

Shi-wook Lee, National Institute of Advanced  
 Industrial Science and Technology, Japan  
 Yun Lei, Alibaba Group  
 Cheung-Chi Leung, Alibaba Group  
 Bo Li, Google Inc.  
 Jinyu Li, Microsoft  
 Ming Li, Duke Kunshan University  
 Shang-Wen Li, Apple Inc.  
 Hank Liao, Google Inc.  
 Bing Liu, Facebook  
 Gang Liu, Alibaba Group (U.S.) Inc.  
 Lijuan Liu, IFLYTEK CO., LTD  
 Xunying Liu, Chinese University of Hong Kong  
 Yuzong Liu, Amazon  
 Eduardo Lleida Solano, Universidad de Zaragoza  
 Timo Lohrenz, Technische Universität Braunschweig  
 Jaime Lorenzo-Trueba, Amazon  
 Heng Lu, Tencent American  
 Liang Lu, Microsoft  
 Xugang Lu, National Institute of Information and  
 Communications Technology, Japan  
 Jeff Ma, Raytheon BBN Technologies, Cambridge,  
 USA  
 Mathew Magimai Doss, Idiap Research Institute  
 Sébastien Le Maguer, Adapt Centre, Trinity College  
 Dublin  
 Rannierey Maia, Federal University of Santa Catarina  
 Man-wai Mak, The Hong Kong Polytechnic University  
 Kazunori Mano, Shibaura Institute of Technology  
 Takashi Masuko, Preferred Networks, Inc.  
 Ryo Masumura, NTT Corporation  
 Marco Matassoni, Fondazione Bruno Kessler  
 Pavel Matejka, Brno University of Technology  
 Shigeki Matsuda, Cotoba Design  
 Tomoko Matsui, The Institute of Statistical  
 Mathematics  
 Iurii Matveev, ITMO University  
 Erik McDermott, Google Inc.  
 Mitchell McLaren, SRI International  
 Zhong Meng, Microsoft  
 Florian Metze, Carnegie Mellon University  
 Antonio Miguel, University of Zaragoza  
 Rosanna Milner, University of Sheffield  
 Masato Mimura, Kyoto University  
 Yasuhiro Minami, The University of Electro-  
 Communications  
 Vikramjit Mitra, Apple Inc.  
 Abdelrahman Mohamed, Facebook AI Research  
 (FAIR)  
 Alexis Moinet, Amazon  
 Juan M Montero, Universidad Politécnica de Madrid  
 Niko Moritz, Mitsubishi Electric Research Labs  
 Petr Motlicek, Idiap Research Institute  
 Athanasios Mouchtaris, Amazon.com  
 Seiichi Nakagawa, Chubu University  
 Satoshi Nakamura, Nara Institute of Science and  
 Technology  
 Tomohiro Nakatani, NTT Corporation  
 Hiroaki Nanjo, Kyoto University  
 Arun Narayanan, Google Inc.  
 Eva Navas, University of the Basque Country (UPV/  
 EHU)  
 Hermann Ney, RWTH Aachen University  
 Trung Hieu Nguyen, Alibaba Group  
 Chongjia Ni, Alibaba Group  
 Masafumi Nishida, Shizuoka University  
 Takanobu Nishiura, Ritsumeikan University  
 Atsunori Ogawa, NTT Corporation  
 Tetsuji Ogawa, Waseda University  
 Alfonso Ortega, University of Zaragoza  
 Douglas O'Shaughnessy, INRS-EMT  
 Slim Ouni, University of Lorraine  
 Mukund Padmanabhan, Renaissance Technologies  
 LLC  
 Ankita Pasad, Toyota Technological Institute at  
 Chicago  
 Vijayaditya Peddinti, Google Inc  
 Michael Picheny, IBM Research  
 Oldrich Plchot, Brno University of Technology  
 Alexandros Potamianos, National Technical University  
 of Athens  
 Gerasimos Potamianos, University of Thessaly  
 Dan Povey, Johns Hopkins University  
 Rohit Prabhavalkar, Google  
 Yanmin Qian, Shanghai Jiao Tong University  
 Yao Qian, Educational Testing Service  
 Anton Ragni, University of Cambridge  
 Tuomo Raitio, Apple Inc.  
 Anirudh Raju, Amazon Alexa  
 Sai Krishna Rallabandi, Carnegie Mellon University  
 Bhuvana Ramabhadran, Google  
 Vikram Ramanarayanan, ETS Research and University  
 of California, San Francisco  
 Daniel Ramos, Universidad Autonoma de Madrid  
 Steve Renals, University of Edinburgh  
 Douglas Reynolds, MIT Lincoln Laboratory  
 Giuseppe Riccardi, University of Trento  
 Korin Richmond, University of Edinburgh  
 Luca Rigazio, Totemic  
 Johan Rohdin, Brno University of Technology  
 Andrew Rosenberg, Google  
 Omid Sadjadi, National Institute of Standards and  
 Technology  
 Md Sahidullah, Inria

Sakriani Sakti, Nara Institute of Science and Technology  
 Giampiero Salvi, KTH Royal Institute of Technology  
 George Saon, IBM  
 Thomas Schaaf, 3M | M\*Modal  
 Ralf Schlüter, RWTH Aachen University  
 Björn Schuller, Imperial College London  
 Frank Seide, Microsoft Research  
 Gregory Sell, Johns Hopkins University  
 Abhinav Sethy, Amazon  
 Izhak Shafran, Google Inc.  
 Koichi Shinoda, Tokyo Institute of Technology  
 Yusuke Shinohara, NTT Corporation  
 Takahiro Shinozaki, Tokyo Institute of Technology  
 Suwon Shon, Massachusetts Institute of Technology  
 Sabato Marco Siniscalchi, Kore University of Enna  
 Olivier Siohan, Google Inc.  
 Hagen Soltau, Google Inc.  
 Themis Stafylakis, Omilia - Conversational Intelligence  
 John Steinberg, MIT Lincoln Laboratory  
 Evgeny Stepanov, University of Trento  
 Andreas Stolcke, Microsoft Research  
 Sebastian Stueker, Karlsruhe Institute of Technology  
 Dan Su, Tencent  
 Ming Sun, Amazon.com  
 Yuuki Tachioka, Denso IT Laboratory  
 Tetsuya Takiguchi, Kobe University  
 Hao Tang, MIT  
 Qingming Tang, Toyota Technological Institute at Chicago  
 Tadahiro Taniguchi, Ritsumeikan University  
 Fei Tao, Uber  
 Jianhua Tao, National Library of Pattern Recognition, Institute of Automation Chinese Academy of Science  
 Louis Ten Bosch, Radboud University  
 Samuel Thomas, IBM Research  
 Tomoki Toda, Nagoya University  
 Massimiliano Todisco, EURECOM  
 Rong Tong, Alibaba Group  
 Pedro Torres-Carrasquillo, MIT Lincoln Laboratory  
 Trang Tran, University of Washington  
 Isabel Trancoso, INESC-ID / IST, University of Lisbon  
 Jan Trmal, Johns Hopkins University  
 Yu Tsao, Academia Sinica  
 Gokhan Tur, Uber AI  
 Oytun Turk, Sensory Inc  
 Zoltán Tüske, IBM Research AI  
 David van Leeuwen, Radboud University Nijmegen  
 Amparo Varona, University of the Basque Country  
 Dimitra Vergyri, SRI International  
 Karel Vesely, Brno University of Technology  
 Ville Vestman, University of Eastern Finland  
 Emmanuel Vincent, Inria  
 Carl Vogel, Trinity College Dublin  
 Oliver Walter, Paderborn University  
 Patrick Wambacq, KU Leuven  
 Guangsen Wang, Tencent AI Lab  
 Jun Wang, Tencent  
 Lei Wang, Institute for Infocomm Research  
 Xin Wang, National Institute of Informatics  
 Nigel Ward, University of Texas at El Paso  
 Shinji Watanabe, Johns Hopkins University  
 Janguo Wei, Tianjin University  
 Chao Weng, Tencent  
 Bo Wu, Tencent  
 Chung-Hsien Wu, National Cheng Kung University  
 Xiong Xiao, Microsoft  
 Lei Xie, Northwestern Polytechnical University  
 Haihua Xu, Temasek Laboratories, Nanyang Technological University, Singapore  
 Hitoshi Yamamoto, NEC Corporation  
 Yonghong Yan, Institute of Acoustics, Chinese Academy of Sciences  
 Bayya Yegnanarayana, International Institute of Information Technology Hyderabad  
 Koichiro Yoshino, Nara Institute of Science and Technology  
 Changhuai You, Institute for Infocomm Research  
 Dong Yu, Tencent AI Lab  
 Kai Yu, Shanghai Jiao Tong University  
 Vicky Zayats, University of Washington  
 Heiga Zen, Google AI  
 Shengkui Zhao, Alibaba Group  
 Yong Zhao, Microsoft  
 Yunxin Zhao, University of Missouri  
 Imed Zitouni, Microsoft

# TABLE OF CONTENTS

## ADV: ASR IN ADVERSE ENVIRONMENTS

### ADV.1: SPEAKER-AWARE SPEECH-TRANSFORMER.....222

Zhiyun Fan, Institute of Automation, Chinese Academy of Sciences, China; Jie Li, Kwai, China; Shiyu Zhou, Bo Xu, Institute of Automation, Chinese Academy of Sciences, China

### ADV.2: SPEECH SEPARATION USING SPEAKER INVENTORY.....230

Peidong Wang, Ohio State University, United States; Zhuo Chen, Xiong Xiao, Zhong Meng, Takuya Yoshioka, Tianyan Zhou, Liang Lu, Jinyu Li, Microsoft, United States

### ADV.3: MIMO-SPEECH: END-TO-END MULTI-CHANNEL MULTI-SPEAKER SPEECH RECOGNITION .....237

Xuankai Chang, Johns Hopkins University, United States; Wangyou Zhang, Yanmin Qian, Shanghai Jiao Tong University, China; Jonathan Le Roux, Mitsubishi Electric Research Laboratories (MERL), United States; Shinji Watanabe, Johns Hopkins University, United States

### ADV.4: JOINT DISTRIBUTION LEARNING IN THE FRAMEWORK OF VARIATIONAL AUTOENCODERS FOR FAR-FIELD SPEECH ENHANCEMENT .....245

Mahesh Kumar Chelimilla, Shashi Kumar, Shakti P. Rath, Samsung Research Institute Bangalore, India

### ADV.5: ANALYZING LARGE RECEPTIVE FIELD CONVOLUTIONAL NETWORKS FOR DISTANT SPEECH RECOGNITION .....252

Salar Jafarlou Jafarlou, Soheil Khorram, Vinay Kothapally, John H.L. Hansen, University of Texas at Dallas, United States

### ADV.6: FASNET: LOW-LATENCY ADAPTIVE BEAMFORMING FOR MULTI-MICROPHONE AUDIO PROCESSING .....260

Yi Luo, Columbia University, United States; Enea Ceolini, University of Zurich and ETH Zurich, Switzerland; Cong Han, Columbia University, United States; Shih-Chii Liu, University of Zurich and ETH Zurich, Switzerland; Nima Mesgarani, Columbia University, United States

### ADV.7: DOMAIN ADAPTATION VIA TEACHER-STUDENT LEARNING FOR END-TO-END SPEECH RECOGNITION .....268

Zhong Meng, Jinyu Li, Yashesh Gaur, Yifan Gong, Microsoft Corporation, United States

### ADV.8: ADVANCES IN ONLINE AUDIO-VISUAL MEETING TRANSCRIPTION .....276

Takuya Yoshioka, Igor Abramovski, Cem Aksoylar, Zhuo Chen, Moshe David, Dimitrios Dimitriadis, Yifan Gong, Ilya Gurvich, Xuedong Huang, Yan Huang, Aviv Hurvitz, Li Jiang, Sharon Koubi, Eyal Krupka, Ido Leichter, Changliang Liu, Partha Parthasarathy, Alon Vinnikov, Lingfeng Wu, Xiong Xiao, Wayne Xiong, Huaming Wang, Zhenghao Wang, Jun Zhang, Yong Zhao, Tianyan Zhou, Microsoft, United States

## ASR-1: AUTOMATIC SPEECH RECOGNITION I

### ASR-1.1: INCREMENTAL LATTICE DETERMINIZATION FOR WFST DECODERS .....1

Zhehuai Chen, Shanghai Jiao Tong University, China; Mahsa Yarmohammadi, Hainan Xu, Johns Hopkins University, United States; Hang Lv, Lei Xie, Northwestern Polytechnical University, China; Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, United States

<b>ASR-1.2: A COMPARISON OF TRANSFORMER AND LSTM ENCODER DECODER MODELS FOR ASR</b>	<b>8</b>
<i>Albert Zeyer, Parnia Bahar, Kazuki Irie, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</i>	
<b>ASR-1.3: A DROPOUT-BASED SINGLE MODEL COMMITTEE APPROACH FOR ACTIVE LEARNING IN ASR</b>	<b>16</b>
<i>Jiayi Fu, Kuang Ru, Zhuiyi Technology Company, China</i>	
<b>ASR-1.4: PERSONALIZATION OF END-TO-END SPEECH RECOGNITION ON MOBILE DEVICES FOR NAMED ENTITIES</b>	<b>23</b>
<i>Khe Chai Sim, Francoise Beaufays, Arnaud Benard, Dhruv Guliani, Andreas Kabel, Nikhil Khare, Tamar Lucassen, Petr Zadrazil, Harry Zhang, Leif Johnson, Giovanni Motta, Lillian Zhou, Google, United States</i>	
<b>ASR-1.5: SIMULTANEOUS SPEECH RECOGNITION AND SPEAKER DIARIZATION FOR MONAURAL DIALOGUE RECORDINGS WITH TARGET-SPEAKER ACOUSTIC MODELS</b>	<b>31</b>
<i>Naoyuki Kanda, Shota Horiguchi, Yusuke Fujita, Yawen Xue, Kenji Nagamatsu, Hitachi, Ltd., Japan; Shinji Watanabe, Johns Hopkins University, United States</i>	
<b>ASR-1.6: INTEGRATING SOURCE-CHANNEL AND ATTENTION-BASED SEQUENCE-TO-SEQUENCE MODELS FOR SPEECH RECOGNITION</b>	<b>39</b>
<i>Qiuqia Li, Chao Zhang, Phil Woodland, University of Cambridge, United Kingdom</i>	
<b>ASR-1.7: AN INVESTIGATION INTO THE EFFECTIVENESS OF ENHANCEMENT IN ASR TRAINING AND TEST FOR CHIME-5 DINNER PARTY TRANSCRIPTION</b>	<b>47</b>
<i>Catalin Zorila, Toshiba Cambridge Research Laboratory, United Kingdom; Christoph Boeddeker, Paderborn University, Germany; Rama Doddipatla, Toshiba Cambridge Research Laboratory, United Kingdom; Reinhold Haeb-Umbach, Paderborn University, Germany</i>	
<b>ASR-1.8: STATE-OF-THE-ART SPEECH RECOGNITION USING MULTI-STREAM SELF-ATTENTION WITH DILATED 1D CONVOLUTIONS</b>	<b>54</b>
<i>Kyu Han, Ramon Prieto, Tao Ma, ASAPP, Inc., United States</i>	
<b>ASR-1.9: HIGHLY EFFICIENT NEURAL NETWORK LANGUAGE MODEL COMPRESSION USING SOFT BINARIZATION TRAINING</b>	<b>62</b>
<i>Rao Ma, Qi Liu, Kai Yu, Shanghai Jiao Tong University, China</i>	
<b>ASR-1.10: IMPROVED MULTI-STAGE TRAINING OF ONLINE ATTENTION-BASED ENCODER-DECODER MODELS</b>	<b>70</b>
<i>Abhinav Garg, Dhananjaya Gowda, Ankur Kumar, Kwangyoun Kim, Mehul Kumar, Chanwoo Kim, Samsung Research, Korea (South)</i>	
<b>ASR-1.11: LEAD2GOLD: TOWARDS EXPLOITING THE FULL POTENTIAL OF NOISY TRANSCRIPTIONS FOR SPEECH RECOGNITION</b>	<b>78</b>
<i>Adrien Dufraux, Facebook AI Research, France; Emmanuel Vincent, INRIA, France; Awni Hannun, Facebook AI Research, United States; Armelle Brun, Université de Lorraine, France; Matthijs Douze, Facebook AI Research, France</i>	
<b>ASR-1.12: ORTHOGONALITY CONSTRAINED MULTI-HEAD ATTENTION FOR KEYWORD SPOTTING</b>	<b>86</b>
<i>Mingu Lee, Jinkyu Lee, Hye Jin Jang, Byeonggeun Kim, Wonil Chang, Kyuwoong Hwang, Qualcomm AI Research, Korea (South)</i>	

<b>ASR-1.13: LEARNING BETWEEN DIFFERENT TEACHER AND STUDENT MODELS IN ASR</b>	<b>93</b>
<i>Jeremy Heng Meng Wong, Microsoft, United States; Mark John Francis Gales, Yu Wang, University of Cambridge, United Kingdom</i>	
<b>ASR-1.14: A UNIFIED ENDPOINTER USING MULTITASK AND MULTIDOMAIN TRAINING</b>	<b>100</b>
<i>Shuo-Yiin Chang, Bo Li, Gabor Simko, Google, United States</i>	
<b>ASR-1.15: DOMAIN EXPANSION IN DNN-BASED ACOUSTIC MODELS FOR ROBUST SPEECH RECOGNITION</b>	<b>107</b>
<i>Shahram Ghorbani, Soheil Khorram, John H.L. Hansen, University of Texas at Dallas, United States</i>	
<b>ASR-1.16: IMPROVING RNN TRANSDUCER MODELING FOR END-TO-END SPEECH RECOGNITION</b>	<b>114</b>
<i>Jinyu Li, Rui Zhao, Hu Hu, Yifan Gong, Microsoft, United States</i>	
<b>ASR-1.17: SIMPLE GATED CONVNET FOR SMALL FOOTPRINT ACOUSTIC MODELING</b>	<b>122</b>
<i>Lukas Lee, Jinhwan Park, Wonyong Sung, Seoul National University, Korea (South)</i>	
<b>ASR-1.18: GANS FOR CHILDREN: A GENERATIVE DATA AUGMENTATION STRATEGY FOR CHILDREN SPEECH RECOGNITION</b>	<b>129</b>
<i>Peiyao Sheng, Zhuolin Yang, Yanmin Qian, Shanghai Jiao Tong University, China</i>	
<b>ASR-1.19: ESPRESSO: A FAST END-TO-END NEURAL SPEECH RECOGNITION TOOLKIT</b>	<b>136</b>
<i>Yiming Wang, Tongfei Chen, Hainan Xu, Shuoyang Ding, Johns Hopkins University, United States; Hang Lv, Northwestern Polytechnical University, China; Yiwen Shao, Johns Hopkins University, United States; Nanyun Peng, University of Southern California, United States; Lei Xie, Northwestern Polytechnical University, China; Shinji Watanabe, Sanjeev Khudanpur, Johns Hopkins University, United States</i>	
<b>ASR-2: AUTOMATIC SPEECH RECOGNITION II</b>	
<b>ASR-2.1: TRAINING LANGUAGE MODELS FOR LONG-SPAN CROSS-SENTENCE EVALUATION</b>	<b>419</b>
<i>Kazuki Irie, Albert Zeyer, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</i>	
<b>ASR-2.2: TRANSFORMER ASR WITH CONTEXTUAL BLOCK PROCESSING</b>	<b>427</b>
<i>Emiru Tsunoo, Yosuke Kashiwagi, Toshiyuki Kumakura, Sony Corporation, Japan; Shinji Watanabe, Johns Hopkins University, United States</i>	
<b>ASR-2.3: A DENSITY RATIO APPROACH TO LANGUAGE MODEL FUSION IN END-TO-END AUTOMATIC SPEECH RECOGNITION</b>	<b>434</b>
<i>Erik McDermott, Hasim Sak, Ehsan Variani, Google Inc, United States</i>	
<b>ASR-2.4: IMPROVING GRAPHEME-TO-PHONEME CONVERSION BY INVESTIGATING COPYING MECHANISM IN RECURRENT ARCHITECTURES</b>	<b>442</b>
<i>Abhishek Niranjan, Mahaboob Ali Basha Shaik, Samsung Research and Development Institute, India</i>	



<b>ASR-2.5: A COMPARATIVE STUDY ON TRANSFORMER VS RNN IN SPEECH APPLICATIONS .....</b>	<b>449</b>
<i>Shigeki Karita, NTT Communication Science Laboratories, Japan; Nanxin Chen, Johns Hopkins University, United States; Tomoki Hayashi, Nagoya University, Japan; Takaaki Hori, Mitsubishi Electric Research Laboratories (MERL), United States; Hirofumi Inaguma, Kyoto University, Japan; Ziyang Jiang, Johns Hopkins University, United States; Masao Someki, Nagoya University, Japan; Nelson Enrique Yalta Soplin, Waseda University, Japan; Ryuichi Yamamoto, LINE Corporation, Japan; Xiaofei Wang, Shinji Watanabe, Johns Hopkins University, United States; Takenori Yoshimura, Nagoya University, Japan; Wangyou Zhang, Shanghai Jiao Tong University, China</i>	
<b>ASR-2.6: FROM SENONES TO CHENONES: TIED CONTEXT-DEPENDENT GRAPHEMES FOR HYBRID SPEECH RECOGNITION .....</b>	<b>457</b>
<i>Duc Le, Xiaohui Zhang, Weiyi Zheng, Christian Fuegen, Geoffrey Zweig, Michael L. Seltzer, Facebook, United States</i>	
<b>ASR-2.7: ATTENTION-BASED SPEECH RECOGNITION USING GAZE INFORMATION .....</b>	<b>465</b>
<i>Osamu Segawa, Chubu Electric Power Co., Inc., Japan; Tomoki Hayashi, Kazuya Takeda, Nagoya University, Japan</i>	
<b>ASR-2.8: LISTENING WHILE SPEAKING AND VISUALIZING: IMPROVING ASR THROUGH MULTIMODAL CHAIN .....</b>	<b>471</b>
<i>Johannes Effendi, Nara Institute of Science and Technology / RIKEN Center for Advanced Intelligence Project AIP, Japan; Andros Tjandra, 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</i>	
<b>ASR-2.9: EMBEDDINGS FOR DNN SPEAKER ADAPTIVE TRAINING .....</b>	<b>479</b>
<i>Joanna Rownicka, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
<b>ASR-2.10: LANGUAGE MODEL BOOTSTRAPPING USING NEURAL MACHINE TRANSLATION FOR CONVERSATIONAL SPEECH RECOGNITION .....</b>	<b>487</b>
<i>Surabhi Punjabi, Harish Arsikere, Sri Garimella, Amazon, India</i>	
<b>ASR-2.11: SPEAKER AND LANGUAGE AWARE TRAINING FOR END-TO-END ASR.....</b>	<b>494</b>
<i>Shubham Bansal, Karan Malhotra, Sriram Ganapathy, Indian Institute of Science, Bangalore, India</i>	
<b>ASR-2.12: DATA AUGMENTATION BASED ON VOWEL STRETCH FOR IMPROVING CHILDREN'S SPEECH RECOGNITION .....</b>	<b>502</b>
<i>Tohru Nagano, Takashi Fukuda, Masayuki Suzuki, Gakuto Kurata, IBM, Japan</i>	
<b>ASR-2.13: MIXED BANDWIDTH ACOUSTIC MODELING LEVERAGING KNOWLEDGE DISTILLATION .....</b>	<b>509</b>
<i>Takashi Fukuda, Samuel Thomas, IBM, Japan</i>	
<b>ASR-2.14: ON TEMPORAL CONTEXT INFORMATION FOR HYBRID BLSTM-BASED PHONEME RECOGNITION .....</b>	<b>516</b>
<i>Timo Lohrenz, Maximilian Strake, Tim Fingscheidt, Technische Universität Braunschweig, Germany</i>	
<b>ASR-2.15: EXPLORING MODEL UNITS AND TRAINING STRATEGIES FOR END-TO-END SPEECH RECOGNITION .....</b>	<b>524</b>
<i>Mingkun Huang, Yi Zhou Lu, Shanghai Jiao Tong University, China; Lan Wang, Chinese Academy of Sciences, China; Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China</i>	

<b>ASR-2.16: QUERY-BY-EXAMPLE ON-DEVICE KEYWORD SPOTTING .....</b>	<b>532</b>
<i>Byeonggeun Kim, Mingu Lee, Jinkyu Lee, Yeonseok Kim, Kyuwoong Hwang, Qualcomm, Korea (South)</i>	
<b>ASR-2.17: SMALL-FOOTPRINT KEYWORD SPOTTING WITH GRAPH .....</b>	<b>539</b>
<b>CONVOLUTIONAL NETWORK</b>	
<i>Xi Chen, Shouyi Yin, Tsinghua University, China; Dandan Song, Peng Ouyang, TsingMicro Co. Ltd., China; Leibo Liu, Shaojun Wei, Tsinghua University, China</i>	
<b>ASR-2.18: SIMPLIFIED LSTMS FOR SPEECH RECOGNITION.....</b>	<b>547</b>
<i>George Saon, Zoltan Tuske, Kartik Audhkhasi, Brian Kingsbury, Michael Picheny, Samuel Thomas, IBM, United States</i>	
<b>ASR-2.19: GENERALIZED LARGE-CONTEXT LANGUAGE MODELS BASED ON .....</b>	<b>554</b>
<b>FORWARD-BACKWARD HIERARCHICAL RECURRENT ENCODER-DECODER MODELS</b>	
<i>Ryo Masumura, Mana Ihori, Tomohiro Tanaka, Itsumi Saito, Kyosuke Nishida, Takanobu Oba, NTT Corporation, Japan</i>	
<b>ASR-2.20: END-TO-END TRAINING OF A LARGE VOCABULARY END-TO-END .....</b>	<b>562</b>
<b>SPEECH RECOGNITION SYSTEM</b>	
<i>Chanwoo Kim, Sungsoo Kim, Kwangyoun Kim, Mehul Kumar, Jiyeon Kim, Kyungmin Lee, Changwoo Han, Abhinav Garg, Eunhyang Kim, Minkyoo Shin, Shatrughan Singh, Larry Heck, Dhananjaya Gowda, Samsung Research, Korea (South)</i>	
 <b>ASR-3: AUTOMATIC SPEECH RECOGNITION III</b>	
<b>ASR-3.1: SEMI-SUPERVISED TRAINING AND DATA AUGMENTATION FOR ADAPTATION .....</b>	<b>867</b>
<b>OF AUTOMATIC BROADCAST NEWS CAPTIONING SYSTEMS</b>	
<i>Yinghui Huang, Samuel Thomas, Masayuki Suzuki, Zoltan Tuske, Larry Sansone, Michael Picheny, IBM, United States</i>	
<b>ASR-3.2: ONLINE BATCH NORMALIZATION ADAPTATION FOR AUTOMATIC SPEECH .....</b>	<b>875</b>
<b>RECOGNITION</b>	
<i>Franco Mana, Felix Weninger, Roberto Gemello, Puming Zhan, NUANCE Communications, Italy</i>	
<b>ASR-3.3: SPEAKER ADAPTIVE TRAINING USING MODEL AGNOSTIC META-LEARNING.....</b>	<b>881</b>
<i>Ondrej Klejch, Joachim Fainberg, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
<b>ASR-3.4: A COMPARISON OF END-TO-END MODELS FOR LONG-FORM SPEECH .....</b>	<b>889</b>
<b>RECOGNITION</b>	
<i>Chung-Cheng Chiu, Wei Han, Yu Zhang, Ruoming Pang, Sergey Kishchenko, Google Inc, United States; Patrick Nguyen, Grab Technologies, United States; Hagen Soltau, Arun Narayanan, Hank Liao, Shuyuan Zhang, Anjuli Kannan, Rohit Prabhavalkar, Zhifeng Chen, Tara Sainath, Yonghui Wu, Google Inc, United States</i>	
<b>ASR-3.5: ACOUSTIC MODEL ADAPTATION FROM RAW WAVEFORMS WITH SINCNET.....</b>	<b>897</b>
<i>Joachim Fainberg, Ondrej Klejch, Erfan Loweimi, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
<b>ASR-3.6: RECURRENT NEURAL NETWORK TRANSDUCER FOR AUDIO-VISUAL .....</b>	<b>905</b>
<b>SPEECH RECOGNITION</b>	
<i>Takaki Makino, Hank Liao, Google Inc., United States; Yannis Assael, Brendan Shillingford, DeepMind, United Kingdom; Basilio Garcia, Otavio Braga, Olivier Siohan, Google Inc., United States</i>	

**ASR-3.7: EXPLICIT ALIGNMENT OF TEXT AND SPEECH ENCODINGS FOR .....913**  
**ATTENTION-BASED END-TO-END SPEECH RECOGNITION**

*Jennifer Drexler, James Glass, Massachusetts Institute of Technology, United States*

**ASR-3.8: RECOGNIZING LONG-FORM SPEECH USING STREAMING END-TO-END .....920**  
**MODELS**

*Arun Narayanan, Rohit Prabhavalkar, Chung-Cheng Chiu, David Rybach, Tara Sainath, Trevor Strohman, Google Inc., United States*

**ASR-3.9: LEVERAGING LANGUAGE ID IN MULTILINGUAL END-TO-END SPEECH .....928**  
**RECOGNITION**

*Austin Waters, Neeraj Gaur, Parisa Haghani, Pedro Moreno, Zhongdi Qu, Google, United States*

**ASR-3.10: STREAMING END-TO-END SPEECH RECOGNITION WITH JOINT .....936**  
**CTC-ATTENTION BASED MODELS**

*Niko Moritz, Takaaki Hori, Jonathan Le Roux, Mitsubishi Electric Research Laboratories (MERL), United States*

**ASR-3.11: MONOTONIC RECURRENT NEURAL NETWORK TRANSDUCER AND .....944**  
**DECODING STRATEGIES**

*Anshuman Tripathi, Han Lu, Hasim Sak, Hagen Soltau, Google, United States*

**ASR-3.12: CHARACTER-AWARE ATTENTION-BASED END-TO-END SPEECH .....949**  
**RECOGNITION**

*Zhong Meng, Yashesh Gaur, Jinyu Li, Yifan Gong, Microsoft Corporation, United States*

**ASR-3.13: ATTENTION BASED ON-DEVICE STREAMING SPEECH RECOGNITION .....956**  
**WITH LARGE SPEECH CORPUS**

*Kwangyoun Kim, Kyungmin Lee, Dhananjaya Gowda, Junmo Park, Sungsoo Kim, Sichen Jin, Young-Yoon Lee, Jinsu Yeo, Daehyun Kim, Seokyeong Jung, Jungin Lee, Myoungji Han, Chanwoo Kim, Samsung Electronics, Korea (South)*

**ASR-3.14: ZERO-SHOT CODE-SWITCHING ASR AND TTS WITH MULTILINGUAL .....964**  
**MACHINE SPEECH CHAIN**

*Sahoko Nakayama, Nara Institute of Science and Technology / RIKEN Center for Advanced Intelligence Project AIP, Japan; Andros Tjandra, 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*

**ASR-3.15: END-TO-END CODE-SWITCHING ASR FOR LOW-RESOURCED .....972**  
**LANGUAGE PAIRS**

*Xianghu Yue, Beijing Institute of Technology / National University of Singapore, Singapore; Grandee Lee, Emre Yilmaz, National University of Singapore, Singapore; Fang Deng, Beijing Institute of Technology, China; Haizhou Li, National University of Singapore, Singapore*

**ASR-3.16: UNSUPERVISED ADAPTATION OF ACOUSTIC MODELS FOR ASR USING .....980**  
**UTTERANCE-LEVEL EMBEDDINGS FROM SQUEEZE AND EXCITATION NETWORKS**

*Hardik Sailor, Salil Deena, Md Asif Jalal, Rasa Lileikyte, Thomas Hain, University of Sheffield, United Kingdom*



<b>ASR-3.17: POWER-LAW NONLINEARITY WITH MAXIMALLY UNIFORM</b>	<b>988</b>
<b>DISTRIBUTION CRITERION FOR IMPROVED NEURAL NETWORK TRAINING IN</b>	
<b>AUTOMATIC SPEECH RECOGNITION</b>	
<i>Chanwoo Kim, Mehul Kumar, Kwangyoun Kim, Dhananjaya Gowda, Samsung Research, Korea (South)</i>	
<b>ASR-3.18: SPEECH RECOGNITION WITH AUGMENTED SYNTHESIZED SPEECH</b>	<b>996</b>
<i>Andrew Rosenberg, Yu Zhang, Bhuvana Ramabhadran, Ye Jia, Pedro Moreno, Yonghui Wu, Zelin Wu, Google, United States</i>	
<b>MLP: MULTILINGUAL LANGUAGE PROCESSING</b>	
<b>MLP.1: SECOND LANGUAGE TRANSFER LEARNING IN HUMANS AND MACHINES</b>	<b>1040</b>
<b>USING IMAGE SUPERVISION</b>	
<i>Kiran Praveen, Indian Institute of Science, India; Anshul Gupta, International Institute of Information Technology, Bangalore, India; Akshara Soman, Sriram Ganapathy, Indian Institute of Science, India</i>	
<b>MLP.2: ZERO-SHOT PRONUNCIATION LEXICONS FOR CROSS-LANGUAGE</b>	<b>1048</b>
<b>ACOUSTIC MODEL TRANSFER</b>	
<i>Matthew Wiesner, Oliver Adams, David Yarowsky, Jan Trmal, Sanjeev Khudanpur, Johns Hopkins University, United States</i>	
<b>NEW: NEW APPLICATIONS OF ASR</b>	
<b>NEW.1: IMPROVING SPEECH ENHANCEMENT WITH PHONETIC EMBEDDING</b>	<b>645</b>
<b>FEATURES</b>	
<i>Bo Wu, Meng Yu, Lianwu Chen, Mingjie Jin, Dan Su, Dong Yu, Tencent, China</i>	
<b>NEW.2: DETECTING DECEPTION IN POLITICAL DEBATES USING ACOUSTIC AND</b>	<b>652</b>
<b>TEXTUAL FEATURES</b>	
<i>Daniel Kopev, Sofia University, Bulgaria; Ahmed Ali, Hamad Bin Khalifa University, Qatar; Ivan Koychev, Sofia University, Bulgaria; Preslav Nakov, Hamad Bin Khalifa University, Qatar</i>	
<b>NEW.3: END-TO-END OVERLAPPED SPEECH DETECTION AND SPEAKER</b>	<b>660</b>
<b>COUNTING WITH RAW WAVEFORM</b>	
<i>Wangyou Zhang, Man Sun, Shanghai Jiao Tong University, China; Lan Wang, Chinese Academy of Sciences, China; Yanmin Qian, Shanghai Jiao Tong University, China</i>	
<b>NEW.4: TIME DOMAIN AUDIO VISUAL SPEECH SEPARATION</b>	<b>667</b>
<i>Jian Wu, Northwestern Polytechnical University, China; Yong Xu, Shi-Xiong Zhang, Lianwu Chen, Meng Yu, Tencent AI Lab, United States; Lei Xie, Northwestern Polytechnical University, China; Dong Yu, Tencent AI Lab, United States</i>	
<b>NEW.5: SPEECH REVEALS FUTURE RISK OF DEVELOPING DEMENTIA:</b>	<b>674</b>
<b>PREDICTIVE DEMENTIA SCREENING FROM BIOGRAPHIC INTERVIEWS</b>	
<i>Jochen Weiner, University of Bremen, Germany; Claudia Frankenberg, Johannes Schröder, University of Heidelberg, Germany; Tanja Schultz, University of Bremen, Germany</i>	
<b>NEW.6: IMPROVING FUNDAMENTAL FREQUENCY GENERATION IN</b>	<b>682</b>
<b>EMG-TO-SPEECH CONVERSION USING A QUANTIZATION APPROACH</b>	
<i>Lorenz Diener, Tejas Umesh, Tanja Schultz, University of Bremen, Germany</i>	

**NEW.7: TOWARDS REAL-TIME MISPRONUNCIATION DETECTION IN KIDS' SPEECH .....690**

*Peter Plantinga, Eric Fosler-Lussier, Ohio State University, United States*

**S2S: SPEECH-TO-SPEECH TRANSLATION**

**S2S.1: MULTILINGUAL END-TO-END SPEECH TRANSLATION .....570**

*Hirofumi Inaguma, Kyoto University, Japan; Kevin Duh, Johns Hopkins University, United States; Tatsuya Kawahara, Kyoto University, Japan; Shinji Watanabe, Johns Hopkins University, United States*

**S2S.2: NEURAL MACHINE TRANSLATION WITH ACOUSTIC EMBEDDING .....578**

*Takatomo Kano, 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*

**S2S.3: ONE-TO-MANY MULTILINGUAL END-TO-END SPEECH TRANSLATION .....585**

*Mattia Antonino Di Gangi, Fondazione Bruno Kessler / University of Trento, Italy; Matteo Negri, Marco Turchi, Fondazione Bruno Kessler, Italy*

**S2S.4: SPEECH-TO-SPEECH TRANSLATION BETWEEN UNTRANSCRIBED UNKNOWN LANGUAGES .....593**

*Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology, Japan*

**SDR: SPOKEN DOCUMENT RETRIEVAL**

**SDR.1: ENHANCED BERT-BASED RANKING MODELS FOR SPOKEN DOCUMENT RETRIEVAL .....601**

*Hsiao-Yun Lin, Tien-Hong Lo, Berlin Chen, National Taiwan Normal University, Taiwan*

**SDR.2: VIRTUAL ADVERSARIAL TRAINING FOR DS-CNN BASED SMALL-FOOTPRINT KEYWORD SPOTTING .....607**

*Xiong Wang, Northwestern Polytechnical University, China; Sining Sun, Northwestern Polytechnical University / Tencent, China; Lei Xie, Northwestern Polytechnical University, China*

**SDR.3: VERIFYING DEEP KEYWORD SPOTTING DETECTION WITH ACOUSTIC WORD EMBEDDINGS .....613**

*Yougen Yuan, Northwestern Polytechnical University, China; Zhiqiang Lv, Shen Huang, Tencent Research, China; Lei Xie, Northwestern Polytechnical University, China*

**SDR.4: MULTILINGUAL BOTTLENECK FEATURES FOR QUERY BY EXAMPLE SPOKEN TERM DETECTION .....621**

*Dhananjay Ram, Lesly Miculicich, Hervé Bourlard, Idiap Research Institute / École Polytechnique Fédérale de Lausanne, Switzerland*

**SDR.5: ADDITIONAL SHARED DECODER ON SIAMESE MULTI-VIEW ENCODERS FOR LEARNING ACOUSTIC WORD EMBEDDINGS .....629**

*Myunghun Jung, Hyungjun Lim, Jahyun Goo, Youngmoon Jung, Hoirin Kim, Korea Advanced Institute of Science and Technology (KAIST), Korea (South)*

**SDR.6: EFFICIENT FREE KEYWORD DETECTION BASED ON CNN AND END-TO-END CONTINUOUS DP-MATCHING .....637**

*Tomohiro Tanaka, Takahiro Shinozaki, Tokyo Institute of Technology, Japan*

## **SDS: SPOKEN DIALOG SYSTEMS**

### **SDS.1: ROBUST BELIEF STATE SPACE REPRESENTATION FOR STATISTICAL .....1055 DIALOGUE MANAGERS USING DEEP AUTOENCODERS**

*Fotios Lygerakis, Vassilios Diakouloulas, Michail Lagoudakis, Technical University of Crete, Greece; Kotti Margarita, Toshiba Research Cambridge, United Kingdom*

### **SDS.2: IMPROVING SPEECH-BASED END-OF-TURN DETECTION VIA .....1062 CROSS-MODAL □ REPRESENTATION LEARNING WITH PUNCTUATED TEXT DATA**

*Ryo Masumura, Mana Ihori, Tomohiro Tanaka, Atsushi Ando, Ryo Ishii, Takanobu Oba, Ryuichiro Higashinaka, NTT Corporation, Japan*

### **SDS.3: DIALOGUE ENVIRONMENTS ARE DIFFERENT FROM GAMES: .....1070 INVESTIGATING VARIANTS OF DEEP Q-NETWORKS FOR DIALOGUE POLICY**

*Yu-An Wang, Yun-Nung Chen, National Taiwan University, Taiwan*

### **SDS.4: EFFICIENT SEMI-SUPERVISED LEARNING FOR NATURAL LANGUAGE .....1077 UNDERSTANDING BY OPTIMIZING DIVERSITY**

*Eunah Cho, He Xie, Alexa AI, United States; John Lalor, University of Notre Dame, United States; Varun Kumar, William M. Campbell, Alexa AI, United States*

## **SLR-1: SPEAKER/LANGUAGE RECOGNITION I**

### **SLR-1.1: JOINT OPTIMIZATION OF CLASSIFICATION AND CLUSTERING FOR DEEP .....284 SPEAKER EMBEDDING**

*Zhiming Wang, Kaisheng Yao, Shuo Fang, Xiaolong Li, Ant Financial, Inc., China*

### **SLR-1.2: EXPLORING EFFECTIVE DATA AUGMENTATION WITH TDNN-LSTM .....291 NEURAL NETWORK EMBEDDING FOR SPEAKER RECOGNITION**

*Chien-Lin Huang, PingAn AI Lab, United States*

### **SLR-1.3: END-TO-END NEURAL SPEAKER DIARIZATION WITH SELF-ATTENTION.....296**

*Yusuke Fujita, Naoyuki Kanda, Shota Horiguchi, Yawen Xue, Kenji Nagamatsu, Hitachi, Ltd., Japan; Shinji Watanabe, Johns Hopkins University, United States*

### **SLR-1.4: A CROSS-CORPUS STUDY ON SPEECH EMOTION RECOGNITION .....304**

*Rosanna Milner, Md Asif Jalal, University of Sheffield, United Kingdom; Raymond W. M. Ng, Emotech Labs, United Kingdom; Thomas Hain, University of Sheffield, United Kingdom*

### **SLR-1.5: ADVERSARIAL ATTACKS ON SPOOFING COUNTERMEASURES OF .....312 AUTOMATIC SPEAKER VERIFICATION**

*Songxiang Liu, Chinese University of Hong Kong, China; Haibin Wu, Hung-yi Lee, National Taiwan University, Taiwan; Helen Meng, Chinese University of Hong Kong, China*

### **SLR-1.6: SPOKEN LANGUAGE IDENTIFICATION USING BIDIRECTIONAL LSTM .....320 BASED LID SEQUENTIAL SENONES.**

*Muralikrishna H, Pulkit Sapra, Anuksha Jain, Dileep Aroor Dinesh, Indian Institute of Technology, Mandi, India*

<b>SLR-1.7: TIME-DOMAIN SPEAKER EXTRACTION NETWORK.....</b>	<b>327</b>
<i>Chenglin Xu, Nanyang Technological University, Singapore; Wei Rao, National University of Singapore, Singapore; Eng Siong Chng, Nanyang Technological University, Singapore; Haizhou Li, National University of Singapore, Singapore</i>	
<b>SLR-1.8: SHORT UTTERANCE COMPENSATION IN SPEAKER VERIFICATION VIA ..... COSINE-BASED TEACHER-STUDENT LEARNING OF SPEAKER EMBEDDINGS</b>	<b>335</b>
<i>Jee-weon Jung, Hee-Soo Heo, Hye-jin Shim, Ha-Jin Yu, University of Seoul, Korea (South)</i>	
<b>SLR-1.9: NOVEL ENHANCED TEAGER ENERGY BASED CEPSTRAL COEFFICIENTS ..... FOR REPLAY SPOOF DETECTION</b>	<b>342</b>
<i>Rajul Acharya, Hemant A. Patil, Harsh Kotta, Dhirubhai Ambani Institute of Information and Communication Technology, India</i>	
<b>SLR-1.10: SYLLABLE-DEPENDENT DISCRIMINATIVE LEARNING FOR SMALL ..... FOOTPRINT TEXT-DEPENDENT SPEAKER VERIFICATION</b>	<b>350</b>
<i>Junyi Peng, Yuexian Zou, Peking University, China; Na Li, Deyi Tuo, Dan Su, Meng Yu, Chunlei Zhang, Dong Yu, Tencent, China</i>	
<b>SLR-1.11: LATENT SPACE REPRESENTATION FOR MULTI-TARGET SPEAKER ..... DETECTION AND IDENTIFICATION WITH A SPARSE DATASET USING TRIPLET NEURAL NETWORKS</b>	<b>358</b>
<i>Kin Wai Cheuk, Balamurali B T, Gemma Roig, Dorien Herremans, Singapore University of Technology and Design, Singapore</i>	
<b>SLR-1.12: SELF-ADAPTIVE SOFT VOICE ACTIVITY DETECTION USING DEEP ..... NEURAL NETWORKS FOR ROBUST SPEAKER VERIFICATION</b>	<b>365</b>
<i>Youngmoon Jung, Yeunju Choi, Hoirin Kim, Korea Advanced Institute of Science and Technology (KAIST), Korea (South)</i>	
<b>SLR-1.13: SPHEREDIAR: AN EFFECTIVE SPEAKER DIARIZATION SYSTEM FOR ..... MEETING DATA</b>	<b>373</b>
<i>Tuomas Kaseva, Aku Rouhe, Mikko Kurimo, Aalto University, Finland</i>	
<b>SLR-1.14: BAYESIAN ADVERSARIAL LEARNING FOR SPEAKER RECOGNITION.....</b>	<b>381</b>
<i>Jen-Tzung Chien, Chun-Lin Kuo, National Chiao Tung University, Taiwan</i>	
<b>SLR-1.15: AN INVESTIGATION OF LSTM-CTC BASED JOINT ACOUSTIC MODEL ..... FOR INDIAN LANGUAGE IDENTIFICATION</b>	<b>389</b>
<i>Tirusha Mandava, Ravi Kumar Vuddagiri, Hari Krishna Vydan, Anil Kumar Vuppala, IIIT Hyderabad, India</i>	
<b>SLR-1.16: A MULTI PURPOSE AND LARGE SCALE SPEECH CORPUS IN PERSIAN AND ..... ENGLISH FOR SPEAKER AND SPEECH RECOGNITION: THE DEEPMINE DATABASE</b>	<b>397</b>
<i>Hossein Zeinali, Lukáš Burget, Jan Cernocky, Brno University of Technology, Czech Republic</i>	
<b>SLR-1.17: NATIVE LANGUAGE IDENTIFICATION FROM RAW WAVEFORMS USING ..... DEEP CONVOLUTIONAL NEURAL NETWORKS WITH ATTENTIVE POOLING</b>	<b>403</b>
<i>Rutuja Ubale, Vikram Ramanarayanan, Yao Qian, Keelan Evanini, Chee Wee Leong, Chong Min Lee, Educational Testing Service, United States</i>	

<b>SLR-1.18: SPEAKER VERIFICATION WITH APPLICATION-AWARE BEAMFORMING .....</b>	<b>411</b>
<i>Ladislav Mošner, Oldřich Plchot, Johan Rohdin, Lukáš Burget, Jan Černocký, Brno University of Technology, Czech Republic</i>	
 <b>SLR-2: SPEAKER/LANGUAGE RECOGNITION II</b>	
<b>SLR-2.1: INCORPORATING PRIOR KNOWLEDGE INTO SPEAKER DIARIZATION AND .....697</b>	
<b>LINKING FOR IDENTIFYING COMMON SPEAKER</b>	
<i>Tsun-Yat Leung, Lahiru Samarakoon, Albert Y.S. Lam, Fano Labs, China</i>	
<b>SLR-2.2: LOGISTIC SIMILARITY METRIC LEARNING VIA AFFINITY MATRIX FOR .....704</b>	
<b>TEXT-INDEPENDENT SPEAKER VERIFICATION</b>	
<i>Junyi Peng, Rongzhi Gu, Yuexian Zou, Peking University, China</i>	
<b>SLR-2.3: LOW-RESOURCE DOMAIN ADAPTATION FOR SPEAKER RECOGNITION .....710</b>	
<b>USING CYCLE-GANS</b>	
<i>Phani Sankar Nidadavolu, Saurabh Kataria, Jesús Villalba, Najim Dehak, Johns Hopkins University, United States</i>	
<b>SLR-2.4: CNN WITH PHONETIC ATTENTION FOR TEXT-INDEPENDENT SPEAKER .....718</b>	
<b>VERIFICATION</b>	
<i>Tianyan Zhou, Yong Zhao, Jinyu Li, Yifan Gong, Jian Wu, Microsoft Corporation, United States</i>	
<b>SLR-2.5: PROBING THE INFORMATION ENCODED IN X-VECTORS.....726</b>	
<i>Desh Raj, David Snyder, Daniel Povey, Sanjeev Khudanpur, The Johns Hopkins University, United States</i>	
<b>SLR-2.6: IN-THE-WILD END-TO-END DETECTION OF SPEECH AFFECTING .....734</b>	
<b>DISEASES</b>	
<i>Joana Correia, Carnegie Mellon University / INESC-ID, Instituto Superior Tecnico / University of Lisbon, United States; Isabel Trancoso, INESC-ID, Instituto Superior Tecnico / University of Lisbon, Portugal; Bhiksha Raj, Carnegie Mellon University, United States</i>	
<b>SLR-2.7: OPTIMIZING NEURAL NETWORK EMBEDDINGS USING A PAIR-WISE .....742</b>	
<b>LOSS FOR TEXT-INDEPENDENT SPEAKER VERIFICATION</b>	
<i>Hira Dharmyal, Tianyan Zhou, Bhiksha Raj, Rita Singh, Carnegie Mellon University, United States</i>	
<b>SLR-2.8: TOWARDS CONTROLLING FALSE ALARM — MISS TRADE-OFF IN .....749</b>	
<b>PERCEPTUAL SPEAKER COMPARISON VIA NON-NEUTRAL LISTENING TASK</b>	
<b>FRAMING</b>	
<i>Rosa González Hautamäki, Tomi H. Kinnunen, University of Eastern Finland, Finland</i>	
<b>SLR-2.9: DOVER: A METHOD FOR COMBINING DIARIZATION OUTPUTS.....757</b>	
<i>Andreas Stolcke, Takuya Yoshioka, Microsoft, United States</i>	
 <b>SLU: SPOKEN LANGUAGE UNDERSTANDING</b>	
<b>SLU.1: USING VERY DEEP CONVOLUTIONAL NEURAL NETWORKS TO .....764</b>	
<b>AUTOMATICALLY DETECT PLAGIARIZED SPOKEN RESPONSES</b>	
<i>Xinhao Wang, Keelan Evanini, Yao Qian, Klaus Zechner, Educational Testing Service, United States</i>	



<b>SLU.2: SPOKEN MULTIPLE-CHOICE QUESTION ANSWERING USING MULTIMODAL CONVOLUTIONAL NEURAL NETWORKS</b>	<b>772</b>
<i>Shang-Bao Luo, National Taiwan University of Science and Technology, Taiwan; Hung-Shin Lee, Institute of Information Science, Academia Sinica, Taiwan; Kuan-Yu Chen, National Taiwan University of Science and Technology, Taiwan; Hsin-Min Wang, Institute of Information Science, Academia Sinica, Taiwan</i>	
<b>SLU.3: TRANSFER LEARNING FOR CONTEXT-AWARE SPOKEN LANGUAGE UNDERSTANDING</b>	<b>779</b>
<i>Qian Chen, Zhu Zhuo, Wen Wang, Qiuyun Xu, Alibaba Group, China</i>	
<b>SLU.4: EMOCEPTION: AN INCEPTION INSPIRED EFFICIENT SPEECH EMOTION RECOGNITION NETWORK</b>	<b>787</b>
<i>Chirag Singh, Abhay Kumar, Ajay Nagar, Suraj Tripathi, Promod Yenigalla, Samsung R&amp;D Institute India - Bangalore, India</i>	
<b>SLU.5: A COMPARATIVE STUDY ON END-TO-END SPEECH TO TEXT TRANSLATION</b>	<b>792</b>
<i>Parnia Bahar, Tobias Bieschke, Hermann Ney, RWTH Aachen University, Germany</i>	
<b>SLU.6: JOINT LEARNING OF WORD AND LABEL EMBEDDINGS FOR SEQUENCE LABELLING IN SPOKEN LANGUAGE UNDERSTANDING</b>	<b>800</b>
<i>Jiewen Wu, Institute for Infocomm Research, Singapore; Luis Fernando D'Haro, Universidad Politecnica de Madrid, Spain; Nancy F. Chen, Pavitra Krishnaswamy, Institute for Infocomm Research, Singapore; Rafael Banchs, Intapp Inc., United States</i>	
<b>SLU.7: MARKOV RECURRENT NEURAL NETWORK LANGUAGE MODEL</b>	<b>807</b>
<i>Jen-Tzung Chien, Che-Yu Kuo, National Chiao Tung University, Taiwan</i>	
<b>SLU.8: TOPIC-AWARE POINTER-GENERATOR NETWORKS FOR SUMMARIZING SPOKEN CONVERSATIONS</b>	<b>814</b>
<i>Zhengyuan Liu, Institute for Infocomm Research, A*STAR, Singapore; Angela Ng, Sheldon Lee, Changi General Hospital, Singapore; Ai Ti Aw, Nancy F. Chen, Institute for Infocomm Research, A*STAR, Singapore</i>	
<b>SLU.9: SLU FOR VOICE COMMAND IN SMART HOME: COMPARISON OF PIPELINE AND END-TO-END APPROACHES</b>	<b>822</b>
<i>Thierry Desot, François Portet, Michel Vacher, Laboratoire d'informatique de Grenoble, France</i>	
<b>SLU.10: SCALABLE NEURAL DIALOGUE STATE TRACKING</b>	<b>830</b>
<i>Vevake Balaraman, Bernardo Magnini, Fondazione Bruno Kessler, Italy</i>	
<b>SLU.11: HIERARCHICAL TRANSFORMERS FOR LONG DOCUMENT CLASSIFICATION</b>	<b>838</b>
<i>Raghavendra Pappagari, Johns Hopkins University, United States; Piotr Zelasko, Avaya Conversational Intelligence, United States; Jesus Villalba, Johns Hopkins University, United States; Yishay Carmiel, Avaya Conversational Intelligence, United States; Najim Dehak, Johns Hopkins University, United States</i>	
<b>SLU.12: ADAPTING PRETRAINED TRANSFORMER TO LATTICES FOR SPOKEN LANGUAGE UNDERSTANDING</b>	<b>845</b>
<i>Chao-Wei Huang, Yun-Nung Chen, National Taiwan University, Taiwan</i>	
<b>SLU.13: SPATIO-TEMPORAL CONTEXT MODELLING FOR SPEECH EMOTION CLASSIFICATION</b>	<b>853</b>
<i>Md Asif Jalal, Roger K. Moore, Thomas Hain, University of Sheffield, United Kingdom</i>	

**SLU.14: PARAPHRASE GENERATION BASED ON VAE AND POINTER-GENERATOR NETWORKS .....860**

*Lohith Ravuru, Hyungtak Choi, Siddarth K.M., Hojung Lee, Inchul Hwang, Samsung Electronics Co. Ltd., Korea (South)*

**SS-1: ASVSPOOF 2019: FUTURE HORIZONS IN SPOOFED AND FAKE AUDIO DETECTION, ANALYSING OPERATIONAL SETTINGS**

**SS-1.1: DEVELOPMENT OF VOICE SPOOFING DETECTION SYSTEMS FOR 2019 EDITION OF AUTOMATIC SPEAKER VERIFICATION AND COUNTERMEASURES CHALLENGE .....1003**

*Joao Monteiro, Institut National de la Recherche Scientifique, Canada; Jahangir Alam, Centre de Recherche Informatique de Montreal, Canada*

**SS-1.2: SPOOF DETECTION USING TIME-DELAY SHALLOW NEURAL NETWORK AND FEATURE SWITCHING .....1011**

*Mari Ganesh Kumar, Indian Institute of Technology, Madras, India; Suvidha Rupesh Kumar, SSN College of Engineering, India; Saranya M S, Indian Institute of Technology, Madras, India; Bharathi B, Shiva Subramaniya Nadar College of Engineering, India; Hema A Murthy, Indian Institute of Technology, Madras, India*

**SS-1.3: LONG RANGE ACOUSTIC AND DEEP FEATURES PERSPECTIVE ON ASVSPOOF 2019 .....1018**

*Rohan Kumar Das, Jichen Yang, Haizhou Li, National University of Singapore, Singapore*

**SS-2: MGB-5: THE FIFTH EDITION OF THE MULTI-GENRE BROADCAST CHALLENGE**

**SS-2.1: THE MGB-5 CHALLENGE: RECOGNITION AND DIALECT IDENTIFICATION OF DIALECTAL ARABIC SPEECH .....1026**

*Ahmed Ali, Hamad Bin Khalifa University, Qatar; Suwon Shon, Massachusetts Institute of Technology, United States; Younes Samih, Hamdy Mubarak, Ahmed Abdelali, Hamad Bin Khalifa University, Qatar; James Glass, Massachusetts Institute of Technology, United States; Steve Renals, University of Edinburgh, United Kingdom; Khalid Choukri, European Language Resources Association, France*

**SS-2.2: RDI-CU SYSTEM FOR THE 2019 ARABIC MULTI-GENRE BROADCAST CHALLENGE .....1034**

*Hany Ahmed, Hazem Mamdouh, Salah Ashraf, Ali Ramadan, Mohsen Rashwan, Research and Development International Company, Egypt*

**TTS: TEXT-TO-SPEECH SYSTEM**

**TTS.1: ON THE STUDY OF GENERATIVE ADVERSARIAL NETWORKS FOR CROSS-LINGUAL VOICE CONVERSION .....144**

*Berrak Sisman, Mingyang Zhang, National University of Singapore, Singapore; Minghui Dong, Institute for Infocomm Research, A\*STAR, Singapore; Haizhou Li, National University of Singapore, Singapore*

<b>TTS.2: WAVENET FACTORIZATION WITH SINGULAR VALUE DECOMPOSITION .....</b>	<b>152</b>
<b>FOR VOICE CONVERSION</b>	
<i>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</i>	
<b>TTS.3: A MODULARIZED NEURAL NETWORK WITH LANGUAGE-SPECIFIC OUTPUT .....</b>	<b>160</b>
<b>LAYERS FOR CROSS-LINGUAL VOICE CONVERSION</b>	
<i>Yi Zhou, Xiaohai Tian, Emre Yilmaz, Rohan Kumar Das, Haizhou Li, National University of Singapore, Singapore</i>	
<b>TTS.4: KNOWLEDGE DISTILLATION FROM BERT IN PRE-TRAINING AND .....</b>	<b>168</b>
<b>FINE-TUNING FOR POLYPHONE DISAMBIGUATION</b>	
<i>Hao Sun, Peking University, China; Xu Tan, Microsoft Research, China; Jun-Wei Gan, Sheng Zhao, Dongxu Han, Microsoft STC Asia, China; Hongzhi Liu, Peking University, China; Tao Qin, Tie-Yan Liu, Microsoft Research, China</i>	
<b>TTS.5: INVESTIGATION OF SHALLOW WAVENET VOCODER WITH LAPLACIAN .....</b>	<b>176</b>
<b>DISTRIBUTION OUTPUT</b>	
<i>Patrick Lumban Tobing, Tomoki Hayashi, Tomoki Toda, Nagoya University, Japan</i>	
<b>TTS.6: LEARNING HIERARCHICAL REPRESENTATIONS FOR EXPRESSIVE .....</b>	<b>184</b>
<b>SPEAKING STYLE IN END-TO-END SPEECH SYNTHESIS</b>	
<i>Xiaochun An, Northwestern Polytechnical University, China; Yuxuan Wang, ByteDance AI Lab, China; Shan Yang, Northwestern Polytechnical University, China; Zejun Ma, ByteDance AI Lab, China; Lei Xie, Northwestern Polytechnical University, China</i>	
<b>TTS.7: CONTROLLING EMOTION STRENGTH WITH RELATIVE ATTRIBUTE FOR .....</b>	<b>192</b>
<b>END-TO-END SPEECH SYNTHESIS</b>	
<i>Xiaolian Zhu, Shan Yang, Geng Yang, Lei Xie, Northwestern Polytechnical University, China</i>	
<b>TTS.8: BOOTSTRAPPING NON-PARALLEL VOICE CONVERSION FROM .....</b>	<b>200</b>
<b>SPEAKER-ADAPTIVE TEXT-TO-SPEECH</b>	
<i>Hieu-Thi Luong, Junichi Yamagishi, National Institute of Informatics, Japan</i>	
<b>TTS.9: IMPROVING MANDARIN END-TO-END SPEECH SYNTHESIS BY .....</b>	<b>208</b>
<b>SELF-ATTENTION AND LEARNABLE GAUSSIAN BIAS</b>	
<i>Fengyu Yang, Shan Yang, Northwestern Polytechnical University, China; Pengcheng Zhu, Pengju Yan, Tongdun, China; Lei Xie, Northwestern Polytechnical University, China</i>	
<b>TTS.10: TACOTRON-BASED ACOUSTIC MODEL USING PHONEME ALIGNMENT .....</b>	<b>214</b>
<b>FOR PRACTICAL NEURAL TEXT-TO-SPEECH SYSTEMS</b>	
<i>Takuma Okamoto, National Institute of Information and Communications Technology, Japan; Tomoki Toda, Nagoya University, Japan; Yoshinori Shiga, Hisashi Kawai, National Institute of Information and Communications Technology, Japan</i>	