2021 24th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA 2021)

Singapore
18 – 20 November 2021

IEEE

IEEE Catalog Number: CFP2108H-POD
ISBN: 978-1-6654-0871-4

Copyright © 2021 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:	CFP2108H-POD
ISBN (Print-On-Demand):	978-1-6654-0871-4
ISBN (Online):	978-1-6654-0870-7
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 Content

About the Conference VII

Message from the O-COCOSDA Convenor VIII

Message from the Conference Chairs IX

Message from the Program Chairs X

Conference Committees XI

Keynote Speeches XIII

Papers

1 Speech Recognition

P12 Multi-Encoder Sequential Attention Network for Context-Aware Speech Recognition in Japanese Dialog Conversation
Nobuya Tachimori, Sakriani Sakti and Satoshi Nakamura 1

P29 Investigation Of a Single-Channel Frequency-Domain Speech Enhancement Network to Improve End-T o-End Bengali Automatic Speech Recognition Under Unseen Noisy Conditions
Md Mahbub E Noor, Yen-Ju Lu, Syu-Siang Wang, Supratip Ghose, Chia-Yu Chang, Ryandhimas E. Zezario, Shafique Ahmed, Wei-Ho Chung, Yu Tsao and Hsin-Min Wang..... 7

P32 A Study on Native American English Speech Recognition by Indian Listeners with Varying Word Familiarity Level
Abhayjeet Singh, Achuth Rao Mv, Rakesh Vaideeswaran, Chiranjeevi Yarra and Prasanta Kumar Ghosh 13

P34 Speech Recognition System for Writing Dentist Medical Records
Dinda Yora Islami and Dessi *Puji* Lestari 19

P42 A Multi-Genre Urdu Broadcast Speech Recognition System
Erbaz Khan, Sahar Rauf, Farah *Adeeba* and Sarmad Hussain 25

P39 An Empirical Study of Speaker Identification System for Mono And Traverse Linguistic Background Using Em And Smem
Amita Dev, Shweta Bansal and *Shyam* Sunder Agrawal 31

P43 Self-Supervised Spoken Question Understanding and Speaking with Automatic Vocabulary Learning
Keisuke Toyoda, Yusuke *Kimura*, Mingxin Zhang, Kent Hino, Kosuke Mori and Takahiro Shinozak 37

2 Emotion and Prosody

P17 Design and Basic Analysis of the TUT Emotional Storytelling Corpus
Hikaru Oishi, Mika Enomoto, Keiko Ochi and Yasunari Obuchi 43

P20 Construction and Analysis of Tibetan Amdo Dialect Speech Dataset for Speech Synthesis
Xinyi Zhang, Wenhuan Lu, Xinyue Zhao, Yi Zhu and Jianguo Wei *49*

P23 INTO-CASS: A Corpus for The Study of Intonation and Prosody in Chinese Dialects and Ethnic Languages
Aijun Li and Ziyu Xiong *53*

P37 Discourse Timing in Children's Rhyme Speech Produced by Prelingually Deaf Mandarin-Speaking Children with Cochlear Implants
Jue Yu and Qianwen Jin *59*

P21 Towards the Development of Segment Level Speech Overlap Detection Using Convolutional Neural Network
Ronald John Cabatic and Angelica De La Cruz *65*

P33 On the Use of Gestures in Dialogue Breakdown Detection
Taiga Mori, Kristiina Jokinen and Yasuharu Den *70*

P18 Aspect-Based Sentiment Analysis of User Created Game Reviews
Ian Michael Urriza and Maria Art Antonette Clariño *76*

3 Language Learning

P01 L2 Accent and intelligibility by Chinese L2 Speakers of English
Yizhou Lan and Tongtong Xie *82*

P02 A Study on English Word-final Coronal Stop Deletion by Chinese EFL Learners
Tong Li and Hui Feng *88*

P08 The Role of High Variability Phonetic Training on Chinese EFL Learners' Perception of English Vowels in Noisy Environment
Qianxi Yu and Ping Tang *94*

P10 The Effect of Overnight Consolidation on English Vowel Perception by Chinese Learners After High Speaker Variability Phonetic Training
Yanan Shen and Ping Tang *100*

P24 Mandarin Speakers' Acquisitions and Representations of Flapping in American English in An ESL Context: A Perception and Production Study
Chuang Chia Wei *106*

P36 Tonal Patterns of Tri-Syllabic Words in The Production of Standard Chinese of Bilingual Teachers
Yuan Jia and Bin Li *111*

4 Multimodal Databases

P03 SPIRE VCV: An Acoustic-Articulatory Corpus with Three Different Speaking Rates
Tilak Purohit, Tejas Umesh, Shankar Narayanan, S Minulakshmi and Prasanta Ghosh *116*

P09 Khmer Speech Translation Corpus of The Extraordinary Chambers in The Courts of Cambodia (ECCC)

Kak Soky, Masato Mimura, Tatsuya Kawahara, Sheng Li, Chenchen Ding, Chenhui Chu and Sethserey Sam *122*

P11 SLoClas: A Database for Joint Sound Localization and Classification
Qian Xinyuan, Bidisha Sharma, Amine El Abridi and Haizhou Li *128*

P16 GAMVA: A Japanese Audio-visual Multi-Angle Speech Corpus
Shinnosuke Isobe, Ryuichi Hirose, Takumi Nishiwaki, Tomohiro Hattori, Satoshi Tamura, Yuuto Gotoh and Masaki Nose *134*

P19 M2ASR-Mongo: A Free Mongolian Speech Database and Accompanied Baselines
Tiankai Zhi, Ying Shi, Wenqiang Du, Guanyu Li and Dong Wang *140*

P44 WSPIRE: A Parallel Multi-Device Corpus in Neutral and Whisper Speech
Abinay Reddy Naini, Bhavuk Singhal and Prasanta Kumar Ghosh *146*

5 Dialects and Accents

P04 Which Phonemes Will Distinguish the Different Regions Within the Same Dialect?
Xuefei Liu, Jianhua Tao, Yurong Han, Chenglong Wang, Xueying Zheng and Zhengqi Wen .. *152*

P05 Comparison of Static and Time-Sequential Features in Automatic Fluency Detection of Spontaneous Speech
Huaijin Deng, Takehito Utsuro, Akio Kobayashi and Hiromitsu Nishizaki *158*

P06 How Do Speakers Pause and Hesitate in English and Japanese? - A Comparison Using Parallel Corpora of English and Japanese Presentation Speeches
Michiko Watanabe, Yuma Shirahata, Ralph Rose and Kikuo Maekawa *164*

P26 Korean Dialect Identification Based on Intonation Modeling
Jooyoung Lee, Kyungwha Kim and Minhwa Chung *168*

P35 Development of Accent Recognition Systems for Vietnamese Speech
Quang Tien Duong and Van Hai Do *174*

P07 A Blind Method for Phone Segmentation and Its Evaluation on Vietnamese Speech Corpus
Dac-Thang Hoang and Tat-Thang Vu *180*

6 Speech Synthesis and Translation

P13 Simultaneous Speech-to-speech Translation System with Transformer-based Incremental ASR, MT, and TTS
Ryo Fukuda, Sashi Novitasari, Yui Oka, Yasumasa Kano, Yuki Yano, Yuka Ko, Hirotaka Tokuyama, Kosuke Doi, Tomoya Yanagita, Sakriani Sakti, Katsuhito Sudoh and Satoshi Nakamura *186*

P14 Using Speech Enhancement to Realize Speech Synthesis of Low-Resource Dungan Languages
Rui Jiang, Hongwu Yang, Sicheng Chen and Xin Shan *193*

P15 A study on neural-network-based Text-to-Speech adaptation techniques for Vietnamese
Phuong Pham Ngoc, Chung Tran Quang, Truong Do Quoc and Mai Luong Chi *199*

P25 Using Local Phrase Dependency Structure Information in Neural Sequence-To-Sequence Speech Synthesis

Nobuyoshi Kaiki, Sakriani Sakti and Satoshi Nakamura *206*

P28 Text-to-Speech Systems for Filipino using Unit Selection and Deep Learning
Edsel Jedd Renovalles and Crisron Rudolf Lucas *212*

P31 Investigation of an Input Sequence on Thai Neural Sequence-to-Sequence Speech Synthesis
Pongsathon Janyoi and Ausdang Thangthai *218*

Country/Region Reports

China 224
Aijun Li, Dong Wang

Hong Kong 227
Tan Lee

India 230
S. S. Agrawal, K. Samudravijaya

Indonesia 233
Hammam Riza

Japan 236
Satoshi Nakamura

Korea 239
Yong-Ju Lee

The Philippines 242
Nathaniel Oco

Singapore 245
Haizhou Li

Taiwan 248
Sin-Horng Chen, Hsin-Min Wang

Thailand 251
N/A

Vietnam 254
Luong Chi Mai