

2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP 2021)

**Hong Kong
24 – 27 January 2021**



**IEEE Catalog Number: CFP21583-POD
ISBN: 978-1-7281-6995-8**

**Copyright © 2021 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:	CFP21583-POD
ISBN (Print-On-Demand):	978-1-7281-6995-8
ISBN (Online):	978-1-7281-6994-1

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

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

USABILITY AND PRACTICALITY OF SPEECH RECORDING BY MOBILE PHONES FOR PHONETIC ANALYSIS.....	1
<i>Yihan Guan, Bin Li</i>	
TOWARDS REALIZING SIGN LANGUAGE TO EMOTIONAL SPEECH CONVERSION BY DEEP LEARNING.....	6
<i>Weizhe Wang, Hongwu Yang</i>	
AUTOMATIC EXTRACTION OF SEMANTIC PATTERNS IN DIALOGS USING CONVEX POLYTOPIC MODEL.....	11
<i>Jingyan Zhou, Xiaoying Zhang, Xiaohan Feng, King Keung Wu, Helen Meng</i>	
CONSONANTAL EFFECTS OF ASPIRATION ON ONSET F0 IN CANTONESE	16
<i>Xinran Ren, Peggy Mok</i>	
COMPLEX PATTERNS OF TONAL REALIZATION IN TAIFENG CHINESE.....	21
<i>Xiaoyan Zhang, Aijun Li, Zhiqiang Li</i>	
PROSODIC PROFILES OF THE MANDARIN SPEECH CONVEYING IRONIC COMPLIMENT.....	26
<i>Shanpeng Li, Wentao Gu</i>	
GAN-BASED INTER-CHANNEL AMPLITUDE RATIO DECODING IN MULTI-CHANNEL SPEECH CODING.....	31
<i>Jinru Zhu, Changchun Bao</i>	
AGE-RELATED DECLINE OF CLASSIFIER USAGE IN SOUTHWESTERN MANDARIN.....	36
<i>Yun Feng, Yan Feng, Chenwei Xie, William Shi-Yuan Wang</i>	
AGE-INVARIANT SPEAKER EMBEDDING FOR DIARIZATION OF COGNITIVE ASSESSMENTS	41
<i>Sean Shensheng Xu, Man-Wai Mak, Ka Ho Wong, Helen Meng, Timothy C. Y. Kwok</i>	
LEVERAGING TEXT DATA USING HYBRID TRANSFORMER-LSTM BASED END-TO-END ASR IN TRANSFER LEARNING	46
<i>Zhiping Zeng, Van Tung Pham, Haihua Xu, Yerbolat Khassanov, Eng Siong Chng, Chongjia Ni, Bin Ma</i>	
APPROACHES TO IMPROVING RECOGNITION OF UNDERREPRESENTED NAMED ENTITIES IN HYBRID ASR SYSTEMS	51
<i>Tingzhi Mao, Yerbolat Khassanov, Van Tung Pham, Haihua Xu, Hao Huang, Eng Siong Chng</i>	
PRODUCTION OF TONE 3 SANDHI BY ADVANCED KOREAN LEARNERS OF MANDARIN.....	56
<i>Xin Li, Yin Huang, Yunheng Xu, Linxin Yi, Yuming Yuan, Min Xiang</i>	
SPEAKER CHARISMA ANALYZED THROUGH THE CULTURAL LENS	61
<i>Anna Gutnyk, Oliver Niebuhr, Wentao Gu</i>	
EFFECTS OF MANDARIN TONES ON ACOUSTIC CUE WEIGHTING PATTERNS FOR PROMINENCE.....	66
<i>Wei Zhang, Meghan Clayards, Jinsong Zhang</i>	

THE ACOUSTIC CORRELATES AND TIME SPAN OF THE NON-MODAL PHONATION IN KUNSHAN WU CHINESE	71
<i>Wenwei Xu, Peggy Mok</i>	
ACOUSTICAL CHARACTERISTICS OF THE CANTONESE VOWELS AND TONES PRODUCED BY HEARING IMPAIRED SPEAKERS	76
<i>Wai-Sum Lee, Irene Ching-Yin Tsoi</i>	
UNET++-BASED MULTI-CHANNEL SPEECH DEREVERBERATION AND DISTANT SPEECH RECOGNITION	81
<i>Tuo Zhao, Yunxin Zhao, Shaojun Wang, Mei Han</i>	
CONTEXT-AWARE RNNLM RESCORING FOR CONVERSATIONAL SPEECH RECOGNITION	86
<i>Kun Wei, Pengcheng Guo, Hang Lv, Zhen Tu, Lei Xie</i>	
LOW-COMPLEXITY POST-PROCESSING METHOD FOR SPEECH ENHANCEMENT	91
<i>Feng Bao, Yuepeng Li, Shidong Shang</i>	
CAPSULE NETWORK BASED END-TO-END SYSTEM FOR DETECTION OF REPLAY ATTACKS	96
<i>Meidan Ouyang, Rohan Kumar Das, Jichen Yang, Haizhou Li</i>	
SAMS-NET: A SLICED ATTENTION-BASED NEURAL NETWORK FOR MUSIC SOURCE SEPARATION	101
<i>Tingle Li, Jiawei Chen, Haowen Hou, Ming Li</i>	
IMPACT OF MISMATCHED SPECTRAL AMPLITUDE LEVELS ON VOWEL IDENTIFICATION IN SIMULATED ELECTRIC-ACOUSTIC HEARING	106
<i>Changjie Pan, Fei Chen</i>	
AN INVESTIGATION OF POSITIONAL ENCODING IN TRANSFORMER-BASED END-TO-END SPEECH RECOGNITION	111
<i>Fengpeng Yue, Tom Ko</i>	
AUDIO CAPTION IN A CAR SETTING WITH A SENTENCE-LEVEL LOSS	116
<i>Xuenan Xu, Heinrich Dinkel, Mengyue Wu, Kai Yu</i>	
RAPID WORD LEARNING OF CHILDREN WITH COCHLEAR IMPLANTS: PHONOLOGICAL STRUCTURE AND MUTUAL EXCLUSIVITY	121
<i>Yu-Chen Hung, Tzu-Hui Lin</i>	
RNN-TRANSDUCER WITH LANGUAGE BIAS FOR END-TO-END MANDARIN-ENGLISH CODE-SWITCHING SPEECH RECOGNITION	125
<i>Shuai Zhang, Jiangyan Yi, Zhengkun Tian, Jianhua Tao, Ye Bai</i>	
IMPROVED END-TO-END DYSARTHIC SPEECH RECOGNITION VIA META-LEARNING BASED MODEL RE-INITIALIZATION	130
<i>Disong Wang, Jianwei Yu, Xixin Wu, Lifa Sun, Xunying Liu, Helen Meng</i>	
UNSUPERVISED CROSS-LINGUAL SPEECH EMOTION RECOGNITION USING DOMAIN ADVERSARIAL NEURAL NETWORK	135
<i>Xiong Cai, Zhiyong Wu, Kuo Zhong, Bin Su, Dongyang Dai, Helen Meng</i>	
PLDA-BASED SPEAKER VERIFICATION IN MULTI-ENROLLMENT SCENARIO USING EXPECTED VECTOR APPROACH	140
<i>Meet Soni, Ashish Panda</i>	

IMPROVES NEURAL ACOUSTIC WORD EMBEDDINGS QUERY BY EXAMPLE SPOKEN TERM DETECTION WITH WAV2VEC PRETRAINING AND CIRCLE LOSS	145
<i>Zhaoqi Li, Long Wu, Ta Li, Yonghong Yan</i>	
A PRACTICAL WAY TO IMPROVE AUTOMATIC PHONETIC SEGMENTATION PERFORMANCE	150
<i>Wenjie Peng, Yingming Gao, Binghuai Lin, Jinsong Zhang</i>	
DEEP TIME DELAY NEURAL NETWORK FOR SPEECH ENHANCEMENT WITH FULL DATA LEARNING	155
<i>Cunhang Fan, Bin Liu, Jianhua Tao, Jiangyan Yi, Zhengqi Wen, Leichao Song</i>	
TRANSFORMER-BASED EMPATHETIC RESPONSE GENERATION USING DIALOGUE SITUATION AND ADVANCED-LEVEL DEFINITION OF EMPATHY	160
<i>Yi-Hsuan Wang, Jia-Hao Hsu, Chung-Hsien Wu, Tsung-Hsien Yang</i>	
ARTICULATORY AND ACOUSTIC FEATURES OF MANDARIN /ɿ/: A PRELIMINARY STUDY.....	165
<i>Shuwen Chen, Peggy Pik Ki Mok</i>	
CHANNEL INTERDEPENDENCE ENHANCED SPEAKER EMBEDDINGS FOR FAR-FIELD SPEAKER VERIFICATION	170
<i>Ling-Jun Zhao, Man-Wai Mak</i>	
NON-PARALLEL SEQUENCE-TO-SEQUENCE VOICE CONVERSION FOR ARBITRARY SPEAKERS	175
<i>Ying Zhang, Hao Che, Xiaorui Wang</i>	
NON-AUTOREGRESSIVE DELIBERATION-ATTENTION BASED END-TO-END ASR.....	180
<i>Changfeng Gao, Gaofeng Cheng, Jun Zhou, Pengyuan Zhang, Yonghong Yan</i>	
A NEW METHOD FOR IMPROVING GENERATIVE ADVERSARIAL NETWORKS IN SPEECH ENHANCEMENT	185
<i>Fan Yang, Junfeng Li, Yonghong Yan</i>	
SPEAKER EMBEDDING AUGMENTATION WITH NOISE DISTRIBUTION MATCHING	190
<i>Xun Gong, Zhengyang Chen, Yexin Yang, Shuai Wang, Lan Wang, Yanmin Qian</i>	
AN ATTENTION-AUGMENTED FULLY CONVOLUTIONAL NEURAL NETWORK FOR MONAURAL SPEECH ENHANCEMENT	195
<i>Zezheng Xu, Ting Jiang, Chao Li, Jiacheng Yu</i>	
ESTIMATING MUTUAL INFORMATION IN PROSODY REPRESENTATION FOR EMOTIONAL PROSODY TRANSFER IN SPEECH SYNTHESIS	200
<i>Guangyan Zhang, Shirong Qiu, Ying Qin, Tan Lee</i>	
HIERARCHICALLY ATTENDING TIME-FREQUENCY AND CHANNEL FEATURES FOR IMPROVING SPEAKER VERIFICATION	205
<i>Chenglong Wang, Jiangyan Yi, Jianhua Tao, Ye Bai, Zhengkun Tian</i>	
BYTESING: A CHINESE SINGING VOICE SYNTHESIS SYSTEM USING DURATION ALLOCATED ENCODER-DECODER ACOUSTIC MODELS AND WAVERNN VOCODERS.....	210
<i>Yu Gu, Xiang Yin, Yonghui Rao, Yuan Wan, Benlai Tang, Yang Zhang, Jitong Chen, Yuxuan Wang, Zejun Ma</i>	

TEXT ENHANCEMENT FOR PARAGRAPH PROCESSING IN END-TO-END CODE-SWITCHING TTS	215
<i>Chunyu Qiang, Jianhua Tao, Ruibo Fu, Zhengqi Wen, Jiangyan Yi, Tao Wang, Shiming Wang</i>	
FREQUENCY-SPECIFIC BRAIN NETWORK DYNAMICS DURING PERCEIVING REAL WORDS AND PSEUDOWORDS	220
<i>Taiyan Guo, Jianwu Dang, Gaoyan Zhang, Bin Zhao, Masashi Unoki</i>	
PHASE SPECTRUM RECOVERY FOR ENHANCING LOW-QUALITY SPEECH CAPTURED BY LASER MICROPHONES	225
<i>Chang Liu, Yang Ai, Zhenhua Ling</i>	
PROSODY AND DIALOGUE ACT: A PERCEPTUAL STUDY ON CHINESE INTERROGATIVES.....	230
<i>Gan Huang, Aijun Li, Sichen Zhang, Liang Zhang</i>	
CONTEXT-DEPENDENT LABEL SMOOTHING REGULARIZATION FOR ATTENTION-BASED END-TO-END CODE-SWITCHING SPEECH RECOGNITION	235
<i>Zheyang Huang, Peng Li, Ji Xu, Pengyuan Zhang, Yonghong Yan</i>	
ORDER-AWARE PAIRWISE INTOXICATION DETECTION.....	240
<i>Meng Ge, Ruixiong Zhang, Wei Zou, Xiangang Li, Cheng Gong, Longbiao Wang, Jianwu Dang</i>	
MULTI-SCALE MODEL FOR MANDARIN TONE RECOGNITION	245
<i>Linkai Peng, Wang Dai, Dengfeng Ke, Jinsong Zhang</i>	
AN EXPERIMENTAL RESEARCH ON TONAL ERRORS IN MONOSYLLABLES OF STANDARD SPOKEN CHINESE LANGUAGE PRODUCED BY UYGHUR LEARNERS.....	250
<i>Qiuyuan Li, Yuan Jia</i>	
A MODEL ENSEMBLE APPROACH FOR SOUND EVENT LOCALIZATION AND DETECTION.....	255
<i>Qing Wang, Huaxin Wu, Zijun Jing, Feng Ma, Yi Fang, Yuxuan Wang, Tairan Chen, Jia Pan, Jun Du, Chin-Hui Lee</i>	
DIALOGUE ACT RECOGNITION USING BRANCH ARCHITECTURE WITH ATTENTION MECHANISM FOR IMBALANCED DATA	260
<i>Mengfei Wu, Longbiao Wang, Yuke Si, Jianwu Dang</i>	
A COMPARISON STUDY ON THE ALIGNMENT OF PROSODIC AND SEMANTIC UNITS AND ITS EFFECTS ON F0 SHIFTING IN L1 AND L2 ENGLISH SPONTANEOUS SPEECH.....	265
<i>Yuqing Zhang, Zhu Li, Jinsong Zhang</i>	
REVISITING THE STATISTICS POOLING LAYER IN DEEP SPEAKER EMBEDDING LEARNING.....	270
<i>Shuai Wang, Yexin Yang, Yanmin Qian, Kai Yu</i>	
AUTOMATIC DETECTION OF WORD-LEVEL READING ERRORS IN NON-NATIVE ENGLISH SPEECH BASED ON ASR OUTPUT	275
<i>Ying Qin, Yao Qian, Anastassia Loukina, Patrick Lange, Abhinav Misra, Keelan Evanini, Tan Lee</i>	
AN EYE-TRACKING STUDY OF TRANSPOSED-LETTER EFFECT IN ENGLISH WORD RECOGNITION BY MANDARIN SPEAKERS.....	280
<i>Huan Lei, Jianwu Dang, Yu Chen</i>	

IMPROVING ATTENTION-BASED END-TO-END ASR BY INCORPORATING AN N-GRAM NEURAL NETWORK.....	285
<i>Junyi Ao, Tom Ko</i>	
ACOUSTIC WORD EMBEDDING SYSTEM FOR CODE-SWITCHING QUERY-BY- EXAMPLE SPOKEN TERM DETECTION	290
<i>Murong Ma, Haiwei Wu, Xuyang Wang, Lin Yang, Junjie Wang, Ming Li</i>	
MOEVC: A MIXTURE OF EXPERTS VOICE CONVERSION SYSTEM WITH SPARSE GATING MECHANISM FOR ONLINE COMPUTATION ACCELERATION.....	295
<i>Yu-Tao Chang, Yuan-Hong Yang, Yu-Huai Peng, Syu-Siang Wang, Tai-Shih Chi, Yu Tsao, Hsin-Min Wang</i>	
SYLLABLE-BASED ACOUSTIC MODELING WITH LATTICE-FREE MMI FOR MANDARIN SPEECH RECOGNITION.....	300
<i>Jie Li, Zhiyun Fan, Xiaorui Wang, Yan Li</i>	
EXPLORING CROSS-LINGUAL SINGING VOICE SYNTHESIS USING SPEECH DATA	305
<i>Yuewen Cao, Songxiang Liu, Shiyin Kang, Na Hu, Peng Liu, Xunying Liu, Dan Su, Dong Yu, Helen Meng</i>	
COMPARING THE RHYTHM OF INSTRUMENTAL MUSIC AND VOCAL MUSIC IN MANDARIN AND ENGLISH	310
<i>Lujia Yang, Hongwei Ding</i>	
AUTOMATIC SPEAKER-LEVEL PRONUNCIATION ASSESSMENT OF L2 SPEECH USING POSTERIOR PROBABILITIES FROM MULTIPLE UTTERANCES	315
<i>Guolei Jiang, Chunhong Liao, Kun Li, Pengfei Liu, Linying Jiang, Helen Meng</i>	
SPOKEN LANGUAGE UNDERSTANDING WITH SEMEME KNOWLEDGE AS DOMAIN KNOWLEDGE	320
<i>Sixia Li, Jianwu Dang, Longbiao Wang</i>	
TOWARDS FINE-GRAINED PROSODY CONTROL FOR VOICE CONVERSION.....	325
<i>Zheng Lian, Rongxiu Zhong, Zhengqi Wen, Bin Liu, Jianhua Tao</i>	
TONE REALIZATION IN MANDARIN SPEECH: A LARGE CORPUS BASED STUDY OF DISYLLABIC WORDS	330
<i>Yaru Wu, Lori Lamel, Martine Adda-Decker</i>	
ON ADAPTIVE LASSO-BASED SPARSE TIME-VARYING COMPLEX AR SPEECH ANALYSIS	335
<i>Keiichi Funaki</i>	
ACCENT AND SPEAKER DISENTANGLEMENT IN MANY-TO-MANY VOICE CONVERSION	340
<i>Zhichao Wang, Wenshuo Ge, Xiong Wang, Shan Yang, Wendong Gan, Haitao Chen, Hai Li, Lei Xie, Xiulin Li</i>	
ADVERSARIAL TRAINING FOR MULTI-DOMAIN SPEAKER RECOGNITION	345
<i>Qing Wang, Wei Rao, Pengcheng Guo, Lei Xie</i>	
SPEECH EMOTION RECOGNITION BASED ON ACOUSTIC SEGMENT MODEL	350
<i>Siyuan Zheng, Jun Du, Hengshun Zhou, Xue Bai, Chin-Hui Lee, Shipeng Li</i>	
CONTROLLABLE EMOTION TRANSFER FOR END-TO-END SPEECH SYNTHESIS.....	355
<i>Tao Li, Shan Yang, Liumeng Xue, Lei Xie</i>	

Author Index