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

# TABLE OF CONTENTS

## VOLUME 1

### ISCA MEDAL TALK

- *From Vocoder to Code-Excited Linear Prediction: Learning How We Hear What We Hear* ................................................................. 1  
  *Bishnu S. Atal*

### END-TO-END SPEECH RECOGNITION

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

### PROSODY MODELING AND GENERATION

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

### SPEAKER VERIFICATION I

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

### SPOKEN TERM DETECTION

- *CNN Based Query by Example Spoken Term Detection* ..................................................................................................................... 92  
  *Dhimanjyot Ram, Lesly Miculschik, Hervé Bourlard*
- *Learning Acoustic Word Embeddings with Temporal Context for Query-by-Example Speech Search* ........................................... 97  
  *Yougen Yuan, Cheung-Chi Leung, Lei Xie, Hongjie Chen, Bin Ma, Haizhou Li*
Siamese Recurrent Auto-Encoder Representation for Query-by-Example Spoken Term Detection ................................................................. 102
Zexin Zhu, Zhijie Wu, Runnan Li, Helen Meng, Limhong Cai
Wei Li, Brian Mak
LSTM Based Attentive Fusion of Spectral and Prosodic Information for Keyword Spotting in Hindi Language ...................................... 112
Laxmi Pandey, Karan Nathwani
Spoken Keyword Detection Using Joint DTW-CNN ........................................... 117
Ravi Shankar, Vikram C M, S R Mshadewa Prasanna

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
Björn Schuller, Stefan Steidl, Anton Bahtner, Peter B. Marschik, Harald Baumeister, Fengguan Dong, Simone Hanke, Florian B. Pokorny, Eva-Maria Rahner, Katrin D. Barrie-Pokorny, Christa Einspieler, Dajie Zhang, Alice Baird, Shahin Amiriparian, Kun Qian, Zhao Ren, Maximilian Schmitt, Panagiotis Tzirakis, Stefanos Zafeiriou
An Ensemble of Transfer, Semi-supervised and Supervised Learning Methods for Pathological Heart Sound Classification ........................................................................................................ 127
Anam Ali Tugtek Tunan, Engin Erzin
Monitoring Infant’s Emotional Cry in Domestic Environments Using the Capsule Network Architecture ...................................................... 132
Neural Network Architecture That Combines Temporal and Summative Features for Infant Cry Classification in the Interspeech 2018 Computational Paralinguistics Challenge ........................................... 137
Mark Huckvale
Evolving Learning for Analysing Mood-Related Infant Vocalisation ................................................................................................................. 142
Zheng Zhang, Jing Han, Kun Qian, Björn Schuller
Deep Learning in Paralinguistic Recognition Tasks: Are Hand-crafted Features Still Relevant? ................................................................. 147
Johannes Wagner, Dominik Schiller, Andreas Seiderer, Elisabeth André
Investigation on Joint Representation Learning for Robust Feature Extraction in Speech Emotion Recognition ........................................ 152
Danqing Luo, Yuexian Zou, Dongyan Huang
Using Voice Quality Supervectors for Affect Identification ............................................................................................................................. 157
Soo Jin Park, Amber Afsan, Zhi Ming Chua, Abeer Alwan
An End-to-End Deep Learning Framework for Speech Emotion Recognition of Atypical Individuals .............................................................. 162
Dengke Tang, Junlin Zeng, Ming Li

SHOW AND TELL 1
DialogOS: Simple and Extensible Dialogue Modeling .................................................. 167
Alexander Faller, Timo Baumann, Arne Köhn
A Framework for Speech Recognition Benchmarking ........................................... 169
Franck Dernoncourt, Trung Bai, Walter Chang
Flexible Tongue Housed in a Static Model of the Vocal Tract With Jaws, Lips and Teeth ........................................................... 171
Takayuki Arata
Voice Analysis Using Acoustic and Throat Microphones for Speech Therapy ................................................................................................. 173
Lani Mathew, K. Gopakumar
A Robust Context-Dependent Speech-to-Speech Phraselator Toolkit for Alexa ................................................................. 175
Manny Rayner, Nikos Tzourakis, Jan Stanek

SPEECH SEGMENTS AND VOICE QUALITY
Discriminating Nasals and Approximants in English Language Using Zero Time Windowing ............................................................ 177
Ravishankar Prasad, Sudarsana Reddy Kadiri, Suryakant V Gangashetty, Bayya Yegnanarayana
Gestural Lenition of Rhotics Captures Variation in Brazilian Portuguese ................................................................................................. 182
Phil Howson, Alexei Kochetov
Identification and Classification of Fricatives in Speech Using Zero Time Windowing Method ................................................................. 187
Ravishankar Prasad, Bayya Yegnanarayana
GlobalTIMIT: Acoustic-Phonetic Datasets for the World’s Languages .......................................................... 192
Nattanun Chanchana, Stephen Cieri, Japhet Debrah, Hongwei Ding, Yue Jiang, Sishi Liu, Mark Liberman, Jonathan Wright, Jiahong Yuan, Juhong Zhan, Yuqing Zhan
Structural Effects on Properties of Consonantal Gestures in Tashlhiyt .......................................................... 197
Anne Hermes, Doris Mücke, Bastian Auris, Rachel Bideau
The Retroflex-dental Contrast in Punjabi Stops and Nasals: A Principal Component Analysis of Ultrasound Images ................................................................................... 202
Alexei Kochetov, Matthew Faytak, Kiranpreet Nara
Vowels and Diphthongs in Hangzhou Wu Chinese Dialect .................................................. 207
Yang Yue, Fang Hu
Resyllabification in Indian Languages and Its Implications in Text-to-speech Systems .......................................................... 212
Mukesh M, Jeena J Prakash, Hema Murthy

Voice Source Contribution to Prominence Perception: Rd Implementation .............................................................. 217
Andy Murphy, Irena Tanaseshkaya, Ailbhe Ni Chasaide, Christe Gobł

On the Relationship between Glottal Pulse Shape and Its Spectrum: Correlations of Open Quotient, Pulse Skew
and Peak Flow with Source Harmonic Amplitudes ........................................................................................................... 222
Christe Gobł, Andy Murphy, Irena Tanaseshkaya, Ailbhe Ni Chasaide

Comparison System ........................................................................................................................................... 227

Breathy to Tense Voice Discrimination using Zero-Time Windowing Cepstral Coefficients (ZTWCCs) ............................................... 232
Sudarsana Reddy Kadiri, Bayya Yegnanarayana

Analysis of Breathiness in Contextual Vowel of Voiceless Nasals in Mizo .................................................................................. 237
Pamir Gogoi, Sithir Kalita, Parsimita Gogoi, Ratnveer Wayland, Priyankoo Sarmah, SR Mahadeva Prusanna

SPEAKER STATE AND TRAIT

Infant Emotional Outbursts Detection in Infant-parent Spoken Interactions ........................................................................... 242
Yiuya Xu, Mark Hasegawa-Johnson, Nancy McElwain

Deep Neural Networks for Emotion Recognition Combining Audio and Transcripts ................................................................. 247
Huang Cho, Bhagavadendra Pappagari, Parva Kulkarni, Jesu Villalba, Yoshik Carnele, Najim Dehak

Preference-Learning with Qualitative Agreement for Sentence Level Emotional Annotations ......................................................... 252
Sriniwas Parthasarathy, Carlos Busso

Transfer Learning for Improving Speech Emotion Classification Accuracy ............................................................................. 257
Sadique Latiff, Rajib Rana, Shazad Younis, Junaid Qadir, Juliet Epps

What Do Classifiers Actually Learn? a Case Study on Emotion Recognition Datasets ................................................................. 262
Patrick Meyer, Eric Buschermohle, Tim Fingscheidt

State of Mind: Classification through Self-reported Affect and Word Use in Speech ................................................................. 267
Evamaría Raffiner, Yannik Terhorst, Nicholas Cunnings, Björn Schuller, Harald Baumeister

Exploring Spatio-Temporal Representations by Integrating Attention-based Bidirectional-LSTM-RNNs and FCNs
for Speech Emotion Recognition ........................................................................................................................................... 272
Ziqing Zhao, Yu Zheng, Zixing Zhang, Haitshuai Wang, Yuqin Zhao, Chao Li

End-to-end Deep Neural Network Age Estimation .................................................................................................................. 277
Pegah Ghatrehmani, Phani Sankar Yadavavolu, Nansin Chen, Jesu Villalba, Daniel Posey, Sanjeev Khudanpur, Najim Dehak

Improving Gender Identification in Movie Audio Using Cross-Domain Data ............................................................................. 282
Bajat Hебbar, Krishna Somandepalli, Shrikant Narayanan

On Learning to Identify Genders from Raw Speech Signal Using CNNs ..................................................................................... 287
Selen Hande Kabili, Hannah Muckenhirn, Mathew Magimai-Doss

Denoising and Raw-waveform Networks for Weakly-Supervised Gender Identification on Noisy Speech ........................................... 292
Jill Sebastian, Manoj Kumar, Pavan Kumar D. S., Mathew Magimai-Doss, Hema Murthy, Shrikant Narayanan

The Effect of Exposure to High Altitude and Heat on Speech Articulatory Coordination ..................................................................... 297
James Williamson, Thomas Quatieri, Adam Lammert, Katherine Mitchell, Katherine Finkelstein, Nicole Ekou, Caitlin Dillon, Robert Kenefick, Kristin Heaton

DEEP LEARNING FOR SOURCE SEPARATION AND PITCH TRACKING

Permutation Invariant Training of Generative Adversarial Network for Monaural Speech Separation .............................................. 302
Lianwu Chen, Meng Yu, Yanmin Qian, Dan Su, Dong Yu

Deep Extractor Network for Target Speaker Recovery from Single Channel Speech Mixtures ......................................................... 307
Jun Wang, Jie Chen, Dan Su, Lianwu Chen, Meng Yu, Yanmin Qian, Dong Yu

Joint Localization and Classification of Multiple Sound Sources Using a Multi-task Neural Network .................................................. 312
Weipeng He, Petur Mósttvek, Jean-Marc Odobez

Detection of Glottal Closure Instants from Speech Signals: A Convolutional Neural Network Based Method ....................................................... 317
Shuai Yang, Zhiyong Wu, Binbin Shen, Helen Meng

Robust TDOA Estimation Based on Time-Frequency Masking and Deep Neural Networks .......................................................... 322
Zhong-Qiu Wang, Xueliang Zhang, Deliang Wang

Waveform to Single Sinusoid Regression to Estimate the F0 Contour from Noisy Speech Using Recurrent Deep
Neural Networks ........................................................................................................................................................................ 327
Atsuhito Kato, Tomi Kinnunen

Reducing Interference with Phase Recovery in DNN-based Monaural Singing Voice Separation .............................................................. 332
Paul Magron, Konstantinos Drossos, Stylianos Ioannis Mimilakis, Tuomas Virtanen

Nebula: F0 Estimation and Voicing Detection by Modeling the Statistical Properties of Feature Extractors ...................................................... 337
Kanru Hua

Real-time Single-channel Dereverberation and Separation with Time-domain Audio Separation ......................................................... 342
Yi Luo, Nima Moghaddam

Music Source Activity Detection and Separation Using Deep Attractor Network ........................................................................... 347
Rajath Kumar, Yi Luo, Nima Moghaddam
Improving Mandarin Tone Recognition Using Convolutional Bidirectional Long Short-Term Memory with Attention ............................................................. 352
Longfei Yang, Yuanta Xie, JinSong Zhang

ACOUSTIC ANALYSIS-SYNTHESIS OF SPEECH DISORDERS

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

ASR SYSTEMS AND TECHNOLOGIES

Cold Fusion: Training Seq2Seq Models Together with Language Models .................................................. 387
Amanose Siriram, HeeSoo Jun, Sanjeev Saithesh, Adam Coates
Investigation on Estimation of Sentence Probability by Combining Forward, Backward and Bi-directional LSTM-RNNs .......................................... 392
Kazuki Irie, Zhikong Lei, Lihua Deng, Ralf Schlüter, Hermann Ney
Subword and Crossword Units for CTC Acoustic Models ............................................................. 396
Thomas Zenkel, Ramon Sanabría, Florian Mette, Alex Waibel
Neural Error Corrective Language Models for Automatic Speech Recognition ...................................... 401
Tomohiro Tanaka, Ryo Masumura, Hirokazu Masatuki, Toshiyuki Aono
Entity-Aware Language Model as an Unsupervised Reranker ............................................................ 406
Mohammad Salegh Rasooli, Sarangarajan Parthasarathy
Character-level Language Modeling with Gated Hierarchical Recurrent Neural Networks .................. 411
Iksoo Choi, Jinhwan Park, Wonnyong Sung

DECEPTION, PERSONALITY, AND CULTURE ATTRIBUTE

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

AUTOMATIC DETECTION AND RECOGNITION OF VOICE AND SPEECH DISORDERS

A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks ........................................ 446
Huayi Wu, John Soraghan, Anja Lovit, Gaetano Di-Caterina
Dysarthric Speech Recognition Using Time-delay Neural Network Based Denoising Autoencoder .................................................. 451
Chitralekha Bhat, Biswajit Das, Bhaskar Vachhani, Sunil Kumar Kopparapu
A Multitask Learning Approach to Assess the Dysarthria Severity in Patients with Parkinson's Disease ................................................ 456
Juan Camilo Vásquez Correa, Tomas Arias, Juan Rafael Orozco-Arroyave, Elnmar Nöth
The Use of Machine Learning and Phonetic Endophenotypes to Discover Genetic Variants Associated with Speech Sound Disorder ................................................................. 461
Jason Lilley, Erin Crowgey, H Timothy Bunnell
Whistle-blowing ASRs: Evaluating the Need for More Inclusive Speech Recognition Systems ........ 466
Meredith Moore, Hemanth Venkateswara, Sethuraman Panchanathan
Data Augmentation Using Healthy Speech for Dysarthric Speech Recognition ........................................................................................................... 471
Bhavik Vaichhani, Chitrakalekha Bhat, Sunil Kumar Koppurapu

VOICE CONVERSION

Improving Sparse Representations in Exemplar-Based Voice Conversion with a Phoneme-Selective Objective

Shaojin Ding, Gianlong Zhao, Christopher Liberatore, Ricardo Gutierrez-Osuna ......................................................... 476

Learning Structured Dictionaries for Exemplar-based Voice Conversion

Shaojin Ding, Christopher Liberatore, Ricardo Gutierrez-Osuna .................................................................................................................. 481

Exemplar-Based Spectral Detail Compensation for Voice Conversion

Yu-Huai Peng, Hsin-Fe Huang, Tichao Wu, Yu Tao, Hsin-Min Wang ......................................................................................... 486

Whispered Speech to Neutral Speech Conversion Using Bidirectional LSTMs

Yu-Huai Peng, Hsin-Fe Huang, Tichao Wu, Yu Tao, Hsin-Min Wang ......................................................................................... 491

Multitarget Voice Conversion without Parallel Data by Adversarially Learning Disentangled Audio Representations

Ju-Chieh Chou, Cheng-Chieh Teh, Hung-Yi Lee, Lin-Shan Lee ......................................................................................... 501

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

Attention-based Sequence Classification for Affect Detection ........................................................................................................... 506

Sankar Mukherjee, Thierry Legou, Leonardo Lancia, Pauline Hilt, Alice Tomassini, Luciano Fadiga, Alessandro D’Ausilio, Joo-Kyung Kim, Young-Bum Kim

Computational Paralinguistics: Automatic Assessment of Emotions, Mood and Behavioural State from Acoustics of Speech

Zafi Sherban Syed, Julien Schroeter, Kirill Sidorov, David Marshall ........................................................................................................... 511

Investigating Utterance Level Representations for Detecting Intent from Acoustics

Saikrishna Rallabandi, Bhavya Karki, Carla Viegas, Eric Nyberg, Alan W Black ........................................................................................................... 516

LSTM Based Cross-corpus and Cross-task Acoustic Emotion Recognition

Bogdan Vlaseko, Jili Sebastian, D. S. Pavun Kumar, Mathes Magmain-Doss ......................................................................................... 521

Implementing Fusion Techniques for the Classification of Paralinguistic Information

Gábor Gosztolya, Tamás Grósz, László Tóth ..................................................................................................................... 526

General Utterance-Level Feature Extraction for Classifying Crying Sounds, Atypical & Self-Assessed Affect and Heart Beats

Gabor Gosztolya, Tamás Grósz, László Tóth ..................................................................................................................... 531

Self-Assessed Affect Recognition Using Fusion of Attentional BLSTM and Static Acoustic Features

Heysem Kaya, Dmitiri Fedotov, Ali Yasilkanan, Oana Verkholysak, Yang Zhang, Jing-Lin Li, Chi-Chun Lee ......................................................................................... 536

Vocalic, Lexical and Prosodic Cues for the INTERSPEECH 2018 Self-Assessed Affect Challenge

Claude Montacié, Marie-José Caraty ..................................................................................................................... 541

SHOW AND TELL 2

Intonation tutor by SPIRE (In-SPIRE): An Online Tool for an Automatic Feedback to the Second Language

P. A. Anand, Chiranjeevi Yarra, N. K. Kausthubha, Prasanta Kumar Ghosh ......................................................................................... 546

Game-based Spoken Dialog Language Learning Applications for Young Students


The IBM Virtual Voice Creator

Alexander Sorin, Slava Shechtman, Zvi Kons, Ron Hooray, Shay Ben-David, Joe Pavitt, Shai Rezenberg, Carmel Rabinovitz, Tal Drory ......................................................................................... 554

Mobile Application for Learning Languages for the Unlettered

G. Gayathri, N. Mohana, Radhika Pal, Hema Murthy ..................................................................................................................... 552

Mandarin-English Code-switching Speech Recognition

Haihua Xu, Van Tung Pham, Zin Tun Kyaw, Zhu Hao Lin, Eng Siong Chng, Haizhou Li ......................................................................................... 558

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

Joo-Kyung Kim, Young-Bum Kim ..................................................................................................................... 556

Analyzing Vocal Tract Movements During Speech Accommodation

Sankar Mukherjee, Thierry Legou, Leonardo Lancia, Pauline Hilt, Alice Tomassini, Luciano Fadiga, Alessandro D’Ausilio, Leonardo Badinos, Noël Nguyen ......................................................................................... 561

Cross-Lingual Multi-Task Neural Architecture for Spoken Language Understanding

Yajiang Li, Xuemin Zhao, Weiquan Xu, Yonghong Yan ......................................................................................... 566
Singing Voice Phoneme Segmentation by Hierarchically Inferring Syllable and Phoneme Onset Positions ................................................................. 716
Bong Gong, Xavier Sorra

Novel Empirical Mode Decomposition Cepstral Features for Replay Spoof Detection ................................................................. 721
Prasad Taphiri, Hemant Patil

Novel Linear Frequency Residual Cepstral Features for Replay Attack Detection ................................................................. 726
Hemlata Tak, Hemant Patil

Analysis of Sparse Representation Based Feature on Speech Mode Classification ................................................................. 731
Kamal Tripathi, K. Sreenivasa Rao

Multicomponent 2-D AM-FM Modeling of Speech Spectrograms ................................................................................. 736
Enea Ceolini, Jithendar Anumula, Adrian Huber, Ilya Kiselev, Shih-Chii Liu

An Optimization Framework for Recovery of Speech from Phase-Encoded Spectrograms .......................................................... 741
Abhilash Sainathan, Sunil Radhesh, Chandra Sekhar Seelamantula

Speaker Recognition with Nonlinear Distortion: Clipping Analysis and Impact ................................................................. 746
Wei Xia, John H. L. Hansen

Linear Prediction Residual based Short-term Cepstral Features for Replay Attacks Detection ................................................................. 751
Madhu Sudan Singh, Debadatta Pati

Analysis of Variational Mode Functions for Robust Detection of Vowels ............................................................................. 756
Surbhi Sakshi, Avinash Kumar, Gayadhar Pradhan

SEQUENCE MODELS FOR ASR

Improving Attention Based Sequence-to-Sequence Models for End-to-End English Conversational Speech Recognition ................................................................. 761
Chao Weng, Jia Cui, Guansen Wang, Jun Wang, Chengzhu Yu, Dan Su, Dong Yu

Segmental Encoder-Decoder Models for Large Vocabulary Automatic Speech Recognition ................................................................. 766
Eugen Beck, Mirko Hannemann, Patrick Dittrich, Ralf Schütte, Hermann Ney

Acoustic Modeling with DFSMN-CTC and Joint CTC-CE Learning .................................................................................. 771
Shilasang Zhang, Ming Lei

End-to-End Speech Command Recognition with Capsule Network ................................................................................ 776
Jaesung Bae, Dae-Shik Kim

End-to-End Speech Recognition from the Raw Waveform ................................................................................ 781
Neil Zeghidour, Nicolas Usunier, Gabriel Synnaeve, Ronan Collobert, Emmanuel Dupoux

A Multistage Training Framework for Acoustic-to-Word Model ................................................................................ 786
Chengzhu Yu, Chantei Zhang, Chao Weng, Jia Cui, Dong Yu

Syllable-Based Sequence-to-Sequence Speech Recognition with the Transformer in Mandarin Chinese ............................................................................. 791
Shiyu Zhou, Linhao Dong, Shuang Xu, Bo Xu

Densely Connected Networks for Conversational Speech Recognition ................................................................................ 796
Kyu Han, Akshay Chandrareshkar, Junguk Kim, Ian Lane

Multi-Head Decoder for End-to-End Speech Recognition ............................................................................. 801
Tomoki Hayashi, Shinji Watanabe, Tomoki Toda, Kazuya Takeda

Compressing End-to-end ASR Networks by Tensor-Train Decomposition ............................................................................. 806
Takuma Mori, Andros Tjandra, Sakriani Sakti, Satoshi Nakamura

Speech2Vec: A Sequence-to-Sequence Framework for Learning Word Embeddings from Speech ............................................................................... 811
Yu-An Chung, James Glass

Extending Recurrent Neural Aligner for Streaming End-to-End Speech Recognition in Mandarin ............................................................................. 816
Linhao Dong, Shiyu Zhou, Wei Chen, Bo Xu

SOURCE SEPARATION AND SPATIAL ANALYSIS

Joint Noise and Reverberation Adaptive Learning for Robust Speaker DOA Estimation with an Acoustic Vector Sensor ............................................................................. 821
Dong Wang, Yuexian Zou

Multiple Concurrent Sound Source Tracking Based on Observation-Guided Adaptive Particle Filter ............................................................................. 826
Hong Liu, Haipeng Lan, Bing Yang, Cheng Pang

Harmonic-Percussive Source Separation of Polyphonic Music by Suppressing Impulsive Noise Events ............................................................................. 831
M. Gurunath Reddy, K. Sreenivasa Rao, Partha Pratim Das

Speaker Activity Detection and Minimum Variance Beamforming for Source Separation ............................................................................. 836
Ennea Ceolini, Jithendar Anumula, Adrian Huber, Ilya Kiselev, Shih-Chii Liu

Sparsity-Constrained Weight Mapping for Head-Related Transfer Functions Individualization from Anthropicometric Features ............................................................................. 841
Xiaohe Qi, Jianhua Tao

Speech Source Separation Using ICA in Constant Q Transform Domain ................................................................................ 846
D. Y. L. N. Dheeraj Sai, K. S. Kishor, Kodukula Sri Ramamurthy

Multi-talker Speech Separation Based on Permutation Invariant Training and Beamforming ................................................................................ 851
Lu Yin, Ziqing Wang, Risheng Xia, Junfeng Li, Yonghong Yan

Expectation-Maximization Algorithms for Iwakura-Saito Nonnegative Matrix Factorization ................................................................................ 856
Paul Magron, Taomas Virtanen
Subband Weighting for Binaural Speech Source Localization ................................................................. 861
Karthik Girija Ramesan, Parth Suresh, Prasanta Kumar Ghosh

PLENARY TALK-1

Universal Tendencies for Cross-Linguistic Prosodic Tendencies: A Review and Some New Proposals .................. 866
Jacqueline Vaissière

ACOUSTIC MODEL ADAPTATION

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

STATISTICAL PARAMETRIC SPEECH SYNTHESIS

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

EMOTION MODELING

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

MODELS OF SPEECH PERCEPTION

Picture Naming or Word Reading: Does the Modality Affect Speech Motor Adaptation and Its Transfer? .................. 956
Tiphaine Caudrelier, Pascal Perrier, Jean-Luc Schwartz, Amélie Rochet-Capellan
Measuring the Band Importance Function for Mandarin Chinese with a Bayesian Adaptive Procedure .................. 961
Yufan Du, Yi Shen, Hongying Yang, Xiaohong Wu, Jing Chen
Wide Learning for Auditory Comprehension .................................................................................................... 966
Elnaz Shahaei-Bajestan, R. Harald Baayen
Analyzing Reaction Time Sequences from Human Participants in Auditory Experiments ......................................................... 971

Prediction of Perceived Speech Quality Using Deep Machine Listening ................................................................. 976

Prediction of Subjective Listening Effort from Acoustic Data with Non-Intrusive Deep Models ........................................... 981

Ziqiang Shi, Huibin Lin, Liu Liu, Rujie Liu

Fahimeh Bahmaninezhad, John H. L. Hansen

Longting Xu, Kong Aik Lee, Haizhou Li, Zhen Yang

Moez Ajili, Jean-François Bonastre, Solange Rossato

Annam Naresh, Rushabh Gandhi, Mallikarjuna Rao Bellamkonda, Mithun Das Gupta

Aku Rouhe, Reima Karhila, Atja Elg, Minna Leena Toivola, Peter Smit, Antta-Riikka Smolander, Mikko Kurimo

Asami Murata, Kiyoharu Kuroki, Yasuhiro Miyahara, Takanori Sato, Kazuhiko Honda, Tatsuki Nishimura

Jasper Ooster, Rainer Huber, Bernd T. Meyer

Paul Kranzusch, Rainer Huber, Melanie Krüger, Birger Kollmeier, Bernd T. Meyer

MULTIMODAL DIALOGUE SYSTEMS

A Case Study on the Importance of Belief State Representation for Dialogue Policy Management ........................................ 986

Margareta Kotti, Fassilios Diakoulakas, Alexandros Papangelis, Michail Lagoudakis, Yannis Stylianou

Prediction of Turn-taking Using Multitask Learning with Prediction of Backchannels and Fillers ........................................ 991

Kohei Haro, Kei Inoue, Katayama Takehisa, Tatsuya Kawanaka

Conversational Analysis Using Utterance-level Attention-based Bidirectional Recurrent Neural Networks ..................... 996

Chandrakant Bohle, Sven Magg, Cornelius Weber, Stefan Wermter

A Comparative Study of Statistical Conversion of Face to Voice Based on Their Subjective Impressions .......................... 1001

Yasuhiko Ohtsu, Daiya Saito, Nobuaki Mimematsu

Follow-up Question Generation Using Pattern-based Seq2seq with a Small Corpus for Interview Coaching ..................... 1006

Ming-Hsiang Su, Chang-Hsien Wu, Kun-Yi Huang, Qun-Bei Hong, Hua-Hung Huang

Coherence Models for Dialogue ................................................................................................................................ 1011

Alessandra Corvone, Evgeny Stepanov, Giuseppe Riccardi

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

K. E. Manjunath, K. Sreenivasa Rao, Dinesh Babu Jayagopi, V Ramasubramanian

Rapid Collection of Spontaneous Speech Corpora Using Telephonic Community Forums ................................................. 1021

Agha Ali Raza, Avvais Athar, Shan Randhawa, Zain Tariq, Muhammad Bilal Saleem, Haris Bin Zia, Umar Saif, Roni Rosenfeld

Effect of TTS Generated Audio on OOV Detection and Word Error Rate in ASR for Low-resource Languages .................... 1026

Savitha Murthy, Dinkar Sitaram, Sanayana Sitaram

Development of Large Vocabulary Speech Recognition System with Keyword Search for Manipuri ...................................... 1031

Tanvina Patel, D. N. Krishna, Noor Fathima, Nihar Shah, C. Mahima, Deepak Kumar, Anurupoy Ivengar

Robust Mizo Continuous Speech Recognition ................................................................................................................. 1036

Abhishek De, Biwejati Dev Sarma, Wendy Laltinglingthui, Lalnuniasami Ngente, Parismitei Gogoi, Priyankoo Sarmah, S R Mahadeva Prasanna, Rohit Sinha, S. R. Nirmala

Semi-supervised and Active-learning Scenarios: Efficient Acoustic Model Refinement for a Low Resource Indian Language .................................................................................................................................................. 1041

Maharaj Prasad Chellappriyadarshini, Anoop Raghavan, V Ramasubramanian

Automatic Speech Recognition with Articulatory Information and a Unified Dictionary for Hindi, Marathi, Bengali and Oriya ........................................................................................................................................... 1046

Debadatta Dash, Myungjung Kim, Kristin Teplantsky, Jun Wang

SHOW AND TELL 3

Captainai: Integrated Pronunciation Practice and Data Collection Portal .................................................................................. 1051

Aku Rouhe, Reima Karhila, Atja Elg, Minna Leena Toivola, Peter Smit, Antta-Riikka Smolander, Mikko Kurimo

auMina™ - Enterprise Speech Analytics ................................................................................................................................. 1053

Umesh Sachdev, RajagopalJayaraman, ZainabMilwala

Holo Companion: An MR Friend for EveryOne .......................................................................................................................... 1055

Annam Nareesh, Rushabh Gandhi, Mallikarijunarao Bellamkonda, MithunDasGupta

akeira™- Virtual Assistant .......................................................................................................................................................... 1057

Umesh Sachdev, RajagopalJayaraman, ZainabMilwala

Brain-Computer Interface using Electroencephalogram Signatures of Eye Blinks ............................................................... 1059

SrihariMaruthachalam, SathishAggarwal, MariGaneshKumar, MriganhaSar, HemaMurthy

SPEAKER VERIFICATION II

Voice Comparison and Rhythm: Behavioral Differences between Target and Non-target Comparisons .............................. 1061

Moez Ajili, Jean-François Bonastre, Solange Rossato

Co-whitening of I-vectors for Short and Long Duration Speaker Verification ........................................................................ 1066

Longting Xu, Kong Aik Lee, Haizhou Li, Zhen Yang

Compensation for Domain Mismatch in Text-independent Speaker Recognition ................................................................. 1071

FahimehBahmaninezhad, John H.L. Hansen

Joint Learning of J-Vector Extractor and Joint Bayesian Model for Text Dependent Speaker Verification ............................ 1076

Ziqiang Shi, Liu Liu, Haibin Lin, Rajie Liu

Latent Factor Analysis of Deep Bottleneck Features for Speaker Verification with Random Digit Strings .......................... 1081

Ziqiang Shi, Haibin Lin, Liu Liu, Rajie Liu
NOVEL APPROACHES TO ENHANCEMENT

A New Framework for Supervised Speech Enhancement in the Time Domain ........................................ 1136
Ashutosh Pandey, Deliang Wang

Speech Enhancement Using the Minimum-probability-of-error Criterion .................................................. 1141
Jishnu Sadasivan, Subhadip Mukherjee, Chandra Sekhar Seelamantula

Exploring the Relationship between Conic Affinity of NMF Dictionaries and Speech Enhancement Metrics ................................................................. 1146
Pavlos Papadopoulos, Colin Fay, Shrkanath Narayanan

Using Shifted Real Spectrum Mask as Training Target for Supervised Speech Separation ............................. 1151
Yun Liu, Hui Zhang, Xueliang Zhang

Enhancement of Noisy Speech Signal by Non-Local Means Estimation of Variational Mode Functions .......... 1156
Nagapuri Srinivas, Gaydhara Pradhan, Syed Shahnawazuddin

Phase-locked Loop (PLL) Based Phase Estimation in Single Channel Speech Enhancement .......................... 1161
Priya Pallavi, Ch Y Rama Rao

Cycle-Consistent Speech Enhancement ........................................................................................... 1165
Zhong Meng, Junyu Li, Yijan Gong, Bing-Huang (Fred) Juang

Visual Speech Enhancement ........................................................................................................ 1170
Ariv-Gabby, Asaph Shamir, Shmuel Felg

Implementation of Digital Hearing Aid as a Smartphone Application ..................................................... 1175
Saketh Sharma, Nitya Tiwari, Prem C. Pandey

Bone-Conduction Sensor Assisted Noise Estimation for Improved Speech Enhancement ............................ 1180
Ching-Hua Lee, Bhaskar D. Rao, Harinath Garudadri

Artificial Bandwidth Extension with Memory Inclusion Using Semi-supervised Stacked Auto-encoders ........ 1185
Pranam Bachhav, Massimiliano Todisco, Nicholas Evans

Large Vocabulary Concatenative Resynthesis ........................................................................................ 1190
Soumi Maiti, Jocy Ching, Michael Mandel

Concatenative Resynthesis with Improved Training Signals for Speech Enhancement ............................... 1195
Ali Raza Syed, Viet Anh Trinh, Michael Mandel

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
Ooko Râsâmen, Seshadri Shreyas, Marisa Casillas

A Comparison of Input Types to a Deep Neural Network-based Forced Aligner ........................................ 1205
Matthew C. Kelley, Benjamin Y. Tucker

Joint Learning Using Denoising Variational Autoencoders for Voice Activity Detection ........................... 1210
Youngmoon Jung, Younggwon Kim, Yeunju Choi, Heorin Kim

Information Bottleneck Based Percussion Instrument Diarization System for Taniavartnam Segments of Carnatic Music Concerts ................................................................................ 1215
Naoman Dawalelalab, Jom Kuriakose, Chandra Sekhar Chella, Hema Murthy

Robust Voice Activity Detection Using Frequency Domain Long-Term Differential Entropy .................... 1220
Debjyant Ghosh, R. Murulishankar, Sanjeev Gurugopinath

Device-directed Utterance Detection ................................................................................................ 1225
Sri Harsh Mallidi, Roland Maas, Kyle Goshner, Ariya Bastrow, Spyros Matsoukas, Björn Hoffmeister
Acoustic-Prosodic Features of Tabla Bol Recitation and Correspondence with the Tabla Imitation ................................................................. 1229

Who Said That? A Comparative Study of Non-negative Matrix Factorization Techniques .......................................................... 1234

Teun Krikke, Frank Gra, David Lane

AVSpeech: A Densely Labeled Dataset of Speech Activity in Movies ........................................................................................................... 1239

Saurish Chaudhuri, Joseph Roth, Daniel P. W. Ellis, Andrew Gallagher, Liat Kaver, Radhika Marwah, Caroline Pantofaru, Nathan Reale, Loretta Guarino Reid, Kevin Wilson, Zhonghua Xi

Audiovisual Speech Activity Detection with Advanced Long Short-Term Memory ........................................................................................................... 1244

Fei Tao, Carlos Busso

Towards Automatic Speech Identification from Vocal Tract Shape Dynamics in Real-time MRI ................................................................. 1249

Pramit Saha, Praneeth Srngarapu, Sidney Fels

SELECTED TOPICS IN NEURAL SPEECH PROCESSING

Structured Word Embedding for Low Memory Neural Network Language Model ................................................................. 1254

Kaiyu Shi, Kai Yu

Role Play Dialogue Aware Language Models Based on Conditional Hierarchical Recurrent Encoder-Decoder ................................................................. 1259

Ryo Masumura, Tomohiro Tanaka, Atsushi Ando, Hirokazu Masatani, Yoshi Aono

Efficient Keyword Spotting Using Time Delay Neural Networks ........................................................................................................... 1264

Samuel Myers, Vihritr Singh Tomar

VOLUME 3

Automatic DNN Node Pruning Using Mixture Distribution-based Group Regularization ........................................................................................................... 1269

Tsukasa Yoshida, Takafumi Morita, Kazuho Watanabe, Yusuke Shinohara, Yoshikazu Yamaguchi, Yoshi Aono

Conditional-Computation-Based Recursion in Neural Networks for Computationally Efficient Acoustic Modelling ........................................................................................................... 1274

Raffaele Tavarone, Leonardo Badino

Leveraging Translations for Speech Recognition in Low-resource Settings ........................................................................................................... 1279

Antonios Anastasopoulou, David Chiang

Sequence-to-sequence Neural Network Model with 2D Attention for Learning Japanese Pitch Accents ........................................................................................................... 1284

Antoine Bruguiere, Heiga Zen, Arkady Arkhangelsky

Task Specific Sentence Embeddings for ASR Error Detection ........................................................................................................... 1288

Sahar Ghannay, Yannick Estève, Nathalie Camelin

Low-Latency Neural Speech Translation ........................................................................................................................................... 1293

Jan Niehuus, Ngoc-Quan Pham, Thanh-Ke Hu, Matthias Sperber, Alex Waibel

Low-Resource Speech-to-Text Translation ........................................................................................................................................... 1298

Sameer Buntal, Herman Kampen, Karen Livescu, Adam Lopez, Sharon Goldwater

VoiceGuard: Secure and Private Speech Processing ........................................................................................................................................... 1303

Ferdinand Brasser, Tommaso Frassetto, Korbinian Riedhammer, Ahmad-Reza Sadeghi, Thomas Schneider, Christian Weinert

PERSPECTIVE TALK-1

Deep Learning based Situated Goal-oriented Dialogue Systems ........................................................................................................... 1308

Dilek Hakkani-Tür

DEREVERBERATION

Single-channel Speech Dereverberation via Generative Adversarial Training ........................................................................................................... 1309

Chenxing Li, Tiejing Wang, Shuang Xu, Bo Xu

Single-Channel Dereverberation Using Direct MMSE Optimization and Bidirectional LSTM Networks ........................................................................................................... 1314

Wolfgang Mack, Soumitro Chakrabarty, Fabian-Robert Stöter, Sebastian Braun, Bernd Edler, Emanuel Habets

Single-channel Late Reverberation Power Spectral Density Estimation Using Denoising Autoencoders ........................................................................................................... 1319

Ina Kedreasi, Hervé Bourlard

A Non-convolutive NMF Model for Speech Dereverberation ........................................................................................................... 1324

Nikhil M, Rajibah Vanmuragan, Preeti Bao

Cross-Corpora Convolutional Deep Neural Network Dereverberation Preprocessing for Speaker Verification and Speech Enhancement ........................................................................................................... 1329

Peter Guziewich, Stephen Zadorian, Xiao Chen, Hao Zhang

Dereverberation and Beamforming in Robust Far-field Speaker Recognition ........................................................................................................... 1334

Ladislav Molnar, Oldrich Pichot, Pavel Matejka, Ondrej Novotny, Jan Cormočky

AUDIO EVENTS AND ACOUSTIC SCENES

Comparing the Max and Noisy-Or Pooling Functions in Multiple Instance Learning for Weakly Supervised Sequence Learning Tasks ........................................................................................................... 1339

Yun Wang, Juncheng Li, Florian Metze
A Simple Model for Detection of Rare Sound Events .................................................. 1344
Weiran Wang, Chieh-Chi Kao, Chao Wang

Temporal Transformer Networks for Acoustic Scene Classification .......................... 1349
Teng Zhang, Kailai Zhang, Ji Wu

Temporal Attentive Pooling for Acoustic Event Detection ....................................... 1354
Xugang Lu, Peng Shen, Sheng Li, Yu Tsao, Hisatani Kawai

R-CRNN: Region-based Convolutional Recurrent Neural Network for Audio Event Detection .......................................................... 1358
Chieh-Chi Kao, Weiran Wang, Ming Sun, Chao Wang

Detecting Media Sound Presence in Acoustic Scenes ............................................... 1363
Constantinos Papayanis, Justice Amoh, Viktor Bozic, Shiva Sundaram, Chao Wang

SPEAKER DIARIZATION

S4D: Speaker Diarization Toolkit in Python ............................................................... 1368
Pierre-Alexandre Brous, Florent Devou, Anthony Larcher, Simon Petitrenaud, Jean Carriere, Sylvain Meignier

Multimodal Speaker Segmentation and Diarization Using Lexical and Acoustic Cues via Sequence to Sequence
Neural Networks .................................................. 1373
Tae Jin Park, Panayiotis Georgiou

Combined Speaker Clustering and Role Recognition in Conversational Speech .................. 1378
Nikolaos Iliou, Pavlos Papadopoulos, James Gibson, Shrinas Narayanan

The ACLF DIII E: An Easy-to-use Diarization Tool ................................................. 1383
Adrien Le Franc, Eric Riebling, Julien Karadayi, Yan Wang, Camila Scaff, Florian Metze, Alejandra Cristina

Automatic Detection of Multi-speaker Fragments with High Time Resolution ............. 1388
Evodia Kasimirova, Andrey Belyaeve

Neural Speech Turn Segmentation and Affinity Propagation for Speaker Diarization .......... 1393
Ruiqing Yin, Herve Bredin, Claude Barras

PHONATION

Pitch or Phonation: on the Glottalization in Tone Productions in the Ruokeng Hui Chinese Dialect .......................................................... 1398
Minghui Zhang, Fang Hu

Speaker-specific Structure in German Voiceless Stop Voice Onset Times ...................... 1403
Marc-Antony Hullebus, Stephen Tobin, Adamantios Gafos

Creak in the Respiratory Cycle .................................................................................. 1408
Kätil Aare, Pärrel Linnus, Marcin Włodarczak, Mattias Heldner

Acoustic Analysis of Whispey Voice Disguise in Mandarin Chinese ......................... 1413
Cuiling Zhang, Bin Li, Si Chen, Yile Yang

The Zurich Corpus of Vowel and Voice Quality, Version 1.0 ...................................... 1417
Dieter Maurer, Christian D’Heureuse, Heidy Suter, Volker Dettbo, Daniel Friedrichs, Thayabaran Kathiresan

Weighting of Coda Voicing Cues: Glottalisation and Vowel Duration ......................... 1422
Joshua Penney, Felicity Cox, Anita Szakay

COGNITION AND BRAIN STUDIES

Revealing Spatiotemporal Brain Dynamics of Speech Production Based on EEG and Eye Movement .......................................................... 1427
Bin Zhao, Jinjeng Huang, Gaoyan Zhang, Jianwu Ding, Minbo Chen, Yingjianfu, Longbao Wang

Neural Response Development During Distributed Learning ...................................... 1432
Natalie Boll-Attiyani, Jessie S. Nixon, Tomas O. Lentz, Liqian Liu, Sandrien Van Ommen, Çagri Coltekin, Jakobien Van Rij

Learning Two Tone Languages Enhances the Brainstem Encoding of Lexical Tones .... 1437
Athey Raj Megge, Wenging Zong, Yina Lu, Patrick C. M. Wong

Perceptual Sensitivity to Spectral Change in Australian English Close Front Vowels: An Electroencephalographic Investigation .............. 1442
Daniel Williams, Paola Escudero, Adamantios Gafos

Effective Acoustic Cue Learning Is Not Just Statistical, It Is Discriminative .................. 1447
Jessie S. Nixon

Kimberley Mulder, Louis Ten Bosch, Lou Boves

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

Information Encoding by Deep Neural Networks: What Can We Learn? ...................... 1457
Louis Ten Bosch, Lou Boves

Scalable Factorized Hierarchical Variational Autoencoder Training ......................... 1462
Wei-Ning Hsu, James Glass

State Gradients for RNN Memory Analysis ............................................................... 1467
Lyam Verwimp, Hugo Van Hamme, Vincent Renkens, Patrick Wambacq
**SHOW AND TELL 4**

**SPICE AND SINGING PRODUCTION**

**FACTS: A Hierarchical Task-based Control Model of Speech Incorporating Sensory Feedback**
Benjamin Parrell, Vikram Ramanarayanan, Srikanth Nagarajan, John Houde

**Sensorimotor Response to Tongue Displacement Imagery by Talkers with Parkinson’s Disease**
William Katz, Patrick Reidy, Divya Prabhakaran

**Automatic Pronunciation Evaluation of Singing**
Chitradehka Gupta, Hatzhou Li, Ye Wang

**Classification of Nonverbal Human Produced Audio Events: A Pilot Study**
Rachel E. Bruecher, Philippe Chabot, Milton Sarria-Pujia, Patrick Cardinal, Jérémie Voix

**UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science**
Lorenzo Spreafico, Michael Pucher, Anna Matosova

**Articulatory Consequences of Vocal Effort Elicitation Method**
Elisabet Eir Cortes, Marcin Wodarczak, Juraj Simko

**Age-related Effects on Sensorimotor Control of Speech Production**
Anne Hermes, Jane Mertens, Doris Mücke

**An Ultrasound Study of Geminination in Coronal Stops in Eastern Oromo**
Maida Percival, Alexei Kochetov, Yoonjung Kang

**Processing Transition Regions of Glottal Stop Substituted /s/ for Intelligibility Enhancement of Cleft Palate Speech**
Pratima Nomo Sudro, Sithir Kaliota, S R Mahadeva Prassanna

**Reconstructing Neutral Speech from Tracheoesophageal Speech**
N. Abinay Reddy, M. V. Achuth Rao, G. Vishu Meenakshi, Prasanta Kumar Ghosh

**Automatic Evaluation of Soft Articulatory Contact for Stuttering Treatment**
Kesko Ochi, Koichi Mori, Naomi Sakai

**Korean Singing Voice Synthesis Based on an LSTM Recurrent Neural Network**
Juntae Kim, Heejin Choi, Jinuk Park, Minsoo Hahn, Sangjin Kim, Jong-Jin Kim

**The Trajectory of Voice Onset Time with Vocal Aging**
Xiaoxia Chen, Ziyu Xiong, Jian Hu

**ROBUST SPEECH RECOGNITION**

**The Fifth ‘CHiME’ Speech Separation and Recognition Challenge: Dataset, Task and Baselines**
Jon Barker, Shinji Watanabe, Emmanuel Vincent, Jan Trmal

**Voices Obscured in Complex Environmental Settings (VOICES) Corpus**
Colleen Richay, Maria A. Barrios, Zeb Armstrong, Chris Bartels, Horacio Franco, Martin Graciarena, Aaron Lawson, Mahesh Kumarfundava, Allen Stauffer, Julien Van Hout, Paul Gamble, Jeffrey Hetherly, Cory Stephenson, Karl N

**Building State-of-the-art Distant Speech Recognition Using the CHiME-4 Challenge with a Setup of Speech Enhancement Baseline**
Szu-Jus Chen, Arwin Shanmugam Subramanian, Haiman Xu, Shinji Watanabe

**Unsupervised Adaptation with Interpretable Disentangled Representations for Distant Conversational Speech Recognition**
Wei-Ning Hsu, Hao Tang, James Glass

**Investigating Generative Adversarial Networks Based Speech Dereverberation for Robust Speech Recognition**
Ke Wang, Junbo Zhang, Sining Sun, Yujun Wang, Fei Xiang, Lei Xie
Monaural Multi-Talker Speech Recognition with Attention Mechanism and Gated Convolutional Networks .......................................................... 1586

Jian Cheng

Weighting Time-Frequency Representation of Speech Using Auditory Saliency for Automatic Speech Recognition ........................................ 1591

Tuka Al Hanai, Mohammad Ghassemi, James Glass

Acoustic Modeling from Frequency Domain Representations of Speech .................................................................................................................. 1596

Pegah Ghahremani, Hossein Hadian, Hang Lv, Daniel Povey, Sanjeev Khudanpur

Non-Uniform Spectral Smoothing for Robust Children’s Speech Recognition ................................................................................................. 1601

Istwara Chandra Yadav, Avinash Kumar, Syed Shahnawazuddin, Gyanabhar Pradhan

Bidirectional Long-Short Term Memory Network-based Estimation of Reliable Spectral Component Locations ......................................................... 1606

Aaron Nicolson, Kuldip K. Paliwal

Speech Emotion Recognition by Combining Amplitude and Phase Information Using Convolutional Neural Network .............................................. 1611

Lili Guo, Longbiao Wang, Jianwu Dang, Lingjuan Zhang, Haotian Guan, Xiangang Li

Bubble Cooperative Networks for Identifying Important Speech Cues ............................................................................................................. 1616

Viet Anh Trinh, Brian McFee, Michael I Mandel

APPLICATIONS IN EDUCATION AND LEARNING

Real-Time Scoring of an Oral Reading Assessment on Mobile Devices .................................................................................................................. 1621

A Deep Learning Approach to Assessing Non-native Pronunciation of English Using Phone Distances ................................................................. 1626

Konstantinos Kyriakopoulos, Kate Knill, Mark Gales

Paired Phone-Posters Approach to ESL Pronunciation Quality Assessment ........................................................................................................ 1631

Yujia Xiao, Frank Soong, Wening Hu

Investigating the Role of L1 in Automatic Pronunciation Evaluation of L2 Speech ........................................................................................... 1636

Ming Tu, Anna Grabek, Julie Liss, Visar Berisha

Impact of ASR Performance on Free Speaking Language Assessment ............................................................................................................. 1641

Kate Knill, Mark Gales, Konstantinos Kyriakopoulos, Andrey Malinin, Anton Boggi, Yu Wang, Andrew Gaines

Automatic Miscue Detection Using RNN Based Models with Data Augmentation ................................................................................................. 1646

Yoon Seok Hong, Kyung Seo Ki, Gahneun Gyeon

A Study of Objective Measurement of Comprehensibility through Native Speakers’ Shadowing of Learners’ Utterances ................................................... 1651

Yusuke Inoue, Suguru Kabashima, Daisuke Saito, Nobuaki Minematsu, Kumi Kamamura, Yutaka Yamashita

Factorized Deep Neural Network Adaptation for Automatic Scoring of L2 Speech in English Speaking Tests .................................................................... 1656

Dean Luo, Chunxiao Zhang, Linzhong Xia, Linjuan Zhang

On the Difficulties of Automatic Speech Recognition for Kindergarten-Aged Children ............................................................................................... 1661

Gary Yeung, Abner Alwan

Improved Acoustic Modelling for Automatic Literacy Assessment of Children ........................................................................................................ 1666

Mauro Nicolao, Michiel Sanders, Thomas Hain

INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS

Anomaly Detection Approach for Pronunciation Verification of Disordered Speech Using Speech Attribute Features ......................................................... 1671

Amber Afzhan, Jinxi Guo, Soo Jin Park, Vijiay Ravi, Jonathan Flint, Abner Alwan

Effectiveness of Voice Quality Features in Detecting Depression ....................................................................................................................... 1676

Prasanna Kothalkar, Johanna Rudolph, Christine Dollaghan, Jennifer McGlothlin, Thomas Campbell, John H. L. Hansen

Fusing Text-dependent Word-level i-Vector Models to Screen ‘at Risk’ Child Speech ............................................................................................. 1681

Hussain Ali, John H. L. Hansen, M. Ram Charan

Testing Paradigms for Assistive Hearing Devices in Diverse Acoustic Environments ................................................................................................. 1686

Detection of Dementia from Responses to Atypical Questions Asked by Embodied Conversational Agents .............................................................. 1691

Analysis of Visual and Auditory Features Associated with Sustained Vowel and Continuous Speech Productions by Chinese Children with

Functional Articulation Disorders ........................................................................................................................................................................... 1696

Wang Zhang, Xiangqun Gu, Tianqi Wang, Manwa Ng, Feng Yang, Lan Wang, Nan Yan

Estimation of Hypernasality Scores from Cleft Lip and Palate Speech ...................................................................................................................... 1701

C. M. Vikram, Ayush Tripathi, Sushir Kalita, S R Mahadeva Prasanna

Detecting Alzheimer’s Disease Using Gated Convolutional Neural Network from Audio Data ....................................................................................... 1706

Yasuaki Chang, Yantin Qian, Tfani Warnita, Nakamura Inoue, Koichi Shinoda

Automatic Detection of Orofacial Impairment in Stroke ........................................................................................................................................ 1711

Andrea Bandini, Jordana Green, Brian Richburg, Yana Yamasouva

Detecting Depression with Audio/Text Sequence Modeling of Interviews ............................................................................................................. 1716

Tuka Al Hanai, Mohammad Ghassemi, James Glass
SPEAKER CHARACTERIZATION AND ANALYSIS

Discourse Marker Detection for Hesitation Events on Mandarin Conversation ................................................................. 1721
Yu-Wen Wang, Jen-Hsin Huang, Kuan-Yu Chen, Hsin-Hsi Chen

Acoustic and Perceptual Characteristics of Mandarin Speech in Homosexual and Heterosexual Male Speakers .................. 1726
Puyang Geng, Wentao Gu, Hiroya Fuzisaki

Automatic Question Detection from Acoustic and Phonetic Features Using Feature-wise Pre-training ............................... 1731
Atsushi Ando, Reine Asakawa, Ryo Matsumura, Hosana Kamiyama, Satoshi Kobashikawa, Yushi Aono

Improving Response Time of Active Speaker Detection Using Visual Prosody Information Prior to Articulation .................. 1736
Fash Halder, Saturnino Lao, Carl Vogel, Nick Campbell

Audio-Visual Prediction of Head-Nod and Turn-Taking Events in Dyadic Interactions .................................................. 1740
Bekir Berker Tüker, Engin Erzin, Yücel Yemez, Mehdi Sezgin

Analyzing Effect of Physical Expression on English Proficiency for Multimodal Computer-Assisted Language Learning .......................................................... 1747
Haoran Wu, Yuwa Chiba, Takashi Nose, Akinori Ito

Analysis of the Effect of Speech-Laugh on Speaker Recognition System ........................................................................... 1751
Sri Harsha Dampala, Ashish Panda, Sanil Kumar Koppurapu

Vocal Biomarkers for Cognitive Performance Estimation in a Working Memory Task ..................................................... 1756
Jennifer Sloboda, Adam Lampert, James Williamson, Christopher Small, Daryush D. Mehta, Col Ian Curry, Kristin Heaton, Jeffrey Palmer, Thomas Quatieri

Lexical and Acoustic Deep Learning Model for Personality Recognition ........................................................................... 1761
Guzhchen An, Rivka Levitan

PERSPECTIVE TALK-2

Open Problems in Speech Recognition ......................................................................................................................... 1766
Bhuvana Ramabhadran

PLENARY TALK-2

Evolution of Neural Network Architectures for Speech Recognition ....................................................................... 1767
Hervé Bourlard

NOVEL NEURAL NETWORK ARCHITECTURES FOR ACOUSTIC MODELLING

Layer Trajectory LSTM .................................................................................................................................................. 1768
Jinyu Li, Changliang Liu, Yifan Gong

Semi-tied Units for Efficient Gating in LSTM and Highway Networks ............................................................................ 1773
Chao Zhang, Philip Woodland

Gaussian Process Neural Networks for Speech Recognition ........................................................................................... 1778
Max W. Y. Lam, Shoukang Hu, Xiaodong Xie, Shinlong Liu, Jianwei Yu, Rongfeng Su, Xuming Liu, Helen Meng

Acoustic Modeling with Densely Connected Residual Network for Multichannel Speech Recognition ............................ 1783
Jian Tang, Yan Song, Lirong Dai, Ian McLoughlin

Gated Recurrent Unit Based Acoustic Modeling with Future Context ............................................................................. 1788
Jie Li, Xiaorui Wang, Yuanxuan Zhao, Yan Li

Output-Gate Projected Gated Recurrent Unit for Speech Recognition ............................................................................ 1793
Gaofeng Cheng, Daniel Povey, Lu Huang, Ji Xu, Sanjeev Khudanpur, Yonghong Yan

LANGUAGE IDENTIFICATION

Performance Analysis of the 2017 NIST Language Recognition Evaluation ......................................................................... 1798
Seyed Omid Sadjadi, Timoteo Kheyrykhah, Craig Greenberg, Elliot Singer, Douglas Reynolds, Lisa Mason, Jaime Hernandez-

Using Deep Neural Networks for Identification of Slavic Languages from Acoustic Signal ............................................... 1803
Lucas Mateju, Petr Cerva, Jindrich Zdansky, Radek Safarik

Adding New Classes without Access to the Original Training Data with Applications to Language Identification .............. 1808
Hagai Taitelbaum, Ehud Ben-Reuven, Jacob Goldberger

Feature Representation of Short Utterances Based on Knowledge Distillation for Spoken Language Identification .......................... 1813
Peng Shen, Xugang Lu, Sheng Li, Hisashi Kawai

Sub-band Envelope Features Using Frequency Domain Linear Prediction for Short Duration Language Identification .................................................. 1818
Sarith Fernando, Vidhyaasaharan Sethu, Elijahamby Ambikairajah

Effectiveness of Single-Channel BLSTM Enhancement for Language Identification ........................................ 1823
Peter Sibbern Frederiksen, Jesús Villalba, Shinji Watanabe, Zheng-Hua Tan, Najim Dehak
PRODUCTION OF PROSODY

Articulation Rate as a Speaker Discriminant in British English .................................................. 1828
Erica Gold

Truncation and Compression in Southern German and Australian English .................................. 1833
Jenny Yu, Katharina Zahner

Prominence-based Evaluation of L2 Prosody ................................................................................. 1838
Heini Kaltilo, Antti Sani, Päivi Virkkunen, Juraj Smko

Length Contrast and Covarying Features: Whistled Speech as a Case Study .............................. 1843
Rachid Boudanoue, Giuseppina Turco, Julien Meyer

Information Structure, Affect and Prenuclear Prominence in American English ...................... 1848
Eleanor Chodruff, Jennifer Cole

Effects of User Controlled Speech Rate on Intelligibility in Noisy Environments ......................... 1853
John S. Novak III, Robert V. Kenyon

SPEECH INTELLIGIBILITY AND QUALITY

Binaural Speech Intelligibility Estimation Using Deep Neural Networks .................................. 1858
Kazuhito Kondo, Kazuya Taira, Toshike Kobayashi

Multi-resolution Gammachirp Envelope Distortion Index for Intelligibility Prediction of Noisy Speech .......................................................... 1863
Katsunori Yamamoto, Toshio Irie, Narumi Ohashi, Shoko Araki, Keisuke Kimishita, Tomohiro Nakatani

Speech Intelligibility Enhancement Based on a Non-causal Wavenet-like Model ...................... 1868
P. V. Muhammad Shifaa, Vassilis Tsiaras, Yannis Stylianou

Quality-Net: An End-to-End Non-intrusive Speech Quality Assessment Model Based on BLSTM .......................................................... 1873
Szu-Wei Fu, Yu Tsao, Hsin-Tz Huang, Hsin-Min Wang

Global SNR Estimation of Speech Signals Using Entropy and Uncertainty Estimates from Dropout Networks .......................................................... 1878
Rohith Aralikatti, Dilip Kumar Margam, Tanay Sharma, Aphinav Thanda, Shankar Venkatesan

Detecting Packet-Loss Concealment Using Formant Features and Decision Tree Learning ........ 1883
Gabriel Mittag, Sebastian Möller

INTEGRATING SPEECH SCIENCE AND TECHNOLOGY FOR CLINICAL APPLICATIONS

UltraSuite: A Repository of Ultrasound and Acoustic Data from Child Speech Therapy Sessions ........................................................................ 1888
Aciel Eshky, Manuel Sam Ribeiro, Joanne Cleland, Korin Richmond, Zoe Roxburgh, James M. Scobie, Alan Wrench

Detecting Signs of Dementia Using Word Vector Representations ............................................. 1893
Bahman Mirheidari, Daniel Blackburn, Traci Walker, Annalena Vennetti, Markus Reuber, Heidi Christensen

Classification of Huntington Disease Using Acoustic and Lexical Features ............................ 1898
Matthew Perez, Wenyu Jin, Duc Le, Noelle Carlucci, Praveen Dayalu, Angela Roberts, Emily Mower Provost

VOLUME 4

The PRIORI Emotion Dataset: Linking Mood to Emotion Detected In-the-Wild .......................... 1903
Soheil Khorsam, Minamia Jaiswal, John Gideon, Melvin McIntire, Emily Mower Provost

Language Features for Automated Evaluation of Cognitive Behavior Psychotherapy Sessions ........................................................................ 1908
Nikolaos Flemotomos, Victor Martinez, James Gibson, David Atkins, Torrey Creed, Shrikanth Narayanan

Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks .......................................................... 1913
Kwanghoom An, Myungjong Kim, Kristin Teplinsky, Jordan Green, Thomas Campbell, Yana Yumusova, Daragh Heitczman, Jun Wang

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
Preeti Rao, Megha Pandya, Kamini Sahu, Kanhaiya Kumar, Nandini Bondale

Building a Unified Code-Switching ASR System for South African Languages ........................ 1923
Emre Yilmaz, Asit Biswas, Ewald van der Westhuizen, Febe De Wet, Thomas Niesler

Study of Semi-supervised Approaches to Improving English-Mandarin Code-Switching Speech Recognition ........................................................................ 1928
Pengcheng Guo, Haixia Xu, Lei Xie, Eng Siong Chng

Acoustic and Textual Data Augmentation for Improved ASR of Code-switched Speech .......... 1933
Emre Yilmaz, Henk Van Der Heuvel, David Van Leeuwen

The Role of Cognate Words, POS Tags and Entrainment in Code-Switching .......................... 1938
Victor Soto, Nishmar Castors, Julia Hirschberg

Homophone Identification and Merging for Code-switched Speech Recognition .................... 1943
Brij Mohan Lal Srivastava, Sanayana Sitaram

Code-switching in Indic Speech Synthesisers .............................................................................. 1948
Anju Leela Thomas, Anusha Prakash, Arun Baby, Hema Murthy
A Novel Approach for Effective Recognition of the Code-Switched Data on Monolingual Language Model .......................................................... 1953
Sreevani Ganji, Rohit Sinha

SHOW AND TELL 5

Hierarchical Accent Determination and Application in a Large Scale ASR System .......................................................... 1958
Banya Virwanathan, Pertyasamy Paramasivam, Jithendra Vepa

Toward Scalable Dialog Technology for Conversational Language Learning: Case Study of the TOEFL® MOOC ............................................. 1960
Vikram Ramanarayanan, David Pauller, Patrick Lange, Eugene Tsugun, Rina M. Ubahe, Keean G. Vanini, David Saendemann-Oeef

Machine Learning Powered Data Platform for High-Quality Speech and NLP Workflows .......................................................... 1962
João Freitas, Jorge Ribeiro, Daan Baldewijns, Sara Oliveira, Daniela Braga

Fully Automatic Speaker Separation System, with Automatic Enrolling of Recurrent Speakers .......................................................... 1964
Raphael Cohen, Orgad Keller, Jason Levy, Russell Levy, Micha Breakstone, Amit Ashkenazi

Online Speech Translation System for Tamil .......................................................... 1966
Ayayya Madhavaraj, H. K. Shiwa Kumar, A. G. Ramakrishnan

VOICE CONVERSION AND SPEECH SYNTHESIS

Unsupervised Vocal Tract Length Warped Posterior Features for Non-Parallel Voice Conversion .......................................................... 1968
Nirmesh Shah, Maulik C. Madhavi, Hemant Patil

Voice Conversion with Conditional SampleRNN .......................................................... 1973
Cong Zhou, Michael Horgan, Vivek Kumar, Cristina Vasco, Dan Darcy

A Voice Conversion Framework with Tandem Feature Sparse Representation and Speaker-Adapted WaveNet Vocoder .......................................................... 1978
Berrak Siyan, Mingyang Zhang, Haizhou Li

WaveNet Vocoder with Limited Training Data for Voice Conversion .......................................................... 1983
Li-Juan Liu, Zhen-Hua Ling, Yuan Jiang, Ming Zhou, Li-Rong Dai

Collapsed Speech Segment Detection and Suppression for WaveNet Vocoder .......................................................... 1988
Yi-Chiao Wu, Kazuhiro Kobayashi, Tomoki Hayashi, Patrick Lamban Togbing, Tomoki Toda

High-quality Voice Conversion Using Spectrogram-Based WaveNet Vocoder .......................................................... 1993
Kuan Chen, Bo Chen, Jishao Lai, Kai Yu

Spanish Statistical Parametric Speech Synthesis Using a Neural Vocoder .......................................................... 1998
Antonio Bonafonte, Santiago Pascual, Georgina Dorca

Experiments with Training Corpora for Statistical Text-to-speech Systems .......................................................... 2002
Monika Podsiadlo, Victor Ungureanu

Yu Gu, Yongguo Kang

Speaker-independent Raw Waveform Model for Glottal Excitation .......................................................... 2012
Laure Javela, Vassilis Triaras, Bajijabu Bollepalli, Mana Airaksinen, Junichi Yamagishi, Pauvo Alho

A New Glottal Neural Vocoder for Speech Synthesis .......................................................... 2017
Yang Cui, Xi Wang, Lei He, Frank K. Soong

Exemplar-based Speech Waveform Generation .......................................................... 2022
Oliver Watts, Cassia Valentin-Botinhao, Felipe Espic, Simon King

Frequency Domain Variants of Velvet Noise and Their Application to Speech Processing and Synthesis .......................................................... 2027
Hideki Kawahara, Ken-Ichi Sakakihara, Masanori Morise, Hideki Banno, Tomoki Toda, Toshio Iriyama

EXTRACTING INFORMATION FROM AUDIO

Joint Learning of Interactive Spoken Content Retrieval and Trainable User Simulator .......................................................... 2032
Pei-Hung Chang, Kun Tung, Ching-Luan Tai, Hung-Yi Lee

Attention-based End-to-End Models for Small-Footprint Keyword Spotting .......................................................... 2037
Changshao Shan, Junbo Zhang, Yujun Wang, Lei Xie

Raghav Rajan, Ashwin Vijayakumar, Deepu Vijayakumar

Topic and Keyword Identification for Low-resourced Speech Using Cross-Language Transfer Learning .......................................................... 2047
Wenda Chen, Mark Ratezusson-Johnson, Nancy F. Chen

Automatic Speech Recognition and Topic Identification from Speech for Almost-Zero-Resource Languages .......................................................... 2052
Matthew Wiesner, Chunxi Liu, Lucas Ondel, Craig Harman, Vimal Manohar, Jun Trmal, Zhongqi Huang, Najim Dehak, Sanjeev Khudanpur

Play Duration Based User-Entity Affinity Modeling in Spoken Dialog System .......................................................... 2057
Bo Xiao, Nicholas Monath, Shankar Ananthakrishnan, Abishek Ravi

Empirical Analysis of Score Fusion Application to Combined Neural Networks for Open Vocabulary Spoken Term Detection .......................................................... 2062
Shi-Wook Lee, Kazuyo Tanaka, Yoshiaki Ishi

Phonological Posterior Hasing for Query by Example Spoken Term Detection .......................................................... 2067
Afsaneh Asaei, Dhananjay Ram, Hervé Bourlard
Term Extraction via Neural Sequence Labeling a Comparative Evaluation of Strategies Using Recurrent Neural Networks ................................................................. 2072
Anjuli Kannan, Kai Chen, Diana Jannuzzi, Alvin Rajkomar

Semi-supervised Learning for Information Extraction from Dialogue .............................................................................. 2077
Youjuan Shi, Kang Min Yoo, Sang-Goo Lee

Slot Filling with Delexicalized Sentence Generation .......................................................................................... 2082
Takaaki Shochi, Jean-Luc Rouas, Marine Guerry, Donna Erickson

Music Genre Recognition Using Deep Neural Networks and Transfer Learning .............................................. 2087
Deepamroy Ghosal, Maheshkumar H. Kolekar

Efficient Voice Trigger Detection for Low Resource Hardware .............................................................................. 2092
Siddharth Sigta, Rob Haynes, Hywel Richards, Erik Marchi, John Bridle

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
Qiguang Lin, Yiwen Shao

Estimation of the Vocal Tract Length of Vowel Sounds Based on the Frequency of the Significant Spectral Valley .............................................................................. 2102
T. V. Ananthapadmanabha, A. G. Ramakrishnan

Deep Learning Techniques for Koala Activity Detection .......................................................................................... 2107
Ivan Herman, Michael Toosey, Bradley Law, Paul Roe

Glottal Closure Instant Detection from Speech Signal Using Voting Classifier and Recursive Feature Elimination .............................................................................. 2112
Jindrich Matousek, Daniel Tihelka

Assessing Speaker Engagement in 2-Person Debates: Overlap Detection in United States Presidential Debates .............................................................................. 2117
Midia Yousefi, Navid Shokouhi, John H. L. Hansen

All-Conv Net for Bird Activity Detection: Significance of Learned Pooling .............................................................................. 2122
Arjun Panajakshie, Anshul Thakur, Daleh Thapar, Padmanabhan Rajan, Aditya Nigam

Deep Convex Representations: Feature Representations for Bioacoustics Classification .............................................. 2127
Anshul Thakur, Vinayak Abrol, Pulkit Sharma, Padmanabhan Rajan

Detection of Glottal Excitation Epochs in Speech Signal Using Hilbert Envelope .............................................................................. 2132
Hirak Dasgupta, Prem C. Pandey, K S Natrajan

Analyzing Thai Tone Distribution through Functional Data Analysis .............................................................................. 2137
Hong Zhang

Articulatory Feature Classification Using Convolutional Neural Networks .............................................................................. 2142
B. Ganga Gowri, K. P. Soman, D. Govind

A New Frequency Coverage Metric and a New Subband Encoding Model, with an Application in Pitch Estimation .............................................................................. 2147
Shoufeng Lin

Improved Epoch Extraction from Telephonic Speech Using Chebfun and Zero Frequency Filtering .............................................................................. 2152
B. Ganga Gowri, K. P. Soman, D. Govind

SPEECH PROSODY

An Empirical Analysis of the Correlation of Syntax and Prosody .............................................................................. 2157
Arne Köhn, Timo Baumann, Oskar Dörfler

Analysing the Focus of a Hierarchical Attention Network: the Importance of Enjambments When Classifying Post-modern Poetry .............................................................................. 2162
Timo Baumann, Hussein Hussein, Burkhard Meyer-Sickendiek

Language-Dependent Melody Embeddings .......................................................................................... 2167
Daniil Kocharov, Alla Menshikova

Stress Distribution of Given Information in Chinese Reading Texts .............................................................................. 2171
Yuan Jia, Xianxiao Ma

Acoustic-prosodic Entrainment in Structural Metadata Events .............................................................................. 2176
Vera Cabarrão, Fernando Batista, Helena Moniz, Isabel Trancoso, Ana Isabel Mata

Formant Measures of Vowels Adjacent to Alveolar and Retroflex Consonants in Arrernte: Stressed and Unstressed Position .............................................................................. 2181
Marina Tahain, Richard Beare, Andrew Butcher

Automatic Assessment of L1 English Word Prosody Using Weighted Distances of F0 and Intensity Contours .............................................................................. 2186
Quy-Thao Truong, Tsuneo Kato, Seiichi Yamamoto

Homogeneity vs Heterogeneity in Indian English: Investigating Influences of L1 on F0 Range .............................................................................. 2191
Olga Maxwell, Elinore Payne, Rosey Billington

Emotional Prosody Perception in Mandarin-speaking Congenital Amusics .............................................................................. 2196
Yixin Zhang, Zianzheng Geng, Jinsong Zhang

Cultural Differences in Pattern Matching: Multisensory Recognition of Socio-affective Prosody .............................................................................. 2201
Takasaki Shoichi, Jean-Luc Rouas, Marine Guerry, Donna Erickson
PERSPECTIVE TALK-3

Speech Processing in the Human Brain Meets Deep Learning .................................................. 2206
Nima Mesgarani

RECURRENT NEURAL MODELS FOR ASR

ESPnet: End-to-End Speech Processing Toolkit ........................................................................... 2207
Shinji Watanabe, Takuaki Horii, Shigeki Karita, Tomoki Hayakshi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalla Soplin, Jahn Heymann, Matthew Wiesner, Nanxin Chen, Adithya Renduchintala, Tsukasa Ochiai

A GPU-based WFST Decoder with Exact Lattice Generation ....................................................... 2212
Zhehuaui Chen, Justin Luijens, Hainan Xu, Yiming Wang, Daniel Povey, Sanjeev Khudanpur

Automatic Speech Recognition System Development in the "Wild" ........................................... 2217
Anton Ragni, Mark Gales

Semantic Lattice Processing in Contextual Automatic Speech Recognition for Google Assistant ................................................. 2222
Leonid Velivskich, Ian Williams, Justin Schelter, Petar Aleksic, Pedro Moreno, Michael Riley

Contextual Speech Recognition in End-to-end Neural Network Systems Using Beam Search ...................................................... 2227
Ian Williams, Anjuli Kannan, Petar Aleksic, David Rybach, Tara Sainath

Forward-Backward Attention Decoder ......................................................................................... 2232
Masaom Mooma, Shinuske Sakai, Tatsuya Kawahara

SPEAKER VERIFICATION USING NEURAL NETWORK METHODS I

Learning Discriminative Features for Speaker Identification and Verification .......................... 2237
Sarthak Yadav, Anil Rai

Triplet Loss Based Cosine Similarity Metric Learning for Text-independent Speaker Recognition .................................................. 2242
Sergey Novoselov, Vadim Shchemelinin, Andrey Shulipa, Alexandre Koslov, Ivan Kremen

Speaker Embedding Extraction with Phonetic Information ......................................................... 2247
Yi Liu, Liang He, Jia Liu, Michael T. Johnson

Attentional Statistics Pooling for Deep Speaker Embedding ...................................................... 2252
Koji Okabe, Takamizu Kazushige, Koichi Shinoda

Robust and Discriminative Speaker Embedding via Intra-Class Distance Variance Regularization .................................................................................................................................................................................................................. 2257

Deep Discriminative Embeddings for Duration Robust Speaker Verification ......................... 2262
Na Li, Deyi Tuo, Dan Su, Zhifeng Li, Dong Yu

SPEECH PERCEPTION IN ADVERSE CONDITIONS

Impact of Different Speech Types on Listening Effort ................................................................... 2267
Olympia Simantiraki, Martin Cooke, Simon King

Who Are You Listening to? Towards a Dynamic Measure of Auditory Attention to Speech-on-speech. .................................................................................................................................................................................................................................................. 2272
Moira-Phoebe Huet, Christophe Michely, Etienne Gaudrain, Etienne Parizet

Investigating the Role of Familiar Face and Voice Cues in Speech Processing in Noise ........ 2276
Jeevan Kim, Sonya Karisma, Vincent Aubanel, Chris Davis

The Conversation Continues: the Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition .......................................................................................................................................................................................................................... 2280
Odette Scharenborg, Martha Larson

Loud and Shouted Speech Perception at Variable Distances in a Forest .................................... 2285
Julien Meyer, Fanny Meunier, Laure Dentel, Noelia Do Carmo Blanco, Frédéric Sèbe

Phoneme Resistance and Phoneme Confusion in Noise: Impact of Dyslexia .......................... 2290
Noelia Do Carmo Blanco, Julien Meyer, Michel Hoen, Fanny Meunier

MEASURING PITCH AND ARTICULATION

Conditional End-to-End Audio Transforms ................................................................................. 2295
Albert Haque, Michelle Guo, Parveek Verma

Detection of Glottal Closure Instants in Degraded Speech Using Single Frequency Filtering Analysis ........................................................................................................................................................................................................................................... 2300
Gunnam Aneeja, Sudarsana Reddy Kadiri, Bayya Yegnunarayana

Tone Recognition Using Lifters and CTC ...................................................................................... 2305
Loren Loganisi, Vickyright Singh Tomar

Epoch Extraction from Pathological Children Speech Using Single Pole Filtering Approach .................................................................................................................................................................................................................................................. 2310
C. M. Vikram, S R Mohadeva Prasanna

Automated Classification of Vowel-Gesture Parameters Using External Broadband Excitation ........................................................................................................................................................................................................................................... 2315
B. T. Balamurali, Jer-Ming Chen

Estimation of Fundamental Frequency from Singing Voice Using Harmonics of Impulse-like Excitation Source ........................................................................................................................................................................................................... 2319
Sudarsana Reddy Kadiri, Bayya Yegnunarayana
SPEECH AND LANGUAGE ANALYTICS FOR MENTAL HEALTH

Investigating the Effect of Audio Duration on Dementia Detection Using Acoustic Features ............................................................ 2324
Jochen Weiser, Miguel Angel Sánchez, Srinivasan Umesh, Tanja Schulze

An Interlocutor-Modulated Attentional LSTM for Differentiating between Subgroups of Autism Spectrum Disorder ........................................ 2329
Yun-Shao Lin, Susan Shar-Fen Gau, Chi-Chun Lee

Recognition of Echolalic Autistic Child Vocalisations Utilising Convolutional Recurrent Neural Networks ........................................ 2334
Shahin Amiriparian, Alice Baird, Sahib Julka, Alyssa Alcorn, Sandra Ohl, Suncica Petrovic, Eloise Ainger, Nicholas Cummins, Björn Schuller

Modeling Interpersonal Influence of Verbal Behavior in Couples Therapy Dyadic Interactions .............................................................. 2339
Sandeep Nalian Chakravarthula, Brian Baumcom, Panayiotis Georgiou

Computational Modeling of Conversational Humor in Psychotherapy .................................................................................................. 2344
Asil Ramakrishna, Timothy Greer, David Atkins, Shriniketh Narayanan

Multimodal I-vectors to Detect and Evaluate Parkinson’s Disease ....................................................................................................... 2349
Nicanor García, Juan Camilo Vásquez Correa, Juan Rafael Orozco-Arroyave, Elmar Nößl

SPOKEN CALL SHARED TASK, SECOND EDITION

Overview of the 2018 Spoken CALL Shared Task ................................................................................................................................. 2354
Claudia Bauer, Andrew Caines, Cathy Chua, Johanna Gerlach, Mengjie Qian, Manny Rayner, Martin Russell, Helmer Strik, Xizi Wei

The CSU-K Rule-Based System for the 2nd Edition Spoken CALL Shared Task ........................................... 2359
Dominik Jülg, Mario Kunstek, Cem Philipp Freimoser, Kay Berkling, Mengjie Qian

Liulishao’s System for the Spoken CALL Shared Task 2018 .................................................................................................................. 2364
Hu Nguyen, Lei Chen, Ramon Prieto, Chuan Wang, Yang Liu

An Optimization Based Approach for Solving Spoken CALL Shared Task ............................................................................................. 2369
Mohammed Atieh, Amadoboul Hamani, Aziz Qarouch

The University of Birmingham 2018 Spoken CALL Shared Task Systems .............................................................................................. 2374
Mengjie Qian, Xizi Wei, Peter Jancovic, Martin Russell

Improvements to an Automated Content Scoring System for Spoken CALL Responses: the ETS Submission to the Second Spoken CALL Shared Task .............................................................. 2379
Keelan Evanini, Matthew Mulholland, Rutuja Ubale, Yao Qian, Robert Pugh, Vikram Ramanarayan, Aoife Cahill

SHOW AND TELL 6

Extracting Speaker’s Gender, Accent, Age and Emotional State from Speech .............................................................................................. 2384
Nagendra Goel, Moumita Sarma, Tejendra Kothwah, Dharmesh Agarwal, Zobra Iqbal, Sarbhi Chauhan

Determining Speaker Location from Speech in a Practical Environment ................................................................................................. 2386
Bhvs Narayananmurthy, J. V. Satyanarayana, Bayya Tegnanarayana

An Automatic Speech Transcription System for Manipuri Language ...................................................................................................... 2388
Tanvina Patel, D. N. Krishna, Noor Fatihma, Nizar Shah, C. Mahima, Deepak Kumar, Anuroop Iyengar

SPIRE-SST: An Automatic Web-based Self-learning Tool for Syllable Stress Tutoring (SST) to the Second Language Learners .................................................................................................................... 2390
Chiranjeevi Amur, Kaushik Prasanta Kumar Ghosh

Glottovibrato Graph: A Device and Method for Recording, Analysis and Visualization of Glottal Activity .............................................. 2392
Kishalay Chakraborty, Senjam Shantirani Devi, Sanjeevan Devnath, S R Mahadeva Prasanna, Priyankoo Sarmah

ADJUSTING TO SPEAKER, ACCENT, AND DOMAIN

Multi-Modal Data Augmentation for End-to-end ASR ................................................................................................................................. 2394
Aditya Renduchintala, Shuoyang Ding, Matthew Wiesner, Shinji Watanabe

Multi-task Learning with Augmentation Strategy for Acoustic-to-word Attention-based Encoder-decoder Speech Recognition .................. 2399
Takafumi Moriya, Sei Ueno, Yusuke Shinohara, Marc Delcroix, Yoshikazu Yamaguchi, Yushi Aono

Training Augmentation with Adversarial Examples for Robust Speech Recognition ................................................................................ 2404
Sining Sun, Ching-Feng Yeh, Mrti Ostendorf, Mei-Yuh Hwang, Lei Xie

Data Augmentation Improves Recognition of Foreign Accented Speech .............................................................................................. 2409
Takashi Fukushima, Raul Fernandez, Andrew Rosenberg, Samuel Thomas, Bhuvana Ramabhadran, Alexander Sorin, Gakuto Kurata

Speaker Adaptive Training and Mixup Regularization for Neural Network Acoustic Models in Automatic Speech Recognition .................. 2414
Natalia Tomashenko, Yuri Khokhlov, Tannick Esteve

Neural Language Codes for Multilingual Acoustic Models ................................................................................................................ 2419
Markus Müller, Sebastian Stüker, Alex Waibel

Encoder Transfer for Attention-based Acoustic-to-word Speech Recognition ........................................................................................ 2424
Sei Ueno, Takafumi Moriya, Masato Mimura, Shinsuke Sakai, Yusuke Shinohara, Yoshikazu Yamaguchi, Yushi Aono, Tatsuya Kawahara
Empirical Evaluation of Speaker Adaptation on DNN Based Acoustic Model ................................................................. 2429
Improving DNNs Trained with Non-Native Transcriptions Using Knowledge Distillation and Target Interpolation .......... 2434
Amit Das, Mark Hasegawa-Johnson
Improving Cross-Linguinal Knowledge Transferability Using Multilingual TDNN-BLSTM with Language-Dependent Pre-Final Layer .............................................................................. 2439
Siyan Fong, Tan Lee
Auxiliary Feature Based Adaptation of End-to-end ASR Systems ................................................................. 2444
Marc Delcroix, Shinji Watanabe, Atsufumi Ogawa, Shigeki Karita, Tomohiro Nakatani
Leveraging Native Language Information for Improved Accented Speech Recognition ................................................... 2449
Shahram Gharbani, John H. L. Hansen
Improved Accented Speech Recognition Using Accent Embeddings and Multi-task Learning ....................................... 2454
Abhinav Jain, Minali Upreti, Preethi Jyothi
Fast Language Adaptation Using Phonological Information .................................................................................. 2459
Sibo Tong, Philip N. Garner, Hervé Bourlard

SPEECH SYNTHESIS PARADIGMS AND METHODS

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

VOLUME 5

SECOND LANGUAGE ACQUISITION AND CODE-SWITCHING

Should Code-switching Models Be Asymmetric? ................................................................................................. 2534
Barbara E. Bullock, Giulietta Gazzmán, Jacqueline Serigou, Almeida Jacqueline Toribio
Cross-language Perception of Mandarin Lexical Tones by Mongolian-speaking Bilinguals in the Inner Mongolia
Autonomous Region, China .................................................................................................................................. 2539
Kimiko Tsukada, Yu Rong
Automatically Measuring L2 Speech Fluency without the Need of ASR: A Proof-of-concept Study with Japanese
Learners of French ............................................................................................................................................. 2544
Lionel Fontain, Maxime Le Coc, Sylvain Deter
Analysis of L2 Learners’ Progress of Distinguishing Mandarin Tone 2 and Tone 3 ...................................................... 2549
Yue Sun, Win Thazar Kyaw, Jinsong Zhang, Yoshinori Sagioka
Unsupervised Discovery of Non-native Phonetic Patterns in L2 English Speech for Mispronunciation Detection
and Diagnosis .................................................................................................................................................. 2554
Xu Li, Shuoguang Mao, Xixin Wu, Kun Li, Xuying Liu, Helen Meng
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wuxi Speakers’ Production and Perception of Coda Nasals in Mandarin</td>
<td>2559</td>
</tr>
<tr>
<td>The Diphthongs of Formal Nigerian English: A Preliminary Acoustic Analysis</td>
<td>2563</td>
</tr>
<tr>
<td>Characterizing Rhythm Differences between Strong and Weak Accented L2 Speech</td>
<td>2568</td>
</tr>
<tr>
<td>Analysis of Phone Errors Attributable to Phonological Effects Associated With Language Acquisition Through Bottleneck Feature Visualisations</td>
<td>2573</td>
</tr>
<tr>
<td>Category Similarity in Multilingual Pronunciation Training</td>
<td>2578</td>
</tr>
<tr>
<td>Talker Diarization in the Wild: the Case of Child-centered Daylong Audio-recordings</td>
<td>2583</td>
</tr>
<tr>
<td>Automated Classification of Children’s Linguistic versus Non-Linguistic Vocalisations</td>
<td>2588</td>
</tr>
<tr>
<td>Pitch Characteristics of L2 English Speech by Chinese Speakers: A Large-scale Study</td>
<td>2593</td>
</tr>
<tr>
<td>TOPICS IN SPEECH RECOGNITION</td>
<td></td>
</tr>
<tr>
<td>Dual Language Models for Code Switched Speech Recognition</td>
<td>2598</td>
</tr>
<tr>
<td>Multilingual Neural Network Acoustic Modelling for ASR of Under-Resourced English-isiZulu Code-Switched Speech</td>
<td>2603</td>
</tr>
<tr>
<td>Fast ASR-free and Almost Zero-resource Keyword Spotting Using DTW and CNNs for Humanitarian Monitoring</td>
<td>2608</td>
</tr>
<tr>
<td>Text-Dependent Speech Enhancement for Small-Footprint Robust Keyword Detection</td>
<td>2613</td>
</tr>
<tr>
<td>Improved ASR for Under-resourced Languages through Multi-task Learning with Acoustic Landmarks</td>
<td>2618</td>
</tr>
<tr>
<td>Cross-language Phoneme Mapping for Low-resource Languages: An Exploration of Benefits and Trade-offs</td>
<td>2623</td>
</tr>
<tr>
<td>User-centric Evaluation of Automatic Punctuation in ASR Closed Captioning</td>
<td>2628</td>
</tr>
<tr>
<td>Punctuation Prediction Model for Conversational Speech</td>
<td>2633</td>
</tr>
<tr>
<td>BUT OpenSAT 2017 Speech Recognition System</td>
<td>2638</td>
</tr>
<tr>
<td>Visual Recognition of Continuous Cued Speech Using a Tandem CNN-HMM Approach</td>
<td>2643</td>
</tr>
<tr>
<td>Building Large-vocabulary Speaker-independent Lipreading Systems</td>
<td>2648</td>
</tr>
<tr>
<td>CRIM’s System for the MGB-3 English Multi-Genre Broadcast Media Transcription</td>
<td>2653</td>
</tr>
<tr>
<td>Sampling Strategies in Siamese Networks for Unsupervised Speech Representation Learning</td>
<td>2658</td>
</tr>
<tr>
<td>Compact Feedforward Sequential Memory Networks for Small-footprint Keyword Spotting</td>
<td>2663</td>
</tr>
<tr>
<td>ZERO-RESOURCE SPEECH RECOGNITION</td>
<td></td>
</tr>
<tr>
<td>Multilingual Bottleneck Features for Subword Modeling in Zero-resource Languages</td>
<td>2668</td>
</tr>
<tr>
<td>Exploiting Speaker and Phonetic Diversity of Mismatched Language Resources for Unsupervised Subword Modeling</td>
<td>2673</td>
</tr>
<tr>
<td>Unsupervised Word Segmentation from Speech with Attention</td>
<td>2678</td>
</tr>
<tr>
<td>Full Bayesian Hidden Markov Model Variational Autoencoder for Acoustic Unit Discovery</td>
<td>2688</td>
</tr>
<tr>
<td>Unsup: Unsupervised Speech Context Embeddings</td>
<td>2693</td>
</tr>
</tbody>
</table>
### Spatial and Phase Cues for Source Separation and Speech Recognition

- Impact of Aliasing on Deep CNN-Based End-to-End Acoustic Models .......................................................... 2698
- Yuan Gong, Christian Poellabauer

- Keyword Based Speaker Localization: Localizing a Target Speaker in a Multi-speaker Environment .......................... 2703
- Sanit Sivanakaran, Emmanuel Vincent, Dominique Fohr

- End-to-End Speech Separation with Unfolded Iterative Phase Reconstruction .................................................. 2708
- Zhong-Qiu Wang, Jonathan Le Roux, Deliang Wang, John Hershey

- PhaseNet: Discretized Phase Modeling with Deep Neural Networks for Audio Source Separation ...................... 2713
- Naoya Takahashi, Purvi Agrawal, Nabarun Goswami, Yuki Mitsufuji

- Integrating Spectral and Spatial Features for Multi-Channel Speaker Separation ........................................... 2718
- Zhong-Qiu Wang, Deliang Wang

- DNN Driven Speaker Independent Audio-Visual Mask Estimation for Speech Separation .................................. 2723
- Mandar Gogate, Ahsan Aseed, Ricard Marxer, Jon Barker, Amir Hussain

### Dialectal Variation

- Exploring Temporal Reduction in Dialectal Spanish: A Large-scale Study of Lenition of Voiced Stops and Coda-s .................. 2728
- Ioana Varliscu, Nidia Hernandez, Bianca Vieru, Lori Lamel

- Dialect-geographical Acoustic-Tonetics: Five Disyllabic Tone Sandhi Patterns in Cognate Words from the Wu Dialects of ZheJiang Province .......................................................... 2733
- Puch Rosé

- Regional Variation of /r/ in Swiss German Dialects .................................................................................. 2738
- Adrian Leemann, Stephan Schmid, Dieter Studer-Joho, Marie-José Kolly

- Variation in the FACE Vowel Across West Yorkshire: Implications for Forensic Speaker Comparisons .................. 2743
- Kate Earnshaw, Erica Gold

- The ‘West Yorkshire Regional English Database’: Investigations into the Generalizability of Reference Populations for Forensic Speaker Comparison Casework ........................................... 2748
- Kate Earnshaw, Erica Gold

- Studying Vowel Variation in French-Algerian Arabic Code-switched Speech ............................................. 2753
- Jane Wostawa, Djezdigiga Amazouz, Martine Adda-Decker, Lori Lamel

### Spoken Corpora and Annotation

- Fearless Steps: Apollo-11 Corpus Advancements for Speech Technologies from Earth to the Moon ......................... 2758

- A Knowledge Driven Structural Segmentation Approach for Play-Talk Classification During Autism Assessment ............ 2763
- Manoj Kumar, Pooja Chebula, So Hyun Kim, Kassandra Martinez, Catherine Lord, Shrikanth Narayanan

- An Open Source Emotional Speech Corpus for Human Robot Interaction Applications ........................................ 2768
- Jesin James, Li Tian, Catherine Inez Watson

- Speech Database and Protocol Validation Using Waveform Entropy .................................................................. 2773
- Ithshak Lapidot, Héctor Delgado, Massimiliano Todisco, Nicholas Evans, Jean-François Bonastre

- A French-Spanish Multimodal Speech Communication Corpus Incorporating Acoustic Data, Facial, Hands and Arms Gestures Information .......................................................... 2778
- Lucas D. Terissi, Gonzalo Sad, Mauricio Cerda, Slim Ouni, Rodrigo Galvez, Juan C. Gómez, Bernard Girau, Nancy Hitzfeld-Kahler

- L2-ARCTIC: A Non-native English Speech Corpus ................................................................................. 2783
- Guanlong Zhao, Sinem Sonsaat, Alif Silpachai, Ivana Lucic, Evgeny Chukharev-Hudilainen, John Levis, Ricardo Gutierrez-Otuna

### The First Dihard Speech Diarization Challenge

- ZCU-NTIS Speaker Diarization System for the Dihard 2018 Challenge .............................................................. 2788
- Zbynek Zajíc, Marie Kunešová, Jan Zelinka, Marek Hraiz

- Speaker Diarization with Enhancing Speech for the First Dihard Challenge ....................................................... 2793
- Lei Sun, Jun Du, Chao Jiang, Xueyang Zhang, Shan He, Bing Yin, Chin-Hui Lee

- BUT System for Dihard Speech Diarization Challenge 2018 .......................................................................... 2798
- Mireia Diez, Federico Lundini, Lukáš Burger, Johan Rohdin, Anna Silnova, Katerína Žmolíková, Ondřej Novotný, Karel Veselý, Ondřej Glembek, Oldřich Pichot, Ludslav Mottern, Pavel Matějka

- Estimation of the Number of Speakers with Variational Bayesian PLDA in the Dihard Diarization Challenge .......... 2803
- Ignacio Viñals, Pablo Gimenéz, Alfonso Ortégá, Antonio Miguel, Eduardo Lledia

- Diarization is Hard: Some Experiences and Lessons Learned for the JHU Team in the Inaugural Dihard Challenge .... 2808
- Gregory Sell, David Snyder, Alan McCree, Daniel Garcia-Romero, Jesis Villalba, Matthew Maciejewski, Vimal Manohar, Najim Dehkordi, Daniel Povey, Sharif Watanabe, Sanjeev Khudanpur

- The Eucoreom Submission to the First Dihard Challenge .................................................................................. 2813
- Jose Patino, Hèctor Delgado, Nicholas Evans
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Discriminative Embedding Learning, Speech Activity and Overlap Detection for the DIHARD Speaker</td>
<td>2818</td>
</tr>
<tr>
<td>Diarization Challenge ........................................................................</td>
<td></td>
</tr>
<tr>
<td>Y. Tang, X. Hu, Y. Li, S. Liu, L. Xu, Y. Ma, W. Xiong, Z. Pu, S. Wang</td>
<td></td>
</tr>
<tr>
<td>Joint Discriminative Embedding Learning, Speech Activity and Overlap Detection for the DIHARD Speaker</td>
<td>2833</td>
</tr>
<tr>
<td>Automatic Speech Recognition ..............................................................</td>
<td></td>
</tr>
<tr>
<td>M. Shalaby, T. Luan, M. A. Al-Mutlaq, J. Xiong, Y. Zhang, L. Zhang, D. Zhang</td>
<td></td>
</tr>
<tr>
<td>H. Li, J. Gu, Y. Li, X. Wang, W. Tian, L. Xue, L. Zhang, Z. Pu, S. Wang</td>
<td></td>
</tr>
<tr>
<td>Joint Discriminative Embedding Learning, Speech Activity and Overlap Detection for the DIHARD Speaker</td>
<td>2848</td>
</tr>
<tr>
<td>Text Analysis, Multilingual Issues and Evaluation in Speech Synthesis</td>
<td></td>
</tr>
<tr>
<td>Multilingual Grapheme-to-Phoneme Conversion with Global Character Vectors</td>
<td>2823</td>
</tr>
<tr>
<td>J. N. Kitani, Y. Shiga, H. Kawasaki ....................................................</td>
<td></td>
</tr>
<tr>
<td>A Hybrid Approach to Grapheme to Phoneme Conversion for Assamese ............</td>
<td>2828</td>
</tr>
<tr>
<td>Somnath Roy, Shakuntala Mahanta .......................................................</td>
<td></td>
</tr>
<tr>
<td>Investigation of Using Disentangled and Interpretable Representations for One-shot Cross-lingual Voice Conversion</td>
<td>2833</td>
</tr>
<tr>
<td>Seyed Hamidreza Mohammadi, Taehwan Kim .............................................</td>
<td></td>
</tr>
<tr>
<td>Using Pupilometry to Measure the Cognitive Load of Synthetic Speech ........</td>
<td>2838</td>
</tr>
<tr>
<td>Avashna Govender, Simon King .............................................................</td>
<td></td>
</tr>
<tr>
<td>Measuring the Cognitive Load of Synthetic Speech Using a Dual Task Paradigm</td>
<td>2843</td>
</tr>
<tr>
<td>Avashna Govender, Simon King .............................................................</td>
<td></td>
</tr>
<tr>
<td>Attentive Sequence-to-Sequence Learning for Diacritic Restoration of Yoruba Language Text</td>
<td>2848</td>
</tr>
<tr>
<td>Iroko Orji ..........................................................</td>
<td></td>
</tr>
<tr>
<td>Gated Convolutional Neural Network for Sentence Matching ......................</td>
<td>2853</td>
</tr>
<tr>
<td>Pezzio Chen, Wu Guo, Zhi Chen, Jian Sun, Lamhau You ...............................</td>
<td></td>
</tr>
<tr>
<td>On Training and Evaluation of Grapheme-to-Phoneme Mappings with Limited Data</td>
<td>2858</td>
</tr>
<tr>
<td>Dravyansh Sharma .....................................................................................</td>
<td></td>
</tr>
<tr>
<td>The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech</td>
<td>2863</td>
</tr>
<tr>
<td>Alice Baird, Emilia Parada-Cabaleiro, Simone Hanke, Felix Burkhardt, Nicholas Cummings, Björn Schuller</td>
<td></td>
</tr>
<tr>
<td>Word Emphasis Prediction for Expressive Text to Speech ........................</td>
<td>2868</td>
</tr>
<tr>
<td>Yossi Mass, Slava Shevchenko, Moran Meromchay, Ron Hoory, Oren Sar Shalom, Guy Lev, David Komolicky</td>
<td></td>
</tr>
<tr>
<td>A Comparison of Speaker-based and Utterance-based Data Selection for Text-to-Speech Synthesis</td>
<td>2873</td>
</tr>
<tr>
<td>Kai-Zhan Lee, Erica Cooper, Julia Hirschberg ........................................</td>
<td></td>
</tr>
<tr>
<td>Data Requirements, Selection and Augmentation for DNN-based Speech Synthesis from Crowdsourced Data</td>
<td>2878</td>
</tr>
<tr>
<td>Markus Toman, Geoffrey S. Melzner, Rapul Patel .....................................</td>
<td></td>
</tr>
<tr>
<td>Neural Network Training Strategies for ASR ............................................</td>
<td></td>
</tr>
<tr>
<td>Lightly Supervised vs. Semi-supervised Training of Acoustic Model on Luxembourg for Low-resource</td>
<td>2883</td>
</tr>
<tr>
<td>Automatic Speech Recognition ................................................................</td>
<td></td>
</tr>
<tr>
<td>Karel Veselý, Carlos Segura, Igor Štěhka, Jordi Luque, Jan Černocký .............</td>
<td></td>
</tr>
<tr>
<td>Inference-Invariant Transformation of Batch Normalization for Domain Adaptation of Acoustic Models</td>
<td>2893</td>
</tr>
<tr>
<td>Masayuki Suzuki, Toshio Nagano, Gakuto Kurata, Samuel Thomas ..................</td>
<td></td>
</tr>
<tr>
<td>Active Learning for LF-MMI Trained Neural Networks in ASR .....................</td>
<td>2898</td>
</tr>
<tr>
<td>Yuhua Long, Hong Ye, Yijie Li, Juean Liang .........................................</td>
<td></td>
</tr>
<tr>
<td>An Investigation of Mixup Training Strategies for Acoustic Models in ASR ....</td>
<td>2903</td>
</tr>
<tr>
<td>Ivan Medennikov, Yuri Khokhlov, Alexei Romanenko, Dmitry Pavlov, Natalia Tomášenková, Maxim Amosov, Andrey Zatvarmovsky</td>
<td></td>
</tr>
<tr>
<td>Comparison of Unsupervised Modulation Filter Learning Methods for ASR ..........</td>
<td>2908</td>
</tr>
<tr>
<td>Pervu Agrawal, Striam Ganapathy .......................................................</td>
<td></td>
</tr>
<tr>
<td>Improved Training for Online End-to-end Speech Recognition Systems ........</td>
<td>2913</td>
</tr>
<tr>
<td>Sayoung Kim, Michael Seltzer, Jinyu Li, Rui Zhao ..................................</td>
<td></td>
</tr>
<tr>
<td>Combining Natural Gradient with Hessian Free Methods for Sequence Training</td>
<td>2918</td>
</tr>
<tr>
<td>Adnan Haider, Philip Woodward ..........................................................</td>
<td></td>
</tr>
<tr>
<td>Lattice-free State-level Minimum Bayes Risk Training of Acoustic Models ..........</td>
<td>2923</td>
</tr>
<tr>
<td>Naoyuki Kanda, Yasuke Fujita, Kenji Nagamatsu ......................................</td>
<td></td>
</tr>
<tr>
<td>A Study of Enhancement, Augmentation and Autoencoder Methods for Domain Adaptation in Distant Speech Recognition ..........</td>
<td>2928</td>
</tr>
<tr>
<td>Hao Tang, Wei-Ning Hsu, François Grondin, James Glass ..........................</td>
<td></td>
</tr>
<tr>
<td>Multilingual Deep Neural Network Training Using Cyclical Learning Rate ........</td>
<td>2933</td>
</tr>
<tr>
<td>Andreas Soeborg Kirkedal, Yeon-Jun Kim ............................................</td>
<td></td>
</tr>
<tr>
<td>Application of ASR in Medical Practice .............................................</td>
<td></td>
</tr>
<tr>
<td>Development of the CUHK Dysarthric Speech Recognition System for the UA Speech Corpus</td>
<td>2938</td>
</tr>
<tr>
<td>Jianwei Yu, Xurong Xie, Shansong Liu, Shouhong Hu, Max W. Y. Lam, Xinzi Wu, Ka Hoi Wong, Xunying Liu, Helen Meng</td>
<td></td>
</tr>
<tr>
<td>Automatic Evaluation of Speech Intelligibility Based on I-vectors in the Context of Head and Neck Cancers</td>
<td>2943</td>
</tr>
<tr>
<td>Imed Larridh, Corinne Fredouille, Alain Ghio, Muriel Lalain, Virginie Woisard</td>
<td></td>
</tr>
<tr>
<td>Dysarthric Speech Recognition Using Convolutional LSTM Neural Network ..........</td>
<td>2948</td>
</tr>
<tr>
<td>Myungjong Kim, Beiming Cao, Kwanghoon An, Jun Wang ..............................</td>
<td></td>
</tr>
</tbody>
</table>
Perceptual and Automatic Evaluations of the Intelligibility of Speech Degraded by Noise Induced Hearing Loss Simulation .......................................................................................................................... 2953

Articulatory Features for ASR of Pathological Speech ............................................................................................................................. 2958
Emre Yilmaz, Vikramjit Mitra, Chris Bartels, Horacio Franco

Mining Multimodal Repositories for Speech Affecting Diseases ........................................................................................................... 2963
Joana Correia, Bhiksia Raj, Isabel Francose, Francisco Teixeira

Long Distance Voice Channel Diagnosis Using Deep Neural Networks .................................................................................................. 2968

Speech Recognition for Medical Conversations ........................................................................................................................................ 2972
Chung-Cheng Chiu, Anshuman Tripathi, Katherine Chou, Chris Co, Navdeep Jaitly, Diana Jauzenikare, Anjuli Kamran, Patrick Nguyen, Hasim Suk, Ananth Sankar, Justin Tansuwan, Nathan Wan, Yonghui Wu, Xuedong Zhang

SOURCE AND SUPRA-SEGMENTS

Prosodic Focus Acquisition in French Early Cochlear Implanted Children ........................................................................................................ 2977
Chadi Farah, Stephanie Roman, Maripauola D’Imperio

The Role of Temporal Variation in Narrative Organization .......................................................................................................................... 2982
Nassima Fezza

Interaction Mechanisms between Glottal Source and Vocal Tract in Pitch Glides .................................................................................. 2987
Tiina Martola, Jarmo Malinen

Relating Articulatory Motions in Different Speaking Rates ...................................................................................................................... 2992
Asha Singh, G. Nitisha Meenakshi, Prasanta Kumar Ghosh

Estimation of the Asymmetry Parameter of the Glottal Flow Waveform Using the Electroglostographic Signal ........................................................................................................ 2997
João Cabral

Classification of Disorders in Vocal Folds Using Electroglostographic Signal .................................................................................................. 3002

Automatic Glottis Localization and Segmentation in Stroboscopic Videos Using Deep Neural Network ...................................................... 3007
Achuth Rao Rao, Mahul Krishnamurthy, Pebbili Gopikishore, Veeramanth Priyadarshini, Prasanta Kumar Ghosh

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
Toshiko Ise-Ishikawa, Keiko Ochi, Keikichi Hirose

A Preliminary Study on Tonal Coarticulation in Continuous Speech ........................................................................................................ 3017
Listia Hao, Wei Zhang, Yanlu Xie, Jinsong Zhang

PLENARY TALK-3

Speech and Language Processing for Learning and Wellbeing ........................................................................................................... 3022
Helen Meng

DISTANT ASR

Far-Field Speech Recognition Using Multivariate Autoregressive Models .............................................................................................. 3023
Sriram Ganapathy, Madhumita Harish

Efficient Implementation of the Room Simulator for Training Deep Neural Network Acoustic Models .......................................................... 3028
Chanwoo Kim, Ehsan Varniani, Arun Narayanan, Michel Bacchiani

Stream Attention for Distributed Multi-Microphone Speech Recognition .................................................................................................. 3033
XiaoFei Wang, Raizhi Li, Hynck Hermansky

Recognizing Overlapped Speech in Meetings: A Multichannel Separation Approach Using Neural Networks ........................................................................................................ 3038
Takuya Yoshikawa, Hakan Erdogan, Zhuo Chen, Xiong Xiao, Fil Allewa

Integrating Neural Network Based Beamforming and Weighted Prediction Error Dereverberation .............................................................. 3043
Lukas Drude, Christoph Boedeker, Jahn Heymann, Reinhold Haeb-Umbach, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani

A Probability Weighted Beamformer for Noise Robust ASR .................................................................................................................. 3048
Suliang Bu, Yunxin Zhao, MeiYah Hwang, SinHeng Sun

EXPRESSIVE SPEECH SYNTHESIS

Effects of Dimensional Input on Paralinguistic Information Perceived from Synthesized Dialogue Speech with Neural Network .......................................................................................................................... 3053
Masaki Yokoyama, Tomohiro Nagata, Hiroki Morita

Neural MultiVoice Models for Expressing Novel Personalities in Dialog .................................................................................................. 3057
Shereen Oraby, Lena Reed, Sharath T. S., Shubhangi Tandon, Marilyn Walker

Expressive Speech Synthesis Using Sentiment Embeddings .......................................................................................................................... 3062
Igor Jauk, Jaime Lorenzo-Trueba, Junichi Yamagishi, Antonio Bonafonte

Expressive Speech Synthesis via Modeling Expressions with Variational Autoencoder ........................................................................... 3067
Kei Akuzawa, Yutaka Iwasawa, Yutaka Matsuo
REPRESENTATION LEARNING FOR EMOTION

Bags in Bag: Generating Context-Aware Bags for Tracking Emotions from Speech ................................................. 3082
Jing Han, Zixing Zhang, Maximilian Schmitt, Zhao Ren, Fabien Ringeval, Björn Schuller

An Attention Pooling Based Representation Learning Method for Speech Emotion Recognition .......................... 3087
Pengcheng Li, Yan Song, Ian McLaughlin, Wu Guo, Lirong Dai

Predicting Arousal and Valence from Waveforms and Spectrograms Using Deep Neural Networks .................. 3092
Zietaofan Yang, Julia Hirschberg

Emotion Identification from Raw Speech Signals Using DNNs ..................................................................................... 3097
Mousmita Sarma, Pegah Ghaemmaghami, Daniel Povey, Nagendra Kumar Goel, Kandarpa Kamar Sarma, Najim Dehak

Encoding Individual Acoustic Features Using Dyad-Augmented Deep Variational Representations for Dialog-level Emotion Recognition .................................................................................................................. 3102
Jeng-Lin Li, Chi-Chun Lee

Variational Autoencoders for Learning Latent Representations of Speech Emotion: A Preliminary Study ............ 3107
Siddique Latif, Rajib Rana, Junaid Qadir, Julien Epps

ARTICULATORY INFORMATION, MODELING AND INVERSION

Phoneme-to-Articulatory Mapping Using Bidirectional Gated RNN ........................................................................ 3112
Théo Bisututto-Lervat, Slim Ouni

Tongue Segmentation with Geometrically Constrained Snake Model ................................................................. 3117
Zhuhua Su, Jianguo Wei, Qiang Fang, Jianrong Wang, Kiyoshi Honda

Low Resource Acoustic-to-articulatory Inversion Using Bi-directional Long Short Term Memory .................... 3122
Arafan Iliia, Prasanta Kumar Ghosh

Automatic Visual Augmentation for Concatenation Based Synthesized Articulatory Videos from Real-time MRI Data for Spoken Language Training .............................................................................................................. 3127
S. Chandana, Chiranjeevi Varra, Bhiu Aggarwal, Sanjeev Kumar Mittal, N. K. Kaushubba, K. T. Raveena, Astha Singh, Prasanta Kumar Ghosh

Air-Tissue Boundary Segmentation in Real-Time Magnetic Resonance Imaging Video Using Semantic Segmentation with Fully Convolutional Networks ................................................................................. 3132
C. A. Valliaspan, Renuka Mannem, Prasanta Kumar Ghosh

Noise Robust Acoustic to Articulatory Speech Inversion ............................................................................................. 3137
Nadee Seneviratne, Ganesh Sivaraman, Vikramjit Mitra, Carol Eppy-Wilson

NOVEL PARADIGMS FOR DIRECT SYNTHESIS BASED ON SPEECH-RELATED BIOSIGNALS

Designing a Pneumatic Bionic Voice Prosthesis - A Statistical Approach for Source Excitation Generation ............ 3142
Farzaneh Ahmad, Tomoki Toda

A Neural Model to Predict Parameters for a Generalized Command Response Model of Intonation .................... 3147
Bastian Schnell, Philip N. Garner

Articulation-to-Speech Synthesis Using Articulatory Flesh Point Sensors’ Orientation Information ..................... 3152
Beiming Cao, Myongjong Kim, Jun R. Wang, Han Van Santen, Ted Mau, Jun Wang

Effectiveness of Generative Adversarial Network for Non-Audible Murmurs-to-Whisper Speech Conversion ........ 3157
Neil Shah, Nimesh Shah, Hemant Patil

Investigating Objective Intelligibility in Real-Time EMG-to-Speech Conversion ...................................................... 3162
Lorenz Diener, Tanja Schultz

VOLUME 6

Domain-Adversarial Training for Session Independent EMG-based Speech Recognition ...................................... 3167
Michael Wand, Tanja Schultz, Jürgen Schmidhuber

Multi-Task Learning of Speech Recognition and Speech Synthesis Parameters for Ultrasound-based Silent Speech Interfaces ................................................................................................................................. 3172
László Tóth, Gábor Gosztolya, Tamás Grósz, Alexandra Markó, Tamás Gábor Csapó

LOW RESOURCE SPEECH RECOGNITION CHALLENGE FOR INDIAN LANGUAGES

Transcription Correction for Indian Languages Using Acoustic Signatures ......................................................... 3177
Joona Jprakash, Golda Brunet Rajan, Hema Murthy

BUT System for Low Resource Indian Language ASR .............................................................................................. 3182
Bhargav Pulugundla, Muruli Karthick Baskar, Santosh Kesiraju, Ekaterina Egorova, Martin Karafiat, Lukáš Burget, Jan Černocký
**DEEP ENHANCEMENT**

Shuai Nie, Shen Liang, Bin Liu, Yaping Zhang, Wenya Liu, Jianhua Tan
A Deep Neural Network Based Harmonic Noise Model for Speech Enhancement ................................................................. 3224
Zhiheng Qiu, Hongjiang Yu, Wei-Ping Zhu, Benoît Champagne
A Convolutional Recurrent Neural Network for Real-Time Speech Enhancement ................................................................. 3229
Ke Tan, Deliang Wang
All-Neural Multi-Channel Speech Enhancement .................................................................................................................. 3234
Zhong-Qiu Wang, Deliang Wang
Deep Learning for Acoustic Echo Cancellation in Noisy and Double-Talk Scenarios ............................................................ 3239
Hao Zhang, Deliang Wang
The Conversation: Deep Audio-Visual Speech Enhancement .................................................................................................. 3244
Triantafyllos Aiforos, Joon Son Chung, Andrew Zisserman
Student-Teacher Learning for BLSTM Mask-based Speech Enhancement ................................................................................ 3249
Arwin Shannumagam Satrnamanian, Szu-Jui Chen, Shinji Watanabe
Speech Enhancement Using Deep Mixture of Experts Based on Hard Expectation Maximization .............................................. 3254
Pavan Kalyal, Prasanta Kumar Ghosh
Adversarial Feature-Mapping for Speech Enhancement ......................................................................................................... 3259
Zhong Meng, Jinyu Li, Yifan Gong, Biing-Hwang (Fred) Juang
Biophysiologically-Inspired Features Improve the Generalizability of Neural Network-based Speech Enhancement Systems .......................................................... 3264
Deepak Baby, Sarah Verhulst
Error Modeling via Asymmetric Laplace Distribution for Deep Neural Network Based Single-Channel Speech Enhancement .................................................................................................................. 3269
Li Chai, Jun Du, Chinh-Hai Lee
A Priori SNR Estimation Based on a Recurrent Neural Network for Robust Speech Enhancement .............................................. 3274
Yangyang Xia, Richard Stern

**ACOUSTIC SCENES AND RARE EVENTS**

Multiple Instance Deep Learning for Weakly Supervised Small-Footprint Audio Event Detection .................................................. 3279
Shao-Yen Tseng, Juncheng Li, Yun Wang, Florian Metze, Joseph Searley, Samaajit Das
Unsupervised Temporal Feature Learning Based on Sparse Coding Embedded BoAW for Acoustic Event Recognition .......................................................... 3284
Lisen Zhang, Jinqing Han, Shiw’en Deng
Data Independent Sequence Augmentation Method for Acoustic Scene Classification ................................................................. 3289
Teng Zhang, Kailai Zhang, Ji Wu
A Compact and Discriminative Feature Based on Auditory Summation Statistics for Acoustic Scene Classification .......................... 3294
Hongwei Song, Jinqing Han, Shiw’en Deng
ASE: Acoustic Scene Embedding Using Deep Archetypal Analysis and GMM ............................................................................. 3299
Pulkit Sharma, Vinayak Abrol, Anshul Thakur
Deep Convolutional Neural Network with Scalogram for Audio Scene Modeling ................................................................. 3304
Jing-Rong Wang, Yuan-Shan Huang, Jia-Xu Wang, Yang-Tao Chen

Time Aggregation Operators for Multi-label Audio Event Detection .................................................................................... 3309
Pankaj Joshi, Dipvijaysingh Gautam, Ganesh Ramakrishnan, Preethi Jyothi

Early Detection of Continuous and Partial Audio Events Using CNN .................................................................................. 3314
Ian McLoughlin, Yan Song, Lam Dang Pham, Ramaswamy Palaniappan, Huy Phan, Yue Lang

Robust Acoustic Event Classification Using Bag-of-Visual-Words ....................................................................................... 3319
Manjunath Mulimani, Shashidhar G Koolagudi

Wavelet Transform Based Mel-scaled Features for Acoustic Scene Classification ................................................................. 3323
Shefali Waldekar, Gautam Saha

Multi-modal Attention Mechanisms in LSTM and Its Application to Acoustic Scene Classification ........................................... 3328
Teng Zhang, Kai-lai Zhang, Ji Wu

LANGUAGE MODELING

Contextual Language Model Adaptation for Conversational Agents ......................................................................................... 3333
Anirudh Raju, Behnam Hedayatnia, Linda Liu, Ankur Gandhe, Chandra Khatri, Angeliki Metallinou, Anu Venkatesh, Ariya Rastrow

Active Memory Networks for Language Modeling ..................................................................................................................... 3338
Oscar Chen, Jatin Bag, Mark Gales, Xie Chen

Unsupervised and Efficient Vocabulary Expansion for Recurrent Neural Network Language Models in ASR .............................. 3343
Yerbolat Khasanov, Eng Siong Chng

Improving Language Modeling with an Adversarial Critic for Automatic Speech Recognition ............................................... 3348
Yike Zhang, Pengyuan Zhang, Yonghong Yan

Training Recurrent Neural Network through Moment Matching for NLP Applications ............................................................ 3353
Yue Deng, Yixin Shen, Kaiwei Chen, Hongxia Jin

Investigation on LSTM Recurrent N-gram Language Models for Speech Recognition ............................................................. 3358
Zehan Tusing, Ralf Schlüter, Hermann Ney

Online Incremental Learning for Speaker-Adaptive Language Models .................................................................................. 3363
Chih Chi Hu, Bing Liu, John Shen, Ian Lane

Efficient Language Model Adaptation with Noise Contrastive Estimation and Kullback-Leibler Regularization ......................... 3368
Jesús Andrés-Ferrer, Nathan Bodenstab, Paul Voice

Recurrent Neural Network Language Model Adaptation for Conversational Speech Recognition .................................................. 3373
Ke Li, Haiman Xu, Yiming Wang, Daniel Procy, Sanjeev Khudanpur

What to Expect from Expected Kneser-Ney Smoothing ................................................................................................................ 3378
Michael Levit, Sarangarajan Partanasarathe, Shuangyu Chang

i-Vectors in Language Modeling: An Efficient Way of Domain Adaptation for Feed-Forward Models ....................................... 3383
Karel Beneš, Santosh Kesiraju, Lukas Burger

SPEECH PATHOLOGY, DEPRESSION, AND MEDICAL APPLICATIONS

How Did You like 2017? Detection of Language Markers of Depression and Narcissism in Personal Narratives .......................... 3388
Eva-Maria Rathi, Julia Djamali, Yannik Terhorst, Bjorn Schuller, Nicholas Cummins, Gerard Salamon, Christina Hunger-Schoppe, Harald Baumeister

Depression Detection from Short Utterances via Diverse Smartphones in Natural Environmental Conditions .......................... 3393
Zhaocheng Huang, Julien Egns, Dale Joachim, Michael Chen

Multi-Lingual Depression-Level Assessment from Conversational Speech Using Acoustic and Text Features ............................. 3398
Yasin Özkancan, Cenk Demiroglu, Atili Besirli, Selim Celik

Dysarthric Speech Classification Using Glottal Features Computed from Non-words, Words and Sentences ................................ 3403
N. P. Narendra, Piavio Akla

Identifying Schizophrenia Based on Temporal Parameters in Spontaneous Speech ............................................................... 3408
Gabor Gezsohdy, Anit Bagi, Szilvia Szalóki, Irszván Szendi, Idoló Hoffmann

Using Prosodic and Lexical Information for Learning Utterance-level Behaviors in Psychotherapy ........................................ 3413
Karan Singla, Zhuxiao Chen, Nikolaos Flemotomos, James Gibson, Dogan Can, David Atkins, Shrikanto Narayanan

Automatic Speech Assessment for People with Aphasia Using TDNN-BLSTM with Multi-Task Learning ............................... 3418
Ying Qin, Tan Lee, Siyuan Feng, Anthony Pak Hin Kong

Towards an Unsupervised Entrainment Distance in Conversational Speech Using Deep Neural Networks ............................... 3423
Md Nasir, Brian Bancon, Shrikant Narayanan, Panayiotis Georgiou

Patient Privacy in Paralinguistic Tasks .................................................................................................................................. 3428
Francisco Teixeira, Alberto Abad, Isabel Trancoso

A Lightly Supervised Approach to Detect Stuttering in Children's Speech .............................................................. 3433
Shefali Waldekar, Gautam Saha

Learning Conditional Acoustic Latent Representation with Gender and Age Attributes for Automatic Pain Level Recognition ...................................................................................................................... 3438
Jeng-Lin Li, Yi-Ming Weng, Chip-Jin Yg, Chi-Chun Lee
PERSPECTIVE TALK-4

Speaker and Language Recognition -- From Laboratory Technologies to the Wild .......................................................... 3443
Sriram Ganapathy

SPOKEN LANGUAGE UNDERSTANDING

A Deep Reinforcement Learning Based Multimodal Coaching Model (DCM) for Slot Filling in Spoken Language Understanding (SLU) ........................................................................................................ 3444
Yu Wang, Abhishek Patel, Yilin Shen, Hongxia Jin

Is ATIS Too Shallow to Go Deeper for Benchmarking Spoken Language Understanding Models? ........................................... 3449
Frédéric Béchet, Christian Raymond

Robust Spoken Language Understanding via Paraphrasing ........................................................................................................... 3454
Avik Ray, Yilin Shen, Hongxia Jin

Spoken SQuAD: A Study of Mitigating the Impact of Speech Recognition Errors on Listening Comprehension ................ 3459
Chia-Hsuan Lee, Zuo-Lin Wu, Chi-Liang Liu, Hung-Yi Lee

User Information Augmented Semantic Frame Parsing Using Progressive Neural Networks ................................................. 3464
Yilin Shen, Xiayang Zeng, Yu Wang, Hongxia Jin

An Efficient Approach to Encoding Context for Spoken Language Understanding ................................................................. 3469
Raghur Gupta, Abhinav Rastogi, Dilek Hakkani-Tür

SOURCE SEPARATION FROM MONOaurAL INPUT

Deep Speech Denoising with Vector Space Projections .................................................................................................................. 3474
Jeffrey Hetherly, Paul Gamble, Maria Alejandro Barrion, Cory Stephenson, Karl Ni

A Shifted Delta Coefficient Objective for Monaural Speech Separation Using Multi-task Learning ........................................... 3479
Chenglin Xu, Wei Ruo, Eng Siong Chng, Hai Zhou Li

A Two-Stage Approach to Noisy Cochannel Speech Separation with Gated Residual Networks ................................................. 3484
Ke Tan, Deliang Wang

Monaural Audio Source Separation Using Variational Autoencoders .......................................................................................... 3489
Lasmi Pandey, Anurendra Kamar, Vinay Nambodiri

Towards Automated Single Channel Source Separation Using Neural Networks ................................................................. 3494

Investigations on Data Augmentation and Loss Functions for Deep Learning Based Speech-Background Separation .............................................................................................................................................. 3499
Hakan Erdogan, Tatsuya Yoshioka

MULTIMODAL SYSTEMS

Annotator Trustability-based Cooperative Learning Solutions for Intelligent Audio Analysis ...................................................... 3504
Simone Hanke, Christoph Stemp, Björn Schuller

Semi-supervised Cross-domain Visual Feature Learning for Audio-Visual Broadcast Speech Transcription .......................... 3509
Rongfeng Su, Xuming Liu, Lan Wang

Deep Lip Reading: A Comparison of Models and an Online Application .................................................................................. 3514
Triantafylllos Afouras, Joan Son Chang, Andrew Zisserman

Iterative Learning of Speech Recognition Models for Air Traffic Control .................................................................................. 3519
Ajay Srinivasasamuthy, Petri Motlicek, Mittal Singh, Yousef Ouailil, Matthias Kleinert, Heiko Ehr, Hartmut Helmke

Speaker Adaptive Audio-Visual Fusion for the Open-Vocabulary Section of AVICAR ............................................................... 3524
Ledia Sari, Mark Hasogava-Johnson, S. Kumarar, Georg Stemmer, Krishnakumar N Nair

Multimodal Name Recognition in Live TV Subtitling ......................................................................................................................... 3529
Marek Hrad, Aleš Průšák, Michal Baluš

CODING

Dithered Quantization for Frequency-Domain Speech and Audio Coding .................................................................................. 3533
Tom Bäckström, Johannes Fischer, Sneha Das

Postfiltering with Complex Spectral Correlations for Speech and Audio Coding ................................................................. 3538
Sneha Das, Tom Bäckström

Postfiltering Using Log-Magnitude Spectrum for Speech and Audio Coding ................................................................. 3543
Sneha Das, Tom Bäckström

Temporal Noise Shaping with Comanding ................................................................................................................................. 3548
Arjii Bivas, Per Hedelin, Lars Villemoes, Vinay Mekate

Multi-frame Quantization of LSF Parameters Using a Deep Autoencoder and Pyramid Vector Quantizer ............................. 3553
Yaxing Li, Estede Derb Emira, Shengwu Xiong, Anna Zhu, Pengfei Duan, Yichang Li

Multi-frame Coding of LSF Parameters Using Block-Constrained Trellis Coded Vector Quantization ................................ 3558
Yaxing Li, Shan Xu, Shengwu Xiong, Anna Zhu, Pengfei Duan, Yueming Ding
SPEAKER VERIFICATION USING NEURAL NETWORK METHODS II

Training Utterance-level Embedding Networks for Speaker Identification and Verification ................................................. 3563
Heejoong Park, Sukhyun Cho, Kyubyong Park, Namju Kim, Jonghan Park

Analysis of Complementary Information Sources in the Speaker Embeddings Framework ................................................. 3568
Mahesh Kumar Nandwana, Mitchell McLaren, Diego Castan, Julien Van Hout, Aaron Lawson

Self-Attentive Speaker Embeddings for Text-Independent Speaker Verification ................................................................. 3573
Yingke Zhu, Tom Ko, David Snyder, Brian Mak, Daniel Povey

An Improved Deep Embedding Learning Method for Short Duration Speaker Verification .............................................. 3578
Zhifu Gao, Yan Song, Ian McLaughlin, Wu Guo, Liqong Dai

Avoiding Speaker Overfitting in End-to-End DNNs Using Raw Waveform for Text-Independent Speaker Verification ........... 3583
Jee-Weon Jung, Hee-Soo Heo, Il-Ho Yang, Hy-Im Min, Hs-Jin Yu

Deeply Fused Speaker Embeddings for Text-Independent Speaker Verification ................................................................. 3588
Gautam Bhattacharya, Jahangir Alam, Vishwa Gupta, Patrick Kenny

Employing Phonetic Information in DNN Speaker Embeddings to Improve Speaker Recognition Performance ................ 3593
Md Hafizur Rahman, Ivan Himawan, Mitchell McLaren, Clinton Fookes, Sridha Sridharan

End-to-end Text-dependent Speaker Verification Using Novel Distance Measures ............................................................ 3598
Subhadeep Deb, Srikant Madikeri, Petr Motlicek

Robust Speaker Clustering using Mixtures of von Mises-Fisher Distributions for Naturalistic Audio Streams ....................... 3603
Harschanddra Dubey, Abhishek Sangwan, John H. L. Hansen

Triplet Network with Attention for Speaker Diarization ........................................................................................................ 3608
Huan Song, Megan Willi, Jayaraman J. Thiagarajan, Visar Berisha, Andreas Spanias

I-vector Transformation Using Conditional Generative Adversarial Networks for Short Utterance Speaker Verification ........ 3613
Jiachen Zhang, Nakamura Inoue, Keichi Shinoda

Analysis of Length Normalization in End-to-End Speaker Verification System ................................................................. 3618
Weicheng Cai, Jinlan Chen, Ming Li

Angular Softmax for Short-Duration Text-independent Speaker Verification ................................................................. 3623
Zili Huang, Shuai Wang, Kai Yu

An End-to-End Text-independent Speaker Identification System on Short Utterances ......................................................... 3628
Ruifang Ji, Yinyun Cai, Xi Bo

MTGAN: Speaker Verification through Multitasking Triplet Generative Adversarial Networks ............................................. 3633
Wenhao Ding, Liang He

EMOTION RECOGNITION AND ANALYSIS

Categorical vs Dimensional Perception of Italian Emotional Speech ...................................................................................... 3638
Emilia Passaro-Cabaleiro, Giovanni Costantini, Anton Batliner, Alice Baird, Björn Schuller

A Three-Layer Emotion Perception Model for Valence and Arousal-Based Detection from Multilingual Speech .................. 3643
Xifeng Li, Masato Akagi

Cross-lingual Speech Emotion Recognition through Factor Analysis ................................................................................. 3648
Breche Dechamps, Kris Demuynck

Modeling Self-Reported and Observed Affect from Speech .................................................................................................. 3653
Jian Cheng, Jared Bernstein, Elizabeth Rosenfeld, Peter W. Foltz, Alex S. Cohen, Terje B. Holmlund, Brita Elvevag

Stochastic Shake-Shake Regularization for Affective Learning from Speech ........................................................................ 3658
Che-Wei Huang, Shrikant Narayanan

Investigating Speech Enhancement and Perceptual Quality for Speech Emotion Recognition .............................................. 3663
Anderson R. Avila, Jahangir Alam, Douglas O'Shaughnessy, Tiago Falk

Demonstrating and Modelling Systematic Time-varying Annotator Disagreement in Continuous Emotion Annotation .............. 3668
Mia Atcheson, Vidhyasaharan Sethu, Julien Epps

Speech Emotion Recognition from Variable-Length Inputs with Triplet Loss Function ......................................................... 3673
Jian Huang, Yi Li, Jianhua Tao, Zhen Lian

Imbalance Learning-based Framework for Fear Recognition in the MediaEval Emotional Impact of Movies Task ................ 3678
Xiaotong Zhang, Xingfeng Cheng, Mingxing Xu, Thomas Fang Zheng

Emotion Recognition from Variable-Length Speech Segments Using Deep Learning on Spectrograms ............................... 3683
Xi Ma, Zhiyoug Wu, Jia Jia, Mingxing Xu, Helen Meng, Lianghong Cai

Speech Emotion Recognition Using Spectrogram & Phoneme Embedding ............................................................................. 3688
Promod Venigalla, Abhay Kumar, Suraj Tripathi, Chirag Singh, Silvambharu Kar, Jithendra Vepa

On Enhancing Speech Emotion Recognition Using Generative Adversarial Networks ......................................................... 3693
Saurabh Sahu, Rahul Gupta, Carol Eddy-Wilson

Ladder Networks for Emotion Recognition: Using Unsupervised Auxiliary Tasks to Improve Predictions of Emotional Attributes ................................................................. 3698
Srinivas Parthasarathy, Carlos Baso
ACOUSTIC MODELLING

Knowledge Distillation for Sequence Model ........................................................................................................... 3703
Mingkun Huang, Yongbin You, Zhehuai Chen, Yanmin Qian, Kai Yu

Improving CTC-based Acoustic Model with Very Deep Residual Time-delay Neural Networks ............................................. 3708
Sheng Li, Xugang Lu, Ryoichitaka Takahata, Peng Shen, Tatsuya Kawahara, Hisashi Kawai

Juax Guo, Ning Xu, Xin Chen, Yang Shi, Kuiyuan Xu, Abeer Alwan

Twin Regularization for Online Speech Recognition .................................................................................................. 3718
Mirok Ravanelli, Dmitriy Serdyuk, Yoshua Bengio

Self-Attentional Acoustic Models ........................................................................................................................................ 3723
Matthias Sperber, Jan Niehues, Graham Neubig, Sebastian Stüber, Alex Watbel

Hierarchical Recurrent Neural Networks for Acoustic Modeling ..................................................................................... 3728
Jinhwan Park, Iksoo Choi, Yoonho Boo, Wonyong Sung

Dictionary Augmented Sequence-to-Sequence Neural Network for Grapheme to Phoneme Prediction .............................................. 3733
Antoine Broquier, Anton Bakhtin, Dravyansh Sharma

Leveraging Second-Order Log-Linear Model for Improved Deep Learning Based ASR Performance .............................................. 3738
Ankit Raj, Shakti P. Rath, Jithendra Vepa

Semi-Orthogonal Low-Rank Matrix Factorization for Deep Neural Networks ........................................................................... 3743
Daorong Liu, Kaun-Yu Chen, Hung-Yi Lee, Lin-Shan Lee

Completely Unsupervised Phoneme Recognition by Adversarially Learning Mapping Relationships from Audio Embeddings ................................................................................................................. 3748

Phone Recognition Using a Non-Linear Manifold with Broad Phone Class Dependent DNNs .................................................. 3753
Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, Richard Socher

A Multi-Discriminator CycleGAN for Unsupervised Non-Parallel Speech Domain Adaptation ..................................................... 3758

SPEECH AND SPEAKER PERCEPTION

Interactions between Vowels and Nasal Codas in Mandarin Speakers’ Perception of Nasal Finals ......................................................... 3763
Chong Cao, Wei Wei, Wei Wang, Yandu Xie, Jinsong Zhang

Weighting Pitch Contour and Loudness Contour in Mandarin Tone Perception in Cochlear Implant Listeners ............................................. 3768
Qinglin Meng, Nengheng Zheng, Amitka Prasad Mishra, Jacinta Dan Luo, Jan W. H. Schnapp

Implementing DIANA to Model Isolated Auditory Word Recognition in English ........................................................................... 3772
Filip Nenadic, Louis Ten Bosch, Benjamin V. Tucker

Effects of Homophone Density on Spoken Word Recognition in Mandarin Chinese .............................................................. 3777
Bhamini Sharma

Visual Timing Information in Audiovisual Speech Perception: Evidence from Lexical Tone Contour .................................................. 3781
Hai Xie, Biao Zeng, Rui Wang

COSMO SylPhon: A Bayesian Perceptuo-motor Model to Assess Phonological Learning ......................................................... 3786
Marie-Lou Barnaud, Juian Diard, Pierre Bessière, Jean-Luc Schwartz

Experience-dependent Influence of Music and Language on Lexical Pitch Learning Is Not Additive ................................................... 3791
Akshay Raj Maggu, Patrick C. M. Wong, Hanjun Liu, Francis C. K. Wong

Influences of Fundamental Oscillation on Speaker Identification in Vocalic Utterances by Humans and Computers ......................... 3795
Volker Dellwo, Thayabaran Kathiresan, Elisa Pellegrino, Lei He, Sandra Schwab, Dieter Maurer

Author Index