2021 IEEE Spoken Language Technology Workshop (SLT 2021)

Shenzhen, China
19 – 22 January 2021

Pages 1-566

IEEE Catalog Number: CFP21SLT-POD
ISBN: 978-1-7281-7067-1

1/2

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:	CFP21SLT-POD
ISBN (Print-On-Demand):	978-1-7281-7067-1
ISBN (Online):	978-1-7281-7066-4

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

SPEECH RECOGNITION: STREAMING END-TO-END MODELS

1: TRANSFORMER-BASED ONLINE SPEECH RECOGNITION WITH DECODER-END 1
ADAPTIVE COMPUTATION STEPS
Mohan Li, Catalin Zorila, Rama Doddipatla, Toshiba Cambridge Research Laboratory, United Kingdom

2: STREAMING ATTENTION-BASED MODELS WITH AUGMENTED MEMORY FOR 8
END-TO-END SPEECH RECOGNITION
Ching-Feng Yeh, Yongqiang Wang, Yangyang Shi, Chunyang Wu, Frank Zhang, Julian Chan, Michael Seltzer, Facebook, United States

3: CASCADE RNN-TRANSDUCER: SYLLABLE BASED STREAMING ON-DEVICE MANDARIN .15
SPEECH RECOGNITION WITH A SYLLABLE-TO-CHARACTER CONVERTER
Xiong Wang, Zhuoyuan Yao, Xian Shi, Lei Xie, Northwestern Polytechnical University, China

4: STREAMING TRANSFORMER ASR WITH BLOCKWISE SYNCHRONOUS BEAM SEARCH.....22
Emiru Tsunoo, Yosuke Kashiwagi, Sony Corporation, Japan; Shinji Watanabe, Johns Hopkins University, United States

5: CONVOLUTION-BASED ATTENTION MODEL WITH POSITIONAL ENCODING FOR 30
STREAMING SPEECH RECOGNITION ON EMBEDDED DEVICES
Jinhwan Park, Seoul National University, Korea (South); Chanwoo Kim, Samsung Research, Korea (South); Wonyong Sung, Seoul National University, Korea (South)

6: LEARNING TO COUNT WORDS IN FLUENT SPEECH ENABLES ONLINE SPEECH 38
RECOGNITION
George Sterpu, Christian Saam, Naomi Harte, Trinity College Dublin, Ireland

7: BENCHMARKING LF-MMI, CTC AND RNN-T CRITERIA FOR STREAMING ASR......................46
Xiaohui Zhang, Frank Zhang, Chunxi Liu, Kjell Schubert, Julian Chan, Pradyot Prakash, Jun Liu, Ching-feng Yeh, Fuchun Peng, Yatharth Saraf, Geoffrey Zweig, Facebook AI, United States

8: ALIGNMENT RESTRICTED STREAMING RECURRENT NEURAL NETWORK 52
TRANSDUCER
Jay Mahadeokar, Yuan Shangguan, Duc Le, Gil Keren, Hang Su, Thong Le, Ching-feng Yeh, Christian Fuegen, Michael L Seltzer, Facebook, United States

SPEECH RECOGNITION: MODEL ARCHITECTURES

1: EFFICIENT NEURAL ARCHITECTURE SEARCH FOR END-TO-END SPEECH 60
RECOGNITION VIA STRAIGHT-THROUGH GRADIENTS
Huahuan Zheng, Keyu An, Zhijian Ou, Tsinghua University, China

2: TRANSFORMER BASED DELIBERATION FOR TWO-PASS SPEECH RECOGNITION 68
Ke Hu, Ruoming Pang, Tara Sainath, Trevor Strohman, Google, United States

3: SIMPLIFIED SELF-ATTENTION FOR TRANSFORMER-BASED END-TO-END SPEECH 75
RECOGNITION
Haoneng Luo, Northwestern Polytechnical University, China; Shiliang Zhang, Ming Lei, alibaba, China; Lei Xie, Northwestern Polytechnical University, China

4: MULTI-QUARTZNET: MULTI-RESOLUTION CONVOLUTION FOR SPEECH 82
RECOGNITION WITH MULTI-LAYER FEATURE FUSION
Jian Luo, Jianzong Wang, Ning Cheng, Guilin Jiang, Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

5: ON THE USEFULNESS OF SELF-ATTENTION FOR AUTOMATIC SPEECH 89
RECOGNITION WITH TRANSFORMERS
Shucong Zhang, Erfan Loweimi, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom

6: LOW-ACTIVITY SUPERVISED CONVOLUTIONAL SPIKING NEURAL NETWORKS 97
APPLIED TO SPEECH COMMANDS RECOGNITION
Thomas Pellegrini, Université de Toulouse III ; IRIT, France; Romain Zimmer, IRIT; CERCO UMR 5549, CNRS, France; Timothée Masquelier, CERCO UMR 5549, CNRS ; Université de Toulouse III, France

7: MULTI-CHANNEL AUTOMATIC SPEECH RECOGNITION USING DEEP COMPLEX UNET ...104
Yuxiang Kong, Jian Wu, Northwestern Polytechnical University, Xi'an, China, China; Quandong Wang, Peng Gao, Weiji Zhuang, Yujun Wang, Xiaomi Inc., China; Lei Xie, Northwestern Polytechnical University, Xi'an, China, China

8: DYNAMICALLY WEIGHTED ENSEMBLE MODELS FOR AUTOMATIC SPEECH 111
RECOGNITION
Kiran Praveen, Abhishek Pandey, Deepak Kumar, Samsung R&D Institute, India - Bangalore, India; Shakti Prasad RathSandip Bapat, Samsung R&D Institute, India - Bangalore, India

SPEECH RECOGNITION: TRAINING METHODS

1: INVESTIGATION OF NODE PRUNING CRITERIA FOR NEURAL NETWORKS MODEL 117
COMPRESSION WITH NON-LINEAR FUNCTION AND NON-UNIFORM NETWORK TOPOLOGY
Kazuhiro Nakadai, Yosuke Fukumoto, Honda Research Inst. Japan Co., Ltd., Japan; Ryu Takeda, Osaka University, Japan

2: SEMI-SUPERVISED END-TO-END SPEECH RECOGNITION VIA LOCAL PRIOR 125
MATCHING
Wei-Ning Hsu, Ann Lee, Gabriel Synnaeve, Awni Hannun, Facebook AI Research, United States

3: METRIC LEARNING FOR KEYWORD SPOTTING..133
Jaesung Huh, Minjae Lee, Heesoo Heo, Seongkyu Mun, Joon Son Chung, Naver Corporation, Korea (South)

4: DATA-FILTERING METHODS FOR SELF-TRAINING OF AUTOMATIC SPEECH 141
RECOGNITION SYSTEMS
Alexandru-Lucian Georgescu, Cristian Manolache, Dan Oneață, Horia Cucu, Corneliu Burileanu, University Politehnica of Bucharest, Romania

5: EFFICIENT LARGE SCALE SEMI-SUPERVISED LEARNING FOR CTC BASED ACOUSTIC .148
MODELS
Prakhar Swarup, Debmalya Chakrabarty, Ashtosh Sapru, Hitesh Tulsiani, Harish Arsikere, Sri Garimella, Amazon, India

6: TOWARDS UNSUPERVISED LEARNING OF SPEECH FEATURES IN THE WILD 156
Morgane Riviere, Emmanuel Dupoux, Facebook, France

7: WHOLE-WORD SEGMENTAL SPEECH RECOGNITION WITH ACOUSTIC WORD 164
EMBEDDINGS
Bowen Shi, Shane Settle, Karen Livescu, Toyota Technological Institute at Chicago, United States

8: IMPROVING RNN TRANSDUCER BASED ASR WITH AUXILIARY TASKS..................................172
Chunxi Liu, Frank Zhang, Duc Le, Facebook, United States; Suyoun KimYatharth Saraf, Geoffrey Zweig, Facebook, United States

SPEECH RECOGNITION: DATA AND FEATURE AUGMENTATION

1: IMPROVING SPEECH RECOGNITION ACCURACY OF LOCAL POI USING 180
GEOGRAPHICAL MODELS
Songjun Cao, Yike Zhang, Xiaobing Feng, Long Ma, Tencent, China

2: END-TO-END WHISPERED SPEECH RECOGNITION WITH FREQUENCY-WEIGHTED 186
APPROACHES AND PSEUDO WHISPER PRE-TRAINING
Heng-Jui Chang, Alexander H. Liu, Hung-yi Lee, Lin-shan Lee, National Taiwan University, Taiwan

3: DATA AUGMENTATION FOR END-TO-END CODE-SWITCHING SPEECH RECOGNITION....194
Chenpeng Du, Hao Li, Yizhou Lu, Shanghai Jiao Tong University, China; Lan Wang, Shenzhen Institutes of Advanced Technology, China; Yanmin Qian, Shanghai Jiao Tong University, China

4: INCORPORATING DISCRIMINATIVE DPGMM POSTERIORGRAMS FOR 201
LOW-RESOURCE ASR
Bin Wu, Nara Institute of Science and Technology, Japan; Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology; RIKEN, Center for Advanced Intelligence Project AIP, Japan

5: FRAME-LEVEL SPECAUGMENT FOR DEEP CONVOLUTIONAL NEURAL NETWORKS IN .209
HYBRID ASR SYSTEMS
Xinwei Li, Yuanyuan Zhang, Xiaodan Zhuang, Daben Liu, Apple, United States

6: DATA AUGMENTING CONTRASTIVE LEARNING OF SPEECH REPRESENTATIONS IN 215
THE TIME DOMAIN
Eugene Kharitonov, Morgane Rivière, Gabriel Synnaeve, Lior Wolf, Pierre-Emmanuel Mazaré, Matthijs Douze, Facebook, France; Emmanuel Dupoux, Facebook/EHESS, CNRS, INRIA, ENS-PSL University, France

7: DUAL APPLICATION OF SPEECH ENHANCEMENT FOR AUTOMATIC SPEECH 223
RECOGNITION
Ashutosh Pandey, The Ohio State University, United States; Chunxi Liu, Yun Wang, Yatharth Saraf, Facebook, United States

8: TWO-STAGE AUGMENTATION AND ADAPTIVE CTC FUSION FOR IMPROVED 229
ROBUSTNESS OF MULTI-STREAM END-TO-END ASR
Ruizhi Li, Gregory Sell, Hynek Hermansky, Johns Hopkins University, United States

SPEECH RECOGNITION: GENERAL TOPICS

1: BLOCK-ONLINE GUIDED SOURCE SEPARATION 236
Shota Horiguchi, Yusuke Fujita, Kenji Nagamatsu, Hitachi, Ltd., Japan

3: INTERNAL LANGUAGE MODEL ESTIMATION FOR DOMAIN-ADAPTIVE END-TO-END 243
SPEECH RECOGNITION
Zhong Meng, Sarangarajan Parthasarathy, Eric Sun, Yashesh Gaur, Naoyuki Kanda, Liang Lu, Xie Chen, Rui Zhao, Jinyu Li, Yifan Gong, Microsoft Corporation, United States

4: DEEP SHALLOW FUSION FOR RNN-T PERSONALIZATION 251
Duc Le, Gil Keren, Julian Chan, Jay Mahadeokar, Christian Fuegen, Michael Seltzer, Facebook, United States

5: AN EVALUATION OF WORD-LEVEL CONFIDENCE ESTIMATION FOR END-TO-END 258
AUTOMATIC SPEECH RECOGNITION
Dan Oneață, Alexandru Caranica, University POLITEHNICA of Bucharest, Romania; Adriana Stan, Technical University of Cluj-Napoca, Romania; Horia Cucu, University POLITEHNICA of Bucharest, Romania

6: INNOVATIVE BERT-BASED RERANKING LANGUAGE MODELS FOR SPEECH 266
RECOGNITION
Shih-Hsuan Chiu, Berlin Chen, National Taiwan Normal University, Taiwan

7: REED: AN APPROACH TOWARDS QUICKLY BOOTSTRAPPING MULTILINGUAL 272
ACOUSTIC MODELS
Bipasha Sen, Aditya Agarwal, Microsoft, India; Mirishkar Ganesh, Anil Vuppala, IIIT Hyderabad, India

8: WORD SIMILARITY BASED LABEL SMOOTHING IN RNNLM TRAINING FOR ASR 280
Minguang Song, Yunxin Zhao, University of Missouri, United States; Shaojun Wang, Mei Han, PAII Inc, United States

SPEAKER RECOGNITION

1: SUPERVISED ATTENTION FOR SPEAKER RECOGNITION...286
Seong Min Kye, Korea Advanced Institute of Science and Technology, Korea (South); Joon Son Chung, Naver Corporation, Korea (South); Hoirin Kim, Korea Advanced Institute of Science and Technology, Korea (South)

2: CROSS ATTENTIVE POOLING FOR SPEAKER VERIFICATION...294
Seong Min Kye, Korea Advanced Institute of Science and Technology, Korea (South); Yoohwan Kwon, Yonsei university, Korea (South); Joon Son Chung, Naver Corporation, Korea (South)

3: RESNEXT AND RES2NET STRUCTURES FOR SPEAKER VERIFICATION 301
Tianyan Zhou, Yong Zhao, Jian Wu, Microsoft corporation, United States

4: EMBEDDING AGGREGATION FOR FAR-FIELD SPEAKER VERIFICATION WITH 308
DISTRIBUTED MICROPHONE ARRAYS
Danwei Cai, Duke University, United States; Ming Li, Duke Kunshan University, China

5: SYNTH2AUG: CROSS-DOMAIN SPEAKER RECOGNITION WITH TTS SYNTHESIZED 316
SPEECH
Yiling Huang, Google, United States; Yutian Chen, DeepMind, United Kingdom; Jason Pelecanos, Quan Wang, Google, United States

6: UIAI SYSTEM FOR SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2020..........323
Md Sahidullah, Inria, France; Achintya Kumar Sarkar, Indian Institute of Information Technology (IIIT), Sri City, India; Ville Vestman, University of Eastern Finland, Finland; Xuechen Liu, Inria, France; Romain Serizel, LORIA, France; Tomi Kinnunen, University of Eastern Finland, Finland; Zheng-Hua Tan, Aalborg University, Denmark; Emmanuel Vincent, Inria, France

7: MULTI-FEATURE LEARNING WITH CANONICAL CORRELATION ANALYSIS CONSTRAINT . 330
FOR TEXT-INDEPENDENT SPEAKER VERIFICATION
Zheng Li, Miao Zhao, Lin Li, Qingyang Hong, Xiamen University, China

8: IMPROVING SPEAKER RECOGNITION WITH QUALITY INDICATORS....................................338
Hrishikesh Rao, Kedar Phatak, Elie Khoury, Pindrop Security, United States

9: AUDIO ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF AUDIO 344
REPRESENTATION
Po-Han Chi, Pei-Hung Chung, Tsung-Han Wu, Chun-Cheng Hsieh, Yen-Hao Chen, National Taiwan University, Taiwan; Shang-Wen Li, Amazon AI, Taiwan; Hung-yi Lee, National Taiwan University, Taiwan

EMOTION RECOGNITION

1: A CONDITIONAL CYCLE EMOTION GAN FOR CROSS CORPUS SPEECH EMOTION 351
RECOGNITION
Bo-Hao Su, Chi-Chun Lee, National Tsing Hua University, Taiwan

2: INVESTIGATIONS ON AUDIOVISUAL EMOTION RECOGNITION IN NOISY 358
CONDITIONS
Michael Neumann, Ngoc Thang Vu, University of Stuttgart, Germany

3: IMPROVING CONVOLUTIONAL RECURRENT NEURAL NETWORKS FOR SPEECH 365
EMOTION RECOGNITION
Patrick Meyer, Ziyi Xu, Tim Fingscheidt, Technische Universität Braunschweig, Germany

4: ON THE USE OF SELF-SUPERVISED PRE-TRAINED ACOUSTIC AND LINGUISTIC 373
FEATURES FOR CONTINUOUS SPEECH EMOTION RECOGNITION
Manon Macary, Laboratoire d'Informatique de l'Université du Mans & Allo-Media, France; Marie Tahon, Laboratoire d'Informatique de l'Université du Mans, France; Yannick Estève, Laboratoire d'Informatique d'Avignon, France; Anthony Rousseau, Allo-Media, France

5: SELF-SUPERVISED LEARNING WITH CROSS-MODAL TRANSFORMERS FOR EMOTION .381
RECOGNITION
Aparna Khare, Srinivas Parthasarathy, Shiva Sundaram, Amazon, United States

6: DOMAIN GENERALIZATION WITH TRIPLET NETWORK FOR CROSS-CORPUS SPEECH ..389
EMOTION RECOGNITION
Shi-wook Lee, National Institute of Advanced Industrial Science and Technology, Japan, Japan

7: EMOTION RECOGNITION IN PUBLIC SPEAKING SCENARIOS UTILISING AN 397
LSTM-RNN APPROACH WITH ATTENTION
Alice Baird, Shahin Amiriparian, Manuel Milling, Björn Schuller, University of Augsburg, Germany

EXPRESSIVE SPEECH SYNTHESIS AND OTHERS

1: CONVERSATIONAL END-TO-END TTS FOR VOICE AGENTS 403
Haohan Guo, Northwestern Polytechnical University, China; Shaofei Zhang, Microsoft, China; Frank Soong, Microsoft Research Asia, China; Lei He, Microsoft, China; Lei Xie, Northwestern Polytechnical University, China

2: CONTROLLABLE EMPHATIC SPEECH SYNTHESIS BASED ON FORWARD ATTENTION 410
FOR EXPRESSIVE SPEECH SYNTHESIS
Liangqi Liu, Jiankun Hu, Zhiyong Wu, Tsinghua University, China; Song Yang, Songfan Yang, TAL Education Group, China; Jia Jia, Tsinghua University, China; Helen Meng, The Chinese University of Hong Kong, China

3: VAW-GAN FOR DISENTANGLEMENT AND RECOMPOSITION OF EMOTIONAL 415
ELEMENTS IN SPEECH
Kun Zhou, National University of Singapore, Singapore; Berrak Sisman, Singapore University of Technology and Design, Singapore; Haizhou Li, National University of Singapore, Singapore

4: FINE-GRAINED EMOTION STRENGTH TRANSFER, CONTROL AND PREDICTION FOR 423
EMOTIONAL SPEECH SYNTHESIS
Yi Lei, Shan Yang, Lei Xie, Northwestern Polytechnical University, China

5: SUPERVISED AND UNSUPERVISED APPROACHES FOR CONTROLLING NARROW 431
LEXICAL FOCUS IN SEQUENCE-TO-SEQUENCE SPEECH SYNTHESIS
Slava Shechtman, Raul Fernandez, David Haws, IBM Research, Israel

6: GRAPHPB: GRAPHICAL REPRESENTATIONS OF PROSODY BOUNDARY IN SPEECH 438
SYNTHESIS
Aolan Sun, Jianzong Wang, Ning Cheng, Huayi Peng, Zhen Zeng, Lingwei Kong, Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

7: HIERARCHICAL PROSODY MODELING FOR NON-AUTOREGRESSIVE SPEECH 446
SYNTHESIS
Chung-Ming Chien, Hung-yi Lee, National Taiwan University, Taiwan

8: WHISPERED AND LOMBARD NEURAL SPEECH SYNTHESIS 454
Qiong Hu, Apple, United States; Tobias Bleisch, Petko Petkov, Tuomo Raitio, Erik Marchi, Varun Lakshminarasimhan, Apple Inc., United States

9: NEURAL MOS PREDICTION FOR SYNTHESIZED SPEECH USING MULTI-TASK 462
LEARNING WITH SPOOFING DETECTION AND SPOOFING TYPE CLASSIFICATION
Yeunju Choi, Youngmoon Jung, Hoirin Kim, KAIST, Korea (South)

NEURAL VOCODER AND OTHERS

1: IMPROVED PARALLEL WAVEGAN VOCODER WITH PERCEPTUALLY WEIGHTED 470
SPECTROGRAM LOSS
Eunwoo Song, NAVER Corp., Korea (South); Ryuichi Yamamoto, LINE Corp., Japan; Min-Jae Hwang, Search Solutions Inc., Korea (South); Jin-Seob Kim, Ohsung Kwon, Jae-Min Kim, NAVER Corp., Korea (South)

2: DENOISING-AND-DEREVERBERATION HIERARCHICAL NEURAL VOCODER FOR 477
ROBUST WAVEFORM GENERATION
Yang Ai, University of Science and Technology of China, China; Haoyu Li, Xin Wang, Junichi Yamagishi, National Institute of Informatics, Japan; Zhenhua Ling, University of Science and Technology of China, China

3: MELGLOW: EFFICIENT WAVEFORM GENERATIVE NETWORK BASED ON 485
LOCATION-VARIABLE CONVOLUTION
Zhen Zeng, Jianzong Wang, Ning Cheng, Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

4: MULTI-BAND MELGAN: FASTER WAVEFORM GENERATION FOR HIGH-QUALITY 492
TEXT-TO-SPEECH
Geng Yang, Shan Yang, Northwestern Polytechnical University, Xi'an, China, China; Kai Liu, Peng Fang, Wei Chen, Sogou, China; Lei Xie, Northwestern Polytechnical University, China

5: LIGHTSPEECH: LIGHTWEIGHT NON-AUTOREGRESSIVE MULTI-SPEAKER 499
TEXT-TO-SPEECH
Song Li, Beibei Ouyang, Lin Li, Qingyang Hong, Xiamen University (China), China

6: OPTIMIZING VOICE CONVERSION NETWORK WITH CYCLE CONSISTENCY LOSS 507
OF SPEAKER IDENTITY
Hongqiang Du, Northwestern Polytechnical University, China; Xiaohai Tian, National University of Singapore, Singapore; Lei Xie, Northwestern Polytechnical University, China; Haizhou Li, National University of Singapore, Singapore

7: HOW FAR ARE WE FROM ROBUST VOICE CONVERSION: A SURVEY.......................................514
Tzu-hsien Huang, Jheng-hao Lin, Hung-yi Lee, National Taiwan University, Taiwan

8: LEARN2SING: TARGET SPEAKER SINGING VOICE SYNTHESIS BY LEARNING FROM A ..522
SINGING TEACHER
Heyang Xue, Shan Yang, Yi Lei, Lei Xie, Northwestern Polytechnical University, China; Xiulin Li, Databaker (Beijing) Technology Co., Ltd., China

9: UNSUPERVISED ACOUSTIC-TO-ARTICULATORY INVERSION NEURAL NETWORK 530
LEARNING BASED ON DETERMINISTIC POLICY GRADIENT
Hayato Shibata, Mingxin Zhang, Takahiro Shinozaki, Tokyo Institute of Technology, Japan

ANTI-SPOOFING AND DIARIZATION IN SPEAKER RECOGNITION

1: SPOOFPRINT: A NEW PARADIGM FOR SPOOFING ATTACKS DETECTION 538
Tianxiang Chen, Elie Khoury, Pindrop Security Inc., United States

2: DETECTION AND EVALUATION OF HUMAN AND MACHINE GENERATED SPEECH IN 544
SPOOFING ATTACKS ON AUTOMATIC SPEAKER VERIFICATION SYSTEMS
Yang Gao, Jiachen Lian, Bhiksha Raj, Rita Singh, Carnegie Mellon University, United States

3: DEFENDING YOUR VOICE: ADVERSARIAL ATTACK ON VOICE CONVERSION.....................552
Chien-yu Huang, Yist Y. Lin, Hung-yi Lee, Lin-shan Lee, National Taiwan University, Taiwan

4: LIGHTWEIGHT VOICE ANONYMIZATION BASED ON DATA-DRIVEN OPTIMIZATION OF 560
CASCADED VOICE MODIFICATION MODULES
Hiroto Kai, Tokyo Metropolitan University, Japan; Shinnosuke Takamichi, The University of Tokyo, Japan; Sayaka Shiota, Hitoshi Kiya, Tokyo Metropolitan University, Japan

5: LOOK WHO'S NOT TALKING 567
Youngki Kwon, Hee Soo Heo, Jaesung Huh, Bong-Jin Lee, Joon Son Chung, Naver Corporation, Korea (South)

6: DISCRIMINATIVE NEURAL CLUSTERING FOR SPEAKER DIARISATION 574
Qiujia Li, Florian Kreyssig, Chao Zhang, Phil Woodland, University of Cambridge, United Kingdom

7: MULTI-CLASS SPECTRAL CLUSTERING WITH OVERLAPS FOR SPEAKER DIARIZATION582
Desh Raj, Zili Huang, Sanjeev Khudanpur, Johns Hopkins University, United States

8: DEVELOPING NEURAL REPRESENTATIONS FOR ROBUST CHILD-ADULT DIARIZATION.590
Suchitra Krishnamachari, Manoj Kumar, University of Southern California, United States; So Hyun Kim, Weill Cornell Medical College, United States; Catherine Lord, University of California, Los Angeles, United States; Shrikanth Narayanan, University of Southern California, United States

MULTIMODAL PROCESSING

1: END-TO-END LIP SYNCHRONISATION BASED ON PATTERN CLASSIFICATION....................598
You Jin Kim, Hee Soo Heo, Naver Corporation, Korea (South); Soo-Whan Chung, Yonsei University, Korea (South); Bong-Jin Lee, Naver Corporation, Korea (South)

2: END-TO-END SILENT SPEECH RECOGNITION WITH ACOUSTIC SENSING 606
Jian Luo, Jianzong Wang, Ning Cheng, Guilin Jiang, Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

3: SPEAKER-INDEPENDENT VISUAL SPEECH RECOGNITION WITH THE INCEPTION V3 613
MODEL
Timothy Israel Santos, Andrew Abel, Xi'an Jiaotong-Liverpool University, China; Nick Wilson, Macquarie University, Australia; Yan Xu, Xi'an Jiaotong-Liverpool University, China

4: LISTEN, LOOK AND DELIBERATE: VISUAL CONTEXT-AWARE SPEECH RECOGNITION ..621
USING PRE-TRAINED TEXT-VIDEO REPRESENTATIONS
Shahram Ghorbani, The University of Texas at Dallas, United States; Yashesh Gaur, Yu Shi, Jinyu Li, Microsoft, United States

5: ANALYSIS OF MULTIMODAL FEATURES FOR SPEAKING PROFICIENCY SCORING IN AN 629
INTERVIEW DIALOGUE
Mao Saeki, Yoichi Matsuyama, Waseda University, Japan; Satoshi Kobashikawa, NTT Corporation, Japan; Tetsuji Ogawa, Tetsunori Kobayashi, Waseda University, Japan

6: DETECTING EXPRESSIONS WITH MULTIMODAL TRANSFORMERS 636
Srinivas Parthasarathy, Shiva Sundaram, Amazon.com, United States

SPEECH ANALYSIS AND ENHANCEMENT

1: ENHANCING LOW-QUALITY VOICE RECORDINGS USING DISENTANGLED CHANNEL 734
FACTOR AND NEURAL WAVEFORM MODEL
Haoyu Li, National Institute of Informatics, Japan; Yang Ai, University of Science and Technology of China, China; Junichi Yamagishi, National Institute of Informatics, Japan

2: CAN WE TRUST DEEP SPEECH PRIOR? 742
Ying Shi, Harbin Institute of Technology, China; Haolin Chen, Zhiyuan Tang, Lantian Li, Dong Wang, Tsinghua University, China; Jiqing Han, Harbin Institute of Technology, China

3: CONTEXTUAL JOINT FACTOR ACOUSTIC EMBEDDINGS 750
Yanpei Shi, Thomas Hain, University of Sheffield, United Kingdom

4: SUPERVISED SPEAKER EMBEDDING DE-MIXING IN TWO-SPEAKER ENVIRONMENT 758
Yanpei Shi, Thomas Hain, University of Sheffield, United Kingdom

5: NEURAL MASK BASED MULTI-CHANNEL CONVOLUTIONAL BEAMFORMING FOR 766
JOINT DEREVERBERATION, ECHO CANCELLATION AND DENOISING
Jianming Liu, Meng Yu, Yong Xu, Chao Weng, Shi-Xiong Zhang, Lianwu Chen, Dong Yu, Tencent AI Lab, United States

6: PERSONALIZING SPEECH START POINT AND END POINT DETECTION IN ASR 771
SYSTEMS FROM SPEAKER EMBEDDINGS
Aditya Jayasimha, Periyasamy Paramasivam, Samsung Research Institute Bangalore, India

7: MULTIMODAL ATTENTION FUSION FOR TARGET SPEAKER EXTRACTION 778
Hiroshi Sato, Tsubasa Ochiai, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani, Shoko Araki, NTT Corporation, Japan

8: ESPNET-SE: END-TO-END SPEECH ENHANCEMENT AND SEPARATION TOOLKIT 785
DESIGNED FOR ASR INTEGRATION
Chenda Li, Shanghai Jiao Tong University, China; Jing Shi, Chinese Academy of Sciences, China; Wangyou Zhang, Shanghai Jiao Tong University, China; Aswin Shanmugam Subramanian, Xuankai Chang, Johns Hopkins University, United States; Naoyuki Kamo, NTT Corporation, Japan; Moto Hira, Facebook, United States; Tomoki Hayashi, Nagoya University;Human Dataware Lab. Co., Ltd., Japan; Christoph Boeddeker, Paderborn University, Germany; Zhuo Chen, Microsoft, United States; Shinji Watanabe, Johns Hopkins University, United States

9: AN INVESTIGATION INTO THE MULTI-CHANNEL TIME DOMAIN SPEAKER 793
EXTRACTION NETWORK
Catalin Zorila, Mohan Li, Rama Doddipatla, Toshiba Cambridge Research Laboratory, United Kingdom

SPECIAL SESSION ON INTEGRATION OF SPEECH SEPARATION, RECOGNITION AND DIARIZATION TOWARDS REAL CONVERSATION PROCESSING 2

1: DUAL-PATH RNN FOR LONG RECORDING SPEECH SEPARATION..865
Chenda Li, Shanghai Jiao Tong University, China; Yi Luo, Cong Han, Columbia University, United States; Jinyu Li, Takuya Yoshioka, Tianyan Zhou, Microsoft, United States; Marc Delcroix, Keisuke Kinoshita, NTT Corporation, Japan; Christoph Boeddeker, Paderborn University, Germany; Yanmin Qian, Shanghai Jiao Tong University, China; Shinji Watanabe, Johns Hopkins University, United States; Zhuo Chen, Microsoft, United States

2: RNN-T MODELS FAIL TO GENERALIZE TO OUT-OF-DOMAIN AUDIO: CAUSES AND 873
SOLUTIONS
Chung-Cheng Chiu, Arun Narayanan, Wei Han, Rohit Prabhavalkar, Yu Zhang, Google, United States; Navdeep Jaitly, D. E. Shaw Group, United States; Ruoming Pang, Tara Sainath, Google, United States; Patrick Nguyen, None, United States; Liangliang Cao, Yonghui Wu, Google, United States

3: DOVER-LAP: A METHOD FOR COMBINING OVERLAP-AWARE DIARIZATION OUTPUTS ..881
Desh Raj, Leibny Paola Garcia Perera, Zili Huang, Shinji Watanabe, Johns Hopkins University, United States; Daniel Povey, Xiaomi Corp., China; Andreas Stolcke, Amazon, United States; Sanjeev Khudanpur, Johns Hopkins University, United States

4: INTEGRATION OF VARIATIONAL AUTOENCODER AND SPATIAL CLUSTERING FOR 889
ADAPTIVE MULTI-CHANNEL NEURAL SPEECH SEPARATION
Katerina Zmolikova, Brno University of Technology, Czech Republic; Marc Delcroix, NTT Communication Science Laboratories, Japan; Lukáš Burget, Brno University of Technology, Czech Republic; Tomohiro Nakatani, NTT Communication Science Laboratories, Japan; Jan Černocký, Brno University of Technology, Czech Republic

5: INTEGRATION OF SPEECH SEPARATION, DIARIZATION, AND RECOGNITION FOR 897
MULTI-SPEAKER MEETINGS: SYSTEM DESCRIPTION, COMPARISON, AND ANALYSIS
Desh Raj, Johns Hopkins University, United States; Pavel Denisov, University of Stuttgart, Germany; Zhuo Chen, Microsoft Corp., United States; Hakan Erdogan, Google Research, United States; Zili Huang, Johns Hopkins University, United States; Maokui He, University of Science and Technology of China, China; Shinji Watanabe, Johns Hopkins University, United States; Jun Du, University of Science and Technology of China, China; Takuya Yoshioka, Microsoft Corp., United States; Yi Luo, Columbia University, United States; Naoyuki Kanda, Jinyu Li, Microsoft Corp., United States; Scott Wisdom, John R. Hershey, Google Research, United States

6: SEQUENTIAL MULTI-FRAME NEURAL BEAMFORMING FOR SPEECH SEPARATION AND 905
ENHANCEMENT
Zhong-Qiu Wang, MERL, United States; Hakan Erdogan, Scott Wisdom, Kevin Wilson, Google, United States; Desh Raj, Shinji Watanabe, Johns Hopkins University, United States; Zhuo Chen, Microsoft, United States; John Hershey, Google, United States

7: ACOUSTIC MODELING FOR MULTI-ARRAY CONVERSATIONAL SPEECH RECOGNITION 912
IN THE CHIME-6 CHALLENGE
Li Chai, Jun Du, University of Science and Technology of China, China; Di-Yuan Liu, iFlytek Research, China; Yan-Hui Tu, University of Science and Technology of China, China; Chin-Hui Lee, Georgia Institute of Technology, United States

INFORMATION RETRIEVAL FROM SPEECH AND SPEECH TRANSLATION

1: ACOUSTIC WORD EMBEDDINGS FOR ZERO-RESOURCE LANGUAGES USING 919
SELF-SUPERVISED CONTRASTIVE LEARNING AND MULTILINGUAL ADAPTATION
Christiaan Jacobs, Stellenbosch University, South Africa; Yevgen Matusevych, University of Edinburgh, United Kingdom; Herman Kamper, Stellenbosch University, South Africa

2: A COMPARISON OF SELF-SUPERVISED SPEECH REPRESENTATIONS AS INPUT 927
FEATURES FOR UNSUPERVISED ACOUSTIC WORD EMBEDDINGS
Lisa van Staden, Herman Kamper, Stellenbosch University, South Africa

3: ACOUSTIC SPAN EMBEDDINGS FOR MULTILINGUAL QUERY-BY-EXAMPLE SEARCH 935
Yushi Hu, University of Chicago, United States; Shane Settle, Karen Livescu, TTI-Chicago, United States

4: UNCERTAINTY-AWARE REPRESENTATIONS FOR SPOKEN QUESTION ANSWERING..........943
Merve Unlu, Bogazici University / Galatasaray University, Turkey; Ebru Arisoy, MEF University, Turkey

5: TIGHT INTEGRATED END-TO-END TRAINING FOR CASCADED SPEECH TRANSLATION .950
Parnia Bahar, Tobias Bieschke, Ralf Schlueter, Hermann Ney, RWTH Aachen University, Germany

6: TRANSFORMER-BASED DIRECT SPEECH-TO-SPEECH TRANSLATION WITH 958
TRANSCODER
Takatomo Kano, Nara Institute of Science and Technology, Japan; Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology and RIKEN, Center for Advanced Intelligence Project AIP, Japan

NATURAL LANGUAGE PROCESSING

1: GO BEYOND PLAIN FINE-TUNING: IMPROVING PRETRAINED MODELS FOR SOCIAL ...1028
COMMONSENSE
Ting-Yun Chang, Taiwan Academia Sinica, Taiwan; Yang Liu, Karthik Gopalakrishnan, Behnam Hedayatnia, Amazon, Alexa AI, United States; Pei Zhou, University of Southern California, United States; Dilek Hakkani-Tur, Amazon, Alexa AI, United States

2: GETTING YOUR CONVERSATION ON TRACK: ESTIMATION OF RESIDUAL LIFE FOR ...1036
CONVERSATIONS
Zexin Lu, Jing Li, The Hong Kong Polytechnic University, China; Yingyi Zhang, Nanjing University of Science and Technology, China; Haisong Zhang, tencent, China

3: PERSONALIZED EXTRACTIVE SUMMARIZATION FOR A NEWS DIALOGUE SYSTEM......1044
Hiroaki Takatsu, Mayu Okuda, Yoichi Matsuyama, Waseda University, Japan; Hiroshi Honda, Honda Motor Co.,Ltd., Japan; Shinya Fujie, Chiba Institute of Technology, Japan; Tetsunori Kobayashi, Waseda University, Japan

4: CROSS-DEMOGRAPHIC PORTABILITY OF DEEP NLP-BASED DEPRESSION MODELS 1052
Tomek Rutowski, Elizabeth Shriberg, Amir Harati, Yang Lu, Piotr Chlebek, Ricardo Oliveira, Ellipsis Health Inc, United States

5: THROUGH THE WORDS OF VIEWERS: USING COMMENT-CONTENT ENTANGLED 1058
NETWORK FOR HUMOR IMPRESSION RECOGNITION
Huan-Yu Chen, Yun-Shao Lin, Chi-Chun Lee, National Tsing Hua University, Taiwan

6: TWO-WAY NEURAL MACHINE TRANSLATION: A PROOF OF CONCEPT FOR 1065
BIDIRECTIONAL TRANSLATION MODELING USING A TWO-DIMENSIONAL GRID
Parnia Bahar, Christopher Brix, Hermann Ney, RWTH Aachen University, Germany

SPOKEN LANGUAGE UNDERSTANDING AND SPOKEN DIALOGUE SYSTEMS

1: PROTODA: EFFICIENT TRANSFER LEARNING FOR FEW-SHOT INTENT 966
CLASSIFICATION
Manoj Kumar, University of Southern California, United States; Varun Kumar, Hadrien Glaude, Cyprien de Lichy, Aman Alok, Rahul Gupta, Amazon, United States

2: VIRAAL: VIRTUAL ADVERSARIAL ACTIVE LEARNING FOR NLU 973
Gregory Senay, xBrain/Augustus Intelligence, United States; Badr Youbi Idrissi, xBrain, United States; Marine Haziza, xBrain/Augustus Intelligence, United States

3: WARPED LANGUAGE MODELS FOR NOISE ROBUST LANGUAGE UNDERSTANDING.........981
Mahdi Namazifar, Gokhan Tur, Dilek Hakkani-Tur, Amazon, United States

4: RNN BASED INCREMENTAL ONLINE SPOKEN LANGUAGE UNDERSTANDING.....................989
Prashanth Gurunath Shivakumar, University of Southern California, United States; Naveen Kumar, Disney Research, United States; Panayiotis Georgiou, Shrikanth Narayanan, University of Southern California, United States

5: A LIGHT TRANSFORMER FOR SPEECH-TO-INTENT APPLICATIONS..997
Pu Wang, Hugo Van hamme, KU Leuven, Belgium

6: META LEARNING TO CLASSIFY INTENT AND SLOT LABELS WITH NOISY FEW SHOT ...1004
EXAMPLES
Shang-Wen Li, Jason Krone, Shuyan Dong, Yi Zhang, Yaser Al-onaizan, Amazon AI, United States

7: LARGE-CONTEXT CONVERSATIONAL REPRESENTATION LEARNING: 1012
SELF-SUPERVISED LEARNING FOR CONVERSATIONAL DOCUMENTS
Ryo Masumura, Naoki Makishima, Mana Ihori, Akihiko Takashima, Tomohiro Tanaka, Shota Orihashi, NTT Corporation, Japan

8: USING PARALINGUISTIC INFORMATION TO DISAMBIGUATE USER INTENTIONS FOR .1020
DISTINGUISHING PHRASE STRUCTURE AND SARCASM IN SPOKEN DIALOG
SYSTEMS
Zhengyu Zhou, Bosch Research and Technology Center, United States; In Gyu Choi, Emory University, United States; Yongliang He, Georgia Institute of Technology, United States; Vikas Yadav, University of Arizona, United States; Chin-Hui Lee, Georgia Institute of Technology, United States

RESOURCES AND EVALUATION

1: A NEW DATASET FOR NATURAL LANGUAGE UNDERSTANDING OF EXERCISE LOGS IN A ... 1071
FOOD AND FITNESS SPOKEN DIALOGUE SYSTEM
Maya Epps, Juan Uribe, Mandy Korpusik, Loyola Marymount University, United States

2: TOWARDS LARGE-SCALE DATA ANNOTATION OF AUDIO FROM WEARABLES: VALIDATING 1079
ZOONIVERSE ANNOTATIONS OF INFANT VOCALIZATION TYPES
Chiara Semenzin, ENS, EHESS, CNRS, PSL University, France; Lisa Hamrick, Amanda Seidl, Bridgette Kelleher, Purdue University, United States; Alejandrina Cristia, ENS, EHESS, CNRS, PSL University, France

3: IDEA: AN ITALIAN DYSARTHRIC SPEECH DATABASE...1086
Marco Marini, University of Pisa, Italy; Mauro Viganò, Massimo Corbo, Casa Cura Policlinico, Italy; Marina Zettin, Gloria Simoncini, University of Torino, Italy; Bruno Fattori, Clelia D’Anna, Massimiliano Donati, Luca Fanucci, University of Pisa, Italy

4: EFFICIENT CORPUS DESIGN FOR WAKE-WORD DETECTION...1094
Delowar Hossain, Yoshinao Sato, Fairy Devices, Japan

5: IEEE SLT 2021 ALPHA-MINI SPEECH CHALLENGE: OPEN DATASETS, TRACKS, RULES .1101
AND BASELINES
Yihui Fu, Zhuoyuan Yao, Northwestern Polytechnical University, China; Weipeng He, Idiap Research Institute, Switzerland; Jian Wu, Microsoft, China; Xiong Wang, Zhanheng Yang, Shimin Zhang, Lei Xie, Northwestern Polytechnical University, China; Dongyan Huang, UBTECH Technology Co., Ltd., China; Hui BuPetr Motlicek, Jean-Marc Odobez, Idiap Research Institute, Switzerland

6: TAL: A SYNCHRONISED MULTI-SPEAKER CORPUS OF ULTRASOUND TONGUE 1109
IMAGING, AUDIO, AND LIP VIDEOS
Manuel Sam Ribeiro, Jennifer Sanger, The University of Edinburgh, United Kingdom; Jing-Xuan Zhang, University of Science and Technology of China, China; Aciel Eshky, The University of Edinburgh, United Kingdom; Alan Wrench, Articulate Instruments Ltd, United Kingdom; Korin Richmond, Steve Renals, The University of Edinburgh, United Kingdom

7: THE SLT 2021 CHILDREN SPEECH RECOGNITION CHALLENGE: OPEN DATASETS, ...1117
RULES AND BASELINES
Fan Yu, Zhuoyuan Yao, Xiong Wang, Northwestern Polytechnical University, China; Keyu An, Tsinghua University, China; Lei Xie, Northwestern Polytechnical University, China; Zhijian Ou, Tsinghua University, China; Bo Liu, Xiulin Li, Guanqiong Miao, Databaker Technology, China

ASSISTIVE TECHNOLOGIES

1: TOWARDS AUTOMATIC ROUTE DESCRIPTION UNIFICATION IN SPOKEN DIALOG 685
SYSTEMS
Yulan Feng, Alan Black, Maxine Eskenazi, Carnegie Mellon University, United States

2: ARTICULATORY COMPARISON OF L1 AND L2 SPEECH FOR MISPRONUNCIATION 693
DIAGNOSIS
Subash Khanal, Michael T. Johnson, Narjes Bozorg, University of Kentucky, United States

3: OPTIMIZED PREDICTION OF FLUENCY OF L2 ENGLISH BASED ON INTERPRETABLE 698
NETWORK USING QUANTITY OF PHONATION AND QUALITY OF PRONUNCIATION
Yang Shen, Ayano Yasukagawa, Daisuke Saito, Nobuaki Minematsu, The University of Tokyo, Japan; Kazuya Saito, University College London, United Kingdom

4: AUTOMATED SCORING OF SPONTANEOUS SPEECH FROM YOUNG LEARNERS OF 705
ENGLISH USING TRANSFORMERS
Xinhao Wang, Keelan Evanini, Educational Testing Service, United States; Yao Qian, Microsoft, United States; Matthew Mulholland, Educational Testing Service, United States

5: IMPROVING L2 ENGLISH RHYTHM EVALUATION WITH AUTOMATIC SENTENCE 713
STRESS DETECTION
Binghuai Lin, Liyuan Wang, Tencent Technology Company, China; Hongwei Ding, School of Foreign Languages Shanghai Jiao Tong University, China; Xiaoli Feng, Beijing Language and Culture University, China

6: ENHANCING THE INTELLIGIBILITY OF CLEFT LIP AND PALATE SPEECH USING 720
CYCLE-CONSISTENT ADVERSARIAL NETWORKS
Protima Nomo Sudro, Indian Institute of Technology Guwahati, India; Rohan Kumar Das, National University of Singapore, Singapore; Rohit Sinha, Indian Institute of Technology Guwahati, India; S. R. Mahadeva Prasanna, Indian Institute of Technology Dharwad, India

7: DEVELOPMENT OF CNN-BASED COCHLEAR IMPLANT AND NORMAL HEARING SOUND 728
RECOGNITION MODELS USING NATURAL AND AURALIZED ENVIRONMENTAL
AUDIO
Ram Charan Chandra Shekar, Chelzy Belitz, John Hansen, UNIVERSITY OF TEXAS AT DALLAS, United States

SPECIAL SESSION ON INTEGRATION OF SPEECH SEPARATION, RECOGNITION AND DIARIZATION TOWARDS REAL CONVERSATION PROCESSING 1

1: EFFECTIVE LOW-COST TIME-DOMAIN AUDIO SEPARATION USING GLOBALLY 801
ATTENTIVE LOCALLY RECURRENT NETWORKS
Max W. Y. Lam, Jun Wang, Dan Su, Dong Yu, Tencent AI Lab, China

2: INVESTIGATION OF END-TO-END SPEAKER-ATTRIBUTED ASR FOR CONTINUOUS 809
MULTI-TALKER RECORDINGS
Naoyuki Kanda, Microsoft, United States; Xuankai Chang, Johns Hopkins University, United States; Yashesh Gaur, Xiaofei Wang, Zhong Meng, Zhuo Chen, Takuya Yoshioka, Microsoft, United States

3: WPD++: AN IMPROVED NEURAL BEAMFORMER FOR SIMULTANEOUS SPEECH 817
SEPARATION AND DEREVERBERATION
Zhaoheng Ni, City University of New York, United States; Yong Xu, Meng Yu, Bo Wu, Shixiong Zhang, Dong Yu, Tencent AI lab, United States; Michael I Mandel, City University of New York, United States

4: DISTORTION-CONTROLLED TRAINING FOR END-TO-END REVERBERANT SPEECH 825
SEPARATION WITH AUXILIARY AUTOENCODING LOSS
Yi Luo, Cong Han, Nima Mesgarani, Columbia University, United States

5: EXPLORING END-TO-END MULTI-CHANNEL ASR WITH BIAS INFORMATION FOR 833
MEETING TRANSCRIPTION
Xiaofei Wang, Naoyuki Kanda, Yashesh Gaur, Zhuo Chen, Zhong Meng, Takuya Yoshioka, Microsoft Corp., United States

6: ONLINE END-TO-END NEURAL DIARIZATION WITH SPEAKER-TRACING BUFFER............841
Yawen Xue, Shota Horiguchi, Yusuke Fujita, Hitachi, Ltd., Japan; Shinji Watanabe, Paola Garcia, Johns Hopkins University, United States; Kenji Nagamatsu, Hitachi, Ltd, United States

7: END-TO-END SPEAKER DIARIZATION CONDITIONED ON SPEECH ACTIVITY AND 849
OVERLAP DETECTION
Yuki Takashima, Yusuke Fujita, Hitachi, Ltd., Japan; Shinji Watanabe, Johns Hopkins University, United States; Shota Horiguchi, Hitachi, Ltd., Japan; Paola Garcia, Johns Hopkins University, United States; Kenji Nagamatsu, Hitachi, Ltd., Japan

8: DESNET: A MULTI-CHANNEL NETWORK FOR SIMULTANEOUS SPEECH 857
DEREVERBERATION, ENHANCEMENT AND SEPARATION
Yihui Fu, Northwestern Polytechnical University, China; Jian Wu, Microsoft, China; Yanxin Hu, Mengtao Xing, Lei Xie, Northwestern Polytechnical University, China

SPOKEN LANGUAGE UNDERSTANDING AND SOKEN DIALOGUE SYSTEMS

HUMAN COMPUTER INTERACTION

1: NOISE-ROBUST SPOKEN LANGUAGE IDENTIFICATION USING LANGUAGE RELEVANCE 644
FACTOR BASED EMBEDDING
Muralikrishna H, Shikha Gupta, Dileep Aroor Dinesh, Padmanabhan Rajan, Indian Institue of Technology Mandi, India

2: VOXLINGUA107: A DATASET FOR SPOKEN LANGUAGE RECOGNITION 652
Jörgen Valk, Tanel Alumäe, Tallinn University of Technology, Estonia

3: STREAMING RESLSTM WITH CAUSAL MEAN AGGREGATION FOR DEVICE-DIRECTED .659
UTTERANCE DETECTION
Xiaosu Tong, Che-Wei Huang, Sri Harish Mallidi, Shaun Joseph, Sonal Pareek, Chander Chandak, Ariya Rastrow, Roland Maas, Amazon, United States

4: REAL-TIME INDEPENDENT VECTOR ANALYSIS WITH A DEEP-LEARNING-BASED 665
SOURCE MODEL
Fang Kang, Feiran Yang, Jun Yang, University of Chinese Academy of Sciences, China

5: AN EXPLORATION OF LOG-MEL SPECTROGRAM AND MFCC FEATURES FOR 670
ALZHEIMER’S DEMENTIA RECOGNITION FROM SPONTANEOUS SPEECH
Amit Meghanani, Anoop C. S., Ramakrishnan A. G., Indian Institute of Science, India

6: FILM QUALITY PREDICTION USING ACOUSTIC, PROSODIC AND LEXICAL CUES 678
Su Ji Park, Alan Rozet, Columbia University, United States