

# **19th Annual Conference of the International Speech Communication Association (INTERSPEECH 2018)**

Speech Research for Emerging Markets in  
Multilingual Societies

Hyderabad, India  
2 - 6 September 2018

Volume 1 of 6

ISBN: 978-1-5108-7221-9

**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© (2018) by International Speech Communication Association  
All rights reserved.

Printed by Curran Associates, Inc. (2019)

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

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

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

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

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

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

# TABLE OF CONTENTS

## VOLUME 1

### ISCA MEDAL TALK

<b>From Vocoders to Code-Excited Linear Prediction: Learning How We Hear What We Hear</b> .....	1
<i>Bishnu S. Atal</i>	

### END-TO-END SPEECH RECOGNITION

<b>Semi-Supervised End-to-End Speech Recognition</b> .....	2
<i>Shigeki Karita, Shinji Watanabe, Tomoharu Iwata, Atsunori Ogawa, Marc Delcroix</i>	
<b>Improved Training of End-to-end Attention Models for Speech Recognition</b> .....	7
<i>Albert Zeyer, Kazuki Irie, Ralf Schlüter, Hermann Ney</i>	
<b>End-to-end Speech Recognition Using Lattice-free MMI</b> .....	12
<i>Hossein Hadian, Hossein Sameti, Daniel Povey, Sanjeev Khudanpur</i>	
<b>Multi-channel Attention for End-to-End Speech Recognition</b> .....	17
<i>Stefan Braun, Daniel Neil, Jithendar Anumula, Enea Ceolini, Shih-Chii Liu</i>	
<b>Quaternion Convolutional Neural Networks for End-to-End Automatic Speech Recognition</b> .....	22
<i>Titouan Parcollet, Ying Zhang, Mohamed Morchid, Chiheb Trabelsi, Georges Linares, Renato De Mori, Yoshua Bengio</i>	
<b>Compression of End-to-End Models</b> .....	27
<i>Ruoming Pang, Tara Sainath, Rohit Prabhavalkar, Suyog Gupta, Yonghui Wu, Shuyuan Zhang, Chung-Cheng Chiu</i>	

### PROSODY MODELING AND GENERATION

<b>Learning Interpretable Control Dimensions for Speech Synthesis by Using External Data</b> .....	32
<i>Zack Hodari, Oliver Watts, Srikanth Ronanki, Simon King</i>	
<b>Investigating Accuracy of Pitch-accent Annotations in Neural Network-based Speech Synthesis and Denoising Effects</b> .....	37
<i>Hieu-Thi Luong, Xin Wang, Junichi Yamagishi, Nobuyuki Nishizawa</i>	
<b>An Exploration of Local Speaking Rate Variations in Mandarin Read Speech</b> .....	42
<i>Guan-Ting Liou, Chen-Yu Chiang, Yih-Ru Wang, Sin-Horng Chen</i>	
<b>BLSTM-CRF Based End-to-End Prosodic Boundary Prediction with Context Sensitive Embeddings in a Text-to-Speech Front-End</b> .....	47
<i>Yibin Zheng, Jianhua Tao, Zhengqi Wen, Ya Li</i>	
<b>Wavelet Analysis of Speaker Dependent and Independent Prosody for Voice Conversion</b> .....	52
<i>Berrak Sisman, Haizhou Li</i>	
<b>Improving Mongolian Phrase Break Prediction by Using Syllable and Morphological Embeddings with BiLSTM Model</b> .....	57
<i>Rui Liu, Feilong Bao, Guanglai Gao, Hui Zhang, Yonghe Wang</i>	

### SPEAKER VERIFICATION I

<b>Improved Supervised Locality Preserving Projection for I-vector Based Speaker Verification</b> .....	62
<i>Lanhua You, Wu Guo, Yan Song, Sheng Zhang</i>	
<b>Double Joint Bayesian Modeling of DNN Local I-Vector for Text Dependent Speaker Verification with Random Digit Strings</b> .....	67
<i>Ziqiang Shi, Huibin Lin, Liu Liu, Rujie Liu</i>	
<b>Fast Variational Bayes for Heavy-tailed PLDA Applied to i-vectors and x-vectors</b> .....	72
<i>Anna Silnova, Niko Brümmer, Daniel Garcia-Romero, David Snyder, Lukáš Burget</i>	
<b>Integrated Presentation Attack Detection and Automatic Speaker Verification: Common Features and Gaussian Back-end Fusion</b> .....	77
<i>Massimiliano Todisco, Héctor Delgado, Kong Aik Lee, Md Sahidullah, Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi</i>	
<b>A Generalization of PLDA for Joint Modeling of Speaker Identity and Multiple Nuisance Conditions</b> .....	82
<i>Luciana Ferrer, Mitchell McLaren</i>	
<b>An Investigation of Non-linear i-vectors for Speaker Verification</b> .....	87
<i>Nanxin Chen, Jesús Villalba, Najim Dehak</i>	

### SPOKEN TERM DETECTION

<b>CNN Based Query by Example Spoken Term Detection</b> .....	92
<i>Dhananjay Ram, Lesly Miculicich, Hervé Bourlard</i>	
<b>Learning Acoustic Word Embeddings with Temporal Context for Query-by-Example Speech Search</b> .....	97
<i>Yougen Yuan, Cheung-Chi Leung, Lei Xie, Hongjie Chen, Bin Ma, Haizhou Li</i>	

<b>Siamese Recurrent Auto-Encoder Representation for Query-by-Example Spoken Term Detection</b> .....	102
<i>Ziwei Zhu, Zhiyong Wu, Runnan Li, Helen Meng, Lianhong Cai</i>	
<b>Fast Derivation of Cross-lingual Document Vectors from Self-attentive Neural Machine Translation Model</b> .....	107
<i>Wei Li, Brian Mak</i>	
<b>LSTM Based Attentive Fusion of Spectral and Prosodic Information for Keyword Spotting in Hindi Language</b> .....	112
<i>Laxmi Pandey, Karan Nathwani</i>	
<b>Spoken Keyword Detection Using Joint DTW-CNN</b> .....	117
<i>Ravi Shankar, Vikram C M, S R Mahadeva Prasanna</i>	

## **THE INTERSPEECH 2018 COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE): ATYPICAL & SELF-ASSESSED AFFECT, CRYING & HEART BEATS 1**

<b>The INTERSPEECH 2018 Computational Paralinguistics Challenge: Atypical &amp; Self-Assessed Affect, Crying &amp; Heart Beats</b> .....	122
<i>Björn Schuller, Stefan Steidl, Anton Batliner, Peter B. Marschik, Harald Baumeister, Fengquan Dong, Simone Hantke, Florian B. Pokorny, Eva-Maria Rathner, Katrin D. Bartl-Pokorny, Christa Einspieler, Dajje Zhang, Alice Baird, Shahin Amiriparian, Kun Qian, Zhao Ren, Maximilian Schmitt, Panagiotis Tzirakis, Stefanos Zafeiriou</i>	
<b>An Ensemble of Transfer, Semi-supervised and Supervised Learning Methods for Pathological Heart Sound Classification</b> .....	127
<i>Ahmed Imtiaz Humayun, Md. Tauhiduzzaman Khan, Shabnam Ghaffarzadegan, Zhe Feng, Taufiq Hasan</i>	
<b>Monitoring Infant's Emotional Cry in Domestic Environments Using the Capsule Network Architecture</b> .....	132
<i>Mehmet Ali Tugtekin Turan, Engin Erzincan</i>	
<b>Neural Network Architecture That Combines Temporal and Summative Features for Infant Cry Classification in the Interspeech 2018 Computational Paralinguistics Challenge</b> .....	137
<i>Mark Huckvale</i>	
<b>Evolving Learning for Analysing Mood-Related Infant Vocalisation</b> .....	142
<i>Zixing Zhang, Jing Han, Kun Qian, Björn Schuller</i>	
<b>Deep Learning in Paralinguistic Recognition Tasks: Are Hand-crafted Features Still Relevant?</b> .....	147
<i>Johannes Wagner, Dominik Schiller, Andreas Seiderer, Elisabeth André</i>	
<b>Investigation on Joint Representation Learning for Robust Feature Extraction in Speech Emotion Recognition</b> .....	152
<i>Danqing Luo, Yuexian Zou, Dongyan Huang</i>	
<b>Using Voice Quality Supervectors for Affect Identification</b> .....	157
<i>Soo Jin Park, Amber Afshan, Zhi Ming Chua, Abeer Alwan</i>	
<b>An End-to-End Deep Learning Framework for Speech Emotion Recognition of Atypical Individuals</b> .....	162
<i>Dengke Tang, Junlin Zeng, Ming Li</i>	

## **SHOW AND TELL 1**

<b>DialogOS: Simple and Extensible Dialogue Modeling</b> .....	167
<i>Alexander Koller, Timo Baumann, Arne Köhn</i>	
<b>A Framework for Speech Recognition Benchmarking</b> .....	169
<i>Franck Dernoncourt, Trung Bui, Walter Chang</i>	
<b>Flexible Tongue Housed in a Static Model of the Vocal Tract With Jaws, Lips and Teeth</b> .....	171
<i>Takayuki Arai</i>	
<b>Voice Analysis Using Acoustic and Throat Microphones for Speech Therapy</b> .....	173
<i>Lani Mathew, K. Gopakumar</i>	
<b>A Robust Context-Dependent Speech-to-Speech Phraselator Toolkit for Alexa</b> .....	175
<i>Manny Rayner, Nikos Tsourakis, Jan Stanek</i>	

## **SPEECH SEGMENTS AND VOICE QUALITY**

<b>Discriminating Nasals and Approximants in English Language Using Zero Time Windowing</b> .....	177
<i>Ravishankar Prasad, Sudarsana Reddy Kadiri, Suryakanth V Gangashetty, Bayya Yegnanarayana</i>	
<b>Gestural Lenition of Rhotics Captures Variation in Brazilian Portuguese</b> .....	182
<i>Phil Howson, Alexei Kochetov</i>	
<b>Identification and Classification of Fricatives in Speech Using Zero Time Windowing Method</b> .....	187
<i>Ravishankar Prasad, Bayya Yegnanarayana</i>	
<b>GlobalTIMIT: Acoustic-Phonetic Datasets for the World's Languages</b> .....	192
<i>Nattanun Chanchaochai, Christopher Cieri, Japhet Debrah, Hongwei Ding, Yue Jiang, Sishi Liao, Mark Liberman, Jonathan Wright, Jiahong Yuan, Juhong Zhan, Yuqing Zhan</i>	
<b>Structural Effects on Properties of Consonantal Gestures in Tashlihiyt</b> .....	197
<i>Anne Hermes, Doris Mücke, Bastian Auris, Rachid Ridouane</i>	
<b>The Retroflex-dental Contrast in Punjabi Stops and Nasals: A Principal Component Analysis of Ultrasound Images</b> .....	202
<i>Alexei Kochetov, Matthew Faytak, Kiranpreet Nara</i>	
<b>Vowels and Diphthongs in Hangzhou Wu Chinese Dialect</b> .....	207
<i>Yang Yue, Fang Hu</i>	

<b>Resyllabification in Indian Languages and Its Implications in Text-to-speech Systems</b> .....	212
<i>Mahesh M, Jeena J Prakash, Hema Murthy</i>	
<b>Voice Source Contribution to Prominence Perception: Rd Implementation</b> .....	217
<i>Andy Murphy, Irena Yanushevskaya, Ailbhe Ní Chasaíde, Christer Gobl</i>	
<b>On the Relationship between Glottal Pulse Shape and Its Spectrum: Correlations of Open Quotient, Pulse Skew and Peak Flow with Source Harmonic Amplitudes</b> .....	222
<i>Christer Gobl, Andy Murphy, Irena Yanushevskaya, Ailbhe Ní Chasaíde</i>	
<b>The Individual and the System: Assessing the Stability of the Output of a Semi-automatic Forensic Voice Comparison System</b> .....	227
<i>Vincent Hughes, Philip Harrison, Paul Foulkes, Peter French, Colleen Kavanagh, Eugenia San Segundo</i>	
<b>Breathly to Tense Voice Discrimination using Zero-Time Windowing Cepstral Coefficients (ZTWCCs)</b> .....	232
<i>Sudarsana Reddy Kadiri, Bayya Yegnanarayana</i>	
<b>Analysis of Breathiness in Contextual Vowel of Voiceless Nasals in Mizo</b> .....	237
<i>Pamir Gogoi, Sishir Kalita, Parismita Gogoi, Ratree Wayland, Priyankoo Sarmah, S R Mahadeva Prasanna</i>	

## **SPEAKER STATE AND TRAIT**

<b>Infant Emotional Outbursts Detection in Infant-parent Spoken Interactions</b> .....	242
<i>Yijia Xu, Mark Hasegawa-Johnson, Nancy McElwain</i>	
<b>Deep Neural Networks for Emotion Recognition Combining Audio and Transcripts</b> .....	247
<i>Jaemin Cho, Raghavendra Pappagari, Purva Kulkarni, Jesús Villalba, Yishay Carmiel, Najim Dehak</i>	
<b>Preference-Learning with Qualitative Agreement for Sentence Level Emotional Annotations</b> .....	252
<i>Srinivas Parthasarathy, Carlos Busso</i>	
<b>Transfer Learning for Improving Speech Emotion Classification Accuracy</b> .....	257
<i>Siddique Latif, Rajib Rana, Shahzad Younis, Junaaid Qadir, Julien Epps</i>	
<b>What Do Classifiers Actually Learn? a Case Study on Emotion Recognition Datasets</b> .....	262
<i>Patrick Meyer, Eric Buschermöhle, Tim Fingscheidt</i>	
<b>State of Mind: Classification through Self-reported Affect and Word Use in Speech.</b> .....	267
<i>Eva-Maria Rathner, Yannik Terhorst, Nicholas Cummins, Björn Schuller, Harald Baumeister</i>	
<b>Exploring Spatio-Temporal Representations by Integrating Attention-based Bidirectional-LSTM-RNNs and FCNs for Speech Emotion Recognition</b> .....	272
<i>Ziping Zhao, Yu Zheng, Zixing Zhang, Haishuai Wang, Yiqin Zhao, Chao Li</i>	
<b>End-to-end Deep Neural Network Age Estimation</b> .....	277
<i>Pegah Ghahremani, Phani Sankar Nidadavolu, Nanxin Chen, Jesús Villalba, Daniel Povey, Sanjeev Khudanpur, Najim Dehak</i>	
<b>Improving Gender Identification in Movie Audio Using Cross-Domain Data</b> .....	282
<i>Rajat Hebbar, Krishna Somandepalli, Shrikanth Narayanan</i>	
<b>On Learning to Identify Genders from Raw Speech Signal Using CNNs</b> .....	287
<i>Selen Hande Kabil, Hannah Muckenhirn, Mathew Magimai.-Doss</i>	
<b>Denoising and Raw-waveform Networks for Weakly-Supervised Gender Identification on Noisy Speech</b> .....	292
<i>Jilt Sebastian, Manoj Kumar, Pavan Kumar D. S., Mathew Magimai.-Doss, Hema Murthy, Shrikanth Narayanan</i>	
<b>The Effect of Exposure to High Altitude and Heat on Speech Articulatory Coordination</b> .....	297
<i>James Williamson, Thomas Quatieri, Adam Lammert, Katherine Mitchell, Katherine Finkelstein, Nicole Ekon, Caitlin Dillon, Robert Kenefick, Kristin Heaton</i>	

## **DEEP LEARNING FOR SOURCE SEPARATION AND PITCH TRACKING**

<b>Permutation Invariant Training of Generative Adversarial Network for Monaural Speech Separation</b> .....	302
<i>Lianwu Chen, Meng Yu, Yanmin Qian, Dan Su, Dong Yu</i>	
<b>Deep Extractor Network for Target Speaker Recovery from Single Channel Speech Mixtures</b> .....	307
<i>Jun Wang, Jie Chen, Dan Su, Lianwu Chen, Meng Yu, Yanmin Qian, Dong Yu</i>	
<b>Joint Localization and Classification of Multiple Sound Sources Using a Multi-task Neural Network</b> .....	312
<i>Weipeng He, Petr Motlicek, Jean-Marc Odobez</i>	
<b>Detection of Glottal Closure Instants from Speech Signals: A Convolutional Neural Network Based Method</b> .....	317
<i>Shuai Yang, Zhiyong Wu, Binbin Shen, Helen Meng</i>	
<b>Robust TDOA Estimation Based on Time-Frequency Masking and Deep Neural Networks</b> .....	322
<i>Zhong-Qiu Wang, Xueliang Zhang, Deliang Wang</i>	
<b>Waveform to Single Sinusoid Regression to Estimate the F0 Contour from Noisy Speech Using Recurrent Deep Neural Networks</b> .....	327
<i>Akihiro Kato, Tomi Kinnunen</i>	
<b>Reducing Interference with Phase Recovery in DNN-based Monaural Singing Voice Separation</b> .....	332
<i>Paul Magron, Konstantinos Drossos, Stylianos Ioannis Mimitakis, Tuomas Virtanen</i>	
<b>Nebula: F0 Estimation and Voicing Detection by Modeling the Statistical Properties of Feature Extractors</b> .....	337
<i>Kanru Hua</i>	
<b>Real-time Single-channel Dereverberation and Separation with Time-domain Audio Separation Network</b> .....	342
<i>Yi Luo, Nima Mesgarani</i>	
<b>Music Source Activity Detection and Separation Using Deep Attractor Network</b> .....	347
<i>Rajath Kumar, Yi Luo, Nima Mesgarani</i>	

<b>Improving Mandarin Tone Recognition Using Convolutional Bidirectional Long Short-Term Memory with Attention</b> .....	352
<i>Longfei Yang, Yanlu Xie, Jinsong Zhang</i>	

### ACOUSTIC ANALYSIS-SYNTHESIS OF SPEECH DISORDERS

<b>Vowel Space as a Tool to Evaluate Articulation Problems</b> .....	357
<i>Rob Van Son, Catherine Middag, Kris Demuyne</i>	
<b>Towards a Better Characterization of Parkinsonian Speech: A Multidimensional Acoustic Study</b> .....	362
<i>Veronique Delvaux, Kathy Huet, Myriam Piccaluga, Sophie Van Malderen, Bernard Harmegnies</i>	
<b>Self-similarity Matrix Based Intelligibility Assessment of Cleft Lip and Palate Speech</b> .....	367
<i>Sishir Kalita, S R Mahadeva Prasanna, Samarendra Dandapat</i>	
<b>Pitch-Adaptive Front-end Feature for Hypernasality Detection</b> .....	372
<i>Akhilesh Kumar Dubey, S R Mahadeva Prasanna, S Dandapat</i>	
<b>Detection of Amyotrophic Lateral Sclerosis (ALS) via Acoustic Analysis</b> .....	377
<i>Raquel Norel, Mary Pietrowicz, Carla Agurto, Shay Rishoni, Guillermo Cecchi</i>	
<b>Detection of Glottal Activity Errors in Production of Stop Consonants in Children with Cleft Lip and Palate</b> .....	382
<i>C. M. Vikram, S R Mahadeva Prasanna, Ajish K Abraham, M. Pushpavathi, K. S. Girish</i>	

### ASR SYSTEMS AND TECHNOLOGIES

<b>Cold Fusion: Training Seq2Seq Models Together with Language Models</b> .....	387
<i>Anuroop Sriram, Heewoo Jun, Sanjeev Satheesh, Adam Coates</i>	
<b>Investigation on Estimation of Sentence Probability by Combining Forward, Backward and Bi-directional LSTM-RNNs</b> .....	392
<i>Kazuki Irie, Zhihong Lei, Liuhui Deng, Ralf Schlüter, Hermann Ney</i>	
<b>Subword and Crossword Units for CTC Acoustic Models</b> .....	396
<i>Thomas Zenkel, Ramon Sanabria, Florian Metze, Alex Waibel</i>	
<b>Neural Error Corrective Language Models for Automatic Speech Recognition</b> .....	401
<i>Tomohiro Tanaka, Ryo Masumura, Hirokazu Masataki, Yushi Aono</i>	
<b>Entity-Aware Language Model as an Unsupervised Reranker</b> .....	406
<i>Mohammad Sadegh Rasooli, Sarangarajan Parthasarathy</i>	
<b>Character-level Language Modeling with Gated Hierarchical Recurrent Neural Networks</b> .....	411
<i>Iksoo Choi, Jinhwan Park, Wonyong Sung</i>	

### DECEPTION, PERSONALITY, AND CULTURE ATTRIBUTE

<b>Acoustic-Prosodic Indicators of Deception and Trust in Interview Dialogues</b> .....	416
<i>Sarah Ita Levitan, Angel Maredia, Julia Hirschberg</i>	
<b>Deep Personality Recognition for Deception Detection</b> .....	421
<i>Guozhen An, Sarah Ita Levitan, Julia Hirschberg, Rivka Levitan</i>	
<b>Cross-cultural (A)symmetries in Audio-visual Attitude Perception</b> .....	426
<i>Hansjörg Mixdorff, Albert Rilliard, Tan Lee, Matthew K. H. Ma, Angelika Hönemann</i>	
<b>An Active Feature Transformation Method for Attitude Recognition of Video Bloggers</b> .....	431
<i>Fasih Haider, Fahim A. Salim, Owen Conlan, Saturnino Luz</i>	
<b>Automatic Assessment of Individual Culture Attribute of Power Distance Using a Social Context-Enhanced Prosodic Network Representation</b> .....	436
<i>Fu-Sheng Tsai, Hao-Chun Yang, Wei-Wen Chang, Chi-Chun Lee</i>	
<b>Analysis and Detection of Phonation Modes in Singing Voice using Excitation Source Features and Single Frequency Filtering Cepstral Coefficients (SFFCC)</b> .....	441
<i>Sudarsana Reddy Kadiri, Bayya Yegnanarayana</i>	

### AUTOMATIC DETECTION AND RECOGNITION OF VOICE AND SPEECH DISORDERS

<b>A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks</b> .....	446
<i>Huiyi Wu, John Soraghan, Anja Lowit, Gaetano Di-Caterina</i>	
<b>Dysarthric Speech Recognition Using Time-delay Neural Network Based Denoising Autoencoder</b> .....	451
<i>Chitrakha Bhat, Biswajit Das, Bhavik Vachhani, Sunil Kumar Kopparapu</i>	
<b>A Multitask Learning Approach to Assess the Dysarthria Severity in Patients with Parkinson's Disease</b> .....	456
<i>Juan Camilo Vásquez Correa, Tomas Arias, Juan Rafael Orozco-Arroyave, Elmar Nöth</i>	
<b>The Use of Machine Learning and Phonetic Endophenotypes to Discover Genetic Variants Associated with Speech Sound Disorder</b> .....	461
<i>Jason Lilley, Erin Crowgey, H Timothy Bunnell</i>	
<b>Whistle-blowing ASRs: Evaluating the Need for More Inclusive Speech Recognition Systems</b> .....	466
<i>Meredith Moore, Hemanth Venkateswara, Sethuraman Panchanathan</i>	

<b>Data Augmentation Using Healthy Speech for Dysarthric Speech Recognition</b> .....	471
<i>Bhavik Vachhani, Chitralkha Bhat, Sunil Kumar Kopparapu</i>	

## **VOICE CONVERSION**

<b>Improving Sparse Representations in Exemplar-Based Voice Conversion with a Phoneme-Selective Objective Function</b> .....	476
<i>Shaojin Ding, Guanlong Zhao, Christopher Liberatoro, Ricardo Gutierrez-Osuna</i>	
<b>Learning Structured Dictionaries for Exemplar-based Voice Conversion</b> .....	481
<i>Shaojin Ding, Christopher Liberatoro, Ricardo Gutierrez-Osuna</i>	
<b>Exemplar-Based Spectral Detail Compensation for Voice Conversion</b> .....	486
<i>Yu-Huai Peng, Hsin-Te Hwang, Yichiao Wu, Yu Tsao, Hsin-Min Wang</i>	
<b>Whispered Speech to Neutral Speech Conversion Using Bidirectional LSTMs</b> .....	491
<i>G. Nisha Meenakshi, Prasanta Kumar Ghosh</i>	
<b>Voice Conversion Across Arbitrary Speakers Based on a Single Target-Speaker Utterance</b> .....	496
<i>Songxiang Liu, Jinghua Zhong, Lijia Sun, Xixin Wu, Xunying Liu, Helen Meng</i>	
<b>Multi-target Voice Conversion without Parallel Data by Adversarially Learning Disentangled Audio Representations</b> .....	501
<i>Ju-Chieh Chou, Cheng-Chieh Yeh, Hung-Yi Lee, Lin-Shan Lee</i>	

## **THE INTERSPEECH 2018 COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE): ATYPICAL & SELF-ASSESSED AFFECT, CRYING & HEART BEATS 2**

<b>Attention-based Sequence Classification for Affect Detection</b> .....	506
<i>Cristina Gorrostieta, Richard Brutti, Kye Taylor, Avi Shapiro, Joseph Moran, Ali Azarbayejani, John Kane</i>	
<b>Computational Paralinguistics: Automatic Assessment of Emotions, Mood and Behavioural State from Acoustics of Speech</b> .....	511
<i>Zafri Sherhan Syed, Julien Schroeter, Kirill Sidorov, David Marshall</i>	
<b>Investigating Utterance Level Representations for Detecting Intent from Acoustics</b> .....	516
<i>Saikrishna Rallabandi, Bhavya Karki, Carla Viegas, Eric Nyberg, Alan W Black</i>	
<b>LSTM Based Cross-corpus and Cross-task Acoustic Emotion Recognition</b> .....	521
<i>Heysem Kaya, Dmitrii Fedotov, Ali Yesilkanat, Oxana Verkholyak, Yang Zhang, Alexey Karpov</i>	
<b>Implementing Fusion Techniques for the Classification of Paralinguistic Information</b> .....	526
<i>Bogdan Vlasenko, Jilt Sebastian, D. S. Pavan Kumar, Mathew Magimai.-Doss</i>	
<b>General Utterance-Level Feature Extraction for Classifying Crying Sounds, Atypical &amp; Self-Assessed Affect and Heart Beats</b> .....	531
<i>Gábor Gosztolya, Tamás Grósz, László Tóth</i>	
<b>Self-Assessed Affect Recognition Using Fusion of Attentional BLSTM and Static Acoustic Features</b> .....	536
<i>Bo-Hao Su, Sung-Lin Yeh, Ming-Ya Ko, Huan-Yu Chen, Shun-Chang Zhong, Jeng-Lin Li, Chi-Chun Lee</i>	
<b>Vocalic, Lexical and Prosodic Cues for the INTERSPEECH 2018 Self-Assessed Affect Challenge</b> .....	541
<i>Claude Montacié, Marie-José Caraty</i>	

## **SHOW AND TELL 2**

<b>Intonation tutor by SPIRE (In-SPIRE): An Online Tool for an Automatic Feedback to the Second Language Learners in Learning Intonation</b> .....	546
<i>P. A. Anand, Chiranjeevi Yarra, N. K. Kausthubha, Prasanta Kumar Ghosh</i>	
<b>Game-based Spoken Dialog Language Learning Applications for Young Students</b> .....	548
<i>Keelan Evanini, Veronika Timpe-Laughlin, Eugene Tsuprun, Ian Blood, Jeremy Lee, James Bruno, Vikram Ramanarayanan, Patrick Lange, David Suendermann-Oefi</i>	
<b>The IBM Virtual Voice Creator</b> .....	550
<i>Alexander Sorin, Slava Shechtman, Zvi Kons, Ron Hoory, Shay Ben-David, Joe Pavitt, Shai Rozenberg, Carmel Rabinovitz, Tal Drory</i>	
<b>Mobile Application for Learning Languages for the Unlettered</b> .....	552
<i>G. Gayathri, N. Mohana, Radhika Pal, Hema Murthy</i>	
<b>Mandarin-English Code-switching Speech Recognition</b> .....	554
<i>Haihua Xu, Van Tung Pham, Zin Tun Kyaw, Zhi Hao Lim, Eng Siong Chng, Haizhou Li</i>	

## **SPOKEN DIALOGUE SYSTEMS AND CONVERSATIONAL ANALYSIS**

<b>Joint Learning of Domain Classification and Out-of-Domain Detection with Dynamic Class Weighting for Satisficing False Acceptance Rates</b> .....	556
<i>Joo-Kyung Kim, Young-Bum Kim</i>	
<b>Analyzing Vocal Tract Movements During Speech Accommodation</b> .....	561
<i>Sankar Mukherjee, Thierry Legou, Leonardo Lancia, Pauline Hilt, Alice Tomassini, Luciano Fadiga, Alessandro D'Ausilio, Leonardo Badino, Noël Nguyen</i>	
<b>Cross-Lingual Multi-Task Neural Architecture for Spoken Language Understanding</b> .....	566
<i>Yujiang Li, Xuemin Zhao, Weiqun Xu, Yonghong Yan</i>	

<b>Statistical Model Compression for Small-Footprint Natural Language Understanding</b> .....	571
<i>Grant P. Srimel, Kanthashree Mysore Sathyendra, Stanislav Peshterliev</i>	
<b>Comparison of an End-to-end Trainable Dialogue System with a Modular Statistical Dialogue System</b> .....	576
<i>Norbert Braunschweiler, Alexandros Papangelis</i>	
<b>A Discriminative Acoustic-Prosodic Approach for Measuring Local Entrainment</b> .....	581
<i>Megan Willi, Stephanie A. Borrie, Tyson S. Barrett, Ming Tu, Visar Berisha</i>	
<b>Investigating Speech Features for Continuous Turn-Taking Prediction Using LSTMs</b> .....	586
<i>Matthew Roddy, Gabriel Skantze, Naomi Harte</i>	
<b>Classification of Correction Turns in Multilingual Dialogue Corpus</b> .....	591
<i>Ivan Kraljevski, Diane Hirschfeld</i>	
<b>Contextual Slot Carryover for Disparate Schemas</b> .....	596
<i>Chetan Naik, Arpit Gupta, Hancheng Ge, Mathias Lambert, Ruhi Sarikaya</i>	
<b>Capsule Networks for Low Resource Spoken Language Understanding</b> .....	601
<i>Vincent Renkens, Hugo Van Hamme</i>	
<b>Intent Discovery Through Unsupervised Semantic Text Clustering</b> .....	606
<i>Padmasundari, Srinivas Bangalore</i>	
<b>Multimodal Polynomial Fusion for Detecting Driver Distraction</b> .....	611
<i>Yulun Du, Alan W Black, Louis-Philippe Morency, Maxine Eskenazi</i>	
<b>Engagement Recognition in Spoken Dialogue via Neural Network by Aggregating Different Annotators' Models</b> .....	616
<i>Koji Inoue, Divesh Lala, Katsuya Takanashi, Tatsuya Kawahara</i>	
<b>A First Investigation of the Timing of Turn-taking in Ruuli</b> .....	621
<i>Tuarik Buanzur, Margaret Zellers, Saudah Namyalo, Alena WitzJack-Makarevich</i>	

## **SPOOFING DETECTION**

<b>Spoofing Detection Using Adaptive Weighting Framework and Clustering Analysis</b> .....	626
<i>Yuanjun Zhao, Roberto Togneri, Victor Sreeram</i>	
<b>Exploration of Compressed ILPR Features for Replay Attack Detection</b> .....	631
<i>Sarfraz Jelil, Sishir Kalita, S R Mahadeva Prasanna, Rohit Sinha</i>	

## **VOLUME 2**

<b>Detection of Replay-Spoofing Attacks Using Frequency Modulation Features</b> .....	636
<i>Tharshini Gunendradasan, Buddhi Wickramasinghe, Ngoc Phu Le, Eliathamby Ambikairajah, Julien Epps</i>	
<b>Effectiveness of Speech Demodulation-Based Features for Replay Detection</b> .....	641
<i>Madhu Kamble, Hemlata Tak, Hemant Patil</i>	
<b>Novel Variable Length Energy Separation Algorithm Using Instantaneous Amplitude Features for Replay Detection</b> .....	646
<i>Madhu Kamble, Hemant Patil</i>	
<b>Feature with Complementarity of Statistics and Principal Information for Spoofing Detection</b> .....	651
<i>Jichen Yang, Changhuai You, Qianhua He</i>	
<b>Multiple Phase Information Combination for Replay Attacks Detection</b> .....	656
<i>Dongbo Li, Longbiao Wang, Jianwu Dang, Meng Liu, Zeyan Oo, Seiichi Nakagawa, Haotian Guan, Xiangang Li</i>	
<b>Frequency Domain Linear Prediction Features for Replay Spoofing Attack Detection</b> .....	661
<i>Buddhi Wickramasinghe, Saad Irtza, Eliathamby Ambikairajah, Julien Epps</i>	
<b>Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection</b> .....	666
<i>Hardik Sailor, Madhu Kamble, Hemant Patil</i>	
<b>Deep Siamese Architecture Based Replay Detection for Secure Voice Biometric</b> .....	671
<i>Kaavya Sriskandaraja, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i>	
<b>A Deep Identity Representation for Noise Robust Spoofing Detection</b> .....	676
<i>Alejandro Gómez Alanís, Antonio M. Peinado, Jose A. Gonzalez, Angel Gomez</i>	
<b>End-To-End Audio Replay Attack Detection Using Deep Convolutional Networks with Attention</b> .....	681
<i>Francis Tom, Mohit Jain, Prasenjit Dey</i>	
<b>Decision-level Feature Switching as a Paradigm for Replay Attack Detection</b> .....	686
<i>M. S. Saranya, Hema Murthy</i>	
<b>Modulation Dynamic Features for the Detection of Replay Attacks</b> .....	691
<i>Gajan Suthokumar, Vidhyasaharan Sethu, Chamith Wijenayake, Eliathamby Ambikairajah</i>	

## **SPEECH ANALYSIS AND REPRESENTATION**

<b>On the Usefulness of the Speech Phase Spectrum for Pitch Extraction</b> .....	696
<i>Erfan Loweimi, Jon Barker, Thomas Hain</i>	
<b>Time-regularized Linear Prediction for Noise-robust Extraction of the Spectral Envelope of Speech</b> .....	701
<i>Manu Airaksinen, Lauri Juvela, Okko Räsänen, Paavo Alku</i>	
<b>Auditory Filterbank Learning Using ConvRBM for Infant Cry Classification</b> .....	706
<i>Hardik B. Sailor, Hemant Patil</i>	
<b>Effectiveness of Dynamic Features in INCA and Temporal Context-INCA</b> .....	711
<i>Nirmesh Shah, Hemant Patil</i>	



<b>Singing Voice Phoneme Segmentation by Hierarchically Inferring Syllable and Phoneme Onset Positions</b> .....	716
<i>Rong Gong, Xavier Serra</i>	
<b>Novel Empirical Mode Decomposition Cepstral Features for Replay Spoof Detection</b> .....	721
<i>Prasad Tapkir, Hemant Patil</i>	
<b>Novel Linear Frequency Residual Cepstral Features for Replay Attack Detection</b> .....	726
<i>Hemlata Tak, Hemant Patil</i>	
<b>Analysis of Sparse Representation Based Feature on Speech Mode Classification</b> .....	731
<i>Kumud Tripathi, K. Sreenivasa Rao</i>	
<b>Multicomponent 2-D AM-FM Modeling of Speech Spectrograms</b> .....	736
<i>Jitendra Kumar Dhiman, Neeraj Sharma, Chandra Sekhar Seelamantula</i>	
<b>An Optimization Framework for Recovery of Speech from Phase-Encoded Spectrograms</b> .....	741
<i>Abhilash Sainathan, Sunil Rudresh, Chandra Sekhar Seelamantula</i>	
<b>Speaker Recognition with Nonlinear Distortion: Clipping Analysis and Impact</b> .....	746
<i>Wei Xia, John H. L. Hansen</i>	
<b>Linear Prediction Residual based Short-term Cepstral Features for Replay Attacks Detection</b> .....	751
<i>Madhusudan Singh, Debadatta Pati</i>	
<b>Analysis of Variational Mode Functions for Robust Detection of Vowels</b> .....	756
<i>Surbhi Sakshi, Avinash Kumar, Gayadhar Pradhan</i>	

## **SEQUENCE MODELS FOR ASR**

<b>Improving Attention Based Sequence-to-Sequence Models for End-to-End English Conversational Speech Recognition</b> .....	761
<i>Chao Weng, Jia Cui, Guangsen Wang, Jun Wang, Chengzhu Yu, Dan Su, Dong Yu</i>	
<b>Segmental Encoder-Decoder Models for Large Vocabulary Automatic Speech Recognition</b> .....	766
<i>Eugen Beck, Mirko Hannemann, Patrick Dötsch, Ralf Schlüter, Hermann Ney</i>	
<b>Acoustic Modeling with DFSSN-CTC and Joint CTC-CE Learning</b> .....	771
<i>Shiliang Zhang, Ming Lei</i>	
<b>End-to-End Speech Command Recognition with Capsule Network</b> .....	776
<i>Jaesung Bae, Dae-Shik Kim</i>	
<b>End-to-End Speech Recognition from the Raw Waveform</b> .....	781
<i>Neil Zeghidour, Nicolas Usunier, Gabriel Synnaeve, Ronan Collobert, Emmanuel Dupoux</i>	
<b>A Multistage Training Framework for Acoustic-to-Word Model</b> .....	786
<i>Chengzhu Yu, Chunlei Zhang, Chao Weng, Jia Cui, Dong Yu</i>	
<b>Syllable-Based Sequence-to-Sequence Speech Recognition with the Transformer in Mandarin Chinese</b> .....	791
<i>Shiyu Zhou, Linhao Dong, Shuang Xu, Bo Xu</i>	
<b>Densely Connected Networks for Conversational Speech Recognition</b> .....	796
<i>Kyu Han, Akshay Chandrashekar, Jungsuk Kim, Ian Lane</i>	
<b>Multi-Head Decoder for End-to-End Speech Recognition</b> .....	801
<i>Tomoki Hayashi, Shinji Watanabe, Tomoki Toda, Kazuya Takeda</i>	
<b>Compressing End-to-end ASR Networks by Tensor-Train Decomposition</b> .....	806
<i>Takuma Mori, Andros Tjandra, Sakriani Sakti, Satoshi Nakamura</i>	
<b>Speech2Vec: A Sequence-to-Sequence Framework for Learning Word Embeddings from Speech</b> .....	811
<i>Yu-An Chung, James Glass</i>	
<b>Extending Recurrent Neural Aligner for Streaming End-to-End Speech Recognition in Mandarin</b> .....	816
<i>Linhao Dong, Shiyu Zhou, Wei Chen, Bo Xu</i>	

## **SOURCE SEPARATION AND SPATIAL ANALYSIS**

<b>Joint Noise and Reverberation Adaptive Learning for Robust Speaker DOA Estimation with an Acoustic Vector Sensor</b> .....	821
<i>Disong Wang, Yuexian Zou</i>	
<b>Multiple Concurrent Sound Source Tracking Based on Observation-Guided Adaptive Particle Filter</b> .....	826
<i>Hong Liu, Haipeng Lan, Bing Yang, Cheng Pang</i>	
<b>Harmonic-Percussive Source Separation of Polyphonic Music by Suppressing Impulsive Noise Events</b> .....	831
<i>M. Gurunath Reddy, K. Sreenivasa Rao, Partha Pratim Das</i>	
<b>Speaker Activity Detection and Minimum Variance Beamforming for Source Separation</b> .....	836
<i>Enea Ceolini, Jithendar Anumula, Adrian Huber, Ilya Kiselev, Shih-Chii Liu</i>	
<b>Sparsity-Constrained Weight Mapping for Head-Related Transfer Functions Individualization from Anthropometric Features</b> .....	841
<i>Xiaoqi Qi, Jianhua Tao</i>	
<b>Speech Source Separation Using ICA in Constant Q Transform Domain</b> .....	846
<i>D. V. L. N. Dheeraj Sai, K. S. Kishor, Kodukula Sri Rama Murthy</i>	
<b>Multi-talker Speech Separation Based on Permutation Invariant Training and Beamforming</b> .....	851
<i>Lu Yin, Ziteng Wang, Risheng Xia, Junfeng Li, Yonghong Yan</i>	
<b>Expectation-Maximization Algorithms for Itakura-Saito Nonnegative Matrix Factorization</b> .....	856
<i>Paul Magron, Tuomas Virtanen</i>	

<b>Subband Weighting for Binaural Speech Source Localization</b> .....	861
<i>Karthik Girija Ramesan, Parth Suresh, Prasanta Kumar Ghosh</i>	

### **PLENARY TALK-1**

<b>Universal Tendencies for Cross-Linguistic Prosodic Tendencies: A Review and Some New Proposals</b> .....	866
<i>Jacqueline Vaissière</i>	

### **ACOUSTIC MODEL ADAPTATION**

<b>Learning to Adapt: A Meta-learning Approach for Speaker Adaptation</b> .....	867
<i>Ondrej Klejch, Joachim Fainberg, Peter Bell</i>	
<b>Speaker Adaptation and Adaptive Training for Jointly Optimised Tandem Systems</b> .....	872
<i>Yu Wang, Chao Zhang, Mark Gales, Philip Woodland</i>	
<b>Comparison of BLSTM-Layer-Specific Affine Transformations for Speaker Adaptation</b> .....	877
<i>Markus Kitzka, Ralf Schlüter, Hermann Ney</i>	
<b>Correlational Networks for Speaker Normalization in Automatic Speech Recognition</b> .....	882
<i>Rini A Sharon, Sandeep Reddy Kothinti, Umesh Srinivasan</i>	
<b>Machine Speech Chain with One-shot Speaker Adaptation</b> .....	887
<i>Andros Tjandra, Sakriani Sakti, Satoshi Nakamura</i>	
<b>Domain Adaptation Using Factorized Hidden Layer for Robust Automatic Speech Recognition</b> .....	892
<i>Khe Chai Sim, Arun Narayanan, Ananya Misra, Anshuman Tripathi, Golan Pundak, Tara Sainath, Parisa Haghani, Bo Li, Michiel Bacchiani</i>	

### **STATISTICAL PARAMETRIC SPEECH SYNTHESIS**

<b>Waveform-Based Speaker Representations for Speech Synthesis</b> .....	897
<i>Moquan Wan, Gilles Degottex, Mark J. F. Gales</i>	
<b>Incremental TTS for Japanese Language</b> .....	902
<i>Tomoya Yanagita, Sakriani Sakti, Satoshi Nakamura</i>	
<b>Transfer Learning Based Progressive Neural Networks for Acoustic Modeling in Statistical Parametric Speech Synthesis</b> .....	907
<i>Ruibo Fu, Jianhua Tao, Yibin Zheng, Zhengqi Wen</i>	
<b>A Unified Framework for the Generation of Glottal Signals in Deep Learning-based Parametric Speech Synthesis Systems</b> .....	912
<i>Min-Jae Hwang, Eunwoo Song, Jin-Seob Kim, Hong-Goo Kang</i>	
<b>Acoustic Modeling Using Adversarially Trained Variational Recurrent Neural Network for Speech Synthesis</b> .....	917
<i>Joun Yeop Lee, Sung Jun Cheon, Byoung Jin Choi, Nam Soo Kim, Eunwoo Song</i>	
<b>On the Application and Compression of Deep Time Delay Neural Network for Embedded Statistical Parametric Speech Synthesis</b> .....	922
<i>Yibin Zheng, Jianhua Tao, Zhengqi Wen, Ruibo Fu</i>	

### **EMOTION MODELING**

<b>Integrating Recurrence Dynamics for Speech Emotion Recognition</b> .....	927
<i>Efthymios Tzinis, Georgios Paraskevopoulos, Christos Baziotis, Alexandros Potamianos</i>	
<b>Towards Temporal Modelling of Categorical Speech Emotion Recognition</b> .....	932
<i>Wenjing Han, Huabin Ruan, Xiaomin Chen, Zhixiang Wang, Haifeng Li, Björn Schuller</i>	
<b>Emotion Recognition from Human Speech Using Temporal Information and Deep Learning</b> .....	937
<i>John Kim, Rif A. Saurous</i>	
<b>Role of Regularization in the Prediction of Valence from Speech</b> .....	941
<i>Kusha Sridhar, Srinivas Parthasarathy, Carlos Busso</i>	
<b>Learning Spontaneity to Improve Emotion Recognition in Speech</b> .....	946
<i>Karttikeya Mangalam, Tanaya Guha</i>	
<b>Predicting Categorical Emotions by Jointly Learning Primary and Secondary Emotions through Multitask Learning</b> .....	951
<i>Reza Lotfian, Carlos Busso</i>	

### **MODELS OF SPEECH PERCEPTION**

<b>Picture Naming or Word Reading: Does the Modality Affect Speech Motor Adaptation and Its Transfer?</b> .....	956
<i>Tiphaine Caudrelier, Pascal Perrier, Jean-Luc Schwartz, Amélie Rochet-Capellan</i>	
<b>Measuring the Band Importance Function for Mandarin Chinese with a Bayesian Adaptive Procedure</b> .....	961
<i>Yufan Du, Yi Shen, Hongying Yang, Xihong Wu, Jing Chen</i>	
<b>Wide Learning for Auditory Comprehension</b> .....	966
<i>Elnaz Shafaei-Bajestan, R. Harald Baayen</i>	

<b>Analyzing Reaction Time Sequences from Human Participants in Auditory Experiments</b> .....	971
<i>Louis Ten Bosch, Mirjam Ernestus, Lou Boves</i>	
<b>Prediction of Perceived Speech Quality Using Deep Machine Listening</b> .....	976
<i>Jasper Ooster, Rainer Huber, Bernd T. Meyer</i>	
<b>Prediction of Subjective Listening Effort from Acoustic Data with Non-Intrusive Deep Models</b> .....	981
<i>Paul Kranzusch, Rainer Huber, Melanie Krüger, Birger Kollmeier, Bernd T. Meyer</i>	

## **MULTIMODAL DIALOGUE SYSTEMS**

<b>A Case Study on the Importance of Belief State Representation for Dialogue Policy Management</b> .....	986
<i>Margarita Kotti, Vassilios Diakouloukas, Alexandros Papangelis, Michail Lagoudakis, Yannis Stylianou</i>	
<b>Prediction of Turn-taking Using Multitask Learning with Prediction of Backchannels and Fillers</b> .....	991
<i>Kohei Hara, Koji Inoue, Katsuya Takanashi, Tatsuya Kawahara</i>	
<b>Conversational Analysis Using Utterance-level Attention-based Bidirectional Recurrent Neural Networks</b> .....	996
<i>Chandrakant Bothe, Sven Magg, Cornelius Weber, Stefan Wermter</i>	
<b>A Comparative Study of Statistical Conversion of Face to Voice Based on Their Subjective Impressions</b> .....	1001
<i>Yasuhito Ohsugi, Daisuke Saito, Nobuaki Minematsu</i>	
<b>Follow-up Question Generation Using Pattern-based Seq2seq with a Small Corpus for Interview Coaching</b> .....	1006
<i>Ming-Hsiang Su, Chung-Hsien Wu, Kun-Yi Huang, Qian-Bei Hong, Huai-Hung Huang</i>	
<b>Coherence Models for Dialogue</b> .....	1011
<i>Alessandra Cervone, Evgeny Stepanov, Giuseppe Riccardi</i>	

## **SPEECH RECOGNITION FOR INDIAN LANGUAGES**

<b>Indian Languages ASR: A Multilingual Phone Recognition Framework with IPA Based Common Phone-set, Predicted Articulatory Features and Feature fusion</b> .....	1016
<i>K. E. Manjunath, K. Sreenivasa Rao, Dinesh Babu Jayagopi, V Ramasubramanian</i>	
<b>Rapid Collection of Spontaneous Speech Corpora Using Telephonic Community Forums</b> .....	1021
<i>Agha Ali Raza, Awais Athar, Shan Randhawa, Zain Tariq, Muhammad Bilal Saleem, Haris Bin Zia, Umar Saif, Roni Rosenfeld</i>	
<b>Effect of TTS Generated Audio on OOV Detection and Word Error Rate in ASR for Low-resource Languages</b> .....	1026
<i>Savitha Murthy, Dinkar Sitaram, Sunayana Sitaram</i>	
<b>Development of Large Vocabulary Speech Recognition System with Keyword Search for Manipuri</b> .....	1031
<i>Tamvina Patel, D. N. Krishna, Noor Fathima, Nisar Shah, C. Mahima, Deepak Kumar, Anuroop Iyengar</i>	
<b>Robust Mizo Continuous Speech Recognition</b> .....	1036
<i>Abhishek Dey, Biswajit Dev Sarma, Wendy Lalhminghlui, Lalnunsiami Ngente, Parismita Gogoi, Priyankoo Sarmah, S R Mahadeva Prasanna, Rohit Sinha, S. R. Nirmala</i>	
<b>Semi-supervised and Active-learning Scenarios: Efficient Acoustic Model Refinement for a Low Resource Indian Language</b> .....	1041
<i>Maharajan Chellapriyadharshini, Anoop Toffy, K. M. Srinivasa Raghavan, V Ramasubramanian</i>	
<b>Automatic Speech Recognition with Articulatory Information and a Unified Dictionary for Hindi, Marathi, Bengali and Oriya</b> .....	1046
<i>Debadatta Dash, Myungjong Kim, Kristin Teplansky, Jun Wang</i>	

## **SHOW AND TELL 3**

<b>Captaina: Integrated Pronunciation Practice and Data Collection Portal</b> .....	1051
<i>Aku Rouhe, Reima Karhila, Aija Elg, Minnaleena Toivola, Peter Smit, Anna-Riikka Smolander, Mikko Kurimo</i>	
<b>auMina™ - Enterprise Speech Analytics</b> .....	1053
<i>Umesh Sachdev, Rajagopal Jayaraman, Zainab Millwala</i>	
<b>HoloCompanion: An MR Friend for Everyone</b> .....	1055
<i>Annam Naresh, Rushabh Gandhi, Mallikarjuna Rao Bellamkonda, Mithun Das Gupta</i>	
<b>akeira™ - Virtual Assistant</b> .....	1057
<i>Umesh Sachdev, Rajagopal Jayaraman, Zainab Millwala</i>	
<b>Brain-Computer Interface using Electroencephalogram Signatures of Eye Blinks</b> .....	1059
<i>Srihari Maruthachalam, Sidharth Aggarwal, Mari Ganesh Kumar, Mriganka Sur, Hema Murthy</i>	

## **SPEAKER VERIFICATION II**

<b>Voice Comparison and Rhythm: Behavioral Differences between Target and Non-target Comparisons</b> .....	1061
<i>Moez Ajili, Jean-François Bonastre, Solange Rossato</i>	
<b>Co-whitening of I-vectors for Short and Long Duration Speaker Verification</b> .....	1066
<i>Longting Xu, Kong Aik Lee, Haizhou Li, Zhen Yang</i>	
<b>Compensation for Domain Mismatch in Text-independent Speaker Recognition</b> .....	1071
<i>Fahimeh Bahmaninezhad, John H. L. Hansen</i>	
<b>Joint Learning of J-Vector Extractor and Joint Bayesian Model for Text Dependent Speaker Verification</b> .....	1076
<i>Ziqiang Shi, Liu Liu, Huibin Lin, Rujie Liu</i>	
<b>Latent Factor Analysis of Deep Bottleneck Features for Speaker Verification with Random Digit Strings</b> .....	1081
<i>Ziqiang Shi, Huibin Lin, Liu Liu, Rujie Liu</i>	

<b>VoxCeleb2: Deep Speaker Recognition</b> .....	1086
<i>Joon Son Chung, Arsha Nagrani, Andrew Senior</i>	
<b>Supervised I-vector Modeling - Theory and Applications</b> .....	1091
<i>Shreyas Ramoji, Sriram Ganapathy</i>	
<b>LOCUST - Longitudinal Corpus and Toolset for Speaker Verification</b> .....	1096
<i>Evgeny Dmitriev, Yulia Kim, Anastasia Matveeva, Claude Montacié, Yannick Boulard, Yadviga Sinyavskaya, Yulia Zhukova, Adam Zarazinski, Egor Akhanov, Ilya Viksnin, Andrei Shlykov, Maria Usova</i>	
<b>Analysis of Language Dependent Front-End for Speaker Recognition</b> .....	1101
<i>Srikanth Madikeri, Subhadeep Dey, Petr Motlicek</i>	
<b>Robust Speaker Recognition from Distant Speech under Real Reverberant Environments Using Speaker Embeddings</b> .....	1106
<i>Mahesh Kumar Nandwana, Julien Van Hout, Mitchell McLaren, Allen Stauffer, Colleen Richey, Aaron Lawson, Martin Graciarena</i>	
<b>Investigation on Bandwidth Extension for Speaker Recognition</b> .....	1111
<i>Phani Sankar Nidadavolu, Cheng-I Lai, Jesús Villalba, Najim Dehak</i>	
<b>On Learning Vocal Tract System Related Speaker Discriminative Information from Raw Signal Using CNNs</b> .....	1116
<i>Hannah Muckenhirn, Mathew Magimai.-Doss, Sebastien Marcel</i>	
<b>On Convolutional LSTM Modeling for Joint Wake-Word Detection and Text Dependent Speaker Verification</b> .....	1121
<i>Rajath Kumar, Vaishnavi Yeruva, Sriram Ganapathy</i>	
<b>Cosine Metric Learning for Speaker Verification in the I-vector Space</b> .....	1126
<i>Zhongxin Bai, Xiao-Lei Zhang, Jingdong Chen</i>	
<b>An Unsupervised Neural Prediction Framework for Learning Speaker Embeddings Using Recurrent Neural Networks</b> .....	1131
<i>Arindam Jati, Panayiotis Georgiou</i>	

## **NOVEL APPROACHES TO ENHANCEMENT**

<b>A New Framework for Supervised Speech Enhancement in the Time Domain</b> .....	1136
<i>Ashutosh Pandey, Deliang Wang</i>	
<b>Speech Enhancement Using the Minimum-probability-of-error Criterion</b> .....	1141
<i>Jishnu Sadasivan, Subhadip Mukherjee, Chandra Sekhar Seelamantula</i>	
<b>Exploring the Relationship between Conic Affinity of NMF Dictionaries and Speech Enhancement Metrics</b> .....	1146
<i>Pavlos Papadopoulos, Colin Vaz, Shrikanth Narayanan</i>	
<b>Using Shifted Real Spectrum Mask as Training Target for Supervised Speech Separation</b> .....	1151
<i>Yun Liu, Hui Zhang, Xueliang Zhang</i>	
<b>Enhancement of Noisy Speech Signal by Non-Local Means Estimation of Variational Mode Functions</b> .....	1156
<i>Nagapuri Srinivas, Gayadhar Pradhan, Syed Shahmawazuddin</i>	
<b>Phase-locked Loop (PLL) Based Phase Estimation in Single Channel Speech Enhancement</b> .....	1161
<i>Priya Pallavi, Ch V Rama Rao</i>	
<b>Cycle-Consistent Speech Enhancement</b> .....	1165
<i>Zhong Meng, Jinyu Li, Yifan Gong, Bing-Hwang (Fred) Juang</i>	
<b>Visual Speech Enhancement</b> .....	1170
<i>Aviv Gabbay, Asaph Shamir, Shmuel Peleg</i>	
<b>Implementation of Digital Hearing Aid as a Smartphone Application</b> .....	1175
<i>Saket Sharma, Nitya Tiwari, Prem C. Pandey</i>	
<b>Bone-Conduction Sensor Assisted Noise Estimation for Improved Speech Enhancement</b> .....	1180
<i>Ching-Hua Lee, Bhaskar D. Rao, Harinath Garudadri</i>	
<b>Artificial Bandwidth Extension with Memory Inclusion Using Semi-supervised Stacked Auto-encoders</b> .....	1185
<i>Pramod Bachhav, Massimiliano Todisco, Nicholas Evans</i>	
<b>Large Vocabulary Concatenative Resynthesis</b> .....	1190
<i>Soumi Maiti, Joey Ching, Michael Mandel</i>	
<b>Concatenative Resynthesis with Improved Training Signals for Speech Enhancement</b> .....	1195
<i>Ali Raza Syed, Viet Anh Trinh, Michael Mandel</i>	

## **SYLLABIFICATION, RHYTHM, AND VOICE ACTIVITY DETECTION**

<b>Comparison of Syllabification Algorithms and Training Strategies for Robust Word Count Estimation across Different Languages and Recording Conditions</b> .....	1200
<i>Okko Räsänen, Seshadri Shreyas, Marisa Casillas</i>	
<b>A Comparison of Input Types to a Deep Neural Network-based Forced Aligner</b> .....	1205
<i>Matthew C. Kelley, Benjamin V. Tucker</i>	
<b>Joint Learning Using Denoising Variational Autoencoders for Voice Activity Detection</b> .....	1210
<i>Youngmoon Jung, Younggwon Kim, Yeunju Choi, Hoirin Kim</i>	
<b>Information Bottleneck Based Percussion Instrument Diarization System for Taniavartanam Segments of Carnatic Music Concerts</b> .....	1215
<i>Nauman Dawlatabad, Jom Kuriakose, Chandra Sekhar Chellu, Hema Murthy</i>	
<b>Robust Voice Activity Detection Using Frequency Domain Long-Term Differential Entropy</b> .....	1220
<i>Debayon Ghosh, R. Muralishankar, Sanjeev Gurugopinath</i>	
<b>Device-directed Utterance Detection</b> .....	1225
<i>Sri Harish Mallidi, Roland Maas, Kyle Goehner, Ariya Rastrow, Spyros Matsoukas, Björn Hoffmeister</i>	

<b>Acoustic-Prosodic Features of Tabla Bol Recitation and Correspondence with the Tabla Imitation</b> .....	1229
<i>M. A. Rohit, Preeti Rao</i>	
<b>Who Said That? A Comparative Study of Non-negative Matrix Factorization Techniques</b> .....	1234
<i>Teun Krikke, Frank Broz, David Lane</i>	
<b>AVA-Speech: A Densely Labeled Dataset of Speech Activity in Movies</b> .....	1239
<i>Sourish Chaudhuri, Joseph Roth, Daniel P. W. Ellis, Andrew Gallagher, Liat Kaver, Radhika Marvin, Caroline Pantofaru, Nathan Reale, Loretta Guarino Reid, Kevin Wilson, Zhonghua Xi</i>	
<b>Audiovisual Speech Activity Detection with Advanced Long Short-Term Memory</b> .....	1244
<i>Fei Tao, Carlos Busso</i>	
<b>Towards Automatic Speech Identification from Vocal Tract Shape Dynamics in Real-time MRI</b> .....	1249
<i>Pramit Saha, Praneeth Srungarapu, Sidney Fels</i>	

### **SELECTED TOPICS IN NEURAL SPEECH PROCESSING**

<b>Structured Word Embedding for Low Memory Neural Network Language Model</b> .....	1254
<i>Kaiyu Shi, Kai Yu</i>	
<b>Role Play Dialogue Aware Language Models Based on Conditional Hierarchical Recurrent Encoder-Decoder</b> .....	1259
<i>Ryo Masumura, Tomohiro Tanaka, Atsushi Ando, Hirokazu Masataki, Yushi Aono</i>	
<b>Efficient Keyword Spotting Using Time Delay Neural Networks</b> .....	1264
<i>Samuel Myer, Vikrant Singh Tomar</i>	

### **VOLUME 3**

<b>Automatic DNN Node Pruning Using Mixture Distribution-based Group Regularization</b> .....	1269
<i>Tsukasa Yoshida, Takafumi Moriya, Kazuo Watanabe, Yusuke Shinohara, Yoshikazu Yamaguchi, Yushi Aono</i>	
<b>Conditional-Computation-Based Recurrent Neural Networks for Computationally Efficient Acoustic Modelling</b> .....	1274
<i>Raffaele Tavarone, Leonardo Badino</i>	
<b>Leveraging Translations for Speech Transcription in Low-resource Settings</b> .....	1279
<i>Antonios Anastasopoulos, David Chiang</i>	
<b>Sequence-to-sequence Neural Network Model with 2D Attention for Learning Japanese Pitch Accents</b> .....	1284
<i>Antoine Bruguier, Heiga Zen, Arkady Arkhangorodsky</i>	
<b>Task Specific Sentence Embeddings for ASR Error Detection</b> .....	1288
<i>Sahar Ghannay, Yannick Estève, Nathalie Camelin</i>	
<b>Low-Latency Neural Speech Translation</b> .....	1293
<i>Jan Niehues, Ngoc-Quan Pham, Thanh-Le Ha, Matthias Sperber, Alex Waibel</i>	
<b>Low-Resource Speech-to-Text Translation</b> .....	1298
<i>Sameer Bansal, Herman Kamper, Karen Livescu, Adam Lopez, Sharon Goldwater</i>	
<b>VoiceGuard: Secure and Private Speech Processing</b> .....	1303
<i>Ferdinand Brasser, Tommaso Frassetto, Korbinian Riedhammer, Ahmad-Reza Sadeghi, Thomas Schneider, Christian Weinert</i>	

### **PERSPECTIVE TALK-1**

<b>Deep Learning based Situated Goal-oriented Dialogue Systems</b> .....	1308
<i>Dilek Hakkani-Tür</i>	

### **DEREVERBERATION**

<b>Single-channel Speech Dereverberation via Generative Adversarial Training</b> .....	1309
<i>Chenxing Li, Tieqiang Wang, Shuang Xu, Bo Xu</i>	
<b>Single-Channel Dereverberation Using Direct MMSE Optimization and Bidirectional LSTM Networks</b> .....	1314
<i>Wolfgang Mack, Soumitro Chakrabarty, Fabian-Robert Stöter, Sebastian Braun, Bernd Edler, Emanuel Habets</i>	
<b>Single-channel Late Reverberation Power Spectral Density Estimation Using Denoising Autoencoders</b> .....	1319
<i>Ina Kodrasi, Hervé Bourlard</i>	
<b>A Non-convolutive NMF Model for Speech Dereverberation</b> .....	1324
<i>Nikhil M, Rajbabu Velmurugan, Preeti Rao</i>	
<b>Cross-Corpora Convolutional Deep Neural Network Dereverberation Preprocessing for Speaker Verification and Speech Enhancement</b> .....	1329
<i>Peter Guzewich, Stephen Zahorian, Xiao Chen, Hao Zhang</i>	
<b>Dereverberation and Beamforming in Robust Far-Field Speaker Recognition</b> .....	1334
<i>Ladislav Mošner, Oldrich Plchot, Pavel Matejka, Ondrej Novomý, Jan Cernocký</i>	

### **AUDIO EVENTS AND ACOUSTIC SCENES**

<b>Comparing the Max and Noisy-Or Pooling Functions in Multiple Instance Learning for Weakly Supervised Sequence Learning Tasks</b> .....	1339
<i>Yun Wang, Juncheng Li, Florian Metze</i>	

<b>A Simple Model for Detection of Rare Sound Events</b> .....	1344
<i>Weiran Wang, Chieh-Chi Kao, Chao Wang</i>	
<b>Temporal Transformer Networks for Acoustic Scene Classification</b> .....	1349
<i>Teng Zhang, Kailai Zhang, Ji Wu</i>	
<b>Temporal Attentive Pooling for Acoustic Event Detection</b> .....	1354
<i>Xugang Lu, Peng Shen, Sheng Li, Yu Tsao, Hisashi Kawai</i>	
<b>R-CRNN: Region-based Convolutional Recurrent Neural Network for Audio Event Detection</b> .....	1358
<i>Chieh-Chi Kao, Weiran Wang, Ming Sun, Chao Wang</i>	
<b>Detecting Media Sound Presence in Acoustic Scenes</b> .....	1363
<i>Constantinos Papayiannis, Justice Amoh, Viktor Rozgic, Shiva Sundaram, Chao Wang</i>	

## **SPEAKER DIARIZATION**

<b>S4D: Speaker Diarization Toolkit in Python</b> .....	1368
<i>Pierre-Alexandre Broux, Florent Desnous, Anthony Larcher, Simon Petitrenaud, Jean Carrive, Sylvain Meignier</i>	
<b>Multimodal Speaker Segmentation and Diarization Using Lexical and Acoustic Cues via Sequence to Sequence Neural Networks</b> .....	1373
<i>Tae Jin Park, Panayiotis Georgiou</i>	
<b>Combined Speaker Clustering and Role Recognition in Conversational Speech</b> .....	1378
<i>Nikolaos Flemotomos, Pavlos Papadopoulos, James Gibson, Shrikanth Narayanan</i>	
<b>The ACLEW DiViMe: An Easy-to-use Diarization Tool</b> .....	1383
<i>Adrien Le Franc, Eric Riebling, Julien Karadayi, Yun Wang, Camila Scaff, Florian Metz, Alejandrina Cristia</i>	
<b>Automatic Detection of Multi-speaker Fragments with High Time Resolution</b> .....	1388
<i>Evdokia Kazimirova, Andrey Belyaev</i>	
<b>Neural Speech Turn Segmentation and Affinity Propagation for Speaker Diarization</b> .....	1393
<i>Ruiqing Yin, Hervé Bredin, Claude Barras</i>	

## **PHONATION**

<b>Pitch or Phonation: on the Glottalization in Tone Productions in the Ruokeng Hui Chinese Dialect</b> .....	1398
<i>Minghui Zhang, Fang Hu</i>	
<b>Speaker-specific Structure in German Voiceless Stop Voice Onset Times</b> .....	1403
<i>Marc Antony Hullebus, Stephen Tobin, Adamantios Gafos</i>	
<b>Creak in the Respiratory Cycle</b> .....	1408
<i>Kätlin Aare, Pärtel Lippus, Marcin Włodarczyk, Mattias Heldner</i>	
<b>Acoustic Analysis of Whispersy Voice Disguise in Mandarin Chinese</b> .....	1413
<i>Cuiling Zhang, Bin Li, Si Chen, Yike Yang</i>	
<b>The Zurich Corpus of Vowel and Voice Quality, Version 1.0</b> .....	1417
<i>Dieter Maurer, Christian D'Heureuse, Heidy Suter, Volker Dellwo, Daniel Friedrichs, Thayabaran Kathiresan</i>	
<b>Weighting of Coda Voicing Cues: Glottalisation and Vowel Duration</b> .....	1422
<i>Joshua Penney, Felicity Cox, Anita Szakay</i>	

## **COGNITION AND BRAIN STUDIES**

<b>Revealing Spatiotemporal Brain Dynamics of Speech Production Based on EEG and Eye Movement</b> .....	1427
<i>Bin Zhao, Jinfeng Huang, Gaoyan Zhang, Jianwu Dang, Minbo Chen, Yingjianfu, Longbiao Wang</i>	
<b>Neural Response Development During Distributional Learning</b> .....	1432
<i>Natalie Boll-Avetisyan, Jessie S. Nixon, Tomas O. Lentz, Liqun Liu, Sandrien Van Ommen, Çağrı Çöltekin, Jacolien Van Rij</i>	
<b>Learning Two Tone Languages Enhances the Brainstem Encoding of Lexical Tones</b> .....	1437
<i>Akshay Raj Maggu, Wenqing Zong, Vina Law, Patrick C. M. Wong</i>	
<b>Perceptual Sensitivity to Spectral Change in Australian English Close Front Vowels: An Electroencephalographic Investigation</b> .....	1442
<i>Daniel Williams, Paola Escudero, Adamantios Gafos</i>	
<b>Effective Acoustic Cue Learning Is Not Just Statistical, It Is Discriminative</b> .....	1447
<i>Jessie S. Nixon</i>	
<b>Analyzing EEG Signals in Auditory Speech Comprehension Using Temporal Response Functions and Generalized Additive Models</b> .....	1452
<i>Kimberley Mulder, Louis Ten Bosch, Lou Boves</i>	

## **DEEP NEURAL NETWORKS: HOW CAN WE INTERPRET WHAT THEY LEARNED?**

<b>Information Encoding by Deep Neural Networks: What Can We Learn?</b> .....	1457
<i>Louis Ten Bosch, Lou Boves</i>	
<b>Scalable Factorized Hierarchical Variational Autoencoder Training</b> .....	1462
<i>Wei-Ning Hsu, James Glass</i>	
<b>State Gradients for RNN Memory Analysis</b> .....	1467
<i>Lyan Verwimp, Hugo Van Hamme, Vincent Renkens, Patrick Wambacq</i>	

<b>Exploring How Phone Classification Neural Networks Learn Phonetic Information by Visualising and Interpreting Bottleneck Features</b> .....	1472
<i>Linxue Bai, Philip Weber, Peter Jancovic, Martin Russell</i>	
<b>Memory Time Span in LSTMs for Multi-Speaker Source Separation</b> .....	1477
<i>Jeroen Zegers, Hugo Van Hamme</i>	
<b>Visualizing Phoneme Category Adaptation in Deep Neural Networks</b> .....	1482
<i>Odetta Scharenborg, Sebastian Tiesmeyer, Mark Hasegawa-Johnson, Najim Dehak</i>	

#### **SHOW AND TELL 4**

<b>Early Vocabulary Development Through Picture-based Software Solutions</b> .....	1487
<i>G. Kasthuri, Prabha Ramanathan, Hema Murthy, Namita Jacob, Anil Prabhakar</i>	
<b>Automatic Detection of Expressiveness in Oral Reading</b> .....	1489
<i>Kamini Sabu, Kanhaiya Kumar, Preeti Rao</i>	
<b>PannoMulloKathan: Voice Enabled Mobile App for Agricultural Commodity Price Dissemination in Bengali Language</b> .....	1491
<i>Madhab Pal, Rajib Roy, Soma Khan, Milton S. Bepari, Joyanta Basu</i>	
<b>Visualizing Punctuation Restoration in Speech Transcripts with Prosograph</b> .....	1493
<i>Alp Öktem, Mireia Farrús, Antonio Bonafonte</i>	
<b>CACTAS - Collaborative Audio Categorization and Transcription for ASR Systems</b> .....	1495
<i>Mithul Mathivanan, Kinnera Saranu, Abhishek Pandey, Jithendra Vepa</i>	

#### **SPEECH AND SINGING PRODUCTION**

<b>FACTS: A Hierarchical Task-based Control Model of Speech Incorporating Sensory Feedback</b> .....	1497
<i>Benjamin Parrell, Vikram Ramanarayanan, Srikantan Nagarajan, John Houde</i>	
<b>Sensorimotor Response to Tongue Displacement Imagery by Talkers with Parkinson's Disease</b> .....	1502
<i>William Katz, Patrick Reidy, Divya Prabhakaran</i>	
<b>Automatic Pronunciation Evaluation of Singing</b> .....	1507
<i>Chitralkha Gupta, Haizhou Li, Ye Wang</i>	
<b>Classification of Nonverbal Human Produced Audio Events: A Pilot Study</b> .....	1512
<i>Rachel E. Bouserhal, Philippe Chabot, Milton Sarria-Paja, Patrick Cardinal, Jérémie Voix</i>	
<b>UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science</b> .....	1517
<i>Lorenzo Spreafico, Michael Pucher, Anna Matosova</i>	
<b>Articulatory Consequences of Vocal Effort Elicitation Method</b> .....	1521
<i>Elisabet Eir Cortes, Marcin Włodarczak, Juraj Šimko</i>	
<b>Age-related Effects on Sensorimotor Control of Speech Production</b> .....	1526
<i>Anne Hermes, Jane Mertens, Doris Mücke</i>	
<b>An Ultrasound Study of Gemination in Coronal Stops in Eastern Oromo</b> .....	1531
<i>Maida Percival, Alexei Kochetov, Yoonjung Kang</i>	
<b>Processing Transition Regions of Glottal Stop Substituted /S/ for Intelligibility Enhancement of Cleft Palate Speech</b> .....	1536
<i>Protima Nomo Sudro, Sishir Kalita, S R Mahadeva Prasanna</i>	
<b>Reconstructing Neutral Speech from Tracheoesophageal Speech</b> .....	1541
<i>N. Abinay Reddy, M. V. Achuth Rao, G. Nisha Meenakshi, Prasanta Kumar Ghosh</i>	
<b>Automatic Evaluation of Soft Articulatory Contact for Stuttering Treatment</b> .....	1546
<i>Keiko Ochi, Koichi Mori, Naomi Sakai</i>	
<b>Korean Singing Voice Synthesis Based on an LSTM Recurrent Neural Network</b> .....	1551
<i>Juntae Kim, Heejin Choi, Jinuk Park, Minsoo Hahn, Sangjin Kim, Jong-Jin Kim</i>	
<b>The Trajectory of Voice Onset Time with Vocal Aging</b> .....	1556
<i>Xuanda Chen, Ziyu Xiong, Jian Hu</i>	

#### **ROBUST SPEECH RECOGNITION**

<b>The Fifth 'CHiME' Speech Separation and Recognition Challenge: Dataset, Task and Baselines</b> .....	1561
<i>Jon Barker, Shinji Watanabe, Emmanuel Vincent, Jan Trmal</i>	
<b>Voices Obscured in Complex Environmental Settings (VOICES) Corpus</b> .....	1566
<i>Colleen Richey, Maria A. Barrios, Zeb Armstrong, Chris Bartels, Horacio Franco, Martin Graciarana, Aaron Lawson, Mahesh Kumar Nandwana, Allen Stauffer, Julien Van Hout, Paul Gamble, Jeffrey Hetherly, Cory Stephenson, Karl Ni</i>	
<b>Building State-of-the-art Distant Speech Recognition Using the CHiME-4 Challenge with a Setup of Speech Enhancement Baseline</b> .....	1571
<i>Szu-Jui Chen, Aswin Shanmugam Subramanian, Hainan Xu, Shinji Watanabe</i>	
<b>Unsupervised Adaptation with Interpretable Disentangled Representations for Distant Conversational Speech Recognition</b> .....	1576
<i>Wei-Ning Hsu, Hao Tang, James Glass</i>	
<b>Investigating Generative Adversarial Networks Based Speech Dereverberation for Robust Speech Recognition</b> .....	1581
<i>Ke Wang, Junbo Zhang, Sining Sun, Yujun Wang, Fei Xiang, Lei Xie</i>	

<b>Monaural Multi-Talker Speech Recognition with Attention Mechanism and Gated Convolutional Networks</b> .....	1586
<i>Xuankai Chang, Yanmin Qian, Dong Yu</i>	
<b>Weighting Time-Frequency Representation of Speech Using Auditory Saliency for Automatic Speech Recognition</b> .....	1591
<i>Cong-Thanh Do, Yannis Stylianou</i>	
<b>Acoustic Modeling from Frequency Domain Representations of Speech</b> .....	1596
<i>Pegah Ghahremani, Hossein Hadian, Hang Lv, Daniel Povey, Sanjeev Khudanpur</i>	
<b>Non-Uniform Spectral Smoothing for Robust Children's Speech Recognition</b> .....	1601
<i>Ishwar Chandra Yadav, Avinash Kumar, Syed Shah Nawazuddin, Gayadhar Pradhan</i>	
<b>Bidirectional Long-Short Term Memory Network-based Estimation of Reliable Spectral Component Locations</b> .....	1606
<i>Aaron Nicolson, Kuldip K. Paliwal</i>	
<b>Speech Emotion Recognition by Combining Amplitude and Phase Information Using Convolutional Neural Network</b> .....	1611
<i>Lili Guo, Longbiao Wang, Jianwu Dang, Linjuan Zhang, Haotian Guan, Xiangang Li</i>	
<b>Bubble Cooperative Networks for Identifying Important Speech Cues</b> .....	1616
<i>Viet Anh Trinh, Brian McFee, Michael I Mandel</i>	

## **APPLICATIONS IN EDUCATION AND LEARNING**

<b>Real-Time Scoring of an Oral Reading Assessment on Mobile Devices</b> .....	1621
<i>Jian Cheng</i>	
<b>A Deep Learning Approach to Assessing Non-native Pronunciation of English Using Phone Distances</b> .....	1626
<i>Konstantinos Kyriakopoulos, Kate Knill, Mark Gales</i>	
<b>Paired Phone-Posteriors Approach to ESL Pronunciation Quality Assessment</b> .....	1631
<i>Yujia Xiao, Frank Soong, Wenping Hu</i>	
<b>Investigating the Role of L1 in Automatic Pronunciation Evaluation of L2 Speech</b> .....	1636
<i>Ming Tu, Anna Grabek, Julie Liss, Visar Berisha</i>	
<b>Impact of ASR Performance on Free Speaking Language Assessment</b> .....	1641
<i>Kate Knill, Mark Gales, Konstantinos Kyriakopoulos, Andrey Malinin, Anton Ragni, Yu Wang, Andrew Caines</i>	
<b>Automatic Miscue Detection Using RNN Based Models with Data Augmentation</b> .....	1646
<i>Yoon Seok Hong, Kyung Seo Ki, Gahgene Gweon</i>	
<b>A Study of Objective Measurement of Comprehensibility through Native Speakers' Shadowing of Learners' Utterances</b> .....	1651
<i>Yusuke Inoue, Suguru Kabashima, Daisuke Saito, Nobuaki Minematsu, Kumi Kanamura, Yutaka Yamauchi</i>	
<b>Factorized Deep Neural Network Adaptation for Automatic Scoring of L2 Speech in English Speaking Tests</b> .....	1656
<i>Dean Luo, Chunxiao Zhang, Linzhong Xia, Lixin Wang</i>	
<b>On the Difficulties of Automatic Speech Recognition for Kindergarten-Aged Children</b> .....	1661
<i>Gary Yeung, Abeer Alwan</i>	
<b>Improved Acoustic Modelling for Automatic Literacy Assessment of Children</b> .....	1666
<i>Mauro Nicolao, Michiel Sanders, Thomas Hain</i>	

## **INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS**

<b>Anomaly Detection Approach for Pronunciation Verification of Disordered Speech Using Speech Attribute Features</b> .....	1671
<i>Mostafa Shahin, Beena Ahmed, Jim X. Ji, Kirrie Ballard</i>	
<b>Effectiveness of Voice Quality Features in Detecting Depression</b> .....	1676
<i>Amber Afshan, Jinxi Guo, Soo Jin Park, Vijay Ravi, Jonathan Flint, Abeer Alwan</i>	
<b>Fusing Text-dependent Word-level i-Vector Models to Screen 'at Risk' Child Speech</b> .....	1681
<i>Prasanna Kothalkar, Johanna Rudolph, Christine Dollaghan, Jennifer McGlothlin, Thomas Campbell, John H. L. Hansen</i>	
<b>Testing Paradigms for Assistive Hearing Devices in Diverse Acoustic Environments</b> .....	1686
<i>Hussnain Ali, John H. L. Hansen, M. C. Ram Charan</i>	
<b>Detection of Dementia from Responses to Atypical Questions Asked by Embodied Conversational Agents</b> .....	1691
<i>Tsuyuki Ujiro, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo, Satoshi Nakamura</i>	
<b>Acoustic Features Associated with Sustained Vowel and Continuous Speech Productions by Chinese Children with Functional Articulation Disorders</b> .....	1696
<i>Wang Zhang, Xiangquan Gui, Tianqi Wang, Manwa Ng, Feng Yang, Lan Wang, Nan Yan</i>	
<b>Estimation of Hypernasality Scores from Cleft Lip and Palate Speech</b> .....	1701
<i>C. M. Vikram, Ayush Tripathi, Sishir Kalita, S R Mahadeva Prasanna</i>	
<b>Detecting Alzheimer's Disease Using Gated Convolutional Neural Network from Audio Data</b> .....	1706
<i>Tifani Warnita, Nakamasa Inoue, Koichi Shinoda</i>	
<b>Automatic Detection of Orofacial Impairment in Stroke</b> .....	1711
<i>Andrea Bandini, Jordan Green, Brian Richburg, Yana Yunusova</i>	
<b>Detecting Depression with Audio/Text Sequence Modeling of Interviews</b> .....	1716
<i>Tuka Al Hanai, Mohammad Ghassemi, James Glass</i>	



## **SPEAKER CHARACTERIZATION AND ANALYSIS**

<b>Discourse Marker Detection for Hesitation Events on Mandarin Conversation</b> .....	1721
<i>Yu-Win Wang, Hen-Hsen Huang, Kuan-Yu Chen, Hsin-Hsi Chen</i>	
<b>Acoustic and Perceptual Characteristics of Mandarin Speech in Homosexual and Heterosexual Male Speakers</b> .....	1726
<i>Puyang Geng, Wentao Gu, Hiroya Fujisaki</i>	
<b>Automatic Question Detection from Acoustic and Phonetic Features Using Feature-wise Pre-training</b> .....	1731
<i>Atsushi Ando, Reine Asakawa, Ryo Masumura, Hosana Kamiyama, Satoshi Kobashikawa, Yushi Aono</i>	
<b>Improving Response Time of Active Speaker Detection Using Visual Prosody Information Prior to Articulation</b> .....	1736
<i>Fasih Haider, Saturnino Luz, Carl Vogel, Nick Campbell</i>	
<b>Audio-Visual Prediction of Head-Nod and Turn-Taking Events in Dyadic Interactions</b> .....	1741
<i>Bekir Berker Türker, Engin Erzin, Yücel Yemez, Metin Sezgin</i>	
<b>Analyzing Effect of Physical Expression on English Proficiency for Multimodal Computer-Assisted Language Learning</b> .....	1746
<i>Haoran Wu, Yuya Chiba, Takashi Nose, Akinori Ito</i>	
<b>Analysis of the Effect of Speech-Laugh on Speaker Recognition System</b> .....	1751
<i>Sri Harsha Dumpala, Ashish Panda, Sunil Kumar Koppurapu</i>	
<b>Vocal Biomarkers for Cognitive Performance Estimation in a Working Memory Task</b> .....	1756
<i>Jennifer Sloboda, Adam Lammert, James Williamson, Christopher Smalt, Daryush D. Mehta, Col Ian Curry, Kristin Heaton, Jeffrey Palmer, Thomas Quatieri</i>	
<b>Lexical and Acoustic Deep Learning Model for Personality Recognition</b> .....	1761
<i>Guozhen An, Rivka Levitan</i>	

## **PERSPECTIVE TALK-2**

<b>Open Problems in Speech Recognition</b> .....	1766
<i>Bhuvana Ramabhadran</i>	

## **PLENARY TALK-2**

<b>Evolution of Neural Network Architectures for Speech Recognition</b> .....	1767
<i>Hervé Bourlard</i>	

## **NOVEL NEURAL NETWORK ARCHITECTURES FOR ACOUSTIC MODELLING**

<b>Layer Trajectory LSTM</b> .....	1768
<i>Jinyu Li, Changliang Liu, Yifan Gong</i>	
<b>Semi-tied Units for Efficient Gating in LSTM and Highway Networks</b> .....	1773
<i>Chao Zhang, Philip Woodland</i>	
<b>Gaussian Process Neural Networks for Speech Recognition</b> .....	1778
<i>Max W. Y. Lam, Shoukang Hu, Xurong Xie, Shansong Liu, Jianwei Yu, Rongfeng Su, Xunying Liu, Helen Meng</i>	
<b>Acoustic Modeling with Densely Connected Residual Network for Multichannel Speech Recognition</b> .....	1783
<i>Jian Tang, Yan Song, Lirong Dai, Ian McLoughlin</i>	
<b>Gated Recurrent Unit Based Acoustic Modeling with Future Context</b> .....	1788
<i>Jie Li, Xiaorui Wang, Yuanyuan Zhao, Yan Li</i>	
<b>Output-Gate Projected Gated Recurrent Unit for Speech Recognition</b> .....	1793
<i>Gaofeng Cheng, Daniel Povey, Lu Huang, Ji Xu, Sanjeev Khudanpur, Yonghong Yan</i>	

## **LANGUAGE IDENTIFICATION**

<b>Performance Analysis of the 2017 NIST Language Recognition Evaluation</b> .....	1798
<i>Seyed Omid Sadjadi, Timothee Kheyrkhan, Craig Greenberg, Elliot Singer, Douglas Reynolds, Lisa Mason, Jaime Hernandez-Cordero</i>	
<b>Using Deep Neural Networks for Identification of Slavic Languages from Acoustic Signal</b> .....	1803
<i>Lukas Mateju, Petr Cerva, Jindrich Zdansky, Radek Safarik</i>	
<b>Adding New Classes without Access to the Original Training Data with Applications to Language Identification</b> .....	1808
<i>Hagai Taitelbaum, Ehud Ben-Reuven, Jacob Goldberger</i>	
<b>Feature Representation of Short Utterances Based on Knowledge Distillation for Spoken Language Identification</b> .....	1813
<i>Peng Shen, Xugang Lu, Sheng Li, Hisashi Kawai</i>	
<b>Sub-band Envelope Features Using Frequency Domain Linear Prediction for Short Duration Language Identification</b> .....	1818
<i>Sarith Fernando, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i>	
<b>Effectiveness of Single-Channel BLSTM Enhancement for Language Identification</b> .....	1823
<i>Peter Sibbern Frederiksen, Jesús Villalba, Shinji Watanabe, Zheng-Hua Tan, Najim Dehak</i>	

## **PRODUCTION OF PROSODY**

<b>Articulation Rate as a Speaker Discriminant in British English</b> .....	1828
<i>Erica Gold</i>	
<b>Truncation and Compression in Southern German and Australian English</b> .....	1833
<i>Jenny Yu, Katharina Zahner</i>	
<b>Prominence-based Evaluation of L2 Prosody</b> .....	1838
<i>Heini Kallio, Antti Suni, Päivi Virkkunen, Juraj Šimko</i>	
<b>Length Contrast and Covarying Features: Whistled Speech as a Case Study</b> .....	1843
<i>Rachid Ridouane, Giuseppina Turco, Julien Meyer</i>	
<b>Information Structure, Affect and Prenuclear Prominence in American English</b> .....	1848
<i>Eleanor Chodroff, Jennifer Cole</i>	
<b>Effects of User Controlled Speech Rate on Intelligibility in Noisy Environments</b> .....	1853
<i>John S. Novak III, Robert V. Kenyon</i>	

## **SPEECH INTELLIGIBILITY AND QUALITY**

<b>Binaural Speech Intelligibility Estimation Using Deep Neural Networks</b> .....	1858
<i>Kazuhiro Kondo, Kazuya Taira, Yosuke Kobayashi</i>	
<b>Multi-resolution Gammachirp Envelope Distortion Index for Intelligibility Prediction of Noisy Speech</b> .....	1863
<i>Katsuhiko Yamamoto, Toshio Irino, Narumi Ohashi, Shoko Araki, Keisuke Kinoshita, Tomohiro Nakatani</i>	
<b>Speech Intelligibility Enhancement Based on a Non-causal Wavenet-like Model</b> .....	1868
<i>P. V. Mohammed Shifas, Vassilis Tsiraras, Yannis Stylianou</i>	
<b>Quality-Net: An End-to-End Non-intrusive Speech Quality Assessment Model Based on BLSTM</b> .....	1873
<i>Szu-Wei Fu, Yu Tsao, Hsin-Te Hwang, Hsin-Min Wang</i>	
<b>Global SNR Estimation of Speech Signals Using Entropy and Uncertainty Estimates from Dropout Networks</b> .....	1878
<i>Rohith Aralikatti, Dilip Kumar Margam, Tanay Sharma, Abhinav Thanda, Shankar Venkatesan</i>	
<b>Detecting Packet-Loss Concealment Using Formant Features and Decision Tree Learning</b> .....	1883
<i>Gabriel Mittag, Sebastian Möller</i>	

## **INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS**

<b>UltraSuite: A Repository of Ultrasound and Acoustic Data from Child Speech Therapy Sessions</b> .....	1888
<i>Aciel Eshky, Manuel Sam Ribeiro, Joanne Cleland, Korin Richmond, Zoe Roxburgh, James M Scobbie, Alan Wrench</i>	
<b>Detecting Signs of Dementia Using Word Vector Representations</b> .....	1893
<i>Bahman Mirheidari, Daniel Blackburn, Traci Walker, Annalena Veneri, Markus Reuber, Heidi Christensen</i>	
<b>Classification of Huntington Disease Using Acoustic and Lexical Features</b> .....	1898
<i>Mathew Perez, Wenyu Jin, Duc Le, Noelle Carlozzi, Praveen Dayalu, Angela Roberts, Emily Mower Provost</i>	

## **VOLUME 4**

<b>The PRIORI Emotion Dataset: Linking Mood to Emotion Detected In-the-Wild</b> .....	1903
<i>Soheil Khorram, Mimansa Jaiswal, John Gideon, Melvin McInnis, Emily Mower Provost</i>	
<b>Language Features for Automated Evaluation of Cognitive Behavior Psychotherapy Sessions</b> .....	1908
<i>Nikolaos Flenotomos, Victor Martinez, James Gibson, David Atkins, Torrey Creed, Shrikanth Narayanan</i>	
<b>Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks</b> .....	1913
<i>Kwanghoon An, Myungjong Kim, Kristin Teplansky, Jordan Green, Thomas Campbell, Yana Yunusova, Daragh Heitzman, Jun Wang</i>	

## **SPEECH TECHNOLOGIES FOR CODE-SWITCHING IN MULTILINGUAL COMMUNITIES**

<b>A Study of Lexical and Prosodic Cues to Segmentation in a Hindi-English Code-switched Discourse</b> .....	1918
<i>Preeti Rao, Mugdha Pandya, Kamini Sabu, Kanhaiya Kumar, Nandini Bondale</i>	
<b>Building a Unified Code-Switching ASR System for South African Languages</b> .....	1923
<i>Emre Yilmaz, Astik Biswas, Ewald Van Der Westhuizen, Febe De Wet, Thomas Niesler</i>	
<b>Study of Semi-supervised Approaches to Improving English-Mandarin Code-Switching Speech Recognition</b> .....	1928
<i>Pengcheng Guo, Haihua Xu, Lei Xie, Eng Siong Chng</i>	
<b>Acoustic and Textual Data Augmentation for Improved ASR of Code-Switching Speech</b> .....	1933
<i>Emre Yilmaz, Henk Van Den Heuvel, David Van Leeuwen</i>	
<b>The Role of Cognate Words, POS Tags and Entrainment in Code-Switching</b> .....	1938
<i>Victor Soto, Nishmar Cestero, Julia Hirschberg</i>	
<b>Homophone Identification and Merging for Code-switched Speech Recognition</b> .....	1943
<i>Brij Mohan Lal Srivastava, Sunayana Sitaram</i>	
<b>Code-switching in Indic Speech Synthesizers</b> .....	1948
<i>Anju Leela Thomas, Anusha Prakash, Arun Baby, Hema Murthy</i>	

<b>A Novel Approach for Effective Recognition of the Code-Switched Data on Monolingual Language Model</b> .....	1953
<i>Sreeram Ganji, Rohit Sinha</i>	

## **SHOW AND TELL 5**

<b>Hierarchical Accent Determination and Application in a Large Scale ASR System</b> .....	1958
<i>Ramya Viswanathan, Periyasamy Paramasivam, Jithendra Vepa</i>	
<b>Toward Scalable Dialog Technology for Conversational Language Learning: Case Study of the TOEFL® MOOC</b> .....	1960
<i>Vikram Ramanarayanan, David Pautler, Patrick Lange, Eugene Tsuprun, Rutuja Ubale, Keelan Evanini, David Suendermann-Oeft</i>	
<b>Machine Learning Powered Data Platform for High-Quality Speech and NLP Workflows</b> .....	1962
<i>João Freitas, Jorge Ribeiro, Daan Balduwijn, Sara Oliveira, Daniela Braga</i>	
<b>Fully Automatic Speaker Separation System, with Automatic Enrolling of Recurrent Speakers</b> .....	1964
<i>Raphael Cohen, Orgad Keller, Jason Levy, Russell Levy, Micha Breakstone, Amit Ashkenazi</i>	
<b>Online Speech Translation System for Tamil</b> .....	1966
<i>Ayyavu Madhavaraj, H. R. Shiva Kumar, A. G. Ramakrishnan</i>	

## **VOICE CONVERSION AND SPEECH SYNTHESIS**

<b>Unsupervised Vocal Tract Length Warped Posterior Features for Non-Parallel Voice Conversion</b> .....	1968
<i>Nirmesh Shah, Maulik C. Madhavi, Hemant Patil</i>	
<b>Voice Conversion with Conditional SampleRNN</b> .....	1973
<i>Cong Zhou, Michael Horgan, Vivek Kumar, Cristina Vasco, Dan Darcy</i>	
<b>A Voice Conversion Framework with Tandem Feature Sparse Representation and Speaker-Adapted WaveNet Vocoder</b> .....	1978
<i>Berrak Sisman, Mingyang Zhang, Haizhou Li</i>	
<b>WaveNet Vocoder with Limited Training Data for Voice Conversion</b> .....	1983
<i>Li-Juan Liu, Zhen-Hua Ling, Yuan Jiang, Ming Zhou, Li-Rong Dai</i>	
<b>Collapsed Speech Segment Detection and Suppression for WaveNet Vocoder</b> .....	1988
<i>Yi-Chiao Wu, Kazuhiro Kobayashi, Tomoki Hayashi, Patrick Lumban Tobing, Tomoki Toda</i>	
<b>High-quality Voice Conversion Using Spectrogram-Based WaveNet Vocoder</b> .....	1993
<i>Kuan Chen, Bo Chen, Jiahao Lai, Kai Yu</i>	
<b>Spanish Statistical Parametric Speech Synthesis Using a Neural Vocoder</b> .....	1998
<i>Antonio Bonafonte, Santiago Pascual, Georgina Dorca</i>	
<b>Experiments with Training Corpora for Statistical Text-to-speech Systems.</b> .....	2002
<i>Monika Podsiadlo, Victor Ungureanu</i>	
<b>Multi-task WaveNet: A Multi-task Generative Model for Statistical Parametric Speech Synthesis without Fundamental Frequency Conditions</b> .....	2007
<i>Yu Gu, Yongguo Kang</i>	
<b>Speaker-independent Raw Waveform Model for Glottal Excitation</b> .....	2012
<i>Lauri Juvela, Vassilis Tsiaras, Bajjaba Bollepalli, Manu Airaksinen, Junichi Yamagishi, Paavo Alku</i>	
<b>A New Glottal Neural Vocoder for Speech Synthesis</b> .....	2017
<i>Yang Cui, Xi Wang, Lei He, Frank K. Soong</i>	
<b>Exemplar-based Speech Waveform Generation</b> .....	2022
<i>Oliver Watts, Cassia Valentini-Botinhao, Felipe Espic, Simon King</i>	
<b>Frequency Domain Variants of Velvet Noise and Their Application to Speech Processing and Synthesis</b> .....	2027
<i>Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, Tomoki Toda, Toshio Irino</i>	

## **EXTRACTING INFORMATION FROM AUDIO**

<b>Joint Learning of Interactive Spoken Content Retrieval and Trainable User Simulator</b> .....	2032
<i>Pei-Hung Chung, Kuan Tung, Ching-Lun Tai, Hung-Yi Lee</i>	
<b>Attention-based End-to-End Models for Small-Footprint Keyword Spotting</b> .....	2037
<i>Changhao Shan, Junbo Zhang, Yujun Wang, Lei Xie</i>	
<b>Prediction of Aesthetic Elements in Karnatic Music: A Machine Learning Approach</b> .....	2042
<i>M. Ragesh Rajan, Ashwin Vijayakumar, Deepu Vijayasanen</i>	
<b>Topic and Keyword Identification for Low-resourced Speech Using Cross-Language Transfer Learning</b> .....	2047
<i>Wenda Chen, Mark Hasegawa-Johnson, Nancy F. Chen</i>	
<b>Automatic Speech Recognition and Topic Identification from Speech for Almost-Zero-Resource Languages</b> .....	2052
<i>Matthew Wiesner, Chunxi Liu, Lucas Ondel, Craig Harman, Vimal Manohar, Jan Trmal, Zhongqiang Huang, Najim Dehak, Sanjeev Khudanpur</i>	
<b>Play Duration Based User-Entity Affinity Modeling in Spoken Dialog System</b> .....	2057
<i>Bo Xiao, Nicholas Monath, Shankar Ananthakrishnan, Abishek Ravi</i>	
<b>Empirical Analysis of Score Fusion Application to Combined Neural Networks for Open Vocabulary Spoken Term Detection</b> .....	2062
<i>Shi-Wook Lee, Kazuyo Tanaka, Yoshiaki Itoh</i>	
<b>Phonological Posterior Hashing for Query by Example Spoken Term Detection</b> .....	2067
<i>Afsaneh Asaei, Dhananjay Ram, Hervé Bourlard</i>	

<b>Term Extraction via Neural Sequence Labeling a Comparative Evaluation of Strategies Using Recurrent Neural Networks</b> .....	2072
<i>Maren Kucza, Jan Niehues, Thomas Zenkel, Alex Waibel, Sebastian Stüker</i>	
<b>Semi-supervised Learning for Information Extraction from Dialogue</b> .....	2077
<i>Anjuli Kannan, Kai Chen, Diana Jaunzeikare, Alvin Rajkomar</i>	
<b>Slot Filling with Delexicalized Sentence Generation</b> .....	2082
<i>Youhyun Shin, Kang Min Yoo, Sang-Goo Lee</i>	
<b>Music Genre Recognition Using Deep Neural Networks and Transfer Learning</b> .....	2087
<i>Deepanway Ghosal, Maheshkumar H. Kolekar</i>	
<b>Efficient Voice Trigger Detection for Low Resource Hardware</b> .....	2092
<i>Siddharth Sigtia, Rob Haynes, Hywel Richards, Erik Marchi, John Bridle</i>	

## **SIGNAL ANALYSIS FOR THE NATURAL, BIOLOGICAL AND SOCIAL SCIENCES**

<b>A Novel Normalization Method for Autocorrelation Function for Pitch Detection and for Speech Activity Detection</b> .....	2097
<i>Qiguang Lin, Yiwen Shao</i>	
<b>Estimation of the Vocal Tract Length of Vowel Sounds Based on the Frequency of the Significant Spectral Valley</b> .....	2102
<i>T. V. Ananthapadmanabha, A. G. Ramakrishnan</i>	
<b>Deep Learning Techniques for Koala Activity Detection</b> .....	2107
<i>Ivan Himawan, Michael Towsey, Bradley Law, Paul Roe</i>	
<b>Glottal Closure Instant Detection from Speech Signal Using Voting Classifier and Recursive Feature Elimination</b> .....	2112
<i>Jindrich Matoušek, Daniel Tihelka</i>	
<b>Assessing Speaker Engagement in 2-Person Debates: Overlap Detection in United States Presidential Debates</b> .....	2117
<i>Midia Yousefi, Navid Shokouhi, John H. L. Hansen</i>	
<b>All-Conv Net for Bird Activity Detection: Significance of Learned Pooling</b> .....	2122
<i>Arjun Pankajakshan, Anshul Thakur, Daksh Thapar, Padmanabhan Rajan, Aditya Nigam</i>	
<b>Deep Convex Representations: Feature Representations for Bioacoustics Classification</b> .....	2127
<i>Anshul Thakur, Vinayak Abrol, Pulkit Sharma, Padmanabhan Rajan</i>	
<b>Detection of Glottal Excitation Epochs in Speech Signal Using Hilbert Envelope</b> .....	2132
<i>Hirak Dasgupta, Prem C. Pandey, K S Nataraj</i>	
<b>Analyzing Thai Tone Distribution through Functional Data Analysis</b> .....	2137
<i>Hong Zhang</i>	
<b>Articulatory Feature Classification Using Convolutional Neural Networks</b> .....	2142
<i>Danny Merckx, Odette Scharenborg</i>	
<b>A New Frequency Coverage Metric and a New Subband Encoding Model, with an Application in Pitch Estimation</b> .....	2147
<i>Shoufeng Lin</i>	
<b>Improved Epoch Extraction from Telephonic Speech Using Chebfun and Zero Frequency Filtering</b> .....	2152
<i>B. Ganga Gowri, K. P. Soman, D. Govind</i>	

## **SPEECH PROSODY**

<b>An Empirical Analysis of the Correlation of Syntax and Prosody</b> .....	2157
<i>Arne Köhn, Timo Baumann, Oskar Dörfler</i>	
<b>Analysing the Focus of a Hierarchical Attention Network: the Importance of Enjambments When Classifying Post-modern Poetry</b> .....	2162
<i>Timo Baumann, Hussein Hussein, Burkhard Meyer-Sickendiek</i>	
<b>Language-Dependent Melody Embeddings</b> .....	2167
<i>Daniil Kocharov, Alla Menshikova</i>	
<b>Stress Distribution of Given Information in Chinese Reading Texts</b> .....	2171
<i>Yuan Jia, Xiaoxiao Ma</i>	
<b>Acoustic-prosodic Entrainment in Structural Metadata Events</b> .....	2176
<i>Vera Cabarrão, Fernando Batista, Helena Moniz, Isabel Trancoso, Ana Isabel Mata</i>	
<b>Formant Measures of Vowels Adjacent to Alveolar and Retroflex Consonants in Arrernte: Stressed and Unstressed Position</b> .....	2181
<i>Marija Tabain, Richard Beare, Andrew Butcher</i>	
<b>Automatic Assessment of L2 English Word Prosody Using Weighted Distances of F0 and Intensity Contours</b> .....	2186
<i>Quy-Thao Truong, Tsuneo Kato, Seiichi Yamamoto</i>	
<b>Homogeneity vs Heterogeneity in Indian English: Investigating Influences of L1 on f0 Range</b> .....	2191
<i>Olga Maxwell, Elinor Payne, Rosey Billington</i>	
<b>Emotional Prosody Perception in Mandarin-speaking Congenital Amusics</b> .....	2196
<i>Yixin Zhang, Tianzhu Geng, Jinsong Zhang</i>	
<b>Cultural Differences in Pattern Matching: Multisensory Recognition of Socio-affective Prosody</b> .....	2201
<i>Takaaki Shochi, Jean-Luc Rouas, Marine Guerry, Donna Erickson</i>	

### **PERSPECTIVE TALK-3**

<b>Speech Processing in the Human Brain Meets Deep Learning</b> .....	2206
<i>Nima Mesgarani</i>	

### **RECURRENT NEURAL MODELS FOR ASR**

<b>ESPnet: End-to-End Speech Processing Toolkit</b> .....	2207
<i>Shinji Watanabe, Takaaki Hori, Shigeki Karita, Tomoki Hayashi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalta Soplín, Jahn Heymann, Matthew Wiesner, Nanxin Chen, Adithya Renduchintala, Tsubasa Ochiai</i>	
<b>A GPU-based WFST Decoder with Exact Lattice Generation</b> .....	2212
<i>Zhehuai Chen, Justin Luijens, Hainan Xu, Yiming Wang, Daniel Povey, Sanjeev Khudanpur</i>	
<b>Automatic Speech Recognition System Development in the "Wild"</b> .....	2217
<i>Anton Ragni, Mark Gales</i>	
<b>Semantic Lattice Processing in Contextual Automatic Speech Recognition for Google Assistant</b> .....	2222
<i>Leonid Velikovich, Ian Williams, Justin Scheiner, Petar Aleksic, Pedro Moreno, Michael Riley</i>	
<b>Contextual Speech Recognition in End-to-end Neural Network Systems Using Beam Search</b> .....	2227
<i>Ian Williams, Anjali Kannan, Petar Aleksic, David Rybach, Tara Sainath</i>	
<b>Forward-Backward Attention Decoder</b> .....	2232
<i>Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara</i>	

### **SPEAKER VERIFICATION USING NEURAL NETWORK METHODS I**

<b>Learning Discriminative Features for Speaker Identification and Verification</b> .....	2237
<i>Sarthak Yadav, Atul Rai</i>	
<b>Triplet Loss Based Cosine Similarity Metric Learning for Text-independent Speaker Recognition</b> .....	2242
<i>Sergey Novoselov, Vadim Shchemelinin, Andrey Shulipa, Alexandr Kozlov, Ivan Krennev</i>	
<b>Speaker Embedding Extraction with Phonetic Information</b> .....	2247
<i>Yi Liu, Liang He, Jia Liu, Michael T. Johnson</i>	
<b>Attentive Statistics Pooling for Deep Speaker Embedding</b> .....	2252
<i>Koji Okabe, Takafumi Koshinaka, Koichi Shinoda</i>	
<b>Robust and Discriminative Speaker Embedding via Intra-Class Distance Variance Regularization</b> .....	2257
<i>Nam Le, Jean-Marc Odobez</i>	
<b>Deep Discriminative Embeddings for Duration Robust Speaker Verification</b> .....	2262
<i>Na Li, Deyi Tuo, Dan Su, Zhifeng Li, Dong Yu</i>	

### **SPEECH PERCEPTION IN ADVERSE CONDITIONS**

<b>Impact of Different Speech Types on Listening Effort</b> .....	2267
<i>Olympia Simantiraki, Martin Cooke, Simon King</i>	
<b>Who Are You Listening to? Towards a Dynamic Measure of Auditory Attention to Speech-on-speech</b> .....	2272
<i>Moira-Phoebé Huet, Christophe Micheyl, Etienne Gaudrain, Etienne Parizet</i>	
<b>Investigating the Role of Familiar Face and Voice Cues in Speech Processing in Noise</b> .....	2276
<i>Jeesun Kim, Sonya Karisma, Vincent Aubanel, Chris Davis</i>	
<b>The Conversation Continues: the Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition</b> .....	2280
<i>Odette Scharenborg, Martha Larson</i>	
<b>Loud and Shouted Speech Perception at Variable Distances in a Forest</b> .....	2285
<i>Julien Meyer, Fanny Meunier, Laure Dentel, Noelia Do Carmo Blanco, Frédéric Sèbe</i>	
<b>Phoneme Resistance and Phoneme Confusion in Noise: Impact of Dyslexia</b> .....	2290
<i>Noelia Do Carmo Blanco, Julien Meyer, Michel Hoen, Fanny Meunier</i>	

### **MEASURING PITCH AND ARTICULATION**

<b>Conditional End-to-End Audio Transforms</b> .....	2295
<i>Albert Haque, Michelle Guo, Prateek Verma</i>	
<b>Detection of Glottal Closure Instants in Degraded Speech Using Single Frequency Filtering Analysis</b> .....	2300
<i>Gunnam Aneja, Sudarsana Reddy Kadiri, Bayya Yegnanarayana</i>	
<b>Tone Recognition Using Lifters and CTC</b> .....	2305
<i>Loren Lugosch, Vikrant Singh Tomar</i>	
<b>Epoch Extraction from Pathological Children Speech Using Single Pole Filtering Approach</b> .....	2310
<i>C. M. Vikram, S R Mahadeva Prasanna</i>	
<b>Automated Classification of Vowel-Gesture Parameters Using External Broadband Excitation</b> .....	2315
<i>B. T. Balamurali, Jer-Ming Chen</i>	
<b>Estimation of Fundamental Frequency from Singing Voice Using Harmonics of Impulse-like Excitation Source</b> .....	2319
<i>Sudarsana Reddy Kadiri, Bayya Yegnanarayana</i>	

## SPEECH AND LANGUAGE ANALYTICS FOR MENTAL HEALTH

<b>Investigating the Effect of Audio Duration on Dementia Detection Using Acoustic Features</b> .....	2324
<i>Jochen Weiner, Miguel Angrick, Srinivasan Umesh, Tanja Schultz</i>	
<b>An Interlocutor-Modulated Attentional LSTM for Differentiating between Subgroups of Autism Spectrum Disorder</b> .....	2329
<i>Yun-Shao Lin, Susan Shur-Fen Gau, Chi-Chun Lee</i>	
<b>Recognition of Echolalic Autistic Child Vocalisations Utilising Convolutional Recurrent Neural Networks</b> .....	2334
<i>Shahin Amiriparian, Alice Baird, Sahib Julka, Alyssa Alcorn, Sandra Orl, Suncica Petrovic, Eloise Ainger, Nicholas Cummins, Björn Schuller</i>	
<b>Modeling Interpersonal Influence of Verbal Behavior in Couples Therapy Dyadic Interactions</b> .....	2339
<i>Sandeep Nallan Chakravarthula, Brian Baucom, Panayiotis Georgiou</i>	
<b>Computational Modeling of Conversational Humor in Psychotherapy</b> .....	2344
<i>Anil Ramakrishna, Timothy Greer, David Atkins, Shrikanth Narayanan</i>	
<b>Multimodal I-vectors to Detect and Evaluate Parkinson's Disease</b> .....	2349
<i>Nicanor Garcia, Juan Camilo Vásquez Correa, Juan Rafael Orozco-Arroyave, Elmar Nöth</i>	

## SPOKEN CALL SHARED TASK, SECOND EDITION

<b>Overview of the 2018 Spoken CALL Shared Task</b> .....	2354
<i>Claudia Baur, Andrew Caines, Cathy Chua, Johanna Gerlach, Mengjie Qian, Manny Rayner, Martin Russell, Helmer Strik, Xizi Wei</i>	
<b>The CSU-K Rule-Based System for the 2nd Edition Spoken CALL Shared Task</b> .....	2359
<i>Dominik Jülg, Mario Kunstek, Cem Philipp Freimoser, Kay Berkling, Mengjie Qian</i>	
<b>Liulishuo's System for the Spoken CALL Shared Task 2018</b> .....	2364
<i>Huy Nguyen, Lei Chen, Ramon Prieto, Chuan Wang, Yang Liu</i>	
<b>An Optimization Based Approach for Solving Spoken CALL Shared Task</b> .....	2369
<i>Mohammad Ateeq, Abualsoud Hanani, Aziz Qaroush</i>	
<b>The University of Birmingham 2018 Spoken CALL Shared Task Systems</b> .....	2374
<i>Mengjie Qian, Xizi Wei, Peter Jancovic, Martin Russell</i>	
<b>Improvements to an Automated Content Scoring System for Spoken CALL Responses: the ETS Submission to the Second Spoken CALL Shared Task</b> .....	2379
<i>Keelan Evanini, Matthew Mulholland, Rutuja Ubale, Yao Qian, Robert Pugh, Vikram Ramanarayanan, Aoife Cahill</i>	

## SHOW AND TELL 6

<b>Extracting Speaker's Gender, Accent, Age and Emotional State from Speech</b> .....	2384
<i>Nagendra Goel, Mousmita Sarma, Tejendra Kushwah, Dharmesh Agarwal, Zikra Iqbal, Surbhi Chauhan</i>	
<b>Determining Speaker Location from Speech in a Practical Environment</b> .....	2386
<i>Bhys Narayanamurthy, J. V. Satyanarayana, Bayya Yegnanarayana</i>	
<b>An Automatic Speech Transcription System for Manipuri Language</b> .....	2388
<i>Tanvina Patel, D. N. Krishna, Noor Fathima, Nisar Shah, C. Mahima, Deepak Kumar, Anuroop Iyengar</i>	
<b>SPIRE-SST: An Automatic Web-based Self-learning Tool for Syllable Stress Tutoring (SST) to the Second Language Learners</b> .....	2390
<i>Chiranjeevi Yarra, P. A. Anand, N. K. Kausthubha, Prasanta Kumar Ghosh</i>	
<b>Glottal Vibrato Graph: A Device and Method for Recording, Analysis and Visualization of Glottal Activity</b> .....	2392
<i>Kishalay Chakraborty, Senjam Shantirani Devi, Sanjeevan Devnath, S R Mahadeva Prasanna, Priyankoo Sarmah</i>	

## ADJUSTING TO SPEAKER, ACCENT, AND DOMAIN

<b>Multi-Modal Data Augmentation for End-to-end ASR</b> .....	2394
<i>Adithya Renduchintala, Shuoyang Ding, Matthew Wiesner, Shinji Watanabe</i>	
<b>Multi-task Learning with Augmentation Strategy for Acoustic-to-word Attention-based Encoder-decoder Speech Recognition</b> .....	2399
<i>Takafumi Moriya, Sei Ueno, Yusuke Shinohara, Marc Delcroix, Yoshikazu Yamaguchi, Yushi Aono</i>	
<b>Training Augmentation with Adversarial Examples for Robust Speech Recognition</b> .....	2404
<i>Sining Sun, Ching-Feng Yeh, Mari Ostendorf, Mei-Yuh Hwang, Lei Xie</i>	
<b>Data Augmentation Improves Recognition of Foreign Accented Speech</b> .....	2409
<i>Takashi Fukuda, Raul Fernandez, Andrew Rosenberg, Samuel Thomas, Bhuvana Ramabhadran, Alexander Sorin, Gakuto Kurata</i>	
<b>Speaker Adaptive Training and Mixup Regularization for Neural Network Acoustic Models in Automatic Speech Recognition</b> .....	2414
<i>Natalia Tomashenko, Yuri Khokhlov, Yannick Estève</i>	
<b>Neural Language Codes for Multilingual Acoustic Models</b> .....	2419
<i>Markus Müller, Sebastian Stüker, Alex Waibel</i>	
<b>Encoder Transfer for Attention-based Acoustic-to-word Speech Recognition</b> .....	2424
<i>Sei Ueno, Takafumi Moriya, Masato Mimura, Shinsuke Sakai, Yusuke Shinohara, Yoshikazu Yamaguchi, Yushi Aono, Tatsuya Kawahara</i>	

<b>Empirical Evaluation of Speaker Adaptation on DNN Based Acoustic Model</b> .....	2429
<i>Ke Wang, Junbo Zhang, Yujun Wang, Lei Xie</i>	
<b>Improving DNNs Trained with Non-Native Transcriptions Using Knowledge Distillation and Target Interpolation</b> .....	2434
<i>Amit Das, Mark Hasegawa-Johnson</i>	
<b>Improving Cross-Lingual Knowledge Transferability Using Multilingual TDNN-BLSTM with Language-Dependent Pre-Final Layer</b> .....	2439
<i>Siyuan Feng, Tan Lee</i>	
<b>Auxiliary Feature Based Adaptation of End-to-end ASR Systems</b> .....	2444
<i>Marc Delcroix, Shinji Watanabe, Atsunori Ogawa, Shigeki Karita, Tomohiro Nakatani</i>	
<b>Leveraging Native Language Information for Improved Accented Speech Recognition</b> .....	2449
<i>Shahram Ghorbani, John H. L. Hansen</i>	
<b>Improved Accented Speech Recognition Using Accent Embeddings and Multi-task Learning</b> .....	2454
<i>Abhinav Jain, Minali Upreti, Preethi Jyothi</i>	
<b>Fast Language Adaptation Using Phonological Information</b> .....	2459
<i>Sibo Tong, Philip N. Garner, Hervé Bourlard</i>	

## **SPEECH SYNTHESIS PARADIGMS AND METHODS**

<b>Naturalness Improvement Algorithm for Reconstructed Glossectomy Patient's Speech Using Spectral Differential Modification in Voice Conversion</b> .....	2464
<i>Hiroki Murakami, Sunao Hara, Masanobu Abe, Masaaki Sato, Shogo Minagi</i>	
<b>Audio-visual Voice Conversion Using Deep Canonical Correlation Analysis for Deep Bottleneck Features</b> .....	2469
<i>Satoshi Tamura, Kento Horio, Hajime Endo, Satoru Hayamizu, Tomoki Toda</i>	
<b>An Investigation of Convolution Attention Based Models for Multilingual Speech Synthesis of Indian Languages</b> .....	2474
<i>Pallavi Baljekar, Saikrishna Rallabandi, Alan W Black</i>	
<b>The Effect of Real-Time Constraints on Automatic Speech Animation</b> .....	2479
<i>Damy Websdale, Sarah Taylor, Ben Milner</i>	
<b>Joint Learning of Facial Expression and Head Pose from Speech</b> .....	2484
<i>David Greenwood, Iain Matthews, Stephen Laycock</i>	
<b>Acoustic-dependent Phonemic Transcription for Text-to-speech Synthesis</b> .....	2489
<i>Kévin Vythelingum, Yannick Estève, Olivier Rosec</i>	
<b>Multimodal Speech Synthesis Architecture for Unsupervised Speaker Adaptation</b> .....	2494
<i>Hieu-Thi Luong, Junichi Yamagishi</i>	
<b>Articulatory-to-speech Conversion Using Bi-directional Long Short-term Memory</b> .....	2499
<i>Fumiaki Taguchi, Tokihiko Kaburagi</i>	
<b>Implementation of Respiration in Articulatory Synthesis Using a Pressure-Volume Lung Model</b> .....	2504
<i>Keisuke Tanihara, Shogo Yonekura, Yasuo Kuniyoshi</i>	
<b>Learning and Modeling Unit Embeddings for Improving HMM-based Unit Selection Speech Synthesis</b> .....	2509
<i>Xiao Zhou, Zhen-Hua Ling, Zhi-Ping Zhou, Li-Rong Dai</i>	
<b>Deep Metric Learning for the Target Cost in Unit-Selection Speech Synthesizer</b> .....	2514
<i>Ruibo Fu, Jianhua Tao, Yibin Zheng, Zhengqi Wen</i>	
<b>DNN-based Speech Synthesis for Small Data Sets Considering Bidirectional Speech-Text Conversion</b> .....	2519
<i>Kentarō Sone, Toru Nakashika</i>	
<b>A Weighted Superposition of Functional Contours Model for Modelling Contextual Prominence of Elementary Prosodic Contours</b> .....	2524
<i>Branislav Gerazov, Gérard Bailly, Yi Xu</i>	
<b>LSTBM: A Novel Sequence Representation of Speech Spectra Using Restricted Boltzmann Machine with Long Short-Term Memory</b> .....	2529
<i>Toru Nakashika</i>	

## **VOLUME 5**

### **SECOND LANGUAGE ACQUISITION AND CODE-SWITCHING**

<b>Should Code-switching Models Be Asymmetric?</b> .....	2534
<i>Barbara E. Bullock, Gualberto Guzmán, Jacqueline Serigos, Almeida Jacqueline Toribio</i>	
<b>Cross-language Perception of Mandarin Lexical Tones by Mongolian-speaking Bilinguals in the Inner Mongolia Autonomous Region, China</b> .....	2539
<i>Kimiko Tsukada, Yu Rong</i>	
<b>Automatically Measuring L2 Speech Fluency without the Need of ASR: A Proof-of-concept Study with Japanese Learners of French</b> .....	2544
<i>Lionel Fontan, Maxime Le Coz, Sylvain Detey</i>	
<b>Analysis of L2 Learners' Progress of Distinguishing Mandarin Tone 2 and Tone 3</b> .....	2549
<i>Yue Sun, Win Thuzar Kyaw, Jinsong Zhang, Yoshinori Sagisaka</i>	
<b>Unsupervised Discovery of Non-native Phonetic Patterns in L2 English Speech for Mispronunciation Detection and Diagnosis</b> .....	2554
<i>Xu Li, Shaoguang Mao, Xixin Wu, Kun Li, Xunying Liu, Helen Meng</i>	

<b>Wuxi Speakers' Production and Perception of Coda Nasals in Mandarin</b> .....	2559
<i>Lei Wang, Jie Cui, Ying Chen</i>	
<b>The Diphthongs of Formal Nigerian English: A Preliminary Acoustic Analysis</b> .....	2563
<i>Natalia Dyrenko, Robert Fuchs</i>	
<b>Characterizing Rhythm Differences between Strong and Weak Accented L2 Speech</b> .....	2568
<i>Chris Davis, Jeusun Kim</i>	
<b>Analysis of Phone Errors Attributable to Phonological Effects Associated With Language Acquisition Through Bottleneck Feature Visualisations</b> .....	2573
<i>Eva Fringi, Martin Russell</i>	
<b>Category Similarity in Multilingual Pronunciation Training</b> .....	2578
<i>Jacques Koreman</i>	
<b>Talker Diarization in the Wild: the Case of Child-centered Daylong Audio-recordings</b> .....	2583
<i>Alejandrina Cristia, Shobhana Ganesh, Marisa Casillas, Sriram Ganapathy</i>	
<b>Automated Classification of Children's Linguistic versus Non-Linguistic Vocalisations</b> .....	2588
<i>Zixing Zhang, Alejandrina Cristia, Anne Warlaumont, Björn Schuller</i>	
<b>Pitch Characteristics of L2 English Speech by Chinese Speakers: A Large-scale Study</b> .....	2593
<i>Jiahong Yuan, Qiusi Dong, Fei Wu, Huan Luan, Xiaofei Yang, Hui Lin, Yang Liu</i>	

## **TOPICS IN SPEECH RECOGNITION**

<b>Dual Language Models for Code Switched Speech Recognition</b> .....	2598
<i>Saurabh Garg, Tanmay Parekh, Preethi Jyothi</i>	
<b>Multilingual Neural Network Acoustic Modelling for ASR of Under-Resourced English-isiZulu Code-Switched Speech</b> .....	2603
<i>Astik Biswas, Febe De Wet, Ewald Van Der Westhuizen, Emre Yilmaz, Thomas Niesler</i>	
<b>Fast ASR-free and Almost Zero-resource Keyword Spotting Using DTW and CNNs for Humanitarian Monitoring</b> .....	2608
<i>Raghav Menon, Herman Kamper, John Quinn, Thomas Niesler</i>	
<b>Text-Dependent Speech Enhancement for Small-Footprint Robust Keyword Detection</b> .....	2613
<i>Meng Yu, Xuan Ji, Yi Gao, Lianwu Chen, Jie Chen, Jimeng Zheng, Dan Su, Dong Yu</i>	
<b>Improved ASR for Under-resourced Languages through Multi-task Learning with Acoustic Landmarks</b> .....	2618
<i>Di He, Boon Pang Lim, Xuesong Yang, Mark Hasegawa-Johnson, Deming Chen</i>	
<b>Cross-language Phoneme Mapping for Low-resource Languages: An Exploration of Benefits and Trade-offs</b> .....	2623
<i>Nick K Chibuye, Todd Rosenstock, Brian Derenzi</i>	
<b>User-centric Evaluation of Automatic Punctuation in ASR Closed Captioning</b> .....	2628
<i>Máté Ákos Tündik, György Szaszák, Gábor Gosztolya, András Beke</i>	
<b>Punctuation Prediction Model for Conversational Speech</b> .....	2633
<i>Piotr Zelasko, Piotr Szymanski, Jan Mizgajski, Adrian Szymczak, Yishay Carmiel, Najim Dehak</i>	
<b>BUT OpenSAT 2017 Speech Recognition System</b> .....	2638
<i>Martin Karafát, Murali Karthick Baskar, Igor Szöke, Vladimír Malenovský, Karel Veselý, František Grézl, Lukáš Burget, Jan Cernocký</i>	
<b>Visual Recognition of Continuous Cued Speech Using a Tandem CNN-HMM Approach</b> .....	2643
<i>Li Liu, Thomas Hueber, Gang Feng, Denis Beutemps</i>	
<b>Building Large-vocabulary Speaker-independent Lipreading Systems</b> .....	2648
<i>Kwanchiva Thangthai, Richard Harvey</i>	
<b>CRIM's System for the MGB-3 English Multi-Genre Broadcast Media Transcription</b> .....	2653
<i>Vishwa Gupta, Gilles Boulianne</i>	
<b>Sampling Strategies in Siamese Networks for Unsupervised Speech Representation Learning</b> .....	2658
<i>Rachid Riad, Corentin Dancette, Julien Karadayi, Neil Zeghidour, Thomas Schatz, Emmanuel Dupoux</i>	
<b>Compact Feedforward Sequential Memory Networks for Small-footprint Keyword Spotting</b> .....	2663
<i>Mengzhe Chen, Shiliang Zhang, Ming Lei, Yong Liu, Haitao Yao, Jie Gao</i>	

## **ZERO-RESOURCE SPEECH RECOGNITION**

<b>Multilingual Bottleneck Features for Subword Modeling in Zero-resource Languages</b> .....	2668
<i>Enno Hermann, Sharon Goldwater</i>	
<b>Exploiting Speaker and Phonetic Diversity of Mismatched Language Resources for Unsupervised Subword Modeling</b> .....	2673
<i>Siyuan Feng, Tan Lee</i>	
<b>Unsupervised Word Segmentation from Speech with Attention</b> .....	2678
<i>Pierre Godard, Marcely Zanon Boito, Lucas Ondel, Alexandre Berard, François Yvon, Aline Villavicencio, Laurent Besacier</i>	
<b>Learning Word Embeddings: Unsupervised Methods for Fixed-size Representations of Variable-length Speech Segments</b> .....	2683
<i>Nils Holzenberger, Mingxing Du, Julien Karadayi, Rachid Riad, Emmanuel Dupoux</i>	
<b>Full Bayesian Hidden Markov Model Variational Autoencoder for Acoustic Unit Discovery</b> .....	2688
<i>Thomas Glarner, Patrick Hanebrink, Janek Ebbes, Reinhold Haeb-Umbach</i>	
<b>Unspeech: Unsupervised Speech Context Embeddings</b> .....	2693
<i>Benjamin Milde, Chris Biemann</i>	



## SPATIAL AND PHASE CUES FOR SOURCE SEPARATION AND SPEECH RECOGNITION

<b>Impact of Aliasing on Deep CNN-Based End-to-End Acoustic Models</b> .....	2698
<i>Yuan Gong, Christian Poellabauer</i>	
<b>Keyword Based Speaker Localization: Localizing a Target Speaker in a Multi-speaker Environment</b> .....	2703
<i>Sunit Sivasankaran, Emmanuel Vincent, Dominique Fohr</i>	
<b>End-to-End Speech Separation with Unfolded Iterative Phase Reconstruction</b> .....	2708
<i>Zhong-Qiu Wang, Jonathan Le Roux, Deliang Wang, John Hershey</i>	
<b>PhaseNet: Discretized Phase Modeling with Deep Neural Networks for Audio Source Separation</b> .....	2713
<i>Naoya Takahashi, Purvi Agrawal, Nabarun Goswami, Yuki Mitsufoji</i>	
<b>Integrating Spectral and Spatial Features for Multi-Channel Speaker Separation</b> .....	2718
<i>Zhong-Qiu Wang, Deliang Wang</i>	
<b>DNN Driven Speaker Independent Audio-Visual Mask Estimation for Speech Separation</b> .....	2723
<i>Mandar Gogate, Ahsan Adeel, Ricard Marxer, Jon Barker, Amir Hussain</i>	

## DIALECTAL VARIATION

<b>Exploring Temporal Reduction in Dialectal Spanish: A Large-scale Study of Lenition of Voiced Stops and Coda-s</b> .....	2728
<i>Ioana Vasilescu, Nidia Hernandez, Bianca Vieru, Lori Lamel</i>	
<b>Dialect-geographical Acoustic-Tonetics: Five Disyllabic Tone Sandhi Patterns in Cognate Words from the Wu Dialects of ZheJiang Province</b> .....	2733
<i>Phil Rose</i>	
<b>Regional Variation of /r/ in Swiss German Dialects</b> .....	2738
<i>Adrian Leemann, Stephan Schmid, Dieter Studer-Joho, Marie-José Kolly</i>	
<b>Variation in the FACE Vowel Across West Yorkshire: Implications for Forensic Speaker Comparisons</b> .....	2743
<i>Kate Earnshaw, Erica Gold</i>	
<b>The ‘West Yorkshire Regional English Database’: Investigations into the Generalizability of Reference Populations for Forensic Speaker Comparison Casework</b> .....	2748
<i>Erica Gold, Sula Ross, Kate Earnshaw</i>	
<b>Studying Vowel Variation in French-Algerian Arabic Code-switched Speech</b> .....	2753
<i>Jane Wottawa, Djegdžiga Amazouz, Martine Adda-Decker, Lori Lamel</i>	

## SPOKEN CORPORA AND ANNOTATION

<b>Fearless Steps: Apollo-11 Corpus Advancements for Speech Technologies from Earth to the Moon</b> .....	2758
<i>John H. L. Hansen, Abhijeet Sangwan, Aditya Joglekar, Ahmet E. Bulut, Lakshmi Kaushik, Chengzhu Yu</i>	
<b>A Knowledge Driven Structural Segmentation Approach for Play-Talk Classification During Autism Assessment</b> .....	2763
<i>Manoj Kumar, Pooja Chebolu, So Hyun Kim, Kassandra Martinez, Catherine Lord, Shrikanth Narayanan</i>	
<b>An Open Source Emotional Speech Corpus for Human Robot Interaction Applications</b> .....	2768
<i>Jesin James, Li Tian, Catherine Inez Watson</i>	
<b>Speech Database and Protocol Validation Using Waveform Entropy</b> .....	2773
<i>Itshak Lapidot, Héctor Delgado, Massimiliano Todisco, Nicholas Evans, Jean-François Bonastre</i>	
<b>A French-Spanish Multimodal Speech Communication Corpus Incorporating Acoustic Data, Facial, Hands and Arms Gestures Information</b> .....	2778
<i>Lucas D. Terissi, Gonzalo Sad, Mauricio Cerda, Slim Ouni, Rodrigo Galvez, Juan C. Gómez, Bernard Girau, Nancy Hirschfeld-Kahler</i>	
<b>L2-ARCTIC: A Non-native English Speech Corpus</b> .....	2783
<i>Guanlong Zhao, Sinem Sonsaat, Alif Silpachai, Ivana Lucic, Evgeny Chukharev-Hudilainen, John Levis, Ricardo Gutierrez-Osuna</i>	

## THE FIRST DIHARD SPEECH DIARIZATION CHALLENGE

<b>ZCU-NTIS Speaker Diarization System for the DIHARD 2018 Challenge</b> .....	2788
<i>Zbynek Zajíc, Marie Kunešová, Jan Zelinka, Marek Hružík</i>	
<b>Speaker Diarization with Enhancing Speech for the First DIHARD Challenge</b> .....	2793
<i>Lei Sun, Jun Du, Chao Jiang, Xueyang Zhang, Shan He, Bing Yin, Chin-Hui Lee</i>	
<b>BUT System for DIHARD Speech Diarization Challenge 2018</b> .....	2798
<i>Mireia Diez, Federico Landini, Lukáš Burget, Johan Rohdin, Anna Šilnova, Katerina Žmolíková, Ondrej Novotný, Karel Veselý, Ondrej Glembek, Oldřich Plchot, Ladislav Mošner, Pavel Matejka</i>	
<b>Estimation of the Number of Speakers with Variational Bayesian PLDA in the DIHARD Diarization Challenge</b> .....	2803
<i>Ignacio Viñals, Pablo Gimeno, Alfonso Ortega, Antonio Miguel, Eduardo Lleida</i>	
<b>Diarization is Hard: Some Experiences and Lessons Learned for the JHU Team in the Inaugural DIHARD Challenge</b> .....	2808
<i>Gregory Sell, David Snyder, Alan McCree, Daniel Garcia-Romero, Jesús Villalba, Matthew Maciejewski, Vimal Manohar, Najim Dehak, Daniel Povey, Shinji Watanabe, Sanjeev Khudanpur</i>	
<b>The EURECOM Submission to the First DIHARD Challenge</b> .....	2813
<i>Jose Patino, Héctor Delgado, Nicholas Evans</i>	

<b>Joint Discriminative Embedding Learning, Speech Activity and Overlap Detection for the DIHARD Speaker Diarization Challenge</b> .....	2818
<i>Valter Akira Miasato Filho, Diego Augusto Silva, Luis Gustavo Depra Cuozzo</i>	

## **TEXT ANALYSIS, MULTILINGUAL ISSUES AND EVALUATION IN SPEECH SYNTHESIS**

<b>Multilingual Grapheme-to-Phoneme Conversion with Global Character Vectors</b> .....	2823
<i>Jinju Ni, Yoshinori Shiga, Hisashi Kawai</i>	
<b>A Hybrid Approach to Grapheme to Phoneme Conversion in Assamese</b> .....	2828
<i>Somnath Roy, Shakuntala Mahanta</i>	
<b>Investigation of Using Disentangled and Interpretable Representations for One-shot Cross-lingual Voice Conversion</b> .....	2833
<i>Seyed Hamidreza Mohammadi, Taehwan Kim</i>	
<b>Using Pupillometry to Measure the Cognitive Load of Synthetic Speech</b> .....	2838
<i>Avashna Govender, Simon King</i>	
<b>Measuring the Cognitive Load of Synthetic Speech Using a Dual Task Paradigm</b> .....	2843
<i>Avashna Govender, Simon King</i>	
<b>Attentive Sequence-to-Sequence Learning for Diacritic Restoration of Yoruba Language Text</b> .....	2848
<i>Iroro Orife</i>	
<b>Gated Convolutional Neural Network for Sentence Matching</b> .....	2853
<i>Peixin Chen, Wu Guo, Zhi Chen, Jian Sun, Lanhua You</i>	
<b>On Training and Evaluation of Grapheme-to-Phoneme Mappings with Limited Data</b> .....	2858
<i>Dravyansh Sharma</i>	
<b>The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech</b> .....	2863
<i>Alice Baird, Emilia Parada-Cabaleiro, Simone Hantke, Felix Burkhardt, Nicholas Cummins, Björn Schuller</i>	
<b>Word Emphasis Prediction for Expressive Text to Speech</b> .....	2868
<i>Yosi Mass, Slava Shechtman, Moran Mordechay, Ron Hoory, Oren Sar Shalom, Guy Lev, David Konopnicki</i>	
<b>A Comparison of Speaker-based and Utterance-based Data Selection for Text-to-Speech Synthesis</b> .....	2873
<i>Kai-Zhan Lee, Erica Cooper, Julia Hirschberg</i>	
<b>Data Requirements, Selection and Augmentation for DNN-based Speech Synthesis from Crowdsourced Data</b> .....	2878
<i>Markus Toman, Geoffrey S. Meltzer, Rupal Patel</i>	

## **NEURAL NETWORK TRAINING STRATEGIES FOR ASR**

<b>Lightly Supervised vs. Semi-supervised Training of Acoustic Model on Luxembourgish for Low-resource Automatic Speech Recognition</b> .....	2883
<i>Karel Veselý, Carlos Segura, Igor Szöke, Jordi Luque, Jan Cernocký</i>	
<b>Investigation on the Combination of Batch Normalization and Dropout in BLSTM-based Acoustic Modeling for ASR</b> .....	2888
<i>Wenjie Li, Gaofeng Cheng, Fengpei Ge, Pengyuan Zhang, Yonghong Yan</i>	
<b>Inference-Invariant Transformation of Batch Normalization for Domain Adaptation of Acoustic Models</b> .....	2893
<i>Masayuki Suzuki, Tohru Nagano, Gakuto Kurata, Samuel Thomas</i>	
<b>Active Learning for LF-MMI Trained Neural Networks in ASR</b> .....	2898
<i>Yanhua Long, Hong Ye, Yijie Li, Jiaen Liang</i>	
<b>An Investigation of Mixup Training Strategies for Acoustic Models in ASR</b> .....	2903
<i>Ivan Medennikov, Yuri Khokhlov, Aleksei Romanenko, Dmitry Popov, Natalia Tomashenko, Ivan Sorokin, Alexander Zatvornitskiy</i>	
<b>Comparison of Unsupervised Modulation Filter Learning Methods for ASR</b> .....	2908
<i>Purvi Agrawal, Sriram Ganapathy</i>	
<b>Improved Training for Online End-to-end Speech Recognition Systems</b> .....	2913
<i>Suyoun Kim, Michael Seltzer, Jinyu Li, Rui Zhao</i>	
<b>Combining Natural Gradient with Hessian Free Methods for Sequence Training</b> .....	2918
<i>Adnan Haider, Philip Woodland</i>	
<b>Lattice-free State-level Minimum Bayes Risk Training of Acoustic Models</b> .....	2923
<i>Naoyuki Kanda, Yusuke Fujita, Kenji Nagamatsu</i>	
<b>A Study of Enhancement, Augmentation and Autoencoder Methods for Domain Adaptation in Distant Speech Recognition</b> .....	2928
<i>Hao Tang, Wei-Ning Hsu, François Grondin, James Glass</i>	
<b>Multilingual Deep Neural Network Training Using Cyclical Learning Rate</b> .....	2933
<i>Andreas Søeberg Kirkedal, Yeon-Jun Kim</i>	

## **APPLICATION OF ASR IN MEDICAL PRACTICE**

<b>Development of the CUHK Dysarthric Speech Recognition System for the UA Speech Corpus</b> .....	2938
<i>Jianwei Yu, Xurong Xie, Shansong Liu, Shoukang Hu, Max W. Y. Lam, Xixin Wu, Ka Ho Wong, Xunying Liu, Helen Meng</i>	
<b>Automatic Evaluation of Speech Intelligibility Based on I-vectors in the Context of Head and Neck Cancers</b> .....	2943
<i>Imed Laaridh, Corinne Fredouille, Alain Ghio, Muriel Lalain, Virginie Woisard</i>	
<b>Dysarthric Speech Recognition Using Convolutional LSTM Neural Network</b> .....	2948
<i>Myungjong Kim, Beiming Cao, Kwanghoon An, Jun Wang</i>	

<b>Perceptual and Automatic Evaluations of the Intelligibility of Speech Degraded by Noise Induced Hearing Loss Simulation</b> .....	2953
<i>Imed Laaridh, Julien Tardieu, Cynthia Magnen, Pascal Gaillard, Jérôme Farinas, Julien Pinquier</i>	
<b>Articulatory Features for ASR of Pathological Speech</b> .....	2958
<i>Emre Yilmaz, Vikramjit Mitra, Chris Bartels, Horacio Franco</i>	
<b>Mining Multimodal Repositories for Speech Affecting Diseases</b> .....	2963
<i>Joana Correia, Bhiksha Raj, Isabel Trancoso, Francisco Teixeira</i>	
<b>Long Distance Voice Channel Diagnosis Using Deep Neural Networks</b> .....	2968
<i>Zhen Qin, Tom Ko, Guangjian Tian</i>	
<b>Speech Recognition for Medical Conversations</b> .....	2972
<i>Chung-Cheng Chiu, Anshuman Tripathi, Katherine Chou, Chris Co, Navdeep Jaitly, Diana Jaunzeikare, Anjali Kannan, Patrick Nguyen, Hasim Sak, Ananth Sankar, Justin Tansuwan, Nathan Wan, Yonghui Wu, Xuedong Zhang</i>	

## **SOURCE AND SUPRA-SEGMENTALS**

<b>Prosodic Focus Acquisition in French Early Cochlear Implanted Children</b> .....	2977
<i>Chadi Farah, Stephane Roman, Mariapaola D'Imperio</i>	
<b>The Role of Temporal Variation in Narrative Organization</b> .....	2982
<i>Nassima Fezza</i>	
<b>Interaction Mechanisms between Glottal Source and Vocal Tract in Pitch Glides</b> .....	2987
<i>Tiina Murtola, Jarmo Malinen</i>	
<b>Relating Articulatory Motions in Different Speaking Rates</b> .....	2992
<i>Astha Singh, G. Nisha Meenakshi, Prasanta Kumar Ghosh</i>	
<b>Estimation of the Asymmetry Parameter of the Glottal Flow Waveform Using the Electroglottographic Signal</b> .....	2997
<i>João Cabral</i>	
<b>Classification of Disorders in Vocal Folds Using Electroglottographic Signal</b> .....	3002
<i>Tanumay Mandal, K. Sreenivasa Rao, Sanjay Kumar Gupta</i>	
<b>Automatic Glottis Localization and Segmentation in Stroboscopic Videos Using Deep Neural Network</b> .....	3007
<i>Achuth Rao Mv, Rahul Krishnamurthy, Pebbili Gopikishore, Veeramani Priyadharshini, Prasanta Kumar Ghosh</i>	
<b>Respiratory and Respiratory Muscular Control in JL1's and JL2's Text Reading Utilizing 4-RSTs and a Soft Respiratory Mask with a Two-Way Bulb</b> .....	3012
<i>Toshiko Isei-Jaakkola, Keiko Ochi, Keikichi Hirose</i>	
<b>A Preliminary Study on Tonal Coarticulation in Continuous Speech</b> .....	3017
<i>Lixia Hao, Wei Zhang, Yanlu Xie, Jinsong Zhang</i>	

## **PLENARY TALK-3**

<b>Speech and Language Processing for Learning and Wellbeing</b> .....	3022
<i>Helen Meng</i>	

## **DISTANT ASR**

<b>Far-Field Speech Recognition Using Multivariate Autoregressive Models</b> .....	3023
<i>Sriram Ganapathy, Madhumita Harish</i>	
<b>Efficient Implementation of the Room Simulator for Training Deep Neural Network Acoustic Models</b> .....	3028
<i>Chanwoo Kim, Ehsan Variansi, Arun Narayanan, Michiel Bacchiani</i>	
<b>Stream Attention for Distributed Multi-Microphone Speech Recognition</b> .....	3033
<i>Xiaofei Wang, Ruizhi Li, Hynek Hermansky</i>	
<b>Recognizing Overlapped Speech in Meetings: A Multichannel Separation Approach Using Neural Networks</b> .....	3038
<i>Takuya Yoshioka, Hakan Erdogan, Zhuo Chen, Xiong Xiao, Fil Alleva</i>	
<b>Integrating Neural Network Based Beamforming and Weighted Prediction Error Dereverberation</b> .....	3043
<i>Lukas Drude, Christoph Boeddeker, Jahn Heymann, Reinhold Haeb-Umbach, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani</i>	
<b>A Probability Weighted Beamformer for Noise Robust ASR</b> .....	3048
<i>Suliang Bu, Yunxin Zhao, Meiyuh Hwang, Sining Sun</i>	

## **EXPRESSIVE SPEECH SYNTHESIS**

<b>Effects of Dimensional Input on Paralinguistic Information Perceived from Synthesized Dialogue Speech with Neural Network</b> .....	3053
<i>Masaki Yokoyama, Tomohiro Nagata, Hiroki Mori</i>	
<b>Neural MultiVoice Models for Expressing Novel Personalities in Dialog</b> .....	3057
<i>Shereen Oraby, Lena Reed, Sharath T. S., Shubhangi Tandon, Marilyn Walker</i>	
<b>Expressive Speech Synthesis Using Sentiment Embeddings</b> .....	3062
<i>Igor Jauk, Jaime Lorenzo-Trueba, Junichi Yamagishi, Antonio Bonafonte</i>	
<b>Expressive Speech Synthesis via Modeling Expressions with Variational Autoencoder</b> .....	3067
<i>Kei Akuzawa, Yusuke Iwasawa, Yutaka Matsuo</i>	

<b>Rapid Style Adaptation Using Residual Error Embedding for Expressive Speech Synthesis</b> .....	3072
<i>Xixin Wu, Yuewen Cao, Mu Wang, Songxiang Liu, Shiyin Kang, Zhiyong Wu, Xunying Liu, Dan Su, Dong Yu, Helen Meng</i>	
<b>EMPHASIS: An Emotional Phoneme-based Acoustic Model for Speech Synthesis System</b> .....	3077
<i>Hao Li, Yongguo Kang, Zhenyu Wang</i>	

### **REPRESENTATION LEARNING FOR EMOTION**

<b>Bags in Bag: Generating Context-Aware Bags for Tracking Emotions from Speech</b> .....	3082
<i>Jing Han, Zixing Zhang, Maximilian Schmitt, Zhao Ren, Fabien Ringeval, Björn Schuller</i>	
<b>An Attention Pooling Based Representation Learning Method for Speech Emotion Recognition</b> .....	3087
<i>Pengcheng Li, Yan Song, Ian McLoughlin, Wu Guo, Lirong Dai</i>	
<b>Predicting Arousal and Valence from Waveforms and Spectrograms Using Deep Neural Networks</b> .....	3092
<i>Zixiaofan Yang, Julia Hirschberg</i>	
<b>Emotion Identification from Raw Speech Signals Using DNNs</b> .....	3097
<i>Mousmita Sarma, Pegah Ghahremani, Daniel Povey, Nagendra Kumar Goel, Kandarpa Kumar Sarma, Najim Dehak</i>	
<b>Encoding Individual Acoustic Features Using Dyad-Augmented Deep Variational Representations for Dialog-level Emotion Recognition</b> .....	3102
<i>Jeng-Lin Li, Chi-Chun Lee</i>	
<b>Variational Autoencoders for Learning Latent Representations of Speech Emotion: A Preliminary Study</b> .....	3107
<i>Siddique Latif, Rajib Rana, Junaid Qadir, Julien Epps</i>	

### **ARTICULATORY INFORMATION, MODELING AND INVERSION**

<b>Phoneme-to-Articulatory Mapping Using Bidirectional Gated RNN</b> .....	3112
<i>Théo Biasutto-Lervat, Slim Ouni</i>	
<b>Tongue Segmentation with Geometrically Constrained Snake Model</b> .....	3117
<i>Zhihua Su, Jianguo Wei, Qiang Fang, Jianrong Wang, Kiyoshi Honda</i>	
<b>Low Resource Acoustic-to-articulatory Inversion Using Bi-directional Long Short Term Memory</b> .....	3122
<i>Aravind Illa, Prasanta Kumar Ghosh</i>	
<b>Automatic Visual Augmentation for Concatenation Based Synthesized Articulatory Videos from Real-time MRI Data for Spoken Language Training</b> .....	3127
<i>S. Chandana, Chiranjeevi Yarra, Ritu Aggarwal, Sanjeev Kumar Mittal, N. K. Kausthubha, K. T. Raseena, Astha Singh, Prasanta Kumar Ghosh</i>	
<b>Air-Tissue Boundary Segmentation in Real-Time Magnetic Resonance Imaging Video Using Semantic Segmentation with Fully Convolutional Networks</b> .....	3132
<i>C. A. Valliappan, Renuka Mannem, Prasanta Kumar Ghosh</i>	
<b>Noise Robust Acoustic to Articulatory Speech Inversion</b> .....	3137
<i>Nadee Seneviratne, Ganesh Sivaraman, Vikramjit Mitra, Carol Espy-Wilson</i>	

### **NOVEL PARADIGMS FOR DIRECT SYNTHESIS BASED ON SPEECH-RELATED BIOSIGNALS**

<b>Designing a Pneumatic Bionic Voice Prosthesis - A Statistical Approach for Source Excitation Generation</b> .....	3142
<i>Farzaneh Ahmadi, Tomoki Toda</i>	
<b>A Neural Model to Predict Parameters for a Generalized Command Response Model of Intonation</b> .....	3147
<i>Bastian Schnell, Philip N. Garner</i>	
<b>Articulation-to-Speech Synthesis Using Articulatory Flesh Point Sensors' Orientation Information</b> .....	3152
<i>Beiming Cao, Myungjong Kim, Jun R. Wang, Jan Van Santen, Ted Mau, Jun Wang</i>	
<b>Effectiveness of Generative Adversarial Network for Non-Audible Murmur-to-Whisper Speech Conversion</b> .....	3157
<i>Neil Shah, Nirmesh Shah, Hemant Patil</i>	
<b>Investigating Objective Intelligibility in Real-Time EMG-to-Speech Conversion</b> .....	3162
<i>Lorenz Diener, Tanja Schultz</i>	

## VOLUME 6

<b>Domain-Adversarial Training for Session Independent EMG-based Speech Recognition</b> .....	3167
<i>Michael Wand, Tanja Schultz, Jürgen Schmidhuber</i>	
<b>Multi-Task Learning of Speech Recognition and Speech Synthesis Parameters for Ultrasound-based Silent Speech Interfaces</b> .....	3172
<i>László Tóth, Gábor Gosztolya, Tamás Grósz, Alexandra Markó, Tamás Gábor Csapó</i>	

### **LOW RESOURCE SPEECH RECOGNITION CHALLENGE FOR INDIAN LANGUAGES**

<b>Transcription Correction for Indian Languages Using Acoustic Signatures</b> .....	3177
<i>Jeena Jprakash, Golda Brunet Rajan, Hema Murthy</i>	
<b>BUT System for Low Resource Indian Language ASR</b> .....	3182
<i>Bhargav Pulugundla, Murali Karthick Baskar, Santosh Kesiraju, Ekaterina Egorova, Martin Karafiát, Lukáš Burget, Jan Cernocký</i>	

<b>DA-IICT/IITV System for Low Resource Speech Recognition Challenge 2018</b> .....	3187
<i>Hardik B. Sailor, Maddala Venkata Siva Krishna, Diksha Chhabra, Ankur T. Patil, Madhu Kamble, Hemant Patil</i>	
<b>An Exploration Towards Joint Acoustic Modeling for Indian Languages: IIT-H Submission for Low Resource Speech Recognition Challenge for Indian Languages, INTERSPEECH 2018</b> .....	3192
<i>Hari Krishna Vydana, Krishna Gurugubelli, Vishnu Vidyadhara Raju, Anil Kumar Vuppala</i>	
<b>TDNN-based Multilingual Speech Recognition System for Low Resource Indian Languages</b> .....	3197
<i>Fathima Noor, Patel Tanvina, C. Mahima, Iyengar Anuroop</i>	
<b>Articulatory and Stacked Bottleneck Features for Low Resource Speech Recognition</b> .....	3202
<i>Vishwas M. Shetty, Rini A Sharon, Basil Abraham, Tejaswi Seeram, Anusha Prakash, Nithya Ravi, S. Umesh</i>	
<b>ISI ASR System for the Low Resource Speech Recognition Challenge for Indian Languages</b> .....	3207
<i>Jayadev Billa</i>	

## **SHOW AND TELL 7**

<b>An Automated Assistant for Medical Scribes</b> .....	3212
<i>Gregory Finley, Erik Edwards, Amanda Robinson, Najmeh Sadoughi, James Fone, Mark Miller, David Suendermann-Oeft, Michael Brenndorfer, Nico Axtmann</i>	
<b>AGROASSAM: A Web Based Assamese Speech Recognition Application for Retrieving Agricultural Commodity Price and Weather Information</b> .....	3214
<i>Abhishek Dey, Abhash Deka, Siddika Imani, Barsha Deka, Rohit Sinha, S R Mahadeva Prasanna, Priyankoo Samah, K. Samudravijaya, S. R. Nirmala</i>	
<b>Voice-powered Solutions with Cloud AI</b> .....	3216
<i>Dan Aharon</i>	
<b>Speech Synthesis in the Wild</b> .....	3217
<i>Ganesh Sivaraman, Parav Nagarsheth, Elie Khoury</i>	

## **DEEP ENHANCEMENT**

<b>Deep Noise Tracking Network: A Hybrid Signal Processing/Deep Learning Approach to Speech Enhancement</b> .....	3219
<i>Shuai Nie, Shan Liang, Bin Liu, Yaping Zhang, Wenju Liu, Jianhua Tao</i>	
<b>A Deep Neural Network Based Harmonic Noise Model for Speech Enhancement</b> .....	3224
<i>Zhiheng Ouyang, Hongjiang Yu, Wei-Ping Zhu, Benoit Champagne</i>	
<b>A Convolutional Recurrent Neural Network for Real-Time Speech Enhancement</b> .....	3229
<i>Ke Tan, Deliang Wang</i>	
<b>All-Neural Multi-Channel Speech Enhancement</b> .....	3234
<i>Zhong-Qiu Wang, Deliang Wang</i>	
<b>Deep Learning for Acoustic Echo Cancellation in Noisy and Double-Talk Scenarios</b> .....	3239
<i>Hao Zhang, Deliang Wang</i>	
<b>The Conversation: Deep Audio-Visual Speech Enhancement</b> .....	3244
<i>Triantafyllos Afouras, Joon Son Chung, Andrew Zisserman</i>	
<b>Student-Teacher Learning for BLSTM Mask-based Speech Enhancement</b> .....	3249
<i>Aswin Shanmugam Subramanian, Szu-Jui Chen, Shinji Watanabe</i>	
<b>Speech Enhancement Using Deep Mixture of Experts Based on Hard Expectation Maximization</b> .....	3254
<i>Pavan Karjol, Prasanta Kumar Ghosh</i>	
<b>Adversarial Feature-Mapping for Speech Enhancement</b> .....	3259
<i>Zhong Meng, Jinyu Li, Yifan Gong, Bing-Hwang (Fred) Juang</i>	
<b>Biophysically-inspired Features Improve the Generalizability of Neural Network-based Speech Enhancement Systems</b> .....	3264
<i>Deepak Baby, Sarah Verhulst</i>	
<b>Error Modeling via Asymmetric Laplace Distribution for Deep Neural Network Based Single-Channel Speech Enhancement</b> .....	3269
<i>Li Chai, Jun Du, Chin-Hui Lee</i>	
<b>A Prior SNR Estimation Based on a Recurrent Neural Network for Robust Speech Enhancement</b> .....	3274
<i>Yangyang Xia, Richard Stern</i>	

## **ACOUSTIC SCENES AND RARE EVENTS**

<b>Multiple Instance Deep Learning for Weakly Supervised Small-Footprint Audio Event Detection</b> .....	3279
<i>Shao-Yen Tseng, Juncheng Li, Yun Wang, Florian Metze, Joseph Szurley, Samarjit Das</i>	
<b>Unsupervised Temporal Feature Learning Based on Sparse Coding Embedded BoAW for Acoustic Event Recognition</b> .....	3284
<i>Liwen Zhang, Jiqing Han, Shiwen Deng</i>	
<b>Data Independent Sequence Augmentation Method for Acoustic Scene Classification</b> .....	3289
<i>Teng Zhang, Kailai Zhang, Ji Wu</i>	
<b>A Compact and Discriminative Feature Based on Auditory Summary Statistics for Acoustic Scene Classification</b> .....	3294
<i>Hongwei Song, Jiqing Han, Shiwen Deng</i>	
<b>ASe: Acoustic Scene Embedding Using Deep Archetypal Analysis and GMM</b> .....	3299
<i>Pulkit Sharma, Vinayak Abrol, Anshul Thakur</i>	

<b>Deep Convolutional Neural Network with Scalogram for Audio Scene Modeling</b> .....	3304
<i>Hangting Chen, Pengyuan Zhang, Haichuan Bai, Qingsheng Yuan, Xiuguo Bao, Yonghong Yan</i>	
<b>Time Aggregation Operators for Multi-label Audio Event Detection</b> .....	3309
<i>Pankaj Joshi, Digvijaysingh Gautam, Ganesh Ramakrishnan, Preethi Jyothi</i>	
<b>Early Detection of Continuous and Partial Audio Events Using CNN</b> .....	3314
<i>Ian McLoughlin, Yan Song, Lam Dang Pham, Ramaswamy Palaniappan, Huy Phan, Yue Lang</i>	
<b>Robust Acoustic Event Classification Using Bag-of-Visual-Words</b> .....	3319
<i>Manjunath Mulimani, Shashidhar G Koolagudi</i>	
<b>Wavelet Transform Based Mel-scaled Features for Acoustic Scene Classification</b> .....	3323
<i>Shefali Waldekar, Goutam Saha</i>	
<b>Multi-modal Attention Mechanisms in LSTM and Its Application to Acoustic Scene Classification</b> .....	3328
<i>Teng Zhang, Kailai Zhang, Ji Wu</i>	

## LANGUAGE MODELING

<b>Contextual Language Model Adaptation for Conversational Agents</b> .....	3333
<i>Anirudh Raju, Behnam Hedayatnia, Linda Liu, Ankur Gandhe, Chandra Khatri, Angeliki Metallinou, Anu Venkatesh, Ariya Rastrow</i>	
<b>Active Memory Networks for Language Modeling</b> .....	3338
<i>Oscar Chen, Anton Ragni, Mark Gales, Xie Chen</i>	
<b>Unsupervised and Efficient Vocabulary Expansion for Recurrent Neural Network Language Models in ASR</b> .....	3343
<i>Yerbolat Khassanov, Eng Siong Chng</i>	
<b>Improving Language Modeling with an Adversarial Critic for Automatic Speech Recognition</b> .....	3348
<i>Yike Zhang, Pengyuan Zhang, Yonghong Yan</i>	
<b>Training Recurrent Neural Network through Moment Matching for NLP Applications</b> .....	3353
<i>Yue Deng, Yilin Shen, Kawai Chen, Hongxia Jin</i>	
<b>Investigation on LSTM Recurrent N-gram Language Models for Speech Recognition</b> .....	3358
<i>Zoltán Tüske, Ralf Schlüter, Hermann Ney</i>	
<b>Online Incremental Learning for Speaker-Adaptive Language Models</b> .....	3363
<i>Chih Chi Hu, Bing Liu, John Shen, Ian Lane</i>	
<b>Efficient Language Model Adaptation with Noise Contrastive Estimation and Kullback-Leibler Regularization</b> .....	3368
<i>Jesús Andrés-Ferrer, Nathan Bodenstab, Paul Vozila</i>	
<b>Recurrent Neural Network Language Model Adaptation for Conversational Speech Recognition</b> .....	3373
<i>Ke Li, Hainan Xu, Yiming Wang, Daniel Povey, Sanjeev Khudanpur</i>	
<b>What to Expect from Expected Kneser-Ney Smoothing</b> .....	3378
<i>Michael Levit, Sarangarajan Parthasarathy, Shuangyu Chang</i>	
<b>i-Vectors in Language Modeling: An Efficient Way of Domain Adaptation for Feed-Forward Models</b> .....	3383
<i>Karel Beneš, Santosh Kesiraju, Lukas Burget</i>	

## SPEECH PATHOLOGY, DEPRESSION, AND MEDICAL APPLICATIONS

<b>How Did You like 2017? Detection of Language Markers of Depression and Narcissism in Personal Narratives</b> .....	3388
<i>Eva-Maria Rathner, Julia Djamali, Yannik Terhorst, Björn Schuller, Nicholas Cummins, Gudrun Salamon, Christina Hunger-Schoppe, Harald Baumeister</i>	
<b>Depression Detection from Short Utterances via Diverse Smartphones in Natural Environmental Conditions</b> .....	3393
<i>Zhaocheng Huang, Julien Epps, Dale Joachim, Michael Chen</i>	
<b>Multi-Lingual Depression-Level Assessment from Conversational Speech Using Acoustic and Text Features</b> .....	3398
<i>Yasin Özkanca, Cenk Demiroglu, Asli Besirli, Selime Celik</i>	
<b>Dysarthric Speech Classification Using Glottal Features Computed from Non-words, Words and Sentences</b> .....	3403
<i>N. P. Narendra, Paavo Alku</i>	
<b>Identifying Schizophrenia Based on Temporal Parameters in Spontaneous Speech</b> .....	3408
<i>Gábor Gosztoya, Anita Bagi, Szilvia Szalóki, István Szendi, Ildikó Hoffmann</i>	
<b>Using Prosodic and Lexical Information for Learning Utterance-level Behaviors in Psychotherapy</b> .....	3413
<i>Karan Singla, Zhuohao Chen, Nikolaos Flemotomos, James Gibson, Dogan Can, David Atkins, Shrikanth Narayanan</i>	
<b>Automatic Speech Assessment for People with Aphasia Using TDNN-BLSTM with Multi-Task Learning</b> .....	3418
<i>Ying Qin, Tan Lee, Siyuan Feng, Anthony Pak Hin Kong</i>	
<b>Towards an Unsupervised Entrainment Distance in Conversational Speech Using Deep Neural Networks</b> .....	3423
<i>Md Nasir, Brian Baucom, Shrikanth Narayanan, Panayiotis Georgiou</i>	
<b>Patient Privacy in Paralinguistic Tasks</b> .....	3428
<i>Francisco Teixeira, Alberto Abad, Isabel Trancoso</i>	
<b>A Lightly Supervised Approach to Detect Stuttering in Children's Speech</b> .....	3433
<i>Sadeen Alharbi, Madina Hasan, Anthony J H Simons, Shelagh Brumfit, Phil Green</i>	
<b>Learning Conditional Acoustic Latent Representation with Gender and Age Attributes for Automatic Pain Level Recognition</b> .....	3438
<i>Jeng-Lin Li, Yi-Ming Weng, Chip-Jin Ng, Chi-Chun Lee</i>	

## **PERSPECTIVE TALK-4**

<b>Speaker and Language Recognition -- From Laboratory Technologies to the Wild</b> .....	3443
<i>Sriram Ganapathy</i>	

## **SPOKEN LANGUAGE UNDERSTANDING**

<b>A Deep Reinforcement Learning Based Multimodal Coaching Model (DCM) for Slot Filling in Spoken Language Understanding(SLU)</b> .....	3444
<i>Yu Wang, Abhishek Patel, Yilin Shen, Hongxia Jin</i>	
<b>Is ATIS Too Shallow to Go Deeper for Benchmarking Spoken Language Understanding Models?</b> .....	3449
<i>Frédéric Béchet, Christian Raymond</i>	
<b>Robust Spoken Language Understanding via Paraphrasing</b> .....	3454
<i>Avik Ray, Yilin Shen, Hongxia Jin</i>	
<b>Spoken SQuAD: A Study of Mitigating the Impact of Speech Recognition Errors on Listening Comprehension</b> .....	3459
<i>Chia-Hsuan Lee, Szu-Lin Wu, Chi-Liang Liu, Hung-Yi Lee</i>	
<b>User Information Augmented Semantic Frame Parsing Using Progressive Neural Networks</b> .....	3464
<i>Yilin Shen, Xiangyu Zeng, Yu Wang, Hongxia Jin</i>	
<b>An Efficient Approach to Encoding Context for Spoken Language Understanding</b> .....	3469
<i>Raghav Gupta, Abhinav Rastogi, Dilek Hakkani-Tür</i>	

## **SOURCE SEPARATION FROM MONAURAL INPUT**

<b>Deep Speech Denoising with Vector Space Projections</b> .....	3474
<i>Jeffrey Hetherly, Paul Gamble, Maria Alejandra Barrios, Cory Stephenson, Karl Ni</i>	
<b>A Shifted Delta Coefficient Objective for Monaural Speech Separation Using Multi-task Learning</b> .....	3479
<i>Chenglin Xu, Wei Rao, Eng Siong Chng, Haizhou Li</i>	
<b>A Two-Stage Approach to Noisy Cochannel Speech Separation with Gated Residual Networks</b> .....	3484
<i>Ke Tan, Deliang Wang</i>	
<b>Monoaural Audio Source Separation Using Variational Autoencoders</b> .....	3489
<i>Laxmi Pandey, Anurendra Kumar, Vinay Namboodiri</i>	
<b>Towards Automated Single Channel Source Separation Using Neural Networks</b> .....	3494
<i>Arpita Gang, Pravesh Biyani, Akshay Soni</i>	
<b>Investigations on Data Augmentation and Loss Functions for Deep Learning Based Speech-Background Separation</b> .....	3499
<i>Hakan Erdogan, Takuya Yoshioka</i>	

## **MULTIMODAL SYSTEMS**

<b>Annotator Trustability-based Cooperative Learning Solutions for Intelligent Audio Analysis</b> .....	3504
<i>Simone Hantke, Christoph Stemp, Björn Schuller</i>	
<b>Semi-supervised Cross-domain Visual Feature Learning for Audio-Visual Broadcast Speech Transcription</b> .....	3509
<i>Rongfeng Su, Xunying Liu, Lan Wang</i>	
<b>Deep Lip Reading: A Comparison of Models and an Online Application</b> .....	3514
<i>Triantafyllos Afouras, Joon Son Chung, Andrew Zisserman</i>	
<b>Iterative Learning of Speech Recognition Models for Air Traffic Control</b> .....	3519
<i>Ajay Srinivasamurthy, Petr Motlicek, Mittul Singh, Youssef Oualil, Matthias Kleinert, Heiko Ehr, Hartmut Helmke</i>	
<b>Speaker Adaptive Audio-Visual Fusion for the Open-Vocabulary Section of AVICAR</b> .....	3524
<i>Leda Sari, Mark Hasegawa-Johnson, S. Kumaran, Georg Stemmer, Krishnakumar N Nair</i>	
<b>Multimodal Name Recognition in Live TV Subtitling</b> .....	3529
<i>Marek Hruš, Aleš Pražák, Michal Bušta</i>	

## **CODING**

<b>Dithered Quantization for Frequency-Domain Speech and Audio Coding</b> .....	3533
<i>Tom Bäckström, Johannes Fischer, Sneha Das</i>	
<b>Postfiltering with Complex Spectral Correlations for Speech and Audio Coding</b> .....	3538
<i>Sneha Das, Tom Bäckström</i>	
<b>Postfiltering Using Log-Magnitude Spectrum for Speech and Audio Coding</b> .....	3543
<i>Sneha Das, Tom Bäckström</i>	
<b>Temporal Noise Shaping with Companding</b> .....	3548
<i>Arijit Biswas, Per Hedelin, Lars Villemoes, Vinay Melkote</i>	
<b>Multi-frame Quantization of LSF Parameters Using a Deep Autoencoder and Pyramid Vector Quantizer</b> .....	3553
<i>Yaxing Li, Eshete Derb Emiru, Shengwu Xiong, Anna Zhu, Pengfei Duan, Yichang Li</i>	
<b>Multi-frame Coding of LSF Parameters Using Block-Constrained Trellis Coded Vector Quantization</b> .....	3558
<i>Yaxing Li, Shan Xu, Shengwu Xiong, Anna Zhu, Pengfei Duan, Yueming Ding</i>	

## SPEAKER VERIFICATION USING NEURAL NETWORK METHODS II

<b>Training Utterance-level Embedding Networks for Speaker Identification and Verification</b> .....	3563
<i>Heewoong Park, Sukhyun Cho, Kyubyong Park, Namju Kim, Jonghun Park</i>	
<b>Analysis of Complementary Information Sources in the Speaker Embeddings Framework</b> .....	3568
<i>Mahesh Kumar Nandwana, Mitchell McLaren, Diego Castan, Julien Van Hout, Aaron Lawson</i>	
<b>Self-Attentive Speaker Embeddings for Text-Independent Speaker Verification</b> .....	3573
<i>Yingke Zhu, Tom Ko, David Snyder, Brian Mak, Daniel Povey</i>	
<b>An Improved Deep Embedding Learning Method for Short Duration Speaker Verification</b> .....	3578
<i>Zhifu Gao, Yan Song, Ian McLoughlin, Wu Guo, Lirong Dai</i>	
<b>Avoiding Speaker Overfitting in End-to-End DNNs Using Raw Waveform for Text-Independent Speaker Verification</b> .....	3583
<i>Jee-Weon Jung, Hee-Soo Heo, Il-Ho Yang, Hye-Jin Shim, Ha-Jin Yu</i>	
<b>Deeply Fused Speaker Embeddings for Text-Independent Speaker Verification</b> .....	3588
<i>Gautam Bhattacharya, Jahangir Alam, Vishwa Gupta, Patrick Kenny</i>	
<b>Employing Phonetic Information in DNN Speaker Embeddings to Improve Speaker Recognition Performance</b> .....	3593
<i>Md Hafizur Rahman, Ivan Himawan, Mitchell McLaren, Clinton Fookes, Sridha Sridharan</i>	
<b>End-to-end Text-dependent Speaker Verification Using Novel Distance Measures</b> .....	3598
<i>Subhadeep Dey, Srikanth Madikeri, Petr Motlicek</i>	
<b>Robust Speaker Clustering using Mixtures of von Mises-Fisher Distributions for Naturalistic Audio Streams</b> .....	3603
<i>Harishchandra Dubey, Abhijeet Sangwan, John H. L. Hansen</i>	
<b>Triplet Network with Attention for Speaker Diarization</b> .....	3608
<i>Huan Song, Megan Willi, Jayaraman J. Thiagarajan, Visar Berisha, Andreas Spanias</i>	
<b>I-vector Transformation Using Conditional Generative Adversarial Networks for Short Utterance Speaker Verification</b> .....	3613
<i>Jiacen Zhang, Nakamasa Inoue, Koichi Shinoda</i>	
<b>Analysis of Length Normalization in End-to-End Speaker Verification System</b> .....	3618
<i>Weicheng Cai, Jinkun Chen, Ming Li</i>	
<b>Angular Softmax for Short-Duration Text-independent Speaker Verification</b> .....	3623
<i>Zili Huang, Shuai Wang, Kai Yu</i>	
<b>An End-to-End Text-Independent Speaker Identification System on Short Utterances</b> .....	3628
<i>Ruifang Ji, Xinyuan Cai, Xu Bo</i>	
<b>MTGAN: Speaker Verification through Multitasking Triplet Generative Adversarial Networks</b> .....	3633
<i>Wenhao Ding, Liang He</i>	

## EMOTION RECOGNITION AND ANALYSIS

<b>Categorical vs Dimensional Perception of Italian Emotional Speech</b> .....	3638
<i>Emilia Parada-Cabaleiro, Giovanni Costantini, Anton Batliner, Alice Baird, Björn Schuller</i>	
<b>A Three-Layer Emotion Perception Model for Valence and Arousal-Based Detection from Multilingual Speech</b> .....	3643
<i>Xingfeng Li, Masato Akagi</i>	
<b>Cross-lingual Speech Emotion Recognition through Factor Analysis</b> .....	3648
<i>Brecht Desplanques, Kris Demuynck</i>	
<b>Modeling Self-Reported and Observed Affect from Speech</b> .....	3653
<i>Jian Cheng, Jared Bernstein, Elizabeth Rosenfeld, Peter W. Foltz, Alex S. Cohen, Terje B. Holmlund, Brita Elvevag</i>	
<b>Stochastic Shake-Shake Regularization for Affective Learning from Speech</b> .....	3658
<i>Che-Wei Huang, Shrikanth Narayanan</i>	
<b>Investigating Speech Enhancement and Perceptual Quality for Speech Emotion Recognition</b> .....	3663
<i>Anderson R. Avila, Jahangir Alam, Douglas O'Shaughnessy, Tiago Falk</i>	
<b>Demonstrating and Modelling Systematic Time-varying Annotator Disagreement in Continuous Emotion Annotation</b> .....	3668
<i>Mia Atcheson, Vidhyasaharan Sethu, Julien Epps</i>	
<b>Speech Emotion Recognition from Variable-Length Inputs with Triplet Loss Function</b> .....	3673
<i>Jian Huang, Ya Li, Jianhua Tao, Zhen Lian</i>	
<b>Imbalance Learning-based Framework for Fear Recognition in the MediaEval Emotional Impact of Movies Task</b> .....	3678
<i>Xiaotong Zhang, Xingliang Cheng, Mingxing Xu, Thomas Fang Zheng</i>	
<b>Emotion Recognition from Variable-Length Speech Segments Using Deep Learning on Spectrograms</b> .....	3683
<i>Xi Ma, Zhiyong Wu, Jia Jia, Mingxing Xu, Helen Meng, Lianhong Cai</i>	
<b>Speech Emotion Recognition Using Spectrogram &amp; Phoneme Embedding</b> .....	3688
<i>Promod Yenigalla, Abhay Kumar, Suraj Tripathi, Chirag Singh, Sibsambhu Kar, Jithendra Vepa</i>	
<b>On Enhancing Speech Emotion Recognition Using Generative Adversarial Networks</b> .....	3693
<i>Saurabh Sahu, Rahul Gupta, Carol Espy-Wilson</i>	
<b>Ladder Networks for Emotion Recognition: Using Unsupervised Auxiliary Tasks to Improve Predictions of Emotional Attributes</b> .....	3698
<i>Srinivas Parthasarathy, Carlos Busso</i>	



## ACOUSTIC MODELLING

<b>Knowledge Distillation for Sequence Model</b> .....	3703
<i>Mingkun Huang, Yongbin You, Zhehuai Chen, Yanmin Qian, Kai Yu</i>	
<b>Improving CTC-based Acoustic Model with Very Deep Residual Time-delay Neural Networks</b> .....	3708
<i>Sheng Li, Xugang Lu, Ryoichi Takashima, Peng Shen, Tatsuya Kawahara, Hisashi Kawai</i>	
<b>Filter Sampling and Combination CNN (FSC-CNN): A Compact CNN Model for Small-footprint ASR Acoustic Modeling Using Raw Waveforms</b> .....	3713
<i>Jinxi Guo, Ning Xu, Xin Chen, Yang Shi, Kaiyuan Xu, Abeer Alwan</i>	
<b>Twin Regularization for Online Speech Recognition</b> .....	3718
<i>Mirco Ravanelli, Dmitriy Serdyuk, Yoshua Bengio</i>	
<b>Self-Attentional Acoustic Models</b> .....	3723
<i>Mathias Sperber, Jan Niehues, Graham Neubig, Sebastian Stüker, Alex Waibel</i>	
<b>Hierarchical Recurrent Neural Networks for Acoustic Modeling</b> .....	3728
<i>Jinhwan Park, Iksoo Choi, Yoonho Boo, Wonyong Sung</i>	
<b>Dictionary Augmented Sequence-to-Sequence Neural Network for Grapheme to Phoneme Prediction</b> .....	3733
<i>Antoine Bruguier, Anton Bakhtin, Dravyansh Sharma</i>	
<b>Leveraging Second-Order Log-Linear Model for Improved Deep Learning Based ASR Performance</b> .....	3738
<i>Ankit Raj, Shakti P Rath, Jithendra Vepa</i>	
<b>Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks</b> .....	3743
<i>Daniel Povey, Gaofeng Cheng, Yiming Wang, Ke Li, Hainan Xu, Mahsa Yarmohammadi, Sanjeev Khudanpur</i>	
<b>Completely Unsupervised Phoneme Recognition by Adversarial Learning Mapping Relationships from Audio Embeddings</b> .....	3748
<i>Da-Rong Liu, Kuan-Yu Chen, Hung-Yi Lee, Lin-Shan Lee</i>	
<b>Phone Recognition Using a Non-Linear Manifold with Broad Phone Class Dependent DNNs</b> .....	3753
<i>Mengjie Qian, Linxue Bai, Peter Jancovic, Martin Russell</i>	
<b>A Multi-Discriminator CycleGAN for Unsupervised Non-Parallel Speech Domain Adaptation</b> .....	3758
<i>Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, Richard Socher</i>	

## SPEECH AND SPEAKER PERCEPTION

<b>Interactions between Vowels and Nasal Coda in Mandarin Speakers' Perception of Nasal Finals</b> .....	3763
<i>Chong Cao, Wei Wei, Wei Wang, Yanlu Xie, Jinsong Zhang</i>	
<b>Weighting Pitch Contour and Loudness Contour in Mandarin Tone Perception in Cochlear Implant Listeners</b> .....	3768
<i>Qinglin Meng, Nengheng Zheng, Ambika Prasad Mishra, Jacinta Dan Luo, Jan W. H. Schnupp</i>	
<b>Implementing DIANA to Model Isolated Auditory Word Recognition in English</b> .....	3772
<i>Filip Nenadic, Louis Ten Bosch, Benjamin V. Tucker</i>	
<b>Effects of Homophone Density on Spoken Word Recognition in Mandarin Chinese</b> .....	3777
<i>Bhamini Sharma</i>	
<b>Visual Timing Information in Audiovisual Speech Perception: Evidence from Lexical Tone Contour</b> .....	3781
<i>Hui Xie, Biao Zeng, Rui Wang</i>	
<b>COSMO SylPhon: A Bayesian Perceptuo-motor Model to Assess Phonological Learning</b> .....	3786
<i>Marie-Lou Barnaud, Juïen Diard, Pierre Bessière, Jean-Luc Schwartz</i>	
<b>Experience-dependent Influence of Music and Language on Lexical Pitch Learning Is Not Additive</b> .....	3791
<i>Akshay Raj Maggu, Patrick C. M. Wong, Hanjun Liu, Francis C. K. Wong</i>	
<b>Influences of Fundamental Oscillation on Speaker Identification in Vocalic Utterances by Humans and Computers</b> .....	3795
<i>Volker Dellwo, Thayabaran Kathiresan, Elisa Pellegrino, Lei He, Sandra Schwab, Dieter Maurer</i>	
<b>Author Index</b>	