

22nd Annual Conference of the International Speech Communication Association (INTERSPEECH 2021)

Brno, Czech Republic
30 August – 3 September 2021

Volume 1 of 6

ISBN: 978-1-7138-3690-2

Printed from e-media with permission by:

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



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

Copyright© (2021) by International Speech Communication Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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

INTERSPEECH 2021 : ACOUSTIC EVENT DETECTION AND ACOUSTIC SCENE CLASSIFICATION

OPTIMIZING LATENCY FOR ONLINE VIDEO CAPTIONING USING AUDIO-VISUAL TRANSFORMERS	1
<i>Chiori Hori, Takaaki Hori, Jonathan Le Roux</i>	
SPECMIX : A MIXED SAMPLE DATA AUGMENTATION METHOD FOR TRAINING WITH TIME-FREQUENCY DOMAIN FEATURES.....	6
<i>Gwantae Kim, David K. Han, Hanseok Ko</i>	
EVENT SPECIFIC ATTENTION FOR POLYPHONIC SOUND EVENT DETECTION.....	11
<i>Harshavardhan Sundar, Ming Sun, Chao Wang</i>	
AN EVALUATION OF DATA AUGMENTATION METHODS FOR SOUND SCENE GEOTAGGING.....	16
<i>Helen L. Bear, Veronica Morfi, Emmanouil Benetos</i>	
SPECAUGMENT++: A HIDDEN SPACE DATA AUGMENTATION METHOD FOR ACOUSTIC SCENE CLASSIFICATION	21
<i>Helin Wang, Yuexian Zou, Wenwu Wang</i>	
ACOUSTIC SCENE CLASSIFICATION USING KERVLUTION-BASED SUBSPECTRALNET.....	26
<i>Ritika Nandi, Shashank Shekhar, Manjunath Mulimani</i>	
VARIATIONAL INFORMATION BOTTLENECK FOR EFFECTIVE LOW-RESOURCE AUDIO CLASSIFICATION	31
<i>Shijing Si, Jianzong Wang, Huiming Sun, Jianhan Wu, Chuanyao Zhang, Xiaoyang Qu, Ning Cheng, Lei Chen, Jing Xiao</i>	
IMPROVING WEAKLY SUPERVISED SOUND EVENT DETECTION WITH SELF-SUPERVISED AUXILIARY TASKS	36
<i>Soham Deshmukh, Bhiksha Raj, Rita Singh</i>	
SHALLOW CONVOLUTION-AUGMENTED TRANSFORMER WITH DIFFERENTIABLE NEURAL COMPUTER FOR LOW-COMPLEXITY CLASSIFICATION OF VARIABLE-LENGTH ACOUSTIC SCENE.....	41
<i>Soonshin Seo, Donghyun Lee, Ji-Hwan Kim</i>	
ACOUSTIC EVENT DETECTION WITH CLASSIFIER CHAINS	46
<i>Tatsuya Komatsu, Shinji Watanabe, Koichi Miyazaki, Tomoki Hayashi</i>	
AN EFFECTIVE MUTUAL MEAN TEACHING BASED DOMAIN ADAPTATION METHOD FOR SOUND EVENT DETECTION	51
<i>Xu Zheng, Yan Song, Li-Rong Dai, Ian McLoughlin, Lin Liu</i>	
AST: AUDIO SPECTROGRAM TRANSFORMER.....	56
<i>Yuan Gong, Yu-An Chung, James Glass</i>	

INTERSPEECH 2021 : APPLICATIONS IN TRANSCRIPTION, EDUCATION AND LEARNING

DEEP FEATURE TRANSFER LEARNING FOR AUTOMATIC PRONUNCIATION ASSESSMENT	61
<i>Binghuai Lin, Liyuan Wang</i>	
WEAKLY-SUPERVISED WORD-LEVEL PRONUNCIATION ERROR DETECTION IN NON-NATIVE ENGLISH SPEECH	66
<i>Daniel Korzekwa, Jaime Lorenzo-Trueba, Thomas Drugman, Shira Calamaro, Bozena Kostek</i>	
UNDERSTANDING MEDICAL CONVERSATIONS: RICH TRANSCRIPTION, CONFIDENCE SCORES & INFORMATION EXTRACTION.....	71
<i>Hagen Soltau, Mingqiu Wang, Izhak Shafran, Laurent El Shafey</i>	
MULTILINGUAL SPEECH EVALUATION: CASE STUDIES ON ENGLISH, MALAY AND TAMIL	76
<i>Huayun Zhang, Ke Shi, Nancy F. Chen</i>	
PHONE-LEVEL PRONUNCIATION SCORING FOR SPANISH SPEAKERS LEARNING ENGLISH USING A GOP-DNN SYSTEM	81
<i>Jazmin Vidal, Cynthia Bonomi, Marcelo Sancinetti, Luciana Ferrer</i>	
A STUDY ON FINE-TUNING WAV2VEC2.0 MODEL FOR THE TASK OF MISPRONUNCIATION DETECTION AND DIAGNOSIS	86
<i>Linkai Peng, Kaiqi Fu, Binghuai Lin, Dengfeng Ke, Jinsong Zhan</i>	
END-TO-END SPEAKER-ATTRIBUTED ASR WITH TRANSFORMER	91
<i>Naoyuki Kanda, Guoli Ye, Yashesh Gaur, Xiaofei Wang, Zhong Meng, Zhuo Chen, Takuya Yoshioka</i>	
“YOU DON'T UNDERSTAND ME!": COMPARING ASR RESULTS FOR L1 AND L2 SPEAKERS OF SWEDISH	96
<i>Ronald Cumbal, Birger Moell, Jose Lopes, Olov Engwall</i>	
IMPROVEMENT OF AUTOMATIC ENGLISH PRONUNCIATION ASSESSMENT WITH SMALL NUMBER OF UTTERANCES USING SENTENCE SPEAKABILITY	101
<i>Satsuki Naijo, Akinori Ito, Takashi Nose</i>	
LEXICAL DENSITY ANALYSIS OF WORD PRODUCTIONS IN JAPANESE ENGLISH USING ACOUSTIC WORD EMBEDDINGS.....	106
<i>Shintaro Ando, Nobuaki Minematsu, Daisuke Saito</i>	
END-TO-END RICH TRANSCRIPTION-STYLE AUTOMATIC SPEECH RECOGNITION WITH SEMI-SUPERVISED LEARNING	111
<i>Tomohiro Tanaka, Ryo Masumura, Mana Ihori, Akihiko Takashima, Shota Orihashi, Naoki Makishima</i>	
EXPLORE WAV2VEC 2.0 FOR MISPRONUNCIATION DETECTION	116
<i>Xiaoshuo Xu, Yueteng Kang, Songjun Cao, Binghuai Lin, Long Ma</i>	
NEMO INVERSE TEXT NORMALIZATION: FROM DEVELOPMENT TO PRODUCTION	121
<i>Yang Zhang, Evelina Bakhturina, Kyle Gorman, Boris Ginsburg</i>	

THE IMPACT OF ASR ON THE AUTOMATIC ANALYSIS OF LINGUISTIC COMPLEXITY AND SOPHISTICATION IN SPONTANEOUS L2 SPEECH.....	126
<i>Yu Qiao, Wei Zhou, Elma Kerz, Ralf Schluter</i>	

INTERSPEECH 2021 : ASR TECHNOLOGIES AND SYSTEMS

FEARLESS STEPS CHALLENGE PHASE-3 (FSC P3): ADVANCING SLT FOR UNSEEN CHANNEL AND MISSION DATA ACROSS NASA APOLLO AUDIO	131
<i>Aditya Joglekar, Seyed Omid Sadjadi, Meena Chandra-Shekar, Christopher Cieri, John H. L. Hansen</i>	
NEURAL TEXT DENORMALIZATION FOR SPEECH TRANSCRIPTS	136
<i>Benjamin Suter, Josef Novak</i>	
ALIGNED CONTRASTIVE PREDICTIVE CODING	141
<i>Jan Chorowski, Grzegorz Ciesielski, Jaroslaw Dzikowski, Adrian Lancucki, Ricard Marxer, Mateusz Opala, Piotr Pusz, Pawel Rychlikowski, Michal Stypulkowski</i>	
INFORMATION RETRIEVAL FOR ZEROSPEECH 2021: THE SUBMISSION BY UNIVERSITY OF WROCLAW	146
<i>Jan Chorowski, Grzegorz Ciesielski, Jaroslaw Dzikowski, Adrian Lancucki, Ricard Marxer, Mateusz Opala, Piotr Pusz, Pawel Rychlikowski, Michal Stypulkowski</i>	

INTERSPEECH 2021 : ASSESSMENT OF PATHOLOGICAL SPEECH AND LANGUAGE I

NEURAL SPEAKER EMBEDDINGS FOR ULTRASOUND-BASED SILENT SPEECH INTERFACES.....	151
<i>Amin Honarmandi Shandiz, Laszlo Toth, Gabor Gosztolya, Alexandra Marko, Tamas Gabor Csapo</i>	
AUTOMATICALLY DETECTING ERRORS AND DISFLUENCIES IN READ SPEECH TO PREDICT COGNITIVE IMPAIRMENT IN PEOPLE WITH PARKINSON'S DISEASE.....	156
<i>Amrit Romana, John Bandon, Matthew Perez, Stephanie Gutierrez, Richard Richter, Angela Roberts, Emily Mower Provost</i>	
LATE FUSION OF THE AVAILABLE LEXICON AND RAW WAVEFORM-BASED ACOUSTIC MODELING FOR DEPRESSION AND DEMENTIA RECOGNITION.....	161
<i>Esau Villatoro-Tello, S. Pavankumar Dubagunta, Julian Fritsch, Gabriela Ramirez-De-La-Rosa, Petr Motliceck, Mathew Magimai-Doss</i>	
SPEECH DISORDER CLASSIFICATION USING EXTENDED FACTORIZED HIERARCHICAL VARIATIONAL AUTO-ENCODERS.....	166
<i>Jinzi Qi, Hugo Van Hamme</i>	
AUTOMATIC EXTRACTION OF SPEECH RHYTHM DESCRIPTORS FOR SPEECH INTELLIGIBILITY ASSESSMENT IN THE CONTEXT OF HEAD AND NECK CANCERS	171
<i>Robin Vaysse, Jerome Farinas, Corine Astesano, Regine Andre-Obrecht</i>	
THE IMPACT OF FORCED-ALIGNMENT ERRORS ON AUTOMATIC PRONUNCIATION EVALUATION	176
<i>Vikram C. Mathad, Tristan J. Mahr, Nancy Scherer, Kathy Chapman, Katherine C. Hustad, Julie Liss, Visar Berisha</i>	

INTERSPEECH 2021 : ASSESSMENT OF PATHOLOGICAL SPEECH AND LANGUAGE II

ASSESSING POSTERIOR-BASED MISPRONUNCIATION DETECTION ON FIELD-COLLECTED RECORDINGS FROM CHILD SPEECH THERAPY SESSIONS.....	181
<i>Adam Hair, Guanlong Zhao, Beena Ahmed, Kirrie J. Ballard, Ricardo Gutierrez-Osuna</i>	
IDENTIFYING COGNITIVE IMPAIRMENT USING SENTENCE REPRESENTATION VECTORS.....	186
<i>Bahman Mirheidari, Yilin Pan, Daniel Blackburn, Ronan O'Malley, Heidi Christensen</i>	
PHONETIC COMPLEXITY, SPEECH ACCURACY AND INTELLIGIBILITY ASSESSMENT OF ITALIAN DYSARTHIC SPEECH.....	191
<i>Barbara Gili Fivela, Vincenzo Sallustio, Silvia Pedo, Danilo Patrocínio</i>	
UNSUPERVISED DOMAIN ADAPTATION FOR DYSARTHIC SPEECH DETECTION VIA DOMAIN ADVERSARIAL TRAINING AND MUTUAL INFORMATION MINIMIZATION	196
<i>Disong Wang, Liqun Deng, Yu Ting Yeung, Xiao Chen, Xunying Liu, Helen Meng</i>	
CLAC: A SPEECH CORPUS OF HEALTHY ENGLISH SPEAKERS	201
<i>R'Mani Haulcy, James Glass</i>	
DETECTION OF CONSONANT ERRORS IN DISORDERED SPEECH BASED ON CONSONANT-VOWEL SEGMENT EMBEDDING	206
<i>Si-Ioi Ng, Cymie Wing-Yee Ng, Jingyu Li, Tan Lee</i>	
SOURCE AND VOCAL TRACT CUES FOR SPEECH-BASED CLASSIFICATION OF PATIENTS WITH PARKINSON'S DISEASE AND HEALTHY SUBJECTS.....	211
<i>Tanuka Bhattacharjee, Jhansi Mallela, Yamini Belur, Nalini Atchayaram, Ravi Yadav, Pradeep Reddy, Dipanjan Gope, Prasanta Kumar Ghosh</i>	
UNCERTAINTY-AWARE COVID-19 DETECTION FROM IMBALANCED SOUND DATA	216
<i>Tong Xia, Jing Han, Lorena Qendro, Ting Dang, Cecilia Mascolo</i>	
VOCALIZATION RECOGNITION OF PEOPLE WITH PROFOUND INTELLECTUAL AND MULTIPLE DISABILITIES (PIMD) USING MACHINE LEARNING ALGORITHMS	221
<i>Waldemar Jesko</i>	
SPEECH INTELLIGIBILITY OF DYSARTHIC SPEECH: HUMAN SCORES AND ACOUSTIC-PHONETIC FEATURES	226
<i>Wei Xue, Roeland Van Hout, Fleur Boogmans, Mario Ganzeboom, Catia Cucchiarini, Helmer Strik</i>	
ANALYZING SHORT TERM DYNAMIC SPEECH FEATURES FOR UNDERSTANDING BEHAVIORAL TRAITS OF CHILDREN WITH AUTISM SPECTRUM DISORDER.....	231
<i>Young-Kyung Kim, Rimita Lahiri, Md. Nasir, So Hyun Kim, Somer Bishop, Catherine Lord, Shrikanth S. Narayanan</i>	
PARENTAL SPOKEN SCAFFOLDING AND NARRATIVE SKILLS IN CROWD-SOURCED STORYTELLING SAMPLES OF YOUNG CHILDREN.....	236
<i>Zhengjun Yue, Jon Barker, Heidi Christensen, Cristina McKean, Elaine Ashton, Yvonne Wren, Swapnil Gadgil, Rebecca Bright</i>	

INTERSPEECH 2021 : AUTOMATIC SPEECH RECOGNITION IN AIR TRAFFIC MANAGEMENT

MODELING THE EFFECT OF MILITARY OXYGEN MASKS ON SPEECH CHARACTERISTICS.....	241
<i>Benjamin Elie, Jodie Gauvain, Jean-Luc Gauvain, Lori Lamel</i>	
DETECTING ENGLISH SPEECH IN THE AIR TRAFFIC CONTROL VOICE COMMUNICATION	246
<i>Igor Szoke, Santosh Kesiraju, Ondrej Novotny, Martin Kocour, Karel Vesely, Jan Cernocky</i>	
CONTEXTUAL SEMI-SUPERVISED LEARNING: AN APPROACH TO LEVERAGE AIR-SURVEILLANCE AND UNTRANSCRIBED ATC DATA IN ASR SYSTEMS	251
<i>Juan Zuluaga-Gomez, Iuliia Nigmatulina, Amrutha Prasad, Petr Motlicek, Karel Vesely, Martin Kocour, Igor Szoke</i>	
BOOSTING OF CONTEXTUAL INFORMATION IN ASR FOR AIR-TRAFFIC CALL-SIGN RECOGNITION.....	256
<i>Martin Kocour, Karel Vesely, Alexander Blatt, Juan Zuluaga Gomez, Igor Szoke, Jan Cernocky, Dietrich Klakow, Petr Motlicek</i>	
TOWARDS AN ACCENT-ROBUST APPROACH FOR ATC COMMUNICATIONS TRANSCRIPTION	261
<i>Nataly Jahchan, Florentin Barbier, Ariyanidevi Dharma Gita, Khaled Khelif, Estelle Delpech</i>	
ROBUST COMMAND RECOGNITION FOR LITHUANIAN AIR TRAFFIC CONTROL TOWER UTTERANCES.....	266
<i>Oliver Ohneiser, Seyyed Saeed Sarfjoo, Hartmut Helmke, Shruthi Shetty, Petr Motlicek, Matthias Kleinert, Heiko Ehr, Sarunas Murauskas</i>	

INTERSPEECH 2021 : COMMUNICATION AND INTERACTION, MULTIMODALITY

EFFECTS OF VOICE TYPE AND TASK ON L2 LEARNERS' AWARENESS OF PRONUNCIATION ERRORS.....	271
<i>Alif Silpachai, Ivana Rehman, Taylor Anne Barriuso, John Levis, Evgeny Chukharev-Hudilainen, Guanlong Zhao, Ricardo Gutierrez-Osuna</i>	
LEXICAL ENTRAINMENT AND INTRA-SPEAKER VARIABILITY IN COOPERATIVE DIALOGUES.....	276
<i>Alla Menshikova, Daniil Kocharov, Tatiana Kachkovskaia</i>	
ANALYSIS OF EYE GAZE REASONS AND GAZE AVERSIONS DURING THREE-PARTY CONVERSATIONS.....	281
<i>Carlos Toshinori Ishi, Taiken Shintani</i>	
A PSYCHOLOGY-DRIVEN COMPUTATIONAL ANALYSIS OF POLITICAL INTERVIEWS	286
<i>Darren Cook, Miri Zilka, Simon Maskell, Laurence Alison</i>	
INVESTIGATING THE INTERPLAY BETWEEN AFFECTIVE, PHONATORY AND MOTORIC SUBSYSTEMS IN AUTISM SPECTRUM DISORDER USING A MULTIMODAL DIALOGUE AGENT.....	291
<i>Hardik Kothare, Vikram Ramanarayanan, Oliver Roesler, Michael Neumann, Jackson Liscombe, William Burke, Andrew Cornish, Doug Habberstad, Alaa Sakallah, Sara Markuson, Seemran Kansara, Afik Faerman, Yasmine Bensidi-Slimane, Laura Fry, Saige Portera, David Suendermann-Oeft, David Pautler, Carly Demopoulos</i>	

CROSS-MODAL LEARNING FOR AUDIO-VISUAL VIDEO PARSING 296
Jatin Lamba, Abhishek, Jayaprakash Akula, Rishabh Dabral, Preethi Jyothi, Ganesh Ramakrishnan

SPEECH EMOTION RECOGNITION BASED ON ATTENTION WEIGHT CORRECTION
USING WORD-LEVEL CONFIDENCE MEASURE..... 301
Jennifer Santoso, Takeshi Yamada, Shoji Makino, Kenkichi Ishizuka, Takekatsu Hiramura

DETECTING ALZHEIMER'S DISEASE USING INTERACTIONAL AND ACOUSTIC
FEATURES FROM SPONTANEOUS SPEECH 306
Shamila Nasreen, Julian Hough, Matthew Purver

INTERSPEECH 2021 : CONFERENCINGSPEECH 2021 CHALLENGE: FAR-FIELD MULTI-CHANNEL SPEECH ENHANCEMENT FOR VIDEO CONFERENCING

REAL-TIME MULTI-CHANNEL SPEECH ENHANCEMENT BASED ON NEURAL
NETWORK MASKING WITH ATTENTION MODEL..... 311
Cheng Xue, Weilong Huang, Weiguang Chen, Jinwei Feng

IMPROVING CHANNEL DECORRELATION FOR MULTI-CHANNEL TARGET SPEECH
EXTRACTION 316
Jiangyu Han, Wei Rao, Yannan Wang, Yanhua Long

INPLACE GATED CONVOLUTIONAL RECURRENT NEURAL NETWORK FOR DUAL-
CHANNEL SPEECH ENHANCEMENT 321
Jinjiang Liu, Xueliang Zhang

SRIB-LEAP SUBMISSION TO FAR-FIELD MULTI-CHANNEL SPEECH ENHANCEMENT
CHALLENGE FOR VIDEO CONFERENCING 326
R. G. Prithvi Raj, Rohit Kumar, M. K. Jayesh, Anurenjan Purushothaman, Sriram Ganapathy, M. A. Basha Shaik

A PARTITIONED-BLOCK FREQUENCY-DOMAIN ADAPTIVE KALMAN FILTER FOR
STEREOPHONIC ACOUSTIC ECHO CANCELLATION 331
Rui Zhu, Feiran Yang, Yuepeng Li, Shidong Shang

REAL-TIME INDEPENDENT VECTOR ANALYSIS USING SEMI-SUPERVISED
NONNEGATIVE MATRIX FACTORIZATION AS A SOURCE MODEL 336
Taihui Wang, Feiran Yang, Rui Zhu, Jun Yang

A CAUSAL U-NET BASED NEURAL BEAMFORMING NETWORK FOR REAL-TIME
MULTI-CHANNEL SPEECH ENHANCEMENT 341
Xinlei Ren, Xu Zhang, Lianwu Chen, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu

INTERSPEECH 2021 : CROSS/MULTI-LINGUAL AND CODE-SWITCHED ASR

UNSUPERVISED CROSS-LINGUAL REPRESENTATION LEARNING FOR SPEECH
RECOGNITION 346
Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli

MUCS 2021: MULTILINGUAL AND CODE-SWITCHING ASR CHALLENGES FOR LOW RESOURCE INDIAN LANGUAGES.....	351
<i>Anuj Diwan, Rakesh Vaideeswaran, Sanket Shah, Ankita Singh, Srinivasa Raghavan, Shreya Khare, Vinit Unni, Saurabh Vyas, Akash Rajpuria, Chiranjeevi Yarra, Ashish Mittal, Prasanta Kumar Ghosh, Preethi Jyothi, Kalika Bali, Vivek Seshadri, Sunayana Sitaram, Samarth Bharadwaj, Jai Nanavati, Raoul Nanavati, Karthik Sankaranarayanan</i>	
DIFFERENTIABLE ALLOPHONE GRAPHS FOR LANGUAGE-UNIVERSAL SPEECH RECOGNITION.....	356
<i>Brian Yan, Siddharth Dalmia, David R. Mortensen, Florian Metze, Shinji Watanabe</i>	
ADAPT-AND-ADJUST: OVERCOMING THE LONG-TAIL PROBLEM OF MULTILINGUAL SPEECH RECOGNITION.....	361
<i>Genta Indra Winata, Guangsen Wang, Caiming Xiong, Steven Hoi</i>	
SRI-B END-TO-END SYSTEM FOR MULTILINGUAL AND CODE-SWITCHING ASR CHALLENGES FOR LOW RESOURCE INDIAN LANGUAGES.....	366
<i>Hardik Sailor, Kiran Praveen T., Vikas Agrawal, Abhinav Jain, Abhishek Pandey</i>	
USING LARGE SELF-SUPERVISED MODELS FOR LOW-RESOURCE SPEECH RECOGNITION.....	371
<i>Krishna D. N., Pinyi Wang, Bruno Bozza</i>	
BOOTSTRAP AN END-TO-END ASR SYSTEM BY MULTILINGUAL TRAINING, TRANSFER LEARNING, TEXT-TO-TEXT MAPPING AND SYNTHETIC AUDIO.....	376
<i>Manuel Giollo, Deniz Gunceler, Yulan Liu, Daniel Willett</i>	
DUAL SCRIPT E2E FRAMEWORK FOR MULTILINGUAL AND CODE-SWITCHING ASR.....	381
<i>Mari Ganesh Kumar, Jom Kuriakose, Anand Thyagachandran, Arun Kumar A., Ashish Seth, Lodagala V. S. V. Durga Prasad, Saish Jaiswal, Anusha Prakash, Hema A. Murthy</i>	
EFFICIENT WEIGHT FACTORIZATION FOR MULTILINGUAL SPEECH RECOGNITION.....	386
<i>Ngoc-Quan Pham, Tuan-Nam Nguyen, Sebastian Stucker, Alex Waibel</i>	
TOWARDS ONE MODEL TO RULE ALL: MULTILINGUAL STRATEGY FOR DIALECTAL CODE-SWITCHING ARABIC ASR.....	391
<i>Shammur Absar Chowdhury, Amir Hussein, Ahmed Abdelali, Ahmed Ali</i>	
LANGUAGE AND SPEAKER-INDEPENDENT FEATURE TRANSFORMATION FOR END-TO-END MULTILINGUAL SPEECH RECOGNITION.....	396
<i>Tomoaki Hayakawa, Chee Siang Leow, Akio Kobayashi, Takehito Utsuro, Hiromitsu Nishizaki</i>	
HIERARCHICAL PHONE RECOGNITION WITH COMPOSITIONAL PHONETICS.....	401
<i>Xinjian Li, Juncheng Li, Florian Metze, Alan W. Black</i>	
<u>INTERSPEECH 2021 : DISORDERED SPEECH</u>	
ON MODELING GLOTTAL SOURCE INFORMATION FOR PHONATION ASSESSMENT IN PARKINSON'S DISEASE.....	406
<i>J. C. Vasquez-Correa, Julian Fritsch, J. R. Orozco-Arroyave, Elmar Noth, Mathew Magimai-Doss</i>	

DISTORTION OF VOICED OBSTRUENTS FOR DIFFERENTIAL DIAGNOSIS BETWEEN PARKINSON'S DISEASE AND MULTIPLE SYSTEM ATROPHY	411
<i>Khalid Daoudi, Biswajit Das, Solange Milhe De Saint Victor, Alexandra Foubert-Samier, Anne Pavy-Le Traon, Olivier Rascol, Wassilios G. Meissner, Virginie Woisard</i>	
A STUDY INTO PRE-TRAINING STRATEGIES FOR SPOKEN LANGUAGE UNDERSTANDING ON DYSARTHIC SPEECH	416
<i>Pu Wang, Bagher Babaali, Hugo Van Hamme</i>	
EASYCALL CORPUS: A DYSARTHIC SPEECH DATASET	421
<i>Rosanna Turrisi, Arianna Braccia, Marco Emanuele, Simone Giulietti, Maura Pugliatti, Mariachiara Sensi, Luciano Fadiga, Leonardo Badino</i>	
ACOUSTIC INDICATORS OF SPEECH MOTOR COORDINATION IN ADULTS WITH AND WITHOUT TRAUMATIC BRAIN INJURY	426
<i>Tanya Talkar, Nancy Pearl Solomon, Douglas S. Brungart, Stefanie E. Kuchinsky, Megan M. Eitel, Sara M. Lippa, Tracey A. Brickell, Louis M. French, Rael T. Lange, Thomas F. Quatieri</i>	
<u>INTERSPEECH 2021 : DIVERSE MODES OF SPEECH ACQUISITION AND PROCESSING</u>	
IMAGE-BASED ASSESSMENT OF JAW PARAMETERS AND JAW KINEMATICS FOR ARTICULATORY SIMULATION: PRELIMINARY RESULTS	431
<i>Ajish K. Abraham, V. Sivaramakrishnan, N. Swapna, N. Manohar</i>	
INVESTIGATING SPEECH RECONSTRUCTION FOR LARYNGECTOMIES FOR SILENT SPEECH INTERFACES	436
<i>Beiming Cao, Nordine Sebki, Arpan Bhavsar, Omer T. Inan, Robin Samlan, Ted Mau, Jun Wang</i>	
RASSPER: RADAR-BASED SILENT SPEECH RECOGNITION	441
<i>David Ferreira, Samuel Silva, Francisco Curado, Antonio Teixeira</i>	
EFFECT OF CARRIER BANDWIDTH ON UNDERSTANDING MANDARIN SENTENCES IN SIMULATED ELECTRIC-ACOUSTIC HEARING	446
<i>Feng Wang, Jing Chen, Fei Chen</i>	
AN ATTENTION SELF-SUPERVISED CONTRASTIVE LEARNING BASED THREE-STAGE MODEL FOR HAND SHAPE FEATURE REPRESENTATION IN CUED SPEECH	451
<i>Jianrong Wang, Nan Gu, Mei Yu, Xuwei Li, Qiang Fang, Li Liu</i>	
REMOTE SMARTPHONE-BASED SPEECH COLLECTION: ACCEPTANCE AND BARRIERS IN INDIVIDUALS WITH MAJOR DEPRESSIVE DISORDER	456
<i>Judith Dineley, Grace Lavelle, Daniel Leightley, Faith Matcham, Sara Siddi, Maria Teresa Penarrubia-Maria, Katie M. White, Alina Ivan, Carolin Oetzmann, Sara Simblett, Erin Dawe-Lane, Stuart Bruce, Daniel Stahl, Yatharth Ranjan, Zulqarnain Rashid, Pauline Conde, Amos A. Folarin, Josep Maria Haro, Til Wykes, Richard J. B. Dobson, Vaibhav A. Narayan, Matthew Hotopf, Bjorn W. Schuller, Nicholas Cummins, The Radar-Cns Consortium</i>	
A COMPARATIVE STUDY OF DIFFERENT EMG FEATURES FOR ACOUSTICS-TO-EMG MAPPING	461
<i>Manthan Sharma, Navaneetha Gaddam, Tejas Umesh, Aditya Murthy, Prasanta Kumar Ghosh</i>	
SILENT VERSUS MODAL MULTI-SPEAKER SPEECH RECOGNITION FROM ULTRASOUND AND VIDEO	466
<i>Manuel Sam Ribeiro, Aciel Eshky, Korin Richmond, Steve Renals</i>	

AN AUTOMATIC, SIMPLE ULTRASOUND BIOFEEDBACK PARAMETER FOR DISTINGUISHING ACCURATE AND MISARTICULATED RHOTIC SYLLABLES.....	471
<i>Sarah R. Li, Colin T. Annand, Sarah Dugan, Sarah M. Schwab, Kathryn J. Eary, Michael Swearengen, Sarah Stack, Suzanne Boyce, Michael A. Riley, T. Douglas Mast</i>	

SEGMENT AND TONE PRODUCTION IN CONTINUOUS SPEECH OF HEARING AND HEARING-IMPAIRED CHILDREN	476
<i>Shu-Chuan Tseng, Yi-Fen Liu</i>	

INTERSPEECH 2021 : EMBEDDING AND NETWORK ARCHITECTURE FOR SPEAKER RECOGNITION

LEVERAGING SPEAKER ATTRIBUTE INFORMATION USING MULTI TASK LEARNING FOR SPEAKER VERIFICATION AND DIARIZATION	481
<i>Chau Luu, Peter Bell, Steve Renals</i>	

ICSPK: INTERPRETABLE COMPLEX SPEAKER EMBEDDING EXTRACTOR FROM RAW WAVEFORM.....	486
<i>Junyi Peng, Xiaoyang Qu, Jianzong Wang, Rongzhi Gu, Jing Xiao, Lukas Burget, Jan Cernocky</i>	

SPINE2NET: SPINENET WITH RES2NET AND TIME-SQUEEZE-AND-EXCITATION BLOCKS FOR SPEAKER RECOGNITION.....	491
<i>Magdalena Rybicka, Jesus Villalba, Piotr Zelasko, Najim Dehak, Konrad Kowalczyk</i>	

SPEAKER EMBEDDINGS BY MODELING CHANNEL-WISE CORRELATIONS	496
<i>Themis Stafylakis, Johan Rohdin, Lukas Burget</i>	

MULTI-TASK NEURAL NETWORK FOR ROBUST MULTIPLE SPEAKER EMBEDDING EXTRACTION	501
<i>Weipeng He, Petr Motlicek, Jean-Marc Odobez</i>	

INTERSPEECH 2021 : EMOTION AND SENTIMENT ANALYSIS I

SPEAKER ATTENTIVE SPEECH EMOTION RECOGNITION	506
<i>Clement Le Moine, Nicolas Obin, Axel Roebel</i>	

M3: MULTIMODAL MASKING APPLIED TO SENTIMENT ANALYSIS	511
<i>Efthymios Georgiou, Georgios Paraskevopoulos, Alexandros Potamianos</i>	

SEPARATION OF EMOTIONAL AND RECONSTRUCTION EMBEDDINGS ON LADDER NETWORK TO IMPROVE SPEECH EMOTION RECOGNITION ROBUSTNESS IN NOISY CONDITIONS.....	516
<i>Seong-Gyun Leem, Daniel Fulford, Jukka-Pekka Onnela, David Gard, Carlos Busso</i>	

INTERSPEECH 2021 : EMOTION AND SENTIMENT ANALYSIS II

ACOUSTIC FEATURES AND NEURAL REPRESENTATIONS FOR CATEGORICAL EMOTION RECOGNITION FROM SPEECH	521
<i>Aaron Keesing, Yun Sing Koh, Michael Witbrock</i>	

AUTOMATIC ANALYSIS OF THE EMOTIONAL CONTENT OF SPEECH IN DAYLONG CHILD-CENTERED RECORDINGS FROM A NEONATAL INTENSIVE CARE UNIT.....	526
<i>Einari Vaaras, Sari Ahlqvist-Bjorkroth, Konstantinos Drossos, Okko Rasanen</i>	

MULTIMODAL SENTIMENT ANALYSIS WITH TEMPORAL MODALITY ATTENTION	531
<i>Fan Qian, Jiqing Han</i>	
LEARNING FINE-GRAINED CROSS MODALITY EXCITEMENT FOR SPEECH EMOTION RECOGNITION.....	536
<i>Hang Li, Wenbiao Ding, Zhongqin Wu, Zitao Liu</i>	
ACTED VS. IMPROVISED: DOMAIN ADAPTATION FOR ELICITATION APPROACHES IN AUDIO-VISUAL EMOTION RECOGNITION.....	541
<i>Haoqi Li, Yelin Kim, Cheng-Hao Kuo, Shrikanth S. Narayanan</i>	
GRAPH ISOMORPHISM NETWORK FOR SPEECH EMOTION RECOGNITION	546
<i>Jiawang Liu, Haoxiang Wang</i>	
EMOTION RECOGNITION FROM SPEECH USING WAV2VEC 2.0 EMBEDDINGS	551
<i>Leonardo Pepino, Pablo Riera, Luciana Ferrer</i>	
STOCHASTIC PROCESS REGRESSION FOR CROSS-CULTURAL SPEECH EMOTION RECOGNITION.....	556
<i>Mani Kumar T., Enrique Sanchez, Georgios Tzimiropoulos, Timo Giesbrecht, Michel Valstar</i>	
APPLYING TDNN ARCHITECTURES FOR ANALYZING DURATION DEPENDENCIES ON SPEECH EMOTION RECOGNITION.....	561
<i>Pooja Kumawat, Aurobinda Routray</i>	
LEVERAGING PRE-TRAINED LANGUAGE MODEL FOR SPEECH SENTIMENT ANALYSIS	566
<i>Suwon Shon, Pablo Brusco, Jing Pan, Kyu J. Han, Shinji Watanabe</i>	
TEMPORAL CONTEXT IN SPEECH EMOTION RECOGNITION.....	571
<i>Yangyang Xia, Li-Wei Chen, Alexander Rudnicky, Richard M. Stern</i>	
<u>INTERSPEECH 2021 : EMOTION AND SENTIMENT ANALYSIS III</u>	
PARAMETRIC DISTRIBUTIONS TO MODEL NUMERICAL EMOTION LABELS	576
<i>Deboshree Bose, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i>	
AFFECT RECOGNITION THROUGH SCALOGRAM AND MULTI-RESOLUTION COCHLEAGRAM FEATURES	581
<i>Fasih Haider, Saturnino Luz</i>	
A SPEECH EMOTION RECOGNITION FRAMEWORK FOR BETTER DISCRIMINATION OF CONFUSIONS.....	586
<i>Jiawang Liu, Haoxiang Wang</i>	
TIME-FREQUENCY REPRESENTATION LEARNING WITH GRAPH CONVOLUTIONAL NETWORK FOR DIALOGUE-LEVEL SPEECH EMOTION RECOGNITION.....	591
<i>Jiaxing Liu, Yaodong Song, Longbiao Wang, Jianwu Dang, Ruiguo Yu</i>	
AUDIO-VISUAL SPEECH EMOTION RECOGNITION BY DISENTANGLING EMOTION AND IDENTITY ATTRIBUTES	596
<i>Koichiro Ito, Takuya Fujioka, Qinghua Sun, Kenji Nagamatsu</i>	
GENERALIZED DILATED CNN MODELS FOR DEPRESSION DETECTION USING INVERTED VOCAL TRACT VARIABLES	601
<i>Nadee Seneviratne, Carol Espy-Wilson</i>	

SPEECH EMOTION RECOGNITION VIA MULTI-LEVEL CROSS-MODAL DISTILLATION	606
<i>Ruichen Li, Jinming Zhao, Qin Jin</i>	
SPEECH EMOTION RECOGNITION WITH MULTI-TASK LEARNING.....	611
<i>Xingyu Cai, Jiahong Yuan, Renjie Zheng, Liang Huang, Kenneth Church</i>	
METRIC LEARNING BASED FEATURE REPRESENTATION WITH GATED FUSION MODEL FOR SPEECH EMOTION RECOGNITION.....	616
<i>Yuan Gao, Jiaying Liu, Longbiao Wang, Jianwu Dang</i>	
LEARNING MUTUAL CORRELATION IN MULTIMODAL TRANSFORMER FOR SPEECH EMOTION RECOGNITION	621
<i>Yuhua Wang, Guang Shen, Yuezhu Xu, Jiahang Li, Zhengdao Zhao</i>	
<u>INTERSPEECH 2021 : FEATURE, EMBEDDING AND NEURAL ARCHITECTURE FOR SPEAKER RECOGNITION</u>	
Y-VECTOR: MULTISCALE WAVEFORM ENCODER FOR SPEAKER EMBEDDING	626
<i>Ge Zhu, Fei Jiang, Zhiyao Duan</i>	
SERIALIZED MULTI-LAYER MULTI-HEAD ATTENTION FOR NEURAL SPEAKER EMBEDDING.....	631
<i>Hongning Zhu, Kong Aik Lee, Haizhou Li</i>	
BIDIRECTIONAL MULTISCALE FEATURE AGGREGATION FOR SPEAKER VERIFICATION.....	636
<i>Jiajun Qi, Wu Guo, Bin Gu</i>	
ADAPTIVE CONVOLUTIONAL NEURAL NETWORK FOR TEXT-INDEPENDENT SPEAKER RECOGNITION	641
<i>Seong-Hu Kim, Yong-Hwa Park</i>	
BINARY NEURAL NETWORK FOR SPEAKER VERIFICATION.....	646
<i>Tinglong Zhu, Xiaoyi Qin, Ming Li</i>	
PHONEME-AWARE AND CHANNEL-WISE ATTENTIVE LEARNING FOR TEXT DEPENDENT SPEAKER VERIFICATION	651
<i>Yan Liu, Zheng Li, Lin Li, Qingyang Hong</i>	
IMPROVING DEEP CNN ARCHITECTURES WITH VARIABLE-LENGTH TRAINING SAMPLES FOR TEXT-INDEPENDENT SPEAKER VERIFICATION	656
<i>Yanfeng Wu, Junan Zhao, Chenkai Guo, Jing Xu</i>	
MUTUAL INFORMATION ENHANCED TRAINING FOR SPEAKER EMBEDDING.....	661
<i>Youzhi Tu, Man-Wai Mak</i>	
IMPROVING TIME DELAY NEURAL NETWORK BASED SPEAKER RECOGNITION WITH CONVOLUTIONAL BLOCK AND FEATURE AGGREGATION METHODS.....	666
<i>Yu-Jia Zhang, Yih-Wen Wang, Chia-Ping Chen, Chung-Li Lu, Bo-Cheng Chan</i>	

INTERSPEECH 2021 : GRAPH AND END-TO-END LEARNING FOR SPEAKER RECOGNITION

REFORMULATING DOVER-LAP LABEL MAPPING AS A GRAPH PARTITIONING PROBLEM.....	671
<i>Desh Raj, Sanjeev Khudanpur</i>	
GRAPH ATTENTION NETWORKS FOR ANTI-SPOOFING.....	676
<i>Hemlata Tak, Jee-Weon Jung, Jose Patino, Massimiliano Todisco, Nicholas Evans</i>	
EFFECTIVE PHASE ENCODING FOR END-TO-END SPEAKER VERIFICATION.....	681
<i>Junyi Peng, Xiaoyang Qu, Rongzhi Gu, Jianzong Wang, Jing Xiao, Lukas Burget, Jan Cernocky</i>	
LOG-LIKELIHOOD-RATIO COST FUNCTION AS OBJECTIVE LOSS FOR SPEAKER VERIFICATION SYSTEMS.....	686
<i>Victoria Mingote, Antonio Miguel, Alfonso Ortega, Eduardo Lleida</i>	

INTERSPEECH 2021 : HEALTH AND AFFECT I

ACOUSTIC-PROSODIC, LEXICAL AND DEMOGRAPHIC CUES TO PERSUASIVENESS IN COMPETITIVE DEBATE SPEECHES.....	691
<i>Huyen Nguyen, Ralph Vente, David Lupea, Sarah Ita Levitan, Julia Hirschberg</i>	
AUDIO-VISUAL RECOGNITION OF EMOTIONAL ENGAGEMENT OF PEOPLE WITH DEMENTIA.....	696
<i>Lars Steinert, Felix Putze, Dennis Kuster, Tanja Schultz</i>	
SPEAKING CORONA? HUMAN AND MACHINE RECOGNITION OF COVID-19 FROM VOICE.....	701
<i>Pascal Hecker, Florian B. Pokorny, Katrin D. Bartl-Pokorny, Uwe Reichel, Zhao Ren, Simone Hantke, Florian Eyben, Dagmar M. Schuller, Bert Arnrich, Bjorn W. Schuller</i>	
MEASURING VOICE QUALITY PARAMETERS AFTER SPEAKER PSEUDONYMIZATION.....	706
<i>Rob J. J. H. Van Son</i>	

INTERSPEECH 2021 : HEALTH AND AFFECT II

EMOTION CARRIER RECOGNITION FROM PERSONAL NARRATIVES.....	711
<i>Aniruddha Tammewar, Alessandra Cervone, Giuseppe Riccardi</i>	
VISUAL SPEECH FOR OBSTRUCTIVE SLEEP APNEA DETECTION.....	716
<i>Catarina Botelho, Alberto Abad, Tanja Schultz, Isabel Trancoso</i>	
TDCA-NET: TIME-DOMAIN CHANNEL ATTENTION NETWORK FOR DEPRESSION DETECTION.....	721
<i>Cong Cai, Mingyue Niu, Bin Liu, Jianhua Tao, Xuefei Liu</i>	
ANALYSIS OF CONTEXTUAL VOICE CHANGES IN REMOTE MEETINGS.....	726
<i>Hector A. Cordourier Maruri, Sinem Aslan, Georg Stemmer, Nese Alyuz, Lama Nachman</i>	
STACKED RECURRENT NEURAL NETWORKS FOR SPEECH-BASED INFERENCE OF ATTACHMENT CONDITION IN SCHOOL AGE CHILDREN.....	731
<i>Huda Alsofyani, Alessandro Vinciarelli</i>	

ROBUST LAUGHTER DETECTION IN NOISY ENVIRONMENTS	736
<i>Jon Gillick, Wesley Deng, Kimiko Ryokai, David Bamman</i>	

VOLUME 2

IMPACT OF EMOTIONAL STATE ON ESTIMATION OF WILLINGNESS TO BUY FROM ADVERTISING SPEECH	741
<i>Mizuki Nagano, Yusuke Ijima, Sadao Hiroya</i>	

SPEECH BASED DEPRESSION SEVERITY LEVEL CLASSIFICATION USING A MULTI- STAGE DILATED CNN-LSTM MODEL	746
<i>Nadee Seneviratne, Carol Espy-Wilson</i>	

LANGUAGE OR PARALANGUAGE, THIS IS THE PROBLEM: COMPARING DEPRESSED AND NON-DEPRESSED SPEAKERS THROUGH THE ANALYSIS OF GATED MULTIMODAL UNITS	751
<i>Nujud Alosban, Anna Esposito, Alessandro Vinciarelli</i>	

NON-VERBAL VOCALISATION AND LAUGHTER DETECTION USING SEQUENCE-TO- SEQUENCE MODELS AND MULTI-LABEL TRAINING	756
<i>Scott Condron, Georgia Clarke, Anita Klementiev, Daniela Morse-Kopp, Jack Parry, Dimitri Palaz</i>	

AUTOMATIC SPEECH RECOGNITION SYSTEMS ERRORS FOR OBJECTIVE SLEEPINESS DETECTION THROUGH VOICE	761
<i>Vincent P. Martin, Jean-Luc Rouas, Florian Boyer, Pierre Philip</i>	

INTERSPEECH 2021 : INTERSPEECH 2021 ACOUSTIC ECHO CANCELLATION CHALLENGE

NONLINEAR ACOUSTIC ECHO CANCELLATION WITH DEEP LEARNING	766
<i>Amir Ivry, Israel Cohen, Baruch Berdugo</i>	

Y2-NET FCNN FOR ACOUSTIC ECHO AND NOISE SUPPRESSION	771
<i>Ernst Seidel, Jan Franzen, Maximilian Strake, Tim Fingscheidt</i>	

ACOUSTIC ECHO CANCELLATION WITH CROSS-DOMAIN LEARNING	776
<i>Lukas Pfeifenberger, Matthias Zoehrer, Franz Pernkopf</i>	

ACOUSTIC ECHO CANCELLATION USING DEEP COMPLEX NEURAL NETWORK WITH NONLINEAR MAGNITUDE COMPRESSION AND PHASE INFORMATION	781
<i>Renhua Peng, Linjuan Cheng, Chengshi Zheng, Xiaodong Li</i>	

INTERSPEECH 2021 ACOUSTIC ECHO CANCELLATION CHALLENGE	786
<i>Ross Cutler, Ando Saabas, Tanel Parnamaa, Markus Loide, Sten Sootla, Marju Purin, Hannes Gamper, Sebastian Braun, Karsten Sorensen, Robert Aichner, Sriram Srinivasan</i>	

F-T-LSTM BASED COMPLEX NETWORK FOR JOINT ACOUSTIC ECHO CANCELLATION AND SPEECH ENHANCEMENT	791
<i>Shimin Zhang, Yuxiang Kong, Shubo Lv, Yanxin Hu, Lei Xie</i>	

INTERSPEECH 2021 : INTERSPEECH 2021 DEEP NOISE SUPPRESSION CHALLENGE

A SIMULTANEOUS DENOISING AND DEREVERBERATION FRAMEWORK WITH TARGET DECOUPLING.....	796
<i>Andong Li, Wenzhe Liu, Xiaoxue Luo, Guochen Yu, Chengshi Zheng, Xiaodong Li</i>	
INTERSPEECH 2021 DEEP NOISE SUPPRESSION CHALLENGE	801
<i>Chandan K. A. Reddy, Harishchandra Dubey, Kazuhito Koishida, Arun Nair, Vishak Gopal, Ross Cutler, Sebastian Braun, Hannes Gamper, Robert Aichner, Sriram Srinivasan</i>	
DBNET: A DUAL-BRANCH NETWORK ARCHITECTURE PROCESSING ON SPECTRUM AND WAVEFORM FOR SINGLE-CHANNEL SPEECH ENHANCEMENT	806
<i>Kanghao Zhang, Shulin He, Hao Li, Xueliang Zhang</i>	
LIGHTWEIGHT CAUSAL TRANSFORMER WITH LOCAL SELF-ATTENTION FOR REAL-TIME SPEECH ENHANCEMENT	811
<i>Koen Oostermeijer, Qing Wang, Jun Du</i>	
DCCRN+: CHANNEL-WISE SUBBAND DCCRN WITH SNR ESTIMATION FOR SPEECH ENHANCEMENT.....	816
<i>Shubo Lv, Yanxin Hu, Shimin Zhang, Lei Xie</i>	
DPCRN: DUAL-PATH CONVOLUTION RECURRENT NETWORK FOR SINGLE CHANNEL SPEECH ENHANCEMENT	821
<i>Xiaohuai Le, Hongsheng Chen, Kai Chen, Jing Lu</i>	
LOW-DELAY SPEECH ENHANCEMENT USING PERCEPTUALLY MOTIVATED TARGET AND LOSS.....	826
<i>Xu Zhang, Xinlei Ren, Xiguang Zheng, Lianwu Chen, Chen Zhang, Liang Guo, Bing Yu</i>	
DEEP NOISE SUPPRESSION WITH NON-INTRUSIVE PESQNET SUPERVISION ENABLING THE USE OF REAL TRAINING DATA.....	831
<i>Ziyi Xu, Maximilian Strake, Tim Fingscheidt</i>	

INTERSPEECH 2021 : KEYWORD SEARCH AND SPOKEN LANGUAGE PROCESSING

END-TO-END OPEN VOCABULARY KEYWORD SEARCH.....	836
<i>Bolaji Yusuf, Alican Gok, Batuhan Gundogdu, Murat Saraclar</i>	
SEMANTIC SENTENCE SIMILARITY: SIZE DOES NOT ALWAYS MATTER	841
<i>Danny Merckx, Stefan L. Frank, Mirjam Ernestus</i>	
TOWARD GENRE ADAPTED CLOSED CAPTIONING.....	846
<i>Francois Buet, Francois Yvon</i>	
SPOKEN TERM DETECTION AND RELEVANCE SCORE ESTIMATION USING DOT-PRODUCT OF PRONUNCIATION EMBEDDINGS.....	851
<i>Jan Svec, Lubos Smidl, Josef V. Psutka, Ales Prazak</i>	
DEVICE PLAYBACK AUGMENTATION WITH ECHO CANCELLATION FOR KEYWORD SPOTTING.....	856
<i>Kuba Lopatka, Katarzyna Kaszuba-Miotke, Piotr Klinke, Pawel Trella</i>	

INTERSPEECH 2021 : LANGUAGE AND ACCENT RECOGNITION

SELF-SUPERVISED PHONOTACTIC REPRESENTATIONS FOR LANGUAGE IDENTIFICATION	861
<i>G. Ramesh, C. Shiva Kumar, K. Sri Rama Murty</i>	
END-TO-END LANGUAGE DIARIZATION FOR BILINGUAL CODE-SWITCHING SPEECH	866
<i>Hexin Liu, Leibny Paola Garcia Perera, Xinyi Zhang, Justin Dauwels, Andy W. H. Khong, Sanjeev Khudanpur, Suzy J. Styles</i>	
A WEIGHT MOVING AVERAGE BASED ALTERNATE DECOUPLED LEARNING ALGORITHM FOR LONG-TAILED LANGUAGE IDENTIFICATION.....	871
<i>Hui Wang, Lin Liu, Yan Song, Lei Fang, Ian McLoughlin, Li-Rong Dai</i>	
E2E-BASED MULTI-TASK LEARNING APPROACH TO JOINT SPEECH AND ACCENT RECOGNITION.....	876
<i>Jicheng Zhang, Yizhou Peng, Van Tung Pham, Haihua Xu, Hao Huang, Eng Siong Chng</i>	
IMPROVING ACCENT IDENTIFICATION AND ACCENTED SPEECH RECOGNITION UNDER A FRAMEWORK OF SELF-SUPERVISED LEARNING.....	881
<i>Keqi Deng, Songjun Cao, Long Ma</i>	
EXCITATION SOURCE FEATURE BASED DIALECT IDENTIFICATION IN AO --- A LOW RESOURCE LANGUAGE	886
<i>Moakala Tzudir, Shikha Baghel, Priyankoo Sarmah, S. R. Mahadeva Prasanna</i>	
MODELING AND TRAINING STRATEGIES FOR LANGUAGE RECOGNITION SYSTEMS	891
<i>Raphael Duroselle, Md. Sahidullah, Denis Jouviet, Irina Illina</i>	
EXPLORING WAV2VEC 2.0 ON SPEAKER VERIFICATION AND LANGUAGE IDENTIFICATION	896
<i>Zhiyun Fan, Meng Li, Shiyu Zhou, Bo Xu</i>	

INTERSPEECH 2021 : LANGUAGE AND LEXICAL MODELING FOR ASR

REVISITING PARITY OF HUMAN VS. MACHINE CONVERSATIONAL SPEECH TRANSCRIPTION	901
<i>Courtney Mansfield, Sara Ng, Gina-Anne Levow, Richard A. Wright, Mari Ostendorf</i>	
IMPROVING CUSTOMIZATION OF NEURAL TRANSDUCERS BY MITIGATING ACOUSTIC MISMATCH OF SYNTHESIZED AUDIO	906
<i>Gakuto Kurata, George Saon, Brian Kingsbury, David Haws, Zoltan Tuske</i>	
CONTEXTUAL DENSITY RATIO FOR LANGUAGE MODEL BIASING OF SEQUENCE TO SEQUENCE ASR SYSTEMS.....	911
<i>Jesus Andres-Ferrer, Dario Albesano, Puming Zhan, Paul Vozila</i>	
CLASS-BASED NEURAL NETWORK LANGUAGE MODEL FOR SECOND-PASS RESCORING IN ASR	916
<i>Lingfeng Dai, Qi Liu, Kai Yu</i>	
CORRECTING AUTOMATED AND MANUAL SPEECH TRANSCRIPTION ERRORS USING WARPED LANGUAGE MODELS.....	921
<i>Mahdi Namazifar, John Malik, Li Erran Li, Gokhan Tur, Dilek Hakkani Tur</i>	

A DISCRIMINATIVE ENTITY-AWARE LANGUAGE MODEL FOR VIRTUAL ASSISTANTS	926
<i>Mandana Saebi, Ernest Pusateri, Aaksha Meghawat, Christophe Van Gysel</i>	
INCORPORATING EXTERNAL POS TAGGER FOR PUNCTUATION RESTORATION.....	931
<i>Ning Shi, Wei Wang, Boxin Wang, Jinfeng Li, Xiangyu Liu, Zhouhan Lin</i>	
TOKEN-LEVEL SUPERVISED CONTRASTIVE LEARNING FOR PUNCTUATION RESTORATION	936
<i>Qiushi Huang, Tom Ko, H. Lilian Tang, Xubo Liu, Bo Wu</i>	
SEMANTIC DISTANCE: A NEW METRIC FOR ASR PERFORMANCE ANALYSIS TOWARDS SPOKEN LANGUAGE UNDERSTANDING.....	941
<i>Suyoun Kim, Abhinav Arora, Duc Le, Ching-Feng Yeh, Christian Fuegen, Ozlem Kalinli, Michael L. Seltzer</i>	
PHONETICALLY INDUCED SUBWORDS FOR END-TO-END SPEECH RECOGNITION.....	946
<i>Vasileios Papadourakis, Markus Muller, Jing Liu, Athanasios Mouchtaris, Maurizio Omologo</i>	
LOOKUP-TABLE RECURRENT LANGUAGE MODELS FOR LONG TAIL SPEECH RECOGNITION.....	951
<i>W. Ronny Huang, Tara N. Sainath, Cal Peysner, Shankar Kumar, David Rybach, Trevor Strohman</i>	
A LIGHT-WEIGHT CONTEXTUAL SPELLING CORRECTION MODEL FOR CUSTOMIZING TRANSDUCER-BASED SPEECH RECOGNITION SYSTEMS.....	956
<i>Xiaoqiang Wang, Yanqing Liu, Sheng Zhao, Jinyu Li</i>	
BART BASED SEMANTIC CORRECTION FOR MANDARIN AUTOMATIC SPEECH RECOGNITION SYSTEM	961
<i>Yun Zhao, Xuerui Yang, Jinchao Wang, Yongyu Gao, Chao Yan, Yuanfu Zhou</i>	
<u>INTERSPEECH 2021 : LANGUAGE MODELING AND TEXT-BASED INNOVATIONS FOR ASR</u>	
BERT-BASED SEMANTIC MODEL FOR RESCORING N-BEST SPEECH RECOGNITION LIST	966
<i>Dominique Fohr, Irina Illina</i>	
FAST TEXT-ONLY DOMAIN ADAPTATION OF RNN-TRANSDUCER PREDICTION NETWORK.....	971
<i>Janne Pyllkonen, Antti Ukkonen, Juho Kilpikoski, Samu Tamminen, Hannes Heikinheimo</i>	
TEXT AUGMENTATION FOR LANGUAGE MODELS IN HIGH ERROR RECOGNITION SCENARIO	976
<i>Karel Benes, Lukas Burget</i>	
ON SAMPLING-BASED TRAINING CRITERIA FOR NEURAL LANGUAGE MODELING.....	981
<i>Yingbo Gao, David Thulke, Alexander Gerstenberger, Khoa Viet Tran, Ralf Schluter, Hermann Ney</i>	

INTERSPEECH 2021 : LINGUISTIC COMPONENTS IN END-TO-END ASR

MODELING DIALECTAL VARIATION FOR SWISS GERMAN AUTOMATIC SPEECH RECOGNITION.....	986
<i>Abbas Khosravani, Philip N. Garner, Alexandros Lazaridis</i>	
OUT-OF-VOCABULARY WORDS DETECTION WITH ATTENTION AND CTC ALIGNMENTS IN AN END-TO-END ASR SYSTEM.....	991
<i>Ekaterina Egorova, Hari Krishna Vydana, Lukas Burget, Jan Cernocky</i>	
TRAINING HYBRID MODELS ON NOISY TRANSLITERATED TRANSCRIPTS FOR CODE-SWITCHED SPEECH RECOGNITION	996
<i>Matthew Wiesner, Mousmita Sarma, Ashish Arora, Desh Raj, Dongji Gao, Ruizhe Huang, Supreet Preet, Moris Johnson, Zikra Iqbal, Nagendra Goel, Jan Trmal, Leibny Paola Garcia Perera, Sanjeev Khudanpur</i>	
THE CSTR SYSTEM FOR MULTILINGUAL AND CODE-SWITCHING ASR CHALLENGES FOR LOW RESOURCE INDIAN LANGUAGES.....	1001
<i>Ondrej Klejch, Electra Wallington, Peter Bell</i>	
EQUIVALENCE OF SEGMENTAL AND NEURAL TRANSDUCER MODELING: A PROOF OF CONCEPT.....	1006
<i>Wei Zhou, Albert Zeyer, Andre Merboldt, Ralf Schluter, Hermann Ney</i>	
ACOUSTIC DATA-DRIVEN SUBWORD MODELING FOR END-TO-END SPEECH RECOGNITION.....	1011
<i>Wei Zhou, Mohammad Zeineldeen, Zuoyun Zheng, Ralf Schluter, Hermann Ney</i>	

INTERSPEECH 2021 : LOW-RESOURCE SPEECH RECOGNITION

AVLNET: LEARNING AUDIO-VISUAL LANGUAGE REPRESENTATIONS FROM INSTRUCTIONAL VIDEOS.....	1016
<i>Andrew Rouditchenko, Angie Boggust, David Harwath, Brian Chen, Dhiraj Joshi, Samuel Thomas, Kartik Audhkhasi, Hilde Kuehne, Rameswar Panda, Rogerio Feris, Brian Kingsbury, Michael Picheny, Antonio Torralba, James Glass</i>	
ANALYZING SPEAKER INFORMATION IN SELF-SUPERVISED MODELS TO IMPROVE ZERO-RESOURCE SPEECH PROCESSING	1021
<i>Benjamin Van Niekerk, Leanne Nortje, Matthew Baas, Herman Kamper</i>	
MULTILINGUAL TRANSFER OF ACOUSTIC WORD EMBEDDINGS IMPROVES WHEN TRAINING ON LANGUAGES RELATED TO THE TARGET ZERO-RESOURCE LANGUAGE.....	1026
<i>Christiaan Jacobs, Herman Kamper</i>	
SPEECH SIMCLR: COMBINING CONTRASTIVE AND RECONSTRUCTION OBJECTIVE FOR SELF-SUPERVISED SPEECH REPRESENTATION LEARNING.....	1031
<i>Dongwei Jiang, Wubo Li, Miao Cao, Wei Zou, Xiangang Li</i>	
THE ZERO RESOURCE SPEECH CHALLENGE 2021: SPOKEN LANGUAGE MODELLING	1036
<i>Ewan Dunbar, Mathieu Bernard, Nicolas Hamilakis, Tu Anh Nguyen, Maureen De Seyssel, Patricia Roze, Morgane Riviere, Eugene Kharitonov, Emmanuel Dupoux</i>	
ZERO-SHOT FEDERATED LEARNING WITH NEW CLASSES FOR AUDIO CLASSIFICATION.....	1041
<i>Gautham Krishna Gudur, Satheesh Kumar Perepu</i>	

TOWARDS UNSUPERVISED PHONE AND WORD SEGMENTATION USING SELF-SUPERVISED VECTOR-QUANTIZED NEURAL NETWORKS..... 1046
Herman Kamper, Benjamin Van Niekerk

LOW RESOURCE ASR: THE SURPRISING EFFECTIVENESS OF HIGH RESOURCE TRANSLITERATION 1051
Shreya Khare, Ashish Mittal, Anuj Diwan, Sunita Sarawagi, Preethi Jyothi, Samarth Bharadwaj

UNSUPERVISED NEURAL-BASED GRAPH CLUSTERING FOR VARIABLE-LENGTH SPEECH REPRESENTATION DISCOVERY OF ZERO-RESOURCE LANGUAGES 1056
Shun Takahashi, Sakriani Sakti, Satoshi Nakamura

UNSUPERVISED ACOUSTIC UNIT DISCOVERY BY LEVERAGING A LANGUAGE-INDEPENDENT SUBWORD DISCRIMINATIVE FEATURE REPRESENTATION 1061
Siyan Feng, Piotr Zelasko, Laureano Moro-Velazquez, Odette Scharenborg

SPEECH REPRESENTATION LEARNING COMBINING CONFORMER CPC WITH DEEP CLUSTER FOR THE ZEROSPEECH CHALLENGE 2021 1066
Takashi Maekaku, Xuankai Chang, Yuya Fujita, Li-Wei Chen, Shinji Watanabe, Alexander Rudnicky

IDENTIFYING INDICATORS OF VULNERABILITY FROM SHORT SPEECH SEGMENTS USING ACOUSTIC AND TEXTUAL FEATURES 1071
Xia Cui, Amila Gamage, Terry Hanley, Tingting Mu

INTERSPEECH 2021 : MISCELLANEOUS TOPICS IN ASR

GOLOS: RUSSIAN DATASET FOR SPEECH RESEARCH 1076
Nikolay Karpov, Alexander Denisenko, Fedor Minkin

SPGISPEECH: 5,000 HOURS OF TRANSCRIBED FINANCIAL AUDIO FOR FULLY FORMATTED END-TO-END SPEECH RECOGNITION 1081
Patrick K. O'Neill, Vitaly Lavrukhin, Somshubra Majumdar, Vahid Noroozi, Yuekai Zhang, Oleksii Kuchaiev, Jagadeesh Balam, Yuliya Dovzhenko, Keenan Freyberg, Michael D. Shulman, Boris Ginsburg, Shinji Watanabe, Georg Kucsko

SELF-SUPERVISED END-TO-END ASR FOR LOW RESOURCE L2 SWEDISH..... 1086
Ragheb Al-Ghezi, Yaroslav Getman, Aku Rouhe, Raili Hilden, Mikko Kurimo

RADICALLY OLD WAY OF COMPUTING SPECTRA: APPLICATIONS IN END-TO-END ASR 1091
Samik Sadhu, Hynek Hermansky

LEBENCHMARK: A REPRODUCIBLE FRAMEWORK FOR ASSESSING SELF-SUPERVISED REPRESENTATION LEARNING FROM SPEECH..... 1096
Solene Evain, Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Esteve, Benjamin Lecouteux, Francois Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, Laurent Besacier

INTERSPEECH 2021 : MULTI- AND CROSS-LINGUAL ASR, OTHER TOPICS IN ASR

ARABIC CODE-SWITCHING SPEECH RECOGNITION USING MONOLINGUAL DATA..... 1101
Ahmed Ali, Shammur Absar Chowdhury, Amir Hussein, Yasser Hifny

REDUCE AND RECONSTRUCT: ASR FOR LOW-RESOURCE PHONETIC LANGUAGES	1106
<i>Anuj Diwan, Preethi Jyothi</i>	
IMPROVING MULTILINGUAL TRANSFORMER TRANSDUCER MODELS BY REDUCING LANGUAGE CONFUSIONS	1111
<i>Eric Sun, Jinyu Li, Zhong Meng, Yu Wu, Jian Xue, Shujie Liu, Yifan Gong</i>	
REDUCING STREAMING ASR MODEL DELAY WITH SELF ALIGNMENT.....	1116
<i>Jaeyoung Kim, Han Lu, Anshuman Tripathi, Qian Zhang, Hasim Sak</i>	
ON MINIMUM WORD ERROR RATE TRAINING OF THE HYBRID AUTOREGRESSIVE TRANSDUCER	1121
<i>Liang Lu, Zhong Meng, Naoyuki Kanda, Jinyu Li, Yifan Gong</i>	
EARNINGS-21: A PRACTICAL BENCHMARK FOR ASR IN THE WILD	1126
<i>Miguel Del Rio, Natalie Delworth, Ryan Westerman, Michelle Huang, Nishchal Bhandari, Joseph Palakapilly, Quinten McNamara, Joshua Dong, Piotr Zelasko, Miguel Jette</i>	
LARGE-SCALE PRE-TRAINING OF END-TO-END MULTI-TALKER ASR FOR MEETING TRANSCRIPTION WITH SINGLE DISTANT MICROPHONE.....	1131
<i>Naoyuki Kanda, Guoli Ye, Yu Wu, Yashesh Gaur, Xiaofei Wang, Zhong Meng, Zhuo Chen, Takuya Yoshioka</i>	
LISTEN WITH INTENT: IMPROVING SPEECH RECOGNITION WITH AUDIO-TO-INTENT FRONT-END	1136
<i>Swayambhu Nath Ray, Minhua Wu, Anirudh Raju, Pegah Ghahremani, Raghavendra Bilgi, Milind Rao, Harish Arsikere, Ariya Rastrow, Andreas Stolcke, Jasha Droppo</i>	
KNOWLEDGE DISTILLATION BASED TRAINING OF UNIVERSAL ASR SOURCE MODELS FOR CROSS-LINGUAL TRANSFER.....	1141
<i>Takashi Fukuda, Samuel Thomas</i>	
CROSS-DOMAIN SPEECH RECOGNITION WITH UNSUPERVISED CHARACTER-LEVEL DISTRIBUTION MATCHING.....	1146
<i>Wenxin Hou, Jindong Wang, Xu Tan, Tao Qin, Takahiro Shinozaki</i>	
EXPLORING TARGETED UNIVERSAL ADVERSARIAL PERTURBATIONS TO END-TO- END ASR MODELS.....	1151
<i>Zhiyun Lu, Wei Han, Yu Zhang, Liangliang Cao</i>	
<u>INTERSPEECH 2021 : MULTI-CHANNEL SPEECH ENHANCEMENT AND HEARING AIDS</u>	
A DEEP LEARNING METHOD TO MULTI-CHANNEL ACTIVE NOISE CONTROL.....	1156
<i>Hao Zhang, Deliang Wang</i>	
LACOPE: LATENCY-CONSTRAINED PITCH ESTIMATION FOR SPEECH ENHANCEMENT.....	1161
<i>Hendrik Schroter, Tobias Rosenkranz, Alberto N. Escalante-B., Andreas Maier</i>	
MULTIPLE SOUND SOURCE LOCALIZATION BASED ON INTERCHANNEL PHASE DIFFERENCES IN ALL FREQUENCIES WITH SPECTRAL MASKS	1166
<i>Hyungchan Song, Jong Won Shin</i>	

ALPHA-STABLE AUTOREGRESSIVE FAST MULTICHANNEL NONNEGATIVE MATRIX FACTORIZATION FOR JOINT SPEECH ENHANCEMENT AND DEREVERBERATION	1171
<i>Mathieu Fontaine, Kouhei Sekiguchi, Aditya Arie Nugraha, Yoshiaki Bando, Kazuyoshi Yoshii</i>	
CANCELLATION OF LOCAL COMPETING SPEAKER WITH NEAR-FIELD LOCALIZATION FOR DISTRIBUTED AD-HOC SENSOR NETWORK	1176
<i>Pablo Perez Zarazaga, Mariem Bouafif Mansali, Tom Backstrom, Zied Lachiri</i>	
CLARITY-2021 CHALLENGES: MACHINE LEARNING CHALLENGES FOR ADVANCING HEARING AID PROCESSING.....	1181
<i>Simone Graetzer, Jon Barker, Trevor J. Cox, Michael Akeroyd, John F. Culling, Graham Naylor, Eszter Porter, Rhoddy Viveros Munoz</i>	
MICROPHONE ARRAY GENERALIZATION FOR MULTICHANNEL NARROWBAND DEEP SPEECH ENHANCEMENT	1186
<i>Siyuan Zhang, Xiaofei Li</i>	
EXPLAINING DEEP LEARNING MODELS FOR SPEECH ENHANCEMENT.....	1191
<i>Sunit Sivasankaran, Emmanuel Vincent, Dominique Fohr</i>	
MINIMUM-NORM DIFFERENTIAL BEAMFORMING FOR LINEAR ARRAY WITH DIRECTIONAL MICROPHONES.....	1196
<i>Weilong Huang, Jinwei Feng</i>	
OPTIMISING HEARING AID FITTINGS FOR SPEECH IN NOISE WITH A DIFFERENTIABLE HEARING LOSS MODEL	1201
<i>Zehai Tu, Ning Ma, Jon Barker</i>	
<u>INTERSPEECH 2021 : MULTIMODAL SYSTEMS</u>	
CASCADED MULTILINGUAL AUDIO-VISUAL LEARNING FROM VIDEOS	1206
<i>Andrew Rouditchenko, Angie Boggust, David Harwath, Samuel Thomas, Hilde Kuehne, Brian Chen, Rameswar Panda, Rogerio Feris, Brian Kingsbury, Michael Picheny, James Glass</i>	
AUTOMATIC LIP-READING WITH HIERARCHICAL PYRAMIDAL CONVOLUTION AND SELF-ATTENTION FOR IMAGE SEQUENCES WITH NO WORD BOUNDARIES	1211
<i>Hang Chen, Jun Du, Yu Hu, Li-Rong Dai, Bao-Cai Yin, Chin-Hui Lee</i>	
A FAST DISCRETE TWO-STEP LEARNING HASHING FOR SCALABLE CROSS-MODAL RETRIEVAL.....	1216
<i>Huan Zhao, Kaili Ma</i>	
CROSS-MODAL KNOWLEDGE DISTILLATION METHOD FOR AUTOMATIC CUED SPEECH RECOGNITION.....	1221
<i>Jianrong Wang, Ziyue Tang, Xuwei Li, Mei Yu, Qiang Fang, Li Liu</i>	
ATTENTION-BASED KEYWORD LOCALISATION IN SPEECH USING VISUAL GROUNDING.....	1226
<i>Kayode Olaleye, Herman Kamper</i>	
EVALUATION OF AUDIO-VISUAL ALIGNMENTS IN VISUALLY GROUNDED SPEECH MODELS	1231
<i>Khazar Khorrami, Okko Rasanen</i>	

DIRECT MULTIMODAL FEW-SHOT LEARNING OF SPEECH AND IMAGES	1236
<i>Leanne Nortje, Herman Kamper</i>	
LIRA: LEARNING VISUAL SPEECH REPRESENTATIONS FROM AUDIO THROUGH SELF-SUPERVISION	1241
<i>Pingchuan Ma, Rodrigo Mira, Stavros Petridis, Bjorn W. Schuller, Maja Pantic</i>	
TALK, DON'T WRITE: A STUDY OF DIRECT SPEECH-BASED IMAGE RETRIEVAL	1246
<i>Ramon Sanabria, Austin Waters, Jason Baldrige</i>	
END-TO-END AUDIO-VISUAL SPEECH RECOGNITION FOR OVERLAPPING SPEECH	1251
<i>Richard Rose, Olivier Siohan, Anshuman Tripathi, Otavio Braga</i>	
AUDIO-VISUAL MULTI-TALKER SPEECH RECOGNITION IN A COCKTAIL PARTY.....	1256
<i>Yifei Wu, Chenda Li, Song Yang, Zhongqin Wu, Yanmin Qian</i>	

INTERSPEECH 2021 : NEURAL NETWORK TRAINING METHODS AND ARCHITECTURES FOR ASR

COMPARING CTC AND LFMMI FOR OUT-OF-DOMAIN ADAPTATION OF WAV2VEC 2.0 ACOUSTIC MODEL.....	1261
<i>Apoorv Vyas, Srikanth Madikeri, Herve Bourlard</i>	
KNOWLEDGE DISTILLATION FOR STREAMING TRANSFORMER--TRANSDUCER.....	1266
<i>Atsushi Kojima</i>	
INVESTIGATING METHODS TO IMPROVE LANGUAGE MODEL INTEGRATION FOR ATTENTION-BASED ENCODER-DECODER ASR MODELS	1271
<i>Mohammad Zeineldeen, Aleksandr Glushko, Wilfried Michel, Albert Zeyer, Ralf Schluter, Hermann Ney</i>	
SELF-PACED ENSEMBLE LEARNING FOR SPEECH AND AUDIO CLASSIFICATION	1276
<i>Nicolae-Catalin Ristea, Radu Tudor Ionescu</i>	
CONDITIONAL INDEPENDENCE FOR PRETEXT TASK SELECTION IN SELF- SUPERVISED SPEECH REPRESENTATION LEARNING	1281
<i>Salah Zaiem, Titouan Parcollet, Slim Essid</i>	
MULTI-ENCODER LEARNING AND STREAM FUSION FOR TRANSFORMER-BASED END-TO-END AUTOMATIC SPEECH RECOGNITION.....	1286
<i>Timo Lohrenz, Zhengyang Li, Tim Fingscheidt</i>	

INTERSPEECH 2021 : NEURAL NETWORK TRAINING METHODS FOR ASR

4-BIT QUANTIZATION OF LSTM-BASED SPEECH RECOGNITION MODELS	1291
<i>Andrea Fasoli, Chia-Yu Chen, Mauricio Serrano, Xiao Sun, Naigang Wang, Swagath Venkataramani, George Saon, Xiaodong Cui, Brian Kingsbury, Wei Zhang, Zoltan Tuske, Kailash Gopalakrishnan</i>	
VARIABLE FRAME RATE ACOUSTIC MODELS USING MINIMUM ERROR REINFORCEMENT LEARNING	1296
<i>Dongcheng Jiang, Chao Zhang, Philip C. Woodland</i>	
REGULARIZING WORD SEGMENTATION BY CREATING MISSPELLINGS.....	1301
<i>Hainan Xu, Kartik Audhkhasi, Yinghui Huang, Jesse Emond, Bhuvana Ramabhadran</i>	

TOWARDS LIFELONG LEARNING OF END-TO-END ASR.....	1306
<i>Heng-Jui Chang, Hung-Yi Lee, Lin-Shan Lee</i>	
MULTI-DOMAIN KNOWLEDGE DISTILLATION VIA UNCERTAINTY-MATCHING FOR END-TO-END ASR MODELS.....	1311
<i>Ho-Gyeong Kim, Min-Joong Lee, Hoshik Lee, Tae Gyoon Kang, Jihyun Lee, Eunho Yang, Sung Ju Hwang</i>	
SELF-ADAPTIVE DISTILLATION FOR MULTILINGUAL SPEECH RECOGNITION: LEVERAGING STUDENT INDEPENDENCE.....	1316
<i>Isabel Leal, Neeraj Gaur, Parisa Haghani, Brian Farris, Pedro J. Moreno, Manasa Prasad, Bhuvana Ramabhadran, Yun Zhu</i>	
SCALING LAWS FOR ACOUSTIC MODELS.....	1321
<i>Jasha Droppo, Oguz Elibol</i>	
LEVERAGING NON-TARGET LANGUAGE RESOURCES TO IMPROVE ASR PERFORMANCE IN A TARGET LANGUAGE.....	1326
<i>Jayadev Billa</i>	
MODEL-AGNOSTIC FAST ADAPTIVE MULTI-OBJECTIVE BALANCING ALGORITHM FOR MULTILINGUAL AUTOMATIC SPEECH RECOGNITION MODEL TRAINING.....	1331
<i>Jiabin Xue, Tieran Zheng, Jiqing Han</i>	
LEARNING A NEURAL DIFF FOR SPEECH MODELS.....	1336
<i>Jonathan Macoskey, Grant P. Strimel, Ariya Rastrow</i>	
MULTITASK TRAINING WITH TEXT DATA FOR END-TO-END SPEECH RECOGNITION.....	1341
<i>Peidong Wang, Tara N. Sainath, Ron J. Weiss</i>	
UNIFIED AUTOREGRESSIVE MODELING FOR JOINT END-TO-END MULTI-TALKER OVERLAPPED SPEECH RECOGNITION AND SPEAKER ATTRIBUTE ESTIMATION.....	1346
<i>Ryo Masumura, Daiki Okamura, Naoki Makishima, Mana Ihori, Akihiko Takashima, Tomohiro Tanaka, Shota Orihashi</i>	
STOCHASTIC ATTENTION HEAD REMOVAL: A SIMPLE AND EFFECTIVE METHOD FOR IMPROVING TRANSFORMER BASED ASR MODELS.....	1351
<i>Shucong Zhang, Erfan Loweimi, Peter Bell, Steve Renals</i>	
EMITTING WORD TIMINGS WITH HMM-FREE END-TO-END SYSTEM IN AUTOMATIC SPEECH RECOGNITION.....	1356
<i>Xianzhao Chen, Hao Ni, Yi He, Kang Wang, Zejun Ma, Zongxia Xie</i>	
MINIMUM WORD ERROR RATE TRAINING WITH LANGUAGE MODEL FUSION FOR END-TO-END SPEECH RECOGNITION.....	1361
<i>Zhong Meng, Yu Wu, Naoyuki Kanda, Liang Lu, Xie Chen, Guoli Ye, Eric Sun, Jinyu Li, Yifan Gong</i>	

INTERSPEECH 2021 : NON-AUTOREGRESSIVE SEQUENTIAL MODELING FOR SPEECH PROCESSING

NON-AUTOREGRESSIVE PREDICTIVE CODING FOR LEARNING SPEECH REPRESENTATIONS FROM LOCAL DEPENDENCIES.....	1366
<i>Alexander H. Liu, Yu-An Chung, James Glass</i>	

PUSHING THE LIMITS OF NON-AUTOREGRESSIVE SPEECH RECOGNITION.....	1371
<i>Edwin G. Ng, Chung-Cheng Chiu, Yu Zhang, William Chan</i>	
VAENAR-TTS: VARIATIONAL AUTO-ENCODER BASED NON-AUTOREGRESSIVE TEXT-TO-SPEECH SYNTHESIS.....	1376
<i>Hui Lu, Zhiyong Wu, Xixin Wu, Xu Li, Shiyin Kang, Xunying Liu, Helen Meng</i>	
LAYER PRUNING ON DEMAND WITH INTERMEDIATE CTC	1381
<i>Jaesong Lee, Jingu Kang, Shinji Watanabe</i>	
RELAXING THE CONDITIONAL INDEPENDENCE ASSUMPTION OF CTC-BASED ASR BY CONDITIONING ON INTERMEDIATE PREDICTIONS	1386
<i>Jumon Nozaki, Tatsuya Komatsu</i>	
ALIGN-DENOISE: SINGLE-PASS NON-AUTOREGRESSIVE SPEECH RECOGNITION	1391
<i>Nanxin Chen, Piotr Zelasko, Laureano Moro-Velazquez, Jesus Villalba, Najim Dehak</i>	
WAVEGRAD 2: ITERATIVE REFINEMENT FOR TEXT-TO-SPEECH SYNTHESIS.....	1396
<i>Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, Najim Dehak, William Chan</i>	
MULTI-SPEAKER ASR COMBINING NON-AUTOREGRESSIVE CONFORMER CTC AND CONDITIONAL SPEAKER CHAIN	1401
<i>Pengcheng Guo, Xuankai Chang, Shinji Watanabe, Lei Xie</i>	
AN IMPROVED SINGLE STEP NON-AUTOREGRESSIVE TRANSFORMER FOR AUTOMATIC SPEECH RECOGNITION	1406
<i>Ruchao Fan, Wei Chu, Peng Chang, Jing Xiao, Abeer Alwan</i>	
REAL-TIME END-TO-END MONAURAL MULTI-SPEAKER SPEECH RECOGNITION.....	1411
<i>Song Li, Beibei Ouyang, Fuchuan Tong, Dexin Liao, Lin Li, Qingyang Hong</i>	
TALKNET: NON-AUTOREGRESSIVE DEPTH-WISE SEPARABLE CONVOLUTIONAL MODEL FOR SPEECH SYNTHESIS.....	1416
<i>Stanislav Beliaev, Boris Ginsburg</i>	
STREAMING END-TO-END ASR BASED ON BLOCKWISE NON-AUTOREGRESSIVE MODELS	1421
<i>Tianzi Wang, Yuya Fujita, Xuankai Chang, Shinji Watanabe</i>	
TOWARD STREAMING ASR WITH NON-AUTOREGRESSIVE INSERTION-BASED MODEL.....	1426
<i>Yuya Fujita, Tianzi Wang, Shinji Watanabe, Motoi Omachi</i>	
 <u>INTERSPEECH 2021 : NON-NATIVE SPEECH</u>	
TESTING ACOUSTIC VOICE QUALITY CLASSIFICATION ACROSS LANGUAGES AND SPEECH STYLES	1431
<i>Bettina Braun, Nicole Dehe, Marieke Einfeldt, Daniela Wochner, Katharina Zahner-Ritter</i>	
A NEURAL NETWORK-BASED NOISE COMPENSATION METHOD FOR PRONUNCIATION ASSESSMENT.....	1436
<i>Binghuai Lin, Liyuan Wang</i>	
L1 IDENTIFICATION FROM L2 SPEECH USING NEURAL SPECTROGRAM ANALYSIS	1441
<i>Calbert Graham</i>	

DETECTION OF LEXICAL STRESS ERRORS IN NON-NATIVE (L2) ENGLISH WITH DATA AUGMENTATION AND ATTENTION.....	1446
<i>Daniel Korzekwa, Roberto Barra-Chicote, Szymon Zaporowski, Grzegorz Beringer, Jaime Lorenzo-Trueba, Alicja Serafinowicz, Jasha Droppo, Thomas Drugman, Bozena Kostek</i>	
F_(0) PATTERNS OF L2 ENGLISH SPEECH BY MANDARIN CHINESE LEARNERS	1451
<i>Hongwei Ding, Binghuai Lin, Liyuan Wang</i>	
PHONETIC DISTANCE AND SURPRISAL IN MULTILINGUAL PRIMING: EVIDENCE FROM SLAVIC	1456
<i>Jacek Kudera, Philip Georgis, Bernd Mobius, Tania Avgustinova, Dietrich Klakow</i>	
CROSS-LINGUISTIC PERCEPTION OF THE JAPANESE SINGLETON/GEMINATE CONTRAST: KOREAN, MANDARIN AND MONGOLIAN COMPARED	1461
<i>Kimiko Tsukada, Yurong, Joo-Yeon Kim, Jeong-Im Han, John Hajek</i>	
ADAPTIVE LISTENING DIFFICULTY DETECTION FOR L2 LEARNERS THROUGH MODERATING ASR RESOURCES	1466
<i>Maryam Sadat Mirzaei, Kourosh Meshgi</i>	
TRANSFORMER BASED END-TO-END MISPRONUNCIATION DETECTION AND DIAGNOSIS	1471
<i>Minglin Wu, Kun Li, Wai-Kim Leung, Helen Meng</i>	
ACQUISITION OF PROSODIC FOCUS MARKING BY THREE- TO SIX-YEAR-OLD CHILDREN LEARNING MANDARIN CHINESE.....	1476
<i>Qianyutong Zhang, Kexin Lyu, Zening Chen, Ping Tang</i>	
A PRELIMINARY STUDY ON DISCOURSE PROSODY ENCODING IN L1 AND L2 ENGLISH SPONTANEOUS NARRATIVES	1480
<i>Yuqing Zhang, Zhu Li, Binghuai Lin, Jinsong Zhang</i>	
<u>INTERSPEECH 2021 : NOVEL NEURAL NETWORK ARCHITECTURES FOR ASR</u>	
LIBRISPEECH TRANSDUCER MODEL WITH INTERNAL LANGUAGE MODEL PRIOR CORRECTION	1485
<i>Albert Zeyer, Andre Merboldt, Wilfried Michel, Ralf Schluter, Hermann Ney</i>	
END TO END TRANSFORMER-BASED CONTEXTUAL SPEECH RECOGNITION BASED ON POINTER NETWORK	1490
<i>Binghuai Lin, Liyuan Wang</i>	
TRANSFORMER-BASED END-TO-END SPEECH RECOGNITION WITH RESIDUAL GAUSSIAN-BASED SELF-ATTENTION	1495
<i>Chengdong Liang, Menglong Xu, Xiao-Lei Zhang</i>	
ONLINE COMPRESSIVE TRANSFORMER FOR END-TO-END SPEECH RECOGNITION.....	1500
<i>Chi-Hang Leong, Yu-Han Huang, Jen-Tzung Chien</i>	
FLEXI-TRANSDUCER: OPTIMIZING LATENCY, ACCURACY AND COMPUTE FOR MULTI-DOMAIN ON-DEVICE SCENARIOS	1505
<i>Jay Mahadeokar, Yangyang Shi, Yuan Shangguan, Chunyang Wu, Alex Xiao, Hang Su, Duc Le, Ozlem Kalinli, Christian Fuegen, Michael L. Seltzer</i>	

DEFORMABLE TDNN WITH ADAPTIVE RECEPTIVE FIELDS FOR SPEECH RECOGNITION.....	1510
<i>Keyu An, Yi Zhang, Zhijian Ou</i>	
TRANSFORMER-BASED ASR INCORPORATING TIME-REDUCTION LAYER AND FINE-TUNING WITH SELF-KNOWLEDGE DISTILLATION.....	1515
<i>Md. Akmal Haidar, Chao Xing, Mehdi Rezagholizadeh</i>	
A DELIBERATION-BASED JOINT ACOUSTIC AND TEXT DECODER	1520
<i>Sepand Mavandadi, Tara N. Sainath, Ke Hu, Zelin Wu</i>	
A COMPARATIVE STUDY ON NEURAL ARCHITECTURES AND TRAINING METHODS FOR JAPANESE SPEECH RECOGNITION.....	1525
<i>Shigeki Karita, Yotaro Kubo, Michiel Adriaan Unico Bacchiani, Llion Jones</i>	
DOMAIN-AWARE SELF-ATTENTION FOR MULTI-DOMAIN NEURAL MACHINE TRANSLATION.....	1530
<i>Shiqi Zhang, Yan Liu, Deyi Xiong, Pei Zhang, Boxing Chen</i>	
ADVANCED LONG-CONTEXT END-TO-END SPEECH RECOGNITION USING CONTEXT-EXPANDED TRANSFORMERS.....	1535
<i>Takaaki Hori, Niko Moritz, Chiori Hori, Jonathan Le Roux</i>	
DYNAMIC ENCODER TRANSDUCER: A FLEXIBLE SOLUTION FOR TRADING OFF ACCURACY FOR LATENCY.....	1540
<i>Yangyang Shi, Varun Nagaraja, Chunyang Wu, Jay Mahadeokar, Duc Le, Rohit Prabhavalkar, Alex Xiao, Ching-Feng Yeh, Julian Chan, Christian Fuegen, Ozlem Kalinli, Michael L. Seltzer</i>	
SPEECHMOE: SCALING TO LARGE ACOUSTIC MODELS WITH DYNAMIC ROUTING MIXTURE OF EXPERTS	1545
<i>Zhao You, Shulin Feng, Dan Su, Dong Yu</i>	
ON THE LIMIT OF ENGLISH CONVERSATIONAL SPEECH RECOGNITION	1550
<i>Zoltan Tuske, George Saon, Brian Kingsbury</i>	

INTERSPEECH 2021 : OPENASR20 AND LOW RESOURCE ASR DEVELOPMENT

ONE SIZE DOES NOT FIT ALL IN RESOURCE-CONSTRAINED ASR	1555
<i>Ethan Morris, Robbie Jimerson, Emily Prud'Hommeaux</i>	
SYSTEMS FOR LOW-RESOURCE SPEECH RECOGNITION TASKS IN OPEN AUTOMATIC SPEECH RECOGNITION AND FORMOSA SPEECH RECOGNITION CHALLENGES	1560
<i>Hung-Pang Lin, Yu-Jia Zhang, Chia-Ping Chen</i>	

VOLUME 3

THE TNT TEAM SYSTEM DESCRIPTIONS OF CANTONESE AND MONGOLIAN FOR IARPA OPENASR20.....	1565
<i>Jing Zhao, Zhiqiang Lv, Ambyera Han, Guan-Bo Wang, Guixin Shi, Jian Kang, Jinghao Yan, Pengfei Hu, Shen Huang, Wei-Qiang Zhang</i>	
OPENASR20: AN OPEN CHALLENGE FOR AUTOMATIC SPEECH RECOGNITION OF CONVERSATIONAL TELEPHONE SPEECH IN LOW-RESOURCE LANGUAGES	1570
<i>Kay Peterson, Audrey Tong, Yan Yu</i>	

AN IMPROVED WAV2VEC 2.0 PRE-TRAINING APPROACH USING ENHANCED LOCAL DEPENDENCY MODELING FOR SPEECH RECOGNITION.....	1575
<i>Qiu-Shi Zhu, Jie Zhang, Ming-Hui Wu, Xin Fang, Li-Rong Dai</i>	
MULTITASK ADAPTATION WITH LATTICE-FREE MMI FOR MULTI-GENRE SPEECH RECOGNITION OF LOW RESOURCE LANGUAGES.....	1580
<i>Srikanth Madikeri, Petr Motlicek, Herve Bourlard</i>	
COMBINING HYBRID AND END-TO-END APPROACHES FOR THE OPENASR20 CHALLENGE.....	1585
<i>Tanel Alumae, Jiaming Kong</i>	

INTERSPEECH 2021 : ORIENTAL LANGUAGE RECOGNITION

AN END-TO-END DIALECT IDENTIFICATION SYSTEM WITH TRANSFER LEARNING FROM A MULTILINGUAL AUTOMATIC SPEECH RECOGNITION MODEL.....	1590
<i>Ding Wang, Shuaishuai Ye, Xinhui Hu, Sheng Li, Xinkang Xu</i>	
LANGUAGE RECOGNITION BASED ON UNSUPERVISED PRETRAINED MODELS	1595
<i>Haibin Yu, Jing Zhao, Song Yang, Zhongqin Wu, Yuting Nie, Wei-Qiang Zhang</i>	
ORIENTAL LANGUAGE RECOGNITION (OLR) 2020: SUMMARY AND ANALYSIS.....	1600
<i>Jing Li, Binling Wang, Yiming Zhi, Zheng Li, Lin Li, Qingyang Hong, Dong Wang</i>	
LANGUAGE RECOGNITION ON UNKNOWN CONDITIONS: THE LORIA-INRIA- MULTISPEECH SYSTEM FOR AP20-OLR CHALLENGE.....	1605
<i>Raphael Duroselle, Md. Sahidullah, Denis Jouviet, Irina Illina</i>	
DYNAMIC MULTI-SCALE CONVOLUTION FOR DIALECT IDENTIFICATION	1610
<i>Tianlong Kong, Shouyi Yin, Dawei Zhang, Wang Geng, Xin Wang, Dandan Song, Jinwen Huang, Huiyu Shi, Xiaorui Wang</i>	
ADDITIVE PHONEME-AWARE MARGIN SOFTMAX LOSS FOR LANGUAGE RECOGNITION.....	1615
<i>Zheng Li, Yan Liu, Lin Li, Qingyang Hong</i>	

INTERSPEECH 2021 : PHONATION AND VOICING

AUTOMATIC CLASSIFICATION OF PHONATION TYPES IN SPONTANEOUS SPEECH: TOWARDS A NEW WORKFLOW FOR THE CHARACTERIZATION OF SPEAKERS' VOICE QUALITY	1620
<i>Anais Chanclu, Imen Ben Amor, Cedric Gendrot, Emmanuel Ferragne, Jean-Francois Bonastre</i>	
CUE INTERACTION IN THE PERCEPTION OF PROSODIC PROMINENCE: THE ROLE OF VOICE QUALITY	1624
<i>Bogdan Ludusan, Petra Wagner, Marcin Wlodarczak</i>	
VOICE QUALITY IN VERBAL IRONY: ELECTROGLOTTOGRAPHIC ANALYSES OF IRONIC UTTERANCES IN STANDARD AUSTRIAN GERMAN	1629
<i>Hannah Leykum</i>	
GLOTTAL STOPS IN UPPER SORBIAN: A DATA-DRIVEN APPROACH	1634
<i>Ivan Kraljevski, Maria Paola Bissiri, Frank Duckhorn, Constanze Tschoepe, Matthias Wolff</i>	

GLOTTAL SOUNDS IN KOREBAJU 1639
Jenifer Vega Rodriguez, Nathalie Vallee

SYNCHRONIC FORTITION IN FIVE ROMANCE LANGUAGES? A LARGE CORPUS-
BASED STUDY OF WORD-INITIAL DEVOICING..... 1643
Mathilde Hutin, Yaru Wu, Adele Jatteau, Ioana Vasilescu, Lori Lamel, Martine Adda-Decker

INTERSPEECH 2021 : PHONETICS I

DIALECT FEATURES IN HETEROGENEOUS AND HOMOGENEOUS GHEG SPEAKING
COMMUNITIES..... 1648
Josiane Riverin-Coutlee, Conceicao Cunha, Enkeleida Kapia, Jonathan Harrington

DOMAIN-INITIAL STRENGTHENING IN TURKISH: ACOUSTIC CUES TO PROSODIC
HIERARCHY IN STOP CONSONANTS 1653
Kubra Bodur, Sweeney Branje, Morgane Peirola, Ingrid Tiscareno, James S. German

AN EXPLORATION OF THE ACOUSTIC SPACE OF RHOTICS AND LATERALS IN
RURUULI 1658
Margaret Zellers, Alena Witzlack-Makarevich, Lilja Saeboe, Saudah Namyalo

PROSODIC ACCOMMODATION IN FACE-TO-FACE AND TELEPHONE DIALOGUES 1663
Pavel Sturm, Radek Skarnitzl, Tomas Nechansky

INTERSPEECH 2021 : PHONETICS II

SPEECH PERCEPTION AND LOANWORD ADAPTATIONS: THE CASE OF COPY-VOWEL
EPENTHESIS 1668
Adriana Guevara-Rukoz, Shi Yu, Sharon Peperkamp

QUANTIFYING VOCAL TRACT SHAPE VARIATION AND ITS ACOUSTIC IMPACT: A
GEOMETRIC MORPHOMETRIC APPROACH..... 1673
Amelia J. Gully

REVISITING RECALL EFFECTS OF FILLER PARTICLES IN GERMAN AND ENGLISH..... 1678
*Beeke Muhlack, Mikey Elmers, Heiner Drenhaus, Jurgen Trouvain, Marjolein Van Os,
Raphael Werner, Margarita Ryzhova, Bernd Mobius*

HOW RELIABLE ARE PHONETIC DATA COLLECTED REMOTELY? COMPARISON OF
RECORDING DEVICES AND ENVIRONMENTS ON ACOUSTIC MEASUREMENTS 1683
Chunyu Ge, Yixuan Xiong, Peggy Mok

DEVELOPMENTAL CHANGES OF VOWEL ACOUSTICS IN ADOLESCENTS..... 1688
Einar Meister, Lya Meister

A CROSS-DIALECTAL COMPARISON OF APICAL VOWELS IN BEIJING MANDARIN,
NORTHEASTERN MANDARIN AND SOUTHWESTERN MANDARIN: AN EMA AND
ULTRASOUND STUDY..... 1693
Jing Huang, Feng-Fan Hsieh, Yueh-Chin Chang

DISSECTING THE AERO-ACOUSTIC PARAMETERS OF OPEN ARTICULATORY
TRANSITIONS..... 1698
Mark Gibson, Oihane Muxika, Marianne Pouplier

LEVERAGING REAL-TIME MRI FOR ILLUMINATING LINGUISTIC VELUM ACTION	1703
<i>Miran Oh, Dani Byrd, Shrikanth S. Narayanan</i>	
EXPLORATION OF WELSH ENGLISH PRE-ASPIRATION: HOW WIDE-SPREAD IS IT?.....	1708
<i>Misa Hejna</i>	
THE PACIFIC EXPANSION: OPTIMIZING PHONETIC TRANSCRIPTION OF ARCHIVAL CORPORA	1713
<i>Rosey Billington, Hywel Stoakes, Nick Thieberger</i>	
CONTEXT AND CO-TEXT INFLUENCE ON THE ACCURACY PRODUCTION OF ITALIAN L2 NON-NATIVE SOUNDS	1718
<i>Sonia D'Apolito, Barbara Gili Fivela</i>	
A NEW VOWEL NORMALIZATION FOR SOCIOPHONETICS	1723
<i>Wilbert Heeringa, Hans Van De Velde</i>	
SPEAKERS COARTICULATE LESS WHEN FACING REAL AND IMAGINED COMMUNICATIVE DIFFICULTIES: AN ANALYSIS OF READ AND SPONTANEOUS SPEECH FROM THE LUCID CORPUS.....	1728
<i>Zhe-Chen Guo, Rajka Smiljanic</i>	
SEGMENTAL ALIGNMENT OF ENGLISH SYLLABLES WITH SINGLETON AND CLUSTER ONSETS	1733
<i>Zirui Liu, Yi Xu</i>	
<u>INTERSPEECH 2021 : PRIVACY-PRESERVING MACHINE LEARNING FOR AUDIO & SPEECH PROCESSING</u>	
SYNTHASR: UNLOCKING SYNTHETIC DATA FOR SPEECH RECOGNITION.....	1738
<i>Amin Fazal, Wei Yang, Yulan Liu, Roberto Barra-Chicote, Yixiong Meng, Roland Maas, Jasha Droppo</i>	
PATE-AAE: INCORPORATING ADVERSARIAL AUTOENCODER INTO PRIVATE AGGREGATION OF TEACHER ENSEMBLES FOR SPOKEN COMMAND CLASSIFICATION.....	1743
<i>Chao-Han Huck Yang, Sabato Marco Siniscalchi, Chin-Hui Lee</i>	
CONTINUAL LEARNING FOR FAKE AUDIO DETECTION	1748
<i>Haoxin Ma, Jiangyan Yi, Jianhua Tao, Ye Bai, Zhengkun Tian, Chenglong Wang</i>	
COMMUNICATION-EFFICIENT AGNOSTIC FEDERATED AVERAGING	1753
<i>Jae Ro, Mingqing Chen, Rajiv Mathews, Mehryar Mohri, Ananda Theertha Suresh</i>	
EVALUATING THE VULNERABILITY OF END-TO-END AUTOMATIC SPEECH RECOGNITION MODELS TO MEMBERSHIP INFERENCE ATTACKS	1758
<i>Muhammad A. Shah, Joseph Szurley, Markus Mueller, Athanasios Mouchtaris, Jasha Droppo</i>	
PRIVACY-PRESERVING VOICE ANTI-SPOOFING USING SECURE MULTI-PARTY COMPUTATION	1763
<i>Oubaida Chouchane, Baptiste Brossier, Jorge Esteban Gamboa Gamboa, Thomas Lardy, Hemlata Tak, Orhan Ermis, Madhu R. Kamble, Jose Patino, Nicholas Evans, Melek Onen, Massimiliano Todisco</i>	
CONFIGURABLE PRIVACY-PRESERVING AUTOMATIC SPEECH RECOGNITION	1768
<i>Ranya Aloufi, Hamed Haddadi, David Boyle</i>	

ADJUNCT-EMERITUS DISTILLATION FOR SEMI-SUPERVISED LANGUAGE MODEL ADAPTATION	1773
<i>Scott Novotney, Yile Gu, Ivan Bulyko</i>	

PRIVACY-PRESERVING FEATURE EXTRACTION FOR CLOUD-BASED WAKE WORD VERIFICATION	1778
<i>Timm Koppelman, Alexandru Nelus, Lea Schonherr, Dorothea Kolossa, Rainer Martin</i>	

INTERSPEECH 2021 : PROSODIC FEATURES AND STRUCTURE

AN EXPERIMENT IN PARATONE DETECTION IN A PROSODICALLY ANNOTATED EAP SPOKEN CORPUS	1783
<i>Adrien Meli, Nicolas Ballier, Achille Falaise, Alice Henderson</i>	

ON THE FEASIBILITY OF THE DANISH MODEL OF INTONATIONAL TRANSCRIPTION: PHONETIC EVIDENCE FROM JUTLANDIC DANISH	1788
<i>Anna Bothe Jespersen, Pavel Sturm, Misa Hejna</i>	

PROSODY OF CASE MARKERS IN URDU	1793
<i>Benazir Mumtaz, Massimiliano Canzi, Miriam Butt</i>	

PROSOBEAST PROSODY ANNOTATION TOOL	1798
<i>Branislav Gerazov, Michael Wagner</i>	

ARTICULATORY CHARACTERISTICS OF ICELANDIC VOICED FRICATIVE LENITION: GRADIENCE, CATEGORICITY, AND SPEAKER/GESTURE-SPECIFIC EFFECTS	1803
<i>Brynhildur Stefansdottir, Francesco Burroni, Sam Tilsen</i>	

THE LF MODEL IN THE FREQUENCY DOMAIN FOR GLOTTAL AIRFLOW MODELLING WITHOUT ALIASING DISTORTION	1808
<i>Christer Gobl</i>	

HOW F0 AND PHRASE POSITION AFFECT PAPUAN MALAY WORD IDENTIFICATION	1813
<i>Constantijn Kaland, Matthew Gordon</i>	

TAIWAN MIN NAN (TAIWANESE) CHECKED TONES SOUND CHANGE	1818
<i>Ho-Hsien Pan, Shao-Ren Lyu</i>	

LEVERAGING THE UNIFORMITY FRAMEWORK TO EXAMINE CROSSLINGUISTIC SIMILARITY FOR LONG-LAG STOPS IN SPONTANEOUS CANTONESE-ENGLISH BILINGUAL SPEECH	1823
<i>Khia A. Johnson</i>	

THE INTERACTION OF WORD COMPLEXITY AND WORD DURATION IN AN AGGLUTINATIVE LANGUAGE	1828
<i>Maria Gosy, Kalman Abari</i>	

PARSING SPEECH FOR GROUPING AND PROMINENCE, AND THE TYPOLOGY OF RHYTHM	1833
<i>Michael Wagner, Alvaro Iturralde Zurita, Sijia Zhang</i>	

IN-GROUP ADVANTAGE IN THE PERCEPTION OF EMOTIONS: EVIDENCE FROM THREE VARIETIES OF GERMAN	1838
<i>Moritz Jakob, Bettina Braun, Katharina Zahner-Ritter</i>	

TARGETED AND TARGETLESS NEUTRAL TONES IN TAIWANESE SOUTHERN MIN 1843
Roger Cheng-Yen Liu, Feng-Fan Hsieh, Yueh-Chin Chang

ASSESSING THE USE OF PROSODY IN CONSTITUENCY PARSING OF IMPERFECT
TRANSCRIPTS 1848
Trang Tran, Mari Ostendorf

INTERSPEECH 2021 : RESOURCE-CONSTRAINED ASR

BROADCASTED RESIDUAL LEARNING FOR EFFICIENT KEYWORD SPOTTING..... 1853
Byeonggeun Kim, Simyung Chang, Jinkyu Lee, Dooyong Sung

COMPRESSING 1D TIME-CHANNEL SEPARABLE CONVOLUTIONS USING SPARSE
RANDOM TERNARY MATRICES 1858
Goncalo Mordido, Matthijs Van Keirsbilck, Alexander Keller

PQK: MODEL COMPRESSION VIA PRUNING, QUANTIZATION, AND KNOWLEDGE
DISTILLATION 1863
Jangho Kim, Simyung Chang, Nojun Kwak

AMORTIZED NEURAL NETWORKS FOR LOW-LATENCY SPEECH RECOGNITION..... 1868
Jonathan Macoskey, Grant P. Strimel, Jinru Su, Ariya Rastrow

WEAKLY SUPERVISED CONSTRUCTION OF ASR SYSTEMS FROM MASSIVE VIDEO
DATA..... 1873
Mengli Cheng, Chengyu Wang, Jun Huang, Xiaobo Wang

TIED & REDUCED RNN-T DECODER 1878
Rami Botros, Tara N. Sainath, Robert David, Emmanuel Guzman, Wei Li, Yanzhang He

CODERT: DISTILLING ENCODER REPRESENTATIONS WITH CO-LEARNING FOR
TRANSDUCER-BASED SPEECH RECOGNITION 1883
*Rupak Vignesh Swaminathan, Brian King, Grant P. Strimel, Jasha Droppo, Athanasios
Mouchtaris*

THE ENERGY AND CARBON FOOTPRINT OF TRAINING END-TO-END SPEECH
RECOGNIZERS 1888
Titouan Parcollet, Mirco Ravanelli

COLLABORATIVE TRAINING OF ACOUSTIC ENCODERS FOR SPEECH RECOGNITION 1893
*Varun Nagaraja, Yangyang Shi, Ganesh Venkatesh, Ozlem Kalinli, Michael L. Seltzer, Vikas
Chandra*

EFFICIENT CONFORMER WITH PROB-SPARSE ATTENTION MECHANISM FOR END-
TO-END SPEECH RECOGNITION 1898
Xiong Wang, Sining Sun, Lei Xie, Long Ma

DISSECTING USER-PERCEIVED LATENCY OF ON-DEVICE E2E SPEECH RECOGNITION..... 1903
*Yuan Shanguan, Rohit Prabhavalkar, Hang Su, Jay Mahadeokar, Yangyang Shi, Jiatong
Zhou, Chunyang Wu, Duc Le, Ozlem Kalinli, Christian Fuegen, Michael L. Seltzer*

EXTREMELY LOW FOOTPRINT END-TO-END ASR SYSTEM FOR SMART DEVICE 1908
Zhifu Gao, Yiwu Yao, Shiliang Zhang, Jun Yang, Ming Lei, Ian McLoughlin

INTERSPEECH 2021 : ROBUST AND FAR-FIELD ASR

AGE-INVARIANT TRAINING FOR END-TO-END CHILD SPEECH RECOGNITION USING ADVERSARIAL MULTI-TASK LEARNING	1913
<i>Lars Rumberg, Hanna Ehlert, Ulrike Ludtke, Jorn Ostermann</i>	
SIMULATING READING MISTAKES FOR CHILD SPEECH TRANSFORMER-BASED PHONE RECOGNITION	1918
<i>Lucile Gelin, Thomas Pellegrini, Julien Piquier, Morgane Daniel</i>	
ETLT 2021: SHARED TASK ON AUTOMATIC SPEECH RECOGNITION FOR NON-NATIVE CHILDREN'S SPEECH	1923
<i>R. Gretter, Marco Matassoni, D. Falavigna, A. Misra, C. W. Leong, K. Knill, L. Wang</i>	
SELF-ATTENTION CHANNEL COMBINATOR FRONTEND FOR END-TO-END MULTICHANNEL FAR-FIELD SPEECH RECOGNITION	1928
<i>Rong Gong, Carl Quillen, Dushyant Sharma, Andrew Goderre, Jose Lainez, Ljubomir Milanovic</i>	
LEARNING TO RANK MICROPHONES FOR DISTANT SPEECH RECOGNITION	1933
<i>Samuele Cornell, Alessio Brutti, Marco Matassoni, Stefano Squartini</i>	

INTERSPEECH 2021 : ROBUST SPEAKER RECOGNITION

UNSUPERVISED BAYESIAN ADAPTATION OF PLDA FOR SPEAKER VERIFICATION	1938
<i>Bengt J. Borgstrom</i>	
VARIATIONAL INFORMATION BOTTLENECK BASED REGULARIZATION FOR SPEAKER RECOGNITION	1943
<i>Dan Wang, Yuanjie Dong, Yaxing Li, Yunfei Zi, Zhihui Zhang, Xiaoqi Li, Shengwu Xiong</i>	
GRADIENT REGULARIZATION FOR NOISE-ROBUST SPEAKER VERIFICATION	1948
<i>Jianchen Li, Jiqing Han, Hongwei Song</i>	
SCALING EFFECT OF SELF-SUPERVISED SPEECH MODELS.....	1953
<i>Jie Pu, Yuguang Yang, Ruirui Li, Oguz Elibol, Jasha Droppo</i>	
SPEAKER ANONYMISATION USING THE MCADAMS COEFFICIENT.....	1958
<i>Jose Patino, Natalia Tomashenko, Massimiliano Todisco, Andreas Nautsch, Nicholas Evans</i>	
MULTI-LEVEL TRANSFER LEARNING FROM NEAR-FIELD TO FAR-FIELD SPEAKER VERIFICATION	1963
<i>Li Zhang, Qing Wang, Kong Aik Lee, Lei Xie, Haizhou Li</i>	
OUT OF A HUNDRED TRIALS, HOW MANY ERRORS DOES YOUR SPEAKER VERIFIER MAKE?	1968
<i>Niko Brummer, Luciana Ferrer, Albert Swart</i>	
SPEAKERSTEW: SCALING TO MANY LANGUAGES WITH A TRIAGED MULTILINGUAL TEXT-DEPENDENT AND TEXT-INDEPENDENT SPEAKER VERIFICATION SYSTEM	1973
<i>Roza Chojnacka, Jason Pelecanos, Quan Wang, Ignacio Lopez Moreno</i>	
DEEP FEATURE CYCLEGANS: SPEAKER IDENTITY PRESERVING NON-PARALLEL MICROPHONE-TELEPHONE DOMAIN ADAPTATION FOR SPEAKER VERIFICATION	1978
<i>Saurabh Kataria, Jesus Villalba, Piotr Zelasko, Laureano Moro-Velazquez, Najim Dehak</i>	

THE DKU-DUKE-LENOVO SYSTEM DESCRIPTION FOR THE FEARLESS STEPS CHALLENGE PHASE III	1983
<i>Weiqing Wang, Danwei Cai, Jin Wang, Qingjian Lin, Xuyang Wang, Mi Hong, Ming Li</i>	
IMPROVED META-LEARNING TRAINING FOR SPEAKER VERIFICATION.....	1988
<i>Yafeng Chen, Wu Guo, Bin Gu</i>	
JOINT FEATURE ENHANCEMENT AND SPEAKER RECOGNITION WITH MULTI- OBJECTIVE TASK-ORIENTED NETWORK	1993
<i>Yibo Wu, Longbiao Wang, Kong Aik Lee, Meng Liu, Jianwu Dang</i>	
ANTVOICE NEURAL SPEAKER EMBEDDING SYSTEM FOR FFSVC 2020.....	1998
<i>Zhiming Wang, Furong Xu, Kaisheng Yao, Yuan Cheng, Tao Xiong, Huijia Zhu</i>	
<u>INTERSPEECH 2021 : SDSV CHALLENGE 2021: ANALYSIS AND EXPLORATION OF NEW IDEAS ON SHORT-DURATION SPEAKER VERIFICATION</u>	
SDSVC CHALLENGE 2021: TIPS AND TRICKS TO BOOST THE SHORT-DURATION SPEAKER VERIFICATION SYSTEM PERFORMANCE	2003
<i>Aleksei Gusev, Alisa Vinogradova, Sergey Novoselov, Sergei Astapov</i>	
THE ID R&D SYSTEM DESCRIPTION FOR SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2021	2008
<i>Alexander Alenin, Anton Okhotnikov, Rostislav Makarov, Nikita Torgashov, Ilya Shigabeev, Konstantin Simonchik</i>	
THE SJTU SYSTEM FOR SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2021	2013
<i>Bing Han, Zhengyang Chen, Zhikai Zhou, Yanmin Qian</i>	
INTEGRATING FREQUENCY TRANSLATIONAL INVARIANCE IN TDNNS AND FREQUENCY POSITIONAL INFORMATION IN 2D RESNETS TO ENHANCE SPEAKER VERIFICATION	2018
<i>Jenthe Thienpondt, Brecht Desplanques, Kris Demuyneck</i>	
THE SOGOU SYSTEM FOR SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2021	2023
<i>Jie Yan, Shengyu Yao, Yiqian Pan, Wei Chen</i>	
INVESTIGATION OF IMU&ELEVOC SUBMISSION FOR THE SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2021.....	2028
<i>Peng Zhang, Peng Hu, Xueliang Zhang</i>	
TEAM02 TEXT-INDEPENDENT SPEAKER VERIFICATION SYSTEM FOR SDSV CHALLENGE 2021	2033
<i>Woo Hyun Kang, Nam Soo Kim</i>	
OUR LEARNED LESSONS FROM CROSS-LINGUAL SPEAKER VERIFICATION: THE CRMI-DKU SYSTEM DESCRIPTION FOR THE SHORT-DURATION SPEAKER VERIFICATION CHALLENGE 2021	2038
<i>Xiaoyi Qin, Chao Wang, Yong Ma, Min Liu, Shilei Zhang, Ming Li</i>	

INTERSPEECH 2021 : SEARCH/DECODING TECHNIQUES AND CONFIDENCE MEASURES FOR ASR

SEQUENCE-LEVEL CONFIDENCE CLASSIFIER FOR ASR UTTERANCE ACCURACY AND APPLICATION TO ACOUSTIC MODELS.....	2043
<i>Amber Afshan, Kshitiz Kumar, Jian Wu</i>	
INSIGHTS ON NEURAL REPRESENTATIONS FOR END-TO-END SPEECH RECOGNITION	2048
<i>Anna Ollerenshaw, Md. Asif Jalal, Thomas Hain</i>	
LT-LM: A NOVEL NON-AUTOREGRESSIVE LANGUAGE MODEL FOR SINGLE-SHOT LATTICE RESCORING.....	2053
<i>Anton Mitrofanov, Mariya Korenevskaya, Ivan Podluzhny, Yuri Khokhlov, Aleksandr Laptev, Andrei Andrusenko, Aleksei Ilin, Maxim Korenevsky, Ivan Medennikov, Aleksei Romanenko</i>	
A HYBRID SEQ-2-SEQ ASR DESIGN FOR ON-DEVICE AND SERVER APPLICATIONS.....	2058
<i>Cyril Allauzen, Ehsan Variiani, Michael Riley, David Rybach, Hao Zhang</i>	
MULTI-TASK LEARNING FOR END-TO-END ASR WORD AND UTTERANCE CONFIDENCE WITH DELETION PREDICTION	2063
<i>David Qiu, Yanzhang He, Qiuqia Li, Yu Zhang, Liangliang Cao, Ian McGraw</i>	
VAD-FREE STREAMING HYBRID CTC/ATTENTION ASR FOR UNSEGMENTED RECORDING.....	2068
<i>Hirofumi Inaguma, Tatsuya Kawahara</i>	
DEEP NEURAL NETWORK CALIBRATION FOR E2E SPEECH RECOGNITION SYSTEM	2073
<i>Mun-Hak Lee, Joon-Hyuk Chang</i>	
RESIDUAL ENERGY-BASED MODELS FOR END-TO-END SPEECH RECOGNITION.....	2078
<i>Qiuqia Li, Yu Zhang, Bo Li, Liangliang Cao, Philip C. Woodland</i>	
CROSS-MODAL TRANSFORMER-BASED NEURAL CORRECTION MODELS FOR AUTOMATIC SPEECH RECOGNITION	2083
<i>Tomohiro Tanaka, Ryo Masumura, Mana Ithori, Akihiko Takashima, Takafumi Moriya, Takanori Ashihara, Shota Orihashi, Naoki Makishima</i>	
FSR: ACCELERATING THE INFERENCE PROCESS OF TRANSDUCER-BASED MODELS BY APPLYING FAST-SKIP REGULARIZATION	2088
<i>Zhengkun Tian, Jiangyan Yi, Ye Bai, Jianhua Tao, Shuai Zhang, Zhengqi Wen</i>	
WENET: PRODUCTION ORIENTED STREAMING AND NON-STREAMING END-TO-END SPEECH RECOGNITION TOOLKIT.....	2093
<i>Zhuoyuan Yao, Di Wu, Xiong Wang, Binbin Zhang, Fan Yu, Chao Yang, Zhendong Peng, Xiaoyu Chen, Lei Xie, Xin Lei</i>	

INTERSPEECH 2021 : SELF-SUPERVISION AND SEMI-SUPERVISION FOR NEURAL ASR TRAINING

A COMPARISON OF SUPERVISED AND UNSUPERVISED PRE-TRAINING OF END-TO-END MODELS	2098
<i>Ananya Misra, Dongseong Hwang, Zhouyuan Huo, Shefali Garg, Nikhil Siddhartha, Arun Narayanan, Khe Chai Sim</i>	

ON THE LEARNING DYNAMICS OF SEMI-SUPERVISED TRAINING FOR ASR	2103
<i>Electra Wallington, Benji Kershenbaum, Ondrej Klejch, Peter Bell</i>	
WAV2VEC-C: A SELF-SUPERVISED MODEL FOR SPEECH REPRESENTATION LEARNING.....	2108
<i>Samik Sadhu, Di He, Che-Wei Huang, Sri Harish Mallidi, Minhua Wu, Ariya Rastrow, Andreas Stolcke, Jasha Droppo, Roland Maas</i>	
IMPROVING STREAMING TRANSFORMER BASED ASR UNDER A FRAMEWORK OF SELF-SUPERVISED LEARNING	2113
<i>Songjun Cao, Yueteng Kang, Yanzhe Fu, Xiaoshuo Xu, Sining Sun, Yike Zhang, Long Ma</i>	
SLIMIPL: LANGUAGE-MODEL-FREE ITERATIVE PSEUDO-LABELING.....	2118
<i>Tatiana Likhomanenko, Qiantong Xu, Jacob Kahn, Gabriel Synnaeve, Ronan Collobert</i>	
ROBUST WAV2VEC 2.0: ANALYZING DOMAIN SHIFT IN SELF-SUPERVISED PRE- TRAINING.....	2123
<i>Wei-Ning Hsu, Anuroop Sriram, Alexei Baevski, Tatiana Likhomanenko, Qiantong Xu, Vineel Pratap, Jacob Kahn, Ann Lee, Ronan Collobert, Gabriel Synnaeve, Michael Auli</i>	
PHONETICALLY MOTIVATED SELF-SUPERVISED SPEECH REPRESENTATION LEARNING.....	2128
<i>Xianghu Yue, Haizhou Li</i>	
IMPROVING RNN-T FOR DOMAIN SCALING USING SEMI-SUPERVISED TRAINING WITH NEURAL TTS	2133
<i>Yan Deng, Rui Zhao, Zhong Meng, Xie Chen, Bing Liu, Jinyu Li, Yifan Gong, Lei He</i>	
MOMENTUM PSEUDO-LABELING FOR SEMI-SUPERVISED SPEECH RECOGNITION.....	2138
<i>Yosuke Higuchi, Niko Moritz, Jonathan Le Roux, Takaaki Hori</i>	
SEMI-SUPERVISION IN ASR: SEQUENTIAL MIXMATCH AND FACTORIZED TTS-BASED AUGMENTATION.....	2143
<i>Zhehuai Chen, Andrew Rosenberg, Yu Zhang, Heiga Zen, Mohammadreza Ghodsi, Yinghui Huang, Jesse Emond, Gary Wang, Bhuvana Ramabhadran, Pedro J. Moreno</i>	
<u>INTERSPEECH 2021 : SHOW AND TELL 1</u>	
THE LIUM HUMAN ACTIVE CORRECTION PLATFORM FOR SPEAKER DIARIZATION.....	2148
<i>Alexandre Flucha, Anthony Larcher, Ambuj Mehrish, Sylvain Meignier, Florian Plaut, Nicolas Poupon, Yevhenii Prokopalo, Adrien Puertolas, Meysam Shamsi, Marie Tahon</i>	
APPLICATION FOR DETECTING DEPRESSION, PARKINSON'S DISEASE AND DYSPHONIC SPEECH	2150
<i>Gabor Kiss, David Sztaho, Miklos Gabriel Tulics</i>	
ADVANCED SEMI-BLIND SPEAKER EXTRACTION AND TRACKING IMPLEMENTED IN EXPERIMENTAL DEVICE WITH REVOLVING DENSE MICROPHONE ARRAY	2152
<i>J. Cmejla, T. Kounovsky, J. Jansky, Jiri Malek, M. Rozkovec, Z. Koldovsky</i>	
BEEY: MORE THAN A SPEECH-TO-TEXT EDITOR.....	2154
<i>Lenka Weingartova, Veronika Volna, Ewa Balejova</i>	
ROXANNE RESEARCH PLATFORM: AUTOMATE CRIMINAL INVESTIGATIONS.....	2156
<i>Mael Fabien, Shantipriya Parida, Petr Motliceck, Dawei Zhu, Aravind Krishnan, Hoang H. Nguyen</i>	

Downsizing of vocal-tract models to line up variations and reduce manufacturing costs.....	2159
<i>Takayuki Arai</i>	

On-device streaming transformer-based end-to-end speech recognition.....	2161
<i>Yoo Rhee Oh, Kiyoun Park</i>	

INTERSPEECH 2021 : SHOW AND TELL 2

Live TV subtitling through respeaking.....	2163
<i>Ales Prazak, Zdenek Loose, Josef V. Psutka, Vlasta Radova, Josef Psutka, Jan Svec</i>	

Expressive robot performance based on facial motion capture.....	2165
<i>Jonas Beskow, Charlie Caper, Johan Ehrenfors, Nils Hagberg, Anne Jansen, Chris Wood</i>	

Audio segmentation based conversational silence detection for contact center calls.....	2167
<i>Krishnachaitanya Gogineni, Tarun Reddy Yadama, Jithendra Vepa</i>	

Themepro 2.0: Showcasing the role of thematic progression in engaging human-computer interaction.....	2169
<i>Monica Dominguez, Juan Soler-Company, Leo Wanner</i>	

Addressing compliance in call centers with entity extraction.....	2171
<i>Sai Gururu, Jithendra Vepa</i>	

Autonomous robot for measuring room impulse responses.....	2173
<i>Stefan Fagner, Tobias Topar, Maximilian Giller, Lukas Pfeifenberger, Franz Pernkopf</i>	

Multi-speaker emotional text-to-speech synthesizer.....	2175
<i>Sungjae Cho, Soo-Young Lee</i>	

INTERSPEECH 2021 : SHOW AND TELL 3

Articulatory data recorder: A framework for real-time articulatory data recording.....	2177
<i>Alexander Wilbrandt, Simon Stone, Peter Birkholz</i>	

MOM: Minutes of Meeting Bot.....	2179
<i>Benjamin Milde, Tim Fischer, Steffen Remus, Chris Biemann</i>	

Expressive Latvian speech synthesis for dialog systems.....	2181
<i>Davis Nicmanis, Askars Salimbajevs</i>	

The ingenious multilingual operations app.....	2183
<i>Joan Codina-Filba, Guillermo Cambara, Alex Peiro-Lilja, Jens Grivolla, Roberto Carlini, Mireia Farrus</i>	

Digital Einstein Experience: Fast text-to-speech for conversational AI.....	2185
<i>Joanna Rownicka, Kilian Sprenkamp, Antonio Tripiana, Volodymyr Gromoglasov, Timo P. Kunz</i>	

VISTAFAE: A VISUAL SPEECH-TRAINING AID WITH FEEDBACK OF ARTICULATORY EFFORTS.....	2187
<i>Pramod H. Kachare, Prem C. Pandey, Vishal Mane, HIRAK DASGUPTA, K. S. Nataraj, Akshada Rathod, Sheetal K. Pathak</i>	

LIVE SUBTITLING FOR BIGBLUEBUTTON WITH OPEN-SOURCE SOFTWARE	2189
<i>Robert Geislinger, Benjamin Milde, Timo Baumann, Chris Biemann</i>	

INTERSPEECH 2021 : SHOW AND TELL 4

DUPLEX CONVERSATION IN OUTBOUND AGENT SYSTEM.....	2191
<i>Chunxiang Jin, Minghui Yang, Zujie Wen</i>	

LALILO: A READING ASSISTANT FOR CHILDREN FEATURING SPEECH RECOGNITION-BASED READING MISTAKE DETECTION.....	2193
<i>Corentin Hembise, Lucile Gelin, Morgane Daniel</i>	

SAVE YOUR VOICE: VOICE BANKING AND TTS FOR ANYONE	2195
<i>Daniel Tihelka, Marketa Rezaczkova, Martin Gruber, Zdenek Hanzlicek, Jakub Vit, Jindrich Matousek</i>	

INTERACTIVE AND REAL-TIME ACOUSTIC MEASUREMENT TOOLS FOR SPEECH DATA ACQUISITION AND PRESENTATION: APPLICATION OF AN EXTENDED MEMBER OF TIME STRETCHED PULSES.....	2197
<i>Hideki Kawahara, Kohei Yatabe, Ken-Ichi Sakakibara, Mitsunori Mizumachi, Masanori Morise, Hideki Banno, Toshio Irino</i>	

WITTYKIDDY: MULTILINGUAL SPOKEN LANGUAGE LEARNING FOR KIDS	2199
<i>Ke Shi, Kye Min Tan, Huayun Zhang, Siti Umairah Md. Salleh, Shikang Ni, Nancy F. Chen</i>	

AUTOMATIC RADIOLOGY REPORT EDITING THROUGH VOICE.....	2201
<i>Manh Hung Nguyen, Vu Hoang, Tu Anh Nguyen, Trung H. Bui</i>	

WEB INTERFACE FOR ESTIMATING ARTICULATORY MOVEMENTS IN SPEECH PRODUCTION FROM ACOUSTICS AND TEXT	2203
<i>Sathvik Udupa, Anwasha Roy, Abhayjeet Singh, Aravind Illa, Prasanta Kumar Ghosh</i>	

NEMO (INVERSE) TEXT NORMALIZATION: FROM DEVELOPMENT TO PRODUCTION.....	2205
<i>Yang Zhang, Evelina Bakhturina, Boris Ginsburg</i>	

INTERSPEECH 2021 : SINGLE-CHANNEL SPEECH ENHANCEMENT

PERSONALIZED SPEECH ENHANCEMENT THROUGH SELF-SUPERVISED DATA AUGMENTATION AND PURIFICATION	2208
<i>Aswin Sivaraman, Sunwoo Kim, Minje Kim</i>	

SE-CONFORMER: TIME-DOMAIN SPEECH ENHANCEMENT USING CONFORMER.....	2213
<i>Eesung Kim, Hyeji Seo</i>	

IMPROVED SPEECH ENHANCEMENT USING A COMPLEX-DOMAIN GAN WITH FUSED TIME-DOMAIN AND TIME-FREQUENCY DOMAIN CONSTRAINTS	2218
<i>Feng Dang, Pengyuan Zhang, Hangting Chen</i>	

DEMUCS-MOBILE : ON-DEVICE LIGHTWEIGHT SPEECH ENHANCEMENT.....	2223
<i>Lukas Lee, Youna Ji, Minjae Lee, Min-Seok Choi</i>	

SPEECH DENOISING WITHOUT CLEAN TRAINING DATA: A NOISE2NOISE APPROACH	2228
<i>Madhav Mahesh Kashyap, Anuj Tambwekar, Krishnamoorthy Manohara, S. Natarajan</i>	
SPEECH DENOISING WITH AUDITORY MODELS	2233
<i>Mark R. Sandler, Andrew Senior, Jenelle Feather, Kaizhi Qian, Yang Zhang, Josh H. McDermott</i>	
SINGLE-CHANNEL SPEECH ENHANCEMENT USING LEARNABLE LOSS MIXUP	2238
<i>Oscar Chang, Dung N. Tran, Kazuhito Koishida</i>	
HUMAN LISTENING AND LIVE CAPTIONING: MULTI-TASK TRAINING FOR SPEECH ENHANCEMENT.....	2243
<i>Sefik Emre Eskimez, Xiaofei Wang, Min Tang, Hemin Yang, Zirun Zhu, Zhuo Chen, Huaming Wang, Takuya Yoshioka</i>	
LEARNING SPEECH STRUCTURE TO IMPROVE TIME-FREQUENCY MASKS	2248
<i>Sulian Bu, Yunxin Zhao, Shaojun Wang, Mei Han</i>	
WHISPER SPEECH ENHANCEMENT USING JOINT VARIATIONAL AUTOENCODER FOR IMPROVED SPEECH RECOGNITION	2253
<i>Vikas Agrawal, Shashi Kumar, Shakti P. Rath</i>	
A MAXIMUM LIKELIHOOD APPROACH TO SNR-PROGRESSIVE LEARNING USING GENERALIZED GAUSSIAN DISTRIBUTION FOR LSTM-BASED SPEECH ENHANCEMENT.....	2258
<i>Xiao-Qi Zhang, Jun Du, Li Chai, Chin-Hui Lee</i>	
MULTI-STAGE PROGRESSIVE SPEECH ENHANCEMENT NETWORK.....	2263
<i>Xinmeng Xu, Yang Wang, Dongxiang Xu, Yiyuan Peng, Cong Zhang, Jie Jia, Binbin Chen</i>	
SPEECH ENHANCEMENT WITH TOPOLOGY-ENHANCED GENERATIVE ADVERSARIAL NETWORKS (GANS)	2268
<i>Xudong Zhang, Liang Zhao, Feng Gu</i>	
<u>INTERSPEECH 2021 : SOURCE SEPARATION I</u>	
CONTINUOUS SPEECH SEPARATION USING SPEAKER INVENTORY FOR LONG RECORDING.....	2273
<i>Cong Han, Yi Luo, Chenda Li, Tianyan Zhou, Keisuke Kinoshita, Shinji Watanabe, Marc Delcroix, Hakan Erdogan, John R. Hershey, Nima Mesgarani, Zhuo Chen</i>	
DUAL-PATH FILTER NETWORK: SPEAKER-AWARE MODELING FOR SPEECH SEPARATION	2278
<i>Fan-Lin Wang, Yu-Huai Peng, Hung-Shin Lee, Hsin-Min Wang</i>	
INVESTIGATION OF PRACTICAL ASPECTS OF SINGLE CHANNEL SPEECH SEPARATION FOR ASR.....	2283
<i>Jian Wu, Zhuo Chen, Sanyuan Chen, Yu Wu, Takuya Yoshioka, Naoyuki Kanda, Shujie Liu, Jinyu Li</i>	
END-TO-END SPEECH SEPARATION USING ORTHOGONAL REPRESENTATION IN COMPLEX AND REAL TIME-FREQUENCY DOMAIN.....	2288
<i>Kai Wang, Hao Huang, Ying Hu, Zhihua Huang, Sheng Li</i>	

GROUP DELAY BASED RE-WEIGHTED SPARSE RECOVERY ALGORITHMS FOR ROBUST AND HIGH-RESOLUTION SOURCE SEPARATION IN DOA FRAMEWORK.....	2293
<i>Murtiza Ali, Ashwani Koul, Karan Nathwani</i>	
ULTRA FAST SPEECH SEPARATION MODEL WITH TEACHER STUDENT LEARNING	2298
<i>Sanyuan Chen, Yu Wu, Zhuo Chen, Jian Wu, Takuya Yoshioka, Shujie Liu, Jinyu Li, Xiangzhan Yu</i>	
STABILIZING LABEL ASSIGNMENT FOR SPEECH SEPARATION BY SELF-SUPERVISED PRE-TRAINING	2303
<i>Sung-Feng Huang, Shun-Po Chuang, Da-Rong Liu, Yi-Chen Chen, Gene-Ping Yang, Hung-Yi Lee</i>	
CROSSFIRE CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS FOR SINGING VOICE EXTRACTION	2308
<i>Weitao Yuan, Shengbei Wang, Xiangrui Li, Masashi Unoki, Wenwu Wang</i>	
IMPLICIT FILTER-AND-SUM NETWORK FOR END-TO-END MULTI-CHANNEL SPEECH SEPARATION	2313
<i>Yi Luo, Nima Mesgarani</i>	
GENERALIZED SPATIO-TEMPORAL RNN BEAMFORMER FOR TARGET SPEECH SEPARATION	2318
<i>Yong Xu, Zhuohuang Zhang, Meng Yu, Shi-Xiong Zhang, Dong Yu</i>	
EFFICIENT AND STABLE ADVERSARIAL LEARNING USING UNPAIRED DATA FOR UNSUPERVISED MULTICHANNEL SPEECH SEPARATION	2323
<i>Yu Nakagome, Masahito Togami, Tetsuji Ogawa, Tetsunori Kobayashi</i>	
<u>INTERSPEECH 2021 : SOURCE SEPARATION II</u>	
ONLINE BLIND AUDIO SOURCE SEPARATION USING RECURSIVE EXPECTATION-MAXIMIZATION.....	2328
<i>Aviad Eisenberg, Boaz Schwartz, Sharon Gannot</i>	
ROBUST SPEAKER EXTRACTION NETWORK BASED ON ITERATIVE REFINED ADAPTATION	2333
<i>Chengyun Deng, Shiqian Ma, Yongtao Sha, Yi Zhang, Hui Zhang, Hui Song, Fei Wang</i>	
BINAURAL SPEECH SEPARATION OF MOVING SPEAKERS WITH PRESERVED SPATIAL CUES.....	2338
<i>Cong Han, Yi Luo, Nima Mesgarani</i>	
TEACHER-STUDENT MIXIT FOR UNSUPERVISED AND SEMI-SUPERVISED SPEECH SEPARATION	2343
<i>Jisi Zhang, Catalin Zorila, Rama Doddipatla, Jon Barker</i>	
FEW-SHOT LEARNING OF NEW SOUND CLASSES FOR TARGET SOUND EXTRACTION	2348
<i>Marc Delcroix, Jorge Bennisar Vazquez, Tsubasa Ochiai, Keisuke Kinoshita, Shoko Araki</i>	
SPEAKER VERIFICATION-BASED EVALUATION OF SINGLE-CHANNEL SPEECH SEPARATION	2353
<i>Matthew Maciejewski, Shinji Watanabe, Sanjeev Khudanpur</i>	
DEEP AUDIO-VISUAL SPEECH SEPARATION BASED ON FACIAL MOTION	2358
<i>Remi Rigal, Jacques Chodorowski, Benoit Zerr</i>	

VOCAL HARMONY SEPARATION USING TIME-DOMAIN NEURAL NETWORKS	2363
<i>Saurjya Sarkar, Emmanouil Benetos, Mark Sandler</i>	
AVATR: ONE-SHOT SPEAKER EXTRACTION WITH TRANSFORMERS	2368
<i>Shell Xu Hu, Md. Rifat Arefin, Viet-Nhat Nguyen, Alish Dipani, Xaq Pitkow, Andreas Savas Talias</i>	
GRAPH-PIT: GENERALIZED PERMUTATION INVARIANT TRAINING FOR CONTINUOUS SEPARATION OF ARBITRARY NUMBERS OF SPEAKERS	2373
<i>Thilo Von Neumann, Keisuke Kinoshita, Christoph Boeddeker, Marc Delcroix, Reinhold Haeb-Umbach</i>	
IMPROVED SPEECH SEPARATION WITH TIME-AND-FREQUENCY CROSS-DOMAIN FEATURE SELECTION	2378
<i>Tian Lan, Yuxin Qian, Yilan Lyu, Refuoe Mokhosi, Wenxin Tai, Qiao Liu</i>	
NEURAL SPEAKER EXTRACTION WITH SPEAKER-SPEECH CROSS-ATTENTION NETWORK	2383
<i>Wupeng Wang, Chenglin Xu, Meng Ge, Haizhou Li</i>	

VOLUME 4

EMPIRICAL ANALYSIS OF GENERALIZED ITERATIVE SPEECH SEPARATION NETWORKS	2388
<i>Yi Luo, Cong Han, Nima Mesgarani</i>	

INTERSPEECH 2021 : SOURCE SEPARATION III

A HANDS-ON COMPARISON OF DNNS FOR DIALOG SEPARATION USING TRANSFER LEARNING FROM MUSIC SOURCE SEPARATION	2393
<i>Martin Strauss, Jouni Paulus, Matteo Torcoli, Bernd Edler</i>	
GLOBALPHONE MIX-TO-SEPARATE OUT OF 2: A MULTILINGUAL 2000 SPEAKERS MIXTURES DATABASE FOR SPEECH SEPARATION	2398
<i>Marvin Borsdorf, Chenglin Xu, Haizhou Li, Tanja Schultz</i>	
COMBATING REVERBERATION IN NTF-BASED SPEECH SEPARATION USING A SUB-SOURCE WEIGHTED MULTICHANNEL WIENER FILTER AND LINEAR PREDICTION	2403
<i>Mieszko Fras, Marcin Witkowski, Konrad Kowalczyk</i>	
MANY-SPEAKERS SINGLE CHANNEL SPEECH SEPARATION WITH OPTIMAL PERMUTATION TRAINING	2408
<i>Shaked Dovrat, Eliya Nachmani, Lior Wolf</i>	

INTERSPEECH 2021 : SOURCE SEPARATION, DEREVERBERATION AND ECHO CANCELLATION

A DEEP LEARNING APPROACH TO MULTI-CHANNEL AND MULTI-MICROPHONE ACOUSTIC ECHO CANCELLATION	2413
<i>Hao Zhang, Deliang Wang</i>	

TECANET: TEMPORAL-CONTEXTUAL ATTENTION NETWORK FOR ENVIRONMENT-AWARE SPEECH DEREVERBERATION	2418
<i>Helin Wang, Bo Wu, Lianwu Chen, Meng Yu, Jianwei Yu, Yong Xu, Shi-Xiong Zhang, Chao Weng, Dan Su, Dong Yu</i>	
SHOULD WE ALWAYS SEPARATE?: SWITCHING BETWEEN ENHANCED AND OBSERVED SIGNALS FOR OVERLAPPING SPEECH RECOGNITION	2423
<i>Hiroshi Sato, Tsubasa Ochiai, Marc Delcroix, Keisuke Kinoshita, Takafumi Moriya, Naoyuki Kamo</i>	
RESIDUAL ECHO AND NOISE CANCELLATION WITH FEATURE ATTENTION MODULE AND MULTI-DOMAIN LOSS FUNCTION	2428
<i>Jianjun Gu, Longbiao Cheng, Xingwei Sun, Junfeng Li, Yonghong Yan</i>	
MANIFOLD-AWARE DEEP CLUSTERING: MAXIMIZING ANGLES BETWEEN EMBEDDING VECTORS BASED ON REGULAR SIMPLEX	2433
<i>Keitaro Tanaka, Ryosuke Sawata, Shusuke Takahashi</i>	
PERSONALIZED PERCEPNET: REAL-TIME, LOW-COMPLEXITY TARGET VOICE SEPARATION AND ENHANCEMENT	2438
<i>Ritwik Giri, Shrikant Venkataramani, Jean-Marc Valin, Umut Isik, Arvindh Krishnaswamy</i>	
MIMO SELF-ATTENTIVE RNN BEAMFORMER FOR MULTI-SPEAKER SPEECH SEPARATION	2443
<i>Xiyun Li, Yong Xu, Meng Yu, Shi-Xiong Zhang, Jiaming Xu, Bo Xu, Dong Yu</i>	
MULTI-STREAM GATED AND PYRAMIDAL TEMPORAL CONVOLUTIONAL NEURAL NETWORKS FOR AUDIO-VISUAL SPEECH SEPARATION IN MULTI-TALKER ENVIRONMENTS	2448
<i>Yiyu Luo, Jing Wang, Liang Xu, Lidong Yang</i>	
SCENE-AGNOSTIC MULTI-MICROPHONE SPEECH DEREVERBERATION.....	2453
<i>Yochai Yemini, Ethan Fetaya, Haggai Maron, Sharon Gannot</i>	
JOINT ONLINE MULTICHANNEL ACOUSTIC ECHO CANCELLATION, SPEECH DEREVERBERATION AND SOURCE SEPARATION.....	2458
<i>Yueyue Na, Ziteng Wang, Zhang Liu, Biao Tian, Qiang Fu</i>	

INTERSPEECH 2021 : SPEAKER DIARIZATION I

END-TO-END SPEAKER SEGMENTATION FOR OVERLAP-AWARE RESEGMENTATION.....	2463
<i>Herve Bredin, Antoine Laurent</i>	
THREE-CLASS OVERLAPPED SPEECH DETECTION USING A CONVOLUTIONAL RECURRENT NEURAL NETWORK	2468
<i>Jee-Weon Jung, Hee-Soo Heo, Youngki Kwon, Joon Son Chung, Bong-Jin Lee</i>	
A THOUSAND WORDS ARE WORTH MORE THAN ONE RECORDING: WORD-EMBEDDING BASED SPEAKER CHANGE DETECTION.....	2473
<i>Or Haim Anidjar, Itshak Lapidot, Chen Hajaj, Amit Dvir</i>	
ONLINE SPEAKER DIARIZATION EQUIPPED WITH DISCRIMINATIVE MODELING AND GUIDED INFERENCE.....	2478
<i>Xucheng Wan, Kai Liu, Huan Zhou</i>	

ONLINE STREAMING END-TO-END NEURAL DIARIZATION HANDLING OVERLAPPING SPEECH AND FLEXIBLE NUMBERS OF SPEAKERS.....	2483
<i>Yawen Xue, Shota Horiguchi, Yusuke Fujita, Yuki Takashima, Shinji Watanabe, Leibny Paola Garcia Perera, Kenji Nagamatsu</i>	
END-TO-END NEURAL DIARIZATION: FROM TRANSFORMER TO CONFORMER	2488
<i>Yi Chieh Liu, Eunjung Han, Chul Lee, Andreas Stolcke</i>	
ADAPTING SPEAKER EMBEDDINGS FOR SPEAKER DIARISATION	2493
<i>Youngki Kwon, Jee-Weon Jung, Hee-Soo Heo, You Jin Kim, Bong-Jin Lee, Joon Son Chung</i>	
SEMI-SUPERVISED TRAINING WITH PSEUDO-LABELING FOR END-TO-END NEURAL DIARIZATION	2498
<i>Yuki Takashima, Yusuke Fujita, Shota Horiguchi, Shinji Watanabe, Leibny Paola Garcia Perera, Kenji Nagamatsu</i>	
SCENARIO-DEPENDENT SPEAKER DIARIZATION FOR DIHARD-III CHALLENGE.....	2503
<i>Yu-Xuan Wang, Jun Du, Maokui He, Shu-Tong Niu, Lei Sun, Chin-Hui Lee</i>	

INTERSPEECH 2021 : SPEAKER DIARIZATION II

ANONYMOUS SPEAKER CLUSTERS: MAKING DISTINCTIONS BETWEEN ANONYMISED SPEECH RECORDINGS WITH CLUSTERING INTERFACE	2508
<i>Benjamin O'Brien, Natalia Tomashenko, Anais Chanclu, Jean-Francois Bonastre</i>	
ADVANCES IN INTEGRATION OF END-TO-END NEURAL AND CLUSTERING-BASED DIARIZATION FOR REAL CONVERSATIONAL SPEECH	2513
<i>Keisuke Kinoshita, Marc Delcroix, Naohiro Tawara</i>	
SPEAKER DIARIZATION USING TWO-PASS LEAVE-ONE-OUT GAUSSIAN PLDA CLUSTERING OF DNN EMBEDDINGS	2518
<i>Kiran Karra, Alan McCree</i>	
TARGET-SPEAKER VOICE ACTIVITY DETECTION WITH IMPROVED I-VECTOR ESTIMATION FOR UNKNOWN NUMBER OF SPEAKER	2523
<i>Maokui He, Desh Raj, Zili Huang, Jun Du, Zhuo Chen, Shinji Watanabe</i>	
ECAPA-TDNN EMBEDDINGS FOR SPEAKER DIARIZATION	2528
<i>Nauman Dawalatabad, Mirco Ravanelli, Francois Grondin, Jenthe Thienpondt, Brecht Desplanques, Hwidong Na</i>	
THE THIRD DIHARD DIARIZATION CHALLENGE	2533
<i>Neville Ryant, Prachi Singh, Venkat Krishnamohan, Rajat Varma, Kenneth Church, Christopher Cieri, Jun Du, Sriram Ganapathy, Mark Liberman</i>	
LEAP SUBMISSION FOR THE THIRD DIHARD DIARIZATION CHALLENGE	2538
<i>Prachi Singh, Rajat Varma, Venkat Krishnamohan, Srikanth Raj Chetupalli, Sriram Ganapathy</i>	
INVESTIGATION OF SPATIAL-ACOUSTIC FEATURES FOR OVERLAPPING SPEECH DETECTION IN MULTIPARTY MEETINGS.....	2543
<i>Shiliang Zhang, Siqi Zheng, Weilong Huang, Ming Lei, Hongbin Suo, Jinwei Feng, Zhijie Yan</i>	
ROBUST END-TO-END SPEAKER DIARIZATION WITH CONFORMER AND ADDITIVE MARGIN PENALTY.....	2548
<i>Tsun-Yat Leung, Lahiru Samarakoon</i>	

INTERSPEECH 2021 : SPEAKER RECOGNITION: APPLICATIONS

PRESENTATION MATTERS: EVALUATING SPEAKER IDENTIFICATION TASKS	2553
<i>Benjamin O'Brien, Christine Meunier, Alain Ghio</i>	
AN INTEGRATED FRAMEWORK FOR TWO-PASS PERSONALIZED VOICE TRIGGER.....	2558
<i>Dexin Liao, Jing Li, Yiming Zhi, Song Li, Qingyang Hong, Lin Li</i>	
CHRONOLOGICAL SELF-TRAINING FOR REAL-TIME SPEAKER DIARIZATION	2563
<i>Dirk Padfield, Daniel J. Liebling</i>	
AUTOMATIC ERROR CORRECTION FOR SPEAKER EMBEDDING LEARNING WITH NOISY LABELS.....	2568
<i>Fuchuan Tong, Yan Liu, Song Li, Jie Wang, Lin Li, Qingyang Hong</i>	
DR-VECTORS: DECISION RESIDUAL NETWORKS AND AN IMPROVED LOSS FOR SPEAKER RECOGNITION	2573
<i>Jason Pelecanos, Quan Wang, Ignacio Lopez Moreno</i>	
MASKED PROXY LOSS FOR TEXT-INDEPENDENT SPEAKER VERIFICATION.....	2578
<i>Jiachan Lian, Aiswarya Vinod Kumar, Hira Dharmyal, Bhiksha Raj, Rita Singh</i>	
GRAPH-BASED LABEL PROPAGATION FOR SEMI-SUPERVISED SPEAKER IDENTIFICATION	2583
<i>Long Chen, Venkatesh Ravichandran, Andreas Stolcke</i>	
FUSION OF EMBEDDINGS NETWORKS FOR ROBUST COMBINATION OF TEXT DEPENDENT AND INDEPENDENT SPEAKER RECOGNITION.....	2588
<i>Ruirui Li, Chelsea J.-T. Ju, Zeya Chen, Hongda Mao, Oguz Elibol, Andreas Stolcke</i>	
ADAPTIVE MARGIN CIRCLE LOSS FOR SPEAKER VERIFICATION.....	2593
<i>Runqiu Xiao, Xiaoxiao Miao, Wenchao Wang, Pengyuan Zhang, Bin Cai, Liuping Luo</i>	
A GENERATIVE MODEL FOR DURATION-DEPENDENT SCORE CALIBRATION.....	2598
<i>Sandro Cumani, Salvatore Sarni</i>	
MULTI-CHANNEL SPEAKER VERIFICATION FOR SINGLE AND MULTI-TALKER SPEECH	2603
<i>Saurabh Kataria, Shi-Xiong Zhang, Dong Yu</i>	

INTERSPEECH 2021 : SPEAKER, LANGUAGE, AND PRIVACY

USING GAMES TO AUGMENT CORPORA FOR LANGUAGE RECOGNITION AND CONFUSABILITY	2608
<i>Christopher Cieri, James Fiumara, Jonathan Wright</i>	
FAIR VOICE BIOMETRICS: IMPACT OF DEMOGRAPHIC IMBALANCE ON GROUP FAIRNESS IN SPEAKER RECOGNITION	2613
<i>Gianni Fenu, Mirko Marras, Giacomo Medda, Giacomo Meloni</i>	
KNOWLEDGE DISTILLATION FROM MULTI-MODALITY TO SINGLE-MODALITY FOR PERSON VERIFICATION.....	2618
<i>Leying Zhang, Zhengyang Chen, Yanmin Qian</i>	

ADVERSARIAL DISENTANGLEMENT OF SPEAKER REPRESENTATION FOR ATTRIBUTE-DRIVEN PRIVACY PRESERVATION	2623
<i>Paul-Gauthier Noe, Mohammad Mohammadamini, Driss Matrouf, Titouan Parcollet, Andreas Nautsch, Jean-Francois Bonastre</i>	

INTERSPEECH 2021 : SPEECH AND AUDIO ANALYSIS

AUDIO RETRIEVAL WITH NATURAL LANGUAGE QUERIES	2628
<i>Andreea-Maria Oncescu, A. Sophia Koepke, Joao F. Henriques, Zeynep Akata, Samuel Albanie</i>	
ORCA-SLANG: AN AUTOMATIC MULTI-STAGE SEMI-SUPERVISED DEEP LEARNING FRAMEWORK FOR LARGE-SCALE KILLER WHALE CALL TYPE IDENTIFICATION.....	2633
<i>Christian Bergler, Manuel Schmitt, Andreas Maier, Helena Symonds, Paul Spong, Steven R. Ness, George Tzanetakis, Elmar Noth</i>	
NON-INTRUSIVE SPEECH QUALITY ASSESSMENT WITH TRANSFER LEARNING AND SUBJECT-SPECIFIC SCALING.....	2638
<i>Natalia Nessler, Milos Cernak, Paolo Prandoni, Pablo Mainar</i>	
EXTENDING THE FULLBAND E-MODEL TOWARDS BACKGROUND NOISE, BURSTY PACKET LOSS, AND CONVERSATIONAL DEGRADATIONS.....	2643
<i>Thilo Michael, Gabriel Mittag, Andreas Butow, Sebastian Moller</i>	
AUDIOVISUAL TRANSFER LEARNING FOR AUDIO TAGGING AND SOUND EVENT DETECTION.....	2648
<i>Wim Boes, Hugo Van Hamme</i>	

INTERSPEECH 2021 : SPEECH CODING AND PRIVACY

IMPROVING THE EXPRESSIVENESS OF NEURAL VOCODING WITH NON-AFFINE NORMALIZING FLOWS	2653
<i>Adam Gabrys, Yunlong Jiao, Viacheslav Klimkov, Daniel Korzekwa, Roberto Barra-Chicote</i>	
DATA QUALITY AS PREDICTOR OF VOICE ANTI-SPOOFING GENERALIZATION	2658
<i>Bhusan Chettri, Rosa Gonzalez Hautamaki, Md. Sahidullah, Tomi Kinnunen</i>	
PROTECTING GENDER AND IDENTITY WITH DISENTANGLED SPEECH REPRESENTATIONS	2663
<i>Dimitrios Stoidis, Andrea Cavallaro</i>	
QISTA-NET-AUDIO: AUDIO SUPER-RESOLUTION VIA NON-CONVEX L _Q -NORM MINIMIZATION	2668
<i>Gang-Xuan Lin, Shih-Wei Hu, Yen-Ju Lu, Yu Tsao, Chun-Shien Lu</i>	
VOICE PRIVACY THROUGH X-VECTOR AND CYCLEGAN-BASED ANONYMIZATION	2673
<i>Gauri P. Prajapati, Dipesh K. Singh, Preet P. Amin, Hemant A. Patil</i>	
EFFECTS OF PROSODIC VARIATIONS ON ACCIDENTAL TRIGGERS OF A COMMERCIAL VOICE ASSISTANT	2678
<i>Ingo Siegert</i>	
HALF-TRUTH: A PARTIALLY FAKE AUDIO DETECTION DATASET	2683
<i>Jiangyan Yi, Ye Bai, Jianhua Tao, Haoxin Ma, Zhengkun Tian, Chenglong Wang, Tao Wang, Ruibo Fu</i>	

DEVELOPMENT OF A PSYCHOACOUSTIC LOSS FUNCTION FOR THE DEEP NEURAL NETWORK (DNN)-BASED SPEECH CODER	2688
<i>Joon Byun, Seungmin Shin, Youngcheol Park, Jongmo Sung, Seungkwon Beack</i>	
A TWO-STAGE APPROACH TO SPEECH BANDWIDTH EXTENSION	2693
<i>Ju Lin, Yun Wang, Kaustubh Kalgaonkar, Gil Keren, Didi Zhang, Christian Fuegen</i>	
NU-WAVE: A DIFFUSION PROBABILISTIC MODEL FOR NEURAL AUDIO UPSAMPLING.....	2698
<i>Junhyeok Lee, Seungu Han</i>	
WSRGLOW: A GLOW-BASED WAVEFORM GENERATIVE MODEL FOR AUDIO SUPER-RESOLUTION	2703
<i>Kexun Zhang, Yi Ren, Changliang Xu, Zhou Zhao</i>	
X-NET: A JOINT SCALE DOWN AND SCALE UP METHOD FOR VOICE CALL.....	2708
<i>Liang Wen, Lizhong Wang, Xue Wen, Yuxing Zheng, Youngo Park, Kwang Pyo Choi</i>	
MULTI-CHANNEL OPUS COMPRESSION FOR FAR-FIELD AUTOMATIC SPEECH RECOGNITION WITH A FIXED BITRATE BUDGET	2713
<i>Lukas Drude, Jahn Heymann, Andreas Schwarz, Jean-Marc Valin</i>	
CODED SPEECH ENHANCEMENT USING NEURAL NETWORK-BASED VECTOR-QUANTIZED RESIDUAL FEATURES	2718
<i>Youngju Cheon, Soojoong Hwang, Sangwook Han, Inseon Jang, Jong Won Shin</i>	

INTERSPEECH 2021 : SPEECH ENHANCEMENT AND CODING

N-MTTL SI MODEL: NON-INTRUSIVE MULTI-TASK TRANSFER LEARNING-BASED SPEECH INTELLIGIBILITY PREDICTION MODEL WITH SCENERY CLASSIFICATION	2723
<i>Lubos Marcinek, Michael Stone, Rebecca Millman, Patrick Gaydecki</i>	
END-TO-END OPTIMIZED MULTI-STAGE VECTOR QUANTIZATION OF SPECTRAL ENVELOPES FOR SPEECH AND AUDIO CODING.....	2728
<i>Mohammad Hassan Vali, Tom Backstrom</i>	
FUSION-NET: TIME-FREQUENCY INFORMATION FUSION Y-NETWORK FOR SPEECH ENHANCEMENT.....	2733
<i>Santhan Kumar Reddy Nareddula, Subrahmanyam Gorthi, Rama Krishna Sai S. Gorthi</i>	

INTERSPEECH 2021 : SPEECH ENHANCEMENT AND INTELLIGIBILITY

A SPECTRO-TEMPORAL GLIMPSING INDEX (STGI) FOR SPEECH INTELLIGIBILITY PREDICTION	2738
<i>Amin Edraki, Wai-Yip Chan, Jesper Jensen, Daniel Fogerty</i>	
COMPARISON OF REMOTE EXPERIMENTS USING CROWDSOURCING AND LABORATORY EXPERIMENTS ON SPEECH INTELLIGIBILITY	2743
<i>Ayako Yamamoto, Toshio Irino, Kenichi Arai, Shoko Araki, Atsunori Ogawa, Keisuke Kinoshita, Tomohiro Nakatani</i>	
PERCEPTUAL CONTRIBUTIONS OF VOWELS AND CONSONANT-VOWEL TRANSITIONS IN UNDERSTANDING TIME-COMPRESSED MANDARIN SENTENCES	2748
<i>Changjie Pan, Feng Yang, Fei Chen</i>	

RESTORING DEGRADED SPEECH VIA A MODIFIED DIFFUSION MODEL.....	2753
<i>Jianwei Zhang, Suren Jayasuriya, Visar Berisha</i>	
INCORPORATING EMBEDDING VECTORS FROM A HUMAN MEAN-OPINION SCORE PREDICTION MODEL FOR MONAURAL SPEECH ENHANCEMENT.....	2758
<i>Khandokar Md. Nayem, Donald S. Williamson</i>	
TEMPORAL CONVOLUTIONAL NETWORK WITH FREQUENCY DIMENSION ADAPTIVE ATTENTION FOR SPEECH ENHANCEMENT.....	2763
<i>Qiquan Zhang, Qi Song, Aaron Nicolson, Tian Lan, Haizhou Li</i>	
SPEECH ENHANCEMENT WITH WEAKLY LABELLED DATA FROM AUDIOSET.....	2768
<i>Qiuqiang Kong, Haohe Liu, Xingjian Du, Li Chen, Rui Xia, Yuxuan Wang</i>	
TRANSFER LEARNING FOR SPEECH INTELLIGIBILITY IMPROVEMENT IN NOISY ENVIRONMENTS	2773
<i>Ritujoy Biswas, Karan Nathwani, Vinayak Abrol</i>	
METRICGAN+: AN IMPROVED VERSION OF METRICGAN FOR SPEECH ENHANCEMENT.....	2778
<i>Szu-Wei Fu, Cheng Yu, Tsun-An Hsieh, Peter Plantinga, Mirco Ravanelli, Xugang Lu, Yu Tsao</i>	
IMPROVING PERCEPTUAL QUALITY BY PHONE-FORTIFIED PERCEPTUAL LOSS USING WASSERSTEIN DISTANCE FOR SPEECH ENHANCEMENT	2783
<i>Tsun-An Hsieh, Cheng Yu, Szu-Wei Fu, Xugang Lu, Yu Tsao</i>	
KNOW YOUR ENEMY, KNOW YOURSELF: A UNIFIED TWO-STAGE FRAMEWORK FOR SPEECH ENHANCEMENT	2788
<i>Wenzhe Liu, Andong Li, Yuxuan Ke, Chengshi Zheng, Xiaodong Li</i>	
SELF-SUPERVISED LEARNING BASED PHONE-FORTIFIED SPEECH ENHANCEMENT	2793
<i>Yuanhang Qiu, Ruili Wang, Satwinder Singh, Zhizhong Ma, Feng Hou</i>	
FUNNEL DEEP COMPLEX U-NET FOR PHASE-AWARE SPEECH ENHANCEMENT	2798
<i>Yuhang Sun, Linju Yang, Huifeng Zhu, Jie Hao</i>	
<u>INTERSPEECH 2021 : SPEECH LOCALIZATION, ENHANCEMENT, AND QUALITY ASSESSMENT</u>	
CNN-BASED PROCESSING OF ACOUSTIC AND RADIO FREQUENCY SIGNALS FOR SPEAKER LOCALIZATION FROM MAVS	2803
<i>Andrea Toma, Daniele Salvati, Carlo Drioli, Gian Luca Foresti</i>	
SUBJECTIVE EVALUATION OF NOISE SUPPRESSION ALGORITHMS IN CROWDSOURCING.....	2808
<i>Babak Naderi, Ross Cutler</i>	
PILOT: INTRODUCING TRANSFORMERS FOR PROBABILISTIC SOUND EVENT LOCALIZATION	2813
<i>Christopher Schymura, Benedikt Bonninghoff, Tsubasa Ochiai, Marc Delcroix, Keisuke Kinoshita, Tomohiro Nakatani, Shoko Araki, Dorothea Kolossa</i>	
NISQA: A DEEP CNN-SELF-ATTENTION MODEL FOR MULTIDIMENSIONAL SPEECH QUALITY PREDICTION WITH CROWDSOURCED DATASETS.....	2818
<i>Gabriel Mittag, Babak Naderi, Assmaa Chehadi, Sebastian Moller</i>	

RELIABLE INTENSITY VECTOR SELECTION FOR MULTI-SOURCE DIRECTION-OF-ARRIVAL ESTIMATION USING A SINGLE ACOUSTIC VECTOR SENSOR	2823
<i>Jianhua Geng, Sifan Wang, Juan Li, Jingwei Li, Xin Lou</i>	
ASSESSMENT OF VON MISES-BERNOULLI DEEP NEURAL NETWORK IN SOUND SOURCE LOCALIZATION	2828
<i>Katsutoshi Itoyama, Yoshiya Morimoto, Shungo Masaki, Ryosuke Kojima, Kenji Nishida, Kazuhiro Nakadai</i>	
SOUND SOURCE LOCALIZATION WITH MAJORIZATION MINIMIZATION	2833
<i>Masahito Togami, Robin Scheibler</i>	
METRICNET: TOWARDS IMPROVED MODELING FOR NON-INTRUSIVE SPEECH QUALITY ASSESSMENT	2838
<i>Meng Yu, Chunlei Zhang, Yong Xu, Shi-Xiong Zhang, Dong Yu</i>	
DIFFERENCE IN PERCEIVED SPEECH SIGNAL QUALITY ASSESSMENT AMONG MONOLINGUAL AND BILINGUAL TEENAGE STUDENTS	2843
<i>Przemyslaw Falkowski-Gilski</i>	
FEATURE FUSION BY ATTENTION NETWORKS FOR ROBUST DOA ESTIMATION	2848
<i>Rongliang Liu, Nengheng Zheng, Xi Chen</i>	
FAR-FIELD SPEAKER LOCALIZATION AND ADAPTIVE GLMB TRACKING	2853
<i>Shoufeng Lin, Zhaojie Luo</i>	
ON THE DESIGN OF DEEP PRIORS FOR UNSUPERVISED AUDIO RESTORATION	2858
<i>Vivek Sivaraman Narayanaswamy, Jayaraman J. Thiagarajan, Andreas Spanias</i>	
CRAMER-RAO LOWER BOUND FOR DOA ESTIMATION WITH AN ARRAY OF DIRECTIONAL MICROPHONES IN REVERBERANT ENVIRONMENTS	2863
<i>Weiguang Chen, Cheng Xue, Xionghu Zhong</i>	

INTERSPEECH 2021 : SPEECH PERCEPTION I

VARIATION IN PERCEPTUAL SENSITIVITY AND COMPENSATION FOR COARTICULATION ACROSS ADULT AND CHILD NATURALLY-PRODUCED AND TTS VOICES	2868
<i>Aleese Block, Michelle Cohn, Georgia Zellou</i>	
TIME-TO-EVENT MODELS FOR ANALYZING REACTION TIME SEQUENCES	2873
<i>Louis Ten Bosch, Lou Boves</i>	
WORD COMPETITION: AN ENTROPY-BASED APPROACH IN THE DIANA MODEL OF HUMAN WORD COMPREHENSION	2878
<i>Louis Ten Bosch, Lou Boves</i>	
EXTRACTING DIFFERENT LEVELS OF SPEECH INFORMATION FROM EEG USING AN LSTM-BASED MODEL	2883
<i>Mohammad Jalilpour Monesi, Bernd Accou, Tom Francart, Hugo Van Hamme</i>	
MODELS OF REACTION TIMES IN AUDITORY LEXICAL DECISION: RTONSET VERSUS RTOFFSET	2888
<i>Sophie Brand, Kimberley Mulder, Louis Ten Bosch, Lou Boves</i>	

PROSODIC DISAMBIGUATION USING CHIRONOMIC STYLIZATION OF INTONATION WITH NATIVE AND NON-NATIVE SPEAKERS.....	2893
<i>Xiao Xiao, Nicolas Audibert, Gregoire Locqueville, Christophe D'Alessandro, Barbara Kuhnert, Claire Pillot-Loiseau</i>	

INTERSPEECH 2021 : SPEECH PERCEPTION II

EXPLORING THE POTENTIAL OF LEXICAL PARAPHRASES FOR MITIGATING NOISE-INDUCED COMPREHENSION ERRORS.....	2898
<i>Anupama Chingacham, Vera Demberg, Dietrich Klakow</i>	
PRIMACY OF MOUTH OVER EYES: EYE MOVEMENT EVIDENCE FROM AUDIOVISUAL MANDARIN LEXICAL TONES AND VOWELS	2903
<i>Biao Zeng, Rui Wang, Guoxing Yu, Christian Dobel</i>	
HUMAN SPOOFING DETECTION PERFORMANCE ON DEGRADED SPEECH.....	2907
<i>Camryn Terblanche, Philip Harrison, Amelia J. Gully</i>	
RELIABLE ESTIMATES OF INTERPRETABLE CUE EFFECTS WITH ACTIVE LEARNING IN PSYCHOLINGUISTIC RESEARCH	2912
<i>Marieke Einfeldt, Rita Sevastjanova, Katharina Zahner-Ritter, Ekaterina Kazak, Bettina Braun</i>	
EFFECTS OF AGING AND AGE-RELATED HEARING LOSS ON TALKER DISCRIMINATION.....	2917
<i>Min Xu, Jing Shao, Lan Wang</i>	
SPEECHADJUSTER: A TOOL FOR INVESTIGATING LISTENER PREFERENCES AND SPEECH INTELLIGIBILITY.....	2922
<i>Olympia Simantiraki, Martin Cooke</i>	
TOWARDS THE EXPLAINABILITY OF MULTIMODAL SPEECH EMOTION RECOGNITION.....	2927
<i>Puneet Kumar, Vishesh Kaushik, Balasubramanian Raman</i>	
VOCALTURK: EXPLORING FEASIBILITY OF CROWDSOURCED SPEAKER IDENTIFICATION	2932
<i>Susumu Saito, Yuta Ide, Teppei Nakano, Tetsuji Ogawa</i>	
INVESTIGATING THE IMPACT OF SPECTRAL AND TEMPORAL DEGRADATION ON END-TO-END AUTOMATIC SPEECH RECOGNITION PERFORMANCE.....	2937
<i>Takanori Ashihara, Takafumi Moriya, Makio Kashino</i>	
THE INFLUENCE OF PARALLEL PROCESSING ON ILLUSORY VOWELS.....	2942
<i>Takeshi Kishiyama</i>	
PERCEPTION OF STANDARD ARABIC SYNTHETIC SPEECH RATE	2947
<i>Yahya Aldholmi, Rawan Aldhafyan, Asma Alqahtani</i>	
RELATIONSHIPS BETWEEN PERCEPTUAL DISTINCTIVENESS, ARTICULATORY COMPLEXITY AND FUNCTIONAL LOAD IN SPEECH COMMUNICATION.....	2951
<i>Yuqing Zhang, Zhu Li, Bin Wu, Yanlu Xie, Binghui Lin, Jinsong Zhang</i>	

INTERSPEECH 2021 : SPEECH PRODUCTION I

CHANGES IN GLOTTAL SOURCE PARAMETER VALUES WITH LIGHT TO MODERATE PHYSICAL LOAD	2956
<i>Heather Weston, Laura L. Koenig, Susanne Fuchs</i>	
LEARNING ROBUST SPEECH REPRESENTATION WITH AN ARTICULATORY-REGULARIZED VARIATIONAL AUTOENCODER	2961
<i>Marc-Antoine Georges, Laurent Girin, Jean-Luc Schwartz, Thomas Hueber</i>	
EFFECTS OF TIME PRESSURE AND SPONTANEITY ON PHONOTACTIC INNOVATIONS IN GERMAN DIALOGUES	2966
<i>Petra Wagner, Sina Zarriess, Joana Cholin</i>	
COMPARISON OF THE FINITE ELEMENT METHOD, THE MULTIMODAL METHOD AND THE TRANSMISSION-LINE MODEL FOR THE COMPUTATION OF VOCAL TRACT TRANSFER FUNCTIONS	2971
<i>Remi Blandin, Marc Arnela, Simon Felix, Jean-Baptiste Doc, Peter Birkholz</i>	
IMPORTANCE OF PARASAGITTAL SENSOR INFORMATION IN TONGUE MOTION CAPTURE THROUGH A DIPHONIC ANALYSIS	2976
<i>Salvador Medina, Sarah Taylor, Mark Tiede, Alexander Hauptmann, Iain Matthews</i>	
TOWARDS THE PREDICTION OF THE VOCAL TRACT SHAPE FROM THE SEQUENCE OF PHONEMES TO BE ARTICULATED	2981
<i>Vinicius Ribeiro, Karyna Isaieva, Justine Leclere, Pierre-Andre Vuissoz, Yves Laprie</i>	

INTERSPEECH 2021 : SPEECH PRODUCTION II

MODEL-BASED EXPLORATION OF LINKING BETWEEN VOWEL ARTICULATORY SPACE AND ACOUSTIC SPACE	2986
<i>Anqi Xu, Daniel Van Niekerk, Branislav Gerazov, Paul Konstantin Krug, Santitham Prom-On, Peter Birkholz, Yi Xu</i>	
MIXTURE OF ORTHOGONAL SEQUENCES MADE FROM EXTENDED TIME-STRETCHED PULSES ENABLES MEASUREMENT OF INVOLUNTARY VOICE FUNDAMENTAL FREQUENCY RESPONSE TO PITCH PERTURBATION	2991
<i>Hideki Kawahara, Toshie Matsui, Kohei Yatabe, Ken-Ichi Sakakibara, Minoru Tsuzaki, Masanori Morise, Toshio Irino</i>	
TAKE A BREATH: RESPIRATORY SOUNDS IMPROVE RECOLLECTION IN SYNTHETIC SPEECH	2996
<i>Mikey Elmers, Raphael Werner, Beeke Muhlack, Bernd Mobius, Jurgen Trouvain</i>	
COMPARISON BETWEEN LUMPED-MASS MODELING AND FLOW SIMULATION OF THE REED-TYPE ARTIFICIAL VOCAL FOLD	3001
<i>Rafia Inaam, Tsukasa Yoshinaga, Takayuki Arai, Hiroshi Yokoyama, Akiyoshi Iida</i>	
INHALATIONS IN SPEECH: ACOUSTIC AND PHYSIOLOGICAL CHARACTERISTICS	3006
<i>Raphael Werner, Susanne Fuchs, Jurgen Trouvain, Bernd Mobius</i>	
USING TRANSPOSED CONVOLUTION FOR ARTICULATORY-TO-ACOUSTIC CONVERSION FROM REAL-TIME MRI DATA	3011
<i>Ryo Tanji, Hidefumi Ohmura, Kouichi Katsurada</i>	

MODELING SENSORIMOTOR ADAPTATION IN SPEECH THROUGH ALTERATIONS TO FORWARD AND INVERSE MODELS 3016
Taijing Chen, Adam Lammert, Benjamin Parrell

VOCAL-TRACT MODELS TO VISUALIZE THE AIRSTREAM OF HUMAN BREATH AND DROPLETS WHILE PRODUCING SPEECH 3021
Takayuki Arai

A SIMPLIFIED MODEL FOR THE VOCAL TRACT OF [S] WITH INCLINED INCISORS 3026
Tsukasa Yoshinaga, Kohei Tada, Kazunori Nozaki, Akiyoshi Iida

INTERSPEECH 2021 : SPEECH RECOGNITION OF ATYPICAL SPEECH

LEARNING EXPLICIT PROSODY MODELS AND DEEP SPEAKER EMBEDDINGS FOR ATYPICAL VOICE CONVERSION 3031
Disong Wang, Songxiang Liu, Lifa Sun, Xixin Wu, Xunying Liu, Helen Meng

HANDLING ACOUSTIC VARIATION IN DYSARTHIC SPEECH RECOGNITION SYSTEMS THROUGH MODEL COMBINATION 3036
Enno Hermann, Mathew Magimai-Doss

AUTOMATIC SEVERITY CLASSIFICATION OF KOREAN DYSARTHIC SPEECH USING PHONEME-LEVEL PRONUNCIATION FEATURES 3041
Eun Jung Yeo, Sunhee Kim, Minhwa Chung

BAYESIAN PARAMETRIC AND ARCHITECTURAL DOMAIN ADAPTATION OF LF-MMI TRAINED TDNNS FOR ELDERLY AND DYSARTHIC SPEECH RECOGNITION 3046
Jiajun Deng, Fabian Ritter Gutierrez, Shoukang Hu, Mengzhe Geng, Xurong Xie, Zi Ye, Shansong Liu, Jianwei Yu, Xunying Liu, Helen Meng

AUTOMATIC SPEECH RECOGNITION OF DISORDERED SPEECH: PERSONALIZED MODELS OUTPERFORMING HUMAN LISTENERS ON SHORT PHRASES 3051
Jordan R. Green, Robert L. Macdonald, Pan-Pan Jiang, Julie Cattiau, Rus Heywood, Richard Cave, Katie Seaver, Marilyn A. Ladewig, Jimmy Tobin, Michael P. Brenner, Philip C. Nelson, Katrin Tomanek

SPECTRO-TEMPORAL DEEP FEATURES FOR DISORDERED SPEECH ASSESSMENT AND RECOGNITION 3056
Mengzhe Geng, Shansong Liu, Jianwei Yu, Xurong Xie, Shoukang Hu, Zi Ye, Zengrui Jin, Xunying Liu, Helen Meng

INVESTIGATING THE UTILITY OF MULTIMODAL CONVERSATIONAL TECHNOLOGY AND AUDIOVISUAL ANALYTIC MEASURES FOR THE ASSESSMENT AND MONITORING OF AMYOTROPHIC LATERAL SCLEROSIS AT SCALE 3061
Michael Neumann, Oliver Roesler, Jackson Liscombe, Hardik Kothare, David Suendermann-Oeft, David Pautler, Indu Navar, Aria Anvar, Jochen Kumm, Raquel Norel, Ernest Fraenkel, Alexander V. Sherman, James D. Berry, Gary L. Pattee, Jun Wang, Jordan R. Green, Vikram Ramanarayanan

DISORDERED SPEECH DATA COLLECTION: LESSONS LEARNED AT 1 MILLION UTTERANCES FROM PROJECT EUPHONIA 3066
Robert L. Macdonald, Pan-Pan Jiang, Julie Cattiau, Rus Heywood, Richard Cave, Katie Seaver, Marilyn A. Ladewig, Jimmy Tobin, Michael P. Brenner, Philip C. Nelson, Jordan R. Green, Katrin Tomanek

SPEAKING WITH A KN95 FACE MASK: ASR PERFORMANCE AND SPEAKER COMPENSATION.....	3071
<i>Sarah E. Gutz, Hannah P. Rowe, Jordan R. Green</i>	
A VOICE-ACTIVATED SWITCH FOR PERSONS WITH MOTOR AND SPEECH IMPAIRMENTS: ISOLATED-VOWEL SPOTTING USING NEURAL NETWORKS.....	3076
<i>Shanqing Cai, Lisie Lillianfeld, Katie Seaver, Jordan R. Green, Michael P. Brenner, Philip C. Nelson, D. Sculley</i>	
COMPARING SUPERVISED MODELS AND LEARNED SPEECH REPRESENTATIONS FOR CLASSIFYING INTELLIGIBILITY OF DISORDERED SPEECH ON SELECTED PHRASES.....	3081
<i>Subhashini Venugopalan, Joel Shor, Manoj Plakal, Jimmy Tobin, Katrin Tomanek, Jordan R. Green, Michael P. Brenner</i>	
ANALYSIS AND TUNING OF A VOICE ASSISTANT SYSTEM FOR DYSFLUENT SPEECH.....	3086
<i>Vikramjit Mitra, Zifang Huang, Colin Lea, Lauren Tooley, Sarah Wu, Darren Botten, Ashwini Palekar, Shrinath Thelapurath, Panayiotis Georgiou, Sachin Kajarekar, Jefferey Bigham</i>	
VARIATIONAL AUTO-ENCODER BASED VARIABILITY ENCODING FOR DYSARTHIC SPEECH RECOGNITION.....	3091
<i>Xurong Xie, Rukiye Ruzi, Xunying Liu, Lan Wang</i>	
ADVERSARIAL DATA AUGMENTATION FOR DISORDERED SPEECH RECOGNITION.....	3096
<i>Zengrui Jin, Mengzhe Geng, Xurong Xie, Jianwei Yu, Shansong Liu, Xunying Liu, Helen Meng</i>	
CONFORMER PARROTRON: A FASTER AND STRONGER END-TO-END SPEECH CONVERSION AND RECOGNITION MODEL FOR ATYPICAL SPEECH.....	3101
<i>Zhehuai Chen, Bhuvana Ramabhadran, Fadi Biadisy, Xia Zhang, Youzheng Chen, Liyang Jiang, Fang Chu, Rohan Doshi, Pedro J. Moreno</i>	
<u>INTERSPEECH 2021 : SPEECH SIGNAL ANALYSIS AND REPRESENTATION I</u>	
RAW SPEECH-TO-ARTICULATORY INVERSION BY TEMPORAL FILTERING AND DECIMATION.....	3106
<i>Abdolreza Sabzi Shahrehabaki, Sabato Marco Siniscalchi, Torbjorn Svendsen</i>	
SPEECH DECOMPOSITION BASED ON A HYBRID SPEECH MODEL AND OPTIMAL SEGMENTATION.....	3111
<i>Alfredo Esquivel Jaramillo, Jesper Kjaer Nielsen, Mads Graesboll Christensen</i>	
DO SOUND EVENT REPRESENTATIONS GENERALIZE TO OTHER AUDIO TASKS? A CASE STUDY IN AUDIO TRANSFER LEARNING.....	3116
<i>Anurag Kumar, Yun Wang, Vamsi Krishna Ithapu, Christian Fuegen</i>	
NOISE ROBUST PITCH STYLIZATION USING MINIMUM MEAN ABSOLUTE ERROR CRITERION.....	3121
<i>Chiranjeevi Yarra, Prasanta Kumar Ghosh</i>	
SYNCHRONISING SPEECH SEGMENTS WITH MUSICAL BEATS IN MANDARIN AND ENGLISH SINGING.....	3126
<i>Cong Zhang, Jian Zhu</i>	
UNSUPERVISED MULTI-TARGET DOMAIN ADAPTATION FOR ACOUSTIC SCENE CLASSIFICATION.....	3131
<i>Dongchao Yang, Helin Wang, Yuexian Zou</i>	

PITCH CONTOUR SEPARATION FROM OVERLAPPING SPEECH.....	3136
<i>Hiroki Mori</i>	
FRILL: A NON-SEMANTIC SPEECH EMBEDDING FOR MOBILE DEVICES	3141
<i>Jacob Peplinski, Joel Shor, Sachin Joglekar, Jake Garrison, Shwetak Patel</i>	
UNSUPERVISED TRAINING OF A DNN-BASED FORMANT TRACKER	3146
<i>Jason Lilley, H. Timothy Bunnell</i>	
DROPOUT REGULARIZATION FOR SELF-SUPERVISED LEARNING OF TRANSFORMER ENCODER SPEECH REPRESENTATION.....	3151
<i>Jian Luo, Jianzong Wang, Ning Cheng, Jing Xiao</i>	
ESTIMATING ARTICULATORY MOVEMENTS IN SPEECH PRODUCTION WITH TRANSFORMER NETWORKS	3156
<i>Sathvik Udupa, Anwasha Roy, Abhayjeet Singh, Aravind Illa, Prasanta Kumar Ghosh</i>	
SUPERB: SPEECH PROCESSING UNIVERSAL PERFORMANCE BENCHMARK.....	3161
<i>Shu-Wen Yang, Po-Han Chi, Yung-Sung Chuang, Cheng-I Jeff Lai, Kushal Lakhota, Yist Y. Lin, Andy T. Liu, Jiatong Shi, Xuankai Chang, Guan-Ting Lin, Tzu-Hsien Huang, Wei-Cheng Tseng, Ko-Tik Lee, Da-Rong Liu, Zili Huang, Shuyan Dong, Shang-Wen Li, Shinji Watanabe, Abdelrahman Mohamed, Hung-Yi Lee</i>	
AN ATTRIBUTE-ALIGNED STRATEGY FOR LEARNING SPEECH REPRESENTATION	3166
<i>Yu-Lin Huang, Bo-Hao Su, Y.-W. Peter Hong, Chi-Chun Lee</i>	

INTERSPEECH 2021 : SPEECH SIGNAL ANALYSIS AND REPRESENTATION II

FRICATIVE PHONEME DETECTION USING DEEP NEURAL NETWORKS AND ITS COMPARISON TO TRADITIONAL METHODS	3171
<i>Metehan Yurt, Pavan Kantharaju, Sascha Disch, Andreas Niedermeier, Alberto N. Escalante-B., Veniamin I. Morgenshtern</i>	
IDENTIFICATION OF F1 AND F2 IN SPEECH USING MODIFIED ZERO FREQUENCY FILTERING	3176
<i>Ravishankar Prasad, Mathew Magimai-Doss</i>	
A BENCHMARK OF DYNAMICAL VARIATIONAL AUTOENCODERS APPLIED TO SPEECH SPECTROGRAM MODELING.....	3181
<i>Xiaoyu Bie, Laurent Girin, Simon Leglaive, Thomas Hueber, Xavier Alameda-Pineda</i>	
PHONEME-TO-AUDIO ALIGNMENT WITH RECURRENT NEURAL NETWORKS FOR SPEAKING AND SINGING VOICE	3186
<i>Yann Teytaut, Axel Roebel</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: LINGUISTIC PROCESSING, PARADIGMS AND OTHER TOPICS

UNSUPERVISED LEARNING OF DISENTANGLED SPEECH CONTENT AND STYLE REPRESENTATION	3191
<i>Andros Tjandra, Ruoming Pang, Yu Zhang, Shigeki Karita</i>	
LABEL EMBEDDING FOR CHINESE GRAPHEME-TO-PHONEME CONVERSION	3196
<i>Eunbi Choi, Hwa-Yeon Kim, Jong-Hwan Kim, Jae-Min Kim</i>	

PDF: POLYPHONE DISAMBIGUATION IN CHINESE BY USING FLAT.....	3201
<i>Haiteng Zhang</i>	
LINEARSPEECH: PARALLEL TEXT-TO-SPEECH WITH LINEAR COMPLEXITY	3206
<i>Haozhe Zhang, Zhihua Huang, Zengqiang Shang, Pengyuan Zhang, Yonghong Yan</i>	
IMPROVING POLYPHONE DISAMBIGUATION FOR MANDARIN CHINESE BY COMBINING MIX-POOLING STRATEGY AND WINDOW-BASED ATTENTION.....	3211
<i>Junjie Li, Zhiyu Zhang, Minchuan Chen, Jun Ma, Shaojun Wang, Jing Xiao</i>	

VOLUME 5

DYNAMICALLY ADAPTIVE MACHINE SPEECH CHAIN INFERENCE FOR TTS IN NOISY ENVIRONMENT: LISTEN AND SPEAK LOUDER.....	3216
<i>Sashi Novitasari, Sakriani Sakti, Satoshi Nakamura</i>	
UNITNET-BASED HYBRID SPEECH SYNTHESIS	3221
<i>Xiao Zhou, Zhen-Hua Ling, Li-Rong Dai</i>	
POLYPHONE DISAMBIGUATION IN MANDARIN CHINESE WITH SEMI-SUPERVISED LEARNING.....	3226
<i>Yi Shi, Congyi Wang, Yu Chen, Bin Wang</i>	
A NEURAL-NETWORK-BASED APPROACH TO IDENTIFYING SPEAKERS IN NOVELS	3231
<i>Yue Chen, Zhen-Hua Ling, Qing-Feng Liu</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: NEURAL WAVEFORM GENERATION

GAN VOCODER: MULTI-RESOLUTION DISCRIMINATOR IS ALL YOU NEED	3236
<i>Jaeseong You, Dalhyun Kim, Gyuhyeon Nam, Geumbyeol Hwang, Gyeongsu Chae</i>	
GLOW-WAVEGAN: LEARNING SPEECH REPRESENTATIONS FROM GAN-BASED VARIATIONAL AUTO-ENCODER FOR HIGH FIDELITY FLOW-BASED SPEECH SYNTHESIS	3241
<i>Jian Cong, Shan Yang, Lei Xie, Dan Su</i>	
FRE-GAN: ADVERSARIAL FREQUENCY-CONSISTENT AUDIO SYNTHESIS	3246
<i>Ji-Hoon Kim, Sang-Hoon Lee, Ji-Hyun Lee, Seong-Whan Lee</i>	
GANSPEECH: ADVERSARIAL TRAINING FOR HIGH-FIDELITY MULTI-SPEAKER SPEECH SYNTHESIS.....	3251
<i>Jinhyeok Yang, Jae-Sung Bae, Taejun Bak, Young-Ik Kim, Hoon-Young Cho</i>	
HARMONIC WAVEGAN: GAN-BASED SPEECH WAVEFORM GENERATION MODEL WITH HARMONIC STRUCTURE DISCRIMINATOR	3256
<i>Kazuki Mizuta, Tomoki Koriyama, Hiroshi Saruwatari</i>	
HIGH-FIDELITY PARALLEL WAVEGAN WITH MULTI-BAND HARMONIC-PLUS-NOISE MODEL.....	3261
<i>Min-Jae Hwang, Ryuichi Yamamoto, Eunwoo Song, Jae-Min Kim</i>	
CONTINUOUS WAVELET VOCODER-BASED DECOMPOSITION OF PARAMETRIC SPEECH WAVEFORM SYNTHESIS.....	3266
<i>Mohammed Salah Al-Radhi, Tamas Gabor Csapo, Csaba Zainko, Geza Nemeth</i>	

HIGH-FIDELITY AND LOW-LATENCY UNIVERSAL NEURAL VOCODER BASED ON MULTIBAND WAVERNN WITH DATA-DRIVEN LINEAR PREDICTION FOR DISCRETE WAVEFORM MODELING	3271
<i>Patrick Lumban Tobing, Tomoki Toda</i>	
UNIFIED SOURCE-FILTER GAN: UNIFIED SOURCE-FILTER NETWORK BASED ON FACTORIZATION OF QUASI-PERIODIC PARALLEL WAVEGAN	3276
<i>Reo Yoneyama, Yi-Chiao Wu, Tomoki Toda</i>	
UNIVNET: A NEURAL VOCODER WITH MULTI-RESOLUTION SPECTROGRAM DISCRIMINATORS FOR HIGH-FIDELITY WAVEFORM GENERATION.....	3281
<i>Won Jang, Dan Lim, Jaesam Yoon, Bongwan Kim, Juntae Kim</i>	
BASIS-MELGAN: EFFICIENT NEURAL VOCODER BASED ON AUDIO DECOMPOSITION	3286
<i>Zhengxi Liu, Yanmin Qian</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: OTHER TOPICS

T5G2P: USING TEXT-TO-TEXT TRANSFER TRANSFORMER FOR GRAPHEME-TO-PHONEME CONVERSION	3291
<i>Marketa Rezackova, Jan Svec, Daniel Tihelka</i>	
CONVERSION OF AIRBORNE TO BONE-CONDUCTED SPEECH WITH DEEP NEURAL NETWORKS.....	3296
<i>Michael Pucher, Thomas Woltron</i>	
EVALUATING THE EXTRAPOLATION CAPABILITIES OF NEURAL VOCODERS TO EXTREME PITCH VALUES	3301
<i>Olivier Perrotin, Hussein El Amouri, Gerard Bailly, Thomas Hueber</i>	
A SYSTEMATIC REVIEW AND ANALYSIS OF MULTILINGUAL DATA STRATEGIES IN TEXT-TO-SPEECH FOR LOW-RESOURCE LANGUAGES.....	3306
<i>Phat Do, Matt Coler, Jelske Dijkstra, Esther Klabbers</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: PROSODY MODELING I

A PROTOTYPICAL NETWORK APPROACH FOR EVALUATING GENERATED EMOTIONAL SPEECH	3311
<i>Alice Baird, Silvan Mertes, Manuel Milling, Lukas Stappen, Thomas Wiest, Elisabeth Andre, Bjorn W. Schuller</i>	
RICH PROSODY DIVERSITY MODELLING WITH PHONE-LEVEL MIXTURE DENSITY NETWORK.....	3316
<i>Chenpeng Du, Kai Yu</i>	
APPLYING THE INFORMATION BOTTLENECK PRINCIPLE TO PROSODIC REPRESENTATION LEARNING.....	3321
<i>Guangyan Zhang, Ying Qin, Daxin Tan, Tan Lee</i>	
IMPROVING MULTI-SPEAKER TTS PROSODY VARIANCE WITH A RESIDUAL ENCODER AND NORMALIZING FLOWS	3326
<i>Ivan Valles-Perez, Julian Roth, Grzegorz Beringer, Roberto Barra-Chicote, Jasha Droppo</i>	

PHONEME DURATION MODELING USING SPEECH RHYTHM-BASED SPEAKER EMBEDDINGS FOR MULTI-SPEAKER SPEECH SYNTHESIS	3331
<i>Kenichi Fujita, Atsushi Ando, Yusuke Ijima</i>	
PHRASE BREAK PREDICTION WITH BIDIRECTIONAL ENCODER REPRESENTATIONS IN JAPANESE TEXT-TO-SPEECH SYNTHESIS	3336
<i>Kosuke Futamata, Byeongseon Park, Ryuichi Yamamoto, Kentaro Tachibana</i>	
INTRA-SENTENTIAL SPEAKING RATE CONTROL IN NEURAL TEXT-TO-SPEECH FOR AUTOMATIC DUBBING	3341
<i>Mayank Sharma, Yogesh Virkar, Marcello Federico, Roberto Barra-Chicote, Robert Enyedi</i>	
FINE-GRAINED PROSODY MODELING IN NEURAL SPEECH SYNTHESIS USING TOBI REPRESENTATION	3346
<i>Yuxiang Zou, Shichao Liu, Xiang Yin, Haopeng Lin, Chunfeng Wang, Haoyu Zhang, Zejun Ma</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: PROSODY MODELING II

ADEPT: A DATASET FOR EVALUATING PROSODY TRANSFER	3351
<i>Alexandra Torresquintero, Tian Huey Teh, Christopher G. R. Wallis, Marlene Staib, Devang S. Ram Mohan, Vivian Hu, Lorenzo Foglianti, Jiameng Gao, Simon King</i>	
ALTERNATE ENDINGS: IMPROVING PROSODY FOR INCREMENTAL NEURAL TTS WITH PREDICTED FUTURE TEXT INPUT	3356
<i>Brooke Stephenson, Thomas Hueber, Laurent Girin, Laurent Besacier</i>	
CTRL-P: TEMPORAL CONTROL OF PROSODIC VARIATION FOR SPEECH SYNTHESIS	3361
<i>Devang S. Ram Mohan, Vivian Hu, Tian Huey Teh, Alexandra Torresquintero, Christopher G. R. Wallis, Marlene Staib, Lorenzo Foglianti, Jiameng Gao, Simon King</i>	
PROSODIC BOUNDARY PREDICTION MODEL FOR VIETNAMESE TEXT-TO-SPEECH	3366
<i>Nguyen Thi Thu Trang, Nguyen Hoang Ky, Albert Rilliard, Christophe D'Alessandro</i>	
EXPLORING EMOTIONAL PROTOTYPES IN A HIGH DIMENSIONAL TTS LATENT SPACE	3371
<i>Pol Van Rijn, Silvan Mertes, Dominik Schiller, Peter M. C. Harrison, Pauline Larrouy-Maestri, Elisabeth Andre, Nori Jacoby</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: SINGING, MULTIMODAL, CROSSLINGUAL SYNTHESIS

CROSS-LINGUAL SPEAKER ADAPTATION USING DOMAIN ADAPTATION AND SPEAKER CONSISTENCY LOSS FOR TEXT-TO-SPEECH SYNTHESIS	3376
<i>Detai Xin, Yuki Saito, Shinnosuke Takamichi, Tomoki Koriyama, Hiroshi Saruwatari</i>	
INVESTIGATING CONTRIBUTIONS OF SPEECH AND FACIAL LANDMARKS FOR TALKING HEAD GENERATION	3381
<i>Ege Kesim, Engin Erzin</i>	
CROSS-LINGUAL LOW RESOURCE SPEAKER ADAPTATION USING PHONOLOGICAL FEATURES	3386
<i>Georgia Maniati, Nikolaos Ellinas, Konstantinos Markopoulos, Georgios Vamvoukakis, June Sig Sung, Hyoungmin Park, Aimilios Chalamandaris, Pirros Tsiakoulis</i>	

N-SINGER: A NON-AUTOREGRESSIVE KOREAN SINGING VOICE SYNTHESIS SYSTEM FOR PRONUNCIATION ENHANCEMENT	3391
<i>Gyeong-Hoon Lee, Tae-Woo Kim, Hanbin Bae, Min-Ji Lee, Young-Ik Kim, Hoon-Young Cho</i>	
IMPROVE CROSS-LINGUAL TEXT-TO-SPEECH SYNTHESIS ON MONOLINGUAL CORPORA WITH PITCH CONTOUR INFORMATION	3396
<i>Haoyue Zhan, Haitong Zhang, Wenjie Ou, Yue Lin</i>	
SPEECH2VIDEO: CROSS-MODAL DISTILLATION FOR SPEECH TO VIDEO GENERATION	3401
<i>Shijing Si, Jianzong Wang, Xiaoyang Qu, Ning Cheng, Wenqi Wei, Xinghua Zhu, Jing Xiao</i>	
INCORPORATING CROSS-SPEAKER STYLE TRANSFER FOR MULTI-LANGUAGE TEXT-TO-SPEECH	3406
<i>Zengqiang Shang, Zhihua Huang, Haozhe Zhang, Pengyuan Zhang, Yonghong Yan</i>	
CROSS-LINGUAL VOICE CONVERSION WITH DISENTANGLED UNIVERSAL LINGUISTIC REPRESENTATIONS	3411
<i>Zhenchuan Yang, Weibin Zhang, Yufei Liu, Xiaofen Xing</i>	
EFFICIENTSING: A CHINESE SINGING VOICE SYNTHESIS SYSTEM USING DURATION-FREE ACOUSTIC MODEL AND HIFI-GAN VOCODER	3416
<i>Zhengchen Liu, Chenfeng Miao, Qingying Zhu, Minchuan Chen, Jun Ma, Shaojun Wang, Jing Xiao</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: SPEAKING STYLE AND EMOTION

FINE-GRAINED STYLE MODELING, TRANSFER AND PREDICTION IN TEXT-TO-SPEECH SYNTHESIS VIA PHONE-LEVEL CONTENT-STYLE DISENTANGLEMENT	3421
<i>Daxin Tan, Tan Lee</i>	
CONTROLLABLE CONTEXT-AWARE CONVERSATIONAL SPEECH SYNTHESIS	3426
<i>Jian Cong, Shan Yang, Na Hu, Guangzhi Li, Lei Xie, Dan Su</i>	
STYLER: STYLE FACTOR MODELING WITH RAPIDITY AND ROBUSTNESS VIA SPEECH DECOMPOSITION FOR EXPRESSIVE AND CONTROLLABLE NEURAL TEXT TO SPEECH	3431
<i>Keon Lee, Kyumin Park, Daeyoung Kim</i>	
EXPRESSIVE TEXT-TO-SPEECH USING STYLE TAG	3436
<i>Minchan Kim, Sung Jun Cheon, Byoung Jin Choi, Jong Jin Kim, Nam Soo Kim</i>	
REINFORCEMENT LEARNING FOR EMOTIONAL TEXT-TO-SPEECH SYNTHESIS WITH IMPROVED EMOTION DISCRIMINABILITY	3441
<i>Rui Liu, Berrak Sisman, Haizhou Li</i>	
EMOTIONAL PROSODY CONTROL FOR SPEECH GENERATION	3446
<i>Sarath Sivaprasad, Saiteja Kosgi, Vineet Gandhi</i>	
CROSS-SPEAKER STYLE TRANSFER WITH PROSODY BOTTLENECK IN NEURAL SPEECH SYNTHESIS	3451
<i>Shifeng Pan, Lei He</i>	

SYNTHESIS OF EXPRESSIVE SPEAKING STYLES WITH LIMITED TRAINING DATA IN A MULTI-SPEAKER, PROSODY-CONTROLLABLE SEQUENCE-TO-SEQUENCE ARCHITECTURE.....	3456
<i>Slava Shechtman, Raul Fernandez, Alexander Sorin, David Haws</i>	
TOWARDS MULTI-SCALE STYLE CONTROL FOR EXPRESSIVE SPEECH SYNTHESIS	3461
<i>Xiang Li, Changhe Song, Jingbei Li, Zhiyong Wu, Jia Jia, Helen Meng</i>	
IMPROVING PERFORMANCE OF SEEN AND UNSEEN SPEECH STYLE TRANSFER IN END-TO-END NEURAL TTS	3466
<i>Xiaochun An, Frank K. Soong, Lei Xie</i>	
ADAPTIVE TEXT TO SPEECH FOR SPONTANEOUS STYLE	3471
<i>Yuzi Yan, Xu Tan, Bohan Li, Guangyan Zhang, Tao Qin, Sheng Zhao, Yuan Shen, Wei-Qiang Zhang, Tie-Yan Liu</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: TOOLS, DATA, EVALUATION

DETECTION AND ANALYSIS OF ATTENTION ERRORS IN SEQUENCE-TO-SEQUENCE TEXT-TO-SPEECH	3476
<i>Cassia Valentini-Botinhao, Simon King</i>	
EMOVIE: A MANDARIN EMOTION SPEECH DATASET WITH A SIMPLE EMOTIONAL TEXT-TO-SPEECH MODEL	3481
<i>Chenye Cui, Yi Ren, Jinglin Liu, Feiyang Chen, Rongjie Huang, Ming Lei, Zhou Zhao</i>	
HI-FI MULTI-SPEAKER ENGLISH TTS DATASET	3486
<i>Evelina Bakhturina, Vitaly Lavrukhin, Boris Ginsburg, Yang Zhang</i>	
CONFIDENCE INTERVALS FOR ASR-BASED TTS EVALUATION	3491
<i>Jason Taylor, Korin Richmond</i>	
COMPARING SPEECH ENHANCEMENT TECHNIQUES FOR VOICE ADAPTATION-BASED SPEECH SYNTHESIS	3496
<i>Nicholas Eng, C. T. Justine Hui, Yusuke Hioka, Catherine I. Watson</i>	
RYANSPEECH: A CORPUS FOR CONVERSATIONAL TEXT-TO-SPEECH SYNTHESIS	3501
<i>Rohola Zandie, Mohammad H. Mahoor, Julia Madsen, Eshrat S. Emamian</i>	
PERCEPTION OF SOCIAL SPEAKER CHARACTERISTICS IN SYNTHETIC SPEECH.....	3506
<i>Sai Sirisha Rallabandi, Abhinav Bharadwaj, Babak Naderi, Sebastian Moller</i>	
KAZAKHTTS: AN OPEN-SOURCE KAZAKH TEXT-TO-SPEECH SYNTHESIS DATASET	3511
<i>Saida Mussakhoyeva, Aigerim Janaliyeva, Almas Mirzakhmetov, Yerbolat Khassanov, Huseyin Atakan Varol</i>	
SPECTRAL AND LATENT SPEECH REPRESENTATION DISTORTION FOR TTS EVALUATION	3516
<i>Thananchai Kongthaworn, Burin Naowarat, Ekapol Chuangsuwanich</i>	
UTILIZING SELF-SUPERVISED REPRESENTATIONS FOR MOS PREDICTION.....	3521
<i>Wei-Cheng Tseng, Chien-Yu Huang, Wei-Tsung Kao, Yist Y. Lin, Hung-Yi Lee</i>	
AISHELL-3: A MULTI-SPEAKER MANDARIN TTS CORPUS	3526
<i>Yao Shi, Hui Bu, Xin Xu, Shaoji Zhang, Ming Li</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: TOWARD END-TO-END SYNTHESIS I

SPEECH RESYNTHESIS FROM DISCRETE DISENTANGLED SELF-SUPERVISED REPRESENTATIONS	3531
<i>Adam Polyak, Yossi Adi, Jade Copet, Eugene Kharitonov, Kushal Lakhotia, Wei-Ning Hsu, Abdelrahman Mohamed, Emmanuel Dupoux</i>	
ZERO-SHOT TEXT-TO-SPEECH FOR TEXT-BASED INSERTION IN AUDIO NARRATION.....	3536
<i>Chuanxin Tang, Chong Luo, Zhiyuan Zhao, Dacheng Yin, Yucheng Zhao, Wenjun Zeng</i>	
A UNIVERSAL MULTI-SPEAKER MULTI-STYLE TEXT-TO-SPEECH VIA DISENTANGLED REPRESENTATION LEARNING BASED ON RENYI DIVERGENCE MINIMIZATION	3541
<i>Dipjyoti Paul, Sankar Mukherjee, Yannis Pantazis, Yannis Stylianou</i>	
SC-GLOWTTS: AN EFFICIENT ZERO-SHOT MULTI-SPEAKER TEXT-TO-SPEECH MODEL.....	3546
<i>Edresson Casanova, Christopher Shulby, Eren Golge, Nicolas Michael Muller, Frederico Santos De Oliveira, Arnaldo Candido Jr., Anderson Da Silva Soares, Sandra Maria Aluisio, Moacir Antonelli Ponti</i>	
LITETTS: A LIGHTWEIGHT MEL-SPECTROGRAM-FREE TEXT-TO-WAVE SYNTHESIZER BASED ON GENERATIVE ADVERSARIAL NETWORKS	3551
<i>Huu-Kim Nguyen, Kihyuk Jeong, Seyun Um, Min-Jae Hwang, Eunwoo Song, Hong-Goo Kang</i>	
REINFORCE-ALIGNER: REINFORCEMENT ALIGNMENT SEARCH FOR ROBUST END-TO-END TEXT-TO-SPEECH	3556
<i>Hyunseung Chung, Sang-Hoon Lee, Seong-Whan Lee</i>	
HIERARCHICAL CONTEXT-AWARE TRANSFORMERS FOR NON-AUTOREGRESSIVE TEXT TO SPEECH.....	3561
<i>Jae-Sung Bae, Taejun Bak, Young-Sun Joo, Hoon-Young Cho</i>	
DIFF-TTS: A DENOISING DIFFUSION MODEL FOR TEXT-TO-SPEECH.....	3566
<i>Myeonghun Jeong, Hyeongju Kim, Sung Jun Cheon, Byoung Jin Choi, Nam Soo Kim</i>	
A LEARNED CONDITIONAL PRIOR FOR THE VAE ACOUSTIC SPACE OF A TTS SYSTEM	3571
<i>Penny Karanasou, Sri Karlapati, Alexis Moinet, Arnaud Joly, Ammar Abbas, Simon Slangen, Jaime Lorenzo-Trueba, Thomas Drugman</i>	
TRIPLE M: A PRACTICAL TEXT-TO-SPEECH SYNTHESIS SYSTEM WITH MULTI-GUIDANCE ATTENTION AND MULTI-BAND MULTI-TIME LPCNET	3576
<i>Shilun Lin, Fenglong Xie, Li Meng, Xinhui Li, Li Lu</i>	
RELATIONAL DATA SELECTION FOR DATA AUGMENTATION OF SPEAKER-DEPENDENT MULTI-BAND MELGAN VOCODER	3581
<i>Yi-Chiao Wu, Cheng-Hung Hu, Hung-Shin Lee, Yu-Huai Peng, Wen-Chin Huang, Yu Tsao, Hsin-Min Wang, Tomoki Toda</i>	
FEDERATED LEARNING WITH DYNAMIC TRANSFORMER FOR TEXT TO SPEECH.....	3586
<i>Zhenhou Hong, Jianzong Wang, Xiaoyang Qu, Jie Liu, Chendong Zhao, Jing Xiao</i>	

INTERSPEECH 2021 : SPEECH SYNTHESIS: TOWARD END-TO-END SYNTHESIS II

TACOLPCNET: FAST AND STABLE TTS BY CONDITIONING LPCNET ON MEL SPECTROGRAM PREDICTIONS.....	3591
<i>Cheng Gong, Longbiao Wang, Ju Zhang, Shaotong Guo, Yuguang Wang, Jianwu Dang</i>	
TRANSFORMER-BASED ACOUSTIC MODELING FOR STREAMING SPEECH SYNTHESIS	3596
<i>Chunyang Wu, Zhiping Xiu, Yangyang Shi, Ozlem Kalinli, Christian Fuegen, Thilo Koehler, Qing He</i>	
PARALLEL TACOTRON 2: A NON-AUTOREGRESSIVE NEURAL TTS MODEL WITH DIFFERENTIABLE DURATION MODELING	3601
<i>Isaac Elias, Heiga Zen, Jonathan Shen, Yu Zhang, Ye Jia, R. J. Skerry-Ryan, Yonghui Wu</i>	
PHONETIC AND PROSODIC INFORMATION ESTIMATION FROM TEXTS FOR GENUINE JAPANESE END-TO-END TEXT-TO-SPEECH	3606
<i>Naoto Kakegawa, Sunao Hara, Masanobu Abe, Yusuke Ijima</i>	
DELIBERATION-BASED MULTI-PASS SPEECH SYNTHESIS	3611
<i>Qingyun Dou, Xixin Wu, Moquan Wan, Yiting Lu, Mark J. F. Gales</i>	
FASTPITCHFORMANT: SOURCE-FILTER BASED DECOMPOSED MODELING FOR SPEECH SYNTHESIS.....	3616
<i>Taejun Bak, Jae-Sung Bae, Hanbin Bae, Young-Ik Kim, Hoon-Young Cho</i>	
SEQUENCE-TO-SEQUENCE LEARNING FOR DEEP GAUSSIAN PROCESS BASED SPEECH SYNTHESIS USING SELF-ATTENTION GP LAYER	3621
<i>Taiki Nakamura, Tomoki Koriyama, Hiroshi Saruwatari</i>	
INFORMATION SIEVE: CONTENT LEAKAGE REDUCTION IN END-TO-END PROSODY TRANSFER FOR EXPRESSIVE SPEECH SYNTHESIS	3626
<i>Xudong Dai, Cheng Gong, Longbiao Wang, Kaili Zhang</i>	
PNG BERT: AUGMENTED BERT ON PHONEMES AND GRAPHEMES FOR NEURAL TTS.....	3631
<i>Ye Jia, Heiga Zen, Jonathan Shen, Yu Zhang, Yonghui Wu</i>	
SPEED UP TRAINING WITH VARIABLE LENGTH INPUTS BY EFFICIENT BATCHING STRATEGIES	3636
<i>Zhenhao Ge, Lakshmish Kaushik, Masanori Omote, Saket Kumar</i>	

INTERSPEECH 2021 : SPEECH TYPE CLASSIFICATION AND DIAGNOSIS

A MULTI-BRANCH DEEP LEARNING NETWORK FOR AUTOMATED DETECTION OF COVID-19	3641
<i>Ahmed Fakhry, Xinyi Jiang, Jaclyn Xiao, Gunvant Chaudhari, Asriel Han</i>	
FAKE AUDIO DETECTION IN RESOURCE-CONSTRAINED SETTINGS USING MICROFEATURES.....	3646
<i>Hira Dharmyal, Ayesha Ali, Ihsan Ayyub Qazi, Agha Ali Raza</i>	
OPEN-SET AUDIO CLASSIFICATION WITH LIMITED TRAINING RESOURCES BASED ON AUGMENTATION ENHANCED VARIATIONAL AUTO-ENCODER GAN WITH DETECTION-CLASSIFICATION JOINT TRAINING	3651
<i>Kah Kuan Teh, Huy Dat Tran</i>	

AN AGENT FOR COMPETING WITH HUMANS IN A DECEPTIVE GAME BASED ON VOCAL CUES	3656
<i>Noa Mansbach, Evgeny Hershkovitch Neiterman, Amos Azaria</i>	
AGE ESTIMATION WITH SPEECH-AGE MODEL FOR HETEROGENEOUS SPEECH DATASETS.....	3661
<i>Ryu Takeda, Kazunori Komatani</i>	
AUTOMATIC DETECTION OF SHOUTED SPEECH SEGMENTS IN INDIAN NEWS DEBATES	3666
<i>Shikha Baghel, Mrinmoy Bhattacharjee, S. R. Mahadeva Prasanna, Prithwjit Guha</i>	
KNOWLEDGE DISTILLATION FOR SINGING VOICE DETECTION.....	3671
<i>Soumava Paul, Gurunath Reddy M., K. Sreenivasa Rao, Partha Pratim Das</i>	
DEEP SPECTRAL-CEPSTRAL FUSION FOR SHOUTED AND NORMAL SPEECH CLASSIFICATION.....	3676
<i>Takahiro Fukumori</i>	
COUGHING-BASED RECOGNITION OF COVID-19 WITH SPATIAL ATTENTIVE CONVLSTM RECURRENT NEURAL NETWORKS	3681
<i>Tianhao Yan, Hao Meng, Emilia Parada-Cabaleiro, Shuo Liu, Meishu Song, Bjorn W. Schuller</i>	
OVERLAPPED SPEECH DETECTION BASED ON SPECTRAL AND SPATIAL FEATURE FUSION	3686
<i>Weiguang Chen, Van Tung Pham, Eng Siong Chng, Xionghu Zhong</i>	
GENERALIZED SPOOFING DETECTION INSPIRED FROM AUDIO GENERATION ARTIFACTS	3691
<i>Yang Gao, Tyler Vuong, Mahsa Elyasi, Gaurav Bharaj, Rita Singh</i>	
RW-RESNET: A NOVEL SPEECH ANTI-SPOOFING MODEL USING RAW WAVEFORM	3696
<i>Youxuan Ma, Zongze Ren, Shugong Xu</i>	

INTERSPEECH 2021 : SPOKEN DIALOGUE SYSTEMS I

LEVERAGING ASR N-BEST IN DEEP ENTITY RETRIEVAL	3701
<i>Haoyu Wang, John Chen, Majid Laali, Kevin Durda, Jeff King, William Campbell, Yang Liu</i>	
USER-INITIATED REPETITION-BASED RECOVERY IN MULTI-UTTERANCE DIALOGUE SYSTEMS	3706
<i>Hoang Long Nguyen, Vincent Renkens, Joris Pelemans, Srividya Pranavi Potharaju, Anil Kumar Nalamalapu, Murat Akbacak</i>	
DOMAIN-SPECIFIC MULTI-AGENT DIALOG POLICY LEARNING IN MULTI-DOMAIN TASK-ORIENTED SCENARIOS	3711
<i>Li Tang, Yuke Si, Longbiao Wang, Jianwu Dang</i>	
SELF-SUPERVISED DIALOGUE LEARNING FOR SPOKEN CONVERSATIONAL QUESTION ANSWERING	3716
<i>Nuo Chen, Chenyu You, Yuexian Zou</i>	
ACT-AWARE SLOT-VALUE PREDICTING IN MULTI-DOMAIN DIALOGUE STATE TRACKING	3721
<i>Ruolin Su, Ting-Wei Wu, Biing-Hwang Juang</i>	

SEMANTIC TRANSPORTATION PROTOTYPICAL NETWORK FOR FEW-SHOT INTENT DETECTION.....	3726
<i>Weiyuan Xu, Peilin Zhou, Chenyu You, Yuexian Zou</i>	
NEURAL SPOKEN-RESPONSE GENERATION USING PROSODIC AND LINGUISTIC CONTEXT FOR CONVERSATIONAL SYSTEMS.....	3731
<i>Yoshihiro Yamazaki, Yuya Chiba, Takashi Nose, Akinori Ito</i>	
DIALOGUE SITUATION RECOGNITION FOR EVERYDAY CONVERSATION USING MULTIMODAL INFORMATION.....	3736
<i>Yuya Chiba, Ryuichiro Higashinaka</i>	

INTERSPEECH 2021 : SPOKEN DIALOGUE SYSTEMS II

ADAPTING LONG CONTEXT NLM FOR ASR RESCORING IN CONVERSATIONAL AGENTS.....	3741
<i>Ashish Shenoy, Sravan Bodapati, Monica Sunkara, Srikanth Ronanki, Katrin Kirchhoff</i>	
CONTEXTUALIZED ATTENTION-BASED KNOWLEDGE TRANSFER FOR SPOKEN CONVERSATIONAL QUESTION ANSWERING.....	3746
<i>Chenyu You, Nuo Chen, Yuexian Zou</i>	
JOINT RETRIEVAL-EXTRACTION TRAINING FOR EVIDENCE-AWARE DIALOG RESPONSE SELECTION.....	3751
<i>Hongyin Luo, James Glass, Garima Lalwani, Yi Zhang, Shang-Wen Li</i>	
HUMAN-TO-HUMAN CONVERSATION DATASET FOR LEARNING FINE-GRAINED TURN-TAKING ACTION.....	3756
<i>Kehan Chen, Zezhong Li, Suyang Dai, Wei Zhou, Haiqing Chen</i>	
CAUSAL CONFUSION REDUCTION FOR ROBUST MULTI-DOMAIN DIALOGUE POLICY.....	3761
<i>Mahdin Rohmatillah, Jen-Tzung Chien</i>	
PHONEMEBERT: JOINT LANGUAGE MODELLING OF PHONEME SEQUENCE AND ASR TRANSCRIPT.....	3766
<i>Mukuntha Narayanan Sundararaman, Ayush Kumar, Jithendra Vepa</i>	
TIMING GENERATING NETWORKS: NEURAL NETWORK BASED PRECISE TURN- TAKING TIMING PREDICTION IN MULTIPARTY CONVERSATION.....	3771
<i>Shinya Fujie, Hayato Katayama, Jin Sakuma, Tetsunori Kobayashi</i>	
INJECTING DESCRIPTIVE META-INFORMATION INTO PRE-TRAINED LANGUAGE MODELS WITH HYPERNETWORKS.....	3776
<i>Wenying Duan, Xiaoxi He, Zimu Zhou, Hong Rao, Lothar Thiele</i>	

INTERSPEECH 2021 : SPOKEN LANGUAGE PROCESSING I

A NOISE ROBUST METHOD FOR WORD-LEVEL PRONUNCIATION ASSESSMENT.....	3781
<i>Binghuai Lin, Liyuan Wang</i>	
ENHANCING SEMANTIC UNDERSTANDING WITH SELF-SUPERVISED METHODS FOR ABSTRACTIVE DIALOGUE SUMMARIZATION.....	3786
<i>Hyunjae Lee, Jaewoong Yun, Hyunjin Choi, Seongho Joe, Youngjune L. Gwon</i>	

DISFLUENCY DETECTION WITH UNLABELED DATA AND SMALL BERT MODELS	3791
<i>Johann C. Rocholl, Vicky Zayats, Daniel D. Walker, Noah B. Murad, Aaron Schneider, Daniel J. Liebling</i>	
TARGETED KEYWORD FILTERING FOR ACCELERATED SPOKEN TOPIC IDENTIFICATION	3796
<i>Jonathan Wintrode</i>	
ZERO-SHOT JOINT MODELING OF MULTIPLE SPOKEN-TEXT-STYLE CONVERSION TASKS USING SWITCHING TOKENS	3801
<i>Mana Ihori, Naoki Makishima, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, Ryo Masumura</i>	
SPEAKER TRANSITION PATTERNS IN THREE-PARTY CONVERSATION: EVIDENCE FROM ENGLISH, ESTONIAN AND SWEDISH.....	3806
<i>Marcin Wlodarczak, Emer Gilmartin</i>	
DISCRIMINATIVE SELF-TRAINING FOR PUNCTUATION PREDICTION.....	3811
<i>Qian Chen, Wen Wang, Mengzhe Chen, Qinglin Zhang</i>	
SMALLER: SCALING NEURAL ENTITY RESOLUTION FOR EDGE DEVICES.....	3816
<i>Ross McGowan, Jinru Su, Vince Dicocco, Thejaswi Muniyappa, Grant P. Strimel</i>	
SPEAKER-CONVERSATION FACTORIAL DESIGNS FOR DIARIZATION ERROR ANALYSIS	3821
<i>Scott Seyfarth, Sundararajan Srinivasan, Katrin Kirchhoff</i>	
MULTIMODAL SPEECH SUMMARIZATION THROUGH SEMANTIC CONCEPT LEARNING.....	3826
<i>Shruti Palaskar, Ruslan Salakhutdinov, Alan W. Black, Florian Metze</i>	

INTERSPEECH 2021 : SPOKEN LANGUAGE PROCESSING II

ACTIVE SPEAKER DETECTION AS A MULTI-OBJECTIVE OPTIMIZATION WITH UNCERTAINTY-BASED MULTIMODAL FUSION.....	3831
<i>Baptiste Pouthier, Laurent Pilati, Leela K. Gudupudi, Charles Bouveyron, Frederic Precioso</i>	
LOST IN INTERPRETING: SPEECH TRANSLATION FROM SOURCE OR INTERPRETER?.....	3836
<i>Dominik Machacek, Matus Zilinec, Ondrej Bojar</i>	
IMPACT OF ENCODING AND SEGMENTATION STRATEGIES ON END-TO-END SIMULTANEOUS SPEECH TRANSLATION	3841
<i>Ha Nguyen, Yannick Esteve, Laurent Besacier</i>	
IT'S NOT WHAT YOU SAID, IT'S HOW YOU SAID IT: DISCRIMINATIVE PERCEPTION OF SPEECH AS A MULTICHANNEL COMMUNICATION SYSTEM	3846
<i>Sarenne Wallbridge, Peter Bell, Catherine Lai</i>	

INTERSPEECH 2021 : SPOKEN LANGUAGE UNDERSTANDING I

DATA AUGMENTATION FOR SPOKEN LANGUAGE UNDERSTANDING VIA PRETRAINED LANGUAGE MODELS.....	3851
<i>Baolin Peng, Chenguang Zhu, Michael Zeng, Jianfeng Gao</i>	

DEXTER: DEEP ENCODING OF EXTERNAL KNOWLEDGE FOR NAMED ENTITY RECOGNITION IN VIRTUAL ASSISTANTS.....	3856
<i>Deepak Muralidharan, Joel Ruben Antony Moniz, Weicheng Zhang, Stephen Pulman, Lin Li, Megan Barnes, Jingjing Pan, Jason Williams, Alex Acero</i>	
INTEGRATING DIALOG HISTORY INTO END-TO-END SPOKEN LANGUAGE UNDERSTANDING SYSTEMS	3861
<i>Jatin Ganhotra, Samuel Thomas, Hong-Kwang J. Kuo, Sachindra Joshi, George Saon, Zoltan Tuske, Brian Kingsbury</i>	
FANS: FUSING ASR AND NLU FOR ON-DEVICE SLU	3866
<i>Martin Radfar, Athanasios Mouchtaris, Siegfried Kunzmann, Ariya Rastrow</i>	
PRE-TRAINING FOR SPOKEN LANGUAGE UNDERSTANDING WITH JOINT TEXTUAL AND PHONETIC REPRESENTATION LEARNING.....	3871
<i>Qian Chen, Wen Wang, Qinglin Zhang</i>	
PREDICTING TEMPORAL PERFORMANCE DROP OF DEPLOYED PRODUCTION SPOKEN LANGUAGE UNDERSTANDING MODELS	3876
<i>Quynh Do, Judith Gaspers, Daniil Sorokin, Patrick Lehnen</i>	
RETHINKING END-TO-END EVALUATION OF DECOMPOSABLE TASKS: A CASE STUDY ON SPOKEN LANGUAGE UNDERSTANDING	3881
<i>Siddhant Arora, Alissa Ostapenko, Vijay Viswanathan, Siddharth Dalmia, Florian Metze, Shinji Watanabe, Alan W. Black</i>	
COREFERENCE AUGMENTATION FOR MULTI-DOMAIN TASK-ORIENTED DIALOGUE STATE TRACKING	3886
<i>Ting Han, Chongxuan Huang, Wei Peng</i>	
A CONTEXT-AWARE HIERARCHICAL BERT FUSION NETWORK FOR MULTI-TURN DIALOG ACT DETECTION	3891
<i>Ting-Wei Wu, Ruolin Su, Bing-Hwang Juang</i>	
SEQUENTIAL END-TO-END INTENT AND SLOT LABEL CLASSIFICATION AND LOCALIZATION	3896
<i>Yiran Cao, Nihal Potdar, Anderson R. Avila</i>	
 <u>INTERSPEECH 2021 : SPOKEN LANGUAGE UNDERSTANDING II</u>	
AUGMENTING SLOT VALUES AND CONTEXTS FOR SPOKEN LANGUAGE UNDERSTANDING WITH PRETRAINED MODELS.....	3901
<i>Haitao Lin, Lu Xiang, Yu Zhou, Jiajun Zhang, Chengqing Zong</i>	
FACTORIZATION-AWARE TRAINING OF TRANSFORMERS FOR NATURAL LANGUAGE UNDERSTANDING ON THE EDGE.....	3906
<i>Hamidreza Saghir, Samridhi Choudhary, Sepehr Eghbali, Clement Chung</i>	
THE IMPACT OF INTENT DISTRIBUTION MISMATCH ON SEMI-SUPERVISED SPOKEN LANGUAGE UNDERSTANDING.....	3911
<i>Judith Gaspers, Quynh Do, Daniil Sorokin, Patrick Lehnen</i>	
INTENT DETECTION AND SLOT FILLING FOR VIETNAMESE	3916
<i>Mai Hoang Dao, Thanh Hung Truong, Dat Quoc Nguyen</i>	

END-TO-END SPOKEN LANGUAGE UNDERSTANDING FOR GENERALIZED VOICE ASSISTANTS	3921
<i>Michael Saxon, Samridhi Choudhary, Joseph P. McKenna, Athanasios Mouchtaris</i>	
THREE-MODULE MODELING FOR END-TO-END SPOKEN LANGUAGE UNDERSTANDING USING PRE-TRAINED DNN-HMM-BASED ACOUSTIC-PHONETIC MODEL.....	3926
<i>Nick J. C. Wang, Lu Wang, Yandan Sun, Haimei Kang, Dejun Zhang</i>	
BI-DIRECTIONAL JOINT NEURAL NETWORKS FOR INTENT CLASSIFICATION AND SLOT FILLING	3931
<i>Soyeon Caren Han, Siqu Long, Huichun Li, Henry Weld, Josiah Poon</i>	
SPEAK OR CHAT WITH ME: END-TO-END SPOKEN LANGUAGE UNDERSTANDING SYSTEM WITH FLEXIBLE INPUTS	3936
<i>Sujeong Cha, Wangrui Hou, Hyun Jung, My Phung, Michael Picheny, Hong-Kwang J. Kuo, Samuel Thomas, Edmilson Morais</i>	
END-TO-END CROSS-LINGUAL SPOKEN LANGUAGE UNDERSTANDING MODEL WITH MULTILINGUAL PRETRAINING	3941
<i>Xianwei Zhang, Liang He</i>	
KNOWLEDGE DISTILLATION FROM BERT TRANSFORMER TO SPEECH TRANSFORMER FOR INTENT CLASSIFICATION	3946
<i>Yidi Jiang, Bidisha Sharma, Maulik Madhavi, Haizhou Li</i>	
<u>INTERSPEECH 2021 : SPOKEN MACHINE TRANSLATION</u>	
TOWARDS SIMULTANEOUS MACHINE INTERPRETATION	3951
<i>Alejandro Perez-Gonzalez-De-Martos, Javier Iranzo-Sanchez, Adria Gimenez Pastor, Javier Jorge, Joan-Albert Silvestre-Cerda, Jorge Civera, Albert Sanchis, Alfons Juan</i>	
COVOST 2 AND MASSIVELY MULTILINGUAL SPEECH TRANSLATION.....	3956
<i>Changhan Wang, Anne Wu, Jiatao Gu, Juan Pino</i>	
LARGE-SCALE SELF- AND SEMI-SUPERVISED LEARNING FOR SPEECH TRANSLATION.....	3961
<i>Changhan Wang, Anne Wu, Juan Pino, Alexei Baevski, Michael Auli, Alexis Conneau</i>	
SUBTITLE TRANSLATION AS MARKUP TRANSLATION.....	3966
<i>Colin Cherry, Naveen Arivazhagan, Dirk Padfield, Maxim Krikun</i>	
LEXICAL MODELING OF ASR ERRORS FOR ROBUST SPEECH TRANSLATION.....	3971
<i>Giuseppe Martucci, Mauro Cettolo, Matteo Negri, Marco Turchi</i>	
TRANSCRIBING PARALINGUISTIC ACOUSTIC CUES TO TARGET LANGUAGE TEXT IN TRANSFORMER-BASED SPEECH-TO-TEXT TRANSLATION.....	3976
<i>Hirotaoka Tokuyama, Sakriani Sakti, Katsuhito Sudoh, Satoshi Nakamura</i>	
WEAKLY-SUPERVISED SPEECH-TO-TEXT MAPPING WITH VISUALLY CONNECTED NON-PARALLEL SPEECH-TEXT DATA USING CYCLIC PARTIALLY-ALIGNED TRANSFORMER	3981
<i>Johanes Effendi, Sakriani Sakti, Satoshi Nakamura</i>	

SPECREC: AN ALTERNATIVE SOLUTION FOR IMPROVING END-TO-END SPEECH-TO-TEXT TRANSLATION VIA SPECTROGRAM RECONSTRUCTION	3986
<i>Junkun Chen, Mingbo Ma, Renjie Zheng, Liang Huang</i>	
OPTIMALLY ENCODING INDUCTIVE BIASES INTO THE TRANSFORMER IMPROVES END-TO-END SPEECH TRANSLATION	3991
<i>Piyush Vyas, Anastasia Kuznetsova, Donald S. Williamson</i>	
END-TO-END SPEECH TRANSLATION VIA CROSS-MODAL PROGRESSIVE TRAINING.....	3996
<i>Rong Ye, Mingxuan Wang, Lei Li</i>	
EFFECTS OF FEATURE SCALING AND FUSION ON SIGN LANGUAGE TRANSLATION	4001
<i>Tejaswini Ananthanarayana, Lipisha Chaudhary, Ifeoma Nwogu</i>	
ALLOST: LOW-RESOURCE SPEECH TRANSLATION WITHOUT SOURCE TRANSCRIPTION	4006
<i>Yao-Fei Cheng, Hung-Shin Lee, Hsin-Min Wang</i>	
ASR POSTERIOR-BASED LOSS FOR MULTI-TASK END-TO-END SPEECH TRANSLATION	4011
<i>Yuka Ko, Katsuhito Sudoh, Sakriani Sakti, Satoshi Nakamura</i>	

INTERSPEECH 2021 : SPOKEN TERM DETECTION & VOICE SEARCH

TEACHING KEYWORD SPOTTERS TO SPOT NEW KEYWORDS WITH LIMITED EXAMPLES	4016
<i>Abhijeet Awasthi, Kevin Kilgour, Hassan Rom</i>	
KEYWORD TRANSFORMER: A SELF-ATTENTION MODEL FOR KEYWORD SPOTTING	4021
<i>Axel Berg, Mark O'Connor, Miguel Tairum Cruz</i>	
DO ACOUSTIC WORD EMBEDDINGS CAPTURE PHONOLOGICAL SIMILARITY? AN EMPIRICAL STUDY	4026
<i>Badr M. Abdullah, Marius Mosbach, Iuliia Zaitova, Bernd Mobius, Dietrich Klakow</i>	
AUXILIARY SEQUENCE LABELING TASKS FOR DISFLUENCY DETECTION	4031
<i>Dongyub Lee, Byeongil Ko, Myeong Cheol Shin, Taesun Whang, Daniel Lee, Eunhwa Kim, Eunggyun Kim, Jaechoon Jo</i>	
ENERGY-FRIENDLY KEYWORD SPOTTING SYSTEM USING ADD-BASED CONVOLUTION	4036
<i>Hang Zhou, Wenchao Hu, Yu Ting Yeung, Xiao Chen</i>	

VOLUME 6

AUTO-KWS 2021 CHALLENGE: TASK, DATASETS, AND BASELINES	4041
<i>Jingsong Wang, Yuxuan He, Chunyu Zhao, Qijie Shao, Wei-Wei Tu, Tom Ko, Hung-Yi Lee, Lei Xie</i>	
TEXT ANCHOR BASED METRIC LEARNING FOR SMALL-FOOTPRINT KEYWORD SPOTTING.....	4046
<i>Li Wang, Rongzhi Gu, Nuo Chen, Yuexian Zou</i>	
FEW-SHOT KEYWORD SPOTTING IN ANY LANGUAGE	4051
<i>Mark Mazumder, Colby Banbury, Josh Meyer, Pete Warden, Vijay Janapa Reddi</i>	

PERSONALIZED KEYPHRASE DETECTION USING SPEAKER AND ENVIRONMENT INFORMATION	4056
<i>Rajeev Rikhye, Quan Wang, Qiao Liang, Yanzhang He, Ding Zhao, Yiteng Huang, Arun Narayanan, Ian McGraw</i>	
STREAMING TRANSFORMER FOR HARDWARE EFFICIENT VOICE TRIGGER DETECTION AND FALSE TRIGGER MITIGATION	4061
<i>Vineet Garg, Wonil Chang, Siddharth Sigtia, Saurabh Adya, Pramod Simha, Pranay Dighe, Chandra Dhir</i>	
THE 2020 PERSONALIZED VOICE TRIGGER CHALLENGE: OPEN DATASETS, EVALUATION METRICS, BASELINE SYSTEM AND RESULTS	4066
<i>Yan Jia, Xingming Wang, Xiaoyi Qin, Yinping Zhang, Xuyang Wang, Junjie Wang, Dong Zhang, Ming Li</i>	
A META-LEARNING APPROACH FOR USER-DEFINED SPOKEN TERM CLASSIFICATION WITH VARYING CLASSES AND EXAMPLES	4071
<i>Yangbin Chen, Tom Ko, Jianping Wang</i>	
PARAPHRASE LABEL ALIGNMENT FOR VOICE APPLICATION RETRIEVAL IN SPOKEN LANGUAGE UNDERSTANDING	4076
<i>Zheng Gao, Radhika Arava, Qian Hu, Xibin Gao, Thahir Mohamed, Wei Xiao, Mohamed Abdelhady</i>	
<u>INTERSPEECH 2021 : STREAMING FOR ASR/RNN TRANSDUCERS</u>	
IMPROVING RNN-T ASR ACCURACY USING CONTEXT AUDIO	4081
<i>Andreas Schwarz, Ilya Sklyar, Simon Wiesler</i>	
CONTEXTUALIZED STREAMING END-TO-END SPEECH RECOGNITION WITH TRIE-BASED DEEP BIASING AND SHALLOW FUSION	4086
<i>Duc Le, Mahaveer Jain, Gil Keren, Suyoun Kim, Yangyang Shi, Jay Mahadeokar, Julian Chan, Yuan Shangguan, Christian Fuegen, Ozlem Kalinli, Yatharth Saraf, Michael L. Seltzer</i>	
STABLEEMIT: SELECTION PROBABILITY DISCOUNT FOR REDUCING EMISSION LATENCY OF STREAMING MONOTONIC ATTENTION ASR.....	4091
<i>Hirofumi Inaguma, Tatsuya Kawahara</i>	
MIXTURE MODEL ATTENTION: FLEXIBLE STREAMING AND NON-STREAMING AUTOMATIC SPEECH RECOGNITION	4096
<i>Kartik Audhkhasi, Tongzhou Chen, Bhuvana Ramabhadran, Pedro J. Moreno</i>	
MULTI-MODE TRANSFORMER TRANSDUCER WITH STOCHASTIC FUTURE CONTEXT	4101
<i>Kwangyoum Kim, Felix Wu, Prashant Sridhar, Kyu J. Han, Shinji Watanabe</i>	
STREAMING MULTI-TALKER SPEECH RECOGNITION WITH JOINT SPEAKER IDENTIFICATION	4106
<i>Liang Lu, Naoyuki Kanda, Jinyu Li, Yifan Gong</i>	
HMM-FREE ENCODER PRE-TRAINING FOR STREAMING RNN TRANSDUCER.....	4111
<i>Lu Huang, Jingyu Sun, Yufeng Tang, Junfeng Hou, Jinkun Chen, Jun Zhang, Zejun Ma</i>	
DUAL CAUSAL/NON-CAUSAL SELF-ATTENTION FOR STREAMING END-TO-END SPEECH RECOGNITION	4116
<i>Niko Moritz, Takaaki Hori, Jonathan Le Roux</i>	

STREAMING END-TO-END SPEECH RECOGNITION FOR HYBRID RNN-T/ATTENTION ARCHITECTURE.....	4121
<i>Takafumi Moriya, Tomohiro Tanaka, Takanori Ashihara, Tsubasa Ochiai, Hiroshi Sato, Atsushi Ando, Ryo Masumura, Marc Delcroix, Taichi Asami</i>	
AN EFFICIENT STREAMING NON-RECURRENT ON-DEVICE END-TO-END MODEL WITH IMPROVEMENTS TO RARE-WORD MODELING.....	4126
<i>Tara N. Sainath, Yanzhang He, Arun Narayanan, Rami Botros, Ruoming Pang, David Rybach, Cyril Allauzen, Ehsan Variani, James Qin, Quoc-Nam Le-The, Shuo-Yiin Chang, Bo Li, Anmol Gulati, Jiahui Yu, Chung-Cheng Chiu, Diamantino Caseiro, Wei Li, Qiao Liang, Pat Rondon</i>	
SUPER-HUMAN PERFORMANCE IN ONLINE LOW-LATENCY RECOGNITION OF CONVERSATIONAL SPEECH.....	4131
<i>Thai-Son Nguyen, Sebastian Stucker, Alex Waibel</i>	
BRIDGING THE GAP BETWEEN STREAMING AND NON-STREAMING ASR SYSTEMS BY DISTILLING ENSEMBLES OF CTC AND RNN-T MODELS	4136
<i>Thibault Doutré, Wei Han, Chung-Cheng Chiu, Ruoming Pang, Olivier Siohan, Liangliang Cao</i>	
MULTIPLE SOFTMAX ARCHITECTURE FOR STREAMING MULTILINGUAL END-TO-END ASR SYSTEMS	4141
<i>Vikas Joshi, Amit Das, Eric Sun, Rupesh R. Mehta, Jinyu Li, Yifan Gong</i>	
REDUCING EXPOSURE BIAS IN TRAINING RECURRENT NEURAL NETWORK TRANSDUCERS	4146
<i>Xiaodong Cui, Brian Kingsbury, George Saon, David Haws, Zoltan Tuske</i>	
<u>INTERSPEECH 2021 : TARGET SPEAKER DETECTION, LOCALIZATION AND SEPARATION</u>	
TIME DELAY ESTIMATION FOR SPEAKER LOCALIZATION USING CNN-BASED PARAMETRIZED GCC-PHAT FEATURES	4151
<i>Daniele Salvati, Carlo Drioli, Gian Luca Foresti</i>	
AUXILIARY LOSS FUNCTION FOR TARGET SPEECH EXTRACTION AND RECOGNITION WITH WEAK SUPERVISION BASED ON SPEAKER CHARACTERISTICS	4156
<i>Katerina Zmolikova, Marc Delcroix, Desh Raj, Shinji Watanabe, Jan Cernocky</i>	
USING X-VECTORS FOR SPEECH ACTIVITY DETECTION IN BROADCAST STREAMS	4161
<i>Lukas Mateju, Frantisek Kynych, Petr Cerva, Jindrich Zdansky, Jiri Malek</i>	
UNIVERSAL SPEAKER EXTRACTION IN THE PRESENCE AND ABSENCE OF TARGET SPEAKERS FOR SPEECH OF ONE AND TWO TALKERS.....	4166
<i>Marvin Borsdorf, Chenglin Xu, Haizhou Li, Tanja Schultz</i>	
REAL-TIME SPEAKER COUNTING IN A COCKTAIL PARTY SCENARIO USING ATTENTION-GUIDED CONVOLUTIONAL NEURAL NETWORK.....	4171
<i>Midia Yousefi, John H. L. Hansen</i>	

INTERSPEECH 2021 : THE ADRESSO CHALLENGE: DETECTING COGNITIVE DECLINE USING SPEECH ONLY

COMPARING ACOUSTIC-BASED APPROACHES FOR ALZHEIMER'S DISEASE DETECTION.....	4176
<i>Aparna Balagopalan, Jekaterina Novikova</i>	
AUTOMATIC DETECTION OF ALZHEIMER'S DISEASE USING SPONTANEOUS SPEECH ONLY	4181
<i>Jun Chen, Jieping Ye, Fengyi Tang, Jiayu Zhou</i>	
ALZHEIMER DISEASE RECOGNITION USING SPEECH-BASED EMBEDDINGS FROM PRE-TRAINED MODELS	4186
<i>Lara Gauder, Leonardo Pepino, Luciana Ferrer, Pablo Riera</i>	
ALZHEIMER'S DEMENTIA RECOGNITION USING ACOUSTIC, LEXICAL, DISFLUENCY AND SPEECH PAUSE FEATURES ROBUST TO NOISY INPUTS	4191
<i>Morteza Rohanian, Julian Hough, Matthew Purver</i>	
MODULAR MULTI-MODAL ATTENTION NETWORK FOR ALZHEIMER'S DISEASE DETECTION USING PATIENT AUDIO AND LANGUAGE DATA	4196
<i>Ning Wang, Yupeng Cao, Shuai Hao, Zongru Shao, K. P. Subbalakshmi</i>	
INFLUENCE OF THE INTERVIEWER ON THE AUTOMATIC ASSESSMENT OF ALZHEIMER'S DISEASE IN THE CONTEXT OF THE ADRESSO CHALLENGE.....	4201
<i>P. A. Perez-Toro, S. P. Bayerl, T. Arias-Vergara, J. C. Vasquez-Correa, P. Klumpp, M. Schuster, Elmar Noth, J. R. Orozco-Arroyave, K. Riedhammer</i>	
AUTOMATIC DETECTION AND ASSESSMENT OF ALZHEIMER DISEASE USING SPEECH AND LANGUAGE TECHNOLOGIES IN LOW-RESOURCE SCENARIOS	4206
<i>Raghavendra Pappagari, Jaejin Cho, Sonal Joshi, Laureano Moro-Velazquez, Piotr Zelasko, Jesus Villalba, Najim Dehak</i>	
DETECTING COGNITIVE DECLINE USING SPEECH ONLY: THE ADRESSO CHALLENGE	4211
<i>Saturnino Luz, Fasih Haider, Sofia De La Fuente, Davida Fromm, Brian Macwhinney</i>	
USING THE OUTPUTS OF DIFFERENT AUTOMATIC SPEECH RECOGNITION PARADIGMS FOR ACOUSTIC- AND BERT-BASED ALZHEIMER'S DEMENTIA DETECTION THROUGH SPONTANEOUS SPEECH.....	4216
<i>Yilin Pan, Bahman Mirheidari, Jennifer M. Harris, Jennifer C. Thompson, Matthew Jones, Julie S. Snowden, Daniel Blackburn, Heidi Christensen</i>	
WAVBERT: EXPLOITING SEMANTIC AND NON-SEMANTIC SPEECH USING WAV2VEC AND BERT FOR DEMENTIA DETECTION	4221
<i>Youxiang Zhu, Abdelrahman Obyat, Xiaohui Liang, John A. Batsis, Robert M. Roth</i>	
ALZHEIMER'S DISEASE DETECTION FROM SPONTANEOUS SPEECH THROUGH COMBINING LINGUISTIC COMPLEXITY AND (DIS)FLUENCY FEATURES WITH PRETRAINED LANGUAGE MODELS.....	4226
<i>Yu Qiao, Xuefeng Yin, Daniel Wiechmann, Elma Kerz</i>	
TACKLING THE ADRESSO CHALLENGE 2021: THE MUET-RMIT SYSTEM FOR ALZHEIMER'S DEMENTIA RECOGNITION FROM SPONTANEOUS SPEECH	4231
<i>Zafi Sherhan Syed, Muhammad Shehram Shah Syed, Margaret Lech, Elena Pirogova</i>	

INTERSPEECH 2021 : THE FIRST DICOVA CHALLENGE: DIAGNOSIS OF COVID-19 USING ACOUSTICS

COUGH-BASED COVID-19 DETECTION WITH CONTEXTUAL ATTENTION CONVOLUTIONAL NEURAL NETWORKS AND GENDER INFORMATION	4236
<i>Adria Mallol-Ragolta, Helena Cuesta, Emilia Gomez, Bjorn W. Schuller</i>	
DICOVA CHALLENGE: DATASET, TASK, AND BASELINE SYSTEM FOR COVID-19 DIAGNOSIS USING ACOUSTICS	4241
<i>Ananya Muguli, Lancelot Pinto, Nirmala R., Neeraj Sharma, Prashant Krishnan, Prasanta Kumar Ghosh, Rohit Kumar, Shrirama Bhat, Srikanth Raj Chetupalli, Sriram Ganapathy, Shreyas Ramoji, Viral Nanda</i>	
INVESTIGATING FEATURE SELECTION AND EXPLAINABILITY FOR COVID-19 DIAGNOSTICS FROM COUGH SOUNDS	4246
<i>Flavio Avila, Amir H. Poorjam, Deepak Mittal, Charles Dognin, Ananya Muguli, Rohit Kumar, Srikanth Raj Chetupalli, Sriram Ganapathy, Maneesh Singh</i>	
THE DICOVA 2021 CHALLENGE --- AN ENCODER-DECODER APPROACH FOR COVID-19 RECOGNITION FROM COUGHING AUDIO.....	4251
<i>Gauri Deshpande, Bjorn W. Schuller</i>	
DETECTING COVID-19 FROM AUDIO RECORDING OF COUGHS USING RANDOM FORESTS AND SUPPORT VECTOR MACHINES	4256
<i>Isabella Sodergren, Maryam Pahlavan Nodeh, Prakash Chandra Chhipa, Konstantina Nikolaidou, Gyorgy Kovacs</i>	
CLASSIFICATION OF COVID-19 FROM COUGH USING AUTOREGRESSIVE PREDICTIVE CODING PRETRAINING AND SPECTRAL DATA AUGMENTATION	4261
<i>John Harvill, Yash R. Wani, Mark Hasegawa-Johnson, Narendra Ahuja, David Beiser, David Chestek</i>	
COVID-19 DETECTION FROM SPECTRAL FEATURES ON THE DICOVA DATASET	4266
<i>Kotra Venkata Sai Ritwik, Shareef Babu Kalluri, Deepu Vijayasenan</i>	
PANACEA COUGH SOUND-BASED DIAGNOSIS OF COVID-19 FOR THE DICOVA 2021 CHALLENGE.....	4271
<i>Madhu R. Kamble, Jose A. Gonzalez-Lopez, Teresa Grau, Juan M. Espin, Lorenzo Cascioli, Yiqing Huang, Alejandro Gomez-Alanis, Jose Patino, Roberto Font, Antonio M. Peinado, Angel M. Gomez, Nicholas Evans, Maria A. Zuluaga, Massimiliano Todisco</i>	
DIAGNOSIS OF COVID-19 USING AUDITORY ACOUSTIC CUES.....	4276
<i>Rohan Kumar Das, Maulik Madhavi, Haizhou Li</i>	
CONTRASTIVE LEARNING OF COUGH DESCRIPTORS FOR AUTOMATIC COVID-19 PRELIMINARY DIAGNOSIS	4281
<i>Swapnil Bhosale, Upasana Tiwari, Rupayan Chakraborty, Sunil Kumar Kopparapu</i>	
RECOGNISING COVID-19 FROM COUGHING USING ENSEMBLES OF SVMs AND LSTMS WITH HANDCRAFTED AND DEEP AUDIO FEATURES	4286
<i>Vincent Karas, Bjorn W. Schuller</i>	

INTERSPEECH 2021 : THE INTERSPEECH 2021 COMPUTATIONAL PARALINGUISTICS CHALLENGE (COMPARE) --- COVID-19 COUGH, COVID-19 SPEECH, ESCALATION & PRIMATES

THE INTERSPEECH 2021 COMPUTATIONAL PARALINGUISTICS CHALLENGE: COVID-19 COUGH, COVID-19 SPEECH, ESCALATION & PRIMATES.....	4291
<i>Bjorn W. Schuller, Anton Batliner, Christian Bergler, Cecilia Mascolo, Jing Han, Iulia Lefter, Heysem Kaya, Shahin Amiriparian, Alice Baird, Lukas Stappen, Sandra Ottl, Maurice Gerczuk, Panagiotis Tzirakis, Chloe Brown, Jagmohan Chauhan, Andreas Grammenos, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, Leon J. M. Rothkrantz, Joeri A. Zwerts, Jelle Treep, Casper S. Kaandorp</i>	
ANALYSIS BY SYNTHESIS: USING AN EXPRESSIVE TTS MODEL AS FEATURE EXTRACTOR FOR PARALINGUISTIC SPEECH CLASSIFICATION.....	4296
<i>Dominik Schiller, Silvan Mertes, Pol Van Rijn, Elisabeth Andre</i>	
TRANSFER LEARNING AND DATA AUGMENTATION TECHNIQUES TO THE COVID-19 IDENTIFICATION TASKS IN COMPARE 2021	4301
<i>Edresson Casanova, Arnaldo Candido Jr., Ricardo Corso Fernandes Jr., Marcelo Finger, Lucas Rafael Stefanel Gris, Moacir Antonelli Ponti, Daniel Peixoto Pinto Da Silva</i>	
MULTI-ATTENTIVE DETECTION OF THE SPIDER MONKEY WHINNY IN THE (ACTUAL) WILD	4306
<i>Georgios Rizos, Jenna Lawson, Zhuoda Han, Duncan Butler, James Rosindell, Krystian Mikolajczyk, Cristina Banks-Leite, Bjorn W. Schuller</i>	
INTRODUCING A CENTRAL AFRICAN PRIMATE VOCALISATION DATASET FOR AUTOMATED SPECIES CLASSIFICATION	4311
<i>Joeri A. Zwerts, Jelle Treep, Casper S. Kaandorp, Floor Meewis, Amparo C. Koot, Heysem Kaya</i>	
IDENTIFYING CONFLICT ESCALATION AND PRIMATES BY USING ENSEMBLE X-VECTORS AND FISHER VECTOR FEATURES.....	4316
<i>Jose Vicente Egas-Lopez, Mercedes Vetrab, Laszlo Toth, Gabor Gosztolya</i>	
ENSEMBLE-WITHIN-ENSEMBLE CLASSIFICATION FOR ESCALATION PREDICTION FROM SPEECH.....	4321
<i>Oxana Verkholiyak, Denis Dresvyanskiy, Anastasia Dvoynikova, Denis Kotov, Elena Ryumina, Alena Velichko, Danila Mamontov, Wolfgang Minker, Alexey Karpov</i>	
THE PHONETIC FOOTPRINT OF COVID-19?	4326
<i>P. Klumpp, T. Bocklet, T. Arias-Vergara, J. C. Vasquez-Correa, P. A. Perez-Toro, S. P. Bayerl, J. R. Orozco-Arroyave, Elmar Noth</i>	
A DEEP AND RECURRENT ARCHITECTURE FOR PRIMATE VOCALIZATION CLASSIFICATION.....	4331
<i>Robert Muller, Steffen Illium, Claudia Linnhoff-Popien</i>	
TRANSFER LEARNING-BASED COUGH REPRESENTATIONS FOR AUTOMATIC DETECTION OF COVID-19.....	4336
<i>Ruben Solera-Urena, Catarina Botelho, Francisco Teixeira, Thomas Rolland, Alberto Abad, Isabel Trancoso</i>	
VISUAL TRANSFORMERS FOR PRIMATES CLASSIFICATION AND COVID DETECTION	4341
<i>Steffen Illium, Robert Muller, Andreas Sedlmeier, Claudia-Linnhoff Popien</i>	

DEEP-LEARNING-BASED CENTRAL AFRICAN PRIMATE SPECIES CLASSIFICATION WITH MIXUP AND SPECAUGMENT	4346
<i>Thomas Pellegrini</i>	

INTERSPEECH 2021 : TOOLS, CORPORA AND RESOURCES

AUSKIDTALK: AN AUDITORY-VISUAL CORPUS OF 3- TO 12-YEAR-OLD AUSTRALIAN CHILDREN'S SPEECH	4351
<i>Beena Ahmed, Kirrie J. Ballard, Denis Burnham, Tharmakulasingham Sirojan, Hadi Mehmood, Dominique Estival, Elise Baker, Felicity Cox, Joanne Arciuli, Titia Benders, Katherine Demuth, Barbara Kelly, Chloe Diskin-Holdaway, Mostafa Shahin, Vidhyasaharan Sethu, Julien Epps, Chwee Beng Lee, Eliathamby Ambikairajah</i>	
TUSOM2021: A PHONETICALLY TRANSCRIBED SPEECH DATASET FROM AN ENDANGERED LANGUAGE FOR UNIVERSAL PHONE RECOGNITION EXPERIMENTS	4356
<i>David R. Mortensen, Jordan Picone, Xinjian Li, Kathleen Siminyu</i>	
ANNOTATION CONFIDENCE VS. TRAINING SAMPLE SIZE: TRADE-OFF SOLUTION FOR PARTIALLY-CONTINUOUS CATEGORICAL EMOTION RECOGNITION.....	4361
<i>Elena Ryumina, Oxana Verkholyak, Alexey Karpov</i>	
THE MULTILINGUAL TEDX CORPUS FOR SPEECH RECOGNITION AND TRANSLATION	4366
<i>Elizabeth Salesky, Matthew Wiesner, Jacob Bremerman, Roldano Cattoni, Matteo Negri, Marco Turchi, Douglas W. Oard, Matt Post</i>	
EUROPARL-ASR: A LARGE CORPUS OF PARLIAMENTARY DEBATES FOR STREAMING ASR BENCHMARKING AND SPEECH DATA FILTERING/VERBATIMIZATION	4371
<i>Goncal V. Garces Diaz-Munio, Joan-Albert Silvestre-Cerda, Javier Jorge, Adria Gimenez Pastor, Javier Iranzo-Sanchez, Pau Baquero-Arnal, Nahuel Rosello, Alejandro Perez- Gonzalez-De-Martos, Jorge Civera, Albert Sanchis, Alfons Juan</i>	
GIGASPEECH: AN EVOLVING, MULTI-DOMAIN ASR CORPUS WITH 10,000 HOURS OF TRANSCRIBED AUDIO	4376
<i>Guoguo Chen, Shuzhou Chai, Guan-Bo Wang, Jiayu Du, Wei-Qiang Zhang, Chao Weng, Dan Su, Daniel Povey, Jan Trmal, Junbo Zhang, Mingjie Jin, Sanjeev Khudanpur, Shinji Watanabe, Shuaijiang Zhao, Wei Zou, Xiangang Li, Xuchen Yao, Yongqing Wang, Zhao You, Zhiyong Yan</i>	
SPOKEN OBJECTNET: A BIAS-CONTROLLED SPOKEN CAPTION DATASET.....	4381
<i>Ian Palmer, Andrew Rouditchenko, Andrei Barbu, Boris Katz, James Glass</i>	
SPEECHOCEAN762: AN OPEN-SOURCE NON-NATIVE ENGLISH SPEECH CORPUS FOR PRONUNCIATION ASSESSMENT	4386
<i>Junbo Zhang, Zhiwen Zhang, Yongqing Wang, Zhiyong Yan, Qiong Song, Yukai Huang, Ke Li, Daniel Povey, Yujun Wang</i>	
TOWARDS AUTOMATIC SPEECH TO SIGN LANGUAGE GENERATION	4391
<i>Parul Kapoor, Rudrabha Mukhopadhyay, Sindhu B. Hegde, Vinay Namboodiri, C. V. Jawahar</i>	
HUMAN-IN-THE-LOOP EFFICIENCY ANALYSIS FOR BINARY CLASSIFICATION IN EDYSON.....	4396
<i>Per Fallgren, Jens Edlund</i>	
KOSP2E: KOREAN SPEECH TO ENGLISH TRANSLATION CORPUS	4401
<i>Won Ik Cho, Seok Min Kim, Hyunchang Cho, Nam Soo Kim</i>	

AISHELL-4: AN OPEN SOURCE DATASET FOR SPEECH ENHANCEMENT, SEPARATION,
RECOGNITION AND SPEAKER DIARIZATION IN CONFERENCE SCENARIO..... 4406
*Yihui Fu, Luyao Cheng, Shubo Lv, Yukai Jv, Yuxiang Kong, Zhuo Chen, Yanxin Hu, Lei Xie,
Jian Wu, Hui Bu, Xin Xu, Jun Du, Jingdong Chen*

LOOK WHO'S TALKING: ACTIVE SPEAKER DETECTION IN THE WILD 4411
*You Jin Kim, Hee-Soo Heo, Soyeon Choe, Soo-Whan Chung, Yoohwan Kwon, Bong-Jin Lee,
Youngki Kwon, Joon Son Chung*

**INTERSPEECH 2021 : TOPICS IN ASR: ADAPTATION, TRANSFER LEARNING,
CHILDREN'S SPEECH, AND LOW-RESOURCE SETTINGS**

THE TAL SYSTEM FOR THE INTERSPEECH2021 SHARED TASK ON AUTOMATIC
SPEECH RECOGNITION FOR NON-NATIVE CHILDRENS SPEECH..... 4416
Gaopeng Xu, Song Yang, Lu Ma, Chengfei Li, Zhongqin Wu

THE TAL SYSTEM FOR THE INTERSPEECH2021 SHARED TASK ON AUTOMATIC
SPEECH RECOGNITION FOR NON-NATIVE CHILDRENS SPEECH..... 4421
Gaopeng Xu, Song Yang, Lu Ma, Chengfei Li, Zhongqin Wu

ZERO-SHOT CROSS-LINGUAL PHONETIC RECOGNITION WITH EXTERNAL
LANGUAGE EMBEDDING 4426
Heting Gao, Junrui Ni, Yang Zhang, Kaizhi Qian, Shiyu Chang, Mark Hasegawa-Johnson

ZERO-SHOT CROSS-LINGUAL PHONETIC RECOGNITION WITH EXTERNAL
LANGUAGE EMBEDDING 4431
Heting Gao, Junrui Ni, Yang Zhang, Kaizhi Qian, Shiyu Chang, Mark Hasegawa-Johnson

SEMANTIC DATA AUGMENTATION FOR END-TO-END MANDARIN SPEECH
RECOGNITION 4436
*Jianwei Sun, Zhiyuan Tang, Hengxin Yin, Wei Wang, Xi Zhao, Shuaijiang Zhao, Xiaoning
Lei, Wei Zou, Xiangang Li*

LOW RESOURCE GERMAN ASR WITH UNTRANSCRIBED DATA SPOKEN BY NON-
NATIVE CHILDREN --- INTERSPEECH 2021 SHARED TASK SPAPL SYSTEM..... 4441
Jinhan Wang, Yunzheng Zhu, Ruchao Fan, Wei Chu, Abeer Alwan

LOW RESOURCE GERMAN ASR WITH UNTRANSCRIBED DATA SPOKEN BY NON-
NATIVE CHILDREN --- INTERSPEECH 2021 SHARED TASK SPAPL SYSTEM..... 4446
Jinhan Wang, Yunzheng Zhu, Ruchao Fan, Wei Chu, Abeer Alwan

ROBUST CONTINUOUS ON-DEVICE PERSONALIZATION FOR AUTOMATIC SPEECH
RECOGNITION 4451
*Khe Chai Sim, Angad Chandorkar, Fan Gao, Mason Chua, Tsendsuren Munkhdalai,
Francoise Beaufays*

ROBUST CONTINUOUS ON-DEVICE PERSONALIZATION FOR AUTOMATIC SPEECH
RECOGNITION 4456
*Khe Chai Sim, Angad Chandorkar, Fan Gao, Mason Chua, Tsendsuren Munkhdalai,
Francoise Beaufays*

BEST OF BOTH WORLDS: ROBUST ACCENTED SPEECH RECOGNITION WITH
ADVERSARIAL TRANSFER LEARNING 4461
*Nilaksh Das, Sravan Bodapati, Monica Sunkara, Sundararajan Srinivasan, Duen Hornq
Chau*

BEST OF BOTH WORLDS: ROBUST ACCENTED SPEECH RECOGNITION WITH ADVERSARIAL TRANSFER LEARNING	4466
<i>Nilaksh Das, Sravan Bodapati, Monica Sunkara, Sundararajan Srinivasan, Duen Hornq Chau</i>	
SPEAKER NORMALIZATION USING JOINT VARIATIONAL AUTOENCODER.....	4471
<i>Shashi Kumar, Shakti P. Rath, Abhishek Pandey</i>	
SPEAKER NORMALIZATION USING JOINT VARIATIONAL AUTOENCODER.....	4476
<i>Shashi Kumar, Shakti P. Rath, Abhishek Pandey</i>	
ON-THE-FLY ALIGNED DATA AUGMENTATION FOR SEQUENCE-TO-SEQUENCE ASR.....	4481
<i>Tsz Kin Lam, Mayumi Ohta, Shigehiko Schamoni, Stefan Riezler</i>	
ON-THE-FLY ALIGNED DATA AUGMENTATION FOR SEQUENCE-TO-SEQUENCE ASR.....	4486
<i>Tsz Kin Lam, Mayumi Ohta, Shigehiko Schamoni, Stefan Riezler</i>	
EXTENDING PRONUNCIATION DICTIONARY WITH AUTOMATICALLY DETECTED WORD MISPRONUNCIATIONS TO IMPROVE PAII'S SYSTEM FOR INTERSPEECH 2021 NON-NATIVE CHILD ENGLISH CLOSE TRACK ASR CHALLENGE.....	4491
<i>Wei Chu, Peng Chang, Jing Xiao</i>	
EXTENDING PRONUNCIATION DICTIONARY WITH AUTOMATICALLY DETECTED WORD MISPRONUNCIATIONS TO IMPROVE PAII'S SYSTEM FOR INTERSPEECH 2021 NON-NATIVE CHILD ENGLISH CLOSE TRACK ASR CHALLENGE.....	4496
<i>Wei Chu, Peng Chang, Jing Xiao</i>	
LAYER-WISE FAST ADAPTATION FOR END-TO-END MULTI-ACCENT SPEECH RECOGNITION.....	4501
<i>Xun Gong, Yizhou Lu, Zhikai Zhou, Yanmin Qian</i>	
LAYER-WISE FAST ADAPTATION FOR END-TO-END MULTI-ACCENT SPEECH RECOGNITION.....	4506
<i>Xun Gong, Yizhou Lu, Zhikai Zhou, Yanmin Qian</i>	
RAPID SPEAKER ADAPTATION FOR CONFORMER TRANSDUCER: ATTENTION AND BIAS ARE ALL YOU NEED	4511
<i>Yan Huang, Guoli Ye, Jinyu Li, Yifan Gong</i>	
RAPID SPEAKER ADAPTATION FOR CONFORMER TRANSDUCER: ATTENTION AND BIAS ARE ALL YOU NEED	4516
<i>Yan Huang, Guoli Ye, Jinyu Li, Yifan Gong</i>	
<u>INTERSPEECH 2021 : VOICE ACTIVITY DETECTION</u>	
VOICE ACTIVITY DETECTION WITH TEACHER-STUDENT DOMAIN EMULATION.....	4521
<i>Jarrold Luckenbaugh, Samuel Abplanalp, Rachel Gonzalez, Daniel Fulford, David Gard, Carlos Busso</i>	
EML ONLINE SPEECH ACTIVITY DETECTION FOR THE FEARLESS STEPS CHALLENGE PHASE-III.....	4526
<i>Omid Ghahabi, Volker Fischer</i>	
UNSUPERVISED REPRESENTATION LEARNING FOR SPEECH ACTIVITY DETECTION IN THE FEARLESS STEPS CHALLENGE 2021	4530
<i>Pablo Gimeno, Alfonso Ortega, Antonio Miguel, Eduardo Lleida</i>	

SPEECH ACTIVITY DETECTION BASED ON MULTILINGUAL SPEECH RECOGNITION SYSTEM 4535
Seyyed Saeed Sarfjoo, Srikanth Madikeri, Petr Motlicek

THE APPLICATION OF LEARNABLE STRF KERNELS TO THE 2021 FEARLESS STEPS PHASE-03 SAD CHALLENGE 4540
Tyler Vuong, Yangyang Xia, Richard M. Stern

INTERSPEECH 2021 : VOICE ACTIVITY DETECTION AND KEYWORD SPOTTING

END-TO-END TRANSFORMER-BASED OPEN-VOCABULARY KEYWORD SPOTTING WITH LOCATION-GUIDED LOCAL ATTENTION 4545
Bo Wei, Meirong Yang, Tao Zhang, Xiao Tang, Xing Huang, Kyuhong Kim, Jaeyun Lee, Kiho Cho, Sung-Un Park

AUDIO-VISUAL INFORMATION FUSION USING CROSS-MODAL TEACHER-STUDENT LEARNING FOR VOICE ACTIVITY DETECTION IN REALISTIC ENVIRONMENTS 4550
Hengshun Zhou, Jun Du, Hang Chen, Zijun Jing, Shifu Xiong, Chin-Hui Lee

NOISY STUDENT-TEACHER TRAINING FOR ROBUST KEYWORD SPOTTING 4555
Hyun-Jin Park, Pai Zhu, Ignacio Lopez Moreno, Niranjana Subrahmanya

ENROLLMENT-LESS TRAINING FOR PERSONALIZED VOICE ACTIVITY DETECTION 4560
Naoki Makishima, Mana Ihori, Tomohiro Tanaka, Akihiko Takashima, Shota Orihashi, Ryo Masumura

MULTI-CHANNEL VAD FOR TRANSCRIPTION OF GROUP DISCUSSION 4565
Osamu Ichikawa, Kaito Nakano, Takahiro Nakayama, Hajime Shirouzu

SEGMENTAL CONTRASTIVE PREDICTIVE CODING FOR UNSUPERVISED WORD SEGMENTATION 4570
Saurabhchand Bhati, Jesus Villalba, Piotr Zelasko, Laureano Moro-Velazquez, Najim Dehak

NOISE-TOLERANT SELF-SUPERVISED LEARNING FOR AUDIO-VISUAL VOICE ACTIVITY DETECTION 4575
Ui-Hyun Kim

A LIGHTWEIGHT FRAMEWORK FOR ONLINE VOICE ACTIVITY DETECTION IN THE WILD 4580
Xuenan Xu, Heinrich Dinkel, Mengyue Wu, Kai Yu

FASTICARL: FAST INCREMENTAL CLASSIFIER AND REPRESENTATION LEARNING WITH EFFICIENT BUDGET ALLOCATION IN AUDIO SENSING APPLICATIONS 4585
Young D. Kwon, Jagmohan Chauhan, Cecilia Mascolo

ATTENTION-BASED CROSS-MODAL FUSION FOR AUDIO-VISUAL VOICE ACTIVITY DETECTION IN MUSICAL VIDEO STREAMS 4590
Yuanbo Hou, Zhesong Yu, Xia Liang, Xingjian Du, Bilei Zhu, Zejun Ma, Dick Botteldooren

VOICE ACTIVITY DETECTION FOR LIVE SPEECH OF BASEBALL GAME BASED ON TANDEM CONNECTION WITH SPEECH/NOISE SEPARATION MODEL 4595
Yuto Nonaka, Chee Siang Leow, Akio Kobayashi, Takehito Utsuro, Hiromitsu Nishizaki

INTERSPEECH 2021 : VOICE AND VOICING

VOICING ASSIMILATIONS BY FRENCH SPEAKERS OF GERMAN IN STOP-FRICATIVE SEQUENCES.....	4600
<i>Anne Bonneau</i>	
“SEE WHAT I MEAN, HUH?” EVALUATING VISUAL INSPECTION OF F ₍₀₎ TRACKING IN NASAL GRUNTS.....	4605
<i>Aurelie Chlebowski, Nicolas Ballier</i>	
SYSTEM PERFORMANCE AS A FUNCTION OF CALIBRATION METHODS, SAMPLE SIZE AND SAMPLING VARIABILITY IN LIKELIHOOD RATIO-BASED FORENSIC VOICE COMPARISON.....	4610
<i>Bruce Xiao Wang, Vincent Hughes</i>	
CROSS-LINGUISTIC SPEAKER INDIVIDUALITY OF LONG-TERM FORMANT DISTRIBUTIONS: PHONETIC AND FORENSIC PERSPECTIVES	4615
<i>Justin J. H. Lo</i>	
MAP ADAPTATION CHARACTERISTICS IN FORENSIC LONG-TERM FORMANT ANALYSIS	4620
<i>Michael Jessen</i>	
VOICING CONTRASTS IN THE SINGLETON STOPS OF PALESTINIAN ARABIC: PRODUCTION AND PERCEPTION.....	4625
<i>Nour Tamim, Silke Hamann</i>	
SOUND CHANGE IN SPONTANEOUS BILINGUAL SPEECH: A CORPUS STUDY ON THE CANTONESE N-L MERGER IN CANTONESE-ENGLISH BILINGUALS	4630
<i>Rachel Soo, Khia A. Johnson, Molly Babel</i>	
ACOUSTIC AND PROSODIC CORRELATES OF EMOTIONS IN URDU SPEECH	4635
<i>Saba Urooj, Benazir Mumtaz, Sarmad Hussain, Ehsan Ul Haq</i>	
A COMPARISON OF THE ACCURACY OF DISSEN AND KESHET'S (2016) DEEPPFORMANTS AND TRADITIONAL LPC METHODS FOR SEMI-AUTOMATIC SPEAKER RECOGNITION	4640
<i>Thomas Coy, Vincent Hughes, Philip Harrison, Amelia J. Gully</i>	
THE FOUR-WAY CLASSIFICATION OF STOPS WITH VOICING AND ASPIRATION FOR NON-NATIVE SPEECH EVALUATION.....	4645
<i>Titus Chakraborty, Vaishali Patil, Preeti Rao</i>	
CHARACTERIZING VOICED AND VOICELESS NASALS IN MIZO	4650
<i>Wendy Lalhminghlui, Priyankoo Sarmah</i>	

INTERSPEECH 2021 : VOICE ANTI-SPOOFING AND COUNTERMEASURE

VOTING FOR THE RIGHT ANSWER: ADVERSARIAL DEFENSE FOR SPEAKER VERIFICATION	4655
<i>Haibin Wu, Yang Zhang, Zhiyong Wu, Dong Wang, Hung-Yi Lee</i>	
ATTENTION-BASED CONVOLUTIONAL NEURAL NETWORK FOR ASV SPOOFING DETECTION.....	4660
<i>Hefei Ling, Leichao Huang, Junrui Huang, Baiyan Zhang, Ping Li</i>	

REPRESENTATION LEARNING TO CLASSIFY AND DETECT ADVERSARIAL ATTACKS AGAINST SPEAKER AND SPEECH RECOGNITION SYSTEMS	4665
<i>Jesus Villalba, Sonal Joshi, Piotr Zelasko, Najim Dehak</i>	
AN INITIAL INVESTIGATION FOR DETECTING PARTIALLY SPOOFED AUDIO	4670
<i>Lin Zhang, Xin Wang, Erica Cooper, Junichi Yamagishi, Jose Patino, Nicholas Evans</i>	
VISUALIZING CLASSIFIER ADJACENCY RELATIONS: A CASE STUDY IN SPEAKER VERIFICATION AND VOICE ANTI-SPOOFING	4675
<i>Tomi Kinnunen, Andreas Nautsch, Md. Sahidullah, Nicholas Evans, Xin Wang, Massimiliano Todisco, Hector Delgado, Junichi Yamagishi, Kong Aik Lee</i>	
PARTIALLY-CONNECTED DIFFERENTIABLE ARCHITECTURE SEARCH FOR DEEPPFAKE AND SPOOFING DETECTION	4680
<i>Wanying Ge, Michele Panariello, Jose Patino, Massimiliano Todisco, Nicholas Evans</i>	
A COMPARATIVE STUDY ON RECENT NEURAL SPOOFING COUNTERMEASURES FOR SYNTHETIC SPEECH DETECTION.....	4685
<i>Xin Wang, Junichi Yamagishi</i>	
CROSS-DATABASE REPLAY DETECTION IN TERMINAL-DEPENDENT SPEAKER VERIFICATION.....	4690
<i>Xingliang Cheng, Mingxing Xu, Thomas Fang Zheng</i>	
CHANNEL-WISE GATED RES2NET: TOWARDS ROBUST DETECTION OF SYNTHETIC SPEECH ATTACKS.....	4695
<i>Xu Li, Xixin Wu, Hui Lu, Xunying Liu, Helen Meng</i>	
SIAMESE NETWORK WITH WAV2VEC FEATURE FOR SPOOFING SPEECH DETECTION	4700
<i>Yang Xie, Zhenchuan Zhang, Yingchun Yang</i>	
AN EMPIRICAL STUDY ON CHANNEL EFFECTS FOR SYNTHETIC VOICE SPOOFING COUNTERMEASURE SYSTEMS	4705
<i>You Zhang, Ge Zhu, Fei Jiang, Zhiyao Duan</i>	
THE EFFECT OF SILENCE AND DUAL-BAND FUSION IN ANTI-SPOOFING SYSTEM.....	4710
<i>Yuxiang Zhang, Wenchao Wang, Pengyuan Zhang</i>	
PAIRING WEAK WITH STRONG: TWIN MODELS FOR DEFENDING AGAINST ADVERSARIAL ATTACK ON SPEAKER VERIFICATION.....	4715
<i>Zhiyuan Peng, Xu Li, Tan Lee</i>	

INTERSPEECH 2021 : VOICE CONVERSION AND ADAPTATION I

VQMIVC: VECTOR QUANTIZATION AND MUTUAL INFORMATION-BASED UNSUPERVISED SPEECH REPRESENTATION DISENTANGLEMENT FOR ONE-SHOT VOICE CONVERSION	4720
<i>Disong Wang, Liqun Deng, Yu Ting Yeung, Xiao Chen, Xunying Liu, Helen Meng</i>	
IMPROVING ROBUSTNESS OF ONE-SHOT VOICE CONVERSION WITH DEEP DISCRIMINATIVE SPEAKER ENCODER.....	4725
<i>Hongqiang Du, Lei Xie</i>	
NORMALIZATION DRIVEN ZERO-SHOT MULTI-SPEAKER SPEECH SYNTHESIS.....	4730
<i>Neeraj Kumar, Srishti Goel, Ankur Narang, Brejesh Lall</i>	

ONE-SHOT VOICE CONVERSION WITH SPEAKER-AGNOSTIC STARGAN	4735
<i>Sefik Emre Eskimez, Dimitrios Dimitriadis, Kenichi Kumatani, Robert Gmyr</i>	
STARGAN-VC+ASR: STARGAN-BASED NON-PARALLEL VOICE CONVERSION REGULARIZED BY AUTOMATIC SPEECH RECOGNITION	4740
<i>Shoki Sakamoto, Akira Taniguchi, Tadahiro Taniguchi, Hirokazu Kameoka</i>	
FINE-TUNING PRE-TRAINED VOICE CONVERSION MODEL FOR ADDING NEW TARGET SPEAKERS WITH LIMITED DATA	4745
<i>Takeshi Koshizuka, Hidefumi Ohmura, Kouichi Katsurada</i>	
CVC: CONTRASTIVE LEARNING FOR NON-PARALLEL VOICE CONVERSION	4750
<i>Tingle Li, Yichen Liu, Chenxu Hu, Hang Zhao</i>	
A PRELIMINARY STUDY OF A TWO-STAGE PARADIGM FOR PRESERVING SPEAKER IDENTITY IN DYSARTHIC VOICE CONVERSION	4755
<i>Wen-Chin Huang, Kazuhiro Kobayashi, Yu-Huai Peng, Ching-Feng Liu, Yu Tsao, Hsin-Min Wang, Tomoki Toda</i>	
TWO-PATHWAY STYLE EMBEDDING FOR ARBITRARY VOICE CONVERSION	4760
<i>Xuexin Xu, Liang Shi, Jinhui Chen, Xunquan Chen, Jie Lian, Pingyuan Lin, Zhihong Zhang, Edwin R. Hancock</i>	
CROSS-LINGUAL VOICE CONVERSION WITH A CYCLE CONSISTENCY LOSS ON LINGUISTIC REPRESENTATION	4765
<i>Yi Zhou, Xiaohai Tian, Zhizheng Wu, Haizhou Li</i>	
STARGANV2-VC: A DIVERSE, UNSUPERVISED, NON-PARALLEL FRAMEWORK FOR NATURAL-SOUNDING VOICE CONVERSION	4770
<i>Yinghao Aaron Li, Ali Zare, Nima Mesgarani</i>	
NON-PARALLEL ANY-TO-MANY VOICE CONVERSION BY REPLACING SPEAKER STATISTICS	4775
<i>Yufei Liu, Chengzhu Yu, Wang Shuai, Zhenchuan Yang, Yang Chao, Weibin Zhang</i>	
<u>INTERSPEECH 2021 : VOICE CONVERSION AND ADAPTATION II</u>	
AN EXEMPLAR SELECTION ALGORITHM FOR NATIVE-NONNATIVE VOICE CONVERSION	4780
<i>Christopher Liberatore, Ricardo Gutierrez-Osuna</i>	
S2VC: A FRAMEWORK FOR ANY-TO-ANY VOICE CONVERSION WITH SELF- SUPERVISED PRETRAINED REPRESENTATIONS	4785
<i>Jheng-Hao Lin, Yist Y. Lin, Chung-Ming Chien, Hung-Yi Lee</i>	
ADVERSARIALLY LEARNING DISENTANGLED SPEECH REPRESENTATIONS FOR ROBUST MULTI-FACTOR VOICE CONVERSION	4790
<i>Jie Wang, Jingbei Li, Xintao Zhao, Zhiyong Wu, Shiyin Kang, Helen Meng</i>	
LIMITED DATA EMOTIONAL VOICE CONVERSION LEVERAGING TEXT-TO-SPEECH: TWO-STAGE SEQUENCE-TO-SEQUENCE TRAINING	4795
<i>Kun Zhou, Berrak Sisman, Haizhou Li</i>	
MANY-TO-MANY VOICE CONVERSION BASED FEATURE DISENTANGLEMENT USING VARIATIONAL AUTOENCODER	4800
<i>Manh Luong, Viet Anh Tran</i>	

INVESTIGATING DEEP NEURAL STRUCTURES AND THEIR INTERPRETABILITY IN THE DOMAIN OF VOICE CONVERSION	4805
<i>Samuel J. Broughton, Md. Asif Jalal, Roger K. Moore</i>	
AN IMPROVED STARGAN FOR EMOTIONAL VOICE CONVERSION: ENHANCING VOICE QUALITY AND DATA AUGMENTATION	4810
<i>Xiangheng He, Junjie Chen, Georgios Rizos, Bjorn W. Schuller</i>	
ADVERSARIAL VOICE CONVERSION AGAINST NEURAL SPOOFING DETECTORS	4815
<i>Yi-Yang Ding, Li-Juan Liu, Yu Hu, Zhen-Hua Ling</i>	
ENRICHING SOURCE STYLE TRANSFER IN RECOGNITION-SYNTHESIS BASED NON-PARALLEL VOICE CONVERSION	4820
<i>Zhichao Wang, Xinyong Zhou, Fengyu Yang, Tao Li, Hongqiang Du, Lei Xie, Wendong Gan, Haitao Chen, Hai Li</i>	
TVQVC: TRANSFORMER BASED VECTOR QUANTIZED VARIATIONAL AUTOENCODER WITH CTC LOSS FOR VOICE CONVERSION	4825
<i>Ziyi Chen, Pengyuan Zhang</i>	
<u>INTERSPEECH 2021 : VOICE QUALITY CHARACTERIZATION FOR CLINICAL VOICE ASSESSMENT: VOICE PRODUCTION, ACOUSTICS, AND AUDITORY PERCEPTION</u>	
INVESTIGATING VOICE FUNCTION CHARACTERISTICS OF GREEK SPEAKERS WITH HEARING LOSS USING AUTOMATIC GLOTTAL SOURCE FEATURE EXTRACTION	4830
<i>Anna Sfakianaki, George P. Kafentzis</i>	
MODELING DYSPHONIA SEVERITY AS A FUNCTION OF ROUGHNESS AND BREATHINESS RATINGS IN THE GRBAS SCALE	4835
<i>Carlos A. Ferrer, Efren Aragon, Maria E. Hdez-Diaz, Marc S. De Bodt, Roman Cmejla, Marina Englert, Mara Behlau, Elmar Noth</i>	
OPTIMIZING AN AUTOMATIC CREAKY VOICE DETECTION METHOD FOR AUSTRALIAN ENGLISH SPEAKING FEMALES	4840
<i>Hannah White, Joshua Penney, Andy Gibson, Anita Szakay, Felicity Cox</i>	
A COMPARISON OF ACOUSTIC CORRELATES OF VOICE QUALITY ACROSS DIFFERENT RECORDING DEVICES: A CAUTIONARY TALE	4845
<i>Joshua Penney, Andy Gibson, Felicity Cox, Michael Proctor, Anita Szakay</i>	
AUTOMATED DETECTION OF VOICE DISORDER IN THE SAARBRUCKEN VOICE DATABASE: EFFECTS OF PATHOLOGY SUBSET AND AUDIO MATERIALS	4850
<i>Mark Huckvale, Catinca Buciuileac</i>	
ARTICULATORY COORDINATION FOR SPEECH MOTOR TRACKING IN HUNTINGTON DISEASE	4855
<i>Matthew Perez, Amrit Romana, Angela Roberts, Noelle Carlozzi, Jennifer Ann Miner, Praveen Dayalu, Emily Mower Provost</i>	
ACCELEROMETER-BASED MEASUREMENTS OF VOICE QUALITY IN CHILDREN DURING SEMI-OCCLUDED VOCAL TRACT EXERCISE WITH A NARROW STRAW IN AIR	4860
<i>Steven M. Lulich, Rita R. Patel</i>	

Author Index