2024 27th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA 2024)

Hsinchu City, Taiwan
17-19 October 2024

IEEE

IEEE Catalog Number: CFP2408H-POD
ISBN: 979-8-3315-0604-9

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:	CFP2408H-POD
ISBN (Print-On-Demand):	979-8-3315-0604-9
ISBN (Online):	979-8-3315-0603-2
ISSN:	2163-3479

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

Table of Contents

Oral Session #1: Speech Corpora

Session Chair: Yi-Wen Liu, National Tsing Hua University, Taiwan

Check Your Audio Data: Nkululeko for Bias Detection P. 1
Felix Burkhardt, Bagus Tris Atmaja, Anna Derington, Florian Eyben, and Bjoern Schuller

CAO Robot for Taiwanese/English Knowledge Graph Application P. 7
Chang-Shing Lee, Mei-Hui Wang, Guan-Ying Tseng, Chao-Cyuan Yue, Hao-Chun Hsieh, Marek Reformat

InStant-EMDB: A multi model spontaneous English and Malayalam speech corpora for depression detection P. 13
Anjali Mathew, Raniya, Harsha Sanjan, Amjith S B, Reni K Cherian, Starlet Ben Alex, Priyanka Srivastava, Chiranjeevi Yarra

Chinese Psychological Counseling Corpus Construction for Valence-Arousal Sentiment Intensity Prediction P. 19
Hsiu-Min Shih, Tzu-Mi Lin, Yu-Wen Tzeng, Jung-Ying Chang, Kuo-Kai Shyu, and Lung-Hao Lee

UCSYSpoof: A Myanmar Language Dataset for Voice Spoofing Detection P. 24
Hay Mar Soe Naing, Win Pa Pa, Aye Mya Hlaing, Myat Aye Aye Aung, Kasorn Galajit, and Candy Olivia Mawalim

Oral Session #2: Speech Corpora

Session Chair: Ming-Hsiang Su, Soochow University, Taiwan

Speech Watermarking for Tampering Detection using Singular Spectrum Analysis with Quantization Index Modulation and Psychoacoustic Model P. 29
Pantarat Vichathai, Puchit Bunpleng, Patharapol Laolakkana, Sasiporn Usanavasin, Phondanai Khanti, Kasorn Galajit, and Jessada Karnjana

IIITSAINT-EMOMDB: Carefully Curated Malayalam Speech Corpus with Emotion and Self-Reported Depression Ratings P. 35
Christa Thomas, Guneesh Vats, Aravind Johnson, Ashin George, Talit Sara George, Reni K Cherian, Priyanka Srivastava, and Chiranjeevi Yarra

CL-CHILD Corpus: The Phonological Development of Putonghua in Children from Dialect-speaking Regions P. 41
Jiewen Zheng, Tianxin Zheng, and Mengxue Cao

WikiTND24: A Chinese Text Normalization Database P. 47
Wu-Hao Li, and Chen Yu Chiang

Poster Session #1: Language Acquisition, Pronunciation, and Speech Learning

A Parameter-efficient Multi-step Fine-tuning of Multilingual and Multi-task Learning Model for Japanese Dialect Speech Recognition P. 52
Yuta Kamiya, Shogo Miwa, and Atsuhiko Kai

A Study on The Acquisition of Triphthong Vowels by Altaic Chinese Learners Under The 'belt And Road' Initiative P. 58
Yuan Jia, and Linjiao Pan

Acoustic Realization of /s/ Across Accents of Urdu P. 64
Iram Fatima, and Sahar Rauf

Age-related and Gender-related Differences in Cantonese Vowels P. 70
Wai-Sum Lee

An Investigation of Chinese Speech Under Alcohol Influence: Database Construction and Phonetic Analysis P.76
Peppina Po-Lun Lee, Mosi He, and Bin Li

Analysis of Pathological Features for Spoof Detection P. 81
Myat Aye Aye Aung, Hay Mar Soe Naing, Aye Mya Hlaing, Win Pa Pa, Kasorn Galajit, and Candy Olivia Mawalim

Benchmarking Cognitive Domains for LLMs: Insights from Taiwanese Hakka Culture P. 89
Chen-Chi Chang, Ching-Yuan Chen, Hung-Shin Lee, and Chi-Cheng Lee

Clapping Hands to Word Stress Improves Children's L2 English Pronunciation Accuracy in A Word Imitation Task: Evidence from A Classroom Study P. 95
Meiyun Chen

Comparative Study on the Phonetic Characteristics of Chinese Vowels between Kyrgyz and Kirgiz Learners P.101
Yuan Jia, and Mingshuai Yin

Computer-Assisted Pronunciation Training System for Atayal, An Indigenous Language in Taiwan P.107
Yu-Lan Chuang, Hsiu-Ray Hsu, Di Tam Luu, Yi-Wen Liu, and Ching-Ting Hsin

Continual Gated Adapter for Bilingual Codec Text-to-Speech P. 113
Li-Jen Yang, and Jen-Tzung Chien

Continual Learning in Machine Speech Chain Using Gradient Episodic Memory P. 119
Geoffrey Tyndall, Kurniawati Azizah, Dipta Tanaya, Ayu Purwarianti, Dessi Puji Lestari, and

Sakriani Sakti

Developing a Robust Mispronunciation Detection by Data Augmentation Based on Automatic Phone Annotation P. 125

Jong In Kim, Sunhee Kim, and Minhwa Chung

Developing a Thai Name Pronunciation Dictionary from Road Signs and Naming websites P. 130

Ausdang Thangthai

Development of an English Oral Assessment System with The GEPT Dataset P. 136

Hao Chien Lu, Chung-Chun Wang, Jhen-Ke Lin, and Berlin Chen

Enhancing Indonesian Automatic Speech Recognition: Evaluating Multilingual Models with Diverse Speech Variabilities P.142

Aulia Adila, Dessi Lestari, Ayu Purwarianti, Dipta Tanaya, Kurniawati Azizah, and Sakriani Sakti

Enhancing Phoneme Recognition in The Bengali Language Through Fine-Tuning of Multilingual Model P. 148

Akash Deep, Puja Bharati, Sabyasachi Chandra, Debolina Pramanik, Korra Siva Naik, and Shayamal Kumar Das Mandal

Exploration of Mongolian Word Stress Research Methods Based on Intonation Synthesis Technology P. 153

Aomin, Dahu Baiyila, and Aijun Li

Fusion of Multiple Audio Descriptors for The Recognition of Dysarthric Speech P. 160

Komal Bharti, and Pradip K. Das

Gated Adapters with Balanced Activation for Effective Contextual Speech Recognition P. 166

Yu-Chun Liu, Yi-Cheng Wang, Li-Ting Pai, Jia-Liang Lu, and Berlin Chen

Improving Speech Recognition by Enhancing Accent Discrimination P. 172

Hao-Tian Zheng, and Berlin Chen

Research on The Temporal Effect of Focus on Trisyllabic Sequences in Leizhou Min P. 178

Maolin Wang, Ying Liu, Han Yu, Ziyu Xiong, and Qiguang Lin

Right-prominent Trisyllabic Tone Sandhi in Taifeng Chinese P. 184

Xiaoyan Zhang, Aijun Li, and Zhiqiang Li

The Development of LOTUS-TRD: A Thai Regional Dialect Speech Corpus P. 189

Sumonmas Thatphithakkul, Kwanchiva Thangthai, and Vataya Chunwijitra

The Effectiveness of Audio-Visual Feedback for L2 Chinese Sentence Stress Perception and Production P. 195

Xingzi Gao, Yujie Gao, and Sichang Gao

Unified Spoken Language Proficiency Assessment System P. 201

Sunil Kumar Kopparapu, and Ashish Panda

Using Automatic Speech Recognition for Speech Comprehension Evaluation in The Cochlear Implant P. 207
Hsin-Li Chang, Enoch Hsin-Ho Huang, Yi-Ching Wang, and Yu Tsao

Oral Session #3: Speech Generation

Session Chair: Ying-Hui Lai, National Yang Ming Chiao Tung University, Taiwan

VoxHakka: A Dialectally Diverse Multi-speaker Text-to-Speech System for Taiwanese Hakka P. 212
Li-Wei Chen, Hung-Shin Lee, and Chen-Chi Chang
Learning Contrastive Emotional Nuances in Speech Synthesis P. 218
Bryan Gautama Ngo, Mahdin Rohmatillah, and Jen-Tzung Chien
Indonesian-English Code-Switching Speech Synthesizer Utilizing Multilingual STEN-TTS and BERT LID P. 224
Ahmad Alfani Handoyo, Chung Tran, Dessi Puji Lestari, and Sakriani Sakti
Exemplar-Based Methods for Mandarin Electrolaryngeal Speech Voice Conversion P. 230
Hsin-Te Hwang, Chia-Hua Wu, Ming-Chi Yen, Yu Tsao, and Hsin-Min Wang

Poster Session #2: Speech Technology, Machine Learning, and AI in Speech Processing

A Deep Learning Based Approach with Data Augmentation for Infant Cry Sound Verification P. 236
Namita Gokavi, Padala Sri Ramulu, Kandregula Nanda Kishore, Sunil Saumya, and Deepak K T
A Preliminary Study on End-To-End Multimodal Subtitle Recognition for Taiwanese TV Programs P. 242
Pei-Chung Su, Cheng-Hsiu Cho, Chih-Chung Kuo, Yen-Chun Lai, Yan-Ming Lin, Chao-Shih Huang, and Yuan-Fu Liao
A Preliminary Study on Taiwanese Pos Taggers: Leveraging Chinese in The Absence of Taiwanese Pos Annotation Datasets P. 248
Chao-Yang Chang, Yan-Ming Lin, Chih-Chung Kuo, Yen-Chun Lai, Chao-Shih Huang, Yuan-Fu Liao, and Tsun-Guan Thiann
Agent-Driven Large Language Models for Mandarin Lyric Generation P. 254
Hong-Hsiang Liu, and Yi-Wen Liu
An Evaluation of Neural Vocoder-based Voice Cloning System for Dysphonia Speech Disorder P. 260
Dhiya Dewangga, Dessi Lestari, Ayu Purwarianti, Dipta Tanaya, Kurniawati Azizah, and Sakriani

Sakti

An N-Best List Selection Framework for Asr N-Best Rescoring P. 267
Chen-Han Wu, and Kuan-Yu Chen

Analysis and Detection of Differences in Spoken User Behaviors between Autonomous and Wizard-of-Oz Systems P. 273
Mikey Elmers, Koji Inoue, Divesh Lala, Keiko Ochi, and Tatsuya Kawahara

Analysis and Discussion of Feature Extraction Technology for Musical Genre Classification P. 279
Shu-Hua Chen, Wei-Ting Huang, Cheng-Hao Lai, Yu-Lun Lin, and Ming-Hsiang Su

Annotation of Addressing Behavior in Multi-Party Conversation P. 283
Keisuke Kadota, Seima Oyama, and Yasuharu Den

Benchmarking Clickbait Detection from News Headlines P. 289
Ying-Lung Lin, Shao-Ying Lu, and Ling-Chih Yu

Chunk Size Scheduling for Optimizing the Quality-Latency Trade-off in Simultaneous Speech Translation P. 294
Iqbal Pahlevi Amin, Haotian Tan, Kurniawati Azizah, and Sakriani Sakti

Comprehensive Benchmarking and Analysis of Open Pre-trained Thai Speech Recognition Models P. 300
Pattara Tipakasorn, Oatsada Chatthong, Ren Yonehana, and Kwanchiva Thangthai

Depression Classification Using Log-Mel Spectrograms: A Comparative Analysis of Window Size-Based Data Augmentation and Deep Learning Models P. 307
Lokesh Kumar, Kumar Kaustubh, Shashaank Aswatha Mattur, and S. R. Mahadeva Prasanna

Effects of Multiple Japanese Datasets for training Voice Activity Projection Models P. 313
Yuki Sato, Yuya Chiba, and Ryuichiro Higashinaka

Exploring Branchformer-based End-to-end Speaker Diarization with Speaker-wise VAD Loss P. 319
Pei-Ying Lee, Hau-Yun Guo, Tien-Hong Lo, and Berlin Chen

IIIT-Speech Twins 1.0: An English-Hindi Parallel Speech Corpora for Speech-to-Speech Machine Translation and Automatic Dubbing P. 325
Anindita Mondal, Anil Kumar Vuppala, and Chiranjeevi Yarra

Improving Real-Time Music Accompaniment Separation with MMDenseNet P. 331
Chun-Hsiang Wang, Chung-Che Wang, Jun-You Wang, Jyh-Shing Roger Jang, and Yen-Hsun Chu

Infant Cry Verification with Multi-View Self-Attention Vision Transformers P. 337
Kartik Jagtap, Namita Gokavi, and Sunil Saumya

Modeling Response Relevance Using Dialog Act and Utterance-Design Features: A Corpus-Based Analysis P. 343

Mika Enomoto, Yuichi Ishimoto, and Yasuharu Den

Multi-resolution Singing Voice Separation P. 349

Yih-Liang Shen, Ya-Ching Lai, and Tai-Shih Chi

Multilingual Speech Translator for Medical Consultation P. 355

Zhe-Jia Xu, Yeou-Jiunn Chen, and Qian-Bei Hong

Overcoming the Impact of Different Materials on Optical Microphones for Speech Capture Using Deep Learning P. 360

Yi-Hao Jiang, Jia-Hui Li, Jia-Wei Chen, Yi-Chang Wu, and Ying-Hui Lai

Robust Audio-Visual Speech Enhancement: Correcting Misassignments in Complex Environments with Advanced Post-Processing P. 365

Wenze Ren, Kuo-Hsuan Hung, Rong Chao, YouJin Li, Hsin-Min Wang, and Yu Tsao

Singer Separation for Karaoke Content Generation P. 371

Hsuan-Yu Lin, Xuanjun Chen, and Jyh-Shing Roger Jang

Uncertainty-Based Ensemble Learning for Speech Classification P. 376

Bagus Tris Atmaja, Akira Sasou, and Felix Burkhardt

A Neural Machine Translation System for The Low-Resource Sixian Hakka Language P. 382

Yi-Hsian Hung, and Yi-Chin Huang

Oral Session #4: Best Paper Candidates

Session Chair: Yu Taso, Academia Sinica, Taiwan

Proposal of Protocols for Speech Materials Acquisition and Presentation Assisted by Tools Based on Structured Test Signals P. 388

Hideki Kawahara, Ken-Ichi Sakakibara, Mitsunori Mizumachi, and Kohei Yatabe

Exploring Impact of Prioritizing Intra-Singer Acoustic Variations on Singer Embedding Extractor Construction for Singer Verification P. 394

Sayaka Toma, Tomoki Ariga, Yosuke Higuchi, Ichiju Hayasaka, Rie Shigyo, and Tetsuji Ogawa

A Feedback-Driven Self-Improvement Strategy and Emotion-Aware Vocoder for Emotional Voice Conversion P. 400

Zhanhang Zhang, and Sakriani Sakti

ConvCounsel: A Conversational Dataset for Student Counseling P. 406

Po-Chaun Chen, Mahdin Rohmatillah, You-Teng Lin, and Jen-Tzung Chien

Construction of Large Language Models for Taigi and Hakka Using Transfer Learning P. 412

Yen-Chun Lai, Yi-Jun Zheng, Wen-Han Hsu, Yan-Ming Lin, Cheng-Hsiu Cho, Chih-Chung Kuo, Chao-Shih Huang, and Yuan-Fu Liao

Country Region Report

China	P. 418
Hong Kong	P. 424
India	P. 425
Indonesia	P. 431
Japan	P. 437
Korea	P. 443
Malaysia	P. 449
Myanmar	P. 456
Philippine	P. 462
Taiwan	P. 468
Thailand	P. 474
Timor-Leste	P. 479
Vietnam	P. 485