

# **18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)**

Situated Interaction

Stockholm, Sweden  
20 - 24 August 2017

Volume 1 of 6

ISBN: 978-1-5108-4876-4

**Printed from e-media with permission by:**

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



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2017) by International Speech Communication Association  
All rights reserved.

Printed by Curran Associates, Inc. (2018)

For permission requests, please contact International Speech Communication Association  
at the address below.

International Speech Communication Association  
c/o Mme Emmanuelle FOXONET  
4 Rue des Fauvettes - Lous Tourils  
F-66390 Baixas, France

Phone: 49 228 735 643  
Fax: 33 468 385 827

[secretariat@isca-speech.org](mailto:secretariat@isca-speech.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## VOLUME 1

### ISCA MEDAL 2017 CEREMONY

<b>ISCA Medal for Scientific Achievement</b> .....	1
<i>Fumitada Itakura</i>	

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

<b>The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection</b> .....	2
<i>Tomi Kinnunen, Md. Sahidullah, Hector Delgado, Massimiliano Todisco, Nicholas Evans, Junichi Yamagishi, Kong Aik Lee</i>	
<b>Experimental Analysis of Features for Replay Attack Detection --- Results on the ASVspoof 2017 Challenge</b> .....	7
<i>Roberto Font, Juan M. Espin, Maria Jose Cano</i>	
<b>Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection</b> .....	12
<i>Hemant A. Patil, Madhu R. Kamble, Tanvina B. Patel, Meet H. Soni</i>	
<b>Countermeasures for Automatic Speaker Verification Replay Spoofing Attack : On Data Augmentation, Feature Representation, Classification and Fusion</b> .....	17
<i>Weicheng Cai, Danwei Cai, Wenbo Liu, Gang Li, Ming Li</i>	
<b>Spoof Detection Using Source, Instantaneous Frequency and Cepstral Features</b> .....	22
<i>Sarfaraz Jellil, Rohan Kumar Das, S. R. Mahadeva Prasanna, Rohit Sinha</i>	
<b>Audio Replay Attack Detection Using High-Frequency Features</b> .....	27
<i>Marcin Witkowski, Stanislaw Kacprzak, Piotr Zelasko, Konrad Kowalczyk, Jakub Galka</i>	
<b>Feature Selection Based on CQCCs for Automatic Speaker Verification Spoofing</b> .....	32
<i>Xianliang Wang, Yanhong Xiao, Xuan Zhu</i>	

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

<b>Longitudinal Speaker Clustering and Verification Corpus with Code-Switching Frisian-Dutch Speech</b> .....	37
<i>Emre Yilmaz, Jelske Dijkstra, Hans Van De Velde, Frederik Kampstra, Jouke Algra, Henk Van Den Heuvel, David Van Leeuwen</i>	
<b>Exploiting Untranscribed Broadcast Data for Improved Code-Switching Detection</b> .....	42
<i>Emre Yilmaz, Henk Van Den Heuvel, David Van Leeuwen</i>	
<b>Jee haan, I'd like both, por favor: Elicitation of a Code-Switched Corpus of Hindi--English and Spanish--English Human--Machine Dialog</b> .....	47
<i>Vikram Ramanarayanan, David Suendermann-Oeft</i>	
<b>On Building Mixed Lingual Speech Synthesis Systems</b> .....	52
<i>Saikrishna Rallabandi, Alan W. Black</i>	
<b>Speech Synthesis for Mixed-Language Navigation Instructions</b> .....	57
<i>Khyathi Raghavi Chandu, Sai Krishna Rallabandi, Sunayana Sitaram, Alan W. Black</i>	
<b>Addressing Code-Switching in French/Algerian Arabic Speech</b> .....	62
<i>Djegdjiiga Amazouz, Martine Adda-Decker, Lori Lamel</i>	
<b>Metrics for Modeling Code-Switching Across Corpora</b> .....	67
<i>Gualberto Guzman, Joseph Ricard, Jacqueline Serigos, Barbara E. Bullock, Almeida Jacqueline Toribio</i>	
<b>Synthesising isiZulu-English Code-Switch Bigrams Using Word Embeddings</b> .....	72
<i>Ewald Van Der Westhuizen, Thomas Niesler</i>	
<b>Crowdsourcing Universal Part-of-Speech Tags for Code-Switching</b> .....	77
<i>Victor Soto, Julia Hirschberg</i>	

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

<b>Audio Replay Attack Detection with Deep Learning Frameworks</b> .....	82
<i>Galina Lavrentyeva, Sergey Novoselov, Egor Malykh, Alexander Kozlov, Oleg Kudashev, Vadim Shchemelinin</i>	
<b>Ensemble Learning for Countermeasure of Audio Replay Spoofing Attack in ASVspoof2017</b> .....	87
<i>Zhe Ji, Zhi-Yi Li, Peng Li, Maobo An, Shengxiang Gao, Dan Wu, Faru Zhao</i>	
<b>A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification</b> .....	92
<i>Lantian Li, Yixiang Chen, Dong Wang, Thomas Fang Zheng</i>	
<b>Replay Attack Detection Using DNN for Channel Discrimination</b> .....	97
<i>Parav Nagarsheth, Elie Khoury, Kailash Patil, Matt Garland</i>	
<b>ResNet and Model Fusion for Automatic Spoofing Detection</b> .....	102
<i>Zhuxin Chen, Zhifeng Xie, Weibin Zhang, Xiangmin Xu</i>	

<b>SFF Anti-Spoofers: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2017</b> .....	107
<i>K. N. R. K. Raju Alluri, Sivanand Achanta, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, Anil Kumar Vuppala</i>	

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

<b>Improved Single System Conversational Telephone Speech Recognition with VGG Bottleneck Features</b> .....	112
<i>William Hartmann, Roger Hsiao, Tim Ng, Jeff Ma, Francis Keith, Man-Hung Siu</i>	
<b>Student-Teacher Training with Diverse Decision Tree Ensembles</b> .....	117
<i>Jeremy H. M. Wong, Mark J. F. Gales</i>	
<b>Embedding-Based Speaker Adaptive Training of Deep Neural Networks</b> .....	122
<i>Xiaodong Cui, Vaibhava Goel, George Saon</i>	
<b>Improving Deliverable Speech-to-Text Systems with Multilingual Knowledge Transfer</b> .....	127
<i>Jeff Ma, Francis Keith, Tim Ng, Man-Hung Siu, Owen Kimball</i>	
<b>English Conversational Telephone Speech Recognition by Humans and Machines</b> .....	132
<i>George Saon, Gakuto Kurata, Tom Sercu, Kartik Audhkhasi, Samuel Thomas, Dimitrios Dimitriadis, Xiaodong Cui, Bhuvana Ramabhadran, Michael Picheny, Lynn-Li Lim, Bergul Roomi, Phil Hall</i>	
<b>Comparing Human and Machine Errors in Conversational Speech Transcription</b> .....	137
<i>Andreas Stolcke, Jasha Droppo</i>	

### **MON-O-1-2: MULTIMODAL PARALINGUISTICS**

<b>Multimodal Makers of Persuasive Speech: Designing a Virtual Debate Coach</b> .....	142
<i>Volha Petukhova, Manoj Raju, Harry Bunt</i>	
<b>Acoustic-Prosodic and Physiological Response to Stressful Interactions in Children with Autism Spectrum Disorder</b> .....	147
<i>Daniel Bone, Julia Mertens, Emily Zane, Sungbok Lee, Shrikanth S. Narayanan, Ruth Grossman</i>	
<b>A Stepwise Analysis of Aggregated Crowdsourced Labels Describing Multimodal Emotional Behaviors</b> .....	152
<i>Alec Burmania, Carlos Busso</i>	
<b>An Information Theoretic Analysis of the Temporal Synchrony Between Head Gestures and Prosodic Patterns in Spontaneous Speech</b> .....	157
<i>Gaurav Fotedar, Prasanta Kumar Ghosh</i>	
<b>Multimodal Prediction of Affective Dimensions via Fusing Multiple Regression Techniques</b> .....	162
<i>D.-Y. Huang, Wan Ding, Mingyu Xu, Huaiping Ming, Minghui Dong, Xinguo Yu, Haizhou Li</i>	
<b>Co-Production of Speech and Pointing Gestures in Clear and Perturbed Interactive Tasks: Multimodal Designation Strategies</b> .....	166
<i>Marion Dohen, Benjamin Roustan</i>	

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

<b>Improving Speaker Verification for Reverberant Conditions with Deep Neural Network Dereverberation Processing</b> .....	171
<i>Peter Guzewich, Stephen A. Zahorian</i>	
<b>Stepsize Control for Acoustic Feedback Cancellation Based on the Detection of Reverberant Signal Periods and the Estimated System Distance</b> .....	176
<i>Philipp Bulling, Klaus Linhard, Arthur Wolf, Gerhard Schmidt</i>	
<b>A Delay-Flexible Stereo Acoustic Echo Cancellation for DFT-Based In-Car Communication (ICC) Systems</b> .....	181
<i>Jan Franzen, Tim Fingscheidt</i>	
<b>Speech Enhancement Based on Harmonic Estimation Combined with MMSE to Improve Speech Intelligibility for Cochlear Implant Recipients</b> .....	186
<i>Dongmei Wang, John H. L. Hansen</i>	
<b>Improving Speech Intelligibility in Binaural Hearing Aids by Estimating a Time-Frequency Mask with a Weighted Least Squares Classifier</b> .....	191
<i>David Ayllon, Roberto Gil-Pita, Manuel Rosa-Zurera</i>	
<b>Simulations of High-Frequency Vocoder on Mandarin Speech Recognition for Acoustic Hearing Preserved Cochlear Implant</b> .....	196
<i>Tsung-Chen Wu, Tai-Shih Chi, Chia-Fone Lee</i>	

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

<b>Phonetic Correlates of Pharyngeal and Pharyngealized Consonants in Saudi, Lebanese, and Jordanian Arabic: An rt-MRI Study</b> .....	201
<i>Zainab Hermes, Marissa Barlaz, Ryan Shosted, Zhi-Pei Liang, Brad Sutton</i>	
<b>Glottal Opening and Strategies of Production of Fricatives</b> .....	206
<i>Benjamin Elie, Yves Laprie</i>	
<b>Acoustics and Articulation of Medial versus Final Coronal Stop Gemination Contrasts in Moroccan Arabic</b> .....	210
<i>Mohamed Yassine Frej, Christopher Carignan, Catherine T. Best</i>	
<b>How are Four-Level Length Distinctions Produced? Evidence from Moroccan Arabic</b> .....	215
<i>Giuseppina Turco, Karim Shoul, Rachid Ridouane</i>	

<b>Vowels in the Barunga Variety of North Australian Kriol</b> .....	219
<i>Caroline Jones, Katherine Demuth, Weicong Li, Andre Almeida</i>	
<b>Nature of Contrast and Coarticulation: Evidence from Mizo Tones and Assamese Vowel Harmony</b> .....	224
<i>Indranil Dutta, S. Irfan, Pamir Gogoi, Priyankoo Sarmah</i>	

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

<b>The Influence of Synthetic Voice on the Evaluation of a Virtual Character</b> .....	229
<i>Joao Paulo Cabral, Benjamin R. Cowan, Katja Zibrek, Rachel McDonnell</i>	
<b>Articulatory Text-to-Speech Synthesis Using the Digital Waveguide Mesh Driven by a Deep Neural Network</b> .....	234
<i>Amelia J. Gully, Takenori Yoshimura, Damian T. Murphy, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda</i>	
<b>An HMM/DNN Comparison for Synchronized Text-to-Speech and Tongue Motion Synthesis</b> .....	239
<i>Sebastien Le Maguer, Ingmar Steiner, Alexander Hewer</i>	
<b>VCV Synthesis Using Task Dynamics to Animate a Factor-Based Articulatory Model</b> .....	244
<i>Rachel Alexander, Tanner Sorensen, Asterios Toutios, Shrikanth S. Narayanan</i>	
<b>Beyond the Listening Test: An Interactive Approach to TTS Evaluation</b> .....	249
<i>Joseph Mendelson, Matthew P. Aylett</i>	
<b>Integrating Articulatory Information in Deep Learning-Based Text-to-Speech Synthesis</b> .....	254
<i>Beiming Cao, Myungjong Kim, Jan Van Santen, Ted Mau, Jun Wang</i>	

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

<b>Approaches for Neural-Network Language Model Adaptation</b> .....	259
<i>Min Ma, Michael Nirschl, Fadi Biadisy, Shankar Kumar</i>	
<b>A Batch Noise Contrastive Estimation Approach for Training Large Vocabulary Language Models</b> .....	264
<i>Youssef Oualil, Dietrich Klakow</i>	
<b>Investigating Bidirectional Recurrent Neural Network Language Models for Speech Recognition</b> .....	269
<i>X. Chen, A. Ragni, X. Liu, Mark J. F. Gales</i>	
<b>Fast Neural Network Language Model Lookups at N-Gram Speeds</b> .....	274
<i>Yinghui Huang, Abhinav Sethy, Bhuvana Ramabhadran</i>	
<b>Empirical Exploration of Novel Architectures and Objectives for Language Models</b> .....	279
<i>Gakuto Kurata, Abhinav Sethy, Bhuvana Ramabhadran, George Saon</i>	
<b>Residual Memory Networks in Language Modeling: Improving the Reputation of Feed-Forward Networks</b> .....	284
<i>Karel Benes, Murali Karthick Baskar, Lukas Burget</i>	

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

<b>Dominant Distortion Classification for Pre-Processing of Vowels in Remote Biomedical Voice Analysis</b> .....	289
<i>Amir Hossein Poorjam, Jesper Rindom Jensen, Max A. Little, Mads Graesboll Christensen</i>	
<b>Automatic Paraphasia Detection from Aphasic Speech: A Preliminary Study</b> .....	294
<i>Duc Le, Keli Licata, Emily Mower Provost</i>	
<b>Evaluation of the Neurological State of People with Parkinson's Disease Using i-Vectors</b> .....	299
<i>N. Garcia, Juan Rafael Orozco-Arroyave, L. F. D'Haro, Najim Dehak, Elmar Noth</i>	
<b>Objective Severity Assessment from Disordered Voice Using Estimated Glottal Airflow</b> .....	304
<i>Yu-Ren Chien, Michal Borsky, Jon Gudnason</i>	
<b>Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach</b> .....	309
<i>Florian B. Pokorny, Bjorn Schuller, Peter B. Marschik, Raymond Brueckner, Par Nystrom, Nicholas Cummins, Sven Bolte, Christa Einspieler, Terje Falck-Ytter</i>	
<b>Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease</b> .....	314
<i>J. C. Vasquez-Correa, Juan Rafael Orozco-Arroyave, Elmar Noth</i>	

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

<b>Phone Classification Using a Non-Linear Manifold with Broad Phone Class Dependent DNNs</b> .....	319
<i>Linxue Bai, Peter Jancovic, Martin Russell, Philip Weber, Steve Houghton</i>	
<b>An Investigation of Crowd Speech for Room Occupancy Estimation</b> .....	324
<i>Siyuan Chen, Julien Epps, Eliathamby Ambikairajah, Phu Ngoc Le</i>	
<b>Time-Frequency Coherence for Periodic-Aperiodic Decomposition of Speech Signals</b> .....	329
<i>Karthika Vijayan, Jitendra Kumar Dhiman, Chandra Sekhar Seelamantula</i>	
<b>Musical Speech: A New Methodology for Transcribing Speech Prosody</b> .....	334
<i>Alexsandro R. Meireles, Antonio R. M. Simoes, Antonio Celso Ribeiro, Beatriz Raposo De Medeiros</i>	
<b>Estimation of Place of Articulation of Fricatives from Spectral Characteristics for Speech Training</b> .....	339
<i>K. S. Nataraj, Prem C. Pandey, Hirak Dasgupta</i>	
<b>Estimation of the Probability Distribution of Spectral Fine Structure in the Speech Source</b> .....	344
<i>Tom Backstrom</i>	

## **MON-O-2-6: PERCEPTION OF DIALECTS AND L2**

<b>End-to-End Acoustic Feedback in Language Learning for Correcting Devoiced French Final-Fricatives</b> .....	349
<i>Sucheta Ghosh, Camille Fauth, Yves Laprie, Aghilas Sini</i>	
<b>Dialect Perception by Older Children</b> .....	354
<i>Ewa Jacewicz, Robert A. Fox</i>	
<b>Perception of Non-Contrastive Variations in American English by Japanese Learners: Flaps are Less Favored Than Stops</b> .....	359
<i>Kiyoko Yoneyama, Mafuyu Kitahara, Keiichi Tajima</i>	
<b>L1 Perceptions of L2 Prosody: The Interplay Between Intonation, Rhythm, and Speech Rate and Their Contribution to Accentedness and Comprehensibility</b> .....	364
<i>Lieke Van Maastricht, Tim Zee, Emiel Kraemer, Marc Swerts</i>	
<b>Effects of Pitch Fall and L1 on Vowel Length Identification in L2 Japanese</b> .....	369
<i>Izumi Takiguchi</i>	
<b>A Preliminary Study of Prosodic Disambiguation by Chinese EFL Learners</b> .....	374
<i>Yuanyuan Zhang, Hongwei Ding</i>	

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

<b>Generation of Large-Scale Simulated Utterances in Virtual Rooms to Train Deep-Neural Networks for Far-Field Speech Recognition in Google Home</b> .....	379
<i>Chanwoo Kim, Ananya Misra, Kean Chin, Thad Hughes, Arun Narayanan, Tara N. Sainath, Michiel Bacchiani</i>	
<b>Neural Network-Based Spectrum Estimation for Online WPE Dereverberation</b> .....	384
<i>Keisuke Kinoshita, Marc Delcroix, Haeyong Kwon, Takuma Mori, Tomohiro Nakatani</i>	
<b>Factorial Modeling for Effective Suppression of Directional Noise</b> .....	389
<i>Osamu Ichikawa, Takashi Fukuda, Gakuto Kurata, Steven J. Rennie</i>	
<b>On Design of Robust Deep Models for CHiME-4 Multi-Channel Speech Recognition with Multiple Configurations of Array Microphones</b> .....	394
<i>Yan-Hui Tu, Jun Du, Lei Sun, Feng Ma, Chin-Hui Lee</i>	
<b>Acoustic Modeling for Google Home</b> .....	399
<i>Bo Li, Tara N. Sainath, Arun Narayanan, Joe Caroselli, Michiel Bacchiani, Ananya Misra, Izhak Shafran, Hasim Sak, Golan Pundak, Kean Chin, Khe Chai Sim, Ron J. Weiss, Kevin W. Wilson, Ehsan Varianti, Chanwoo Kim, Olivier Siohan, Mitchel Weintraub, Erik McDermott, Richard Rose, Matt Shannon</i>	
<b>On Multi-Domain Training and Adaptation of End-to-End RNN Acoustic Models for Distant Speech Recognition</b> .....	404
<i>Seyedmahdad Mirsamadi, John H. L. Hansen</i>	

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

<b>Low-Dimensional Representation of Spectral Envelope Without Deterioration for Full-Band Speech Analysis/Synthesis System</b> .....	409
<i>Masanori Morise, Genta Miyashita, Kenji Ozawa</i>	
<b>Robust Source-Filter Separation of Speech Signal in the Phase Domain</b> .....	414
<i>Erfan Loweimi, Jon Barker, Oscar Saz Torralba, Thomas Hain</i>	
<b>A Time-Warping Pitch Tracking Algorithm Considering Fast <math>F_0</math> Changes</b> .....	419
<i>Simon Stone, Peter Steiner, Peter Birkholz</i>	
<b>A Modulation Property of Time-Frequency Derivatives of Filtered Phase and its Application to Aperiodicity and <math>f_0</math> Estimation</b> .....	424
<i>Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, Tomoki Toda</i>	
<b>Non-Local Estimation of Speech Signal for Vowel Onset Point Detection in Varied Environments</b> .....	429
<i>Avinash Kumar, S. Shahmawazuddin, Gayadhar Pradhan</i>	
<b>Time-Domain Envelope Modulating the Noise Component of Excitation in a Continuous Residual-Based Vocoder for Statistical Parametric Speech Synthesis</b> .....	434
<i>Mohammed Salah Al-Radhi, Tamas Gabor Csapo, Geza Nemeth</i>	
<b>Wavelet Speech Enhancement Based on Robust Principal Component Analysis</b> .....	439
<i>Chia-Lung Wu, Hsiang-Ping Hsu, Syu-Siang Wang, Jieh-Weih Hung, Ying-Hui Lai, Hsin-Min Wang, Yu Tsao</i>	
<b>Vowel Onset Point Detection Using Sonority Information</b> .....	444
<i>Bidisha Sharma, S. R. Mahadeva Prasanna</i>	
<b>Analytic Filter Bank for Speech Analysis, Feature Extraction and Perceptual Studies</b> .....	449
<i>Unto K. Laine</i>	
<b>Learning the Mapping Function from Voltage Amplitudes to Sensor Positions in 3D-EMA Using Deep Neural Networks</b> .....	454
<i>Christian Kroos, Mark D. Plumbley</i>	

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

<b>Multilingual i-Vector Based Statistical Modeling for Music Genre Classification</b> .....	459
<i>Jia Dai, Wei Xue, Wenju Liu</i>	
<b>Indoor/Outdoor Audio Classification Using Foreground Speech Segmentation</b> .....	464
<i>Banriskhem K. Khonglah, K. T. Deepak, S. R. Mahadeva Prasanna</i>	

<b>Attention Based CLDNNs for Short-Duration Acoustic Scene Classification</b> .....	469
<i>Jinxi Guo, Ning Xu, Li-Jia Li, Abeer Alwan</i>	
<b>Frame-Wise Dynamic Threshold Based Polyphonic Acoustic Event Detection</b> .....	474
<i>Xianjun Xia, Roberto Togneri, Ferdous Sohel, David Huang</i>	
<b>Enhanced Feature Extraction for Speech Detection in Media Audio</b> .....	479
<i>Inseon Jang, Chunghyun Ahn, Jeongil Seo, Younseon Jang</i>	
<b>Audio Classification Using Class-Specific Learned Descriptors</b> .....	484
<i>Sukanya Sonowal, Tushar Sandhan, Inkyu Choi, Nam Soo Kim</i>	
<b>Hidden Markov Model Variational Autoencoder for Acoustic Unit Discovery</b> .....	488
<i>Janek Ebberts, Jahn Heymann, Lukas Drude, Thomas Glarner, Reinhold Haeb-Umbach, Bhiksha Raj</i>	
<b>Virtual Adversarial Training and Data Augmentation for Acoustic Event Detection with Gated Recurrent Neural Networks</b> .....	493
<i>Matthias Zohrer, Franz Pernkopf</i>	
<b>Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi</b> .....	498
<i>Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, Morgan Sonderegger</i>	
<b>A Robust Voiced/Unvoiced Phoneme Classification from Whispered Speech Using the `Color' of Whispered Phonemes and Deep Neural Network</b> .....	503
<i>G. Nisha Meenakshi, Prasanta Kumar Ghosh</i>	

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

<b>Rescoring-Aware Beam Search for Reduced Search Errors in Contextual Automatic Speech Recognition</b> .....	508
<i>Ian Williams, Petar Aleksic</i>	
<b>Comparison of Decoding Strategies for CTC Acoustic Models</b> .....	513
<i>Thomas Zenkel, Ramon Sanabria, Florian Metzke, Jan Niehues, Matthias Sperber, Sebastian Stuker, Alex Waibel</i>	
<b>Phone Duration Modeling for LVCSR Using Neural Networks</b> .....	518
<i>Hossein Hadian, Daniel Povey, Hossein Sameti, Sanjeev Khudanpur</i>	
<b>Towards Better Decoding and Language Model Integration in Sequence to Sequence Models</b> .....	523
<i>Jan Chorowski, Navdeep Jaitly</i>	
<b>Empirical Evaluation of Parallel Training Algorithms on Acoustic Modeling</b> .....	528
<i>Wenpeng Li, Binbin Zhang, Lei Xie, Dong Yu</i>	
<b>Binary Deep Neural Networks for Speech Recognition</b> .....	533
<i>Xu Xiang, Yanmin Qian, Kai Yu</i>	
<b>Hierarchical Constrained Bayesian Optimization for Feature, Acoustic Model and Decoder Parameter Optimization</b> .....	538
<i>Akshay Chandrashekar, Ian Lane</i>	
<b>Use of Global and Acoustic Features Associated with Contextual Factors to Adapt Language Models for Spontaneous Speech Recognition</b> .....	543
<i>Shohei Toyama, Daisuke Saito, Nobuaki Minematsu</i>	
<b>Joint Learning of Correlated Sequence Labeling Tasks Using Bidirectional Recurrent Neural Networks</b> .....	548
<i>Vardaan Pahuja, Anirban Laha, Shachar Mirkin, Vikas Raykar, Lili Kotlerman, Guy Lev</i>	
<b>Estimation of Gap Between Current Language Models and Human Performance</b> .....	553
<i>Xiaoyu Shen, Youssef Oualil, Clayton Greenberg, Mittul Singh, Dietrich Klakow</i>	
<b>A Phonological Phrase Sequence Modelling Approach for Resource Efficient and Robust Real-Time Punctuation Recovery</b> .....	558
<i>Anna Moro, Gyorgy Szaszak</i>	

#### **MON-P-2-1: SPEECH PERCEPTION**

<b>Factors Affecting the Intelligibility of Low-Pass Filtered Speech</b> .....	563
<i>Lei Wang, Fei Chen</i>	
<b>Phonetic Restoration of Temporally Reversed Speech</b> .....	567
<i>Shi-Yu Wang, Fei Chen</i>	
<b>Simultaneous Articulatory and Acoustic Distortion in L1 and L2 Listening: Locally Time-Reversed "Fast" Speech</b> .....	571
<i>Mako Ishida</i>	
<b>Lexically Guided Perceptual Learning in Mandarin Chinese</b> .....	576
<i>L. Ann Burchfield, San-Hei Kenny Luk, Mark Antoniou, Anne Cutler</i>	
<b>The Effect of Spectral Profile on the Intelligibility of Emotional Speech in Noise</b> .....	581
<i>Chris Davis, Chee Seng Chong, Jeeseun Kim</i>	
<b>Whether Long-term Tracking of Speech Rate Affects Perception Depends on Who is Talking</b> .....	586
<i>Merel Maslowski, Antje S. Meyer, Hand Rutger Bosker</i>	
<b>Emotional Thin-Slicing: A Proposal for a Short- and Long-Term Division of Emotional Speech</b> .....	591
<i>Daniel Oliveira Peres, Dominic Watt, Waldemar Ferreira Netto</i>	
<b>Predicting Epenthetic Vowel Quality from Acoustics</b> .....	596
<i>Adriana Guevara-Rukoz, Erika Parlato-Oliveira, Shi Yu, Yuki Hirose, Sharon Peperkamp, Emmanuel Dupoux</i>	
<b>The Effect of Spectral Tilt on Size Discrimination of Voiced Speech Sounds</b> .....	601
<i>Toshie Matsui, Toshio Irino, Kodai Yamamoto, Hideki Kawahara, Roy D. Patterson</i>	
<b>Misperceptions of the Emotional Content of Natural and Vocoded Speech in a Car</b> .....	606
<i>Jaime Lorenzo-Trueba, Cassia Valentini Botinhao, Gustav Eje Henter, Junichi Yamagishi</i>	

<b>The Relative Cueing Power of F0 and Duration in German Prominence Perception</b> .....	611
<i>Oliver Niebuhr, Jana Winkler</i>	
<b>Perception and Acoustics of Vowel Nasality in Brazilian Portuguese</b> .....	616
<i>Luciana Marques, Rebecca Scarborough</i>	
<b>Sociophonetic Realizations Guide Subsequent Lexical Access</b> .....	621
<i>Jonny Kim, Katie Drager</i>	

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

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

## **VOLUME 2**

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

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

<b>Multilingual Recurrent Neural Networks with Residual Learning for Low-Resource Speech Recognition</b> .....	704
<i>Shiyu Zhou, Yuanyuan Zhao, Shuang Xu, Bo Xu</i>	
<b>CTC Training of Multi-Phone Acoustic Models for Speech Recognition</b> .....	709
<i>Olivier Siohan</i>	
<b>An Investigation of Deep Neural Networks for Multilingual Speech Recognition Training and Adaptation</b> .....	714
<i>Sibo Tong, Philip N. Garner, Herve Bourlard</i>	
<b>2016 BUT Babel System: Multilingual BLSTM Acoustic Model with i-Vector Based Adaptation</b> .....	719
<i>Martin Karafiat, Murali Karthick Baskar, Pavel Matejka, Karel Vesely, Frantisek Grezl, Lukas Burget, Jan Cernocky</i>	
<b>Optimizing DNN Adaptation for Recognition of Enhanced Speech</b> .....	724
<i>Marco Matassoni, Alessio Brutti, Daniele Falavigna</i>	
<b>Deep Least Squares Regression for Speaker Adaptation</b> .....	729
<i>Younggwan Kim, Hyungjun Lim, Jahyun Goo, Hoirin Kim</i>	
<b>Multi-Task Learning Using Mismatched Transcription for Under-Resourced Speech Recognition</b> .....	734
<i>Van Hai Do, Nancy F. Chen, Boon Pang Lim, Mark Hasegawa-Johnson</i>	
<b>Generalized Distillation Framework for Speaker Normalization</b> .....	739
<i>Neethu Mariam Joy, Sandeep Reddy Kothinti, S. Umesh, Basil Abraham</i>	
<b>Learning Factorized Transforms for Unsupervised Adaptation of LSTM-RNN Acoustic Models</b> .....	744
<i>Lahiru Samarakoon, Brian Mak, Khe Chai Sim</i>	



<b>Factorised Representations for Neural Network Adaptation to Diverse Acoustic Environments</b> .....	749
<i>Joachim Fainberg, Steve Renals, Peter Bell</i>	

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

<b>An RNN Model of Text Normalization</b> .....	754
<i>Richard Sproat, Navdeep Jaitly</i>	
<b>Weakly-Supervised Phrase Assignment from Text in a Speech-Synthesis System Using Noisy Labels</b> .....	759
<i>Asaf Rendel, Raul Fernandez, Zvi Kons, Andrew Rosenberg, Ron Hoory, Bhuvana Ramabhadran</i>	
<b>Prosody Aware Word-Level Encoder Based on BLSTM-RNNs for DNN-Based Speech Synthesis</b> .....	764
<i>Yusuke Ijima, Nobukatsu Hojo, Ryo Masumura, Taichi Asami</i>	
<b>Global Syllable Vectors for Building TTS Front-End with Deep Learning</b> .....	769
<i>Jinfu Ni, Yoshinori Shiga, Hisashi Kawai</i>	
<b>Prosody Control of Utterance Sequence for Information Delivering</b> .....	774
<i>Ishin Fukuoka, Kazuhiko Iwata, Tetsunori Kobayashi</i>	
<b>Multi-Task Learning for Prosodic Structure Generation Using BLSTM RNN with Structured Output Layer</b> .....	779
<i>Yuchen Huang, Zhiyong Wu, Runnan Li, Helen Meng, Lianhong Cai</i>	
<b>Investigating Efficient Feature Representation Methods and Training Objective for BLSTM-Based Phone Duration Prediction</b> .....	784
<i>Yibin Zheng, Jianhua Tao, Zhengqi Wen, Ya Li, Bin Liu</i>	
<b>Discrete Duration Model for Speech Synthesis</b> .....	789
<i>Bo Chen, Tianling Bian, Kai Yu</i>	
<b>Comparison of Modeling Target in LSTM-RNN Duration Model</b> .....	794
<i>Bo Chen, Jiahao Lai, Kai Yu</i>	
<b>Learning Word Vector Representations Based on Acoustic Counts</b> .....	799
<i>M. Sam Ribeiro, Oliver Watts, Junichi Yamagishi</i>	
<b>Synthesising Uncertainty: The Interplay of Vocal Effort and Hesitation Disfluencies</b> .....	804
<i>Eva Szekeley, Joseph Mendelson, Joakim Gustafson</i>	

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

<b>Prosograph: A Tool for Prosody Visualisation of Large Speech Corpora</b> .....	809
<i>Alp Oktem, Mireia Farrus, Leo Wanner</i>	
<b>ChunkitApp: Investigating the Relevant Units of Online Speech Processing</b> .....	811
<i>Svetlana Vetchimikova, Anna Mauranen, Nina Mikusova</i>	
<b>Extending the EMU Speech Database Management System: Cloud Hosting, Team Collaboration, Automatic Revision Control</b> .....	813
<i>Markus Jochim</i>	
<b>HomeBank: A Repository for Long-Form Real-World Audio Recordings of Children</b> .....	815
<i>Anne S. Warlaumont, Mark Vandam, Erika Bergelson, Alejandrina Cristia</i>	
<b>A System for Real Time Collaborative Transcription Correction</b> .....	817
<i>Peter Bell, Joachim Fainberg, Catherine Lai, Mark Sinclair</i>	
<b>MoPAREST --- Mobile Phone Assisted Remote Speech Therapy Platform</b> .....	819
<i>Chitralkha Bhat, Anjali Kant, Bhavik Vachhani, Sarita Rautara, Ashok Kumar Sinha, Sunil Kumar Koppurapu</i>	

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

<b>An Apparatus to Investigate Western Opera Singing Skill Learning Using Performance and Result Biofeedback, and Measuring its Neural Correlates</b> .....	821
<i>Aurore Jaumard-Hakoun, Samy Chikhi, Takfarinas Medani, Angelika Nair, Gerard Dreyfus, Francois-Benoit Vialatte</i>	
<b>PercyConfigurator --- Perception Experiments as a Service</b> .....	823
<i>Christoph Draxler</i>	
<b>System for Speech Transcription and Post-Editing in Microsoft Word</b> .....	825
<i>Askars Salimbajevs, Indra Ikauniece</i>	
<b>Emojiive! Collecting Emotion Data from Speech and Facial Expression Using Mobile Game App</b> .....	827
<i>Ji Ho Park, Nayeon Lee, Dario Bertero, Anik Dey, Pascale Fung</i>	
<b>Myly --- The Mill: A New Platform for Processing Speech and Text Corpora Easily and Efficiently</b> .....	829
<i>Mietta Lennes, Jussi Piitulainen, Martin Matthiesen</i>	
<b>Visual Learning 2: Pronunciation App Using Ultrasound, Video, and MRI</b> .....	831
<i>Kyori Suzuki, Ian Wilson, Hayato Watanabe</i>	

### **KEYNOTE 1: JAMES ALLEN**

<b>Dialogue as Collaborative Problem Solving</b> .....	833
<i>James Allen</i>	

## **TUE-SS-3-11: SPECIAL SESSION: SPEECH AND HUMAN-ROBOT INTERACTION**

<b>Elicitation Design for Acoustic Depression Classification: An Investigation of Articulation Effort, Linguistic Complexity, and Word Affect</b> .....	834
<i>Brian Stasak, Julien Epps, Roland Goecke</i>	
<b>Robustness Over Time-Varying Channels in DNN-HMM ASR Based Human-Robot Interaction</b> .....	839
<i>Jose Novoa, Jorge Wuth, Juan Pablo Escudero, Josue Fredes, Rodrigo Mahu, Richard M. Stern, Nestor Becerra Yoma</i>	
<b>Analysis of Engagement and User Experience with a Laughter Responsive Social Robot</b> .....	844
<i>Bekir Berker Turker, Zana Bucinca, Engin Erzin, Yucel Yemez, Metin Sezgin</i>	
<b>Automatic Classification of Autistic Child Vocalisations: A Novel Database and Results</b> .....	849
<i>Alice Baird, Shahin Amiriparian, Nicholas Cummins, Alyssa M. Alcorn, Anton Batliner, Sergey Pugachevskiy, Michael Freitag, Maurice Gerczuk, Bjorn Schuller</i>	
<b>Crowd-Sourced Design of Artificial Attentive Listeners</b> .....	854
<i>Catharine Oertel, Patrik Jonell, Dimosthenis Kontogiorgos, Joseph Mendelson, Jonas Beskow, Joakim Gustafson</i>	
<b>Studying the Link Between Inter-Speaker Coordination and Speech Imitation Through Human-Machine Interactions</b> .....	859
<i>Leonardo Lancia, Thierry Chaminade, Noel Nguyen, Laurent Prevot</i>	

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

<b>Adjusting the Frame: Biphasic Performative Control of Speech Rhythm</b> .....	864
<i>Samuel Delalez, Christophe D'Alessandro</i>	
<b>Attentional Factors in Listeners' Uptake of Gesture Cues During Speech Processing</b> .....	869
<i>Raheleh Saryazdi, Craig G. Chambers</i>	
<b>Motion Analysis in Vocalized Surprise Expressions</b> .....	874
<i>Carlos Ishi, Takashi Minato, Hiroshi Ishiguro</i>	
<b>Enhancing Backchannel Prediction Using Word Embeddings</b> .....	879
<i>Robin Ruede, Markus Muller, Sebastian Stuker, Alex Waibel</i>	
<b>A Computational Model for Phonetically Responsive Spoken Dialogue Systems</b> .....	884
<i>Eran Raveh, Ingmar Steiner, Bernd Mobius</i>	
<b>Incremental Dialogue Act Recognition: Token- vs Chunk-Based Classification</b> .....	889
<i>Eustace Ebbotemhen, Volha Petukhova, Dietrich Klakow</i>	

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

<b>Clear Speech --- Mere Speech? How Segmental and Prosodic Speech Reduction Shape the Impression That Speakers Create on Listeners</b> .....	894
<i>Oliver Niebuhr</i>	
<b>Relationships Between Speech Timing and Perceived Hostility in a French Corpus of Political Debates</b> .....	899
<i>Charlotte Kouklia, Nicolas Audibert</i>	
<b>Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution</b> .....	904
<i>Laura Fernandez Gallardo, Benjamin Weiss</i>	
<b>Prosodic Analysis of Attention-Drawing Speech</b> .....	909
<i>Carlos Ishi, Jun Arai, Norihiro Hagita</i>	
<b>Perceptual and Acoustic Correlates of Gender in the Perpubertal Voice</b> .....	914
<i>Adrian P. Simpson, Riccarda Funk, Frederik Palmer</i>	
<b>To See or Not to See: Interlocutor Visibility and Likeability Influence Convergence in Intonation</b> .....	919
<i>Katrin Schweitzer, Michael Walsh, Antje Schweitzer</i>	
<b>Acoustic Correlates of Parental Role and Gender Identity in the Speech of Expecting Parents</b> .....	924
<i>Melanie Weirich, Adrian P. Simpson</i>	
<b>A Semi-Supervised Learning Approach for Acoustic-Prosodic Personality Perception in Under-Resourced Domains</b> .....	929
<i>Ruben Solera-Urena, Helena Moniz, Fernando Batista, Vera Cabarrao, Anna Pompili, Ramon Fernandez Astudillo, Joana Campos, Ana Paiva, Isabel Trancoso</i>	
<b>Effects of Talker Dialect, Gender &amp; Race on Accuracy of Bing Speech and YouTube Automatic Captions</b> .....	934
<i>Rachael Tatman, Conner Kasten</i>	

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

<b>A Comparison of Sequence-to-Sequence Models for Speech Recognition</b> .....	939
<i>Rohit Prabhavalkar, Kanishka Rao, Tara N. Sainath, Bo Li, Leif Johnson, Navdeep Jaitly</i>	
<b>CTC in the Context of Generalized Full-Sum HMM Training</b> .....	944
<i>Albert Zeyer, Eugen Beck, Ralf Schluter, Hermann Ney</i>	
<b>Advances in Joint CTC-Attention Based End-to-End Speech Recognition with a Deep CNN Encoder and RNN-LM</b> .....	949
<i>Takaaki Hori, Shinji Watanabe, Yu Zhang, William Chan</i>	
<b>Multitask Learning with CTC and Segmental CRF for Speech Recognition</b> .....	954
<i>Liang Lu, Lingpeng Kong, Chris Dyer, Noah A. Smith</i>	
<b>Direct Acoustics-to-Word Models for English Conversational Speech Recognition</b> .....	959
<i>Kartik Audhkhasi, Bhuvana Ramabhadran, George Saon, Michael Picheny, David Nahamoo</i>	

<b>Reducing the Computational Complexity of Two-Dimensional LSTMs</b> .....	964
<i>Bo Li, Tara N. Sainath</i>	

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

<b>Functional Principal Component Analysis of Vocal Tract Area Functions</b> .....	969
<i>Jorge C. Lucero</i>	
<b>Analysis of Acoustic-to-Articulatory Speech Inversion Across Different Accents and Languages</b> .....	974
<i>Ganesh Sivaraman, Carol Espy-Wilson, Martijn Wieling</i>	
<b>Integrated Mechanical Model of [r]-[l] and [b]-[m]-[w] Producing Consonant Cluster [br]</b> .....	979
<i>Takayuki Arai</i>	
<b>A Speaker Adaptive DNN Training Approach for Speaker-Independent Acoustic Inversion</b> .....	984
<i>Leonardo Badino, Luca Franceschi, Raman Arora, Michele Donini, Massimiliano Pontil</i>	
<b>Acoustic-to-Articulatory Mapping Based on Mixture of Probabilistic Canonical Correlation Analysis</b> .....	989
<i>Hidetsugu Uchida, Daisuke Saito, Nobuaki Minematsu</i>	
<b>Test-retest Repeatability of Articulatory Strategies Using Real-time Magnetic Resonance Imaging</b> .....	994
<i>Tanner Sorensen, Asterios Toutios, Johannes Toger, Louis Goldstein, Shrikanth Narayanan</i>	

### **TUE-O-3-4: SPEAKER RECOGNITION**

<b>Deep Neural Network Embeddings for Text-Independent Speaker Verification</b> .....	999
<i>David Snyder, Daniel Garcia-Romero, Daniel Povey, Sanjeev Khudanpur</i>	
<b>Tied Variational Autoencoder Backends for i-Vector Speaker Recognition</b> .....	1004
<i>Jesus Villalba, Niko Brummer, Najim Dehak</i>	
<b>Improved Gender Independent Speaker Recognition Using Convolutional Neural Network Based Bottleneck Features</b> .....	1009
<i>Shivesh Ranjan, John H. L. Hansen</i>	
<b>Autoencoder Based Domain Adaptation for Speaker Recognition Under Insufficient Channel Information</b> .....	1014
<i>Suwon Shon, Seongkyu Mun, Wooil Kim, Hanseok Ko</i>	
<b>Nonparametrically Trained Probabilistic Linear Discriminant Analysis for i-Vector Speaker Verification</b> .....	1019
<i>Abbas Khosravani, Mohammad Mehdi Homayounpour</i>	
<b>DNN Bottleneck Features for Speaker Clustering</b> .....	1024
<i>Jesus Jorin, Paola Garcia, Luis Buera</i>	

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

<b>Creak as a Feature of Lexical Stress in Estonian</b> .....	1029
<i>Katlin Aare, Partel Lippus, Juraj Simko</i>	
<b>Cross-Speaker Variation in Voice Source Correlates of Focus and Deaccentuation</b> .....	1034
<i>Irena Yanushevskaya, Ailbhe Ni Chasaide, Christer Gobl</i>	
<b>Acoustic Characterization of Word-final Glottal Stops in Mizo and Assam Sora</b> .....	1039
<i>Sishir Kalita, Wendy Lalhminghlui, Luke Horo, Priyankoo Sarmah, S. R. Mahadeva Prasanna, Samarendra Dandapat</i>	
<b>Iterative Optimal Preemphasis for Improved Glottal-Flow Estimation by Iterative Adaptive Inverse Filtering</b> .....	1044
<i>Parham Mokhtari, Hiroshi Ando</i>	
<b>Automatic Measurement of Pre-aspiration</b> .....	1049
<i>Yaniv Sheena, Misa Hejna, Yossi Adi, Joseph Keshet</i>	
<b>Acoustic and Electroglossographic Study of Breathy and Modal Vowels as Produced by Heritage and Native Gujarati Speakers</b> .....	1054
<i>Kiranpreet Nara</i>	

### **TUE-O-3-8: SPEECH SYNTHESIS PROSODY**

<b>An RNN-Based Quantized F0 Model with Multi-Tier Feedback Links for Text-to-Speech Synthesis</b> .....	1059
<i>Xin Wang, Shinji Takaki, Junichi Yamagishi</i>	
<b>Phrase Break Prediction for Long-Form Reading TTS: Exploiting Text Structure Information</b> .....	1064
<i>Viacheslav Klimkov, Adam Nadolski, Alexis Moinet, Bartosz Putrycz, Roberto Barra-Chicote, Thomas Merritt, Thomas Drugman</i>	
<b>Physically Constrained Statistical F<sub>0</sub> Prediction for Electrolaryngeal Speech Enhancement</b> .....	1069
<i>Kou Tanaka, Hirokazu Kameoka, Tomoki Toda, Satoshi Nakamura</i>	
<b>DNN-SPACE: DNN-HMM-Based Generative Model of Voice F<sub>0</sub> Contours for Statistical Phrase/Accent Command Estimation</b> .....	1074
<i>Nobukatsu Hojo, Yasuhito Ohsugi, Yusuke Ijima, Hirokazu Kameoka</i>	
<b>Controlling Prominence Realisation in Parametric DNN-Based Speech Synthesis</b> .....	1079
<i>Zofia Malisz, Harald Berthelsen, Jonas Beskow, Joakim Gustafson</i>	
<b>Increasing Recall of Lengthening Detection via Semi-Automatic Classification</b> .....	1084
<i>Simon Betz, Jana Vosse, Sina Zarriess, Petra Wagner</i>	

## **TUE-O-3-10: EMOTION RECOGNITION**

<b>Efficient Emotion Recognition from Speech Using Deep Learning on Spectrograms</b> .....	1089
<i>Aharon Satt, Shai Rozenberg, Ron Hoory</i>	
<b>Interaction and Transition Model for Speech Emotion Recognition in Dialogue</b> .....	1094
<i>Ruo Zhang, Ando Atsushi, Satoshi Kobashikawa, Yushi Aono</i>	
<b>Progressive Neural Networks for Transfer Learning in Emotion Recognition</b> .....	1098
<i>John Gideon, Soheil Khorram, Zakaria Aldeneh, Dimitrios Dimitriadis, Emily Mower Provost</i>	
<b>Jointly Predicting Arousal, Valence and Dominance with Multi-Task Learning</b> .....	1103
<i>Srinivas Parthasarathy, Carlos Busso</i>	
<b>Discretized Continuous Speech Emotion Recognition with Multi-Task Deep Recurrent Neural Network</b> .....	1108
<i>Duc Le, Zakaria Aldeneh, Emily Mower Provost</i>	
<b>Towards Speech Emotion Recognition "in the Wild" Using Aggregated Corpora and Deep Multi-Task Learning</b> .....	1113
<i>Jaebok Kim, Gwenn Englebienne, Khiet P. Truong, Vanessa Evers</i>	

## **TUE-O-4-1: WAVENET AND NOVEL PARADIGMS**

<b>Speaker-Dependent WaveNet Vocoder</b> .....	1118
<i>Akira Tamamori, Tomoki Hayashi, Kazuhiro Kobayashi, Kazuya Takeda, Tomoki Toda</i>	
<b>Waveform Modeling Using Stacked Dilated Convolutional Neural Networks for Speech Bandwidth Extension</b> .....	1123
<i>Yu Gu, Zhen-Hua Ling</i>	
<b>Direct Modeling of Frequency Spectra and Waveform Generation Based on Phase Recovery for DNN-Based Speech Synthesis</b> .....	1128
<i>Shinji Takaki, Hirokazu Kameoka, Junichi Yamagishi</i>	
<b>A Hierarchical Encoder-Decoder Model for Statistical Parametric Speech Synthesis</b> .....	1133
<i>Srikanth Ronanki, Oliver Watts, Simon King</i>	
<b>Statistical Voice Conversion with WaveNet-Based Waveform Generation</b> .....	1138
<i>Kazuhiro Kobayashi, Tomoki Hayashi, Akira Tamamori, Tomoki Toda</i>	
<b>Google's Next-Generation Real-Time Unit-Selection Synthesizer Using Sequence-to-Sequence LSTM-Based Autoencoders</b> .....	1143
<i>Vincent Wan, Yannis Agiomyrgiannakis, Hanna Silen, Jakub Vit</i>	

## **TUE-O-4-2: MODELS OF SPEECH PERCEPTION**

<b>A Comparison of Sentence-Level Speech Intelligibility Metrics</b> .....	1148
<i>Alexander Kain, Max Del Giudice, Kris Tjaden</i>	
<b>An Auditory Model of Speaker Size Perception for Voiced Speech Sounds</b> .....	1153
<i>Toshio Irino, Eri Takimoto, Toshie Matsui, Roy D. Patterson</i>	
<b>The Recognition of Compounds: A Computation Account</b> .....	1158
<i>L. Ten Bosch, L. Boves, M. Ernestus</i>	
<b>Humans Do Not Maximize the Probability of Correct Decision When Recognizing DANTALE Words in Noise</b> .....	1163
<i>Mohsen Zareian Jahromi, Jan Ostergaard, Jesper Jensen</i>	
<b>Single-Ended Prediction of Listening Effort Based on Automatic Speech Recognition</b> .....	1168
<i>Rainer Huber, Constantin Spille, Bernd T. Meyer</i>	
<b>Modeling Categorical Perception with the Receptive Fields of Auditory Neurons</b> .....	1173
<i>Chris Neufeld</i>	

## **TUE-O-4-4: SOURCE SEPARATION AND AUDITORY SCENE ANALYSIS**

<b>A Maximum Likelihood Approach to Deep Neural Network Based Nonlinear Spectral Mapping for Single-Channel Speech Separation</b> .....	1178
<i>Yaman Wang, Jun Du, Li-Rong Dai, Chin-Hui Lee</i>	
<b>Deep Clustering-Based Beamforming for Separation with Unknown Number of Sources</b> .....	1183
<i>Takuya Higuchi, Keisuke Kinoshita, Marc Delcroix, Katerina Zmolikova, Tomohiro Nakatani</i>	
<b>Time-Frequency Masking for Blind Source Separation with Preserved Spatial Cues</b> .....	1188
<i>Shadi Pirhossainloo, Kostas Kokkinakis</i>	
<b>Variational Recurrent Neural Networks for Speech Separation</b> .....	1193
<i>Jen-Tzung Chien, Kuan-Ting Kuo</i>	
<b>Detecting Overlapped Speech on Short Timeframes Using Deep Learning</b> .....	1198
<i>Valentin Andrei, Horia Cucu, Corneliu Burileanu</i>	
<b>Ideal Ratio Mask Estimation Using Deep Neural Networks for Monaural Speech Segregation in Noisy Reverberant Conditions</b> .....	1203
<i>Xu Li, Junfeng Li, Yonghong Yan</i>	

## **TUE-O-4-6: PROSODY: TONE AND INTONATION**

<b>The Vocative Chant and Beyond: German Calling Melodies Under Routine and Urgent Contexts</b> .....	1208
<i>Sergio I. Quiroz, Marzena Zygis</i>	

<b>Comparing Languages Using Hierarchical Prosodic Analysis</b> .....	1213
<i>Juraj Simko, Antti Suni, Katri Hiovain, Martti Vainio</i>	
<b>Intonation Facilitates Prediction of Focus Even in the Presence of Lexical Tones</b> .....	1218
<i>Martin Ho Kwan Ip, Anne Cutler</i>	
<b>Mind the Peak: When Museum is Temporarily Understood as Musical in Australian English</b> .....	1223
<i>Katharina Zahner, Heather Kember, Bettina Braun</i>	
<b>Pashto Intonation Patterns</b> .....	1228
<i>Luca Rognoni, Judith Bishop, Miriam Corris</i>	
<b>A New Model of Final Lowering in Spontaneous Monologue</b> .....	1233
<i>Kikuo Maekawa</i>	

#### **TUE-O-4-8: EMOTION MODELING**

<b>Speech Emotion Recognition with Emotion-Pair Based Framework Considering Emotion Distribution Information in Dimensional Emotion Space</b> .....	1238
<i>Xi Ma, Zhiyong Wu, Jia Jia, Mingxing Xu, Helen Meng, Lianhong Cai</i>	
<b>Adversarial Auto-Encoders for Speech Based Emotion Recognition</b> .....	1243
<i>Saurabh Sahu, Rahul Gupta, Ganesh Sivaraman, Wael Abdalmageed, Carol Espy-Wilson</i>	
<b>An Investigation of Emotion Prediction Uncertainty Using Gaussian Mixture Regression</b> .....	1248
<i>Ting Dang, Vidhyasaharan Sethu, Julien Epps, Eliathamby Ambikairajah</i>	
<b>Capturing Long-Term Temporal Dependencies with Convolutional Networks for Continuous Emotion Recognition</b> .....	1253
<i>Soheil Khorram, Zakaria Aldeneh, Dimitrios Dimitriadis, Melvin McInnis, Emily Mower Provost</i>	
<b>Voice-to-Affect Mapping: Inferences on Language Voice Baseline Settings</b> .....	1258
<i>Ailbhe Ni Chasaide, Irena Yanushevskaya, Christer Gobl</i>	
<b>Attentive Convolutional Neural Network Based Speech Emotion Recognition: A Study on the Impact of Input Features, Signal Length, and Acted Speech</b> .....	1263
<i>Michael Neumann, Ngoc Thang Vu</i>	

#### **TUE-O-4-10: VOICE CONVERSION 1**

<b>Voice Conversion Using Sequence-to-Sequence Learning of Context Posterior Probabilities</b> .....	1268
<i>Hiroyuki Miyoshi, Yuki Saito, Shinnosuke Takamichi, Hiroshi Saruwatari</i>	
<b>Learning Latent Representations for Speech Generation and Transformation</b> .....	1273
<i>Wei-Ning Hsu, Yu Zhang, James Glass</i>	
<b>Parallel-Data-Free Many-to-Many Voice Conversion Based on DNN Integrated with Eigenspace Using a Non-Parallel Speech Corpus</b> .....	1278
<i>Tetsuya Hashimoto, Hidetsugu Uchida, Daisuke Saito, Nobuaki Minematsu</i>	
<b>Sequence-to-Sequence Voice Conversion with Similarity Metric Learned Using Generative Adversarial Networks</b> .....	1283
<i>Takuhiro Kaneko, Hirokazu Kameoka, Kaoru Hiramatsu, Kunio Kashino</i>	
<b>A Mouth Opening Effect Based on Pole Modification for Expressive Singing Voice Transformation</b> .....	1288
<i>Luc Ardaillon, Axel Roebel</i>	
<b>Siamese Autoencoders for Speech Style Extraction and Switching Applied to Voice Identification and Conversion</b> .....	1293
<i>Seyed Hamidreza Mohammadi, Alexander Kain</i>	

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

<b>Recurrent Neural Aligner: An Encoder-Decoder Neural Network Model for Sequence to Sequence Mapping</b> .....	1298
<i>Hasim Sak, Matt Shannon, Kanishka Rao, Françoise Beaufays</i>	
<b>Highway-LSTM and Recurrent Highway Networks for Speech Recognition</b> .....	1303
<i>Golan Pundak, Tara N. Sainath</i>	
<b>Improving Speech Recognition by Revising Gated Recurrent Units</b> .....	1308
<i>Mirco Ravanelli, Philemon Brakel, Maurizio Omologo, Yoshua Bengio</i>	
<b>Stochastic Recurrent Neural Network for Speech Recognition</b> .....	1313
<i>Jen-Tzung Chien, Chen Shen</i>	
<b>Frame and Segment Level Recurrent Neural Networks for Phone Classification</b> .....	1318
<i>Martin Ratajczak, Sebastian Tschjatschek, Franz Pernkopf</i>	
<b>Deep Learning-Based Telephony Speech Recognition in the Wild</b> .....	1323
<i>Kyu J. Han, Seongjun Hahn, Byung-Hak Kim, Jungsuk Kim, Ian Lane</i>	

## VOLUME 3

### TUE-O-5-2: SPEAKER RECOGNITION EVALUATION

<b>The I4U Mega Fusion and Collaboration for NIST Speaker Recognition Evaluation 2016</b> .....	1328
<i>Kong Aik Lee, V. Hautamaki, T. Kinnunen, A. Larcher, C. Zhang, A. Nautsch, T. Stafylakis, G. Liu, M. Rouvier, W. Rao, F. Alegre, J. Ma, M. W. Mak, A. K. Sarkar, H. Delgado, R. Saeidi, H. Aronowitz, A. Sizov, H. Sun, T. H. Nguyen, G. Wang, B. Ma, V. Vestman, M. Sahidullah, M. Halonen, A. Kanervisto, G. Le Lan, F. Bahmaninezhad, S. Isadskiy, C. Rathgeb, C. Busch, G. Tzimiropoulos, Q. Qian, Z. Wang, Q. Zhao, T. Wang, H. Li, J. Xue, S. Zhu, R. Jin, T. Zhao, P.-M. Bousquet, M. Ajili, W. B. Kheder, D. Matrouf, Z. H. Lim, C. Xu, H. Xu, X. Xiao, E. S. Chng, B. Fauve, K. Sriskandaraja, V. Sethu, W. W. Lin, D. A. L. Thomsen, Z.-H. Tan, M. Todisco, N. Evans, H. Li, J. H. L. Hansen, J.-F. Bonastre, E. Ambikairajah</i>	
<b>The MIT-LL, JHU and LRDE NIST 2016 Speaker Recognition Evaluation System</b> .....	1333
<i>Pedro A. Torres-Carrasquillo, Fred Richardson, Shahan Nercessian, Douglas Sturim, William Campbell, Youngjune Gwon, Swaroop Vattam, Najim Dehak, Harish Mallidi, Phani Sankar Nidadavolu, Ruizhi Li, Reda Dehak</i>	
<b>Nuance - Politecnico di Torino's 2016 NIST Speaker Recognition Evaluation System</b> .....	1338
<i>Daniele Colibro, Claudio Vair, Emanuele Dalmaso, Kevin Farrell, Gennady Karvitsky, Sandro Cumani, Pietro Laface</i>	
<b>UTD-CRSS Systems for 2016 NIST Speaker Recognition Evaluation</b> .....	1343
<i>Chunlei Zhang, Fahimeh Bahmaninezhad, Shivesh Ranjan, Chengzhu Yu, Navid Shokouhi, John H. L. Hansen</i>	
<b>Analysis and Description of ABC Submission to NIST SRE 2016</b> .....	1348
<i>Oldrich Plchot, Pavel Matejka, Anna Silnova, Ondrej Novotny, Mireia Diez Sanchez, Johan Rohdin, Ondrej Glembek, Niko Brummer, Albert Swart, Jesus Jorin-Prieto, Paola Garcia, Luis Buera, Patrick Kenny, Jahangir Alam, Gautam Bhattacharya</i>	
<b>The 2016 NIST Speaker Recognition Evaluation</b> .....	1353
<i>Seyed Omid Sadjadi, Timothee Kheyrkhan, Audrey Tong, Craig Greenberg, Douglas Reynolds, Elliot Singer, Lisa Mason, Jaime Hernandez-Cordero</i>	

### TUE-O-5-4: GLOTTAL SOURCE MODELING

<b>A New Cosine Series Antialiasing Function and its Application to Aliasing-Free Glottal Source Models for Speech and Singing Synthesis</b> .....	1358
<i>Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, Tomoki Toda, Toshio Irino</i>	
<b>Speaking Style Conversion from Normal to Lombard Speech Using a Glottal Vocoder and Bayesian GMMs</b> .....	1363
<i>Ana Ramirez Lopez, Shreyas Seshadri, Lauri Juvela, Okko Rasanen, Paavo Alku</i>	
<b>Reducing Mismatch in Training of DNN-Based Glottal Excitation Models in a Statistical Parametric Text-to-Speech System</b> .....	1368
<i>Lauri Juvela, Bajjibabu Bollepalli, Junichi Yamagishi, Paavo Alku</i>	
<b>Semi Parametric Concatenative TTS with Instant Voice Modification Capabilities</b> .....	1373
<i>Alexander Sorin, Slava Shechtman, Asaf Rendel</i>	
<b>Modeling Laryngeal Muscle Activation Noise for Low-Order Physiological Based Speech Synthesis</b> .....	1378
<i>Rodrigo Manriquez, Sean D. Peterson, Pavel Prado, Patricio Orio, Mattias Zanartu</i>	
<b>Direct Modelling of Magnitude and Phase Spectra for Statistical Parametric Speech Synthesis</b> .....	1383
<i>Felipe Espic, Cassia Valentini Botinhao, Simon King</i>	

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

<b>Similar Prosodic Structure Perceived Differently in German and English</b> .....	1388
<i>Heather Kember, Ann-Kathrin Grohe, Katharina Zahner, Bettina Braun, Andrea Weber, Anne Cutler</i>	
<b>Disambiguate or not? --- The Role of Prosody in Unambiguous and Potentially Ambiguous Anaphora Production in Strictly Mandarin Parallel Structures</b> .....	1393
<i>Luying Hou, Bert Le Bruyn, Rene Kager</i>	
<b>Acoustic Properties of Canonical and Non-Canonical Stress in French, Turkish, Armenian and Brazilian Portuguese</b> .....	1398
<i>Angeliki Athanasopoulou, Irene Vogel, Hossep Dolatian</i>	
<b>Phonological Complexity, Segment Rate and Speech Tempo Perception</b> .....	1403
<i>Leendert Plug, Rachel Smith</i>	
<b>On the Duration of Mandarin Tones</b> .....	1407
<i>Jing Yang, Yu Zhang, Aijun Li, Li Xu</i>	
<b>The Formant Dynamics of Long Close Vowels in Three Varieties of Swedish</b> .....	1412
<i>Otto Ewald, Eva Liina Asu, Susanne Schotz</i>	

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

<b>Bidirectional LSTM-RNN for Improving Automated Assessment of Non-Native Children's Speech</b> .....	1417
<i>Yao Qian, Keelan Evanini, Xinhao Wang, Chong Min Lee, Matthew Mulholland</i>	
<b>Automatic Scoring of Shadowing Speech Based on DNN Posteriors and Their DTW</b> .....	1422
<i>Junwei Yue, Fumiya Shiozawa, Shohei Toyama, Yutaka Yamauchi, Kayoko Ito, Daisuke Saito, Nobuaki Minematsu</i>	
<b>Off-Topic Spoken Response Detection Using Siamese Convolutional Neural Networks</b> .....	1427
<i>Chong Min Lee, Su-Youn Yoon, Xihao Wang, Matthew Mulholland, Ikkyu Choi, Keelan Evanini</i>	
<b>Phonological Feature Based Mispronunciation Detection and Diagnosis Using Multi-Task DNNs and Active Learning</b> .....	1432
<i>Vipul Arora, Aditi Lahiri, Henning Reetz</i>	

<b>Detection of Mispronunciations and Disfluencies in Children Reading Aloud</b> .....	1437
<i>Jorge Proenca, Carla Lopes, Michael Tjalve, Andreas Stolcke, Sara Candeias, Fernando Perdigao</i>	
<b>Automatic Assessment of Non-Native Prosody by Measuring Distances on Prosodic Label Sequences</b> .....	1442
<i>David Escudero-Mancebo, Cesar Gonzalez-Ferreras, Lourdes Aguilar, Eva Estebas-Vilaplana</i>	

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

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

### **TUE-P-3-1: SHORT UTTERANCES SPEAKER RECOGNITION**

<b>A Generative Model for Score Normalization in Speaker Recognition</b> .....	1477
<i>Albert Swart, Niko Brummer</i>	
<b>Content Normalization for Text-Dependent Speaker Verification</b> .....	1482
<i>Subhadeep Dey, Srikanth Madikeri, Petr Motlicek, Marc Ferras</i>	
<b>End-to-End Text-Independent Speaker Verification with Triplet Loss on Short Utterances</b> .....	1487
<i>Chunlei Zhang, Kazuhito Koishida</i>	
<b>Adversarial Network Bottleneck Features for Noise Robust Speaker Verification</b> .....	1492
<i>Hong Yu, Zheng-Hua Tan, Zhanyu Ma, Jun Guo</i>	
<b>What Does the Speaker Embedding Encode?</b> .....	1497
<i>Shuai Wang, Yanmin Qian, Kai Yu</i>	
<b>Incorporating Local Acoustic Variability Information into Short Duration Speaker Verification</b> .....	1502
<i>Jianbo Ma, Vidhyasaharan Sethu, Eliathamby Ambikairajah, Kong Aik Lee</i>	
<b>DNN i-Vector Speaker Verification with Short, Text-Constrained Test Utterances</b> .....	1507
<i>Jinghua Zhong, Wenping Hu, Frank K. Soong, Helen Meng</i>	
<b>Time-Varying Autoregressions for Speaker Verification in Reverberant Conditions</b> .....	1512
<i>Ville Vestman, Dhananjaya Gowda, Md. Sahidullah, Paavo Alku, Tomi Kinnunen</i>	
<b>Deep Speaker Embeddings for Short-Duration Speaker Verification</b> .....	1517
<i>Gautam Bhattacharya, Jahangir Alam, Patrick Kenny</i>	
<b>Using Voice Quality Features to Improve Short-Utterance, Text-Independent Speaker Verification Systems</b> .....	1522
<i>Soo Jin Park, Gary Yeung, Jody Kreiman, Patricia A. Keating, Abeer Alwan</i>	
<b>Gain Compensation for Fast i-Vector Extraction Over Short Duration</b> .....	1527
<i>Kong Aik Lee, Haizhou Li</i>	
<b>Joint Training of Expanded End-to-End DNN for Text-Dependent Speaker Verification</b> .....	1532
<i>Hee-Soo Heo, Jee-Weon Jung, Il-Ho Yang, Sung-Hyun Yoon, Ha-Jin Yu</i>	

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

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

<b>Domain Mismatch Modeling of Out-Domain i-Vectors for PLDA Speaker Verification .....</b>	<b>1581</b>
<i>Md. Hafizur Rahman, Ivan Himawan, David Dean, Sridha Sridharan</i>	

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

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

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

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

### **TUE-P-4-3: DIALOG MODELING**

<b>Online End-of-Turn Detection from Speech Based on Stacked Time-Asynchronous Sequential Networks .....</b>	<b>1661</b>
<i>Ryo Masumura, Taichi Asami, Hirokazu Masataki, Ryo Ishii, Ryuichiro Higashinaka</i>	
<b>Improving Prediction of Speech Activity Using Multi-Participant Respiratory State .....</b>	<b>1666</b>
<i>Marcin Wlodarczyk, Kornel Laskowski, Mattias Heldner, Katlin Aare</i>	
<b>Turn-Taking Offsets and Dialogue Context .....</b>	<b>1671</b>
<i>Peter A. Heeman, Rebecca Lunsford</i>	
<b>Towards Deep End-of-Turn Prediction for Situated Spoken Dialogue Systems .....</b>	<b>1676</b>
<i>Angelika Maier, Julian Hough, David Schlangen</i>	
<b>End-of-Utterance Prediction by Prosodic Features and Phrase-Dependency Structure in Spontaneous Japanese Speech .....</b>	<b>1681</b>
<i>Yuichi Ishimoto, Takehiro Teraoka, Mika Enomoto</i>	
<b>Turn-Taking Estimation Model Based on Joint Embedding of Lexical and Prosodic Contents .....</b>	<b>1686</b>
<i>Chaoran Liu, Carlos Ishi, Hiroshi Ishiguro</i>	
<b>Social Signal Detection in Spontaneous Dialogue Using Bidirectional LSTM-CTC .....</b>	<b>1691</b>
<i>Hirofumi Inaguma, Koji Inoue, Masato Mimura, Tatsuya Kawahara</i>	
<b>Entrainment in Multi-Party Spoken Dialogues at Multiple Linguistic Levels .....</b>	<b>1696</b>
<i>Zahra Rahimi, Anish Kumar, Diane Litman, Susannah Paletz, Mingzhi Yu</i>	
<b>Measuring Synchrony in Task-Based Dialogues .....</b>	<b>1701</b>
<i>Justine Reverdy, Carl Vogel</i>	
<b>Sequence to Sequence Modeling for User Simulation in Dialog Systems .....</b>	<b>1706</b>
<i>Paul Crook, Alex Marin</i>	
<b>Human and Automated Scoring of Fluency, Pronunciation and Intonation During Human--Machine Spoken Dialog Interactions .....</b>	<b>1711</b>
<i>Vikram Ramanarayanan, Patrick L. Lange, Keelan Evanini, Hillary R. Molloy, David Suendermann-Oeft</i>	
<b>Hierarchical LSTMs with Joint Learning for Estimating Customer Satisfaction from Contact Center Calls .....</b>	<b>1716</b>
<i>Atsushi Ando, Ryo Masumura, Hosana Kamiyama, Satoshi Kobashikawa, Yushi Aono</i>	



<b>Domain-Independent User Satisfaction Reward Estimation for Dialogue Policy Learning</b> .....	1721
<i>Stefan Ultes, Pawel Budzianowski, Inigo Casanueva, Nikola Mrksic, Lina Rojas-Barahona, Pei-Hao Su, Tsung-Hsien Wen, Milica Gasic, Steve Young</i>	
<b>Analysis of the Relationship Between Prosodic Features of Fillers and its Forms or Occurrence Positions</b> .....	1726
<i>Shizuka Nakamura, Ryosuke Nakanishi, Katsuya Takanashi, Tatsuya Kawahara</i>	
<b>Cross-Subject Continuous Emotion Recognition Using Speech and Body Motion in Dyadic Interactions</b> .....	1731
<i>Syeda Narjis Fatima, Engin Erzin</i>	

### **TUE-P-5-1: L1 AND L2 ACQUISITION**

<b>An Automatically Aligned Corpus of Child-Directed Speech</b> .....	1736
<i>Micha Elsner, Kiwako Ito</i>	
<b>A Comparison of Danish Listeners' Processing Cost in Judging the Truth Value of Norwegian, Swedish, and English Sentences</b> .....	1741
<i>Ocke-Schwen Bohn, Trine Askjaer-Jorgensen</i>	
<b>On the Role of Temporal Variability in the Acquisition of the German Vowel Length Contrast</b> .....	1745
<i>Felicitas Kleber</i>	
<b>A Data-driven Approach for Perceptually Validated Acoustic Features for Children's Sibilant Fricative Productions</b> .....	1750
<i>Patrick F. Reidy, Mary E. Beckman, Jan Edwards, Benjamin Munson</i>	
<b>Proficiency Assessment of ESL Learner's Sentence Prosody with TTS Synthesized Voice as Reference</b> .....	1755
<i>Yujia Xiao, Frank K. Soong</i>	
<b>Mechanisms of Tone Sandhi Rule Application by Non-Native Speakers</b> .....	1760
<i>Si Chen, Yunjuan He, Chun Wah Yuen, Bei Li, Yike Yang</i>	
<b>Changes in Early L2 Cue-Weighting of Non-Native Speech: Evidence from Learners of Mandarin Chinese</b> .....	1765
<i>Seth Wiener</i>	
<b>Directing Attention During Perceptual Training: A Preliminary Study of Phonetic Learning in Southern Min by Mandarin Speakers</b> .....	1770
<i>Ying Chen, Eric Pederson</i>	
<b>Prosody Analysis of L2 English for Naturalness Evaluation Through Speech Modification</b> .....	1775
<i>Dean Luo, Ruxin Luo, Lixin Wang</i>	
<b>Measuring Encoding Efficiency in Swedish and English Language Learner Speech Production</b> .....	1779
<i>Gintare Grigonyte, Gerold Schneider</i>	
<b>Lexical Adaptation to a Novel Accent in German: A Comparison Between German, Swedish, and Finnish Listeners</b> .....	1784
<i>Adriana Hanulikova, Jenny Ekstrom</i>	
<b>Qualitative Differences in L3 Learners' Neurophysiological Response to L1 versus L2 Transfer</b> .....	1789
<i>Alejandra Keidel Fernandez, Thomas Horberg</i>	
<b>Articulation Rate in Swedish Child-Directed Speech Increases as a Function of the Age of the Child Even When Surprisal is Controlled for</b> .....	1794
<i>Johan Sjons, Thomas Horberg, Robert Ostling, Johannes Bjerva</i>	
<b>The Relationship Between the Perception and Production of Non-Native Tones</b> .....	1799
<i>Kaile Zhang, Gang Peng</i>	
<b>MMN Responses in Adults After Exposure to Bimodal and Unimodal Frequency Distributions of Rotated Speech</b> .....	1804
<i>Ellen Marklund, Elisabet Eir Cortes, Johan Sjons</i>	

### **TUE-P-5-2: VOICE, SPEECH AND HEARING DISORDERS**

<b>Float Like a Butterfly Sting Like a Bee: Changes in Speech Preceded Parkinsonism Diagnosis for Muhammad Ali</b> .....	1809
<i>Visar Berisha, Julie Liss, Timothy Huston, Alan Wisler, Yishan Jiao, Jonathan Eig</i>	
<b>Cepstral and Entropy Analyses in Vowels Excerpted from Continuous Speech of Dysphonic and Control Speakers</b> .....	1814
<i>Antonella Castellana, Andreas Selamtzis, Giampiero Salvi, Alessio Carullo, Arianna Astolfi</i>	
<b>Classification of Bulbar ALS from Kinematic Features of the Jaw and Lips: Towards Computer-Mediated Assessment</b> .....	1819
<i>Andrea Bandini, Jordan R. Green, Lorne Zinman, Yana Yunusova</i>	
<b>Zero Frequency Filter Based Analysis of Voice Disorders</b> .....	1824
<i>Nagaraj Adiga, Vikram C. M., Keerthi Pulela, S. R. Mahadeva Prasanna</i>	
<b>Hypernasality Severity Analysis in Cleft Lip and Palate Speech Using Vowel Space Area</b> .....	1829
<i>Nikitha K., Sishir Kalita, C. M. Vikram, M. Pushpavathi, S. R. Mahadeva Prasanna</i>	
<b>Automatic Prediction of Speech Evaluation Metrics for Dysarthric Speech</b> .....	1834
<i>Imed Laaridh, Waad Ben Kheder, Corinne Fredouille, Christine Meunier</i>	
<b>Apkinson --- A Mobile Monitoring Solution for Parkinson's Disease</b> .....	1839
<i>Philipp Klumpp, Tomas Janu, Tomas Arias-Vergara, J. C. Vasquez-Correa, Juan Rafael Orozco-Arroyave, Elmar Noth</i>	
<b>Dysprosody Differentiate Between Parkinson's Disease, Progressive Supranuclear Palsy, and Multiple System Atrophy</b> .....	1844
<i>Jan Hlavnicka, Tereza Tykalova, Roman Cmejla, Jiri Klempir, Evzen Ruzicka, Jan Ruzs</i>	
<b>Interpretable Objective Assessment of Dysarthric Speech Based on Deep Neural Networks</b> .....	1849
<i>Ming Tu, Visar Berisha, Julie Liss</i>	
<b>Deep Autoencoder Based Speech Features for Improved Dysarthric Speech Recognition</b> .....	1854
<i>Bhavik Vachhani, Chitralekha Bhat, Biswajit Das, Sunil Kumar Kopparapu</i>	

<b>Prediction of Speech Delay from Acoustic Measurements</b> .....	1859
<i>Jason Lilley, Madhavi Ratnagiri, H. Timothy Bunnell</i>	
<b>The Frequency Range of "The Ling Six Sounds" in Standard Chinese</b> .....	1864
<i>Aijun Li, Hua Zhang, Wen Sun</i>	
<b>Production of Sustained Vowels and Categorical Perception of Tones in Mandarin Among Cochlear-Implanted Children</b> .....	1869
<i>Wentao Gu, Jiao Yin, James Mahshie</i>	

### **TUE-P-5-3: SOURCE SEPARATION AND VOICE ACTIVITY DETECTION**

<b>Audio Content Based Geotagging in Multimedia</b> .....	1874
<i>Anurag Kumar, Benjamin Elizalde, Bhiksha Raj</i>	
<b>Time Delay Histogram Based Speech Source Separation Using a Planar Array</b> .....	1879
<i>Zhaoqiong Huang, Zhanzhong Cao, Dongwen Ying, Jielin Pan, Yonghong Yan</i>	
<b>Excitation Source Features for Improving the Detection of Vowel Onset and Offset Points in a Speech Sequence</b> .....	1884
<i>Gayadhar Pradhan, Avinash Kumar, S. Shahnawazuddin</i>	
<b>A Contrast Function and Algorithm for Blind Separation of Audio Signals</b> .....	1889
<i>Wei Gao, Roberto Togneri, Victor Sreeram</i>	
<b>Weighted Spatial Covariance Matrix Estimation for MUSIC Based TDOA Estimation of Speech Source</b> .....	1894
<i>Chenglin Xu, Xiong Xiao, Sining Sun, Wei Rao, Eng Siong Chng, Haizhou Li</i>	
<b>Speaker Direction-of-Arrival Estimation Based on Frequency-Independent Beampattern</b> .....	1899
<i>Feng Guo, Yuhang Cao, Zheng Liu, Jiaen Liang, Baoqing Li, Xiaobing Yuan</i>	
<b>A Mask Estimation Method Integrating Data Field Model for Speech Enhancement</b> .....	1904
<i>Xianyun Wang, Changchun Bao, Feng Bao</i>	
<b>Improved End-of-Query Detection for Streaming Speech Recognition</b> .....	1909
<i>Matt Shannon, Gabor Simko, Shuo-Yiin Chang, Carolina Parada</i>	
<b>Using Approximated Auditory Roughness as a Pre-Filtering Feature for Human Screaming and Affective Speech AED</b> .....	1914
<i>Di He, Zuofu Cheng, Mark Hasegawa-Johnson, Deming Chen</i>	
<b>Improving Source Separation via Multi-Speaker Representations</b> .....	1919
<i>Jeroen Zegers, Hugo Van Hamme</i>	
<b>Multiple Sound Source Counting and Localization Based on Spatial Principal Eigenvector</b> .....	1924
<i>Bing Yang, Hong Liu, Cheng Pang</i>	
<b>Subband Selection for Binaural Speech Source Localization</b> .....	1929
<i>Girija Ramesan Karthik, Prasanta Kumar Ghosh</i>	
<b>Unmixing Convolutional Mixtures by Exploiting Amplitude Co-Modulation: Methods and Evaluation on Mandarin Speech Recordings</b> .....	1934
<i>Bo-Rui Chen, Huang-Yi Lee, Yi-Wen Liu</i>	
<b>Bimodal Recurrent Neural Network for Audiovisual Voice Activity Detection</b> .....	1938
<i>Fei Tao, Carlos Busso</i>	
<b>Domain-Specific Utterance End-Point Detection for Speech Recognition</b> .....	1943
<i>Roland Maas, Ariya Rastrow, Kyle Goehner, Gautam Tiwari, Shaun Joseph, Bjorn Hoffmeister</i>	
<b>Speech Detection and Enhancement Using Single Microphone for Distant Speech Applications in Reverberant Environments</b> .....	1948
<i>Vinay Kothapally, John H. L. Hansen</i>	

### **TUE-P-5-4: SPEECH-ENHANCEMENT**

<b>A Post-Filtering Approach Based on Locally Linear Embedding Difference Compensation for Speech Enhancement</b> .....	1953
<i>Yi-Chiao Wu, Hsin-Te Hwang, Syu-Siang Wang, Chin-Cheng Hsu, Yu Tsao, Hsin-Min Wang</i>	
<b>Multi-Target Ensemble Learning for Monaural Speech Separation</b> .....	1958
<i>Hui Zhang, Xueliang Zhang, Guanglai Gao</i>	
<b>Improved Example-Based Speech Enhancement by Using Deep Neural Network Acoustic Model for Noise Robust Example Search</b> .....	1963
<i>Atsunori Ogawa, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani</i>	
<b>Subjective Intelligibility of Deep Neural Network-Based Speech Enhancement</b> .....	1968
<i>Femke B. Gelderblom, Tron V. Tronstad, Erlend Magnus Viggen</i>	
<b>Real-Time Modulation Enhancement of Temporal Envelopes for Increasing Speech Intelligibility</b> .....	1973
<i>Maria Koutsogiannaki, Holly Francois, Kihyun Choo, Eunmi Oh</i>	
<b>On the Influence of Modifying Magnitude and Phase Spectrum to Enhance Noisy Speech Signals</b> .....	1978
<i>Hans-Gunter Hirsch, Michael Gref</i>	
<b>MixMax Approximation as a Super-Gaussian Log-Spectral Amplitude Estimator for Speech Enhancement</b> .....	1983
<i>Robert Rehr, Timo Gerkmann</i>	
<b>Binary Mask Estimation Strategies for Constrained Imputation-Based Speech Enhancement</b> .....	1988
<i>Ricard Marxer, Jon Barker</i>	
<b>A Fully Convolutional Neural Network for Speech Enhancement</b> .....	1993
<i>Se Rim Park, Jin Won Lee</i>	

<b>Speech Enhancement Using Non-Negative Spectrogram Models with Mel-Generalized Cepstral Regularization .....</b>	<b>1998</b>
<i>Li Li, Hirokazu Kameoka, Tomoki Toda, Shoji Makino</i>	

## VOLUME 4

<b>A Comparison of Perceptually Motivated Loss Functions for Binary Mask Estimation in Speech Separation .....</b>	<b>2003</b>
<i>Danny Websdale, Ben Milner</i>	
<b>Conditional Generative Adversarial Networks for Speech Enhancement and Noise-Robust Speaker Verification .....</b>	<b>2008</b>
<i>Daniel Michelsanti, Zheng-Hua Tan</i>	
<b>Speech Enhancement Using Bayesian Wavenet .....</b>	<b>2013</b>
<i>Kaizhi Qian, Yang Zhang, Shiyu Chang, Xuesong Yang, Dinei Florencio, Mark Hasegawa-Johnson</i>	
<b>Binaural Reverberant Speech Separation Based on Deep Neural Networks .....</b>	<b>2018</b>
<i>Xueliang Zhang, Deliang Wang</i>	
<b>On the Quality and Intelligibility of Noisy Speech Processed for Near-End Listening Enhancement .....</b>	<b>2023</b>
<i>Tudor-Catalin Zorila, Yannis Stylianou</i>	

### TUE-S&T-3/4-A: SHOW & TELL 3

<b>Applications of the BBN Sage Speech Processing Platform .....</b>	<b>2028</b>
<i>Ralf Meermeier, Sean Colbath</i>	
<b>Bob Speaks Kaldi .....</b>	<b>2030</b>
<i>Milos Cernak, Alain Komaty, Amir Mohammadi, Andre Anjos, Sebastien Marcel</i>	
<b>Real Time Pitch Shifting with Formant Structure Preservation Using the Phase Vocoder .....</b>	<b>2032</b>
<i>Michal Lenarczyk</i>	
<b>A Signal Processing Approach for Speaker Separation Using SFF Analysis .....</b>	<b>2034</b>
<i>Nivedita Chennupati, B. H. V. S. Narayana Murthy, B. Yegnanarayana</i>	
<b>Speech Recognition and Understanding on Hardware-Accelerated DSP .....</b>	<b>2036</b>
<i>Georg Stemmer, Munir Georges, Joachim Hofer, Piotr Rozen, Josef Bauer, Jakub Nowicki, Tobias Bocklet, Hannah R. Colett, Ohad Falik, Michael Deisher, Sylvia J. Downing</i>	
<b>MetaLab: A Repository for Meta-Analyses on Language Development, and More .....</b>	<b>2038</b>
<i>Sho Tsuji, Christina Bergmann, Molly Lewis, Mika Braginsky, Page Piccinini, Michael C. Frank, Alejandrina Cristia</i>	

### TUE-S&T-3/4-B: SHOW & TELL 4

<b>Evolving Recurrent Neural Networks That Process and Classify Raw Audio in a Streaming Fashion .....</b>	<b>2040</b>
<i>Adrien Daniel</i>	
<b>Combining Gaussian Mixture Models and Segmental Feature Models for Speaker Recognition .....</b>	<b>2042</b>
<i>Milana Milosevic, Ulrike Glavitsch</i>	
<b>"Did You Laugh Enough Today?" --- Deep Neural Networks for Mobile and Wearable Laughter Trackers .....</b>	<b>2044</b>
<i>Gerhard Hagerer, Nicholas Cummins, Florian Eyben, Bjorn Schuller</i>	
<b>Low-Frequency Ultrasonic Communication for Speech Broadcasting in Public Transportation .....</b>	<b>2046</b>
<i>Kwang Myung Jeon, Nam Kyun Kim, Chan Woong Kwak, Jung Min Moon, Hong Kook Kim</i>	
<b>Real-Time Speech Enhancement with GCC-NMF: Demonstration on the Raspberry Pi and NVIDIA Jetson .....</b>	<b>2048</b>
<i>Sean U. N. Wood, Jean Rouat</i>	
<b>Reading Validation for Pronunciation Evaluation in the Digitala Project .....</b>	<b>2050</b>
<i>Aku Rouhe, Reima Karhila, Peter Smit, Mikko Kurimo</i>	

### KEYNOTE 2: CATHERINE PELACHAUD

<b>Conversing with Social Agents That Smile and Laugh .....</b>	<b>2052</b>
<i>Catherine Pelachaud</i>	

### WED-SS-6-2: SPECIAL SESSION: DIGITAL REVOLUTION FOR UNDER-RESOURCED LANGUAGES 1

<b>Team ELISA System for DARPA LORELEI Speech Evaluation 2016 .....</b>	<b>2053</b>
<i>Pavlos Papadopoulos, Ruchir Travadi, Colin Vaz, Nikolaos Malandrakis, Ulf Hermjakob, Nima Pourdamghani, Michael Pust, Boliang Zhang, Xiaoman Pan, Di Lu, Ying Lin, Ondrej Glembek, Murali Karthick Baskar, Martin Karafiat, Lukas Burget, Mark Hasegawa-Johnson, Heng Ji, Jonathan May, Kevin Knight, Shrikanth S. Narayanan</i>	
<b>First Results in Developing a Medieval Latin Language Charter Dictation System for the East-Central Europe Region .....</b>	<b>2058</b>
<i>Peter Mihajlik, Lili Szabo, Balazs Tarjan, Andras Balog, Krisztina Rabai</i>	
<b>The Motivation and Development of MP<i>A</i>i, a Maori Pronunciation Aid .....</b>	<b>2063</b>
<i>C. I. Watson, P. J. Keegan, M. A. Maclagan, R. Harlow, J. King</i>	
<b>On the Linguistic Relevance of Speech Units Learned by Unsupervised Acoustic Modeling .....</b>	<b>2068</b>
<i>Siyuan Feng, Tan Lee</i>	
<b>Deep Auto-Encoder Based Multi-Task Learning Using Probabilistic Transcriptions .....</b>	<b>2073</b>
<i>Amit Das, Mark Hasegawa-Johnson, Karel Vesely</i>	

<b>Areal and Phylogenetic Features for Multilingual Speech Synthesis</b> .....	2078
<i>Alexander Gutkin, Richard Sproat</i>	

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

<b>SLPAnnotator: Tools for Implementing Sign Language Phonetic Annotation</b> .....	2083
<i>Kathleen Currie Hall, Scott Mackie, Michael Fry, Oksana Tkachman</i>	
<b>The LENA™ System Applied to Swedish: Reliability of the Adult Word Count Estimate</b> .....	2088
<i>Iris-Corinna Schwarz, Noor Botros, Aleksandra Lord, Amelie Marcusson, Henrik Tidelius, Ellen Marklund</i>	
<b>What do Babies Hear? Analyses of Child- and Adult-Directed Speech</b> .....	2093
<i>Marisa Casillas, Andrei Amatum, Amanda Seidl, Melanie Soderstrom, Anne S. Warlaumont, Erika Bergelson</i>	
<b>A New Workflow for Semi-Automated Annotations: Tests with Long-Form Naturalistic Recordings of Childrens Language Environments</b> .....	2098
<i>Marisa Casillas, Erika Bergelson, Anne S. Warlaumont, Alejandrina Cristia, Melanie Soderstrom, Mark Vandam, Han Sloetjes</i>	
<b>Top-Down versus Bottom-Up Theories of Phonological Acquisition: A Big Data Approach</b> .....	2103
<i>Christina Bergmann, Sho Tsuji, Alejandrina Cristia</i>	
<b>Which Acoustic and Phonological Factors Shape Infants' Vowel Discrimination? Exploiting Natural Variation in InPhonDB</b> .....	2108
<i>Sho Tsuji, Alejandrina Cristia</i>	

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

<b>The ABAIR Initiative: Bringing Spoken Irish into the Digital Space</b> .....	2113
<i>Ailbhe Ni Chasaide, Neasa Ni Chiarain, Christoph Wendler, Harald Berthelsen, Andy Murphy, Christer Gobl</i>	
<b>Very Low Resource Radio Browsing for Agile Developmental and Humanitarian Monitoring</b> .....	2118
<i>Armin Saeb, Raghav Menon, Hugh Cameron, William Kibira, John Quinn, Thomas Niesler</i>	
<b>Extracting Situation Frames from Non-English Speech: Evaluation Framework and Pilot Results</b> .....	2123
<i>Nikolaos Malandrakis, Ondrej Glembek, Shrikanth S. Narayanan</i>	
<b>Eliciting Meaningful Units from Speech</b> .....	2128
<i>Daniil Kocharov, Tatiana Kachkovskaia, Pavel Skrelin</i>	
<b>Unsupervised Speech Signal to Symbol Transformation for Zero Resource Speech Applications</b> .....	2133
<i>Saurabhchand Bhati, Shekhar Nayak, K. Sri Rama Murty</i>	
<b>Machine Assisted Analysis of Vowel Length Contrasts in Wolof</b> .....	2138
<i>Elodie Gauthier, Laurent Besacier, Sylvie Voisin</i>	
<b>Leveraging Text Data for Word Segmentation for Underresourced Languages</b> .....	2143
<i>Thomas Glarner, Benedikt Boenninghoff, Oliver Walter, Reinhold Haeb-Umbach</i>	
<b>Improving DNN Bluetooth Narrowband Acoustic Models by Cross-Bandwidth and Cross-Lingual Initialization</b> .....	2148
<i>Xiaodan Zhuang, Arnab Ghoshal, Antti-Veikko Rosti, Matthias Paulik, Daben Liu</i>	
<b>Joint Estimation of Articulatory Features and Acoustic Models for Low-Resource Languages</b> .....	2153
<i>Basil Abraham, S. Umesh, Neethu Mariam Joy</i>	
<b>Transfer Learning and Distillation Techniques to Improve the Acoustic Modeling of Low Resource Languages</b> .....	2158
<i>Basil Abraham, Tejaswi Seeram, S. Umesh</i>	
<b>Building an ASR Corpus Using Althingi's Parliamentary Speeches</b> .....	2163
<i>Inga Run Helgadóttir, Robert Kjaran, Anna Bjork Nikulasdóttir, Jon Gudnason</i>	
<b>Implementation of a Radiology Speech Recognition System for Estonian Using Open Source Software</b> .....	2168
<i>Tanel Alumae, Andrus Paats, Ivo Fridolin, Einar Meister</i>	
<b>Building ASR Corpora Using Eyra</b> .....	2173
<i>Jon Gudnason, Matthias Petursson, Robert Kjaran, Simon Kluppel, Anna Bjork Nikulasdóttir</i>	
<b>Rapid Development of TTS Corpora for Four South African Languages</b> .....	2178
<i>Daniel Van Niekerk, Charl Van Heerden, Marelie Davel, Neil Kleynhans, Oddur Kjartansson, Martin Jansche, Linne Ha</i>	
<b>Uniform Multilingual Multi-Speaker Acoustic Model for Statistical Parametric Speech Synthesis of Low-Resourced Languages</b> .....	2183
<i>Alexander Gutkin</i>	
<b>Nativization of Foreign Names in TTS for Automatic Reading of World News in Swahili</b> .....	2188
<i>Joseph Mendelson, Pilar Oplustil, Oliver Watts, Simon King</i>	

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

<b>Multi-Task Learning for Mispronunciation Detection on Singapore Children's Mandarin Speech</b> .....	2193
<i>Rong Tong, Nancy F. Chen, Bin Ma</i>	
<b>Relating Unsupervised Word Segmentation to Reported Vocabulary Acquisition</b> .....	2198
<i>Elin Larsen, Alejandrina Cristia, Emmanuel Dupoux</i>	
<b>Modelling the Informativeness of Non-Verbal Cues in Parent-Child Interaction</b> .....	2203
<i>Mats Wiren, Kristina N. Bjorkenstam, Robert Ostling</i>	
<b>Computational Simulations of Temporal Vocalization Behavior in Adult-Child Interaction</b> .....	2208
<i>Ellen Marklund, David Pagmar, Tove Gerholm, Lisa Gustavsson</i>	
<b>Approximating Phonotactic Input in Children's Linguistic Environments from Orthographic Transcripts</b> .....	2213
<i>Sofia Strombergsson, Jens Edlund, Jana Gotze, Kristina Nilsson Bjorkenstam</i>	

<b>Learning Weakly Supervised Multimodal Phoneme Embeddings .....</b>	<b>2218</b>
<i>Rahma Chaabouni, Ewan Dunbar, Neil Zeghidour, Emmanuel Dupoux</i>	

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

<b>Personalized Quantification of Voice Attractiveness in Multidimensional Merit Space.....</b>	<b>2223</b>
<i>Yasunari Obuchi</i>	
<b>The Role of Temporal Amplitude Modulations in the Political Arena: Hillary Clinton vs. Donald Trump .....</b>	<b>2228</b>
<i>Hans Rutger Bosker</i>	
<b>Perceptual Ratings of Voice Likability Collected Through In-Lab Listening Tests vs. Mobile-Based Crowdsourcing.....</b>	<b>2233</b>
<i>Laura Fernandez Gallardo, Rafael Zequeira Jimenez, Sebastian Moller</i>	
<b>Attractiveness of French Voices for German Listeners --- Results from Native and Non-Native Read Speech.....</b>	<b>2238</b>
<i>Jurgen Trouvain, Frank Zimmerer</i>	
<b>Social Attractiveness in Dialogs .....</b>	<b>2243</b>
<i>Antje Schweitzer, Natalie Lewandowski, Daniel Duran</i>	
<b>A Gender Bias in the Acoustic-Melodic Features of Charismatic Speech? .....</b>	<b>2248</b>
<i>Eszter Novak-Tot, Oliver Niebuhr, Aoju Chen</i>	
<b>Pitch Convergence as an Effect of Perceived Attractiveness and Likability .....</b>	<b>2253</b>
<i>Jan Michalsky, Heike Schoormann</i>	
<b>Does Posh English Sound Attractive? .....</b>	<b>2257</b>
<i>Li Jiao, Chengxia Wang, Cristiane Hsu, Peter Birkholz, Yi Xu</i>	
<b>Large-Scale Speaker Ranking from Crowdsourced Pairwise Listener Ratings .....</b>	<b>2262</b>
<i>Timo Baumann</i>	

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

<b>Aerodynamic Features of French Fricatives .....</b>	<b>2267</b>
<i>Rosario Signorello, Sergio Hassid, Didier Demolin</i>	
<b>Inter-Speaker Variability: Speaker Normalisation and Quantitative Estimation of Articulatory Invariants in Speech Production for French .....</b>	<b>2272</b>
<i>Antoine Serrurier, Pierre Badin, Louis-Jean Boe, Laurent Lamalle, Christiane Neuschaefer-Rube</i>	
<b>Comparison of Basic Beatboxing Articulations Between Expert and Novice Artists Using Real-Time Magnetic Resonance Imaging.....</b>	<b>2277</b>
<i>Nimisha Patil, Timothy Greer, Reed Blaylock, Shrikanth S. Narayanan</i>	
<b>Speaker-specific Biomechanical Model-based Investigation of a Simple Speech Task based on Tagged-MRI .....</b>	<b>2282</b>
<i>Keyi Tang, Negar M. Harandi, Jonghye Woo, Georges El Fakhri, Mavreen Stone, Sidney Fels</i>	
<b>Sounds of the Human Vocal Tract .....</b>	<b>2287</b>
<i>Reed Blaylock, Nimisha Patil, Timothy Greer, Shrikanth S. Narayanan</i>	
<b>A Simulation Study on the Effect of Glottal Boundary Conditions on Vocal Tract Formants .....</b>	<b>2292</b>
<i>Yasufumi Uezu, Tokihiko Kaburagi</i>	

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

<b>A Robust and Alternative Approach to Zero Frequency Filtering Method for Epoch Extraction.....</b>	<b>2297</b>
<i>P. Gangamohan, B. Yegnanarayana</i>	
<b>Improving YANGsaf F0 Estimator with Adaptive Kalman Filter .....</b>	<b>2301</b>
<i>Kanru Hua</i>	
<b>A Spectro-Temporal Demodulation Technique for Pitch Estimation .....</b>	<b>2306</b>
<i>Jitendra Kumar Dhiman, Nagaraj Adiga, Chandra Sekhar Seelamantula</i>	
<b>Robust Method for Estimating F0 of Complex Tone Based on Pitch Perception of Amplitude Modulated Signal.....</b>	<b>2311</b>
<i>Kenichiro Miwa, Masashi Unoki</i>	
<b>Low-Complexity Pitch Estimation Based on Phase Differences Between Low-Resolution Spectra .....</b>	<b>2316</b>
<i>Simon Graf, Tobias Herbig, Markus Buck, Gerhard Schmidt</i>	
<b>Harvest: A High-Performance Fundamental Frequency Estimator from Speech Signals .....</b>	<b>2321</b>
<i>Masanori Morise</i>	

**WED-O-6-6: DIALOG AND PROSODY**

<b>Prosodic Event Recognition Using Convolutional Neural Networks with Context Information .....</b>	<b>2326</b>
<i>Sabrina Stehwien, Ngoc Thang Vu</i>	
<b>Prosodic Facilitation and Interference While Judging on the Veracity of Synthesized Statements .....</b>	<b>2331</b>
<i>Ramiro H. Galvez, Stefan Benus, Agustin Gravano, Marian Trnka</i>	
<b>An Investigation of Pitch Matching Across Adjacent Turns in a Corpus of Spontaneous German .....</b>	<b>2336</b>
<i>Margaret Zellers, Antje Schweitzer</i>	
<b>The Relationship Between F0 Synchrony and Speech Convergence in Dyadic Interaction .....</b>	<b>2341</b>
<i>Sankar Mukherjee, Alessandro D'Ausilio, Noel Nguyen, Luciano Fadiga, Leonardo Badino</i>	

<b>The Role of Linguistic and Prosodic Cues on the Prediction of Self-Reported Satisfaction in Contact Centre Phone Calls</b> .....	2346
<i>Jordi Luque, Carlos Segura, Ariadna Sanchez, Marti Umbert, Luis Angel Galindo</i>	
<b>Cross-Linguistic Study of the Production of Turn-Taking Cues in American English and Argentine Spanish</b> .....	2351
<i>Pablo Brusco, Juan Manuel Perez, Agustin Gravano</i>	

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

<b>Emotional Features for Speech Overlaps Classification</b> .....	2356
<i>Olga Egorow, Andreas Wendemuth</i>	
<b>Computing Multimodal Dyadic Behaviors During Spontaneous Diagnosis Interviews Toward Automatic Categorization of Autism Spectrum Disorder</b> .....	2361
<i>Chin-Po Chen, Xian-Hong Tseng, Susan Shur-Fen Gau, Chi-Chun Lee</i>	
<b>Deriving Dyad-Level Interaction Representation Using Interlocutors Structural and Expressive Multimodal Behavior Features</b> .....	2366
<i>Yun-Shao Lin, Chi-Chun Lee</i>	
<b>Spotting Social Signals in Conversational Speech over IP: A Deep Learning Perspective</b> .....	2371
<i>Raymond Brueckner, Maximilian Schmitt, Maja Pantic, Bjorn Schuller</i>	
<b>Optimized Time Series Filters for Detecting Laughter and Filler Events</b> .....	2376
<i>Gabor Goszolya</i>	
<b>Visual, Laughter, Applause and Spoken Expression Features for Predicting Engagement Within TED Talks</b> .....	2381
<i>Fasih Haider, Fahim A. Salim, Saturnino Luz, Carl Vogel, Owen Conlan, Nick Campbell</i>	

### **WED-O-6-10: ACOUSTIC MODEL ADAPTATION**

<b>Large-Scale Domain Adaptation via Teacher-Student Learning</b> .....	2386
<i>Jinyu Li, Michael L. Seltzer, Xi Wang, Rui Zhao, Yifan Gong</i>	
<b>Improving Children's Speech Recognition Through Explicit Pitch Scaling Based on Iterative Spectrogram Inversion</b> .....	2391
<i>W. Ahmad, S. Shah Nawazuddin, H. K. Kathania, Gayadhar Pradhan, A. B. Samaddar</i>	
<b>RNN-LDA Clustering for Feature Based DNN Adaptation</b> .....	2396
<i>Xurong Xie, Xunying Liu, Tan Lee, Lan Wang</i>	
<b>Robust Online i-Vectors for Unsupervised Adaptation of DNN Acoustic Models: A Study in the Context of Digital Voice Assistants</b> .....	2401
<i>Harish Arsikere, Sri Garimella</i>	
<b>Semi-Supervised Learning with Semantic Knowledge Extraction for Improved Speech Recognition in Air Traffic Control</b> .....	2406
<i>Ajay Srinivasanurthy, Petr Motlicek, Ivan Himawan, Gyorgy Szaszak, Youssef Oualil, Hartmut Helmke</i>	
<b>Dynamic Layer Normalization for Adaptive Neural Acoustic Modeling in Speech Recognition</b> .....	2411
<i>Taesup Kim, Inchul Song, Yoshua Bengio</i>	

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

<b>An Entrained Rhythm's Frequency, Not Phase, Influences Temporal Sampling of Speech</b> .....	2416
<i>Hans Rutger Bosker, Anne Kosem</i>	
<b>Context Regularity Indexed by Auditory N1 and P2 Event-Related Potentials</b> .....	2421
<i>Xiao Wang, Yanhui Zhang, Gang Peng</i>	
<b>Discovering Language in Marmoset Vocalization</b> .....	2426
<i>Sakshi Verma, K. L. Prateek, Karthik Pandia, Nauman Dawalatabad, Rogier Landman, Jitendra Sharma, Mriganka Sur, Hema A. Murthy</i>	
<b>Subject-Independent Classification of Japanese Spoken Sentences by Multiple Frequency Bands Phase Pattern of EEG Response During Speech Perception</b> .....	2431
<i>Hiroki Watanabe, Hiroki Tanaka, Sakriani Sakti, Satoshi Nakamura</i>	
<b>The Phonological Status of the French Initial Accent and its Role in Semantic Processing: An Event-Related Potentials Study</b> .....	2436
<i>Noemie Te Rietmolen, Radouane El Yagoubi, Alain Ghio, Corine Astesano</i>	
<b>A Neuro-Experimental Evidence for the Motor Theory of Speech Perception</b> .....	2441
<i>Bin Zhao, Jianwu Dang, Gaoyan Zhang</i>	

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

<b>Speech Representation Learning Using Unsupervised Data-Driven Modulation Filtering for Robust ASR</b> .....	2446
<i>Purvi Agrawal, Sriram Ganapathy</i>	
<b>Combined Multi-Channel NMF-Based Robust Beamforming for Noisy Speech Recognition</b> .....	2451
<i>Masato Mimura, Yoshiaki Bando, Kazuki Shimada, Shinsuke Sakai, Kazuyoshi Yoshii, Tatsuya Kawahara</i>	
<b>Recognizing Multi-Talker Speech with Permutation Invariant Training</b> .....	2456
<i>Dong Yu, Xuankai Chang, Yanmin Qian</i>	

<b>Coupled Initialization of Multi-Channel Non-Negative Matrix Factorization Based on Spatial and Spectral Information</b> .....	2461
<i>Yuuki Tachioka, Tomohiro Narita, Iori Miura, Takanobu Uramoto, Natsuki Monta, Shingo Uenohara, Ken'Ichi Furuya, Shinji Watanabe, Jonathan Le Roux</i>	
<b>Channel Compensation in the Generalised Vector Taylor Series Approach to Robust ASR</b> .....	2466
<i>Erfan Loweimi, Jon Barker, Thomas Hain</i>	
<b>Robust Speech Recognition via Anchor Word Representations</b> .....	2471
<i>Brian King, I-Fan Chen, Yonatan Vaizman, Yuzong Liu, Roland Maas, Sree Hari Krishnan Parthasarathi, Bjorn Hoffmeister</i>	

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

<b>Towards Zero-Shot Frame Semantic Parsing for Domain Scaling</b> .....	2476
<i>Ankur Bapna, Gokhan Tur, Dilek Hakkani-Tur, Larry Heck</i>	
<b>Clockwork-RNN Based Architectures for Slot Filling</b> .....	2481
<i>Despoina Georgiadou, Vassilios Diakouloukas, Vassilios Tsiaras, Vassilios Digalakis</i>	
<b>Investigating the Effect of ASR Tuning on Named Entity Recognition</b> .....	2486
<i>Mohamed Ameer Ben Jannet, Olivier Galibert, Martine Adda-Decker, Sophie Rosset</i>	
<b>Label-Dependency Coding in Simple Recurrent Networks for Spoken Language Understanding</b> .....	2491
<i>Marco Dinarelli, Vedran Vukotic, Christian Raymond</i>	
<b>Minimum Semantic Error Cost Training of Deep Long Short-Term Memory Networks for Topic Spotting on Conversational Speech</b> .....	2496
<i>Zhong Meng, Bing-Hwang Juang</i>	
<b>Topic Identification for Speech Without ASR</b> .....	2501
<i>Chunxi Liu, Jan Trmal, Matthew Wiesner, Craig Harman, Sanjeev Khudanpur</i>	

#### **WED-O-7-6: DIALOG SYSTEMS**

<b>An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog</b> .....	2506
<i>Bing Liu, Ian Lane</i>	
<b>Deep Reinforcement Learning of Dialogue Policies with Less Weight Updates</b> .....	2511
<i>Heriberto Cuayahuitl, Seunghak Yu</i>	
<b>Towards End-to-End Spoken Dialogue Systems with Turn Embeddings</b> .....	2516
<i>Ali Orkan Bayer, Evgeny A. Stepanov, Giuseppe Riccardi</i>	
<b>Speech and Text Analysis for Multimodal Addressee Detection in Human-Human-Computer Interaction</b> .....	2521
<i>Oleg Akhltamov, Maxim Sidorov, Alexey A. Karpov, Wolfgang Minker</i>	
<b>Rushing to Judgement: How Do Laypeople Rate Caller Engagement in Thin-Slice Videos of Human-Machine Dialog?</b> .....	2526
<i>Vikram Ramanarayanan, Chee Wee Leong, David Suendermann-Oeft</i>	
<b>Hyperarticulation of Corrections in Multilingual Dialogue Systems</b> .....	2531
<i>Ivan Kraljevski, Diane Hirschfeld</i>	

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

<b>Multitask Sequence-to-Sequence Models for Grapheme-to-Phoneme Conversion</b> .....	2536
<i>Benjamin Milde, Christoph Schmidt, Joachim Kohler</i>	
<b>Acoustic Data-Driven Lexicon Learning Based on a Greedy Pronunciation Selection Framework</b> .....	2541
<i>Xiaohui Zhang, Vimal Manohar, Daniel Povey, Sanjeev Khudanpur</i>	
<b>Semi-Supervised Learning of a Pronunciation Dictionary from Disjoint Phonemic Transcripts and Text</b> .....	2546
<i>Takahiro Shinozaki, Shinji Watanabe, Daichi Mochihashi, Graham Neubig</i>	
<b>Improved Subword Modeling for WFST-Based Speech Recognition</b> .....	2551
<i>Peter Smit, Sami Virpioja, Mikko Kurimo</i>	
<b>Pronunciation Learning with RNN-Transducers</b> .....	2556
<i>Antoine Bruguier, Danushen Gnanpragasam, Leif Johnson, Kanishka Rao, Françoise Beaufays</i>	
<b>Learning Similarity Functions for Pronunciation Variations</b> .....	2561
<i>Einat Naaman, Yossi Adi, Joseph Keshet</i>	

#### **WED-O-7-10: LANGUAGE RECOGNITION**

<b>Spoken Language Identification Using LSTM-Based Angular Proximity</b> .....	2566
<i>G. Gelly, J. L. Gauvain</i>	
<b>End-to-End Language Identification Using High-Order Utterance Representation with Bilinear Pooling</b> .....	2571
<i>Ma Jin, Yan Song, Ian McLoughlin, Wu Guo, Li-Rong Dai</i>	
<b>Dialect Recognition Based on Unsupervised Bottleneck Features</b> .....	2576
<i>Qian Zhang, John H. L. Hansen</i>	
<b>Investigating Scalability in Hierarchical Language Identification System</b> .....	2581
<i>Saad Irta, Vidhyasaharan Sethu, Eliathamby Ambikairajah, Haizhou Li</i>	
<b>Improving Sub-Phone Modeling for Better Native Language Identification with Non-Native English Speech</b> .....	2586
<i>Yao Qian, Keelan Evanini, Xinhao Wang, David Suendermann-Oeft, Robert A. Pugh, Patrick L. Lange, Hillary R. Molloy, Frank K. Soong</i>	

<b>QMDIS: QCRI-MIT Advanced Dialect Identification System</b> .....	2591
<i>Sameer Khurana, Maryam Najafian, Ahmed Ali, Tuka Al Hanai, Yonatan Belinkov, James Glass</i>	

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

<b>Detection of Replay Attacks Using Single Frequency Filtering Cepstral Coefficients</b> .....	2596
<i>K. N. R. K. Raju Alluri, Sivanand Achanta, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, Anil Kumar Vuppala</i>	
<b>Unsupervised Representation Learning Using Convolutional Restricted Boltzmann Machine for Spoof Speech Detection</b> .....	2601
<i>Hardik B. Sailor, Madhu R. Kamble, Hemant A. Patil</i>	
<b>Independent Modelling of High and Low Energy Speech Frames for Spoofing Detection</b> .....	2606
<i>Gajan Suthokumar, Kaavya Sriskandaraja, Vidhyasaharan Sethu, Chamith Wijenayake, Eliathamby Ambikarajah</i>	
<b>Improving Speaker Verification Performance in Presence of Spoofing Attacks Using Out-of-Domain Spoofed Data</b> .....	2611
<i>Achintya Kr. Sarkar, Md. Sahidullah, Zheng-Hua Tan, Tomi Kinnunen</i>	
<b>VoxCeleb: A Large-Scale Speaker Identification Dataset</b> .....	2616
<i>Arsha Nagrani, Joon Son Chung, Andrew Senior</i>	
<b>Call My Net Corpus: A Multilingual Corpus for Evaluation of Speaker Recognition Technology</b> .....	2621
<i>Karen Jones, Stephanie Strassel, Kevin Walker, David Graff, Jonathan Wright</i>	

### **WED-O-8-4: SPEECH TRANSLATION**

<b>Sequence-to-Sequence Models Can Directly Translate Foreign Speech</b> .....	2625
<i>Ron J. Weiss, Jan Chorowski, Navdeep Jaitly, Yonghui Wu, Zhifeng Chen</i>	
<b>Structured-Based Curriculum Learning for End-to-End English-Japanese Speech Translation</b> .....	2630
<i>Takatomo Kano, Sakriani Sakti, Satoshi Nakamura</i>	
<b>Assessing the Tolerance of Neural Machine Translation Systems Against Speech Recognition Errors</b> .....	2635
<i>Nicholas Ruiz, Mattia Antonino Di Gangi, Nicola Bertoldi, Marcello Federico</i>	
<b>Toward Expressive Speech Translation: A Unified Sequence-to-Sequence LSTMs Approach for Translating Words and Emphasis</b> .....	2640
<i>Quoc Truong Do, Sakriani Sakti, Satoshi Nakamura</i>	
<b>NMT-Based Segmentation and Punctuation Insertion for Real-Time Spoken Language Translation</b> .....	2645
<i>Eunah Cho, Jan Niehues, Alex Waibel</i>	

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

<b>Tight Integration of Spatial and Spectral Features for BSS with Deep Clustering Embeddings</b> .....	2650
<i>Lukas Drude, Reinhold Haeb-Umbach</i>	
<b>Speaker-Aware Neural Network Based Beamformer for Speaker Extraction in Speech Mixtures</b> .....	2655
<i>Katerina Zmolikova, Marc Delcroix, Keisuke Kinoshita, Takuya Higuchi, Atsunori Ogawa, Tomohiro Nakatani</i>	
<b>Eigenvector-Based Speech Mask Estimation Using Logistic Regression</b> .....	2660
<i>Lukas Pfeifenberger, Matthias Zohrer, Franz Pernkopf</i>	
<b>Real-Time Speech Enhancement with GCC-NMF</b> .....	2665
<i>Sean U. N. Wood, Jean Rouat</i>	
<b>Coherence-Based Dual-Channel Noise Reduction Algorithm in a Complex Noisy Environment</b> .....	2670
<i>Youna Ji, Jun Byun, Young-Cheol Park</i>	
<b>Glottal Model Based Speech Beamforming for Ad-Hoc Microphone Arrays</b> .....	2675
<i>Yang Zhang, Dinei Florencio, Mark Hasegawa-Johnson</i>	

## VOLUME 5

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

<b>Acoustic Assessment of Disordered Voice with Continuous Speech Based on Utterance-Level ASR Posterior Features</b> .....	2680
<i>Yuanyuan Liu, Tan Lee, P. C. Ching, Thomas K. T. Law, Kathy Y. S. Lee</i>	
<b>Multi-Stage DNN Training for Automatic Recognition of Dysarthric Speech</b> .....	2685
<i>Emre Yilmaz, Mario Ganzeboom, Catia Cucchiari, Helmer Strik</i>	
<b>Improving Child Speech Disorder Assessment by Incorporating Out-of-Domain Adult Speech</b> .....	2690
<i>Daniel Smith, Alex Sneddon, Lauren Ward, Andreas Duenser, Jill Freyne, David Silvera-Tawil, Angela Morgan</i>	
<b>On Improving Acoustic Models for TORGO Dysarthric Speech Database</b> .....	2695
<i>Neethu Mariam Joy, S. Umesh, Basil Abraham</i>	
<b>Glottal Source Features for Automatic Speech-Based Depression Assessment</b> .....	2700
<i>Olympia Simantiraki, Paulos Charonyktakis, Anastasia Pampouchidou, Manolis Tsiknakis, Martin Cooke</i>	
<b>Speech Processing Approach for Diagnosing Dementia in an Early Stage</b> .....	2705
<i>Roohbeh Sadeghian, J. David Schaffer, Stephen A. Zahorian</i>	



## WED-O-8-10: LANGUAGE MODELS FOR ASR

<b>Effectively Building Tera Scale MaxEnt Language Models Incorporating Non-Linguistic Signals</b> .....	2710
<i>Fadi Biadisy, Mohammadreza Ghodsi, Diamantino Caseiro</i>	
<b>Semi-Supervised Adaptation of RNNLMs by Fine-Tuning with Domain-Specific Auxiliary Features</b> .....	2715
<i>Salil Deena, Raymond W. M. Ng, Pranava Madhyastha, Lucia Specia, Thomas Hain</i>	
<b>Approximated and Domain-Adapted LSTM Language Models for First-Pass Decoding in Speech Recognition</b> .....	2720
<i>Mittul Singh, Youssef Oualil, Dietrich Klakow</i>	
<b>Sparse Non-Negative Matrix Language Modeling: Maximum Entropy Flexibility on the Cheap</b> .....	2725
<i>Ciprian Chelba, Diamantino Caseiro, Fadi Biadisy</i>	
<b>Multi-Scale Context Adaptation for Improving Child Automatic Speech Recognition in Child-Adult Spoken Interactions</b> .....	2730
<i>Manoj Kumar, Daniel Bone, Kelly McWilliams, Shanna Williams, Thomas D. Lyon, Shrikanth S. Narayanan</i>	
<b>Using Knowledge Graph and Search Query Click Logs in Statistical Language Model for Speech Recognition</b> .....	2735
<i>Weiwu Zhu</i>	

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

<b>Developing On-Line Speaker Diarization System</b> .....	2739
<i>Dimitrios Dimitriadis, Petr Fousek</i>	
<b>Comparison of Non-Parametric Bayesian Mixture Models for Syllable Clustering and Zero-Resource Speech Processing</b> .....	2744
<i>Shreyas Seshadri, Ulpu Remes, Okko Rasanen</i>	
<b>Automatic Evaluation of Children Reading Aloud on Sentences and Pseudowords</b> .....	2749
<i>Jorge Proenca, Carla Lopes, Michael Tjalve, Andreas Stolcke, Sara Candeias, Fernando Perdigao</i>	
<b>Off-Topic Spoken Response Detection with Word Embeddings</b> .....	2754
<i>Su-Youn Yoon, Chong Min Lee, Ikkyu Choi, Xinhao Wang, Matthew Mulholland, Keelan Evanini</i>	
<b>Improving Mispronunciation Detection for Non-Native Learners with Multisource Information and LSTM-Based Deep Models</b> .....	2759
<i>Wei Li, Nancy F. Chen, Sabato Marco Siniscalchi, Chin-Hui Lee</i>	
<b>Automatic Explanation Spot Estimation Method Targeted at Text and Figures in Lecture Slides</b> .....	2764
<i>Shoko Tsujimura, Kazumasa Yamamoto, Seiichi Nakagawa</i>	
<b>Multiview Representation Learning via Deep CCA for Silent Speech Recognition</b> .....	2769
<i>Myungjong Kim, Beiming Cao, Ted Mau, Jun Wang</i>	
<b>Use of Graphemic Lexicons for Spoken Language Assessment</b> .....	2774
<i>K. M. Knill, Mark J. F. Gales, K. Kyriakopoulos, A. Ragni, Y. Wang</i>	
<b>Distilling Knowledge From an Ensemble of Models for Punctuation Prediction</b> .....	2779
<i>Jiangyan Yi, Jianhua Tao, Zhengqi Wen, Ya Li</i>	
<b>A Mostly Data-Driven Approach to Inverse Text Normalization</b> .....	2784
<i>Ernest Puserari, Bharat Ram Ambati, Elizabeth Brooks, Ondrej Plátek, Donald McAllaster, Venki Nagesha</i>	
<b>Mismatched Crowdsourcing from Multiple Annotator Languages for Recognizing Zero-Resourced Languages: A Nullspace Clustering Approach</b> .....	2789
<i>Wenda Chen, Mark Hasegawa-Johnson, Nancy F. Chen, Boon Pang Lim</i>	
<b>Experiments in Character-Level Neural Network Models for Punctuation</b> .....	2794
<i>William Gale, Sarangarajan Parthasarathy</i>	
<b>Multi-Channel Apollo Mission Speech Transcripts Calibration</b> .....	2799
<i>Lakshmi Kaushik, Abhijeet Sangwan, John H. L. Hansen</i>	

## WED-P-6-2: SPEAKER AND LANGUAGE RECOGNITION APPLICATIONS

<b>Calibration Approaches for Language Detection</b> .....	2804
<i>Mitchell McLaren, Luciana Ferrer, Diego Castan, Aaron Lawson</i>	
<b>Bidirectional Modelling for Short Duration Language Identification</b> .....	2809
<i>Sarith Fernando, Vidhyasaharan Sethu, Eliathamby Ambikairajah, Julien Epps</i>	
<b>Conditional Generative Adversarial Nets Classifier for Spoken Language Identification</b> .....	2814
<i>Peng Shen, Xugang Lu, Sheng Li, Hisashi Kawai</i>	
<b>Tied Hidden Factors in Neural Networks for End-to-End Speaker Recognition</b> .....	2819
<i>Antonio Miguel, Jorge Lombart, Alfonso Ortega, Eduardo Lleida</i>	
<b>Speaker Clustering by Iteratively Finding Discriminative Feature Space and Cluster Labels</b> .....	2824
<i>Sungrack Yum, Hye Jin Jang, Taesu Kim</i>	
<b>Domain Adaptation of PLDA Models in Broadcast Diarization by Means of Unsupervised Speaker Clustering</b> .....	2829
<i>Ignacio Vinals, Alfonso Ortega, Jesus Villalba, Antonio Miguel, Eduardo Lleida</i>	
<b>LSTM Neural Network-Based Speaker Segmentation Using Acoustic and Language Modelling</b> .....	2834
<i>Miquel India, Jose A. R. Fonollosa, Javier Hernandez</i>	
<b>Acoustic Pairing of Original and Dubbed Voices in the Context of Video Game Localization</b> .....	2839
<i>Adrien Gresse, Mickael Rouvier, Richard Dufour, Vincent Labatut, Jean-Francois Bonastre</i>	
<b>Homogeneity Measure Impact on Target and Non-Target Trials in Forensic Voice Comparison</b> .....	2844
<i>Moez Ajili, Jean-Francois Bonastre, Waad Ben Kheder, Solange Rossato, Juliette Kahn</i>	

<b>Null-Hypothesis LLR: A Proposal for Forensic Automatic Speaker Recognition</b> .....	2849
<i>Yosef A. Solewicz, Michael Jessen, David Van Der Vloed</i>	
<b>The Opensesame NIST 2016 Speaker Recognition Evaluation System</b> .....	2854
<i>Gang Liu, Qi Qian, Zhibin Wang, Qingen Zhao, Tianzhou Wang, Hao Li, Jian Xue, Shenghuo Zhu, Rong Jin, Tuo Zhao</i>	
<b>IITG-Indigo System for NIST 2016 SRE Challenge</b> .....	2859
<i>Nagendra Kumar, Rohan Kumar Das, Sarfaraz Jelil, Dhanush B. K., H. Kashyap, K. Sri Rama Murty, Sriram Ganapathy, Rohit Sinha, S. R. Mahadeva Prasanna</i>	
<b>Locally Weighted Linear Discriminant Analysis for Robust Speaker Verification</b> .....	2864
<i>Abhinav Misra, Shivesh Ranjan, John H. L. Hansen</i>	
<b>Recursive Whitening Transformation for Speaker Recognition on Language Mismatched Condition</b> .....	2869
<i>Suwon Shon, Seongkyu Mun, Hanseok Ko</i>	

### **WED-P-6-3: SPOKEN DOCUMENT PROCESSING**

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

### **WED-P-6-4: SPEECH INTELLIGIBILITY**

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

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

<b>Mental Representation of Japanese Mora; Focusing on its Intrinsic Duration</b> .....	2973
<i>Kosuke Sugai</i>	
<b>Temporal Dynamics of Lateral Channel Formation in /l/: 3D EMA Data from Australian English</b> .....	2978
<i>Jia Ying, Christopher Carignan, Jason A. Shaw, Michael Proctor, Donald Derrick, Catherine T. Best</i>	

<b>Vowel and Consonant Sequences in Three Bavarian Dialects of Austria</b> .....	2983
<i>Nicola Klingler, Sylvia Moosmuller, Hannes Scheutz</i>	
<b>Acoustic Cues to the Singleton-Geminate Contrast: The Case of Libyan Arabic Sonorants</b> .....	2988
<i>Amel Issa</i>	
<b>Mel-cepstral Distortion of German Vowels in Different Information Density Contexts</b> .....	2993
<i>Erika Brandt, Frank Zimmerer, Bistra Andreeva, Bernd Mobius</i>	
<b>Effect of Formant and F0 Discontinuity on Perceived Vowel Duration: Impacts for Concatenative Speech Synthesis</b> .....	2998
<i>Tomas Boril, Pavel Sturm, Radek Skarnitzl, Jan Volin</i>	
<b>An Ultrasound Study of Alveolar and Retroflex Consonants in Arrernte: Stressed and Unstressed Syllables</b> .....	3003
<i>Marija Tabain, Richard Beare</i>	
<b>Reshaping the Transformed LF Model: Generating the Glottal Source from the Waveshape Parameter <math>R_d</math></b> .....	3008
<i>Christer Gobl</i>	
<b>Kinematic Signatures of Prosody in Lombard Speech</b> .....	3013
<i>Stefan Benus, Juraj Simko, Mona Lehtinen</i>	
<b>What do Finnish and Central Bavarian Have in Common? Towards an Acoustically Based Quantity Typology</b> .....	3018
<i>Markus Jochim, Felicitas Kleber</i>	
<b>Locating Burst Onsets Using SFF Envelope and Phase Information</b> .....	3023
<i>Bhanu Teja Nellore, Ravishankar Prasad, Sudarsana Reddy Kadiri, Suryakanth V. Gangashetty, B. Yegnanarayana</i>	
<b>A Preliminary Phonetic Investigation of Alphabetic Words in Mandarin Chinese</b> .....	3028
<i>Hongwei Ding, Yuan Yuan Zhang, Hongchao Liu, Chu-Ren Huang</i>	
<b>A Quantitative Measure of the Impact of Coarticulation on Phone Discriminability</b> .....	3033
<i>Thomas Schatz, Rory Turnbull, Francis Bach, Emmanuel Dupoux</i>	

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

<b>Sinusoidal Partial Tracking for Singing Analysis Using the Heuristic of the Minimal Frequency and Magnitude Difference</b> .....	3038
<i>Kin Wah Edward Lin, Hans Anderson, Clifford So, Simon Lui</i>	
<b>Audio Scene Classification with Deep Recurrent Neural Networks</b> .....	3043
<i>Huy Phan, Philipp Koch, Fabrice Katzberg, Marco Maass, Radoslaw Mazur, Alfred Mertins</i>	
<b>Automatic Time-Frequency Analysis of Echolocation Signals Using the Matched Gaussian Multitaper Spectrogram</b> .....	3048
<i>Maria Sandsten, Isabella Reinhold, Josefín Starkhammar</i>	
<b>Classification-Based Detection of Glottal Closure Instants from Speech Signals</b> .....	3053
<i>Jindrich Matousek, Daniel Tihelka</i>	
<b>A Domain Knowledge-Assisted Nonlinear Model for Head-Related Transfer Functions Based on Bottleneck Deep Neural Network</b> .....	3058
<i>Xiaoke Qi, Jianhua Tao</i>	
<b>Laryngeal Articulation During Trumpet Performance: An Exploratory Study</b> .....	3063
<i>Luis M. T. Jesus, Bruno Rocha, Andreia Hall</i>	
<b>Matrix of Polynomials Model Based Polynomial Dictionary Learning Method for Acoustic Impulse Response Modeling</b> .....	3068
<i>Jian Guan, Xuan Wang, Pengming Feng, Jing Dong, Wenwu Wang</i>	
<b>Acoustic Scene Classification Using a CNN-SuperVector System Trained with Auditory and Spectrogram Image Features</b> .....	3073
<i>Rakib Hyder, Shabnam Ghaffarzadegan, Zhe Feng, John H. L. Hansen, Taufiq Hasan</i>	
<b>An Environmental Feature Representation for Robust Speech Recognition and for Environment Identification</b> .....	3078
<i>Xue Feng, Brigitte Richardson, Scott Amman, James Glass</i>	
<b>Attention and Localization Based on a Deep Convolutional Recurrent Model for Weakly Supervised Audio Tagging</b> .....	3083
<i>Yong Xu, Qiuqiang Kong, Qiang Huang, Wenwu Wang, Mark D. Plumbley</i>	
<b>An Audio Based Piano Performance Evaluation Method Using Deep Neural Network Based Acoustic Modeling</b> .....	3088
<i>Jing Pan, Ming Li, Zhanmei Song, Xin Li, Xiaolin Liu, Hua Yi, Manman Zhu</i>	
<b>Music Tempo Estimation Using Sub-Band Synchrony</b> .....	3093
<i>Shreyan Chowdhury, Tanaya Guha, Rajesh M. Hegde</i>	
<b>A Transfer Learning Based Feature Extractor for Polyphonic Sound Event Detection Using Connectionist Temporal Classification</b> .....	3097
<i>Yun Wang, Florian Metze</i>	
<b>A Note Based Query By Humming System Using Convolutional Neural Network</b> .....	3102
<i>Naziba Mostafa, Pascale Fung</i>	
<b>Unsupervised Filterbank Learning Using Convolutional Restricted Boltzmann Machine for Environmental Sound Classification</b> .....	3107
<i>Hardik B. Sailor, Dharmesh M. Agrawal, Hemant A. Patil</i>	
<b>Novel Shifted Real Spectrum for Exact Signal Reconstruction</b> .....	3112
<i>Meet H. Soni, Rishabh Tak, Hemant A. Patil</i>	

## **WED-P-7-4: DISORDERS RELATED TO SPEECH AND LANGUAGE**

<b>Manual and Automatic Transcriptions in Dementia Detection from Speech</b> .....	3117
<i>Jochen Weiner, Mathis Engelbart, Tanja Schultz</i>	
<b>An Affect Prediction Approach Through Depression Severity Parameter Incorporation in Neural Networks</b> .....	3122
<i>Rahul Gupta, Saurabh Sahu, Carol Espy-Wilson, Shrikanth S. Narayanan</i>	
<b>Cross-Database Models for the Classification of Dysarthria Presence</b> .....	3127
<i>Stephanie Gillespie, Yash-Yee Logan, Elliot Moore, Jacqueline Laures-Gore, Scott Russell, Rupal Patel</i>	
<b>Acoustic Evaluation of Nasality in Cerebellar Syndromes</b> .....	3132
<i>M. Novotny, Jan Ruzs, K. Spalenka, Jiri Klempir, D. Horakova, Evzen Ruzicka</i>	
<b>Emotional Speech of Mentally and Physically Disabled Individuals: Introducing the EmotAsS Database and First Findings</b> .....	3137
<i>Simone Hantke, Hesam Sagha, Nicholas Cummins, Bjorn Schuller</i>	
<b>Phonological Markers of Oxytocin and MDMA Ingestion</b> .....	3142
<i>Carla Agurto, Raquel Norel, Rachel Ostrand, Gillinder Bedi, Harriet De Wit, Matthew J. Baggott, Matthew G. Kirkpatrick, Margaret Wardle, Guillermo A. Cecchi</i>	
<b>An Avatar-Based System for Identifying Individuals Likely to Develop Dementia</b> .....	3147
<i>Bahman Mirheidari, Daniel Blackburn, Kirsty Harkness, Traci Walker, Annalena Venneri, Markus Reuber, Heidi Christensen</i>	
<b>Cross-Domain Classification of Drowsiness in Speech: The Case of Alcohol Intoxication and Sleep Deprivation</b> .....	3152
<i>Yue Zhang, Felix Weninger, Bjorn Schuller</i>	
<b>Depression Detection Using Automatic Transcriptions of De-Identified Speech</b> .....	3157
<i>Paula Lopez-Otero, Laura Docio-Fernandez, Alberto Abad, Carmen Garcia-Mateo</i>	
<b>An N-Gram Based Approach to the Automatic Diagnosis of Alzheimer's Disease from Spoken Language</b> .....	3162
<i>Sebastian Wankler, Elmar Noth, Stefan Evert</i>	
<b>Exploiting Intra-Annotator Rating Consistency Through Copeland's Method for Estimation of Ground Truth Labels in Couples' Therapy</b> .....	3167
<i>Karel Mundnich, Md. Nasir, Panayiotis Georgiou, Shrikanth S. Narayanan</i>	
<b>Rhythmic Characteristics of Parkinsonian Speech: A Study on Mandarin and Polish</b> .....	3172
<i>Massimo Pettorino, Wentao Gu, Pawel Polrola, Ping Fan</i>	

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

<b>Trisyllabic Tone 3 Sandhi Patterns in Mandarin Produced by Cantonese Speakers</b> .....	3177
<i>Jung-Yueh Tu, Janice Wing-Sze Wong, Jih-Ho Cha</i>	
<b>Intonation of Contrastive Topic in Estonian</b> .....	3181
<i>Heete Sakkai, Meelis Mihkla</i>	
<b>Reanalyze Fundamental Frequency Peak Delay in Mandarin</b> .....	3186
<i>Lixia Hao, Wei Zhang, Yanlu Xie, Jinsong Zhang</i>	
<b>How Does the Absence of Shared Knowledge Between Interlocutors Affect the Production of French Prosodic Forms?</b> .....	3191
<i>Amandine Michelas, Cecile Cau, Maud Champagne-Lavau</i>	
<b>Three Dimensions of Sentence Prosody and Their (Non-)Interactions</b> .....	3196
<i>Michael Wagner, Michael McAuliffe</i>	
<b>Using Prosody to Classify Discourse Relations</b> .....	3201
<i>Janine Kleinhans, Mireia Farrus, Agustin Gravano, Juan Manuel Perez, Catherine Lai, Leo Wanner</i>	
<b>Canonical Correlation Analysis and Prediction of Perceived Rhythmic Prominences and Pitch Tones in Speech</b> .....	3206
<i>Elizabeth Godoy, James R. Williamson, Thomas F. Quatieri</i>	
<b>Evaluation of Spectral Tilt Measures for Sentence Prominence Under Different Noise Conditions</b> .....	3211
<i>Sofoklis Kakouros, Okko Rasanen, Paavo Alku</i>	
<b>Creaky Voice as a Function of Tonal Categories and Prosodic Boundaries</b> .....	3216
<i>Jianjing Kuang</i>	
<b>The Acoustics of Word Stress in Czech as a Function of Speaking Style</b> .....	3221
<i>Radek Skarnitzl, Anders Eriksson</i>	
<b>What You See is What You Get Prosodically Less --- Visibility Shapes Prosodic Prominence Production in Spontaneous Interaction</b> .....	3226
<i>Petra Wagner, Nataliya Bryhadyr</i>	
<b>Focus Acoustics in Mandarin Nominals</b> .....	3231
<i>Yu-Yin Hsu, Anqi Xu</i>	
<b>Exploring Multidimensionality: Acoustic and Articulatory Correlates of Swedish Word Accents</b> .....	3236
<i>Malin Svensson Lundmark, Gilbert Ambrazaitis, Otto Ewald</i>	
<b>The Perception of English Intonation Patterns by German L2 Speakers of English</b> .....	3241
<i>Karin Puga, Robert Fuchs, Jane Setter, Peggy Mok</i>	

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

<b>The Perception of Emotions in Noisified Nonsense Speech</b> .....	3246
<i>Emilia Parada-Cabaleiro, Alice Baird, Anton Batliner, Nicholas Cummins, Simone Hantke, Bjorn Schuller</i>	
<b>Attention Networks for Modeling Behaviors in Addiction Counseling</b> .....	3251
<i>James Gibson, Dogan Can, Panayiotis Georgiou, David C. Atkins, Shrikanth S. Narayanan</i>	

<b>Computational Analysis of Acoustic Descriptors in Psychotic Patients</b> .....	3256
<i>Torsten Wortwein, Tadas Baltrusaitis, Eugene Laksana, Luciana Pennant, Elizabeth S. Liebson, Dost Ongur, Justin T. Baker, Louis-Philippe Morency</i>	
<b>Modeling Perceivers Neural-Responses Using Lobe-Dependent Convolutional Neural Network to Improve Speech Emotion Recognition</b> .....	3261
<i>Ya-Tse Wu, Hsuan-Yu Chen, Yu-Hsien Liao, Li-Wei Kuo, Chi-Chun Lee</i>	
<b>Implementing Gender-Dependent Vowel-Level Analysis for Boosting Speech-based Depression Recognition</b> .....	3266
<i>Bogdan Vlasenko, Hesam Sagha, Nicholas Cummins, Bjorn Schuller</i>	
<b>Bilingual Word Embeddings for Cross-Lingual Personality Recognition Using Convolutional Neural Nets</b> .....	3271
<i>Farhad Bin Siddique, Pascale Fung</i>	
<b>Emotion Category Mapping to Emotional Space by Cross-Corpus Emotion Labeling</b> .....	3276
<i>Yoshiko Arimoto, Hiroki Mori</i>	
<b>Big Five vs. Prosodic Features as Cues to Detect Abnormality in SSPNET-Personality Corpus</b> .....	3281
<i>Cedric Fayet, Arnaud Delhay, Damien Lolive, Pierre-Francois Marteau</i>	
<b>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</b> .....	3286
<i>Hayakawa Akira, Carl Vogel, Saturnino Luz, Nick Campbell</i>	
<b>Approaching Human Performance in Behavior Estimation in Couples Therapy Using Deep Sentence Embeddings</b> .....	3291
<i>Shao-Yen Tseng, Brian Baucom, Panayiotis Georgiou</i>	
<b>Complexity in Speech and its Relation to Emotional Bond in Therapist-Patient Interactions During Suicide Risk Assessment Interviews</b> .....	3296
<i>Md. Nasir, Brian Baucom, Craig J. Bryan, Shrikanth S. Narayanan, Panayiotis Georgiou</i>	
<b>An Investigation of Emotion Dynamics and Kalman Filtering for Speech-based Emotion Prediction</b> .....	3301
<i>Zhaocheng Huang, Julien Epps</i>	

### **WED-P-8-3: LANGUAGE UNDERSTANDING AND GENERATION**

<b>Zero-Shot Learning for Natural Language Understanding Using Domain-Independent Sequential Structure and Question Types</b> .....	3306
<i>Kugatsu Sadamitsu, Yukinori Homma, Ryuichiro Higashinaka, Yoshihiro Matsuo</i>	
<b>Parallel Hierarchical Attention Networks with Shared Memory Reader for Multi-Stream Conversational Document Classification</b> .....	3311
<i>Naoki Sawada, Ryo Masumura, Hiromitsu Nishizaki</i>	
<b>Internal Memory Gate for Recurrent Neural Networks with Application to Spoken Language Understanding</b> .....	3316
<i>Mohamed Morchid</i>	
<b>Character-Based Embedding Models and Reranking Strategies for Understanding Natural Language Meal Descriptions</b> .....	3320
<i>Mandy Korpusik, Zachary Collins, James Glass</i>	
<b>Quaternion Denoising Encoder-Decoder for Theme Identification of Telephone Conversations</b> .....	3325
<i>Titouan Parcollet, Mohamed Morchid, Georges Linares</i>	
<b>ASR Error Management for Improving Spoken Language Understanding</b> .....	3329
<i>Edwin Simonnet, Sahar Ghannay, Nathalie Camelin, Yannick Esteve, Renato De Mori</i>	
<b>Jointly Trained Sequential Labeling and Classification by Sparse Attention Neural Networks</b> .....	3334
<i>Mingbo Ma, Kai Zhao, Liang Huang, Bing Xiang, Bowen Zhou</i>	
<b>To Plan or Not to Plan? Discourse Planning in Slot-Value Informed Sequence to Sequence Models for Language Generation</b> .....	3339
<i>Neha Nayak, Dilek Hakkani-Tur, Marilyn Walker, Larry Heck</i>	
<b>Online Adaptation of an Attention-Based Neural Network for Natural Language Generation</b> .....	3344
<i>Mathieu Riou, Bassam Jabaian, Stephane Huet, Fabrice Lefevre</i>	
<b>Spanish Sign Language Recognition with Different Topology Hidden Markov Models</b> .....	3349
<i>Carlos-D. Martinez-Hinarejos, Zuzanna Parcheta</i>	
<b>OpenMM: An Open-Source Multimodal Feature Extraction Tool</b> .....	3354
<i>Michelle Renee Morales, Stefan Scherer, Rivka Levitan</i>	

### **VOLUME 6**

<b>Speaker Dependency Analysis, Audiovisual Fusion Cues and a Multimodal BLSTM for Conversational Engagement Recognition</b> .....	3359
<i>Yuyun Huang, Emer Gilmartin, Nick Campbell</i>	

### **WED-P-8-4: VOICE CONVERSION 2**

<b>Voice Conversion from Unaligned Corpora Using Variational Autoencoding Wasserstein Generative Adversarial Networks</b> .....	3364
<i>Chin-Cheng Hsu, Hsin-Te Hwang, Yi-Chiao Wu, Yu Tsao, Hsin-Min Wang</i>	
<b>CAB: An Energy-Based Speaker Clustering Model for Rapid Adaptation in Non-Parallel Voice Conversion</b> .....	3369
<i>Toru Nakashika</i>	
<b>Phoneme-Discriminative Features for Dysarthric Speech Conversion</b> .....	3374
<i>Ryo Aihara, Tetsuya Takiguchi, Yasuo Ariki</i>	

<b>Denosing Recurrent Neural Network for Deep Bidirectional LSTM Based Voice Conversion</b> .....	3379
<i>Jie Wu, D.-Y. Huang, Lei Xie, Haizhou Li</i>	
<b>Speaker Dependent Approach for Enhancing a Glossectomy Patient's Speech via GMM-Based Voice Conversion</b> .....	3384
<i>Kei Tanaka, Sunao Hara, Masanobu Abe, Masaaki Sato, Shogo Minagi</i>	
<b>Generative Adversarial Network-Based Postfilter for STFT Spectrograms</b> .....	3389
<i>Takuhiro Kaneko, Shinji Takaki, Hirokazu Kameoka, Junichi Yamagishi</i>	
<b>Generative Adversarial Network-Based Glottal Waveform Model for Statistical Parametric Speech Synthesis</b> .....	3394
<i>Bajibabu Bollepalli, Lauri Juvela, Paavo Alku</i>	
<b>Emotional Voice Conversion with Adaptive Scales F0 Based on Wavelet Transform Using Limited Amount of Emotional Data</b> .....	3399
<i>Zhaojie Luo, Jinhui Chen, Tetsuya Takiguchi, Yasuo Arikki</i>	
<b>Speaker Adaptation in DNN-Based Speech Synthesis Using d-Vectors</b> .....	3404
<i>Rama Doddipatla, Norbert Braunschweiler, Rannieri Maia</i>	
<b>Spectro-Temporal Modelling with Time-Frequency LSTM and Structured Output Layer for Voice Conversion</b> .....	3409
<i>Ruman Li, Zhiyong Wu, Yishuang Ning, Lifa Sun, Helen Meng, Lianhong Cai</i>	
<b>Segment Level Voice Conversion with Recurrent Neural Networks</b> .....	3414
<i>Miguel Varela Ramos, Alan W. Black, Ramon Fernandez Astudillo, Isabel Trancoso, Nuno Fonseca</i>	

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

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

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

<b>Integrating the Talkamatic Dialogue Manager with Alexa</b> .....	3431
<i>Staffan Larsson, Alex Berman, Andreas Krona, Fredrik Kronlid</i>	
<b>A Robust Medical Speech-to-Speech/Speech-to-Sign Phraselator</b> .....	3433
<i>Farhia Ahmed, Pierrette Bouillon, Chelle Destefano, Johanna Gerlach, Sonia Halimi, Angela Hooper, Manny Rayner, Herve Spechbach, Irene Strasly, Nikos Tsourakis</i>	
<b>Towards an Autarkic Embedded Cognitive User Interface</b> .....	3435
<i>Frank Duckhorn, Markus Huber, Werner Meyer, Oliver Jokisch, Constanze Tschope, Matthias Wolff</i>	
<b>Nora the Empathetic Psychologist</b> .....	3437
<i>Genta Indra Winata, Onno Kampman, Yang Yang, Anik Dey, Pascale Fung</i>	
<b>Modifying Amazon's Alexa ASR Grammar and Lexicon --- A Case Study</b> .....	3439
<i>Hassan Alam, Aman Kumar, Manan Vyas, Tina Werner, Rachmat Hartono</i>	

### **KEYNOTE 3: BJORN LINDBLOM**

<b>Re-Inventing Speech --- The Biological Way</b> .....	3441
<i>Bjorn Lindblom</i>	

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

<b>The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addressee, Cold &amp; Snoring</b> .....	3442
<i>Bjorn Schuller, Stefan Steidl, Anton Batliner, Erika Bergelson, Jarek Krajewski, Christoph Janott, Andrei Amatu, Marisa Casillas, Amanda Seidl, Melanie Soderstrom, Anne S. Warlaumont, Guillermo Hidalgo, Sebastian Schlieder, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Maximilian Schmitt, Kun Qian, Yue Zhang, George Trigeorgis, Panagiotis Tzirakis, Stefanos Zafeiriou</i>	
<b>It Sounds Like You Have a Cold! Testing Voice Features for the Interspeech 2017 Computational Paralinguistics Cold Challenge</b> .....	3447
<i>Mark Huckvale, Andras Beke</i>	
<b>End-to-End Deep Learning Framework for Speech Paralinguistics Detection Based on Perception Aware Spectrum</b> .....	3452
<i>Danwei Cai, Zhidong Ni, Wenbo Liu, Weicheng Cai, Gang Li, Ming Li</i>	
<b>Infected Phonemes: How a Cold Impairs Speech on a Phonetic Level</b> .....	3457
<i>Johannes Wagner, Thiago Fraga-Silva, Yvan Josse, Dominik Schiller, Andreas Seiderer, Elisabeth Andre</i>	

<b>Phoneme State Posteriorgram Features for Speech Based Automatic Classification of Speakers in Cold and Healthy Condition</b> .....	3462
<i>Akshay Kalkunte Suresh, K. M. Srinivasa Raghavan, Prasanta Kumar Ghosh</i>	
<b>An Integrated Solution for Snoring Sound Classification Using Bhattacharyya Distance Based GMM Supervectors with SVM, Feature Selection with Random Forest and Spectrogram with CNN</b> .....	3467
<i>Tin Lay Nwe, Huy Dat Tran, Wen Zheng Terence Ng, Bin Ma</i>	

### **THU-SS-9-11: SPECIAL SESSION: STATE OF THE ART IN PHYSICS-BASED VOICE SIMULATION**

<b>Acoustic Analysis of Detailed Three-Dimensional Shape of the Human Nasal Cavity and Paranasal Sinuses</b> .....	3472
<i>Tatsuya Kitamura, Hironori Takemoto, Hisanori Makinae, Tetsutaro Yamaguchi, Kotaro Maki</i>	
<b>A Semi-Polar Grid Strategy for the Three-Dimensional Finite Element Simulation of Vowel-Vowel Sequences</b> .....	3477
<i>Marc Arneta, Saeed Dabbaghchian, Oriol Guasch, Olov Engwall</i>	
<b>A Fast Robust 1D Flow Model for a Self-Oscillating Coupled 2D FEM Vocal Fold Simulation</b> .....	3482
<i>Arvind Vasudevan, Victor Zappi, Peter Anderson, Sidney Fels</i>	
<b>Waveform Patterns in Pitch Glides Near a Vocal Tract Resonance</b> .....	3487
<i>Tiina Murtola, Jarmo Malinen</i>	
<b>A Unified Numerical Simulation of Vowel Production That Comprises Phonation and the Emitted Sound</b> .....	3492
<i>Niyazi Cem Degirmenci, Johan Jansson, Johan Hoffman, Marc Arneta, Patricia Sanchez-Martin, Oriol Guasch, Sten Ternstrom</i>	
<b>Synthesis of VV Utterances from Muscle Activation to Sound with a 3D Model</b> .....	3497
<i>Saeed Dabbaghchian, Marc Arneta, Olov Engwall, Oriol Guasch</i>	

### **THU-SS-10-10: SPECIAL SESSION: INTERSPEECH 2017 COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE) 2**

<b>A Dual Source-Filter Model of Snore Audio for Snorer Group Classification</b> .....	3502
<i>Achuth Rao M. V., Shivani Yadav, Prasanta Kumar Ghosh</i>	
<b>An 'End-to-Evolution' Hybrid Approach for Snore Sound Classification</b> .....	3507
<i>Michael Freitag, Shahin Amiriparian, Nicholas Cummins, Maurice Gerczuk, Bjorn Schuller</i>	
<b>Snore Sound Classification Using Image-Based Deep Spectrum Features</b> .....	3512
<i>Shahin Amiriparian, Maurice Gerczuk, Sandra Ottl, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, Bjorn Schuller</i>	
<b>Exploring Fusion Methods and Feature Space for the Classification of Paralinguistic Information</b> .....	3517
<i>David Tavarez, Xabier Sarasola, Agustin Alonso, Jon Sanchez, Luis Serrano, Eva Navas, Inma Hernaez</i>	
<b>DNN-Based Feature Extraction and Classifier Combination for Child-Directed Speech, Cold and Snoring Identification</b> .....	3522
<i>Gabor Gosztolya, Robert Busa-Fekete, Tamas Grosz, Laszlo Toth</i>	
<b>Introducing Weighted Kernel Classifiers for Handling Imbalanced Paralinguistic Corpora: Snoring, Addressee and Cold</b> .....	3527
<i>Heysem Kaya, Alexey A. Karpov</i>	

### **THU-O-9-1: DISCRIMINATIVE TRAINING FOR ASR**

<b>Multitask Learning with Low-Level Auxiliary Tasks for Encoder-Decoder Based Speech Recognition</b> .....	3532
<i>Shubham Toshniwal, Hao Tang, Liang Lu, Karen Livescu</i>	
<b>Optimizing Expected Word Error Rate via Sampling for Speech Recognition</b> .....	3537
<i>Matt Shannon</i>	
<b>Annealed f-Smoothing as a Mechanism to Speed up Neural Network Training</b> .....	3542
<i>Tara N. Sainath, Vijayaditya Peddinti, Olivier Stohan, Arun Narayanan</i>	
<b>Non-Uniform MCE Training of Deep Long Short-Term Memory Recurrent Neural Networks for Keyword Spotting</b> .....	3547
<i>Zhong Meng, Biing-Hwang Juang</i>	
<b>Exploiting Eigenposteriors for Semi-Supervised Training of DNN Acoustic Models with Sequence Discrimination</b> .....	3552
<i>Pranay Dighe, Afsaneh Asaei, Herve Bourlard</i>	
<b>Discriminative Autoencoders for Acoustic Modeling</b> .....	3557
<i>Ming-Han Yang, Hung-Shin Lee, Yu-Ding Lu, Kuan-Yu Chen, Yu Tsao, Berlin Chen, Hsin-Min Wang</i>	

### **THU-O-9-2: SPEAKER DIARIZATION**

<b>Speaker Diarization Using Convolutional Neural Network for Statistics Accumulation Refinement</b> .....	3562
<i>Zbynek Zajic, Marek Hruz, Ludek Muller</i>	
<b>Speaker2Vec: Unsupervised Learning and Adaptation of a Speaker Manifold Using Deep Neural Networks with an Evaluation on Speaker Segmentation</b> .....	3567
<i>Arindam Jati, Panayiotis Georgiou</i>	
<b>A Triplet Ranking-Based Neural Network for Speaker Diarization and Linking</b> .....	3572
<i>Gael Le Lan, Delphine Charlet, Anthony Larcher, Sylvain Meignier</i>	
<b>Estimating Speaker Clustering Quality Using Logistic Regression</b> .....	3577
<i>Yishai Cohen, Itshak Lapidot</i>	

<b>Combining Speaker Turn Embedding and Incremental Structure Prediction for Low-Latency Speaker Diarization</b> .....	3582
<i>Guillaume Wisniewski, Herve Bredin, G. Gelly, Claude Barras</i>	
<b>Family Pyannote.Metrics: A Toolkit for Reproducible Evaluation, Diagnostic, and Error Analysis of Speaker Diarization Systems</b> .....	3587
<i>Herve Bredin</i>	

#### **THU-O-9-4: SPOKEN TERM DETECTION**

<b>A Rescoring Approach for Keyword Search Using Lattice Context Information</b> .....	3592
<i>Zhipeng Chen, Ji Wu</i>	
<b>The Kaldi OpenKWS System: Improving Low Resource Keyword Search</b> .....	3597
<i>Jan Trmal, Matthew Wiesner, Vijayaditya Peddinti, Xiaohui Zhang, Pegah Ghahremani, Yiming Wang, Vimal Manohar, Hainan Xu, Daniel Povey, Sanjeev Khudanpur</i>	
<b>The STC Keyword Search System for OpenKWS 2016 Evaluation</b> .....	3602
<i>Yuri Khokhlov, Ivan Medennikov, Aleksei Romanenko, Valentin Mendeleev, Maxim Korenevsky, Alexey Prudnikov, Natalia Tomashenko, Alexander Zlatovitskiy</i>	
<b>Compressed Time Delay Neural Network for Small-Footprint Keyword Spotting</b> .....	3607
<i>Ming Sun, David Snyder, Yixin Gao, Varun Nagaraja, Mike Rodehorst, Sankaran Panchapagesan, Nikko Strom, Spyros Matsoukas, Shiv Vitaladevuni</i>	
<b>Symbol Sequence Search from Telephone Conversation</b> .....	3612
<i>Masayuki Suzuki, Gakuto Kurata, Abhinav Sethy, Bhuvana Ramabhadran, Kenneth W. Church, Mark Drake</i>	
<b>Similarity Learning Based Query Modeling for Keyword Search</b> .....	3617
<i>Batuhan Gundogdu, Murat Saraclar</i>	

#### **THU-O-9-6: NOISE REDUCTION**

<b>Deep Recurrent Neural Network Based Monaural Speech Separation Using Recurrent Temporal Restricted Boltzmann Machines</b> .....	3622
<i>Suman Samui, Indrajit Chakrabarti, Soumya K. Ghosh</i>	
<b>Improved Codebook-Based Speech Enhancement Based on MBE Model</b> .....	3627
<i>Qizheng Huang, Changchun Bao, Xianyun Wang</i>	
<b>Improving Mask Learning Based Speech Enhancement System with Restoration Layers and Residual Connection</b> .....	3632
<i>Zhuo Chen, Yan Huang, Jinyu Li, Yifan Gong</i>	
<b>Exploring Low-Dimensional Structures of Modulation Spectra for Robust Speech Recognition</b> .....	3637
<i>Bi-Cheng Yan, Chin-Hong Shih, Shih-Hung Liu, Berlin Chen</i>	
<b>SEGAN: Speech Enhancement Generative Adversarial Network</b> .....	3642
<i>Santiago Pascual, Antonio Bonafonte, Joan Serra</i>	
<b>Concatenative Resynthesis Using Twin Networks</b> .....	3647
<i>Soumi Maiti, Michael I. Mandel</i>	

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

<b>Combining Residual Networks with LSTMs for Lipreading</b> .....	3652
<i>Themis Stafylakis, Georgios Tzimiropoulos</i>	
<b>Improving Computer Lipreading via DNN Sequence Discriminative Training Techniques</b> .....	3657
<i>Kwanchiva Thangthai, Richard Harvey</i>	
<b>Improving Speaker-Independent Lipreading with Domain-Adversarial Training</b> .....	3662
<i>Michael Wand, Jurgen Schmidhuber</i>	
<b>Turbo Decoders for Audio-Visual Continuous Speech Recognition</b> .....	3667
<i>Ahmed Hussien Abdelaziz</i>	
<b>DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface</b> .....	3672
<i>Tamas Gabor Csapo, Tamas Grosz, Gabor Gosztolya, Laszlo Toth, Alexandra Marko</i>	
<b>Visually Grounded Learning of Keyword Prediction from Untranscribed Speech</b> .....	3677
<i>Herman Kamper, Shane Settle, Gregory Shakhmarovich, Karen Livescu</i>	

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

<b>Deep Neural Factorization for Speech Recognition</b> .....	3682
<i>Jen-Tzung Chien, Chen Shen</i>	
<b>Semi-Supervised DNN Training with Word Selection for ASR</b> .....	3687
<i>Karel Vesely, Lukas Burget, Jan Cernocky</i>	
<b>Gaussian Prediction Based Attention for Online End-to-End Speech Recognition</b> .....	3692
<i>Junfeng Hou, Shiliang Zhang, Li-Rong Dai</i>	
<b>Efficient Knowledge Distillation from an Ensemble of Teachers</b> .....	3697
<i>Takashi Fukuda, Masayuki Suzuki, Gakuto Kurata, Samuel Thomas, Jia Cui, Bhuvana Ramabhadran</i>	
<b>An Analysis of "Attention" in Sequence-to-Sequence Models</b> .....	3702
<i>Rohit Prabhavalkar, Tara N. Sainath, Bo Li, Kanishka Rao, Navdeep Jaitly</i>	



<b>Neural Speech Recognizer: Acoustic-to-Word LSTM Model for Large Vocabulary Speech Recognition</b> .....	3707
<i>Hagen Soltau, Hank Liao, Hasim Sak</i>	

### **THU-O-10-2: ROBUST SPEAKER RECOGNITION**

<b>CNN-Based Joint Mapping of Short and Long Utterance i-Vectors for Speaker Verification Using Short Utterances</b> .....	3712
<i>Jinxi Guo, Usha Amrutha Nookala, Abeer Alwan</i>	
<b>Curriculum Learning Based Probabilistic Linear Discriminant Analysis for Noise Robust Speaker Recognition</b> .....	3717
<i>Shivesh Ranjan, Abhinav Misra, John H. L. Hansen</i>	
<b>i-Vector Transformation Using a Novel Discriminative Denoising Autoencoder for Noise-Robust Speaker Recognition</b> .....	3722
<i>Shivangi Mahto, Hitoshi Yamamoto, Takafumi Koshinaka</i>	
<b>Unsupervised Discriminative Training of PLDA for Domain Adaptation in Speaker Verification</b> .....	3727
<i>Qiongqiong Wang, Takafumi Koshinaka</i>	
<b>Speaker Verification Under Adverse Conditions Using i-Vector Adaptation and Neural Networks</b> .....	3732
<i>Jahangir Alam, Patrick Kenny, Gautam Bhattacharya, Marcel Kockmann</i>	
<b>Improving Robustness of Speaker Recognition to New Conditions Using Unlabeled Data</b> .....	3737
<i>Diego Castan, Mitchell McLaren, Luciana Ferrer, Aaron Lawson, Alicia Lozano-Diez</i>	

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

<b>CALYOU: A Comparable Spoken Algerian Corpus Harvested from YouTube</b> .....	3742
<i>K. Abidi, M. A. Menacer, Kamel Smaili</i>	
<b>PRAV: A Phonetically Rich Audio Visual Corpus</b> .....	3747
<i>Abhishek Narwekar, Prasanta Kumar Ghosh</i>	
<b>NTCD-TIMIT: A New Database and Baseline for Noise-Robust Audio-Visual Speech Recognition</b> .....	3752
<i>Ahmed Hussen Abdelaziz</i>	
<b>The Extended SPaRky Restaurant Corpus: Designing a Corpus with Variable Information Density</b> .....	3757
<i>David M. Howcroft, Dietrich Klakow, Vera Demberg</i>	
<b>Automatic Construction of the Finnish Parliament Speech Corpus</b> .....	3762
<i>Andre Mansikkaniemi, Peter Smit, Mikko Kurimo</i>	
<b>Building Audio-Visual Phonetically Annotated Arabic Corpus for Expressive Text to Speech</b> .....	3767
<i>Omnia Abdo, Sherif Abdou, Mervat Fashal</i>	

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

<b>What is the Relevant Population? Considerations for the Computation of Likelihood Ratios in Forensic Voice Comparison</b> .....	3772
<i>Vincent Hughes, Paul Foulkes</i>	
<b>Voice Disguise vs. Impersonation: Acoustic and Perceptual Measurements of Vocal Flexibility in Non Experts</b> .....	3777
<i>Veronique Delvaux, Lise Caucheteux, Kathy Huet, Myriam Piccaluga, Bernard Harmegnies</i>	
<b>Schwa Realization in French: Using Automatic Speech Processing to Study Phonological and Socio-Linguistic Factors in Large Corpora</b> .....	3782
<i>Yaru Wu, Martine Adda-Decker, Cecile Fougeron, Lori Lamel</i>	
<b>The Social Life of Setswana Ejectives</b> .....	3787
<i>Daniel Duran, Jagoda Bruni, Grzegorz Dogil, Justus Roux</i>	
<b>How Long is Too Long? How Pause Features After Requests Affect the Perceived Willingness of Affirmative Answers</b> .....	3792
<i>Lea S. Kohtz, Oliver Niebuhr</i>	
<b>Shadowing Synthesized Speech - Segmental Analysis of Phonetic Convergence</b> .....	3797
<i>Iona Gessinger, Eran Raveh, Sebastien Le Maguer, Bernd Mobius, Ingmar Steiner</i>	

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

<b>Occupancy Detection in Commercial and Residential Environments Using Audio Signal</b> .....	3802
<i>Shabnam Ghaffarzadegan, Attila Reiss, Mirko Ruhs, Robert Duerichen, Zhe Feng</i>	
<b>Data Augmentation, Missing Feature Mask and Kernel Classification for Through-the-Wall Acoustic Surveillance</b> .....	3807
<i>Huy Dat Tran, Wen Zheng Terence Ng, Yi Ren Leng</i>	
<b>Endpoint Detection Using Grid Long Short-Term Memory Networks for Streaming Speech Recognition</b> .....	3812
<i>Shuo-Yiin Chang, Bo Li, Tara N. Sainath, Gabor Simko, Carolina Parada</i>	
<b>Deep Learning Techniques in Tandem with Signal Processing Cues for Phonetic Segmentation for Text to Speech Synthesis in Indian Languages</b> .....	3817
<i>Arun Baby, Jeena J. Prakash, Rupak Vignesh, Hema A. Murthy</i>	
<b>Gate Activation Signal Analysis for Gated Recurrent Neural Networks and its Correlation with Phoneme Boundaries</b> .....	3822
<i>Yu-Hsuan Wang, Cheng-Tao Chung, Hung-Yi Lee</i>	

<b>Speaker Change Detection in Broadcast TV Using Bidirectional Long Short-Term Memory Networks</b> .....	3827
<i>Ruiqing Yin, Herve Bredin, Claude Barras</i>	

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

<b>Improved Automatic Speech Recognition Using Subband Temporal Envelope Features and Time-Delay Neural Network Denoising Autoencoder</b> .....	3832
<i>Cong-Thanh Do, Yannis Stylianou</i>	
<b>Factored Deep Convolutional Neural Networks for Noise Robust Speech Recognition</b> .....	3837
<i>Masakiyo Fujimoto</i>	
<b>Global SNR Estimation of Speech Signals for Unknown Noise Conditions Using Noise Adapted Non-Linear Regression</b> .....	3842
<i>Pavlos Papadopoulos, Ruchir Travadi, Shrikanth S. Narayanan</i>	
<b>Joint Training of Multi-Channel-Condition Dereverberation and Acoustic Modeling of Microphone Array Speech for Robust Distant Speech Recognition</b> .....	3847
<i>Fengpei Ge, Kehuang Li, Bo Wu, Sabato Marco Siniscalchi, Yonghong Yan, Chin-Hui Lee</i>	
<b>Uncertainty Decoding with Adaptive Sampling for Noise Robust DNN-Based Acoustic Modeling</b> .....	3852
<i>Dung T. Tran, Marc Delcroix, Atsunori Ogawa, Tomohiro Nakatani</i>	
<b>Attention-Based LSTM with Multi-Task Learning for Distant Speech Recognition</b> .....	3857
<i>Yu Zhang, Pengyuan Zhang, Yonghong Yan</i>	
<b>To Improve the Robustness of LSTM-RNN Acoustic Models Using Higher-Order Feedback from Multiple Histories</b> .....	3862
<i>Hengguan Huang, Brian Mak</i>	
<b>End-to-End Speech Recognition with Auditory Attention for Multi-Microphone Distance Speech Recognition</b> .....	3867
<i>Suyoun Kim, Ian Lane</i>	
<b>Robust Speech Recognition Based on Binaural Auditory Processing</b> .....	3872
<i>Anjali Menon, Chanwoo Kim, Richard M. Stern</i>	
<b>Adaptive Multichannel Dereverberation for Automatic Speech Recognition</b> .....	3877
<i>Joe Caroselli, Izhak Shafran, Arun Narayanan, Richard Rose</i>	

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

<b>The Effects of Real and Placebo Alcohol on Deaffrication</b> .....	3882
<i>Urban Zihlmann</i>	
<b>Polyglot and Speech Corpus Tools: A System for Representing, Integrating, and Querying Speech Corpora</b> .....	3887
<i>Michael McAuliffe, Elias Stengel-Eskin, Michaela Socolof, Morgan Sonderegger</i>	
<b>Mapping Across Feature Spaces in Forensic Voice Comparison: The Contribution of Auditory-Based Voice Quality to (Semi-)Automatic System Testing</b> .....	3892
<i>Vincent Hughes, Philip Harrison, Paul Foulkes, Peter French, Colleen Kavanagh, Eugenia San Segundo</i>	
<b>Effect of Language, Speaking Style and Speaker on Long-Term F0 Estimation</b> .....	3897
<i>Pablo Arantes, Anders Eriksson, Suska Gutzeit</i>	
<b>Stability of Prosodic Characteristics Across Age and Gender Groups</b> .....	3902
<i>Jan Volin, Tereza Tykalova, Tomas Boril</i>	
<b>Electrophysiological Correlates of Familiar Voice Recognition</b> .....	3907
<i>Julien Plante-Hebert, Victor J. Boucher, Boutheina Jemel</i>	
<b>Developing an Embosi (Bantu C25) Speech Variant Dictionary to Model Vowel Elision and Morpheme Deletion</b> .....	3911
<i>Jamison Cooper-Leavitt, Lori Lamel, Annie Rialland, Martine Adda-Decker, Gilles Adda</i>	
<b>R<sub>a</sub> as a Control Parameter to Explore Affective Correlates of the Tense-Lax Continuum</b> .....	3916
<i>Andy Murphy, Irena Yanushevskaya, Ailbhe Ni Chasaide, Christer Gobl</i>	
<b>Cross-Linguistic Distinctions Between Professional and Non-Professional Speaking Styles</b> .....	3921
<i>Plinio A. Barbosa, Sandra Madureira, Philippe Boula De Mareuil</i>	
<b>Perception and Production of Word-Final /R/ in French</b> .....	3926
<i>Cedric Gendrot</i>	
<b>Glottal Source Estimation from Coded Telephone Speech Using a Deep Neural Network</b> .....	3931
<i>N. P. Narendra, Manu Airaksinen, Paavo Alku</i>	
<b>Automatic Labelling of Prosodic Prominence, Phrasing and Disfluencies in French Speech by Simulating the Perception of Naive and Expert Listeners</b> .....	3936
<i>George Christodoulides, Mathieu Avanzi, Anne Catherine Simon</i>	
<b>Don't Count on ASR to Transcribe for You: Breaking Bias with Two Crowds</b> .....	3941
<i>Michael Levit, Yan Huang, Shuangyu Chang, Yifan Gong</i>	
<b>Effects of Training Data Variety in Generating Glottal Pulses from Acoustic Features with DNNs</b> .....	3946
<i>Manu Airaksinen, Paavo Alku</i>	
<b>Towards Intelligent Crowdsourcing for Audio Data Annotation: Integrating Active Learning in the Real World</b> .....	3951
<i>Simone Hantke, Zixing Zhang, Bjorn Schuller</i>	

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

<b>Principles for Learning Controllable TTS from Annotated and Latent Variation</b> .....	3956
<i>Gustav Eje Henter, Jaime Lorenzo-Trueba, Xin Wang, Junichi Yamagishi</i>	

<b>Sampling-Based Speech Parameter Generation Using Moment-Matching Networks .....</b>	<b>3961</b>
<i>Shinnosuke Takamichi, Tomoki Koriyama, Hiroshi Saruwatari</i>	
<b>Unit Selection with Hierarchical Cascaded Long Short Term Memory Bidirectional Recurrent Neural Nets .....</b>	<b>3966</b>
<i>Vincent Pollet, Enrico Zovato, Sufian Irhimeh, Pier Batzu</i>	
<b>Utterance Selection for Optimizing Intelligibility of TTS Voices Trained on ASR Data .....</b>	<b>3971</b>
<i>Erica Cooper, Xinyue Wang, Alison Chang, Yocheved Levitan, Julia Hirschberg</i>	
<b>Bias and Statistical Significance in Evaluating Speech Synthesis with Mean Opinion Scores .....</b>	<b>3976</b>
<i>Andrew Rosenberg, Bhuvana Ramabhadran</i>	
<b>Phase Modeling Using Integrated Linear Prediction Residual for Statistical Parametric Speech Synthesis .....</b>	<b>3981</b>
<i>Nagaraj Adiga, S. R. Mahadeva Prasanna</i>	
<b>Evaluation of a Silent Speech Interface Based on Magnetic Sensing and Deep Learning for a Phonetically Rich Vocabulary .....</b>	<b>3986</b>
<i>Jose A. Gonzalez, Lam A. Cheah, Phil D. Green, James M. Gilbert, Stephen R. Ell, Roger K. Moore, Ed Holdsworth</i>	
<b>Predicting Head Pose from Speech with a Conditional Variational Autoencoder .....</b>	<b>3991</b>
<i>David Greenwood, Stephen Laycock, Iain Matthews</i>	
<b>Real-Time Reactive Speech Synthesis: Incorporating Interruptions .....</b>	<b>3996</b>
<i>Mirjam Wester, David A. Braude, Blaise Potard, Matthew P. Aylett, Francesca Shaw</i>	
<b>A Neural Parametric Singing Synthesizer .....</b>	<b>4001</b>
<i>Merlijn Blaauw, Jordi Bonada</i>	
<b>Tacotron: Towards End-to-End Speech Synthesis .....</b>	<b>4006</b>
<i>Yuxuan Wang, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio, Quoc Le, Yannis Agiomyrigiannakis, Rob Clark, Rif A. Saurous</i>	
<b>Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System .....</b>	<b>4011</b>
<i>Tim Capes, Paul Coles, Alistair Conkie, Ladan Golipour, Abie Hadjitarkhani, Qiong Hu, Nancy Huddleston, Melynn Hunt, Jiangchuan Li, Matthias Neeracher, Kishore Prahallad, Tuomo Raitio, Ramya Rasipuram, Greg Townsend, Becci Williamson, David Winarsky, Zhizheng Wu, Hepeng Zhang</i>	
<b>An Expanded Taxonomy of Semiotic Classes for Text Normalization .....</b>	<b>4016</b>
<i>Daan Van Esch, Richard Sproat</i>	
<b>Complex-Valued Restricted Boltzmann Machine for Direct Learning of Frequency Spectra .....</b>	<b>4021</b>
<i>Toru Nakashika, Shinji Takaki, Junichi Yamagishi</i>	

#### **THU-S&T-9/10-A: SHOW & TELL 7**

<b>Soundtracing for Realtime Speech Adjustment to Environmental Conditions in 3D Simulations .....</b>	<b>4026</b>
<i>Bartosz Ziolko, Tomasz Pedzima, Szymon Palka</i>	
<b>Vocal-Tract Model with Static Articulators: Lips, Teeth, Tongue, and More .....</b>	<b>4028</b>
<i>Takayuki Arai</i>	
<b>Remote Articulation Test System Based on WebRTC .....</b>	<b>4030</b>
<i>Ikuyo Masuda-Katsuse</i>	
<b>The ModelTalker Project: A Web-Based Voice Banking Pipeline for ALS/MND Patients .....</b>	<b>4032</b>
<i>H. Timothy Bunnell, Jason Lilley, Kathleen McGrath</i>	
<b>Visible Vowels: A Tool for the Visualization of Vowel Variation .....</b>	<b>4034</b>
<i>Wilbert Heeringa, Hans Van De Velde</i>	
<b>Author Index</b>	