

2024 14th International Symposium on Chinese Spoken Language Processing (ISCSLP 2024)

**Beijing, China
7-10 November 2024**



**IEEE Catalog Number: CFP24583-POD
ISBN: 979-8-3315-1683-3**

**Copyright © 2024 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP24583-POD
ISBN (Print-On-Demand):	979-8-3315-1683-3
ISBN (Online):	979-8-3315-1682-6

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Content

About the Conference	XVI
Message from the Conference Chairs	XVII
Message from the Program Chairs	XVIII
Conference Committees	XIX
Keynote Speeches	XXI
Tutorials	XXV
Industry Forum	XXIX
Author Index	XXXI
Papers	
Oral 1: Automatic Speech Recognition	
OS1.1 Evaluating Speech Intelligibility for Cochlear Implants Using Automatic Speech Recognition <i>Hengzhi Zhou , Mingyue Shi , Qinglin Meng</i>	<i>N/A</i>
OS1.2 OPC-KWS:Optimizing Keyword Spotting with Path Retrieval Decoding and Contrastive Learning <i>Jingchen Li , Xin Liu , Xueliang Zhang</i>	<i>6</i>
OS1.3 CUSIDE-T: Chunking, Simulating Future and Decoding for Transducer based Streaming ASR <i>Wenbo Zhao , Ziwei Li , Chuan Yu , Zhijian Ou</i>	<i>11</i>
OS1.4 Dual-strategy Fusion Method in Noise-Robust Speech Recognition <i>Jiahao Li , Cunhang Fan , Enrui Liu , Jian Zhou , Zhao Lv</i>	<i>16</i>
OS1.5 A Preliminary Study on Constructing Mandarin Personalized Speech Recognition Systems for the Speech Impaired <i>Jia-Jyu Su , Chen Yu Chiang , Yue-Shan Chang , Chao-Yin Lin , Jiunn-Horng Kang , Min-Yuh Day</i>	<i>21</i>
OS1.6 Unveiling the Potential of LLM-Based ASR on Chinese Open-Source Datasets <i>Xuelong Geng , Tianyi Xu , Kun Wei , Bingshen Mu , Hongfei Xue , He Wang , Yangze Li , Pengcheng Guo , Yuhang Dai , Longhao Li , Mingchen Shao , Lei Xie</i>	<i>26</i>
Special Session 1: Recent Advances in speech encoding and decoding based on the brain signals	
SS1.1 Naturalistic Language-related Movie-Watching fMRI Task for Detecting Neurocognitive Decline and Disorder <i>Yuejiao Wang , Xianmin Gong , Xixin Wu , Patrick Wong , Helene Fung , Man-Wai Mak , Helen Meng</i>	<i>31</i>

SS1.2 A Study on Depression Detection through Explainable Features of Speech <i>Di Zhou , Daisuke Mizuguchi , Takeshi Yamamoto , Yasuhiro Omiya</i>	36
SS1.3 A Hybrid DFSMN and Mamba Architecture for Low Bitrate Neural Speech Coding <i>Yuhao Zhao , Maoshen Jia , Jiawei Ru , Junqi Tai</i>	41
SS1.4 Binary Temporal Convolutional Neural Network for Multi-class Auditory Spatial Attention Detection <i>Peng Zhao , Ruicong Wang , Xueyi Zhang , Mingrui Lao , Siqi Cai</i>	46
SS1.5 Robustness and Generalization Capability Validation of Convolutional Neural Network on a Chinese EEG Auditory Attention Decoding Dataset <i>Yuanming Zhang , Zeyan Song , Haoliang Du , Xia Gao , Jing Lu</i>	51
SS1.6 Bi-directional Oscillatory Interaction in the Neural Networks Engaged in Sentence Oral Reading <i>Bin Zhao , Gaoyan Zhang , Jianwu Dang , Aijun Li</i>	56
SS1.7 Spatial Attention in Interfering Speech Perception <i>Zelin Qiu , Junfeng Li , Yingyi Luo</i>	61
Special Session 2: Recent Advances in Speech Processing and Binaural Signal Processing for VR Conference	
SS2.1 The Feasibility of Sound Zone Control Using an Array of Parametric Array Loudspeakers <i>Tao Zhuang , Jiaxin Zhong , Jing Lu</i>	66
SS2.2 A Differential Quantization Based end-to-end Neural Speech Codec <i>Pincheng Lu , Liang Xu , Jing Wang</i>	71
SS2.3 Non-Intrusive Audio Quality Assessment Based on Deep Neural Network for Subjective MOS Prediction <i>Xinwen Yue , Yupei Zhang , Jianqian Zhang , Zhiyu Li , Jing Wang , Shenghui Zhao</i>	76
SS2.4 EnvFake: An Initial Environmental-Fake Audio Dataset for Scene-Consistency Detection <i>Kangyue Li , Haonan Cheng , Long Ye , Jingling Wang</i>	81
SS2.5 The Impact of Dynamic Cue and Audio Stimulus Type on Subjective Localization in VR Headsets <i>Cheng Chi , Xiaoyu Li , Rui Zhang , Xiaodong Li , Chengshi Zheng</i>	86
Oral 2: General Speech Science and Linguistics	
OS2.1 A Fresh Review on Chinese Pronunciation Acquisition: Insights and Recommendations for L2 Foreign Children <i>Mewlude Nijat , Dong Wang , Askar Hamdulla</i>	91
OS2.2 Multiple Patterns of Merging Guangzhou Cantonese Tones in Production and Perception: Study on Youth Groups <i>Lishan Li , Xiaoying Xu</i>	96

OS2.3 Production of Mandarin Chinese R-suffix by Mandarin-speaking Children: A preliminary study <i>Shuwen Chen , Jun Gao , Zixuan Jia</i>	101
--	-----

OS2.4 A Study on the Effectiveness of Mandarin Seven-sound Test across Multiple Speakers <i>Yu Chen , Yu Bai , Ju Zhang</i>	106
---	-----

OS2.5 Gradiency in Obstruent Devoicing in the Varieties of Wu Chinese <i>Fang Hu</i>	111
--	-----

OS2.6 Acoustic Features of Consonants Production in Standard Chinese by Primary School Teachers <i>Yuan Jia , Xintong Zuo</i>	116
---	-----

Special Session 3: Multimodal Dysarthria Severity Classification Challenge

SS3.1 Tongue Model-Driven Method Based on Fully Connected Neural Network <i>Shaochuan Zhang , Fengji Li , Li Wang , Jie Zhou , Haijun Niu</i>	121
---	-----

SS3.2 Multi-Modal Dysarthria Severity Assessment Using Dual-Branch Feature Decoupling Network and Mixed Expert Framework <i>Shuang Liang , Yu Gu</i>	126
--	-----

SS3.3 The Open-Access Mandarin Subacute Stroke Dysarthria Multimodal (MSDM) Database for Intelligent Assessment <i>Juan Liu , Xiaokang Liu , Yudong Yang , Rukiye Ruzi , Xiaoyi Zuo , Changqing Xu , Chaojinzi Li , Xinyu Li , Rongfeng Su , Anming Hu , Yumei Zhang , Shaofeng Zhao , Xiaoxia Du , Lan Wang , Nan Yan</i>	131
--	-----

SS3.4 The ISCSLP 2024 Multimodal Dysarthria Severity Assessment (MDSA) Challenge : Dataset, Tracts, Baseline and Results <i>Juan Liu , Yudong Yang , Xiaokang Liu , Xiaoyi Zuo , Junfeng Li , Lan Wang , Nan Yan</i> ..	136
---	-----

SS3.5 Investigation of Cross Modality Feature Fusion for Audio-Visual Dysarthric Speech Assessment <i>Yicong Jiang , Youjun Chen , Tianzi Wang , Zengrui Jin , Xurong Xie , Hui Chen , Xunying Liu , Feng Tian</i>	141
--	-----

SS3.6 Constant Q Transform for Audio-visual Dysarthria Severity Assessment <i>Guolun Sun , Li Wang</i>	146
--	-----

SS3.7 Fine-Tuning Pre-Trained Audio Models for Dysarthria Severity Classification: A Second Place Solution in the Multimodal Dysarthria Severity Classification Challenge <i>Wei Dai , Menglong Li , Yingqi He , Yongqiang Zhu</i>	151
--	-----

Poster Session 1: Speech Synthesis, Enhancement, and Conversion

PS1.1 Arti-Invar: A Pre-trained Model for Enhancing Acoustic-to-Articulatory Inversion Performance <i>Yifeng Sun , Yanlu Xie , Jinsong Zhang , Dengfeng Ke</i>	154
--	-----

PS1.2 CATAD: Conformer-Based Adversarial Training with Adaptive Diffusion for Bone-Conducted Speech Enhancement <i>Zhiqiang Duan , Jian Zhou , Cunhang Fan , Liang Tao , Zhao Lv</i>	159
PS1.3 RobustSVC: HuBERT-based Melody Extractor and Adversarial Learning for Robust Singing Voice Conversion <i>Wei Chen , Xintao Zhao , Jun Chen , Binzhu Sha , Zhiwei Lin , Zhiyong Wu</i>	164
PS1.4 Wav2f0: Exploring the Potential of Wav2vec 2.0 for Speech Fundamental Frequency Extraction <i>Rui Feng , Yin-Long Liu , Zhen-Hua Ling , Jiahong Yuan</i>	169
PS1.5 GM-LPC Based Multiband Analysis and Enhancement of Pathological Voice <i>Yaqin Wu , Yan Chang , Yanzhang Geng , Xiaofeng Cao , Jiawei Zhao</i>	174
PS1.6 Knowledge Distillation from Discriminative Model to Generative Model with Parallel Architecture for Speech Enhancement <i>Tingxiao Zhou , Leying Zhang , Yanmin Qian</i>	179
PS1.7 COMOSVC: Consistency Model-based Singing Voice Conversion <i>Yiwen Lu , Zhen Ye , Wei Xue , Xu Tan , Qifeng Liu , Yike Guo</i>	184
PS1.8 Enhancing Mispronunciation Detection with WavLM and Mixture-of-Experts Network <i>Wenqian Bao , Yan Yuchen , Jinsong Zhang</i>	189
PS1.9 Quantization-Error-Free Soft Label for 2D Sound Source Localization <i>Linfeng Feng , Xiaolei Zhang , Xuelong Li</i>	194
PS1.10 SAMOS: A Neural MOS Prediction Model Leveraging Semantic Representations and Acoustic Features <i>Yu-Fei Shi , Ye-Xin Lu , Yang Ai , Hui-Peng Du , Zhen-Hua Ling</i>	199
Oral 3: Multimodal Processing	
OS3.1 Exploring the Role of Audio in Multimodal Misinformation Detection <i>Moyang Liu , Yukun Liu , Ruibo Fu , Zhengqi Wen , Jianhua Tao , Xuefei Liu , Guanjun Li</i>	204
OS3.2 Independent Feature Enhanced Crossmodal Fusion for Match-Mismatch Classification of Speech Stimulus and EEG Response <i>Shitong Fan , Wenbo Wang , Feiyang Xiao , Shiheng Zhang , Qiaoxi Zhu , Jian Guan</i>	209
OS3.3 G2DiaR: Enhancing Commonsense Reasoning of LLMs with Graph-to-Dialogue & Reasoning <i>Huijun Lian , Keqi Chen , Zekai Sun , Yingming Gao , Ya Li</i>	214
OS.4 Representation of Articulatory Features in EEG During Speech Production Tasks <i>Sinan Sun , Longxiang Zhang , Bo Wang , Xihong Wu , Jing Chen</i>	219
OS3.5 Encoding and Decoding of Chinese Phonemes Based on MEG Signals <i>Jinghua Liang , Bo Wang , Xihong Wu , Jing Chen</i>	224

OS3.6 Bridging Facial Imagery and Vocal Reality: Stable Diffusion-Enhanced Voice Generation Yueqian Lin , Dong Liu , Yunfei Xu , Hongbin Suo , Ming Li	229
--	-----

Special Session 2: AI-Powered Advances in Speech, Auditory, and Cognitive Processing

PS2.1 Advancing Music Therapy: Integrating Eastern Five-Element Music Theory and Western Techniques with AI in the Novel Five-Element Harmony System Weizhen Bian , Yubo Zhou , Kaitai Zhang , Xiaohan Gu	234
---	-----

PS2.2 A Study of Brain Mechanisms by Which Sound Source Location and Amount of Masking Affect Target Perception Dawei Xiang , Yong Ma , Yiming Yang	239
---	-----

PS2.3 Unification of Balti and Trans-border Sister Dialects in the Essence of LLMs and AI Technology Muhammad Sharif , Jiangyan Yi , Muhammad Shoaib	244
--	-----

PS2.4 An Optimizer for Conformer based on Conjugate Gradient Method Wenyi Yu , Chao Zhang	249
---	-----

PS2.5 Not All Errors Are Equal: Investigation of Speech Recognition Errors in Alzheimer's Disease Detection Jiawen Kang , Junan Li , Jinchao Li , Xixin Wu , Helen Meng	254
---	-----

PS2.6 A Noval Feature via Color Quantisation for Fake Audio Detection Zhiyong Wang , Xiaopeng Wang , Yuankun Xie , Ruibo Fu , Zhengqi Wen , Jianhua Tao , Yukun Liu , Guanjun Li , Xin Qi , Yi Lu , Xuefei Liu , Yongwei Li	259
---	-----

PS2.7 Low-Resourced Speech Recognition for Iu Mien Language via Weakly-Supervised Phoneme-based Multilingual Pre-training Lukuan Dong , Donghong Qin , Fengbo Bai , Fanhua Song , Yan Liu , Chen Xu , Zhijian Ou	264
--	-----

PS2.8 Individual Differences in Tone Perception and Production in Emerging Dialect: A Case Study of Elementary School Children in Changsha Yi Zhang , Lishan Li , Xiaoying Xu	269
---	-----

PS2.9 The Development of Speech Rhythm in Mandarin-speaking Children Aijun Li , Zhiwei Wang , Sichen Zhang , Jun Gao	274
--	-----

PS2.10 Cross-language Forensic Voice Comparison of Hong Kong Trilingual Speakers Using Filled Pauses and an Automatic Speaker Recognition System Grace Wenling Cao , Vincent Hughes , Bruce Wang , Peggy Mok	N/A
--	-----

Oral 4: Speech Synthesis and Generation

OS4.1 Diffusion-Based Method with TTS Guidance for Foreign Accent Conversion Qibing Bai , Shuai Wang , Zhijun Liu , Mingyang Zhang , Wei Rao , Yannan Wang , Haizhou Li	284
---	-----

OS4.2 A Systematic Exploration of Joint-training for Singing Voice Synthesis <i>Yuning Wu , Yifeng Yu , Jiatong Shi , Tao Qian , Qin Jin</i>	289
--	-----

OS4.3 EELE: Exploring Efficient and Extensible LoRA Integration in Emotional Text-to-Speech <i>Xin Qi , Ruibo Fu , Zhengqi Wen , Jianhua Tao , Zhiyong Wang , Xiaopeng Wang , Shuchen Shi , Yi Lu , Yuankun Xie , Yukun Liu , Guanjin Li , Xuefei Liu , Yongwei Li</i>	294
--	-----

OS4.4 FCTalker: Fine and Coarse Grained Context Modeling for Expressive Conversational Speech Synthesis <i>Yifan Hu , Rui Liu , Guanglai Gao , Haizhou Li</i>	299
---	-----

OS4.5 ExpressiveSinger: Synthesizing Expressive Singing Voices as An Instrument <i>Fengping Wang , Bingsong Bai , Yayue Deng , Jinlong Xue , Yingming Gao , Ya Li</i>	304
---	-----

OS4.6 LCM-SVC: Latent Diffusion Model Based Singing Voice Conversion with Inference Acceleration via Latent Consistency Distillation <i>Shihao Chen , Yu Gu , Jianwei Cui , Jie Zhang , Rilin Chen , Lirong Dai</i>	309
---	-----

Special Session 4: The first Chinese Auditory Attention Decoding Challenge

SS4.1 StreamAAD: Decoding Spatial Auditory Attention with a Streaming Architecture <i>Zelin Qiu , Dingding Yao , Junfeng Li</i>	314
---	-----

SS4.2 The NERCSLIP-USTC System for Track2 of The First Chinese Auditory Attention Decoding Challenge <i>Yubang Zhang , Jie Zhang , Zhen-Hua Ling</i>	319
--	-----

SS4.3 ViT-Based EEG Analysis Method for Auditory Attention Detection <i>Fengyu Xu , Yongxiong Xiao , Qiang Fu</i>	324
---	-----

SS4.4 The NWPU-BBIC System for The ISCSLP 2024 Chinese Auditory Attention Decoding Challenge <i>Qilong Yuan , Di Zhu , Enze Shi , Kui Zhao</i>	N/A
--	-----

SS4.5 Enhancing Spatio-Temporal Auditory Attention Decoding with ST-AADNet <i>Ruofan Yan , Shu Peng , Zhige Chen , Zhi-An Huang , Rui Liu , Kay Chen Tan , Jibin Wu</i>	334
---	-----

SS4.6 Ensemble Deep Learning Models for EEG-based Auditory Attention Decoding <i>Peng Zhao , Zijie Lin , Ruicong Wang , Zexu Pan , Haizhou Li , Xueyi Zhang</i>	339
---	-----

Special Session 5: Multimodal Large Language Models in Human – Robot Interaction

SS5.1 Mitigating Hallucination in Visual Language Model Segmentation with Negative Sampling <i>Chung Tong Lee , Jianxin Pang , Jun Cheng</i>	344
--	-----

SS5.2 Fusion Pruning for Large Language Models <i>Shixin Jiang , Ming Liu , Bing Qin</i>	349
--	-----

SS5.3 Semantic Search Using LLM-aided Topic Generation on Knowledge Graphs for Paper Discovery	
---	--

<i>Sabrina Chow , Lilian Guo , Jonathan Chow , Chelsea Chia , Sarah Li , Dong-Yan Huang</i>	<i>353</i>
---	------------

SS5.4 Empowering Robots with Multimodal Language Models for Task Planning with Interaction <i>Chung Tong Lee , Jianxin Pang , Jun Cheng</i>	<i>358</i>
---	------------

SS5.5 Zh-Paral: Benchmark Dataset for Comprehension of Chinese Paralinguistic Speech <i>Dekun Chen , Zhizheng Wu</i>	<i>363</i>
--	------------

Oral 5: Speaker Diarization and Verification

OS5.1 Speaker Diarization for Unlimited Number of Speakers using Dynamic Linear <i>Siyin Wang , Chao Zhang</i>	<i>368</i>
--	------------

OS5.2 Online Neural Speaker Diarization with Spectral Clustering for Meeting Scenarios <i>Tianyou Cheng , Mao-Kui He , Gaobin Yang , Shutong Niu , Yanqiang Lei , Limei Peng , Jun Du</i>	<i>373</i>
---	------------

OS5.3 Optimizing Deep Speaker Embeddings with a Dynamic Cross Triplet Framework <i>Jin Li , Lirong Dai</i>	<i>378</i>
--	------------

OS5.4 Prototype and Instance Contrastive Learning for Unsupervised Domain Adaptation in Speaker Verification <i>Wen Huang , Bing Han , Zhengyang Chen , Shuai Wang , Yanmin Qian</i>	<i>383</i>
--	------------

OS5.5 An Investigation of Reprogramming for Cross-Language Adaptation in Speaker Verification Systems <i>Jingyu Li , Aemon Yat Fei Chiu , Tan Lee</i>	<i>388</i>
---	------------

OS5.6 LE-CAM++: A Lighter and More Efficient CAM++ for Speaker Verification <i>Shuanghong Liu , Zhida Song , Zhihua Fang , Liang He</i>	<i>393</i>
---	------------

Special Session 6: Induced Emotion Recognition with Personality Characteristics Challenge 2024

SS6.1 Speech Emotion Recognition Based on Shallow Structure of Wav2vec 2.0 and Attention Mechanism <i>Yumei Zhang , Maoshen Jia , Xuan Cao , Zichen Zhao</i>	<i>398</i>
--	------------

SS6.2 Speech Emotion Recognition Using Fine-Tuned DWFormer: A Study on Track 1 of the IERP Challenge 2024 <i>Honghong Wang , Xupeng Jia , Jing Deng , Rong Zheng</i>	<i>403</i>
--	------------

SS6.3 Personalized Multimodal Emotion Recognition: Integrating Temporal Dynamics and Individual Traits for Enhanced Performance <i>Binqiang Wang , Gang Dong , Yaqian Zhao , Rengang Li</i>	<i>408</i>
---	------------

SS6.4 IERP 2024: Induced Emotion Recognition with Personality Characteristics Challenge 2024 <i>Cong Cai , Shan Liang , Xuefei Liu , Kang Zhu , Zhenhua Cheng , Zheng Lian , Bin Liu ...</i>	<i>413</i>
--	------------

Special Session 3: Cross-Modal and Multimodal Speech Systems

PS3.1 EmoSpeech: A Corpus of Emotionally Rich and Contextually Detailed Speech Annotations <i>Weizhen Bian , Yobo Zhou , Kaitai Zhang , Xiaohan Gu</i>	<i>417</i>
PS3.2 Inferring Agent Speaking Styles for Auditory-Visual User-Agent Conversation <i>Zhihan Yang , Chunfeng Wang , Zhiyong Wu , Jia Jia</i>	<i>421</i>
PS3.3 Quantitative Analysis of Audio-Visual Tasks: An Information-Theoretic Perspective <i>Chen Chen , Xiaolou Li , Zehua Liu , Lantian Li , Dong Wang</i>	<i>426</i>
PS3.4 Transferring Personality Knowledge to Multimodal Sentiment Analysis <i>Kang Zhu , Xuefei Liu , Heng Xie , Cong Cai , Ruibo Fu , Guanjun Li , Zhengqi Wen , Jianhua Tao , Cunhang Fan , Zhao Lv , Le Wang , Hao Lin</i>	<i>431</i>
PS3.5 An Empirical Study of Retrieval Augmented Generation with Chain-of-Thought <i>Yuetong Zhao , Hongyu Cao , Xianyu Zhao , Zhijian Ou</i>	<i>436</i>
PS3.6 NLPP: A Natural Language Prosodic Prominence Dataset Assisted by ChatGPT <i>Rui Niu , Changhe Song , Zhiyong Wu</i>	<i>441</i>
PS3.7 Temporal Shift for Personality Recognition with Pre-trained Representations <i>Hanzhe Xu , Xuefei Liu , Cong Cai , Kang Zhu , Jizhou Cui , Ruibo Fu , Heng Xie , Jianhua Tao , Zhengqi Wen , Ziping Zhao , Guanjun Li , Le Wang , Hao Lin.....</i>	<i>446</i>
PS3.8 CSDA: Cross-Session Domain Adaptation in Auditory Attention Decoding of EEG for a Single Subject <i>Xingguang Dong , Cunhang Fan , Hongyu Zhang , Xiaoke Yang , Sheng Zhang , Jian Zhou , Zhao Lv</i>	<i>451</i>
PS3.9 ERVQ: Leverage Residual Vector Quantization for Speech Emotion Recognition <i>Jingran Xie , Yang Xiang , Hui Wang , Xixin Wu , Zhiyong Wu , Helen Meng</i>	<i>456</i>
PS3.10 Cross-Subject Domain Adaptation for EEG-Based Auditory Attention Decoding via Prototypical Representation <i>Xiaoke Yang , Cunhang Fan , Hongyu Zhang , Xingguang Dong , Jian Zhou , Xinhui Li , Zhao Lv.....</i>	<i>461</i>
Oral 6: Deepfake Detection and Disease Detection	
OS6.1 MADD: A Multi-lingual Multi-speaker Audio Deepfake Detection Dataset <i>Xiaoke Qi , Hao Gu , Jiangyan Yi , Jianhua Tao , Yong Ren , Jiayi He , Siding Zeng.....</i>	<i>466</i>
OS6.2 Devising a Set of Compact and Explainable Spoken Language Feature for Screening Alzheimer's Disease <i>Junan Li , Yunxiang Li , Yuren Wang , Xixin Wu , Helen Meng</i>	<i>471</i>
OS6.3 Reject Threshold Adaptation for Open-Set Model Attribution of Deepfake Audio <i>Xinrui Yan , Jiangyan Yi , Jianhua Tao , Yujie Chen , Hao Gu , Guanjun Li , Junzuo Zhou , Yong Ren , Tao Xu</i>	<i>476</i>
OS6.4 Does Current Deepfake Audio Detection Model Effectively Detect ALM-based Deepfake Audio? <i>Yuankun Xie , Chenxu Xiong , Xiaopeng Wang , Zhiyong Wang , Yi Lu , Xin Qi , Ruibo Fu , Yukun Liu , Zhengqi Wen , Jianhua Tao , Guanjun Li , Long Ye</i>	<i>481</i>

OS6.5 Leveraging Prompt Learning and Pause Encoding for Alzheimer's Disease Detection <i>Yin-Long Liu , Rui Feng , Jiahong Yuan , Zhen-Hua Ling</i>	486
---	-----

OS6.6 Cross-Lingual Alzheimer's Disease Detection Based on Scale Criteria <i>Yuancheng Wang , Haoran Zheng , Qi Sun , Yong Ma , Shihu Zhu , Le Zhang , Wei-Qiang Zhang</i>	491
--	-----

Special Session 7: Conversational Voice Clone Challenge

SS7.1 The Codec Language Model-based Zero-Shot Spontaneous Style TTS System for CoVoC Challenge 2024 <i>Shuoyi Zhou , Yixuan Zhou , Weiqin Li , Jun Chen , Runchuan Ye , Weihao Wu , Zijian Lin , Shun Lei , Zhiyong Wu</i>	496
---	-----

SS7.2 Lightweight Language Model for Speech Synthesis: Attempts and Analysis <i>Zhuojun Wu , Dong Liu , Ming Li</i>	501
---	-----

SS7.3 The ISCSLP 2024 Conversational Voice Clone (CoVoC) Challenge: Tasks, Results and Findings <i>Kangxiang Xia , Dake Guo , Liumeng Xue , Jixun Yao , Hanzhao Li , Shuai Wang , Zhao Guo , Lei Xie , Lei Luo , Qingqing Zhang , Minghui Dong , Peng Sun</i>	506
---	-----

Special Session 4: Speech Prosody, Tones, and Perception

PS4.1 Focus and Gender Affect Sentence Type Perception: Observations on Mandarin Sentence-Final Particle Ba <i>Qixin Li , Gaowu Wang</i>	N/A
--	-----

PS4.2 Contributions of Acoustic Factors to Tone Identification in Whispered Mandarin <i>Fuqian Wu , Xiyu Wu</i>	516
---	-----

PS4.3 Speech Rate Influence on Rhythm Alterations in Mandarin <i>Zhongxuan Mao , Chenyu Li , Shanpeng Li</i>	521
--	-----

PS4.4 Analysis of Normal and Slow Speech Rate on the F0 Contour of Tones in Mandarin Broadcasting Speech <i>Chenyu Li , Zhongxuan Mao , Shanpeng Li</i>	526
---	-----

PS4.5 The Contributions of Formants to the Intelligibility in Uyghur Sine-wave Sentences <i>Xiaoming Liang , Zhihua Huang</i>	531
---	-----

PS4.6 The Role of F0 in the Recognition of Aspiration Contrasts in Mandarin <i>Chong Cao , Qian Li</i>	536
--	-----

PS4.7 Statistical Analysis of F0 Characteristics of "Grade A Level 1" Mandarin Tones: on the Application of the T-Value Method <i>Xiyao Lu , Yukai Wan , Ruishan Li , Jinsong Zhang</i>	541
---	-----

PS4.8 A Study on the Information Mechanism of the 3rd Tone Sandhi Rule in Mandarin Across Word Boundaries <i>Xiaowang Liu , Xiaolin Wu , Jinsong Zhang</i>	546
--	-----

PS4.9 Categorical Perception of Tone 2 and Tone 3 of Standard Chinese by Bilingual Korean Ethnic Speakers in China <i>Shuang Zhou , Yinghao Li</i>	551
--	-----

PS4.10 The Effect of Focus Position on Downstep in Chinese Non-interrogative Sentences <i>Yuting Zhang , Xiaoying Xu</i>	556
--	-----

Oral 7: Speech Emotion Analysis and Speech enhancement

OS7.1 Insights from Hyperparameter Scaling of Online Speech Separation <i>Xin Zhou , Wangyou Zhang , Chenda Li , Yanmin Qian</i>	561
--	-----

OS7.2 Unlocking the Power of Emotions: Enhancing Personality Trait Recognition through Utilization of Emotional Cues <i>Jizhou Cui , Xuefei Liu , Yongwei Li , Xiaoying Xu , Ruibo Fu , Jianhua Tao , Zhengqi Wen , Yukun Liu , Guanjin Li , Le Wang , Hao Lin</i>	566
--	-----

OS7.3 A Spectral Change Enhancement Method Based on Self-supervised Learning Framework <i>Nan Li , Yadong Niu , Liushuai Yuan , Xihong Wu , Jing Chen</i>	571
---	-----

OS7.4 Fast Sampling Based on Policy Gradient for Diffusion-based Speech Enhancement <i>Yubo Jiang , Zhihua Huang</i>	576
--	-----

OS7.5 TF-DiffuSE: Time-Frequency Prior-Conditioned Diffusion Model for Speech Enhancement <i>Wenjun Ding , Xinsheng Wang , Lijian Gao , Qirong Mao</i>	581
--	-----

OS7.6 E-chat: Emotion-sensitive Spoken Dialogue System with Large Language Models <i>Hongfei Xue , Yuhao Liang , Bingshen Mu , Shiliang Zhang , Mengzhe Chen , Qian Chen , Lei Xie</i>	586
--	-----

Special Session 8: Inspirational and Convincing Audio Generation Challenge 2024 ICAGC 2024

SS8.1 Convincing Audio Generation Based on LLM and Speech Tokenization <i>Ruibo Liu , Shuaixin Wang , Zhuangzhuang Liu , Jiangjiang Zhao , Yuling Ren , Yu Liu</i> ...	591
--	-----

SS8.2 LLM-based Expressive Text-to-Speech Synthesizer with Style and Timbre Disentanglement <i>Yuanyuan Zhu , Jiaxu He , Ruihao Jing , Yaodong Song , Jie Lian , Xiaolei Zhang , Jie Li</i>	596
---	-----

SS8.3 Hola-TTS: A Cross-Lingual Zero-Shot Text-to-Speech System for Chinese, English, Japanese, and Korean <i>Hongwu Ding , Yiquan Zhou , Wenyu Wang , Jiacheng Xu , Jiaqi Mei</i>	601
--	-----

SS8.4 Expressive Speech Synthesis with Theme-oriented Few-shot Learning in ICAGC 2024 <i>Haoyu Wang , Tianrui Wang , Cheng Gong , Yu Jiang , Qiuyu Liu , Longbiao Wang , Jianwu Dang</i>	606
--	-----

SS8.5 Expressive Text-to-Speech with Contextual Background for ICAGC 2024 <i>Yu Jiang , Tianrui Wang , Haoyu Wang , Cheng Gong , Qiuyu Liu , Zikang Huang , Longbiao Wang , Jianwu Dang</i>	611
---	-----

SS8.6 The NPU-HWC System for the ISCSLP 2024 Inspirational and Convincing Audio Generation Challenge <i>Dake Guo , Jixun Yao , Xinfu Zhu , Kangxiang Xia , Zhao Guo , Ziyu Zhang , Yao Wang , Jie Liu , Lei Xie</i>	616
---	-----

SS8.7 Vivid Background Audio Generation based on Large Language Models and AudioLDM <i>Yiwei Liang , Ming Li</i>	621
--	-----

SS8.8 ICAGC 2024: Inspirational and Convincing Audio Generation Challenge 2024 <i>Ruibo Fu , Rui Liu , Qiang Chunyu , Yingming Gao , Yi Lu , Shuchen Shi , Tao Wang , Ya Li , Zhengqi Wen , Chen Zhang , Hui Bu , Yukun Liu , Xin Qi , Guanjun Li</i>	626
---	-----

Special Session 9: Speech Processing in LLM Era

SS9.1 Improving Grapheme-to-Phoneme Conversion through In-Context Knowledge Retrieval with Large Language Models <i>Dongrui Han , Mingyu Cui , Jiawen Kang , Xixin Wu , Xunying Liu , Helen Meng</i>	631
--	-----

SS9.2 Improving Emotion Recognition with Pre-trained Models, Multimodality, and Contextual Information <i>Zhengshun Xia , Ziyang Ma , Zhisheng Zheng , Xie Chen</i>	636
---	-----

SS9.3 The X-LANCE Technical Report for Interspeech 2024 Speech Processing Using Discrete Speech Unit Challenge <i>Yiwei Guo , Chenrun Wang , Yifan Yang , Hankun Wang , Ziyang Ma , Chenpeng Du , Shuai Wang , Hanzheng Li , Xu Li , Shuai Fan , Hui Zhang , Xie Chen , Kai Yu</i>	641
--	-----

SS9.4 Bridging Speech and Text: Enhancing ASR with Pinyin-to-Character Pre-training in LLMs <i>Yuhang Yang , Yizhou Peng , Eng Siong Chng , Xionghu Zhong</i>	646
---	-----

Oral 8: Speech Coding and Evaluation

OS8.1 An Exploration on Singing MOS Prediction <i>Yuxun Tang , Jiatong Shi , Yuning Wu , Qin Jin</i>	651
--	-----

OS8.2 CMAST: Efficient Speech-Text Joint Training Method to Enhance Linguistic Features Learning of Speech Representations <i>Jingran Xie , Changhe Song , Yang Xiang , Hui Wang , Xixin Wu , Zhiyong Wu , Helen Meng</i>	656
---	-----

OS8.3 A Dual-path Conformer-based Network for Neural Speech Coding <i>Jiawei Ru , Maoshen Jia , Yuhao Zhao , Liang Tao</i>	661
--	-----

OS8.4 Comparing Human-Labeled and LLM-Generated Semantic Features via Cortical Neural Representation <i>Boda Xiao , Bo Wang , Xuning Chen , Xiran Xu , Xihong Wu , Jing Chen</i>	666
--	-----

OS8.5 Acoustic Features at Intonational Phrase Boundaries: Comparative Study of Native Speakers and L2 Learners of Chinese Mandarin <i>Ruishan Li , Yanlu Xie</i>	671
---	-----

OS8.6 APCodec+: A Spectrum-Coding-Based High-Fidelity and High-Compression-Rate Neural Audio Codec with Staged Training Paradigm <i>Hui-Peng Du , Yang Ai , Rui-Chen Zheng , Zhen-Hua Ling</i>	676
--	-----

Poster Session 5: Speech Recognition and Speaker Verification

PS5.1 Band-wise Front-end Distortion Suppression for Robust Speech Recognition <i>Siyi Zhao , Wei Wang , Yanmin Qian</i>	681
--	-----

PS5.2 Wav2Nas: An Exploratory Approach to Nasalance Estimation in Speech <i>Rui Feng , Yuang Chen , Yin-Long Liu , Jiahong Yuan , Zhen-Hua Ling</i>	686
---	-----

PS5.3 Simplified Skip-connected UNet for Robust Speaker Verification under Noisy Environments <i>Zonghui Wang , Zhihua Fang , Zhida Song , Liang He</i>	691
---	-----

PS5.4 Layer-Adaptive Low-Rank Adaptation of Large ASR Model for Low-Resource Multilingual Scenarios <i>Yi Han , Hang Chen , Jun Du , Changqing Kong , Shifu Xiong , Jia Pan</i>	696
---	-----

PS5.5 Combining the Self-supervised Learning and Adversarial Training based Domain Adaptation for Speaker Verification <i>Zhengyang Chen , Shuai Wang , Bing Han , Yanmin Qian</i>	701
--	-----

PS5.6 Robust Coherent sources Localization Based on Hankel Matrix Reconstruction <i>Chengyuan Qin , Wenmeng Xiong , Maoshen Jia , Haoyang Zhou , Jing Zhang , Xianhong Chen , Qi Wang</i>	706
---	-----

PS5.7 ConMamba: A Convolution-Augmented Mamba Encoder Model for Efficient End-to-End ASR Systems <i>Haoliang Hou , Xun Gong , Yanmin Qian</i>	711
---	-----

PS5.8 Ensemble Knowledge Distillation from Speech SSL Models Considering Inter-teacher Differences <i>Pei Jun Liao , Hung-Yi Lee , Hsin-Min Wang</i>	716
--	-----

PS5.9 CUSIDE-array: A Streaming Multi-Channel End-to-End Speech Recognition System with Realistic Evaluations <i>Xiangzhu Kong , Tianqi Ning , Hao Huang , Zhijian Ou</i>	721
---	-----

PS5.10 Query-by-Example speech search using Mamba and Random Offset Mixed Padding <i>Kaixin Yang , Gaofeng Cheng , Ta Li , Qingwei Zhao , Yonghong Yan</i>	726
--	-----