

23rd Annual Conference of the International Speech Communication Association (INTERSPEECH 2022)

Human and Humanizing Speech
Technology

Incheon, South Korea
18-22 September 2022

Volume 1 of 7

ISBN: 978-1-7138-8879-6

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© (2022) by International Speech Communication Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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

SPEECH SYNTHESIS: TOWARD END-TO-END SYNTHESIS

SANE-TTS: Stable and Natural End-To-End Multilingual Text-To-Speech.....	1
<i>Hyunjae Cho, Wonbin Jung, Junhyeok Lee, Sang Hoon Woo</i>	
Enhancement of Pitch Controllability Using Timbre-Preserving Pitch Augmentation in FastPitch.....	6
<i>Hanbin Bae, Young-Sun Joo</i>	
Speaking Rate Control of End-To-End TTS Models by Direct Manipulation of the Encoder's Output Embeddings.....	11
<i>Martin Lenglet, Olivier Perrotin, Gérard Bailly</i>	
TriniTTS: Pitch-Controllable End-To-End TTS Without External Aligner.....	16
<i>Yooncheol Ju, Ilhwan Kim, Hongsun Yang, Ji-Hoon Kim, Byeongyeol Kim, Soumi Maiti, Shinji Watanabe</i>	
JETS: Jointly Training FastSpeech2 and HiFi-GAN for End to End Text to Speech.....	21
<i>Dan Lim, Sunghee Jung, Eesung Kim</i>	

TECHNOLOGY FOR DISORDERED SPEECH

Interpretable Dysarthric Speaker Adaptation Based on Optimal-Transport.....	26
<i>Rosanna Turrisi, Leonardo Badino</i>	
Dysarthric Speech Recognition from Raw Waveform with Parametric CNNs.....	31
<i>Zhengjun Yue, Erfan Loweimi, Heidi Christensen, Jon Barker, Zoran Cvetkovic</i>	
The Effectiveness of Time Stretching for Enhancing Dysarthric Speech for Improved Dysarthric Speech Recognition.....	36
<i>Luke Prananta, Bence Halpern, Siyuan Feng, Odette Scharenborg</i>	
Investigating Self-Supervised Pretraining Frameworks for Pathological Speech Recognition.....	41
<i>Lester Phillip Violeta, Wen Chin Huang, Tomoki Toda</i>	
Improved ASR Performance for Dysarthric Speech Using Two-Stage DataAugmentation.....	46
<i>Chitralkha Bhat, Ashish Panda, Helmer Strik</i>	
Cross-Lingual Self-Supervised Speech Representations for Improved Dysarthric Speech Recognition.....	51
<i>Abner Hernandez, Paula Andrea Pérez-Toro, Elmar Noeth, Juan Rafael Orozco-Arroyave, Andreas Maier, Seung Hee Yang</i>	

NEURAL NETWORK TRAINING METHODS FOR ASR I

Regularizing Transformer-Based Acoustic Models by Penalizing Attention Weights.....	56
<i>Munhak Lee, Joon-Hyuk Chang, Sang-Eon Lee, Ju-Seok Seong, Chanhee Park, Haeyoung Kwon</i>	

Content-Context Factorized Representations for Automated Speech Recognition	61
<i>David Chan, Shalini Ghosh</i>	
Comparison and Analysis of New Curriculum Criteria for End-To-End ASR	66
<i>Georgios Karakasidis, Tamás Grósz, Mikko Kurimo</i>	
Incremental Learning for RNN-Transducer Based Speech Recognition Models	71
<i>Deepak Baby, Pasquale D'Alterio, Valentin Mendelev</i>	
Production Federated Keyword Spotting Via Distillation, Filtering, and Joint Federated-Centralized Training	76
<i>Andrew Hard, Kurt Partridge, Neng Chen, Sean Augenstein, Aishanee Shah, Hyun Jin Park, Alex Park, Sara Ng, Jessica Nguyen, Ignacio Lopez-Moreno, Rajiv Mathews, Francoise Beaufays</i>	

ACOUSTIC PHONETICS AND PROSODY

Use of Prosodic and Lexical Cues for Disambiguating Wh-Words in Korean	81
<i>Jieun Song, Hae-Sung Jeon, Jieun Kiaer</i>	
Autoencoder-Based Tongue Shape Estimation During Continuous Speech.....	86
<i>Vinicius Ribeiro, Yves Laprie</i>	
Phonetic Erosion and Information Structure in Function Words: The Case of Mia	91
<i>Giuseppe Magistro, Claudia Crocco</i>	
Dynamic Vertical Larynx Actions Under Prosodic Focus	96
<i>Miran Oh, Yoonjeong Lee</i>	
Fundamental Frequency Variability Over Time in Telephone Interactions	101
<i>Leah Bradshaw, Eleanor Chodroff, Lena Jäger, Volker Dellwo</i>	

SPOKEN MACHINE TRANSLATION

SHAS: Approaching Optimal Segmentation for End-To-End Speech Translation.....	106
<i>Ioannis Tsiamas, Gerard I. Gállego, José A. R. Fonollosa, Marta R. Costa-Jussà</i>	
M-Adapter: Modality Adaptation for End-To-End Speech-To-Text Translation.....	111
<i>Jinming Zhao, Hao Yang, Gholamreza Haffari, Ehsan Shareghi</i>	
Cross-Modal Decision Regularization for Simultaneous Speech Translation	116
<i>Mohd Abbas Zaidi, Beomseok Lee, Sangha Kim, Chanwoo Kim</i>	
Speech Segmentation Optimization Using Segmented Bilingual Speech Corpus for End-To-End Speech Translation	121
<i>Ryo Fukuda, Katsuhito Sudoh, Satoshi Nakamura</i>	
Generalized Keyword Spotting Using ASR Embeddings.....	126
<i>Kirandevraj R, Vinod Kumar Kurmi, Vinay Namboodiri, C V Jawahar</i>	

(MULTIMODAL) SPEECH EMOTION RECOGNITION I

Multi-Corpus Speech Emotion Recognition for Unseen Corpus Using Corpus-Wise Weights in Classification Loss	131
<i>Youngdo Ahn, Sung Joo Lee, Jong Won Shin</i>	
Improving Speech Emotion Recognition Through Focus and Calibration Attention Mechanisms	136
<i>Junghun Kim, Yoojin An, Jihie Kim</i>	
The Emotion is Not One-Hot Encoding: Learning with Grayscale Label for Emotion Recognition in Conversation.....	141
<i>Joosung Lee</i>	
Probing Speech Emotion Recognition Transformers for Linguistic Knowledge.....	146
<i>Andreas Triantafyllopoulos, Johannes Wagner, Hagen Wierstorf, Maximilian Schmitt, Uwe Reichel, Florian Eyben, Felix Burkhardt, Björn W. Schuller</i>	
End-To-End Label Uncertainty Modeling for Speech-Based Arousal Recognition Using Bayesian Neural Networks.....	151
<i>Navin Raj Prabhu, Guillaume Carbajal, Nale Lehmann-Willenbrock, Timo Gerkmann</i>	
Mind the Gap: On the Value of Silence Representations to Lexical-Based Speech Emotion Recognition	156
<i>Matthew Perez, Mimansa Jaiswal, Minxue Niu, Cristina Gorrostieta, Matthew Roddy, Kye Taylor, Reza Lotfian, John Kane, Emily Mower Provost</i>	
Exploiting Co-Occurrence Frequency of Emotions in Perceptual Evaluations to Train a Speech Emotion Classifier	161
<i>Huang-Cheng Chou, Chi-Chun Lee, Carlos Busso</i>	
Positional Encoding for Capturing Modality Specific Cadence for Emotion Detection	166
<i>Hira Dharmyal, Bhiksha Raj, Rita Singh</i>	

DEREVERBERATION, NOISE REDUCTION, AND SPEAKER EXTRACTION

Speak Like a Professional: Increasing Speech Intelligibility by Mimicking Professional Announcer Voice with Voice Conversion.....	171
<i>Tuan Vu Ho, Maori Kobayashi, Masato Akagi</i>	
Vector-Quantized Variational Autoencoder for Phase-Aware Speech Enhancement.....	176
<i>Tuan Vu Ho, Quoc Huy Nguyen, Masato Akagi, Masashi Unoki</i>	
IDeepMMSE: An Improved Deep Learning Approach to MMSE Speech and Noise Power Spectrum Estimation for Speech Enhancement.....	181
<i>Minseung Kim, Hyungchan Song, Sein Cheong, Jong Won Shin</i>	
Boosting Self-Supervised Embeddings for Speech Enhancement.....	186
<i>Kuo-Hsuan Hung, Szu-Wei Fu, Huan-Hsin Tseng, Hsin-Tien Chiang, Yu Tsao, Chii-Wann Lin</i>	
Monoaural Speech Enhancement Using a Nested U-Net with Two-Level Skip Connections.....	191
<i>Seorim Hwang, Youngcheol Park, Sungwook Park</i>	
CycleGAN-Based Unpaired Speech Dereverberation	196
<i>Hannah Muckenhirn, Aleksandr Safin, Hakan Erdogan, Felix De Chaumont Quitry, Marco Tagliasacchi, Scott Wisdom, John R. Hershey</i>	

Attentive Training: A New Training Framework for Talker-Independent Speaker Extraction	201
<i>Ashutosh Pandey, Deliang Wang</i>	
Improved Modulation-Domain Loss for Neural-Network-Based Speech Enhancement.....	206
<i>Tyler Vuong, Richard Stern</i>	
Perceptual Characteristics Based Multi-Objective Model for Speech Enhancement	211
<i>Chiang-Jen Peng, Yun-Ju Chan, Yih-Liang Shen, Cheng Yu, Yu Tsao, Tai-Shih Chi</i>	
Listen Only to Me! How Well Can Target Speech Extraction Handle False Alarms?	216
<i>Marc Delcroix, Keisuke Kinoshita, Tsubasa Ochiai, Katerina Zmolikova, Hiroshi Sato, Tomohiro Nakatani</i>	
Monaural Speech Enhancement Based on Spectrogram Decomposition for Convolutional Neural Network-Sensitive Feature Extraction.....	221
<i>Hao Shi, Longbiao Wang, Sheng Li, Jianwu Dang, Tatsuya Kawahara</i>	
Neural Network-Augmented Kalman Filtering for Robust Online Speech Dereverberation in Noisy Reverberant Environments	226
<i>Jean-Marie Lemercier, Joachim Thiemann, Raphael Koning, Timo Gerkmann</i>	
<u>SOURCE SEPARATION II</u>	
PodcastMix: A Dataset for Separating Music and Speech in Podcasts.....	231
<i>Nicolás Schmidt, Jordi Pons, Marius Miron</i>	
Independence-Based Joint Dereverberation and Separation with Neural Source Model	236
<i>Kohei Saijo, Robin Scheibler</i>	
Spatial Loss for Unsupervised Multi-Channel Source Separation.....	241
<i>Kohei Saijo, Robin Scheibler</i>	
Effect of Head Orientation on Speech Directivity.....	246
<i>Samuel Bellows, Timothy W. Leishman</i>	
Unsupervised Training of Sequential Neural Beamformer Using Coarsely-Separated and Non-Separated Signals	251
<i>Kohei Saijo, Tetsuji Ogawa</i>	
Blind Language Separation: Disentangling Multilingual Cocktail Party Voices by Language	256
<i>Marvin Borsdorf, Kevin Scheck, Haizhou Li, Tanja Schultz</i>	
NTF of Spectral and Spatial Features for Tracking and Separation of Moving Sound Sources in Spherical Harmonic Domain	261
<i>Mateusz Guzik, Konrad Kowalczyk</i>	
Modelling Turn-Taking in Multispeaker Parties for Realistic Data Simulation	266
<i>Jack Deadman, Jon Barker</i>	
An Initialization Scheme for Meeting Separation with Spatial Mixture Models.....	271
<i>Christoph Boeddeker, Tobias Cord-Landwehr, Thilo Von Neumann, Reinhold Haeb-Umbach</i>	
Prototypical Speaker-Interference Loss for Target Voice Separation Using Non-Parallel Audio Samples	276
<i>Seongkyu Mun, Dhananjaya Gowda, Jihwan Lee, Changwoo Han, Dokyun Lee, Chanwoo Kim</i>	

EMBEDDING AND NETWORK ARCHITECTURE FOR SPEAKER RECOGNITION

Reliability Criterion Based on Learning-Phase Entropy for Speaker Recognition with Neural Network.....	281
<i>Pierre-Michel Bousquet, Mickael Rouvier, Jean-Francois Bonastre</i>	
Attentive Feature Fusion for Robust Speaker Verification.....	286
<i>Bei Liu, Zhengyang Chen, Yanmin Qian</i>	
Dual Path Embedding Learning for Speaker Verification with Triplet Attention.....	291
<i>Bei Liu, Zhengyang Chen, Yanmin Qian</i>	
DF-ResNet: Boosting Speaker Verification Performance with Depth-First Design.....	296
<i>Bei Liu, Zhengyang Chen, Shuai Wang, Haoyu Wang, Bing Han, Yanmin Qian</i>	
Adaptive Rectangle Loss for Speaker Verification.....	301
<i>Li Ruida, Fang Shuo, Ma Chenguang, Li Liang</i>	
MFA-Conformer: Multi-Scale Feature Aggregation Conformer for Automatic Speaker Verification.....	306
<i>Yang Zhang, Zhiqiang Lv, Haibin Wu, Shanshan Zhang, Pengfei Hu, Zhiyong Wu, Hung-Yi Lee, Helen Meng</i>	
Enroll-Aware Attentive Statistics Pooling for Target Speaker Verification.....	311
<i>Leying Zhang, Zhengyang Chen, Yanmin Qian</i>	
Transport-Oriented Feature Aggregation for Speaker Embedding Learning.....	316
<i>Yusheng Tian, Jingyu Li, Tan Lee</i>	
Multi-Frequency Information Enhanced Channel Attention Module for Speaker Representation Learning.....	321
<i>Mufan Sang, John H. L. Hansen</i>	
CS-CTCSConvID: Small Footprint Speaker Verification with Channel Split Time-Channel-Time Separable 1-Dimensional Convolution.....	326
<i>Linjun Cai, Yuhong Yang, Xufeng Chen, Weiping Tu, Hongyang Chen</i>	
Reliable Visualization for Deep Speaker Recognition.....	331
<i>Pengqi Li, Lantian Li, Askar Hamdulla, Dong Wang</i>	
Unifying Cosine and PLDA Back-Ends for Speaker Verification.....	336
<i>Zhiyuan Peng, Xuanji He, Ke Ding, Tan Lee, Guanglu Wan</i>	
CTFALite: Lightweight Channel-Specific Temporal and Frequency Attention Mechanism for Enhancing the Speaker Embedding Extractor.....	341
<i>Yuheng Wei, Junzhao Du, Hui Liu, Qian Wang</i>	

SPEECH REPRESENTATION II

SpeechFormer: A Hierarchical Efficient Framework Incorporating the Characteristics of Speech.....	346
<i>Weidong Chen, Xiaofen Xing, Xiangmin Xu, Jianxin Pang, Lan Du</i>	
VoiceLab: Software for Fully Reproducible Automated Voice Analysis.....	351
<i>David Feinberg</i>	

TRILLsson: Distilled Universal Paralinguistic Speech Representations.....	356
<i>Joel Shor, Subhashini Venugopalan</i>	
Global Signal-To-Noise Ratio Estimation Based on Multi-Subband Processing Using Convolutional Neural Network	361
<i>Nan Li, Meng Ge, Longbiao Wang, Masashi Unoki, Sheng Li, Jianwu Dang</i>	
A Sparsity-Promoting Dictionary Model for Variational Autoencoders	366
<i>Mostafa Sadeghi, Paul Magron</i>	
Deep Transductive Transfer Regression Network for Cross-Corpus Speech Emotion Recognition	371
<i>Yan Zhao, Jincen Wang, Ru Ye, Yuan Zong, Wenming Zheng, Li Zhao</i>	
Audio Anti-Spoofing Using Simple Attention Module and Joint Optimization Based on Additive Angular Margin Loss and Meta-Learning	376
<i>John H. L. Hansen, Zhenyu Wang</i>	
PEAF: Learnable Power Efficient Analog Acoustic Features for Audio Recognition.....	381
<i>Boris Bergsma, Minhao Yang, Milos Cernak</i>	
Hybrid Handcrafted and Learnable Audio Representation for Analysis of Speech Under Cognitive and Physical Load	386
<i>Gasser Elbanna, Alice Biryukov, Neil Scheidwasser-Clow, Lara Orlandic, Pablo Mainar, Mikolaj Kegler, Pierre Beckmann, Milos Cernak</i>	
Generative Data Augmentation Guided by Triplet Loss for Speech Emotion Recognition.....	391
<i>Shijun Wang, Hamed Hemati, Jón Guðnason, Damian Borth</i>	
Learning Neural Audio Features Without Supervision.....	396
<i>Sarthak Yadav, Neil Zeghidour</i>	
Densely-Connected Convolutional Recurrent Network for Fundamental Frequency Estimation in Noisy Speech.....	401
<i>Yixuan Zhang, Heming Wang, Deliang Wang</i>	
Predicting Label Distribution Improves Non-Intrusive Speech Quality Estimation.....	406
<i>Abu Zaher Md Faridee, Hannes Gamper</i>	
Deep Versus Wide: An Analysis of Student Architectures for Task-Agnostic Knowledge Distillation of Self-Supervised Speech Models.....	411
<i>Takanori Ashihara, Takafumi Moriya, Kohei Matsuura, Tomohiro Tanaka</i>	
Dataset Pruning for Resource-Constrained Spoofed Audio Detection	416
<i>Abdul Hameed Azeemi, Ihsan Ayyub Qazi, Agha Ali Raza</i>	

SPEECH SYNTHESIS: LINGUISTIC PROCESSING, PARADIGMS AND OTHER TOPICS II

EdiTTS: Score-Based Editing for Controllable Text-To-Speech.....	421
<i>Jaesung Tae, Hyeongju Kim, Taesu Kim</i>	
Improving Mandarin Prosodic Structure Prediction with Multi-Level Contextual Information	426
<i>Jie Chen, Changhe Song, Deyi Tuo, Xixin Wu, Shiyin Kang, Zhiyong Wu, Helen Meng</i>	
SpeechPainter: Text-Conditioned Speech Inpainting	431
<i>Zalan Borsos, Matthew Sharifi, Marco Tagliasacchi</i>	

A Polyphone BERT for Polyphone Disambiguation in Mandarin Chinese	436
<i>Song Zhang, Ken Zheng, Xiaoxu Zhu, Baoxiang Li</i>	
Neural Lexicon Reader: Reduce Pronunciation Errors in End-To-End TTS by Leveraging External Textual Knowledge.....	441
<i>Mutian He, Jingzhou Yang, Lei He, Frank Soong</i>	
ByT5 Model for Massively Multilingual Grapheme-To-Phoneme Conversion	446
<i>Jian Zhu, Cong Zhang, David Jurgens</i>	
DocLayoutTTS: Dataset and Baselines for Layout-Informed Document-Level Neural Speech Synthesis.....	451
<i>Puneet Mathur, Franck Dérnoncourt, Quan Hung Tran, Jiuxiang Gu, Ani Nenkova, Vlad Morariu, Rajiv Jain, Dinesh Manocha</i>	
Mixed-Phoneme BERT: Improving BERT with Mixed Phoneme and Sup-Phoneme Representations for Text to Speech.....	456
<i>Guangyan Zhang, Kaitao Song, Xu Tan, Daxin Tan, Yuzi Yan, Yanqing Liu, Gang Wang, Wei Zhou, Tao Qin, Tan Lee, Sheng Zhao</i>	
Unsupervised Text-To-Speech Synthesis by Unsupervised Automatic Speech Recognition.....	461
<i>Junrui Ni, Liming Wang, Heting Gao, Kaizhi Qian, Yang Zhang, Shiyu Chang, Mark Hasegawa-Johnson</i>	
An Efficient and High Fidelity Vietnamese Streaming End-To-End Speech Synthesis	466
<i>Tho Nguyen Duc Tran, The Chuong Chu, Vu Hoang, Trung Huu Bui, Hung Quoc Truong</i>	
Predicting Pairwise Preferences Between TTS Audio Stimuli Using Parallel Ratings Data and Anti- Symmetric Twin Neural Networks	471
<i>Cassia Valentini-Botinhao, Manuel Sam Ribeiro, Oliver Watts, Korin Richmond, Gustav Eje Henter</i>	
An Automatic Soundtracking System for Text-To-Speech Audiobooks.....	476
<i>Zikai Chen, Lin Wu, Junjie Pan, Xiang Yin</i>	
Environment Aware Text-To-Speech Synthesis.....	481
<i>Daxin Tan, Guangyan Zhang, Tan Lee</i>	
SoundChoice: Grapheme-To-Phoneme Models with Semantic Disambiguation	486
<i>Artem Ploujnikov, Mirco Ravanelli</i>	
Shallow Fusion of Weighted Finite-State Transducer and Language Model for Text Normalization	491
<i>Evelina Bakhturina, Yang Zhang, Boris Ginsburg</i>	
Prosodic Alignment for Off-Screen Automatic Dubbing.....	496
<i>Yogesh Virkar, Marcello Federico, Robert Enyedi, Roberto Barra-Chicote</i>	
A Study of Modeling Rising Intonation in Cantonese Neural Speech Synthesis	501
<i>Qibing Bai, Tom Ko, Yu Zhang</i>	
CAUSE: Crossmodal Action Unit Sequence Estimation from Speech.....	506
<i>Hirokazu Kameoka, Takuhiro Kaneko, Shogo Seki, Kou Tanaka</i>	
Visualising Model Training Via Vowel Space for Text-To-Speech Systems	511
<i>Binu Nisal Abeyasinghe, Jesin James, Catherine Watson, Felix Marattukalam</i>	

OTHER TOPICS IN SPEECH RECOGNITION

Binary Early-Exit Network for Adaptive Inference on Low-Resource Devices.....	516
<i>Aaqib Saeed</i>	
Streaming Speaker-Attributed ASR with Token-Level Speaker Embeddings	521
<i>Naoyuki Kanda, Jian Wu, Yu Wu, Xiong Xiao, Zhong Meng, Xiaofei Wang, Yashesh Gaur, Zhuo Chen, Jinyu Li, Takuya Yoshioka</i>	
Speaker Consistency Loss and Step-Wise Optimization for Semi-Supervised Joint Training of TTS and ASR Using Unpaired Text Data	526
<i>Naoki Makishima, Satoshi Suzuki, Atsushi Ando, Ryo Masumura</i>	
Audio-Visual Generalized Few-Shot Learning with Prototype-Based Co-Adaptation	531
<i>Yi-Kai Zhang, Da-Wei Zhou, Han-Jia Ye, De-Chuan Zhan</i>	
Federated Domain Adaptation for ASR with Full Self-Supervision.....	536
<i>Junteng Jia, Jay Mahadeokar, Weiyi Zheng, Yuan Shangguan, Ozlem Kalinli, Frank Seide</i>	
Augmented Adversarial Self-Supervised Learning for Early-Stage Alzheimer's Speech Detection	541
<i>Longfei Yang, Wenqing Wei, Sheng Li, Jiyi Li, Takahiro Shinozaki</i>	
Extending RNN-T-Based Speech Recognition Systems with Emotion and Language Classification.....	546
<i>Zvi Kons, Hagai Aronowitz, Edmilson Morais, Matheus Damasceno, Hong-Kwang Kuo, Samuel Thomas, George Saon</i>	
Thutmose Tagger: Single-Pass Neural Model for Inverse Text Normalization	550
<i>Alexandra Antonova, Evelina Bakhturina, Boris Ginsburg</i>	
Leveraging Prosody for Punctuation Prediction of Spontaneous Speech.....	555
<i>Yeonjin Cho, Sara Ng, Trang Tran, Mari Ostendorf</i>	
A Comparative Study on Speaker-Attributed Automatic Speech Recognition in Multi-Party Meetings	560
<i>Fan Yu, Zhihao Du, Shiliang Zhang, Yuxiao Lin, Lei Xie</i>	

AUDIO DEEP PLC (PACKET LOSS CONCEALMENT) CHALLENGE

TMGAN-PLC: Audio Packet Loss Concealment Using Temporal Memory Generative Adversarial Network.....	565
<i>Yuansheng Guan, Guochen Yu, Andong Li, Chengshi Zheng, Jie Wang</i>	
Real-Time Packet Loss Concealment with Mixed Generative and Predictive Model	570
<i>Jean-Marc Valin, Ahmed Mustafa, Christopher Montgomery, Timothy B. Terriberry, Michael Klingbeil, Paris Smaragdis, Arvinth Krishnaswamy</i>	
PLCNet: Real-Time Packet Loss Concealment with Semi-Supervised Generative Adversarial Network.....	575
<i>Baiyun Liu, Qi Song, Mingxue Yang, Wuwen Yuan, Tianbao Wang</i>	
INTERSPEECH 2022 Audio Deep Packet Loss Concealment Challenge	580
<i>Lorenz Diener, Sten Sootla, Solomiya Branets, Ando Saabas, Robert Aichner, Ross Cutler</i>	
End-To-End Multi-Loss Training for Low Delay Packet Loss Concealment.....	585
<i>Nan Li, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu</i>	

ROBUST SPEAKER RECOGNITION

Extended U-Net for Speaker Verification in Noisy Environments	590
<i>Ju-Ho Kim, Jungwoo Heo, Hye-Jin Shim, Ha-Jin Yu</i>	
Domain Agnostic Few-Shot Learning for Speaker Verification	595
<i>Seunghan Yang, Debasmith Das, Janghoon Cho, Hyounghoo Park, Sungrack Yun</i>	
Scoring of Large-Margin Embeddings for Speaker Verification: Cosine Or PLDA?.....	600
<i>Qionqiong Wang, Kong Aik Lee, Tianchi Liu</i>	
Training Speaker Embedding Extractors Using Multi-Speaker Audio with Unknown Speaker Boundaries.....	605
<i>Themis Stafylakis, Ladislav Mosner, Oldrich Plchot, Johan Rohdin, Anna Silnova, Lukas Burget, Jan Cernocký</i>	
Investigating the Contribution of Speaker Attributes to Speaker Separability Using Disentangled Speaker Representations.....	610
<i>Chau Luu, Steve Renals, Peter Bell</i>	
Joint Domain Adaptation and Speech Bandwidth Extension Using Time-Domain GANs for Speaker Verification	615
<i>Saurabh Kataria, Jesús Villalba, Laureano Moro-Velázquez, Najim Dehak</i>	

SPEECH PRODUCTION

Variability in Production of Non-Sibilant Fricative [ç] in /Hi/	620
<i>Tsukasa Yoshinaga, Kikuo Maekawa, Akiyoshi Iida</i>	
Streaming Model for Acoustic to Articulatory Inversion with Transformer Networks	625
<i>Sathvik Udupa, Aravind Illa, Prasanta Ghosh</i>	
Trajectories Predicted by Optimal Speech Motor Control Using LSTM Networks	630
<i>Tsiky Rakotomalala, Pierre Baraduc, Pascal Perrier</i>	
Exploration Strategies for Articulatory Synthesis of Complex Syllable Onsets.....	635
<i>Daniel Van Niekerc, Anqi Xu, Branislav Gerazov, Paul Konstantin Krug, Peter Birkholz, Yi Xu</i>	
Linguistic Versus Biological Factors Governing Acoustic Voice Variation.....	640
<i>Yoonjeong Lee, Jody Kreiman</i>	
Acquisition of Allophonic Variation in Second Language Speech: An Acoustic and Articulatory Study of English Laterals by Japanese Speakers	644
<i>Takayuki Nagamine</i>	

SPEECH QUALITY ASSESSMENT

SAQAM: Spatial Audio Quality Assessment Metric.....	649
<i>Pranay Manocha, Anurag Kumar, Buye Xu, Anjali Menon, Israel Degene Gebru, Vamsi Krishna Ithapu, Paul Calamia</i>	
Speech Quality Assessment Through MOS Using Non-Matching References	654
<i>Pranay Manocha, Anurag Kumar</i>	

An Objective Test Tool for Pitch Extractors' Response Attributes	659
<i>Hideki Kawahara, Kohei Yatabe, Ken-Ichi Sakakibara, Tatsuya Kitamura, Hideki Banno, Masanori Morise</i>	
Data Augmentation Using McAdams-Coefficient-Based Speaker Anonymization for Fake Audio Detection	664
<i>Kai Li, Sheng Li, Xugang Lu, Masato Akagi, Meng Liu, Lin Zhang, Chang Zeng, Longbiao Wang, Jianwu Dang, Masashi Unoki</i>	
Automatic Data Augmentation Selection and Parametrization in Contrastive Self-Supervised Speech Representation Learning	669
<i>Salah Zaiem, Titouan Parcollet, Slim Essid</i>	
Transformer-Based Quality Assessment Model for Generalized User-Generated Multimedia Audio Content	674
<i>Deebha Mumtaz, Ajit Jena, Vinit Jakhetiya, Karan Nathwani, Sharath Chandra Guntuku</i>	

LANGUAGE MODELING AND LEXICAL MODELING FOR ASR

Space-Efficient Representation of Entity-Centric Query Language Models	679
<i>Christophe Van Gysel, Mirko Hannemann, Ernest Pusateri, Youssef Oualil, Ilya Oparin</i>	
Domain Prompts: Towards Memory and Compute Efficient Domain Adaptation of ASR Systems	684
<i>Saket Dingliwal, Ashish Shenoy, Sravan Bodapati, Ankur Gandhe, Ravi Teja Gadde, Katrin Kirchhoff</i>	
Sentence-Select: Large-Scale Language Model Data Selection for Rare-Word Speech Recognition	689
<i>W. Ronny Huang, Cal Peysers, Tara Sainath, Ruoming Pang, Trevor D. Strohman, Shankar Kumar</i>	

VOLUME 2

UserLibri: A Dataset for ASR Personalization Using Only Text	694
<i>Theresa Breiner, Swaroop Ramaswamy, Ehsan Variansi, Shefali Garg, Rajiv Mathews, Khe Chai Sim, Kilol Gupta, Mingqing Chen, Lara McConnaughey</i>	
A BERT-Based Language Modeling Framework	699
<i>Chin-Yueh Chien, Kuan-Yu Chen</i>	

CHALLENGES AND OPPORTUNITIES FOR SIGNAL PROCESSING AND MACHINE LEARNING FOR MULTIPLE SMART DEVICES

Joint Optimization of Sampling Rate Offsets Based on Entire Signal Relationship Among Distributed Microphones	704
<i>Yoshiki Masuyama, Kouei Yamaoka, Nobutaka Ono</i>	
Challenges and Opportunities in Multi-Device Speech Processing	709
<i>Gregory Ciccarelli, Jarred Barber, Arun Nair, Israel Cohen, Tao Zhang</i>	
Practical Over-The-Air Perceptual Acoustic Watermarking	714
<i>Ameya Agaskar</i>	

Clustering-Based Wake Word Detection in Privacy-Aware Acoustic Sensor Networks.....	719
<i>Timm Koppelman, Luca Becker, Alexandru Nelus, Rene Glitza, Lea Schönherr, Rainer Martin</i>	
Relative Acoustic Features for Distance Estimation in Smart-Homes	724
<i>Francesco Nespola, Daniel Barreda, Patrick Naylor</i>	
Time-Domain Ad-Hoc Array Speech Enhancement Using a Triple-Path Network.....	729
<i>Ashutosh Pandey, Buye Xu, Anurag Kumar, Jacob Donley, Paul Calamia, Deliang Wang</i>	

SPEECH PROCESSING & MEASUREMENT

Relationship Between the Acoustic Time Intervals and Tongue Movements of German Diphthongs	734
<i>Arne-Lukas Fietkau, Simon Stone, Peter Birkholz</i>	
Development of Allophonic Realization Until Adolescence: A Production Study of the Affricate-Fricative Variation of /Z/ Among Japanese Children.....	739
<i>Sanae Matsui, Kyoji Iwamoto, Reiko Mazuka</i>	
Recurrent Multi-Head Attention Fusion Network for Combining Audio and Text for Speech Emotion Recognition.....	744
<i>Chung-Soo Ahn, Chamara Kasun, Sunil Sivadas, Jagath Rajapakse</i>	
Low-Level Physiological Implications of End-To-End Learning for Speech Recognition	749
<i>Louise Coppieters De Gibson, Philip N. Garner</i>	
Idiosyncratic Lingual Articulation of American English /æ/ and /ɑ/ Using Network Analysis	754
<i>Carolina Lins Machado, Volker Dellwo, Lei He</i>	
Method for Improving the Word Intelligibility of Presented Speech Using Bone-Conduction Headphones	759
<i>Teruki Toya, Wenyu Zhu, Maori Kobayashi, Kenichi Nakamura, Masashi Unoki</i>	
Three-Dimensional Finite-Difference Time-Domain Acoustic Analysis of Simplified Vocal Tract Shapes.....	764
<i>Debasish Mohapatra, Mario Fleischer, Victor Zappi, Peter Birkholz, Sidney Fels</i>	
Speech Imitation Skills Predict Automatic Phonetic Convergence: A GMM-UBM Study on L2.....	769
<i>Dorina De Jong, Aldo Pastore, Noël Nguyen, Alessandro D'Ausilio</i>	
Self-Supervised Speech Unit Discovery from Articulatory and Acoustic Features Using VQ-VAE.....	774
<i>Marc-Antoine Georges, Jean-Luc Schwartz, Thomas Hueber</i>	
Deep Speech Synthesis from Articulatory Representations.....	779
<i>Peter Wu, Shinji Watanabe, Louis Goldstein, Alan W Black, Gopala Krishna Anumanchipalli</i>	
Orofacial Somatosensory Inputs in Speech Perceptual Training Modulate Speech Production.....	784
<i>Monica Ashokumar, Jean-Luc Schwartz, Takayuki Ito</i>	

SPEECH SYNTHESIS: ACOUSTIC MODELING AND NEURAL WAVEFORM GENERATION

I

Transfer Learning Framework for Low-Resource Text-To-Speech Using a Large-Scale Unlabeled Speech Corpus.....	788
<i>Minchan Kim, Myeonghun Jeong, Byoung Jin Choi, Sunghwan Ahn, Joun Yeop Lee, Nam Soo Kim</i>	
DRSpeech: Degradation-Robust Text-To-Speech Synthesis with Frame-Level and Utterance-Level Acoustic Representation Learning.....	793
<i>Takaaki Saeki, Kentaro Tachibana, Ryuichi Yamamoto</i>	
MSR-NV: Neural Vocoder Using Multiple Sampling Rates.....	798
<i>Kentaro Mitsui, Kei Sawada</i>	
SpecGrad: Diffusion Probabilistic Model Based Neural Vocoder with Adaptive Noise Spectral Shaping.....	803
<i>Yuma Koizumi, Heiga Zen, Kohei Yatabe, Nanxin Chen, Michiel Bacchiani</i>	
Bunched LPCNet2: Efficient Neural Vocoders Covering Devices from Cloud to Edge	808
<i>Sangjun Park, Kihyun Choo, Joohyung Lee, Anton V. Porov, Konstantin Osipov, June Sig Sung</i>	
Hierarchical and Multi-Scale Variational Autoencoder for Diverse and Natural Non-Autoregressive Text-To-Speech.....	813
<i>Jaesung Bae, Jinhyeok Yang, Taejun Bak, Young-Sun Joo</i>	
End-To-End LPCNet: A Neural Vocoder with Fully-Differentiable LPC Estimation	818
<i>Krishna Subramani, Jean-Marc Valin, Umut Isik, Paris Smaragdis, Arvinth Krishnaswamy</i>	
EPIC TTS Models: Empirical Pruning Investigations Characterizing Text-To-Speech Models	823
<i>Perry Lam, Huayun Zhang, Nancy Chen, Berrak Sisman</i>	
Fine-Grained Noise Control for Multispeaker Speech Synthesis	828
<i>Karolos Nikitaras, Georgios Vamvoukakis, Nikolaos Ellinas, Konstantinos Klapsas, Konstantinos Markopoulos, Spyros Raptis, June Sig Sung, Gunu Jho, Aimilios Chalamandaris, Pirros Tsiakoulis</i>	
WavThruVec: Latent Speech Representation as Intermediate Features for Neural Speech Synthesis.....	833
<i>Hubert Siuzdak, Piotr Dura, Pol Van Rijn, Nori Jacoby</i>	
Fast Grad-TTS: Towards Efficient Diffusion-Based Speech Generation on CPU.....	838
<i>Ivan Vovk, Tasnima Sadekova, Vladimir Gogoryan, Vadim Popov, Mikhail Kudinov, Jiansheng Wei</i>	
Simple and Effective Unsupervised Speech Synthesis.....	843
<i>Alexander H. Liu, Cheng-I Lai, Wei-Ning Hsu, Michael Auli, Alexei Baevski, James Glass</i>	
Unified Source-Filter GAN with Harmonic-Plus-Noise Source Excitation Generation.....	848
<i>Reo Yoneyama, Yi-Chiao Wu, Tomoki Toda</i>	

SHOW AND TELL I

NeMo Open Source Speaker Diarization System.....	853
<i>Tae Jin Park, Nithin Rao Koluguri, Fei Jia, Jagadeesh Balam, Boris Ginsburg</i>	

Voice2Alliance: Automatic Speaker Diarization and Quality Assurance of Conversational Alignment.....	855
<i>Baihan Lin</i>	
VAgyojaka: An Annotating and Post-Editing Tool for Automatic Speech Recognition.....	857
<i>Rishabh Kumar, Devaraja Adiga, Mayank Kothari, Jatin Dalal, Ganesh Ramakrishnan, Preethi Jyothi</i>	
SKYE: More than a Conversational AI	859
<i>Alzahra Badi, ChungHo Park, Minseok Keum, Miguel Alba, Youngsuk Ryu, Jeongmin Bae</i>	

SPATIAL AUDIO

Training Data Generation with DOA-Based Selecting and Remixing for Unsupervised Training of Deep Separation Models.....	861
<i>Hokuto Munakata, Ryu Takeda, Kazunori Komatani</i>	
Beam-Guided TasNet: An Iterative Speech Separation Framework with Multi-Channel Output	866
<i>Hangting Chen, Yi Yang, Feng Dang, Pengyuan Zhang</i>	
Joint Estimation of Direction-Of-Arrival and Distance for Arrays with Directional Sensors Based on Sparse Bayesian Learning	871
<i>Feifei Xiong, Pengyu Wang, Zhongfu Ye, Jinwei Feng</i>	
How to Listen? Rethinking Visual Sound Localization.....	876
<i>Ho-Hsiang Wu, Magdalena Fuentes, Prem Seetharaman, Juan Pablo Bello</i>	
Small Footprint Neural Networks for Acoustic Direction of Arrival Estimation	881
<i>Zhiheng Ouyang, Miao Wang, Wei-Ping Zhu</i>	
Multi-Modal Multi-Correlation Learning for Audio-Visual Speech Separation.....	886
<i>Xiaoyu Wang, Xiangyu Kong, Xiulian Peng, Yan Lu</i>	
MIMO-DoAnet: Multi-Channel Input and Multiple Outputs DoA Network with Unknown Number of Sound Sources.....	891
<i>Haoran Yin, Meng Ge, Yanjie Fu, Gaoyan Zhang, Longbiao Wang, Lei Zhang, Lin Qiu, Jianwu Dang</i>	
Iterative Sound Source Localization for Unknown Number of Sources	896
<i>Yanjie Fu, Meng Ge, Haoran Yin, Xinyuan Qian, Longbiao Wang, Gaoyan Zhang, Jianwu Dang</i>	
Distance-Based Sound Separation.....	901
<i>Katharine Patterson, Kevin Wilson, Scott Wisdom, John R. Hershey</i>	
VCSE: Time-Domain Visual-Contextual Speaker Extraction Network	906
<i>Junjie Li, Meng Ge, Zexu Pan, Longbiao Wang, Jianwu Dang</i>	
TRUNet: Transformer-Recurrent-U Network for Multi-Channel Reverberant Sound Source Separation.....	911
<i>Ali Aroudi, Stefan Uhlich, Marc Ferras Font</i>	

SINGLE-CHANNEL SPEECH ENHANCEMENT II

PercepNet+: A Phase and SNR Aware PercepNet for Real-Time Speech Enhancement.....	916
<i>Xiaofeng Ge, Jiangyu Han, Yanhua Long, Haixin Guan</i>	
Lightweight Full-Band and Sub-Band Fusion Network for Real Time Speech Enhancement	921
<i>Zhuangqi Chen, Pingjian Zhang</i>	
Cross-Layer Similarity Knowledge Distillation for Speech Enhancement.....	926
<i>Jiaming Cheng, Ruiyu Liang, Yue Xie, Li Zhao, Björn Schuller, Jie Jia, Yiyuan Peng</i>	
Spectro-Temporal SubNet for Real-Time Monaural Speech Denoising and Dereverberation	931
<i>Feifei Xiong, Weiguang Chen, Pengyu Wang, Xiaofei Li, Jinwei Feng</i>	
CMGAN: Conformer-Based Metric GAN for Speech Enhancement.....	936
<i>Ruizhe Cao, Sherif Abdulatif, Bin Yang</i>	
Model Compression by Iterative Pruning with Knowledge Distillation and Its Application to Speech Enhancement.....	941
<i>Zeyuan Wei, Li Hao, Xueliang Zhang</i>	
Single-Channel Speech Enhancement Using Graph Fourier Transform.....	946
<i>Chenhui Zhang, Xiang Pan</i>	
Joint Optimization of the Module and Sign of the Spectral Real Part Based on CRN for Speech Denoising	951
<i>Zilu Guo, Xu Xu, Zhongfu Ye</i>	
Attentive Recurrent Network for Low-Latency Active Noise Control.....	956
<i>Hao Zhang, Ashutosh Pandey, Deliang Wang</i>	
Memory-Efficient Multi-Step Speech Enhancement with Neural ODE.....	961
<i>Jen-Hung Huang, Chung-Hsien Wu</i>	
GLD-Net: Improving Monaural Speech Enhancement by Learning Global and Local Dependency Features with GLD Block.....	966
<i>Xinmeng Xu, Yang Wang, Jie Jia, Binbin Chen, Jianjun Hao</i>	
Improving Visual Speech Enhancement Network by Learning Audio-Visual Affinity with Multi- Head Attention.....	971
<i>Xinmeng Xu, Yang Wang, Jie Jia, Binbin Chen, Dejun Li</i>	
Speech Enhancement with Fullband-Subband Cross-Attention Network	976
<i>Jun Chen, Wei Rao, Zilin Wang, Zhiyong Wu, Yannan Wang, Tao Yu, Shidong Shang, Helen Meng</i>	
OSSEM: One-Shot Speaker Adaptive Speech Enhancement Using Meta Learning	981
<i>Cheng Yu, Szu-Wei Fu, Tsun-An Hsieh, Yu Tsao, Mirco Ravanelli</i>	
Efficient Speech Enhancement with Neural Homomorphic Synthesis.....	986
<i>Wenbin Jiang, Tao Liu, Kai Yu</i>	
Fast Real-Time Personalized Speech Enhancement: End-To-End Enhancement Network (E3Net) and Knowledge Distillation	991
<i>Manthan Thakker, Sefik Emre Eskimez, Takuya Yoshioka, Huaming Wang</i>	

Strategies to Improve Robustness of Target Speech Extraction to Enrollment Variations	996
<i>Hiroshi Sato, Tsubasa Ochiai, Marc Delcroix, Keisuke Kinoshita, Takafumi Moriya, Naoki Makishima, Mana Ihori, Tomohiro Tanaka, Ryo Masumura</i>	

NOVEL MODELS AND TRAINING METHODS FOR ASR II

FedNST: Federated Noisy Student Training for Automatic Speech Recognition.....	1001
<i>Haaris Mehmood, Agnieszka Dobrowolska, Karthikeyan Saravanan, Mete Ozay</i>	
SCaLa: Supervised Contrastive Learning for End-To-End Speech Recognition.....	1006
<i>Li Fu, Xiaoxiao Li, Runyu Wang, Lu Fan, Zhengchen Zhang, Meng Chen, Youzheng Wu, Xiaodong He</i>	
NAS-SCAE: Searching Compact Attention-Based Encoders for End-To-End Automatic Speech Recognition	1011
<i>Yukun Liu, Ta Li, Pengyuan Zhang, Yonghong Yan</i>	
Leveraging Acoustic Contextual Representation by Audio-Textual Cross-Modal Learning for Conversational ASR	1016
<i>Kun Wei, Yike Zhang, Sining Sun, Lei Xie, Long Ma</i>	
PM-MMUT: Boosted Phone-Mask Data Augmentation Using Multi-Modeling Unit Training for Phonetic-Reduction-Robust E2E Speech Recognition	1021
<i>Guodong Ma, Pengfei Hu, Nurmamet Yolwas, Shen Huang, Hao Huang</i>	
Analysis of Self-Attention Head Diversity for Conformer-Based Automatic Speech Recognition	1026
<i>Kartik Audhkhasi, Yinghui Huang, Bhuvana Ramabhadran, Pedro J. Moreno</i>	
Improving Rare Word Recognition with LM-Aware MWER Training	1031
<i>Wang Weiran, Tongzhou Chen, Tara Sainath, Ehsan Variiani, Rohit Prabhavalkar, W. Ronny Huang, Bhuvana Ramabhadran, Neeraj Gaur, Sepand Mavandadi, Cal Peyser, Trevor Strohman, Yanzhang He, David Rybach</i>	
Improving the Training Recipe for a Robust Conformer-Based Hybrid Model	1036
<i>Mohammad Zeineldeen, Jingjing Xu, Christoph Lüscher, Ralf Schlüter, Hermann Ney</i>	
CTC Variations Through New WFST Topologies	1041
<i>Aleksandr Laptev, Somshubra Majumdar, Boris Ginsburg</i>	
Dealing with Unknowns in Continual Learning for End-To-End Automatic Speech Recognition	1046
<i>Martin Sustek, Samik Sadhu, Hynek Hermansky</i>	
Towards Efficiently Learning Monotonic Alignments for Attention-Based End-To-End Speech Recognition	1051
<i>Chenfeng Miao, Kun Zou, Ziyang Zhuang, Tao Wei, Jun Ma, Shaojun Wang, Jing Xiao</i>	
On Monoaural Speech Enhancement for Automatic Recognition of Real Noisy Speech Using Mixture Invariant Training	1056
<i>Jisi Zhang, Catalin Zorila, Rama Doddipatla, Jon Barker</i>	
From Undercomplete to Sparse Overcomplete Autoencoders to Improve LF-MMI Based Speech Recognition	1061
<i>Selen Hande Kabil, Herve Bourlard</i>	

Domain Adversarial Self-Supervised Speech Representation Learning for Improving Unknown Domain Downstream Tasks.....	1066
<i>Tomohiro Tanaka, Ryo Masumura, Hiroshi Sato, Mana Ihori, Kohei Matsuura, Takanori Ashihara, Takafumi Moriya</i>	

Attention Weight Smoothing Using Prior Distributions for Transformer-Based End-To-End ASR.....	1071
<i>Takashi Maekaku, Yuya Fujita, Yifan Peng, Shinji Watanabe</i>	

SPOKEN DIALOGUE SYSTEMS AND MULTIMODALITY

Reducing Offensive Replies in Open Domain Dialogue Systems.....	1076
<i>Naokazu Uchida, Takeshi Homma, Makoto Iwayama, Yasuhiro Sogawa</i>	

Induce Spoken Dialog Intents Via Deep Unsupervised Context Contrastive Clustering.....	1081
<i>Ting-Wei Wu, Biing Juang</i>	

Dialogue Acts Aided Important Utterance Detection Based on Multiparty and Multimodal Information.....	1086
<i>Fumio Nihei, Ryo Ishii, Yukiko Nakano, Kyosuke Nishida, Ryo Masumura, Atsushi Fukayama, Takao Nakamura</i>	

Contextual Acoustic Barge-In Classification for Spoken Dialog Systems.....	1091
<i>Dhanush Bekal, Sundararajan Srinivasan, Srikanth Ronanki, Sravan Bodapati, Katrin Kirchhoff</i>	

Calibrate and Refine! a Novel and Agile Framework for ASR Error Robust Intent Detection.....	1096
<i>Peilin Zhou, Dading Chong, Helin Wang, Qingcheng Zeng</i>	

ASR-Robust Natural Language Understanding on ASR-GLUE Dataset.....	1101
<i>Lingyun Feng, Jianwei Yu, Yan Wang, Songxiang Liu, Deng Cai, Haitao Zheng</i>	

From Disfluency Detection to Intent Detection and Slot Filling.....	1106
<i>Mai Hoang Dao, Thinh Truong, Dat Quoc Nguyen</i>	

Audio-Visual Wake Word Spotting in MISP2021 Challenge: Dataset Release and Deep Analysis.....	1111
<i>Hengshun Zhou, Jun Du, Gongzhen Zou, Zhaoxu Nian, Chin-Hui Lee, Sabato Marco Siniscalchi, Shinji Watanabe, Odette Scharenborg, Jingdong Chen, Shifu Xiong, Jian-Qing Gao</i>	

Extending Compositional Attention Networks for Social Reasoning in Videos.....	1116
<i>Christina Sartzetaki, Georgios Paraskevopoulos, Alexandros Potamianos</i>	

TopicKS: Topic-Driven Knowledge Selection for Knowledge-Grounded Dialogue Generation.....	1121
<i>Shiquan Wang, Yuke Si, Xiao Wei, Longbiao Wang, Zhiqiang Zhuang, Xiaowang Zhang, Jianwu Dang</i>	

Bottom-Up Discovery of Structure and Variation in Response Tokens ('backchannels') Across Diverse Languages.....	1126
<i>Andreas Liesenfeld, Mark Dingemanse</i>	

Cross-Modal Transfer Learning Via Multi-Grained Alignment for End-To-End Spoken Language Understanding.....	1131
<i>Yi Zhu, Zexun Wang, Hang Liu, Peiyang Wang, Mingchao Feng, Meng Chen, Xiaodong He</i>	

Use of Nods Less Synchronized with Turn-Taking and Prosody During Conversations in Adults with Autism	1136
<i>Keiko Ochi, Nobutaka Ono, Keiho Owada, Kuroda Miho, Shigeki Sagayama, Hidenori Yamasue</i>	

SHOW AND TELL I(VR)

DAVIS: Driver’s Audio-Visual Speech Recognition	1141
<i>Denis Ivanko, Dmitry Ryumin, Alexey Kashevnik, Alexandr Axyonov, Andrey Kitenko, Igor Lashkov, Alexey Karpov</i>	

SPEECH EMOTION RECOGNITION I

Analysis of Self-Supervised Learning and Dimensionality Reduction Methods in Clustering-Based Active Learning for Speech Emotion Recognition.....	1143
<i>Einari Vaaras, Manu Airaksinen, Okko Räsänen</i>	
Emotion-Shift Aware CRF for Decoding Emotion Sequence in Conversation	1148
<i>Chun-Yu Chen, Yun-Shao Lin, Chi-Chun Lee</i>	
Vaccinating SER to Neutralize Adversarial Attacks with Self-Supervised Augmentation Strategy	1153
<i>Bo-Hao Su, Chi-Chun Lee</i>	
Speech Emotion Recognition in the Wild Using Multi-Task and Adversarial Learning	1158
<i>Jack Parry, Eric Demattos, Anita Klementiev, Axel Ind, Daniela Morse-Kopp, Georgia Clarke, Dimitri Palaz</i>	
The Magnitude and Phase Based Speech Representation Learning Using Autoencoder for Classifying Speech Emotions Using Deep Canonical Correlation Analysis	1163
<i>Ashishkumar Gudmalwar, Biplove Basel, Anirban Dutta, Ch V Rama Rao</i>	
Improving Speech Emotion Recognition Using Self-Supervised Learning with Domain-Specific Audiovisual Tasks	1168
<i>Lucas Goncalves, Carlos Busso</i>	

SINGLE-CHANNEL SPEECH ENHANCEMENT I

SNRi Target Training for Joint Speech Enhancement and Recognition	1173
<i>Yuma Koizumi, Shigeki Karita, Arun Narayanan, Sankaran Panchapagesan, Michiel Bacchiani</i>	
Deep Self-Supervised Learning of Speech Denoising from Noisy Speeches.....	1178
<i>Yutaro Sanada, Takumi Nakagawa, Yuichiro Wada, Kosaku Takanashi, Yuhui Zhang, Kiichi Tokuyama, Takafumi Kanamori, Tomonori Yamada</i>	
NASTAR: Noise Adaptive Speech Enhancement with Target-Conditional Resampling.....	1183
<i>Chi-Chang Lee, Cheng-Hung Hu, Yu-Chen Lin, Chu-Song Chen, Hsin-Min Wang, Yu Tsao</i>	
FFC-SE: Fast Fourier Convolution for Speech Enhancement.....	1188
<i>Ivan Shechekotov, Pavel K. Andreev, Oleg Ivanov, Aibek Alanov, Dmitry Vetrov</i>	
A Systematic Comparison of Phonetic Aware Techniques for Speech Enhancement.....	1193
<i>Or Tal, Moshe Mandel, Felix Kreuk, Yossi Adi</i>	

Multi-View Attention Transfer for Efficient Speech Enhancement.....	1198
<i>Wooseok Shin, Hyun Joon Park, Jin Sob Kim, Byung Hoon Lee, Sung Won Han</i>	

SPEECH SYNTHESIS: NEW APPLICATIONS

SATTS: Speaker Attractor Text to Speech, Learning to Speak by Learning to Separate.....	1203
<i>Nabarun Goswami, Tatsuya Harada</i>	
Correcting Mispronunciations in Speech Using Spectrogram Inpainting.....	1208
<i>Talia Ben Simon, Felix Kreuk, Faten Awwad, Jacob T. Cohen, Joseph Keshet</i>	
Speech Audio Corrector: Using Speech from Non-Target Speakers for One-Off Correction of Mispronunciations in Grapheme-Input Text-To-Speech.....	1213
<i>Jason Fong, Daniel Lyth, Gustav Eje Henter, Hao Tang, Simon King</i>	
End-To-End Binaural Speech Synthesis.....	1218
<i>Wen Chin Huang, Dejan Markovic, Alexander Richard, Israel Dejene Gebru, Anjali Menon</i>	
PoeticTTS - Controllable Poetry Reading for Literary Studies.....	1223
<i>Julia Koch, Florian Lux, Nadja Schauffler, Toni Bernhart, Felix Dieterle, Jonas Kuhn, Sandra Richter, Gabriel Viehhauser, Ngoc Thang Vu</i>	
Articulatory Synthesis for Data Augmentation in Phoneme Recognition.....	1228
<i>Paul Konstantin Krug, Peter Birkholz, Branislav Gerazov, Daniel Rudolph Van Niekerk, Anqi Xu, Yi Xu</i>	

SPOKEN LANGUAGE UNDERSTANDING I

SF-DST: Few-Shot Self-Feeding Reading Comprehension Dialogue State Tracking with Auxiliary Task.....	1233
<i>Jihyun Lee, Gary Geunbae Lee</i>	
Benchmarking Transformers-Based Models on French Spoken Language Understanding Tasks.....	1238
<i>Oralie Cattan, Sahar Ghannay, Christophe Servan, Sophie Rosset</i>	
McBERT: Momentum Contrastive Learning with BERT for Zero-Shot Slot Filling.....	1243
<i>Seong-Hwan Heo, Wonkee Lee, Jong-Hyeok Lee</i>	
Bottleneck Low-Rank Transformers for Low-Resource Spoken Language Understanding.....	1248
<i>Pu Wang, Hugo Van Hamme</i>	
On Joint Training with Interfaces for Spoken Language Understanding.....	1253
<i>Anirudh Raju, Milind Rao, Gautam Tiwari, Pranav Dheram, Bryan Anderson, Zhe Zhang, Chul Lee, Bach Bui, Ariya Rastrow</i>	
Device-Directed Speech Detection: Regularization Via Distillation for Weakly-Supervised Models.....	1258
<i>Vineet Garg, Ognjen Rudovic, Pranay Dighe, Ahmed Hussen Abdelaziz, Erik Marchi, Saurabh Adya, Chandra Dhir, Ahmed Tewfik</i>	

INCLUSIVE AND FAIR SPEECH TECHNOLOGIES I

Building African Voices.....	1263
<i>Perez Ogayo, Graham Neubig, Alan W Black</i>	

Toward Fairness in Speech Recognition: Discovery and Mitigation of Performance Disparities 1268
Pranav Dheram, Murugesan Ramakrishnan, Anirudh Raju, I-Fan Chen, Brian King, Katherine Powell, Melissa Saboowala, Karan Shetty, Andreas Stolcke

Training and Typological Bias in ASR Performance for World Englishes 1273
May Pik Yu Chan, June Choe, Aini Li, Yiran Chen, Xin Gao, Nicole Holliday

INCLUSIVE AND FAIR SPEECH TECHNOLOGIES II

A Study of Gender Impact in Self-Supervised Models for Speech-To-Text Systems..... 1278
Marceley Zanon Boito, Laurent Besacier, Natalia Tomashenko, Yannick Estève

Automatic Dialect Density Estimation for African American English 1283
Alexander Johnson, Kevin Everson, Vijay Ravi, Anissa Gladney, Mari Ostendorf, Abeer Alwan

Improving Language Identification of Accented Speech..... 1288
Kunnar Kukk, Tanel Alumäe

Design Guidelines for Inclusive Speaker Verification Evaluation Datasets 1293
Wiebke Toussaint, Lauriane Gorce, Aaron Yi Ding

Reducing Geographic Disparities in Automatic Speech Recognition Via Elastic Weight Consolidation..... 1298
Viet Anh Trinh, Pegah Ghahremani, Brian King, Jasha Droppo, Andreas Stolcke, Roland Maas

PHONETICS I

Gradual Improvements Observed in Learners' Perception and Production of L2 Sounds Through Continuing Shadowing Practices on a Daily Basis..... 1303
Takuya Kunihara, Chuanbo Zhu, Nobuaki Minematsu, Noriko Nakanishi

Spoofed Speech from the Perspective of a Forensic Phonetician 1308
Christin Kirchhübel, Georgina Brown

Investigating Prosodic Variation in British English Varieties Using ProPer..... 1313
Hae-Sung Jeon, Stephen Nichols

Perceived Prominence and Downstep in Japanese 1318
Hyun Kyung Hwang, Manami Hirayama, Takaomi Kato

The Discrimination of [zi]-[dzi] by Japanese Listeners and the Prospective Phonologization of /zi/ 1322
Andrea Alicehajic, Silke Hamann

Glottal Inverse Filtering Based on Articulatory Synthesis and Deep Learning 1327
Ingo Langheinrich, Simon Stone, Xinyu Zhang, Peter Birkholz

Investigating Phonetic Convergence of Laughter in Conversation 1332
Bogdan Ludusan, Marin Schröer, Petra Wagner

Telling Self-Defining Memories: An Acoustic Study of Natural Emotional Speech Productions..... 1337
Veronique Delvaux, Audrey Lavallée, Fanny Degouis, Xavier Saloppe, Jean-Louis Nandrino, Thierry Pham

Voicing Neutralization in Romanian Fricatives Across Different Speech Styles	1342
<i>Laura Spinu, Ioana Vasilescu, Lori Lamel, Jason Lilley</i>	
Nasal Coda Loss in the Chengdu Dialect of Mandarin: Evidence from RT-MRI.....	1347
<i>Sishi Liao, Phil Hoole, Conceição Cunha, Esther Kunay, Aletheia Cui, Lia Saki Bucar Shigemori, Felicitas Kleber, Dirk Voit, Jens Frahm, Jonathan Harrington</i>	
Ema2wav: Doing Articulation by Praat.....	1352
<i>Philipp Buech, Simon Roessig, Lena Pagel, Doris Muecke, Anne Hermes</i>	

MULTI-, CROSS-LINGUAL AND OTHER TOPICS IN ASR I

Improving Phonetic Transcriptions of Children’s Speech by Pronunciation Modelling with Constrained CTC-Decoding	1357
<i>Lars Rumberg, Christopher Gebauer, Hanna Ehlert, Ulrike Lüdtko, Jörn Ostermann</i>	
Leveraging Simultaneous Translation for Enhancing Transcription of Low-Resource Language Via Cross Attention Mechanism	1362
<i>Kak Soky, Sheng Li, Masato Mimura, Chenhui Chu, Tatsuya Kawahara</i>	
KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus	1367
<i>Saida Mussakhajayeva, Yerbolat Khassanov, Huseyin Atakan Varol</i>	
Knowledge of Accent Differences Can Be Used to Predict Speech Recognition.....	1372
<i>Tuende Szalay, Mostafa Shahin, Beena Ahmed, Kirrie Ballard</i>	
Lombard Effect for Bilingual Speakers in Cantonese and English: Importance of Spectro-Temporal Features	1377
<i>Maximilian Karl Scharf, Sabine Hochmuth, Lena L. N. Wong, Birger Kollmeier, Anna Warzybok</i>	
End-To-End Speech Recognition Modeling from De-Identified Data	1382
<i>Martin Flechl, Shou-Chun Yin, Junho Park, Peter Skala</i>	
Multi-Task End-To-End Model for Telugu Dialect and Speech Recognition.....	1387
<i>Aditya Yadavalli, Ganesh Mirishkar, Anil Kumar Vuppala</i>	
DEFORMER: Coupling Deformed Localized Patterns with Global Context for Robust End-To-End Speech Recognition	1392
<i>Jiamin Xie, John H. L. Hansen</i>	

ZERO, LOW-RESOURCE AND MULTI-MODAL SPEECH RECOGNITION I

Keyword Spotting with Synthetic Data Using Heterogeneous Knowledge Distillation.....	1397
<i>Yuna Lee, Seung Jun Baek</i>	
Probing Phoneme, Language and Speaker Information in Unsupervised Speech Representations.....	1402
<i>Maureen De Seyssel, Marvin Lavechin, Yossi Adi, Emmanuel Dupoux, Guillaume Wisniewski</i>	
Automatic Detection of Reactive Attachment Disorder Through Turn-Taking Analysis in Clinical Child-Caregiver Sessions	1407
<i>Andrei Birladeanu, Helen Minnis, Alessandro Vinciarelli</i>	
Automatic Pronunciation Assessment Using Self-Supervised Speech Representation Learning	1411
<i>Eesung Kim, Jae-Jin Jeon, Hyeji Seo, Hoon Kim</i>	

Exploring Few-Shot Fine-Tuning Strategies for Models of Visually Grounded Speech.....	1416
<i>Tyler Miller, David Harwath</i>	
Pseudo Label is Better than Human Label	1421
<i>Dongseong Hwang, Khe Chai Sim, Zhouyuan Huo, Trevor Strohman</i>	
A Temporal Extension of Latent Dirichlet Allocation for Unsupervised Acoustic Unit Discovery.....	1426
<i>Werner Van Der Merwe, Herman Kamper, Johan Adam Du Preez</i>	

SPEAKER EMBEDDING AND DIARIZATION

PRISM: Pre-Trained Indeterminate Speaker Representation Model for Speaker Diarization and Speaker Verification	1431
<i>Siqi Zheng, Hongbin Suo, Qian Chen</i>	
Cross-Age Speaker Verification: Learning Age-Invariant Speaker Embeddings	1436
<i>Xiaoyi Qin, Na Li, Weng Chao, Dan Su, Ming Li</i>	
Online Target Speaker Voice Activity Detection for Speaker Diarization	1441
<i>Weiqing Wang, Ming Li, Qingjian Lin</i>	
Probabilistic Spherical Discriminant Analysis: An Alternative to PLDA for Length-Normalized Embeddings	1446
<i>Niko Brummer, Albert Swart, Ladislav Mosner, Anna Silnova, Oldrich Plchot, Themos Stafylakis, Lukas Burget</i>	
Deep Speaker Embedding with Frame-Constrained Training Strategy for Speaker Verification	1451
<i>Bin Gu</i>	
Interrelate Training and Searching: A Unified Online Clustering Framework for Speaker Diarization.....	1456
<i>Yifan Chen, Yifan Guo, Qingxuan Li, Gaofeng Cheng, Pengyuan Zhang, Yonghong Yan</i>	
End-To-End Audio-Visual Neural Speaker Diarization.....	1461
<i>Mao-Kui He, Jun Du, Chin-Hui Lee</i>	
Online Speaker Diarization with Core Samples Selection	1466
<i>Yanyan Yue, Jun Du, Mao-Kui He, Yuting Yeung, Renyu Wang</i>	
Robust End-To-End Speaker Diarization with Generic Neural Clustering	1471
<i>Chenyu Yang, Yu Wang</i>	
MSDWild: Multi-Modal Speaker Diarization Dataset in the Wild.....	1476
<i>Tao Liu, Shuai Fan, Xu Xiang, Hongbo Song, Shaoxiong Lin, Jiaqi Sun, Tianyuan Han, Siyuan Chen, Binwei Yao, Sen Liu, Yifei Wu, Yanmin Qian, Kai Yu</i>	
Unsupervised Speaker Diarization that is Agnostic to Language, Overlap-Aware, and Tuning Free.....	1481
<i>Md Iftekhhar Tanveer, Diego Casabuena, Jussi Karlgren, Rosie Jones</i>	
Utterance-By-Utterance Overlap-Aware Neural Diarization with Graph-PIT.....	1486
<i>Keisuke Kinoshita, Thilo Von Neumann, Marc Delcroix, Christoph Boeddeker, Reinhold Haeb-Umbach</i>	
Spatial-Aware Speaker Diarization for Multi-Channel Multi-Party Meeting.....	1491
<i>Jie Wang, Yuji Liu, Binling Wang, Yiming Zhi, Song Li, Shipeng Xia, Jiayang Zhang, Feng Tong, Lin Li, Qingyang Hong</i>	

ACOUSTIC EVENT DETECTION AND CLASSIFICATION

Selective Pseudo-Labeling and Class-Wise Discriminative Fusion for Sound Event Detection..... 1496
Yunhao Liang, Yanhua Long, Yijie Li, Jiaen Liang

VOLUME 3

An End-To-End Macaque Voiceprint Verification Method Based on Channel Fusion Mechanism 1501
Peng Liu, Songbin Li, Jigang Tang

Human Sound Classification Based on Feature Fusion Method with Air and Bone Conducted
Signal..... 1506
Liang Xu, Jing Wang, Lizhong Wang, Sijun Bi, Jianqian Zhang, Qiuyue Ma

RaDur: A Reference-Aware and Duration-Robust Network for Target Sound Detection.....1511
Dongchao Yang, Helin Wang, Zhongjie Ye, Yuexian Zou, Wenwu Wang

Temporal Self Attention-Based Residual Network for Environmental Sound Classification..... 1516
Achyut Tripathi, Konark Paul

AudioTagging Done Right: 2nd Comparison of Deep Learning Methods for Environmental Sound
Classification 1521
Juncheng Li, Shuhui Qu, Po-Yao Huang, Florian Metze

Improving Target Sound Extraction with Timestamp Information..... 1526
Helin Wang, Dongchao Yang, Chao Weng, Jianwei Yu, Yuexian Zou

A Multi-Grained Based Attention Network for Semi-Supervised Sound Event Detection 1531
Ying Hu, Xiujuan Zhu, Yunlong Li, Hao Huang, Liang He

Temporal Coding with Magnitude-Phase Regularization for Sound Event Detection 1536
Sangwook Park, Sandeep Reddy Kothinti, Mounya Elhilali

RCT: Random Consistency Training for Semi-Supervised Sound Event Detection 1541
Nian Shao, Erfan Loweimi, Xiaofei Li

Audio Pyramid Transformer with Domain Adaption for Weakly Supervised Sound Event Detection
and Audio Classification..... 1546
Yifei Xin, Dongchao Yang, Yuexian Zou

Active Few-Shot Learning for Sound Event Detection..... 1551
Yu Wang, Mark Cartwright, Juan Pablo Bello

Uncertainty Calibration for Deep Audio Classifiers..... 1556
Tong Ye, Shijing Si, Jianzong Wang, Ning Cheng, Jing Xiao

Event-Related Data Conditioning for Acoustic Event Classification 1561
Yuanbo Hou, Dick Botteldooren

SPEECH SYNTHESIS: ACOUSTIC MODELING AND NEURAL WAVEFORM GENERATION

II

A Multi-Scale Time-Frequency Spectrogram Discriminator for GAN-Based Non-Autoregressive TTS.....	1566
<i>Haohan Guo, Hui Lu, Xixin Wu, Helen Meng</i>	
RetrieverTTS: Modeling Decomposed Factors for Text-Based Speech Insertion.....	1571
<i>Dacheng Yin, Chuanxin Tang, Yanqing Liu, Xiaoqiang Wang, Zhiyuan Zhao, Yucheng Zhao, Zhiwei Xiong, Sheng Zhao, Chong Luo</i>	
FlowVocoder: A Small Footprint Neural Vocoder Based Normalizing Flow for Speech Synthesis.....	1576
<i>Manh Luong, Viet Anh Tran</i>	
DelightfulTTS 2: End-To-End Speech Synthesis with Adversarial Vector-Quantized Auto-Encoders.....	1581
<i>Yanqing Liu, Ruiqing Xue, Lei He, Xu Tan, Sheng Zhao</i>	
AdaVocoder: Adaptive Vocoder for Custom Voice.....	1586
<i>Xin Yuan, Robin Feng, Mingming Ye, Cheng Tuo, Minhang Zhang</i>	
RefineGAN: Universally Generating Waveform Better than Ground Truth with Highly Accurate Pitch and Intensity Responses.....	1591
<i>Shengyuan Xu, Wenxiao Zhao, Jing Guo</i>	
VQTTS: High-Fidelity Text-To-Speech Synthesis with Self-Supervised VQ Acoustic Feature.....	1596
<i>Chenpeng Du, Yiwei Guo, Xie Chen, Kai Yu</i>	
Improving GAN-Based Vocoder for Fast and High-Quality Speech Synthesis.....	1601
<i>He Mengnan, Tingwei Guo, Zhenxing Lu, Zhang Ruixiong, Gong Caixia</i>	
SoftSpeech: Unsupervised Duration Model in FastSpeech 2.....	1606
<i>Yuan-Hao Yi, Lei He, Shifeng Pan, Xi Wang, Yuchao Zhang</i>	
A Multi-Stage Multi-Codebook VQ-VAE Approach to High-Performance Neural TTS.....	1611
<i>Haohan Guo, Feng-Long Xie, Frank Soong, Xixin Wu, Helen Meng</i>	
SiD-WaveFlow: A Low-Resource Vocoder Independent of Prior Knowledge.....	1616
<i>Yuhan Li, Ying Shen, Dongqing Wang, Lin Zhang</i>	
Text-To-Speech Synthesis Using Spectral Modeling Based on Non-Negative Autoencoder.....	1621
<i>Takeru Gorai, Daisuke Saito, Nobuaki Minematsu</i>	
Joint Modeling of Multi-Sample and Subband Signals for Fast Neural Vocoding on CPU.....	1626
<i>Hiroki Kanagawa, Yusuke Ijima, Hiroyuki Toda</i>	
MISRNet: Lightweight Neural Vocoder Using Multi-Input Single Shared Residual Blocks.....	1631
<i>Takuhiro Kaneko, Hirokazu Kameoka, Kou Tanaka, Shogo Seki</i>	
A Compact Transformer-Based GAN Vocoder.....	1636
<i>Chenfeng Miao, Ting Chen, Minchuan Chen, Jun Ma, Shaojun Wang, Jing Xiao</i>	
Diffusion Generative Vocoder for Fullband Speech Synthesis Based on Weak Third-Order SDE Solver.....	1641
<i>Hideyuki Tachibana, Muneyoshi Inahara, Mocho Go, Yotaro Katayama, Yotaro Watanabe</i>	

ASR: ARCHITECTURE AND SEARCH

On the Optimal Interpolation Weights for Hybrid Autoregressive Transducer Model	1646
<i>Ehsan Variani, Michael Riley, David Rybach, Cyril Allauzen, Tongzhou Chen, Bhuvana Ramabhadran</i>	
Learning to Rank with BERT-Based Confidence Models in ASR Rescoring	1651
<i>Ting-Wei Wu, I-Fan Chen, Ankur Gandhe</i>	
VQ-T: RNN Transducers Using Vector-Quantized Prediction Network States	1656
<i>Jiatong Shi, George Saon, David Haws, Shinji Watanabe, Brian Kingsbury</i>	
WeNet 2.0: More Productive End-To-End Speech Recognition Toolkit	1661
<i>Binbin Zhang, Di Wu, Zhendong Peng, Xingchen Song, Zhuoyuan Yao, Hang Lv, Lei Xie, Chao Yang, Fuping Pan, Jianwei Niu</i>	
Internal Language Model Estimation Through Explicit Context Vector Learning for Attention-Based Encoder-Decoder ASR	1666
<i>Yufei Liu, Rao Ma, Haihua Xu, Yi He, Zejun Ma, Weibin Zhang</i>	
Improving Streaming End-To-End ASR on Transformer-Based Causal Models with Encoder States Revision Strategies	1671
<i>Zehan Li, Haoran Miao, Keqi Deng, Gaofeng Cheng, Sanli Tian, Ta Li, Yonghong Yan</i>	
Parameter-Efficient Conformers Via Sharing Sparsely-Gated Experts for End-To-End Speech Recognition	1676
<i>Ye Bai, Jie Li, Wenjing Han, Hao Ni, Kaituo Xu, Zhuo Zhang, Cheng Yi, Xiaorui Wang</i>	
CaTT-KWS: A Multi-Stage Customized Keyword Spotting Framework Based on Cascaded Transducer-Transformer	1681
<i>Zhanheng Yang, Sining Sun, Jin Li, Xiaoming Zhang, Xiong Wang, Long Ma, Lei Xie</i>	
LightHuBERT: Lightweight and Configurable Speech Representation Learning with Once-For-All Hidden-Unit BERT	1686
<i>Rui Wang, Qibing Bai, Junyi Ao, Long Zhou, Zhixiang Xiong, Zhihua Wei, Yu Zhang, Tom Ko, Haizhou Li</i>	
Multi-Stage Progressive Compression of Conformer Transducer for On-Device Speech Recognition	1691
<i>Jash Rathod, Nauman Dawalatabad, Shatrughan Singh, Dhananjaya Gowda</i>	
Streaming Align-Refine for Non-Autoregressive Deliberation	1696
<i>Wang Weiran, Ke Hu, Tara Sainath</i>	
Federated Pruning: Improving Neural Network Efficiency with Federated Learning	1701
<i>Rongmei Lin, Yonghui Xiao, Tien-Ju Yang, Ding Zhao, Li Xiong, Giovanni Motta, Francoise Beaufays</i>	
A Unified Cascaded Encoder ASR Model for Dynamic Model Sizes	1706
<i>Shaojin Ding, Wang Weiran, Ding Zhao, Tara Sainath, Yanzhang He, Robert David, Rami Botros, Xin Wang, Rina Panigrahy, Qiao Liang, Dongseong Hwang, Ian McGraw, Rohit Prabhavalkar, Trevor Strohman</i>	
4-Bit Conformer with Native Quantization Aware Training for Speech Recognition	1711
<i>Shaojin Ding, Phoenix Meadowlark, Yanzhang He, Lukasz Lew, Shivani Agrawal, Oleg Rybakov</i>	

Self-Distillation Based on High-Level Information Supervision for Compressing End-To-End ASR Model	1716
<i>Qiang Xu, Tongtong Song, Longbiao Wang, Hao Shi, Yuqin Lin, Yongjie Lv, Meng Ge, Qiang Yu, Jianwu Dang</i>	

SPOKEN LANGUAGE PROCESSING II

Leveraging Unsupervised and Weakly-Supervised Data to Improve Direct Speech-To-Speech Translation	1721
<i>Ye Jia, Yifan Ding, Ankur Bapna, Colin Cherry, Yu Zhang, Alexis Conneau, Nobu Morioka</i>	
A High-Quality and Large-Scale Dataset for English-Vietnamese Speech Translation	1726
<i>Linh The Nguyen, Nguyen Luong Tran, Long Doan, Manh Luong, Dat Quoc Nguyen</i>	
Investigating Parameter Sharing in Multilingual Speech Translation	1731
<i>Qian Wang, Chen Wang, Jiajun Zhang</i>	
Open Source MagicData-RAMC: A Rich Annotated Mandarin Conversational(RAMC) Speech Dataset	1736
<i>Zehui Yang, Yifan Chen, Lei Luo, Runyan Yang, Lingxuan Ye, Gaofeng Cheng, Ji Xu, Yaohui Jin, Qingqing Zhang, Pengyuan Zhang, Lei Xie, Yonghong Yan</i>	
TALCS: An Open-Source Mandarin-English Code-Switching Corpus and a Speech Recognition Baseline	1741
<i>Chengfei Li, Shuhao Deng, Yaoping Wang, Guangjing Wang, Yaguang Gong, Changbin Chen, Jinfeng Bai</i>	
Blockwise Streaming Transformer for Spoken Language Understanding and Simultaneous Speech Translation	1746
<i>Keqi Deng, Shinji Watanabe, Jiatong Shi, Siddhant Arora</i>	
BARTpho: Pre-Trained Sequence-To-Sequence Models for Vietnamese	1751
<i>Nguyen Luong Tran, Duong Le, Dat Quoc Nguyen</i>	
Biometric Russian Audio-Visual Extended MASKS (BRAVE-MASKS) Corpus: Multimodal Mask Type Recognition Task	1756
<i>Maxim Markitantov, Elena Ryumina, Dmitry Ryumin, Alexey Karpov</i>	
Bayesian Transformer Using Disentangled Mask Attention	1761
<i>Jen-Tzung Chien, Yu-Han Huang</i>	
Audio-Visual Speech Recognition in MISP2021 Challenge: Dataset Release and Deep Analysis	1766
<i>Hang Chen, Jun Du, Yusheng Dai, Chin-Hui Lee, Sabato Marco Siniscalchi, Shinji Watanabe, Odette Scharenborg, Jingdong Chen, Baocai Yin, Jia Pan</i>	
From Start to Finish: Latency Reduction Strategies for Incremental Speech Synthesis in Simultaneous Speech-To-Speech Translation	1771
<i>Danni Liu, Changhan Wang, Hongyu Gong, Xutai Ma, Yun Tang, Juan Pino</i>	
Isochrony-Aware Neural Machine Translation for Automatic Dubbing	1776
<i>Derek Tam, Surafel M. Lakew, Yogesh Virkar, Prashant Mathur, Marcello Federico</i>	
Leveraging Pseudo-Labeled Data to Improve Direct Speech-To-Speech Translation	1781
<i>Qianqian Dong, Fengpeng Yue, Tom Ko, Mingxuan Wang, Qibing Bai, Yu Zhang</i>	

SOURCE SEPARATION I

A Hybrid Continuity Loss to Reduce Over-Suppression for Time-Domain Target Speaker Extraction	1786
<i>Zexu Pan, Meng Ge, Haizhou Li</i>	
Extending GCC-PHAT Using Shift Equivariant Neural Networks	1791
<i>Axel Berg, Mark O'Connor, Kalle Åström, Magnus Oskarsson</i>	
Heterogeneous Target Speech Separation.....	1796
<i>Efthymios Tzinis, Gordon Wichern, Aswin Shanmugam Subramanian, Paris Smaragdis, Jonathan Le Roux</i>	
Separate What You Describe: Language-Queried Audio Source Separation.....	1801
<i>Xubo Liu, Haohe Liu, Qiuqiang Kong, Xinhao Mei, Jinzheng Zhao, Qiushi Huang, Mark D. Plumbley, Wenwu Wang</i>	
Implicit Neural Spatial Filtering for Multichannel Source Separation in the Waveform Domain.....	1806
<i>Dejan Markovic, Alexandre Defossez, Alexander Richard</i>	

ASR TECHNOLOGIES AND SYSTEMS

End-To-End Speech-To-Punctuated-Text Recognition.....	1811
<i>Jumon Nozaki, Tatsuya Kawahara, Kenkichi Ishizuka, Taiichi Hashimoto</i>	
End-To-End Dependency Parsing of Spoken French	1816
<i>Adrien Pupier, Maximin Coavoux, Benjamin Lecouteux, Jerome Goulian</i>	
Turn-Taking Prediction for Natural Conversational Speech.....	1821
<i>Shuo-Yiin Chang, Bo Li, Tara Sainath, Chao Zhang, Trevor Strohman, Qiao Liang, Yanzhang He</i>	
Streaming Intended Query Detection Using E2E Modeling for Continued Conversation	1826
<i>Shuo-Yiin Chang, Guru Prakash, Zelin Wu, Tara Sainath, Bo Li, Qiao Liang, Adam Stambler, Shyam Upadhyay, Manaal Faruqui, Trevor Strohman</i>	
Exploring Capabilities of Monolingual Audio Transformers Using Large Datasets in Automatic Speech Recognition of Czech.....	1831
<i>Jan Lehecka, Jan Švec, Ales Prazak, Josef Psutka</i>	
SVTS: Scalable Video-To-Speech Synthesis.....	1836
<i>Rodrigo Schoburg Carrillo De Mira, Alexandros Haliassos, Stavros Petridis, Björn W. Schuller, Maja Pantic</i>	

SPEECH PERCEPTION

One-Step Models in Pitch Perception: Experimental Evidence from Japanese.....	1841
<i>Takeshi Kishiyama, Chuyu Huang, Yuki Hirose</i>	
Generating Iso-Accented Stimuli for Second Language Research: Methodology and a Dataset for Spanish-Accented English.....	1846
<i>Rubén Pérez Ramón, Martin Cooke, Maria Luisa Garcia Lecumberri</i>	

Factors Affecting the Percept of Yanny V. Laurel (or Mixed): Insights from a Large-Scale Study on Swiss German Listeners	1851
<i>Adrian Leemann, Péter Jeszenszky, Carina Steiner, Corinne Lanthemann</i>	
Effects of Laryngeal Manipulations on Voice Gender Perception	1856
<i>Zhaoyan Zhang, Jason Zhang, Jody Kreiman</i>	
Why is Korean Lenis Stop Difficult to Perceive for L2 Korean Learners?	1861
<i>Boram Lee, Naomi Yamaguchi, Cécile Fougeron</i>	
Lexical Stress in Spanish Word Segmentation	1866
<i>Alvaro Martin Iturralde Zurita, Meghan Clayards</i>	

SPOKEN TERM DETECTION AND VOICE SEARCH

Learning Audio-Text Agreement for Open-Vocabulary Keyword Spotting	1871
<i>Hyeon-Kyeong Shin, Hyewon Han, Doyeon Kim, Soo-Whan Chung, Hong-Goo Kang</i>	
Integrating Form and Meaning: A Multi-Task Learning Model for Acoustic Word Embeddings	1876
<i>Badr M. Abdullah, Bernd Möbius, Dietrich Klakow</i>	
Personalized Keyword Spotting Through Multi-Task Learning	1881
<i>Seunghan Yang, Byeonggeun Kim, Inseop Chung, Simyung Chang</i>	
Deep LSTM Spoken Term Detection Using Wav2Vec 2.0 Recognizer	1886
<i>Jan Švec, Jan Lehecka, Luboš Šmídl</i>	
Latency Control for Keyword Spotting	1891
<i>Christin Jose, Joe Wang, Grant Strimel, Mohammad Omar Khurshed, Yuriy Mishchenko, Brian Kulis</i>	
Improving Voice Trigger Detection with Metric Learning	1896
<i>Prateeth Nayak, Takuya Higuchi, Anmol Gupta, Shivesh Ranjan, Stephen Shum, Siddharth Sigtia, Erik Marchi, Varun Lakshminarasimhan, Minsik Cho, Saurabh Adya, Chandra Dhir, Ahmed Tewfik</i>	

SPEECH AND LANGUAGE IN HEALTH: FROM REMOTE MONITORING TO MEDICAL CONVERSATIONS I

RNN Transducers for Named Entity Recognition with Constraints on Alignment for Understanding Medical Conversations	1901
<i>Hagen Soltau, Izhak Shafran, Mingqiu Wang, Laurent El Shafey</i>	
Towards Automated Counselling Decision-Making: Remarks on Therapist Action Forecasting on the AnnoMI Dataset	1906
<i>Zixiu Wu, Rim Helaoui, Diego Reforgiato Recupero, Daniele Riboni</i>	
Speech and the n-Back Task as a Lens into Depression. How Combining Both May Allow Us to Isolate Different Core Symptoms of Depression	1911
<i>Salvatore Fara, Stefano Gorla, Emilia Molimpakis, Nicholas Cummins</i>	
Enabling Off-The-Shelf Disfluency Detection and Categorization for Pathological Speech	1916
<i>Amrit Romana, Minxue Niu, Matthew Perez, Angela Roberts, Emily Mower Provost</i>	

Challenges of Using Longitudinal and Cross-Domain Corpora on Studies of Pathological Speech..... 1921
Catarina Botelho, Tanja Schultz, Alberto Abad, Isabel Trancoso

SPEECH SYNTHESIS: LINGUISTIC PROCESSING, PARADIGMS AND OTHER TOPICS I

g2pW: A Conditional Weighted Softmax BERT for Polyphone Disambiguation in Mandarin 1926
Yi-Chang Chen, Yu-Chuan Steven, Yen-Cheng Chang, Yi-Ren Yeh

A Unified Accent Estimation Method Based on Multi-Task Learning for Japanese Text-To-Speech 1931
Byeongseon Park, Ryuichi Yamamoto, Kentaro Tachibana

Vocal Effort Modeling in Neural TTS for Improving the Intelligibility of Synthetic Speech in Noise 1936
Tuomo Raitio, Petko Petkov, Jiangchuan Li, Muhammed Shifas, Andrea Davis, Yannis Stylianou

TTS-By-TTS 2: Data-Selective Augmentation for Neural Speech Synthesis Using Ranking Support
Vector Machine with Variational Autoencoder 1941
Eunwoo Song, Ryuichi Yamamoto, Ohsung Kwon, Chan-Ho Song, Min-Jae Hwang, Suhyeon Oh, Hyun-Wook Yoon, Jin-Seob Kim, Jae-Min Kim

Low-Data? No Problem: Low-Resource, Language-Agnostic Conversational Text-To-Speech Via
F0-Conditioned Data Augmentation 1946
Giulia Comini, Goeric Huybrechts, Manuel Sam Ribeiro, Adam Gabrys, Jaime Lorenzo-Trueba

SHOW AND TELL II

Real-Time Monitoring of Silences in Contact Center Conversations..... 1951
Digvijay Ingle, Ayush Kumar, Krishnachaitanya Gogineni, Jithendra Vepa

Humanizing Bionic Voice: Interactive Demonstration of Aesthetic Design and Control Factors
Influencing the Devices Assembly and Waveshape Engineering 1953
Konrad Zielinski, Marek Grzelec, Martin Hagmüller

Application for Real-Time Personalized Speaker Extraction 1955
Damien Ronssin, Milos Cernak

Coswara: A Website Application Enabling COVID-19 Screening by Analysing Respiratory Sound
Samples and Health Symptoms 1957
Debarpan Bhattacharya, Debottam Dutta, Neeraj Kumar Sharma, Srikanth Raj Chetupalli, Pravin Mote, Sriram Ganapathy, Chandrakiran C, Sahiti Nori, Suhail K K, Sadhana Gonuguntla, Murali Alagesan

CoachLea: An Android Application to Evaluate the Speech Production and Perception of Children
with Hearing Loss 1959
P. Schäfer, P. A. Pérez-Toro, P. Klumpp, J. R. Orozco-Arroyave, E. Nöth, K. Maier, A. Abad, M. Schuster, T. Arias-Vergara

An Automated Mood Diary for Older User's Using Ambient Assisted Living Recorded Speech..... 1961
Fasih Haider, Saturnino Luz

MULTIMODAL SPEECH EMOTION RECOGNITION AND PARALINGUISTICS

Differential Time-Frequency Log-Mel Spectrogram Features for Vision Transformer Based Infant Cry Recognition	1963
<i>Hai-Tao Xu, Jie Zhang, Li-Rong Dai</i>	
Towards Automated Dialog Personalization Using MBTI Personality Indicators	1968
<i>Daniel Fernau, Stefan Hillmann, Nils Feldhus, Tim Polzehl</i>	
Word-Wise Sparse Attention for Multimodal Sentiment Analysis	1973
<i>Fan Qian, Hongwei Song, Jiqing Han</i>	
Estimation of Speaker Age and Height from Speech Signal Using Bi-Encoder Transformer Mixture Model	1978
<i>Tarun Gupta, Tuan Duc Truong, Tran The Anh, Eng Siong Chng</i>	
Exploring Multi-Task Learning Based Gender Recognition and Age Estimation for Class-Imbalanced Data	1983
<i>Wei qiao Zheng, Ping Yang, Rongfeng Lai, Kongyang Zhu, Tao Zhang, Junpeng Zhang, Hongcheng Fu</i>	
Audio-Visual Domain Adaptation Feature Fusion for Speech Emotion Recognition	1988
<i>Jie Wei, Guanyu Hu, Xinyu Yang, Anh Tuan Luu, Yizhuo Dong</i>	
Impact of Background Noise and Contribution of Visual Information in Emotion Identification by Native Mandarin Speakers	1993
<i>Minyue Zhang, Hongwei Ding</i>	
Exploiting Fine-Tuning of Self-Supervised Learning Models for Improving Bi-Modal Sentiment Analysis and Emotion Recognition	1998
<i>Wei Yang, Satoru Fukayama, Panikos Heracleous, Jun Ogata</i>	
Characterizing Therapist's Speaking Style in Relation to Empathy in Psychotherapy	2003
<i>Dehua Tao, Tan Lee, Harold Chui, Sarah Luk</i>	
Hierarchical Attention Network for Evaluating Therapist Empathy in Counseling Session	2008
<i>Dehua Tao, Tan Lee, Harold Chui, Sarah Luk</i>	
Context-Aware Multimodal Fusion for Emotion Recognition.....	2013
<i>Jinchao Li, Shuai Wang, Yang Chao, Xunying Liu, Helen Meng</i>	
Unsupervised Instance Discriminative Learning for Depression Detection from Speech Signals	2018
<i>Jinhan Wang, Vijay Ravi, Jonathan Flint, Abeer Alwan</i>	
How Do Our Eyebrows Respond to Masks and Whispering? the Case of Persians	2023
<i>Nasim Mahdinazhad Sardhaei, Marzena Zygis, Hamid Sharifzadeh</i>	
State & Trait Measurement from Nonverbal Vocalizations: A Multi-Task Joint Learning Approach	2028
<i>Alice Baird, Panagiotis Tzirakis, Jeff Brooks, Lauren Kim, Michael Opara, Chris Gregory, Jacob Metrick, Garrett Boseck, Dacher Keltner, Alan Cowen</i>	
Confidence Measure for Automatic Age Estimation from Speech	2033
<i>Amruta Saraf, Ganesh Sivaraman, Elie Khoury</i>	

NEURAL TRANSDUCERS, STREAMING ASR AND NOVEL ASR MODELS

Accelerating Inference and Language Model Fusion of Recurrent Neural Network Transducers Via End-To-End 4-Bit Quantization.....	2038
<i>Andrea Fasoli, Chia-Yu Chen, Mauricio Serrano, Swagath Venkataramani, George Saon, Xiaodong Cui, Brian Kingsbury, Kailash Gopalakrishnan</i>	
Tree-Constrained Pointer Generator with Graph Neural Network Encodings for Contextual Speech Recognition	2043
<i>Guangzhi Sun, Chao Zhang, Phil Woodland</i>	
Bring Dialogue-Context into RNN-T for Streaming ASR.....	2048
<i>Junfeng Hou, Jinkun Chen, Wanyu Li, Yufeng Tang, Jun Zhang, Zejun Ma</i>	
Conformer with Dual-Mode Chunked Attention for Joint Online and Offline ASR	2053
<i>Felix Weninger, Marco Gaudesi, Md Akmal Haidar, Nicola Ferri, Jesús Andrés-Ferrer, Puming Zhan</i>	
Efficient Training of Neural Transducer for Speech Recognition	2058
<i>Wei Zhou, Wilfried Michel, Ralf Schlüter, Hermann Ney</i>	
Paraformer: Fast and Accurate Parallel Transformer for Non-Autoregressive End-To-End Speech Recognition	2063
<i>Zhifu Gao, Shiliang Zhang, Ian McLoughlin, Zhijie Yan</i>	
Pruned RNN-T for Fast, Memory-Efficient ASR Training.....	2068
<i>Fangjun Kuang, Liyong Guo, Wei Kang, Long Lin, Mingshuang Luo, Zengwei Yao, Daniel Povey</i>	
Deep Sparse Conformer for Speech Recognition	2073
<i>Xianchao Wu</i>	
Chain-Based Discriminative Autoencoders for Speech Recognition	2078
<i>Hung-Shin Lee, Pin-Tuan Huang, Yao-Fei Cheng, Hsin-Min Wang</i>	
Streaming Parallel Transducer Beam Search with Fast Slow Cascaded Encoders	2083
<i>Jay Mahadeokar, Yangyang Shi, Ke Li, Duc Le, Jiedan Zhu, Vikas Chandra, Ozlem Kalinli, Michael Seltzer</i>	
Self-Regularised Minimum Latency Training for Streaming Transformer-Based Speech Recognition	2088
<i>Mohan Li, Rama Sanand Doddipatla, Catalin Zorila</i>	
On the Prediction Network Architecture in RNN-T for ASR	2093
<i>Dario Albesano, Jesús Andrés-Ferrer, Nicola Ferri, Puming Zhan</i>	
Minimum Latency Training of Sequence Transducers for Streaming End-To-End Speech Recognition	2098
<i>Yusuke Shinohara, Shinji Watanabe</i>	
CUSIDE: Chunking, Simulating Future Context and Decoding for Streaming ASR.....	2103
<i>Keyu An, Huahuan Zheng, Zhijian Ou, Hongyu Xiang, Ke Ding, Guanglu Wan</i>	
Attention Enhanced Citrinet for Speech Recognition.....	2108
<i>Xianchao Wu</i>	

ZERO, LOW-RESOURCE AND MULTI-MODAL SPEECH RECOGNITION II

Simple and Effective Zero-Shot Cross-Lingual Phoneme Recognition	2113
<i>Qiantong Xu, Alexei Baeovski, Michael Auli</i>	
Robust Self-Supervised Audio-Visual Speech Recognition	2118
<i>Bowen Shi, Wei-Ning Hsu, Abdelrahman Mohamed</i>	
Speech Sequence Embeddings Using Nearest Neighbors Contrastive Learning.....	2123
<i>Robin Algayres, Adel Nabli, Benoît Sagot, Emmanuel Dupoux</i>	
Towards Green ASR: Lossless 4-Bit Quantization of a Hybrid TDNN System on the 300-Hr Switchover Corpus	2128
<i>Junhao Xu, Shoukang Hu, Xunying Liu, Helen Meng</i>	
Finer-Grained Modeling Units-Based Meta-Learning for Low-Resource Tibetan Speech Recognition	2133
<i>Siqing Qin, Longbiao Wang, Sheng Li, Yuqin Lin, Jianwu Dang</i>	

ATYPICAL SPEECH ANALYSIS AND DETECTION

Adversarial-Free Speaker Identity-Invariant Representation Learning for Automatic Dysarthric Speech Classification.....	2138
<i>Parvaneh Janbakhshi, Ina Kodrasi</i>	
Automated Detection of Wilson’s Disease Based on Improved Mel-Frequency Cepstral Coefficients with Signal Decomposition	2143
<i>Zhenglin Zhang, Li-Zhuang Yang, Xun Wang, Hai Li</i>	
The Effect of Backward Noise on Lexical Tone Discrimination in Mandarin-Speaking Amusics.....	2148
<i>Zixia Fan, Jing Shao, Weigong Pan, Min Xu, Lan Wang</i>	
Automatic Selection of Discriminative Features for Dementia Detection in Cantonese-Speaking People.....	2153
<i>Xiaoquan Ke, Man-Wai Mak, Helen M. Meng</i>	
Automated Voice Pathology Discrimination from Continuous Speech Benefits from Analysis by Phonetic Context	2158
<i>Zhuoya Liu, Mark Huckvale, Julian McGlashan</i>	
Multi-Type Outer Product-Based Fusion of Respiratory Sounds for Detecting COVID-19	2163
<i>Adria Mallol-Ragolta, Helena Cuesta, Emilia Gomez, Björn Schuller</i>	
Robust Cough Feature Extraction and Classification Method for COVID-19 Cough Detection Based on Vocalization Characteristics	2168
<i>Xueshuai Zhang, Jiakun Shen, Jun Zhou, Pengyuan Zhang, Yonghong Yan, Zhihua Huang, Yanfen Tang, Yu Wang, Fujie Zhang, Shaoxing Zhang, Aijun Sun</i>	
Comparing 1-Dimensional and 2-Dimensional Spectral Feature Representations in Voice Pathology Detection Using Machine Learning and Deep Learning Classifiers.....	2173
<i>Farhad Javanmardi, Sudarsana Reddy Kadiri, Manila Kodali, Paavo Alku</i>	
Zero-Shot Cross-Lingual Aphasia Detection Using Automatic Speech Recognition	2178
<i>Gerasimos Chatzoudis, Manos Plitsis, Spyridoula Stamouli, Athanasia-lida Dimou, Nassos Katsamanis, Vassilis Katsouros</i>	

Domain-Aware Intermediate Pretraining for Dementia Detection with Limited Data 2183
Youxiang Zhu, Xiaohui Liang, John A. Batsis, Robert M. Roth

Comparison of 5 Methods for the Evaluation of Intelligibility in Mild to Moderate French
Dysarthric Speech..... 2188
*Cécile Fougeron, Nicolas Audibert, Ina Kodrasi, Parvaneh Janbakhshi, Michaela Pernon,
Nathalie Leveque, Stephanie Borel, Marina Laganaro, Herve Bourlard, Frederic Assal*

ADAPTATION, TRANSFER LEARNING, AND DISTILLATION FOR ASR

Improving Distortion Robustness of Self-Supervised Speech Processing Tasks with Domain
Adaptation 2193
Kuan Po Huang, Yu-Kuan Fu, Yu Zhang, Hung-Yi Lee

Listen, Adapt, Better WER: Source-Free Single-Utterance Test-Time Adaptation for Automatic
Speech Recognition 2198
Guan-Ting Lin, Shang-Wen Li, Hung-Yi Lee

Distilling a Pretrained Language Model to a Multilingual ASR Model 2203
Kwanghee Choi, Hyung-Min Park

Text-Only Domain Adaptation Based on Intermediate CTC 2208
*Hiroaki Sato, Tomoyasu Komori, Takeshi Mishima, Yoshihiko Kawai, Takahiro Mochizuki,
Shoei Sato, Tetsuji Ogawa*

Transfer Learning for Robust Low-Resource Children's Speech ASR with Transformers and
Source-Filter Warping 2213
Jenthe Thienpondt, Kris Demuynck

Updating Only Encoders Prevents Catastrophic Forgetting of End-To-End ASR Models 2218
*Yuki Takashima, Shota Horiguchi, Shinji Watanabe, Leibny Paola Garcia Perera, Yohei
Kawaguchi*

SPEAKER AND LANGUAGE RECOGNITION I

Improved CNN-Transformer Using Broadcasted Residual Learning for Text-Independent Speaker
Verification 2223
Jeong-Hwan Choi, Joon-Young Yang, Ye-Rin Jeoung, Joon-Hyuk Chang

Pushing the Limits of Raw Waveform Speaker Recognition..... 2228
Jee-Weon Jung, Youjin Kim, Hee-Soo Heo, Bong-Jin Lee, Youngki Kwon, Joon Son Chung

PHO-LID: A Unified Model Incorporating Acoustic-Phonetic and Phonotactic Information for
Language Identification 2233
Hexin Liu, Leibny Paola Garcia Perera, Andy Khong, Suzy Styles, Sanjeev Khudanpur

Prosodic Information in Dialect Identification of a Tonal Language: The Case of Ao..... 2238
Moakala Tzudir, Priyankoo Sarmah, S R Mahadeva Prasanna

A Multimodal Strategy for Singing Language Identification 2243
Wo Jae Lee, Emanuele Coviello

PATHOLOGICAL SPEECH ANALYSIS

- A Comparative Study on Vowel Articulation in Parkinson's Disease and Multiple System Atrophy 2248
Khalid Daoudi, Biswajit Das, Solange Milh  De Saint Victor, Alexandra Foubert-Samier, Margherita Fabbri, Anne Pavy-Le Traon, Olivier Rascol, Virginie Woisard, Wassilios G. Meissner
- Voicing Decision Based on Phonemes Classification and Spectral Moments for Whisper-To-Speech Conversion..... 2253
Luc Ardaillon, Nathalie Henrich, Olivier Perrotin
- Speech Acoustics in Mild Cognitive Impairment and Parkinson's Disease with and Without Concurrent Drawing Tasks 2258
Tanya Talkar, Christina Manxhari, James Williamson, Kara M. Smith, Thomas Quatieri
- Investigating the Impact of Speech Compression on the Acoustics of Dysarthric Speech..... 2263
Kelvin Tran, Lingfeng Xu, Gabriela Stegmann, Julie Liss, Visar Berisha, Rene Utianski
- Speaker Trait Enhancement for Cochlear Implant Users: A Case Study for Speaker Emotion Perception..... 2268
Avamarie Brueggeman, John H. L. Hansen
- Optimal Thyroplasty Implant Shape and Stiffness for Treatment of Acute Unilateral Vocal Fold Paralysis: Evidence from a Canine in Vivo Phonation Model..... 2273
Neha Reddy, Yoonjeong Lee, Zhaoyan Zhang, Dinesh K. Chhetri

CROSS/MULTI-LINGUAL ASR

- XLS-R: Self-Supervised Cross-Lingual Speech Representation Learning at Scale 2278
Arun Babu, Changan Wang, Andros Tjandra, Kushal Lakhota, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick Von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, Michael Auli
- Semantically Meaningful Metrics for Norwegian ASR Systems..... 2283
Janine Rugayan, Torbj rn Svendsen, Giampiero Salvi
- Deciphering Speech: A Zero-Resource Approach to Cross-Lingual Transfer in ASR..... 2288
Ondrej Klejch, Electra Wallington, Peter Bell
- Linguistically Informed Post-Processing for ASR Error Correction in Sanskrit 2293
Rishabh Kumar, Devaraja Adiga, Rishav Ranjan, Amrith Krishna, Ganesh Ramakrishnan, Pawan Goyal, Preethi Jyothi
- Cross-Lingual Articulatory Feature Information Transfer for Speech Recognition Using Recurrent Progressive Neural Networks 2298
Mahir Morshed, Mark Hasegawa-Johnson

SPEAKING STYLES AND INTERACTION STYLES I

- Comparison of Models for Detecting Off-Putting Speaking Styles..... 2303
Diego Aguirre, Nigel Ward, Jonathan E. Avila, Heike Lehnert-Lehouillier

VOLUME 4

Multimodal Persuasive Dialogue Corpus Using Teleoperated Android	2308
<i>Seiya Kawano, Muteki Arioka, Akishige Yuguchi, Kenta Yamamoto, Koji Inoue, Tatsuya Kawahara, Satoshi Nakamura, Koichiro Yoshino</i>	
Text-Driven Emotional Style Control and Cross-Speaker Style Transfer in Neural TTS.....	2313
<i>Yookyung Shin, Younggun Lee, Suhee Jo, Yeongtae Hwang, Taesu Kim</i>	
Strategies for Developing a Conversational Speech Dataset for Text-To-Speech Synthesis.....	2318
<i>Adaeze O. Adigwe, Esther Klabbers</i>	
Deep CNN-Based Inductive Transfer Learning for Sarcasm Detection in Speech.....	2323
<i>Xiyuan Gao, Shekhar Nayak, Matt Coler</i>	

SPEAKING STYLES AND INTERACTION STYLES II

End-To-End Text-To-Speech Based on Latent Representation of Speaking Styles Using Spontaneous Dialogue.....	2328
<i>Kentaro Mitsui, Tianyu Zhao, Kei Sawada, Yukiya Hono, Yoshihiko Nankaku, Keiichi Tokuda</i>	
Attention-Based Conditioning Methods Using Variable Frame Rate for Style-Robust Speaker Verification.....	2333
<i>Amber Afshan, Abeer Alwan</i>	
Learning from Human Perception to Improve Automatic Speaker Verification in Style-Mismatched Conditions	2338
<i>Amber Afshan, Abeer Alwan</i>	
Exploring Audio-Based Stylistic Variation in Podcasts.....	2343
<i>Katariina Martikainen, Jussi Karlgren, Khiet Truong</i>	

SPEECH SYNTHESIS: TOOLS, DATA, AND EVALUATION

Automatic Evaluation of Speaker Similarity	2348
<i>Kamil Deja, Ariadna Sanchez, Julian Roth, Marius Cotescu</i>	
Mix and Match: An Empirical Study on Training Corpus Composition for Polyglot Text-To-Speech (TTS).....	2353
<i>Ziyao Zhang, Alessio Falai, Ariadna Sanchez, Orazio Angelini, Kayoko Yanagisawa</i>	
J-MAC: Japanese Multi-Speaker Audiobook Corpus for Speech Synthesis	2358
<i>Shinnosuke Takamichi, Wataru Nakata, Naoko Tanji, Hiroshi Saruwatari</i>	
REYD – the First Yiddish Text-To-Speech Dataset and System	2363
<i>Jacob Webber, Samuel K. Lo, Isaac L. Bleaman</i>	
Data-Augmented Cross-Lingual Synthesis in a Teacher-Student Framework.....	2368
<i>Marcel De Korte, Jaebok Kim, Aki Kunikoshi, Adaeze Adigwe, Esther Klabbers</i>	
Production Characteristics of Obstruents in WaveNet and Older TTS Systems.....	2373
<i>Ayushi Pandey, Sébastien Le Maguer, Julie Carson-Berndsen, Naomi Harte</i>	

Back to the Future: Extending the Blizzard Challenge 2013.....	2378
<i>Sébastien Le Maguer, Simon King, Naomi Harte</i>	
BibleTTS: A Large, High-Fidelity, Multilingual, and Uniquely African Speech Corpus	2383
<i>Josh Meyer, David Adelani, Edresson Casanova, Alp Öktem, Daniel Whitenack, Julian Weber, Salomon Kabongo Kabenamualu, Elizabeth Salesky, Iroko Orife, Colin Leong, Perez Ogayo, Chris Chinenye Emezue, Jonathan Mukiibi, Salomey Osei, Apelete Agbolo, Victor Akinode, Bernard Opoku, Olanrewaju Samuel, Jesujoba Alabi, Shamsuddeen Hassan Muhammad</i>	
SOMOS: The Samsung Open MOS Dataset for the Evaluation of Neural Text-To-Speech Synthesis	2388
<i>Georgia Maniati, Alexandra Vioni, Nikolaos Ellinas, Karolos Nikitaras, Konstantinos Klapsas, June Sig Sung, Gunu Jho, Aimilios Chalamandaris, Pirros Tsiakoulis</i>	

ACOUSTIC SIGNAL REPRESENTATION AND ANALYSIS II

Domain Generalization with Relaxed Instance Frequency-Wise Normalization for Multi-Device Acoustic Scene Classification	2393
<i>Byeonggeun Kim, Seunghan Yang, Jangho Kim, Hyunsin Park, Juntae Lee, Simyung Chang</i>	
Couple Learning for Semi-Supervised Sound Event Detection.....	2398
<i>Tao Rui, Yan Long, Ouchi Kazushige, Xiangdong Wang</i>	
Oktoechos Classification in Liturgical Music Using SBU-LSTM/GRU.....	2403
<i>Rajeev Rajan, Ananya Ayasi</i>	
SoundDoA: Learn Sound Source Direction of Arrival and Semantics from Sound Raw Waveforms.....	2408
<i>Yuhang He, Andrew Markham</i>	
ORCA-WHISPER: An Automatic Killer Whale Sound Type Generation Toolkit Using Deep Learning	2413
<i>Christian Bergler, Alexander Barnhill, Dominik Perrin, Manuel Schmitt, Andreas Maier, Elmar Nöth</i>	
Convolutional Recurrent Neural Network with Auxiliary Stream for Robust Variable-Length Acoustic Scene Classification	2418
<i>Joon-Hyuk Chang, Won-Gook Choi</i>	
Unsupervised Symbolic Music Segmentation Using Ensemble Temporal Prediction Errors.....	2423
<i>Shahaf Bassan, Yossi Adi, Jeffrey Rosenschein</i>	
Visually-Aware Acoustic Event Detection Using Heterogeneous Graphs.....	2428
<i>Amir Shirian, Krishna Somandepalli, Victor Sanchez, Tanaya Guha</i>	
A Passive Similarity Based CNN Filter Pruning for Efficient Acoustic Scene Classification.....	2433
<i>Arshdeep Singh, Mark D. Plumbley</i>	
MAE-AST: Masked Autoencoding Audio Spectrogram Transformer.....	2438
<i>Alan Baade, Puyuan Peng, David Harwath</i>	

SPEECH AND LANGUAGE IN HEALTH: FROM REMOTE MONITORING TO MEDICAL CONVERSATIONS II

What Can Speech and Language Tell Us About the Working Alliance in Psychotherapy.....	2443
<i>Sebastian Peter Bayerl, Gabriel Roccabruna, Shammur Absar Chowdhury, Tommaso Ciulli, Morena Danieli, Korbinian Riedhammer, Giuseppe Riccardi</i>	

TB Or Not TB? Acoustic Cough Analysis for Tuberculosis Classification	2448
<i>Geoffrey T. Frost, Grant Theron, Thomas Niesler</i>	
Are Reported Accuracies in the Clinical Speech Machine Learning Literature Overoptimistic?.....	2453
<i>Visar Berisha, Chelsea Krantsevich, Gabriela Stegmann, Shira Hahn, Julie Liss</i>	
Automatic Detection of Expressed Emotion from Five-Minute Speech Samples: Challenges and Opportunities	2458
<i>Bahman Mirheidari, Andre Bittar, Nicholas Cummins, Johnny Downs, Helen L. Fisher, Heidi Christensen</i>	
Automatic Cognitive Assessment: Combining Sparse Datasets with Disparate Cognitive Scores.....	2463
<i>Bahman Mirheidari, Daniel Blackburn, Heidi Christensen</i>	
Exploring Semi-Supervised Learning for Audio-Based COVID-19 Detection Using FixMatch	2468
<i>Ting Dang, Thomas Quinnell, Cecilia Mascolo</i>	
Analyzing the Impact of SARS-CoV-2 Variants on Respiratory Sound Signals	2473
<i>Debarpan Bhattacharya, Debottam Dutta, Neeraj Sharma, Srikanth Raj Chetupalli, Pravin Mote, Sriram Ganapathy, Chandrakiran C, Sahiti Nori, Suhail K K, Sadhana Gonuguntla, Murali Alagesan</i>	
Automated Evaluation of Standardized Dementia Screening Tests.....	2478
<i>Franziska Braun, Markus Förstel, Bastian Oppermann, Andreas Erzigkeit, Hartmut Lehfeld, Thomas Hillemacher, Korbinian Riedhammer</i>	
Alzheimer's Detection from English to Spanish Using Acoustic and Linguistic Embeddings	2483
<i>Paula Andrea Pérez-Toro, Philipp Klumpp, Abner Hernandez, Tomas Arias, Patricia Lillo, Andrea Slachevsky, Adolfo Martín García, Maria Schuster, Andreas K. Maier, Elmar Noeth, Juan Rafael Orozco-Arroyave</i>	
Extract and Abstract with BART for Clinical Notes from Doctor-Patient Conversations	2488
<i>Jing Su, Longxiang Zhang, Hamid Reza Hassanzadeh, Thomas Schaaf</i>	
Dyadic Interaction Assessment from Free-Living Audio for Depression Severity Assessment	2493
<i>Bishal Lamichhane, Nidal Moukaddam, Ankit B. Patel, Ashutosh Sabharwal</i>	
COVID-19 Detection Based on Respiratory Sensing from Speech.....	2498
<i>Venkata Srikanth Nallanthighal, Aki Harma, Helmer Strik</i>	

DEREVERBERATION AND ECHO CANCELLATION

Bifurcation and Reunion: A Loss-Guided Two-Stage Approach for Monaural Speech Dereverberation	2503
<i>Xiaoxue Luo, Chengshi Zheng, Andong Li, Yuxuan Ke, Xiaodong Li</i>	
A Deep Complex Multi-Frame Filtering Network for Stereophonic Acoustic Echo Cancellation.....	2508
<i>Linjuan Cheng, Chengshi Zheng, Andong Li, Yuquan Wu, Renhua Peng, Xiaodong Li</i>	
Speaker- And Phone-Aware Convolutional Transformer Network for Acoustic Echo Cancellation.....	2513
<i>Chang Han, Weiping Tu, Yuhong Yang, Jingyi Li, Xinhong Li</i>	
Personalized Acoustic Echo Cancellation for Full-Duplex Communications	2518
<i>Shimin Zhang, Ziteng Wang, Yukai Ju, Yihui Fu, Yueyue Na, Qiang Fu, Lei Xie</i>	

LCSM: A Lightweight Complex Spectral Mapping Framework for Stereophonic Acoustic Echo Cancellation.....	2523
<i>Chenggang Zhang, Jinjiang Liu, Xueliang Zhang</i>	
Joint Neural AEC and Beamforming with Double-Talk Detection	2528
<i>Vinay Kothapally, Yong Xu, Meng Yu, Shi-Xiong Zhang, Dong Yu</i>	
Clock Skew Robust Acoustic Echo Cancellation	2533
<i>Karim Helwani, Erfan Soltanmohammadi, Michael Mark Goodwin, Arvinth Krishnaswamy</i>	
A Conformer-Based Waveform-Domain Neural Acoustic Echo Canceller Optimized for ASR Accuracy.....	2538
<i>Sankaran Panchapagesan, Arun Narayanan, Turaj Zakizadeh Shabestary, Shuai Shao, Nathan Howard, Alex Park, James Walker, Alexander Gruenstein</i>	
Complex-Valued Time-Frequency Self-Attention for Speech Dereverberation	2543
<i>Vinay Kothapally, John H. L. Hansen</i>	

VOICE CONVERSION AND ADAPTATION III

Learning Noise-Independent Speech Representation for High-Quality Voice Conversion for Noisy Target Speakers	2548
<i>Liumeng Xue, Shan Yang, Na Hu, Dan Su, Lei Xie</i>	
Speech Representation Disentanglement with Adversarial Mutual Information Learning for One-Shot Voice Conversion	2553
<i>Sicheng Yang, Methawee Tantrawenith, Haolin Zhuang, Zhiyong Wu, Aolan Sun, Jianzong Wang, Ning Cheng, Huaizhen Tang, Xintao Zhao, Jie Wang, Helen Meng</i>	
FlowCPCVC: A Contrastive Predictive Coding Supervised Flow Framework for Any-To-Any Voice Conversion.....	2558
<i>Jiahong Huang, Wen Xu, Yule Li, Junshi Liu, Dongpeng Ma, Wei Xiang</i>	
Glow-WaveGAN 2: High-Quality Zero-Shot Text-To-Speech Synthesis and Any-To-Any Voice Conversion.....	2563
<i>Yi Lei, Shan Yang, Jian Cong, Lei Xie, Dan Su</i>	
AdaSpeech 4: Adaptive Text to Speech in Zero-Shot Scenarios	2568
<i>Yihan Wu, Xu Tan, Bohan Li, Lei He, Sheng Zhao, Ruihua Song, Tao Qin, Tie-Yan Liu</i>	
Content-Dependent Fine-Grained Speaker Embedding for Zero-Shot Speaker Adaptation in Text-To-Speech Synthesis.....	2573
<i>Yixuan Zhou, Changhe Song, Xiang Li, Luwen Zhang, Zhiyong Wu, Yanyao Bian, Dan Su, Helen Meng</i>	
Streamable Speech Representation Disentanglement and Multi-Level Prosody Modeling for Live One-Shot Voice Conversion	2578
<i>Haoquan Yang, Liqun Deng, Yu Ting Yeung, Nianzu Zheng, Yong Xu</i>	
Accent Conversion Using Pre-Trained Model and Synthesized Data from Voice Conversion	2583
<i>Tuan Nam Nguyen, Ngoc-Quan Pham, Alexander Waibel</i>	
VoiceMe: Personalized Voice Generation in TTS.....	2588
<i>Pol Van Rijn, Silvan Mertes, Dominik Schiller, Piotr Dura, Hubert Siuzdak, Peter M. C. Harrison, Elisabeth André, Nori Jacoby</i>	

DeID-VC: Speaker De-Identification Via Zero-Shot Pseudo Voice Conversion.....	2593
<i>Ruibin Yuan, Yuxuan Wu, Jacob Li, Jaxter Kim</i>	
Towards Improved Zero-Shot Voice Conversion with Conditional DSVAE	2598
<i>Jiachen Lian, Chunlei Zhang, Gopala Krishna Anumanchipalli, Dong Yu</i>	
Disentanglement of Emotional Style and Speaker Identity for Expressive Voice Conversion.....	2603
<i>Zongyang Du, Berrak Sisman, Kun Zhou, Haizhou Li</i>	

NOVEL MODELS AND TRAINING METHODS FOR ASR III

Internal Language Model Adaptation with Text-Only Data for End-To-End Speech Recognition	2608
<i>Zhong Meng, Yashesh Gaur, Naoyuki Kanda, Jinyu Li, Xie Chen, Yu Wu, Yifan Gong</i>	
A Complementary Joint Training Approach Using Unpaired Speech and Text a Complementary Joint Training Approach Using Unpaired Speech and Text.....	2613
<i>Yeqian Du, Jie Zhang, Qiu-Shi Zhu, Lirong Dai, Minghui Wu, Xin Fang, Zhouwang Yang</i>	
Knowledge Transfer and Distillation from Autoregressive to Non-Autoregressive Speech Recognition	2618
<i>Xun Gong, Zhikai Zhou, Yanmin Qian</i>	
Confidence Score Based Conformer Speaker Adaptation for Speech Recognition.....	2623
<i>Jiajun Deng, Xurong Xie, Tianzi Wang, Mingyu Cui, Boyang Xue, Zengrui Jin, Mengzhe Geng, Guinan Li, Xunying Liu, Helen Meng</i>	
Decoupled Federated Learning for ASR with Non-IID Data	2628
<i>Han Zhu, Jindong Wang, Gaofeng Cheng, Pengyuan Zhang, Yonghong Yan</i>	
Knowledge Distillation for CTC-Based Speech Recognition Via Consistent Acoustic Representation Learning.....	2633
<i>Sanli Tian, Keqi Deng, Zehan Li, Lingxuan Ye, Gaofeng Cheng, Ta Li, Yonghong Yan</i>	
Improving Generalization of Deep Neural Network Acoustic Models with Length Perturbation and N-Best Based Label Smoothing	2638
<i>Xiaodong Cui, George Saon, Tohru Nagano, Masayuki Suzuki, Takashi Fukuda, Brian Kingsbury, Gakuto Kurata</i>	
Supervision-Guided Codebooks for Masked Prediction in Speech Pre-Training.....	2643
<i>Chengyi Wang, Yiming Wang, Yu Wu, Sanyuan Chen, Jinyu Li, Shujie Liu, Furu Wei</i>	
Speech Pre-Training with Acoustic Piece.....	2648
<i>Shuo Ren, Shujie Liu, Yu Wu, Long Zhou, Furu Wei</i>	
Censer: Curriculum Semi-Supervised Learning for Speech Recognition Based on Self-Supervised Pre-Training.....	2653
<i>Bowen Zhang, Songjun Cao, Xiaoming Xiang, Yike Zhang, Long Ma, Takahiro Shinozaki</i>	
Pre-Training Transformer Decoder for End-To-End ASR Model with Unpaired Speech Data	2658
<i>Junyi Ao, Ziqiang Zhang, Long Zhou, Shujie Liu, Haizhou Li, Tom Ko, Lirong Dai, Jinyu Li, Yao Qian, Furu Wei</i>	
PISA: PoIncaré Saliency-Aware Interpolative Augmentation	2663
<i>Ramit Sawhney, Megh Thakkar, Vishwa Shah, Puneet Mathur, Vasu Sharma, Dinesh Manocha</i>	

Online Continual Learning of End-To-End Speech Recognition Models	2668
<i>Muqiao Yang, Ian Lane, Shinji Watanabe</i>	
Streaming Target-Speaker ASR with Neural Transducer	2673
<i>Takafumi Moriya, Hiroshi Sato, Tsubasa Ochiai, Marc Delcroix, Takahiro Shinozaki</i>	
SPLICEOUT: A Simple and Efficient Audio Augmentation Method	2678
<i>Arjit Jain, Pranay Reddy Samala, Deepak Mittal, Preethi Jyothi, Maneesh Singh</i>	

SPOKEN LANGUAGE MODELING AND UNDERSTANDING

Tokenwise Contrastive Pretraining for Finer Speech-To-BERT Alignment in End-To-End Speech-To-Intent Systems	2683
<i>Vishal Sunder, Eric Fosler-Lussier, Samuel Thomas, Hong-Kwang Kuo, Brian Kingsbury</i>	
Japanese ASR-Robust Pre-Trained Language Model with Pseudo-Error Sentences Generated by Grapheme-Phoneme Conversion	2688
<i>Yasuhito Ohsugi, Itsumi Saito, Kyosuke Nishida, Sen Yoshida</i>	
Improving Spoken Language Understanding with Cross-Modal Contrastive Learning	2693
<i>Jingjing Dong, Jiayi Fu, Peng Zhou, Hao Li, Xiaorui Wang</i>	
Low-Bit Shift Network for End-To-End Spoken Language Understanding	2698
<i>Anderson R. Avila, Khalil Bibi, Rui Heng Yang, Xinlin Li, Chao Xing, Xiao Chen</i>	
Meta Auxiliary Learning for Low-Resource Spoken Language Understanding	2703
<i>Yingying Gao, Junlan Feng, Chao Deng, Shilei Zhang</i>	
Adversarial Knowledge Distillation for Robust Spoken Language Understanding	2708
<i>Ye Wang, Baishun Ling, Yanmeng Wang, Junhao Xue, Shaojun Wang, Jing Xiao</i>	
Incorporating Dual-Aware with Hierarchical Interactive Memory Networks for Task-Oriented Dialogue	2713
<i>Yangyang Ou, Peng Zhang, Jing Zhang, Hui Gao, Xing Ma</i>	
Pay More Attention to History: A Context Modeling Strategy for Conversational Text-To-SQL	2718
<i>Yuntao Li, Hanchu Zhang, Yutian Li, Sirui Wang, Wei Wu, Yan Zhang</i>	
Small Changes Make Big Differences: Improving Multi-Turn Response Selection in Dialogue Systems Via Fine-Grained Contrastive Learning	2723
<i>Yuntao Li, Can Xu, Huang Hu, Lei Sha, Yan Zhang, Daxin Jiang</i>	
Toward Low-Cost End-To-End Spoken Language Understanding	2728
<i>Marco Dinarelli, Marco Naguib, François Portet</i>	
A Multi-Task BERT Model for Schema-Guided Dialogue State Tracking	2733
<i>Eleftherios Kapelonis, Efthymios Georgiou, Alexandros Potamianos</i>	
WavPrompt: Towards Few-Shot Spoken Language Understanding with Frozen Language Models	2738
<i>Heting Gao, Junrui Ni, Kaizhi Qian, Yang Zhang, Shiyu Chang, Mark Hasegawa-Johnson</i>	
Analysis of Praising Skills Focusing on Utterance Contents	2743
<i>Asahi Ogushi, Toshiki Onishi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, Akihiro Miyata</i>	

Speech2Slot: A Limited Generation Framework with Boundary Detection for Slot Filling from Speech	2748
<i>Pengwei Wang, Yinpei Su, Xiaohuan Zhou, Xin Ye, Liangchen Wei, Ming Liu, Yuan You, Feijun Jiang</i>	

ACOUSTIC SIGNAL REPRESENTATION AND ANALYSIS I

Efficient Training of Audio Transformers with Patchout	2753
<i>Khaled Koutini, Jan Schlüter, Hamid Eghbal-Zadeh, Gerhard Widmer</i>	
CNN-Based Audio Event Recognition for Automated Violence Classification and Rating for Prime Video Content.....	2758
<i>Mayank Sharma, Tarun Gupta, Kenny Qiu, Xiang Hao, Raffay Hamid</i>	
Frequency Dynamic Convolution: Frequency-Adaptive Pattern Recognition for Sound Event Detection	2763
<i>Hyeonuk Nam, Seong-Hu Kim, Byeong-Yun Ko, Yong-Hwa Park</i>	
On Breathing Pattern Information in Synthetic Speech.....	2768
<i>Zohreh Mostaani, Mathew Magimai Doss</i>	
Interactive Audio-Text Representation for Automated Audio Captioning with Contrastive Learning	2773
<i>Chen Chen, Nana Hou, Yuchen Hu, Heqing Zou, Xiaofeng Qi, Eng Siong Chng</i>	
Deformable CNN and Imbalance-Aware Feature Learning for Singing Technique Classification	2778
<i>Yuya Yamamoto, Juhan Nam, Hiroko Terasawa</i>	

PRIVACY AND SECURITY IN SPEECH COMMUNICATION

Does Audio Deepfake Detection Generalize?	2783
<i>Nicolas Müller, Pavel Czempin, Franziska Diekmann, Adam Froghyar, Konstantin Böttinger</i>	
Attacker Attribution of Audio Deepfakes	2788
<i>Nicolas Müller, Franziska Diekmann, Jennifer Williams</i>	
Are Disentangled Representations All You Need to Build Speaker Anonymization Systems?	2793
<i>Champion Pierre, Anthony Larcher, Denis Jouvét</i>	
Towards End-To-End Private Automatic Speaker Recognition.....	2798
<i>Francisco Teixeira, Alberto Abad, Bhiksha Raj, Isabel Trancoso</i>	
Extracting Targeted Training Data from ASR Models, and How to Mitigate it	2803
<i>Ehsan Amid, Om Dipakbhai Thakkar, Arun Narayanan, Rajiv Mathews, Françoise Beaufays</i>	
Detecting Unintended Memorization in Language-Model-Fused ASR.....	2808
<i>W. Ronny Huang, Steve Chien, Om Dipakbhai Thakkar, Rajiv Mathews</i>	

MULTIMODAL SYSTEMS

Transformer-Based Automatic Speech Recognition with Auxiliary Input of Source Language Text Toward Transcribing Simultaneous Interpretation.....	2813
<i>Shuta Taniguchi, Tsuneo Kato, Akihiro Tamura, Keiji Yasuda</i>	

AVATAR: Unconstrained Audiovisual Speech Recognition.....	2818
<i>Valentin Gabeur, Paul Hongsuck Seo, Arsha Nagrani, Chen Sun, Karteek Alahari, Cordelia Schmid</i>	
Word Discovery in Visually Grounded, Self-Supervised Speech Models	2823
<i>Puyuan Peng, David Harwath</i>	
End-To-End Multi-Talker Audio-Visual ASR Using an Active Speaker Attention Module	2828
<i>Richard Rose, Olivier Siohan</i>	
Transformer-Based Video Front-Ends for Audio-Visual Speech Recognition for Single and Multi-Person Video.....	2833
<i>Dmitriy Serdyuk, Otavio Braga, Olivier Siohan</i>	
Visual Context-Driven Audio Feature Enhancement for Robust End-To-End Audio-Visual Speech Recognition	2838
<i>Joanna Hong, Minsu Kim, Daehun Yoo, Yong Man Ro</i>	

ATYPICAL SPEECH DETECTION

Frame-Level Stutter Detection	2843
<i>John Harvill, Mark Hasegawa-Johnson, Chang D. Yoo</i>	
Detecting Heart Failure Through Voice Analysis Using Self-Supervised Mode-Based Memory Fusion.....	2848
<i>Darshana Priyasad, Andi Partovi, Sridha Sridharan, Maryam Kashefpoor, Tharindu Fernando, Simon Denman, Clinton Fookes, Jia Tang, David Kaye</i>	
Automatic Detection of Speech Sound Disorder in Child Speech Using Posterior-Based Speaker Representations	2853
<i>Si-Ioi Ng, Cymie Wing-Yee Ng, Jiarui Wang, Tan Lee</i>	
Data Augmentation for Dementia Detection in Spoken Language.....	2858
<i>Dominika Woszczyk, Anna Hedlikova, Alican Akman, Soteris Demetriou, Björn Schuller</i>	
Interpretable Acoustic Representation Learning on Breathing and Speech Signals for COVID-19 Detection	2863
<i>Debottam Dutta, Debarpan Bhattacharya, Sriram Ganapathy, Amir Hossein Poorjam, Deepak Mittal, Maneesh Singh</i>	
Detecting Dysfluencies in Stuttering Therapy Using Wav2vec 2.0.....	2868
<i>Sebastian Peter Bayerl, Dominik Wagner, Elmar Noeth, Korbinian Riedhammer</i>	

SPOOFING-AWARE AUTOMATIC SPEAKER VERIFICATION (SASV) I

HYU Submission for the SASV Challenge 2022: Reforming Speaker Embeddings with Spoofing-Aware Conditioning.....	2873
<i>Jeong-Hwan Choi, Joon-Young Yang, Ye-Rin Jeoung, Joon-Hyuk Chang</i>	
Two Methods for Spoofing-Aware Speaker Verification: Multi-Layer Perceptron Score Fusion Model and Integrated Embedding Projector.....	2878
<i>Jungwoo Heo, Ju-Ho Kim, Hyun-Seo Shin</i>	

Spoofing-Aware Attention Based ASV Back-End with Multiple Enrollment Utterances and a Sampling Strategy for the SASV Challenge 2022.....	2883
<i>Chang Zeng, Lin Zhang, Meng Liu, Junichi Yamagishi</i>	
A Subnetwork Approach for Spoofing Aware Speaker Verification.....	2888
<i>Alexander Alenin, Nikita Torgashov, Anton Okhotnikov, Rostislav Makarov, Ivan Yakovlev</i>	
SASV 2022: The First Spoofing-Aware Speaker Verification Challenge.....	2893
<i>Jee-Weon Jung, Hemlata Tak, Hye-Jin Shim, Hee-Soo Heo, Bong-Jin Lee, Soo-Whan Chung, Ha-Jin Yu, Nicholas Evans, Tomi Kinnunen</i>	
Representation Selective Self-Distillation and Wav2vec 2.0 Feature Exploration for Spoof-Aware Speaker Verification	2898
<i>Jin Woo Lee, Eungbeom Kim, Junghyun Koo, Kyogu Lee</i>	

SINGLE-CHANNEL AND MULTI-CHANNEL SPEECH ENHANCEMENT

TPLCnet: Real-Time Deep Packet Loss Concealment in the Time Domain Using a Short Temporal Context	2903
<i>Nils L. Westhausen, Bernd T. Meyer</i>	
On the Role of Spatial, Spectral, and Temporal Processing for DNN-Based Non-Linear Multi-Channel Speech Enhancement.....	2908
<i>Kristina Tesch, Nils-Hendrik Mohrmann, Timo Gerkmann</i>	
DDS: A New Device-Degraded Speech Dataset for Speech Enhancement.....	2913
<i>Haoyu Li, Junichi Yamagishi</i>	
Direction-Aware Joint Adaptation of Neural Speech Enhancement and Recognition in Real Multiparty Conversational Environments.....	2918
<i>Yicheng Du, Aditya Arie Nugraha, Kouhei Sekiguchi, Yoshiaki Bando, Mathieu Fontaine, Kazuyoshi Yoshii</i>	
Refining DNN-Based Mask Estimation Using CGMM-Based EM Algorithm for Multi-Channel Noise Reduction	2923
<i>Julitta Bartolewska, Stanislaw Kacprzak, Konrad Kowalczyk</i>	
Speech Enhancement with Score-Based Generative Models in the Complex STFT Domain	2928
<i>Simon Welker, Julius Richter, Timo Gerkmann</i>	
Enhancing Embeddings for Speech Classification in Noisy Conditions	2933
<i>Mohamed Nabih Ali, Alessio Brutti, Falavigna Daniele</i>	
Deep Audio Waveform Prior	2938
<i>Arnon Turetzky, Tzvi Michelson, Yossi Adi, Shmuel Peleg</i>	
Convolutional Weighted Multichannel Wiener Filter Front-End for Distant Automatic Speech Recognition in Reverberant Multispeaker Scenarios.....	2943
<i>Mieszko Frasz, Marcin Witkowski, Konrad Kowalczyk</i>	
Efficient Transformer-Based Speech Enhancement Using Long Frames and STFT Magnitudes	2948
<i>Danilo De Oliveira, Tal Peer, Timo Gerkmann</i>	
Improving Speech Enhancement Through Fine-Grained Speech Characteristics	2953
<i>Muqiao Yang, Joseph Konan, David Bick, Anurag Kumar, Shinji Watanabe, Bhiksha Raj</i>	

VOICE CONVERSION AND ADAPTATION II

Creating New Voices Using Normalizing Flows	2958
<i>Piotr Bilinski, Thomas Merritt, Abdelhamid Ezzerg, Kamil Pokora, Sebastian Cygert, Kayoko Yanagisawa, Roberto Barra-Chicote, Daniel Korzekwa</i>	
Unify and Conquer: How Phonetic Feature Representation Affects Polyglot Text-To-Speech (TTS)	2963
<i>Ariadna Sanchez, Alessio Falai, Ziyao Zhang, Orazio Angelini, Kayoko Yanagisawa</i>	
Human-In-The-Loop Speaker Adaptation for DNN-Based Multi-Speaker TTS	2968
<i>Kenta Udagawa, Yuki Saito, Hiroshi Saruwatari</i>	
GlowVC: Mel-Spectrogram Space Disentangling Model for Language-Independent Text-Free Voice Conversion.....	2973
<i>Magdalena Proszewska, Grzegorz Beringer, Daniel Sáez-Trigueros, Thomas Merritt, Abdelhamid Ezzerg, Roberto Barra-Chicote</i>	
One-Shot Speaker Adaptation Based on Initialization by Generative Adversarial Networks for TTS.....	2978
<i>Jaeuk Lee, Joon-Hyuk Chang</i>	
Zero-Shot Voice Conditioning for Denoising Diffusion TTS Models	2983
<i>Alon Levkovitch, Eliya Nachmani, Lior Wolf</i>	
Advanced Speaker Embedding with Predictive Variance of Gaussian Distribution for Speaker Adaptation in TTS	2988
<i>Jaeuk Lee, Joon-Hyuk Chang</i>	
Karaoker: Alignment-Free Singing Voice Synthesis with Speech Training Data	2993
<i>Panagiotis Kakoulidis, Nikolaos Ellinas, Georgios Vamvoukakis, Konstantinos Markopoulos, June Sig Sung, Gunu Jho, Pirros Tsiakoulis, Aimilios Chalamandaris</i>	
ACNN-VC: Utilizing Adaptive Convolution Neural Network for One-Shot Voice Conversion.....	2998
<i>Ji Sub Um, Yeunju Choi, Hoi Rin Kim</i>	
A Unified System for Voice Cloning and Voice Conversion Through Diffusion Probabilistic Modeling	3003
<i>Tasnima Sadekova, Vladimir Gogoryan, Ivan Vovk, Vadim Popov, Mikhail Kudinov, Jiansheng Wei</i>	
Adversarial Multi-Task Learning for Disentangling Timbre and Pitch in Singing Voice Synthesis.....	3008
<i>Tae-Woo Kim, Min-Su Kang, Gyeong-Hoon Lee</i>	
Leveraging Symmetrical Convolutional Transformer Networks for Speech to Singing Voice Style Transfer	3013
<i>Shrutina Agarwal, Naoya Takahashi, Sriram Ganapathy</i>	
Cross-Speaker Emotion Transfer for Low-Resource Text-To-Speech Using Non-Parallel Voice Conversion with Pitch-Shift Data Augmentation	3018
<i>Ryo Terashima, Ryuichi Yamamoto, Eunwoo Song, Yuma Shirahata, Hyun-Wook Yoon, Jae-Min Kim, Kentaro Tachibana</i>	

RESOURCE-CONSTRAINED ASR

Deep Residual Spiking Neural Network for Keyword Spotting in Low-Resource Settings	3023
<i>Qu Yang, Qi Liu, Haizhou Li</i>	

Reducing Domain Mismatch in Self-Supervised Speech Pre-Training	3028
<i>Murali Karthick Baskar, Andrew Rosenberg, Bhuvana Ramabhadran, Yu Zhang, Nicolás Serrano</i>	
Sub-8-Bit Quantization Aware Training for 8-Bit Neural Network Accelerator with On-Device Speech Recognition	3033
<i>Kai Zhen, Hieu Duy Nguyen, Raviteja Chinta, Nathan Susanj, Athanasios Mouchtaris, Tariq Afzal, Ariya Rastrow</i>	
W2V2-Light: A Lightweight Version of Wav2vec 2.0 for Automatic Speech Recognition.....	3038
<i>Dong-Hyun Kim, Jae-Hong Lee, Ji-Hwan Mo, Joon-Hyuk Chang</i>	
Compute Cost Amortized Transformer for Streaming ASR	3043
<i>Yi Xie, Jonathan J. Macoskey, Martin Radfar, Feng-Ju Chang, Brian King, Ariya Rastrow, Athanasios Mouchtaris, Grant Strimel</i>	
On-Demand Compute Reduction with Stochastic Wav2vec 2.0	3048
<i>Apoorv Vyas, Wei-Ning Hsu, Michael Auli, Alexei Baevski</i>	
Transfer Learning from Multi-Lingual Speech Translation Benefits Low-Resource Speech Recognition	3053
<i>Geoffroy Vanderreydt, François Remy, Kris Demuynck</i>	
FeaRLESS: Feature Refinement Loss for Ensembling Self-Supervised Learning Features in Robust End-To-End Speech Recognition	3058
<i>Szu-Jui Chen, Jiamin Xie, John H. L. Hansen</i>	

SPEECH PRODUCTION, PERCEPTION AND MULTIMODALITY

Perceptual Evaluation of Penetrating Voices Through a Semantic Differential Method	3063
<i>Tatsuya Kitamura, Naoki Kunimoto, Hideki Kawahara, Shigeaki Amano</i>	
Non-Native Perception of Japanese Singleton/Geminate Contrasts: Comparison of Mandarin and Mongolian Speakers Differing in Japanese Experience	3068
<i>Kimiko Tsukada, Yurong Yurong</i>	
Evaluating the Effects of Modified Speech on Perceptual Speaker Identification Performance	3073
<i>Benjamin O'Brien, Christine Meunier, Alain Ghio</i>	
Mandarin Lombard Grid: A Lombard-Grid-Like Corpus of Standard Chinese.....	3078
<i>Yuhong Yang, Xufeng Chen, Qingmu Liu, Weiping Tu, Hongyang Chen, Linjun Cai</i>	
Syllable Sequence of /a+/ta/ Can Be Heard as /atta/ in Japanese with Visual Or Tactile Cues	3083
<i>Takayuki Arai, Miho Yamada, Megumi Okusawa</i>	
InQSS: A Speech Intelligibility and Quality Assessment Model Using a Multi-Task Learning Network.....	3088
<i>Yu-Wen Chen, Yu Tsao</i>	
Investigating the Influence of Personality on Acoustic-Prosodic Entrainment	3093
<i>Andreas Weise, Rivka Levitan</i>	
Common and Differential Acoustic Representation of Interpersonal and Tactile Iconic Perception of Mandarin Vowels.....	3098
<i>Yi Li, Xiaoming Jiang</i>	

Effects of Noise on Speech Perception and Spoken Word Comprehension	3103
<i>Jovan Eranovic, Daniel Pape, Magda Stroinska, Elisabet Service, Marijana Matkovski</i>	
Acquisition of Two Consecutive Neutral Tones in Mandarin-Speaking Preschoolers: Phonological Representation and Phonetic Realization	3108
<i>Sichen Zhang, Aijun Li</i>	
Air Tissue Boundary Segmentation Using Regional Loss in Real-Time Magnetic Resonance Imaging Video for Speech Production.....	3113
<i>Anwasha Roy, Varun Belagali, Prasanta Ghosh</i>	

VOLUME 5

Language-Specific Interactions of Vowel Discrimination in Noise.....	3118
<i>Mark Gibson, Marcel Schlechtweg, Beatriz Blecua Falgueras, Judit Ayala Alcalde</i>	
An Improved Transformer Transducer Architecture for Hindi-English Code Switched Speech Recognition	3123
<i>Ansen Antony, Sumanth Reddy Kota, Akhilesh Lade, Spoorthy V, Shashidhar G. Koolagudi</i>	
VocaLiST: An Audio-Visual Synchronisation Model for Lips and Voices	3128
<i>Venkatesh Shenoy Kadandale, Juan F. Montesinos, Gloria Haro</i>	

MULTI-, CROSS-LINGUAL AND OTHER TOPICS IN ASR II

Cross-Lingual Transfer Learning Approach to Phoneme Error Detection Via Latent Phonetic Representation	3133
<i>Jovan M. Dalhouse, Katunobu Itou</i>	
Global RNN Transducer Models for Multi-Dialect Speech Recognition	3138
<i>Takashi Fukuda, Samuel Thomas, Masayuki Suzuki, Gakuto Kurata, George Saon, Brian Kingsbury</i>	
Acoustic Stress Detection in Isolated English Words for Computer-Assisted Pronunciation Training	3143
<i>Vera Bernhard, Sandra Schwab, Jean-Philippe Goldman</i>	
On-The-Fly ASR Corrections with Audio Exemplars	3148
<i>Golan Pundak, Tsendsuren Munkhdalai, Khe Chai Sim</i>	
FFM: A Frame Filtering Mechanism to Accelerate Inference Speed for Conformer in Speech Recognition	3153
<i>Zongfeng Quan, Nick J. C. Wang, Wei Chu, Tao Wei, Shaojun Wang, Jing Xiao</i>	
Two-Pass Decoding and Cross-Adaptation Based System Combination of End-To-End Conformer and Hybrid TDNN ASR Systems	3158
<i>Mingyu Cui, Jiajun Deng, Shoukang Hu, Xurong Xie, Tianzi Wang, Shujie Hu, Mengzhe Geng, Boyang Xue, Xunying Liu, Helen Meng</i>	
Improving Recognition of Out-Of-Vocabulary Words in E2E Code-Switching ASR by Fusing Speech Generation Methods	3163
<i>Lingxuan Ye, Gaofeng Cheng, Runyan Yang, Zehui Yang, Sanli Tian, Pengyuan Zhang, Yonghong Yan</i>	
Mitigating Bias Against Non-Native Accents.....	3168
<i>Yuanyuan Zhang, Yixuan Zhang, Bence Halpern, Tanvina Patel, Odette Scharenborg</i>	

A Multi-Level Acoustic Feature Extraction Framework for Transformer Based End-To-End Speech Recognition	3173
<i>Jin Li, Rongfeng Su, Xurong Xie, Lan Wang, Nan Yan</i>	
LAE: Language-Aware Encoder for Monolingual and Multilingual ASR	3178
<i>Jinchuan Tian, Jianwei Yu, Chunlei Zhang, Yuexian Zou, Dong Yu</i>	
Significance of Single Frequency Filter for the Development of Children’s KWS System	3183
<i>Biswaranjan Pattanayak, Gayadhar Pradhan</i>	
A Language Agnostic Multilingual Streaming On-Device ASR System.....	3188
<i>Bo Li, Tara Sainath, Ruoming Pang, Shuo-Yiin Chang, Qiumin Xu, Trevor Strohman, Vince Chen, Qiao Liang, Heguang Liu, Yanzhang He, Parisa Haghani, Sameer Bidichandani</i>	
Minimizing Sequential Confusion Error in Speech Command Recognition.....	3193
<i>Zhanheng Yang, Hang Lv, Xiong Wang, Ao Zhang, Lei Xie</i>	
Homophone Disambiguation Profits from Durational Information.....	3198
<i>Barbara Schuppler, Emil Berger, Xenia Kogler, Franz Pernkopf</i>	
Speaker-Specific Utterance Ensemble Based Transfer Attack on Speaker Identification	3203
<i>Chu-Xiao Zuo, Jia-Yi Leng, Wu-Jun Li</i>	
Complex Frequency Domain Linear Prediction: A Tool to Compute Modulation Spectrum of Speech	3208
<i>Samik Sadhu, Hynek Hermansky</i>	
Spectral Modification Based Data Augmentation for Improving End-To-End ASR for Children’s Speech	3213
<i>Vishwanath Pratap Singh, Hardik Sailor, Supratik Bhattacharya, Abhishek Pandey</i>	
End-To-End Joint Modeling of Conversation History-Dependent and Independent ASR Systems with Multi-History Training	3218
<i>Ryo Masumura, Yoshihiro Yamazaki, Saki Mizuno, Naoki Makishima, Mana Ihori, Mihiro Uchida, Hiroshi Sato, Tomohiro Tanaka, Akihiko Takashima, Satoshi Suzuki, Shota Orihashi, Takafumi Moriya, Nobukatsu Hojo, Atsushi Ando</i>	
Streaming End-To-End Multilingual Speech Recognition with Joint Language Identification.....	3223
<i>Chao Zhang, Bo Li, Tara Sainath, Trevor Strohman, Sepand Mavandadi, Shuo-Yiin Chang, Parisa Haghani</i>	

SPOKEN LANGUAGE PROCESSING III

An Anchor-Free Detector for Continuous Speech Keyword Spotting.....	3228
<i>Zhiyuan Zhao, Chuanxin Tang, Chengdong Yao, Chong Luo</i>	
Low-Complex and Highly-Performed Binary Residual Neural Network for Small-Footprint Keyword Spotting.....	3233
<i>Xiao Wang, Song Cheng, Jun Li, Shushan Qiao, Yumei Zhou, Yi Zhan</i>	
UniKW-AT: Unified Keyword Spotting and Audio Tagging.....	3238
<i>Heinrich Dinkel, Yongqing Wang, Zhiyong Yan, Junbo Zhang, Yujun Wang</i>	
ESSumm: Extractive Speech Summarization from Untranscribed Meeting	3243
<i>Jun Wang</i>	

XTREME-S: Evaluating Cross-Lingual Speech Representations	3248
<i>Alexis Conneau, Ankur Bapna, Yu Zhang, Min Ma, Patrick Von Platen, Anton Lozhkov, Colin Cherry, Ye Jia, Clara Rivera, Mihir Kale, Daan Van Esch, Vera Axelrod, Simran Khanuja, Jonathan Clark, Orhan Firat, Michael Auli, Sebastian Ruder, Jason Riesa, Melvin Johnson</i>	
Negative Guided Abstractive Dialogue Summarization	3253
<i>Junpeng Liu, Yanyan Zou, Yuxuan Xi, Shengjie Li, Mian Ma, Zhuoye Ding, Bo Long</i>	
Exploring Representation Learning for Small-Footprint Keyword Spotting.....	3258
<i>Fan Cui, Liyong Guo, Quandong Wang, Peng Gao, Yujun Wang</i>	
Large-Scale Streaming End-To-End Speech Translation with Neural Transducers.....	3263
<i>Jian Xue, Peidong Wang, Jinyu Li, Matt Post, Yashesh Gaur</i>	
Phonetic Embedding for ASR Robustness in Entity Resolution	3268
<i>Xiaozhou Zhou, Ruying Bao, William M. Campbell</i>	
Hierarchical Tagger with Multi-Task Learning for Cross-Domain Slot Filling.....	3273
<i>Xiao Wei, Yuke Si, Shiquan Wang, Longbiao Wang, Jianwu Dang</i>	
Multi-Class AUC Optimization for Robust Small-Footprint Keyword Spotting with Limited Training Data.....	3278
<i>Menglong Xu, Shengqiang Li, Chengdong Liang, Xiao-Lei Zhang</i>	
Weak Supervision for Question Type Detection with Large Language Models	3283
<i>Jiri Martinek, Christophe Cerisara, Pavel Kral, Ladislav Lenc, Josef Baloun</i>	

NON-INTRUSIVE OBJECTIVE SPEECH QUALITY ASSESSMENT (NISQA) CHALLENGE FOR ONLINE CONFERENCING APPLICATIONS

BIT-MI Deep Learning-Based Model to Non-Intrusive Speech Quality Assessment Challenge in Online Conferencing Applications	3288
<i>Miao Liu, Jing Wang, Liang Xu, Jianqian Zhang, Shicong Li, Fei Xiang</i>	
MOS Prediction Network for Non-Intrusive Speech Quality Assessment in Online Conferencing.....	3293
<i>Wenjing Liu, Chuan Xie</i>	
Non-Intrusive Speech Quality Assessment with a Multi-Task Learning Based Subband Adaptive Attention Temporal Convolutional Neural Network	3298
<i>Xiaofeng Shu, Yanjie Chen, Chuxiang Shang, Yan Zhao, Chengshuai Zhao, Yehang Zhu, Chuanzeng Huang, Yuxuan Wang</i>	
Soft-Label Learn for No-Intrusive Speech Quality Assessment.....	3303
<i>Junyong Hao, Shunzhou Ye, Cheng Lu, Fei Dong, Jingang Liu, Dong Pi</i>	
ConferencingSpeech 2022 Challenge: Non-Intrusive Objective Speech Quality Assessment (NISQA) Challenge for Online Conferencing Applications.....	3308
<i>Gaoxiong Yi, Wei Xiao, Yiming Xiao, Babak Naderi, Sebastian Möller, Wafaa Wardah, Gabriel Mittag, Ross Culter, Zhuohuang Zhang, Donald S. Williamson, Fei Chen, Fuzheng Yang, Shidong Shang</i>	
MOSRA: Joint Mean Opinion Score and Room Acoustics Speech Quality Assessment	3313
<i>Karl El Hajal, Milos Cernak, Pablo Mainar</i>	

CCATMos: Convolutional Context-Aware Transformer Network for Non-Intrusive Speech Quality Assessment	3318
<i>Yuchen Liu, Li-Chia Yang, Alexander Pawlicki, Marko Stamenovic</i>	

Impairment Representation Learning for Speech Quality Assessment.....	3323
<i>Lianwu Chen, Xinlei Ren, Xu Zhang, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu</i>	

SPEECH AND LANGUAGE IN HEALTH: FROM REMOTE MONITORING TO MEDICAL CONVERSATIONS III

Exploring Linguistic Feature and Model Combination for Speech Recognition Based Automatic AD Detection.....	3328
<i>Yi Wang, Tianzi Wang, Zi Ye, Lingwei Meng, Shoukang Hu, Xixin Wu, Xunying Liu, Helen Meng</i>	

ECAPA-TDNN Based Depression Detection from Clinical Speech	3333
<i>Dong Wang, Yanhui Ding, Qing Zhao, Peilin Yang, Shuping Tan, Ya Li</i>	

A Step Towards Preserving Speakers' Identity While Detecting Depression Via Speaker Disentanglement.....	3338
<i>Vijay Ravi, Jinhan Wang, Jonathan Flint, Abeer Alwan</i>	

Toward Corpus Size Requirements for Training and Evaluating Depression Risk Models Using Spoken Language	3343
<i>Tomasz Rutowski, Amir Harati, Elizabeth Shriberg, Yang Lu, Piotr Chlebek, Ricardo Oliveira</i>	

Deep Learning Approaches for Detecting Alzheimer's Dementia from Conversational Speech of ILSE Study	3348
<i>Ayimmisagul Ablimit, Karen Scholz, Tanja Schultz</i>	

Multimodal Depression Severity Score Prediction Using Articulatory Coordination Features and Hierarchical Attention Based Text Embeddings.....	3353
<i>Nadee Seneviratne, Carol Espy-Wilson</i>	

ASR Error Detection Via Audio-Transcript Entailment	3358
<i>Nimshi Venkat Meripo, Sandeep Konam</i>	

SPEECH SYNTHESIS: PROSODY MODELING

CopyCat2: A Single Model for Multi-Speaker TTS and Many-To-Many Fine-Grained Prosody Transfer	3363
<i>Sri Karlapati, Penny Karanasou, Mateusz Lajszczak, Syed Ammar Abbas, Alexis Moinet, Peter Makarov, Ray Li, Arent Van Korlaar, Simon Slangen, Thomas Drugman</i>	

Simple and Effective Multi-Sentence TTS with Expressive and Coherent Prosody	3368
<i>Peter Makarov, Syed Ammar Abbas, Mateusz Lajszczak, Arnaud Joly, Sri Karlapati, Alexis Moinet, Thomas Drugman, Penny Karanasou</i>	

Acoustic Modeling for End-To-End Empathetic Dialogue Speech Synthesis Using Linguistic and Prosodic Contexts of Dialogue History	3373
<i>Yuto Nishimura, Yuki Saito, Shinnosuke Takamichi, Kentaro Tachibana, Hiroshi Saruwatari</i>	

Emphasis Control for Parallel Neural TTS.....	3378
<i>Shreyas Seshadri, Tuomo Raitio, Dan Castellani, Jiangchuan Li</i>	

BERT, Can HE Predict Contrastive Focus? Predicting and Controlling Prominence in Neural TTS
Using a Language Model..... 3383
Brooke Stephenson, Laurent Besacier, Laurent Girin, Thomas Hueber

Combining Conversational Speech with Read Speech to Improve Prosody in Text-To-Speech
Synthesis..... 3388
Johannah O'Mahony, Catherine Lai, Simon King

SELF-SUPERVISED, SEMI-SUPERVISED, ADAPTATION AND DATA AUGMENTATION FOR ASR

Unsupervised Data Selection Via Discrete Speech Representation for ASR..... 3393
Zhiyun Lu, Yongqiang Wang, Yu Zhang, Wei Han, Zhehuai Chen, Parisa Haghani

CTRL: Continual Representation Learning to Transfer Information of Pre-Trained for WAV2VEC
2.0..... 3398
*Jae-Hong Lee, Chae-Won Lee, Jin-Seong Choi, Joon-Hyuk Chang, Woo Kyeong Seong,
Jeonghan Lee*

Speaker Adaptation for Wav2vec2 Based Dysarthric ASR..... 3403
*Murali Karthick Baskar, Tim Herzig, Diana Nguyen, Mireia Diez, Tim Polzehl, Lukas Burget,
Jan Cernocký*

Non-Parallel Voice Conversion for ASR Augmentation..... 3408
*Gary Wang, Andrew Rosenberg, Bhuvana Ramabhadran, Fadi Biadisy, Jesse Emond, Yinghui
Huang, Pedro J. Moreno*

Improved Consistency Training for Semi-Supervised Sequence-To-Sequence ASR Via Speech
Chain Reconstruction and Self-Transcribing..... 3413
Heli Qi, Sashi Novitasari, Sakriani Sakti, Satoshi Nakamura

Joint Encoder-Decoder Self-Supervised Pre-Training for ASR..... 3418
A Arunkumar, Srinivasan Umesh

PHONETICS AND PHONOLOGY

An Overview of Discourse Clicks in Central Swedish..... 3423
Margaret Zellers

VOT and F0 Perturbations for the Realization of Voicing Contrast in Tohoku Japanese 3428
*Hiroto Noguchi, Sanae Matsui, Naoya Watabe, Chuyu Huang, Ayako Hashimoto, Ai
Mizoguchi, Mafuyu Kitahara*

Complex Sounds and Cross-Language Influence: The Case of Ejectives in Omani Mehri..... 3433
Rachid Ridouane, Philipp Buech

When Phonetics Meets Morphology: Intervocalic Voicing Within and Across Words in Romance
Languages..... 3438
Mathilde Hutin, Martine Adda-Decker, Lori Lamel, Ioana Vasilescu

The Mapping Between Syntactic and Prosodic Phrasing in English and Mandarin..... 3443
Jianjing Kuang, May Pik Yu Chan, Nari Rhee, Mark Liberman, Hongwei Ding

Pharyngealization in Amazigh: Acoustic and Articulatory Marking Over Time 3448
Philipp Buech, Rachid Ridouane, Anne Hermes

SPOKEN LANGUAGE UNDERSTANDING II

ASR-Generated Text for Language Model Pre-Training Applied to Speech Tasks.....	3453
<i>Valentin Pelloin, Franck Dary, Nicolas Hervé, Benoit Favre, Nathalie Camelin, Antoine Laurent, Laurent Besacier</i>	
Contrastive Learning for Improving ASR Robustness in Spoken Language Understanding	3458
<i>Ya-Hsin Chang, Yun-Nung Chen</i>	
Learning Under Label Noise for Robust Spoken Language Understanding Systems	3463
<i>Anoop Kumar, Pankaj Kumar Sharma, Aravind Illa, Sriram Venkatapathy, Subhrangshu Nandi, Pritam Varma, Anurag Dwarakanath, Aram Galstyan</i>	
Deliberation Model for On-Device Spoken Language Understanding.....	3468
<i>Duc Le, Akshat Shrivastava, Paden D. Tomasello, Suyoun Kim, Aleksandr Livshits, Ozlem Kalinli, Michael Seltzer</i>	
Intent Classification Using Pre-Trained Language Agnostic Embeddings for Low Resource Languages.....	3473
<i>Hemant Yadav, Akshat Gupta, Sai Krishna Rallabandi, Alan W Black, Rajiv Ratn Shah</i>	
Two-Pass Low Latency End-To-End Spoken Language Understanding.....	3478
<i>Siddhant Arora, Siddharth Dalmia, Xuankai Chang, Brian Yan, Alan W Black, Shinji Watanabe</i>	

SPEECH INTELLIGIBILITY PREDICTION FOR HEARING-IMPAIRED LISTENERS I

Non-Intrusive Speech Intelligibility Metric Prediction for Hearing Impaired Individuals.....	3483
<i>George Close, Samuel Hollands, Stefan Goetze, Thomas Hain</i>	
Exploiting Hidden Representations from a DNN-Based Speech Recogniser for Speech Intelligibility Prediction in Hearing-Impaired Listeners.....	3488
<i>Zehai Tu, Ning Ma, Jon Barker</i>	
Unsupervised Uncertainty Measures of Automatic Speech Recognition for Non-Intrusive Speech Intelligibility Prediction.....	3493
<i>Zehai Tu, Ning Ma, Jon Barker</i>	
Speech Intelligibility Prediction for Hearing-Impaired Listeners with the LEAP Model.....	3498
<i>Jana Roßbach, Rainer Huber, Saskia Röttges, Christopher F. Hauth, Thomas Biberger, Thomas Brand, Bernd T. Meyer, Jan Rennies</i>	
Predicting Speech Intelligibility Using the Spike Activity Mutual Information Index.....	3503
<i>Franklin Alvarez Cardinale, Waldo Nogueira</i>	
The 1st Clarity Prediction Challenge: A Machine Learning Challenge for Hearing Aid Intelligibility Prediction	3508
<i>Jon Barker, Michael Akeroyd, Trevor J. Cox, John F. Culling, Jennifer Firth, Simone Graetzer, Holly Griffiths, Lara Harris, Graham Naylor, Zuzanna Podwinska, Eszter Porter, Rhoddy Viveros Munoz</i>	

LOW-RESOURCE ASR DEVELOPMENT I

Voice Conversion Can Improve ASR in Very Low-Resource Settings.....	3513
<i>Matthew Baas, Herman Kamper</i>	
Data Augmentation for Low-Resource Quechua ASR Improvement.....	3518
<i>Rodolfo Zevallos, Nùria Bel, Guillermo Cámbara, Mireia Farrús, Jordi Luque</i>	
ScoutWav: Two-Step Fine-Tuning on Self-Supervised Automatic Speech Recognition for Low-Resource Environments.....	3523
<i>Kavan Fatehi, Mercedes Torres Torres, Ayse Kucukyilmaz</i>	
Semi-Supervised Acoustic and Language Modeling for Hindi ASR.....	3528
<i>Tarun Sai Bandarupalli, Shakti Rath, Nirmesh Shah, Onoe Naoyuki, Sriram Ganapathy</i>	
Combining Spectral and Self-Supervised Features for Low Resource Speech Recognition and Translation.....	3533
<i>Dan Berrebbi, Jiatong Shi, Brian Yan, Osbel López-Francisco, Jonathan Amith, Shinji Watanabe</i>	
When is TTS Augmentation Through a Pivot Language Useful?.....	3538
<i>Nathaniel Romney Robinson, Perez Ogayo, Swetha R. Gangu, David R. Mortensen, Shinji Watanabe</i>	
Low Resource Comparison of Attention-Based and Hybrid ASR Exploiting Wav2vec 2.0.....	3543
<i>Aku Rouhe, Anja Virkkunen, Juho Leinonen, Mikko Kurimo</i>	
Gram Vaani ASR Challenge on Spontaneous Telephone Speech Recordings in Regional Variations of Hindi	3548
<i>Anish Bhanushali, Grant Bridgman, Deekshitha G, Prasanta Ghosh, Pratik Kumar, Saurabh Kumar, Adithya Raj Kolladath, Nithya Ravi, Aaditeshwar Seth, Ashish Seth, Abhayjeet Singh, Vrunda Sukhadia, Umesh S, Sathvik Udupa, Lodagala V. S. V. Durga Prasad</i>	

SPEECH REPRESENTATION I

Audio Similarity is Unreliable as a Proxy for Audio Quality	3553
<i>Pranay Manocha, Zeyu Jin, Adam Finkelstein</i>	
Overlapped Frequency-Distributed Network: Frequency-Aware Voice Spoofing Countermeasure.....	3558
<i>Sunmook Choi, Il-Youp Kwak, Seungsang Oh</i>	
Formant Estimation and Tracking Using Probabilistic Heat-Maps	3563
<i>Yosi Shrem, Felix Kreuk, Joseph Keshet</i>	
Anti-Spoofing Using Transfer Learning with Variational Information Bottleneck	3568
<i>Youngsik Eom, Yeonghyeon Lee, Ji Sub Um, Hoi Rin Kim</i>	
Robust Pitch Estimation Using Multi-Branch CNN-LSTM and 1-Norm LP Residual	3573
<i>Mudit D. Batra, Jayesh, C. S. Ramalingam</i>	
DeepFry: Identifying Vocal Fry Using Deep Neural Networks.....	3578
<i>Bronya Roni Chernyak, Talia Ben Simon, Yael Segal, Jeremy Steffman, Eleanor Chodroff, Jennifer Cole, Joseph Keshet</i>	

Phonetic Analysis of Self-Supervised Representations of English Speech	3583
<i>Dan Wells, Hao Tang, Korin Richmond</i>	
FitHuBERT: Going Thinner and Deeper for Knowledge Distillation of Speech Self-Supervised Models.....	3588
<i>Yeonghyeon Lee, Kangwook Jang, Jahyun Goo, Youngmoon Jung, Hoi Rin Kim</i>	
On Combining Global and Localized Self-Supervised Models of Speech	3593
<i>Sri Harsha Dumpala, Chandramouli Shama Sastry, Rudolf Uher, Sageev Oore</i>	
Self-Supervised Representation Fusion for Speech and Wearable Based Emotion Recognition.....	3598
<i>Vipula Dissanayake, Sachith Seneviratne, Hussel Suriyaarachchi, Elliott Wen, Suranga Nanayakkara</i>	
Towards Disentangled Speech Representations	3603
<i>Cal Peysers, W. Ronny Huang, Andrew Rosenberg, Tara Sainath, Michael Picheny, Kyunghyun Cho</i>	

PATHOLOGICAL SPEECH ASSESSMENT

Automatic Assessment of Speech Intelligibility Using Consonant Similarity for Head and Neck Cancer.....	3608
<i>Sebastião Quintas, Julie Mauclair, Virginie Woisard, Julien Pinquier</i>	
Compensation in Verbal and Nonverbal Communication After Total Laryngectomy	3613
<i>Marise Neijman, Femke Hof, Noelle Oosterom, Roland Pfau, Bertus Van Rooy, Rob J. J. H. Van Son, Michiel M. W. M. Van Den Brekel</i>	
Wav2vec2-Based Speech Rating System for Children with Speech Sound Disorder	3618
<i>Yaroslav Getman, Ragheb Al-Ghezi, Katja Voskoboinik, Tamás Grósz, Mikko Kurimo, Giampiero Salvi, Torbjørn Svendsen, Sofia Strömbergsson</i>	
Distinguishing Between Pre- And Post-Treatment in the Speech of Patients with Chronic Obstructive Pulmonary Disease.....	3623
<i>Andreas Triantafyllopoulos, Markus Fendler, Anton Batliner, Maurice Gerczuk, Shahin Amiriparian, Thomas Berghaus, Björn W. Schuller</i>	
A Study on the Phonetic Inventory Development of Children with Cochlear Implants for 5 Years After Implantation	3628
<i>Seonwoo Lee, Sunhee Kim, Minhwa Chung</i>	
Evaluation of Different Antenna Types and Positions in a Stepped Frequency Continuous-Wave Radar-Based Silent Speech Interface.....	3633
<i>Joao Vitor Menezes, Pouriya Amini Digebsara, Christoph Wagner, Marco Mütze, Michael Bärhold, Petr Schaffer, Dirk Plettemeier, Peter Birkholz</i>	
Validation of the Neuro-Concept Detector Framework for the Characterization of Speech Disorders: A Comparative Study Including Dysarthria and Dysphonia	3638
<i>Sondes Abderrazek, Corinne Fredouille, Alain Ghio, Muriel Lalain, Christine Meunier, Virginie Woisard</i>	
Nonwords Pronunciation Classification in Language Development Tests for Preschool Children.....	3643
<i>Ilja Baumann, Dominik Wagner, Sebastian Bayerl, Tobias Bocklet</i>	

PERCEPT-R: An Open-Access American English Child/Clinical Speech Corpus Specialized for the
Audio Classification of /ɹ/ 3648
*Nina Benway, Jonathan L. Preston, Elaine Hitchcock, Asif Salekin, Harshit Sharma, Tara
McAllister*

Data Augmentation for End-To-End Silent Speech Recognition for Laryngectomees..... 3653
*Beiming Cao, Kristin Teplansky, Nordine Sebkh, Arpan Bhavsar, Omer Inan, Robin Samlan,
Ted Mau, Jun Wang*

Statistical and Clinical Utility of Multimodal Dialogue-Based Speech and Facial Metrics for
Parkinson's Disease Assessment..... 3658
*Hardik Kothare, Michael Neumann, Jackson Liscombe, Oliver Roesler, William Burke,
Andrew Exner, Sandy Snyder, Andrew Cornish, Doug Habberstad, David Pautler, David
Suendermann-Oeft, Jessica Huber, Vikram Ramanarayanan*

SHOW AND TELL III

Evaluation of Call Centre Conversations Based on a High-Level Symbolic Representation..... 3663
Leticia Arco, Carlos Mosquera, Fabjola Braho, Yisel Clavel, Johan Loeckx

Evoc-Learn — High Quality Simulation of Early Vocal Learning..... 3665
*Yi Xu, Anqi Xu, Daniel R. Van Niekerk, Branislav Gerazov, Peter Birkholz, Paul Konstantin
Krug, Santitham Prom-On, Lorna F. Halliday*

Watch Me Speak: 2D Visualization of Human Mouth During Speech..... 3667
C Siddarth, Sathvik Udupa, Prasanta Kumar Ghosh

SPEAKER AND LANGUAGE RECOGNITION II

Classification of Accented English Using CNN Model Trained on Amplitude Mel-Spectrograms 3669
*Mariia Lesnichaia, Veranika Mikhailava, Natalia Bogach, Iurii Lezhenin, John Blake, Evgeny
Pyshkin*

MIM-DG: Mutual Information Minimization-Based Domain Generalization for Speaker
Verification 3674
Woohyun Kang, Md Jahangir Alam, Abderrahim Fathan

Multi-Channel Far-Field Speaker Verification with Large-Scale Ad-Hoc Microphone Arrays..... 3679
Chengdong Liang, Yijiang Chen, Jiadi Yao, Xiao-Lei Zhang

Ant Multilingual Recognition System for OLR 2021 Challenge 3684
Anqi Lyu, Zhiming Wang, Huijia Zhu

Class-Aware Distribution Alignment Based Unsupervised Domain Adaptation for Speaker
Verification 3689
Hang-Rui Hu, Yan Song, Li-Rong Dai, Ian McLoughlin, Lin Liu

EDITnet: A Lightweight Network for Unsupervised Domain Adaptation in Speaker Verification 3694
Jingyu Li, Wei Liu, Tan Lee

Why Does Self-Supervised Learning for Speech Recognition Benefit Speaker Recognition? 3699
*Sanyuan Chen, Yu Wu, Chengyi Wang, Shujie Liu, Zhuo Chen, Peidong Wang, Gang Liu,
Jinyu Li, Jian Wu, Xiangzhan Yu, Furu Wei*

Audio Visual Multi-Speaker Tracking with Improved GCF and PMBM Filter	3704
<i>Jinzheng Zhao, Peipei Wu, Xubo Liu, Shidrokh Goudarzi, Haohe Liu, Yong Xu, Wenwu Wang</i>	
The HCCL System for the NIST SRE21	3709
<i>Zhuo Li, Runqiu Xiao, Hangting Chen, Zhenduo Zhao, Zihan Zhang, Wenchao Wang</i>	
UNet-DenseNet for Robust Far-Field Speaker Verification	3714
<i>Zhenke Gao, Manwai Mak, Weiwei Lin</i>	
Linguistic-Acoustic Similarity Based Accent Shift for Accent Recognition	3719
<i>Qijie Shao, Jinghao Yan, Jian Kang, Pengcheng Guo, Xian Shi, Pengfei Hu, Lei Xie</i>	
Transducer-Based Language Embedding for Spoken Language Identification.....	3724
<i>Peng Shen, Xugang Lu, Hisashi Kawai</i>	
Oriental Language Recognition (OLR) 2021: Summary and Analysis	3729
<i>Binling Wang, Feng Wang, Wenxuan Hu, Qiulin Wang, Jing Li, Dong Wang, Lin Li, Qingyang Hong</i>	

SPEECH SEGMENTATION II

Mixup Regularization Strategies for Spoofing Countermeasure System	3734
<i>Woohyun Kang, Md Jahangir Alam, Abderrahim Fathan</i>	
Low-Resource Low-Footprint Wake-Word Detection Using Knowledge Distillation	3739
<i>Arindam Ghosh, Mark Fuhs, Deblin Bagchi, Bahman Farahani, Monika Woszczyna</i>	
Personal VAD 2.0: Optimizing Personal Voice Activity Detection for On-Device Speech Recognition	3744
<i>Shaojin Ding, Rajeev Rikhye, Qiao Liang, Yanzhang He, Quan Wang, Arun Narayanan, Tom O'Malley, Ian McGraw</i>	
Token-Level Speaker Change Detection Using Speaker Difference and Speech Content Via Continuous Integrate-And-Fire.....	3749
<i>Zhiyun Fan, Zhenlin Liang, Linhao Dong, Yi Liu, Shiyu Zhou, Meng Cai, Jun Zhang, Zejun Ma, Bo Xu</i>	
NAS-VAD: Neural Architecture Search for Voice Activity Detection	3754
<i>Daniel Rho, Jinhyeok Park, Jong Hwan Ko</i>	
Adversarial Multi-Task Deep Learning for Noise-Robust Voice Activity Detection with Low Algorithmic Delay	3759
<i>Claus Larsen, Peter Koch, Zheng-Hua Tan</i>	
Rainbow Keywords: Efficient Incremental Learning for Online Spoken Keyword Spotting	3764
<i>Yang Xiao, Nana Hou, Eng Siong Chng</i>	
Filler Word Detection and Classification: A Dataset and Benchmark	3769
<i>Ge Zhu, Juan-Pablo Caceres, Justin Salamon</i>	

ROBUST ASR, AND FAR-FIELD/MULTI-TALKER ASR

Streaming Multi-Talker ASR with Token-Level Serialized Output Training	3774
<i>Naoyuki Kanda, Jian Wu, Yu Wu, Xiong Xiao, Zhong Meng, Xiaofei Wang, Yashesh Gaur, Zhuo Chen, Jinyu Li, Takuya Yoshioka</i>	

PMCT: Patched Multi-Condition Training for Robust Speech Recognition	3779
<i>Pablo Peso Parada, Agnieszka Dobrowolska, Karthikeyan Saravanan, Mete Ozay</i>	
Improving ASR Robustness in Noisy Condition Through VAD Integration	3784
<i>Sashi Novitasari, Takashi Fukuda, Gakuto Kurata</i>	
Empirical Sampling from Latent Utterance-Wise Evidence Model for Missing Data ASR Based on Neural Encoder-Decoder Model.....	3789
<i>Ryu Takeda, Yui Sudo, Kazuhiro Nakadai, Kazunori Komatani</i>	
Coarse-Grained Attention Fusion with Joint Training Framework for Complex Speech Enhancement and End-To-End Speech Recognition	3794
<i>Xuyi Zhuang, Lu Zhang, Zehua Zhang, Yukun Qian, Mingjiang Wang</i>	
DENT-DDSP: Data-Efficient Noisy Speech Generator Using Differentiable Digital Signal Processors for Explicit Distortion Modelling and Noise-Robust Speech Recognition.....	3799
<i>Zixun Guo, Chen Chen, Eng Siong Chng</i>	
Improving Transformer-Based Conversational ASR by Inter-Sentential Attention Mechanism	3804
<i>Kun Wei, Pengcheng Guo, Ning Jiang</i>	
Federated Self-Supervised Speech Representations: Are We There Yet?	3809
<i>Yan Gao, Javier Fernandez-Marques, Titouan Parcollet, Abhinav Mehrotra, Nicholas Lane</i>	
Leveraging Real Conversational Data for Multi-Channel Continuous Speech Separation	3814
<i>Xiaofei Wang, Dongmei Wang, Naoyuki Kanda, Sefik Emre Eskimez, Takuya Yoshioka</i>	
End-To-End Integration of Speech Recognition, Speech Enhancement, and Self-Supervised Learning Representation.....	3819
<i>Xuankai Chang, Takashi Maekaku, Yuya Fujita, Shinji Watanabe</i>	
Weakly-Supervised Neural Full-Rank Spatial Covariance Analysis for a Front-End System of Distant Speech Recognition.....	3824
<i>Yoshiaki Bando, Takahiro Aizawa, Katsutoshi Itoyama, Kazuhiro Nakadai</i>	
A Universally-Deployable ASR Frontend for Joint Acoustic Echo Cancellation, Speech Enhancement, and Voice Separation.....	3829
<i>Thomas R. O'Malley, Arun Narayanan, Quan Wang</i>	
Speaker Conditioned Acoustic Modeling for Multi-Speaker Conversational ASR.....	3834
<i>Srikanth Raj Chetupalli, Sriram Ganapathy</i>	
Hear No Evil: Towards Adversarial Robustness of Automatic Speech Recognition Via Multi-Task Learning	3839
<i>Nilaksh Das, Polo Chau</i>	
Tandem Multitask Training of Speaker Diarisation and Speech Recognition for Meeting Transcription.....	3844
<i>Xianrui Zheng, Chao Zhang, Phil Woodland</i>	

ASR: LINGUISTIC COMPONENTS

Investigating the Impact of Crosslingual Acoustic-Phonetic Similarities on Multilingual Speech Recognition	3849
<i>Muhammad Umar Farooq, Thomas Hain</i>	

An Improved Deliberation Network with Text Pre-Training for Code-Switching Automatic Speech Recognition	3854
<i>Zhijie Shen, Wu Guo</i>	
CyclicAugment: Speech Data Random Augmentation with Cosine Annealing Scheduler for Automatic Speech Recognition	3859
<i>Zhihan Wang, Feng Hou, Yuanhang Qiu, Zhizhong Ma, Satwinder Singh, Ruili Wang</i>	
Prompt-Based Re-Ranking Language Model for ASR.....	3864
<i>Mengxi Nie, Ming Yan, Caixia Gong</i>	
Avoid Overfitting User Specific Information in Federated Keyword Spotting	3869
<i>Xin-Chun Li, Jin-Lin Tang, Shaoming Song, Bingshuai Li, Yinchuan Li, Yunfeng Shao, Le Gan, De-Chuan Zhan</i>	
ASR Error Correction with Constrained Decoding on Operation Prediction	3874
<i>Jingyuan Yang, Rongjun Li, Wei Peng</i>	
Adaptive Multilingual Speech Recognition with Pretrained Models	3879
<i>Ngoc-Quan Pham, Alexander Waibel, Jan Niehues</i>	
Vietnamese Capitalization and Punctuation Recovery Models	3884
<i>Hoang Thi Thu Uyen, Nguyen Anh Tu, Ta Duc Huy</i>	
Non-Autoregressive Error Correction for CTC-Based ASR with Phone-Conditioned Masked LM.....	3889
<i>Hayato Futami, Hirofumi Inaguma, Sei Ueno, Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara</i>	
Reducing Multilingual Context Confusion for End-To-End Code-Switching Automatic Speech Recognition	3894
<i>Shuai Zhang, Jiangyan Yi, Zhengkun Tian, Jianhua Tao, Yu Ting Yeung, Liqun Deng</i>	
Residual Language Model for End-To-End Speech Recognition.....	3899
<i>Emiru Tsunoo, Yosuke Kashiwagi, Chaitanya Prasad Narisetty, Shinji Watanabe</i>	
An Empirical Study of Language Model Integration for Transducer Based Speech Recognition.....	3904
<i>Huahuan Zheng, Keyu An, Zhijian Ou, Chen Huang, Ke Ding, Guanglu Wan</i>	
Self-Normalized Importance Sampling for Neural Language Modeling.....	3909
<i>Zijian Yang, Yingbo Gao, Alexander Gerstenberger, Jintao Jiang, Ralf Schlüter, Hermann Ney</i>	
Improving Contextual Recognition of Rare Words with an Alternate Spelling Prediction Model	3914
<i>Jennifer Fox, Natalie Delworth</i>	
Effect and Analysis of Large-Scale Language Model Rescoring on Competitive ASR Systems	3919
<i>Takuma Udagawa, Masayuki Suzuki, Gakuto Kurata, Nobuyasu Itoh, George Saon</i>	
Language-Specific Characteristic Assistance for Code-Switching Speech Recognition.....	3924
<i>Tongtong Song, Qiang Xu, Meng Ge, Longbiao Wang, Hao Shi, Yongjie Lv, Yuqin Lin, Jianwu Dang</i>	

VOLUME 6

SPEECH INTELLIGIBILITY PREDICTION FOR HEARING-IMPAIRED LISTENERS II

- Speech Intelligibility of Simulated Hearing Loss Sounds and Its Prediction Using the Gammachirp Envelope Similarity Index (GESI) 3929
Toshio Irino, Honoka Tamaru, Ayako Yamamoto
- ELO-SPHERES Intelligibility Prediction Model for the Clarity Prediction Challenge 2022 3934
Mark Huckvale, Gaston Hilkhuisen
- Listening with Googlears: Low-Latency Neural Multiframe Beamforming and Equalization for Hearing Aids 3939
Samuel Yang, Scott Wisdom, Chet Gnegy, Richard F. Lyon, Sagar Savla
- MBI-Net: A Non-Intrusive Multi-Branched Speech Intelligibility Prediction Model for Hearing Aids 3944
Ryandhimas Edo Zezario, Fei Chen, Chiou-Shann Fuh, Hsin-Min Wang, Yu Tsao

SHOW AND TELL III(VR)

- A Deep Learning Platform for Language Education Research and Development..... 3949
Kye Min Tan, Richeng Duan, Xin Huang, Bowei Zou, Xuan Long Do
- A VR Interactive 3D Mandarin Pronunciation Teaching Model 3951
Yujia Jin, Yanlu Xie, Jinsong Zhang

SUMMARIZATION, ENTITY EXTRACTION, EVALUATION AND OTHERS

- Squashed Weight Distribution for Low Bit Quantization of Deep Models 3953
Nikko Strom, Haidar Khan, Wael Hamza
- Evaluating the Performance of State-Of-The-Art ASR Systems on Non-Native English Using Corpora with Extensive Language Background Variation..... 3958
Samuel Hollands, Daniel Blackburn, Heidi Christensen
- Seq-2-Seq Based Refinement of ASR Output for Spoken Name Capture..... 3963
Karan Singla, Shahab Jalalvand, Yeon-Jun Kim, Ryan Price, Daniel Pressel, Srinivas Bangalore
- Qualitative Evaluation of Language Model Rescoring in Automatic Speech Recognition 3968
Thibault Bañeras Roux, Mickael Rouvier, Jane Wottawa, Richard Dufour
- Toward Zero Oracle Word Error Rate on the Switchboard Benchmark 3973
Arlo Faria, Adam Janin, Sidhi Adkoli, Korbinian Riedhammer
- Evaluating User Perception of Speech Recognition System Quality with Semantic Distance Metric 3978
Suyoun Kim, Duc Le, Weiyi Zheng, Tarun Singh, Abhinav Arora, Xiaoyu Zhai, Christian Fuegen, Ozlem Kalinli, Michael Seltzer

AUTOMATIC ANALYSIS OF PARALINGUISTICS

Predicting Emotional Intensity in Political Debates Via Non-Verbal Signals	3983
<i>Jeewoo Yoon, Jinyoung Han, Erik Bucy, Jungseock Joo</i>	
Confusion Detection for Adaptive Conversational Strategies of an Oral Proficiency Assessment Interview Agent	3988
<i>Mao Saeki, Kotoka Miyagi, Shinya Fujie, Shungo Suzuki, Tetsuji Ogawa, Tetsunori Kobayashi, Yoichi Matsuyama</i>	
Deep Learning for Prosody-Based Irony Classification in Spontaneous Speech	3993
<i>Helen Gent, Chase Adams, Yan Tang, Chilin Shih</i>	
Span Classification with Structured Information for Disfluency Detection in Spoken Utterances	3998
<i>Sreyan Ghosh, Sonal Kumar, Yaman Kumar, Rajiv Ratn Shah, Srinivasan Umesh</i>	
Example-Based Explanations with Adversarial Attacks for Respiratory Sound Analysis	4003
<i>Yi Chang, Zhao Ren, Thanh Tam Nguyen, Wolfgang Nejdl, Björn W. Schuller</i>	
Which Model is Best: Comparing Methods and Metrics for Automatic Laughter Detection in a Naturalistic Conversational Dataset	4008
<i>Gordon Rennie, Olga Perepelkina, Alessandro Vinciarelli</i>	

SELF SUPERVISION AND ANTI-SPOOFING

Self-Supervised Speaker Diarization	4013
<i>Yehoshua Dissen, Felix Kreuk, Joseph Keshet</i>	
Label-Efficient Self-Supervised Speaker Verification with Information Maximization and Contrastive Learning	4018
<i>Theo Lepage, Reda Dehak</i>	
Attack Agnostic Dataset: Towards Generalization and Stabilization of Audio DeepFake Detection	4023
<i>Piotr Kawa, Marcin Plata, Piotr Syga</i>	
Non-Contrastive Self-Supervised Learning of Utterance-Level Speech Representations	4028
<i>Jaemin Cho, Raghavendra Pappagari, Piotr Zelasko, Laureano Moro Velazquez, Jesus Villalba, Najim Dehak</i>	
Barlow Twins Self-Supervised Learning for Robust Speaker Recognition	4033
<i>Mohammad Mohammadamini, Driss Matrouf, Jean-Francois Bonastre, Sandipana Dowerah, Romain Serizel, Denis Jouviet</i>	

SPEECH ARTICULATION & NEURAL PROCESSING

Relating the Fundamental Frequency of Speech with EEG Using a Dilated Convolutional Network	4038
<i>Corentin Puffay, Jana Van Canneyt, Jonas Vanthornhout, Hugo Van Hamme, Tom Francart</i>	
Prediction of L2 Speech Proficiency Based on Multi-Level Linguistic Features	4043
<i>Verdiana De Fino, Lionel Fontan, Julien Pinquier, Isabelle Ferrané, Sylvain Detey</i>	
The Effect of Increasing Acoustic and Linguistic Complexity on Auditory Processing: An EEG Study	4048
<i>Fareeha S. Rana, Daniel Pape, Elisabet Service</i>	

Recording and Timing Vocal Responses in Online Experimentation 4053
Katrina Kechun Li, Julia Schwarz, Jasper Hong Sim, Yixin Zhang, Elizabeth Buchanan-Worster, Brechtje Post, Kirsty McDougall

Neural Correlates of Acoustic and Semantic Cues During Speech Segmentation in French..... 4058
Maria Del Mar Cordero, Ambre Denis-Noël, Elsa Spinelli, Fanny Meunier

Evidence of Onset and Sustained Neural Responses to Isolated Phonemes from Intracranial Recordings in a Voice-Based Cursor Control Task..... 4063
Kevin Meng, Seo-Hyun Lee, Farhad Goodarzy, Simon Vogrin, Mark J. Cook, Seong-Whan Lee, David B. Grayden

LOW RESOURCE SPOKEN LANGUAGE UNDERSTANDING

End-To-End Model for Named Entity Recognition from Speech Without Paired Training Data..... 4068
Salima Mdhaffar, Jarod Duret, Titouan Parcollet, Yannick Estève

Multitask Learning for Low Resource Spoken Language Understanding..... 4073
Quentin Meeus, Marie Francine Moens, Hugo Van Hamme

NON-INTRUSIVE OBJECTIVE SPEECH QUALITY ASSESSMENT (NISQA) CHALLENGE FOR ONLINE CONFERENCING APPLICATIONS

Transformer Networks for Non-Intrusive Speech Quality Prediction 4078
M K Jayesh, Mukesh Sharma, Praneeth Vonteddu, Mahaboob Ali Basha Shaik, Sriram Ganapathy

Pre-Trained Speech Representations as Feature Extractors for Speech Quality Assessment in Online Conferencing Applications 4083
Bastiaan Tamm, Helena Balabin, Rik Vandenberghe, Hugo Van Hamme

Exploring the Influence of Fine-Tuning Data on Wav2vec 2.0 Model for Blind Speech Quality Prediction 4088
Helard Becerra, Alessandro Ragano, Andrew Hines

NOVEL MODELS AND TRAINING METHODS FOR ASR I

MAESTRO: Matched Speech Text Representations Through Modality Matching..... 4093
Zhehuai Chen, Yu Zhang, Andrew Rosenberg, Bhuvana Ramabhadran, Pedro J. Moreno, Ankur Bapna, Heiga Zen

FiLM Conditioning with Enhanced Feature to the Transformer-Based End-To-End Noisy Speech Recognition 4098
Da-Hee Yang, Joon-Hyuk Chang

SepTr: Separable Transformer for Audio Spectrogram Processing 4103
Nicolaea Catalin Ristea, Radu Tudor Ionescu, Fahad Shahbaz Khan

End-To-End Spontaneous Speech Recognition Using Disfluency Labeling 4108
Koharu Horii, Meiko Fukuda, Kengo Ohta, Ryota Nishimura, Atsunori Ogawa, Norihide Kitaoka

Recent Improvements of ASR Models in the Face of Adversarial Attacks4113
Raphael Olivier, Bhiksha Raj

Similarity and Content-Based Phonetic Self Attention for Speech Recognition	4118
<i>Kyuhong Shim, Wonyong Sung</i>	
Generalizing RNN-Transducer to Out-Domain Audio Via Sparse Self-Attention Layers.....	4123
<i>Juntae Kim, Jeehye Lee</i>	
Knowledge Distillation for In-Memory Keyword Spotting Model	4128
<i>Zeyang Song, Qi Liu, Qu Yang, Haizhou Li</i>	
Automatic Learning of Subword Dependent Model Scales	4133
<i>Felix Meyer, Wilfried Michel, Mohammad Zeineldeen, Ralf Schlüter, Hermann Ney</i>	
Bayesian Recurrent Units and the Forward-Backward Algorithm	4137
<i>Alexandre Bittar, Philip N. Garner</i>	

ACOUSTIC SCENE ANALYSIS

On Metric Learning for Audio-Text Cross-Modal Retrieval	4142
<i>Xinhao Mei, Xubo Liu, Jianyuan Sun, Mark Plumbley, Wenwu Wang</i>	
CT-SAT: Contextual Transformer for Sequential Audio Tagging	4147
<i>Yuanbo Hou, Zhaoyi Liu, Bo Kang, Yun Wang, Dick Botteldooren</i>	
ADFF: Attention Based Deep Feature Fusion Approach for Music Emotion Recognition	4152
<i>Zi Huang, Shulei Ji, Zhilan Hu, Chuangjian Cai, Jing Luo, Xinyu Yang</i>	
Audio-Visual Scene Classification Based on Multi-Modal Graph Fusion	4157
<i>Han Lei, Ning Chen</i>	
MusicNet: Compact Convolutional Neural Network for Real-Time Background Music Detection	4162
<i>Chandan Reddy, Vishak Gopal, Harishchandra Dubey, Ross Cutler, Sergiy Matushevych, Robert Aichner</i>	
iCNN-Transformer: An Improved CNN-Transformer with Channel-Spatial Attention and Keyword Prediction for Automated Audio Captioning	4167
<i>Kun Chen, Jun Wang, Feng Deng, Xiaorui Wang</i>	
ATST: Audio Representation Learning with Teacher-Student Transformer	4172
<i>Xian Li, Xiaofei Li</i>	
Deep Segment Model for Acoustic Scene Classification	4177
<i>Yajian Wang, Jun Du, Hang Chen, Qing Wang, Chin-Hui Lee</i>	
Novel Augmentation Schemes for Device Robust Acoustic Scene Classification	4182
<i>Sukanya Sonowal, Anish Tamse</i>	
WideResNet with Joint Representation Learning and Data Augmentation for Cover Song Identification	4187
<i>Shichao Hu, Bin Zhang, Jinhong Lu, Yiliang Jiang, Wucheng Wang, Lingcheng Kong, Weifeng Zhao, Tao Jiang</i>	
Impact of Acoustic Event Tagging on Scene Classification in a Multi-Task Learning Framework	4192
<i>Rahil Parikh, Harshavardhan Sundar, Ming Sun, Chao Wang, Spyros Matsoukas</i>	
Introducing Auxiliary Text Query-Modifier to Content-Based Audio Retrieval	4197
<i>Daiki Takeuchi, Yasunori Ohishi, Daisuke Niizumi, Noboru Harada, Kunio Kashino</i>	

SPEECH CODING AND PRIVACY

Speaker Recognition-Assisted Robust Audio Deepfake Detection	4202
<i>Jiahui Pan, Shuai Nie, Hui Zhang, Shulin He, Kanghao Zhang, Shan Liang, Xueliang Zhang, Jianhua Tao</i>	
Preventing Sensitive-Word Recognition Using Self-Supervised Learning to Preserve User-Privacy for Automatic Speech Recognition.....	4207
<i>Yuchen Liu, Apu Kapadia, Donald Williamson</i>	
NESC: Robust Neural End-2-End Speech Coding with GANs.....	4212
<i>Nicola Pia, Kishan Gupta, Srikanth Korse, Markus Multrus, Guillaume Fuchs</i>	
Towards Error-Resilient Neural Speech Coding.....	4217
<i>Huaying Xue, Xiulian Peng, Xue Jiang, Yan Lu</i>	
Cross-Scale Vector Quantization for Scalable Neural Speech Coding	4222
<i>Xue Jiang, Xiulian Peng, Huaying Xue, Yuan Zhang, Yan Lu</i>	
Neural Vocoder is All You Need for Speech Super-Resolution.....	4227
<i>Haohe Liu, Woosung Choi, Xubo Liu, Qiuqiang Kong, Qiao Tian, Deliang Wang</i>	
VoiceFixer: A Unified Framework for High-Fidelity Speech Restoration.....	4232
<i>Haohe Liu, Xubo Liu, Qiuqiang Kong, Qiao Tian, Yan Zhao, Deliang Wang, Chuanzeng Huang, Yuxuan Wang</i>	
Generating Gender-Ambiguous Voices for Privacy-Preserving Speech Recognition	4237
<i>Dimitrios Stoidis, Andrea Cavallaro</i>	

SPEECH SYNTHESIS: SINGING, MULTIMODAL, CROSSLINGUAL SYNTHESIS

Opencpop: A High-Quality Open Source Chinese Popular Song Corpus for Singing Voice Synthesis	4242
<i>Yu Wang, Xinsheng Wang, Pengcheng Zhu, Jie Wu, Hanzhao Li, Heyang Xue, Yongmao Zhang, Lei Xie, Mengxiao Bi</i>	
Exploring Timbre Disentanglement in Non-Autoregressive Cross-Lingual Text-To-Speech	4247
<i>Haoyue Zhan, Xinyuan Yu, Haitong Zhang, Yang Zhang, Yue Lin</i>	
WeSinger: Data-Augmented Singing Voice Synthesis with Auxiliary Losses.....	4252
<i>Zewang Zhang, Yibin Zheng, Xinhui Li, Li Lu</i>	
Decoupled Pronunciation and Prosody Modeling in Meta-Learning-Based Multilingual Speech Synthesis.....	4257
<i>Yukun Peng, Zhenhua Ling</i>	
KaraTuner: Towards End-To-End Natural Pitch Correction for Singing Voice in Karaoke	4262
<i>Xiaobin Zhuang, Huiran Yu, Weifeng Zhao, Tao Jiang, Peng Hu</i>	
Learn2Sing 2.0: Diffusion and Mutual Information-Based Target Speaker SVS by Learning from Singing Teacher	4267
<i>Heyang Xue, Xinsheng Wang, Yongmao Zhang, Lei Xie, Pengcheng Zhu, Mengxiao Bi</i>	
SingAug: Data Augmentation for Singing Voice Synthesis with Cycle-Consistent Training Strategy.....	4272
<i>Shuai Guo, Jiatong Shi, Tao Qian, Shinji Watanabe, Qin Jin</i>	

Muskits: An End-To-End Music Processing Toolkit for Singing Voice Synthesis	4277
<i>Jiatong Shi, Shuai Guo, Tao Qian, Tomoki Hayashi, Yuning Wu, Fangzheng Xu, Xuankai Chang, Huazhe Li, Peter Wu, Shinji Watanabe, Qin Jin</i>	
Pronunciation Dictionary-Free Multilingual Speech Synthesis by Combining Unsupervised and Supervised Phonetic Representations	4282
<i>Chang Liu, Zhen-Hua Ling, Ling-Hui Chen</i>	
Towards High-Fidelity Singing Voice Conversion with Acoustic Reference and Contrastive Predictive Coding	4287
<i>Chao Wang, Zhonghao Li, Benlai Tang, Xiang Yin, Yuan Wan, Yibiao Yu, Zejun Ma</i>	
Towards Improving the Expressiveness of Singing Voice Synthesis with BERT Derived Semantic Information	4292
<i>Shaohuan Zhou, Shun Lei, Weiya You, Deyi Tuo, Yuren You, Zhiyong Wu, Shiyin Kang, Helen Meng</i>	
Normalization of Code-Switched Text for Speech Synthesis	4297
<i>Sreeram Manghat, Sreeja Manghat, Tanja Schultz</i>	
Synthesizing Near Native-Accented Speech for a Non-Native Speaker by Imitating the Pronunciation and Prosody of a Native Speaker	4302
<i>Raymond Chung, Brian Mak</i>	
A Hierarchical Speaker Representation Framework for One-Shot Singing Voice Conversion	4307
<i>Xu Li, Shansong Liu, Ying Shan</i>	

APPLICATIONS IN TRANSCRIPTION, EDUCATION AND LEARNING II

Self-Supervised Learning with Multi-Target Contrastive Coding for Non-Native Acoustic Modeling of Mispronunciation Verification	4312
<i>Longfei Yang, Jinsong Zhang, Takahiro Shinozaki</i>	
L2-GEN: A Neural Phoneme Paraphrasing Approach to L2 Speech Synthesis for Mispronunciation Diagnosis	4317
<i>Daniel Zhang, Ashwinkumar Ganesan, Sarah Campbell, Daniel Korzekwa</i>	
Challenges Remain in Building ASR for Spontaneous Preschool Children Speech in Naturalistic Educational Environments	4322
<i>Satwik Dutta, Sarah Anne Tao, Jacob C. Reyna, Rebecca Elizabeth Hacker, Dwight W. Irvin, Jay F. Buzhardt, John H. L. Hansen</i>	
End-To-End Mispronunciation Detection with Simulated Error Distance	4327
<i>Zhan Zhang, Yuehai Wang, Jianyi Yang</i>	
BiCAPT: Bidirectional Computer-Assisted Pronunciation Training with Normalizing Flows	4332
<i>Zhan Zhang, Yuehai Wang, Jianyi Yang</i>	
Using Fluency Representation Learned from Sequential Raw Features for Improving Non-Native Fluency Scoring	4337
<i>Kaiqi Fu, Shaojun Gao, Xiaohai Tian, Wei Li, Ma Zejun</i>	
An Alignment Method Leveraging Articulatory Features for Mispronunciation Detection and Diagnosis in L2 English	4342
<i>Qi Chen, Binghuai Lin, Yanlu Xie</i>	

RefTextLAS: Reference Text Biased Listen, Attend, and Spell Model for Accurate Reading Evaluation.....	4347
<i>Phani Sankar Nidadavolu, Na Xu, Nick Jutila, Ravi Teja Gadde, Aswarth Abhilash Dara, Joseph Savold, Sapan Patel, Aaron Hoff, Veerdhawal Pande, Kevin Crews, Ankur Gandhe, Ariya Rastrow, Roland Maas</i>	

CoCA-MDD: A Coupled Cross-Attention Based Framework for Streaming Mispronunciation Detection and Diagnosis.....	4352
<i>Nianzu Zheng, Liqun Deng, Wenyong Huang, Yu Ting Yeung, Baohua Xu, Yuanyuan Guo, Yasheng Wang, Xiao Chen, Xin Jiang, Qun Liu</i>	

SPOOFING-AWARE AUTOMATIC SPEAKER VERIFICATION (SASV) II

Spoofing-Aware Speaker Verification by Multi-Level Fusion	4357
<i>Haibin Wu, Lingwei Meng, Jiawen Kang, Jinchao Li, Xu Li, Xixin Wu, Hung-Yi Lee, Helen Meng</i>	

End-To-End Framework for Spoof-Aware Speaker Verification.....	4362
<i>Woohyun Kang, Md Jahangir Alam, Abderrahim Fathan</i>	

The CLIPS System for 2022 Spoofing-Aware Speaker Verification Challenge.....	4367
<i>Jucai Lin, Tingwei Chen, Jingbiao Huang, Ruidong Fang, Jun Yin, Yuanping Yin, Wei Shi, Weizhen Huang, Yapeng Mao</i>	

Norm-Constrained Score-Level Ensemble for Spoofing Aware Speaker Verification	4371
<i>Peng Zhang, Peng Hu, Xueliang Zhang</i>	

SASV Based on Pre-Trained ASV System and Integrated Scoring Module.....	4376
<i>Yuxiang Zhang, Zhuo Li, Wenchao Wang, Pengyuan Zhang</i>	

Backend Ensemble for Speaker Verification and Spoofing Countermeasure.....	4381
<i>Li Zhang, Yue Li, Huan Zhao, Qing Wang, Lei Xie</i>	

NRI-FGSM: An Efficient Transferable Adversarial Attack for Speaker Recognition Systems.....	4386
<i>Hao Tan, Junjian Zhang, Huan Zhang, Le Wang, Yaguan Qian, Zhaoquan Gu</i>	

SA-SASV: An End-To-End Spoof-Aggregated Spoofing-Aware Speaker Verification System.....	4391
<i>Zhongwei Teng, Quchen Fu, Jules White, Maria Powell, Douglas Schmidt</i>	

The DKU-OPPO System for the 2022 Spoofing-Aware Speaker Verification Challenge.....	4396
<i>Xingming Wang, Xiaoyi Qin, Yikang Wang, Yunfei Xu, Ming Li</i>	

SPEECH CODING AND RESTORATION

NU-Wave 2: A General Neural Audio Upsampling Model for Various Sampling Rates.....	4401
<i>Seungu Han, Junhyeok Lee</i>	

SelfRemaster: Self-Supervised Speech Restoration with Analysis-By-Synthesis Approach Using Channel Modeling	4406
<i>Takaaki Saeki, Shinnosuke Takamichi, Tomohiko Nakamura, Naoko Tanji, Hiroshi Saruwatari</i>	

Optimization of Deep Neural Network (DNN) Speech Coder Using a Multi Time Scale Perceptual Loss Function	4411
<i>Joon Byun, Seungmin Shin, Jongmo Sung, Seungkwon Beack, Youngcheol Park</i>	

Phase Vocoder for Time Stretch Based on Center Frequency Estimation	4416
<i>Donghyeon Kim, Bowon Lee</i>	
Ultra-Low-Bitrate Speech Coding with Pretrained Transformers	4421
<i>Ali Siahkoobi, Michael Chinen, Tom Denton, W. Bastiaan Kleijn, Jan Skoglund</i>	
Analyzing Language-Independent Speaker Anonymization Framework Under Unseen Conditions	4426
<i>Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, Natalia Tomashenko</i>	

STREAMING ASR

ConvRNN-T: Convolutional Augmented Recurrent Neural Network Transducers for Streaming Speech Recognition	4431
<i>Martin Radfar, Rohit Barnwal, Rupak Vignesh Swaminathan, Feng-Ju Chang, Grant P. Strimel, Nathan Susanj, Athanasios Mouchtaris</i>	
Knowledge Distillation Via Module Replacing for Automatic Speech Recognition with Recurrent Neural Network Transducer	4436
<i>Kaiqi Zhao, Hieu Nguyen, Animesh Jain, Nathan Susanj, Athanasios Mouchtaris, Lokesh Gupta, Ming Zhao</i>	
Memory-Efficient Training of RNN-Transducer with Sampled Softmax	4441
<i>Jaesong Lee, Lukas Lee, Shinji Watanabe</i>	
Multiple-Hypothesis RNN-T Loss for Unsupervised Fine-Tuning and Self-Training of Neural Transducer	4446
<i>Cong-Thanh Do, Mohan Li, Rama Doddipatla</i>	
Separator-Transducer-Segmenter: Streaming Recognition and Segmentation of Multi-Party Speech	4451
<i>Ilya Sklyar, Anna Piunova, Christian Osendorfer</i>	

APPLICATIONS IN TRANSCRIPTION, EDUCATION AND LEARNING I

Variations of Multi-Task Learning for Spoken Language Assessment	4456
<i>Jeremy Heng Meng Wong, Huayun Zhang, Nancy Chen</i>	
Detection of Learners' Listening Breakdown with Oral Dictation and Its Use to Model Listening Skill Improvement Exclusively Through Shadowing	4461
<i>Takuya Kunihara, Chuanbo Zhu, Daisuke Saito, Nobuaki Minematsu, Noriko Nakanishi</i>	
Automatic Prosody Evaluation of L2 English Read Speech in Reference to Accent Dictionary with Transformer Encoder	4466
<i>Yu Suzuki, Tsuneo Kato, Akihiro Tamura</i>	
View-Specific Assessment of L2 Spoken English	4471
<i>Stefano Bannò, Bhanu Balusu, Mark Gales, Kate Knill, Konstantinos Kyriakopoulos</i>	
The Effects of Implicit and Explicit Feedback in an ASR-Based Reading Tutor for Dutch First-Graders	4476
<i>Yu Bai, Ferdy Hubers, Catia Cucchiarini, Roeland Van Hout, Helmer Strik</i>	
Improving Mispronunciation Detection with Wav2vec2-Based Momentum Pseudo-Labeling for Accentedness and Intelligibility Assessment	4481
<i>Mu Yang, Kevin Hirschi, Stephen Daniel Looney, Okim Kang, John H. L. Hansen</i>	

SPOKEN DIALOGUE SYSTEMS

Response Timing Estimation for Spoken Dialog System Using Dialog Act Estimation.....	4486
<i>Jin Sakuma, Shinya Fujie, Tetsunori Kobayashi</i>	
Hesitations in Urdu/Hindi: Distribution and Properties of Fillers & Silences.....	4491
<i>Farhat Jabeen, Simon Betz</i>	
Interpretability of Speech Emotion Recognition Modelled Using Self-Supervised Speech and Text Pre-Trained Embeddings	4496
<i>K V Vijay Girish, Srikanth Konjeti, Jithendra Vepa</i>	
Does Utterance Entails Intent?: Evaluating Natural Language Inference Based Setup for Few-Shot Intent Detection	4501
<i>Ayush Kumar, Vijit Malik, Jithendra Vepa</i>	
Investigating Perception of Spoken Dialogue Acceptability Through Surprisal	4506
<i>Sarenne Carrol Wallbridge, Catherine Lai, Peter Bell</i>	
Low-Latency Online Streaming VideoQA Using Audio-Visual Transformers.....	4511
<i>Chiori Hori, Takaaki Hori, Jonathan Le Roux</i>	

THE VOICEMOS CHALLENGE

The ZovoMOS Entry to VoiceMOS Challenge 2022.....	4516
<i>Adriana Stan</i>	
UTMOS: UTokyo-SaruLab System for VoiceMOS Challenge 2022	4521
<i>Takaaki Saeki, Detai Xin, Wataru Nakata, Tomoki Koriyama, Shinnosuke Takamichi, Hiroshi Saruwatari</i>	
Automatic Mean Opinion Score Estimation with Temporal Modulation Features on Gammatone Filterbank for Speech Assessment.....	4526
<i>Huy Nguyen, Kai Li, Masashi Unoki</i>	
Using Rater and System Metadata to Explain Variance in the VoiceMOS Challenge 2022 Dataset.....	4531
<i>Michael Chinen, Jan Skoglund, Chandan K. A. Reddy, Alessandro Ragano, Andrew Hines</i>	
The VoiceMOS Challenge 2022	4536
<i>Wen Chin Huang, Erica Cooper, Yu Tsao, Hsin-Min Wang, Tomoki Toda, Junichi Yamagishi</i>	
DDOS: A MOS Prediction Framework Utilizing Domain Adaptive Pre-Training and Distribution of Opinion Scores	4541
<i>Wei-Cheng Tseng, Wei-Tsung Kao, Hung-Yi Lee</i>	

SPEECH SYNTHESIS: SPEAKING STYLE, EMOTION AND ACCENTS I

Expressive, Variable, and Controllable Duration Modelling in TTS	4546
<i>Syed Ammar Abbas, Thomas Merritt, Alexis Moinet, Sri Karlapati, Ewa Muszynska, Simon Slangen, Elia Gatti, Thomas Drugman</i>	

Predicting VQVAE-Based Character Acting Style from Quotation-Annotated Text for Audiobook Speech Synthesis	4551
<i>Wataru Nakata, Tomoki Koriyama, Shinnosuke Takamichi, Yuki Saito, Yusuke Ijima, Ryo Masumura, Hiroshi Saruwatari</i>	
Adversarial and Sequential Training for Cross-Lingual Prosody Transfer TTS.....	4556
<i>Min-Kyung Kim, Joon-Hyuk Chang</i>	
FluentTTS: Text-Dependent Fine-Grained Style Control for Multi-Style TTS.....	4561
<i>Changhwan Kim, Seyun Um, Hyungchan Yoon, Hong-Goo Kang</i>	
Few Shot Cross-Lingual TTS Using Transferable Phoneme Embedding.....	4566
<i>Wei-Ping Huang, Po-Chun Chen, Sung-Feng Huang, Hung-Yi Lee</i>	
Training Text-To-Speech Systems from Synthetic Data: A Practical Approach for Accent Transfer Tasks.....	4571
<i>Lev Finkelstein, Heiga Zen, Norman Casagrande, Chun-An Chan, Ye Jia, Tom Kenter, Alex Petelin, Jonathan Shen, Vincent Wan, Yu Zhang, Yonghui Wu, Robert Clark</i>	
Spoken-Text-Style Transfer with Conditional Variational Autoencoder and Content Word Storage	4576
<i>Daiki Yoshioka, Yusuke Yasuda, Noriyuki Matsunaga, Yamato Ohtani, Tomoki Toda</i>	
Analysis of Expressivity Transfer in Non-Autoregressive End-To-End Multispeaker TTS Systems.....	4581
<i>Ajinkya Kulkarni, Vincent Colotte, Denis Jouviet</i>	
Cross-Lingual Style Transfer with Conditional Prior VAE and Style Loss	4586
<i>Dino Rattcliff, You Wang, Alex Mansbridge, Penny Karanasou, Alexis Moinet, Marius Cotesco</i>	
Daft-Exprt: Cross-Speaker Prosody Transfer on Any Text for Expressive Speech Synthesis	4591
<i>Julian Zaidi, Hugo Seuté, Benjamin Van Niekerk, Marc-André Carbonneau</i>	
Language Model-Based Emotion Prediction Methods for Emotional Speech Synthesis Systems	4596
<i>Hyun-Wook Yoon, Ohsung Kwon, Hoyeon Lee, Ryuichi Yamamoto, Eunwoo Song, Jae-Min Kim, Min-Jae Hwang</i>	
Text Aware Emotional Text-To-Speech with BERT	4601
<i>Arijit Mukherjee, Shubham Bansal, Sandeepkumar Satpal, Rupesh Mehta</i>	

SPEECH SEGMENTATION I

Overlapped Speech Detection in Broadcast Streams Using X-Vectors	4606
<i>Lukas Mateju, Frantisek Kynych, Petr Cerva, Jiri Malek, Jindrich Zdansky</i>	
DDKtor: Automatic Diadochokinetic Speech Analysis.....	4611
<i>Yael Segal, Kasia Hitczenko, Matt Goldrick, Adam Buchwald, Angela Roberts, Joseph Keshet</i>	
SiDi KWS: A Large-Scale Multilingual Dataset for Keyword Spotting.....	4616
<i>Michel Cardoso Meneses, Rafael Bérnago Holanda, Luis Vasconcelos Peres, Gabriela Dantas Rocha</i>	
Dummy Prototypical Networks for Few-Shot Open-Set Keyword Spotting.....	4621
<i>Byeonggeun Kim, Seunghan Yang, Inseop Chung, Simyung Chang</i>	

Unsupervised Voice Activity Detection by Modeling Source and System Information Using Zero Frequency Filtering	4626
<i>Eklavya Sarkar, Ravishankar Prasad, Mathew Magimai Doss</i>	
Multilingual and Multimodal Abuse Detection	4631
<i>Rini Sharon, Heet Shah, Debdoot Mukherjee, Vikram Gupta</i>	
Microphone Array Channel Combination Algorithms for Overlapped Speech Detection.....	4636
<i>Theo Mariotte, Anthony Larcher, Silvio Montrésor, Jean-Hugh Thomas</i>	
Streaming Automatic Speech Recognition with Re-Blocking Processing Based on Integrated Voice Activity Detection	4641
<i>Yui Sudo, Shakeel Muhammad, Kazuhiro Nakadai, Jiatong Shi, Shinji Watanabe</i>	
Unsupervised Word Segmentation Using K Nearest Neighbors.....	4646
<i>Tzeviya Fuchs, Yedid Hoshen, Yossi Keshet</i>	

HUMAN SPEECH & SIGNAL PROCESSING

Investigation on the Band Importance of Phase-Aware Speech Enhancement.....	4651
<i>Zhuohuang Zhang, Donald Williamson, Yi Shen</i>	
Unsupervised Acoustic-To-Articulatory Inversion with Variable Vocal Tract Anatomy	4656
<i>Yifan Sun, Qinlong Huang, Xihong Wu</i>	
Unsupervised Inference of Physiologically Meaningful Articulatory Trajectories with VocalTractLab	4661
<i>Yifan Sun, Qinlong Huang, Xihong Wu</i>	
Radio2Speech: High Quality Speech Recovery from Radio Frequency Signals.....	4666
<i>Running Zhao, Jiangtao Yu, Tingle Li, Hang Zhao, Edith C. H. Ngai</i>	
Isochronous is Beautiful? Syllabic Event Detection in a Neuro-Inspired Oscillatory Model is Facilitated by Isochrony in Speech.....	4671
<i>Mamady Nabe, Julien Diard, Jean-Luc Schwartz</i>	
An Investigation of Regression-Based Prediction of the Femininity Or Masculinity in Speech of Transgender People	4676
<i>Leon Liebig, Christoph Wagner, Alexander Mainka, Peter Birkholz</i>	
Acoustic to Articulatory Speech Inversion Using Multi-Resolution Spectro-Temporal Representations of Speech Signals.....	4681
<i>Rahil Parikh, Nadee Seneviratne, Ganesh Sivaraman, Shihab Shamma, Carol Espy-Wilson</i>	
Deep Neural Convolutional Matrix Factorization for Articulatory Representation Decomposition	4686
<i>Jiachen Lian, Alan W Black, Louis Goldstein, Gopala Krishna Anumanchipalli</i>	
Vocal-Tract Area Functions with Articulatory Reality for Tract Opening	4691
<i>Zhao Zhang, Ju Zhang, Jianguo Wei, Kiyoshi Honda, Tatsuya Kitamura</i>	

SPEECH EMOTION RECOGNITION II

Coupled Discriminant Subspace Alignment for Cross-Database Speech Emotion Recognition.....	4695
<i>Shaokai Li, Peng Song, Keke Zhao, Wenjing Zhang, Wenming Zheng</i>	

Performance Improvement of Speech Emotion Recognition by Neutral Speech Detection Using Autoencoder and Intermediate Representation.....	4700
<i>Jennifer Santos, Takeshi Yamada, Kenkichi Ishizuka, Taiichi Hashimoto, Shoji Makino</i>	
A Graph Isomorphism Network with Weighted Multiple Aggregators for Speech Emotion Recognition	4705
<i>Ying Hu, Yuwu Tang, Hao Huang, Liang He</i>	
Speech Emotion Recognition Via Generation Using an Attention-Based Variational Recurrent Neural Network	4710
<i>Murchana Baruah, Bonny Banerjee</i>	
Speech Emotion: Investigating Model Representations, Multi-Task Learning and Knowledge Distillation	4715
<i>Vikramjit Mitra, Hsiang-Yun Sherry Chien, Vasudha Kowtha, Joseph Yitan Cheng, Erdrin Azemi</i>	
Multiple Enhancements to LSTM for Learning Emotion-Salient Features in Speech Emotion Recognition	4720
<i>Desheng Hu, Xinhui Hu, Xinkang Xu</i>	
Multi-Level Fusion of Wav2vec 2.0 and BERT for Multimodal Emotion Recognition.....	4725
<i>Zihan Zhao, Yanfeng Wang, Yu Wang</i>	
CTA-RNN: Channel and Temporal-Wise Attention RNN Leveraging Pre-Trained ASR Embeddings for Speech Emotion Recognition.....	4730
<i>Chengxin Chen, Pengyuan Zhang</i>	

VOLUME 7

Complex Paralinguistic Analysis of Speech: Predicting Gender, Emotions and Deception in a Hierarchical Framework	4735
<i>Alena Velichko, Maxim Markitantov, Heysem Kaya, Alexey Karpov</i>	
Interactive Co-Learning with Cross-Modal Transformer for Audio-Visual Emotion Recognition	4740
<i>Akihiko Takashima, Ryo Masumura, Atsushi Ando, Yoshihiro Yamazaki, Mihiro Uchida, Shota Orihashi</i>	
SpeechEQ: Speech Emotion Recognition Based on Multi-Scale Unified Datasets and Multitask Learning	4745
<i>Zuheng Kang, Junqing Peng, Jianzong Wang, Jing Xiao</i>	
Discriminative Feature Representation Based on Cascaded Attention Network with Adversarial Joint Loss for Speech Emotion Recognition	4750
<i>Yang Liu, Haoqin Sun, Wenbo Guan, Yuqi Xia, Zhen Zhao</i>	
Intra-Speaker Phonetic Variation in Read Speech: Comparison with Inter-Speaker Variability in a Controlled Population.....	4755
<i>Nicolas Audibert, Cécile Fougeron</i>	

SPEAKER RECOGNITION AND ANTI-SPOOFING

Training Speaker Recognition Systems with Limited Data.....	4760
<i>Nik Vaessen, David Van Leeuwen</i>	

A Deep One-Class Learning Method for Replay Attack Detection.....	4765
<i>Yijie Lou, Shiliang Pu, Jianfeng Zhou, Xin Qi, Qinbo Dong, Hongwei Zhou</i>	
A Universal Identity Backdoor Attack Against Speaker Verification Based on Siamese Network	4770
<i>Haodong Zhao, Wei Du, Junjie Guo, Gongshen Liu</i>	
A Novel Phoneme-Based Modeling for Text-Independent Speaker Identification.....	4775
<i>Xin Wang, Chuan Xie, Qiang Wu, Huayi Zhan, Ying Wu</i>	
Self-Supervised Speaker Verification Using Dynamic Loss-Gate and Label Correction	4780
<i>Bing Han, Zhengyang Chen, Yanmin Qian</i>	
Learning Lip-Based Audio-Visual Speaker Embeddings with AV-HuBERT	4785
<i>Bowen Shi, Abdelrahman Mohamed, Wei-Ning Hsu</i>	
Acoustic Feature Shuffling Network for Text-Independent Speaker Verification	4790
<i>Jin Li, Xin Fang, Fan Chu, Tian Gao, Yan Song, Rong Li Dai</i>	
Multi-Path GMM-MobileNet Based on Attack Algorithms and Codecs for Synthetic Speech and Deepfake Detection	4795
<i>Yan Wen, Zhenchun Lei, Yingen Yang, Changhong Liu, Minglei Ma</i>	
Adversarial Reweighting for Speaker Verification Fairness.....	4800
<i>Minho Jin, Chelsea Ju, Zeya Chen, Yi Chieh Liu, Jasha Droppo, Andreas Stolcke</i>	
Graph-Based Multi-View Fusion and Local Adaptation: Mitigating Within-Household Confusability for Speaker Identification	4805
<i>Long Chen, Yixiong Meng, Venkatesh Ravichandran, Andreas Stolcke</i>	

MISCELLANEOUS TOPICS IN SPEECH, VOICE AND HEARING DISORDERS

Local Context-Aware Self-Attention for Continuous Sign Language Recognition.....	4810
<i>Ronglai Zuo, Brian Mak</i>	
Disentangled Latent Speech Representation for Automatic Pathological Intelligibility Assessment.....	4815
<i>Tobias Weise, Philipp Klumpp, Andreas Maier, Elmar Nöth, Björn Heismann, Maria Schuster, Seung Hee Yang</i>	
Improving Hypernasality Estimation with Automatic Speech Recognition in Cleft Palate Speech	4820
<i>Kaitao Song, Teng Wan, Bixia Wang, Huiqiang Jiang, Luna Qiu, Jiahang Xu, Liping Jiang, Qun Lou, Yuqing Yang, Dongsheng Li, Xudong Wang, Lili Qiu</i>	
Conformer Based Elderly Speech Recognition System for Alzheimer’s Disease Detection.....	4825
<i>Tianzi Wang, Jiajun Deng, Mengzhe Geng, Zi Ye, Shoukang Hu, Yi Wang, Mingyu Cui, Zengrui Jin, Xunying Liu, Helen Meng</i>	
Revisiting Visuo-Spatial Processing in Individuals with Congenital Amusia	4830
<i>Zixia Fan, Jing Shao, Weigong Pan, Lan Wang</i>	
A User-Friendly Headset for Radar-Based Silent Speech Recognition	4835
<i>Pouriya Amini Digehsara, João Vítor Possamai De Menezes, Christoph Wagner, Michael Bärhold, Petr Schaffer, Dirk Plettemeier, Peter Birkholz</i>	
A Study of Production Error Analysis for Mandarin-Speaking Children with Hearing Impairment.....	4840
<i>Jingwen Cheng, Yuchen Yan, Yingming Gao, Xiaoli Feng, Yannan Wang, Jinsong Zhang</i>	

LOW-RESOURCE ASR DEVELOPMENT II

Incremental Layer-Wise Self-Supervised Learning for Efficient Unsupervised Speech Domain Adaptation on Device.....	4845
<i>Zhouyuan Huo, Dongseong Hwang, Khe Chai Sim, Shefali Garg, Ananya Misra, Nikhil Siddhartha, Trevor Strohman, Francoise Beaufays</i>	
Non-Linear Pairwise Language Mappings for Low-Resource Multilingual Acoustic Model Fusion	4850
<i>Muhammad Umar Farooq, Darshan Adiga Haniya Narayana, Thomas Hain</i>	
The THUEE System Description for the IARPA OpenASR21 Challenge	4855
<i>Jing Zhao, Haoyu Wang, Jinpeng Li, Shuzhou Chai, Guanbo Wang, Guoguo Chen, Wei-Qiang Zhang</i>	
External Text Based Data Augmentation for Low-Resource Speech Recognition in the Constrained Condition of OpenASR21 Challenge	4860
<i>Guolong Zhong, Hongyu Song, Ruoyu Wang, Lei Sun, Diyuan Liu, Jia Pan, Xin Fang, Jun Du, Jie Zhang, Lirong Dai</i>	
Cross-Dialect Lexicon Optimisation for an Endangered Language ASR System: The Case of Irish.....	4865
<i>Liam Lonergan, Mengjie Qian, Neasa Ní Chiaráin, Christer Gobl, Ailbhe Ní Chasaide</i>	
Wav2vec-S: Semi-Supervised Pre-Training for Low-Resource ASR.....	4870
<i>Han Zhu, Li Wang, Gaofeng Cheng, Jindong Wang, Pengyuan Zhang, Yonghong Yan</i>	
Comparison of Unsupervised Learning and Supervised Learning with Noisy Labels for Low-Resource Speech Recognition	4875
<i>Yanick Schraner, Christian Scheller, Michel Plüss, Lukas Neukom, Manfred Vogel</i>	
Using Cross-Model Learnings for the Gram Vaani ASR Challenge 2022.....	4880
<i>Tanvina Patel, Odette Scharenborg</i>	
ASR2K: Speech Recognition for Around 2000 Languages Without Audio	4885
<i>Xinjian Li, Florian Metze, David R. Mortensen, Alan W Black, Shinji Watanabe</i>	
Combining Simple but Novel Data Augmentation Methods for Improving Conformer ASR.....	4890
<i>Ronit Damania, Christopher Homan, Emily Prud'Hommeaux</i>	
OpenASR21: The Second Open Challenge for Automatic Speech Recognition of Low-Resource Languages.....	4895
<i>Kay Peterson, Audrey Tong, Yan Yu</i>	
DRAFT: A Novel Framework to Reduce Domain Shifting in Self-Supervised Learning and Its Application to Children's ASR	4900
<i>Ruchao Fan, Abeer Alwan</i>	
Plugging a Neural Phoneme Recognizer into a Simple Language Model: A Workflow for Low-Resource Setting.....	4905
<i>Séverine Guillaume, Guillaume Wisniewski, Benjamin Galliot, Minh-Châu Nguyễn, Maxime Fily, Guillaume Jacques, Alexis Michaud</i>	

VOICE CONVERSION AND ADAPTATION I

An Evaluation of Three-Stage Voice Conversion Framework for Noisy and Reverberant Conditions	4910
<i>Yeonjong Choi, Chao Xie, Tomoki Toda</i>	

An Overview & Analysis of Sequence-To-Sequence Emotional Voice Conversion	4915
<i>Zijiang Yang, Xin Jing, Andreas Triantafyllopoulos, Meishu Song, Ilhan Aslan, Björn W. Schuller</i>	
Zero-Shot Foreign Accent Conversion Without a Native Reference.....	4920
<i>Waris Quamer, Anurag Das, John Levis, Evgeny Chukharev-Hudilainen, Ricardo Gutierrez-Osuna</i>	
Speaker Anonymization with Phonetic Intermediate Representations	4925
<i>Sarina Meyer, Florian Lux, Pavel Denisov, Julia Koch, Pascal Tilli, Ngoc Thang Vu</i>	
Investigation into Target Speaking Rate Adaptation for Voice Conversion.....	4930
<i>Michael Kuhlmann, Fritz Seebauer, Janek Ebbers, Petra Wagner, Reinhold Haeb-Umbach</i>	
Self Supervised Learning for Robust Voice Cloning.....	4935
<i>Konstantinos Klapsas, Nikolaos Ellinas, Karolos Nikitaras, Georgios Vamvoukakis, Panagiotis Kakoulidis, Konstantinos Markopoulos, Spyros Raptis, June Sig Sung, Gunu Jho, Aimilios Chalamandaris, Pirros Tsiakoulis</i>	

SEARCH/DECODING ALGORITHMS FOR ASR

Improving Deliberation by Text-Only and Semi-Supervised Training	4940
<i>Ke Hu, Tara Sainath, Yanzhang He, Rohit Prabhavalkar, Trevor Strohman, Sepand Mavandadi, Weiran Wang</i>	
K-Wav2vec 2.0: Automatic Speech Recognition Based on Joint Decoding of Graphemes and Syllables	4945
<i>Jounghee Kim, Pilsung Kang</i>	
Wav2Vec-Aug: Improved Self-Supervised Training with Limited Data	4950
<i>Anuroop Sriram, Michael Auli, Alexei Baevski</i>	
Revisiting Joint Decoding Based Multi-Talker Speech Recognition with DNN Acoustic Model.....	4955
<i>Martin Kocour, Katerina Zmolikova, Lucas Ondel, Jan Svec, Marc Delcroix, Tsubasa Ochiai, Lukas Burget, Jan Cernocky</i>	
RNN-T Lattice Enhancement by Grafting of Pruned Paths.....	4960
<i>Mirek Novak, Pavlos Papadopoulos</i>	
Better Intermediates Improve CTC Inference	4965
<i>Tatsuya Komatsu, Yusuke Fujita, Jaesong Lee, Lukas Lee, Shinji Watanabe, Yusuke Kida</i>	

EMOTIONAL SPEECH PRODUCTION AND PERCEPTION

Cross-Cultural Comparison of Gradient Emotion Perception: Human Vs. Alexa TTS Voices	4970
<i>Iona Gessinger, Michelle Cohn, Georgia Zellou, Bernd Möbius</i>	
Discriminative Adversarial Learning for Speaker Independent Emotion Recognition.....	4975
<i>Chamara Kasun, Chung Soo Ahn, Jagath Rajapakse, Zhiping Lin, Guang-Bin Huang</i>	
Representing 'How You Say' with 'What You Say': English Corpus of Focused Speech and Text Reflecting Corresponding Implications.....	4980
<i>Naoaki Suzuki, Satoshi Nakamura</i>	

Production Strategies of Vocal Attitudes	4985
<i>Léane Salais, Pablo Arias, Clément Le Moine, Victor Rosi, Yann Teytaut, Nicolas Obin, Axel Roebel</i>	

Where's the Uh, Hesitation? the Interplay Between Filled Pause Location, Speech Rate and Fundamental Frequency in Perception of Confidence.....	4990
<i>Ambika Kirkland, Harm Lameris, Eva Szekely, Joakim Gustafson</i>	

SPEECH ANALYSIS

E2E Segmenter: Joint Segmenting and Decoding for Long-Form ASR.....	4995
<i>W. Ronny Huang, Shuo-Yiin Chang, David Rybach, Tara Sainath, Rohit Prabhavalkar, Cal Peyser, Zhiyun Lu, Cyril Allauzen</i>	

Autoregressive Co-Training for Learning Discrete Speech Representation.....	5000
<i>Sung-Lin Yeh, Hao Tang</i>	

An Exploration of Prompt Tuning on Generative Spoken Language Model for Speech Processing Tasks.....	5005
<i>Kai-Wei Chang, Wei-Cheng Tseng, Shang-Wen Li, Hung-Yi Lee</i>	

Overlapped Speech and Gender Detection with WavLM Pre-Trained Features.....	5010
<i>Martin Lebourdais, Marie Tahon, Antoine Laurent, Sylvain Meignier</i>	

A Study on Constraining Connectionist Temporal Classification for Temporal Audio Alignment	5015
<i>Yann Teytaut, Baptiste Bouvier, Axel Roebel</i>	

Acoustic-To-Articulatory Speech Inversion with Multi-Task Learning	5020
<i>Yashish M. Siriwardena, Ganesh Sivaraman, Carol Espy-Wilson</i>	

TRUSTWORTHY SPEECH PROCESSING

Enhancing Speech Privacy with Slicing.....	5025
<i>Mohamed Maouche, Brij Mohan Lal Srivastava, Nathalie Vauquier, Aurélien Bellet, Marc Tommasi, Emmanuel Vincent</i>	

An Attention-Based Method for Guiding Attribute-Aligned Speech Representation Learning	5030
<i>Yu-Lin Huang, Bo-Hao Su, Y.-W. Peter Hong, Chi-Chun Lee</i>	

Defense Against Adversarial Attacks on Hybrid Speech Recognition System Using Adversarial Fine-Tuning with Denoiser.....	5035
<i>Sonal Joshi, Saurabh Kataria, Yiwen Shao, Piotr Zelasko, Jesús Villalba, Sanjeev Khudanpur, Najim Dehak</i>	

Membership Inference Attacks Against Self-Supervised Speech Models	5040
<i>Wei-Cheng Tseng, Wei-Tsung Kao, Hung-Yi Lee</i>	

Chunking Defense for Adversarial Attacks on ASR.....	5045
<i>Yiwen Shao, Jesus Villalba, Sonal Joshi, Saurabh Kataria, Sanjeev Khudanpur, Najim Dehak</i>	

Semi-FedSER: Semi-Supervised Learning for Speech Emotion Recognition on Federated Learning Using Multiview Pseudo-Labeling.....	5050
<i>Tiantian Feng, Shrikanth Narayanan</i>	

User-Level Differential Privacy Against Attribute Inference Attack of Speech Emotion Recognition on Federated Learning..... 5055
Tiantian Feng, Raghuvveer Peri, Shrikanth Narayanan

AdvEst: Adversarial Perturbation Estimation to Classify and Detect Adversarial Attacks Against Speaker Identification..... 5060
Sonal Joshi, Saurabh Kataria, Jesús Villalba, Najim Dehak

SPEAKER RECOGNITION AND DIARIZATION

Online Learning of Open-Set Speaker Identification by Active User-Registration..... 5065
Eunkyung Yoo, Hyeonseop Song, Taehyeong Kim, Chul Lee

Automatic Speaker Verification System for Dysarthria Patients..... 5070
Shinimol Salim, Syed Shahnawazuddin, Waquar Ahmad

Multimodal Clustering with Role Induced Constraints for Speaker Diarization..... 5075
Nikolaos Flemotomos, Shrikanth Narayanan

Multi-Scale Speaker Diarization with Dynamic Scale Weighting..... 5080
Tae Jin Park, Nithin Rao Koluguri, Jagadeesh Balam, Boris Ginsburg

Improved Relation Networks for End-To-End Speaker Verification and Identification..... 5085
Ashutosh Chaubey, Sparsh Sinha, Susmita Ghose

End-To-End Neural Speaker Diarization with an Iterative Refinement of Non-Autoregressive Attention-Based Attractors..... 5090
Magdalena Rybicka, Jesus Villalba, Najim Dehak, Konrad Kowalczyk

From Simulated Mixtures to Simulated Conversations as Training Data for End-To-End Neural Diarization..... 5095
Federico Landini, Alicia Lozano-Diez, Mireia Diez, Lukáš Burget

Can Humans Correct Errors from System? Investigating Error Tendencies in Speaker Identification Using Crowdsourcing..... 5100
Yuta Ide, Susumu Saito, Teppei Nakano, Tetsuji Ogawa

Light-Weight Speaker Verification with Global Context Information..... 5105
Miseul Kim, Zhenyu Piao, Seyun Um, Ran Lee, Jaemin Joh, Seungshin Lee, Hong-Goo Kang

Learnable Sparse Filterbank for Speaker Verification.....5110
Junyi Peng, Rongzhi Gu, Ladislav Mošner, Oldrich Plchot, Lukas Burget, Jan Cernocký

SELF-SUPERVISED, SEMI-SUPERVISED, ADAPTATION AND DATA AUGMENTATION FOR ASR II

Using Data Augmentation and Consistency Regularization to Improve Semi-Supervised Speech Recognition.....5115
Ashtosh Sapru

Unsupervised Domain Adaptation for Speech Recognition with Unsupervised Error Correction..... 5120
Long Mai, Julie Carson-Berndsen

A Scalable Model Specialization Framework for Training and Inference Using Submodels and Its Application to Speech Model Personalization.....	5125
<i>Fadi Biadsy, Youzheng Chen, Xia Zhang, Oleg Rybakov, Andrew Rosenberg, Pedro Moreno</i>	
Wav2vec Behind the Scenes: How End2end Models Learn Phonetics	5130
<i>Teena Tom Dieck, Paula Andrea Pérez-Toro, Tomas Arias, Elmar Noeth, Philipp Klumpp</i>	
Scaling ASR Improves Zero and Few Shot Learning.....	5135
<i>Weiyi Zheng, Alex Xiao, Gil Keren, Duc Le, Frank Zhang, Christian Fuegen, Ozlem Kalinli, Yatharth Saraf, Abdelrahman Mohamed</i>	
InterAug: Augmenting Noisy Intermediate Predictions for CTC-Based ASR.....	5140
<i>Yu Nakagome, Tatsuya Komatsu, Yusuke Fujita, Shuta Ichimura, Yusuke Kida</i>	
Investigation of Ensemble Features of Self-Supervised Pretrained Models for Automatic Speech Recognition	5145
<i>A Arunkumar, Vrunda Nileshkumar Sukhadia, Srinivasan Umesh</i>	

SPOKEN LANGUAGE PROCESSING I

Dynamic Sliding Window Modeling for Abstractive Meeting Summarization.....	5150
<i>Zhengyuan Liu, Nancy Chen</i>	
STUDIES: Corpus of Japanese Empathetic Dialogue Speech Towards Friendly Voice Agent	5155
<i>Yuki Saito, Yuto Nishimura, Shinnosuke Takamichi, Kentaro Tachibana, Hiroshi Saruwatari</i>	
KidsTALC: A Corpus of 3- To 11-Year-Old German Children’s Connected Natural Speech	5160
<i>Lars Rumberg, Christopher Gebauer, Hanna Ehlert, Maren Wallbaum, Lena Bornholt, Jörn Ostermann, Ulrike Lüdtke</i>	
DUAL: Discrete Spoken Unit Adaptive Learning for Textless Spoken Question Answering	5165
<i>Guan-Ting Lin, Yung-Sung Chuang, Ho-Lam Chung, Shu-Wen Yang, Hsuan-Jui Chen, Shuyan Annie Dong, Shang-Wen Li, Abdelrahman Mohamed, Hung-Yi Lee, Lin-Shan Lee</i>	
Asymmetric Proxy Loss for Multi-View Acoustic Word Embeddings.....	5170
<i>Myunghun Jung, Hoi Rin Kim</i>	
Exploring Continuous Integrate-And-Fire for Adaptive Simultaneous Speech Translation.....	5175
<i>Chih-Chiang Chang, Hung-Yi Lee</i>	
Building Vietnamese Conversational Smart Home Dataset and Natural Language Understanding Model	5180
<i>Thi Thu Trang Nguyen, Trung Duc Anh Dang, Quoc Viet Vu, Woomyoung Park</i>	
DeToxy: A Large-Scale Multimodal Dataset for Toxicity Classification in Spoken Utterances	5185
<i>Sreyana Ghosh, Samden Lepcha, S Sakshi, Rajiv Ratn Shah, Srinivasan Umesh</i>	
Voice Activity Projection: Self-Supervised Learning of Turn-Taking Events	5190
<i>Erik Ekstedt, Gabriel Skantze</i>	
Enhanced Direct Speech-To-Speech Translation Using Self-Supervised Pre-Training and Data Augmentation	5195
<i>Sravya Popuri, Peng-Jen Chen, Changhan Wang, Juan Pino, Yossi Adi, Jiatao Gu, Wei-Ning Hsu, Ann Lee</i>	

QbyE-MLPMixer: Query-By-Example Open-Vocabulary Keyword Spotting Using MLPMixer.....	5200
<i>Jinmiao Huang, Waseem Gharbieh, Qianhui Wan, Han Suk Shim, Hyun Chul Lee</i>	
DyConvMixer: Dynamic Convolution Mixer Architecture for Open-Vocabulary Keyword Spotting	5205
<i>Waseem Gharbieh, Jinmiao Huang, Qianhui Wan, Han Suk Shim, Hyun Chul Lee</i>	
Challenges in Metadata Creation for Massive Naturalistic Team-Based Audio Data	5210
<i>Chelzy Belitz, John H. L. Hansen</i>	

SHOW AND TELL IV

Spoken Dialogue System for Call Centers with Expressive Speech Synthesis	5215
<i>Davis Nicmanis, Askars Salimbajevs</i>	
OCTRA – an Innovative Approach to Orthographic Transcription	5217
<i>Christoph Draxler, Julian Pomp</i>	
Voice Puppetry with FastPitch.....	5219
<i>Emelie Van De Vreken, Korin Richmond, Catherine Lai</i>	
Improving Data Driven Inverse Text Normalization Using Data Augmentation and Machine Translation.....	5221
<i>Debjyoti Paul, Yutong Pang, Szu-Jui Chen, Xuedong Zhang</i>	

PHONETICS II

Native Phonotactic Interference in L2 Vowel Processing: Mouse-Tracking Reveals Cognitive Conflicts During Identification.....	5223
<i>Yizhou Wang, Rikke Bundgaard-Nielsen, Brett Baker, Olga Maxwell</i>	
Mandarin Nasal Place Assimilation Revisited: An Acoustic Study.....	5228
<i>Mingqiong Luo</i>	
Bending the String: Intonation Contour Length as a Correlate of Macro-Rhythm.....	5233
<i>Constantijn Kaland</i>	
Eliciting and Evaluating Likelihood Ratios for Speaker Recognition by Human Listeners Under Forensically Realistic Channel-Mismatched Conditions.....	5238
<i>Vincent Hughes, Carmen Llamas, Thomas Kettig</i>	
Reducing Uncertainty at the score-To-LR Stage in Likelihood Ratio-Based Forensic Voice Comparison Using Automatic Speaker Recognition Systems	5243
<i>Bruce Xiao Wang, Vincent Hughes</i>	
Durational Patterning at Discourse Boundaries in Relation to Therapist Empathy in Psychotherapy	5248
<i>Jonathan Him Nok Lee, Dehua Tao, Harold Chui, Tan Lee, Sarah Luk, Nicolette Wing Tung Lee, Koonkan Fung</i>	
Convolutional Neural Networks for Classification of Voice Qualities from Speech and Neck Surface Accelerometer Signals.....	5253
<i>Sudarsana Reddy Kadiri, Farhad Javanmardi, Paavo Alku</i>	
Applying Syntax–Prosody Mapping Hypothesis and Prosodic Well-Formedness Constraints to Neural Sequence-To-Sequence Speech Synthesis	5258
<i>Kei Furukawa, Takeshi Kishiyama, Satoshi Nakamura</i>	

Effects of Language Contact on Vowel Nasalization in Wenzhou and Rugao Dialects.....	5263
<i>Yan Li, Ying Chen, Xinya Zhang, Yanyang Chen, Jiazheng Wang</i>	
A Blueprint for Using Deepfakes in Sociolinguistic Matched-Guise Experiments.....	5268
<i>Nathan Joel Young, David Britain, Adrian Leemann</i>	
Mandarin Tone Sandhi Realization: Evidence from Large Speech Corpora	5273
<i>Zuoyu Tian, Xiao Dong, Feier Gao, Haining Wang, Charles Lin</i>	
A Laryngographic Study on the Voice Quality of Northern Vietnamese Tones Under the Lombard Effect.....	5278
<i>Giang Le, Chilin Shih, Yan Tang</i>	
The Prosody of Cheering in Sport Events	5283
<i>Marzena Zygis, Sarah Wesolek, Nina Hosseini-Kivanani, Manfred Krifka</i>	
Contribution of the Glottal Flow Residual in Affect-Related Voice Transformation.....	5288
<i>Zihan Wang, Christer Gobl</i>	
High Level Feature Fusion in Forensic Voice Comparison	5293
<i>Michael Carne, Yuko Kinoshita, Shunichi Ishihara</i>	
Modeling Speech Recognition and Synthesis Simultaneously: Encoding and Decoding Lexical and Sublexical Semantic Information into Speech with No Direct Access to Speech Data.....	5298
<i>Gasper Begus, Alan Zhou</i>	
Paraguayan Guarani: Tritonal Pitch Accent and Accentual Phrase.....	5303
<i>Sun-Ah Jun, Maria Luisa Zubizarreta</i>	
Low-Resource Accent Classification in Geographically-Proximate Settings: A Forensic and Sociophonetics Perspective	5308
<i>Qingcheng Zeng, Dading Chong, Peilin Zhou, Jie Yang</i>	
 <u>SOURCE SEPARATION III</u> 	
Tiny-Sepformer: A Tiny Time-Domain Transformer Network for Speech Separation.....	5313
<i>Jian Luo, Jianzong Wang, Ning Cheng, Edward Xiao, Xulong Zhang, Jing Xiao</i>	
Speaker-Aware Mixture of Mixtures Training for Weakly Supervised Speaker Extraction	5318
<i>Zifeng Zhao, Rongzhi Gu, Dongchao Yang, Jinchuan Tian, Yuexian Zou</i>	
SepIt: Approaching a Single Channel Speech Separation Bound.....	5323
<i>Shahar Lutati, Eliya Nachmani, Lior Wolf</i>	
On the Use of Deep Mask Estimation Module for Neural Source Separation Systems	5328
<i>Kai Li, Xiaolin Hu, Yi Luo</i>	
Target Confusion in End-To-End Speaker Extraction: Analysis and Approaches	5333
<i>Zifeng Zhao, Dongchao Yang, Rongzhi Gu, Haoran Zhang, Yuexian Zou</i>	
Embedding Recurrent Layers with Dual-Path Strategy in a Variant of Convolutional Network for Speaker-Independent Speech Separation.....	5338
<i>Xue Yang, Changchun Bao</i>	
Disentangling the Impacts of Language and Channel Variability on Speech Separation Networks.....	5343
<i>Fan-Lin Wang, Hung-Shin Lee, Yu Tsao, Hsin-Min Wang</i>	

Objective Metrics to Evaluate Residual-Echo Suppression During Double-Talk in the Stereophonic Case	5348
<i>Amir Ivry, Israel Cohen, Baruch Berdugo</i>	
QDPN - Quasi-Dual-Path Network for Single-Channel Speech Separation	5353
<i>Joel Rixen, Matthias Renz</i>	
Conformer Space Neural Architecture Search for Multi-Task Audio Separation.....	5358
<i>Shun Lu, Yang Wang, Peng Yao, Chenxing Li, Jianchao Tan, Feng Deng, Xiaorui Wang, Chengru Song</i>	
ResectNet: An Efficient Architecture for Voice Activity Detection on Mobile Devices	5363
<i>Okan Köpüklü, Maja Taseska</i>	
Gated Convolutional Fusion for Time-Domain Target Speaker Extraction Network.....	5368
<i>Wenjing Liu, Chuan Xie</i>	
WA-Transformer: Window Attention-Based Transformer with Two-Stage Strategy for Multi-Task Audio Source Separation.....	5373
<i>Yang Wang, Chenxing Li, Feng Deng, Shun Lu, Peng Yao, Jianchao Tan, Chengru Song, Xiaorui Wang</i>	
Multichannel Speech Separation with Narrow-Band Conformer	5378
<i>Changsheng Quan, Xiaofei Li</i>	
Separating Long-Form Speech with Group-Wise Permutation Invariant Training	5383
<i>Wangyou Zhang, Zhuo Chen, Naoyuki Kanda, Shujie Liu, Jinyu Li, Sefik Emre Eskimez, Takuya Yoshioka, Xiong Xiao, Zhong Meng, Yanmin Qian, Furu Wei</i>	
Directed Speech Separation for Automatic Speech Recognition of Long Form Conversational Speech	5388
<i>Rohit Paturi, Sundararajan Srinivasan, Katrin Kirchhoff, Daniel Garcia-Romero</i>	
Speech Separation for an Unknown Number of Speakers Using Transformers with Encoder-Decoder Attractors.....	5393
<i>Srikanth Raj Chetupalli, Emanuël Habets</i>	
Cooperative Speech Separation with a Microphone Array and Asynchronous Wearable Devices.....	5398
<i>Ryan Corey, Manan Mittal, Kanad Sarkar, Andrew C. Singer</i>	
Text-Driven Separation of Arbitrary Sounds	5403
<i>Kevin Kilgour, Beat Gfeller, Qingqing Huang, Aren Jansen, Scott Wisdom, Marco Tagliasacchi</i>	
An Empirical Analysis on the Vulnerabilities of End-To-End Speech Segregation Models	5408
<i>Rahil Parikh, Gaspar Rochette, Carol Espy-Wilson, Shihab Shamma</i>	

SPEECH ENHANCEMENT AND INTELLIGIBILITY

TaylorBeamformer: Learning All-Neural Beamformer for Multi-Channel Speech Enhancement from Taylor's Approximation Theory.....	5413
<i>Andong Li, Guochen Yu, Chengshi Zheng, Xiaodong Li</i>	
How Bad Are Artifacts?: Analyzing the Impact of Speech Enhancement Errors on ASR.....	5418
<i>Kazuma Iwamoto, Tsubasa Ochiai, Marc Delcroix, Rintaro Ikeshita, Hiroshi Sato, Shoko Araki, Shigeru Katagiri</i>	

Multi-Source Wideband DOA Estimation Method by Frequency Focusing and Error Weighting	5423
<i>Jing Zhou, Changchun Bao</i>	
Convolutional Recurrent Smart Speech Enhancement Architecture for Hearing Aids.....	5428
<i>Soha Nossier, Julie Wall, Mansour Moniri, Cornelius Glackin, Nigel Cannings</i>	
Fully Automatic Balance Between Directivity Factor and White Noise Gain for Large-Scale Microphone Arrays in Diffuse Noise Fields	5433
<i>Weixin Meng, Chengshi Zheng, Xiaodong Li</i>	
A Transfer and Multi-Task Learning Based Approach for MOS Prediction.....	5438
<i>Xiaohai Tian, Kaiqi Fu, Shaojun Gao, Yiwei Gu, Kai Wang, Wei Li, Zejun Ma</i>	
Fusion of Self-Supervised Learned Models for MOS Prediction	5443
<i>Zhengdong Yang, Wangjin Zhou, Chenhui Chu, Sheng Li, Raj Dabre, Raphael Rubino, Yi Zhao</i>	
Perceptual Contrast Stretching on Target Feature for Speech Enhancement.....	5448
<i>Rong Chao, Cheng Yu, Szu-Wei Fu, Xugang Lu, Yu Tsao</i>	
A Speech Enhancement Method for Long-Range Speech Acquisition Task	5453
<i>Yanzhang Geng, Heng Wang, Tao Zhang, Xin Zhao</i>	
ESPnet-SE++: Speech Enhancement for Robust Speech Recognition, Translation, and Understanding	5458
<i>Yen-Ju Lu, Xuankai Chang, Chenda Li, Wangyou Zhang, Samuele Cornell, Zhaoheng Ni, Yoshiki Masuyama, Brian Yan, Robin Scheibler, Zhong-Qiu Wang, Yu Tsao, Yanmin Qian, Shinji Watanabe</i>	
MTI-Net: A Multi-Target Speech Intelligibility Prediction Model.....	5463
<i>Ryandhimas Edo Zezario, Szu-Wei Fu, Fei Chen, Chiou-Shann Fuh, Hsin-Min Wang, Yu Tsao</i>	
Steering Vector Correction in MVDR Beamformer for Speech Enhancement.....	5468
<i>Suliang Bu, Yunxin Zhao, Tuo Zhao</i>	
Speech Modification for Intelligibility in Cochlear Implant Listeners: Individual Effects of Vowel- And Consonant-Boosting	5473
<i>Juliana N. Saba, John H. L. Hansen</i>	
DCTCN:Deep Complex Temporal Convolutional Network for Long Time Speech Enhancement.....	5478
<i>Ren Jigang, Mao Qirong</i>	
Improve Speech Enhancement Using Perception-High-Related Time-Frequency Loss	5483
<i>Ding Zhao, Zhan Zhang, Bin Yu, Yuehai Wang</i>	

SPEECH SYNTHESIS: SPEAKING STYLE, EMOTION AND ACCENTS II

Transplantation of Conversational Speaking Style with Interjections in Sequence-To-Sequence Speech Synthesis	5488
<i>Raul Fernandez, David Haws, Guy Lorberbom, Slava Shechtman, Alexander Sorin</i>	
Accurate Emotion Strength Assessment for Seen and Unseen Speech Based on Data-Driven Deep Learning	5493
<i>Rui Liu, Berrak Sisman, Björn Schuller, Guanglai Gao, Haizhou Li</i>	

Cross-Speaker Emotion Transfer Based on Prosody Compensation for End-To-End Speech Synthesis.....	5498
<i>Tao Li, Xinsheng Wang, Qicong Xie, Zhichao Wang, Mingqi Jiang, Lei Xie</i>	
Self-Supervised Context-Aware Style Representation for Expressive Speech Synthesis.....	5503
<i>Yihan Wu, Xi Wang, Shaofei Zhang, Lei He, Ruihua Song, Jian-Yun Nie</i>	
Integrating Discrete Word-Level Style Variations into Non-Autoregressive Acoustic Models for Speech Synthesis	5508
<i>Zhaoci Liu, Ningqian Wu, Yajie Zhang, Zhenhua Ling</i>	
Automatic Prosody Annotation with Pre-Trained Text-Speech Model.....	5513
<i>Ziqian Dai, Jianwei Yu, Yan Wang, Nuo Chen, Yanyao Bian, Guangzhi Li, Deng Cai, Dong Yu</i>	
Enhancing Word-Level Semantic Representation Via Dependency Structure for Expressive Text-To-Speech Synthesis.....	5518
<i>Yixuan Zhou, Changhe Song, Jingbei Li, Zhiyong Wu, Yanyao Bian, Dan Su, Helen Meng</i>	
Towards Multi-Scale Speaking Style Modelling with Hierarchical Context Information for Mandarin Speech Synthesis.....	5523
<i>Shun Lei, Yixuan Zhou, Liyang Chen, Jiankun Hu, Zhiyong Wu, Shiyin Kang, Helen Meng</i>	
Towards Cross-Speaker Reading Style Transfer on Audiobook Dataset.....	5528
<i>Xiang Li, Changhe Song, Xianhao Wei, Zhiyong Wu, Jia Jia, Helen Meng</i>	
CALM: Constrastive Cross-Modal Speaking Style Modeling for Expressive Text-To-Speech Synthesis.....	5533
<i>Yi Meng, Xiang Li, Zhiyong Wu, Tingtian Li, Zixun Sun, Xinyu Xiao, Chi Sun, Hui Zhan, Helen Meng</i>	
Improve Emotional Speech Synthesis Quality by Learning Explicit and Implicit Representations with Semi-Supervised Training	5538
<i>Jiaxu He, Cheng Gong, Longbiao Wang, Di Jin, Xiaobao Wang, Junhai Xu, Jianwu Dang</i>	

SHOW & TELL IV(VR)

A Vietnamese-English Neural Machine Translation System	5543
<i>Tuan-Duy H. Nguyen, Duy Phung, Duy Tran-Cong Nguyen, Hieu Minh Tran, Manh Luong, Tin Duy Vo, Hung Hai Bui, Dinh Phung, Dat Quoc Nguyen</i>	

Author Index