

25th Annual Conference of the International Speech Communication Association (INTERSPEECH 2024)

Kos, Greece
1-5 September 2024

Volume 1 of 7

ISBN: 979-8-3313-0506-2

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2024) by International Speech Communication Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact International Speech Communication Association
at the address below.

International Speech Communication Association
c/o Mme Emmanuelle FOXONET
4 Rue des Fauvettes - Lous Tourils
F-66390 Baixas, France

Phone: 49 228 735 643
Fax: 33 468 385 827

secretariat@isca-speech.org

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
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

VOLUME 1

KEYNOTE 1 ISCA MEDALLIST

Towards Responsible Speech Processing	1
<i>Isabel Trancoso</i>	

L2 SPEECH, BILINGUALISM AND CODE-SWITCHING

The Influence of L2 Accent Strength and Different Error Types on Personality Trait Ratings	2
<i>Sarah Wesolek, Piotr Gulowski, Joanna Blaszcak, Marzena Zygis</i>	
Characterizing Code-Switching: Applying Linguistic Principles for Metric Assessment and Development	7
<i>Jie Chi, Electra Wallington, Peter Bell</i>	
Towards a Better Understanding of Receptive Multilingualism: Listening Conditions and Priming Effects.....	12
<i>Wei Xue, Ivan Yuen, Bernd Möbius</i>	
2.5D Vocal Tract Modeling: Bridging Low-Dimensional Efficiency with 3D Accuracy	17
<i>Debashish Ray Mohapatra, Victor Zappi, Sidney Fels</i>	

SPEAKER DIARIZATION 1

Investigating Confidence Estimation Measures for Speaker Diarization	22
<i>Anurag Chowdhury, Abhinav Misra, Mark C. Fuhs, Monika Woszczyna</i>	
Speakers Unembedded: Embedding-Free Approach to Long-Form Neural Diarization	27
<i>Xiang Li, Vivek Govindan, Rohit Paturi, Sundararajan Srinivasan</i>	
On the Success and Limitations of Auxiliary Network Based Word-Level End-To-End Neural Speaker Diarization	32
<i>Yiling Huang, Weiran Wang, Guanlong Zhao, Hank Liao, Wei Xia, Quan Wang</i>	
EEND-M2F: Masked-Attention Mask Transformers for Speaker Diarization	37
<i>Marc Härkönen, Samuel J. Broughton, Lahiru Samarakoon</i>	
AFL-Net: Integrating Audio, Facial, and Lip Modalities with a Two-Step Cross-Attention for Robust Speaker Diarization in the Wild	42
<i>Yongkang Yin, Xu Li, Ying Shan, Yuexian Zou</i>	
Exploiting Wavelet Scattering Transform for an Unsupervised Speaker Diarization in Deep Neural Network Framework.....	47
<i>Arunav Arya, Murtiza Ali, Karan Nathwani</i>	

SPEECH AND AUDIO ANALYSIS AND REPRESENTATIONS

MINT: Boosting Audio-Language Model Via Multi-Target Pre-Training and Instruction Tuning.....	52
<i>Hang Zhao, Yifei Xin, Zhesong Yu, Bilei Zhu, Lu Lu, Zejun Ma</i>	
M2D-CLAP: Masked Modeling Duo Meets CLAP for Learning General-Purpose Audio-Language Representation	57
<i>Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, Noboru Harada, Masahiro Yasuda, Shunsuke Tsubaki, Keisuke Imoto</i>	
Audio Fingerprinting with Holographic Reduced Representations	62
<i>Yusuke Fujita, Tatsuya Komatsu</i>	
RAST: A Reference-Audio Synchronization Tool for Dubbed Content	67
<i>David Meyer, Eitan Abecassis, Clara Fernandez-Labrador, Christopher Schroers</i>	
YOLOPitch: A Time-Frequency Dual-Branch YOLO Model for Pitch Estimation	72
<i>Xuefei Li, Hao Huang, Ying Hu, Liang He, Jiabao Zhang, Yuyi Wang</i>	
Reduce, Reuse, Recycle: Is Perturbed Data Better than Other Language Augmentation for Low Resource Self-Supervised Speech Models	77
<i>Asad Ullah, Alessandro Ragano, Andrew Hines</i>	
AlignNet: Learning Dataset Score Alignment Functions to Enable Better Training of Speech Quality Estimators	82
<i>Jaden Pieper, Stephen Voran</i>	

ACOUSTIC EVENT DETECTION AND CLASSIFICATION 2

Improving Audio Classification with Low-Sampled Microphone Input: An Empirical Study Using Model Self-Distillation.....	87
<i>Dawei Liang, Alice Zhang, David Harwath, Edison Thomaz</i>	
MFF-EINV2: Multi-Scale Feature Fusion Across Spectral-Spatial-Temporal Domains for Sound Event Localization and Detection.....	92
<i>Da Mu, Zhicheng Zhang, Haobo Yue</i>	
Diversifying and Expanding Frequency-Adaptive Convolution Kernels for Sound Event Detection.....	97
<i>Hyeonuk Nam, Seong-Hu Kim, Deokki Min, Junhyeok Lee, Yong-Hwa Park</i>	
Stream-Based Active Learning for Anomalous Sound Detection in Machine Condition Monitoring.....	102
<i>Tuan Vu Ho, Kota Dohi, Yohei Kawaguchi</i>	
AnoPatch: Towards Better Consistency in Machine Anomalous Sound Detection.....	107
<i>Anbai Jiang, Bing Han, Zhiqiang Lv, Yufeng Deng, Wei-Qiang Zhang, Xie Chen, Yanmin Qian, Jia Liu, Pingyi Fan</i>	
FakeSound: Deepfake General Audio Detection	112
<i>Zeyu Xie, Baihan Li, Xuenan Xu, Zheng Liang, Kai Yu, Mengyue Wu</i>	
Sound of Traffic: A Dataset for Acoustic Traffic Identification and Counting	117
<i>Shabnam Ghaffarzadegan, Luca Bondi, Wei-Chang Lin, Abinaya Kumar, Ho-Hsiang Wu, Hans-Georg Horst, Samarjit Das</i>	

DETECTION AND CLASSIFICATION OF BIOACOUSTIC SIGNALS

Vision Transformer Segmentation for Visual Bird Sound Denoising	122
<i>Sahil Kumar, Jialu Li, Youshan Zhang</i>	
DB3V: A Dialect Dominated Dataset of Bird Vocalisation for Cross-Corpus Bird Species Recognition	127
<i>Xin Jing, Luyang Zhang, Jiangjian Xie, Alexander Gebhard, Alice Baird, Björn Schuller</i>	
Investigating Self-Supervised Speech Models' Ability to Classify Animal Vocalizations: The Case of Gibbon's Vocal Signatures.....	132
<i>Jules Cauzinille, Benoît Favre, Ricard Marxer, Dena Clink, Abdul Hamid Ahmad, Arnaud Rey</i>	
Study Selectively: An Adaptive Knowledge Distillation Based on a Voting Network for Heart Sound Classification.....	137
<i>Xihang Qiu, Lixian Zhu, Zikai Song, Zeyu Chen, Haojie Zhang, Kun Qian, Ye Zhang, Bin Hu, Yoshiharu Yamamoto, Björn W. Schuller</i>	
SimuSOE: A Simulated Snoring Dataset for Obstructive Sleep Apnea-Hypopnea Syndrome Evaluation During Wakefulness	142
<i>Jie Lin, Xiuping Yang, Li Xiao, Xinhong Li, Weiyan Yi, Yuhong Yang, Weiping Tu, Xiong Chen</i>	

ACOUSTIC ECHO CANCELLATION

Multi-Mic Echo Cancellation Coalesced with Beamforming for Real World Adverse Acoustic Conditions	147
<i>Premanand Nayak, Kamini Sabu, M. Ali Basha Shaik</i>	
Interference Aware Training Target for DNN Based Joint Acoustic Echo Cancellation and Noise Suppression	152
<i>Vahid Khanagha, Dimitris Koutsaidis, Kaustubh Kalgaonkar, Sriram Srinivasan</i>	
Low Complexity Echo Delay Estimator Based on Binarized Feature Matching.....	157
<i>Yi Gao, Xiang Su</i>	
MSA-DPCRN: A Multi-Scale Asymmetric Dual-Path Convolution Recurrent Network with Attentional Feature Fusion for Acoustic Echo Cancellation.....	162
<i>Ye Ni, Cong Pang, Chengwei Huang, Cairong Zou</i>	
Efficient Joint Beamforming and Acoustic Echo Cancellation Structure for Conference Call Scenarios	167
<i>Ofer Schwartz, Sharon Gannot</i>	
SDAEC: Signal Decoupling for Advancing Acoustic Echo Cancellation	172
<i>Fei Zhao, Jinjiang Liu, Xueliang Zhang</i>	

SPEECH SYNTHESIS: VOICE CONVERSION 1

Spatial Voice Conversion: Voice Conversion Preserving Spatial Information and Non-Target Signals	177
<i>Kentaro Seki, Shinnosuke Takamichi, Norihiro Takamune, Yuki Saito, Kanami Imamura, Hiroshi Saruwatari</i>	

Neural Codec Language Models for Disentangled and Textless Voice Conversion	182
<i>Alan Baade, Puyuan Peng, David Harwath</i>	
Fine-Grained and Interpretable Neural Speech Editing.....	187
<i>Max Morrison, Cameron Churchwell, Nathan Pruyne, Bryan Pardo</i>	
FastVoiceGrad: One-Step Diffusion-Based Voice Conversion with Adversarial Conditional Diffusion Distillation	192
<i>Takuhiko Kaneko, Hirokazu Kameoka, Kou Tanaka, Yuto Kondo</i>	
DualVC 3: Leveraging Language Model Generated Pseudo Context for End-To-End Low Latency Streaming Voice Conversion	197
<i>Ziqian Ning, Shuai Wang, Pengcheng Zhu, Zhichao Wang, Jixun Yao, Lei Xie, Mengxiao Bi</i>	
Towards Realistic Emotional Voice Conversion Using Controllable Emotional Intensity	202
<i>Tianhua Qi, Shiyang Wang, Cheng Lu, Yan Zhao, Yuan Zong, Wenming Zheng</i>	

NEURAL NETWORK ARCHITECTURES FOR ASR 2

InterBiasing: Boost Unseen Word Recognition Through Biasing Intermediate Predictions	207
<i>Yu Nakagome, Michael Hentschel</i>	
SEQ-Former: A Context-Enhanced and Efficient Automatic Speech Recognition Framework.....	212
<i>Qinglin Meng, Min Liu, Kaixun Huang, Kun Wei, Lei Xie, Zongfeng Quan, Weihong Deng, Quan Lu, Ning Jiang, Guoqing Zhao</i>	
How Much Context Does My Attention-Based ASR System Need?.....	217
<i>Robert Flynn, Anton Ragni</i>	
Rich Speech Signal: Exploring and Exploiting End-To-End Automatic Speech Recognizers' Ability to Model Hesitation Phenomena.....	222
<i>Vincenzo Norman Vitale, Loredana Schettino, Francesco Cutugno</i>	
Transmitted and Aggregated Self-Attention for Automatic Speech Recognition	227
<i>Tian-Hao Zhang, Xinyuan Qian, Feng Chen, Xu-Cheng Yin</i>	
MULTI-CONVFORMER: Extending Conformer with Multiple Convolution Kernels	232
<i>Darshan Prabhu, Yifan Peng, Preethi Jyothi, Shinji Watanabe</i>	
Exploring the Capability of Mamba in Speech Applications.....	237
<i>Koichi Miyazaki, Yoshiki Masuyama, Masato Murata</i>	
Robust Voice Activity Detection Using Locality-Sensitive Hashing and Residual Frequency- Temporal Attention	242
<i>Shu Li, Peng Zhang, Ye Li</i>	
Lightweight Transducer Based on Frame-Level Criterion	247
<i>Genshun Wan, Mengzhi Wang, Tingzhi Mao, Hang Chen, Zhongfu Ye</i>	
Exploring the Limits of Decoder-Only Models Trained on Public Speech Recognition Corpora.....	252
<i>Ankit Gupta, George Saon, Brian Kingsbury</i>	
Contextual Biasing Speech Recognition in Speech-Enhanced Large Language Model.....	257
<i>Xun Gong, Anqi Lv, Zhiming Wang, Yanmin Qian</i>	

DECODING ALGORITHMS

Towards Effective and Efficient Non-Autoregressive Decoding Using Block-Based Attention Mask	262
<i>Tianzi Wang, Xurong Xie, Zhaoqing Li, Shoukang Hu, Zengrui Jin, Jiajun Deng, Mingyu Cui, Shujie Hu, Mengzhe Geng, Guinan Li, Helen Meng, Xunying Liu</i>	
E-Paraformer: A Faster and Better Parallel Transformer for Non-Autoregressive End-To-End Mandarin Speech Recognition.....	267
<i>Kun Zou, Fengyun Tan, Ziyang Zhuang, Chenfeng Miao, Tao Wei, Shaodan Zhai, Zijian Li, Wei Hu, Shaojun Wang, Jing Xiao</i>	
Beam-Search SIEVE for Low-Memory Speech Recognition	272
<i>Martino Ciaperoni, Athanasios Katsamanis, Aristides Gionis, Panagiotis Karras</i>	
Speed of Light Exact Greedy Decoding for RNN-T Speech Recognition Models on GPU	277
<i>Daniel Galvez, Vladimir Bataev, Hainan Xu, Tim Kaldewey</i>	
Contextual Biasing with the Knuth-Morris-Pratt Matching Algorithm	282
<i>Weiran Wang, Zelin Wu, Diamantino Caseiro, Tsendsuren Munkhdalai, Khe Chai Sim, Pat Rondon, Golan Pundak, Gan Song, Rohit Prabhavalkar, Zhong Meng, Ding Zhao, Tara Sainath, Yanzhang He, Pedro Moreno Mengibar</i>	
Text-Only Domain Adaptation for CTC-Based Speech Recognition Through Substitution of Implicit Linguistic Information in the Search Space	287
<i>Tatsunari Takagi, Yukoh Wakabayashi, Atsunori Ogawa, Norihide Kitaoka</i>	

PRONUNCIATION ASSESSMENT

Pitch-Aware RNN-T for Mandarin Chinese Mispronunciation Detection and Diagnosis	292
<i>Xintong Wang, Mingqian Shi, Ye Wang</i>	
MultiPA: A Multi-Task Speech Pronunciation Assessment Model for Open Response Scenarios	297
<i>Yu-Wen Chen, Zhou Yu, Julia Hirschberg</i>	
A Framework for Phoneme-Level Pronunciation Assessment Using CTC	302
<i>Xinwei Cao, Zijian Fan, Torbjørn Svendsen, Giampiero Salvi</i>	
Phonological-Level Mispronunciation Detection and Diagnosis	307
<i>Mostafa Shahin, Beena Ahmed</i>	
Acoustic Feature Mixup for Balanced Multi-Aspect Pronunciation Assessment.....	312
<i>Heejin Do, Wonjun Lee, Gary Geunbae Lee</i>	
Automated Content Assessment and Feedback for Finnish L2 Learners in a Picture Description Speaking Task.....	317
<i>Nhan Phan, Anna Von Zansen, Maria Kautonen, Ekaterina Voskoboinik, Tamas Grosz, Raili Hilden, Mikko Kurimo</i>	

SPOKEN LANGUAGE PROCESSING

Query-By-Example Keyword Spotting Using Spectral-Temporal Graph Attentive Pooling and Multi-Task Learning	322
<i>Zhenyu Wang, Shuyu Kong, Li Wan, Biqiao Zhang, Yiteng Huang, Mumin Jin, Ming Sun, Xin Lei, Zhaojun Yang</i>	

Relational Proxy Loss for Audio-Text Based Keyword Spotting	327
<i>Youngmoon Jung, Seungjin Lee, Joon-Young Yang, Jaeyoung Roh, Chang Woo Han, Hoon-Young Cho</i>	
CTC-Aligned Audio-Text Embedding for Streaming Open-Vocabulary Keyword Spotting.....	332
<i>Sichen Jin, Youngmoon Jung, Seungjin Lee, Jaeyoung Roh, Changwoo Han, Hoonyoung Cho</i>	
Text-Aware Speech Separation for Multi-Talker Keyword Spotting.....	337
<i>Haoyu Li, Baochen Yang, Yu Xi, Linfeng Yu, Tian Tan, Hao Li, Kai Yu</i>	
Language-Universal Speech Attributes Modeling for Zero-Shot Multilingual Spoken Keyword Recognition	342
<i>Hao Yen, Pin-Jui Ku, Sabato Marco Siniscalchi, Chin-Hui Lee</i>	
Adding User Feedback to Enhance CB-Whisper	347
<i>Raul Monteiro</i>	
OWSM V3.1: Better and Faster Open Whisper-Style Speech Models Based on E-Branchformer	352
<i>Yifan Peng, Jinchuan Tian, William Chen, Siddhant Arora, Brian Yan, Yui Sudo, Muhammad Shakeel, Kwanghee Choi, Jiatong Shi, Xuankai Chang, Jee-Weon Jung, Shinji Watanabe</i>	

SPOKEN MACHINE TRANSLATION 2

Parameter-Efficient Adapter Based on Pre-Trained Models for Speech Translation.....	357
<i>Nan Chen, Yonghe Wang, Feilong Bao</i>	
Wave to Interlingua: Analyzing Representations of Multilingual Speech Transformers for Spoken Language Translation	362
<i>Badr M. Abdullah, Mohammed Maqsood Shaik, Dietrich Klakow</i>	
Knowledge-Preserving Pluggable Modules for Multilingual Speech Translation Tasks.....	367
<i>Nan Chen, Yonghe Wang, Feilong Bao</i>	
Navigating the Minefield of MT Beam Search in Cascaded Streaming Speech Translation.....	372
<i>Rastislav Rabatin, Frank Seide, Ernie Chang</i>	
Soft Language Identification for Language-Agnostic Many-To-One End-To-End Speech Translation.....	377
<i>Peidong Wang, Jian Xue, Jinyu Li, Junkun Chen, Aswin Shanmugam Subramanian</i>	
A Unit-Based System and Dataset for Expressive Direct Speech-To-Speech Translation	382
<i>Anna Min, Chenxu Hu, Yi Ren, Hang Zhao</i>	
Translating Speech with Just Images.....	387
<i>Dan Oneata, Herman Kamper</i>	
ZeroST: Zero-Shot Speech Translation	392
<i>Sameer Khurana, Chiori Hori, Antoine Laurent, Gordon Wichern, Jonathan Le Roux</i>	

BIOSIGNAL-ENABLED SPOKEN COMMUNICATION

A Multimodal Approach to Study the Nature of Coordinative Patterns Underlying Speech Rhythm.....	397
<i>Jinyu Li, Leonardo Lancia</i>	

Towards EMG-To-Speech with Necklace Form Factor.....	402
<i>Peter Wu, Ryan Kaveh, Raghav Nautiyal, Christine Zhang, Albert Guo, Anvitha Kachinthaya, Tavish Mishra, Bohan Yu, Alan W Black, Rikky Muller, Gopala Krishna Anumanchipalli</i>	
Using Articulated Speech EEG Signals for Imagined Speech Decoding	407
<i>Chris Bras, Tanvina Patel, Odette Scharenborg</i>	
Direct Speech Synthesis from Non-Invasive, Neuromagnetic Signals.....	412
<i>Jinuk Kwon, David Harwath, Debadatta Dash, Paul Ferrari, Jun Wang</i>	
Optical Flow Guided Tongue Trajectory Generation for Diffusion-Based Acoustic to Articulatory Inversion.....	417
<i>Yudong Yang, Rongfeng Su, Rukiye Ruzi, Manwa Ng, Shaofeng Zhao, Nan Yan, Lan Wang</i>	
Multimodal Segmentation for Vocal Tract Modeling	422
<i>Rishi Jain, Bohan Yu, Peter Wu, Tejas Prabhune, Gopala Anumanchipalli</i>	
Articulatory Synthesis Using Representations Learnt Through Phonetic Label-Aware Contrastive Loss	427
<i>Jesuraj Bandekar, Sathvik Udupa, Prasanta Kumar Ghosh</i>	
Auditory Attention Decoding in Four-Talker Environment with EEG.....	432
<i>Yujie Yan, Xiran Xu, Haolin Zhu, Pei Tian, Zhongshu Ge, Xihong Wu, Jing Chen</i>	
ASA: An Auditory Spatial Attention Dataset with Multiple Speaking Locations	437
<i>Zijie Lin, Tianyu He, Siqu Cai, Haizhou Li</i>	
Leveraging Graphic and Convolutional Neural Networks for Auditory Attention Detection with EEG	442
<i>Saurav Pahuja, Gabriel Ivucic, Pascal Himmelmann, Siqu Cai, Tanja Schultz, Haizhou Li</i>	

INDIVIDUAL AND SOCIAL FACTORS IN PHONETICS

Echoes of Implicit Bias Exploring Aesthetics and Social Meanings of Swiss German Dialect Features	447
<i>Tillmann Pistor, Adrian Leemann</i>	
In Search of Structure and Correspondence in Intra-Speaker Trial-To-Trial Variability	452
<i>Vivian G. Li</i>	
Modelled Multivariate Overlap: A Method for Measuring Vowel Merger	457
<i>Irene Smith, Morgan Sonderegger, The Spade Consortium</i>	
Entrainment Analysis and Prosody Prediction of Subsequent Interlocutor's Backchannels in Dialogue	462
<i>Keiko Ochi, Koji Inoue, Divesh Lala, Tatsuya Kawahara</i>	
Exploring the Anatomy of Articulation Rate in Spontaneous English Speech: Relationships Between Utterance Length Effects and Social Factors.....	467
<i>James Tanner, Morgan Sonderegger, Jane Stuart-Smith, Tyler Kendall, Jeff Mielke, Robin Dodsworth, Erik Thomas</i>	
Familiar and Unfamiliar Speaker Identification in Speech and Singing	472
<i>Katelyn Taylor, Amelia Gully, Helena Daffern</i>	

PARALINGUISTICS

Cross-Transfer Knowledge Between Speech and Text Encoders to Evaluate Customer Satisfaction	477
<i>Luis Felipe Parra-Gallego, Tilak Purohit, Bogdan Vlasenko, Juan Rafael Orozco-Arroyave, Mathew Magimai.-Doss</i>	
Fine-Tuning of Pre-Trained Models for Classification of Vocal Intensity Category from Speech Signals	482
<i>Manila Kodali, Sudarsana Reddy Kadiri, Paavo Alku</i>	
Real-World PTSD Recognition: A Cross-Corpus and Cross-Linguistic Evaluation.....	487
<i>Alexander Kathan, Martin Bürger, Andreas Triantafyllopoulos, Sabrina Milkus, Jonas Hohmann, Pauline Muderlak, Jürgen Schottdorf, Richard Musil, Björn Schuller, Shahin Amiriparian</i>	
Switching Tongues, Sharing Hearts: Identifying the Relationship Between Empathy and Code-Switching in Speech	492
<i>Debasmita Bhattacharya, Eleanor Lin, Run Chen, Julia Hirschberg</i>	

SPEAKER RECOGNITION: ADVERSARIAL AND SPOOFING ATTACKS

Anti-Spoofing Ensembling Model: Dynamic Weight Allocation in Ensemble Models for Improved Voice Biometrics Security	497
<i>Eros Rosello, Angel M. Gomez, Iván López-Espejo, Antonio M. Peinado, Juan M. Martín-Doñas</i>	
Spoof Diarization: "What Spoofed When" in Partially Spoofed Audio	502
<i>Lin Zhang, Xin Wang, Erica Cooper, Mireia Diez, Federico Landini, Nicholas Evans, Junichi Yamagishi</i>	
Spoofing Speech Detection by Modeling Local Spectro-Temporal and Long-Term Dependency	507
<i>Haochen Wu, Wu Guo, Zhentao Zhang, Wenting Zhao, Shengyu Peng, Jie Zhang</i>	
Improving Copy-Synthesis Anti-Spoofing Training Method with Rhythm and Speaker Perturbation.....	512
<i>Jingze Lu, Yuxiang Zhang, Zhuo Li, Zengqiang Shang, Wenchao Wang, Pengyuan Zhang</i>	
VoiceDefense: Protecting Automatic Speaker Verification Models Against Black-Box Adversarial Attacks.....	517
<i>Yip Keng Kan, Ke Xu, Hao Li, Jie Shi</i>	
Neural Codec-Based Adversarial Sample Detection for Speaker Verification	522
<i>Xuanjun Chen, Jiawei Du, Haibin Wu, Jyh-Shing Roger Jang, Hung-Yi Lee</i>	
Textual-Driven Adversarial Purification for Speaker Verification.....	527
<i>Sizhou Chen, Yibo Bai, Jiadi Yao, Xiao-Lei Zhang, Xuelong Li</i>	
Boosting the Transferability of Adversarial Examples with Gradient-Aligned Ensemble Attack for Speaker Recognition.....	532
<i>Zhuhai Li, Jie Zhang, Wu Guo, Haochen Wu</i>	
Temporal-Channel Modeling in Multi-Head Self-Attention for Synthetic Speech Detection	537
<i>Duc-Tuan Truong, Ruijie Tao, Tuan Nguyen, Hieu-Thi Luong, Kong Aik Lee, Eng Siong Chng</i>	

AUDIO EVENT DETECTION AND CLASSIFICATION 1

Can Synthetic Audio from Generative Foundation Models Assist Audio Recognition and Speech Modeling?.....	542
<i>Tiantian Feng, Dimitrios Dimitriadis, Shrikanth S. Narayanan</i>	
Scaling Up Masked Audio Encoder Learning for General Audio Classification.....	547
<i>Heinrich Dinkel, Zhiyong Yan, Yongqing Wang, Junbo Zhang, Yujun Wang, Bin Wang</i>	
Audio Mamba: Selective State Spaces for Self-Supervised Audio Representations	552
<i>Sarthak Yadav, Zheng-Hua Tan</i>	
MAT-SED: A Masked Audio Transformer with Masked-Reconstruction Based Pre-Training for Sound Event Detection	557
<i>Pengfei Cai, Yan Song, Kang Li, Haoyu Song, Ian McLoughlin</i>	
Sound Event Bounding Boxes	562
<i>Janek Ebbers, François G. Germain, Gordon Wichern, Jonathan Le Roux</i>	
Low-Complexity Acoustic Scene Classification Using Parallel Attention-Convolution Network	567
<i>Yanxiong Li, Jiaxin Tan, Guoqing Chen, Jialong Li, Yongjie Si, Qianhua He</i>	

SOURCE SEPARATION 2

Towards Explainable Monaural Speaker Separation with Auditory-Based Training	572
<i>Hassan Taherian, Vahid Ahmadi Kalkhorani, Ashutosh Pandey, Daniel Wong, Buye Xu, Deliang Wang</i>	
Does the Lombard Effect Matter in Speech Separation? Introducing the Lombard-GRID-2mix Dataset.....	577
<i>Iva Ewert, Marvin Borsdorf, Haizhou Li, Tanja Schultz</i>	
PARIS: Pseudo-AutoRegressive Siamese Training for Online Speech Separation.....	582
<i>Zexu Pan, Gordon Wichern, François G. Germain, Kohei Saijo, Jonathan Le Roux</i>	
OR-TSE: An Overlap-Robust Speaker Encoder for Target Speech Extraction	587
<i>Yiru Zhang, Linyu Yao, Qun Yang</i>	
Multimodal Representation Loss Between Timed Text and Audio for Regularized Speech Separation.....	592
<i>Tsun-An Hsieh, Heeyoul Choi, Minje Kim</i>	
SA-WavLM: Speaker-Aware Self-Supervised Pre-Training for Mixture Speech.....	597
<i>Jingru Lin, Meng Ge, Junyi Ao, Liqun Deng, Haizhou Li</i>	
TSE-PI: Target Sound Extraction Under Reverberant Environments with Pitch Information.....	602
<i>Yiwen Wang, Xihong Wu</i>	
Enhanced Reverberation as Supervision for Unsupervised Speech Separation.....	607
<i>Kohei Saijo, Gordon Wichern, François G. Germain, Zexu Pan, Jonathan Le Roux</i>	

NOISE REDUCTION, DEREVERBERATION, AND ECHO CANCELLATION

Deep Echo Path Modeling for Acoustic Echo Cancellation	612
<i>Fei Zhao, Chenggang Zhang, Shulin He, Jinjiang Liu, Xueliang Zhang</i>	
Graph Attention Based Multi-Channel U-Net for Speech Dereverberation with Ad-Hoc Microphone Arrays.....	617
<i>Hongmei Guo, Yijiang Chen, Xiao-Lei Zhang, Xuelong Li</i>	
Speech Dereverberation Constrained on Room Impulse Response Characteristics	622
<i>Louis Bahrman, Mathieu Fontaine, Jonathan Le Roux, Gaël Richard</i>	
DeWinder: Single-Channel Wind Noise Reduction Using Ultrasound Sensing.....	627
<i>Kuang Yuan, Shuo Han, Swarun Kumar, Bhiksha Raj</i>	
ANIMAL-CLEAN – a Deep Denoising Toolkit for Animal-Independent Signal Enhancement.....	632
<i>Alexander Barnhill, Elmar Noeth, Andreas Maier, Christian Bergler</i>	
Elucidating Clock-Drift Using Real-World Audios in Wireless Mode for Time-Offset Insensitive End-To-End Asynchronous Acoustic Echo Cancellation	637
<i>Premanand Nayak, M. Ali Basha Shaik</i>	
QMIXCAT: Unsupervised Speech Enhancement Using Quality-Guided Signal Mixing and Competitive Alternating Model Training.....	642
<i>Shilin Wang, Haixin Guan, Yanhua Long</i>	

VOLUME 2

COMPUTATIONALLY-EFFICIENT SPEECH ENHANCEMENT

Speech Boosting: Low-Latency Live Speech Enhancement for TWS Earbuds	647
<i>Hanbin Bae, Pavel Andreev, Azat Saginbaev, Nicholas Babaev, Wonjun Lee, Hosang Sung, Hoon-Young Cho</i>	
Knowledge Distillation for Tiny Speech Enhancement with Latent Feature Augmentation	652
<i>Behnam Gholami, Mostafa El-Khamy, Keebong Song</i>	
Sub-PNWR: Speech Enhancement Based on Signal Sub-Band Splitting and Pseudo Noisy Waveform Reconstruction Loss.....	657
<i>Yuewei Zhang, Huanbin Zou, Jie Zhu</i>	
Streamlining Speech Enhancement DNNs: An Automated Pruning Method Based on Dependency Graph with Advanced Regularized Loss Strategies.....	662
<i>Zugang Zhao, Jinghong Zhang, Yonghui Liu, Jianbing Liu, Kai Niu, Zhiqiang He</i>	
Lightweight Dynamic Sparse Transformer for Monaural Speech Enhancement.....	667
<i>Zehua Zhang, Xuyi Zhuang, Yukun Qian, Mingjiang Wang</i>	
MUSE: Flexible Voiceprint Receptive Fields and Multi-Path Fusion Enhanced Taylor Transformer for U-Net-Based Speech Enhancement	672
<i>Zizhen Lin, Xiaoting Chen, Junyu Wang</i>	
Dynamic Gated Recurrent Neural Network for Compute-Efficient Speech Enhancement	677
<i>Longbiao Cheng, Ashutosh Pandey, Buye Xu, Tobi Delbruck, Shih-Chii Liu</i>	

ZERO-SHOT TTS

Improving Audio Codec-Based Zero-Shot Text-To-Speech Synthesis with Multi-Modal Context and Large Language Model.....	682
<i>Jinlong Xue, Yayue Deng, Yicheng Han, Yingming Gao, Ya Li</i>	
An Investigation of Noise Robustness for Flow-Matching-Based Zero-Shot TTS	687
<i>Xiaofei Wang, Sefik Emre Eskimez, Manthan Thakker, Hemin Yang, Zirun Zhu, Min Tang, Yufei Xia, Jinzhu Li, Sheng Zhao, Jinyu Li, Naoyuki Kanda</i>	
Lightweight Zero-Shot Text-To-Speech with Mixture of Adapters.....	692
<i>Kenichi Fujita, Takanori Ashihara, Marc Delcroix, Yusuke Ijima</i>	
DINO-VITS: Data-Efficient Zero-Shot TTS with Self-Supervised Speaker Verification Loss for Noise Robustness	697
<i>Vikentii Pankov, Valeria Pronina, Alexander Kuzmin, Maksim Borisov, Nikita Usoltsev, Xingshan Zeng, Alexander Golubkov, Nikolai Ermolenko, Aleksandra Shirshova, Yulia Matveeva</i>	

NOISE ROBUSTNESS, FAR-FIELD, AND MULTI-TALKER ASR

LibriheavyMix: A 20,000-Hour Dataset for Single-Channel Reverberant Multi-Talker Speech Separation, ASR and Speaker Diarization.....	702
<i>Zengrui Jin, Yifan Yang, Mohan Shi, Wei Kang, Xiaoyu Yang, Zengwei Yao, Fangjun Kuang, Liyong Guo, Lingwei Meng, Long Lin, Yong Xu, Shi-Xiong Zhang, Daniel Povey</i>	
A Joint Noise Disentanglement and Adversarial Training Framework for Robust Speaker Verification	707
<i>Xujiang Xing, Mingxing Xu, Thomas Fang Zheng</i>	
Serialized Output Training by Learned Dominance	712
<i>Ying Shi, Lantian Li, Shi Yin, Dong Wang, Jiqing Han</i>	
SOT Triggered Neural Clustering for Speaker Attributed ASR.....	717
<i>Xianrui Zheng, Guangzhi Sun, Chao Zhang, Philip C. Woodland</i>	
Neural Blind Source Separation and Diarization for Distant Speech Recognition.....	722
<i>Yoshiaki Bando, Tomohiko Nakamura, Shinji Watanabe</i>	
Unified Multi-Talker ASR with and Without Target-Speaker Enrollment.....	727
<i>Ryo Masumura, Naoki Makishima, Tomohiro Tanaka, Mana Ihori, Naotaka Kawata, Shota Orihashi, Kazutoshi Shinoda, Taiga Yamane, Saki Mizuno, Keita Suzuki, Satoshi Suzuki, Nobukatsu Hojo, Takafumi Moriya, Atsushi Ando</i>	

CONTEXTUAL BIASING AND ADAPTATION

Keyword-Guided Adaptation of Automatic Speech Recognition	732
<i>Aviv Shamsian, Aviv Navon, Neta Glazer, Gill Hetz, Joseph Keshet</i>	
Improving Speech Recognition with Prompt-Based Contextualized ASR and LLM-Based Re-Predictor	737
<i>Nguyen Manh Tien Anh, Thach Ho Sy</i>	

Incorporating Class-Based Language Model for Named Entity Recognition in Factorized Neural Transducer	742
<i>Peng Wang, Yifan Yang, Zheng Liang, Tian Tan, Shiliang Zhang, Xie Chen</i>	
Contextual Biasing with Confidence-Based Homophone Detector for Mandarin End-To-End Speech Recognition	747
<i>Chengxu Yang, Lin Zheng, Sanli Tian, Gaofeng Cheng, Sujie Xiao, Ta Li</i>	
Improving Neural Biasing for Contextual Speech Recognition by Early Context Injection and Text Perturbation	752
<i>Ruizhe Huang, Mahsa Yarmohammadi, Sanjeev Khudanpur, Daniel Povey</i>	
Fast Context-Biasing for CTC and Transducer ASR Models with CTC-Based Word Spotter	757
<i>Andrei Andrusenko, Aleksandr Laptev, Vladimir Bataev, Vitaly Lavrukhin, Boris Ginsburg</i>	
Prompt Tuning for Speech Recognition on Unknown Spoken Name Entities	762
<i>Xizi Wei, Stephen McGregor</i>	
Improved Factorized Neural Transducer Model for Text-Only Domain Adaptation	767
<i>Junzhe Liu, Jianwei Yu, Xie Chen</i>	
Modality Translation Learning for Joint Speech-Text Model	772
<i>Pin-Yen Liu, Jen-Tzung Chien</i>	
SAML: Speaker Adaptive Mixture of LoRA Experts for End-To-End ASR	777
<i>Qiuming Zhao, Guangzhi Sun, Chao Zhang, Mingxing Xu, Thomas Fang Zheng</i>	
Factor-Conditioned Speaking-Style Captioning	782
<i>Atsushi Ando, Takafumi Moriya, Shota Horiguchi, Ryo Masumura</i>	
Dual-Pipeline with Low-Rank Adaptation for New Language Integration in Multilingual ASR	787
<i>Yerbolat Khassanov, Zhipeng Chen, Tianfeng Chen, Tze Yuang Chong, Wei Li, Jun Zhang, Lu Lu, Yuxuan Wang</i>	
Speculative Speech Recognition by Audio-Prefixed Low-Rank Adaptation of Language Models	792
<i>Bolaji Yusuf, Murali Karthick Baskar, Andrew Rosenberg, Bhuvana Ramabhadran</i>	
Domain-Aware Data Selection for Speech Classification Via Meta-Reweighting	797
<i>Junghun Kim, Ka Hyun Park, Hoyoung Yoon, U Kang</i>	

SPOKEN LANGUAGE UNDERSTANDING

Finding Task-Specific Subnetworks in Multi-Task Spoken Language Understanding Model	802
<i>Hayato Futami, Siddhant Arora, Yosuke Kashiwagi, Emiru Tsunoo, Shinji Watanabe</i>	
Out-Of-Distribution Generalisation in Spoken Language Understanding	807
<i>Dejan Porjazovski, Anssi Moisio, Mikko Kurimo</i>	
A Dual Task Learning Approach to Fine-Tune a Multilingual Semantic Speech Encoder for Spoken Language Understanding	812
<i>Gaëlle Laperrière, Sahar Ghannay, Bassam Jabaian, Yannick Estève</i>	
Speech-MASSIVE: A Multilingual Speech Dataset for SLU and Beyond	817
<i>Beomseok Lee, Ioan Calapodescu, Marco Gaido, Matteo Negri, Laurent Besacier</i>	

Using Large Language Model for End-To-End Chinese ASR and NER	822
<i>Yuang Li, Jiawei Yu, Min Zhang, Mengxin Ren, Yanqing Zhao, Xiaofeng Zhao, Shimin Tao, Jinsong Su, Hao Yang</i>	

A Contrastive Learning Approach to Mitigate Bias in Speech Models	827
<i>Alkis Koudounas, Flavio Giobergia, Eliana Pastor, Elena Baralis</i>	

SPOKEN MACHINE TRANSLATION 1

Investigating Decoder-Only Large Language Models for Speech-To-Text Translation	832
<i>Chao-Wei Huang, Hui Lu, Hongyu Gong, Hirofumi Inaguma, Ilia Kulikov, Ruslan Mavlyutov, Sravya Popuri</i>	

Diffusion Synthesizer for Efficient Multilingual Speech to Speech Translation	837
<i>Nameer Hirschkind, Xiao Yu, Mahesh Kumar Nandwana, Joseph Liu, Eloi Dubois, Dao Le, Nicolas Thiebaut, Colin Sinclair, Kyle Spence, Charles Shang, Zoe Abrams, Morgan McGuire</i>	

Sign Value Constraint Decomposition for Efficient 1-Bit Quantization of Speech Translation Tasks	842
<i>Nan Chen, Yonghe Wang, Feilong Bao</i>	

Lightweight Audio Segmentation for Long-Form Speech Translation.....	847
<i>Jaesong Lee, Soyeon Kim, Hanbyul Kim, Joon Son Chung</i>	

Contrastive Feedback Mechanism for Simultaneous Speech Translation	852
<i>Haotian Tan, Sakriani Sakti</i>	

Towards Speech-To-Pictograms Translation	857
<i>Cécile Macaire, Chloé Dion, Didier Schwab, Benjamin Lecouteux, Emmanuelle Esperança-Rodier</i>	

HEARING DISORDERS

Automatic Assessment of Speech Production Skills for Children with Cochlear Implants Using Wav2Vec2.0 Acoustic Embeddings	862
<i>Seonwoo Lee, Sunhee Kim, Minhwa Chung</i>	

SyncVSR: Data-Efficient Visual Speech Recognition with End-To-End Crossmodal Audio Token Synchronization	867
<i>Young Jin Ahn, Jungwoo Park, Sangha Park, Jonghyun Choi, Kee-Eung Kim</i>	

Evaluating a 3-Factor Listener Model for Prediction of Speech Intelligibility to Hearing-Impaired Listeners	872
<i>Mark Huckvale, Gaston Hilkhuisen</i>	

Production of Fricative Consonants in French-Speaking Children with Cochlear Implants and Typical Hearing: Acoustic and Phonological Analyses.	877
<i>Sophie Fagniard, Brigitte Charlier, Véronique Delvaux, Bernard Harmegnies, Anne Huberlant, Myriam Piccaluga, Kathy Huet</i>	

Signal Processing Algorithm Effective for Sound Quality of Hearing Loss Simulators	882
<i>Toshio Irino, Shintaro Doan, Minami Ishikawa</i>	

Auditory Spatial Attention Detection Based on Feature Disentanglement and Brain Connectivity-Informed Graph Neural Networks.....	887
<i>Yixiang Niu, Ning Chen, Hongqing Zhu, Zhiying Zhu, Guangqiang Li, Yibo Chen</i>	

Automatic Detection of Hearing Loss from Children's Speech Using Wav2vec 2.0 Features	892
<i>Jessica Monaghan, Arun Sebastian, Nicky Chong-White, Vicky Zhang, Vijayalakshmi Easwar, Pdraig Kitterick</i>	

SPEECH DISORDERS 2

Whister: Using Whisper's Representations for Stuttering Detection	897
<i>Vrushank Changawala, Frank Rudzicz</i>	
Improving Speech-Based Dysarthria Detection Using Multi-Task Learning with Gradient Projection	902
<i>Yan Xiong, Visar Berisha, Julie Liss, Chaitali Chakrabarti</i>	
Cascaded Transfer Learning Strategy for Cross-Domain Alzheimer's Disease Recognition Through Spontaneous Speech	907
<i>Guanlin Chen, Yun Jin</i>	
A Cross-Attention Layer Coupled with Multimodal Fusion Methods for Recognizing Depression from Spontaneous Speech	912
<i>Loukas Ilias, Dimitris Askounis</i>	
Segmental and Suprasegmental Speech Foundation Models for Classifying Cognitive Risk Factors: Evaluating Out-Of-The-Box Performance	917
<i>Si-Ioi Ng, Lingfeng Xu, Kimberly D. Mueller, Julie Liss, Visar Berisha</i>	
Multimodal Continuous Fingerspelling Recognition Via Visual Alignment Learning	922
<i>Katerina Papadimitriou, Gerasimos Potamianos</i>	
Contrastive Learning Approach for Assessment of Phonological Precision in Patients with Tongue Cancer Using MRI Data	927
<i>Tomas Arias-Vergara, Paula Andrea Pérez-Toro, Xiaofeng Liu, Fangxu Xing, Maureen Stone, Jiachen Zhuo, Jerry L. Prince, Maria Schuster, Elmar Noeth, Jonghye Woo, Andreas Maier</i>	
DysArinVox: DYSpHonia & DYSarthria mandARIN Speech Corpus	932
<i>Haojie Zhang, Tao Zhang, Ganjun Liu, Dehui Fu, Xiaohui Hou, Ying Lv</i>	
YOLO-Stutter: End-To-End Region-Wise Speech Dysfluency Detection	937
<i>Xuanru Zhou, Anshul Kashyap, Steve Li, Ayati Sharma, Brittany Morin, David Baquirin, Jet Vonk, Zoe Ezzes, Zachary Miller, Maria Tempini, Jiachen Lian, Gopala Anumanchipalli</i>	
Automatic Longitudinal Investigation of Multiple Sclerosis Subjects	942
<i>Gábor Gosztolya, Veronika Svindt, Judit Bóna, Ildikó Hoffmann</i>	

TAUKADIAL CHALLENGE: SPEECH-BASED COGNITIVE ASSESSMENT IN CHINESE AND ENGLISH (SPECIAL SESSION)

Connected Speech-Based Cognitive Assessment in Chinese and English.....	947
<i>Saturnino Luz, Sofia De La Fuente Garcia, Fasih Haider, Davida Fromm, Brian Macwhinney, Alyssa Lanzi, Ya-Ning Chang, Chia-Ju Chou, Yi-Chien Liu</i>	
Cognitive Insights Across Languages: Enhancing Multimodal Interview Analysis	952
<i>David Ortiz-Perez, Jose Garcia-Rodriguez, David Tomás</i>	

Combining Acoustic Feature Sets for Detecting Mild Cognitive Impairment in the Interspeech'24 TAUKADIAL Challenge.....	957
<i>Gábor Gosztolya, László Tóth</i>	
Pre-Trained Feature Fusion and Matching for Mild Cognitive Impairment Detection	962
<i>Junwen Duan, Fangyuan Wei, Hong-Dong Li, Jin Liu</i>	
The Interspeech 2024 TAUKADIAL Challenge: Multilingual Mild Cognitive Impairment Detection with Multimodal Approach.....	967
<i>Benjamin Barrera-Altuna, Daeun Lee, Zaima Zarnaz, Jinyoung Han, Seungbae Kim</i>	
Leveraging Universal Speech Representations for Detecting and Assessing the Severity of Mild Cognitive Impairment Across Languages.....	972
<i>Anna Favaro, Tianyu Cao, Najim Dehak, Laureano Moro-Velazquez</i>	
Translingual Language Markers for Cognitive Assessment from Spontaneous Speech.....	977
<i>Bao Hoang, Yijiang Pang, Hiroko Dodge, Jiayu Zhou</i>	
Multilingual Speech and Language Analysis for the Assessment of Mild Cognitive Impairment: Outcomes from the Taukadial Challenge.....	982
<i>Paula Andrea Pérez-Toro, Tomas Arias-Vergara, Philipp Klumpp, Tobias Weise, Maria Schuster, Elmar Noeth, Juan Rafael Orozco-Aroyave, Andreas Maier</i>	
<u>SHOW AND TELL 1</u>	
Production of Phrases by Mechanical Models of the Human Vocal Tract.....	987
<i>Takayuki Arai, Ryohei Suzuki, Chandler Earp, Shinya Tsuji, Keiko Ochi</i>	
Faster Vocoder: A Multi Threading Approach to Achieve Low Latency During TTS Inference.....	989
<i>Vishal Gourav, Ankit Tyagi, Phanindra Mankale</i>	
A Powerful and Modern AAC Composition Tool for Impaired Speakers	991
<i>Aanchan Mohan, Monideep Chakraborti, Katelyn Eng, Nailia Kushaeva, Mirjana Prpa, Jordan Lewis, Tianyi Zhang, Vince Geisler, Carol Geisler</i>	
VoxFlow AI: Wearable Voice Converter for Atypical Speech	993
<i>Grzegorz P. Mika, Konrad Zielinski, Pawel Cyrt, Marek Grzelec</i>	
Stress Transfer in Speech-To-Speech Machine Translation.....	995
<i>Sai Akarsh, Vamshiraghusimha Narasinga, Anil Kumar Vuppala</i>	
Mobile PresenTra: NICT Fast Neural Text-To-Speech System on Smartphones with Incremental Inference of MS-FC-HiFi-GAN for Low-Latency Synthesis	997
<i>Takuma Okamoto, Yamato Ohtani, Hisashi Kawai</i>	
Multi-Speaker and Multi-Dialectal Catalan TTS Models for Video Gaming	999
<i>Alex Peiró-Lilja, José Giraldo, Martí Llopart-Font, Carme Armentano-Oller, Baybars Külebi, Mireia Farrús</i>	
ConnecTone: A Modular AAC System Prototype with Contextual Generative Text Prediction and Style-Adaptive Conversational TTS.....	1001
<i>Juliana Francis, Éva Székely, Joakim Gustafson</i>	
Reliable Dialogue System for Facilitating Student-Counselor Communication	1003
<i>Mahdin Rohmatillah, Bryan Gautama Ngo, Willianto Sulaiman, Po-Chuan Chen, Jen-Tzung Chien</i>	

CreakVC: A Voice Conversion Tool for Modulating Creaky Voice.....	1005
<i>Harm Lameris, Joakim Gustafson, Éva Székely</i>	

EZTalking: English Assessment Platform for Teachers and Students	1007
<i>Yu-Sheng Tsao, Yung-Chang Hsu, Jiun-Ting Li, Siang-Hong Weng, Tien-Hong Lo, Berlin Chen</i>	

KEYNOTE 2

Frontier of Frontend for Conversational Speech Processing	1009
<i>Shoko Araki</i>	

PHONETICS AND PHONOLOGY OF SECOND LANGUAGE ACQUISITION

Mmm Whatcha Say? Uncovering Distal and Proximal Context Effects in First and Second- Language Word Perception Using Psychophysical Reverse Correlation.....	1010
<i>Paige Tuttösi, H. Henny Yeung, Yue Wang, Fenqi Wang, Guillaume Denis, Jean-Julien Aucouturier, Angelica Lim</i>	

Automatic Speech Recognition with Parallel L1 and L2 Acoustic Phone Models to Evaluate /L/ Allophony in L2 English Speech Production	1015
<i>Anisia Popescu, Lori Lamel, Ioana Vasilescu, Laurence Devillers</i>	

Analysis of Articulatory Setting for L1 and L2 English Speakers Using MRI Data	1020
<i>Kevin Huang, Jack Goldberg, Louis Goldstein, Shrikanth Narayanan</i>	

Bilingual Rhotic Production Patterns: A Generational Comparison of Spanish-English Bilingual Speakers in Canada	1025
<i>Ioana Colgiu, Laura Spinu, Rajiv Rao, Yasaman Rafat</i>	

Exploring Impact of Pausing and Lexical Stress Patterns on L2 English Comprehensibility in Real Time.....	1030
<i>Sylvain Coulange, Tsuneo Kato, Solange Rossato, Monica Masperi</i>	

Mandarin T3 Production by Chinese and Japanese Native Speakers	1035
<i>Qi Wu</i>	

CORPORA-BASED APPROACHES IN AUTOMATIC EMOTION RECOGNITION

Reinforcement Learning Based Data Augmentation for Noise Robust Speech Emotion Recognition.....	1040
<i>Sumit Ranjan, Rupayan Chakraborty, Sunil Kumar Kopparapu</i>	

Unsupervised Domain Adaptation for Speech Emotion Recognition Using K-Nearest Neighbors Voice Conversion.....	1045
<i>Pravin Mote, Berrak Sisman, Carlos Busso</i>	

Confidence-Aware Hypothesis Transfer Networks for Source-Free Cross-Corpus Speech Emotion Recognition	1050
<i>Jincen Wang, Yan Zhao, Cheng Lu, Hailun Lian, Hongli Chang, Yuan Zong, Wenming Zheng</i>	

An Effective Local Prototypical Mapping Network for Speech Emotion Recognition.....	1055
<i>Yuxuan Xi, Yan Song, Lirong Dai, Haoyu Song, Ian McLoughlin</i>	

Speech Emotion Recognition with Multi-Level Acoustic and Semantic Information Extraction and Interaction.....	1060
<i>Yuan Gao, Hao Shi, Chenhui Chu, Tatsuya Kawahara</i>	

ANALYSIS OF SPEAKERS STATES AND TRAITS

How Rhythm Metrics Are Linked to Produced and Perceived Speaker Charisma.....	1065
<i>Oliver Niebuhr, Nafiseh Taghva</i>	
A Functional Trade-Off Between Prosodic and Semantic Cues in Conveying Sarcasm	1070
<i>Zhu Li, Xiyuan Gao, Yuqing Zhang, Shekhar Nayak, Matt Coler</i>	
Multimodal Belief Prediction.....	1075
<i>John Murzaku, Adil Soubki, Owen Rambow</i>	
Detecting Empathy in Speech.....	1080
<i>Run Chen, Haozhe Chen, Anushka Kulkarni, Eleanor Lin, Linda Pang, Divya Tadimeti, Jun Shin, Julia Hirschberg</i>	
Learning Representation of Therapist Empathy in Counseling Conversation Using Siamese Hierarchical Attention Network.....	1085
<i>Dehua Tao, Tan Lee, Harold Chui, Sarah Luk</i>	
Modelling Lexical Characteristics of the Healthy Aging Population: A Corpus-Based Study	1090
<i>Han Kunmei</i>	
Exploring Gender-Specific Speech Patterns in Automatic Suicide Risk Assessment.....	1095
<i>Maurice Gerczuk, Shahin Amiriparian, Justina Lutz, Wolfgang Strube, Irina Papazova, Alkomiet Hasan, Björn W. Schuller</i>	

SPOOFING AND DEEPPAKE DETECTION

Source Tracing of Audio Deepfake Systems	1100
<i>Nicholas Klein, Tianxiang Chen, Hemlata Tak, Ricardo Casal, Elie Khouiry</i>	
How Do Neural Spoofing Countermeasures Detect Partially Spoofed Audio?	1105
<i>Tianchi Liu, Lin Zhang, Rohan Kumar Das, Yi Ma, Ruijie Tao, Haizhou Li</i>	
Revisiting and Improving Scoring Fusion for Spoofing-Aware Speaker Verification Using Compositional Data Analysis	1110
<i>Xin Wang, Tomi Kinnunen, Kong Aik Lee, Paul-Gauthier Noé, Junichi Yamagishi</i>	
SecureSpectra: Safeguarding Digital Identity from Deep Fake Threats Via Intelligent Signatures	1115
<i>Oguzhan Baser, Kaan Kale, Sandeep P. Chinchali</i>	
Interpretable Temporal Class Activation Representation for Audio Spoofing Detection	1120
<i>Menglu Li, Xiao-Ping Zhang</i>	
DGPN: A Dual Graph Prototypical Network for Few-Shot Speech Spoofing Algorithm Recognition.....	1125
<i>Zirui Ge, Xinzhou Xu, Haiyan Guo, Tingting Wang, Zhen Yang, Björn W. Schuller</i>	

AUDIO CAPTIONING, TAGGING, AND AUDIO-TEXT RETRIEVAL

PFCA-Net: Pyramid Feature Fusion and Cross Content Attention Network for Automated Audio Captioning	1130
<i>Jianyuan Sun, Wenwu Wang, Mark D. Plumbley</i>	
Enhancing Automated Audio Captioning Via Large Language Models with Optimized Audio Encoding.....	1135
<i>Jizhong Liu, Gang Li, Junbo Zhang, Heinrich Dinkel, Yongqing Wang, Zhiyong Yan, Yujun Wang, Bin Wang</i>	
Audio-Text Retrieval with Transformer-Based Hierarchical Alignment and Disentangled Cross-Modal Representation.....	1140
<i>Yifei Xin, Zhihong Zhu, Xuxin Cheng, Xusheng Yang, Yuexian Zou</i>	
Streaming Audio Transformers for Online Audio Tagging.....	1145
<i>Heinrich Dinkel, Zhiyong Yan, Yongqing Wang, Junbo Zhang, Yujun Wang, Bin Wang</i>	
Efficient CNNs with Quaternion Transformations and Pruning for Audio Tagging.....	1150
<i>Aryan Chaudhary, Arshdeep Singh, Vinayak Abrol, Mark D. Plumbley</i>	
ParaCLAP – Towards a General Language-Audio Model for Computational Paralinguistic Tasks.....	1155
<i>Xin Jing, Andreas Triantafyllopoulos, Björn Schuller</i>	
Efficient Audio Captioning with Encoder-Level Knowledge Distillation	1160
<i>Xuenan Xu, Haohe Liu, Mengyue Wu, Wenwu Wang, Mark D. Plumbley</i>	

GENERATIVE SPEECH ENHANCEMENT

Universal Score-Based Speech Enhancement with High Content Preservation.....	1165
<i>Robin Scheibler, Yusuke Fujita, Yuma Shirahata, Tatsuya Komatsu</i>	
Genhancer: High-Fidelity Speech Enhancement Via Generative Modeling on Discrete Codec Tokens	1170
<i>Haici Yang, Jiaqi Su, Minje Kim, Zeyu Jin</i>	
Schrödinger Bridge for Generative Speech Enhancement.....	1175
<i>Ante Jukic, Roman Korostik, Jagadeesh Balam, Boris Ginsburg</i>	
Thunder : Unified Regression-Diffusion Speech Enhancement with a Single Reverse Step Using Brownian Bridge	1180
<i>Thanapat Trachu, Chawan Piansaddhayanon, Ekapol Chuangsuwanich</i>	
Pre-Training Feature Guided Diffusion Model for Speech Enhancement	1185
<i>Yiyuan Yang, Niki Trigoni, Andrew Markham</i>	
Guided Conditioning with Predictive Network on Score-Based Diffusion Model for Speech Enhancement	1190
<i>Dail Kim, Da-Hee Yang, Donghyun Kim, Joon-Hyuk Chang, Jeonghwan Choi, Moa Lee, Jaemo Yang, Han-Gil Moon</i>	

SPEECH SYNTHESIS: EVALUATION

SVSNet+: Enhancing Speaker Voice Similarity Assessment Models with Representations from Speech Foundation Models	1195
<i>Chun Yin, Tai-Shih Chi, Yu Tsao, Hsin-Min Wang</i>	
Enhancing Out-Of-Vocabulary Performance of Indian TTS Systems for Practical Applications Through Low-Effort Data Strategies	1200
<i>Srija Anand, Praveen Srinivasa Varadhan, Ashwin Sankar, Giri Raju, Mitesh M. Khapra</i>	
Assessing the Impact of Contextual Framing on Subjective TTS Quality	1205
<i>Jens Edlund, Christina Tännander, Sébastien Le Maguer, Petra Wagner</i>	
What Do People Hear? Listeners' Perception of Conversational Speech	1210
<i>Adaeze Adigwe, Sarenne Wallbridge, Simon King</i>	
Uncertainty-Aware Mean Opinion Score Prediction	1215
<i>Hui Wang, Shiwan Zhao, Jiaming Zhou, Xiguang Zheng, Haoqin Sun, Xuechen Wang, Yong Qin</i>	
Lifelong Learning MOS Prediction for Synthetic Speech Quality Evaluation	1220
<i>Félix Saget, Meysam Shamsi, Marie Tahon</i>	

MULTILINGUAL ASR

Continual Learning Optimizations for Auto-Regressive Decoder of Multilingual ASR Systems	1225
<i>Chin Yuen Kwok, Jia Qi Yip, Eng Siong Chng</i>	
ML-SUPERB 2.0: Benchmarking Multilingual Speech Models Across Modeling Constraints, Languages, and Datasets	1230
<i>Jiatong Shi, Shih-Heng Wang, William Chen, Martijn Bartelds, Vanya Bannihatti Kumar, Jinchuan Tian, Xuankai Chang, Dan Jurafsky, Karen Livescu, Hung-Yi Lee, Shinji Watanabe</i>	
Weighted Cross-Entropy for Low-Resource Languages in Multilingual Speech Recognition	1235
<i>Andrés Piñeiro-Martín, Carmen García-Mateo, Laura Docio-Fernandez, María Del Carmen López-Pérez, Georg Rehm</i>	
M ² ASR: Multilingual Multi-Task Automatic Speech Recognition Via Multi-Objective Optimization	1240
<i>A F M Saif, Lisha Chen, Xiaodong Cui, Songtao Lu, Brian Kingsbury, Tianyi Chen</i>	
MSR-86K: An Evolving, Multilingual Corpus with 86,300 Hours of Transcribed Audio for Speech Recognition Research	1245
<i>Song Li, Yongbin You, Xuezhi Wang, Zhengkun Tian, Ke Ding, Guanglu Wan</i>	
Improving Multilingual ASR Robustness to Errors in Language Input	1250
<i>Brady Houston, Omid Sadjadi, Zejiang Hou, Srikanth Vishnubhotla, Kyu J. Han</i>	

GENERAL TOPICS IN ASR

Improving Domain-Specific ASR with LLM-Generated Contextual Descriptions	1255
<i>Jiwon Suh, Injae Na, Woohwan Jung</i>	

A Multitask Training Approach to Enhance Whisper with Open-Vocabulary Keyword Spotting.....	1260
<i>Yuang Li, Min Zhang, Chang Su, Yinglu Li, Xiaosong Qiao, Mengxin Ren, Miaomiao Ma, Daimeng Wei, Shimin Tao, Hao Yang</i>	
CrisperWhisper: Accurate Timestamps on Verbatim Speech Transcriptions.....	1265
<i>Mario Zúsg, Laurin Wagner, Bernhad Thallinger</i>	
On Disfluency and Non-Lexical Sound Labeling for End-To-End Automatic Speech Recognition.....	1270
<i>Peter Mihajlik, Yan Meng, Mate S Kadar, Julian Linke, Barbara Schuppler, Katalin Mády</i>	
Inclusive ASR for Disfluent Speech: Cascaded Large-Scale Self-Supervised Learning with Targeted Fine-Tuning and Data Augmentation.....	1275
<i>Dena Mujtaba, Nihar R. Mahapatra, Megan Arney, J. Scott Yaruss, Caryn Herring, Jia Bin</i>	
DualPure: An Efficient Adversarial Purification Method for Speech Command Recognition.....	1280
<i>Hao Tan, Xiaochen Liu, Huan Zhang, Junjian Zhang, Yaguan Qian, Zhaoquan Gu</i>	
A Comparative Analysis of Bilingual and Trilingual Wav2Vec Models for Automatic Speech Recognition in Multilingual Oral History Archives	1285
<i>Jan Lehecka, Josef V. Psutka, Lubos Smidl, Pavel Ircing, Josef Psutka</i>	
A Layer-Wise Analysis of Mandarin and English Suprasegmentals in SSL Speech Models	1290
<i>Anton De La Fuente, Dan Jurafsky</i>	
Fine-Tuning Strategies for Dutch Dysarthric Speech Recognition: Evaluating the Impact of Healthy, Disease-Specific, and Speaker-Specific Data.....	1295
<i>Spyretta Leivaditi, Tatsunari Matsushima, Matt Coler, Shekhar Nayak, Vass Verkhodanova</i>	
Dysarthric Speech Recognition Using Curriculum Learning and Articulatory Feature Embedding.....	1300
<i>I-Ting Hsieh, Chung-Hsien Wu</i>	
Enhancing Dysarthric Speech Recognition for Unseen Speakers Via Prototype-Based Adaptation	1305
<i>Shiyao Wang, Shiwan Zhao, Jiaming Zhou, Aobo Kong, Yong Qin</i>	
An Efficient Text Augmentation Approach for Contextualized Mandarin Speech Recognition	1310
<i>Naijun Zheng, Xucheng Wan, Kai Liu, Ziqing Du, Zhou Huan</i>	
Investigating ASR Error Correction with Large Language Model and Multilingual 1-Best Hypotheses	1315
<i>Sheng Li, Chen Chen, Chin Yuen Kwok, Chenhui Chu, Eng Siong Chng, Hisashi Kawai</i>	
Efficiently Train ASR Models that Memorize Less and Perform Better with Per-Core Clipping	1320
<i>Lun Wang, Om Thakkar, Zhong Meng, Nicole Rafidi, Rohit Prabhavalkar, Arun Narayanan</i>	

SPOKEN LANGUAGE UNDERSTANDING

AR-NLU: A Framework for Enhancing Natural Language Understanding Model Robustness Against ASR Errors	1325
<i>Emmy Phung, Harsh Deshpande, Ahmad Emami, Kanishk Singh</i>	
Prompting Whisper for QA-Driven Zero-Shot End-To-End Spoken Language Understanding	1330
<i>Mohan Li, Simon Keizer, Rama Doddipatla</i>	
VN-SLU: A Vietnamese Spoken Language Understanding Dataset.....	1335
<i>Tuyen Tran, Khanh Le, Ngoc Dang Nguyen, Minh Vu, Huyen Ngo, Woomyoung Park, Thi Thu Trang Nguyen</i>	

Textless Dependency Parsing by Labeled Sequence Prediction.....	1340
<i>Shunsuke Kando, Yusuke Miyao, Jason Naradowsky, Shinnosuke Takamichi</i>	
Towards Speech Classification from Acoustic and Vocal Tract Data in Real-Time MRI.....	1345
<i>Yaoyao Yue, Michael Proctor, Luping Zhou, Rijul Gupta, Tharinda Piyadasa, Amelia Gully, Kirrie Ballard, Craig Jin</i>	
Efficient SQA from Long Audio Contexts: A Policy-Driven Approach.....	1350
<i>Alexander Johnson, Peter Plantinga, Pheobe Sun, Swaroop Gadiyaram, Abenezer Girma, Ahmad Emami</i>	

SPEECH AND MULTIMODAL RESOURCES

BESST Dataset: A Multimodal Resource for Speech-Based Stress Detection and Analysis.....	1355
<i>Jan Pešán, Vojtech Jurik, Martin Karafiát, Jan Cernocký</i>	
HebDB: A Weakly Supervised Dataset for Hebrew Speech Processing.....	1360
<i>Arnon Turetzky, Or Tal, Yael Segal, Yehoshua Dissen, Ella Zeldes, Amit Roth, Eyal Cohen, Yosi Shrem, Bronya R. Chernyak, Olga Seleznova, Joseph Keshet, Yossi Adi</i>	
GLOBE: A High-Quality English Corpus with Global Accents for Zero-Shot Speaker Adaptive Text-To-Speech.....	1365
<i>Wenbin Wang, Yang Song, Sanjay Jha</i>	
STraDa: A Singer Traits Dataset.....	1370
<i>Yuexuan Kong, Viet-Anh Tran, Romain Hennequin</i>	
MaViLS, a Benchmark Dataset for Video-To-Slide Alignment, Assessing Baseline Accuracy with a Multimodal Alignment Algorithm Leveraging Speech, OCR, and Visual Features	1375
<i>Katharina Anderer, Andreas Reich, Matthias Wölfel</i>	
MultiTalk: Enhancing 3D Talking Head Generation Across Languages with Multilingual Video Dataset.....	1380
<i>Kim Sung-Bin, Lee Chae-Yeon, Gihun Son, Oh Hyun-Bin, Janghoon Ju, Suekyeong Nam, Tae-Hyun Oh</i>	
Towards Measuring Fairness in Speech Recognition: Fair-Speech Dataset.....	1385
<i>Irina-Elena Veliche, Zhuangqun Huang, Vineeth Ayyat Kochaniyan, Fuchun Peng, Ozlem Kalinli, Michael L. Seltzer</i>	
Codecfake: An Initial Dataset for Detecting LLM-Based Deepfake Audio	1390
<i>Yi Lu, Yuankun Xie, Ruibo Fu, Zhengqi Wen, Jianhua Tao, Zhiyong Wang, Xin Qi, Xuefei Liu, Yongwei Li, Yukun Liu, Xiaopeng Wang, Shuchen Shi</i>	
SER Evals: In-Domain and Out-Of-Domain Benchmarking for Speech Emotion Recognition	1395
<i>Mohamed Osman, Daniel Z. Kaplan, Tamer Nadeem</i>	

PATHOLOGICAL SPEECH ANALYSIS 1

The MARRYS Helmet: A New Device for Researching and Training “Jaw Dancing”.....	1400
<i>Vidar Freyr Gudmundsson, Keve Márton Gönczi, Malin Svensson Lundmark, Donna Erickson, Oliver Niebuhr</i>	

Exploiting Foundation Models and Speech Enhancement for Parkinson's Disease Detection from Speech in Real-World Operative Conditions.....	1405
<i>Moreno La Quatra, Maria Francesca Turco, Torbjørn Svendsen, Giampiero Salvi, Juan Rafael Orozco-Arroyave, Sabato Marco Siniscalchi</i>	

VOLUME 3

Sustained Vowels for Pre- Vs Post-Treatment COPD Classification.....	1410
<i>Andreas Triantafyllopoulos, Anton Batliner, Wolfgang Mayr, Markus Fendler, Florian Pokorny, Maurice Gerczuk, Shahin Amiriparian, Thomas Berghaus, Björn Schuller</i>	
Adversarial Robustness Analysis in Automatic Pathological Speech Detection Approaches.....	1415
<i>Mahdi Amiri, Ina Kodrasi</i>	
Automatic Children Speech Sound Disorder Detection with Age and Speaker Bias Mitigation.....	1420
<i>Gahye Kim, Yujung Eom, Selina S. Sung, Seunghee Ha, Tae-Jin Yoon, Jungmin So</i>	

SPEECH AND LANGUAGE IN HEALTH: FROM REMOTE MONITORING TO MEDICAL CONVERSATIONS - 1 (SPECIAL SESSION)

Reference-Free Estimation of the Quality of Clinical Notes Generated from Doctor-Patient Conversations	1425
<i>Mojtaba Kadkhodaie Elyaderani, John Glover, Thomas Schaaf</i>	
Developing an End-To-End Framework for Predicting the Social Communication Severity Scores of Children with Autism Spectrum Disorder	1430
<i>Jihyun Mun, Sunhee Kim, Minhwa Chung</i>	
Multimodal Fusion for Vocal Biomarkers Using Vector Cross-Attention	1435
<i>Vladimir Despotovic, Abir Elbéji, Petr V. Nazarov, Guy Fagherazzi</i>	
Revealing Confounding Biases: A Novel Benchmarking Approach for Aggregate-Level Performance Metrics in Health Assessments.....	1440
<i>Stefano Gorla, Roseline Polle, Salvatore Fara, Nicholas Cummins</i>	
Developing Multi-Disorder Voice Protocols: A Team Science Approach Involving Clinical Expertise, Bioethics, Standards, and DEI	1445
<i>Yael Bensoussan, Satrajit Ghosh, Anais Rameau, Micah Boyer, Ruth Bahr, Stephanie Watts, Frank Rudzicz, Don Bolser, Jordan Lerner-Ellis, Shaheen Awan, Maria Powell, Jean-Christophe Belisle-Pipon, Vardit Ravitsky, Alistair Johnson, Alexandros Sigaras, Olivier Elemento, David Dorr, Philip Payne</i>	
Self-Supervised Embeddings for Detecting Individual Symptoms of Depression	1450
<i>Sri Harsha Dumpala, Katerina Dikaos, Abraham Nunes, Frank Rudzicz, Rudolf Uher, Sageev Oore</i>	
Comparing Ambulatory Voice Measures During Daily Life with Brief Laboratory Assessments in Speakers with and Without Vocal Hyperfunction	1455
<i>Daryush D. Mehta, Jarrad H. Van Stan, Hamzeh Ghasemzadeh, Robert E. Hillman</i>	
Predicting Acute Pain Levels Implicitly from Vocal Features	1460
<i>Jennifer Williams, Eike Schneiders, Henry Card, Tina Seabrooke, Beatrice Pakenham-Walsh, Tayyaba Azim, Lucy Valls-Reed, Ganesh Vigneswaran, John Robert Bautista, Rohan Chandra, Arya Farahi</i>	

Towards Intelligent Speech Assistants in Operating Rooms: A Multimodal Model for Surgical Workflow Analysis	1465
<i>Kubilay Can Demir, Belén Lojo Rodríguez, Tobias Weise, Andreas Maier, Seung Hee Yang</i>	

A Multimodal Framework for the Assessment of the Schizophrenia Spectrum	1470
<i>Gowtham Premananth, Yashish M. Siriwardena, Philip Resnik, Sonia Bansal, Deanna L. Kelly, Carol Espy-Wilson</i>	

SPEECH AND BRAIN

Exploring the Complementary Nature of Speech and Eye Movements for Profiling Neurological Disorders	1475
<i>Yuzhe Wang, Anna Favaro, Thomas Thebaud, Jesus Villalba, Najim Dehak, Laureano Moro-Velazquez</i>	

Refining Self-Supervised Learnt Speech Representation Using Brain Activations.....	1480
<i>Hengyu Li, Kangdi Mei, Zhaoci Liu, Yang Ai, Liping Chen, Jie Zhang, Zhenhua Ling</i>	

Large Language Model-Based fMRI Encoding of Language Functions for Subjects with Neurocognitive Disorder	1485
<i>Yuejiao Wang, Xianmin Gong, Lingwei Meng, Xixin Wu, Helen Meng</i>	

From Sound to Meaning in the Auditory Cortex: A Neuronal Representation and Classification Analysis.....	1490
<i>Kumar Neelabh, Vishnu Sreekumar</i>	

Towards an End-To-End Framework for Invasive Brain Signal Decoding with Large Language Models.....	1495
<i>Sheng Feng, Heyang Liu, Yu Wang, Yanfeng Wang</i>	

Toward Fully-End-To-End Listened Speech Decoding from EEG Signals.....	1500
<i>Jihwan Lee, Aditya Kommineni, Tiantian Feng, Kleanthis Avramidis, Xuan Shi, Sudarsana Reddy Kadiri, Shrikanth Narayanan</i>	

INNOVATIVE METHODS IN PHONETICS AND PHONOLOGY

The Use of Phone Categories and Cross-Language Modeling for Phone Alignment of Panāra.....	1505
<i>Emily P. Ahn, Eleanor Chodroff, Myriam Lapierre, Gina-Anne Levow</i>	

Deciphering Assamese Vowel Harmony with Featural InfoWaveGAN	1510
<i>Sneha Ray Barman, Shakuntala Mahanta, Neeraj Kumar Sharma</i>	

Phonological Feature Detection for US English Using the Phonet Library.....	1515
<i>Harsha Veena Tadavarthy, Austin Jones, Margaret E. L. Renwick</i>	

K-Means and Hierarchical Clustering of F0 Contours	1520
<i>Constantijn Kaland, Jeremy Steffman, Jennifer Cole</i>	

Tradition Or Innovation: A Comparison of Modern ASR Methods for Forced Alignment	1525
<i>Rotem Rouso, Eyal Cohen, Joseph Keshet, Eleanor Chodroff</i>	

Using Wav2vec 2.0 for Phonetic Classification Tasks: Methodological Aspects	1530
<i>Lila Kim, Cédric Gendrot</i>	

The Sub-Band Cepstrum as a Tool for Locating Local Spectral Regions of Phonetic Sensitivity: A First Attempt with Multi-Speaker Vowel Data	1535
<i>Michael Lambropoulos, Frantz Clermont, Shunichi Ishihara</i>	
Speaker-Independent Acoustic-To-Articulatory Inversion Through Multi-Channel Attention Discriminator.....	1540
<i>Woo-Jin Chung, Hong-Goo Kang</i>	
Speaker- And Text-Independent Estimation of Articulatory Movements and Phoneme Alignments from Speech.....	1545
<i>Tobias Weise, Philipp Klumpp, Kubilay Can Demir, Paula Andrea Pérez-Toro, Maria Schuster, Elmar Noeth, Bjoern Heismann, Andreas Maier, Seung Hee Yang</i>	
Preprocessing for Acoustic-To-Articulatory Inversion Using Real-Time MRI Movies of Japanese Speech	1550
<i>Anna Oura, Hideaki Kikuchi, Tetsunori Kobayashi</i>	

VOICE, TONES AND F0

Impact of the Tonal Factor on Diphthong Realizations in Standard Mandarin with Generalized Additive Mixed Models.....	1555
<i>Chenyu Li, Jalal Al-Tamimi</i>	
A Study on the Information Mechanism of the 3rd Tone Sandhi Rule in Mandarin Disyllabic Words	1560
<i>Liu Xiaowang, Jinsong Zhang</i>	
Gender and Age Based f_0 -Variation in the German Plapper Corpus.....	1565
<i>Melanie Weirich, Daniel Duran, Stefanie Jannedy</i>	
Voice Quality in Telephone Speech: Comparing Acoustic Measures Between VoIP Telephone and High-Quality Recordings.....	1570
<i>Chenzi Xu, Jessica Wormald, Paul Foulkes, Philip Harrison, Vincent Hughes, Poppy Welch, Finnian Kelly, David Van Der Vloed</i>	
The Use of Modifiers and F0 in Remote Referential Communication with Human and Computer Partners.....	1575
<i>Iona Gessinger, Bistra Andreeva, Benjamin R. Cowan</i>	

EMOTION RECOGNITION: RESOURCES AND BENCHMARKS

EmoBox: Multilingual Multi-Corpus Speech Emotion Recognition Toolkit and Benchmark	1580
<i>Ziyang Ma, Mingjie Chen, Hezhao Zhang, Zhisheng Zheng, Wenxi Chen, Xiquan Li, Jiaxin Ye, Xie Chen, Thomas Hain</i>	
INTERSPEECH 2009 Emotion Challenge Revisited: Benchmarking 15 Years of Progress in Speech Emotion Recognition	1585
<i>Andreas Triantafyllopoulos, Anton Batliner, Simon Rampp, Manuel Milling, Björn Schuller</i>	
What Does it Take to Generalize SER Model Across Datasets? a Comprehensive Benchmark	1590
<i>Adham Ibrahim, Shady Shehata, Ajinkya Kulkarni, Mukhtar Mohamed, Muhammad Abdul-Mageed</i>	
WHiSER: White House Tapes Speech Emotion Recognition Corpus.....	1595
<i>Abinay Reddy Naini, Lucas Goncalves, Mary A. Kohler, Donita Robinson, Elizabeth Richerson, Carlos Busso</i>	

Evaluating Transformer-Enhanced Deep Reinforcement Learning for Speech Emotion Recognition	1600
<i>Siddique Latif, Raja Jurdak, Björn W. Schuller</i>	
Boosting Cross-Corpus Speech Emotion Recognition Using CycleGAN with Contrastive Learning	1605
<i>Jincen Wang, Yan Zhao, Cheng Lu, Chuangao Tang, Sunan Li, Yuan Zong, Wenming Zheng</i>	

SPEAKER AND LANGUAGE IDENTIFICATION AND DIARIZATION

Multi-Latency Look-Ahead for Streaming Speaker Segmentation	1610
<i>Bilal Rahou, Hervé Bredin</i>	
Once More Diarization: Improving Meeting Transcription Systems Through Segment-Level Speaker Reassignment.....	1615
<i>Christoph Boeddeker, Tobias Cord-Landwehr, Reinhold Haeb-Umbach</i>	
ASoBO: Attentive Beamformer Selection for Distant Speaker Diarization in Meetings	1620
<i>Théo Mariotte, Anthony Larcher, Silvio Montrésor, Jean-Hugh Thomas</i>	
Hybrid-Diarization System with Overlap Post-Processing for the DISPLACE 2024 Challenge	1625
<i>Gabriel Pîrlogeanu, Octavian Pascu, Alexandru-Lucian Georgescu, Horia Cucu</i>	
The Second DISPLACE Challenge: Diarization of SPeaker and LAnguage in Conversational Environments.....	1630
<i>Shareef Babu Kalluri, Prachi Singh, Pratik Roy Chowdhuri, Apoorva Kulkarni, Shikha Baghel, Pradyoth Hegde, Swapnil Sontakke, Deepak K T, S. R. Mahadeva Prasanna, Deepu Vijayasanen, Sriram Ganapathy</i>	
TalTech-IRIT-LIS Speaker and Language Diarization Systems for DISPLACE 2024	1635
<i>Joonas Kalda, Tanel Alumae, Martin Lebourdais, Hervé Bredin, Séverin Baroudi, Ricard Marxer</i>	
Exploring Energy-Based Models for Out-Of-Distribution Detection in Dialect Identification.....	1640
<i>Yaqian Hao, Chenguang Hu, Yingying Gao, Shilei Zhang, Junlan Feng</i>	
Exploring Spoken Language Identification Strategies for Automatic Transcription of Multilingual Broadcast and Institutional Speech.....	1645
<i>Martina Valente, Fabio Brugnara, Giovanni Morrone, Enrico Zovato, Leonardo Badino</i>	
AG-LSEC: Audio Grounded Lexical Speaker Error Correction	1650
<i>Rohit Paturi, Xiang Li, Sundararajan Srinivasan</i>	
Speaker Change Detection with Weighted-Sum Knowledge Distillation Based on Self-Supervised Pre-Trained Models	1655
<i>Hang Su, Yuxiang Kong, Lichun Fan, Peng Gao, Yujun Wang, Zhiyong Wu</i>	
SOMSRED: Sequential Output Modeling for Joint Multi-Talker Overlapped Speech Recognition and Speaker Diarization	1660
<i>Naoki Makishima, Naotaka Kawata, Mana Ihori, Tomohiro Tanaka, Shota Orihashi, Atsushi Ando, Ryo Masumura</i>	
Song Data Cleansing for End-To-End Neural Singer Diarization Using Neural Analysis and Synthesis Framework	1665
<i>Hokuto Munakata, Ryo Terashima, Yusuke Fujita</i>	

AUDIO-TEXT RETRIEVAL

DiffATR: Diffusion-Based Generative Modeling for Audio-Text Retrieval.....	1670
<i>Yifei Xin, Xuxin Cheng, Zhihong Zhu, Xusheng Yang, Yuexian Zou</i>	
Bridging Language Gaps in Audio-Text Retrieval	1675
<i>Zhiyong Yan, Heinrich Dinkel, Yongqing Wang, Jizhong Liu, Junbo Zhang, Yujun Wang, Bin Wang</i>	
Domain Adaptation for Contrastive Audio-Language Models	1680
<i>Soham Deshmukh, Rita Singh, Bhiksha Raj</i>	
TinyCLAP: Distilling Contrastive Language-Audio Pretrained Models	1685
<i>Francesco Paissan, Elisabetta Farella</i>	
BTS: Bridging Text and Sound Modalities for Metadata-Aided Respiratory Sound Classification.....	1690
<i>June-Woo Kim, Miika Toikkanen, Yera Choi, Seoung-Eun Moon, Ho-Young Jung</i>	
Enhanced Feature Learning with Normalized Knowledge Distillation for Audio Tagging	1695
<i>Yuwu Tang, Ziang Ma, Haitao Zhang</i>	

SPEECH ENHANCEMENT

RaD-Net 2: A Causal Two-Stage Repairing and Denoising Speech Enhancement Network with Knowledge Distillation and Complex Axial Self-Attention	1700
<i>Mingshuai Liu, Zhuangqi Chen, Xiaopeng Yan, Yuanjun Lv, Xianjun Xia, Chuanzeng Huang, Yijian Xiao, Lei Xie</i>	
DNN-Based Monaural Speech Enhancement Using Alternate Analysis Windows for Phase and Magnitude Modification	1705
<i>Xi Liu, John H. L. Hansen</i>	
Improved Remixing Process for Domain Adaptation-Based Speech Enhancement by Mitigating Data Imbalance in Signal-To-Noise Ratio	1710
<i>Li Li, Shogo Seki</i>	
Neural Network Augmented Kalman Filter for Robust Acoustic Howling Suppression.....	1715
<i>Yixuan Zhang, Hao Zhang, Meng Yu, Dong Yu</i>	
Improving Speech Enhancement by Integrating Inter-Channel and Band Features with Dual-Branch Conformer	1720
<i>Jizhen Li, Xinmeng Xu, Weiping Tu, Yuhong Yang, Rong Zhu</i>	
An Exploration of Length Generalization in Transformer-Based Speech Enhancement.....	1725
<i>Qiquan Zhang, Hongxu Zhu, Xinyuan Qian, Eliathamby Ambikairajah, Haizhou Li</i>	
Reducing Speech Distortion and Artifacts for Speech Enhancement by Loss Function	1730
<i>Haixin Guan, Wei Dai, Guangyong Wang, Xiaobin Tan, Peng Li, Jiaen Liang</i>	
Are Recent Deep Learning-Based Speech Enhancement Methods Ready to Confront Real-World Noisy Environments?	1735
<i>Candy Olivia Mawalim, Shogo Okada, Masashi Unoki</i>	
Beyond Performance Plateaus: A Comprehensive Study on Scalability in Speech Enhancement	1740
<i>Wangyou Zhang, Kohei Saijo, Jee-Weon Jung, Chenda Li, Shinji Watanabe, Yanmin Qian</i>	

SPEECH CODING

TD-PLC: A Semantic-Aware Speech Encoding for Improved Packet Loss Concealment	1745
<i>Jinghong Zhang, Zugang Zhao, Yonghui Liu, Jianbing Liu, Zhiqiang He, Kai Niu</i>	
BS-PLCNet 2: Two-Stage Band-Split Packet Loss Concealment Network with Intra-Model Knowledge Distillation.....	1750
<i>Zihan Zhang, Xianjun Xia, Chuanzeng Huang, Yijian Xiao, Lei Xie</i>	
On Improving Error Resilience of Neural End-To-End Speech Coders	1755
<i>Kishan Gupta, Nicola Pia, Srikanth Korse, Andreas Brendel, Guillaume Fuchs, Markus Multrus</i>	
Speech Quality Evaluation of Neural Audio Codecs.....	1760
<i>Thomas Muller, Stephane Ragot, Laetitia Gros, Pierrick Philippe, Pascal Scalart</i>	
A Low-Bitrate Neural Audio Codec Framework with Bandwidth Reduction and Recovery for High-Sampling-Rate Waveforms.....	1765
<i>Yang Ai, Ye-Xin Lu, Xiao-Hang Jiang, Zheng-Yan Sheng, Rui-Chen Zheng, Zhen-Hua Ling</i>	
CodecFake: Enhancing Anti-Spoofing Models Against Deepfake Audios from Codec-Based Speech Synthesis Systems	1770
<i>Haibin Wu, Yuan Tseng, Hung-Yi Lee</i>	

SPEECH SYNTHESIS: EXPRESSIVITY AND EMOTION

GTR-Voice: Articulatory Phonetics Informed Controllable Expressive Speech Synthesis	1775
<i>Zehua Kcriss Li, Meiying Melissa Chen, Yi Zhong, Pinxin Liu, Zhiyao Duan</i>	
TSP-TTS: Text-Based Style Predictor with Residual Vector Quantization for Expressive Text-To-Speech	1780
<i>Donghyun Seong, Hoyoung Lee, Joon-Hyuk Chang</i>	
Spontaneous Style Text-To-Speech Synthesis with Controllable Spontaneous Behaviors Based on Language Models	1785
<i>Wei qin Li, Peiji Yang, Yicheng Zhong, Yixuan Zhou, Zhisheng Wang, Zhiyong Wu, Xixin Wu, Helen Meng</i>	
Text-Aware and Context-Aware Expressive Audiobook Speech Synthesis.....	1790
<i>Dake Guo, Xinfa Zhu, Liumeng Xue, Yongmao Zhang, Wenjie Tian, Lei Xie</i>	
Controlling Emotion in Text-To-Speech with Natural Language Prompts	1795
<i>Thomas Bott, Florian Lux, Ngoc Thang Vu</i>	
Retrieval Augmented Generation in Prompt-Based Text-To-Speech Synthesis with Context-Aware Contrastive Language-Audio Pretraining	1800
<i>Jinlong Xue, Yayue Deng, Yingming Gao, Ya Li</i>	
Emotion Arithmetic: Emotional Speech Synthesis Via Weight Space Interpolation	1805
<i>Pavan Kalyan, Preeti Rao, Preethi Jyothi, Pushpak Bhattacharyya</i>	
EmoSphere-TTS: Emotional Style and Intensity Modeling Via Spherical Emotion Vector for Controllable Emotional Text-To-Speech	1810
<i>Deok-Hyeon Cho, Hyung-Seok Oh, Seung-Bin Kim, Sang-Hoon Lee, Seong-Whan Lee</i>	

Expressive Paragraph Text-To-Speech Synthesis with Multi-Step Variational Autoencoder	1815
<i>Xuyuan Li, Zengqiang Shang, Peiyang Shi, Hua Hua, Ta Li, Pengyuan Zhang</i>	
Differentiable Time-Varying Linear Prediction in the Context of End-To-End Analysis-By-Synthesis.....	1820
<i>Chin-Yun Yu, György Fazekas</i>	

SPEECH SYNTHESIS: TOOLS AND DATA

SRC4VC: Smartphone-Recorded Corpus for Voice Conversion Benchmark	1825
<i>Yuki Saito, Takuto Igarashi, Kentaro Seki, Shinnosuke Takamichi, Ryuichi Yamamoto, Kentaro Tachibana, Hiroshi Saruwatari</i>	
Rasa: Building Expressive Speech Synthesis Systems for Indian Languages in Low-Resource Settings	1830
<i>Praveen Srinivasa Varadhan, Ashwin Sankar, Giri Raju, Mitesh M Khapra</i>	
FLEURS-R: A Restored Multilingual Speech Corpus for Generation Tasks.....	1835
<i>Min Ma, Yuma Koizumi, Shigeki Karita, Heiga Zen, Jason Riesa, Haruko Ishikawa, Michiel Bacchiani</i>	
WenetSpeech4TTS: A 12,800-Hour Mandarin TTS Corpus for Large Speech Generation Model Benchmark.....	1840
<i>Linhan Ma, Dake Guo, Kun Song, Yuepeng Jiang, Shuai Wang, Liumeng Xue, Weiming Xu, Huan Zhao, Binbin Zhang, Lei Xie</i>	
MSceneSpeech: A Multi-Scene Speech Dataset for Expressive Speech Synthesis	1845
<i>Qian Yang, Jialong Zuo, Zhe Su, Ziyue Jiang, Mingze Li, Zhou Zhao, Feiyang Chen, Zhefeng Wang, Baoxing Huai</i>	
LibriTTS-P: A Corpus with Speaking Style and Speaker Identity Prompts for Text-To-Speech and Style Captioning	1850
<i>Masaya Kawamura, Ryuichi Yamamoto, Yuma Shirahata, Takuya Hasumi, Kentaro Tachibana</i>	
1000 African Voices: Advancing Inclusive Multi-Speaker Multi-Accent Speech Synthesis.....	1855
<i>Sewade Ogun, Abraham T. Owodunni, Tobi Olatunji, Eniola Alese, Babatunde Oladimeji, Tejumade Afonja, Kayode Olaleye, Naome A. Etori, Tosin Adewumi</i>	
SaSLaW: Dialogue Speech Corpus with Audio-Visual Egocentric Information Toward Environment-Adaptive Dialogue Speech Synthesis	1860
<i>Osamu Take, Shinnosuke Takamichi, Kentaro Seki, Yoshiaki Bando, Hiroshi Saruwatari</i>	

SPEECH SYNTHESIS: SINGING VOICE SYNTHESIS

MakeSinger: A Semi-Supervised Training Method for Data-Efficient Singing Voice Synthesis Via Classifier-Free Diffusion Guidance	1865
<i>Semin Kim, Myeonghun Jeong, Hyeonseung Lee, Minchan Kim, Byoung Jin Choi, Nam Soo Kim</i>	
Challenge of Singing Voice Synthesis Using Only Text-To-Speech Corpus with FIRNet Source-Filter Neural Vocoder	1870
<i>Takuma Okamoto, Yamato Ohtani, Sota Shimizu, Tomoki Toda, Hisashi Kawai</i>	

Period Singer: Integrating Periodic and Aperiodic Variational Autoencoders for Natural-Sounding End-To-End Singing Voice Synthesis	1875
<i>Taewoo Kim, Choonsang Cho, Young Han Lee</i>	
Singing Voice Data Scaling-Up: An Introduction to ACE-Opencpop and ACE-KiSing	1880
<i>Jiatong Shi, Yueqian Lin, Xinyi Bai, Keyi Zhang, Yuning Wu, Yuxun Tang, Yifeng Yu, Qin Jin, Shinji Watanabe</i>	
X-Singer: Code-Mixed Singing Voice Synthesis Via Cross-Lingual Learning	1885
<i>Ji-Sang Hwang, Hyeonrae Noh, Yoonseok Hong, Insoo Oh</i>	
An End-To-End Approach for Chord-Conditioned Song Generation	1890
<i>Shuochen Gao, Shun Lei, Fan Zhuo, Hangyu Liu, Feng Liu, Boshi Tang, Qiaochu Huang, Shiyin Kang, Zhiyong Wu</i>	

LLM IN ASR

Speech Prefix-Tuning with RNNT Loss for Improving LLM Predictions	1895
<i>Murali Karthick Baskar, Andrew Rosenberg, Bhuvana Ramabhadran, Neeraj Gaur, Zhong Meng</i>	
Speech ReaLLM – Real-Time Speech Recognition with Multimodal Language Models by Teaching the Flow of Time	1900
<i>Frank Seide, Yangyang Shi, Morrie Doulaty, Yashesh Gaur, Junteng Jia, Chunyang Wu</i>	
A Transcription Prompt-Based Efficient Audio Large Language Model for Robust Speech Recognition	1905
<i>Yangze Li, Xiong Wang, Songjun Cao, Yike Zhang, Long Ma, Lei Xie</i>	
Pinyin Regularization in Error Correction for Chinese Speech Recognition with Large Language Models	1910
<i>Zhiyuan Tang, Dong Wang, Shen Huang, Shidong Shang</i>	

VISION AND SPEECH

AVCap: Leveraging Audio-Visual Features as Text Tokens for Captioning.....	1915
<i>Jongsuk Kim, Jiwon Shin, Junmo Kim</i>	
LipGER: Visually-Conditioned Generative Error Correction for Robust Automatic Speech Recognition	1920
<i>Sreyan Ghosh, Sonal Kumar, Ashish Seth, Purva Chiniya, Utkarsh Tyagi, Ramani Duraiswami, Dinesh Manocha</i>	
Joint Speaker Features Learning for Audio-Visual Multichannel Speech Separation and Recognition	1925
<i>Guinan Li, Jiajun Deng, Youjun Chen, Mengzhe Geng, Shujie Hu, Zhe Li, Zengrui Jin, Tianzi Wang, Xurong Xie, Helen Meng, Xunying Liu</i>	
CNVSRC 2023: The First Chinese Continuous Visual Speech Recognition Challenge	1930
<i>Chen Chen, Zehua Liu, Xiaolou Li, Lantian Li, Dong Wang</i>	

SPOKEN DOCUMENT SUMMARIZATION

Optimizing the Role of Human Evaluation in LLM-Based Spoken Document Summarization Systems.....	1935
<i>Margaret Kroll, Kelsey Kraus</i>	
Key-Element-Informed sLLM Tuning for Document Summarization	1940
<i>Sangwon Ryu, Heejin Do, Yunsu Kim, Gary Geunbae Lee, Jungseul Ok</i>	
Sentence-Wise Speech Summarization: Task, Datasets, and End-To-End Modeling with LM Knowledge Distillation.....	1945
<i>Kohei Matsuura, Takanori Ashihara, Takafumi Moriya, Masato Mimura, Takatomo Kano, Atsunori Ogawa, Marc Delcroix</i>	
An End-To-End Speech Summarization Using Large Language Model	1950
<i>Hengchao Shang, Zongyao Li, Jiaxin Guo, Shaojun Li, Zhiqiang Rao, Yuanchang Luo, Daimeng Wei, Hao Yang</i>	
Prompting Large Language Models with Audio for General-Purpose Speech Summarization.....	1955
<i>Wonjune Kang, Deb Roy</i>	
Real-Time Speech Summarization for Medical Conversations	1960
<i>Khai Le-Duc, Khai-Nguyen Nguyen, Long Vo-Dang, Truong-Son Hy</i>	

SPEECH AND LANGUAGE IN HEALTH: FROM REMOTE MONITORING TO MEDICAL CONVERSATIONS - 2 (SPECIAL SESSIONS)

It's Time to Take Action: Acoustic Modeling of Motor Verbs to Detect Parkinson's Disease	1965
<i>Daniel Escobar-Grisales, Cristian David Rios-Urrego, Ilja Baumann, Korbinian Riedhammer, Elmar Noeth, Tobias Bocklet, Adolfo M. Garcia, Juan Rafael Orozco-Arroyave</i>	
Towards Objective and Interpretable Speech Disorder Assessment: A Comparative Analysis of CNN and Transformer-Based Models	1970
<i>Malo Maisonneuve, Corinne Fredouille, Muriel Lalain, Alain Ghio, Virginie Woisard</i>	
Macro-Descriptors for Alzheimer's Disease Detection Using Large Language Models	1975
<i>Catarina Botelho, John Mendonça, Anna Pompili, Tanja Schultz, Alberto Abad, Isabel Trancoso</i>	
Infusing Acoustic Pause Context into Text-Based Dementia Assessment	1980
<i>Franziska Braun, Sebastian P. Bayerl, Florian Hönig, Hartmut Lehfeld, Thomas Hillemacher, Tobias Bocklet, Korbinian Riedhammer</i>	
Towards Scalable Remote Assessment of Mild Cognitive Impairment Via Multimodal Dialog.....	1985
<i>Oliver Roesler, Jackson Liscombe, Michael Neumann, Hardik Kothare, Abhishek Hosamath, Lakshmi Arbatti, Doug Habberstad, Christiane Suendermann-Oeft, Meredith Bartlett, Cathy Zhang, Nikhil Sukhdev, Kolja Wilms, Anusha Badathala, Sandrine Istas, Steve Ruhmel, Bryan Hansen, Madeline Hannan, David Henley, Arthur Wallace, Ira Shoulson, David Suendermann-Oeft, Vikram Ramanarayanan</i>	
Automatic Recognition and Detection of Aphasic Natural Speech	1990
<i>Mara Barberis, Pieter De Clercq, Bastiaan Tamm, Hugo Van Hamme, Maaike Vandermosten</i>	

When Whisper Listens to Aphasia: Advancing Robust Post-Stroke Speech Recognition.....	1995
<i>Giulia Sanguedolce, Sophie Brook, Dragos C. Gruia, Patrick A. Naylor, Fatemeh Geranmayeh</i>	
Automatic Prediction of Amyotrophic Lateral Sclerosis Progression Using Longitudinal Speech Transformer	2000
<i>Liming Wang, Yuan Gong, Nauman Dawalatabad, Marco Vilela, Katerina Placek, Brian Tracey, Yishu Gong, Alan Premasiri, Fernando Vieira, James Glass</i>	
How Consistent Are Speech-Based Biomarkers in Remote Tracking of ALS Disease Progression Across Languages? a Case Study of English and Dutch	2005
<i>Hardik Kothare, Michael Neumann, Cathy Zhang, Jackson Liscombe, Jordi W J Van Unnik, Lianne C M Botman, Leonard H Van Den Berg, Ruben P A Van Eijk, Vikram Ramanarayanan</i>	
“So . . . My Child . . .” – How Child ADHD Influences the Way Parents Talk	2010
<i>Anika A. Spiesberger, Andreas Triantafyllopoulos, Alexander Kathan, Anastasia Semertzidou, Caterina Gawrilow, Tilman Reinelt, Wolfgang A. Rauch, Björn Schuller</i>	
Variability of Speech Timing Features Across Repeated Recordings: A Comparison of Open-Source Extraction Techniques	2015
<i>Judith Dineley, Ewan Carr, Lauren L. White, Catriona Lucas, Zahia Rahman, Tian Pan, Faith Matcham, Johnny Downs, Richard J. Dobson, Thomas F. Quatieri, Nicholas Cummins</i>	
Zero-Shot End-To-End Spoken Question Answering in Medical Domain	2020
<i>Yanis Labrak, Adel Moumen, Richard Dufour, Mickael Rouvier</i>	
Perceiver-Prompt: Flexible Speaker Adaptation in Whisper for Chinese Disordered Speech Recognition	2025
<i>Yicong Jiang, Tianzi Wang, Xurong Xie, Juan Liu, Wei Sun, Nan Yan, Hui Chen, Lan Wang, Xunying Liu, Feng Tian</i>	

SHOW AND TELL 2

Custom Wake Word Detection.....	2030
<i>Kesavaraj V, Charan Devarkonda, Vamshiraghusimha Narasinga, Anil Kumar Vuppala</i>	
Edged Based Audio-Visual Speech Enhancement Demonstrator	2032
<i>Song Chen, Mandar Gogate, Kia Dashtipour, Jasper Kirton-Wingate, Adeel Hussain, Faiyaz Doctor, Tughrul Arslan, Amir Hussain</i>	
Real-Time Gaze-Directed Speech Enhancement for Audio-Visual Hearing-Aids	2034
<i>Arif Reza Anway, Bryony Buck, Mandar Gogate, Kia Dashtipour, Michael Akeroyd, Amir Hussain</i>	
Detection of Background Agents Speech in Contact Centers.....	2036
<i>Abhishek Kumar, Srikanth Konjeti, Jithendra Vepa</i>	
Leveraging Large Language Models for Post-Transcription Correction in Contact Centers.....	2038
<i>Bramhendra Koilakuntla, Prajesh Rana, Paras Ahuja, Srikanth Konjeti, Jithendra Vepa</i>	
Understanding “understanding”: Presenting a Richly Annotated Multimodal Corpus of Dyadic Interaction.....	2040
<i>Leonie Schade, Nico Dallmann, Olcay Tük, Stefan Lazarov, Petra Wagner</i>	

A Demonstrator for Articulation-Based Command Word Recognition	2042
<i>Joao Vitor Possamai De Menezes, Arne-Lukas Fietkau, Tom Diener, Steffen Kurbis, Peter Birkholz</i>	
Pragmatically Similar Utterance Finder Demonstration.....	2044
<i>Nigel G. Ward, Andres Segura</i>	
Real-Time Scheme for Rapid Extraction of Speaker Embeddings in Challenging Recording Conditions	2046
<i>Kai Liu, Ziqing Du, Zhou Huan, Xucheng Wan, Naijun Zheng</i>	
TEEMI: A Speaking Practice Tool for L2 English Learners	2048
<i>Szu-Yu Chen, Tien-Hong Lo, Yao-Ting Sung, Ching-Yu Tseng, Berlin Chen</i>	

PROSODY

Automatic Pitch Accent Classification Through Image Classification.....	2050
<i>Na Hu, Hugo Schnack, Amalia Arvaniti</i>	
Form and Function in Prosodic Representation: In the Case of 'ma' in Tianjin Mandarin	2055
<i>Tianqi Geng, Hui Feng</i>	
On Comparing Time- And Frequency-Domain Rhythm Measures in Classifying Assamese Dialects	2060
<i>Joyshree Chakraborty, Leena Dihingia, Priyankoo Sarmah, Rohit Sinha</i>	
The Prosody of the Verbal Prefix Ge-: Historical and Experimental Evidence	2065
<i>Chiara Riegger, Tina Bögel, George Walkden</i>	
Influences of Morphosyntax and Semantics on the Intonation of Mandarin Chinese Wh-Indeterminates	2070
<i>Hongchen Wu, Jiwon Yun</i>	
Urdu Alternative Questions: A Hat Pattern.....	2075
<i>Benazir Mumtaz, Miriam Butt</i>	

FOUNDATIONAL MODELS FOR DEEFAKE AND SPOOFED SPEECH DETECTION

Spoofed Speech Detection with a Focus on Speaker Embedding	2080
<i>Hoan My Tran, David Guennec, Philippe Martin, Aghilas Sini, Damien Lolive, Arnaud Delhay, Pierre-François Marteau</i>	
Exploring Self-Supervised Embeddings and Synthetic Data Augmentation for Robust Audio Deepfake Detection	2085
<i>Juan M. Martín-Doñas, Aitor Álvarez, Eros Rosello, Angel M. Gomez, Antonio M. Peinado</i>	
Attentive Merging of Hidden Embeddings from Pre-Trained Speech Model for Anti-Spoofing Detection	2090
<i>Zihan Pan, Tianchi Liu, Hardik B. Sailor, Qiongqiong Wang</i>	
Adapter Learning from Pre-Trained Model for Robust Spoof Speech Detection.....	2095
<i>Haochen Wu, Wu Guo, Shengyu Peng, Zhu hai Li, Jie Zhang</i>	
Speech Formants Integration for Generalized Detection of Synthetic Speech Spoofing Attacks.....	2100
<i>Kexu Liu, Yuanxin Wang, Shengchen Li, Xi Shao</i>	

Balance, Multiple Augmentation, and Re-Synthesis: A Triad Training Strategy for Enhanced Audio Deepfake Detection	2105
<i>Thien-Phuc Doan, Long Nguyen-Vu, Kihun Hong, Souhwan Jung</i>	

SPEAKER RECOGNITION 1

Fine-Tune Pre-Trained Models with Multi-Level Feature Fusion for Speaker Verification.....	2110
<i>Shengyu Peng, Wu Guo, Haochen Wu, Zuoliang Li, Jie Zhang</i>	
Speaker Conditional Sinc-Extractor for Personal VAD	2115
<i>En-Lun Yu, Kuan-Hsun Ho, Jie-Wei Hung, Shih-Chieh Huang, Berlin Chen</i>	
Enhancing ECAPA-TDNN with Feature Processing Module and Attention Mechanism for Speaker Verification	2120
<i>Shiu-Hsiang Liou, Po-Cheng Chan, Chia-Ping Chen, Tzu-Chieh Lin, Chung-Li Lu, Yu-Han Cheng, Hsiang-Feng Chuang, Wei-Yu Chen</i>	
MR-RawNet: Speaker Verification System with Multiple Temporal Resolutions for Variable Duration Utterances Using Raw Waveforms	2125
<i>Seung-Bin Kim, Chan-Yeong Lim, Jungwoo Heo, Ju-Ho Kim, Hyun-Seo Shin, Kyo-Won Koo, Ha-Jin Yu</i>	
Disentangled Representation Learning for Environment-Agnostic Speaker Recognition.....	2130
<i>Kihyun Nam, Hee-Soo Heo, Jee-Weon Jung, Joonson Chung</i>	
Multi-Channel Extension of Pre-Trained Models for Speaker Verification.....	2135
<i>Ladislav Mořner, Romain Serizel, Lukáš Burget, Oldřich Plchot, Emmanuel Vincent, Junyi Peng, Jan Cernocký</i>	
Efficient Integrated Features Based on Pre-Trained Models for Speaker Verification	2140
<i>Yishuang Li, Wenhao Guan, Hukai Huang, Shiyu Miao, Qi Su, Lin Li, Qingyang Hong</i>	
SE/BN Adapter: Parametric Efficient Domain Adaptation for Speaker Recognition	2145
<i>Tianhao Wang, Lantian Li, Dong Wang</i>	
DB-PMMAE: Dual-Branch Prototypical Masked AutoEncoder with Locality for Domain Robust Speaker Verification	2150
<i>Wei-Lin Xie, Yu-Xuan Xi, Yan Song, Jian-Tao Zhang, Hao-Yu Song, Ian McLoughlin</i>	
Evaluating the Santa Barbara Corpus: Challenges of the Breadth of Conversational Spoken Language	2155
<i>Matthew Maciejewski, Dominik Klement, Ruizhe Huang, Matthew Wiesner, Sanjeev Khudanpur</i>	
A Comprehensive Investigation on Speaker Augmentation for Speaker Recognition.....	2160
<i>Zhenyu Zhou, Shibiao Xu, Shi Yin, Lantian Li, Dong Wang</i>	

SOURCE SEPARATION 1

Noise-Robust Speech Separation with Fast Generative Correction.....	2165
<i>Helin Wang, Jesús Villalba, Laureano Moro-Velazquez, Jiarui Hai, Thomas Thebaud, Najim Dehak</i>	

VOLUME 4

MSDET: Multitask Speaker Separation and Direction-Of-Arrival Estimation Training	2170
<i>Roland Hartanto, Sakriani Sakti, Koichi Shinoda</i>	
Unsupervised Improved MVDR Beamforming for Sound Enhancement	2175
<i>Jacob Kealey, John R. Hershey, François Grondin</i>	
Improving Generalization of Speech Separation in Real-World Scenarios: Strategies in Simulation, Optimization, and Evaluation	2180
<i>Ke Chen, Jiaqi Su, Taylor Berg-Kirkpatrick, Shlomo Dubnov, Zeyu Jin</i>	
Enhanced Deep Speech Separation in Clustered Ad Hoc Distributed Microphone Environments	2185
<i>Jihyun Kim, Stijn Kindt, Nilesch Madhu, Hong-Goo Kang</i>	
Towards Audio Codec-Based Speech Separation	2190
<i>Jia Qi Yip, Shengkui Zhao, Dianwen Ng, Eng Siong Chng, Bin Ma</i>	

AUDIO-VISUAL AND GENERATIVE SPEECH ENHANCEMENT

Locally Aligned Rectified Flow Model for Speech Enhancement Towards Single-Step Diffusion	2195
<i>Zhengxiao Li, Nakamasa Inoue</i>	
Diffusion Gaussian Mixture Audio Denoise	2200
<i>Pu Wang, Junhui Li, Jialu Li, Liangdong Guo, Youshan Zhang</i>	
An Analysis of the Variance of Diffusion-Based Speech Enhancement	2205
<i>Bunlong Lay, Timo Gerkmann</i>	
FlowAVSE: Efficient Audio-Visual Speech Enhancement with Conditional Flow Matching	2210
<i>Chaeyoung Jung, Suyeon Lee, Ji-Hoon Kim, Joon Son Chung</i>	
RT-LA-VocE: Real-Time Low-SNR Audio-Visual Speech Enhancement	2215
<i>Honglie Chen, Rodrigo Mira, Stavros Petridis, Maja Pantic</i>	
Complex Image-Generative Diffusion Transformer for Audio Denoising	2220
<i>Junhui Li, Pu Wang, Jialu Li, Youshan Zhang</i>	
Noise-Aware Speech Enhancement Using Diffusion Probabilistic Model	2225
<i>Yuchen Hu, Chen Chen, Ruizhe Li, Qiushi Zhu, Eng Siong Chng</i>	

SPEECH PRIVACY AND BANDWIDTH EXPANSION

Privacy PORCUPINE: Anonymization of Speaker Attributes Using Occurrence Normalization for Space-Filling Vector Quantization	2230
<i>Mohammad Hassan Vali, Tom Bäckström</i>	
SilentCipher: Deep Audio Watermarking	2235
<i>Mayank Kumar Singh, Naoya Takahashi, Weihsiang Liao, Yuki Mitsufoji</i>	
Frequency-Mix Knowledge Distillation for Fake Speech Detection	2240
<i>Cunhang Fan, Shunbo Dong, Jun Xue, Yujie Chen, Jiangyan Yi, Zhao Lv</i>	

A New Approach to Voice Authenticity.....	2245
<i>Nicolas M. Müller, Piotr Kawa, Shen Hu, Matthias Neu, Jennifer Williams, Philip Sperl, Konstantin Böttinger</i>	
TraceableSpeech: Towards Proactively Traceable Text-To-Speech with Watermarking.....	2250
<i>Junzuo Zhou, Jiangyan Yi, Tao Wang, Jianhua Tao, Ye Bai, Chu Yuan Zhang, Yong Ren, Zhengqi Wen</i>	
HarmoNet: Partial DeepFake Detection Network Based on Multi-Scale HarmoF0 Feature Fusion.....	2255
<i>Liwei Liu, Huihui Wei, Dongya Liu, Zhonghua Fu</i>	
Unmasking Neural Codecs: Forensic Identification of AI-Compressed Speech	2260
<i>Denise Moussa, Sandra Bergmann, Christian Riess</i>	
SWiBE: A Parameterized Stochastic Diffusion Process for Noise-Robust Bandwidth Expansion.....	2265
<i>Yin-Tse Lin, Shreya G. Upadhyay, Bo-Hao Su, Chi-Chun Lee</i>	
MultiStage Speech Bandwidth Extension with Flexible Sampling Rate Control.....	2270
<i>Ye-Xin Lu, Yang Ai, Zheng-Yan Sheng, Zhen-Hua Ling</i>	
MaskSR: Masked Language Model for Full-Band Speech Restoration	2275
<i>Xu Li, Qirui Wang, Xiaoyu Liu</i>	

SPEECH SYNTHESIS: PROSODY

Word-Level Text Markup for Prosody Control in Speech Synthesis.....	2280
<i>Yuliya Korotkova, Ilya Kalinovskiy, Tatiana Vakhrusheva</i>	
Should You Use a Probabilistic Duration Model in TTS? Probably! Especially for Spontaneous Speech	2285
<i>Shivam Mehta, Harm Lameris, Rajiv Punmiya, Jonas Beskow, Eva Szekely, Gustav Eje Henter</i>	
Total-Duration-Aware Duration Modeling for Text-To-Speech Systems	2290
<i>Sefik Emre Eskimez, Xiaofei Wang, Manthan Thakker, Chung-Hsien Tsai, Canrun Li, Zhen Xiao, Hemin Yang, Zirun Zhu, Min Tang, Jinyu Li, Sheng Zhao, Naoyuki Kanda</i>	
A Human-In-The-Loop Approach to Improving Cross-Text Prosody Transfer.....	2295
<i>Himanshu Maurya, Atli Sigurgeirsson</i>	
Towards Expressive Zero-Shot Speech Synthesis with Hierarchical Prosody Modeling.....	2300
<i>Yuepeng Jiang, Tao Li, Fengyu Yang, Lei Xie, Meng Meng, Yujun Wang</i>	
Multi-Modal Automatic Prosody Annotation with Contrastive Pretraining of Speech-Silence and Word-Punctuation.....	2305
<i>Jinzuomu Zhong, Yang Li, Hui Huang, Korin Richmond, Jie Liu, Zhibo Su, Jing Guo, Benlai Tang, Fengjie Zhu</i>	

ACCENTED SPEECH, PROSODIC FEATURES, DIALECT, EMOTION, SOUND CLASSIFICATION

Improving Self-Supervised Pre-Training Using Accent-Specific Codebooks.....	2310
<i>Darshan Prabhu, Abhishek Gupta, Omkar Nitsure, Preethi Jyothi, Sriram Ganapathy</i>	

Performant ASR Models for Medical Entities in Accented Speech.....	2315
<i>Tejumade Afonja, Tobi Olatunji, Sewade Ogun, Naome A. Etori, Abraham Owodunni, Moshood Yekini</i>	
LAHAJA: A Robust Multi-Accent Benchmark for Evaluating Hindi ASR Systems	2320
<i>Tahir Javed, Janki Nawale, Sakshi Joshi, Eldho George, Kaushal Bhogale, Deovrat Mehendale, Mitesh M. Khapra</i>	
LearnerVoice: A Dataset of Non-Native English Learners' Spontaneous Speech	2325
<i>Haechan Kim, Junho Myung, Seoyoung Kim, Sungpah Lee, Dongyeop Kang, Juho Kim</i>	
MinSpeech: A Corpus of Southern Min Dialect for Automatic Speech Recognition	2330
<i>Jiayan Lin, Shenghui Lu, Hukai Huang, Wenhao Guan, Binbin Xu, Hui Bu, Qingyang Hong, Lin Li</i>	
Cross-Modal Features Interaction-And-Aggregation Network with Self-Consistency Training for Speech Emotion Recognition	2335
<i>Ying Hu, Huamin Yang, Hao Huang, Liang He</i>	
Exploring Multilingual Unseen Speaker Emotion Recognition: Leveraging Co-Attention Cues in Multitask Learning	2340
<i>Arnav Goel, Medha Hira, Anubha Gupta</i>	
SELM: Enhancing Speech Emotion Recognition for Out-Of-Domain Scenarios	2345
<i>Hazim Bukhari, Soham Deshmukh, Hira Dharmyal, Bhiksha Raj, Rita Singh</i>	
The Processing of Stress in End-To-End Automatic Speech Recognition Models.....	2350
<i>Martijn Bentum, Louis Ten Bosch, Tom Lentz</i>	
LingWav2Vec2: Linguistic-Augmented Wav2vec 2.0 for Vietnamese Mispronunciation Detection	2355
<i>Tuan Nguyen, Huy Dat Tran</i>	
Learning from Memory-Based Models	2360
<i>Rhiannon Mogridge, Anton Ragni</i>	
Towards End-To-End Unified Recognition for Mandarin and Cantonese.....	2365
<i>Meiling Chen, Pengjie Liu, Heng Yang, Haofeng Wang</i>	

NEURAL NETWORK ADAPTATION

Shared-Adapters: A Novel Transformer-Based Parameter Efficient Transfer Learning Approach for Children's Automatic Speech Recognition.....	2370
<i>Thomas Rolland, Alberto Abad</i>	
AdaRA: Adaptive Rank Allocation of Residual Adapters for Speech Foundation Model.....	2375
<i>Zhouyuan Huo, Dongseong Hwang, Gan Song, Khe Chai Sim, Weiran Wang</i>	
Leveraging Adapter for Parameter-Efficient ASR Encoder.....	2380
<i>Kyuhong Shim, Jinkyu Lee, Hyunjae Kim</i>	
Whisper Multilingual Downstream Task Tuning Using Task Vectors	2385
<i>Ji-Hun Kang, Jae-Hong Lee, Mun-Hak Lee, Joon-Hyuk Chang</i>	
Speaker-Smoothed kNN Speaker Adaptation for End-To-End ASR	2390
<i>Shaojun Li, Daimeng Wei, Hengchao Shang, Jiaxin Guo, Zongyao Li, Zhanglin Wu, Zhiqiang Rao, Yuanchang Luo, Xianghui He, Hao Yang</i>	

Qifusion-Net: Layer-Adapted Stream/Non-Stream Model for End-To-End Multi-Accent Speech Recognition	2395
<i>Jinming Chen, Jingyi Fang, Yuanzhong Zheng, Yaoxuan Wang, Haojun Fei</i>	

ASR AND LLMS

HuBERT-EE: Early Exiting HuBERT for Efficient Speech Recognition	2400
<i>Ji Won Yoon, Beom Jun Woo, Nam Soo Kim</i>	
MaLa-ASR: Multimedia-Assisted LLM-Based ASR	2405
<i>Guanrou Yang, Ziyang Ma, Fan Yu, Zhifu Gao, Shiliang Zhang, Xie Chen</i>	
Spoken-To-Written Text Conversion with Large Language Model.....	2410
<i>Hyunjung Choi, Muyeol Choi, Yohan Lim, Minkyu Lee, Seonhui Kim, Seung Yun, Donghyun Kim, Sanghun Kim</i>	
MM-KWS: Multi-Modal Prompts for Multilingual User-Defined Keyword Spotting.....	2415
<i>Zhiqi Ai, Zhiyong Chen, Shugong Xu</i>	
Whisper-Flamingo: Integrating Visual Features into Whisper for Audio-Visual Speech Recognition and Translation	2420
<i>Andrew Rouditchenko, Yuan Gong, Samuel Thomas, Leonid Karlinsky, Hilde Kuehne, Rogerio Feris, James Glass</i>	
Speech Recognition Models Are Strong Lip-Readers	2425
<i>K R Prajwal, Triantafyllos Afouras, Andrew Zisserman</i>	

PATHOLOGICAL SPEECH ANALYSIS 3

Towards Self-Attention Understanding for Automatic Articulatory Processes Analysis in Cleft Lip and Palate Speech	2430
<i>Ilja Baumann, Dominik Wagner, Maria Schuster, Korbinian Riedhammer, Elmar Noeth, Tobias Bocklet</i>	
Clever Hans Effect Found in Automatic Detection of Alzheimer's Disease Through Speech	2435
<i>Yin-Long Liu, Rui Feng, Jia-Hong Yuan, Zhen-Hua Ling</i>	
Leveraging Phonemic Transcription and Whisper Toward Clinically Significant Indices for Automatic Child Speech Assessment	2440
<i>Yeh-Sheng Lin, Shu-Chuan Tseng, Jyh-Shing Roger Jang</i>	
Developing Vocal System Impaired Patient-Aimed Voice Quality Assessment Approach Using ASR Representation-Included Multiple Features.....	2445
<i>Shaoxiang Dang, Tetsuya Matsumoto, Yoshinori Takeuchi, Takashi Tsuboi, Yasuhiro Tanaka, Daisuke Nakatsubo, Satoshi Maesawa, Ryuta Saito, Masahisa Katsuno, Hiroaki Kudo</i>	
A Cluster-Based Personalized Federated Learning Strategy for End-To-End ASR of Dementia Patients	2450
<i>Wei-Tung Hsu, Chin-Po Chen, Yun-Shao Lin, Chi-Chun Lee</i>	
A Comparative Analysis of Federated Learning for Speech-Based Cognitive Decline Detection	2455
<i>Stefan Kalabakov, Monica Gonzalez-Machorro, Florian Eyben, Björn W. Schuller, Bert Arnrich</i>	

Multimodal Digital Biomarkers for Longitudinal Tracking of Speech Impairment Severity in ALS: An Investigation of Clinically Important Differences	2460
<i>Michael Neumann, Hardik Kothare, Jackson Liscombe, Emma C. L. Leschly, Oliver Roesler, Vikram Ramanarayanan</i>	

SPEECH DISORDERS 3

Enhancing Voice Wake-Up for Dysarthria: Mandarin Dysarthria Speech Corpus Release and Customized System Design	2465
<i>Ming Gao, Hang Chen, Jun Du, Xin Xu, Hongxiao Guo, Hui Bu, Jianxing Yang, Ming Li, Chin-Hui Lee</i>	
Towards Improving NAM-To-Speech Synthesis Intelligibility Using Self-Supervised Speech Models	2470
<i>Neil Shah, Shirish Karande, Vineet Gandhi</i>	
PARAN: Variational Autoencoder-Based End-To-End Articulation-To-Speech System for Speech Intelligibility	2475
<i>Seyun Um, Doyeon Kim, Hong-Goo Kang</i>	
Acoustic Changes in Speech Prosody Produced by Children with Autism After Robot-Assisted Speech Training	2480
<i>Si Chen, Bruce Xiao Wang, Yitian Hong, Fang Zhou, Angel Chan, Po-Yi Tang, Bin Li, Chunyi Wen, James Cheung, Yan Liu, Zhuoming Chen</i>	
Fine-Tuning Automatic Speech Recognition for People with Parkinson's: An Effective Strategy for Enhancing Speech Technology Accessibility	2485
<i>Xiuwen Zheng, Bornali Phukon, Mark Hasegawa-Johnson</i>	
Learnings from Curating a Trustworthy, Well-Annotated, and Useful Dataset of Disordered English Speech	2490
<i>Pan-Pan Jiang, Jimmy Tobin, Katrin Tomanek, Robert Macdonald, Katie Seaver, Richard Cave, Marilyn Ladewig, Rus Heywood, Jordan Green</i>	
Training Data Augmentation for Dysarthric Automatic Speech Recognition by Text-To-Dysarthric- Speech Synthesis	2494
<i>Wing-Zin Leung, Mattias Cross, Anton Ragni, Stefan Goetze</i>	
Wav2vec 2.0 Embeddings Are No Swiss Army Knife -- A Case Study for Multiple Sclerosis	2499
<i>Gábor Gosztolya, Mercedes Vetráb, Veronika Svindt, Judit Bóna, Ildikó Hoffmann</i>	

SPEECH RECOGNITION WITH LARGE PRETRAINED SPEECH MODELS FOR UNDER-REPRESENTED LANGUAGES (SPECIAL SESSION)

Interface Design for Self-Supervised Speech Models	2504
<i>Yi-Jen Shih, David Harwath</i>	
Comparing Discrete and Continuous Space LLMs for Speech Recognition	2509
<i>Yaoxun Xu, Shi-Xiong Zhang, Jianwei Yu, Zhiyong Wu, Dong Yu</i>	
Improving Whisper's Recognition Performance for Under-Represented Language Kazakh Leveraging Unpaired Speech and Text	2514
<i>Jinpeng Li, Yu Pu, Qi Sun, Wei-Qiang Zhang</i>	

Empowering Low-Resource Language ASR Via Large-Scale Pseudo Labeling.....	2519
<i>Kaushal Santosh Bhogale, Deovrat Mehendale, Niharika Parasa, Sathish Kumar Reddy G, Tahir Javed, Pratyush Kumar, Mitesh M. Khapra</i>	
Interleaved Audio/Audiovisual Transfer Learning for AV-ASR in Low-Resourced Languages	2524
<i>Zhengyang Li, Patrick Blumenberg, Jing Liu, Thomas Graave, Timo Lohrenz, Siegfried Kunzmann, Tim Fingscheidt</i>	
Adapter Pre-Training for Improved Speech Recognition in Unseen Domains Using Low Resource Adapter Tuning of Self-Supervised Models	2529
<i>Sathvik Udupa, Jesuraj Bandekar, Saurabh Kumar, Deekshitha G, Sandhya B, Abhayjeet S, Savitha Murthy, Priyanka Pai, Srinivasa Raghavan, Raoul Nanavati, Prasanta Kumar Ghosh</i>	
Towards Rehearsal-Free Multilingual ASR: A LoRA-Based Case Study on Whisper	2534
<i>Tianyi Xu, Kaixun Huang, Pengcheng Guo, Yu Zhou, Longtao Huang, Hui Xue, Lei Xie</i>	
Exploring Adaptation Techniques of Large Speech Foundation Models for Low-Resource ASR: A Case Study on Northern Sámi	2539
<i>Yaroslav Getman, Tamas Grosz, Katri Hiovain-Asikainen, Mikko Kurimo</i>	
Learn and Don't Forget: Adding a New Language to ASR Foundation Models.....	2544
<i>Mengjie Qian, Siyuan Tang, Rao Ma, Kate M. Knill, Mark J. F. Gales</i>	

SPEECH PROCESSING USING DISCRETE SPEECH UNITS (SPECIAL SESSION)

TokSing: Singing Voice Synthesis Based on Discrete Tokens.....	2549
<i>Yuning Wu, Chunlei Zhang, Jiatong Shi, Yuxun Tang, Shan Yang, Qin Jin</i>	
How Should We Extract Discrete Audio Tokens from Self-Supervised Models?	2554
<i>Pooneh Mousavi, Jarod Duret, Salah Zaiem, Luca Della Libera, Artem Ploujnikov, Cem Subakan, Mirco Ravanelli</i>	
The Interspeech 2024 Challenge on Speech Processing Using Discrete Units	2559
<i>Xuankai Chang, Jiatong Shi, Jinchuan Tian, Yuning Wu, Yuxun Tang, Yihan Wu, Shinji Watanabe, Yossi Adi, Xie Chen, Qin Jin</i>	
SingOMD: Singing Oriented Multi-Resolution Discrete Representation Construction from Speech Models.....	2564
<i>Yuxun Tang, Yuning Wu, Jiatong Shi, Qin Jin</i>	
MMM: Multi-Layer Multi-Residual Multi-Stream Discrete Speech Representation from Self-Supervised Learning Model	2569
<i>Jiatong Shi, Xutai Ma, Hirofumi Inaguma, Anna Sun, Shinji Watanabe</i>	
Codec-ASR: Training Performant Automatic Speech Recognition Systems with Discrete Speech Representations	2574
<i>Kunal Dhawan, Nithin Rao Koluguri, Ante Jukic, Ryan Langman, Jagadeesh Balam, Boris Ginsburg</i>	

KEYNOTE 3

Analysis of Pathological Speech – Pitfalls Along the Way.....	2579
<i>Elmar Noeth</i>	

DATABASES AND PROGRESS IN METHODOLOGY

VoxSim: A Perceptual Voice Similarity Dataset	2580
<i>Junseok Ahn, Youkyum Kim, Yeunju Choi, Doyeop Kwak, Ji-Hoon Kim, Seongkyu Mun, Joon Son Chung</i>	
UY/CH-CHILD -- A Public Chinese L2 Speech Database of Uyghur Children	2585
<i>Mewlude Nijat, Chen Chen, Dong Wang, Askar Hamdulla</i>	
State-Of-The-Art Speech Production MRI Protocol for New 0.55 Tesla Scanners.....	2590
<i>Prakash Kumar, Ye Tian, Yongwan Lim, Sophia X. Cui, Christina Hagedorn, Dani Byrd, Uttam K. Sinha, Shrikanth Narayanan, Krishna S. Nayak</i>	
DBD-CI: Doubling the Band Density for Bilateral Cochlear Implants	2595
<i>Mingyue Shi, Huali Zhou, Qinglin Meng, Nengheng Zheng</i>	
Leveraging Large Language Models to Refine Automatic Feedback Generation at Articulatory Level in Computer Aided Pronunciation Training.....	2600
<i>Huihang Zhong, Yanlu Xie, Zijin Yao</i>	
Decoding Human Language Acquisition: EEG Evidence for Predictive Probabilistic Statistics in Word Segmentation	2605
<i>Bin Zhao, Mingxuan Huang, Chenlu Ma, Jinyi Xue, Aijun Li, Kunyu Xu</i>	

ARTICULATION, CONVERGENCE AND PERCEPTION

Behavioral Evidence for Higher Speech Rate Convergence Following Natural than Artificial Time Altered Speech	2610
<i>Jérémy Giroud, Jessica Lei, Kirsty Phillips, Matthew H. Davis</i>	
A Novel Experimental Design for the Study of Listener-To-Listener Convergence in Phoneme Categorization	2615
<i>Qingye Shen, Leonardo Lancia, Noel Nguyen</i>	
Cross-Attention-Guided WaveNet for EEG-To-MEL Spectrogram Reconstruction	2620
<i>Hao Li, Yuan Fang, Xueliang Zhang, Fei Chen, Guanglai Gao</i>	
What if HAL Breathed? Enhancing Empathy in Human-AI Interactions with Breathing Speech Synthesis.....	2625
<i>Nicolò Loddo, Francisca Pessanha, Almila Akdag</i>	
Magnitude and Timing of Acceleration Peaks in Stressed and Unstressed Syllables	2630
<i>Malin Svensson Lundmark</i>	

SPEECH EMOTION RECOGNITION

ExHuBERT: Enhancing HuBERT Through Block Extension and Fine-Tuning on 37 Emotion Datasets	2635
<i>Shahin Amiriparian, Filip Packan, Maurice Gerczuk, Björn W. Schuller</i>	
Dataset-Distillation Generative Model for Speech Emotion Recognition.....	2640
<i>Fabian Ritter-Gutierrez, Kuan-Po Huang, Jeremy H. M. Wong, Dianwen Ng, Hung-Yi Lee, Nancy F. Chen, Eng-Siong Chng</i>	

DropFormer: A Dynamic Noise-Dropping Transformer for Speech Emotion Recognition	2645
<i>Jialong Mai, Xiaofen Xing, Weidong Chen, Xiangmin Xu</i>	
From Text to Emotion: Unveiling the Emotion Annotation Capabilities of LLMs	2650
<i>Minxue Niu, Mimansa Jaiswal, Emily Mower Provost</i>	

SELF-SUPERVISED MODELS IN SPEAKER RECOGNITION

Self-Supervised Speaker Verification with Relational Mask Prediction.....	2655
<i>Ju-Ho Kim, Hee-Soo Heo, Bong-Jin Lee, Youngki Kwon, Minjae Lee, Ha-Jin Yu</i>	
Towards Supervised Performance on Speaker Verification with Self-Supervised Learning by Leveraging Large-Scale ASR Models	2660
<i>Victor Miara, Theo Lepage, Reda Dehak</i>	
Improving Noise Robustness in Self-Supervised Pre-Trained Model for Speaker Verification	2665
<i>Chan-Yeong Lim, Hyun-Seo Shin, Ju-Ho Kim, Jungwoo Heo, Kyo-Won Koo, Seung-Bin Kim, Ha-Jin Yu</i>	
On the Impact of Several Regularization Techniques on Label Noise Robustness of Self-Supervised Speaker Verification Systems	2670
<i>Abderrahim Fathan, Xiaolin Zhu, Jahangir Alam</i>	
Parameter-Efficient Fine-Tuning of Speaker-Aware Dynamic Prompts for Speaker Verification.....	2675
<i>Zhe Li, Man-Wai Mak, Hung-Yi Lee, Helen Meng</i>	
Whisper-PMFA: Partial Multi-Scale Feature Aggregation for Speaker Verification Using Whisper Models	2680
<i>Yiyang Zhao, Shuai Wang, Guangzhi Sun, Zehua Chen, Chao Zhang, Mingxing Xu, Thomas Fang Zheng</i>	

SPEECH QUALITY ASSESSMENT

Embedding Learning for Preference-Based Speech Quality Assessment.....	2685
<i>Chenghung Hu, Yusuke Yasuda, Tomoki Toda</i>	
IndicMOS: Multilingual MOS Prediction for 7 Indian Languages	2690
<i>Sathvik Udupa, Soumi Maiti, Prasanta Kumar Ghosh</i>	
Experimental Evaluation of MOS, AB and BWS Listening Test Designs.....	2695
<i>Dan Wells, Andrea Lorena Aldana Blanco, Cassia Valentini, Erica Cooper, Aidan Pine, Junichi Yamagishi, Korin Richmond</i>	
Enhancing No-Reference Speech Quality Assessment with Pairwise, Triplet Ranking Losses, and ASR Pretraining	2700
<i>Bao Thang Ta, Minh Tu Le, Van Hai Do, Huynh Thi Thanh Binh</i>	

PRIVACY AND SECURITY IN SPEECH COMMUNICATION 1

Harder Or Different? Understanding Generalization of Audio Deepfake Detection	2705
<i>Nicolas M. Müller, Nicholas Evans, Hemlata Tak, Philip Sperl, Konstantin Böttinger</i>	

Prompt Tuning for Audio Deepfake Detection: Computationally Efficient Test-Time Domain Adaptation with Limited Target Dataset.....	2710
<i>Hideyuki Oiso, Yuto Matsunaga, Kazuya Kakizaki, Taiki Miyagawa</i>	
Robust Spread Spectrum Speech Watermarking Using Linear Prediction and Deep Spectral Shaping.....	2715
<i>David Looney, Nikolay D. Gaubitch</i>	
RawBMamba: End-To-End Bidirectional State Space Model for Audio Deepfake Detection.....	2720
<i>Yujie Chen, Jiangyan Yi, Jun Xue, Chenglong Wang, Xiaohui Zhang, Shunbo Dong, Siding Zeng, Jianhua Tao, Zhao Lv, Cunhang Fan</i>	
How Private is Low-Frequency Speech Audio in the Wild? an Analysis of Verbal Intelligibility by Humans and Machines	2725
<i>Ailin Liu, Pepijn Vunderink, Jose Vargas Quiros, Chirag Raman, Hayley Hung</i>	
RW-VoiceShield: Raw Waveform-Based Adversarial Attack on One-Shot Voice Conversion.....	2730
<i>Ching-Yu Yang, Shreya G. Upadhyay, Ya-Tse Wu, Bo-Hao Su, Chi-Chun Lee</i>	

SPEECH SYNTHESIS: VOICE CONVERSION 2

Improvement Speaker Similarity for Zero-Shot Any-To-Any Voice Conversion of Whispered and Regular Speech.....	2735
<i>Aleksei Gusev, Anastasia Avdeeva</i>	
Utilizing Adaptive Global Response Normalization and Cluster-Based Pseudo Labels for Zero-Shot Voice Conversion.....	2740
<i>Ji Sub Um, Hoirin Kim</i>	
Vec-Tok-VC+: Residual-Enhanced Robust Zero-Shot Voice Conversion with Progressive Constraints in a Dual-Mode Training Strategy	2745
<i>Linhan Ma, Xinfu Zhu, Yuanjun Lv, Zhichao Wang, Ziqian Wang, Wendi He, Hongbin Zhou, Lei Xie</i>	
Noise-Robust Voice Conversion by Conditional Denoising Training Using Latent Variables of Recording Quality and Environment.....	2750
<i>Takuto Igarashi, Yuki Saito, Kentaro Seki, Shinnosuke Takamichi, Ryuichi Yamamoto, Kentaro Tachibana, Hiroshi Saruwatari</i>	
Pre-Training Neural Transducer-Based Streaming Voice Conversion for Faster Convergence and Alignment-Free Training	2755
<i>Hiroki Kanagawa, Takafumi Moriya, Yusuke Ijima</i>	
Residual Speaker Representation for One-Shot Voice Conversion	2760
<i>Le Xu, Jiangyan Yi, Tao Wang, Yong Ren, Rongxiu Zhong, Zhengqi Wen, Jianhua Tao</i>	
Disentangling Prosody and Timbre Embeddings Via Voice Conversion.....	2765
<i>Nicolas Gengembre, Olivier Le Blouch, Cédric Gendrot</i>	
LDM-SVC: Latent Diffusion Model Based Zero-Shot Any-To-Any Singing Voice Conversion with Singer Guidance	2770
<i>Shihao Chen, Yu Gu, Jie Zhang, Na Li, Rilun Chen, Liping Chen, Lirong Dai</i>	

SPEECH SYNTHESIS: TEXT PROCESSING

A Language Modeling Approach to Diacritic-Free Hebrew TTS	2775
<i>Amit Roth, Arnon Turetzky, Yossi Adi</i>	
Exploring the Benefits of Tokenization of Discrete Acoustic Units	2780
<i>Avihu Dekel, Raul Fernandez</i>	
Homograph Disambiguation with Text-To-Text Transfer Transformer	2785
<i>Markéta Rezáčková, Daniel Tihelka, Jindrich Matoušek</i>	
Enhancing Japanese Text-To-Speech Accuracy with a Novel Combination Transformer-BERT- Based G2P: Integrating Pronunciation Dictionaries and Accent Sandhi	2790
<i>Kiyoshi Kurihara, Masanori Sano</i>	
Audio-Conditioned Phonemic and Prosodic Annotation for Building Text-To-Speech Models from Unlabeled Speech Data.....	2795
<i>Yuma Shirahata, Byeongseon Park, Ryuichi Yamamoto, Kentaro Tachibana</i>	
G2PA: G2P with Aligned Audio for Mandarin Chinese	2800
<i>Xingxing Yang</i>	
Learning Pronunciation from Other Accents Via Pronunciation Knowledge Transfer	2805
<i>Siqi Sun, Korin Richmond</i>	
Positional Description for Numerical Normalization	2810
<i>Deepanshu Gupta, Javier Latorre</i>	
Beyond Graphemes and Phonemes: Continuous Phonological Features in Neural Text-To-Speech Synthesis.....	2815
<i>Christina Tännander, Shivam Mehta, Jonas Beskow, Jens Edlund</i>	

TRAINING METHODS, SELF-SUPERVISED LEARNING, ADAPTATION

MSRS: Training Multimodal Speech Recognition Models from Scratch with Sparse Mask Optimization	2820
<i>Adriana Fernandez-Lopez, Honglie Chen, Pingchuan Ma, Lu Yin, Qiao Xiao, Stavros Petridis, Shiwei Liu, Maja Pantic</i>	
Speech and Language Recognition with Low-Rank Adaptation of Pretrained Models.....	2825
<i>Amrutha Prasad, Srikanth Madikeri, Driss Khalil, Petr Motlicek, Christof Schuepbach</i>	
Convolution-Augmented Parameter-Efficient Fine-Tuning for Speech Recognition.....	2830
<i>Kwangyoun Kim, Suwon Shon, Yi-Te Hsu, Prashant Sridhar, Karen Livescu, Shinji Watanabe</i>	
LASER: Learning by Aligning Self-Supervised Representations of Speech for Improving Content- Related Tasks.....	2835
<i>Amit Meghanani, Thomas Hain</i>	
Self-Train Before You Transcribe	2840
<i>Robert Flynn, Anton Ragni</i>	
Unsupervised Online Continual Learning for Automatic Speech Recognition	2845
<i>Steven Vander Eeckt, Hugo Van Hamme</i>	

Dual-Path Adaptation of Pretrained Feature Extraction Module for Robust Automatic Speech Recognition	2850
<i>Hao Shi, Tatsuya Kawahara</i>	
Hierarchical Multi-Task Learning with CTC and Recursive Operation	2855
<i>Nahomi Kusunoki, Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi</i>	
Boosting CTC-Based ASR Using Inter-Layer Attention-Based CTC Loss	2860
<i>Keigo Hojo, Yukoh Wakabayashi, Kengo Ohta, Atsunori Ogawa, Norihide Kitaoka</i>	
Self-Training ASR Guided by Unsupervised ASR Teacher	2865
<i>Hyung Yong Kim, Byeong-Yeol Kim, Yunkyu Lim, Jihwan Park, Shukjae Choi, Yooncheol Ju, Jinseok Park, Youshin Lim, Seung Woo Yu, Hanbin Lee, Shinji Watanabe</i>	
Personality-Memory Gated Adaptation: An Efficient Speaker Adaptation for Personalized End-To-End Automatic Speech Recognition	2870
<i>Yue Gu, Zhihao Du, Shiliang Zhang, Jiqing Han, Yongjun He</i>	
Speaker Personalization for Automatic Speech Recognition Using Weight-Decomposed Low-Rank Adaptation	2875
<i>George Joseph, Arun Baby</i>	
Online Subloop Search Via Uncertainty Quantization for Efficient Test-Time Adaptation	2880
<i>Jae-Hong Lee, Sang-Eon Lee, Dong-Hyun Kim, Dohee Kim, Joon-Hyuk Chang</i>	
ROAR: Reinforcing Original to Augmented Data Ratio Dynamics for Wav2vec2.0 Based ASR	2885
<i>Vishwanath Pratap Singh, Federico Malato, Ville Hautamäki, Md. Sahidullah, Tomi Kinnunen</i>	
Online Knowledge Distillation of Decoder-Only Large Language Models for Efficient Speech Recognition	2890
<i>Jeehye Lee, Hyeji Seo</i>	

NOVEL ARCHITECTURES FOR ASR

Efficient and Robust Long-Form Speech Recognition with Hybrid H3-Conformer	2895
<i>Tomoki Honda, Shinsuke Sakai, Tatsuya Kawahara</i>	
Rapid Language Adaptation for Multilingual E2E Speech Recognition Using Encoder Prompting	2900
<i>Yosuke Kashiwagi, Hayato Futami, Emiru Tsunoo, Siddhant Arora, Shinji Watanabe</i>	
Quantifying Unintended Memorization in BEST-RQ ASR Encoders	2905
<i>Virat Shejwalkar, Om Thakkar, Arun Narayanan</i>	
SWAN: SubWord Alignment Network for HMM-Free Word Timing Estimation in End-To-End Automatic Speech Recognition	2910
<i>Woo Hyun Kang, Srikanth Vishnubhotla, Rudolf Braun, Yogesh Virkar, Raghuvveer Peri, Kyu J. Han</i>	

MULTIMODALITY AND FOUNDATION MODELS

Spontaneous Speech-Based Suicide Risk Detection Using Whisper and Large Language Models	2915
<i>Ziyun Cui, Chang Lei, Wen Wu, Yinan Duan, Diyang Qu, Ji Wu, Runsen Chen, Chao Zhang</i>	
Spoken Word2Vec: Learning Skipgram Embeddings from Speech	2920
<i>Mohammad Amaan Sayeed, Hanan Aldarmaki</i>	

SAMSEMO: New Dataset for Multilingual and Multimodal Emotion Recognition.....	2925
<i>Pawel Bujnowski, Bartłomiej Kuzma, Bartłomiej Paziewski, Jacek Rutkowski, Joanna Marhula, Zuzanna Bordzicka, Piotr Andruszkiewicz</i>	

VOLUME 5

LLM-Driven Multimodal Opinion Expression Identification	2930
<i>Bonian Jia, Huiyao Chen, Yueheng Sun, Meishan Zhang, Min Zhang</i>	
Zero-Shot Fake Video Detection by Audio-Visual Consistency.....	2935
<i>Xiaolou Li, Zehua Liu, Chen Chen, Lantian Li, Li Guo, Dong Wang</i>	
Enhancing Speech-Driven 3D Facial Animation with Audio-Visual Guidance from Lip Reading Expert	2940
<i>Han Eungi, Oh Hyun-Bin, Kim Sung-Bin, Corentin Nivelet Etcheberry, Suekyeong Nam, Janghoon Ju, Tae-Hyun Oh</i>	

SPOKEN DIALOGUE SYSTEMS AND CONVERSATIONAL ANALYSIS 1

Autoregressive Cross-Interlocutor Attention Scores Meaningfully Capture Conversational Dynamics.....	2945
<i>Matthew McNeill, Rivka Levitan</i>	
ConvoCache: Smart Re-Use of Chatbot Responses	2950
<i>Conor Atkins, Ian Wood, Mohamed Ali Kaafar, Hassan Asghar, Nardine Basta, Michal Kepkowski</i>	
Joint Learning of Context and Feedback Embeddings in Spoken Dialogue.....	2955
<i>Livia Qian, Gabriel Skantze</i>	
DubWise: Video-Guided Speech Duration Control in Multimodal LLM-Based Text-To-Speech for Dubbing.....	2960
<i>Neha Sahipjohn, Ashishkumar Gudmalwar, Nirmesh Shah, Pankaj Wasnik, Rajiv Ratn Shah</i>	
Contextual Interactive Evaluation of TTS Models in Dialogue Systems	2965
<i>Siyang Wang, Éva Székely, Joakim Gustafson</i>	
GSQA: An End-To-End Model for Generative Spoken Question Answering	2970
<i>Min-Han Shih, Ho-Lam Chung, Yu-Chi Pai, Ming-Hao Hsu, Guan-Ting Lin, Shang-Wen Li, Hung-Yi Lee</i>	

SPEECH TECHNOLOGY

Resource-Efficient Speech Quality Prediction Through Quantization Aware Training and Binary Activation Maps	2975
<i>Mattias Nilsson, Riccardo Miccini, Clement Laroche, Tobias Piechowiak, Friedemann Zenke</i>	
Towards Interfacing Large Language Models with ASR Systems Using Confidence Measures and Prompting	2980
<i>Maryam Naderi, Enno Hermann, Alexandre Nanchen, Sevada Hovsepyan, Mathew Magimai.-Doss</i>	

Text Injection for Neural Contextual Biasing.....	2985
<i>Zhong Meng, Zelin Wu, Rohit Prabhavalkar, Cal Peyser, Weiran Wang, Nanxin Chen, Tara N. Sainath, Bhuvana Ramabhadran</i>	
Prompting Large Language Models with Mispronunciation Detection and Diagnosis Abilities.....	2990
<i>Minglin Wu, Jing Xu, Xixin Wu, Helen Meng</i>	
Acceleration of Posteriorgram-Based DTW by Distilling the Class-To-Class Distances Encoded in the Classifier Used to Calculate Posteriors.....	2995
<i>Haitong Sun, Jaehyun Choi, Nobuaki Minematsu, Daisuke Saito</i>	
VECL-TTS: Voice Identity and Emotional Style Controllable Cross-Lingual Text-To-Speech.....	3000
<i>Ashishkumar Gudmalwar, Nirmesh Shah, Sai Akarsh, Pankaj Wasnik, Rajiv Ratn Shah</i>	
Transferable Speech-To-Text Large Language Model Alignment Module	3005
<i>Boyong Wu, Chao Yan, Haoran Pu</i>	
Sparse Binarization for Fast Keyword Spotting	3010
<i>Jonathan Svirsky, Uri Shaham, Ofir Lindenbaum</i>	

PATHOLOGICAL SPEECH ANALYSIS 2

Quantifying the Effect of Speech Pathology on Automatic and Human Speaker Verification	3015
<i>Bence Mark Halpern, Thomas Tienkamp, Wen-Chin Huang, Lester Phillip Violeta, Teja Rebernik, Sebastiaan De Visscher, Max Witjes, Martijn Wieling, Defne Abur, Tomoki Toda</i>	
Investigation of Layer-Wise Speech Representations in Self-Supervised Learning Models: A Cross-Lingual Study in Detecting Depression	3020
<i>Bubai Maji, Rajlakshmi Guha, Aurobinda Routray, Shazia Nasreen, Debabrata Majumdar</i>	
Detection of Cognitive Impairment and Alzheimer's Disease Using a Speech- And Language-Based Protocol	3025
<i>Tanya Talkar, Sherman Charles, Chelsea Krantsevich, Kan Kawabata</i>	
Analyzing Multimodal Features of Spontaneous Voice Assistant Commands for Mild Cognitive Impairment Detection.....	3030
<i>Nana Lin, Youxiang Zhu, Xiaohui Liang, John A. Batsis, Caroline Summerour</i>	
Prosody-Driven Privacy-Preserving Dementia Detection	3035
<i>Dominika Woszczyk, Ranya Aloufi, Soteris Demetriou</i>	
Voice Disorder Analysis: A Transformer-Based Approach.....	3040
<i>Alkis Koudounas, Gabriele Ciravegna, Marco Fantini, Erika Crosetti, Giovanni Succo, Tania Cerquitelli, Elena Baralis</i>	

SPEECH SCIENCE, SPEECH TECHNOLOGY, AND GENDER (SPECIAL SESSION)

Challenges of German Speech Recognition: A Study on Multi-Ethnolectal Speech Among Adolescents	3045
<i>Martha Schubert, Daniel Duran, Ingo Siegert</i>	
Just Because We Camp, Doesn't Mean We Should: The Ethics of Modelling Queer Voices.	3050
<i>Atli Sigurgeirsson, Eddie L. Ungless</i>	

Automatic Classification of News Subjects in Broadcast News: Application to a Gender Bias Representation Analysis	3055
<i>Valentin Pelloin, Léna Dodson, Émile Chapuis, Nicolas Hervé, David Doukhan</i>	
Gender Representation in TV and Radio: Automatic Information Extraction Methods Versus Manual Analyses	3060
<i>David Doukhan, Lena Dodson, Manon Conan, Valentin Pelloin, Aurélien Clamouse, Mélina Lepape, Géraldine Van Hille, Cécile Méadel, Marlène Coulomb-Gully</i>	
Acoustic Effects of Facial Feminisation Surgery on Speech and Singing: A Case Study	3065
<i>Clíodhna Hughes, Guy Brown, Ning Ma, Nicola Dibben</i>	
An Inclusive Approach to Creating a Palette of Synthetic Voices for Gender Diversity	3070
<i>Eva Szekely, Maxwell Hope</i>	
Speech After Gender: A Trans-Feminine Perspective on Next Steps for Speech Science and Technology	3075
<i>Robin Netzorg, Alyssa Cote, Sumi Koshin, Klo Vivienne Garoute, Gopala Krishna Anumanchipalli</i>	
Voice Quality Variation in AAE: An Additional Challenge for Addressing Bias in ASR Models?	3080
<i>Li-Fang Lai, Nicole Holliday</i>	
Articulatory Configurations Across Genders and Periods in French Radio and TV Archives	3085
<i>Benjamin Elie, David Doukhan, Rémi Uro, Lucas Ondel-Yang, Albert Rilliard, Simon Devauchelle</i>	
On the Encoding of Gender in Transformer-Based ASR Representations	3090
<i>Aravind Krishnan, Badr M. Abdullah, Dietrich Klakow</i>	

SPEECH PRODUCTION AND PERCEPTION

Towards a Quantitative Analysis of Coarticulation with a Phoneme-To-Articulatory Model	3095
<i>Chaofei Fan, Jaimie M. Henderson, Chris Manning, Francis R. Willett</i>	
A Comparative Study of the Impact of Voiceless Alveolar and Palato-Alveolar Sibilants in English on Lip Aperture and Protrusion During VCV Production	3100
<i>Chetan Sharma, Vaishnavi Chandwanshi, Prasanta Kumar Ghosh</i>	
Measurement and Simulation of Pressure Losses Due to Airflow in Vocal Tract Models.....	3105
<i>Peter Birkholz, Patrick Häsner</i>	
On the Performance of EMA-Synchronized Speech and Stand-Alone Speech in Acoustic-To-Articulatory Inversion	3110
<i>Qiang Fang</i>	
Glottal Inverse Filtering and Vocal Tract Tuning for the Numerical Simulation of Vowel /A/ with Different Levels of Vocal Effort	3115
<i>Marc Freixes, Marc Arnela, Joan Claudi Socoró, Luis Joglar-Ongay, Oriol Guasch, Francesc Alias-Pujol</i>	
Temporal Co-Registration of Simultaneous Electromagnetic Articulography and Electroencephalography for Precise Articulatory and Neural Data Alignment	3120
<i>Daniel Friedrichs, Monica Lancheros, Sam Kirkham, Lei He, Andrew Clark, Clemens Lutz, Volker Dellwo, Steven Moran</i>	

PHONETICS AND PHONOLOGY: SEGMENTALS AND SUPRASEGMENTALS

Frication Noise Features of Polish Voiceless Dental Fricative and Affricate Produced by Children with and Without Speech Disorder	3125
<i>Zuzanna Miodonska, Michal Krecichwost, Ewa Kwasniok, Agata Sage, Pawel Badura</i>	
Key Acoustic Cues for the Realization of Metrical Prominence in Tone Languages: A Cross-Dialect Study.....	3130
<i>Yiying Hu, Hui Feng</i>	
Revisiting Pitch Jumps: F0 Ratio in Seoul Korean.....	3135
<i>Michaela Watkins, Paul Boersma, Silke Hamann</i>	
Aerodynamics of Sakata Labial-Velar Oral Stops	3140
<i>Lorenzo Maselli, Véronique Delvaux</i>	
Collecting Mandible Movement in Brazilian Portuguese.....	3145
<i>Donna Erickson, Albert Rilliard, Malin Svensson Lundmark, Adelaide Silva, Leticia Rebollo Couto, Oliver Niebuhr, João Antonio De Moraes</i>	
Pitch-Driven Adjustments in Tongue Positions: Insights from Ultrasound Imaging.....	3150
<i>May Pik Yu Chan, Jianjing Kuang</i>	

TOPICS IN PARALINGUISTICS

Speaking of Health: Leveraging Large Language Models to Assess Exercise Motivation and Behavior of Rehabilitation Patients	3155
<i>Suhas Bn, Amanda Rebar, Saeed Abdullah</i>	
Confidence Estimation for Automatic Detection of Depression and Alzheimer's Disease Based on Clinical Interviews	3160
<i>Wen Wu, Chao Zhang, Philip C. Woodland</i>	
Who Finds This Voice Attractive? a Large-Scale Experiment Using In-The-Wild Data.....	3165
<i>Hitoshi Suda, Aya Watanabe, Shinnosuke Takamichi</i>	
Acoustical Analysis of the Initial Phones in Speech-Laugh	3170
<i>Ryo Setoguchi, Yoshiko Arimoto</i>	
On Calibration of Speech Classification Models: Insights from Energy-Based Model Investigations	3175
<i>Yaqian Hao, Chenguang Hu, Yingying Gao, Shilei Zhang, Junlan Feng</i>	
Emotion-Aware Speech Self-Supervised Representation Learning with Intensity Knowledge	3180
<i>Rui Liu, Zening Ma</i>	

EMOTION RECOGNITION: FAIRNESS, VARIABILITY, UNCERTAINTY

Dual-Constrained Dynamical Neural ODEs for Ambiguity-Aware Continuous Emotion Prediction	3185
<i>Jingyao Wu, Ting Dang, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i>	
An Inter-Speaker Fairness-Aware Speech Emotion Regression Framework.....	3190
<i>Hsing-Hang Chou, Woan-Shiuan Chien, Ya-Tse Wu, Chi-Chun Lee</i>	

The Whole is Bigger than the Sum of Its Parts: Modeling Individual Annotators to Capture Emotional Variability.....	3195
<i>James Tavernor, Yara El-Tawil, Emily Mower Provost</i>	
Iterative Prototype Refinement for Ambiguous Speech Emotion Recognition	3200
<i>Haoqin Sun, Shiwan Zhao, Xiangyu Kong, Xuechen Wang, Hui Wang, Jiaming Zhou, Yong Qin</i>	
An Investigation of Group Versus Individual Fairness in Perceptually Fair Speech Emotion Recognition	3205
<i>Woan-Shiuan Chien, Chi-Chun Lee</i>	
Are You Sure? Analysing Uncertainty Quantification Approaches for Real-World Speech Emotion Recognition	3210
<i>Oliver Schröder, Manuel Milling, Felix Burkhardt, Florian Eyben, Björn Schuller</i>	
Speech Emotion Recognition with Deep Learning Beamforming on a Distant Human-Robot Interaction Scenario.....	3215
<i>Ricardo García, Rodrigo Mahu, Nicolás Grágeda, Alejandro Luzanto, Nicolas Bohmer, Carlos Busso, Néstor Becerra Yoma</i>	

SPEAKER VERIFICATION

Challenging Margin-Based Speaker Embedding Extractors by Using the Variational Information Bottleneck.....	3220
<i>Themis Stafylakis, Anna Silnova, Johan Rohdin, Oldrich Plchot, Lukáš Burget</i>	
Collaborative Contrastive Learning for Hypothesis Domain Adaptation	3225
<i>Jen-Tzung Chien, I-Ping Yeh, Man-Wai Mak</i>	
Extraction of Interpretable and Shared Speaker-Specific Speech Attributes Through Binary Auto-Encoder.....	3230
<i>Imen Ben-Amor, Jean-Francois Bonastre, Salima Mdhaffar</i>	
Reshape Dimensions Network for Speaker Recognition.....	3235
<i>Ivan Yakovlev, Rostislav Makarov, Andrei Balykin, Pavel Malov, Anton Okhotnikov, Nikita Torgashov</i>	
To What Extent Can ASV Systems Naturally Defend Against Spoofing Attacks?.....	3240
<i>Jee-Weon Jung, Xin Wang, Nicholas Evans, Shinji Watanabe, Hye-Jin Shim, Hemlata Tak, Siddhant Arora, Junichi Yamagishi, Joon Son Chung</i>	
ERes2NetV2: Boosting Short-Duration Speaker Verification Performance with Computational Efficiency	3245
<i>Yafeng Chen, Siqi Zheng, Hui Wang, Luyao Cheng, Qian Chen, Shiliang Zhang, Junjie Li</i>	

SPATIAL AUDIO AND ACOUSTICS

Classification of Room Impulse Responses and Its Application for Channel Verification and Diarization	3250
<i>Yuri Khokhlov, Tatiana Prisyach, Anton Mitrofanov, Dmitry Dutov, Igor Agafonov, Tatiana Timofeeva, Aleksei Romanenko, Maxim Korenevsky</i>	

RIR-In-A-Box: Estimating Room Acoustics from 3D Mesh Data Through Shoebox Approximation.....	3255
<i>Liam Kelley, Diego Di Carlo, Aditya Arie Nugraha, Mathieu Fontaine, Yoshiaki Bando, Kazuyoshi Yoshii</i>	
Novel-View Acoustic Synthesis from 3D Reconstructed Rooms.....	3260
<i>Byeongjoo Ahn, Karren Yang, Brian Hamilton, Jonathan Sheaffer, Anurag Ranjan, Miguel Sarabia, Oncel Tuzel, Jen-Hao Rick Chang</i>	
Spatial Acoustic Enhancement Using Unbiased Relative Harmonic Coefficients	3265
<i>Liang Tao, Maoshen Jia, Yonggang Hu, Changchun Bao</i>	
Design of Feedback Active Noise Cancellation Filter Using Nested Recurrent Neural Networks.....	3270
<i>Alireza Bayestehtashk, Amit Kumar, Mike Wurtz</i>	
Neuromorphic Keyword Spotting with Pulse Density Modulation MEMS Microphones	3275
<i>Sidi Yaya Arnaud Yarga, Sean U N Wood</i>	
RevRIR: Joint Reverberant Speech and Room Impulse Response Embedding Using Contrastive Learning with Application to Room Shape Classification.....	3280
<i>Jacob Bitterman, Daniel Levi, Hilel Hagai Diamandi, Sharon Gannot, Tal Rosenwein</i>	

GENERATIVE MODELS FOR SPEECH AND AUDIO

ConsistencyTTA: Accelerating Diffusion-Based Text-To-Audio Generation with Consistency Distillation.....	3285
<i>Yatong Bai, Trung Dang, Dung Tran, Kazuhito Koishida, Somayeh Sojoudi</i>	
Audio Editing with Non-Rigid Text Prompts	3290
<i>Francesco Paissan, Luca Della Libera, Zhepei Wang, Paris Smaragdis, Mirco Ravanelli, Cem Subakan</i>	
Phoneme Discretized Saliency Maps for Explainable Detection of AI-Generated Voice.....	3295
<i>Shubham Gupta, Mirco Ravanelli, Pascal Germain, Cem Subakan</i>	
Exploring Compressibility of Transformer Based Text-To-Music (TTM) Models	3300
<i>Vasileios Moschopoulos, Thanasis Kotsiopoulos, Pablo Peso Parada, Konstantinos Nikiforidis, Alexandros Stergiadis, Gerasimos Papakostas, Md Asif Jalal, Jisi Zhang, Anastasios Drosou, Karthikeyan Saravanan</i>	
Sound of Vision: Audio Generation from Visual Text Embedding Through Training Domain Discriminator.....	3305
<i>Jaewon Kim, Won-Gook Choi, Seyun Ahn, Joon-Hyuk Chang</i>	
Retrieval-Augmented Classifier Guidance for Audio Generation	3310
<i>Ho-Young Choi, Won-Gook Choi, Joon-Hyuk Chang</i>	
Efficient Fine-Tuning of Audio Spectrogram Transformers Via Soft Mixture of Adapters.....	3315
<i>Umberto Cappellazzo, Daniele Falavigna, Alessio Brutti</i>	
PAM: Prompting Audio-Language Models for Audio Quality Assessment	3320
<i>Soham Deshmukh, Dareen Alharthi, Benjamin Elizalde, Hannes Gamper, Mahmoud Al Ismail, Rita Singh, Bhiksha Raj, Huaming Wang</i>	

SPEECH AND AUDIO MODELLING

GenDistiller: Distilling Pre-Trained Language Models Based on an Autoregressive Generative Model	3325
<i>Yingying Gao, Shilei Zhang, Chao Deng, Junlan Feng</i>	
Gender and Language Identification in Multilingual Models of Speech: Exploring the Genericity and Robustness of Speech Representations.....	3330
<i>S��verine Guillaume, Maxime Fily, Alexis Michaud, Guillaume Wisniewski</i>	
Neural Compression Augmentation for Contrastive Audio Representation Learning.....	3335
<i>Zhaoyu Wang, Haohe Liu, Harry Coppock, Bj��rn Schuller, Mark D. Plumbley</i>	
Post-Net: A Linguistically Inspired Sequence-Dependent Transformed Neural Architecture for Automatic Syllable Stress Detection	3340
<i>Sai Harshitha Aluru, Jhansi Mallela, Chiranjeevi Yarra</i>	

MULTI-CHANNEL SPEECH ENHANCEMENT

Array Geometry-Robust Attention-Based Neural Beamformer for Moving Speakers	3345
<i>Marvin Tammen, Tsubasa Ochiai, Marc Delcroix, Tomohiro Nakatani, Shoko Araki, Simon Doclo</i>	
FoVNet: Configurable Field-Of-View Speech Enhancement with Low Computation and Distortion for Smart Glasses.....	3350
<i>Zhongweiyang Xu, Ali Aroudi, Ke Tan, Ashutosh Pandey, Jung-Suk Lee, Buye Xu, Francesco Nesta</i>	
Audio Enhancement from Multiple Crowdsourced Recordings: A Simple and Effective Baseline	3355
<i>Shiran Aziz, Yossi Adi, Shmuel Peleg</i>	
DeFTAN-AA: Array Geometry Agnostic Multichannel Speech Enhancement.....	3360
<i>Dongheon Lee, Jung-Woo Choi</i>	
SA-MF: A Novel Self-Attention Mechanism for Multifeature Fusion in Speech Enhancement Networks	3365
<i>Ruizhe Wang</i>	
PLDNet: PLD-Guided Lightweight Deep Network Boosted by Efficient Attention for Handheld Dual-Microphone Speech Enhancement	3370
<i>Nan Zhou, Youhai Jiang, Jialin Tan, Chongmin Qi</i>	

SPEECH SYNTHESIS: PARADIGMS AND METHODS 1

Highly Intelligible Speaker-Independent Articulatory Synthesis	3375
<i>Charles McGhee, Kate Knill, Mark Gales</i>	
An Attribute Interpolation Method in Speech Synthesis by Model Merging	3380
<i>Masato Murata, Koichi Miyazaki, Tomoki Koriyama</i>	
Low-Dimensional Style Token Control for Hyperarticulated Speech Synthesis	3385
<i>Miku Nishihara, Dan Wells, Korin Richmond, Aidan Pine</i>	

Single-Codec: Single-Codebook Speech Codec Towards High-Performance Speech Generation.....	3390
<i>Hanzhao Li, Liumeng Xue, Haohan Guo, Xinfu Zhu, Yuanjun Lv, Lei Xie, Yunlin Chen, Hao Yin, Zhifei Li</i>	
LiveSpeech: Low-Latency Zero-Shot Text-To-Speech Via Autoregressive Modeling of Audio Discrete Codes.....	3395
<i>Trung Dang, David Aponte, Dung Tran, Kazuhito Koishida</i>	
ClariTTS: Feature-Ratio Normalization and Duration Stabilization for Code-Mixed Multi-Speaker Speech Synthesis	3400
<i>Changhwan Kim</i>	
Multi-Modal Adversarial Training for Zero-Shot Voice Cloning	3405
<i>John Janiczek, Dading Chong, Dongyang Dai, Arlo Faria, Chao Wang, Tao Wang, Yuzong Liu</i>	
Learning Fine-Grained Controllability on Speech Generation Via Efficient Fine-Tuning	3410
<i>Chung-Ming Chien, Andros Tjandra, Apoorv Vyas, Matt Le, Bowen Shi, Wei-Ning Hsu</i>	
Modeling Vocal Tract Like Acoustic Tubes Using the Immersed Boundary Method.....	3415
<i>Rongshuai Wu, Debasish Ray Mohapatra, Sidney Fels</i>	

SPEECH SYNTHESIS: PARADIGMS AND METHODS 2

Small-E: Small Language Model with Linear Attention for Efficient Speech Synthesis	3420
<i>Théodor Lemerle, Nicolas Obin, Axel Roebel</i>	
Improving Robustness of LLM-Based Speech Synthesis by Learning Monotonic Alignment	3425
<i>Paarth Neekhara, Shehzeen Hussain, Subhankar Ghosh, Jason Li, Boris Ginsburg</i>	
Synthesizing Long-Form Speech Merely from Sentence-Level Corpus with Content Extrapolation and LLM Contextual Enrichment.....	3430
<i>Shijie Lai, Minglu He, Zijing Zhao, Kai Wang, Hao Huang, Jichen Yang</i>	
FluentEditor: Text-Based Speech Editing by Considering Acoustic and Prosody Consistency	3435
<i>Rui Liu, Jiatian Xi, Ziyue Jiang, Haizhou Li</i>	
Phonetic Enhanced Language Modeling for Text-To-Speech Synthesis	3440
<i>Kun Zhou, Shengkui Zhao, Yukun Ma, Chong Zhang, Hao Wang, Dianwen Ng, Chongjia Ni, Trung Hieu Nguyen, Jia Qi Yip, Bin Ma</i>	
High Fidelity Text-To-Speech Via Discrete Tokens Using Token Transducer and Group Masked Language Model.....	3445
<i>Joun Yeop Lee, Myeonghun Jeong, Minchan Kim, Ji-Hyun Lee, Hoon-Young Cho, Nam Soo Kim</i>	
FastLips: An End-To-End Audiovisual Text-To-Speech System with Lip Features Prediction for Virtual Avatars	3450
<i>Martin Lenglet, Olivier Perrotin, Gerard Bailly</i>	

NEURAL NETWORK ARCHITECTURES FOR ASR 1

Contemplative Mechanism for Speech Recognition: Speech Encoders Can Think	3455
<i>Tien-Ju Yang, Andrew Rosenberg, Bhuvana Ramabhadran</i>	

SummaryMixing: A Linear-Complexity Alternative to Self-Attention for Speech Recognition and Understanding	3460
<i>Titouan Parcollet, Rogier Van Dalen, Shucong Zhang, Sourav Bhattacharya</i>	
Boosting Hybrid Autoregressive Transducer-Based ASR with Internal Acoustic Model Training and Dual Blank Thresholding.....	3465
<i>Takafumi Moriya, Takanori Ashihara, Masato Mimura, Hiroshi Sato, Kohei Matsuura, Ryo Masumura, Taichi Asami</i>	
RepCNN: Micro-Sized, Mighty Models for Wakeword Detection	3470
<i>Arnav Kundu, Prateeth Nayak, Priyanka Padmanabhan, Devang Naik</i>	
Conformer Without Convolutions	3475
<i>Matthijs Van Keirsbilck, Alexander Keller</i>	
Linear-Complexity Self-Supervised Learning for Speech Processing.....	3480
<i>Shucong Zhang, Titouan Parcollet, Rogier Van Dalen, Sourav Bhattacharya</i>	

ERROR CORRECTION AND RESCORING

SALSA: Speedy ASR-LLM Synchronous Aggregation.....	3485
<i>Ashish Mittal, Darshan Prabhu, Sunita Sarawagi, Preethi Jyothi</i>	
LI-TTA: Language Informed Test-Time Adaptation for Automatic Speech Recognition	3490
<i>Eunseop Yoon, Hee Suk Yoon, John Harvill, Mark Hasegawa-Johnson, Chang D. Yoo</i>	
HypR: A Comprehensive Study for ASR Hypothesis Revising with a Reference Corpus	3495
<i>Yi-Wei Wang, Ke-Han Lu, Kuan-Yu Chen</i>	
Error Correction by Paying Attention to Both Acoustic and Confidence References for Automatic Speech Recognition	3500
<i>Yuchun Shu, Bo Hu, Yifeng He, Hao Shi, Longbiao Wang, Jianwu Dang</i>	
Transformer-Based Model for ASR N-Best Rescoring and Rewriting	3505
<i>Iwen E Kang, Christophe Van Gysel, Man-Hung Siu</i>	
RASU: Retrieval Augmented Speech Understanding Through Generative Modeling	3510
<i>Hao Yang, Min Zhang, Minghan Wang, Jiaxin Guo</i>	

SPOKEN LANGUAGE UNDERSTANDING

Towards Unified Evaluation of Continual Learning in Spoken Language Understanding.....	3515
<i>Muqiao Yang, Xiang Li, Umberto Cappellazzo, Shinji Watanabe, Bhiksha Raj</i>	
Convolutional Gated MLP and Attention Improve End-To-End Spoken Language Understanding	3520
<i>Beida Zheng, Mijit Ablimit, Hankiz Yilahun, Askar Hamdulla</i>	
An Uyghur Extension to the MASSIVE Multi-Lingual Spoken Language Understanding Corpus with Comprehensive Evaluations	3525
<i>Ainikaerjiang Aimaiti, Di Wu, Liting Jiang, Gulnigeer Abudouwaili, Hao Huang, Wushour Silamu</i>	

This Paper Had the Smartest Reviewers - Flattery Detection Utilising an Audio-Textual Transformer-Based Approach.....	3530
<i>Lukas Christ, Shahin Amiriparian, Friederike Hawighorst, Ann-Kathrin Schill, Angelo Boutalidakis, Lorenz Graf-Vlachy, Andreas König, Björn Schuller</i>	
Unified Framework for Spoken Language Understanding and Summarization in Task-Based Human Dialog Processing	3535
<i>Eunice Akani, Frederic Bechet, Benoît Favre, Romain Gemignani</i>	
Automated Human-Readable Label Generation in Open Intent Discovery.....	3540
<i>Grant Anderson, Emma Hart, Dimitra Gkatzia, Ian Beaver</i>	
Applying Reinforcement Learning and Multi-Generators for Stage Transition in an Emotional Support Dialogue System	3545
<i>Jeremy Chang, Kuan-Yu Chen, Chung-Hsien Wu</i>	

SPOKEN DIALOGUE SYSTEMS AND CONVERSATIONAL ANALYSIS 2

Target Conversation Extraction: Source Separation Using Turn-Taking Dynamics.....	3550
<i>Tuochao Chen, Qirui Wang, Bohan Wu, Malek Itani, Emre Sefik Eskimez, Takuya Yoshioka, Shyamnath Gollakota</i>	
Investigating the Influence of Stance-Taking on Conversational Timing of Task-Oriented Speech	3555
<i>Sara Ng, Gina-Anne Levow, Mari Ostendorf, Richard Wright</i>	
Detecting the Terminality of Speech-Turn Boundary for Spoken Interactions in French TV and Radio Content.....	3560
<i>Rémi Uro, Marie Tahon, David Doukhan, Antoine Laurent, Albert Rilliard</i>	
Utilization of Text Data for Response Timing Detection in Attentive Listening.....	3565
<i>Yu Watanabe, Koichiro Ito, Shigeki Matsubara</i>	
Backchannel Prediction, Based on Who, When and What	3570
<i>Yo-Han Park, Wencke Liermann, Yong-Seok Choi, Seung Hi Kim, Jeong-Uk Bang, Seung Yun, Kong Joo Lee</i>	
Uh, Um and Mh: Are Filled Pauses Prone to Conversational Converge?	3575
<i>Mathilde Hutin, Junfei Hu, Liesbeth Degand</i>	
Investigation of Look-Ahead Techniques to Improve Response Time in Spoken Dialogue System.....	3580
<i>Masaya Ohagi, Tomoya Mizumoto, Katsumasa Yoshikawa</i>	

COMPUTATIONAL MODELS OF HUMAN LANGUAGE ACQUISITION, PERCEPTION, AND PRODUCTION (SPECIAL SESSION)

Information-Theoretic Hypothesis Generation of Relative Cue Weighting for the Voicing Contrast.....	3585
<i>Annika Heuser, Jianjing Kuang</i>	
Neurocomputational Model of Speech Recognition for Pathological Speech Detection: A Case Study on Parkinson's Disease Speech Detection	3590
<i>Sevada Hovsepyan, Mathew Magimai.-Doss</i>	
Simulating Articulatory Trajectories with Phonological Feature Interpolation	3595
<i>Angelo Ortiz Tandazo, Thomas Schatz, Thomas Hueber, Emmanuel Dupoux</i>	

A Pilot Study of GSLM-Based Simulation of Foreign Accentuation Only Using Native Speech Corpora.....	3600
<i>Kentaro Onda, Joonyong Park, Nobuaki Minematsu, Daisuke Saito</i>	
Dirichlet Process Mixture Model Based on Topologically Augmented Signal Representation for Clustering Infant Vocalizations.....	3605
<i>Guillem Bonafos, Clara Bourot, Pierre Pudlo, Jean-Marc Freyermuth, Laurence Reboul, Samuel Tronçon, Arnaud Rey</i>	
A Data-Driven Model of Acoustic Speech Intelligibility for Optimization-Based Models of Speech Production	3610
<i>Benjamin Elie, Juraj Simko, Alice Turk</i>	
The Difficulty and Importance of Estimating the Lower and Upper Bounds of Infant Speech Exposure.....	3615
<i>Joseph Coffey, Okko Räsänen, Camila Scaff, Alejandrina Cristia</i>	
Spoken-Term Discovery Using Discrete Speech Units	3620
<i>Benjamin Van Niekerk, Julian Zaidi, Marc-André Carbonneau, Herman Kamper</i>	
Orthogonality and Isotropy of Speaker and Phonetic Information in Self-Supervised Speech Representations	3625
<i>Mukhtar Mohamed, Oli Danyi Liu, Hao Tang, Sharon Goldwater</i>	

SHOW AND TELL 3

OCEAN-AI: Open Multimodal Framework for Personality Traits Assessment and HR-Processes Automatization	3630
<i>Elena Ryumina, Dmitry Ryumin, Alexey Karpov</i>	
VoxMed: One-Step Respiratory Disease Classifier Using Digital Stethoscope Sounds.....	3632
<i>Paridhi Mundra, Manik Sharma, Yashwardhan Chaudhuri, Orchid Chetia Phukan, Arun Balaji Buduru</i>	
AVR: Synergizing Foundation Models for Audio-Visual Humor Detection	3634
<i>Sarthak Sharma, Orchid Chetia Phukan, Drishti Singh, Arun Balaji Buduru, Rajesh Sharma</i>	
ASGIR: Audio Spectrogram Transformer Guided Classification and Information Retrieval for Birds	3636
<i>Yashwardhan Chaudhuri, Paridhi Mundra, Arnesh Batra, Orchid Chetia Phukan, Arun Balaji Buduru</i>	
PERSONA: An Application for Emotion Recognition, Gender Recognition and Age Estimation	3638
<i>Devyani Koshal, Orchid Chetia Phukan, Sarthak Jain, Arun Balaji Buduru, Rajesh Sharma</i>	
NeuRO: An Application for Code-Switched Autism Detection in Children.....	3640
<i>Mohd Mujtaba Akhtar, Girish, Orchid Chetia Phukan, Muskaan Singh</i>	
ComFeAT: Combination of Neural and Spectral Features for Improved Depression Detection	3642
<i>Orchid Chetia Phukan, Sarthak Jain, Shubham Singh, Muskaan Singh, Arun Balaji Buduru, Rajesh Sharma</i>	
The Reasonable Effectiveness of Speaker Embeddings for Violence Detection	3644
<i>Sarthak Jain, Orchid Chetia Phukan, Arun Balaji Buduru, Rajesh Sharma</i>	
ATTEST: An Analytics Tool for the Testing and Evaluation of Speech Technologies	3646
<i>Dmitrii Obukhov, Marcel De Korte, Andrey Adaschik</i>	

PhoneViz: Exploring Alignments at a Glance	3648
<i>Margot Masson, Erfan A. Shams, Iona Gessinger, Julie Carson-Berndsen</i>	
Gryannote Open-Source Speaker Diarization Labeling Tool	3650
<i>Clément Pages, Hervé Bredin</i>	
A Toolkit for Joint Speaker Diarization and Identification with Application to Speaker-Attributed ASR	3652
<i>Giovanni Morrone, Enrico Zovato, Fabio Brugnara, Enrico Sartori, Leonardo Badino</i>	

PHONETICS, PHONOLOGY AND PROSODY

Speaker Detection by the Individual Listener and the Crowd: Parametric Models Applicable to Bonafide and Deepfake Speech	3654
<i>Tomi H. Kinnunen, Rosa Gonzalez Hautamäki, Xin Wang, Junichi Yamagishi</i>	
NumberLie: A Game-Based Experiment to Understand the Acoustics of Deception and Truthfulness	3659
<i>Alessandro De Luca, Andrew Clark, Volker Dellwo</i>	
Preservation, Conservation and Phonetic Study of the Voices of Italian Poets: A Study on the Seven Years of the VIP Archive	3664
<i>Federico Lo Iacono, Valentina Colonna, Antonio Romano</i>	
Do Speaker-Dependent Vowel Characteristics Depend on Speech Style?.....	3669
<i>Nicolas Audibert, Cecile Fougeron, Christine Meunier</i>	
A Comparison of Voice Similarity Through Acoustics, Human Perception and Deep Neural Network (DNN) Speaker Verification Systems	3674
<i>Suyuan Liu, Molly Babel, Jian Zhu</i>	
Evaluating Italian Vowel Variation with the Recurrent Neural Network Phonet.....	3679
<i>Austin Jones, Margaret E. L. Renwick</i>	
Prosodic Marking of Syntactic Boundaries in Khoekhoe.....	3684
<i>Kira Tulchynska, Sylvanus Job, Alena Witzlack-Makarevich, Margaret Zellers</i>	

SEGMENTALS

Crosslinguistic Comparison of Acoustic Variation in the Vowel Sequences /ia/ and /io/ in Four Romance Languages.....	3689
<i>Johanna Cronenberg, Ioana Chitoran, Lori Lamel, Ioana Vasilescu</i>	

VOLUME 6

Nasal Air Flow During Speech Production in Korebaju.....	3694
<i>Jenifer Vega Rodriguez, Nathalie Vallée, Christophe Savariaux, Silvain Gerber</i>	
Affricates in Lushootseed.....	3699
<i>Ted Kye</i>	
Voiced and Voiceless Laterals in Angami.....	3704
<i>Viyazonuo Terhiija, Priyankoo Sarmah</i>	

Intrusive Schwa Within French Stop-Liquid Clusters: An Acoustic Analysis	3709
<i>Minmin Yang, Rachid Ridouane</i>	

NEW AVENUES IN EMOTION RECOGNITION

Can Modelling Inter-Rater Ambiguity Lead to Noise-Robust Continuous Emotion Predictions?	3714
<i>Ya-Tse Wu, Jingyao Wu, Vidhyasaharan Sethu, Chi-Chun Lee</i>	
MFDR: Multiple-Stage Fusion and Dynamically Refined Network for Multimodal Emotion Recognition	3719
<i>Ziping Zhao, Tian Gao, Haishuai Wang, Björn Schuller</i>	
Multimodal Fusion of Music Theory-Inspired and Self-Supervised Representations for Improved Emotion Recognition.....	3724
<i>Xiaohan Shi, Xingfeng Li, Tomoki Toda</i>	
Enrolment-Based Personalisation for Improving Individual-Level Fairness in Speech Emotion Recognition	3729
<i>Andreas Triantafyllopoulos, Björn Schuller</i>	
Keep, Delete, Or Substitute: Frame Selection Strategy for Noise-Robust Speech Emotion Recognition	3734
<i>Seong-Gyun Leem, Daniel Fulford, Jukka-Pekka Onnela, David Gard, Carlos Busso</i>	
Hierarchical Distribution Adaptation for Unsupervised Cross-Corpus Speech Emotion Recognition	3739
<i>Cheng Lu, Yuan Zong, Yan Zhao, Hailun Lian, Tianhua Qi, Björn Schuller, Wenming Zheng</i>	

SPEAKER DIARIZATION 2

Variable Segment Length and Domain-Adapted Feature Optimization for Speaker Diarization	3744
<i>Chenyuan Zhang, Linkai Luo, Hong Peng, Wei Wen</i>	
Efficient Speaker Embedding Extraction Using a Twofold Sliding Window Algorithm for Speaker Diarization.....	3749
<i>Jeong-Hwan Choi, Ye-Rin Jeoung, Ilseok Kim, Joon-Hyuk Chang</i>	
DiarizationLM: Speaker Diarization Post-Processing with Large Language Models	3754
<i>Quan Wang, Yiling Huang, Guanlong Zhao, Evan Clark, Wei Xia, Hank Liao</i>	
Joint Vs Sequential Speaker-Role Detection and Automatic Speech Recognition for Air-Traffic Control.....	3759
<i>Alexander Blatt, Aravind Krishnan, Dietrich Klakow</i>	
On the Calibration of Powerset Speaker Diarization Models	3764
<i>Alexis Plaquet, Hervé Bredin</i>	
Specializing Self-Supervised Speech Representations for Speaker Segmentation	3769
<i>Séverin Baroudi, Thomas Pellegrini, Hervé Bredin</i>	

SPEAKER RECOGNITION 2

On the Usefulness of Speaker Embeddings for Speaker Retrieval in the Wild: A Comparative Study of X-Vector and ECAPA-TDNN Models	3774
<i>Erfan Loweimi, Mengjie Qian, Kate Knill, Mark Gales</i>	

W-GVKT: Within-Global-View Knowledge Transfer for Speaker Verification.....	3779
<i>Zezhong Jin, Youzhi Tu, Man-Wai Mak</i>	
CEC: A Noisy Label Detection Method for Speaker Recognition	3784
<i>Yao Shen, Yingying Gao, Yaqian Hao, Chenguang Hu, Fulin Zhang, Junlan Feng, Shilei Zhang</i>	
Disentangling Age and Identity with a Mutual Information Minimization for Cross-Age Speaker Verification	3789
<i>Fengrun Zhang, Wangjin Zhou, Yiming Liu, Wang Geng, Yahui Shan, Chen Zhang</i>	
Contrastive Learning and Inter-Speaker Distribution Alignment Based Unsupervised Domain Adaptation for Robust Speaker Verification	3794
<i>Zuoliang Li, Wu Guo, Bin Gu, Shengyu Peng, Jie Zhang</i>	
Identifying Speakers in Dialogue Transcripts: A Text-Based Approach Using Pretrained Language Models	3799
<i>Minh Nguyen, Franck DERNONCOURT, Seunghyun Yoon, Hanieh Deilamsalehy, Hao Tan, Ryan Rossi, Quan Hung Tran, Trung Bui, Thien Huu Nguyen</i>	
Attention-Augmented X-Vectors for the Evaluation of Mimicked Speech Using Sparse Autoencoder-LSTM Framework	3804
<i>Bhasi K. C., Rajeev Rajan, Noumida A</i>	

SPEECH AND AUDIO ANALYSIS

Predefined Prototypes for Intra-Class Separation and Disentanglement	3809
<i>Antonio Almudévar, Théo Mariotte, Alfonso Ortega, Marie Tahon, Luis Vicente, Antonio Miguel, Eduardo Lleida</i>	
VAE-Based Phoneme Alignment Using Gradient Annealing and SSL Acoustic Features	3814
<i>Tomoki Koriyama</i>	
A Transformer-Based Voice Activity Detector	3819
<i>Biswajit Karan, Joshua Jansen Van Vuren, Febe De Wet, Thomas Niesler</i>	
XANE: eXplainable Acoustic Neural Embeddings	3824
<i>Sri Harsha Dumpala, Dushyant Sharma, Chandramouli Shama Sastry, Stanislav Kruchinin, James Fosburgh, Patrick A. Naylor</i>	
A Comparative Analysis of Sequential Models that Integrate Syllable Dependency for Automatic Syllable Stress Detection	3829
<i>Jhansi Mallela, Sai Harshitha Aluru, Chiranjeevi Yarra</i>	
Motion Based Audio-Visual Segmentation	3834
<i>Jiahao Li, Miao Liu, Shu Yang, Jing Wang, Xiang Xie</i>	

SPEECH QUALITY AND INTELLIGIBILITY: PREDICTION AND ENHANCEMENT

Transfer Learning from Whisper for Microscopic Intelligibility Prediction	3839
<i>Paul Best, Santiago Cuervo, Ricard Marxer</i>	
Non-Intrusive Speech Intelligibility Prediction for Hearing Aids Using Whisper and Metadata	3844
<i>Ryandhimas E. Zezario, Fei Chen, Chiou-Shann Fuh, Hsin-Min Wang, Yu Tsao</i>	

No-Reference Speech Intelligibility Prediction Leveraging a Noisy-Speech ASR Pre-Trained Model	3849
<i>Haolan Wang, Amin Edraki, Wai-Yip Chan, Iván López-Espejo, Jesper Jensen</i>	
The PESQetarian: On the Relevance of Goodhart's Law for Speech Enhancement.....	3854
<i>Danilo De Oliveira, Simon Welker, Julius Richter, Timo Gerkmann</i>	
Enhancing Non-Matching Reference Speech Quality Assessment Through Dynamic Weight Adaptation	3859
<i>Bao Thang Ta, Van Hai Do, Huynh Thi Thanh Binh</i>	
Exploring Sentence Type Effects on the Lombard Effect and Intelligibility Enhancement: A Comparative Study of Natural and Grid Sentences	3864
<i>Hongyang Chen, Yuhong Yang, Zhongyuan Wang, Weiping Tu, Haojun Ai, Cedar Lin</i>	

SPEECH SYNTHESIS: VOCODERS

FreeV: Free Lunch for Vocoders Through Pseudo Inversed Mel Filter	3869
<i>Yuanjun Lv, Hai Li, Ying Yan, Junhui Liu, Danming Xie, Lei Xie</i>	
QGAN: Low Footprint Quaternion Neural Vocoder for Speech Synthesis	3874
<i>Aryan Chaudhary, Vinayak Abrol</i>	
JenGAN: Stacked Shifted Filters in GAN-Based Speech Synthesis	3879
<i>Hyunjae Cho, Junhyeok Lee, Wonbin Jung</i>	
FA-GAN: Artifacts-Free and Phase-Aware High-Fidelity GAN-Based Vocoder	3884
<i>Rubing Shen, Yanzhen Ren, Zongkun Sun</i>	
QHM-GAN: Neural Vocoder Based on Quasi-Harmonic Modeling	3889
<i>Shaowen Chen, Tomoki Toda</i>	
BiVocoder: A Bidirectional Neural Vocoder Integrating Feature Extraction and Waveform Generation	3894
<i>Hui-Peng Du, Ye-Xin Lu, Yang Ai, Zhen-Hua Ling</i>	

ASR MODEL TRAINING METHODS

Investigating the Effect of Label Topology and Training Criterion on ASR Performance and Alignment Quality	3899
<i>Tina Raissi, Christoph Lüscher, Simon Berger, Ralf Schlüter, Hermann Ney</i>	
ASTRA: Aligning Speech and Text Representations for Asr Without Sampling	3904
<i>Neeraj Gaur, Rohan Agrawal, Gary Wang, Parisa Haghani, Andrew Rosenberg, Bhuvana Ramabhadran</i>	
Contextualized End-To-End Automatic Speech Recognition with Intermediate Biasing Loss	3909
<i>Muhammad Shakeel, Yui Sudo, Yifan Peng, Shinji Watanabe</i>	
Learnable Layer Selection and Model Fusion for Speech Self-Supervised Learning Models	3914
<i>Sheng-Chieh Chiu, Chia-Hua Wu, Jih-Kang Hsieh, Yu Tsao, Hsin-Min Wang</i>	
Sequential Editing for Lifelong Training of Speech Recognition Models.....	3919
<i>Devang Kulshreshtha, Nikolaos Pappas, Brady Houston, Saket Dingliwal, Srikanth Ronanki</i>	

Cross-Modality Diffusion Modeling and Sampling for Speech Recognition	3924
<i>Chia-Kai Yeh, Chih-Chun Chen, Ching-Hsien Hsu, Jen-Tzung Chien</i>	

CROSS-LINGUAL AND MULTILINGUAL PROCESSING

A Parameter-Efficient Language Extension Framework for Multilingual ASR	3929
<i>Wei Liu, Jingyong Hou, Dong Yang, Muyong Cao, Tan Lee</i>	
LoRA-Whisper: Parameter-Efficient and Extensible Multilingual ASR	3934
<i>Zheshu Song, Jianheng Zhuo, Yifan Yang, Ziyang Ma, Shixiong Zhang, Xie Chen</i>	
MHuBERT-147: A Compact Multilingual HuBERT Model	3939
<i>Marcely Zanon Boito, Vivek Iyer, Nikolaos Lagos, Laurent Besacier, Ioan Calapodescu</i>	
All Ears: Building Self-Supervised Learning Based ASR Models for Indian Languages at Scale.....	3944
<i>Vasista Sai Lodagala, Abhishek Biswas, Shoutrik Das, Jordan F, S Umesh</i>	
A Unified Approach to Multilingual Automatic Speech Recognition with Improved Language Identification for Indic Languages	3949
<i>Nikhil Jakhar, Sudhanshu Srivastava, Arun Baby</i>	
Integrating Speech Self-Supervised Learning Models and Large Language Models for ASR.....	3954
<i>Ling Dong, Zhengtao Yu, Wenjun Wang, Yuxin Huang, Shengxiang Gao, Guojiang Zhou</i>	
On the Effects of Heterogeneous Data Sources on Speech-To-Text Foundation Models.....	3959
<i>Jinchuan Tian, Yifan Peng, William Chen, Kwanghee Choi, Karen Livescu, Shinji Watanabe</i>	
Less is More: Accurate Speech Recognition & Translation Without Web-Scale Data	3964
<i>Krishna C. Puvvada, Piotr Zelasko, He Huang, Oleksii Hrinchuk, Nithin Rao Koluguri, Kunal Dhawan, Somshubra Majumdar, Elena Rastorgueva, Zhehuai Chen, Vitaly Lavrukhin, Jagadeesh Balam, Boris Ginsburg</i>	
The Greek Podcast Corpus: Competitive Speech Models for Low-Resourced Languages with Weakly Supervised Data.....	3969
<i>Georgios Paraskevopoulos, Chara Tsoukala, Athanasios Katsamanis, Vassilis Katsouros</i>	
Speech Recognition for Greek Dialects: A Challenging Benchmark	3974
<i>Socrates Vakirtzian, Chara Tsoukala, Stavros Bompolas, Katerina Mouzou, Vivian Stamou, Georgios Paraskevopoulos, Antonios Dimakis, Stella Markantonatou, Angela Ralli, Antonios Anastasopoulos</i>	
LUPET: Incorporating Hierarchical Information Path into Multilingual ASR.....	3979
<i>Wei Liu, Jingyong Hou, Dong Yang, Muyong Cao, Tan Lee</i>	
Whispering in Norwegian: Navigating Orthographic and Dialectic Challenges	3984
<i>Per E Kummervold, Javier De La Rosa, Freddy Wetjen, Rolv-Arild Braaten, Per Erik Solberg</i>	
EFFUSE: Efficient Self-Supervised Feature Fusion for E2E ASR in Low Resource and Multilingual Scenarios.....	3989
<i>Tejes Srivastava, Jiatong Shi, William Chen, Shinji Watanabe</i>	
Enhancing Neural Transducer for Multilingual ASR with Synchronized Language Diarization	3994
<i>Amir Hussein, Desh Raj, Matthew Wiesner, Daniel Povey, Paola Garcia, Sanjeev Khudanpur</i>	

SC-MoE: Switch Conformer Mixture of Experts for Unified Streaming and Non-Streaming Code-Switching ASR	3999
<i>Shuaishuai Ye, Shunfei Chen, Xinhui Hu, Xinkang Xu</i>	

SPEECH ASSESSMENT

Optimizing Automatic Speech Assessment: W-RankSim Regularization and Hybrid Feature Fusion Strategies	4004
<i>Chung-Wen Wu, Berlin Chen</i>	
Context-Aware Speech Recognition Using Prompts for Language Learners	4009
<i>Jian Cheng</i>	
A Dataset and Two-Pass System for Reading Miscue Detection.....	4014
<i>Raj Gothi, Rahul Kumar, Mildred Pereira, Nagesh Nayak, Preeti Rao</i>	
Oversampling, Augmentation and Curriculum Learning for Speaking Assessment with Limited Training Data.....	4019
<i>Tin Mei Lun, Ekaterina Voskoboinik, Ragheb Al-Ghezi, Tamas Grosz, Mikko Kurimo</i>	
Analysis and Visualization of Directional Diversity in Listening Fluency of World Englishes Speakers in the Framework of Mutual Shadowing.....	4024
<i>Yu Tomita, Yingxiang Gao, Nobuaki Minematsu, Noriko Nakanishi, Daisuke Saito</i>	
Quantifying the Role of Textual Predictability in Automatic Speech Recognition	4029
<i>Sean Robertson, Gerald Penn, Ewan Dunbar</i>	

QUESTION ANSWERING FROM SPEECH AND SPOKEN DIALOGUE SYSTEMS

TM-PATHVQA: 90000+ Textless Multilingual Questions for Medical Visual Question Answering	4034
<i>Tonmoy Rajkhowa, Amartya Roy Chowdhury, Sankalp Nagaonkar, Achyut Mani Tripathi, Mahadeva Prasanna</i>	
Towards Multilingual Audio-Visual Question Answering	4039
<i>Orchid Chetia Phukan, Priyabrata Mallick, Swarup Ranjan Behera, Aalekhya Satya Narayani, Arun Balaji Buduru, Rajesh Sharma</i>	
Reinforcement Learning from Answer Reranking Feedback for Retrieval-Augmented Answer Generation	4044
<i>Minh Nguyen, Toan Quoc Nguyen, Kishan Kc, Zeyu Zhang, Thuy Vu</i>	
Instruction Data Generation and Unsupervised Adaptation for Speech Language Models	4049
<i>Vahid Noroozi, Zhehuai Chen, Somshubra Majumdar, Steve Huang, Jagadeesh Balam, Boris Ginsburg</i>	
On the Use of Plausible Arguments in Explainable Conversational AI.....	4054
<i>Martina Di Bratto, Maria Di Maro, Antonio Origlia</i>	
Rapport-Driven Virtual Agent: Rapport Building Dialogue Strategy for Improving User Experience at First Meeting	4059
<i>Muhammad Yeza Baihaqi, Angel Garcia Contreras, Seiya Kawano, Koichiro Yoshino</i>	
Cross-Modal Denoising: A Novel Training Paradigm for Enhancing Speech-Image Retrieval	4064
<i>Lifeng Zhou, Yuke Li, Rui Deng, Yuting Yang, Haoqi Zhu</i>	

SPOKEN DIALOGUE SYSTEMS AND CONVERSATIONAL ANALYSIS 3

MM-NodeFormer: Node Transformer Multimodal Fusion for Emotion Recognition in Conversation	4069
<i>Zilong Huang, Man-Wai Mak, Kong Aik Lee</i>	
Emotional Cues Extraction and Fusion for Multi-Modal Emotion Prediction and Recognition in Conversation.....	4074
<i>Haoxiang Shi, Ziqi Liang, Jun Yu</i>	
Participant-Pair-Wise Bottleneck Transformer for Engagement Estimation from Video Conversation.....	4079
<i>Keita Suzuki, Nobukatsu Hojo, Kazutoshi Shinoda, Saki Mizuno, Ryo Masumura</i>	
Well, What Can You Do with Messy Data? Exploring the Prosody and Pragmatic Function of the Discourse Marker "well" with Found Data and Speech Synthesis	4084
<i>Johannah O'Mahony, Catherine Lai, Éva Székely</i>	
Learning from Multiple Annotator Biased Labels in Multimodal Conversation.....	4089
<i>Kazutoshi Shinoda, Nobukatsu Hojo, Saki Mizuno, Keita Suzuki, Satoshi Kobashikawa, Ryo Masumura</i>	
Non-Linear Inference Time Intervention: Improving LLM Truthfulness.....	4094
<i>Jakub Hoscilowicz, Adam Wiacek, Jan Chojnacki, Adam Cieslak, Leszek Michon, Artur Janicki</i>	
Evaluating Speech Recognition Performance Towards Large Language Model Based Voice Assistants.....	4099
<i>Zhe Liu, Suyoun Kim, Ozlem Kalinli</i>	

DYSARTHRIC SPEECH ASSESSMENT

Automatic Assessment of Dysarthria Using Speech and Synthetically Generated Electroglottograph Signal.....	4104
<i>Fathima Zaheera, Supritha Shetty, Gayadhar Pradhan, Deepak K T</i>	
CDSD: Chinese Dysarthria Speech Database.....	4109
<i>Yan Wan, Mengyi Sun, Xinchun Kang, Jingting Li, Pengfei Guo, Ming Gao, Su-Jing Wang</i>	
Exploring Syllable Discriminability During Diadochokinetic Task with Increasing Dysarthria Severity for Patients with Amyotrophic Lateral Sclerosis.....	4114
<i>Neelesh Samptur, Tanuka Bhattacharjee, Anirudh Chakravarty K, Seena Vengalil, Yamini Belur, Achayaram Nalini, Prasanta Kumar Ghosh</i>	
Beyond Binary: Multiclass Paraphasia Detection with Generative Pretrained Transformers and End-To-End Models	4119
<i>Matthew Perez, Aneesha Sampath, Minxue Niu, Emily Mower Provost</i>	
Electroglottography for the Assessment of Dysphonia in Parkinson's Disease and Multiple System Atrophy.....	4124
<i>Khalid Daoudi, Solange Milhé De Saint Victor, Alexandra Foubert-Samier, Margherita Fabbri, Anne Pavy-Le Traon, Olivier Rascol, Virginie Woisard, Wassilios G. Meissner</i>	
CoLM-DSR: Leveraging Neural Codec Language Modeling for Multi-Modal Dysarthric Speech Reconstruction.....	4129
<i>Xueyuan Chen, Dongchao Yang, Dingdong Wang, Xixin Wu, Zhiyong Wu, Helen Meng</i>	

SPOKEN LANGUAGE MODELS FOR UNIVERSAL SPEECH PROCESSING (SPECIAL SESSION)

On the Effectiveness of Acoustic BPE in Decoder-Only TTS.....	4134
<i>Bohan Li, Feiyu Shen, Yiwei Guo, Shuai Wang, Xie Chen, Kai Yu</i>	
Exploring In-Context Learning of Textless Speech Language Model for Speech Classification Tasks.....	4139
<i>Kai-Wei Chang, Ming-Hao Hsu, Shan-Wen Li, Hung-Yi Lee</i>	
Understanding Sounds, Missing the Questions: The Challenge of Object Hallucination in Large Audio-Language Models	4144
<i>Chun-Yi Kuan, Wei-Ping Huang, Hung-Yi Lee</i>	
Can Large Language Models Understand Spatial Audio?	4149
<i>Changli Tang, Wenyi Yu, Guangzhi Sun, Xianzhao Chen, Tian Tan, Wei Li, Jun Zhang, Lu Lu, Zejun Ma, Yuxuan Wang, Chao Zhang</i>	
DiscreteSLU: A Large Language Model with Self-Supervised Discrete Speech Units for Spoken Language Understanding.....	4154
<i>Suwon Shon, Kwangyoun Kim, Yi-Te Hsu, Prashant Sridhar, Shinji Watanabe, Karen Livescu</i>	
DeSTA: Enhancing Speech Language Models Through Descriptive Speech-Text Alignment.....	4159
<i>Ke-Han Lu, Zhehuai Chen, Szu-Wei Fu, He Huang, Boris Ginsburg, Yu-Chiang Frank Wang, Hung-Yi Lee</i>	
COSMIC: Data Efficient Instruction-Tuning for Speech In-Context Learning.....	4164
<i>Jing Pan, Jian Wu, Yashesh Gaur, Sunit Sivasankaran, Zhuo Chen, Shujie Liu, Jinyu Li</i>	
NAST: Noise Aware Speech Tokenization for Speech Language Models.....	4169
<i>Shoval Messica, Yossi Adi</i>	
Low Bitrate High-Quality RVQGAN-Based Discrete Speech Tokenizer.....	4174
<i>Slava Shechtman, Avihu Dekel</i>	

KEYNOTE 4

Perception of Music and Speech: Focus on Rhythm Processing.....	4179
<i>Barbara Tillmann</i>	

L1/L2 ACQUISITION AND CROSS-LINGUISTIC FACTORS

Acquisition of High Vowel Devoicing in Japanese: A Production Experiment with Three and Four Year Olds.....	4180
<i>Hyun Kyung Hwang, Manami Hirayama</i>	
The Production of Contrastive Focus by 7 to 13-Year-Olds Learning Mandarin Chinese	4184
<i>Zimeng Li, Zhongxuan Mao, Shengting Shen, Ivan Yuen, Ping Tang</i>	
Cross-Linguistic Intelligibility of Non-Compositional Expressions in Spoken Context.....	4189
<i>Iuliia Zaitova, Irina Stenger, Wei Xue, Tania Avgustinova, Bernd Möbius, Dietrich Klakow</i>	
On the Relationship Between Speech Production and Vocabulary Size in 3-5 Year Olds.....	4194
<i>Alexis Demaere, Nicole Van Rootselaar, Fangfang Li, Robbin Gibb, Claudia L. R. Gonzalez</i>	

Towards Classifying Mother Tongue from Infant Cries - Findings Substantiating Prenatal Learning Theory	4199
<i>Tim Polzehl, Tim Herzig, Friedrich Wicke, Kathleen Wermke, Razieh Khamsehashari, Michiko Dahlem, Sebastian Möller</i>	

Effect of Complex Boundary Tones on Tone Identification: An Experimental Study with Mandarin-Speaking Preschool Children.....	4204
<i>Aijun Li, Jun Gao, Zhiwei Wang</i>	

Ethnolinguistic Identification of Vietnamese-German Heritage Speech	4209
<i>Thanh Lan Truong, Andrea Weber</i>	

SPEAKER STANCE, EMOTION AND LANGUAGE-EXTERNAL FACTORS

Joint Prediction of Subjective Listening Effort and Speech Intelligibility Based on End-To-End Learning	4214
<i>Dirk Eike Hoffner, Jana Roßbach, Bernd T. Meyer</i>	

Depression Enhances Internal Inconsistency Between Spoken and Semantic Emotion: Evidence from the Analysis of Emotion Expression in Conversation.....	4219
<i>Xinyi Wu, Changqing Xu, Nan Li, Rongfeng Su, Lan Wang, Nan Yan</i>	

Listeners' F0 Preferences in Quiet and Stationary Noise.....	4224
<i>Olympia Simantiraki, Martin Cooke</i>	

Effects of Talker and Playback Rate of Reverberation-Induced Speech on Speech Intelligibility of Older Adults	4229
<i>Nao Hodoshima</i>	

EXPERIMENTAL PHONETICS AND LABORATORY PHONOLOGY

Age-Related Differences in Acoustic Cues for the Perception of Checked Syllables in Shengzhou Wu	4233
<i>Bingliang Zhao, Jiangping Kong, Xiyu Wu</i>	

Quantity-Sensitivity Affects Recall Performance of Word Stress	4238
<i>Constantijn Kaland, Maria Lialiou</i>	

Phonological Symmetry Does Not Predict Generalization of Perceptual Adaptation to Vowels.....	4243
<i>Zuheyra Tokac, Jennifer Cole</i>	

Perceptual Learning in Lexical Tone: Phonetic Similarity Vs. Phonological Categories.....	4248
<i>Ariëlle Reitsema, Chenxin Li, Leanne Van Lambalgen, Laura Preining, Saskia Galindo Jong, Qing Yang, Xinyi Wen, Yiya Chen</i>	

Modeling Probabilistic Reduction Across Domains with Naive Discriminative Learning	4253
<i>Anna Stein, Kevin Tang</i>	

Do We EXPECT to Find Phonetic Traces for Syntactic Traces?	4258
<i>Jonathan Him Nok Lee, Mark Liberman, Martin Salzmann</i>	

SPEAKER RECOGNITION EVALUATION AND RESOURCES

VoxBlink2: A 100K+ Speaker Recognition Corpus and the Open-Set Speaker-Identification Benchmark.....	4263
<i>Yuke Lin, Ming Cheng, Fulin Zhang, Yingying Gao, Shilei Zhang, Ming Li</i>	
As Biased as You Measure: Methodological Pitfalls of Bias Evaluations in Speaker Verification Research	4268
<i>Wiebke Hutiri, Tanvina Patel, Aaron Yi Ding, Odette Scharenborg</i>	
WeSep: A Scalable and Flexible Toolkit Towards Generalizable Target Speaker Extraction	4273
<i>Shuai Wang, Ke Zhang, Shaoxiong Lin, Junjie Li, Xuefei Wang, Meng Ge, Jianwei Yu, Yanmin Qian, Haizhou Li</i>	
ESPnet-SPK: Full Pipeline Speaker Embedding Toolkit with Reproducible Recipes, Self-Supervised Front-Ends, and Off-The-Shelf Models	4278
<i>Jee-Weon Jung, Wangyou Zhang, Jiatong Shi, Zakaria Aldeneh, Takuya Higuchi, Alex Gichamba, Barry-John Theobald, Ahmed Hussien Abdelaziz, Shinji Watanabe</i>	
Active Speaker Detection in Fisheye Meeting Scenes with Scene Spatial Spectrums	4283
<i>Xinghao Huang, Weiwei Jiang, Long Rao, Wei Xu, Wenqing Cheng</i>	
VSASV: A Vietnamese Dataset for Spoofing-Aware Speaker Verification	4288
<i>Vu Hoang, Viet Thanh Pham, Hoa Nguyen Xuan, Pham Nhi, Phuong Dat, Thi Thu Trang Nguyen</i>	

SPEECH TYPE CLASSIFICATION

E-ODN: An Emotion Open Deep Network for Generalised and Adaptive Speech Emotion Recognition	4293
<i>Liuxian Ma, Lin Shen, Ruobing Li, Haojie Zhang, Kun Qian, Bin Hu, Björn W. Schuller, Yoshiharu Yamamoto</i>	
Enhancing Multilingual Voice Toxicity Detection with Speech-Text Alignment	4298
<i>Joseph Liu, Mahesh Kumar Nandwana, Janne Pyllkkönen, Hannes Heikinheimo, Morgan McGuire</i>	
AraOffence: Detecting Offensive Speech Across Dialects in Arabic Media	4303
<i>Youssef Nafea, Shady Shehata, Zeerak Talat, Ahmed Aboeitta, Ahmed Sharshar, Preslav Nakov</i>	
CogniVoice: Multimodal and Multilingual Fusion Networks for Mild Cognitive Impairment Assessment from Spontaneous Speech.....	4308
<i>Jiali Cheng, Mohamed Elgaar, Nidhi Vakil, Hadi Amiri</i>	
Speech Topic Classification Based on Multi-Scale and Graph Attention Networks	4313
<i>Fangjing Niu, Xiaozhe Qi, Xinya Chen, Liang He</i>	
Enhancing Speech and Music Discrimination Through the Integration of Static and Dynamic Features	4318
<i>Liangwei Chen, Xiren Zhou, Qiang Tu, Huanhuan Chen</i>	

TARGET SPEAKER EXTRACTION

Binaural Selective Attention Model for Target Speaker Extraction.....	4323
<i>Hanyu Meng, Qiquan Zhang, Xiangyu Zhang, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i>	
All Neural Low-Latency Directional Speech Extraction.....	4328
<i>Ashutosh Pandey, Sanha Lee, Juan Azcarreta, Daniel Wong, Buye Xu</i>	
Centroid Estimation with Transformer-Based Speaker Embedder for Robust Target Speaker Extraction	4333
<i>Woon-Haeng Heo, Joongyu Maeng, Yoseb Kang, Namhyun Cho</i>	
Knowledge Boosting During Low-Latency Inference.....	4338
<i>Vidya Srinivas, Malek Itani, Tuochao Chen, Emre Sefik Eskimez, Takuya Yoshioka, Shyamnath Gollakota</i>	
Unified Audio Visual Cues for Target Speaker Extraction	4343
<i>Tianci Wu, Shulin He, Jiahui Pan, Haifeng Huang, Zhijian Mo, Xueliang Zhang</i>	
Target Speaker Extraction with Curriculum Learning.....	4348
<i>Yun Liu, Xuechen Liu, Xiaoxiao Miao, Junichi Yamagishi</i>	

SPEECH SYNTHESIS: VOICE CONVERSION 3

SPA-SVC: Self-Supervised Pitch Augmentation for Singing Voice Conversion.....	4353
<i>Bingsong Bai, Fengping Wang, Yingming Gao, Ya Li</i>	
Towards Naturalistic Voice Conversion: NaturalVoices Dataset with an Automatic Processing Pipeline.....	4358
<i>Ali N. Salman, Zongyang Du, Shreeram Suresh Chandra, Ismail Rasim Ülgen, Carlos Busso, Berrak Sisman</i>	
PRVAE-VC2: Non-Parallel Voice Conversion by Distillation of Speech Representations	4363
<i>Kou Tanaka, Hirokazu Kameoka, Takuhiro Kaneko, Yuto Kondo</i>	
HybridVC: Efficient Voice Style Conversion with Text and Audio Prompts	4368
<i>Xinlei Niu, Jing Zhang, Charles Patrick Martin</i>	
DreamVoice: Text-Guided Voice Conversion.....	4373
<i>Jiarui Hai, Karan Thakkar, Helin Wang, Zengyi Qin, Mounya Elhilali</i>	
Hear Your Face: Face-Based Voice Conversion with F0 Estimation.....	4378
<i>Jaejun Lee, Yoori Oh, Injune Hwang, Kyogu Lee</i>	
Accent Conversion with Articulatory Representations.....	4383
<i>Yashish M. Siriwardena, Nathan Swedlow, Audrey Howard, Evan Gitterman, Dan Darcy, Carol Espy-Wilson, Andrea Fanelli</i>	
USD-AC: Unsupervised Speech Disentanglement for Accent Conversion.....	4388
<i>Jen-Hung Huang, Wei-Tsung Lee, Chung-Hsien Wu</i>	
Knowledge Distillation from Self-Supervised Representation Learning Model with Discrete Speech Units for Any-To-Any Streaming Voice Conversion	4393
<i>Hiroki Kanagawa, Yusuke Ijima</i>	

SPEECH SYNTHESIS: PARADIGMS AND METHODS 3

SimpleSpeech: Towards Simple and Efficient Text-To-Speech with Scalar Latent Transformer Diffusion Models.....	4398
<i>Dongchao Yang, Dingdong Wang, Haohan Guo, Xueyuan Chen, Xixin Wu, Helen Meng</i>	
Sample-Efficient Diffusion for Text-To-Speech Synthesis.....	4403
<i>Justin Lovelace, Soham Ray, Kwangyoun Kim, Kilian Q. Weinberger, Felix Wu</i>	
Exploring the Robustness of Text-To-Speech Synthesis Based on Diffusion Probabilistic Models to Heavily Noisy Transcriptions	4408
<i>Jingyi Feng, Yusuke Yasuda, Tomoki Toda</i>	
VoiceTailor: Lightweight Plug-In Adapter for Diffusion-Based Personalized Text-To-Speech	4413
<i>Heeseung Kim, Sang-Gil Lee, Jiheum Yeom, Che Hyun Lee, Sungwon Kim, Sungroh Yoon</i>	
PitchFlow: Adding Pitch Control to a Flow-Matching Based TTS Model.....	4418
<i>Tasnima Sadekova, Mikhail Kudinov, Vadim Popov, Assel Yermekova, Artem Khrapov</i>	
DualSpeech: Enhancing Speaker-Fidelity and Text-Intelligibility Through Dual Classifier-Free Guidance.....	4423
<i>Jinhyeok Yang, Junhyeok Lee, Hyeong-Seok Choi, Seunghoon Ji, Hyeongju Kim, Juheon Lee</i>	
Generating Speakers by Prompting Listener Impressions for Pre-Trained Multi-Speaker Text-To- Speech Systems	4428
<i>Zhengyang Chen, Xuechen Liu, Erica Cooper, Junichi Yamagishi, Yanmin Qian</i>	
TacoLM: GaTed Attention Equipped Codec Language Model Are Efficient Zero-Shot Text to Speech Synthesizers	4433
<i>Yakun Song, Zhuo Chen, Xiaofei Wang, Ziyang Ma, Guanrou Yang, Xie Chen</i>	

PRIVACY AND SECURITY IN SPEECH COMMUNICATION 2

Anonymising Elderly and Pathological Speech: Voice Conversion Using DDSP and Query-By- Example.....	4438
<i>Suhita Ghosh, Melanie Jouaiti, Arnab Das, Yamini Sinha, Tim Polzehl, Ingo Siegert, Sebastian Stober</i>	
Asynchronous Voice Anonymization Using Adversarial Perturbation on Speaker Embedding	4443
<i>Rui Wang, Liping Chen, Kong Aik Lee, Zhen-Hua Ling</i>	
Probing the Feasibility of Multilingual Speaker Anonymization	4448
<i>Sarina Meyer, Florian Lux, Ngoc Thang Vu</i>	

VOLUME 7

DiffVC+: Improving Diffusion-Based Voice Conversion for Speaker Anonymization.....	4453
<i>Fan Huang, Kun Zeng, Wei Zhu</i>	

STREAMING ASR

Learning from Back Chunks: Acquiring More Future Knowledge for Streaming ASR Models Via Self Distillation.....	4458
<i>Yuting Yang, Guodong Ma, Yuke Li, Binbin Du, Haoqi Zhu, Liang Ruan</i>	
Decoder-Only Architecture for Streaming End-To-End Speech Recognition	4463
<i>Emiru Tsunoo, Hayato Futami, Yosuke Kashiwagi, Siddhant Arora, Shinji Watanabe</i>	
Streaming Decoder-Only Automatic Speech Recognition with Discrete Speech Units: A Pilot Study	4468
<i>Peikun Chen, Sining Sun, Changhao Shan, Qing Yang, Lei Xie</i>	
TfCleanformer: A Streaming, Array-Agnostic, Full- And Sub-Band Modeling Front-End for Robust ASR	4473
<i>Jens Heitkaemper, Joe Caroselli, Arun Narayanan, Nathan Howard</i>	
Improving Streaming Speech Recognition with Time-Shifted Contextual Attention and Dynamic Right Context Masking.....	4478
<i>Khanh Le, Duc Chau</i>	
Simul-Whisper: Attention-Guided Streaming Whisper with Truncation Detection	4483
<i>Haoyu Wang, Guoqiang Hu, Guodong Lin, Wei-Qiang Zhang, Jian Li</i>	

COMPUTATIONAL RESOURCE CONSTRAINED ASR

Dynamic Data Pruning for Automatic Speech Recognition	4488
<i>Qiao Xiao, Pingchuan Ma, Adriana Fernandez-Lopez, Boqian Wu, Lu Yin, Stavros Petridis, Mykola Pechenizkiy, Maja Pantic, Decebal Constantin Mocanu, Shiwei Liu</i>	
Mitigating Overfitting in Structured Pruning of ASR Models with Gradient-Guided Parameter Regularization	4493
<i>Dong-Hyun Kim, Joon-Hyuk Chang</i>	
SparseWAV: Fast and Accurate One-Shot Unstructured Pruning for Large Speech Foundation Models.....	4498
<i>Tianteng Gu, Bei Liu, Hang Shao, Yanmin Qian</i>	
One-Pass Multiple Conformer and Foundation Speech Systems Compression and Quantization Using an All-In-One Neural Model.....	4503
<i>Zhaoqing Li, Haoning Xu, Tianzi Wang, Shoukang Hu, Zengrui Jin, Shujie Hu, Jiajun Deng, Mingyu Cui, Mengzhe Geng, Xunying Liu</i>	
USM RNN-T Model Weights Binarization	4508
<i>Oleg Rybakov, Dmitriy Serdyuk, Chengjian Zheng</i>	
DAISY: Data Adaptive Self-Supervised Early Exit for Speech Representation Models.....	4513
<i>Tzu-Quan Lin, Hung-Yi Lee, Hao Tang</i>	
RepTor: Re-Parameterizable Temporal Convolution for Keyword Spotting Via Differentiable Kernel Search	4518
<i>Eunuk Park, Daehyun Ahn, Hyungjun Kim</i>	
Global-Local Convolution with Spiking Neural Networks for Energy-Efficient Keyword Spotting	4523
<i>Shuai Wang, Dehao Zhang, Kexin Shi, Yuchen Wang, Wenjie Wei, Jibin Wu, Malu Zhang</i>	

ED-sKWS: Early-Decision Spiking Neural Networks for Rapid, and Energy-Efficient Keyword Spotting	4528
<i>Zeyang Song, Qianhui Liu, Qu Yang, Yizhou Peng, Haizhou Li</i>	

A Small and Fast BERT for Chinese Medical Punctuation Restoration	4533
<i>Tongtao Ling, Yutao Lai, Lei Chen, Shilei Huang, Yi Liu</i>	

EVALUATION OF SPEECH TECHNOLOGY SYSTEMS

Quantification of Stylistic Differences in Human- And ASR-Produced Transcripts of African American English	4538
<i>Annika Heuser, Tyler Kendall, Miguel Del Rio, Quinn McNamara, Nishchal Bhandari, Corey Miller, Migüel Jetté</i>	

Beyond Levenshtein: Leveraging Multiple Algorithms for Robust Word Error Rate Computations and Granular Error Classifications	4543
<i>Korbinian Kuhn, Verena Kersken, Gottfried Zimmermann</i>	

Comparing ASR Systems in the Context of Speech Disfluencies	4548
<i>Maria Teleki, Xiangjue Dong, Soohwan Kim, James Caverlee</i>	

Deep Prosodic Features in Tandem with Perceptual Judgments of Word Reduction for Tone Recognition in Conversed Speech.....	4553
<i>Xiang-Li Lu, Yi-Fen Liu</i>	

SeMaScore: A New Evaluation Metric for Automatic Speech Recognition Tasks.....	4558
<i>Zitha Sasindran, Harsha Yelchuri, T. V. Prabhakar</i>	

NEURAL NETWORK TRAINING FOR SPEECH RECOGNITION

Dynamic Encoder Size Based on Data-Driven Layer-Wise Pruning for Speech Recognition	4563
<i>Jingjing Xu, Wei Zhou, Zijian Yang, Eugen Beck, Ralf Schlüter</i>	

Revisiting Convolution-Free Transformer for Speech Recognition	4568
<i>Zejiang Hou, Goeric Huybrechts, Anshu Bhatia, Daniel Garcia-Romero, Kyu J. Han, Katrin Kirchhoff</i>	

Optimizing Large-Scale Context Retrieval for End-To-End ASR.....	4573
<i>Zhiqi Huang, Diamantino Caseiro, Kandarp Joshi, Christopher Li, Pat Rondon, Zelin Wu, Petr Zadrzail, Lillian Zhou</i>	

Self-Supervised Speech Representations Are More Phonetic than Semantic	4578
<i>Kwanghee Choi, Ankita Pasad, Tomohiko Nakamura, Satoru Fukayama, Karen Livescu, Shinji Watanabe</i>	

Enhancing CTC-Based Speech Recognition with Diverse Modeling Units	4583
<i>Shiyi Han, Mingbin Xu, Zhihong Lei, Zhen Huang, Xingyu Na</i>	

Guiding Frame-Level CTC Alignments Using Self-Knowledge Distillation	4588
<i>Eungbeom Kim, Hantae Kim, Kyogu Lee</i>	

LEVERAGING LARGE LANGUAGE MODELS AND CONTEXTUAL FEATURES FOR PHONETIC ANALYSIS (SPECIAL SESSION)

Human-Like Linguistic Biases in Neural Speech Models: Phonetic Categorization and Phonotactic Constraints in Wav2Vec2.0.....	4593
<i>Marianne De Heer Kloots, Willem Zuidema</i>	
Exploring Pre-Trained Speech Model for Articulatory Feature Extraction in Dysarthric Speech Using ASR.....	4598
<i>Yugin Lin, Longbiao Wang, Jianwu Dang, Nobuaki Minematsu</i>	
Exploring Self-Supervised Speech Representations for Cross-Lingual Acoustic-To-Articulatory Inversion.....	4603
<i>Yun Hao, Reihaneh Amooie, Wietse De Vries, Thomas Tienkamp, Rik Van Noord, Martijn Wieling</i>	
Are Articulatory Feature Overlaps Shrouded in Speech Embeddings?	4608
<i>Erfan A. Shams, Iona Gessinger, Patrick Cormac English, Julie Carson-Berndsen</i>	
Searching for Structure: Appraising the Organisation of Speech Features in Wav2vec 2.0 Embeddings.....	4613
<i>Patrick Cormac English, John D. Kelleher, Julie Carson-Berndsen</i>	

RESPONSIBLE SPEECH FOUNDATION MODELS (SPECIAL SESSION)

Speech Foundation Models in Healthcare: Effect of Layer Selection on Pathological Speech Feature Prediction.....	4618
<i>Daniela A. Wiepert, Rene L. Utianski, Joseph R. Duffy, John L. Stricker, Leland R. Barnard, David T. Jones, Hugo Botha</i>	
Outlier Reduction with Gated Attention for Improved Post-Training Quantization in Large Sequence-To-Sequence Speech Foundation Models	4623
<i>Dominik Wagner, Ilja Baumann, Korbinian Riedhammer, Tobias Bocklet</i>	
Unveiling Biases While Embracing Sustainability: Assessing the Dual Challenges of Automatic Speech Recognition Systems.....	4628
<i>Ajinkya Kulkarni, Atharva Kulkarni, Miguel Couceiro, Isabel Trancoso</i>	
Emo-Bias: A Large Scale Evaluation of Social Bias on Speech Emotion Recognition.....	4633
<i>Yi-Cheng Lin, Haibin Wu, Huang-Cheng Chou, Chi-Chun Lee, Hung-Yi Lee</i>	
On the Social Bias of Speech Self-Supervised Models	4638
<i>Yi-Cheng Lin, Tzu-Quan Lin, Hsi-Che Lin, Andy T. Liu, Hung-Yi Lee</i>	
Self-Supervised Speech Representations Still Struggle with African American Vernacular English.....	4643
<i>Kalvin Chang, Yi-Hui Chou, Jiatong Shi, Hsuan-Ming Chen, Nicole Holliday, Odette Scharenborg, David R. Mortensen</i>	
Can You Remove the Downstream Model for Speaker Recognition with Self-Supervised Speech Features?.....	4648
<i>Zakaria Aldeneh, Takuya Higuchi, Jee-Weon Jung, Skylar Seto, Tatiana Likhomanenko, Stephen Shum, Ahmed Hussien Abdelaziz, Shinji Watanabe, Barry-John Theobald</i>	
Empowering Whisper as a Joint Multi-Talker and Target-Talker Speech Recognition System	4653
<i>Lingwei Meng, Jiawen Kang, Yuejiao Wang, Zengrui Jin, Xixin Wu, Xunying Liu, Helen Meng</i>	

MULTIMODAL PARALINGUISTICS

LoRA-MER: Low-Rank Adaptation of Pre-Trained Speech Models for Multimodal Emotion Recognition Using Mutual Information.....	4658
<i>Yunrui Cai, Zhiyong Wu, Jia Jia, Helen Meng</i>	
Enhancing Modal Fusion by Alignment and Label Matching for Multimodal Emotion Recognition.....	4663
<i>Qifei Li, Yingming Gao, Yuhua Wen, Cong Wang, Ya Li</i>	
Prompt Link Multimodal Fusion in Multimodal Sentiment Analysis.....	4668
<i>Kang Zhu, Cunhang Fan, Jianhua Tao, Zhao Lv</i>	
A Multimodal Analysis of Different Types of Laughter Expression in Conversational Dialogues	4673
<i>Kexin Wang, Carlos Ishi, Ryoko Hayashi</i>	
Tackling Missing Modalities in Audio-Visual Representation Learning Using Masked Autoencoders.....	4678
<i>Georgios Chochlakis, Chandrashekhar Lavania, Prashant Mathur, Kyu J. Han</i>	
Enhancing Multimodal Emotion Recognition Through ASR Error Compensation and LLM Fine-Tuning	4683
<i>Jehyun Kyung, Serin Heo, Joon-Hyuk Chang</i>	
Bridging Emotions Across Languages: Low Rank Adaptation for Multilingual Speech Emotion Recognition	4688
<i>Lucas Goncalves, Donita Robinson, Elizabeth Richerson, Carlos Busso</i>	

AUTOMATIC EMOTION RECOGNITION

A Layer-Anchoring Strategy for Enhancing Cross-Lingual Speech Emotion Recognition.....	4693
<i>Shreya G. Upadhyay, Carlos Busso, Chi-Chun Lee</i>	
Are Paralinguistic Representations All that is Needed for Speech Emotion Recognition?	4698
<i>Orchid Chetia Phukan, Gautam Siddharth Kashyap, Arun Balaji Buduru, Rajesh Sharma</i>	
MFSN: Multi-Perspective Fusion Search Network for Pre-Training Knowledge in Speech Emotion Recognition	4703
<i>Haiyang Sun, Fulin Zhang, Yingying Gao, Shilei Zhang, Zheng Lian, Junlan Feng</i>	
Exploring Self-Supervised Multi-View Contrastive Learning for Speech Emotion Recognition with Limited Annotations	4708
<i>Bulat Khaertdinov, Pedro Jeruis, Annanda Sousa, Enrique Hortal</i>	

SELF AND WEAKLY-LABELLED SPEAKER VERIFICATION

Self-Supervised Speaker Verification with Mini-Batch Prediction Correction	4713
<i>Junxu Wang, Zhihua Fang, Liang He</i>	
SCDNet: Self-Supervised Learning Feature Based Speaker Change Detection.....	4718
<i>Yue Li, Xinsheng Wang, Li Zhang, Lei Xie</i>	
Self-Supervised Learning with Multi-Head Multi-Mode Knowledge Distillation for Speaker Verification	4723
<i>Ze Zhong Jin, Youzhi Tu, Man-Wai Mak</i>	

Getting More for Less: Using Weak Labels and AV-Mixup for Robust Audio-Visual Speaker Verification	4728
<i>Anith Selvakumar, Homa Fashandi</i>	

ACOUSTIC EVENT DETECTION, SEGMENTATION AND CLASSIFICATION

FastAST: Accelerating Audio Spectrogram Transformer Via Token Merging and Cross-Model Knowledge Distillation.....	4733
<i>Swarup Ranjan Behera, Abhishek Dhiman, Karthik Gowda, Aalekhya Satya Narayani</i>	
LungAdapter: Efficient Adapting Audio Spectrogram Transformer for Lung Sound Classification.....	4738
<i>Li Xiao, Lucheng Fang, Yuhong Yang, Weiping Tu</i>	
ElasticAST: An Audio Spectrogram Transformer for All Length and Resolutions	4743
<i>Jiu Feng, Mehmet Hamza Erol, Joon Son Chung, Arda Senocak</i>	
Robust Laughter Segmentation with Automatic Diverse Data Synthesis	4748
<i>Taisei Omine, Kenta Akita, Reiji Tsuruno</i>	
Explainable By-Design Audio Segmentation Through Non-Negative Matrix Factorization and Probing	4753
<i>Martin Lebourdais, Théo Mariotte, Antonio Almudévar, Marie Tahon, Alfonso Ortega</i>	
Predicting Heart Activity from Speech Using Data-Driven and Knowledge-Based Features	4758
<i>Gasser Elbanna, Zohreh Mostaani, Mathew Magimai.-Doss</i>	
Measuring Acoustic Dissimilarity of Hierarchical Markers in Task-Oriented Dialogue with MFCC-Based Dynamic Time Warping.....	4763
<i>Natalia Morozova, Guanghao You, Sabine Stoll, Adrian Bangerter</i>	
Comparative Analysis of Personalized Voice Activity Detection Systems: Assessing Real-World Effectiveness	4768
<i>Sai Srujana Buddi, Satyam Kumar, Utkarsh Sarawgi, Vineet Garg, Shivesh Ranjan, Ognjen Rudovic, Ahmed Hussen Abdelaziz, Saurabh Adya</i>	
Generalized Fake Audio Detection Via Deep Stable Learning	4773
<i>Zhiyong Wang, Ruibo Fu, Zhengqi Wen, Yuankun Xie, Yukun Liu, Xiaopeng Wang, Xuefei Liu, Yongwei Li, Jianhua Tao, Xin Qi, Yi Lu, Shuchen Shi</i>	
Multimodal Large Language Models with Fusion Low Rank Adaptation for Device Directed Speech Detection.....	4778
<i>Shruti Palaskar, Ognjen Rudovic, Sameer Dharur, Florian Pesce, Gautam Krishna, Aswin Sivaraman, Jack Berkowitz, Ahmed Hussen Abdelaziz, Saurabh Adya, Ahmed Tewfik</i>	
CtrSVDD: A Benchmark Dataset and Baseline Analysis for Controlled Singing Voice Deepfake Detection	4783
<i>Yongyi Zang, Jiatong Shi, You Zhang, Ryuichi Yamamoto, Jionghao Han, Yuxun Tang, Shengyuan Xu, Wenxiao Zhao, Jing Guo, Tomoki Toda, Zhiyao Duan</i>	
Fully Few-Shot Class-Incremental Audio Classification Using Expandable Dual-Embedding Extractor	4788
<i>Yongjie Si, Yanxiong Li, Jialong Li, Jiaxin Tan, Qianhua He</i>	
Multi-Label Bird Species Classification from Field Recordings Using Mel_Graph-GCN Framework.....	4793
<i>Noumida A, Rajeev Rajan</i>	

SPEECH AND AUDIO MODELLING

DiveSound: LLM-Assisted Automatic Taxonomy Construction for Diverse Audio Generation.....	4798
<i>Baihan Li, Zeyu Xie, Xuenan Xu, Yiwei Guo, Ming Yan, Ji Zhang, Kai Yu, Mengyue Wu</i>	
Leveraging Language Model Capabilities for Sound Event Detection	4803
<i>Hualei Wang, Jianguo Mao, Zhifang Guo, Jiarui Wan, Hong Liu, Xiangdong Wang</i>	
Enhancing Zero-Shot Audio Classification Using Sound Attribute Knowledge from Large Language Models	4808
<i>Xuenan Xu, Pingyue Zhang, Ming Yan, Ji Zhang, Mengyue Wu</i>	
LAFMA: A Latent Flow Matching Model for Text-To-Audio Generation.....	4813
<i>Wenhao Guan, Kaidi Wang, Wangjin Zhou, Yang Wang, Feng Deng, Hui Wang, Lin Li, Qingyang Hong, Yong Qin</i>	
DNSMOS Pro: A Reduced-Size DNN for Probabilistic MOS of Speech.....	4818
<i>Fredrik Cumlin, Xinyu Liang, Victor Ungureanu, Chandan K. A. Reddy, Christian Schüldt, Saikat Chatterjee</i>	
Blind Zero-Shot Audio Restoration: A Variational Autoencoder Approach for Denoising and Inpainting	4823
<i>Veranika Boukun, Jakob Dreßs, Jörg Lücke</i>	

FAKE AUDIO DETECTION

Towards Generalisable and Calibrated Audio Deepfake Detection with Self-Supervised Representations	4828
<i>Octavian Pascu, Adriana Stan, Dan Oneata, Elisabeta Oneata, Horia Cucu</i>	
Generalized Source Tracing: Detecting Novel Audio Deepfake Algorithm with Real Emphasis and Fake Dispersion Strategy.....	4833
<i>Yuankun Xie, Ruibo Fu, Zhengqi Wen, Zhiyong Wang, Xiaopeng Wang, Haonnan Cheng, Long Ye, Jianhua Tao</i>	
Enhancing Partially Spoofed Audio Localization with Boundary-Aware Attention Mechanism	4838
<i>Jiafeng Zhong, Bin Li, Jiangyan Yi</i>	
Singing Voice Graph Modeling for SingFake Detection	4843
<i>Xuanjun Chen, Haibin Wu, Roger Jang, Hung-Yi Lee</i>	
Genuine-Focused Learning Using Mask AutoEncoder for Generalized Fake Audio Detection.....	4848
<i>Xiaopeng Wang, Ruibo Fu, Zhengqi Wen, Zhiyong Wang, Yuankun Xie, Yukun Liu, Jianhua Tao, Xuefei Liu, Yongwei Li, Xin Qi, Yi Lu, Shuchen Shi</i>	
One-Class Learning with Adaptive Centroid Shift for Audio Deepfake Detection.....	4853
<i>Hyun Myung Kim, Kangwook Jang, Hoirin Kim</i>	

DEEP LEARNING-BASED SPEECH ENHANCEMENT: APPROACHES, SCALABILITY, AND EVALUATION

VoiCor: A Residual Iterative Voice Correction Framework for Monaural Speech Enhancement.....	4858
<i>Rui Cao, Tianrui Wang, Meng Ge, Andong Li, Longbiao Wang, Jianwu Dang, Yungang Jia</i>	

Personalized Speech Enhancement Without a Separate Speaker Embedding Model	4863
<i>Tanel Pärnamaa, Ando Saabas</i>	
URGENT Challenge: Universality, Robustness, and Generalizability for Speech Enhancement	4868
<i>Wangyou Zhang, Robin Scheibler, Kohei Saijo, Samuele Cornell, Chenda Li, Zhaocheng Ni, Jan Pirklbauer, Marvin Sach, Shinji Watanabe, Tim Fingscheidt, Yanmin Qian</i>	
EARS: An Anechoic Fullband Speech Dataset Benchmarked for Speech Enhancement and Dereverberation	4873
<i>Julius Richter, Yi-Chiao Wu, Steven Krenn, Simon Welker, Bunlong Lay, Shinji Watanabe, Alexander Richard, Timo Gerkmann</i>	

SPEECH SYNTHESIS: OTHER TOPICS 1

LiteFocus: Accelerated Diffusion Inference for Long Audio Synthesis	4878
<i>Zhenxiong Tan, Xinyin Ma, Gongfan Fang, Xinchao Wang</i>	
Speak in the Scene: Diffusion-Based Acoustic Scene Transfer Toward Immersive Speech Generation	4883
<i>Miseul Kim, Soo-Whan Chung, Youna Ji, Hong-Goo Kang, Min-Seok Choi</i>	
PL-TTS: A Generalizable Prompt-Based Diffusion TTS Augmented by Large Language Model	4888
<i>Shuhua Li, Qirong Mao, Jiatong Shi</i>	
Towards Realtime Co-Speech Gestures Synthesis Using STARGATE	4893
<i>Louis Abel, Vincent Colotte, Slim Ouni</i>	
PPPR: Portable Plug-In Prompt Refiner for Text to Audio Generation	4898
<i>Shuchen Shi, Ruiibo Fu, Zhengqi Wen, Jianhua Tao, Tao Wang, Chunyu Qiang, Yi Lu, Xin Qi, Xuefei Liu, Yukun Liu, Yongwei Li, Zhiyong Wang, Xiaopeng Wang</i>	
Neural ATSM: Fully Neural Network-Based Adaptive Time-Scale Modification Using Sentence-Specific Dynamic Control	4903
<i>Jaeuk Lee, Sohee Jang, Joon-Hyuk Chang</i>	
FLY-TTS: Fast, Lightweight and High-Quality End-To-End Text-To-Speech Synthesis	4908
<i>Yinlin Guo, Yening Lv, Jinqiao Dou, Yan Zhang, Yuehai Wang</i>	
Zero-Shot Out-Of-Domain is No Joke: Lessons Learned in the VoiceMOS 2023 MOS Prediction Challenge	4913
<i>Marie Kunešová, Jan Lehecka, Josef Michálek, Jindrich Matousek, Jan Švec</i>	
Towards a General-Purpose Model of Perceived Pragmatic Similarity	4918
<i>Nigel G. Ward, Andres Segura, Alejandro Ceballos, Divette Marco</i>	

SPEECH SYNTHESIS: OTHER TOPICS 2

Enabling Conversational Speech Synthesis Using Noisy Spontaneous Data	4923
<i>Liisa Rätsep, Rasmus Lellep, Mark Fishel</i>	
Frame-Wise Breath Detection with Self-Training: An Exploration of Enhancing Breath Naturalness in Text-To-Speech	4928
<i>Dong Yang, Tomoki Koriyama, Yuki Saito</i>	

H4C-TTS: Leveraging Multi-Modal Historical Context for Conversational Text-To-Speech	4933
<i>Donghyun Seong, Joon-Hyuk Chang</i>	
Bilingual and Code-Switching TTS Enhanced with Denoising Diffusion Model and GAN.....	4938
<i>Huai-Zhe Yang, Chia-Ping Chen, Shan-Yun He, Cheng-Ruei Li</i>	
SpeechBERTScore: Reference-Aware Automatic Evaluation of Speech Generation Leveraging NLP Evaluation Metrics	4943
<i>Takaaki Saeki, Soumi Maiti, Shinnosuke Takamichi, Shinji Watanabe, Hiroshi Saruwatari</i>	
UNIQUE : Unsupervised Network for Integrated Speech Quality Evaluation.....	4948
<i>Juhwan Yoon, Wooseok Ko, Seyun Um, Sungwoong Hwang, Soojoong Hwang, Changhwan Kim, Hong-Goo Kang</i>	
FVTTS : Face Based Voice Synthesis for Text-To-Speech.....	4953
<i>Minyoung Lee, Eunil Park, Sungeun Hong</i>	

SPEECH SYNTHESIS: CROSS-LINGUAL AND MULTILINGUAL ASPECTS

Meta Learning Text-To-Speech Synthesis in Over 7000 Languages.....	4958
<i>Florian Lux, Sarina Meyer, Lyonel Behringer, Frank Zalkow, Phat Do, Matt Coler, Emanuel A. P. Habets, Ngoc Thang Vu</i>	
An Initial Investigation of Language Adaptation for TTS Systems Under Low-Resource Scenarios.....	4963
<i>Cheng Gong, Erica Cooper, Xin Wang, Chunyu Qiang, Mengzhe Geng, Dan Wells, Longbiao Wang, Jianwu Dang, Marc Tessier, Aidan Pine, Korin Richmond, Junichi Yamagishi</i>	
Improving Multilingual Text-To-Speech with Mixture-Of-Language-Experts and Accent Disentanglement.....	4968
<i>Jing Wu, Ting Chen, Minchuan Chen, Wei Hu, Shaojun Wang, Jing Xiao</i>	
Seamless Language Expansion: Enhancing Multilingual Mastery in Self-Supervised Models	4973
<i>Jing Xu, Minglin Wu, Xixin Wu, Helen Meng</i>	
XTTS: A Massively Multilingual Zero-Shot Text-To-Speech Model.....	4978
<i>Edresson Casanova, Kelly Davis, Eren Gölge, Gökem Gökner, Iulian Gulea, Logan Hart, Aya Aljafari, Joshua Meyer, Reuben Morais, Samuel Olayemi, Julian Weber</i>	
X-E-Speech: Joint Training Framework of Non-Autoregressive Cross-Lingual Emotional Text-To- Speech and Voice Conversion	4983
<i>Houjian Guo, Chaoran Liu, Carlos Toshinori Ishi, Hiroshi Ishiguro</i>	

NOISE, FAR-FIELD, MULTI-TALKER, ENHANCEMENT, AUDIO CLASSIFICATION

RIR-SF: Room Impulse Response Based Spatial Feature for Target Speech Recognition in Multi- Channel Multi-Speaker Scenarios	4988
<i>Yiwen Shao, Shi-Xiong Zhang, Dong Yu</i>	
Multi-Channel Multi-Speaker ASR Using Target Speaker's Solo Segment.....	4993
<i>Yiwen Shao, Shi-Xiong Zhang, Yong Xu, Meng Yu, Dong Yu, Daniel Povey, Sanjeev Khudanpur</i>	

Transcription-Free Fine-Tuning of Speech Separation Models for Noisy and Reverberant Multi-Speaker Automatic Speech Recognition.....	4998
<i>William Ravenscroft, George Close, Stefan Goetze, Thomas Hain, Mohammad Soleymanpour, Anurag Chowdhury, Mark C. Fuhs</i>	
NOTSOFAR-1 Challenge: New Datasets, Baseline, and Tasks for Distant Meeting Transcription	5003
<i>Alon Vinnikov, Amir Ivry, Aviv Hurvitz, Igor Abramovski, Sharon Koubi, Ilya Gurvich, Shai Peer, Xiong Xiao, Benjamin Martinez Elizalde, Naoyuki Kanda, Xiaofei Wang, Shalev Shaer, Stav Yagev, Yossi Asher, Sunit Sivasankaran, Yifan Gong, Min Tang, Huaming Wang, Eyal Krupka</i>	
Enhanced ASR Robustness to Packet Loss with a Front-End Adaptation Network	5008
<i>Yehoshua Dissen, Shiry Yonash, Israel Cohen, Joseph Keshet</i>	
Hold Me Tight: Stable Encoder-Decoder Design for Speech Enhancement	5013
<i>Daniel Haider, Felix Perfler, Vincent Lostanlen, Martin Ehler, Peter Balazs</i>	
DGSRN: Noise-Robust Speech Recognition Method with Dual-Path Gated Spectral Refinement Network.....	5018
<i>Wenjun Wang, Shangbin Mo, Ling Dong, Zhengtao Yu, Junjun Guo, Yuxin Huang</i>	
Towards Robust Few-Shot Class Incremental Learning in Audio Classification Using Contrastive Representation	5023
<i>Riyansha Singh, Parinita Nema, Vinod K Kurmi</i>	
Bird Whisperer: Leveraging Large Pre-Trained Acoustic Model for Bird Call Classification	5028
<i>Muhammad Umer Sheikh, Hassan Abid, Bhuiyan Sanjid Shafique, Asif Hanif, Muhammad Haris Khan</i>	
SpeakerBeam-SS: Real-Time Target Speaker Extraction with Lightweight Conv-TasNet and State Space Modeling	5033
<i>Hiroshi Sato, Takafumi Moriya, Masato Mimura, Shota Horiguchi, Tsubasa Ochiai, Takanori Ashihara, Atsushi Ando, Kentaro Shinayama, Marc Delcroix</i>	
wTIMIT2mix: A Cocktail Party Mixtures Database to Study Target Speaker Extraction for Normal and Whispered Speech.....	5038
<i>Marvin Borsdorf, Zexu Pan, Haizhou Li, Tanja Schultz</i>	

SELF-SUPERVISED LEARNING FOR ASR

What Happens in Continued Pre-Training? Analysis of Self-Supervised Speech Models with Continued Pre-Training for Colloquial Finnish ASR	5043
<i>Yaroslav Getman, Tamas Grosz, Mikko Kurimo</i>	
Self-Supervised Learning for ASR Pre-Training with Uniquely Determined Target Labels and Controlling Cepstrum Truncation for Speech Augmentation	5048
<i>Akihiro Kato, Hiroyuki Nagano, Kohei Chike, Masaki Nose</i>	
MS-HuBERT: Mitigating Pre-Training and Inference Mismatch in Masked Language Modelling Methods for Learning Speech Representations	5053
<i>Hemant Yadav, Sunayana Sitaram, Rajiv Ratn Shah</i>	
Balanced-Wav2Vec: Enhancing Stability and Robustness of Representation Learning Through Sample Reweighting Techniques.....	5058
<i>Mun-Hak Lee, Jae-Hong Lee, Dohee Kim, Ye-Eun Ko, Joon-Hyuk Chang</i>	

SPOKEN TERM DETECTION AND SPEECH RETRIEVAL

Few-Shot Keyword Spotting from Mixed Speech.....	5063
<i>Junming Yuan, Ying Shi, Lantian Li, Dong Wang, Askar Hamdulla</i>	
Pretraining End-To-End Keyword Search with Automatically Discovered Acoustic Units	5068
<i>Bolaji Yusuf, Jan Honza Cernocky, Murat Saraçlar</i>	
2DP-2MRC: 2-Dimensional Pointer-Based Machine Reading Comprehension Method for Multimodal Moment Retrieval	5073
<i>Jiajun He, Tomoki Toda</i>	
GPA: Global and Prototype Alignment for Audio-Text Retrieval	5078
<i>Yuxin Xie, Zhihong Zhu, Xianwei Zhuang, Liming Liang, Zhichang Wang, Yuexian Zou</i>	
Few-Shot Keyword-Incremental Learning with Total Calibration	5083
<i>Ilseok Kim, Ju-Seok Seong, Joon-Hyuk Chang</i>	
Leveraging Speech Data Diversity to Document Indigenous Heritage and Culture.....	5088
<i>Allahsera Tapo, Éric Le Ferrand, Zoey Liu, Christopher Homan, Emily Prud'Hommeaux</i>	

SPEECH DISORDERS 1

Missingness-Resilient Video-Enhanced Multimodal Disfluency Detection.....	5093
<i>Payal Mohapatra, Shamika Likhite, Subrata Biswas, Bashima Islam, Qi Zhu</i>	
AS-70: A Mandarin Stuttered Speech Dataset for Automatic Speech Recognition and Stuttering Event Detection	5098
<i>Rong Gong, Hongfei Xue, Lezhi Wang, Xin Xu, Qisheng Li, Lei Xie, Hui Bu, Shaomei Wu, Jiaming Zhou, Yong Qin, Binbin Zhang, Jun Du, Jia Bin, Ming Li</i>	
Analyzing Speech Motor Movement Using Surface Electromyography in Minimally Verbal Adults with Autism Spectrum Disorder	5103
<i>Wazeer Zulfikar, Nishat Protyasha, Camila Canales, Heli Patel, James Williamson, Laura Sarnie, Lisa Nowinski, Nataliya Kosmyna, Paige Townsend, Sophia Yuditskaya, Tanya Talkar, Utkarsh Oggy Sarawgi, Christopher McDougale, Thomas Quatieri, Pattie Maes, Maria Mody</i>	
Prosody of Speech Production in Latent Post-Stroke Aphasia	5108
<i>Cong Zhang, Tong Li, Gayle Dede, Christos Salis</i>	
MMSD-Net: Towards Multi-Modal Stuttering Detection	5113
<i>Liangyu Nie, Sudarsana Reddy Kadiri, Ruchit Agrawal</i>	
Large Language Models for Dysfluency Detection in Stuttered Speech	5118
<i>Dominik Wagner, Sebastian P. Bayerl, Ilja Baumann, Elmar Noeth, Korbinian Riedhammer, Tobias Bocklet</i>	

CONNECTING SPEECH-SCIENCE AND SPEECH-TECHNOLOGY FOR CHILDREN'S SPEECH (SPECIAL SESSION)

Preliminary Investigation of Psychometric Properties of a Novel Multimodal Dialog Based Affect Production Task in Children and Adolescents with Autism.....	5123
<i>Carly Demopoulos, Linnea Lampinen, Cristian Preciado, Hardik Kothare, Vikram Ramanarayanan</i>	

Training Speech-Breathing Coordination in Computer-Assisted Reading	5128
<i>Delphine Charuau, Andrea Briglia, Erika Godde, Gérard Bailly</i>	
How Does Alignment Error Affect Automated Pronunciation Scoring in Children's Speech?.....	5133
<i>Prad Kadambi, Tristan Mahr, Lucas Annear, Henry Nomeland, Julie Liss, Katherine Hustad, Visar Berisha</i>	
Examining Vocal Tract Coordination in Childhood Apraxia of Speech with Acoustic-To-Articulatory Speech Inversion Feature Sets	5138
<i>Nina R. Benway, Jonathan L. Preston, Carol Espy-Wilson</i>	
Children's Speech Recognition Through Discrete Token Enhancement	5143
<i>Vrunda N. Sukhadia, Shammur Absar Chowdhury</i>	
Bridging Child-Centered Speech Language Identification and Language Diarization Via Phonetics.....	5148
<i>Yujia Wang, Hexin Liu, Leibny Paola Garcia</i>	
Reading Miscue Detection in Primary School Through Automatic Speech Recognition.....	5153
<i>Lingyun Gao, Cristian Tejedor-Garcia, Helmer Strik, Catia Cucchiarini</i>	
Automatic Evaluation of a Sentence Memory Test for Preschool Children	5158
<i>Ilja Baumann, Nicole Unger, Dominik Wagner, Korbinian Riedhammer, Tobias Bocklet</i>	
Enhancing Child Vocalization Classification with Phonetically-Tuned Embeddings for Assisting Autism Diagnosis	5163
<i>Jialu Li, Mark Hasegawa-Johnson, Karrie Karahalios</i>	
Self-Supervised Models for Phoneme Recognition: Applications in Children's Speech for Reading Learning	5168
<i>Lucas Block Medin, Thomas Pellegrini, Lucile Gelin</i>	
Benchmarking Children's ASR with Supervised and Self-Supervised Speech Foundation Models.....	5173
<i>Ruchao Fan, Natarajan Balaji Shankar, Abeer Alwan</i>	
Introduction to Partial Fine-Tuning: A Comprehensive Evaluation of End-To-End Children's Automatic Speech Recognition Adaptation.....	5178
<i>Thomas Rolland, Alberto Abad</i>	
Improving Child Speech Recognition with Augmented Child-Like Speech	5183
<i>Yuan Yuan Zhang, Zhengjun Yue, Tanvina Patel, Odette Scharenborg</i>	
Mixed Children/Adult/Childrenized Fine-Tuning for Children's ASR: How to Reduce Age Mismatch and Speaking Style Mismatch	5188
<i>Thomas Graave, Zhengyang Li, Timo Lohrenz, Tim Fingscheidt</i>	
Exploring Speech Foundation Models for Speaker Diarization in Child-Adult Dyadic Interactions.....	5193
<i>Anfeng Xu, Kevin Huang, Tiantian Feng, Lue Shen, Helen Tager-Flusberg, Shrikanth Narayanan</i>	

SHOW AND TELL 4

IIITH Uchar e-Sudharak: An Automatic English Pronunciation Corrector for School-Going Children with a Teacher in the Loop	5198
<i>Meenakshi Sirigiraju, Arjun Rajasekar, Abhishikth Meejuri, Chiranjeevi Yarra</i>	

Speech Enabled Visual Acuity Test	5200
<i>Boon Peng Yap, Kok Liang Tan, Zhenghao Li, Rong Tong</i>	
A ChatGPT-Based Oral Q&A Practice System for First-Time Student Participants in International Conferences	5202
<i>Mayuko Aiba, Daisuke Saito, Nobuaki Minematsu</i>	
Visual Scene Display Application for Augmentative and Alternative Communication.....	5204
<i>Karthik Venkat Sridaran, Raja Praveen, Reuben T Varghese, Ajish K Abraham, Shankar R, Winnie Rachel Cherian</i>	
CALL System Using Pitch-Accent Feature Representations Reflecting Listeners' Subjective Adequacy.....	5206
<i>Ikuyo Masuda-Katsuse, Ayako Shirose</i>	
The Speech Motor Chaining Web App for Speech Motor Learning.....	5208
<i>Jonathan L Preston, Nina R Benway, Nathan Prestopnik, Nathan Preston</i>	
Visualization for Improving Foreign Language Pronunciation	5210
<i>Charlotte Yoder, Karrie Karahalios, Mark Hasegawa-Johnson, Shreyansh Agrawal</i>	
CaptainA Self-Study Mobile App for Practising Speaking: Task Completion Assessment and Feedback with Generative AI	5212
<i>Nhan Phan, Anna Von Zansen, Maria Kautonen, Tamás Grósz, Mikko Kurimo</i>	

Author Index