

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

Additional copies of this publication are available from:

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

TABLE OF CONTENTS

VOLUME 1

ISCA MEDAL TALK

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

END-TO-END SPEECH RECOGNITION

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

PROSODY MODELING AND GENERATION

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

SPEAKER VERIFICATION I

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

SPOKEN TERM DETECTION

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

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

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

SHOW AND TELL 1

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

SPEECH SEGMENTS AND VOICE QUALITY

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

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

SPEAKER STATE AND TRAIT

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

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

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

ACOUSTIC ANALYSIS-SYNTHESIS OF SPEECH DISORDERS

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

ASR SYSTEMS AND TECHNOLOGIES

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

DECEPTION, PERSONALITY, AND CULTURE ATTRIBUTE

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

AUTOMATIC DETECTION AND RECOGNITION OF VOICE AND SPEECH DISORDERS

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

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

VOICE CONVERSION

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

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

SHOW AND TELL 2

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

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

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

SPOOFING DETECTION

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

VOLUME 2

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

SPEECH ANALYSIS AND REPRESENTATION

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

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

SEQUENCE MODELS FOR ASR

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

SOURCE SEPARATION AND SPATIAL ANALYSIS

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

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

PLENARY TALK-1

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

ACOUSTIC MODEL ADAPTATION

Learning to Adapt: A Meta-learning Approach for Speaker Adaptation	867
<i>Ondrej Klejch, Joachim Fainberg, Peter Bell</i>	
Speaker Adaptation and Adaptive Training for Jointly Optimised Tandem Systems	872
<i>Yu Wang, Chao Zhang, Mark Gales, Philip Woodland</i>	
Comparison of BLSTM-Layer-Specific Affine Transformations for Speaker Adaptation	877
<i>Markus Kitzka, Ralf Schlüter, Hermann Ney</i>	
Correlational Networks for Speaker Normalization in Automatic Speech Recognition	882
<i>Rini A Sharon, Sandeep Reddy Kothinti, Umesh Srinivasan</i>	
Machine Speech Chain with One-shot Speaker Adaptation	887
<i>Andros Tjandra, Sakriani Sakti, Satoshi Nakamura</i>	
Domain Adaptation Using Factorized Hidden Layer for Robust Automatic Speech Recognition	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

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

EMOTION MODELING

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

MODELS OF SPEECH PERCEPTION

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

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

MULTIMODAL DIALOGUE SYSTEMS

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

SPEECH RECOGNITION FOR INDIAN LANGUAGES

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

SHOW AND TELL 3

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

SPEAKER VERIFICATION II

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

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

NOVEL APPROACHES TO ENHANCEMENT

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

SYLLABIFICATION, RHYTHM, AND VOICE ACTIVITY DETECTION

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

Acoustic-Prosodic Features of Tabla Bol Recitation and Correspondence with the Tabla Imitation	1229
<i>M. A. Rohit, Preeti Rao</i>	
Who Said That? A Comparative Study of Non-negative Matrix Factorization Techniques	1234
<i>Teun Krikke, Frank Broz, David Lane</i>	
AVA-Speech: A Densely Labeled Dataset of Speech Activity in Movies	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>	
Audiovisual Speech Activity Detection with Advanced Long Short-Term Memory	1244
<i>Fei Tao, Carlos Busso</i>	
Towards Automatic Speech Identification from Vocal Tract Shape Dynamics in Real-time MRI	1249
<i>Pramit Saha, Praneeth Srungarapu, Sidney Fels</i>	

SELECTED TOPICS IN NEURAL SPEECH PROCESSING

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

VOLUME 3

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

PERSPECTIVE TALK-1

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

DEREVERBERATION

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

AUDIO EVENTS AND ACOUSTIC SCENES

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

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

SPEAKER DIARIZATION

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

PHONATION

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

COGNITION AND BRAIN STUDIES

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

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

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

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

SHOW AND TELL 4

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

SPEECH AND SINGING PRODUCTION

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

ROBUST SPEECH RECOGNITION

The Fifth 'CHiME' Speech Separation and Recognition Challenge: Dataset, Task and Baselines	1561
<i>Jon Barker, Shinji Watanabe, Emmanuel Vincent, Jan Trmal</i>	
Voices Obscured in Complex Environmental Settings (VOICES) Corpus	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>	
Building State-of-the-art Distant Speech Recognition Using the CHiME-4 Challenge with a Setup of Speech Enhancement Baseline	1571
<i>Szu-Jui Chen, Aswin Shanmugam Subramanian, Hainan Xu, Shinji Watanabe</i>	
Unsupervised Adaptation with Interpretable Disentangled Representations for Distant Conversational Speech Recognition	1576
<i>Wei-Ning Hsu, Hao Tang, James Glass</i>	
Investigating Generative Adversarial Networks Based Speech Dereverberation for Robust Speech Recognition	1581
<i>Ke Wang, Junbo Zhang, Sining Sun, Yujun Wang, Fei Xiang, Lei Xie</i>	

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

APPLICATIONS IN EDUCATION AND LEARNING

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

INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS

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

SPEAKER CHARACTERIZATION AND ANALYSIS

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

PERSPECTIVE TALK-2

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

PLENARY TALK-2

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

NOVEL NEURAL NETWORK ARCHITECTURES FOR ACOUSTIC MODELLING

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

LANGUAGE IDENTIFICATION

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

PRODUCTION OF PROSODY

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

SPEECH INTELLIGIBILITY AND QUALITY

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

INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS

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

VOLUME 4

The PRIORI Emotion Dataset: Linking Mood to Emotion Detected In-the-Wild	1903
<i>Soheil Khorram, Mimansa Jaiswal, John Gideon, Melvin McInnis, Emily Mower Provost</i>	
Language Features for Automated Evaluation of Cognitive Behavior Psychotherapy Sessions	1908
<i>Nikolaos Flenotomos, Victor Martinez, James Gibson, David Atkins, Torrey Creed, Shrikanth Narayanan</i>	
Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks	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

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

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

SHOW AND TELL 5

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

VOICE CONVERSION AND SPEECH SYNTHESIS

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

EXTRACTING INFORMATION FROM AUDIO

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

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

SIGNAL ANALYSIS FOR THE NATURAL, BIOLOGICAL AND SOCIAL SCIENCES

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

SPEECH PROSODY

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

PERSPECTIVE TALK-3

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

RECURRENT NEURAL MODELS FOR ASR

ESPnet: End-to-End Speech Processing Toolkit	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>	
A GPU-based WFST Decoder with Exact Lattice Generation	2212
<i>Zhehuai Chen, Justin Luijens, Hainan Xu, Yiming Wang, Daniel Povey, Sanjeev Khudanpur</i>	
Automatic Speech Recognition System Development in the "Wild"	2217
<i>Anton Ragni, Mark Gales</i>	
Semantic Lattice Processing in Contextual Automatic Speech Recognition for Google Assistant	2222
<i>Leonid Velikovich, Ian Williams, Justin Scheiner, Petar Aleksic, Pedro Moreno, Michael Riley</i>	
Contextual Speech Recognition in End-to-end Neural Network Systems Using Beam Search	2227
<i>Ian Williams, Anjali Kannan, Petar Aleksic, David Rybach, Tara Sainath</i>	
Forward-Backward Attention Decoder	2232
<i>Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara</i>	

SPEAKER VERIFICATION USING NEURAL NETWORK METHODS I

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

SPEECH PERCEPTION IN ADVERSE CONDITIONS

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

MEASURING PITCH AND ARTICULATION

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

SPEECH AND LANGUAGE ANALYTICS FOR MENTAL HEALTH

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

SPOKEN CALL SHARED TASK, SECOND EDITION

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

SHOW AND TELL 6

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

ADJUSTING TO SPEAKER, ACCENT, AND DOMAIN

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

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

SPEECH SYNTHESIS PARADIGMS AND METHODS

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

VOLUME 5

SECOND LANGUAGE ACQUISITION AND CODE-SWITCHING

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

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

TOPICS IN SPEECH RECOGNITION

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

ZERO-RESOURCE SPEECH RECOGNITION

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

SPATIAL AND PHASE CUES FOR SOURCE SEPARATION AND SPEECH RECOGNITION

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

DIALECTAL VARIATION

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

SPOKEN CORPORA AND ANNOTATION

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

ZCU-NTIS Speaker Diarization System for the DIHARD 2018 Challenge	2788
<i>Zbynek Zajíc, Marie Kunešová, Jan Zelinka, Marek Hruš</i>	
Speaker Diarization with Enhancing Speech for the First DIHARD Challenge	2793
<i>Lei Sun, Jun Du, Chao Jiang, Xueyang Zhang, Shan He, Bing Yin, Chin-Hui Lee</i>	
BUT System for DIHARD Speech Diarization Challenge 2018	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>	
Estimation of the Number of Speakers with Variational Bayesian PLDA in the DIHARD Diarization Challenge	2803
<i>Ignacio Viñals, Pablo Gimeno, Alfonso Ortega, Antonio Miguel, Eduardo Lleida</i>	
Diarization is Hard: Some Experiences and Lessons Learned for the JHU Team in the Inaugural DIHARD Challenge	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>	
The EURECOM Submission to the First DIHARD Challenge	2813
<i>Jose Patino, Héctor Delgado, Nicholas Evans</i>	

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

TEXT ANALYSIS, MULTILINGUAL ISSUES AND EVALUATION IN SPEECH SYNTHESIS

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

NEURAL NETWORK TRAINING STRATEGIES FOR ASR

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

APPLICATION OF ASR IN MEDICAL PRACTICE

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

Perceptual and Automatic Evaluations of the Intelligibility of Speech Degraded by Noise Induced Hearing Loss Simulation	2953
<i>Imed Laaridh, Julien Tardieu, Cynthia Magnen, Pascal Gaillard, Jérôme Farinas, Julien Pinquier</i>	
Articulatory Features for ASR of Pathological Speech	2958
<i>Emre Yilmaz, Vikramjit Mitra, Chris Bartels, Horacio Franco</i>	
Mining Multimodal Repositories for Speech Affecting Diseases	2963
<i>Joana Correia, Bhiksha Raj, Isabel Trancoso, Francisco Teixeira</i>	
Long Distance Voice Channel Diagnosis Using Deep Neural Networks	2968
<i>Zhen Qin, Tom Ko, Guangjian Tian</i>	
Speech Recognition for Medical Conversations	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

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

PLENARY TALK-3

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

DISTANT ASR

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

EXPRESSIVE SPEECH SYNTHESIS

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

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

REPRESENTATION LEARNING FOR EMOTION

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

ARTICULATORY INFORMATION, MODELING AND INVERSION

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

NOVEL PARADIGMS FOR DIRECT SYNTHESIS BASED ON SPEECH-RELATED BIOSIGNALS

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

VOLUME 6

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

LOW RESOURCE SPEECH RECOGNITION CHALLENGE FOR INDIAN LANGUAGES

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

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

SHOW AND TELL 7

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

DEEP ENHANCEMENT

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

ACOUSTIC SCENES AND RARE EVENTS

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

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

LANGUAGE MODELING

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

SPEECH PATHOLOGY, DEPRESSION, AND MEDICAL APPLICATIONS

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

PERSPECTIVE TALK-4

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

SPOKEN LANGUAGE UNDERSTANDING

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

SOURCE SEPARATION FROM MONAURAL INPUT

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

MULTIMODAL SYSTEMS

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

CODING

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

SPEAKER VERIFICATION USING NEURAL NETWORK METHODS II

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

EMOTION RECOGNITION AND ANALYSIS

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

ACOUSTIC MODELLING

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

SPEECH AND SPEAKER PERCEPTION

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