

2022 IEEE Spoken Language Technology Workshop (SLT 2022)

**Doha, Qatar
9-12 January 2023**

Pages 1-570



**IEEE Catalog Number: CFP23SLT-POD
ISBN: 979-8-3503-9691-1**

**Copyright © 2023 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:	CFP23SLT-POD
ISBN (Print-On-Demand):	979-8-3503-9691-1
ISBN (Online):	979-8-3503-9690-4

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

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

Automatic speech recognition

1-1-16-ASR: CCC-WAV2VEC 2.0: CLUSTERING AIDED CROSS CONTRASTIVE SELF-SUPERVISED LEARNING OF SPEECH REPRESENTATIONS	1
<i>Vasista Sai Lodagala, Indian Institute of Technology, Madras, India; Sreyan Ghosh, University of Maryland, College Park, United States; Srinivasan Umesh, IIT Chennai, India</i>	
1-1-2-ASR: ASBERT: ASR-SPECIFIC SELF-SUPERVISED LEARNING WITH SELF-TRAINING	9
<i>Hyung Yong Kim, Byeong-Yeol Kim, Seung Woo Yu, Youshin Lim, Yunkyu Lim, Hanbin Lee, 42dot, Korea</i>	
1-1-3-ASR: SUB-8-BIT QUANTIZATION FOR ON-DEVICE SPEECH RECOGNITION: A REGULARIZATION-FREE APPROACH	15
<i>Kai Zhen, Martin Radfar, Hieu D Nguyen, Grant Strimel, Athanasios Mouchtaris, Nathan Susanj, Amazon, United States</i>	
1-1-4-ASR: G-AUGMENT: SEARCHING FOR THE META-STRUCTURE OF DATA AUGMENTATION POLICIES FOR ASR	23
<i>Yuan Wang, Bhuvana Ramabhadran, Pedro Moreno, Google, United States; Ekin D Cubuk, Quoc Le, Daniel S Park, Google Brain, United States; Andrew Rosenberg, Google LLC, United States; Shuyang Cheng, Waymo LLC, United States; Ron J Weiss, Google, Inc., United States</i>	
1-1-7-ASR: CONTEXT-AWARE NEURAL CONFIDENCE ESTIMATION FOR RARE WORD SPEECH RECOGNITION	31
<i>David Qiu, Tsendsuren Munkhdalai, Yanzhang He, Google, United States; Khe C Sim, Google Inc., United States</i>	
1-1-8-ASR: FLICKERING REDUCTION WITH PARTIAL HYPOTHESIS RERANKING FOR STREAMING ASR	38
<i>Antoine Bruguier, David Qiu, Trevor Strohman, Yanzhang He, Google, United States</i>	
1-1-9-ASR: INTERDECODER: USING ATTENTION DECODERS AS INTERMEDIATE REGULARIZATION FOR CTC-BASED SPEECH RECOGNITION	46
<i>Tatsuya Komatsu, Yusuke Fujita, LINE Corporation, Japan</i>	
1-2-1-ASR: JOIST: A JOINT SPEECH AND TEXT STREAMING MODEL FOR ASR	52
<i>Tara Sainath, Rohit Prabhavalkar, Yu Zhang, Zhouyuan Huo, Zhehuai Chen, Bo Li, Weiran Wang, Trevor Strohman, Google, United States; Ankur Bapna, Google Research, United States</i>	
1-2-3-ASR: A CONTEXT-AWARE KNOWLEDGE TRANSFERRING STRATEGY FOR CTC-BASED ASR	60
<i>Ke-Han Lu, Kuan-Yu Chen, National Taiwan University of Science and Technology, Taiwan</i>	
1-2-4-ASR: MAESTRO-U: LEVERAGING JOINT SPEECH-TEXT REPRESENTATION LEARNING FOR ZERO SUPERVISED SPEECH ASR	68
<i>Zhehuai Chen, Yu Zhang, Bhuvana Ramabhadran, Pedro Moreno, Nanxin Chen, Google, United States; Ankur Bapna, Google Research, United States; Andrew Rosenberg, Google LLC, United States</i>	

1-2-6-ASR: ALTERNATE INTERMEDIATE CONDITIONING WITH SYLLABLE-LEVEL AND CHARACTER-LEVEL TARGETS FOR JAPANESE ASR	76
<i>Yusuke Fujita, Tatsuya Komatsu, LINE Corporation, Japan; Yusuke Kida, LINE Corp, Japan</i>	
1-2-7-ASR: E-BRANCHFORMER: BRANCHFORMER WITH ENHANCED MERGING FOR SPEECH RECOGNITION	84
<i>Kwangyoun Kim, Felix Wu, Jing Pan, Prashant Sridhar, Kyu Jeong Han, ASAPP, United States; Yifan Peng, Shinji Watanabe, Carnegie Mellon University, United States</i>	
1-2-8-ASR: CONFORMER-BASED ON-DEVICE STREAMING SPEECH RECOGNITION WITH KD COMPRESSION AND TWO-PASS ARCHITECTURE	92
<i>Jinhwan Park, Junmo Park, Dhairya Sandhyana, Samsung Research, Korea; Sichen Jin, Samsung, Korea; Sungsoo Kim, Changheon Lee, Myoungji Han, Jungin Lee, Seokyeong Jung, Chanwoo Kim, Samsung Electronics, United States; Chang Woo Han, Samsung Reserch, Korea</i>	
1-2-9-ASR: ACCELERATOR-AWARE TRAINING FOR TRANSDUCER-BASED SPEECH RECOGNITION	100
<i>Rupak Vignesh Swaminathan, Suhaila Mumtaj Shakiah, Hieu D Nguyen, Raviteja Chinta, Tariq Afzal, Nathan Susanj, Athanasios Mouchtaris, Grant Strimel, Amazon, United States; Ariya Rastrow, Amazon Alexa, United States</i>	
2-1-1-ASR: UNTIED POSITIONAL ENCODINGS FOR EFFICIENT TRANSFORMER-BASED SPEECH RECOGNITION	108
<i>Lahiru T Samarakoon, Ivan Fung, Fano Labs, Hong Kong, Hong Kong SAR</i>	
2-1-2-ASR: MATCH TO WIN: ANALYSING SEQUENCES LENGTHS FOR EFFICIENT SELF-SUPERVISED LEARNING IN SPEECH AND AUDIO	115
<i>Yan Gao, Pedro Gusmao, University of Cambridge, United Kingdom; Javier Fernandez-Marques, Samsung AI, Cambridge, United Kingdom; Titouan Parcollet, Avignon University, United Kingdom; Nicholas Lane, University of Cambridge and Samsung AI, United Kingdom</i>	
2-1-3-ASR: PRONUNCIATION-AWARE UNIQUE CHARACTER ENCODING FOR RNN TRANSDUCER-BASED MANDARIN SPEECH RECOGNITION	123
<i>Peng Shen, Xugang Lu, Hisashi Kawai, NICT, Japan</i>	
2-1-4-ASR: DAMAGE CONTROL DURING DOMAIN ADAPTATION FOR TRANSDUCER BASED AUTOMATIC SPEECH RECOGNITION	130
<i>Somshubra Majumdar, Shantanu Acharya, Vitaly Lavrukhin, Boris Ginsburg, NVIDIA, United States</i>	
2-1-5-ASR: PADA: PRUNING ASSISTED DOMAIN ADAPTATION FOR SELF-SUPERVISED SPEECH REPRESENTATIONS	136
<i>Vasista Sai Lodagala, Indian Institute of Technology, Madras, India; Sreyan Ghosh, University of Maryland, College Park, United States; Srinivasan Umesh, IIT Chennai, India</i>	
2-1-6-ASR: MFCCA: MULTI-FRAME CROSS-CHANNEL ATTENTION FOR MULTI-SPEAKER ASR IN MULTI-PARTY MEETING SCENARIO	144
<i>Fan Yu, Pengcheng Guo, Yuhao Liang, Lei Xie, Northwestern Polytechnical University, China; Shiliang Zhang, Alibaba Group, China; Zhihao Du, Speech Lab, Alibaba Group, China; Yuxiao Lin, Zhejiang University, China</i>	
2-1-7-ASR: FAST ENTROPY-BASED METHODS OF WORD-LEVEL CONFIDENCE ESTIMATION FOR END-TO-END AUTOMATIC SPEECH RECOGNITION	152
<i>Aleksandr Laptev, NVIDIA, ITMO University, Armenia; Boris Ginsburg, NVIDIA, United States</i>	

2-1-9-ASR: RESIDUAL ADAPTERS FOR TARGETED UPDATES IN RNN-TRANSDUCER BASED SPEECH RECOGNITION SYSTEM	.160
<i>Sungjun Han, University of Stuttgart, Germany; Deepak Baby, Valentin Mendeleev, Amazon Alexa, Germany</i>	
2-2-1-ASR: IMPROVED NOISY ITERATIVE PSEUDO-LABELING FOR SEMI-SUPERVISED SPEECH RECOGNITION167
<i>Tian Li, Qingliang Meng, Yujian Sun, Shumei AI Research Institute, China</i>	
2-2-2-ASR: GUIDED CONTRASTIVE SELF-SUPERVISED PRE-TRAINING FOR AUTOMATIC SPEECH RECOGNITION174
<i>Aparna Khare, Amazon, United States; Minhua Wu, Jasha Droppo, Roland Maas, Amazon Inc., United States; Saurabhchand Bhati, Johns Hopkins University, United States</i>	
2-2-3-ASR: LEARNING TO JOINTLY TRANSCRIBE AND SUBTITLE FOR END-TO-END SPONTANEOUS SPEECH RECOGNITION182
<i>Jakob Poncelet, Hugo Van Hamme, KU Leuven, Belgium</i>	
2-2-4-ASR: NAM + : TOWARDS SCALABLE END-TO-END CONTEXTUAL BIASING FOR ADAPTIVE ASR190
<i>Zelin Wu, Jiayang Li, Pat Rondon, Google LLC, United States; Tsendsuren Munkhdalai, Golan Pundak, Tara Sainath, Google, United States; Khe C Sim, Google Inc., United States</i>	
2-2-6-ASR: MODULAR HYBRID AUTOREGRESSIVE TRANSDUCER197
<i>Zhong Meng, Jesse Emond, Google LLC, United States; Tongzhou Chen, Rohit Prabhavalkar, Yu Zhang, Yuan Wang, Kartik Audhkhasi, Trevor Strohman, Bhuvana Ramabhadran, W. Ronny Huang, Ehsan Variiani, Yinghui Huang, Pedro Moreno, Google, United States</i>	
2-2-7-ASR: HOW DOES PRE-TRAINED WAV2VEC 2.0 PERFORM ON DOMAIN-SHIFTED ASR?205
AN EXTENSIVE BENCHMARK ON AIR TRAFFIC CONTROL COMMUNICATIONS	
<i>Juan Pablo Zuluaga Gomez, Amrutha Prasad, Iuliia Nigmatulina, Seyyed Saeed Sarfjoo, Idiap Research Institute, Switzerland; Petr Motlicek, Idiap, Switzerland; Matthias Kleinert, Hartmut Helmke, Oliver Ohneiser, DLR, Germany; Qingran Zhan, Beijing Institute of Technology, China</i>	
2-2-9-ASR: INTERNAL LANGUAGE MODEL PERSONALIZATION OF E2E AUTOMATIC SPEECH RECOGNITION USING RANDOM ENCODER FEATURES	...213
<i>Adam Stooke, Mason Chua, Tsendsuren Munkhdalai, Trevor Strohman, Google, United States; Khe C Sim, Google Inc., United States</i>	
3-1-1-ASR: TOWARDS END-TO-END UNSUPERVISED SPEECH RECOGNITION221
<i>Alexander H Liu, MIT, United States; Wei-Ning Hsu, Massachusetts Institute of Technology, United States; Michael Auli, Facebook, United States; Alexei Baevski, Facebook AI Research, United States</i>	
3-1-3-ASR: MONOTONIC SEGMENTAL ATTENTION FOR AUTOMATIC SPEECH RECOGNITION229
<i>Albert Zeyer, Robin Schmitt, Wei Zhou, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</i>	

3-1-4-ASR: STREAMING, FAST AND ACCURATE ON-DEVICE INVERSE TEXT	237
NORMALIZATION FOR AUTOMATIC SPEECH RECOGNITION	
<i>Yashesh Gaur, Nick Kibre, Kangyuan Shu, Yuhui Wang, Issac Alphonso, Jinyu Li, Yifan Gong, Microsoft, United States; Jian Xue, Microsoft Corporation, United States</i>	
3-1-5-ASR: DUAL LEARNING FOR LARGE VOCABULARY ON-DEVICE ASR.....	245
<i>Charles C Peyser, Google Inc., United States; W. Ronny Huang, Tara Sainath, Rohit Prabhavalkar, Google, United States; Michael Picheny, NYU, United States; Kyunghyun Cho, New York University, United States</i>	
3-1-6-ASR: STREAMING BILINGUAL END TO END ASR MODEL USING ATTENTION OVER	252
MULTIPLE SOFTMAX	
<i>Aditya R Patil, Vikas V Joshi, Purvi Agrawal, Rupesh Mehta, Microsoft, Australia</i>	
3-1-7-ASR: END-TO-END INTEGRATION OF SPEECH RECOGNITION,	260
DEREVERBERATION, BEAMFORMING, AND SELF-SUPERVISED LEARNING REPRESENTATION	
<i>Yoshiki Masuyama, Nobutaka Ono, Tokyo Metropolitan University, Japan; Xuankai Chang, Shinji Watanabe, Carnegie Mellon University, United States; Samuele Cornell, Università Politecnica delle Marche, Italy</i>	
3-1-8-ASR: FULLY UNSUPERVISED TRAINING OF FEW-SHOT KEYWORD SPOTTING.....	266
<i>Minchan Kim, Dongjune Lee, Sung Hwan Mun, Min Hyun Han, Nam Soo Kim, Seoul National University, Korea</i>	
3-1-9-ASR: LEARNING A DUAL-MODE SPEECH RECOGNITION MODEL VIA SELF-PRUNING ..	273
<i>Chunxi Liu, Yuan Shangguan, Ozlem Kalinli, Meta AI, United States; Haichuan Yang, Meta, United States; Yangyang Shi, Raghuraman Krishnamoorthi, Facebook, United States</i>	
4-1-1-ASR: INTER-KD: INTERMEDIATE KNOWLEDGE DISTILLATION FOR CTC-BASED	280
AUTOMATIC SPEECH RECOGNITION	
<i>Ji Won Yoon, Beom Jun Woo, Sunghwan Ahn, Hyeonseung Lee, Nam Soo Kim, Seoul National University, Korea</i>	
4-1-2-ASR: HMM VS. CTC FOR AUTOMATIC SPEECH RECOGNITION: COMPARISON BASED ..	287
ON FULL-SUM TRAINING FROM SCRATCH	
<i>Tina Raissi, Wei Zhou, Simon Berger, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</i>	
4-1-3-ASR: DOMAIN ADAPTATION OF LOW-RESOURCE TARGET-DOMAIN MODELS USING ..	295
WELL-TRAINED ASR CONFORMER MODELS	
<i>Vrunda N Sukhadia, Indian Institute Of Technology Madras, India; Srinivasan Umesh, IIT Chennai, India</i>	
4-1-4-ASR: PERSONALIZATION OF CTC SPEECH RECOGNITION MODELS	302
<i>Saket Dingliwal, Monica Sunkara, Sravan Babu Bodapati, Srikanth Ronanki, Jeff Farris, Katrin Kirchhoff, Amazon, United States</i>	
4-1-6-ASR: UNIFIED END-TO-END SPEECH RECOGNITION AND ENDPOINTING FOR FAST ...	310
AND EFFICIENT SPEECH SYSTEMS	
<i>Shaan Bijwadia, Shuo-yiin Chang, Tara Sainath, Bo Li, Chao Zhang, Yanzhang He, Google, United States</i>	

4-1-7-ASR: LEARNING MASK SCALARS FOR IMPROVED ROBUST AUTOMATIC SPEECH317
RECOGNITION

Arun Narayanan, Google Inc., United States; James Walker, Nathan Howard, Google Llc., United States; Sankaran Panchapagesan, Google, LLC, United States; Yuma Koizumi, Google, Japan

4-1-8-ASR: AN INVESTIGATION OF MONOTONIC TRANSDUCERS FOR LARGE-SCALE324
AUTOMATIC SPEECH RECOGNITION

Niko Moritz, Frank Seide, Duc Le, Meta, United Kingdom; Jay Mahadeokar, Meta AI, United States; Christian Fuegen, Facebook, United Kingdom

4-1-9-ASR: MACRO-BLOCK DROPOUT FOR IMPROVED REGULARIZATION IN TRAINING331
END-TO-END SPEECH RECOGNITION MODELS

Chanwoo Kim, Samsung Electronics, Korea; Sathish Indurti, Jinhwan Park, Samsung Research, Korea; Wonyong Sung, Seoul national university, Korea

Spoken language processing

1-1-10-SLP: AUTOMATIC RATING OF SPONTANEOUS SPEECH FOR LOW-RESOURCE339
LANGUAGES

Yaroslav Getman, Ragheb Al-Ghezi, Ekaterina Voskoboinik, Mikko Kurimo, Aalto University, Finland; Mittul Singh, Silo AI, Finland

1-1-11-SLP: MIXTURE OF DOMAIN EXPERTS FOR LANGUAGE UNDERSTANDING: AN346
ANALYSIS OF MODULARITY, TASK PERFORMANCE, AND MEMORY TRADEOFFS

Benjamin Kleiner, AWS AI Labs, United States; Jack FitzGerald, Amazon Alexa Artificial Intelligence, United States; Haidar Khan, Gokhan Tur, Amazon Alexa AI, United States

1-2-10-SLP: A DATA-DRIVEN INVESTIGATION OF NOISE-ADAPTIVE UTTERANCE353
GENERATION WITH LINGUISTIC MODIFICATION

Anupama Chingacham, Dietrich Klakow, Saarland University, Germany; Vera Demberg, Dept. of Mathematics and Computer Science, Saarland University, Germany

1-2-11-SLP: ON THE USE OF SEMANTICALLY-ALIGNED SPEECH REPRESENTATIONS FOR ...361
SPOKEN LANGUAGE UNDERSTANDING

Gaëlle Laperrière, Mickael Rouvier, Yannick Estève, LIA - Avignon University, France; Valentin Pelloin, LIUM, Le Mans Université, France; Themis Stafylakis, Omilia - Conversational Intelligence, Greece

2-1-10-SLP: RESPONSE TIMING ESTIMATION FOR SPOKEN DIALOG SYSTEMS BASED ON ...369
SYNTACTIC COMPLETENESS PREDICTION

Jin Sakuma, Tetsunori Kobayashi, Waseda University, Japan; Shinya Fujie, Chiba Institute of Technology, Japan

2-1-11-SLP: WEAK-SUPERVISED DYSARTHRIA-INVARIANT FEATURES FOR SPOKEN375
LANGUAGE UNDERSTANDING USING AN FHVAE AND ADVERSARIAL TRAINING

Jinzi Qi, KU Leuven, Belgium; Hugo Van Hamme, KU Leuven, Belgium

2-2-10-SLP: BUILDING MARKOVIAN GENERATIVE ARCHITECTURES OVER PRETRAINED ...382
LM BACKBONES FOR EFFICIENT TASK-ORIENTED DIALOG SYSTEMS

Hong Liu, Yucheng Cai, Zhijian Ou, Tsinghua University, China; Yi Huang, Junlan Feng, China Mobile Research, China

2-2-11-SLP: NON-AUTOREGRESSIVE END-TO-END APPROACHES FOR JOINT AUTOMATIC ..390
SPEECH RECOGNITION AND SPOKEN LANGUAGE UNDERSTANDING
Mohan Li, Toshiba Europe Ltd, United Kingdom; Rama S Doddipatla, Toshiba Europe LTD, United Kingdom

3-1-10-SLP: IMPROVING NOISE ROBUSTNESS FOR SPOKEN CONTENT RETRIEVAL398
USING SEMI-SUPERVISED ASR AND N-BEST TRANSCRIPTS FOR BERT-BASED
RANKING MODELS
Yasufumi Moriya, Gareth Jones, Dublin City University, Ireland

3-1-11-SLP: A STUDY ON THE INTEGRATION OF PRE-TRAINED SSL, ASR, LM AND SLU406
MODELS FOR SPOKEN LANGUAGE UNDERSTANDING
Yifan Peng, Siddhant Arora, Yushi Ueda, Sujay Kumar, Karthik Ganesan, Siddharth Dalmia, Xuankai Chang, Shinji Watanabe, Carnegie Mellon University, United States; Yosuke Higuchi, Waseda University, Japan

4-1-10-SLP: ON THE EFFICIENCY OF INTEGRATING SELF-SUPERVISED LEARNING AND414
META-LEARNING FOR USER-DEFINED FEW-SHOT KEYWORD SPOTTING
Yuan-Kuei Wu, Wei-Tsung Kao, Hung-yi Lee, National Taiwan University, Taiwan; Chia-Ping Chen, Zhi-Sheng Chen, Yu-Pao Tsai, intelliGo Technology inc., Taiwan

Speech enhancement and separation

1-1-12-SES: MULTI-STAGE PROGRESSIVE AUDIO BANDWIDTH EXTENSION422
Liang Wen, Samsung electronics, China; Lizhong Wang, Samsung, China; Ying Zhang, Kwang Pyo Choi, Samsung Electronics, China

1-1-13-SES: JOINT OPTIMIZATION OF DIFFUSION PROBABILISTIC-BASED428
MULTICHANNEL SPEECH ENHANCEMENT WITH FAR-FIELD SPEAKER VERIFICATION
Sandipana Dowerah, Inria, France; Romain Serizel, Université de Lorraine, France; Denis Jouvet, LORIA, France; Mohammad Mohammadamini, Driss Matrouf, Laboratoire Informatique d'Avignon, University of Avignon, France

1-2-12-SES: SPATIAL-DCCRN: DCCRN EQUIPPED WITH FRAME-LEVEL ANGLE FEATURE436
AND HYBRID FILTERING FOR MULTI-CHANNEL SPEECH ENHANCEMENT
Shubo Lv, Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, China; Yihui Fu, Yukai Ju, Lei Xie, Northwestern Polytechnical University, China; Weixin Zhu, Wei Rao, Yannan Wang, Tencent, China

1-2-13-SES: IMPROVED NORMALIZING FLOW-BASED SPEECH ENHANCEMENT USING AN ..444
ALL-POLE GAMMATONE FILTERBANK FOR CONDITIONAL INPUT REPRESENTATION
Martin Strauss, Matteo Torcoli, Bernd Edler, International Audio Laboratories Erlangen, Germany

2-1-12-SES: EXPLORING WAVLM ON SPEECH ENHANCEMENT.....451
Hyungchan Song, Jong Won Shin, Gwangju Institute of Science and Technology, Korea; Sanyuan Chen, Harbin Institute of Technology, China; Zhuo Chen, Takuya Yoshioka, Min Tang, Microsoft, United States; Yu Wu, Shujie Liu, Microsoft Research Asia, China

2-1-13-SES: ADAPTIVE-FSN: INTEGRATING FULL-BAND EXTRACTION AND ADAPTIVE458
SUB-BAND ENCODING FOR MONAURAL SPEECH ENHANCEMENT
Yu-Sheng Tsao, Berlin Chen, National Taiwan Normal University, Taiwan; Kuan-Hsun Ho, NTNU, Taiwan; Jieh-weih Hung, National Chi Nan University, Taiwan

2-1-26-SES: AVSE CHALLENGE: AUDIO-VISUAL SPEECH ENHANCEMENT CHALLENGE	465
<i>Andrea L Aldana, Edinburgh University, United Kingdom; Cassia Valentini, Ondrej Klejch, Peter Bell, University of Edinburgh, United Kingdom; Mandar Gogate, Kia K Dashtipour, Amir Hussain, Edinburgh Napier University, United Kingdom</i>	
2-2-12-SES: TEA-PSE 2.0: SUB-BAND NETWORK FOR REAL-TIME PERSONALIZED SPEECH ..	472
ENHANCEMENT	
<i>Yukai Ju, Shimin Zhang, Lei Xie, Northwestern Polytechnical University, China; Wei Rao, Yannan Wang, Tao Yu, Shi-dong Shang, Tencent, China</i>	
2-2-13-SES: EEND-SS: JOINT END-TO-END NEURAL SPEAKER DIARIZATION AND SPEECH ..	480
SEPARATION FOR FLEXIBLE NUMBER OF SPEAKERS	
<i>Soumi Maiti, CMU, United States; Yushi Ueda, Shinji Watanabe, Carnegie Mellon University, United States; Chunlei Zhang, Tencent AI Lab, United States; Meng Yu, Shixiong Zhang, Tencent, United States; Yong Xu, Tecom, United States</i>	
3-1-12-SES: LIMUSE: LIGHTWEIGHT MULTI-MODAL SPEAKER EXTRACTION	488
<i>Qinghua Liu, Tianjin University, China; Yating Huang, Institute of Automation, Chinese Academy of Sciences (CASIA), China; Yunzhe Hao, Institute of Automation, Chinese Academy of Science, China; Jiaming Xu, Institute of Automation Chinese Academy of Sciences, China; Bo Xu, Institute of Automation, Chinese Academy of Sciences, China</i>	
4-1-11-SES: END-TO-END MULTI-SPEAKER ASR WITH INDEPENDENT VECTOR ANALYSIS....	496
<i>Robin Scheibler, LINE Corporation, Japan; Wangyou Zhang, Yanmin Qian, Shanghai Jiao Tong University, China; Xuankai Chang, Shinji Watanabe, Carnegie Mellon University, United States</i>	
4-1-12-SES: A HYBRID ACOUSTIC ECHO REDUCTION APPROACH USING KALMAN	502
FILTERING AND INFORMED SOURCE EXTRACTION WITH IMPROVED TRAINING	
<i>Wolfgang Mack, Emanuel Habets, AudioLabs Erlangen, Germany</i>	

Speech analysis

1-1-14-ANA: LEARNING INVARIANT REPRESENTATION AND RISK MINIMIZED FOR	509
UNSUPERVISED ACCENT DOMAIN ADAPTATION	
<i>Chendong Zhao, The Shenzhen International Graduate School, Tsinghua University, China, China; Jianzong Wang, Xiaoyang Qu, Ping An Technology (Shenzhen) Co., Ltd, China; Haoqian Wang, Tsinghua Shenzhen International Graduate School, Tsinghua University, China; Jing Xiao, Ping An Insurance (Group) Company of China, China</i>	
1-2-14-ANA: VSAMETER: EVALUATION OF A NEW OPEN-SOURCE TOOL TO MEASURE	517
VOWEL SPACE AREA AND RELATED METRICS	
<i>Tianyu Cao, Laureano Moro-Velazquez, Jesús Villalba, Najim Dehak, Johns Hopkins University, United States; Piotr Żelasko, Meaning, United States</i>	
2-1-14-ANA: INVESTIGATING THE IMPORTANT TEMPORAL MODULATIONS FOR	525
DEEP-LEARNING-BASED SPEECH ACTIVITY DETECTION	
<i>Tyler Vuong, Nikhil Madaan, Rohan Panda, Richard M Stern, Carnegie Mellon University, United States</i>	

3-1-13-ANA: A MULTI-MODAL ARRAY OF INTERPRETABLE FEATURES TO EVALUATE532
LANGUAGE AND SPEECH PATTERNS IN DIFFERENT NEUROLOGICAL DISORDERS
Anna Favaro, Chelsie Motley, Tianyu Cao, Miguel Iglesias, Ankur Butala, Esther S. Oh, Jesús Villalba, Najim Dehak, Laureano Moro-Velazquez, Johns Hopkins University, United States; Robert Stevens, Johns Hopkins Hospital, United States

4-1-13-ANA: EFFICIENT DYNAMIC FILTER FOR ROBUST AND LOW COMPUTATIONAL540
FEATURE EXTRACTION
Donghyeon Kim, Korea university, Korea; Jeong-gi Kwak, Hanseok Ko, Korea University, Korea

Speaker and language recognition

1-2-15-SLR: FREQUENCY AND MULTI-SCALE SELECTIVE KERNEL ATTENTION FOR548
SPEAKER VERIFICATION
Sung Hwan Mun, Min Hyun Han, Nam Soo Kim, Seoul National University, Korea; Jee-weon Jung, Naver Corporation, Korea

2-1-15-SLR: AN ATTENTION-BASED BACKEND ALLOWING EFFICIENT FINE-TUNING OF555
TRANSFORMER MODELS FOR SPEAKER VERIFICATION
Junyi Peng, Oldrich Plchot, Ladislav Mošner, Lukas Burget, Jan Cernocky, Brno University of Technology, Czechia; Themis Stafylakis, Omilia - Conversational Intelligence, Greece

2-2-14-SLR: FLOW-ER: A FLOW-BASED EMBEDDING REGULARIZATION STRATEGY FOR563
ROBUST SPEECH REPRESENTATION LEARNING
Woo Hyun Kang, Computer Research Institute of Montreal, Canada; Jahangir Alam, Computer Research Institute of Montreal (CRIM), Montreal (Quebec) Canada, Canada; Abderrahim Fathan, Computer Research Institute of Montreal (CRIM), Montreal, Quebec, Canada, Canada

2-2-15-SLR: UNSUPERVISED DOMAIN ADAPTATION OF NEURAL PLDA USING SEGMENT ...571
PAIRS FOR SPEAKER VERIFICATION
İsmail Rasim Ülgen, Sestek - Boğaziçi University, Turkey; Mustafa Levent Arslan, Sestek - Boğaziçi Üniversitesi, Turkey

3-1-14-SLR: THE CLEVER HANS EFFECT IN VOICE SPOOFING DETECTION577
Bhusan Chettri, Borac Solutions, India

3-1-15-SLR: INVESTIGATING ACTIVE-LEARNING-BASED TRAINING DATA SELECTION FOR .585
SPEECH SPOOFING COUNTERMEASURE
Xin Wang, Junichi Yamagishi, National Institute of Informatics, Japan

4-1-14-SLR: HOW TO BOOST ANTI-SPOOFING WITH X-VECTORS593
Xinyue Ma, Liang He, Tsinghua University, China; Shanshan Zhang, Shen Huang, Ji Gao, Tencent Research, China; Ying Hu, Xinjiang University, China

4-1-15-SLR: A COMPREHENSIVE STUDY ON SELF-SUPERVISED DISTILLATION FOR599
SPEAKER REPRESENTATION LEARNING
Zhengyang Chen, Bing Han, Yanmin Qian, Shanghai Jiao Tong University, China; Yao Qian, Michael Zeng, Microsoft, United States

Speaker diarization

1-2-16-DIA: JOINT SPEAKER DIARISATION AND TRACKING IN SWITCHING STATE-SPACE MODEL ..605

Jeremy H. M. Wong, Institute for Infocomm Research, Singapore; Yifan Gong, Microsoft, United States

2-1-16-DIA: DIARISATION USING LOCATION TRACKING WITH AGGLOMERATIVE CLUSTERING613

Jeremy H. M. Wong, Institute for Infocomm Research, Singapore; Igor Abramovski, Xiong Xiao, Yifan Gong, Microsoft, United States

2-2-16-DIA: MUTUAL LEARNING OF SINGLE- AND MULTI-CHANNEL END-TO-END NEURAL DIARIZATION 620

Shota Horiguchi, Yuki Takashima, Hitachi, Ltd., Japan; Shinji Watanabe, Carnegie Mellon University, United States; Paola Garcia, Johns Hopkins University, United States

2-2-5-DIA: CONTINUAL SELF-SUPERVISED DOMAIN ADAPTATION FOR END-TO-END SPEAKER DIARIZATION626

Juan Manuel Coria, Sahar Ghannay, Université Paris-Saclay CNRS, LISN, France; Hervé Bredin, CNRS, France; Sophie Rosset, LISN, France

3-1-16-DIA: BERTRAFFIC: BERT-BASED JOINT SPEAKER ROLE AND SPEAKER CHANGE DETECTION FOR AIR TRAFFIC CONTROL COMMUNICATIONS633

Juan Pablo Zuluaga Gomez, Seyyed Saeed Sarfjoo, Amrutha Prasad, Iuliia Nigmatulina, Idiap Research Institute, Switzerland; Petr Motlicek, Idiap, Switzerland; Karel Ondrej, BUT, Czechia; Oliver Ohneiser, Hartmut Helmke, DLR, Germany

4-1-16-DIA: LOW-LATENCY SPEECH SEPARATION GUIDED DIARIZATION FOR TELEPHONE CONVERSATIONS 641

Giovanni Morrone, Samuele Cornell, Luca Serafini, Stefano Squartini, Università Politecnica delle Marche, Italy; Desh Raj, Johns Hopkins University, United States; Enrico Zovato, PerVoice S.p.A., Italy; Alessio Brutti, FBK, Italy

Text-only language processing

1-1-17-TLP: FINE GRAINED SPOKEN DOCUMENT SUMMARIZATION THROUGH TEXT SEGMENTATION647

Samantha Kotey, Naomi Harte, Trinity College Dublin, Ireland; Rozenn Dahyot, Maynooth University, Ireland

1-2-17-TLP: AN ANALYSIS OF THE EFFECTS OF DECODING ALGORITHMS ON FAIRNESS IN OPEN-ENDED LANGUAGE GENERATION655

Jwala Dhamala, Amazon Alexa AI, United States; Varun Kumar, Amazon Alexa, United States; Rahul Gupta, Amazon, United States; Kai-Wei Chang, UCLA, United States; Aram Galstyan, USC Information Sciences Institute, United States

2-1-17-TLP: N-BEST HYPOTHESES RERANKING FOR TEXT-TO-SQL SYSTEMS663

Lu Zeng, Sree Hari Krishnan Parthasarathi, Amazon, Canada; Dilek Z Hakkani-Tur, Amazon Alexa AI, United States

3-1-17-TLP: EFFICIENT TEXT ANALYSIS WITH PRE-TRAINED NEURAL NETWORK MODELS671

Jia Cui, Shiyin Kang, Liqiang He, Guangzhi Li, Tencent, United States; Heng Lu, Dong Yu, Tencent AI Lab, China; Wenjie Wang, Emory University, United States

3-1-24-TLP: FOUR-IN-ONE: A JOINT APPROACH TO INVERSE TEXT NORMALIZATION, PUNCTUATION, CAPITALIZATION, AND DISFLUENCY FOR AUTOMATIC SPEECH RECOGNITION677

Sharman W Tan, Piyush Behre, Nick Kibre, Issac Alphonso, Shawn Chang, Microsoft, United States

4-1-17-TLP: EMPIRICAL ANALYSIS OF TRAINING STRATEGIES OF TRANSFORMER-BASED JAPANESE CHIT-CHAT SYSTEMS ..685

Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, Toyomi Meguro, NTT, Japan

Multimodal speech processing

1-1-18-MMP: PUSH-PULL: CHARACTERIZING THE ADVERSARIAL ROBUSTNESS FOR AUDIO-VISUAL ACTIVE SPEAKER DETECTION692

Xuanjun Chen, Haibin Wu, Hung-yi Lee, Roger Jang, National Taiwan University, China; Helen Meng, The Chinese University of Hong Kong, Hong Kong SAR

1-1-19-MMP: TOWARDS VISUALLY PROMPTED KEYWORD LOCALISATION FOR ZERO-RESOURCE SPOKEN LANGUAGES700

Leanne Nortje, Herman Kamper, Stellenbosch University, South Africa

1-2-18-MMP: EXPLOITING INFORMATION FROM NATIVE DATA FOR NON-NATIVE AUTOMATIC PRONUNCIATION ASSESSMENT708

Binghuai Lin, MIG, Tencent Science and Technology Ltd., China; Liyuan Wang, Tencent Technology Co., Ltd, China

2-1-18-MMP: SPEECHCLIP: INTEGRATING SPEECH WITH PRE-TRAINED VISION AND LANGUAGE MODEL715

Yi-Jen Shih, Hung-yi Lee, National Taiwan University, Taiwan; Hsuan-Fu Wang, Academia Sinica, Taiwan; Heng-Jui Chang, Massachusetts Institute of Technology, United States; Layne Berry, University of Texas at Austin, United States; David Harwath, The University of Texas at Austin, United States

2-1-8-MMP: TRANSFORMER-BASED LIP-READING WITH REGULARIZED DROPOUT AND RELAXED ATTENTION723

Zhengyang Li, Timo Lohrenz, Matthias Dunkelberg, Tim Fingscheidt, Technische Universität Carolo-Wilhelmina Braunschweig, Germany

2-2-18-MMP: YFACC: A YORÙBÁ SPEECH-IMAGE DATASET FOR CROSS-LINGUAL KEYWORD LOCALISATION THROUGH VISUAL GROUNDING ... 731

LOCALISATION THROUGH VISUAL GROUNDING

Kayode K Olaleye, University of Stellenbosch, South Africa; Dan Oneață, University Politehnica of Bucharest, Romania; Herman Kamper, Stellenbosch University, South Africa

3-1-18-MMP: ON THE USE OF MODALITY-SPECIFIC LARGE-SCALE PRE-TRAINED739
ENCODERS FOR MULTIMODAL SENTIMENT ANALYSIS

Atsushi Ando, Ryo Masumura, Naoki Makishima, Keita Suzuki, Takafumi Moriya, Takanori Ashihara, Hiroshi Sato, NTT Corporation, Japan; Akihiko Takashima, NTT, Japan; Satoshi Suzuki, NTT Computer and Data Science Laboratories / The University of Electro-Communications, Japan

4-1-18-MMP: AN ANALYSIS OF SEMANTICALLY-ALIGNED SPEECH-TEXT EMBEDDINGS.....747

Muhammad Huzaifah, Ivan Kukanov, Institute for Infocomm Research, ASTAR, Singapore

Multilingual processing

1-1-1-MLP: EXPLORATION OF LANGUAGE-SPECIFIC SELF-ATTENTION PARAMETERS FOR ..755
MULTILINGUAL END-TO-END SPEECH RECOGNITION

Brady Houston, AWS AI Labs, United States; Katrin Kirchhoff, Amazon, United States

1-1-5-MLP: HOW DO PHONOLOGICAL PROPERTIES AFFECT BILINGUAL AUTOMATIC763
SPEECH RECOGNITION?

Shelly Jain, Aditya Yadavalli, International Institute of Information Technology, Hyderabad, India; Sai Ganesh Mirishkar, IIIT Hyderabad, India; Anil Vuppala, International Institute of Information Technology Hyderabad, India

1-1-6-MLP: SCALING UP DELIBERATION FOR MULTILINGUAL ASR771

Ke Hu, Tara Sainath, Bo Li, Google, United States

1-2-19-MLP: TEXTUAL DATA AUGMENTATION FOR ARABIC-ENGLISH CODE-SWITCHING ...777
SPEECH RECOGNITION

Amir Hussein, Najim Dehak, Sanjeev Khudanpur, Johns Hopkins University, United States; Shammur Chowdhury, Ahmed Abdelali, QCRI, Qatar; Ahmed Ali, Qatar Computing Research Institute, HBKU, Qatar

1-2-2-MLP: CODE-SWITCHED LANGUAGE MODELLING USING A CODE PREDICTIVE785
LSTM IN UNDER-RESOURCED SOUTH AFRICAN LANGUAGES

Joshua Miles Jansen Van Vüren, Thomas Niesler, Stellenbosch University, South Africa

1-2-5-MLP: IMPROVING LUXEMBOURGISH SPEECH RECOGNITION WITH792
CROSS-LINGUAL SPEECH REPRESENTATIONS

Le Minh Nguyen, Shekhar Nayak, Matt Coler, University of Groningen, Luxembourg

2-1-23-MLP: FLEURS: FEW-SHOT LEARNING EVALUATION OF UNIVERSAL798
REPRESENTATIONS OF SPEECH

Alexis Conneau, FAIR, France; Min Ma, Ankur Bapna, Google Research, United States; Simran Khanuja, Yu Zhang, Jason Riesa, Clara Rivera, Google, United States; Vera Axelrod, Google, Inc, United States; Siddharth Dalmia, Carnegie Mellon University, United States

2-2-19-MLP: MULTILINGUAL SPEECH EMOTION RECOGNITION WITH MULTI-GATING806
MECHANISM AND NEURAL ARCHITECTURE SEARCH

Zihan Wang, Qi Meng, Haifeng Lan, Xinrui Zhang, Kehao Guo, Columbia University, United States; Akshat Gupta, JPMorgan, United States

2-2-22-MLP: DISENTANGLED SPEECH REPRESENTATION LEARNING FOR ONE-SHOT	814
CROSS-LINGUAL VOICE CONVERSION USING B-VAE	
<i>Hui Lu, Disong Wang, Xixin Wu, Xunying Liu, Helen Meng, The Chinese University of Hong Kong, Hong Kong SAR; Zhiyong Wu, Tsinghua University, China</i>	
2-2-8-MLP: IMPROVING SEMI-SUPERVISED E2E ASR USING CYCLEGAN AND	822
INTER-DOMAIN LOSSES	
<i>Chia-Yu Li, Institute for Natural Language Processing (IMS), University of Stuttgart, Germany; Ngoc Thang Vu, University of Stuttgart, Germany</i>	
3-1-2-MLP: EXPLORING A UNIFIED ASR FOR MULTIPLE SOUTH INDIAN LANGUAGES	830
LEVERAGING MULTILINGUAL ACOUSTIC AND LANGUAGE MODELS	
<i>C. S. Anoop, Indian Institute of Science, Bengaluru, India; A G Ramakrishnan, INDIAN INSTITUTE OF SCIENCE, India</i>	
4-1-5-MLP: A TRULY MULTILINGUAL FIRST PASS AND MONOLINGUAL SECOND PASS	838
STREAMING ON-DEVICE ASR SYSTEM	
<i>Sepand Mavandadi, Bo Li, Chao Zhang, Brian Farris, Tara Sainath, Trevor Strohman, Google, United States</i>	

Emotion recognition and paralinguistics

1-1-20-EMR: SPEECH EMOTION RECOGNITION WITH COMPLEMENTARY ACOUSTIC	846
REPRESENTATIONS	
<i>Xiaoming Zhang, Nanjing University of Technology, China; Fan Zhang, IBM Massachusetts Laboratory, United States; Xiaodong Cui, IBM T. J. Watson Research Center, United States; Wei Zhang, Wayfair, United States</i>	
1-2-20-EMR: A ZERO-SHOT APPROACH TO IDENTIFYING CHILDREN'S SPEECH IN	853
AUTOMATIC GENDER CLASSIFICATION	
<i>Amruta Saraf, Ganesh Sivaraman, Elie Khoury, Pindrop, United States</i>	
2-1-19-EMR: DISTRIBUTION-BASED EMOTION RECOGNITION IN CONVERSATION	860
<i>Wen Wu, Chao Zhang, University of Cambridge, United Kingdom; Phil Woodland, Machine Intelligence Laboratory, Cambridge University Department of Engineering, United Kingdom</i>	
3-1-19-EMR: EXPLORATION OF A SELF-SUPERVISED SPEECH MODEL: A STUDY ON	868
EMOTIONAL CORPORA	
<i>Yuanhao Li, Yumnah Mohamied, Peter Bell, Catherine Lai, University of Edinburgh, United Kingdom</i>	
4-1-19-EMR: COMBINING CONTRASTIVE AND NON-CONTRASTIVE LOSSES FOR	876
FINE-TUNING PRETRAINED MODELS IN SPEECH ANALYSIS	
<i>Florian Lux, Ching-Yi Chen, Ngoc Thang Vu, University of Stuttgart, Germany</i>	

Speech synthesis and spoken language generation

1-1-21-TTS: WAVEFIT: AN ITERATIVE AND NON-AUTOREGRESSIVE NEURAL VOCODER	884
BASED ON FIXED-POINT ITERATION	
<i>Yuma Koizumi, Heiga Zen, Michiel Bacchiani, Google, Japan; Kohei Yatabe, Tokyo University of Agriculture and Technology, Japan</i>	

1-1-22-TTS: ON GRANULARITY OF PROSODIC REPRESENTATIONS IN EXPRESSIVE TEXT-TO-SPEECH	892
<i>Mikolaj Babianski, Kamil Pokora, Raahil Shah, Rafał Sienkiewicz, Daniel Korzekwa, Viacheslav Klimkov, Amazon, Poland</i>	
1-1-23-TTS: CAN WE USE COMMON VOICE TO TRAIN A MULTI-SPEAKER TTS SYSTEM?	900
<i>Sewade O Ogun, Emmanuel Vincent, Inria, France; Vincent Colotte, LORIA, France</i>	
1-2-21-TTS: GAN YOU HEAR ME? RECLAIMING UNCONDITIONAL SPEECH SYNTHESIS FROM DIFFUSION MODELS	906
<i>Matthew Baas, Herman Kamper, Stellenbosch University, South Africa</i>	
1-2-22-TTS: ANONYMIZING SPEECH WITH GENERATIVE ADVERSARIAL NETWORKS TO PRESERVE SPEAKER PRIVACY	912
<i>Sarina Meyer, Pascal Tilli, Pavel Denisov, Florian Lux, Julia Koch, Ngoc Thang Vu, University of Stuttgart, Germany</i>	
2-1-20-TTS: STYLETTS-VC: ONE-SHOT VOICE CONVERSION BY KNOWLEDGE TRANSFER FROM STYLE-BASED TTS MODELS	920
<i>Yinghao A Li, Nima Mesgarani, Columbia University, United States; Cong Han, Columbia Univeristy, United States</i>	
2-1-21-TTS: LEARNING ACCENT REPRESENTATION WITH MULTI-LEVEL VAE TOWARDS CONTROLLABLE SPEECH SYNTHESIS	928
<i>Jan Melechovsky, Dorien Herremans, Singapore University of Technology and Design, Singapore; Ambuj Mehrish, SUTD, Singapore; Berrak Sisman, Singapore University of Technology and Design (SUTD), Singapore</i>	
2-1-22-TTS: VTTS: VISUAL-TEXT TO SPEECH	936
<i>Yoshifumi Nakano, Takaaki Saeki, Shinnosuke Takamichi, Hiroshi Saruwatari, The University of Tokyo, Japan; Katsuhito Sudoh, Nara Institute of Science and Techonology, Japan</i>	
2-2-20-TTS: GENERATIVE MODELS FOR IMPROVED NATURALNESS, INTELLIGIBILITY, AND VOICING OF WHISPERED SPEECH	943
<i>Dominik Wagner, Sebastian P Bayerl, Technische Hochschule Nürnberg Georg Simon Ohm, Germany; Hector Cordourier, Intel, Mexico; Tobias Bocklet, TH Nürnberg, Germany</i>	
2-2-21-TTS: TWO-STAGE TRAINING METHOD FOR JAPANESE ELECTROLARYNGEAL SPEECH ENHANCEMENT BASED ON SEQUENCE-TO-SEQUENCE VOICE CONVERSION	949
<i>Ding Ma, Lester Phillip G Violeta, Kazuhiro Kobayashi, Tomoki Toda, Nagoya University, Japan</i>	
3-1-20-TTS: SIMD-SIZE AWARE WEIGHT REGULARIZATION FOR FAST NEURAL VOCODING ON CPU	955
<i>Hiroki Kanagawa, Yusuke Ijima, NTT Corporation, Japan</i>	
3-1-21-TTS: EXACT PROSODY CLONING IN ZERO-SHOT MULTISPEAKER TEXT-TO-SPEECH	962
<i>Florian Lux, Julia Koch, Ngoc Thang Vu, University of Stuttgart, Germany</i>	

3-1-22-TTS: NIX-TTS: LIGHTWEIGHT AND END-TO-END TEXT-TO-SPEECH VIA970
MODULE-WISE DISTILLATION

Rendi Chevi, Radityo Eko Prasajo, Kata.ai, Indonesia; Alham Fikri Aji, Amazon, United Kingdom; Andros Tjandra, Meta AI, US, United States; Sakriani Sakti, Japan Advanced Institute of Science and Technology, Japan

4-1-20-TTS: REGOTRON: REGULARIZING THE TACOTRON2 ARCHITECTURE VIA977
MONOTONIC ALIGNMENT LOSS

Efthymios Georgiou, Georgios Paraskevopoulos, Alexandros Potamianos, National Technical University of Athens, Greece; Kosmas Kritis, Vassilis Katsouros, Athena Research Center, Greece; Athanasios Katsamanis, ATHENA R.C., Behavioral Signal Technologies, Greece

4-1-21-TTS: REMAP, WARP AND ATTEND: NON-PARALLEL MANY-TO-MANY ACCENT984
CONVERSION WITH NORMALIZING FLOWS

Abdelhamid Ezzerg, Thomas Merritt, Kayoko Yanagisawa, Piotr Bilinski, Kamil Pokora, Renard Korzeniowski, Roberto Barra-Chicote, Daniel Korzekwa, Amazon, United Kingdom; Magdalena Proszewska, Jagiellonian University, Poland

Resources (new corpora, toolkits, evaluation metrics, etc.)

1-2-23-RES: STOP: A DATASET FOR SPOKEN TASK ORIENTED SEMANTIC PARSING.....991

Paden Tomasello, Akshat Shrivastava, Daniel A Lazar, Po-chun Hsu, Duc Le, Ali Elkahky, Jade Copet, Robin Algayres, Tu Anh Nguyen, Meta, United States; Adithya Sagar, Facebook AI, United States; Wei-Ning Hsu, Massachusetts Institute of Technology, United States; Yossi Adi, Emmanuel Dupoux, Facebook AI Research, Israel; Luke Zettlemoyer, Facebook, United States; Abdel-rahman Mohamed, Facebook AI Research (FAIR), United States

2-2-23-RES: BENCHMARKING EVALUATION METRICS FOR CODE-SWITCHING999
AUTOMATIC SPEECH RECOGNITION

Injy Hamed, New York University Abu Dhabi, Stuttgart University, United Arab Emirates; Amir Hussein, Johns Hopkins University, United States; Oumnia Chellah, Stanford University, United States; Shammur Chowdhury, QCRI, Qatar; Hamdy Mubarak, Ahmed Ali, Qatar Computing Research Institute, HBKU, Qatar; Sunayana Sitaram, Microsoft Research, India; Nizar Habash, New York University Abu Dhabi, United Arab Emirates

4-1-22-RES: MASC: MASSIVE ARABIC SPEECH CORPUS1006

Mohammad Al-Fetyani, Mohammad AlBarham, Appswave, Jordan; Gheith A. Abandah, Adham Alsharkawi, The University of Jordan, Jordan; Maha Dawas, Planning and Statistics Authority, Qatar

Machine learning for speech applications

1-1-15-MLS: SPEED-ROBUST KEYWORD SPOTTING VIA SOFT SELF-ATTENTION ON1014
MULTI-SCALE FEATURES

Chaoyue Ding, Jiakui Li, Martin Zong, Baoxiang Li, SenseTime Group Limited, China

1-1-24-MLS: DISTILLING SEQUENCE-TO-SEQUENCE VOICE CONVERSION MODELS1022
FOR STREAMING CONVERSION APPLICATIONS

Kou Tanaka, NTT corporation, Japan; Hirokazu Kameoka, NTT Communication Science Laboratories, NTT Corporation, Japan; Takuhiro Kaneko, Shogo Seki, NTT Corporation, Japan

1-1-25-MLS: AUTOMATIC PREDICTION OF INTELLIGIBILITY OF WORDS AND PHONEMES 1029
PRODUCED ORALLY BY JAPANESE LEARNERS OF ENGLISH
Nobuaki Minematsu, Chuanbo Zhu, Takuya Kunihara, Daisuke Saito, The University of Tokyo, Japan; Noriko Nakanishi, Kobe Gakuin University, Japan

1-2-24-MLS: SVLDL: IMPROVED SPEAKER AGE ESTIMATION USING SELECTIVE1037
VARIANCE LABEL DISTRIBUTION LEARNING
Zuheng Kang, Jianzong Wang, Junqing Peng, Ping An Technology (Shenzhen) Co., Ltd, China; Jing Xiao, Ping An Insurance (Group) Company of China, China

1-2-25-MLS: PEPPANET: EFFECTIVE MISPRONUNCIATION DETECTION AND DIAGNOSIS ..1045
LEVERAGING PHONETIC, PHONOLOGICAL, AND ACOUSTIC CUES
Bi-Cheng Yan, Hsin-Wei Wang, Berlin Chen, National Taiwan Normal University, Taiwan

2-1-24-MLS: IMPLICIT ACOUSTIC ECHO CANCELLATION FOR KEYWORD SPOTTING AND .1052
DEVICE-DIRECTED SPEECH DETECTION
Samuele Cornell, Università Politecnica delle Marche, Italy; Thomas Balestri, Thibaud Senechal, Amazon, United States

2-2-17-MLS: TDOA ESTIMATION OF SPEECH SOURCE IN NOISY REVERBERANT1059
ENVIRONMENTS
Sulian Bu, Tuo Zhao, Yunxin Zhao, University of Missouri, United States

2-2-24-MLS: PHONEME SEGMENTATION USING SELF-SUPERVISED SPEECH MODELS1067
Luke Strgar, University of Texas, Austin, United States; David Harwath, The University of Texas at Austin, United States

3-1-23-MLS: AN EXPERIMENTAL STUDY ON PRIVATE AGGREGATION OF TEACHER1074
ENSEMBLE LEARNING FOR END-TO-END SPEECH RECOGNITION
Chao-Han Huck Yang, Chin-hui Lee, Georgia Institute of Technology, United States; I-Fan Chen, Amazon Inc., United States; Andreas Stolcke, Amazon, United States; Sabato M Siniscalchi, Kore University of Enna, Italy

4-1-23-MLS: PHONE-LEVEL PRONUNCIATION SCORING FOR L1 USING1081
WEIGHTED-DYNAMIC TIME WARPING
Aghilas Sini, Antoine Perquin, Damien Lolive, Univ Rennes, CNRS, IRISA, France; Arnaud Delhay, IRISA, France

4-1-24-MLS: PROFICIENCY ASSESSMENT OF L2 SPOKEN ENGLISH USING WAV2VEC 2.0....1088
Stefano Bannò, University of Trento, Italy; Marco Matassoni, Fondazione Bruno Kessler, Italy

SUPERB challenge

4-2-1-SUP: SUPERB @ SLT 2022: CHALLENGE ON GENERALIZATION AND EFFICIENCY OF 1096
SELF-SUPERVISED SPEECH REPRESENTATION LEARNING
Tzu-hsun Feng, Shu-wen Yang, Tzu-Quan Lin, Kai-Wei Chang, Haibin Wu, Hung-yi Lee, National Taiwan University, Taiwan; Annie Dong, Shang-Wen Li, Meta, United States; Ching-Feng Yeh, Facebook, United States; Jiatong Shi, Xuankai Chang, Shinji Watanabe, Carnegie Mellon University, United States; Zili Huang, Johns Hopkins University, United States; Abdel-rahman Mohamed, Facebook AI Research (FAIR), United States

1-1-26-SUP: ON THE UTILITY OF SELF-SUPERVISED MODELS FOR PROSODY-RELATED ...1104
TASKS

Guan-Ting Lin, Chi Luen Feng, Wei-Ping Huang, Yuan Tseng, Chen An Li, Tzu-Han Lin, Hung-yi Lee, National Taiwan University, Taiwan; Nigel Ward, UTEP, United States

2-1-25-SUP: IMPROVING GENERALIZABILITY OF DISTILLED SELF-SUPERVISED SPEECH ...1112
PROCESSING MODELS UNDER DISTORTED SETTINGS

Kuan-Po Huang, Tsu-Yuan Hsu, Liang-Hsuan Tseng, Hung-yi Lee, National Taiwan University, Taiwan; Yu-kuan Fu, NTU, Taiwan; Fabian Alejandro Ritter Gutierrez, National University of Singapore, Singapore; Fan-Lin Wang, Academia Sinica, Taiwan; Yu Zhang, Google, United States

2-2-25-SUP: EXPLORING EFFICIENT-TUNING METHODS IN SELF-SUPERVISED SPEECH1120
MODELS

Zih-Ching Chen, Chin-Lun Fu, Chih Ying Liu, Hung-yi Lee, National Taiwan University, Taiwan; Shang-Wen Li, AWS AI, United States

3-1-25-SUP: ON COMPRESSING SEQUENCES FOR SELF-SUPERVISED SPEECH MODELS1128

Yen Meng, Hsuan-Jui Chen, Hung-yi Lee, National Taiwan University, Taiwan; Jiatong Shi, Shinji Watanabe, Carnegie Mellon University, United States; Paola Garcia, Johns Hopkins University, United States; Hao Tang, The University of Edinburgh, United Kingdom

4-1-25-SUP: EXTRACTING SPEAKER AND EMOTION INFORMATION FROM1136
SELF-SUPERVISED SPEECH MODELS VIA CHANNEL-WISE CORRELATIONS

Themis Stafylakis, Omilia - Conversational Intelligence, Greece; Ladislav Mošner, Plchot Oldřich, Lukas Burget, Jan Cernocky, Brno University of Technology, Czechia; Sofoklis Kakouros, University of Helsinki, Finland