

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

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
Web: www.proceedings.com

TABLE OF CONTENTS

VOLUME 1

ISCA MEDAL 2017 CEREMONY

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

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

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

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

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

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

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

SFF Anti-Spoofers: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2017	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

Improved Single System Conversational Telephone Speech Recognition with VGG Bottleneck Features	112
<i>William Hartmann, Roger Hsiao, Tim Ng, Jeff Ma, Francis Keith, Man-Hung Siu</i>	
Student-Teacher Training with Diverse Decision Tree Ensembles	117
<i>Jeremy H. M. Wong, Mark J. F. Gales</i>	
Embedding-Based Speaker Adaptive Training of Deep Neural Networks	122
<i>Xiaodong Cui, Vaibhava Goel, George Saon</i>	
Improving Deliverable Speech-to-Text Systems with Multilingual Knowledge Transfer	127
<i>Jeff Ma, Francis Keith, Tim Ng, Man-Hung Siu, Owen Kimball</i>	
English Conversational Telephone Speech Recognition by Humans and Machines	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>	
Comparing Human and Machine Errors in Conversational Speech Transcription	137
<i>Andreas Stolcke, Jasha Droppo</i>	

MON-O-1-2: MULTIMODAL PARALINGUISTICS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Generation of Large-Scale Simulated Utterances in Virtual Rooms to Train Deep-Neural Networks for Far-Field Speech Recognition in Google Home	379
<i>Chanwoo Kim, Ananya Misra, Kean Chin, Thad Hughes, Arun Narayanan, Tara N. Sainath, Michiel Bacchiani</i>	
Neural Network-Based Spectrum Estimation for Online WPE Dereverberation	384
<i>Keisuke Kinoshita, Marc Delcroix, Haeyong Kwon, Takuma Mori, Tomohiro Nakatani</i>	
Factorial Modeling for Effective Suppression of Directional Noise	389
<i>Osamu Ichikawa, Takashi Fukuda, Gakuto Kurata, Steven J. Rennie</i>	
On Design of Robust Deep Models for CHiME-4 Multi-Channel Speech Recognition with Multiple Configurations of Array Microphones	394
<i>Yan-Hui Tu, Jun Du, Lei Sun, Feng Ma, Chin-Hui Lee</i>	
Acoustic Modeling for Google Home	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>	
On Multi-Domain Training and Adaptation of End-to-End RNN Acoustic Models for Distant Speech Recognition	404
<i>Seyedmahdad Mirsamadi, John H. L. Hansen</i>	

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

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

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

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

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

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

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

MON-P-2-1: SPEECH PERCEPTION

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

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

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

Critical Articulators Identification from RT-MRI of the Vocal Tract	626
<i>Samuel Silva, Antonio Teixeira</i>	
Semantic Edge Detection for Tracking Vocal Tract Air-Tissue Boundaries in Real-Time Magnetic Resonance Images	631
<i>Krishna Somandepalli, Asterios Toutios, Shrikanth S. Narayanan</i>	
Vocal Tract Airway Tissue Boundary Tracking for rtMRI Using Shape and Appearance Priors	636
<i>Sasan Asadiabadi, Engin Erzin</i>	
An Objective Critical Distance Measure Based on the Relative Level of Spectral Valley	641
<i>T. V. Ananthapadmanabha, A. G. Ramakrishnan, Shubham Sharma</i>	
Database of Volumetric and Real-time Vocal Tract MRI for Speech Science	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

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

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

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

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

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

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

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

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

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

KEYNOTE 1: JAMES ALLEN

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

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

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

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

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

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

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

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

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

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

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

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

TUE-O-3-4: SPEAKER RECOGNITION

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

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

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

TUE-O-3-8: SPEECH SYNTHESIS PROSODY

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

TUE-O-3-10: EMOTION RECOGNITION

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

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

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

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

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

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

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

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

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

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

TUE-O-4-8: EMOTION MODELING

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

TUE-O-4-10: VOICE CONVERSION 1

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

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

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

VOLUME 3

TUE-O-5-2: SPEAKER RECOGNITION EVALUATION

The I4U Mega Fusion and Collaboration for NIST Speaker Recognition Evaluation 2016	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>	
The MIT-LL, JHU and LRDE NIST 2016 Speaker Recognition Evaluation System	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>	
Nuance - Politecnico di Torino's 2016 NIST Speaker Recognition Evaluation System	1338
<i>Daniele Colibro, Claudio Vair, Emanuele Dalmaso, Kevin Farrell, Gennady Karvitsky, Sandro Cumani, Pietro Laface</i>	
UTD-CRSS Systems for 2016 NIST Speaker Recognition Evaluation	1343
<i>Chunlei Zhang, Fahimeh Bahmaninezhad, Shivesh Ranjan, Chengzhu Yu, Navid Shokouhi, John H. L. Hansen</i>	
Analysis and Description of ABC Submission to NIST SRE 2016	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>	
The 2016 NIST Speaker Recognition Evaluation	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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

TUE-P-4-3: DIALOG MODELING

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

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

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

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

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

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

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

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

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

TUE-P-5-4: SPEECH-ENHANCEMENT

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

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

VOLUME 4

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

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

Applications of the BBN Sage Speech Processing Platform	2028
<i>Ralf Meermeier, Sean Colbath</i>	
Bob Speaks Kaldi	2030
<i>Milos Cernak, Alain Komaty, Amir Mohammadi, Andre Anjos, Sebastien Marcel</i>	
Real Time Pitch Shifting with Formant Structure Preservation Using the Phase Vocoder	2032
<i>Michal Lenarczyk</i>	
A Signal Processing Approach for Speaker Separation Using SFF Analysis	2034
<i>Nivedita Chennupati, B. H. V. S. Narayana Murthy, B. Yegnanarayana</i>	
Speech Recognition and Understanding on Hardware-Accelerated DSP	2036
<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>	
MetaLab: A Repository for Meta-Analyses on Language Development, and More	2038
<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

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

KEYNOTE 2: CATHERINE PELACHAUD

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

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

Team ELISA System for DARPA LORELEI Speech Evaluation 2016	2053
<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>	
First Results in Developing a Medieval Latin Language Charter Dictation System for the East-Central Europe Region	2058
<i>Peter Mihajlik, Lili Szabo, Balazs Tarjan, Andras Balog, Krisztina Rabai</i>	
The Motivation and Development of MP<i>A</i>i, a Maori Pronunciation Aid	2063
<i>C. I. Watson, P. J. Keegan, M. A. Maclagan, R. Harlow, J. King</i>	
On the Linguistic Relevance of Speech Units Learned by Unsupervised Acoustic Modeling	2068
<i>Siyuan Feng, Tan Lee</i>	
Deep Auto-Encoder Based Multi-Task Learning Using Probabilistic Transcriptions	2073
<i>Amit Das, Mark Hasegawa-Johnson, Karel Vesely</i>	

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

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

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

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

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

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

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

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

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

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

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

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

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

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

WED-O-6-6: DIALOG AND PROSODY

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

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

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

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

WED-O-6-10: ACOUSTIC MODEL ADAPTATION

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

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

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

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

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

Coupled Initialization of Multi-Channel Non-Negative Matrix Factorization Based on Spatial and Spectral Information	2461
<i>Yuuki Tachioka, Tomohiro Narita, Iori Miura, Takanobu Uramoto, Natsuki Monta, Shingo Uenohara, Ken'Ichi Furuya, Shinji Watanabe, Jonathan Le Roux</i>	
Channel Compensation in the Generalised Vector Taylor Series Approach to Robust ASR	2466
<i>Erfan Loweimi, Jon Barker, Thomas Hain</i>	
Robust Speech Recognition via Anchor Word Representations	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

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

WED-O-7-6: DIALOG SYSTEMS

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

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

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

WED-O-7-10: LANGUAGE RECOGNITION

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

QMDIS: QCRI-MIT Advanced Dialect Identification System	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

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

WED-O-8-4: SPEECH TRANSLATION

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

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

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

VOLUME 5

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

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

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

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

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

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

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

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

Null-Hypothesis LLR: A Proposal for Forensic Automatic Speaker Recognition	2849
<i>Yosef A. Solewicz, Michael Jessen, David Van Der Vloed</i>	
The Opensesame NIST 2016 Speaker Recognition Evaluation System	2854
<i>Gang Liu, Qi Qian, Zhibin Wang, Qingen Zhao, Tianzhou Wang, Hao Li, Jian Xue, Shenghuo Zhu, Rong Jin, Tuo Zhao</i>	
IITG-Indigo System for NIST 2016 SRE Challenge	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>	
Locally Weighted Linear Discriminant Analysis for Robust Speaker Verification	2864
<i>Abhinav Misra, Shivesh Ranjan, John H. L. Hansen</i>	
Recursive Whitening Transformation for Speaker Recognition on Language Mismatched Condition	2869
<i>Suwon Shon, Seongkyu Mun, Hanseok Ko</i>	

WED-P-6-3: SPOKEN DOCUMENT PROCESSING

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

WED-P-6-4: SPEECH INTELLIGIBILITY

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

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

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

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

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

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

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

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

WED-P-8-1: PROSODY

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

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

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

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

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

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

VOLUME 6

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

WED-P-8-4: VOICE CONVERSION 2

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

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

Creating a Voice for MiRo, the World's First Commercial Biomimetic Robot	3419
<i>Roger K. Moore, Ben Mitchinson</i>	
A Thematicity-Based Prosody Enrichment Tool for CTS	3421
<i>Monica Dominguez, Mireia Farrus, Leo Wanner</i>	
WebSubDub --- Experimental System for Creating High-Quality Alternative Audio Track for TV Broadcasting	3423
<i>Martin Gruber, Jindrich Matousek, Zdenek Hanzlicek, Jakub Vit, Daniel Tihelka</i>	
Voice Conservation and TTS System for People Facing Total Laryngectomy	3425
<i>Marketa Juzova, Daniel Tihelka, Jindrich Matousek, Zdenek Hanzlicek</i>	
TBT (Toolkit to Build TTS): A High Performance Framework to Build Multiple Language HTS Voice	3427
<i>Atish Shankar Ghone, Rachana Nerpagar, Pranaw Kumar, Arun Baby, Aswin Shanmugam, M. Sasikumar, Hema A. Murthy</i>	
SIAK --- A Game for Foreign Language Pronunciation Learning	3429
<i>Reima Karhila, Sari Ylinoen, 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

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

KEYNOTE 3: BJORN LINDBLOM

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

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

The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addressee, Cold & Snoring	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>	
It Sounds Like You Have a Cold! Testing Voice Features for the Interspeech 2017 Computational Paralinguistics Cold Challenge	3447
<i>Mark Huckvale, Andras Beke</i>	
End-to-End Deep Learning Framework for Speech Paralinguistics Detection Based on Perception Aware Spectrum	3452
<i>Danwei Cai, Zhidong Ni, Wenbo Liu, Weicheng Cai, Gang Li, Ming Li</i>	
Infected Phonemes: How a Cold Impairs Speech on a Phonetic Level	3457
<i>Johannes Wagner, Thiago Fraga-Silva, Yvan Josse, Dominik Schiller, Andreas Seiderer, Elisabeth Andre</i>	

Phoneme State Posteriorgram Features for Speech Based Automatic Classification of Speakers in Cold and Healthy Condition	3462
<i>Akshay Kalkunte Suresh, K. M. Srinivasa Raghavan, Prasanta Kumar Ghosh</i>	
An Integrated Solution for Snoring Sound Classification Using Bhattacharyya Distance Based GMM Supervectors with SVM, Feature Selection with Random Forest and Spectrogram with CNN	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

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

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

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

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

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

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

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

THU-O-9-4: SPOKEN TERM DETECTION

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

THU-O-9-6: NOISE REDUCTION

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

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

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

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

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

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

THU-O-10-2: ROBUST SPEAKER RECOGNITION

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

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

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

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

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

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

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

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

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

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

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

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

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

Sampling-Based Speech Parameter Generation Using Moment-Matching Networks	3961
<i>Shinnosuke Takamichi, Tomoki Koriyama, Hiroshi Saruwatari</i>	
Unit Selection with Hierarchical Cascaded Long Short Term Memory Bidirectional Recurrent Neural Nets	3966
<i>Vincent Pollet, Enrico Zovato, Sufian Irhimeh, Pier Batzu</i>	
Utterance Selection for Optimizing Intelligibility of TTS Voices Trained on ASR Data	3971
<i>Erica Cooper, Xinyue Wang, Alison Chang, Yocheved Levitan, Julia Hirschberg</i>	
Bias and Statistical Significance in Evaluating Speech Synthesis with Mean Opinion Scores	3976
<i>Andrew Rosenberg, Bhuvana Ramabhadran</i>	
Phase Modeling Using Integrated Linear Prediction Residual for Statistical Parametric Speech Synthesis	3981
<i>Nagaraj Adiga, S. R. Mahadeva Prasanna</i>	
Evaluation of a Silent Speech Interface Based on Magnetic Sensing and Deep Learning for a Phonetically Rich Vocabulary	3986
<i>Jose A. Gonzalez, Lam A. Cheah, Phil D. Green, James M. Gilbert, Stephen R. Ell, Roger K. Moore, Ed Holdsworth</i>	
Predicting Head Pose from Speech with a Conditional Variational Autoencoder	3991
<i>David Greenwood, Stephen Laycock, Iain Matthews</i>	
Real-Time Reactive Speech Synthesis: Incorporating Interruptions	3996
<i>Mirjam Wester, David A. Braude, Blaise Potard, Matthew P. Aylett, Francesca Shaw</i>	
A Neural Parametric Singing Synthesizer	4001
<i>Merlijn Blaauw, Jordi Bonada</i>	
Tacotron: Towards End-to-End Speech Synthesis	4006
<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>	
Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System	4011
<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>	
An Expanded Taxonomy of Semiotic Classes for Text Normalization	4016
<i>Daan Van Esch, Richard Sproat</i>	
Complex-Valued Restricted Boltzmann Machine for Direct Learning of Frequency Spectra	4021
<i>Toru Nakashika, Shinji Takaki, Junichi Yamagishi</i>	

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

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