swedish and English Language Learner Speech Production</td>
<td>1779</td>
</tr>
<tr>
<td>Gintare Grigonyte, Gerald Schneider</td>
<td></td>
</tr>
<tr>
<td>Lexical Adaptation to a Novel Accent in German: A Comparison Between German, Swedish, and Finnish Listeners</td>
<td>1784</td>
</tr>
<tr>
<td>Adriana Hunalikova, JennyEkstrom</td>
<td></td>
</tr>
<tr>
<td>Qualitative Differences in L3 Learners’ Neurophysiological Response to L1 versus L2 Transfer</td>
<td>1789</td>
</tr>
<tr>
<td>Alejandra Keidel Fernandez, Thomas Horberg</td>
<td></td>
</tr>
<tr>
<td>Articulation Rate in Swedish Child-Directed Speech Increases as a Function of the Age of the Child Even When Surprisal is Controlled for</td>
<td>1794</td>
</tr>
<tr>
<td>Kaile Zhang, Gang Peng</td>
<td></td>
</tr>
<tr>
<td>The Relationship Between the Perception and Production of Non-Native Tones</td>
<td>1799</td>
</tr>
<tr>
<td>Ellen Marklund, Elisabeth Eir Cortes, Johan Sjöns</td>
<td></td>
</tr>
<tr>
<td>MMN Responses in Adults After Exposure to Bimodal and Unimodal Frequency Distributions of Rotated Speech</td>
<td>1804</td>
</tr>
<tr>
<td>Stefan Ultes, Pawel Budzianowski, Roman Cmejla, Jan Rusz, Trine Askjaer-Jorgensen</td>
<td></td>
</tr>
<tr>
<td><strong>TUE-P-5-2: VOICE, SPEECH AND HEARING DISORDERS</strong></td>
<td></td>
</tr>
<tr>
<td>Float Like a Butterfly Sting Like a Bee: Changes in Speech Preceded Parkinsonism Diagnosis for Muhammad Ali</td>
<td>1809</td>
</tr>
<tr>
<td>Visar Berisha, Julie Liss, Timothy Huston, Alan Wisler, Tishan Jiao, Jonathan Eig</td>
<td></td>
</tr>
<tr>
<td>Cepstral and Entropy Analyses in Vowels Excerpted from Continuous Speech of Dysphonic and Control Speakers</td>
<td>1814</td>
</tr>
<tr>
<td>Antonella Castellana, Andreas Selamatiz, Giampiero Savini, Alessio Carullo, Arianna Astolfi</td>
<td></td>
</tr>
<tr>
<td>Classification of Bulbar ALS from Kinematic Features of the Jaw and Lips: Towards Computer-Mediated Assessment</td>
<td>1819</td>
</tr>
<tr>
<td>Andrea Bandini, Jordan R. Green, Lorne Zinman, Yana Yunussova</td>
<td></td>
</tr>
<tr>
<td>Zero Frequency Filter Based Analysis of Voice Disorders</td>
<td>1824</td>
</tr>
<tr>
<td>Nagaraj Adiga, Vikram C. M., Keerthi Pullerla, S. R. Mahadeva Prasanna</td>
<td></td>
</tr>
<tr>
<td>Hypersalinity Severity Analysis in Cleft Lip and Palate Speech Using Vowel Space Area</td>
<td>1829</td>
</tr>
<tr>
<td>Nikisha K., Shirsh Kalia, C. M. Vikram, M. Pushpavathi, S. R. Mahadeva Prasanna</td>
<td></td>
</tr>
<tr>
<td>Automatic Prediction of Speech Evaluation Metrics for Dysarthric Speech</td>
<td>1834</td>
</tr>
<tr>
<td>Imed Llaaridh, Waad Ben Kheder, Corinne Freduelle, Christine Meunier</td>
<td></td>
</tr>
<tr>
<td>Akpionk — A Mobile Monitoring Solution for Parkinson’s Disease</td>
<td>1839</td>
</tr>
<tr>
<td>Philipp Klumpp, Thomas Janu, Tomas Arias-Vergara, J. C. Vasquez-Correa, Juan Rafael Orozco-Arroyave, Elmar Noth</td>
<td></td>
</tr>
<tr>
<td>Dysprosody Differentiate Between Parkinson’s Disease, Progressive Supranuclear Palsy, and Multiple System Atrophy</td>
<td>1844</td>
</tr>
<tr>
<td>Ming Yu, Visar Berisha, Julie Liss</td>
<td></td>
</tr>
<tr>
<td>Interpretable Objective Assessment of Dysarthric Speech Based on Deep Neural Networks</td>
<td>1849</td>
</tr>
<tr>
<td>Bhavik Vachhani, Chitravelaka Bhat, Biswajit Das, Sunil Kumar Kopparapu</td>
<td></td>
</tr>
<tr>
<td>Deep Autoencoder Based Speech Features for Improved Dysarthric Speech Recognition</td>
<td>1854</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Prediction of Speech Delay from Acoustic Measurements</td>
<td>1859</td>
</tr>
<tr>
<td>Jason Liley, Madhavi Ramagiri, H. Timothy Bunnell</td>
<td></td>
</tr>
<tr>
<td>The Frequency Range of &quot;The Ling Six Sounds&quot; in Standard Chinese</td>
<td>1864</td>
</tr>
<tr>
<td>Aijun Li, Hua Zhang, Wen Sun</td>
<td></td>
</tr>
<tr>
<td>Production of Sustained Vowels and Categorical Perception of Tones in Mandarin Among Cochlear-Impalented Children</td>
<td>1869</td>
</tr>
<tr>
<td>Wentao Gu, Jiao Yin, James Mahshie</td>
<td></td>
</tr>
<tr>
<td><strong>TUE-P-5:3: SOURCE SEPARATION AND VOICE ACTIVITY DETECTION</strong></td>
<td></td>
</tr>
<tr>
<td>Audio Content Based Geotagging in Multimedia</td>
<td>1874</td>
</tr>
<tr>
<td>Anurag Kumar, Benjamin Elizalde, Bhiksha Raj</td>
<td></td>
</tr>
<tr>
<td>Time Delay Histogram Based Speech Source Separation Using a Planar Array</td>
<td>1879</td>
</tr>
<tr>
<td>Feng Guo, Yuhang Cao, Zheng Liu, Jiajiu Li, Xiaotong Li, Xingbo Yuan</td>
<td></td>
</tr>
<tr>
<td>Excitation Source Features for Improving the Detection of Vowel Onset and Offset Points in a Speech Sequence</td>
<td>1884</td>
</tr>
<tr>
<td>Gayyadhari Pradhan, Avinash Kumar, S. Shumavasuuddin</td>
<td></td>
</tr>
<tr>
<td>A Contrast Function and Algorithm for Blind Separation of Audio Signals</td>
<td>1889</td>
</tr>
<tr>
<td>Wei Gao, Roberto Tognoni, Victor Sreevan</td>
<td></td>
</tr>
<tr>
<td>Weighted Spatial Covariance Matrix Estimation for MUSIC Based TDOA Estimation of Speech Source</td>
<td>1894</td>
</tr>
<tr>
<td>Chenglin Xu, Xiong Xiao, Sining Sun, Wei Rao, Eng Song Cheng, Haizhao L, Bjorn Hedeberget</td>
<td></td>
</tr>
<tr>
<td>Speaker Direction-of-Arrival Estimation Based on Frequency-Independent Beampattern</td>
<td>1899</td>
</tr>
<tr>
<td>Xinyun Wang, Changchun Bao, Feng Bao</td>
<td></td>
</tr>
<tr>
<td>Improved End-of-Query Detection for Streaming Speech Recognition</td>
<td>1909</td>
</tr>
<tr>
<td>Matt Shannon, Gabo Simko, Shao-Ting Chang, Carolina Paradua</td>
<td></td>
</tr>
<tr>
<td>Using Approximated Auditory Roughness as a Pre-Filtering Feature for Human Screaming and Affective Speech AED</td>
<td>1914</td>
</tr>
<tr>
<td>Di He, Zaojiu Cheng, Mark Hasegawas-Johnston, Deming Chen</td>
<td></td>
</tr>
<tr>
<td>Improving Source Separation via Multi-Speaker Representations</td>
<td>1919</td>
</tr>
<tr>
<td>Jeroen Zegers, Hugo Van Hamme</td>
<td></td>
</tr>
<tr>
<td>Multiple Sound Source Counting and Localization Based on Spatial Principal Eigenvector</td>
<td>1924</td>
</tr>
<tr>
<td>Bing Yang, Hong Liu, Cheng Pang</td>
<td></td>
</tr>
<tr>
<td>Subband Selection for Binaural Speech Source Localization</td>
<td>1929</td>
</tr>
<tr>
<td>Girija Ramesan Karthik, Prasanta Kumar Ghosh</td>
<td></td>
</tr>
<tr>
<td>Unmixing Convolutive Mixtures by Exploiting Amplitude Co-Modulation: Methods and Evaluation on Mandarin Speech Recordings</td>
<td>1934</td>
</tr>
<tr>
<td>Bo-Bai Chen, Huang-Yi Lee, Yi-Wen Liu</td>
<td></td>
</tr>
<tr>
<td>Bimodal Recurrent Neural Network for Audiovisual Voice Activity Detection</td>
<td>1938</td>
</tr>
<tr>
<td>Fei Tao, Carlos Busso</td>
<td></td>
</tr>
<tr>
<td>Domain-Specific Utterance End-Point Detection for Speech Recognition</td>
<td>1943</td>
</tr>
<tr>
<td>Roland Mou, Arvind Rastrow, Kyle Goehner, Gustav Tissot, Shawn Joseph, Bjorn Hedeberget</td>
<td></td>
</tr>
<tr>
<td>Speech Detection and Enhancement Using Single Microphone for Distant Speech Applications in Reverberant Environments</td>
<td>1948</td>
</tr>
<tr>
<td>Vinay Kothapally, John H. L. Hansen</td>
<td></td>
</tr>
<tr>
<td><strong>TUE-P-5:4: SPEECH-ENHANCEMENT</strong></td>
<td></td>
</tr>
<tr>
<td>A Post-Filtering Approach Based on Locally Linear Embedding Difference Compensation for Speech Enhancement</td>
<td>1953</td>
</tr>
<tr>
<td>Yi-Chiao Wu, Hsin-Tai Huang, Syu-Siung Wang, Chin-Cheng Hsu, Yu Tsao, Hsin-Min Wang</td>
<td></td>
</tr>
<tr>
<td>Multi-Target Ensemble Learning for Monaural Speech Separation</td>
<td>1958</td>
</tr>
<tr>
<td>Hui Zhang, Xuelong Zhang, Guanglai Guo</td>
<td></td>
</tr>
<tr>
<td>Improved Example-Based Speech Enhancement by Using Deep Neural Network Acoustic Model for Noise Robust</td>
<td>1963</td>
</tr>
<tr>
<td>Example Search</td>
<td></td>
</tr>
<tr>
<td>Ansumori Ogawa, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani</td>
<td></td>
</tr>
<tr>
<td>Subjective Intelligibility of Deep Neural Network-Based Speech Enhancement</td>
<td>1968</td>
</tr>
<tr>
<td>Femke B. Gelderblom, Tron T. Trondstad, Erlend Magnus Viggen</td>
<td></td>
</tr>
<tr>
<td>Real-Time Modulation Enhancement of Temporal Envelopes for Increasing Speech Intelligibility</td>
<td>1973</td>
</tr>
<tr>
<td>Maria Koutsogiannaki, Holly Francois, Kihyun Choo, Eunmi Oh</td>
<td></td>
</tr>
<tr>
<td>On the Influence of Modifying Magnitude and Phase Spectrum to Enhance Noisy Speech Signals</td>
<td>1978</td>
</tr>
<tr>
<td>Hans-Gunter Hirsch, Michael Gref</td>
<td></td>
</tr>
<tr>
<td>MixMax Approximation as a Super-Gaussian Log-Spectral Amplitude Estimator for Speech Enhancement</td>
<td>1983</td>
</tr>
<tr>
<td>Robert Rehr, Timo Gerkmann</td>
<td></td>
</tr>
<tr>
<td>Binary Mask Estimation Strategies for Constrained Imputation-Based Speech Enhancement</td>
<td>1988</td>
</tr>
<tr>
<td>Ricard Marxer, Jon Barker</td>
<td></td>
</tr>
<tr>
<td>A Fully Convolutional Neural Network for Speech Enhancement</td>
<td>1993</td>
</tr>
<tr>
<td>Se Rim Park, Jin Won Lee</td>
<td></td>
</tr>
</tbody>
</table>
Areal and Phylogenetic Features for Multilingual Speech Synthesis .......................................................... 2078
Alexander Grahn, Richard Sproat

WED-SS-6-11: SPECIAL SESSION: DATA COLLECTION, TRANSCRIPTION AND ANNOTATION ISSUES IN CHILD LANGUAGE ACQUISITION

S.L.PAnnotator: Tools for Implementing Sign Language Phonetic Annotation ........................................... 2083
Kathleen Currie Hall, Scott Mackie, Michael Fry, Oksana Tkachman

The LENA™ System Applied to Swedish: Reliability of the Adult Word Count Estimate ....................... 2088
Iris-Corinna Schwarz, Noor Botros, Aleksandra Lourd, Amelie Marcussen, Henrik Tidellius, Ellen Marklund

What do Babies Hear? Analyses of Child- and Adult-Directed Speech .................................................... 2093
Marisa Casillas, Andrei Amatuni, Amanda Seidl, Melanie Soderstrom, Anne S. Warlaumont, Erika Bergelson

A New Workflow for Semi-Automatized Annotations: Tests with Long-Form Naturalistic Recordings of Childrens Language Environments ...................................................................................... 2098
Marisa Casillas, Erika Bergelson, Anne S. Warlaumont, Alejandrina Cristia, Melanie Soderstrom, Mark Pendam, Han Slotjes

Top-Down versus Bottom-Up Theories of Phonological Acquisition: A Big Data Approach .................... 2103
Christina Bergmann, Sho Tsuji, Alejandrina Cristia

Which Acoustic and Phonological Factors Shape Infants' Vowel Discrimination? Exploiting Natural Variation in InPhonDB ........................................................................................................................................ 2108
Sho Tsuji, Alejandrina Cristia

WED-SS-7-1: SPECIAL SESSION: DIGITAL REVOLUTION FOR UNDER-RESOURCED LANGUAGES 2

The ABAIR Initiative: Bringing Spoken Irish into the Digital Space ............................................................ 2113
Ashlie Ni Chasaide, Neasa Ni Chiarain, Christophe Wendler, Harold Berthelsen, Andy Murphy, Chris Wohl

Very Low Resource Radio Browsing for Agile Developmental and Humanitarian Monitoring .................. 2118
Armin Sach, Raghab Menon, Hugh Cameron, William Kibira, John Quinn, Thomas Niesler

Extracting Situation Frames from Non-English Speech: Evaluation Framework and Pilot Results ........ 2123
Nikolaos Malandrakis, Ondrej Glembek, Shrikant S. Narayanan

Eliciting Meaningful Units from Speech ........................................................................................................ 2128
Danny Kocharov, Tanjana Kachkovskaya, Pavel Srelin

Unsupervised Speech Signal to Symbol Transformation for Zero Resource Speech Applications .......... 2133
Saurabhchand Bhatti, Shekhur Nayak, K. Sri Rama Murty

Machine Assisted Analysis of Vowel Length Contrasts in Wolof ............................................................... 2138
Elodie Gautier, Laurent Besacier, Sylvie Vosin

Leveraging Text Data for Word Segmentation for Underresourced Languages ........................................ 2143
Thomas Gairner, Benedikt Boeninghoff, Oliver Walter, Reinhild Haeb-Umbach

Improving DNN Bluetooth Narrowband Acoustic Models by Cross-Bandwidth and Cross-Lingual Initialization ........................................................... 2148
Xiaodan Zhuang, Arnab Ghoshal, Antti-Veikko Rosti, Matthias Paulik, Daben Liu

Joint Estimation of Articulatory Features and Acoustic Models for Low-Resource Languages .................. 2153
Basil Abraham, S. Umesh, Neetha Mariman Joy

Transfer Learning and Distillation Techniques to Improve the Acoustic Modeling of Low Resource Languages .............................................................. 2158
Basil Abraham, Tejaswi Seeram, S. Umesh

Building an ASR Corpus Using Althingi’s Parliamentary Speeches .......................................................... 2163
Inga Rún Helgadóttir, Robert Kjaran, Anna Björk Niklasdóttir, Jon Gudnason

Implementation of a Radiology Speech Recognition System for Estonian Using Open Source Software .... 2168
Tanel Alumae, Andrus Paats, Ivo Fridolin, Einar Meister

Building ASR Corpora Using Eyra .................................................................................................................. 2173
Jon Gudnason, Matthias Petersson, Robert Kjaran, Simon Klupfel, Anna Björk Niklasdottir

Rapid Development of TTS Corpora for Four South African Languages ................................................. 2178
Daniel Van Niekerk, Chari Van Heerden, Marelie Davel, Neil Kleyhans, Oddur Kjaranssnn, Martin Jansche, Linne Ha

Uniform Multilingual Multi-Speaker Acoustic Model for Statistical Parametric Speech Synthesis of Low-Resourced Languages ................................................................. 2183
Alexander Gauthier

Nativization of Foreign Names in TTS for Automatic Reading of World News in Swahili ......................... 2188
Joseph Mendelson, Pilar Oplustil, Oliver Watts, Simon King

WED-SS-7-11: SPECIAL SESSION: COMPUTATIONAL MODELS IN CHILD LANGUAGE ACQUISITION

Multi-Task Learning for Mispronunciation Detection on Singapore Children’s Mandarin Speech .............. 2193
Rong Tong, Nancy F. Chen, Bin Ma

Relating Unsupervised Word Segmentation to Reported Vocabulary Acquisition .................................. 2198
Elin Larsen, Alejandrina Cristia, Emmanuel Dupoux

Modelling the Informativeness of Non-Verbal Cues in Parent-Child Interaction ....................................... 2203
Mats Wren, Kristina N. Bjorkenstam, Robert Ouellet

Computational Simulations of Temporal Vocalization Behavior in Adult-Child Interaction ...................... 2208
Ellen Marklund, David Pagnan, Tove Gerdhun, Lisa Gustavsson

Approximating Phonotactic Input in Children’s Linguistic Environments from Orthographic Transcripts .... 2213
Sofia Strombergson, Jens Efland, Jana Gotze, Kristina Nilsson Bjorkenstam
Learning Weakly Supervised Multimodal Phoneme Embeddings .................................................................................................................. 2218
Rahma Chaabouni, Ewan Dunbar, Neil Zeghidour, Emmanuel Dupoux

WED-SS-8-11: SPECIAL SESSION: VOICE ATTRACTIVENESS

Personalized Quantification of Voice Attractiveness in Multidimensional Merit Space .................................................................................. 2223
Yasunari Obuchi

The Role of Temporal Amplitude Modulations in the Political Arena: Hillary Clinton vs. Donald Trump .......................................................... 2228
Hans Ruiger Bosker

Perceptual Ratings of Voice Likability Collected Through In-Lab Listening Tests vs. Mobile-Based Crowdsourcing .................................................. 2233
Laura Fernandez Gallardo, Rafael Zoqueria Jimenez, Sebastian Moller

Attractiveness of French Voices for German Listeners --- Results from Native and Non-Native Read Speech .......................................................... 2238
Jürgen Trouvain, Frank Zimmerer

Social Attractiveness in Dialogs ............................................................................................................................................................................. 2243
Antje Schweitzer, Natalie Lewandowski, Daniel Duran

A Gender Bias in the Acoustic-Melodic Features of Charismatic Speech? ............................................................................................... 2248
eszter Novak-Toth, Oliver Niebuhr, Aaja Chen

Pitch Convergence as an Effect of Perceived Attractiveness and Likability .................................................................................................. 2253
Jan Michalsky, Heike Schoormann

Does Posh English Sound Attractive? ......................................................................................................................................................... 2257
Li Juo, Chengxia Wang, Cristiano Hsu, Peter Birkholz, Yi Xu

Large-Scale Speaker Ranking from Crowdsourced Pairwise Listener Ratings .......................................................................................... 2262
Timo Baumann

WED-O-6-1: SPEECH PRODUCTION AND PHYSIOLOGY

Aerodynamic Features of French Fricatives .................................................................................................................................................... 2267
Rosario Signorelli, Sergio Hassid, Didier Demotin

Inter-Speaker Variability: Speaker Normalisation and Quantitative Estimation of Articulatory Invariants in Speech Production for French ........................................................................................................... 2272
Antoine Serrariere, Pierre Badino, Louis-Jean Boc, Laurent Lamalle, Christiane Neuschaefer-Rube

Comparison of Basic Beatboxing Articulations Between Expert and Novice Artists Using Real-Time Magnetic Resonance Imaging .................................................................................................................. 2277
Nimisha Patel, Timothy Greer, Reed Blaylock, Shrikanth S. Narayanan

Speaker-specific Biomechanical Model-based Investigation of a Simple Speech Task based on Tagged-MRI ........................................ 2282
Keyi Yang, Negan M. Harunadi, Jonghye Woo, Georges El Fakhri, Mavreen Stone, Sidney Fels

Sounds of the Human Vocal Tract ................................................................................................................................................................. 2287
Alex Eglinton

A Simulation Study on the Effect of Glottal Boundary Conditions on Vocal Tract Formants ........................................................................... 2292
Yasufumi Uezu, Tokihiko Kaburagi

WED-O-6-4: SPEECH AND HARMONIC ANALYSIS

A Robust and Alternative Approach to Zero Frequency Filtering Method for Epoch Extraction ................................................................. 2297
P. Gangamohan, B. Yegnarayana

Improving YANGsaf F0 Estimator with Adaptive Kalman Filter .................................................................................................................. 2301
Kanru Hua

A Spatio-Temporal Demodulation Technique for Pitch Estimation .............................................................................................................. 2306
Jisuendra Kumar Dham, Nagaraj Adiga, Chandra Sekhar Seelamantula

Robust Method for Estimating F0 of Complex Tone Based on Pitch Perception of Amplitude Modulated Signal .................................................. 2311
Kenichiro Miwa, Masashi Unoki

Low-Complexity Pitch Estimation Based on Phase Differences Between Low-Resolution Spectra ................................................................... 2316
Simon Graf, Tobias Herbig, Markus Buck, Gerhard Schmidt

Harvest: A High-Performance Fundamental Frequency Estimator from Speech Signals ........................................................................... 2321
Masanori Morise

WED-O-6-6: DIALOG AND PROSODY

Prosodic Event Recognition Using Convolutional Neural Networks with Context Information ................................................................. 2326
Sahrina Stehven, Ngoc Thang Vu

Prosodic Facilitation and Interference While Judging on the Veracity of Synthesized Statements ............................................................. 2331
Ramiro H. Galvez, Stefan Bemus, Agustin Gravano, Marian Trnka

An Investigation of Pitch Matching Across Adjacent Turns in a Corpus of Spontaneous German ........................................................................ 2336
Margaret Zellers, Antje Schweitzer

The Relationship Between F0 Synchrony and Speech Convergence in Dyadic Interaction ............................................................................. 2341
Sankar Mukherjee, Alessandro D’Ascoli, Noel Nguyen, Luciano Fadiga, Leonardo Badino
The Role of Linguistic and Prosodic Cues on the Prediction of Self-Reported Satisfaction in Contact Centre Phone Calls

Jordi Lluque, Carlos Sepúlveda, Martí Umbert, Luis Ángel Gallardo

Cross-Linguistic Study of the Production of Turn-Taking Cues in American English and Argentine Spanish

Pablo Brusco, Juan Manuel Perez, Agustín Gravano

WED-O-6-8: SOCIAL SIGNALS, STYLES, AND INTERACTION

Emotional Features for Speech Overlaps Classification

Olga Egorow, Andreas Wendemuth

Computing Multimodal Dyadic Behaviors During Spontaneous Diagnosis Interviews Toward Automatic Categorization of Autism Spectrum Disorder

Chin-Po Chen, Xian-Hong Tseng, Susan Shiu-Fer Gau, Chi-Chun Lee

Deriving Dyad-Level Interaction Representation Using Interlocutors Structural and Expressive Multimodal Behavior Features

Yun-Shao Lin, Chi-Chun Lee

Spotting Social Signals in Conversational Speech over IP: A Deep Learning Perspective

Raymond Brueckner, Maximilian Schmitt, Maja Pantic, Bjorn Schuller

Optimized Time Series Filters for Detecting Laughter and Filler Events

Gabor Gazdagya

Visual, Laughter, Applause and Spoken Expression Features for Predicting Engagement Within TED Talks

Fasih Haider, Fahim A. Salim, Saturnino Laz, Carl Vogel, Owen Conlan, Nick Campbell

WED-O-6-10: ACOUSTIC MODEL ADAPTATION

Large-Scale Domain Adaptation via Teacher-Student Learning

Jinyu Li, Michael L. Seltzer, Xi Wang, Rui Zhao, Yifan Gong

Improving Children’s Speech Recognition Through Explicit Pitch Scaling Based on Iterative Spectrogram Inversion

W. Ahmad, S. Shahnawazuddin, H. K. Kathania, Gayadhar Pradhan, A. B. Samaaddar

RNN-LDA Clustering for Feature Based DNN Adaptation

Yiwen Xie, Xunying Liu, Tan Lee, Lan Wang

Robust Online i-Vectors for Unsupervised Adaptation of DNN Acoustic Models: A Study in the Context of Digital Voice Assistants

Harish Arisikere, Sri Garimella

Semi-Supervised Learning with Semantic Knowledge Extraction for Improved Speech Recognition in Air Traffic Control

Ajay Srinivasamurthy, Petr Motlicek, Ivan Himawan, Gyorgy Szaszak, Youssef Oualil, Hartmut Helmke

Dynamic Layer Normalization for Adaptive Neural Acoustic Modeling in Speech Recognition

Taeung Kim, Inchul Song, Yoshua Bengio

WED-O-7-1: COGNITION AND BRAIN STUDIES

An Entrained Rhythm’s Frequency, Not Phase, Influences Temporal Sampling of Speech

Hans Rutger Bosker, Anne Kostem

Context Regularity Indexed by Auditory N1 and P2 Event-Related Potentials

Xiao Wang, Yanhui Zhang, Gang Peng

Discovering Language in Marmoset Vocalization

Sakshi Verma, K. L. Prateek, Karthik Pandia, Nauman Dwawalatabad, Rogier Landman, Jitendra Sharma, Mriganka Sur, Hema A. Murthy

Subject-Independent Classification of Japanese Spoken Sentences by Multiple Frequency Bands Phase Pattern of EEG Response During Speech Perception

Hiroki Watanabe, Hiroki Tanaka, Sakuriemi Sakti, Satoshi Nakamura

The Phonological Status of the French Initial Accent and its Role in Semantic Processing: An Event-Related Potentials Study

Noemie Te Rietmeijer, Radowne El Yogoubi, Alain Ghis, Corine Astesano

A Neuro-Experimental Evidence for the Motor Theory of Speech Perception

Bin Zhao, Jianwu Dong, Guoyan Zhang

WED-O-7-2: NOISE ROBUST SPEECH RECOGNITION

Speech Representation Learning Using Unsupervised Data-Driven Modulation Filtering for Robust ASR

Purvi Agrawal, Sridram Ganapathy

Combined Multi-Channel NMF-Based Robust Beamforming for Noisy Speech Recognition

Masato Mimura, Yoshiaki Bando, Kazuki Shimada, Shinuake Sakai, Kazuyoshi Yoshii, Tatsumi Kawahara

Recognizing Multi-Talker Speech with Permutation Invariant Training

Dong Yu, Xuankai Chang, Yannix Qian
Coupled Initialization of Multi-Channel Non-Negative Matrix Factorization Based on Spatial and Spectral Information .......................................................... 2461
Yuski Tachioda, Tomohira Narita, Iori Mura, Takahiro Uramoto, Natsumi Monta, Shingo Uenojaru, Keiichi Paruyt, Shingo Watanabe, Jonathan Le Roux

Channel Compensation in the Generalised Vector Taylor Series Approach to Robust ASR .......................................................... 2462
Erfan Loseime, Jon Barker, Thomas Hain

Robust Speech Recognition via Anchor Word Representations .......................................................... 2471
Brian King, J-Fun Chen, Tonantage Vizitman, Tizhang Liu, Roland Maus, Sree Hari Krishnan Parthasarathi, Bjorn Hoffmeister

WED-O-7-4: TOPIC SPOTTING, ENTITY EXTRACTION AND SEMANTIC ANALYSIS

Towards Zero-Shot Frame Semantic Parsing for Domain Scaling .......................................................... 2476
Ankar Bajana, Mohan Tur, Dilek Hakkani-Tur, Larry Heck

ClockWork-RNN Based Architectures for Slot Filling ......................................................................... 2481
Despoina Georgiadou, Vassilios Diakoloukou, Vassilios Tsararas, Vassilios Digalaklis

Investigating the Effect of ASR Tuning on Named Entity Recognition .......................................................... 2486
Mohamed Amer Ben Jannet, Olivier Galibert, Martinne Addas-Decker, Sophie Bosset

Label-Dependency Coding in Simple Recurrent Networks for Spoken Language Understanding ......................... 2491
Marco Dinarelli, Vedran Kukolic, Christian Raymond

Minimum Semantic Error Cost Training of Deep Long Short-Term Memory Networks for Topic Spotting on Conversational Speech ......................................................................... 2496
Zhong Meng, Biqing-Huang Jiang

Topic Identification for Speech Without ASR .......................................................................................... 2501
Chunxi Liu, Jun Trmal, Matthew Wiener, Craig Harman, Sanjeev Khudanpur

WED-O-7-6: DIALOG SYSTEMS

An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog .......................................................................................... 2506
Bing Liu, Ian Lane

Deep Reinforcement Learning of Dialogue Policies with Less Weight Updates .......................................................... 2511
Herberto Cunha-Silva, Senghak Yu

Towards End-to-End Spoken Dialogue Systems with Turn Embeddings .......................................................... 2516
Ali Orkan Bayer, Evgeny A. Stepanov, Giuseppe Riccardi

Speech and Text Analysis for Multimodal Addressee Detection in Human-Human-Computer Interaction .......................................................................................... 2521
Oleg Akhtiamov, Maxim Sidorenko, Alxsey A. Karpov, Wolfgang Minker

Rushing to Judgement: How Do Laypeople Rate Caller Engagement in Thin-Slice Videos of Human–Machine Dialog? .......................................................................................... 2526
Ivan Kraljevski, Diane Hirschfeld

Hyperarticulation of Corrections in Multilingual Dialogue Systems .......................................................................................... 2531
Antonie Bruguer, Danushen Gnanpragasam, Leif Johnson, Kantishka Rao, Francoise Beaufays

WED-O-7-8: LEXICAL AND PRONUNCIATION MODELING

Multitask Sequence-to-Sequence Models for Grapheme-to-Phoneme Conversion .......................................................... 2536
Benjamin Milde, Christoph Schmidt, Joachim Kohler

Acoustic Data-Driven Lexicon Learning Based on a Greedy Pronunciation Selection Framework .......................................................... 2541
Xiaohui Zhang, Vimla Manohar, Daniel Povey, Sanjeev Khudanpur

Semi-Supervised Learning of a Pronunciation Dictionary from Disjoint Phonemic Transcripts and Text .......................................................................................... 2546
Takahiro Shinozaki, Shingji Watanabe, Daichi Mochihashi, Graham Neubig

Improved Subword Modeling for WEST-Based Speech Recognition .......................................................................................... 2551
Peter Smit, Sami Virpioja, Mikko Kurimo

Pronunciation Learning with RNN-Transducers .......................................................................................... 2556
Antoine Bruguer, Danushen Gnanpragasam, Leif Johnson, Kantishka Rao, Francoise Beaufays

Learning Similarity Functions for Pronunciation Variations .......................................................................................... 2561
Einat Naaman, Yoysi Adi, Joseph Kesht

WED-O-7-10: LANGUAGE RECOGNITION

Spoken Language Identification Using LSTM-Based Angular Proximity .......................................................... 2566
G. Gelly, J. L. Gauchat

End-to-End Language Identification Using High-Order Utterance Representation with Bilinear Pooling .......................................................... 2571
Ma Jin, Yan Song, Ian McLoughlin, Wu Guo, Li-Rong Wang

Dialect Recognition Based on Unsupervised Bottleneck Features .......................................................................................... 2576
Qian Zhang, John H. L. Hansen

Investigating Scalability in Hierarchical Language Identification System .......................................................................................... 2581
Santhi Prasad, Vidhyasagar Sethu, Eluchamy Ambikairajah, Hachi Li

Improving Sub-Phone Modeling for Better Native Language Identification with Non-Native English Speech .......................................................................................... 2586
Yao Qian, Keelan Esmari, Xinshao Wang, David Suenndermann-Oeft, Robert A. Pugh, Patrick L. Lange, Hillary R. Mollony, Frank K. Soong
QMDIS: QCRI-MIT Advanced Dialect Identification System ....................................................... 2591
Sameer Khurana, Maryam Najafian, Ahmed Ali, Taka Al Hanai, Yonatan Belinkov, James Glass

WED-O-8-1: SPEAKER DATABASE AND ANTI-SPOOFING

Detection of Replay Attacks Using Single Frequency Filtering Cepstral Coefficients ............................................................ 2596
K. N. R. K. Raju, Aliiuri, Sivamani Achanta, Sadasivam Reddy Kadiri, Suryakantam V. Gangashetty, Anil Kumar Pappula

Unsupervised Representation Learning Using Convolutional Restricted Boltzmann Machine for Spoof Speech Detection ............................................................ 2601
Hari B. Swap, Madhu R. Kamble, Hemant A. Patil

Independent Modelling of High and Low Energy Speech Frames for Spoofing Detection ............................................................ 2606
Gajan Sathokumar, Kavya Srinivasan, Vidhyasagar Sethu, Chamith Wijenayake, Elathambu Ambikairajah

Improving Speaker Verification Performance in Presence of Spoofing Attacks Using Out-of-Domain Spoofed Data ............................................................ 2611
Achintya Kr. Sarkar, Md. Sahidullah, Zheng-Hua Tan, Tomi Kinnunen

VoxCeleb: A Large-Scale Speaker Identification Dataset ............................................................ 2616
Arsha Nagrai, Joon Son Chung, Andrew Zisserman

Call My Net Corpus: A Multilingual Corpus for Evaluation of Speaker Recognition Technology ............................................................ 2621
Karen Jones, Stephanie Strassel, Kevin Walker, David Graff, Jonathan Wright

WED-O-8-4: SPEECH TRANSLATION

Sequence-to-Sequence Models Can Directly Translate Foreign Speech ............................................................ 2625
Ron J. Weiss, Jan Chorowski, Naveen Jaitly, Yonghai Wu, Zhifeng Chen

Structured-Based Curriculum Learning for End-to-End English-Japanese Speech Translation ............................................................ 2630
Takasi Aono, Kato, Sakibani Saki, Satoshi Nakamura

Assessing the Tolerance of Neural Machine Translation Systems Against Speech Recognition Errors ............................................................ 2635
Nicholas Ruiz, Mattia Antonino Di Gangi, Nicola Bertoldi, Marcello Federico

Toward Expressive Speech Translation: A Unified Sequence-to-Sequence LSTMs Approach for Translating Words and Emphasis ............................................................ 2640
Quoc Truong Do, Sakibani Saki, Satoshi Nakamura

NMT-Based Segmentation and Punctuation Insertion for Real-Time Spoken Language Translation ............................................................ 2645
Eunah Cho, Jan Niehaus, Alex Waibel

WED-O-8-6: MULTI-CHANNEL SPEECH ENHANCEMENT

Tight Integration of Spatial and Spectral Features for BSS with Deep Clustering Embeddings ............................................................ 2650
Lukas Drude, Reinhold Haeb-Umbach

Speaker-Aware Neural Network Based Beamformer for Speaker Extraction in Speech Mixtures ............................................................ 2655
Katerina Zmolkova, Marc Deroce, Keiichi Kinozaka, Takuya Higuchi, Asumnori Ogawa, Tomohiro Nakatani

Eigenvector-Based Speech Mask Estimation Using Logistic Regression ............................................................ 2660
Lukas Pfeifchenberger, Matthias Zohrer, Franz Perdneapolis

Real-Time Speech Enhancement with GCC-NMF ............................................................ 2665
Sean C. N. Wood, Jean Rouat

Coherence-Based Dual-Channel Noise Reduction Algorithm in a Complex Noisy Environment ............................................................ 2670
Youna J, Jun Byun, Young-Cheol Park

Glottal Model Based Speech Beamforming for Ad-Hoc Microphone Arrays ............................................................ 2675
Yang Zhang, Dina Fiorencio, Mark Hasegawa-Johnson

VOLUME 5

WED-O-8-8: SPEECH RECOGNITION: APPLICATIONS IN MEDICAL PRACTICE

Acoustic Assessment of Disordered Voice with Continuous Speech Based on Utterance-Level ASR Posterior Features ............................................................ 2680
Yuanyuan Liu, Tao Lee, P. C. Ching, Thomas K. T. Law, Kathy Y. S. Lee

Multi-Stage DNN Training for Automatic Recognition of Dysarthric Speech ............................................................ 2685
Emre Yilmaz, Mario Ganchevym, Catia Cucchiarelli, Helmer Strik

Improving Child Speech Disorder Assessment by Incorporating Out-of-Domain Adult Speech ............................................................ 2690
Daniel Smith, Alex Sneddon, Lauren Ward, Andreas Dansner, Jill Freyne, David Silvera-Tawil, Angela Morgan

On Improving Acoustic Models for TORG Dysarthric Speech Database ............................................................ 2695
Neetish Maram Joy, S. Umesh, Basil Abraham

Glottal Source Features for Automatic Speech-Based Depression Assessment ............................................................ 2700
Olympia Simantiraki, Paulos Choronytakis, Anastasia Pampouchidou, Manolis Tsoknakis, Martin Cooke

Speech Processing Approach for Diagnosing Dementia in an Early Stage ............................................................ 2705
Roozbeh Sadeghian, J. David Schaffer, Stephen A. Zoharian
WED-D-8-10: LANGUAGE MODELS FOR ASR

Effectively Building Tera Scale MaxEnt Language Models Incorporating Non-Linguistic Signals .......................................................... 2710
Fadi Budiay, Mohamadreza Ghodsi, Dimitri Antonov

Semi-Supervised Adaptation of RNNLMs by Fine-Tuning with Domain-Specific Auxiliary Features .......................................................... 2715
Salil Deena, Raymond W. M. Ng, Pranava Mudigastha, Lucia Specia, Thomas Hain

Approximated and Domain-Adapted LSTM Language Models for First-Pass Decoding in Speech Recognition ........................................ 2720
Mital Singh, Yousaf Ouali, Dietrich Klakow

Sparse Non-Negative Matrix Language Modeling: Maximum Entropy Flexibility on the Cheap .......................................................... 2725
Ciprian Chelba, Dimitri Antonov, Fadi Budiay

Multi-Scale Context Adaptation for Improving Child Automatic Speech Recognition in Child-Adult Spoken Interactions .................................. 2730
Mangol Kumar, Daniel Bone, Kelly Mc-Williams, Shanna Williams, Thomas D. Lyon, Shrishanct S. Narayanan

Using Knowledge Graph and Search Query Click Logs in Statistical Language Model for Speech Recognition ........................................ 2735
Weiwei Zhu

WED-D-6-1: SPEECH RECOGNITION: TECHNOLOGIES FOR NEW APPLICATIONS AND PARADIGMS

Developing On-Line Speaker Diarization System ........................................................................................................... 2739
Dimitrios Aronditadis, Peer Fousek

Comparison of Non-Parametric Bayesian Mixture Models for Syllable Clustering and Zero-Resource Speech Processing ......................... 2744
Shreyas Sehadrir, Ulpa Remes, Okko Rassinen

Automatic Evaluation of Children Reading Aloud on Sentences and Pseudowords................................................................. 2749
Jorge Proenca, Carla Lopes, Michael Tjalve, Andreas Stolcke, Sara Candeias, Fernando Perdigao

Off-Topic Spoken Response Detection with Word Embeddings .................................................................................................... 2754
Su-Tsun Tsoo, Chong Min Lee, Ikyu Choi, Xinhao Wang, Matthew Multifrond, Keelan Evenini

Improving Mispriunciation Detection for Non-Native Learners with Multisource Information and LSTM-Based Deep Models .................. 2759
Wei Li, Nancy F. Chen, Sabato Marco Sintistan, Chin-Hui Lee

Automatic Explanation Spot Estimation Method Targeted at Text and Figures in Lecture Slides ......................................................... 2764
Shoko Tsujimura, Kazumasa Yamamoto, Seichi Nakagawa

Multiview Representation Learning via Deep CCA for Silent Speech Recognition ........................................................................... 2769
Myungseong Kim, Beiming Cao, Ted Mai, Jun Wang

Use of Graphemic Lexicons for Spoken Language Assessment ...................................................................................................... 2774

Distilling Knowledge From an Ensemble of Models for Punctuation Prediction ................................................................................... 2779
Jiangyan Yi, Jianhua Tao, Zhengyi Wen, Ya Li

A Mostly Data-Driven Approach to Inverse Text Normalization ...................................................................................................... 2784
Ernest Pusateri, Bharat Ram Ambati, Elizabeth Brooks, Ondrej Platek, Donald McAllister, Venki Nagesha

Mismatched Crowdsourcing from Multiple Annotator Languages for Recognizing Zero-Resource Languages: A Nullspace Clustering Approach .................................................................................................................. 2789
Wenda Chen, Mark Hasegawa-Johnson, Nancy F. Chen, B.ong Poong Lim

Experiments in Character-Level Neural Network Models for Punctuation .......................................................................................... 2794
William Gale, Sarangarajan Parthasarathy

Multi-Channel Apollo Mission Speech Transcripts Calibration ................................................................................................... 2799
Lakshmin Naushik, Abhijeet Sangwan, John H. L. Hansen

WED-D-6-2: SPEAKER AND LANGUAGE RECOGNITION APPLICATIONS

Calibration Approaches for Language Detection .......................................................................................................................... 2804
Mitchell McLaren, Luciana Ferrer, Diego Castan, Aaron Lawson

Bidirectional Modelling for Short Duration Language Identification .................................................................................................. 2809
Sarita Fernando, Vidhyanaswaran Sethu, Elaishanthy Ambakurajah, Julien Epps

Conditional Generative Adversarial Nets Classifier for Spoken Language Identification ................................................................. 2814
Peng Shen, Xugang Lu, Sheng Li, Hisashi Kawai

Tied Hidden Factors in Neural Networks for End-to-End Speaker Recognition .................................................................................. 2819
Antonio Miguel, Jorge Llombart, Alfonso Ortega, Eduardo Lleida

Speaker Clustering by Iteratively Finding Discriminative Feature Space and Cluster Labels .............................................................. 2824
Sunghack Yun, Hye-Jin Kang, Taeja Kim

Domain Adaptation of PLDA Models in Broadcast Diarization by Means of Unsupervised Speaker Clustering .................................................. 2829
Ignacio Vinals, Alfonso Ortega, Jesus Villalba, Antonio Miguel, Eduardo Lleida

LSTM Neural Network-Based Speaker Segmentation Using Acoustic and Language Modelling .......................................................... 2834
Miquel Indias, Jose A. R. Fonollensa, Javier Hernandez

Acoustic Pairing of Original and Dubbed Voices in the Context of Video Game Localization ................................................................. 2839
Adrien Gresse, Michael Rovner, Richard Dufour, Vincent Labatut, Jean-Francois Bonastre

Homogeneity Measure Impact on Target and Non-Target Trials in Forensic Voice Comparison ........................................................... 2844
Moez Afii, Jean-Francois Bonastre, Waaid Ben Khaled, Solange Rossato, Juliette Kuhn
Null-Hypothesis LLR: A Proposal for Forensic Automatic Speaker Recognition ........................................................................................................... 2849
The Opensesame NIST 2016 Speaker Recognition Evaluation System .......................................................................................................................... 2854
Gang Liu, Qi Qian, Zhilin Wang, Qingyan Zhao, Tianzou Wang, Hao Li, Jian Xue, Shenghao Zhu, Rong Jin, Tuo Zhao
IITG-Indigo System for NIST 2016 SRE Challenge ............................................................................................................................................... 2859
Nagendra Kumar, Rohan Kumar Das, Sarfaraz Jelil, Dhamosh B. K., H. Kashyap, K. Sri Rama Murty, Sreeram Ganapathy, Rohit Sinha, S. R. Mahadeva Prasanna
Locally Weighted Linear Discriminant Analysis for Robust Speaker Verification ........................................................................................................... 2864
Abhinav Misra, Shivesh Banjan, John H. L. Hansen
Recursive Whitening Transformation for Speaker Recognition on Language Mismatched Condition ................................................................................ 2869
Sawon Shon, Seongkyu Mun, Hanseok Ko

WED-P-6-3: SPOKEN DOCUMENT PROCESSING

Query-by-Example Search with Discriminative Neural Acoustic Word Embeddings .......................................................................................... 2874
Shane Settle, Keith Levin, Herman Kamper, Karen Livescu
Constructing Acoustic Distances Between Subwords and States Obtained from a Deep Neural Network for Spoken Term Detection .................. 2879
Daizuke Kaneko, Ryuta Konno, Kizunori Kojima, Kazuya Tanaka, Shi-Wook Lee, Yoshiaki Itoh
Fast and Accurate OOV Decoder on High-Level Features ..................................................................................................................... 2884
Yuri Khokhlov, Natalia Tomashenko, Ivan Medennikov, Aleksei Romanenko
Exploring the Use of Significant Words Language Modeling for Spoken Document Retrieval ................................................................ 2889
Ying-Wen Chen, Kuan-Yu Chen, Hsin-Min Wang, Berlin Chen
Incorporating Acoustic Features for Spontaneous Speech Driven Content Retrieval ..................................................................................... 2894
Hiroto Tasaki, Tomoyoshi Akih
Order-Preserving Abstractive Summarization for Spoken Content Based on Connectionist Temporal Classification .......................... 2899
Bo-Bu Lu, Frank Shyu, Yun-Nung Chen, Hung-Yi Lee, Lin-Shan Lee
Automatic Alignment Between Classroom Lecture Utterances and Slide Components ..................................................................................... 2904
Masatoshii Tsuchiya, Ryo Minamiguchi
Compensating Gender Variability in Query-by-Example Search on Speech Using Voice Conversion ................................................................. 2909
Paula Lopez-Otero, Laura Docio-Fernandez, Carmen Garcia-Mateo
Zero-Shot Learning Across Heterogeneous Overlapping Domains ........................................................................................................ 2914
Anjishna Kumar, Pavankumar Reddy Muddireddy, Markus Dreyer, Bjørn Hoffmeister
Hierarchical Recurrent Neural Network for Story Segmentation ........................................................................................................ 2919
Emiri Tsunao, Peter Bell, Steve Renals
Evaluating Automatic Topic Segmentation as a Segment Retrieval Task ................................................................................................. 2924
Abdessalam Boucheik, Delphine Charlet, Geraldine Dammati, Nathalie Camelin, Yannick Esteve
Improving Speech Recognizers by Refining Broadcast Data with Inaccurate Subtitle Timestamps ................................................................. 2929
Jeong-Uk Bang, Mu-Teol Cho, Sang-Hun Kim, Oh-Wook Kwon
A Relevance Score Estimation for Spoken Term Detection Based on RNN-Generated Pronunciation Embeddings .................................. 2934
Jan Svec, Josef V. Psutka, Lubos Smidl, Jan Trmal

WED-P-6-4: SPEECH INTELLIGIBILITY

Predicting Automatic Speech Recognition Performance Over Communication Channels from Instrumental Speech Quality and Intelligibility Scores .............................................................................................................................. 2939
Laura Fernandez Gallardo, Sebastian Moller, John Beerends
Speech Intelligibility in Cars: The Effect of Speaking Style, Noise and Listener Age ..................................................................................... 2944
Cassia Valenti Botinhao, Junichi Yamagishi
Predicting Speech Intelligibility Using a Gammachirp Envelope Distortion Index Based on the Signal-to-Distortion Ratio ............................................................ 2949
Katsuho Yamamoto, Toshio Primo, Toshibe Matsumi, Shoko Araki, Kenuke Kinoshita, Tomohiro Nakatani
Intelligibilities of Mandarin Chinese Sentences with Spectral "Holes" ......................................................................................................................... 2954
Yafan Chen, Yong Xu, Jun Yang
The Effect of Situation-Specific Non-Speech Acoustic Cues on the Intelligibility of Speech in Noise ................................................................. 2958
Lauern Ward, Ben Shirley, Yan Tang, William J. Davies
On the Use of Band Importance Weighting in the Short-Time Objective Intelligibility Measure ................................................................. 2963
Asger Heidemann Andersen, Jan Mark De Huan, Zheng-Hua Tan, Jesper Jensen
Listening in the Dips: Comparing Relevant Features for Speech Recognition in Humans and Machines ................................................................. 2968
Constantin Spille, Bernd T. Meyer

WED-P-7-2: ARTICULATORY AND ACOUSTIC PHONETICS

Mental Representation of Japanese Mora; Focusing on its Intrinsic Duration ........................................................................................ 2973
Kosuke Sugai
Temporal Dynamics of Lateral Channel Formation in /l/: 3D EMA Data from Australian English ................................................................. 2978
Jia Ying, Christopher Carignan, Jason A. Shaw, Michael Proctor, Donald Derrick, Catherine T. Best
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel and Consonant Sequences in Three Bavarian Dialects of Austria</td>
<td>2983</td>
</tr>
<tr>
<td>Acoustic Cues to the Singleton-Geminate Contrast: The Case of Libyan</td>
<td>2988</td>
</tr>
<tr>
<td>Arabic Sonorants</td>
<td></td>
</tr>
<tr>
<td>Mel-cepstral Distortion of German Vowels in Different Information</td>
<td>2993</td>
</tr>
<tr>
<td>Density Contexts</td>
<td></td>
</tr>
<tr>
<td>Effect of Formant and F0 Discontinuity on Perceived Vowel Duration:</td>
<td>2998</td>
</tr>
<tr>
<td>Impacts for Concatenative Speech Synthesis</td>
<td></td>
</tr>
<tr>
<td>An Ultrasound Study of Alveolar and Retroflex Consonants in Arrentne:</td>
<td>3003</td>
</tr>
<tr>
<td>Stressed and Unstressed Syllables</td>
<td></td>
</tr>
<tr>
<td>Reshaping the Transformed LF Model: Generating the Glottal Source</td>
<td>3008</td>
</tr>
<tr>
<td>from the Waveshape Parameter R_a</td>
<td></td>
</tr>
<tr>
<td>Kinematic Signatures of Prosody in Lombard Speech</td>
<td>3013</td>
</tr>
<tr>
<td>What do Finnish and Central Bavarian Have in Common? Towards an</td>
<td>3018</td>
</tr>
<tr>
<td>Acoustically Based Quantity Typology</td>
<td></td>
</tr>
<tr>
<td>Locating Burst Onsets Using SFF Envelope and Phase Information</td>
<td>3023</td>
</tr>
<tr>
<td>A Preliminary Phonetic Investigation of Alphabetic Words in Mandarin</td>
<td>3028</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
</tr>
<tr>
<td>A Quantitative Measure of the Impact of Coarticulation on Phone</td>
<td>3033</td>
</tr>
<tr>
<td>Discriminability</td>
<td></td>
</tr>
<tr>
<td>Sinusoidal Partials Tracking for Singing Analysis Using the</td>
<td>3038</td>
</tr>
<tr>
<td>Heuristic of the Minimal Frequency and Magnitude Difference</td>
<td></td>
</tr>
<tr>
<td>Audio Scene Classification with Deep Recurrent Neural Networks</td>
<td>3043</td>
</tr>
<tr>
<td>Automatic Time-Frequency Analysis of Echolocation Signals Using the</td>
<td>3048</td>
</tr>
<tr>
<td>Matched Gaussian Multitaper Spectrogram</td>
<td></td>
</tr>
<tr>
<td>Classification-Based Detection of Glottal Closure Instants from Speech</td>
<td>3053</td>
</tr>
<tr>
<td>Signals</td>
<td></td>
</tr>
<tr>
<td>A Domain Knowledge-Assisted Nonlinear Model for Head-Related Transfer</td>
<td>3058</td>
</tr>
<tr>
<td>Functions Based on Bottleneck Deep Neural Network</td>
<td></td>
</tr>
<tr>
<td>Laryngeal Articulation During Trumpet Performance: An Exploratory</td>
<td>3063</td>
</tr>
<tr>
<td>Study</td>
<td></td>
</tr>
<tr>
<td>Matrix of Polynomials Model Based Polynomial Dictionary Learning</td>
<td>3068</td>
</tr>
<tr>
<td>Method for Acoustic Impulse Response Modeling</td>
<td></td>
</tr>
<tr>
<td>Acoustic Scene Classification Using a CNN-SuperVector System Trained</td>
<td>3073</td>
</tr>
<tr>
<td>with Auditory and Spectrogram Features</td>
<td></td>
</tr>
<tr>
<td>An Environmental Feature Representation for Robust Speech Recognition</td>
<td>3078</td>
</tr>
<tr>
<td>and for Environment Identification</td>
<td></td>
</tr>
<tr>
<td>Attention and Localization Based on a Deep Convolutional Recurrent</td>
<td>3083</td>
</tr>
<tr>
<td>Model for Weakly Supervised Audio</td>
<td></td>
</tr>
<tr>
<td>An Audio Based Piano Performance Evaluation Method Using Deep Neural</td>
<td>3088</td>
</tr>
<tr>
<td>Network Based Acoustic Modeling</td>
<td></td>
</tr>
<tr>
<td>Music Tempo Estimation Using Sub-Band Synchrony</td>
<td>3093</td>
</tr>
<tr>
<td>A Transfer Learning Based Feature Extractor for Polyphonic Sound</td>
<td>3097</td>
</tr>
<tr>
<td>Event Detection Using Connectionist Temporal Classification</td>
<td></td>
</tr>
<tr>
<td>A Note Based Query By Humming System Using Convolutional Neural</td>
<td>3102</td>
</tr>
<tr>
<td>Network</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Filterbank Learning Using Convolutional Restricted</td>
<td>3107</td>
</tr>
<tr>
<td>Boltzmann Machine for Environmental Sound Classification</td>
<td></td>
</tr>
<tr>
<td>Novel Shifted Real Spectrum for Exact Signal Reconstruction</td>
<td>3112</td>
</tr>
</tbody>
</table>

WED-P-7-3: MUSIC AND AUDIO PROCESSING

Meet H. Soni, Rishabh Tak, Hemant A. Patil

Nizaba Montafa, Pascale Fang

Hardik B. Sairar, Dharmesh M. Agrawal, Hemant A. Patil

Xue Feng, Brigitte Richardson, Scott Amman, James Glass

Xiaoke Qi, Jianhua Tao

Jian Guan, Xuan Wang, Pengming Feng, Jing Dong, Wenwu Wang

Yong Xu, Qiuqiang Kong, Qiang Huang, Wenwu Wang, Mark D. Plumbley

Jing Pan, Ming Li, Zhanmei Song, Xin Li, Xiaolin Liu, Hua Yi, Mannan Zhu

Yan Wang, Florian Mette

Hardik B. Sairar, Dharmesh M. Agrawal, Hemant A. Patil
**WED-P-7-4: DISORDERS RELATED TO SPEECH AND LANGUAGE**

**Manual and Automatic Transcriptions in Dementia Detection from Speech**
Jochen Weiner, Mathis Engelbart, Tanja Schulte .......................................................... 3117

**An Affect Prediction Approach Through Depression Severity Parameter Incorporation in Neural Networks**
Rahul Gupta, Saurabh Sahu, Carol Espy-Wilson, Shrikanth S. Narayanan .......................... 3122

**Cross-Database Models for the Classification of Dysarthria Presence**
Stephanie Gillespie, Tash-Yee Logan, Elliot Moore, Jacqueline Laurens-Gore, Scott Russell, Bupal Patel .......................... 3127

**Acoustic Evaluation of Nasality in Cerebellar Syndromes**
M. Novotny, Jan Buc, K. Spalenka, Jiri Klempir, D. Horakova, Esvan Ruzicka .......................... 3132

**Emotional Speech of Mentally and Physically Disabled Individuals: Introducing the EmotASs Database and First Findings**
Simone Hancke, Hesam Sagha, Nicholas Cummins, Bjorn Schuller .................................. 3137

**Phonological Markers of Oxytocin and MDMA Ingestion**
Carla Agurto, Raquel Norel, Rachel Ostrund, Gillinder Bedi, Harriet De Wit, Matthew J. Baggott, Matthew G. Kirkpatrick, Margaret Wardle, Guillermo A. Cecchi .......................... 3142

**An Avatar-Based System for Identifying Individuals Likely to Develop Dementia**
Bahman Morheideri, Daniel Blackburn, Kirsty Harkness, Traci Walker, Anulena Venneri, Markus Reuber, Heidi Christensen .................. 3147

**Cross-Domain Classification of Drowsiness in Speech: The Case of Alcohol Intoxication and Sleep Deprivation**
Yue Zhang, Felix Weninger, Bjorn Schuller ........................................................................ 3152

**Depression Detection Using Automatic Transcriptions of De-Identified Speech**
Paula Lopez-Otero, Laura Docio-Fernandez, Alberto Abad, Carmen Garcia-Mateo .................. 3157

**An N-Gram Based Approach to the Automatic Diagnosis of Alzheimer's Disease from Spoken Language**
Sebastian Wankell, Elmar Noth, Stefan Dvert ..................................................................... 3162

**Exploiting Intra-Annotator Rating Consistency Through Copeland's Method for Estimation of Ground Truth Labels in Couples' Therapy**
Karel Mundrich, Md. Nasir, Panayiotis Georgiou, Shrikanth S. Narayanan ......................... 3167

**Rhythmic Characteristics of Parkinsonian Speech: A Study on Mandarin and Polish**
Massimo Pettorino, Wentao Gu, Pawel Polrola, Ping Fan ................................................. 3172

**WED-P-8-1: PROSODY**

**Trisyllabic Tone 3 Sandhi Patterns in Mandarin Produced by Cantonese Speakers**
Jung-Tsueh Tu, Janice Wing-Sze Wong, Jih-Ho Cha .......................................................... 3177

**Intonation of Contrastive Topic in Estonian**
Hetsy Sakhai, Meelis Mihkla ............................................................................................. 3181

**Reanalyze Fundamental Frequency Peak Delay in Mandarin**
Lixia Hao, Wei Zhang, Yantie Xie, Jinsong Zhang ................................................................ 3186

**How Does the Absence of Shared Knowledge Between Interlocutors Affect the Production of French Prosodic Forms?**
Amandine Michelas, Cecile Cau, Maud Champagne-Lavaux ............................................... 3191

**Three Dimensions of Sentence Prosody and Their (Non-)Interactions**
Michael Wagner, Michael McAskill .................................................................................. 3196

**Using Prosody to Classify Discourse Relations**
Janine Kleinhaus, Mireia Farrus, Agustín Graviano, Juan Manuel Perez, Catherine Lai, Leo Wanner .......................................................... 3201

**Canonical Correlation Analysis and Prediction of Perceived Rhythmic Prominences and Pitch Tones in Speech**
Elizabeth Godoy, James R. Williamson, Thomas F. Quartieri ............................................ 3206

**Evaluation of Spectral Tilt Measures for Sentence Prominence Under Different Noise Conditions**
Sofoklis Kakouros, Okko Rasanen, Pauo Alka .................................................................... 3211

**Creaky Voice as a Function of Tonal Categories and Prosodic Boundaries**
Jianjing Kuang .................................................................................................................... 3216

**The Acoustics of Word Stress in Czech as a Function of Speaking Style**
Barelk Skarntiel, Anders Eriksson ..................................................................................... 3221

**What You See is What You Get Prosodically Less --- Visibility Shapes Prosodic Prominence Production in Spontaneous Interaction**
Petra Wagner, Nataliya Bryhady .......................................................... 3226

**Focus Acoustics in Mandarin Nominals**
Yu-Yin Hsu, Anqi Xu ......................................................................................................... 3231

**Exploring Multidimensionality: Acoustic and Articulatory Correlates of Swedish Word Accents**
Malin Svensson Lundmark, Gilbert Ambruzzitis, Otto Ewald ............................................ 3236

**The Perception of English Intonation Patterns by German L2 Speakers of English**
Karin Pagi, Robert Fuchs, Jane Setter, Peggy Mok .................................................................. 3241

**WED-P-8-2: SPEAKER STATES AND TRAITS**

**The Perception of Emotions in Noisified Nonsense Speech**
Emilia Parada-Cabaleiro, Alice Baird, Anton Baltiner, Nicholas Cummins, Simone Hancke, Bjorn Schuller .......................................................... 3246

**Attention Networks for Modeling Behaviors in Addiction Counseling**
James Gibson, Dogan Can, Panayiotis Georgiou, David C. Atkins, Shrikanth S. Narayanan .......................................................... 3251
Computational Analysis of Acoustic Descriptors in Psychotic Patients .......................................................... 3256
Torsten Wortwein, Tadas Bulaitisaitis, Eugene Lukasana, Luciana Pennant, Elizabeth S. Liebson, Dori Ongar, Justin T. Baker, Louis-Philippe Morin

Modeling Perceivers Neural-Responses Using Lobe-Dependent Convolutional Neural Network to Improve Speech

Emotion Recognition

Ya-Tse Wu, Hsuan-Yu Chen, Yu-Wei Liao, Li-Wen Kuo, Chi-Chun Lee .......................................................... 3261

Implementing Gender-Dependent Vowel-Level Analysis for Boosting Speech-based Depression Recognition .......................................................... 3266
Bogdan Vasenko, Hexam Sagha, Nicholas Cummins, Bjorn Schaller

Bilingual Word Embeddings for Cross-Linguial Personality Recognition Using Convolutional Neural Nets .......................................................... 3271
Farhad Bin Siddiquie, Pascale Fang

Emotion Category Mapping to Emotional Space by Cross-Corpus Emotion Labeling .......................................................... 3276
Yoshiko Arimoto, Hiroki Mori

Big Five vs. Prosodic Features as Cues to Detect Abnormality in SSPNET-Personality Corpus .......................................................... 3281
Cedric Fayet, Arnaud Delhuy, Damien Lolive, Pierre-Francois Marieau

Speech Rate Comparison When Talking to a System and Talking to a Human: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task .............................................................................................................. 3286
Hayakawa Akira, Carl Vogel, Saravino Lu, Nick Campbell

Approaching Human Performance in Behavior Estimation in Couples Therapy Using Deep Sentence Embeddings .............................................................................................................. 3291
Shao-Yen Tseng, Brian Baucom, Panayiotis Georgiou

Complexity in Speech and its Relation to Emotional Bond in Therapist-Patient Interactions During Suicide Risk Assessment Interviews .............................................................................................................. 3296
Md. Nasir, Brian Baucom, Craig J. Bryan, Shrikanth S. Narayanan, Panayiotis Georgiou

An Investigation of Emotion Dynamics and Kalman Filtering for Speech-based Emotion Prediction .............................................................................................................. 3301
Zhaoceng Huang, Julien Epps

VOLUME 6

Speaker Dependency Analysis, Audiovisual Fusion Cues and a Multimodal BLSTM for Conversational Engagement Recognition .............................................................................................................. 3359
Yuyan Huang, Enno Gilmarion, Nick Campbell

WED-P-8-4: VOICE CONVERSION 2

Voice Conversion from Unaligned Corpora Using Variational Autoencoding Wasserstein Generative Adversarial Networks .............................................................................................................. 3364
Chin-Cheng Hu, Heim-Te Hwang, Yi-Chiao Wu, Yu Tsao, Hsin-Min Wang

CAB: An Energy-Based Speaker Clustering Model for Rapid Adaptation in Non-Parallel Voice Conversion .............................................................................................................. 3369
Toru Nakashika

Phoneme-Discriminative Features for Dysarthric Speech Conversion .............................................................................................................. 3374
Ryo Ishara, Tetsuya Takeguchi, Yasso Ariti
Denosing Recurrent Neural Network for Deep Bidirectional LSTM Based Voice Conversion .............................................................. 3379
Speaker Dependent Approach for Enhancing a Glossectomy Patient's Speech via GMM-Based Voice Conversion .......................... 3384
Kei Tanaka, Sania Hara, Masanobu Abe, Masashi Sato, Shogo Minagi
Generative Adversarial Network-Based Prefilter for STFT Spectrograms ................................................................................ 3389
Takahiro Kaneko, Shinnji Takaki, Hirokazu Kameoka, Junichi Yamagishi
Generative Adversarial Network-Based Glottal Waveform Model for Statistical Parametric Speech Synthesis ............................. 3394
Bajjibolu Bollapalli, Lauri Juvela, Paavo Alku
Emotional Voice Conversion with Adaptive Scales F0 Based on Wavelet Transform Using Limited Amount of Emotional Data ............................................................................................................. 3399
Zhaojie Luo, Jinhui Chen, Tetsuya Takeda, Yasuo Ariki
Speaker Adaptation in DNN-Based Speech Synthesis Using d-Vectors ......................................................................................... 3404
Rama Dodipatula, Norbert Braunshcweiger, Rannier Maia
Spectro-Temporal Modelling with Time-Frequency LSTM and Structured Output Layer for Voice Conversion .......................... 3409
Runnan Li, Zhiyong Wu, Tishuang Ning, Liya Sun, Helen Meng, Lianhong Cai
Segment Level Voice Conversion with Recurrent Neural Networks .............................................................................................. 3414
Miguel Varela Ramos, Alan W. Black, Ramon Fernandez Ansadillo, Isabel Trancoso, Nuno Fontecca

WED-S&T-6/7-A: SHOW & TELL 5

Creating a Voice for MiRo, the World's First Commercial Biomimetic Robot ............................................................................... 3419
Roger K. Moore, Ben Mitchinson
A Thematicity-Based Prosody Enrichment Tool for CTS ....................................................................................................................... 3421
Monica Dominguez, Mireia Farrus, Leo Wanner
WebSubDub — Experimental System for Creating High-Quality Alternative Audio Track for TV Broadcasting ................................. 3423
Martin Gruber, Jindrich Matousek, Zdenek Hanzlcek, Jekub Vitz, Daniel Tihelka
Voice Conservation and TTS System for People Facing Total Laryngectomy ........................................................................ 3425
Marketa Javou, Daniel Tihelka, Jindrich Matousek, Zdenek Hanzlcek
TBT (Toolkit to Build TTS): A High Performance Framework to Build Multiple Language HTS Voice ...................................................... 3427
Atish Shankar Ghone, Rachana Nerpagar, Pranav Kumar, Arun Bhat, Awin Shunmugam, M. Sasikumar, Hema A. Murthy
SIAK — A Game for Foreign Language Pronunciation Learning ........................................................................................................ 3429
Reima Karhila, Sari Ylenen, Seppo Enarvi, Kalle Palomaki, Aleksander Nikulin, Olli Rantula, Vertti Viitanen, Krupakar Dinakaran, Anna-Riikka Smolander, Heini Kallio, Katja Junttila, Maria Uther, Perttu Hamalainen, Mikko Karinno

WED-S&T-6/7-B: SHOW & TELL 6

Integrating the Talkamatic Dialogue Manager with Alexa .................................................................................................................... 3431
Stefan Larsson, Alex Berman, Andreas Krona, Fredrik Kronlid
A Robust Medical Speech-to-Speech/Speech-to-Sign Phraselator ....................................................................................................... 3433
Farhia Ahmed, Pierreute Bousillon, Chelle Destefano, Johanna Gerlach, Sonia Halimi, Angela Hooper, Manny Raymer, Herve Spechbach, Irene Strain, Nikos Tzourakis
Towards an Autarkic Embedded Cognitive User Interface .................................................................................................................. 3435
Frank Duckhorn, Markus Huber, Werner Meyer, Oliver Jokisch, Constance Tshope, Matthias Wolff
Nora the Empathetic Psychologist ......................................................................................................................................................... 3437
Genta Indra Winata, Onno Kampman, Yung Yang, Anik Dey, Pascale Fung
Modifying Amazon’s Alexa ASR Grammar and Lexicon — A Case Study ........................................................................................ 3439
Hashan Alam, Aman Kumar, Manan Vyas, Tina Werner, Rachmat Hartono

KEYNOTE 3: BJORN LINDBLOM

Re-Inventing Speech — The Biological Way ........................................................................................................................ 3441
Bjorn Lindblom

THU-SS-9-10: SPECIAL SESSION: INTERSPEECH 2017 COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE) 1

The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addresslee, Cold & Snorring .............................................................. 3442
Bjorn Schall, Stefan Steinl, Anton Battiner, Erika Bergelson, Jarek Krajewski, Christoph Janott, Andrei Amatuni, Martin Casillas, Amanda Steinl, Melanie Soderstrom, Anne S. Warram, Guillermo Hidalgo, Sebastian Schneider, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Maximilian Schmitt, Kun Qian, Yue Zhang, George Trigeorgis, Panagistas Tzirakis, Stefanos Zaferiou
It Sounds Like You Have a Cold! Testing Voice Features for the Interspeech 2017 Computational Paralinguistics Cold Challenge .................................................................................................................................................. 3447
Mark Hackvalde, Andreas Beke
End-to-End Deep Learning Framework for Speech Paralinguistics Detection Based on Perception Aware Spectrum ............................................................................................................................................. 3452
Danwei Cai, Zhidong Ni, Wenbo Liu, Weicheng Cai, Gang Li, Ming Li
Infected Phonemes: How a Cold Impairs Speech on a Phonetic Level .............................................................................................. 3457
Johannes Wagner, Thiago Fraga-Silva, Ivan Josses, Dominik Schiller, Andreas Seiderer, Elisabeth Andre
THU-SS-9-11: SPECIAL SESSION: STATE OF THE ART IN PHYSICS-BASED VOICE SIMULATION

Acoustic Analysis of Detailed Three-Dimensional Shape of the Human Nasal Cavity and Paranasal Sinuses ......................................................
Tatsuya Kitamura, Hirokore Takemoto, Hidemori Makinae, Tetsuhiro Yamaguchi, Kotoo Makii ..........................3472

A Semi-Polar Grid Strategy for the Three-Dimensional Finite Element Simulation of Vowel-Vowel Sequences .........................................................
Arvind Vasudevan, Victor Zappi, Peter Anderson, Sidney Fels ........................3477

A Fast Robust 1D Flow Model for a Self-Oscillating Coupled 2D FEM Vocal Fold Simulation ....................
Marc Arnela, Saeed Dabbaghchian, Oriol Guasch, Olav Engwall .............................................................3482

Waveform Patterns in Pitch Glides Near a Vocal Tract Resonance ..........................................................
Tiina Murtola, Jarmo Malinen ..................................................................................................................3487

A Unified Numerical Simulation of Vowel Production That Compares Phonation and the Emitted Sound ...........................
Niayzi Cem Degermenci, Johan Jansson, Johan Hoffman, Marc Arnela, Patricia Sanchez-Martin, Oriol Guasch, Sten Ternstrom ..........................................................3492

Synthesis of VV Utterances from Muscle Activation to Sound with a 3D Model ........................................
Saeed Dabbaghchian, Marc Arnela, Olav Engwall, Oriol Guasch ................................................................3497

THU-SS-10-10: SPECIAL SESSION: INTERSPEECH 2017 COMPUTATIONAL PARALINGUISTICS

CHALLENGE (COMPARE) 2

A Dual Source-Filter Model of Snore Audio for Snorer Group Classification ........................................
Achuth Rao M. V., Shivani Yadav, Prasanta Kumar Ghosh .......................................................................3502

An 'End-to-Evolution' Hybrid Approach for Snore Sound Classification ........................................
Michael Freitag, Shahin Amiriparian, Nicholas Cummins, Maurice Gerckzuk, Bjorn Schuller .....................3507

Snore Sound Classification Using Image-Based Deep Spectrum Features ................................................
Shahin Amiriparian, Maurice Gerckzuk, Sandra Oht, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, Bjorn Schuller .............................3512

Exploring Fusion Methods and Feature Space for the Classification of Paralinguistic Information ................
David Tavares, Xabier Sarasola, Agustin Alonso, Jon Sanchez, Luis Serrano, Eva Navas, Inma Hernaez ...............3517

DNN-Based Feature Extraction and Classifier Combination for Child-Directed Speech, Cold and Snoring
Identification .................................................................................................................................3522
Gabor Gosztolya, Robert Bass-Fekete, Tomasz Grosz, Laszlo Toth

Introducing Weighted Kernel Classifiers for Handling Imbalanced Paralinguistic Corpora: Snoring, Addresssee and Cold .................................................................3527
Heysem Kaya, Alexey A. Karpov

THU-O-1: DISCRIMINATIVE TRAINING FOR ASR

Multitask Learning with Low-Level Auxiliary Tasks for Encoder-Decoder Based Speech Recognition ........................3532
Shubham Toshniwal, Hao Tang, Liang Lu, Kaven Livescu

Optimizing Expected Word Error Rate via Sampling for Speech Recognition ........................................3537
Matt Shannon

Annealed f-Smoothing as a Mechanism to Speed up Neural Network Training ........................................
Tara N. Sainath, Vijayanidya Peddinti, Olivier Sisanchor, Arun Narayanan ................................................3542

Non-Uniform MCE Training of Deep Long Short-Term Memory Recurrent Neural Networks for Keyword
Spotting ................................................................................................................................................3547
Zhong Meng, Biing-Hwang Juang

Exploiting Eigenposteriors for Semi-Supervised Training of DNN Acoustic Models with Sequence Discrimination ........3552
Pranay Dighi, Aftabneh Aasai, Herve Bourlard

Discriminative Autoencoders for Acoustic Modeling ...............................................................................3557

THU-O-2: SPEAKER DIARIZATION

Speaker Diarization Using Convolutional Neural Network for Statistics Accumulation Refinement ........3562
Zbynek Zajic, Marek Hruz, Ludek Muller

Speaker2Vec: Unsupervised Learning and Adaptation of a Speaker Manifold Using Deep Neural Networks with an Evaluation on Speaker Segmentation ...............................................................3567
Arindam Jati, Panayiotis Georgiou

A Triplet Ranking-Based Neural Network for Speaker Diarization and Linking ........................................3572
Gael Le Lant, Delphine Charlet, Anthony Larcher, Sylvain Meignier

Estimating Speaker Clustering Quality Using Logistic Regression ..........................................................3577
Yishai Cohen, Izhak Lapidoth
Combining Speaker Turn Embedding and Incremental Structure Prediction for Low-Latency Speaker Diarization .......................................................... 3582
Guillaume Wiesniewski, Herve Bredin, G. Gelly, Claude Barra

Family Pyannote.Metrics: A Toolkit for Reproducible Evaluation, Diagnostic, and Error Analysis of Speaker Diarization Systems .................................................................................................................. 3587
Herve Bredin

THU-O-9-4: SPOKEN TERM DETECTION

A Rescoring Approach for Keyword Search Using Lattice Context Information .............................................................................................................. 3592
Zhipeng Chen, Ji Wu

The Kaldi OpenKWS System: Improving Low Resource Keyword Search ............................................................................................................ 3597
Jan Trmal, Matthew Wiener, Vijayaditya Peddinti, Xiaohui Zhang, Pegah Ghahremani, Timing Wang, Vimal Manohar, Hainan Yu, Daniel Povey, Sanjeev Khudanpur

The STC Keyword Search System for OpenKWS 2016 Evaluation ................................................................................................................... 3602
Yuri Khokhlov, Ivan Medennikov, Aleksei Romanenko, Valentin Mendelev, Maxim Korenevsky, Alexey Prudnikov, Natalia Tomashenko, Alexander Zatvornitsky

Compressed Time Delay Neural Network for Small-Footprint Keyword Spotting ................................................................................................. 3607
Ming Sun, David Snyder, Yicen Guo, Yarun Nagaraja, Mike Rodehorst, Sankaran Panchapagesan, Nikko Strom, Spyros Matsoukas, Shiv Vithaladevan

Symbol Sequence Search from Telephone Conversation ......................................................................................................................... 3612
Masayuki Suzuki, Gakuto Kurata, Abhinav Sethy, Bhuvana Ramabhadran, Kenneth W. Church, Mark Drake

Similarity Learning Based Query Modeling for Keyword Search ........................................................................................................... 3617
Batuhan Gundogdu, Murat Saracilar

THU-O-9-6: NOISE REDUCTION

Deep Recurrent Neural Network Based Monaural Speech Separation Using Recurrent Temporal Restricted Boltzmann Machines .............................................................................................................. 3622
Tamaan Sami, Iraj Jalili Chakrabarti, Soumya K. Ghosh

Improved Codebook-Based Speech Enhancement Based on MBE Model ........................................................................................................... 3627
Qizheng Huang, Changchun Bao, Xiyuan Wang

Improving Mask Learning Based Speech Enhancement System with Restoration Layers and Residual Connection ......................................................... 3632
Zhao Chen, Yan Huang, Jinyu Li, Yifan Gong

Exploring Low-Dimensional Structures of Modulation Spectra for Robust Speech Recognition ......................................................................................... 3637
Bi-Cheng Tan, Chin-Hong Shih, Shih-Hsiang Liu, Berlin Chen

SEGAN: Speech Enhancement Generative Adversarial Network .......................................................................................................................... 3642
Santiago Pascual, Antonio Bonafonte, Juan Serra

Concatenative Resynthesis Using Twin Networks ........................................................................................................................................ 3647
Soumi Maiti, Michael I. Mandel

THU-O-9-8: SPEECH RECOGNITION: MULTIMODAL SYSTEMS

Combining Residual Networks with LSTMs for Lipreading .......................................................................................................................... 3652
Themos Staflidakis, Georgios Tzimiropoulos

Improving Computer Lipreading via DNN Sequence Discriminative Training Techniques .................................................................................. 3657
Kwanchiva Thangphai, Richard Harvey

Improving Speaker-Independent Lipreading with Domain-Adversarial Training ................................................................................................. 3662
Michael Wand, Jurgen Schmidhuber

Turbo Decoders for Audio-Visual Continuous Speech Recognition ........................................................................................................... 3667
Ahmed Hassen Abdelaziz

DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface ........................................................................................................ 3672
Tamas Gabor Csapo, Tamas Grosz, Gabor Gaszolya, Laszlo Toth, Alexandra Marko

Visually Grounded Learning of Keyword Prediction from Untranscribed Speech .......................................................................................... 3677
Herman Kamber, Shane Settle, Gregory Shakhnarovich, Karen Livescu

THU-O-10-1: NEURAL NETWORK ACOUSTIC MODELS FOR ASR 3

Deep Neural Factorization for Speech Recognition ........................................................................................................................................ 3682
Jen-Tzung Chien, Chen Shen

Semi-Supervised DNN Training with Word Selection for ASR ........................................................................................................... 3687
Karel Vesely, Lukas Burger, Jan Cernocky

Gaussian Prediction Based Attention for Online End-to-End Speech Recognition .......................................................................................... 3692
Junfeng Hou, Shiliang Zhang, Li-Tong Dai

Efficient Knowledge Distillation from an Ensemble of Teachers .......................................................................................................................... 3697
Takahiro Fukuda, Masayuki Suzuki, Gakuto Kurata, Samuel Thomas, Jia Cui, Bhuvana Ramabhadran

An Analysis of "Attention" in Sequence-to-Sequence Models ..................................................................................................................... 3702
Rohit Prabhavalkar, Tara N. Sainath, Bo Li, Kanishka Rao, Navdeep Jaitly
THU-O-10-2: ROBUST SPEAKER RECOGNITION

CNN-Based Joint Mapping of Short and Long Utterance i-Vectors for Speaker Verification Using Short Utterances ................................................................. 3712
Jiaxi Guo, Usha Amrutha Nookala, Abeer Alwan

Curriculum Learning Based Probabilistic Linear Discriminant Analysis for Noise Robust Speaker Recognition ................................................................. 3717
Shiv shev Ranjan, Abhinav Misra, John H. L. Hansen

i-Vector Transformation Using a Novel Discriminative Denoising Autoencoder for Noise-Robust Speaker Recognition ............................................................. 3722
Shivangi Mahato, Hitoshi Yamamoto, Takafumi Koshinaka

Unsupervised Discriminative Training of PLDA for Domain Adaptation in Speaker Verification ................................................................. 3727
Qiongqiong Wang, Takafumi Koshinaka

Speaker Verification Under Adverse Conditions Using i-Vector Adaptation and Neural Networks ................................................................. 3732
Jahangir Alam, Patrick Kenny, Gautam Bhattacharya, Marcel Kockmann

Improving Robustness of Speaker Recognition to New Conditions Using Unlabeled Data ................................................................. 3737
Diego Cuxian, Mitchell McLaren, Luciana Ferrer, Aaron Lawson, Alicia Lozano-Diez

THU-O-10-4: MULTIMODAL RESOURCES AND ANNOTATION

CALYOU: A Comparable Spoken Algerian Corpus Harvested from YouTube................................................................. 3742
K. Abidi, M. A. Menacer, Kamel Smalii

PRAV: A Phonetically Rich Audio Visual Corpus ................................................................. 3747
Abhishek Naverekar, Prasanta Kumar Ghosh

NTCD-TIMIT: A New Database and Baseline for Noise-Robust Audio-Visual Speech Recognition ................................................................. 3752
Ahmed Hassan Abdelaziz

The Extended SPaRKy Restaurant Corpus: Designing a Corpus with Variable Information Density ................................................................. 3757
David M. Howcroft, Dietrich Klakow, Vera Demberg

Automatic Construction of the Finnish Parliament Speech Corpus ................................................................. 3762
Andre Mansikkaniemi, Peter Smil, Mikko Kurimo

Building Audio-Visual Phonetically Annotated Arabic Corpus for Expressive Text to Speech ................................................................. 3767
Omnia Abdou, Sherif Abdou, Mervat Fashal

THU-O-10-8: FORENSIC PHONETICS AND SOCIOPHONETIC VARIETIES

What is the Relevant Population? Considerations for the Computation of Likelihood Ratios in Forensic Voice Comparison ................................................................. 3772
Vincent Hughes, Paul Foulkes

Voice Disguise vs. Impersonation: Acoustic and Perceptual Measurements of Vocal Flexibility in Non Experts ................................................................. 3777
Veronique Delvaux, Lise Cauchois, Kathy Haert, Myriam Piccardou, Bernard Herveman

Schwa Realization in French: Using Automatic Speech Processing to Study Phonological and Socio-Linguistic Factors in Large Corpora ................................................................. 3782
Yaru Wu, Martine Adda-Decker, Cecile Fougereon, Lori Lamel

The Social Life of Setswana Ejectives ................................................................. 3787
Daniel Duran, Jagoda Bruni, Grzegorz Dogil, Justus Roux

How Long is Too Long? How Pause Features After Requests Affect the Perceived Willingness of Affirmative Answers ................................................................. 3792
Lea S. Kohz, Oliver Niebuhr

Shadowing Synthesized Speech - Segmental Analysis of Phonetic Convergence ................................................................. 3797
Iona Gessinger, Eran Raveh, Sebastien Le Maguer, Bernd Mobius, Ingmar Steiner

THU-O-10-11: SPEECH AND AUDIO SEGMENTATION AND CLASSIFICATION 1

Occupancy Detection in Commercial and Residential Environments Using Audio Signal ................................................................. 3802
Shabnam Ghaffarzadeh, Attila Reiss, Mirko Ruhs, Robert Duerrchen, Zhe Feng

Data Augmentation, Missing Feature Mask and Kernel Classification for Through-the-Wall Acoustic Surveillance ................................................................. 3807
Huy Dat Tran, Wen Zheng Terence Ng, Yen Long

Endpoint Detection Using Grid Long Short-Term Memory Networks for Streaming Speech Recognition ................................................................. 3812
Shuo-Yin Chang, Bo Li, Taka N. Sainath, Gabor Simko, Carolina Parad

Deep Learning Techniques in Tandem with Signal Processing Cues for Phonetic Segmentation for Text to Speech Synthesis in Indian Languages ................................................................. 3817
Arun Baby, Jeena J. Prakash, Rapak Vignesh, Hema A. Murthy

Gate Activation Signal Analysis for Gated Recurrent Neural Networks and its Correlation with Phoneme Boundaries ................................................................. 3822
Yu-Hsuan Wang, Cheng-Tao Chang, Hung-Yi Lee
Speaker Change Detection in Broadcast TV Using Bidirectional Long Short-Term Memory Networks
Ruiqing Yin, Herve Bredin, Claude Barrus

THU-P-9-1: NOISE ROBUST AND FAR-FIELD ASR

Improved Automatic Speech Recognition Using Subband Temporal Envelope Features and Time-Delay Neural Network Denoising Autoencoder
Cong-Thanh Do, Yannis Stylianou

Factored Deep Convolutional Neural Networks for Noise Robust Speech Recognition
Masakizo Fujimoto

Global SNR Estimation of Speech Signals for Unknown Noise Conditions Using Noise Adapted Non-Linear Regression

Joint Training of Multi-Channel-Condition Dereverberation and Acoustic Modeling of Microphone Array Speech for Robust Distant Speech Recognition

Genghei Ge, Kehuang Li, Bo Wu, Sabato Marco Siniscalchi, Yonghong Yan, Chin-Hui Lee

Uncertainty Decoding with Adaptive Sampling for Noise Robust DNN-Based Acoustic Modeling
Dung T. Tran, Marc Delcroix, Atsunori Ogawa, Tomohiro Nakatani

Attention-Based LSTM with Multi-Task Learning for Distant Speech Recognition
Yu Zhang, Pengwen Zhang, Yonghong Yan

To Improve the Robustness of LSTM-RNN Acoustic Models Using Higher-Order Feedback from Multiple Histories
Hengguan Huang, Brian Mak

End-to-End Speech Recognition with Auditory Attention for Multi-Microphone Distance Speech Recognition
Suyoun Kim, Jun Lane

Robust Speech Recognition Based on Binaural Auditory Processing

Adaptive Multichannel Dereverberation for Automatic Speech Recognition
Joe Carasselli, Izak Shaftrun, Arun Narayanan, Richard Rose

THU-P-9-3: SYLES, VARIETIES, FORENSICS AND TOOLS

The Effects of Real and Placebo Alcohol on Deaffrication
Urban Zihlmann

Polyglot and Speech Corpus Tools: A System for Representing, Integrating, and Querying Speech Corpora
Michael McIntilfe, Elias Stengel-Eskin, Michaela Syclof, Morgan Sonderegger

Mapping Across Feature Spaces in Forensic Voice Comparison: The Contribution of Auditory-Based Voice Quality to (Semi-)Automatic System Testing

Effect of Language, Speaking Style and Speaker on Long-Term F0 Estimation

Stability of Prosodic Characteristics Across Age and Gender Groups
Jan Volin, Tereza Tykalova, Tomas Boril

Electrophysiological Correlates of Familiar Voice Recognition
Julien Plante-Hebert, Victor J. Bouchet, Boasheima Jimel

Developing an Emboshi (Bantu C25) Speech Variant Dictionary to Model Vowel Elision and Morpheme Deletion
Jamison Cooper-Leavitt, Lari Lamel, Annie Rialland, Martine Adde-Decker, Gilles Adda

Rd as a Control Parameter to Explore Affective Correlates of the Tense-Lax Continuum
Andy Murphy, Irena Tanashinskaya, Aliibe Ni Chusaide, Christer Gobli

Cross-Linguistic Distinctions Between Professional and Non-Professional Speaking Styles
Plimo A. Barbosa, Sandra Madureira, Philippe Boila De Mareuil

Perception and Production of Word-Final /R/ in French
Cedric Gerflet

Glottal Source Estimation from Coded Telephone Speech Using a Deep Neural Network
N. P. Narendra, Manu Airaksinen, Paavo Alku

Automatic Labelling of Prosodic Prominence, Phrasing and Disfluencies in French Speech by Simulating the Perception of Naive and Expert Listeners
George Christodoulides, Mathieu Avanzi, Anne Catherine Simon

Don’t Count on ASR to Transcribe for You: Breaking Bias with Two Crowds
Michael Levit, Yan Huang, Shuangyu Chang, Yijan Gong

Effects of Training Data Variety in Generating Glottal Pulses from Acoustic Features with DNNs
Manu Airaksinen, Paavo Alku

Towards Intelligent Crowdsourcing for Audio Data Annotation: Integrating Active Learning in the Real World
Simone Hunke, Zixing Zhang, Bjorn Schuller

THU-P-9-4: SPEECH SYNTHESIS: DATA, EVALUATION, AND NOVEL PARADIGMS

Principles for Learning Controllable TTS from Annotated and Latent Variation
Gustav Eje Henter, Jaime Lorenzo-Trueba, Xin Wang, Junichi Yamagishi