2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2021)

Cartagena, Colombia
13-17 December 2021

Pages 1-593

IEEE Catalog Number: CFP21SRW-POD
ISBN: 978-1-6654-3740-0

****** 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: CFP21SRW-POD
ISBN (Print-On-Demand): 978-1-6654-3740-0
ISBN (Online): 978-1-6654-3739-4

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

AUTOMATIC SPEECH RECOGNITION 1

Paper #1: INSTANT ONE-SHOT WORD-LEARNING FOR CONTEXT-SPECIFIC NEURAL SEQUENCE-TO-SEQUENCE SPEECH RECOGNITION 1
Christian Huber, Juan Hussain, Sebastian Stüker, Alexander Waibel, KIT Karlsruhe, Germany

Paper #2: EFFICIENT CONFORMER: PROGRESSIVE DOWNSAMPLING AND GROUPED ATTENTION FOR AUTOMATIC SPEECH RECOGNITION 8
Maxime Burchi, Valentin Vielzeuf, Orange Labs, France

Paper #3: A STUDY OF TRANSDUCER BASED END-TO-END ASR WITH ESPNET: ARCHITECTURE, AUXILIARY LOSS AND DECODING STRATEGIES 16
Florian Boyer, Airudit / LaBRI, France; Yusuke Shinohara, Takaaki Ishii, Yahoo Japan Corporation, Japan; Hirofumi Inaguma, Kyoto University, Japan; Shinji Watanabe, Carnegie Mellon University / Johns Hopkins University, United States

SPEECH EMOTION RECOGNITION 1

Paper #1: A STUDY ON CROSS-CORPUS SPEECH EMOTION RECOGNITION AND DATA AUGMENTATION 24
Norbert Braunschweiler, Rama Doddipatla, Simon Keizer, Svetlana Stoyanchev, Toshiba Europe Limited, United Kingdom

Paper #2: DETECTING EMOTION CARRIERS BY COMBINING ACOUSTIC AND LEXICAL REPRESENTATIONS 31
Sebastian Peter Bayerl, Technische Hochschule Nürnberg Georg Simon Ohm, Germany; Aniruddha Tammewar, University of Trento, Italy; Korbinian Riedhammer, Technische Hochschule Nürnberg Georg Simon Ohm, Germany; Giuseppe Riccardi, University of Trento, Italy

Paper #3: BEYOND ISOLATED UTTERANCES: CONVERSATIONAL EMOTION RECOGNITION 39
Raghavendra Pappagari, Piotr Zelasko, Johns Hopkins University, United States; Jesus Villalba, Laureano Moro-Velazquez, Johns Hopkins Univeristy, United States; Najim Dehak, Johns Hopkins University, United States

AUTOMATIC SPEECH RECOGNITION 2

Paper #1: A COMPARATIVE STUDY ON NON-AUTOREGRESSIVE MODELINGS FOR SPEECH-TO-TEXT GENERATION 47
Yosuke Higuchi, Waseda University, Japan; Nanxin Chen, Johns Hopkins University, United States; Yuya Fujita, Yahoo Japan Corporation, Japan; Hirofumi Inaguma, Kyoto University, Japan; Tatsuya Komatsu, LINE Corporation, Japan; Jaesong Lee, Naver Corporation, South Korea; Jumon Nozaki, LINE Corporation, Kyoto University, Japan; Tianzi Wang, Johns Hopkins University, United States; Shinji Watanabe, Carnegie Mellon University, United States

Paper #2: TENET: A TIME-REVERSAL ENHANCEMENT NETWORK FOR NOISE-ROBUST ASR 55
Fu-An Chao, Shao-wei Fan Jiang, Bi-Cheng Yan, National Taiwan Normal University, Taiwan; Jeih-weih Hung, National Chi Nan University, Taiwan; Berlin Chen, National Taiwan Normal University, Taiwan

Paper #3: LATENCY-CONTROLLED NEURAL ARCHITECTURE SEARCH FOR STREAMING SPEECH RECOGNITION 62
Liqiang He, Shulin Feng, Dan Su, Dong Yu, Tencent, China

Paper #4: DATA AUGMENTATION FOR ASR USING TTS VIA A DISCRETE REPRESENTATION 68
Sei Ueno, Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara, Kyoto University, Japan

Paper #5: IMPROVING HYBRID CTC/ATTENTION END-TO-END SPEECH RECOGNITION WITH PRETRAINED ACOUSTIC AND LANGUAGE MODELS 76
Keqi Deng, University of Chinese Academy of Sciences, China, China; Songjun Cao, Yike Zhang, Long Ma, Tencent Cloud Xiaowei, Beijing, China, China

Paper #6: IMPROVING ASR ERROR CORRECTION USING N-BEST HYPOTHESES 83
Linchen Zhu, Wenjie Liu, Linquan Liu, Edward Lin, Microsoft, China

SPEAKER & LANGUAGE RECOGNITION 1

Paper #1: SELF-SUPERVISED METRIC LEARNING WITH GRAPH CLUSTERING FOR SPEAKER DIARIZATION 90
Prachi Singh, Sriram Ganapathy, Indian Institute of Science, India

Paper #2: TOWARDS NEURAL DIARIZATION FOR UNLIMITED NUMBERS OF SPEAKERS USING GLOBAL AND LOCAL ATTRACTORS 98
Shota Horiguchi, Hitachi, Ltd., Japan; Shinji Watanabe, Carnegie Mellon University, United States; Paola Garcia, Johns Hopkins University, United States; Yawen Xue, Yuki Takashima, Yohei Kawaguchi, Hitachi, Ltd., Japan

Paper #3: PL-EESR: PERCEPTUAL LOSS BASED END-TO-END ROBUST SPEAKERREPRESENTATION EXTRACTION 106
*Yi Ma, National University of Singapore, Singapore; Kong Aik Lee, A*STAR, Singapore; Ville Hautamaki, University of Eastern Finland, Finland; Haizhou Li, National University of Singapore, Singapore*

Paper #4: ROBUST SPEECH AGE ESTIMATION USING LOCAL MAXIMUM MEAN DISCREPANCY UNDER MISMATCHED RECORDING CONDITIONS 114
Naohiro Tawara, Atsunori Ogawa, Yuki Kitagishi, Hosana Kamiyama, Yusuke Ijima, Nippon Telegraph and Telephone Corporation, Japan

Paper #5: DEEPLIP: A BENCHMARK FOR DEEP LEARNING-BASED AUDIO-VISUAL LIP BIOMETRICS 122
*Meng Liu, Longbiao Wang, Tianjin University, China; Kong Aik Lee, Institute for Infocomm Research, A*STAR, China; Hanyi Zhang, Tianjin University, China; Chang Zeng, National Institute of Informatics, China; Jianwu Dang, Tianjin University, China*

Paper #6: SHORT-UTTERANCE EMBEDDING ENHANCEMENT METHOD BASED ON TIME SERIES FORECASTING TECHNIQUE FOR TEXT-INDEPENDENT SPEAKER VERIFICATION 130
Jeong-Hwan Choi, Joon-Young Yang, Joon-Hyuk Chang, Hanyang University, South Korea

AUTOMATIC SPEECH RECOGNITION 3

Paper #1: DISTILLING KNOWLEDGE FROM ENSEMBLES OF ACOUSTIC MODELS FOR JOINT CTC-ATTENTION END-TO-END SPEECH RECOGNITION 138
Yan Gao, University of Cambridge, United Kingdom; Titouan Parcollet, Avignon University, France; Nicholas D. Lane, University of Cambridge, Samsung AI Cambridge, United Kingdom

Paper #2: EFFICIENT KEYWORD SPOTTING BY CAPTURING LONG-RANGE INTERACTIONS WITH TEMPORAL LAMBDA NETWORKS 146
Biel Tura Vecino, Universitat Politècnica de Catalunya, Spain; Ferran Diego, Carlos Segura, Jordi Luque, Telefónica, Spain; Santiago Escuder, Universitat Politècnica de Catalunya, Spain

Paper #3: IMPROVING HS-DACS BASED STREAMING TRANSFORMER ASR WITH DEEP REINFORCEMENT LEARNING 154
Mohan Li, Rama Doddipatla, Toshiba Cambridge Research Laboratory, Toshiba Europe Ltd, United Kingdom

Paper #4: ADAPTING GPT, GPT-2 AND BERT LANGUAGE MODELS FOR SPEECH RECOGNITION 162
Xianrui Zheng, Chao Zhang, Philip Woodland, University of Cambridge, United Kingdom

Paper #5: COMPARISON OF SELF-SUPERVISED SPEECH PRE-TRAINING METHODS ON FLEMISH DUTCH 169
Jakob Poncelet, Hugo Van hamme, KU Leuven, Belgium

Paper #6: RELAXED ATTENTION: A SIMPLE METHOD TO BOOST PERFORMANCE OF END-TO-END AUTOMATIC SPEECH RECOGNITION 177
Timo Lohrenz, Patrick Schwarz, Zhengyang Li, Tim Fingscheidt, Technische Universität Braunschweig, Germany

SPEAKER & LANGUAGE RECOGNITION 2

Paper #1: OPTIMIZED POWER NORMALIZED CEPSTRAL COEFFICIENTS TOWARDS ROBUST DEEP SPEAKER VERIFICATION 185
Xuechen Liu, MULTISPEECH, Inria Nancy Grand Est; School of Computing, University of Eastern Finland, France; Md Sahidullah, MULTISPEECH, Inria Nancy Grand Est, France; Tomi Kinnunen, School of Computing, University of Eastern Finland, Finland

Paper #2: ON THE INVERTIBILITY OF A VOICE PRIVACY SYSTEM USING EMBEDDING ALIGNMENT 191
Pierre Champion, INRIA, France; Thomad Thebaud, Gaël Le Lan, Orange, France; Anthony Larcher, LIUM, France; Denis Jouvet, INRIA, France

Paper #3: IMPROVING TEXT-INDEPENDENT SPEAKER VERIFICATION WITH AUXILIARY SPEAKERS USING GRAPH 198
Jingyu Li, Si-Ioi Ng, Tan Lee, The Chinese University of Hong Kong, Hong Kong SAR China

Paper #4: DUALITY TEMPORAL-CHANNEL-FREQUENCY ATTENTION ENHANCED SPEAKER REPRESENTATION LEARNING 206
Li Zhang, Qing Wang, Xie Lei, Northwestern Polytechnical University, China

Paper #5: MACCIF-TDNN: MULTI ASPECT AGGREGATION OF CHANNEL AND CONTEXT INTERDEPENDENCE FEATURES IN TDNN BASED SPEAKER VERIFICATION 214
fangyuan wang, Institute of Automation, Chinese Academy of Sciences, China; zhigang song, Beijing University of Technology, China; hongchen jiang, bo xu, Institute of Automation, Chinese Academy of Sciences, China

Paper #6: SI-NET: MULTI-SCALE CONTEXT-AWARE CONVOLUTIONAL BLOCK FOR SPEAKER VERIFICATION 220
Zhuo Li, Ce Fang, Runqiu Xiao, Wenchao Wang, Yonghong Yan, Institute of Acoustics, Chinese Academy of Sciences, China

AUTOMATIC SPEECH RECOGNITION 4

Paper #1: AN EXPLORATION OF SELF-SUPERVISED PRETRAINED REPRESENTATIONS FOR END-TO-END SPEECH RECOGNITION 228
Xuankai Chang, Carnegie Mellon University, United States; Takashi Maekaku, Yahoo Japan Corporation, Japan; Pengcheng Guo, Northwestern Polytechnical University, China; Jing Shi, Institute of Automation, Chinese Academy of Sciences, China; Yen-Ju Lu, Academia Sinica, Taiwan; Aswin Shanmugam Subramanian, Tianzi Wang, Johns Hopkins University, United States; Shu-wen Yang, National Taiwan University, Taiwan; Yu Tsao, Academia Sinica, Taiwan; Hung-yi Lee, National Taiwan University, Taiwan; Shinji Watanabe, Carnegie Mellon University, United States

Paper #2: REMEMBER THE CONTEXT! ASR SLOT ERROR CORRECTION THROUGH MEMORIZATION 236
Dhanush Bekal, Ashish Shenoy, Monica Sunkara, Sravan Bodapati, Katrin Kirchhoff, Amazon, United States

Paper #3: W2V-BERT: COMBINING CONTRASTIVE LEARNING AND MASKED LANGUAGE MODELING FOR SELF-SUPERVISED SPEECH PRE-TRAINING 244
Yu-An Chung, MIT, United States; Yu Zhang, Wei Han, Chung-Cheng Chiu, James Qin, Ruoming Pang, Yonghui Wu, Google, United States

Paper #4: INJECTING TEXT IN SELF-SUPERVISED SPEECH PRE-TRAINING 251
Zhehuai Chen, Yu Zhang, Andrew Rosenberg, Bhuvana Ramabhadran, Gary Wang, Pedro Moreno, Google, United States

Paper #5: TS-RIR: TRANSLATED SYNTHETIC ROOM IMPULSE RESPONSES FOR SPEECH AUGMENTATION 259
Anton Jeran Ratnarajah, Zhenyu Tang, Dinesh Manocha, UNIVERSITY OF MARYLAND COLLEGE PARK, United States

Paper #6: ON ARCHITECTURES AND TRAINING FOR RAW WAVEFORM FEATURE EXTRACTION IN ASR 267
Peter Vieting, Christoph Lüscher, Wilfried Michel, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

ASR IN ADVERSE ENVIRONMENTS 1

Paper #1: MULTI-USER VOICEFILTER-LITE VIA ATTENTIVE SPEAKER EMBEDDING 275
Quan Wang, Rajeev Rikhye, Qiao Liang, Yanzhang He, Ian McGraw, Google, United States

Paper #2: SPEAKER CONDITIONING OF ACOUSTIC MODELS USING AFFINE TRANSFORMATION FOR MULTI-SPEAKER SPEECH RECOGNITION 283
Midia Yousefi, Research Associate at University of Texas at Dallas, United States; John Hansen, Professor at University of Texas at Dallas, United States

Paper #3: SCENARIO AWARE SPEECH RECOGNITION: ADVANCEMENTS FOR APOLLO FEARLESS STEPS & CHIME-4 CORPORA 289
Szu-Jui Chen, Wei Xia, John H.L. Hansen, University of Texas at Dallas, United States

Paper #4: A COMPARATIVE STUDY OF MODULAR AND JOINT APPROACHES FOR SPEAKER-ATTRIBUTED ASR ON MONAURAL LONG-FORM AUDIO 296
Naoyuki Kanda, Xiong Xiao, Jian Wu, Tianyan Zhou, Yashesh Gaur, Xiaofei Wang, Zhong Meng, Zhuo Chen, Takuya Yoshioka, Microsoft, United States

Paper #5: A CONFORMER-BASED ASR FRONTEND FOR JOINT ACOUSTIC ECHO CANCELLATION, SPEECH ENHANCEMENT AND SPEECH SEPARATION 304
Tom O'Malley, Arun Narayanan, Quan Wang, Alex Park, James Walker, Nathan Howard, Google, United States

Paper #6: CROSS-ATTENTION CONFORMER FOR CONTEXT MODELING IN SPEECH ENHANCEMENT FOR ASR 312
Arun Narayanan, Chung-Cheng Chiu, Tom O'Malley, Quan Wang, Yanzhang He, Google Inc., United States

AUTOMATIC SPEECH RECOGNITION 5

Paper #1: INCREMENTAL LEARNING FOR END-TO-END AUTOMATIC SPEECH RECOGNITION 320
Li Fu, Xiaoxiao Li, Libo Zi, Zhengchen Zhang, Youzheng Wu, Xiaodong He, Bowen Zhou, JD AI Research, China

Paper #2: BOUNDARY AND CONTEXT AWARE TRAINING FOR CIF-BASED NON-AUTOREGRESSIVE END-TO-END ASR 328
Fan Yu, Haoneng Luo, Pengcheng Guo, Yuhao Liang, Zhuoyuan Yao, Lei Xie, Northwestern Polytechnical University, China; Yingying Gao, Leijing Hou, Shilei Zhang, China Mobile Research, China

Paper #3: AUTOMATIC SPEECH RECOGNITION FOR LOW-RESOURCE LANGUAGES: THE THUEE SYSTEMS FOR THE IARPA OPENASR20 EVALUATION 335
Jing Zhao, Guixin Shi, Guan-Bo Wang, Wei-Qiang Zhang, Tsinghua University, China

Paper #4: UNSUPERVISED DOMAIN ADAPTATION SCHEMES FOR BUILDING ASR IN LOW-RESOURCE LANGUAGES 342
Anoop C S, Indian Institute of Science, Bangalore, India; Prathosh A P, Indian Institute of Technology, Delhi, India; Ramakrishnan A G, Indian Institute of Science, Bangalore, India

SPEECH EMOTION RECOGNITION 2

Paper #1: MULTIMODAL EMOTION RECOGNITION WITH HIGH-LEVEL SPEECH AND TEXT FEATURES 350
Mariana Rodrigues Makiuchi, Kuniaki Uto, Koichi Shinoda, Tokyo Institute of Technology, Japan

Paper #2: SPEECH EMOTION RECOGNITION USING SEMI-SUPERVISED LEARNING WITH EFFICIENT LABELING STRATEGIES 358
Zhi Zhu, Yoshinao Sato, Fairy Devices Inc., Japan

Paper #3: UNSUPERVISED CROSS-LINGUAL SPEECH EMOTION RECOGNITION USING PSEUDO MULTILABEL 366
Jin Li, Nan Yan, Lan Wang, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

Paper #4: ENSEMBLE OF DOMAIN ADVERSARIAL NEURAL NETWORKS FOR SPEECH EMOTION RECOGNITION 374
SHI-WOOK LEE, National Institute of Advanced Industrial Science and Technology, Japan, Japan

AUTOMATIC SPEECH RECOGNITION 6

Paper #1: ASR RESCORING AND CONFIDENCE ESTIMATION WITH ELECTRA 380
Hayato Futami, Hirofumi Inaguma, Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara, Kyoto University, Japan

Paper #2: COMPARATIVE STUDY OF DIFFERENT TOKENIZATION STRATEGIES FOR STREAMING END-TO-END ASR 388
Sachin Singh, Ashutosh Gupta, Aman Maghan, Dhananyaya Nagaraj Gowda, Shatrughan Singh, Chanwoo Kim, Samsung, India

Paper #3: HITNET: BYTE-TO-BPE HIERARCHICAL TRANSCRIPTION NETWORK FOR END-TO-END SPEECH RECOGNITION 395
Dhananjaya Gowda, Abhinav Garg, Sachin Singh, Ashutosh Gupta, Jiyeon Kim, Mehul Kumar, Ankur Kumar, Nauman Dawalatabad, Aman Maghan, Shatrughan Singh, Chanwoo Kim, Samsung Research, South Korea

Paper #4: TWO-PASS END-TO-END ASR MODEL COMPRESSION 403
Nauman Dawalatabad, Samsung Research, India; Tushar Vatsal, Amazon, India; Ashutosh Gupta, Sungsoo Kim, Shatrughan Singh, Dhananjaya Gowda, Chanwoo kim, Samsung Research, India

SPOKEN LANGUAGE UNDERSTANDING 1

Paper #1: SEQUENCE MODEL WITH SELF-ADAPTIVE SLIDING WINDOW FOR EFFICIENT SPOKEN DOCUMENT SEGMENTATION 411
Qinglin Zhang, Qian Chen, Yali Li, Alibaba Group, China; Jiaqing Liu, Renmin University of China, China; Wen Wang, Alibaba Group, United States

Paper #2: EXPLORING TEACHER-STUDENT LEARNING APPROACH FOR MULTI-LINGUAL SPEECH-TO-INTENT CLASSIFICATION 419
Bidisha Sharma, Maulik Madhavi, Xuehao Zhou, Haizhou Li, Department of Electrical and Computer Engineering,National University of Singapore, Singapore, Singapore

Paper #3: TOPIC CLASSIFICATION ON SPOKEN DOCUMENTS USING DEEP ACOUSTIC AND LINGUISTIC FEATURES 427
Tan Liu, Wu Guo, University of Science and Technology of China, China

Paper #4: HIERARCHICAL KNOWLEDGE DISTILLATION FOR DIALOGUE SEQUENCE LABELING 433
Shota Orihashi, Yoshihiro Yamazaki, Naoki Makishima, Mana Ihori, Akihiko Takashima, Tomohiro Tanaka, Ryo Masumura, NTT Corporation, Japan

AUTOMATIC SPEECH RECOGNITION 7

Paper #1: LEARNING HOW LONG TO WAIT: ADAPTIVELY-CONSTRAINED MONOTONIC MULTIHEAD ATTENTION FOR STREAMING ASR 441
Jaeyun Song, Hajin Shim, Eunho Yang, KAIST, South Korea

Paper #2: UTTERANCE-LEVEL NEURAL CONFIDENCE MEASURE FOR END-TO-END CHILDREN SPEECH RECOGNITION 449
Wei Liu, Tan Lee, The Chinese University of Hong Kong, Hong Kong SAR China

Paper #3: WARPED ENSEMBLES: A NOVEL TECHNIQUE FOR IMPROVING CTC BASED END-TO-END SPEECH RECOGNITION 457
Kiran Praveen, Hardik Sailor, Abhishek Pandey, Samsung R&D Institute, Bangalore, India

Paper #4: NON-AUTOREGRESSIVE MANDARIN-ENGLISH CODE-SWITCHING SPEECH RECOGNITION 465
Shun-Po Chuang, Heng-Jui Chang, Sung-Feng Huang, Hung-yi Lee, National Taiwan University, Taiwan

SPOKEN LANGUAGE UNDERSTANDING 2

Paper #1: VOICE TO ACTION : SPOKEN LANGUAGE UNDERSTANDING FOR MEMORY-CONSTRAINED SYSTEMS 473
Ashutosh Gupta, Aditya Jayasimha, Aman Maghan, Shatrughan Singh, Dhananjaya Gowda, Chanwoo Kim, Samsung Research, India

Paper #2: VARIATIONAL SEQUENTIAL MODELING, LEARNING AND UNDERSTANDING 480
Jen-Tzung Chien, Chih-Jung Tsai, National Yang Ming Chiao Tung University, Taiwan

Paper #3: ATTENTION-BASED MULTI-HYPOTHESIS FUSION FOR SPEECH SUMMARIZATION 487
Takatomo Kano, Atsunori Ogawa, Marc Delcroix, NTT Corporation, Japan; Shinji Watanabe, Language Technologies Institute, Carnegie Mellon University, United States

Paper #4: ESTIMATING THE GENERATION TIMING OF RESPONSIVE UTTERANCES BY ACTIVE LISTENERS OF SPOKEN NARRATIVES 495
Koichiro Ito, Nagoya University, Japan; Masaki Murata, National Institute of Technology, Toyota College, Japan; Tomohiro Ohno, Tokyo Denki University, Japan; Shigeki Matsubara, Nagoya University, Japan

AUTOMATIC SPEECH RECOGNITION 8

Paper #1: CONTEXT-AWARE TRANSFORMER TRANSDUCER FOR SPEECH RECOGNITION 503
Feng-Ju Chang, Jing Liu, Martin Radfar, Athanasios Mouchtaris, Maurizio Omologo, Ariya Rastrow, Siegfried Kunzmann, Amazon, United States

Paper #2: PSVD: POST-TRAINING COMPRESSION OF LSTM-BASED RNN-T MODELS 511
Suwa Xu, Jinwon Lee, Jim Steele, Amazon, United States

Paper #3: KAIZEN: CONTINUOUSLY IMPROVING TEACHER USING EXPONENTIAL MOVING AVERAGE FOR SEMI-SUPERVISED SPEECH RECOGNITION 518
Vimal Manohar, Tatiana Likhomanenko, Qiantong Xu, Wei-Ning Hsu, Ronan Collobert, Yatharth Saraf, Geoffrey Zweig, Abdelrahman Mohamed, Facebook AI, United States

Paper #4: ON ADDRESSING PRACTICAL CHALLENGES FOR RNN-TRANSDUCER 526
Rui Zhao, Jian Xue, Jinyu Li, Wenning Wei, Lei He, Yifan Gong, Microsoft, United States

Paper #5: DUAL-ENCODER ARCHITECTURE WITH ENCODER SELECTION FOR JOINT CLOSE-TALK AND FAR-TALK SPEECH RECOGNITION 534
Felix Weninger, Marco Gaudesi, Ralf Leibold, Roberto Gemello, Puming Zhan, Nuance Communications, United States

Paper #6: TINY-CRNN: STREAMING WAKEWORD DETECTION IN A LOW FOOTPRINT SETTING 541
Mohammad Omar Khursheed, Christin Jose, Rajath Kumar, Gengshen Fu, Brian Kulis, Santosh Kumar Cheekatmalla, Amazon, United States

OTHER TOPICS 1

Paper #1: TEXTUAL ECHO CANCELLATION 548
Shaojin Ding, Ye Jia, Ke Hu, Quan Wang, Google, United States

Paper #2: COLOMBIAN DIALECT RECOGNITION BASED ON INFORMATION EXTRACTED FROM SPEECH AND TEXT SIGNALS 556
Daniel Escobar-Grisales, Cristian David Ríos-Urrego, Diego Alexánder López-Santander, Jeferson David Gallo-Aristizabal, University of Antioquia, Colombia; Juan Camilo Vásquez-Correa, Pratech Group, Colombia; Elmar Nöth, Friedrich Alexander Universität Erlangen Nürnberg, Germany; Juan Rafael Orozco-Arroyave, University of Antioquia, Colombia

Paper #3: INCORPORATING REAL-WORLD NOISY SPEECH IN NEURAL-NETWORK-BASED SPEECH ENHANCEMENT SYSTEMS 564
Yangyang Xia, Carnegie Mellon University, United States; Anurag Kumar, Buye Xu, Facebook, Inc., United States

Paper #4: MULTI-TASK LEARNING WITH CROSS ATTENTION FOR KEYWORD SPOTTING 571
Takuya Higuchi, Apple, United States; Anmol Gupta, The University of Hong Kong, Hong Kong SAR China; Chandra Dhir, Apple, United States

Paper #5: AUTOMATIC GENERATION OF DIAGNOSTIC CONTENT FEEDBACK IN SPOKEN LANGUAGE LEARNING AND ASSESSMENT 579
Xinhao Wang, Christopher Hamill, Educational Testing Service, United States

Paper #6: ARE YOU DICTATING TO ME? DETECTING EMBEDDED DICTATIONS IN DOCTOR-PATIENT CONVERSATIONS 587
Thomas Schaaf, Longxiang Zhang, Alireza Bayestehtashk, Shahid Durrani, Susanne Burger, Monika Woszczyna, Thomas Polzin, 3M, United States

TEXT-TO-SPEECH SYNTHESIS 1

Paper #1: EXPRESSIVE VOICE CONVERSION: A JOINT FRAMEWORK FOR SPEAKER IDENTITY AND EMOTIONAL STYLE TRANSFER 594
Zongyang Du, Berrak Sisman, Singapore University of Technology and Design, Singapore; Kun Zhou, Haizhou Li, National University of Singapore, Singapore, Singapore

Paper #2: LEARNING LANGUAGE AND SPEAKER INFORMATION FOR CODE-SWITCH SPEECH SYNTHESIS WITH LIMITED DATA 602
Mengxin Chai, Shaotong Guo, Cheng Gong, Longbiao Wang, Tianjin University, China; Jianwu Dang, Tianjin University, Japan Advanced Institute of Science and Technology, China; Ju Zhang, Huiyan Technology Co., Ltd, China

Paper #3: MULTI-STREAM HIFI-GAN WITH DATA-DRIVEN WAVEFORM DECOMPOSITION 610
Takuma Okamoto, National Institute of Information and Communications Technology, Japan; Tomoki Toda, Nagoya University, Japan; Hisashi Kawai, National Institute of Information and Communications Technology, Japan

Paper #4: DEEPA: A DEEP NEURAL ANALYZER FOR SPEECH AND SINGING VOCODING 618
Sergey Nikonorov, National University of Singapore, Singapore; Berrak Sisman, Singapore University of Technology and Design, Singapore; Mingyang Zhang, Haizhou Li, National University of Singapore, Singapore

Paper #5: EDITSPEECH: A TEXT BASED SPEECH EDITING SYSTEM USING PARTIAL INFERENCE AND BIDIRECTIONAL FUSION 626
Daxin Tan, The Chinese University of Hong Kong, Hong Kong SAR China; Liqun Deng, Yu Ting Yeung, Xin Jiang, Xiao Chen, Huawei Noah's Ark Lab, China; Tan Lee, The Chinese University of Hong Kong, Hong Kong SAR China

Paper #6: ON-THE-FLY DATA AUGMENTATION FOR TEXT-TO-SPEECH STYLE TRANSFER 634
Raymond Chung, Brian Mak, The Hong Kong University of Science and Technology, Australia

OTHER TOPICS 2

Paper #1: ON PROSODY MODELING FOR ASR+TTS BASED VOICE CONVERSION 642
Wen-Chin Huang, Tomoki Hayashi, Nagoya University, Japan; Xinjian Li, Shinji Watanabe, Carnegie Mellon University, United States; Tomoki Toda Toda, Nagoya University, Japan

Paper #2: MANDARIN ELECTROLARYNGEAL SPEECH VOICE CONVERSION WITH SEQUENCE-TO-SEQUENCE MODELING 650
Ming-Chi Yen, Academia Sinica, Taiwan; Wen-Chin Huang, Kazuhiro Kobayashi, Nagoya University, Japan; Yu-Huai Peng, Academia Sinica, Taiwan; Shu-Wei Tsai, National Cheng Kung University Hospital, Taiwan; Yu Tsao, Academia Sinica, Taiwan; Tomoki Toda, Nagoya University, Japan; Jyh-Shing Jang, National Taiwan University, Taiwan; Hsin-Min Wang, Academia Sinica, Taiwan

Paper #3: ATTENTION-BASED SCALING ADAPTATION FOR TARGET SPEECH EXTRACTION 658
Jiangyu Han, Shanghai Normal University, China; Wei Rao, Tencent Ethereal Audio Lab, China; Yanhua Long, Shanghai Normal University, China; Jiaen Liang, Unisound AI Technology Co., Ltd., China

Paper #4: GLMSNET: SINGLE CHANNEL SPEECH SEPARATION FRAMEWORK IN NOISY AND REVERBERANT ENVIRONMENTS 663
Huiyu Shi, Tsinghua University, China; Xi Chen, Lenovo Research, China; Tianlong Kong, Shouyi Yin, Tsinghua University, China; Peng Ouyang, TsingMicro Co. Ltd, China

Paper #5: MULTI-TASK AUDIO SOURCE SEPARATION 671
Lu Zhang, Harbin Institute of Technology, Shenzhen, China; Chenxing Li, Feng Deng, Xiaorui Wang, Kuai Shou Technology Co., Beijing, China

Paper #6: CONFERENCINGSPEECH CHALLENGE: TOWARDS FAR-FIELD MULTI-CHANNEL SPEECH ENHANCEMENT FOR VIDEO CONFERENCING 679
Wei Rao, Tencent Ethereal Audio Lab, China; Yihui Fu, Yanxin Hu, Northwestern Polytechnical University, China; Xin Xu, Beijing Shell Shell Technology Co., LTD., China; Yvkai Jv, Northwestern Polytechnical University, China; Jiangyu Han, Zhongjie Jiang, Tencent Ethereal Audio Lab, China; Lei Xie, Northwestern Polytechnical University, China; Yannan Wang, Tencent Ethereal Audio Lab, China; Shinji Watanabe, Carnegie Mellon University, United States; Zheng-Hua Tan, Aalborg University, Denmark; Hui Bu, Beijing Shell Shell Technology Co., LTD., China; Tao Yu, Shidong Shang, Tencent Ethereal Audio Lab, United States

OTHER TOPICS 3

Paper #1: VOXCELEB ENRICHMENT FOR AGE AND GENDER RECOGNITION 687
Khaled Hechmi, Università degli Studi di Milano-Bicocca, Italy; Trung Ngo Trung, Ville Hautamaki, Tomi Kinnunen, University of Eastern Finland, Finland

Paper #2: ENABLING ZERO-SHOT MULTILINGUAL SPOKEN LANGUAGE TRANSLATION WITH LANGUAGE-SPECIFIC ENCODERS AND DECODERS 694
Carlos Escolano, Marta R. Costa-jussà, José A. R. Fonollosa, Universitat Politècnica de Catalunya, Spain; Carlos Segura, Telefónica Research, Spain

Paper #3: DIVE: END-TO-END SPEECH DIARIZATION VIA ITERATIVE SPEAKER EMBEDDING 702
Neil Zeghidour, Olivier Teboul, David Grangier, Google, France

Paper #4: AC-VC: NON-PARALLEL LOW LATENCY PHONETIC POSTERIORGRAMS BASED VOICE CONVERSION 710
Damien Ronssin, Ecole Polytechnique Fédérale de Lausanne, Switzerland; Milos Cernak, Logitech Europe SA, Switzerland

Paper #5: TARGET LANGUAGE EXTRACTION AT MULTILINGUAL COCKTAIL PARTIES 717
Marvin Borsdorf, University of Bremen, Germany; Haizhou Li, National University of Singapore, Singapore; Tanja Schultz, University of Bremen, Germany

Paper #6: ATTENTION BASED MODEL FOR SEGMENTAL PRONUNCIATION ERROR DETECTION 725
Jose Antonio Lopez Saenz, Md Asif Jalal, Rosanna Milner, Thomas Hain, University of Sheffield, United Kingdom

TEXT-TO-SPEECH SYNTHESIS 2

Paper #1: EVALUATION OF TRANSLATION WITH SYNTHESIS: HIGH-RESOURCE LANGUAGES AND DIALECTAL VARIANTS 733
Elizabeth Salesky, Johns Hopkins University, United States; Julian Mäder, Severin Klinger, ETH Zürich, Switzerland

Paper #2: DIFFSVC: A DIFFUSION PROBABILISTIC MODEL FOR SINGING VOICE CONVERSION 741
Songxiang Liu, Yuewen Cao, The Chinese University of Hong Kong, Hong Kong SAR China; Dan Su, Tencent AI Lab, China; Helen Meng, The Chinese University of Hong Kong, Hong Kong SAR China

Paper #3: LOW-LATENCY INCREMENTAL TEXT-TO-SPEECH SYNTHESIS WITH DISTILLED CONTEXT PREDICTION NETWORK 749
Takaaki Saeki, Shinnosuke Takamichi, Hiroshi Saruwatari, The University of Tokyo, Japan

Paper #4: HEARING FACES: TARGET SPEAKER TEXT-TO-SPEECH SYNTHESIS FROM A FACE 757
Björn Plüster, Cornelius Weber, Leyuan Qu, Stefan Wermter, University of Hamburg, Germany

Paper #5: ANALYSIS OF CONVERSATIONAL SPEECH WITH APPLICATION TO VOICE ADAPTATION 765
Bhagyashree Mukherjee, Anusha Prakash, Hema A Murthy, Indian Institute of Technology, Madras, India

Paper #6: VIBRATO LEARNING IN MULTI-SINGER SINGING VOICE SYNTHESIS 773
Ruolan Liu, Xue Wen, Chunhui Lu, Liming Song, Samsung Research China-Beijing, China; June Sig Sung, Samsung Electronics, South Korea

AUTOMATIC SPEECH RECOGNITION 9

Paper #1: TREE-CONSTRAINED POINTER GENERATOR FOR END-TO-END CONTEXTUAL SPEECH RECOGNITION 780
Guangzhi Sun, Chao Zhang, Phil Woodland, Cambridge University Engineering Department, United Kingdom

Paper #2: COMPARING THE BENEFIT OF SYNTHETIC TRAINING DATA FOR VARIOUS AUTOMATIC SPEECH RECOGNITION ARCHITECTURES 788
Nick Rossenbach, Mohammad Zeineldeen, RWTH Aachen University / AppTek GmbH, Germany; Benedikt Hilmes, RWTH Aachen University, Germany; Ralf Schlüter, Hermann Ney, RWTH Aachen University / AppTek GmbH, Germany

Paper #3: AUDIO-VISUAL SPEECH RECOGNITION IS WORTH 32X32X8 VOXELS 796
Dmitriy Serdyuk, Otavio Braga, Olivier Siohan, Google, United States

Paper #4: LEVERAGING LINGUISTIC KNOWLEDGE FOR ACCENT ROBUSTNESS OF END-TO-END MODELS 803
Andrea Carmantini, Steve Renals, Peter Bell, University of Edinburgh, United Kingdom

AUTOMATIC SPEECH RECOGNITION 10

Paper #1: AN EVALUATION BENCHMARK FOR AUTOMATIC SPEECH RECOGNITION OF GERMAN-ENGLISH CODE-SWITCHING 811
Abbas Khosravani, Philip N. Garner, Idiap Research Institute, Switzerland; Alexandros Lazaridis, Swisscom AG, Switzerland

Paper #2: LEARNING TO TRANSLATE LOW-RESOURCED SWISS GERMAN DIALECTAL SPEECH INTO STANDARD GERMAN TEXT 817
Abbas Khosravani, Philip N. Garner, Idiap Research Institute, Switzerland; Alexandros Lazaridis, Swisscom AG, Switzerland

Paper #3: CHANNELAUGMENT: IMPROVING GENERALIZATION OF MULTI-CHANNEL ASR BY TRAINING WITH INPUT CHANNEL RANDOMIZATION 824
Marco Gaudesi, Felix Weninger, Dushyant Sharma, Puming Zhan, Nuance Communications, Italy

Paper #4: IMPROVING SPEECH RECOGNITION ON NOISY SPEECH VIA SPEECH ENHANCEMENT WITH MULTI-DISCRIMINATORS CYCLEGAN 830
Chia-Yu Li, Thang Vu, Institute for Natural Language Processing (IMS), University of Stuttgart, Germany

SPEECH-TO-SPEECH TRANSLATION 1

Paper #1: ATTENTIVE CONTEXTUAL CARRYOVER FOR MULTI-TURN END-TO-END SPOKEN LANGUAGE UNDERSTANDING 837
Kai Wei, Thanh Tran, Feng-Ju Chang, Kanthashree Mysore Sathyendra, Thejaswi Muniyappa, Jing Liu, Anirudh Raju, Ross McGowan, Nathan Susanj, Ariya Rastrow, Grant P. Strimel, Amazon, United States

Paper #2: X-SHOT: LEARNING TO RANK VOICE APPLICATIONS VIA CROSS-LOCALE SHARD-BASED CO-TRAINING 845
Zheng Gao, Radhika Arava, Qian Hu, Xibin Gao, Wei Xiao, Thahir Mohamed, Mohamed AbdelHady, Amazon Alexa AI, United States

Paper #3: INTENT RECOGNITION AND UNSUPERVISED SLOT IDENTIFICATION FOR LOW RESOURCED SPOKEN DIALOG SYSTEMS 853
Akshat Gupta, Olivia Deng, Akruti Kushwaha, Saloni Mittal, William Zeng, SaiKrishna Rallabandi, Alan Black, Carnegie Mellon University, United States

Paper #4: ACTION ITEM DETECTION IN MEETINGS USING PRETRAINED TRANSFORMERS 861
Kishan Sachdeva, Joshua Maynez, Olivier Siohan, Google Inc., United States

Paper #5: DECIDING WHETHER TO ASK CLARIFYING QUESTIONS IN LARGE-SCALE SPOKEN LANGUAGE UNDERSTANDING 869
Joo-Kyung Kim, Guoyin Wang, Sungjin Lee, Young-Bum Kim, Amazon, United States

OTHER TOPICS 4

Paper #1: HUMAN-AGENT COLLABORATION STRATEGIES FOR VISION-GROUNDED INSTRUCTION FOLLOWING 877
Guan-Lin Chao, Ian Lane, Carnegie Mellon University, United States

Paper #2: UNCERTAINTY-AWARE PSEUDO-LABELING FOR SPOKEN LANGUAGE ASSESSMENT 885
Binghuai Lin, Liyuan Wang, Tencent Technology Co., Ltd, China

Paper #3: AN END-TO-END FAR-FIELD KEYWORD SPOTTING SYSTEM WITH NEURAL BEAMFORMING 892
Xuan Ji, Lu Lu, Fuming Fang, Jianbo Ma, Lei Zhu, Jinke Li, Dongdi Zhao, Ming Liu, Feijun Jiang, Alibaba Group, China

Paper #4: IMPROVING REVERBERANT SPEECH SEPARATION WITH SYNTHETIC ROOM IMPULSE RESPONSES 900
Rohith Aralikatti, Anton Ratnarajah, Zhenyu Tang, Dinesh Manocha, University of Maryland, College Park, India

Paper #5: HASA-NET: A NON-INTRUSIVE HEARING-AID SPEECH ASSESSMENT NETWORK 907
Hsin-Tien Chiang, Research Center for Information Technology Innovation, Academia Sinica, Taiwan; Yi-Chiao Wu, Information Technology Center, Nagoya University,, Japan; Cheng Yu, Research Center for Information Technology Innovation, Academia Sinica, Taiwan; Tomoki Toda, Information Technology Center, Nagoya University, Japan; Hsin-Min Wang, Institute of Information Science, Academia Sinica, Taiwan, Taiwan; Yih-Chun Hu, University of Illinois at Urbana-Champaign, United States; Yu Tsao, Research Center for Information Technology Innovation, Academia Sinica, Taiwan

Paper #6: LAYER-WISE ANALYSIS OF A SELF-SUPERVISED SPEECH REPRESENTATION MODEL 914
Ankita Pasad, Ju-Chieh Chou, Karen Livescu, Toyota Technological Institute at Chicago, United States

SPEECH-TO-SPEECH TRANSLATION 2

Paper #1: FAST-MD: FAST MULTI-DECODER END-TO-END SPEECH TRANSLATION WITH NON-AUTOREGRESSIVE HIDDEN INTERMEDIATES 922
Hirofumi Inaguma, Kyoto University, Japan; Siddharth Dalmia, Brian Yan, Shinji Watanabe, Carnegie Mellon University, United States

Paper #2: CYCLEGEAN: CYCLE GENERATIVE ENHANCED ADVERSARIAL NETWORK FOR VOICE CONVERSION 930
Xulong Zhang, Jianzong Wang, Ning Cheng, Ping An Technology (Shenzhen) Co., Ltd., China; Edward Xiao, Aquinas International Academy, United States; Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

Paper #3: TGAVC: IMPROVING AUTOENCODER VOICE CONVERSION WITH TEXT-GUIDED AND ADVERSARIAL TRAINING 938
Huaizhen Tang, University of Science and Technology of China, China; Xulong Zhang, Jianzong Wang, Ning Cheng, Zhen Zeng, Ping An Technology (Shenzhen) Co., Ltd., China; Edward Xiao, Aquinas International Academy, CA, USA, United States; Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

Paper #4: RECONSTRUCTING DUAL LEARNING FOR NEURAL VOICE CONVERSION USING RELATIVELY FEW SAMPLES 946
Aolan Sun, Jianzong Wang, Ning Cheng, Ping An Technology (Shenzhen) Co., Ltd., China; Methawee Tantrawenith, Ping An Technology (Shenzhen) Co., Ltd., Tsinghua University, China; Zhiyong Wu, Helen Meng, Tsinghua University, The Chinese University of Hong Kong, China; Edward Xiao, Aquinas International Academy, United States; Jing Xiao, Ping An Technology (Shenzhen) Co., Ltd., China

SPOKEN DIALOG SYSTEMS 1

Paper #1: MULTITASK GENERATIVE ADVERSARIAL IMITATION LEARNING FOR MULTI-DOMAIN DIALOGUE SYSTEM 954
Chuan-En Hsu, Mahdin Rohmatillah, Jen-Tzung Chien, National Yang Ming Chiao Tung University, Taiwan

Paper #2: AUDIO EMBEDDINGS HELP TO LEARN BETTER DIALOGUE POLICIES 962
Asier López Zorrilla, M. Inés Torres, University of the Basque Country, Spain; Heriberto Cuayáhuitl, University of Lincoln, United Kingdom

Paper #3: WHAT DOES THE USER WANT? INFORMATION GAIN FOR HIERARCHICAL DIALOGUE POLICY OPTIMISATION 969
Christian Geishauser, Heinrich Heine University Düsseldorf, Germany; Songbo Hu, University of Cambridge, United Kingdom; Hsien-chin Lin, Nurul Lubis, Michael Heck, Shutong Feng, Carel van Niekerk, Milica Gašić, Heinrich Heine University Düsseldorf, Germany

Paper #4: DIALOGUE STRATEGY ADAPTATION TO NEW ACTION SETS USING MULTI-DIMENSIONAL MODELLING 977
Simon Keizer, Norbert Braunschweiler, Svetlana Stoyanchev, Rama Doddipatla, Toshiba Europe Limited, United Kingdom

ASR IN ADVERSE ENVIRONMENTS 2

Paper #1: SEMI-SUPERVISED TRANSFER LEARNING FOR LANGUAGE EXPANSION OF END-TO-END SPEECH RECOGNITION MODELS TO LOW-RESOURCE LANGUAGES 984
Jiyeon Kim, Mehul Kumar, Dhananjaya Gowda, Abhinav Garg, Chanwou Kim, Samsung Research, South Korea

Paper #2: A COMPARISON OF STREAMING MODELS AND DATA AUGMENTATION METHODS FOR ROBUST SPEECH RECOGNITION 989
Jiyeon Kim, Samsung Research, South Korea

Paper #3: 3D SPATIAL FEATURES FOR MULTI-CHANNEL TARGET SPEECH SEPARATION 996
Rongzhi Gu, Peking University Shenzhen Graduate School, China; Shi-Xiong Zhang, Meng Yu, Dong Yu, Tencent AI Lab, United States

Paper #4: FAR-FIELD SPEECH RECOGNITION BASED ON COMPLEX-VALUED NEURAL NETWORKS AND INTER-FRAME SIMILARITY DIFFERENCE METHOD 1003
Yifan Guo, Yifan Chen, University of Chinese Academy of Sciences, China; Gaofeng Cheng, Pengyuan Zhang, Yonghong Yan, Chinese Academy of Sciences, China

AUTOMATIC SPEECH RECOGNITION 11

Paper #1: SCALING END-TO-END MODELS FOR LARGE-SCALE MULTILINGUAL ASR 1011
Bo Li, Ruoming Pang, Tara Sainath, Anmol Gulati, Yu Zhang, James Qin, Parisa Haghani, W. Ronny Huang, Min Ma, Junwen Bai, Google, United States

Paper #2: DECOUPLING RECOGNITION AND TRANSCRIPTION IN MANDARIN ASR 1019
Jiahong Yuan, Xingyu Cai, Baidu Research USA, United States; Dongji Gao, Johns Hopkins University, United States; Renjie Zheng, Liang Huang, Kenneth Church, Baidu Research USA, United States

Paper #3: ON LATTICE-FREE BOOSTED MMI TRAINING OF HMM AND CTC-BASED FULL-CONTEXT ASR MODELS 1026
Xiaohui Zhang, Vimal Manohar, David Zhang, Frank Zhang, Yangyang Shi, Nayan Singhal, Julian Chan, Fuchun Peng, Yatharth Saraf, Mike Seltzer, Facebook AI, United States

MULTILINGUAL MODELS AND RESOURCES 1

Paper #1: MULTILINGUAL AND CROSSLINGUAL SPEECH RECOGNITION USING PHONOLOGICAL-VECTOR BASED PHONE EMBEDDINGS 1034
Chengrui Zhu, Keyu An, Huahuan Zheng, Zhijian Ou, Tsinghua University, China

Paper #2: IN PURSUIT OF BABEL - MULTILINGUAL END-TO-END SPOKEN LANGUAGE UNDERSTANDING 1042
Markus Müller, Samridhi Choudhary, Clement Chung, Athanasios Mouchtaris, Siegfried Kunzmann, Amazon Alexa AI, United States

Paper #3: CROSS-LINGUAL TRANSFER FOR SPEECH PROCESSING USING ACOUSTIC LANGUAGE SIMILARITY 1050
Peter Wu, Jiatong Shi, Yifan Zhong, Shinji Watanabe, Alan Black, Carnegie Mellon University, United States

NEW APPLICATIONS OF ASR 1

Paper #1: AN ASR N-BEST TRANSCRIPT NEURAL RANKING MODEL FOR SPOKEN CONTENT RETRIEVAL 1058
Yasufumi Moriya, Gareth Jones, Dublin City University, Ireland

Paper #2: TOWARDS ROBUST MISPRONUNCIATION DETECTION AND DIAGNOSIS FOR L2 ENGLISH LEANERS WITH ACCENT-MODULATING METHODS 1065
Shao-Wei Fan Jiang, Bi-Cheng Yan, Tien-Hong Lo, Fu-An Chao, Berlin Chen, National Taiwan Normal University, Taiwan

Paper #3: MULTI-GRANULARITY ANNOTATION OF INSTANTANEOUS INTELLIGIBILITY OF LEARNERS' UTTERANCES BASED ON SHADOWING TECHNIQUES 1071
Chuanbo Zhu, Ryo Hakoda, Daisuke Saito, Nobuaki Minematsu, The University of Tokyo, Japan; Noriko Nakanishi, Kobe Gakuin University, Japan; Tazuko Nishimura, The University of Tokyo, Japan

Paper #4: APPLYING X-VECTORS ON PATHOLOGICAL SPEECH AFTER LARYNX REMOVAL 1079
Ralph Scheuerer, Nuremberg Institute of Technology, Germany; Tino Haderlein, Elmar Nöth, Universität Erlangen-Nürnberg, Germany; Tobias Bocklet, Nuremberg Institute of Technology, Germany

Paper #5: MULTI-TASK LANGUAGE MODELING FOR IMPROVING SPEECH RECOGNITION OF RARE WORDS 1087
Chao-Han Huck Yang, Georgia Institute of Technology, United States; Linda Liu, Ankur Gandhe, Yile Gu, Anirudh Raju, Denis Filimonov, Ivan Bulyko, Amazon Alexa, United States

Paper #6: LEVERAGING PRE-TRAINED REPRESENTATIONS TO IMPROVE ACCESS TO UNTRANSCRIBED SPEECH FROM ENDANGERED LANGUAGES 1094
Nay San, Stanford University, United States; Martijn Bartelds, University of Groningen, Netherlands; Mitchell Browne, University of Queensland, Australia; Lily Clifford, Stanford University, United States; Fiona Gibson, Batchelor Institute, Australia; John Mansfield, University of Melbourne, Australia; David Nash, Jane Simpson, Australian National University, Australia; Myfany Turpin, University of Sydney, Australia; Maria Vollmer, University of Freiburg, Germany; Sasha Wilmoth, University of Melbourne, Australia; Dan Jurafsky, Stanford University, United States

SPEAKER & LANGUAGE RECOGNITION 3

Paper #1: SPEECHNAS: STATE-OF-THE-ART LARGE-SCALE SPEAKER VERIFICATION WITH NEURAL ARCHITECTURE SEARCH 1102
Wentao Zhu, Shun Lu, Tianlong Kong, Jixiang Li, Dawei Zhang, Feng Deng, Xiaorui Wang, Sen Yang, Ji Liu, Kuaishou Technology, United States

Paper #2: STUDYING SQUEEZE-AND-EXCITATION USED IN CNN FOR SPEAKER VERIFICATION 1110
Mickael Rouvier, Pierre-Michel Bousquet, Avignon University, France

Paper #3: HYBRID NETWORK WITH MULTI-LEVEL GLOBAL-LOCAL STATISTICS POOLING FOR ROBUST TEXT-INDEPENDENT SPEAKER RECOGNITION 1116
Woo Hyun Kang, Jahangir Alam, Abderrahim Fathan, CRIM, Canada

Paper #4: IMPROVING SPEAKER IDENTIFICATION FOR SHARED DEVICES BY ADAPTING EMBEDDINGS TO SPEAKER SUBSETS 1124
Zhenning Tan, Yuguang Yang, Eunjung Han, Andreas Stolcke, Amazon Alexa Speech, United States

Paper #5: PARAMETERIZED CHANNEL NORMALIZATION FOR FAR-FIELD DEEP SPEAKER VERIFICATION 1132
Xuechen Liu, MULTISPEECH, Inria Nancy Grand Est; School of Computing, University of Eastern Finland, France; Md Sahidullah, MULTISPEECH, Inria Nancy Grand Est, France; Tomi Kinnunen, School of Computing, University of Eastern Finland, Finland

Paper #6: OVERLAP-AWARE LOW-LATENCY ONLINE SPEAKER DIARIZATION BASED ON END-TO-END LOCAL SEGMENTATION 1139
Juan Manuel Coria, Université Paris-Saclay CNRS, LISN, France; Hervé Bredin, IRIT, Université de Toulouse, CNRS, France; Sahar Ghannay, Sophie Rosset, Université Paris-Saclay CNRS, LISN, France

SPOKEN DIALOG SYSTEMS 2 AND TEXT-TO-SPEECH SYNTHESIS 3

Paper #1: "HOW ROBUST R U?": EVALUATING TASK-ORIENTED DIALOGUE SYSTEMS ON SPOKEN CONVERSATIONS 1147
Seokhwan Kim, Yang Liu, Di Jin, Alexandros Papangelis, Behnam Hedayatnia, Karthik Gopalakrishnan, Dilek Hakkani-Tur, Amazon Alexa AI, United States

Paper #2: ON-DEVICE NEURAL SPEECH SYNTHESIS 1155
Sivanand Achanta, Albert Antony, Ladan Golipour, Jiangchuan Li, Tuomo Raitio, Ramya Rasipuram, Francesco Rossi, Jennifer Shi, Jaimin Upadhyay, David Winarsky, Hepeng Zhang, Apple, United States

Paper #3: TOWARDS USING HETEROGENEOUS RELATION GRAPHS FOR END-TO-END TTS 1162
Amrith Setlur, Aman Madaan, Tanmay Parekh, Yiming Yang, Alan W Black, Carnegie Mellon University, United States

OTHER TOPICS 5

Paper #1: WORD-LEVEL CONFIDENCE ESTIMATION FOR RNN TRANSDUCERS 1170
Mingqiu Wang, Hagen Soltau, Laurent El Shafey, Izhak Shafran, Google Inc, United States

Paper #2: USING SELF ATTENTION DNNS TO DISCOVER PHONEMIC FEATURES FOR AUDIO DEEP FAKE DETECTION 1178
Hira Yasin Dhamyal, Ayesha Ali, Ihsan Ayyub Qazi, Agha Ali Raza, Lahore University of Management Sciences, Pakistan

Paper #3: JOINT PREDICTION OF TRUECASING AND PUNCTUATION FOR CONVERSATIONAL SPEECH IN LOW-RESOURCE SCENARIOS 1185
Raghavendra Reddy Pappagari, Piotr Zelasko, Johns Hopkins University, United States; Agnieszka Mikołajczyk, Piotr Pezik, VoiceLab, Poland; Najim Dehak, Johns Hopkins University, United States