2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP 2018)

Taipei City, Taiwan
26-29 November 2018

IEEE

IEEE Catalog Number: CFP18583-POD
ISBN: 978-1-5386-5628-0

Copyright © 2018 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP18583-POD
ISBN (Print-On-Demand): 978-1-5386-5628-0
ISBN (Online): 978-1-5386-5627-3

Additional Copies of This Publication Are Available From:

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

Table of Contents

A Front-End Speech Enhancement System for Robust Automotive Speech Recognition **P. 1**
Haikun WANG, Zhongfu YE, Jingdong CHEN

Speech Enhancement using Convolutional Neural Network with Skip Connections **P. 6**
Yupeng SHI, Weicong RONG, Nengheng ZHENG

A Novel Unified Framework for Speech Enhancement and Bandwidth Extension Based on Jointly Trained Neural Networks **P. 11**
Bin LIU, Jianhua TAO, Yibin ZHENG

Speech Enhancement Based on Reducing the Detail Portion of Speech Spectrograms in Modulation Domain via Discrete Wavelet Transform **P. 16**
Shih-kuang LEE, Syu-Siang WANG, Yu TSAO, Jeih-weih HUNG

Two-Stage Enhancement of Noisy and Reverberant Microphone Array Speech for Automatic Speech Recognition Systems Trained with Only Clean Speech **P. 21**
Quandong WANG, Sicheng WANG, Fengpei GE, Chang Woo HAN, Jaewon LEE, Lianghao GUO, Chin-Hui LEE

Utterance-level Permutation Invariant Training with Discriminative Learning for Single Channel Speech Separation **P. 26**
Cunhang FAN, Bin LIU, Jianhua TAO, Zhengqi WEN, Jiangyan YI, Ye BAI

A Method for Emotional Speech Synthesis Based on Speaker Adaptive Training **P. 31**
Xiaoyong LU, Yanqin LI, Hongwu YANG

Investigation of Stacked Deep Neural Networks and Mixture Density Networks for Acoustic-to-Articulatory Inversion **P. 36**
Xurong XIE, Xunying LIU, Tan LEE, Lan WANG

Emotional Speech Synthesis Based on DNN and PAD Emotional State Model **P. 41**
Weizhao ZHANG, Hongwu YANG, Pengpeng ZHI

Research on Dungan Speech Synthesis Based on Deep Neural Network **P. 46**
LiJia CHEN, Hongwu YANG, Hui WANG

Voice Conversion Based on Cross-Domain Features Using Variational Auto Encoders **P. 51**

Wen-Chin HUANG, Hsin-Te HWANG, Yu-Huai PENG, Yu TSAO, Hsin-Min WANG

Frame Selection in SI-DNN Phonetic Space with WaveNet Vocoder for Voice Conversion without Parallel Training Data **P. 56**

Feng-Long XIE, Frank K. SOONG, Xi WANG, Lei HE, Haifeng LI

Disordered Speech Assessment Using Kullback-Leibler Divergence Features with Multi-Task Acoustic Modeling **P. 61**

Yuanyuan LIU, Ying QIN, Siyuan FENG, Tan LEE, P.-C. CHING

An End-to-End Approach to Automatic Speech Assessment for People with Aphasia **P. 66**

Ying QIN, Tan LEE, Yuzhong WU, Anthony Pak Hin KONG

Non-intrusive Speech Quality Assessment Using Deep Belief Network and Backpropagation Neural Network **P. 71**

Yahui SHAN, Jing WANG, Xiang XIE, Liuchen MENG, Jingming KUANG

A Progressive Deep Learning Approach to Child Speech Separation **P. 76**

Xin WANG, Jun DU, Lei SUN, Qing WANG, Chin-Hui LEE

Convolutional Neural Turing Machine for Speech Separation **P. 81**

Jen-Tzung CHIEN, Kai-Wei TSOU

Multilingual Speech Recognition Training and Adaptation with Language-Specific Gate Units **P. 86**

Danyang LIU, Xinxin WAN, Ji XU, Pengyuan ZHANG

Acquisition of English Tense-lax Vowels by Chinese EFL Learners from Different Dialectal Regions **P. 91**

Yuan JIA, Cuiping LI

An Acoustic Study of English Monophthongs Acquisition by Chinese EFL Learners from Northeast Region **P. 96**

Yuan JIA, Huimin ZHANG

Chinese EFL Learners' Acquisition of English Monophthongs-A Typological Study of Fuzhou, Ningbo, and Beijing **P. 101**

Yuan JIA, Xinyin SUN

An Empirical Study of English Vowels Acquisition of EFL Learners in Tianjin and Zibo **P. 106**
Bin LI, Yuan JIA

A Refined Query-by-Example Approach to Spoken-Term-Detection on ESL Learners' Speech **P. 111**
Jingyong HOU, Wenping HU, Frank K. SOONG, Lei XIE

Improve the Accuracy of Non-native Speech Annotation with a Semi-automatic Approach **P. 116**
Wei WANG, Wei WEI, Yanlu XIE, Minghao GUO, Jinsong ZHANG

Data Augmentation Using Conditional Generative Adversarial Networks for Robust Speech Recognition **P. 121**
Peiyao SHENG, Zhuolin YANG, Hu HU, Tian TAN, Yanmin QIAN

Improving Gated Recurrent Unit Based Acoustic Modeling with Batch Normalization and Enlarged Context **P. 126**
Jie LI, Yahui SHAN, Xiaorui WANG, Yan LI

Gated Module Neural Network for Multilingual Speech Recognition **P. 131**
Yuan-Fu LIAO, Matus PLEVA, Daniel HLADEK, Jan STAS, Peter VISZLAY, Martin LOJKA, Jozef JUHAR

Subspace Based Sequence Discriminative Training of LSTM Acoustic Models with Feed-Forward Layers **P. 136**
Lahiru SAMARAKOON, Brian MAK, Albert Y.S. LAM

WaveNet MH-SRU: Deep and Wide Multiple-history Simple Recurrent Unit for Speech Recognition **P. 141**
Hengguang HUANG, Brian MAK

Hybrid CTC-Attention based End-to-End Speech Recognition Using Subword Units **P. 146**
Zhangyu XIAO, Zhijian OU, Wei CHU, Hui LIN

Combining Phase-based Features for Replay Spoof Detection System **P. 151**
Kantheti SRINIVAS, Rohan Kumar DAS, Hemant A. PATIL

Pitch Synchronized Relative Phase with Peak Error Detection For Noise-robust Speaker Recognition **P. 156**

Meng GE, Longbiao WANG, Seiichi NAKAGAWA, Yuta KAWAKAMI, Jianwu DANG, Xiangang LI

Visual Information Affects Auditory Frequency Discrimination with Random Stimulus Sequences: Evidence from ERPs **P. 161**
Lei WANG, Fei CHEN

Investigation of the Comprehension Process during Silent Reading based on Eye Movements **P. 165**
Di ZHOU, Jinfeng HUANG, Jianwu DANG

A Multi-modal Soft Targets Approach for Pronunciation Erroneous Tendency Detection **P. 170**
Ju LIN, Wei ZHANG, Linxuan WEI, Yanlu XIE, Jinsong ZHANG

A Study on Landmark Verification of Mandarin Alveolar-palatal Consonants **P. 175**
Zhenyu WANG, Qi ZHANG, Shuang ZHENG, Jinsong ZHANG, Yanlu XIE

DNN i-vector based Fishervoice and PLDA SVM Scoring for NIST SRE 2016 **P. 180**
Jinghua ZHONG, Helen MENG

Novel Amplitude Weighted Frequency Modulation Features for Replay Spoof Detection **P. 185**
Madhu R. KAMBLE, Hemant A. PATIL

Angular Softmax Loss for End-to-end Speaker Verification **P. 190**
Yutian LI, Feng GAO, Zhijian OU, Jiasong SUN

Deep Discriminant Analysis for i-vector Based Robust Speaker Recognition **P. 195**
Shuai WANG, Zili HUANG, Yanmin QIAN, Kai YU

Exploring a Unified Attention-Based Pooling Framework for Speaker Verification **P. 200**
Yi LIU, Liang HE, Weiwei LIU, Jia LIU

Generative Adversarial Networks based X-vector Augmentation for Robust Probabilistic Linear Discriminant Analysis in Speaker Verification **P. 205**
Yexin YANG, Shuai WANG, Man SUN, Yanmin QIAN, Kai YU

Emphasis Detection for Voice Dialogue Applications Using Multi-channel Convolutional Bidirectional Long Short-Term Memory Network **P. 210**

Long ZHANG, Jia JIA, Fanbo MENG, Suping ZHOU, Wei CHEN, Cunjun ZHANG, Runnan LI

Topic and Prosody Interaction in Chinese Discourse **P. 215**
Yuan JIA, Xiaoxiao MA

Measuring Prosodic Transfer in Vector Space by Weighted Tonal Events **P. 220**
Xuanda CHEN, Yuan JIA, Ziyu XIONG

An ERP Study to Evaluate the Quality of Speech Processed by Wiener Filtering **P. 225**
Fang YU, Chin-Tuan TAN, Fei CHEN

Estimation of glottal source waveforms and vocal tract shapes from speech signals based on ARX-LF model **P. 230**
Yongwei LI, Ken-Ichi SAKAKIBARA, Masato AKAGI

The DKU-JNU-EMA Electromagnetic Articulography Database on Mandarin and Chinese Dialects with Tandem Feature based Acoustic-to-Articulatory Inversion **P. 235**
Zexin CAI, Xiaoyi QIN, Danwei CAI, Ming LI, Xinzhong LIU, Haibin ZHONG

GTDNN-Based Voice Conversion Using DAEs with Binary Distributed Hidden Units **P. 240**
Yi-Yang DING, Ya-Jun HU, Zhen-Hua LING

Unsupervised query by example spoken term detection using features concatenated with Self-Organizing Map distances **P. 245**
Haiwei WU, Ming LI, Zexin CAI, Haibin ZHONG

Multi-Head Attention for End-to-End Neural Machine Translation **P. 250**
Ivan FUNG, Brian MAK

Unusable Spoken Response Detection with BLSTM Neural Networks **P. 255**
Zhaoheng NI, Rutuja UBALE, Yao QIAN, Michael MANDEL, Su-Youn YOON, Abhinav MISRA, David SUENDERMANN-OEFT

Speech Super-Resolution Using Parallel WaveNet **P. 260**
Mu WANG, Zhiyong WU, Shiyin KANG, Xixin WU, Jia JIA, Dan SU, Dong YU, Helen MENG

Speech Emotion Recognition Using Convolutional Neural Network with Audio Word-based Embedding **P. 265**

Kun-Yi HUANG, Chung-Hsien WU, Qian-Bei HONG, Ming-Hsiang SU, Yuan-Rong ZENG

Formosa Speech Recognition Challenge 2018: Data, Plan and Baselines **P. 270**
Yuan-Fu LIAO, Wu-Hua HSU, Yu-Chen LIN, Yung-Hsiang Shawn CHANG, Matus PLEVA, Jozef JUHAR, Guang-Feng DENG

CLMAD: A Chinese Language Model Adaptation Dataset **P. 275**
Ye BAI, Jianhua TAO, Jiangyan YI, Zhengqi WEN, Cunhang FAN

From Speech Signals to Semantics – Tagging Performance at Acoustic, Phonetic and Word Levels **P. 280**
Yao QIAN, Rutuja UBALE, Patrick LANGE, Keelan EVANINI, Frank K. SOONG

Using Dempster-Shafer Evidence Theory for Dialog State Tracking **P. 285**
Minglu LIU, Miao LI, Ji WU, Xiangling FU, Ji GAO

Prediction of Voice Disorder Severity: Contributions from Sustained Vowels and Continuous Speech **P. 290**
Yuanyuan LIU, Tan LEE, Thomas LAW, Kathy LEE, P.C. CHING

A Maximum Likelihood Approach to Masking-based Speech Enhancement Using Deep Neural Network **P. 295**
Qing WANG, Jun DU, Li CHAI, Li-Rong DAI, Chin-Hui LEE

Manifold-based Incremental Community Detection Method for Online Speaker Identification **P. 300**
Hongcui WANG, Dongxiao HE, Jianwu DANG, Xi LIANG

Max Margin Cosine Loss for Speaker Identification on Short Utterances **P. 304**
Ruifang JI, Junhua CAO, Xinyuan CAI, Bo XU

Automatic Personality Perception from Speech in Mandarin **P. 309**
Minxian ZHU, Xiang XIE, Liqiang ZHANG, Jing WANG

Text-dependent Speaker Verification Using Word-based Scoring **P. 314**
Shengyu YAO, Houjun HUANG, Ruohua ZHOU, Yonghong YAN

End-to-end Language Identification Using NetFV and NetVLAD **P. 319**
Jinkun CHEN, Weicheng CAI, Danwei CAI, Zexin CAI, Haibin ZHONG, Ming LI

Robust Front-End Processing for Emotion Recognition in Noisy Speech **P. 324**
Meghna PANDHARIPANDE, Rupayan CHAKRABORTY, Ashish PANDA, Sunil Kumar KOPPARAPU

Replay Attacks Detection Using Phase and Magnitude Features with Various Frequency Resolutions **P. 329**
Meng LIU, Longbiao WANG, Zeyan OO, Jianwu DANG, Dongbo LI, Seiichi NAKAGAWA

Novel Demodulation-Based Features Using Classifier-level Fusion of GMM and CNN for Replay Detection **P. 334**
Madhu R. KAMBLE, Hemlata TAK, Maddala V. Siva KRISHNA, Hemant A. PATIL

Chinese Causal Relation: Conjunction, Order and Focus-to-Stress Assignment **P. 339**
Liang ZHANG, Aijun LI, Yingyi LUO

Parallel Double Audio Fingerprinting **P. 344**
Tianyu LIANG, Xianhong CHEN, Can XU, Liang HE

LSTM-Based Pitch Range Estimation from Spectral Information of Brief Speech Input **P. 349**
Wei ZHANG, Qi ZHANG, Yanlu XIE, Jinsong ZHANG

Acoustic and Kinematic Examination of Dysarthria in Cantonese patients of Parkinson's disease **P. 354**
Yue SUN, Manwa L. NG, Chongyuan LIAN, Lan WANG, Feng YANG, Nan YAN

Enhanced Denoising Auto-Encoder for Robust Speech Recognition in Unseen Noise Conditions **P. 359**
Sonal JOSHI, Ashish PANDA, Biswajit DAS

Bidirectional LSTM with Extended Input Context **P. 364**
Gaofeng CHENG, Lu HUANG, Jiasong SUN, Yonghong YAN

A Comparable Study of Modeling Units for End-to-End Mandarin Speech Recognition **P. 369**
Wei ZOU, Dongwei JIANG, Shuaijiang ZHAO, Guilin YANG, Xiangang LI

Keyword Spotting Based on CTC and RNN for Mandarin Chinese Speech **P. 374**
Yiyan WANG, Yanhua LONG

Space-Time Residual LSTM Architechture for Distant Speech Recognition **P. 379**

Long WU, Li WANG, Pengyuan ZHANG, Ta LI, Yonghong YAN

An Analysis of Decoding for Attention-based End-to-end Mandarin Speech Recognition **P. 384**
Dongwei JIANG, Wei ZOU, Shuaijiang ZHAO, Guilin YANG, Xiangang LI

A Study on Acoustic Modeling for Child Speech Based on Multi-Task Learning **P. 389**
Jiarui WANG, Si Ioi NG, Dehua TAO, Wing Yee NG, Tan LEE

Distant-talking Speech Recognition Based on Multi-objective Learning using Phase and Magnitude-based Feature **P. 394**
Dongbo LI, Longbiao WANG, Jianwu DANG, Meng GE, Haotian GUAN

Speech Enhancement Based on A New Architecture of Wasserstein Generative Adversarial Networks **P. 399**
ShuaiShuai YE, Ting JIANG, Shan QIN, Weixia ZOU, Chengyun DENG

An Investigation of Transfer Learning Mechanism for Acoustic Scene Classification **P. 404**
Hengshun ZHOU, Xue BAI, Jun DU

Microphone Array Acoustic Source Localization System Based on Deep Learning **P. 409**
Junhao DING, Bin REN, Nengheng ZHENG

Evaluating Modeling Units and Sub-word Features in Language Models for Turkish ASR **P. 414**
Chang LIU, Yike ZHANG, Pengyuan ZHANG, Yaofeng WANG

Chinese Poetry Generation with Flexible Styles **P. 419**
Jiyuan ZHANG, Dong WANG

Perceivable Information Structure in Discourse Prosody–Detecting Prominent Prosodic Words in Spoken Discourse Using F0 Contour **P. 424**
Chao-yu SU, Chiu-yu TSENG

Declination and boundary effect in Cantonese declarative sentence **P. 429**
Chunyu GE, Aijun LI

Interaction of Syntax, Semantics and Pragmatics on Discourse Prosody in Standard Chinese **P. 434**
Xinyi WEN, Yuan JIA, Aijun LI

A Preliminary Study on Quantitative Calculation of Prosodic Strength in Mandarin Speech **P. 439**
Wei ZHANG, Yanlu XIE, Jinsong ZHANG

L2 Mispronunciation Verification Based on Acoustic Phone Embedding and Siamese Networks **P. 444**
Zhenyu WANG, Jinsong ZHANG, Yanlu XIE

Comparing Mandarin Lexical Stress Produced by Native Speakers and L2 Learners in Hong Kong **P. 449**
Lei LIU, Xuemei ZHAI, Wentao GU

A study on the pitch realization of focus in Chinese **P. 454**
Ziyu XIONG, Maolin WANG

Effect of Anticipatory Vowel-to-Vowel Coarticulation at Different Prosodic Boundaries in Chinese **P. 458**
Ziyu XIONG, Maolin WANG

Co-articulation between Consonant and Vowel in Cantonese and Taiwanese CVC Syllables **P. 463**
Wai-Sum LEE, Yueh-chin CHANG, Feng-fan HSIEH

Cross-Dialectal Perception of the Third-Tone Sandhi in Standard Chinese - Evidence from Eye Movements **P. 468**
Qian LI, Yingyi LUO, Aijun LI

An Acoustic Comparison between Two Pairs of Assimilatory and Dissimilatory Tone Sandhi Processes in Nanjing Mandarin in Categoricalness/Gradience **P. 473**
Xin LI, René KAGER

Response Acts in Chinese Conversation: the Coding Scheme and Analysis **P. 478**
Aijun LI

End-to-End Mongolian Text-to-Speech System **P. 483**
Jingdong LI, Hui ZHANG, Rui LIU, Xueliang ZHANG, Feilong BAO

Syntactic Structure and Communicative Function of Echo Questions in Chinese Dialogues **P. 488**

Gan HUANG, Lin ZHU, Aijun LI

An Automated Assessment Tool for Child Speech Disorders **P. 493**

Si Ioi NG, Dehua TAO, Jiarui WANG, Yi JIANG, Wing Yee NG, Tan LEE

Hearing aids APP design based on deep learning technology **P. 495**

Ji Yan HAN, Wei Zhong ZHENG, Ren Jie HUANG, Tsao YU, Ying-Hui LAI

IOS-based Ear Scale application for Clinical Audiology and Otology Usage **P. 497**

Wen-Huei LIAO, Pei-Chun LI, Shuenn Tsong YOUNG, Ying-Hui LAI, Yu TSAO