

2024 IEEE Spoken Language Technology Workshop (SLT 2024)

**Macao
2-5 December 2024**

Pages 1-629



**IEEE Catalog Number: CFP24SLT-POD
ISBN: 979-8-3503-9226-5**

**Copyright © 2024 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:	CFP24SLT-POD
ISBN (Print-On-Demand):	979-8-3503-9226-5
ISBN (Online):	979-8-3503-9225-8

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 Content

Welcome	XIX
Conference Committees	XX
Keynote Speeches	XXXI
Invited Talks	XXXVII
Panel Discussion	XLII
Recent Breakthrough	XLIV
Sponsors	XLVI

Papers

Poster Session 1: Speech Recognition

P1-01-ASR PERSONALIZING LARGE SEQUENCE-TO-SEQUENCE SPEECH FOUNDATION MODELS WITH SPEAKER REPRESENTATIONS	1
<i>Dominik Wagner, Ilja Baumann, Thomas Ranzenberger, Korbinian Riedhammer, Tobias Bocklet</i>	
P1-02-ASR LABEL-LOOPING: HIGHLY EFFICIENT DECODING FOR TRANSDUCERS	7
<i>Vladimir Bataev, Hainan Xu, Daniel Galvez, Vitaly Lavrukhin, Boris Ginsburg</i>	
P1-03-ASR ADVANCING MULTI-TALKER ASR PERFORMANCE WITH LARGE LANGUAGE MODELS	14
<i>Mohan Shi, Zengrui Jin, Yaoxun Xu, Yong Xu, Shi-Xiong Zhang, Kun Wei, Yiwon Shao, Chunlei Zhang, Dong Yu</i>	
P1-04-ASR TOKEN-WEIGHTED RNN-T FOR LEARNING FROM FLAWED DATA	22
<i>Gil Keren, Wei Zhou, Ozlem Kalinli</i>	
P1-05-ASR ENHANCING CODE-SWITCHING SPEECH RECOGNITION WITH LID-BASED COLLABORATIVE MIXTURE OF EXPERTS MODEL	30
<i>Hukai Huang, Jiayan Lin, Kaidi Wang, Yishuang Li, Wenhao Guan, Lin Li, Qingyang Hong</i>	
P1-06-ASR LANGUAGE BIAS IN SELF-SUPERVISED LEARNING FOR AUTOMATIC SPEECH RECOGNITION	37
<i>Edward Storey, Naomi Harte, Peter Bell</i>	
P1-07-ASR ROBUST AUDIOVISUAL SPEECH RECOGNITION MODELS WITH MIXTURE-OF- EXPERTS	43
<i>Yihan Wu, Yifan Peng, Yichen Lu, Xuankai Chang, Ruihua Song, Shinji Watanabe</i>	

P1-08-ASR HYBRID ATTENTION-BASED ENCODER-DECODER MODEL FOR EFFICIENT LANGUAGE MODEL ADAPTATION	49
<i>Shaoshi Ling, Guoli Ye, Rui Zhao, Yifan Gong</i>	
P1-09-ASR SPATIALEMB: EXTRACT AND ENCODE SPATIAL INFORMATION FOR 1-STAGE MULTI-CHANNEL MULTI-SPEAKER ASR ON ARBITRARY MICROPHONE ARRAYS	56
<i>Yiwen Shao, Yong Xu, Sanjeev Khudanpur, Dong Yu</i>	
P1-10-ASR EFFECTIVE TEXT ADAPTATION FOR LLM-BASED ASR THROUGH SOFT PROMPT FINE-TUNING	64
<i>Yingyi Ma, Zhe Liu, Ozlem Kalinli</i>	
P1-11-ASR TEMPORAL ORDER PRESERVED OPTIMAL TRANSPORT-BASED CROSS-MODAL KNOWLEDGE TRANSFER LEARNING FOR ASR	70
<i>Xugang Lu, Peng Shen, Yu Tsao, Hisashi Kawai</i>	
P1-12-ASR CONTEXTUALIZED AUTOMATIC SPEECH RECOGNITION WITH DYNAMIC VOCABULARY	78
<i>Yui Sudo, Yosuke Fukumoto, Muhammad Shakeel, Yifan Peng, Shinji Watanabe</i>	
P1-13-ASR DO PROMPTS REALLY PROMPT? EXPLORING THE PROMPT UNDERSTANDING CAPABILITY OF WHISPER	86
<i>Chih-Kai Yang, Kuan-Po Huang, Hung-yi Lee</i>	
P1-14-ASR AN EFFECTIVE CONTEXT-BALANCED ADAPTATION APPROACH FOR LONG-TAILED SPEECH RECOGNITION	94
<i>Yi-Cheng Wang, Li-Ting Pai, Bi-Cheng Yan, Hsin-Wei Wang, Chi-Han Lin, Berlin Chen</i>	
P1-15-ASR TRAINING LARGE ASR ENCODERS WITH DIFFERENTIAL PRIVACY	102
<i>Geeticka Chauhan, Steve Chien, Om Thakkar, Abhradeep Thakurta, Arun Narayanan</i>	
P1-16-ASR TRANSDUCER CONSISTENCY REGULARIZATION FOR SPEECH TO TEXT APPLICATIONS	110
<i>Cindy Tseng, Yun Tang, Vijendra Raj Apsingekar</i>	
P1-17-ASR LEAVE NO KNOWLEDGE BEHIND DURING KNOWLEDGE DISTILLATION: TOWARDS PRACTICAL AND EFFECTIVE KNOWLEDGE DISTILLATION FOR CODE-SWITCHING ASR USING REALISTIC DATA	118
<i>Liang-Hsuan Tseng, Zih-Ching Chen, Wei-Shun Chang, Cheng-Kuang Lee, Tsung-Ren Huang, Hung-yi Lee</i>	
P1-18-ASR CTC-ASSISTED LLM-BASED CONTEXTUAL ASR	126
<i>Guanrou Yang, Ziyang Ma, Zhifu Gao, Shiliang Zhang, Xie Chen</i>	
P1-19-ASR AUTOMATIC TIME ALIGNMENT GENERATION FOR END-TO-END ASR USING ACOUSTIC PROBABILITY MODELLING	132
<i>Dongcheng Jiang, Chao Zhang, Philip C. Woodland</i>	

P1-20-ASR CONTINUAL LEARNING WITH EMBEDDING LAYER SURGERY AND TASK-WISE BEAM SEARCH USING WHISPER	140
<i>Chin Yuen Kwok, Jia Qi Yip, Eng Siong Chng</i>	

P1-21-ASR BESTOW: EFFICIENT AND STREAMABLE SPEECH LANGUAGE MODEL WITH THE BEST OF TWO WORLDS IN GPT AND T5	147
<i>Zhehuai Chen, He Huang, Oleksii Hrinchuk, Krishna C. Puvvada, Nithin Rao Koluguri, Piotr Żelasko, Jagadeesh Balam, Boris Ginsburg</i>	

P1-22-ASR COMBINING TF-GRIDNET AND MIXTURE ENCODER FOR CONTINUOUS SPEECH SEPARATION FOR MEETING TRANSCRIPTION	155
<i>Peter Vieting, Simon Berger, Thilo Von Neumann, Christoph Boeddeker, Ralf Schlüter, Reinhold Haeb-Umbach</i>	

P1-23-ASR MAMBA-BASED DECODER-ONLY APPROACH WITH BIDIRECTIONAL SPEECH MODELING FOR SPEECH RECOGNITION	163
<i>Yoshiki Masuyama, Koichi Miyazaki, Masato Murata</i>	

P1-24-ASR AN ANALYSIS OF LINEAR COMPLEXITY ATTENTION SUBSTITUTES WITH BEST-RQ	169
<i>Ryan Whetten, Titouan Parcollet, Adel Moumen, Marco Dinarelli, Yannick Estève</i>	

P1-25-ASR SPEECH-MAMBA: LONG-CONTEXT SPEECH RECOGNITION WITH SELECTIVE STATE SPACES MODELS	177
<i>Xiaoxue Gao, Nancy F. Chen</i>	

P1-26-ASR LITE ASR TRANSFORMER: A LIGHT WEIGHT TRANSFORMER ARCHITECTURE FOR AUTOMATIC SPEECH RECOGNITION	185
<i>Metilda Sagaya Mary N J, S Umesh</i>	

Poster Session 2: Speech Recognition and Enhancement

P2-01-ASR SERIALIZED SPEECH INFORMATION GUIDANCE WITH OVERLAPPED ENCODING SEPARATION FOR MULTI-SPEAKER AUTOMATIC SPEECH RECOGNITION	193
<i>Hao Shi, Yuan Gao, Zhaoheng Ni, Tatsuya Kawahara</i>	

P2-02-ASR EFFICIENT EXTRACTION OF NOISE-ROBUST DISCRETE UNITS FROM SELF-SUPERVISED SPEECH MODELS	200
<i>Jakob Poncelet, Yujun Wang, Hugo Van Hamme</i>	

P2-03-ASR CONTROLLING WHISPER: UNIVERSAL ACOUSTIC ADVERSARIAL ATTACKS TO CONTROL MULTI-TASK AUTOMATIC SPEECH RECOGNITION MODELS	208
<i>Vyas Raina, Mark Gales</i>	

P2-04-ASR IMPROVING RARE-WORD RECOGNITION OF WHISPER IN ZERO-SHOT SETTINGS	216
<i>Yash Jogi, Vaibhav Aggarwal, Shabari S Nair, Yash Verma, Aayush Kubba</i>	

P2-05-ASR AUGMENTING AUTOMATIC SPEECH RECOGNITION MODELS WITH DISFLUENCY DETECTION	224
<i>Robin Amann, Zhaolin Li, Barbara Bruno, Jan Niehues</i>	
P2-06-ASR ENHANCING UNIFIED STREAMING AND NON-STREAMING ASR THROUGH CURRICULUM LEARNING WITH EASY-TO-HARD TASKS	232
<i>Yuting Yang, Yuke Li, Lifeng Zhou, Binbin Du, Haoqi Zhu</i>	
P2-07-ASR DQ-WHISPER: JOINT DISTILLATION AND QUANTIZATION FOR EFFICIENT MULTILINGUAL SPEECH RECOGNITION	240
<i>Hang Shao, Bei Liu, Wei Wang, Xun Gong, Yanmin Qian</i>	
P2-08-ASR FUSION OF DISCRETE REPRESENTATIONS AND SELF-AUGMENTED REPRESENTATIONS FOR MULTILINGUAL AUTOMATIC SPEECH RECOGNITION	247
<i>Shih-Heng Wang, Jiatong Shi, Chien-yu Huang, Shinji Watanabe, Hung-yi Lee</i>	
P2-09-ASR LONGER IS (NOT NECESSARILY) STRONGER: PUNCTUATED LONG-SEQUENCE TRAINING FOR ENHANCED SPEECH RECOGNITION AND TRANSLATION	255
<i>Nithin Rao Koluguri, Travis Bartley, Hainan Xu, Oleksii Hrinchuk, Jagadeesh Balam, Boris Ginsburg, Georg Kucsiko</i>	
P2-10-ASR SEMI-SUPERVISED LEARNING FOR CODE-SWITCHING ASR WITH LARGE LANGUAGE MODEL FILTER	263
<i>Yu Xi, Wen Ding, Kai Yu, Junjie Lai</i>	
P2-11-ASR PARAMETER AVERAGING IS ALL YOU NEED TO PREVENT FORGETTING	271
<i>Peter Plantinga, Jaekwon Yoo, Abenezer Girma, Chandra Dhir</i>	
P2-12-ASR ADVANCING CTC MODELS FOR BETTER SPEECH ALIGNMENT: A TOPOLOGICAL APPROACH	279
<i>Zeyu Zhao, Peter Bell</i>	
P2-13-SES DUALSEP: A LIGHT-WEIGHT DUAL-ENCODER CONVOLUTIONAL RECURRENT NETWORK FOR REAL-TIME IN-CAR SPEECH SEPARATION	286
<i>Ziqian Wang, Jiayao Sun, Zihan Zhang, Xingchen Li, Jie Liu, Lei Xie</i>	
P2-14-SES DDTSE: DISCRIMINATIVE DIFFUSION MODEL FOR TARGET SPEECH EXTRACTION	294
<i>Leying Zhang, Yao Qian, Linfeng Yu, Heming Wang, Hemin Yang, Shujie Liu, Long Zhou, Yanmin Qian</i>	
P2-15-SES AN INVESTIGATION OF INCORPORATING MAMBA FOR SPEECH ENHANCEMENT	302
<i>Rong Chao, Wen-Huang Cheng, Moreno La Quatra, Sabato Marco Siniscalchi, Chao-Han Huck Yang, Szu-Wei Fu, Yu Tsao</i>	

P2-16-SES EFFECTIVE NOISE-AWARE DATA SIMULATION FOR DOMAIN-ADAPTIVE SPEECH ENHANCEMENT LEVERAGING DYNAMIC STOCHASTIC PERTURBATION	309
<i>Chien-Chun Wang, Li-Wei Chen, Hung-Shin Lee, Berlin Chen, Hsin-Min Wang</i>	
P2-17-SES SMRU: SPLIT-AND-MERGE RECURRENT-BASED UNET FOR ACOUSTIC ECHO CANCELLATION AND NOISE SUPPRESSION	317
<i>Zhihang Sun, Andong Li, Rilin Chen, Hao Zhang, Meng Yu, Yi Zhou, Dong Yu</i>	
P2-18-SES ON THE EFFECTIVENESS OF ENROLLMENT SPEECH AUGMENTATION FOR TARGET SPEAKER EXTRACTION	325
<i>Junjie Li, Ke Zhang, Shuai Wang, Haizhou Li, Man-Wai Mak, Kong Aik Lee</i>	
P2-19-SES DIFFUSION-BASED GENERATIVE MODELING WITH DISCRIMINATIVE GUIDANCE FOR STREAMABLE SPEECH ENHANCEMENT	333
<i>Chenda Li, Samuele Cornell, Shinji Watanabe, Yanmin Qian</i>	
P2-20-SES NEUROSPEX: NEURO-GUIDED SPEAKER EXTRACTION WITH CROSS-MODAL FUSION	341
<i>Dashanka De Silva, Siqi Cai, Saurav Pahuja, Tanja Schultz, Haizhou Li</i>	
P2-21-SES ENHANCING SPEAKER EXTRACTION THROUGH RECTIFYING TARGET CONFUSION	349
<i>Jiahe Wang, Shuai Wang, Junjie Li, Ke Zhang, Yanmin Qian, Haizhou Li</i>	
P2-22-SES DIFF-PLC: A DIFFUSION-BASED APPROACH FOR EFFECTIVE PACKET LOSS CONCEALMENT	357
<i>Da-Hee Yang, Joon-Hyuk Chang</i>	
P2-23-SES IMPROVING CURRICULUM LEARNING FOR TARGET SPEAKER EXTRACTION WITH SYNTHETIC SPEAKERS	364
<i>Yun Liu, Xuechen Liu, Junichi Yamagishi</i>	
P2-24-SS02 LARGE LANGUAGE MODEL BASED GENERATIVE ERROR CORRECTION: A CHALLENGE AND BASELINES FOR SPEECH RECOGNITION, SPEAKER TAGGING, AND EMOTION RECOGNITION	371
<i>Chao-Han Huck Yang, Taejin Park, Yuan Gong, Yuanchao Li, Zhehuai Chen, Yen-Ting Lin, Chen Chen, Yuchen Hu, Kunal Dhawan, Piotr Żelasko, Chao Zhang, Yun-Nung Chen, Yu Tsao, Jagadeesh Balam, Boris Ginsburg, Sabato Marco Siniscalchi, Eng Siong Chng, Peter Bell, Catherine Lai, Shinji Watanabe, Andreas Stolcke</i>	
P2-25-SS06 FGCL: FINE-GRAINED CONTRASTIVE LEARNING FOR MANDARIN STUTTERING EVENT DETECTION	379
<i>Han Jiang, Wenyu Wang, Yiquan Zhou, Hongwu Ding, Jiacheng Xu, Jihua Zhu</i>	
P2-26-SS06 FINDINGS OF THE 2024 MANDARIN STUTTERING EVENT DETECTION AND AUTOMATIC SPEECH RECOGNITION CHALLENGE	385
<i>Hongfei Xue, Rong Gong, Mingchen Shao, Xin Xu, Lezhi Wang, Lei Xie, Hui Bu, Jiaming Zhou, Yong Qin, Jun Du, Ming Li, Binbin Zhang, Bin Jia</i>	

P2-27-SS06 ENHANCED ASR FOR STUTTERING SPEECH: COMBINING ADVERSARIAL AND SIGNAL-BASED DATA AUGMENTATION	393
<i>Shangkun Huang, Dejun Zhang, Jing Deng, Rong Zheng</i>	

Poster Session 3: Speech Processing

P3-01-ANA PROPERTY NEURONS IN SELF-SUPERVISED SPEECH TRANSFORMERS	401
<i>Tzu-Quan Lin, Guan-Ting Lin, Hung-yi Lee, Hao Tang</i>	

P3-02-ANA PRIVACY VERSUS EMOTION PRESERVATION TRADE-OFFS IN EMOTION-PRESERVING SPEAKER ANONYMIZATION	409
<i>Zexin Cai, Henry Li Xinyuan, Ashi Garg, Leibny Paola García-Perera, Kevin Duh, Sanjeev Khudanpur, Nicholas Andrews, Matthew Wiesner</i>	

P3-03-ANA ESTIMATING THE COMPLETENESS OF DISCRETE SPEECH UNITS	415
<i>Sung-Lin Yeh, Hao Tang</i>	

P3-04-ANA INVESTIGATION OF SPEAKER REPRESENTATION FOR TARGET-SPEAKER SPEECH PROCESSING	423
<i>Takanori Ashihara, Takafumi Moriya, Shota Horiguchi, Junyi Peng, Tsubasa Ochiai, Marc Delcroix, Kohei Matsuura, Hiroshi Sato</i>	

P3-05-MMP CROSSMODAL ASR ERROR CORRECTION WITH DISCRETE SPEECH UNITS ...	431
<i>Yuanchao Li, Pinzhen Chen, Peter Bell, Catherine Lai</i>	

P3-06-MMP LISTEN AND SPEAK FAIRLY: A STUDY ON SEMANTIC GENDER BIAS IN SPEECH INTEGRATED LARGE LANGUAGE MODELS	439
<i>Yi-Cheng Lin, Tzu-Quan Lin, Chih-Kai Yang, Ke-Han Lu, Wei-Chih Chen, Chun-Yi Kuan, Hung-yi Lee</i>	

P3-07-MMP LEARNING VIDEO TEMPORAL DYNAMICS WITH CROSS-MODAL ATTENTION FOR ROBUST AUDIO-VISUAL SPEECH RECOGNITION	447
<i>Sungnyun Kim, Kangwook Jang, Sangmin Bae, Hoirin Kim, Se-Young Yun</i>	

P3-08-MMP DATA EFFICIENT REFLOW FOR FEW STEP AUDIO GENERATION	455
<i>Lemeng Wu, Zhaoheng Ni, Bowen Shi, Gael Le Lan, Anurag Kumar, Varun Nagaraja, Xinhao Mei, Yongyang Xiong, Bilge Soran, Raghuraman Krishnamoorthi, Wei-Ning Hsu, Yangyang Shi, Vikas Chandra</i>	

P3-09-MLS OPTIMIZING BYTE-LEVEL REPRESENTATION FOR END-TO-END ASR	462
<i>Roger Hsiao, Lihui Deng, Erik Mcdermott, Ruchir Travadi, Xiaodan Zhuang</i>	

P3-10-MLS ROMANIZATION ENCODING FOR MULTILINGUAL ASR	468
<i>Wen Ding, Fei Jia, Hainan Xu, Yu Xi, Junjie Lai, Boris Ginsburg</i>	

P3-11-MLS ENHANCING CODE-SWITCHING ASR LEVERAGING NON-PEAKY CTC LOSS AND DEEP LANGUAGE POSTERIOR INJECTION	476
<i>Tzu-Ting Yang, Hsin-Wei Wang, Yi-Cheng Wang, Berlin Chen</i>	

P3-12-MLS LANGUAGE-INDEPENDENT PROSODY-ENHANCED SPEECH REPRESENTATIONS FOR MULTILINGUAL SPEECH SYNTHESIS	482
<i>Chang Liu, Zhen-Hua Ling, Ya-Jun Hu</i>	
P3-13-MLS CLASSIFICATION OF SPONTANEOUS AND SCRIPTED SPEECH FOR MULTILINGUAL AUDIO	489
<i>Shahar Elisha, Andrew McDowell, Mariano Beguerisse-Díaz, Emmanouil Benetos</i>	
P3-14-EMR GMP-TL: GENDER-AUGMENTED MULTI-SCALE PSEUDO-LABEL ENHANCED TRANSFER LEARNING FOR SPEECH EMOTION RECOGNITION	496
<i>Yu Pan, Yuguang Yang, Yuheng Huang, Tiancheng Jin, Jingjing Yin, Yanni Hu, Heng Lu, Lei Ma, Jianjun Zhao</i>	
P3-15-EMR EMBRACING AMBIGUITY AND SUBJECTIVITY USING THE ALL-INCLUSIVE AGGREGATION RULE FOR EVALUATING MULTI-LABEL SPEECH EMOTION RECOGNITION SYSTEMS	502
<i>Huang-Cheng Chou, Haibin Wu, Lucas Goncalves, Seong-Gyun Leem, Ali Salman, Carlos Busso, Hung-yi Lee, Chi-Chun Lee</i>	
P3-16-EMR OPEN-EMOTION: A REPRODUCIBLE EMO-SUPERB FOR SPEECH EMOTION RECOGNITION SYSTEMS	510
<i>Haibin Wu, Huang-Cheng Chou, Kai-Wei Chang, Lucas Goncalves, Jiawei Du, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-yi Lee</i>	
P3-17-EMR SPEECH EMOTION RECOGNITION WITH ASR TRANSCRIPTS: A COMPREHENSIVE STUDY ON WORD ERROR RATE AND FUSION TECHNIQUES	518
<i>Yuanchao Li, Peter Bell, Catherine Lai</i>	
P3-18-EMR BEYOND THE BINARY: LIMITATIONS AND POSSIBILITIES OF GENDER-RELATED SPEECH TECHNOLOGY RESEARCH	526
<i>Ariadna Sanchez, Alice Ross, Nina Markl</i>	
P3-19-EMR ENHANCING DOMAIN GENERALIZATION IN SPEECH EMOTION RECOGNITION BY COMBINING DOMAIN-VARIANT REPRESENTATIONS AND DOMAIN-INVARIANT CLASSIFIERS	533
<i>Shi-wook Lee</i>	
P3-20-SS07 MDCTCODEC: A LIGHTWEIGHT MDCT-BASED NEURAL AUDIO CODEC TOWARDS HIGH SAMPLING RATE AND LOW BITRATE SCENARIOS	540
<i>Xiao-Hang Jiang, Yang Ai, Rui-Chen Zheng, Hui-Peng Du, Ye-Xin Lu, Zhen-Hua Ling</i>	
P3-21-SS07 ADDRESSING INDEX COLLAPSE OF LARGE-CODEBOOK SPEECH TOKENIZER WITH DUAL-DECODING PRODUCT-QUANTIZED VARIATIONAL AUTO-ENCODER	548
<i>Haohan Guo, Fenglong Xie, Dongchao Yang, Hui Lu, Xixin Wu, Helen Meng</i>	
P3-22-SS07 INVESTIGATING NEURAL AUDIO CODECS FOR SPEECH LANGUAGE MODEL-BASED SPEECH GENERATION	554
<i>Jiaqi Li, Dongmei Wang, Xiaofei Wang, Yao Qian, Long Zhou, Shujie Liu, Midia Yousefi,</i>	

Canrun Li, Chung-Hsien Tsai, Zhen Xiao, Yanqing Liu, Junkun Chen, Sheng Zhao, Jinyu Li, Zhizheng Wu, Michael Zeng

P3-23-SS07 ESPNET-CODEC: COMPREHENSIVE TRAINING AND EVALUATION OF NEURAL CODECS FOR AUDIO, MUSIC, AND SPEECH 562
Jiatong Shi, Jinchuan Tian, Yihan Wu, Jee-weon Jung, Jia Qi Yip, Yoshiki Masuyama, William Chen, Yuning Wu, Yuxun Tang, Massa Baali, Dareen Alharthi, Dong Zhang, Ruifan Deng, Tejes Srivastava, Haibin Wu, Alexander Liu, Bhiksha Raj, Qin Jin, Ruihua Song, Shinji Watanabe

P3-24-SS07 CODEC-SUPERB @ SLT 2024: A LIGHTWEIGHT BENCHMARK FOR NEURAL AUDIO CODEC MODELS 570
Haibin Wu, Xuanjun Chen, Yi-Cheng Lin, Kaiwei Chang, Jiawei Du, Ke-Han Lu, Alexander H. Liu, Ho-Lam Chung, Yuan-Kuei Wu, Dongchao Yang, Songxiang Liu, Yi-Chiao Wu, Xu Tan, James Glass, Shinji Watanabe, Hung-yi Lee

P3-25-SS08 OPTIMIZING DYSARTHRIA WAKE-UP WORD SPOTTING: AN END-TO-END APPROACH FOR SLT 2024 LRDWWS CHALLENGE 578
Shuiyun Liu, Yuxiang Kong, Pengcheng Guo, Weiwei Zhuang, Peng Gao, Yujun Wang, Lei Xie

P3-26-SS08 PB-LRDWWS SYSTEM FOR THE SLT 2024 LOW-RESOURCE DYSARTHRIA WAKE-UP WORD SPOTTING CHALLENGE 586
Shiyao Wang, Jiaming Zhou, Shiwan Zhao, Yong Qin

P3-27-SS08 SUMMARY OF LOW-RESOURCE DYSARTHRIA WAKE-UP WORD SPOTTING CHALLENGE 592
Ming Gao, Hang Chen, Jun Du, Xin Xu, Hongxiao Guo, Hui Bu, Ming Li, Chin-Hui Lee

P3-28-SLP PROGRES: PROMPTED GENERATIVE RESCORING ON ASR N-BEST 600
Ada Defne Tur, Adel Moumen, Mirco Ravanelli

P3-29-SS02 FLANEC: EXPLORING FLAN-T5 FOR POST-ASR ERROR CORRECTION 608
Moreno La Quatra, Valerio Mario Salerno, Yu Tsao, Sabato Marco Siniscalchi

Poster Session 4: Speech Synthesis

P4-01-TTS AS-SPEECH: ADAPTIVE STYLE FOR SPEECH SYNTHESIS 616
Zhipeng Li, Xiaofen Xing, Jun Wang, Shuaiqi Chen, Guoqiao Yu, Guanglu Wan, Xiangmin Xu

P4-02-TTS ROOM IMPULSE RESPONSES HELP ATTACKERS TO EVADE DEEP FAKE DETECTION 623
Hieu-Thi Luong, Duc-Tuan Truong, Kong Aik Lee, Eng Siong Chng

P4-03-TTS ATTENTION-CONSTRAINED INFERENCE FOR ROBUST DECODER-ONLY TEXT-TO-SPEECH 630
Hankun Wang, Chenpeng Du, Yiwei Guo, Shuai Wang, Xie Chen, Kai Yu

P4-04-TTS STAGE-WISE AND PRIOR-AWARE NEURAL SPEECH PHASE PREDICTION	638
<i>Fei Liu, Yang Ai, Hui-Peng Du, Ye-Xin Lu, Rui-Chen Zheng, Zhen-Hua Ling</i>	
P4-05-TTS SOCODEC: A SEMANTIC-ORDERED MULTI-STREAM SPEECH CODEC FOR EFFICIENT LANGUAGE MODEL BASED TEXT-TO-SPEECH SYNTHESIS	645
<i>Haohan Guo, Fenglong Xie, Kun Xie, Dongchao Yang, Dake Guo, Xixin Wu, Helen Meng</i>	
P4-06-TTS DETECTING THE UNDETECTABLE: ASSESSING THE EFFICACY OF CURRENT SPOOF DETECTION METHODS AGAINST SEAMLESS SPEECH EDITS	652
<i>Sung-Feng Huang, Heng-Cheng Kuo, Zhehuai Chen, Xuesong Yang, Chao-Han Huck Yang, Yu Tsao, Yu-Chiang Frank Wang, Hung-yi Lee, Szu-Wei Fu</i>	
P4-07-TTS DNN-BASED ENSEMBLE SINGING VOICE SYNTHESIS WITH INTERACTIONS BETWEEN SINGERS	660
<i>Hiroaki Hyodo, Shinnosuke Takamichi, Tomohiko Nakamura, Junya Koguchi, Hiroshi Saruwatari</i>	
P4-08-TTS INVESTIGATING DISENTANGLEMENT IN A PHONEME-LEVEL SPEECH CODEC FOR PROSODY MODELING	668
<i>Sotirios Karapiperis, Nikolaos Ellinas, Alexandra Vioni, Junkwang Oh, Gunu Jho, Inchul Hwang, Spyros Raptis</i>	
P4-09-TTS INSTRUCTSING: HIGH-FIDELITY SINGING VOICE GENERATION VIA INSTRUCTING YOURSELF	675
<i>Chang Zeng, Chunhui Wang, Xiaoxiao Miao, Jian Zhao, Zhonglin Jiang, Yong Chen</i>	
P4-10-TTS E2 TTS: EMBARRASSINGLY EASY FULLY NON-AUTOREGRESSIVE ZERO-SHOT TTS	682
<i>Sefik Emre Eskimez, Xiaofei Wang, Manthan Thakker, Canrun Li, Chung-Hsien Tsai, Zhen Xiao, Hemin Yang, Zirun Zhu, Min Tang, Xu Tan, Yanqing Liu, Sheng Zhao, Naoyuki Kanda</i>	
P4-11-TTS LAUGH NOW CRY LATER: CONTROLLING TIME-VARYING EMOTIONAL STATES OF FLOW-MATCHING-BASED ZERO-SHOT TEXT-TO-SPEECH	690
<i>Haibin Wu, Xiaofei Wang, Sefik Emre Eskimez, Manthan Thakker, Daniel Tompkins, Chung- Hsien Tsai, Canrun Li, Zhen Xiao, Sheng Zhao, Jinyu Li, Naoyuki Kanda</i>	
P4-12-TTS DISENTANGLING THE PROSODY AND SEMANTIC INFORMATION WITH PRE- TRAINED MODEL FOR IN-CONTEXT LEARNING BASED ZERO-SHOT VOICE CONVERSION	698
<i>Zhengyang Chen, Shuai Wang, Mingyang Zhang, Xuechen Liu, Junichi Yamagishi, Yanmin Qian</i>	
P4-13-TTS NDVQ: ROBUST NEURAL AUDIO CODEC WITH NORMAL DISTRIBUTION-BASED VECTOR QUANTIZATION	705
<i>Zhikang Niu, Sanyuan Chen, Long Zhou, Ziyang Ma, Xie Chen, Shujie Liu</i>	

P4-14-TTS FAST, HIGH-QUALITY AND PARAMETER-EFFICIENT ARTICULATORY SYNTHESIS USING DIFFERENTIABLE DSP	711
<i>Yisi Liu, Bohan Yu, Drake Lin, Peter Wu, Cheol Jun Cho, Gopala Krishna Anumanchipalli</i>	
P4-15-TTS VISINGER2+: END-TO-END SINGING VOICE SYNTHESIS AUGMENTED BY SELF-SUPERVISED LEARNING REPRESENTATION	719
<i>Yifeng Yu, Jiatong Shi, Yuning Wu, Yuxun Tang, Shinji Watanabe</i>	
P4-16-TTS END-TO-END STREAMING MODEL FOR LOW-LATENCY SPEECH ANONYMIZATION	727
<i>Waris Quamer, Ricardo Gutierrez-Osuna</i>	
P4-17-TTS EMOTION-COHERENT SPEECH DATA AUGMENTATION AND SELF-SUPERVISED CONTRASTIVE STYLE TRAINING FOR ENHANCING KIDS'S STORY SPEECH SYNTHESIS	735
<i>Raymond Chung</i>	
P4-18-TTS DISCRETE UNIT BASED MASKING FOR IMPROVING DISENTANGLEMENT IN VOICE CONVERSION	742
<i>Philip H. Lee, Ismail Rasim Ulgen, Berrak Sisman</i>	
P4-19-TTS CROSS-DIALECT TEXT-TO-SPEECH IN PITCH-ACCENT LANGUAGE INCORPORATING MULTI-DIALECT PHONEME-LEVEL BERT	750
<i>Kazuki Yamauchi, Yuki Saito, Hiroshi Saruwatari</i>	
P4-20-TTS LEVERAGING DIVERSE SEMANTIC-BASED AUDIO PRETRAINED MODELS FOR SINGING VOICE CONVERSION	758
<i>Xueyao Zhang, Zihao Fang, Yicheng Gu, Haopeng Chen, Lexiao Zou, Junan Zhang, Liumeng Xue, Zhizheng Wu</i>	
P4-21-TTS TTSDS: TEXT-TO-SPEECH DISTRIBUTION SCORE	766
<i>Christoph Minixhofer, Ondřej Klejch, Peter Bell</i>	
P4-22-SS03 SPEECH FOUNDATION MODEL ENSEMBLES FOR THE CONTROLLED SINGING VOICE DEEPFAKE DETECTION (CTRSVDD) CHALLENGE 2024	774
<i>Anmol Guragain, Tianchi Liu, Zihan Pan, Hardik B. Sailor, Qiongqiong Wang</i>	
P4-23-SS03 SVDD 2024: THE INAUGURAL SINGING VOICE DEEPFAKE DETECTION CHALLENGE	782
<i>You Zhang, Yongyi Zang, Jiatong Shi, Ryuichi Yamamoto, Tomoki Toda, Zhiyao Duan</i>	
P4-24-SS03 XWSB: A BLEND SYSTEM UTILIZING XLS-R AND WAVLM WITH SLS CLASSIFIER DETECTION SYSTEM FOR SVDD 2024 CHALLENGE	788
<i>Qishan Zhang, Shuangbing Wen, Fangke Yan, Tao Hu, Jun Li</i>	
P4-25-SS03 INTEGRATING SELF-SUPERVISED PRE-TRAINING WITH ADVERSARIAL LEARNING FOR SYNTHESIZED SONG DETECTION	795
<i>Yankai Wang, Yuxuan Du, Dejun Zhang, Rong Zheng, Jing Deng</i>	

P4-26-SS05 THE VOICEMOS CHALLENGE 2024: BEYOND SPEECH QUALITY PREDICTION	803
<i>Wen-Chin Huang, Szu-Wei Fu, Erica Cooper, Ryandhimas E Zezario, Tomoki Toda, Hsin-Min Wang, Junichi Yamagishi, Yu Tsao</i>	
P4-27-SS05 PITCH-AND-SPECTRUM-AWARE SINGING QUALITY ASSESSMENT WITH BIAS CORRECTION AND MODEL FUSION	811
<i>Yu-Fei Shi, Yang Ai, Ye-Xin Lu, Hui-Peng Du, Zhen-Hua Ling</i>	
P4-28-SS05 THE T05 SYSTEM FOR THE VOICEMOS CHALLENGE 2024: TRANSFER LEARNING FROM DEEP IMAGE CLASSIFIER TO NATURALNESS MOS PREDICTION OF HIGH-QUALITY SYNTHETIC SPEECH	818
<i>Kaito Baba, Wataru Nakata, Yuki Saito, Hiroshi Saruwatari</i>	
Poster Session 5: Machine Learning & Resources	
P5-01-TLP AUTOMATED SPEAKING ASSESSMENT OF CONVERSATION TESTS WITH NOVEL GRAPH-BASED MODELING ON SPOKEN RESPONSE COHERENCE	825
<i>Jiun-Ting Li, Bi-Cheng Yan, Tien-Hong Lo, Yi-Cheng Wang, Yung-Chang Hsu, Berlin Chen</i>	
P5-02-TLP CONDITIONAL LABEL SMOOTHING FOR LLM-BASED DATA AUGMENTATION IN MEDICAL TEXT CLASSIFICATION	833
<i>Luca Becker, Philip Pracht, Peter Sertdal, Jil Uboreck, Alexander Bendel, Rainer Martin</i>	
P5-03-TLP PLAN, GENERATE AND OPTIMIZE: EXTENDING LARGE LANGUAGE MODELS FOR DIALOGUE SYSTEMS VIA PROMPT-BASED COLLABORATIVE METHOD	841
<i>Mengfei Guo, Si Chen, Yi Huang, Junlan Feng</i>	
P5-04-TLP TAMING NLU NOISE: STUDENT-TEACHER LEARNING FOR ROBUST DIALOGUE POLICY	849
<i>Mahdin Rohmatillah, Jen-Tzung Chien</i>	
P5-05-RES HEIGHTCELEB - AN ENRICHMENT OF VOXCELEB DATASET WITH SPEAKER HEIGHT INFORMATION	857
<i>Stanisław Kacprzak, Konrad Kowalczyk</i>	
P5-06-RES ESPNET-EZ: PYTHON-ONLY ESPNET FOR EASY FINE-TUNING AND INTEGRATION	863
<i>Masao Someki, Kwanghee Choi, Siddhant Arora, William Chen, Samuele Cornell, Jionghao Han, Yifan Peng, Jiatong Shi, Vaibhav Srivastav, Shinji Watanabe</i>	
P5-07-RES SPOKEN STEREOSET: ON EVALUATING SOCIAL BIAS TOWARD SPEAKER IN SPEECH LARGE LANGUAGE MODELS	871
<i>Yi-Cheng Lin, Wei-Chih Chen, Hung-yi Lee</i>	
P5-08-RES AMPHION: AN OPEN-SOURCE AUDIO, MUSIC, AND SPEECH GENERATION TOOLKIT	879
<i>Xueyao Zhang, Liumeng Xue, Yicheng Gu, Yuancheng Wang, Jiaqi Li, Haorui He, Chaoren</i>	

Wang, Songting Liu, Xi Chen, Junan Zhang, Zihao Fang, Haopeng Chen, Tze Ying Tang, Lexiao Zou, Mingxuan Wang, Jun Han, Kai Chen, Haizhou Li, Zhizheng Wu

P5-09-RES EMILIA: AN EXTENSIVE, MULTILINGUAL, AND DIVERSE SPEECH DATASET FOR LARGE-SCALE SPEECH GENERATION 885
Haorui He, Zengqiang Shang, Chaoren Wang, Xuyuan Li, Yicheng Gu, Hua Hua, Liwei Liu, Chen Yang, Jiaqi Li, Peiyang Shi, Yuancheng Wang, Kai Chen, Pengyuan Zhang, Zhizheng Wu

P5-10-RES FLORAS 50: A MASSIVELY MULTILINGUAL MULTITASK BENCHMARK FOR LONG-FORM CONVERSATIONAL SPEECH 891
William Chen, Brian Yan, Chih-Chen Chen, Shinji Watanabe

P5-11-RES MASSIVELY MULTILINGUAL FORCED ALIGNER LEVERAGING SELF-SUPERVISED DISCRETE UNITS 899
Hirofumi Inaguma, Ilia Kulikov, Zhaoheng Ni, Sravya Popuri, Paden Tomasello

P5-12-RES SPEECH RECOGNITION FOR ANALYSIS OF POLICE RADIO COMMUNICATION 906
Tejes Srivastava, Ju-Chieh Chou, Priyank Shroff, Karen Livescu, Christopher Graziul

P5-13-RES LARGE LANGUAGE MODELS AS USER-AGENTS FOR EVALUATING TASK-ORIENTED-DIALOGUE SYSTEMS 913
Taaha Kazi, Ruiliang Lyu, Sizhe Zhou, Dilek Hakkani-Tür, Gokhan Tur

P5-14-RES DFADD: THE DIFFUSION AND FLOW-MATCHING BASED AUDIO DEEPPFAKE DATASET 921
Jiawei Du, I-Ming Lin, I-Hsiang Chiu, Xuanjun Chen, Haibin Wu, Wenzhe Ren, Yu Tsao, Hung-yi Lee, Jyh-Shing Roger Jang

P5-15-RES SPMIS: AN INVESTIGATION OF SYNTHETIC SPOKEN MISINFORMATION DETECTION 929
Peizhuo Liu, Li Wang, Renqiang He, Haorui He, Lei Wang, Huadi Zheng, Jie Shi, Tong Xiao, Zhizheng Wu

P5-16-MLS SELF-SUPERVISED SPEECH MODELS FOR WORD-LEVEL STUTTERED SPEECH DETECTION 937
Yi-Jen Shih, Zoi Gkalitsiou, Alexandros G. Dimakis, David Harwath

P5-17-MLS ENHANCING AUTOMATIC SPEECH ASSESSMENT LEVERAGING HETEROGENEOUS FEATURES AND SOFT LABELS FOR ORDINAL CLASSIFICATION 945
Wen-Hsuan Peng, Sally Chen, Berlin Chen

P5-18-MLS SPEECH RECOGNITION-BASED FEATURE EXTRACTION FOR ENHANCED AUTOMATIC SEVERITY CLASSIFICATION IN DYSARTHIC SPEECH 953
Yerin Choi, Jeehyun Lee, Myoung-Wan Koo

P5-19-MLS EFFICIENT TRAINING OF SELF-SUPERVISED SPEECH FOUNDATION MODELS ON A COMPUTE BUDGET	961
<i>Andy T. Liu, Yi-Cheng Lin, Haibin Wu, Stefan Winkler, Hung-yi Lee</i>	
P5-20-MLS IMPROVING ANOMALOUS SOUND DETECTION VIA LOW-RANK ADAPTATION FINE-TUNING OF PRE-TRAINED AUDIO MODELS	969
<i>Xinhu Zheng, Anbai Jiang, Bing Han, Yanmin Qian, Pingyi Fan, Jia Liu, Wei-Qiang Zhang</i>	
P5-21-MLS EXPLORING ASR-BASED WAV2VEC2 FOR AUTOMATED SPEECH DISORDER ASSESSMENT: INSIGHTS AND ANALYSIS	975
<i>Tuan Nguyen, Corinne Fredouille, Alain Ghio, Mathieu Balaguer, Virginie Woisard</i>	
P5-22-MLS HIERARCHICAL MULTI-PATH AND MULTI-MODEL SELECTION FOR FAKE SPEECH DETECTION	983
<i>Chang Feng, Yiyang Zhao, Guangzhi Sun, Zehua Chen, Shuai Wang, Chao Zhang, Mingxing Xu, Thomas Fang Zheng</i>	
P5-23-MLS SEMI-SUPERVISED LEARNING FOR ROBUST SPEECH EVALUATION	991
<i>Huayun Zhang, Jeremy H. M. Wong, Geyu Lin, Nancy F. Chen</i>	
P5-24-MLS GE2E-KWS: GENERALIZED END-TO-END TRAINING AND EVALUATION FOR ZERO-SHOT KEYWORD SPOTTING	999
<i>Pai Zhu, Jacob W. Bartel, Dhruuv Agarwal, Kurt Partridge, Hyun Jin Park, Quan Wang</i>	
P5-25-MLS A SIMPLE HMM WITH SELF-SUPERVISED REPRESENTATIONS FOR PHONE SEGMENTATION	1007
<i>Gene-Ping Yang, Hao Tang</i>	
P5-26-MLS DASS: DISTILLED AUDIO STATE SPACE MODELS ARE STRONGER AND MORE DURATION-SCALABLE LEARNERS	1015
<i>Saurabhchand Bhati, Yuan Gong, Leonid Karlinsky, Hilde Kuehne, Rogerio Feris, James Glass</i>	
P5-27-MLS RAND: ROBUSTNESS AWARE NORM DECAY FOR QUANTIZED NEURAL NETWORKS	1023
<i>David Qiu, David Rim, Shaojin Ding, Oleg Rybakov, Yanzhang He</i>	
P5-28-MLS SWIM: SHORT-WINDOW CNN INTEGRATED WITH MAMBA FOR EEG-BASED AUDITORY SPATIAL ATTENTION DECODING	1031
<i>Ziyang Zhang, Andrew Thwaites, Alexandra Woolgar, Brian Moore, Chao Zhang</i>	

Poster Session 6: Spoken Language Processing

P6-01-SLP STUTTER-SOLVER: END-TO-END MULTI-LINGUAL DYSFLUENCY DETECTION	1039
<i>Xuanru Zhou, Cheol Jun Cho, Ayati Sharma, Brittany Morin, David Baquirin, Jet Vonk, Zoe Ezzes, Zachary Miller, Boon Lead Tee, Maria Luisa Gorno-Tempini, Jiachen Lian, Gopala Anumanchipalli</i>	

P6-02-SS09 DOMAIN ADAPTION AND UNIFIED KNOWLEDGE BASE MOTIVATE BETTER RETRIEVAL MODELS IN DIALOG SYSTEMS WITH RAG	1047
<i>Huadong Lin, Yirong Chen, Wenyu Tao, Mingyu Chen, Xiangmin Xu, Xiaofen Xing</i>	
P6-03-SLP SSAMBA: SELF-SUPERVISED AUDIO REPRESENTATION LEARNING WITH MAMBA STATE SPACE MODEL	1053
<i>Siavash Shams, Sukru Samet Dindar, Xilin Jiang, Nima Mesgarani</i>	
P6-04-SLP SPEECH-COPILOT: LEVERAGING LARGE LANGUAGE MODELS FOR SPEECH PROCESSING VIA TASK DECOMPOSITION, MODULARIZATION, AND PROGRAM GENERATION	1060
<i>Chun-Yi Kuan, Chih-Kai Yang, Wei-Ping Huang, Ke-Han Lu, Hung-yi Lee</i>	
P6-05-SLP CTC-GMM: CTC GUIDED MODALITY MATCHING FOR FAST AND ACCURATE STREAMING SPEECH TRANSLATION	1068
<i>Rui Zhao, Jinyu Li, Ruchao Fan, Matt Post</i>	
P6-06-SLP LONG-FORM END-TO-END SPEECH TRANSLATION VIA LATENT ALIGNMENT SEGMENTATION	1076
<i>Peter Polák, Ondřej Bojar</i>	
P6-07-SLP CONFIDENCE ESTIMATION FOR LLM-BASED DIALOGUE STATE TRACKING ..	1083
<i>Yi-Jyun Sun, Suvodip Dey, Dilek Hakkani-Tür, Gokhan Tur</i>	
P6-08-SLP THE 2ND FUTUREDIAL CHALLENGE: DIALOG SYSTEMS WITH RETRIEVAL AUGMENTED GENERATION (FUTUREDIAL-RAG)	1091
<i>Yucheng Cai, Si Chen, Yuxuan Wu, Yi Huang, Junlan Feng, Zhijian Ou</i>	
P6-09-SLP ZERO-SHOT AUDIO TOPIC RERANKING USING LARGE LANGUAGE MODELS	1099
<i>Mengjie Qian, Rao Ma, Adian Liusie, Erfan Loweimi, Kate M. Knill, Mark J.F. Gales</i>	
P6-10-SLP CLEAN LABEL ATTACKS AGAINST SLU SYSTEMS	1107
<i>Henry Li Xinyuan, Sonal Joshi, Thomas Thebaud, Jesus Villalba, Najim Dehak, Sanjeev Khudanpur</i>	
P6-11-SLP WHISMA: A SPEECH-LLM TO PERFORM ZERO-SHOT SPOKEN LANGUAGE UNDERSTANDING	1115
<i>Mohan Li, Cong-Thanh Do, Simon Keizer, Youmna Farag, Svetlana Stoyanchev, Rama Doddipatla</i>	
P6-12-SLP IMPROVING TRANSDUCER-BASED SPOKEN LANGUAGE UNDERSTANDING WITH SELF-CONDITIONED CTC AND KNOWLEDGE TRANSFER	1123
<i>Vishal Sunder, Eric Fosler-Lussier</i>	
P6-13-SLP SELF-SUPERVISED SYLLABLE DISCOVERY BASED ON SPEAKER-DISENTANGLED HUBERT	1131
<i>Ryota Komatsu, Takahiro Shinozaki</i>	

P6-14-SLP JUST ASR + LLM? A STUDY ON SPEECH LARGE LANGUAGE MODELS' ABILITY TO IDENTIFY AND UNDERSTAND SPEAKER IN SPOKEN DIALOGUE	1137
<i>Junkai Wu, Xulin Fan, Bo-Ru Lu, Xilin Jiang, Nima Mesgarani, Mark Hasegawa-Johnson, Mari Ostendorf</i>	
P6-15-SLR ENHANCING OPEN-SET SPEAKER IDENTIFICATION THROUGH RAPID TUNING WITH SPEAKER RECIPROCAL POINTS AND NEGATIVE SAMPLE	1144
<i>Zhiyong Chen, Zhiqi Ai, Xinnuo Li, Shugong Xu</i>	
P6-16-SLR SPOOFING-AWARE SPEAKER VERIFICATION ROBUST AGAINST DOMAIN AND CHANNEL MISMATCHES	1150
<i>Chang Zeng, Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi</i>	
P6-17-SLR ADVERSARIAL PURIFICATION FOR SPEAKER VERIFICATION BY TWO-STAGE DIFFUSION MODELS	1158
<i>Yibo Bai, Xiao-Lei Zhang, Xuelong Li</i>	
P6-18-SLR MEASURING SOUND SYMBOLISM IN AUDIO-VISUAL MODELS	1165
<i>Wei-Cheng Tseng, Yi-Jen Shih, David Harwath, Raymond Mooney</i>	
P6-19-SLR META-LEARNING APPROACHES FOR IMPROVING DETECTION OF UNSEEN SPEECH DEEPPAKES	1173
<i>Ivan Kukanov, Janne Laakkonen, Tomi Kinnunen, Ville Hautamäki</i>	
P6-20-SLR ON THE GENERATION AND REMOVAL OF SPEAKER ADVERSARIAL PERTURBATION FOR VOICE-PRIVACY PROTECTION	1179
<i>Chenyang Guo, Liping Chen, Zhuhai Li, Kong Aik Lee, Zhen-Hua Ling, Wu Guo</i>	
P6-21-SLR TOWARDS QUANTIFYING AND REDUCING LANGUAGE MISMATCH EFFECTS IN CROSS-LINGUAL SPEECH ANTI-SPOOFING	1185
<i>Tianchi Liu, Ivan Kukanov, Zihan Pan, Qiongqiong Wang, Hardik B Sailor, Kong Aik Lee</i>	
P6-22-SLR ENHANCING LOW-RESOURCE SPOKEN LANGUAGE IDENTIFICATION VIA CROSS-MODALITY RETRIEVAL AND CROSS-LINGUAL TEXT-TO-SPEECH SYNTHESIS	1193
<i>Min Ma, Gary Wang, Kyle Kastner, Isaac Caswell, Charles Yoon, Andrew Rosenberg</i>	
P6-23-SLR RECURSIVE ATTENTIVE POOLING FOR EXTRACTING SPEAKER EMBEDDINGS FROM MULTI-SPEAKER RECORDINGS	1201
<i>Shota Horiguchi, Atsushi Ando, Takafumi Moriya, Takanori Ashihara, Hiroshi Sato, Naohiro Tawara, Marc Delcroix</i>	
P6-24-SLR PDAF: A PHONETIC DEBIASING ATTENTION FRAMEWORK FOR SPEAKER VERIFICATION	1209
<i>Massa Baali, Abdulhamid Aldoobi, Hira Dharmyal, Rita Singh, Bhiksha Raj</i>	
P6-25-SLR INX-SPEAKERHUB: A 2000-HOUR INDIAN MULTILINGUAL SPEAKER VERIFICATION CORPUS	1217
<i>Metilda Sagaya Mary N J, S Umesh</i>	

P6-26-DIA RESOURCE-EFFICIENT ADAPTATION OF SPEECH FOUNDATION MODELS FOR MULTI-SPEAKER ASR	1224
<i>Weiqing Wang, Kunal Dhawan, Taejin Park, Krishna C. Puvvada, Ivan Medennikov, Somshubra Majumdar, He Huang, Jagadeesh Balam, Boris Ginsburg</i>	
P6-27-SS01 EXPLORING SELF-SUPERVISED REPRESENTATIONS FOR TEXT-DEPENDENT SPEAKER VERIFICATION	1232
<i>Sankala Sreekanth</i>	
P6-28-SS04 DISTILLATION-BASED FEATURE EXTRACTION ALGORITHM FOR SOURCE SPEAKER VERIFICATION	1240
<i>Xinlei Ma, Wenhuan Lu, Ruiteng Zhang, Junhai Xu, Xugang Lu, Jianguo Wei</i>	
P6-29-SS04 SPEAKER CONTRASTIVE LEARNING FOR SOURCE SPEAKER TRACING	1247
<i>Qing Wang, Hongmei Guo, Jian Kang, Mengjie Du, Jie Li, Xiao-Lei Zhang, Lei Xie</i>	
P6-30-SS04 THE DATABASE AND BENCHMARK FOR THE SOURCE SPEAKER TRACING CHALLENGE 2024	1254
<i>Ze Li, Yuke Lin, Tian Yao, Hongbin Suo, Pengyuan Zhang, Yanzhen Ren, Zexin Cai, Hiromitsu Nishizaki, Ming Li</i>	

Author Index