

20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019)

Crossroads of Speech and Language

Graz, Austria
15-19 September 2019

Volume 1 of 7

ISBN: 978-1-5108-9683-3

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© (2019) by International Speech Communication Association
All rights reserved.

Printed with permission by Curran Associates, Inc. (2019)

For permission requests, please contact International Speech Communication Association
at the address below.

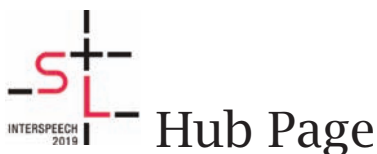
International Speech Communication Association
c/o Mme Emmanuelle FOXONET
4 Rue des Fauvettes - Lous Tourils
F-66390 Baixas, France

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



20th Annual Conference of
the International Speech Communication Association
September 15–19, 2019 ■ Graz, Austria

Proceedings of the 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019). ISSN 2308-457X.
© 2019 International Speech Communication Association. All rights reserved. For technical support please contact Causal Productions
(info@causalproductions.com).

INTERSPEECH 2019 Session List

Monday, 16 September 2019

[Mon-K-1](#) ISCA Medal 2019 Keynote Speech

[Mon-SS-1-6](#) Spoken Language Processing for Children's Speech

[Mon-SS-2-6](#) Dynamics of Emotional Speech Exchanges in Multimodal Communication

[Mon-O-1-1](#) End-to-End Speech Recognition

[Mon-O-1-2](#) Speech Enhancement: Multi-Channel

[Mon-O-1-3](#) Speech Production: Individual Differences and the Brain

[Mon-O-1-4](#) Speech Signal Characterization 1

[Mon-O-1-5](#) Neural Waveform Generation

[Mon-O-2-1](#) Attention Mechanism for Speaker State Recognition

[Mon-O-2-2](#) ASR Neural Network Training — 1

[Mon-O-2-3](#) Zero-Resource ASR

[Mon-O-2-4](#) Sociophonetics

[Mon-O-2-5](#) Resources - Annotation - Evaluation

[Mon-P-1-A](#) Speaker Recognition and Diarization

[Mon-P-1-B](#) ASR for Noisy and Far-Field Speech

[Mon-P-1-C](#) Social Signals Detection and Speaker Traits Analysis

[Mon-P-1-D](#) Applications of Language Technologies

[Mon-P-1-E](#) Speech and Audio Characterization and Segmentation

[Mon-P-2-A](#) Neural Techniques for Voice Conversion and Waveform Generation

[Mon-P-2-B](#) Model Adaptation for ASR

[Mon-P-2-C](#) Dialogue Speech Understanding

[Mon-P-2-D](#) Speech Production and Silent Interfaces

[Mon-P-2-E](#) Speech Signal Characterization 2

[Mon-S&T-1](#) Applications in Language Learning and Healthcare

INTERSPEECH 2019 Session List

Tuesday, 17 September 2019

[Tue-K-2](#) Keynote 2: Tanja Schultz

[Tue-SS-3-6](#) The Second DIHARD Speech Diarization Challenge (DIHARD II)

[Tue-SS-4-4](#) The 2019 Automatic Speaker Verification Spoofing and Countermeasures Challenge: ASVspoof Challenge — O

[Tue-SS-4-A](#) The 2019 Automatic Speaker Verification Spoofing and Countermeasures Challenge: ASVspoof Challenge — P

[Tue-SS-5-6](#) The Zero Resource Speech Challenge 2019: TTS Without T

[Tue-O-3-1](#) Speech Translation

[Tue-O-3-2](#) Speaker Recognition 1

[Tue-O-3-3](#) Dialogue Understanding

[Tue-O-3-4](#) Speech in the Brain

[Tue-O-3-5](#) Far-Field Speech Recognition

[Tue-O-4-1](#) Speaker and Language Recognition 1

[Tue-O-4-2](#) Speech Synthesis: Towards End-to-End

[Tue-O-4-3](#) Semantic Analysis and Classification

[Tue-O-4-5](#) Speech and Audio Source Separation and Scene Analysis 1

[Tue-O-5-1](#) Speech Intelligibility

[Tue-O-5-2](#) ASR Neural Network Architectures 1

[Tue-O-5-3](#) Speech and Language Analytics for Mental Health

[Tue-O-5-4](#) Dialogue Modelling

[Tue-O-5-5](#) Speaker Recognition Evaluation

[Tue-P-3-A](#) Speech Synthesis: Data and Evaluation

[Tue-P-3-B](#) Model Training for ASR

[Tue-P-3-C](#) Network Architectures for Emotion and Paralinguistics Recognition

[Tue-P-3-D](#) Acoustic Phonetics

[Tue-P-3-E](#) Speech Enhancement: Noise Attenuation

Tuesday sessions continue next page ...

INTERSPEECH 2019 Session List

Tuesday sessions continued ...

[Tue-P-4-B](#) Language Learning and Databases

[Tue-P-4-C](#) Emotion and Personality in Conversation

[Tue-P-4-D](#) Voice Quality, Speech Perception, and Prosody

[Tue-P-4-E](#) Speech Signal Characterization 3

[Tue-P-5-A](#) Speech Synthesis: Pronunciation, Multilingual, and Low Resource

[Tue-P-5-B](#) Cross-Lingual and Multilingual ASR

[Tue-P-5-C](#) Spoken Term Detection, Confidence Measure, and End-to-End Speech Recognition

[Tue-P-5-D](#) Speech Perception

[Tue-P-5-E](#) Topics in Speech and Audio Signal Processing

[Tue-S&T-2](#) Speech Processing and Analysis

INTERSPEECH 2019 Session List

Wednesday, 18 September 2019

Wed-K-3 Keynote 3: Manfred Kaltenbacher

Wed-SS-6-4 The Interspeech 2019 Computational Paralinguistics Challenge (ComParE)

Wed-SS-7-3 The VOICES from a Distance Challenge — O

Wed-SS-7-A The VOICES from a Distance Challenge — P

Wed-SS-8-6 Voice Quality Characterization for Clinical Voice Assessment: Voice Production, Acoustics, and Auditory Perception

Wed-O-6-1 Prosody

Wed-O-6-2 Speech and Audio Classification 1

Wed-O-6-3 Singing and Multimodal Synthesis

Wed-O-6-5 ASR Neural Network Training — 2

Wed-O-7-1 Bilingualism, L2, and Non-Nativeness

Wed-O-7-2 Spoken Term Detection

Wed-O-7-4 Speech and Audio Source Separation and Scene Analysis 2

Wed-O-7-5 Speech Enhancement: Single Channel 2

Wed-O-8-1 Multimodal ASR

Wed-O-8-2 ASR Neural Network Architectures 2

Wed-O-8-3 Training Strategy for Speech Emotion Recognition

Wed-O-8-4 Voice Conversion for Style, Accent, and Emotion

Wed-O-8-5 Speaker Recognition 2

Wed-P-6-A Speaker Recognition and Anti-Spoofing

Wed-P-6-B Rich Transcription and ASR Systems

Wed-P-6-C Speech and Language Analytics for Medical Applications

Wed-P-6-D Speech Perception in Adverse Listening Conditions

Wed-P-6-E Speech Enhancement: Single Channel 1

Wednesday sessions continue next page ...

INTERSPEECH 2019 Session List

Wednesday sessions continued ...

Wed-P-7-B Speech Recognition and Beyond

Wed-P-7-C Emotion Modeling and Analysis

Wed-P-7-D Articulatory Phonetics

Wed-P-7-E Speech and Audio Classification 2

Wed-P-8-A Speech Coding and Evaluation

Wed-P-8-B Feature Extraction for ASR

Wed-P-8-C Lexicon and Language Model for Speech Recognition

Wed-P-8-D First and Second Language Acquisition

Wed-P-8-E Speech and Audio Classification 3

Wed-S&T-3 Speech and Speaker Recognition

Wed-S&T-4 Speech Annotation and Labelling

Wed-S&T-5 Speech Synthesis

INTERSPEECH 2019 Session List

Thursday, 19 September 2019

- Thu-K-4 Keynote 4: Mirella Lapata
- Thu-SS-9-6 Privacy in Speech and Audio Interfaces
- Thu-SS-10-5 Speech Technologies for Code-Switching in Multilingual Communities
- Thu-O-9-1 Speech Synthesis: Articulatory and Physical Approaches
- Thu-O-9-2 Sequence-to-Sequence Speech Recognition
- Thu-O-9-3 Search Methods for Speech Recognition
- Thu-O-9-4 Audio Signal Characterization
- Thu-O-9-5 Speech and Voice Disorders 1
- Thu-O-10-1 Neural Networks for Language Modeling
- Thu-O-10-2 Representation Learning of Emotion and Paralinguistics
- Thu-O-10-3 World's Languages and Varieties
- Thu-O-10-4 Adaptation and Accommodation in Conversation
- Thu-P-9-A Speaker and Language Recognition 2
- Thu-P-9-B Medical Applications and Visual ASR
- Thu-P-9-C Turn Management in Dialogue
- Thu-P-9-D Corpus Annotation and Evaluation
- Thu-P-9-E Speech Enhancement: Multi-Channel and Intelligibility
- Thu-P-10-A Speaker Recognition 3
- Thu-P-10-B NN Architectures for ASR
- Thu-P-10-C Speech Synthesis: Text Processing, Prosody, and Emotion
- Thu-P-10-D Speech and Voice Disorders 2
- Thu-P-10-E Speech and Audio Source Separation and Scene Analysis 3
- Thu-S&T-6 Speech-to-Text and Speech Assessment

INTERSPEECH 2019 Table of Contents

Mon-K-1 : ISCA Medal 2019 Keynote Speech

Chair: To be confirmed

Time 09:30-10:30, Monday 16 Sept 2019, Main Hall

Mon-K-1
09:30-10:30
N/A

Statistical Approach to Speech Synthesis: Past, Present and Future
Keiichi Tokuda, Nagoya Institute of Technology, Japan

Mon-SS-1-6: Spoken Language Processing for Children's Speech

Chair: Keelan Evanini

Time 11:00-13:00, Monday 16 Sept 2019, Hall 3

Mon-SS-1-6-1 11:00-11:12 N/A	Introduction: SIG-CHILD Special Interest Group
PAGE 1 Mon-SS-1-6-2 11:12-11:30	Advances in Automatic Speech Recognition for Child Speech Using Factored Time Delay Neural Network <i>Fei Wu, Leibny Paola García-Perera, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA</i>
PAGE 6 Mon-SS-1-6-3 11:30-11:48	A Frequency Normalization Technique for Kindergarten Speech Recognition Inspired by the Role of f_0 in Vowel Perception <i>Gary Yeung, Abeer Alwan, University of California at Los Angeles, USA</i>
PAGE 11 Mon-SS-1-6-4 11:48-12:06	Improving ASR Systems for Children with Autism and Language Impairment Using Domain-Focused DNN Transfer Techniques <i>Robert Gale, Liu Chen, Jill Dolata, Jan van Santen, Meysam Asgari, Oregon Health & Science University, USA</i>
PAGE 16 Mon-SS-1-6-5 12:06-12:24	Ultrasound Tongue Imaging for Diarization and Alignment of Child Speech Therapy Sessions <i>Manuel Sam Ribeiro, Aciel Eshky, Korin Richmond, Steve Renals, University of Edinburgh, UK</i>
PAGE 21 Mon-SS-1-6-6 12:24-12:42	Automated Estimation of Oral Reading Fluency During Summer Camp e-Book Reading with MyTurnToRead <i>Anastassia Loukina, Beata Beigman Klebanov, Patrick Lange, Yao Qian, Binod Gyawali, Nitin Madnani, Abhinav Misra, Klaus Zechner, Zuowei Wang, John Sabatini, Educational Testing Service, USA</i>
PAGE 26 Mon-SS-1-6-7 12:42-13:00	Sustained Vowel Game: A Computer Therapy Game for Children with Dysphonia <i>Vanessa Lopes, João Magalhães, Sofia Cavaco, Universidade Nova de Lisboa, Portugal</i>

Mon-SS-2-6: Dynamics of Emotional Speech Exchanges in Multimodal Communication

Chairs: Anna Esposito, Gennaro Cordasco

Time 14:30-16:30, Monday 16 Sept 2019, Hall 3

PAGE 31 Mon-SS-2-6-1 14:30-14:42	The Dependability of Voice on Elders' Acceptance of Humanoid Agents <i>Anna Esposito¹, Terry Amorese¹, Marialucia Cuciniello¹, Maria Teresa Riviello¹, Antonietta M. Esposito², Alda Troncone¹, Gennaro Cordasco¹</i> <i>¹Università della Campania "Luigi Vanvitelli", Italy; ²INGV, Italy</i>
PAGE 36 Mon-SS-2-6-2 14:42-14:54	God as Interlocutor — Real or Imaginary? Prosodic Markers of Dialogue Speech and Expected Efficacy in Spoken Prayer <i>Oliver Niebuhr¹, Uffe Schjoedt²</i> <i>¹University of Southern Denmark, Denmark; ²Aarhus University, Denmark</i>
PAGE 41 Mon-SS-2-6-3 14:54-15:06	Expressiveness Influences Human Vocal Alignment Toward voice-AI <i>Michelle Cohn, Georgia Zellou, University of California at Davis, USA</i>
PAGE 46 Mon-SS-2-6-4 15:06-15:18	Detecting Topic-Oriented Speaker Stance in Conversational Speech <i>Catherine Lai¹, Beatrice Alex¹, Johanna D. Moore¹, Leimin Tian², Tatsuro Hori³, Gianpiero Francesca⁴</i> <i>¹University of Edinburgh, UK; ²Monash University, Australia; ³Toyota, Japan; ⁴Toyota, Belgium</i>
PAGE 51 Mon-SS-2-6-5 15:18-15:30	Fusion Techniques for Utterance-Level Emotion Recognition Combining Speech and Transcripts <i>Jilt Sebastian, Piero Pierucci, Telepathy Labs, Switzerland</i>
PAGE 56 Mon-SS-2-6-6 15:30-15:42	Explaining Sentiment Classification <i>Marvin Rajwadi¹, Cornelius Glackin², Julie Wall¹, Gérard Chollet², Nigel Cunnings²</i> <i>¹University of East London, UK; ²Intelligent Voice, UK</i>

Mon-SS-2-6 continued...

PAGE 61
Mon-SS-2-6-7
15:42-15:54

Predicting Group-Level Skin Attention to Short Movies from Audio-Based LSTM-Mixture of Experts Models

Ricardo Kleinlein¹, Cristina Luna Jiménez¹, Juan Manuel Montero¹, Zoraida Callejas², Fernando Fernández-Martínez¹

¹Universidad Politécnica de Madrid, Spain; ²Universidad de Granada, Spain

Mon-SS-2-6-8
15:54-16:30
N/A

Discussion

Mon-O-1-1 : End-to-End Speech Recognition

Chairs: Brian Mak, Florian Metze

Time 11:00-13:00, Monday 16 Sept 2019, Main Hall

Mon-O-1-1-1
11:00-11:40
N/A

Survey Talk: Modeling in Automatic Speech Recognition: Beyond Hidden Markov Models

Ralf Schlüter, RWTH Aachen University, Germany

PAGE 66
Mon-O-1-1-2
11:40-12:00

Very Deep Self-Attention Networks for End-to-End Speech Recognition

Ngoc-Quan Pham, Thai-Son Nguyen, Jan Niehues, Markus Müller, Alex Waibel, KIT, Germany

PAGE 71
Mon-O-1-1-3
12:00-12:20

Jasper: An End-to-End Convolutional Neural Acoustic Model

Jason Li¹, Vitaly Lavrukhin¹, Boris Ginsburg¹, Ryan Leary¹, Oleksii Kuchaiev¹, Jonathan M. Cohen¹, Huyen Nguyen¹, Ravi Teja Gadde²

¹NVIDIA, USA; ²New York University, USA

PAGE 76
Mon-O-1-1-4
12:20-12:40

Unidirectional Neural Network Architectures for End-to-End Automatic Speech Recognition

Niko Moritz, Takaaki Hori, Jonathan Le Roux, MERL, USA

PAGE 81
Mon-O-1-1-5
12:40-13:00

Analyzing Phonetic and Graphemic Representations in End-to-End Automatic Speech Recognition

Yonatan Belinkov¹, Ahmed Ali², James Glass¹

¹MIT, USA; ²HBKU, Qatar

Mon-O-1-2 : Speech Enhancement: Multi-Channel

Chairs: *Hong-Goo Kang, Ina Kodrasi*

Time 11:00-13:00, Monday 16 Sept 2019, Hall 1

PAGE 86 Mon-O-1-2-1 11:00-11:20	Multi-Channel Speech Enhancement Using Time-Domain Convolutional Denoising Autoencoder <i>Naohiro Tawara, Tetsunori Kobayashi, Tetsuji Ogawa, Waseda University, Japan</i>
PAGE 91 Mon-O-1-2-2 11:20-11:40	On Nonlinear Spatial Filtering in Multichannel Speech Enhancement <i>Kristina Tesch, Robert Rehr, Timo Gerkmann, Universität Hamburg, Germany</i>
PAGE 96 Mon-O-1-2-3 11:40-12:00	Multi-Channel Block-Online Source Extraction Based on Utterance Adaptation <i>Juan M. Martín-Doñas¹, Jens Heitkaemper², Reinhold Haeb-Umbach², Angel M. Gomez¹, Antonio M. Peinado¹</i> ¹ Universidad de Granada, Spain; ² Universität Paderborn, Germany
PAGE 101 Mon-O-1-2-4 12:00-12:20	Exploiting Multi-Channel Speech Presence Probability in Parametric Multi-Channel Wiener Filter <i>Saeed Bagheri, Daniele Giacobello, Sonos, USA</i>
PAGE 106 Mon-O-1-2-5 12:20-12:40	Variational Bayesian Multi-Channel Speech Dereverberation Under Noisy Environments with Probabilistic Convolutional Transfer Function <i>Masahito Togami, Tatsuya Komatsu, LINE, Japan</i>
PAGE 111 Mon-O-1-2-6 12:40-13:00	Simultaneous Denoising and Dereverberation for Low-Latency Applications Using Frame-by-Frame Online Unified Convolutional Beamformer <i>Tomohiro Nakatani, Keisuke Kinoshita, NTT, Japan</i>

Mon-O-1-3 : Speech Production: Individual Differences and the Brain

Chairs: *Samuel Silva, Tanja Schultz*

Time 11:00-13:00, Monday 16 Sept 2019, Hall 2

PAGE 116 Mon-O-1-3-1 11:00-11:20	Individual Variation in Cognitive Processing Style Predicts Differences in Phonetic Imitation of Device and Human Voices <i>Cathryn Snyder, Michelle Cohn, Georgia Zellou, University of California at Davis, USA</i>
PAGE 121 Mon-O-1-3-2 11:20-11:40	An Investigation on Speaker Specific Articulatory Synthesis with Speaker Independent Articulatory Inversion <i>Aravind Illa, Prasanta Kumar Ghosh, Indian Institute of Science, India</i>
PAGE 126 Mon-O-1-3-3 11:40-12:00	Individual Difference of Relative Tongue Size and its Acoustic Effects <i>Xiaohan Zhang, Chongke Bi, Kiyoshi Honda, Wenhuan Lu, Jianguo Wei, Tianjin University, China</i>
PAGE 131 Mon-O-1-3-4 12:00-12:20	Individual Differences of Airflow and Sound Generation in the Vocal Tract of Sibilant /s/ <i>Tsukasa Yoshinaga¹, Kazunori Nozaki², Shigeo Wada³</i> ¹ Toyohashi University of Technology, Japan; ² Osaka University Dental Hospital, Japan; ³ Osaka University, Japan
PAGE 136 Mon-O-1-3-5 12:20-12:40	Hush-Hush Speak: Speech Reconstruction Using Silent Videos <i>Shashwat Uttam¹, Yaman Kumar², Dhruva Sahrawat³, Mansi Aggarwal⁴, Rajiv Ratn Shah³, Debanjan Mahata⁵, Amanda Stent⁵</i> ¹ NSUT, India; ² Adobe, India; ³ IIT Delhi, India; ⁴ DTU, India; ⁵ Bloomberg, USA
PAGE 141 Mon-O-1-3-6 12:40-13:00	SPEAK YOUR MIND! Towards Imagined Speech Recognition with Hierarchical Deep Learning <i>Pramit Saha, Muhammad Abdul-Mageed, Sidney Fels, University of British Columbia, Canada</i>

Mon-O-1-4: Speech Signal Characterization 1

Chairs: Koichi Shinoda, Shrikanth Narayanan

Time 11:00-13:00, Monday 16 Sept 2019, Hall 11

PAGE 146 Mon-O-1-4-1 11:00-11:20	An Unsupervised Autoregressive Model for Speech Representation Learning <i>Yu-An Chung, Wei-Ning Hsu, Hao Tang, James Glass, MIT, USA</i>
PAGE 151 Mon-O-1-4-2 11:20-11:40	Harmonic-Aligned Frame Mask Based on Non-Stationary Gabor Transform with Application to Content-Dependent Speaker Comparison <i>Feng Huang, Peter Balazs, Austrian Academy of Sciences, Austria</i>
PAGE 156 Mon-O-1-4-3 11:40-12:00	Glottal Closure Instants Detection from Speech Signal by Deep Features Extracted from Raw Speech and Linear Prediction Residual <i>Gurunath Reddy M., K. Sreenivasa Rao, Partha Pratim Das, IIT Kharagpur, India</i>
PAGE 161 Mon-O-1-4-4 12:00-12:20	Learning Problem-Agnostic Speech Representations from Multiple Self-Supervised Tasks <i>Santiago Pascual¹, Mirco Ravanelli², Joan Serra³, Antonio Bonafonte¹, Yoshua Bengio²</i> <i>¹Universitat Politècnica de Catalunya, Spain; ²Université de Montréal, Canada; ³Telefónica I+D, Spain</i>
PAGE 166 Mon-O-1-4-5 12:20-12:40	Excitation Source and Vocal Tract System Based Acoustic Features for Detection of Nasals in Continuous Speech <i>Bhanu Teja Nellore¹, Sri Harsha Dumpala², Karan Nathwani³, Suryakanth V. Gangashetty¹</i> <i>¹IIT Hyderabad, India; ²TCS Innovation Labs Mumbai, India; ³IIT Jammu, India</i>
PAGE 171 Mon-O-1-4-6 12:40-13:00	Data Augmentation Using GANs for Speech Emotion Recognition <i>Aggelina Chatziagapi, Georgios Paraskevopoulos, Dimitris Sgouropoulos, Georgios Pantazopoulos, Malvina Nikandrou, Theodoros Giannakopoulos, Athanasios Katsamanis, Alexandros Potamianos, Shrikanth Narayanan, Behavioral Signal Technologies, USA</i>

Mon-O-1-5: Neural Waveform Generation

Chairs: Hema Murthy, Alan W. Black

Time 11:00-13:00, Monday 16 Sept 2019, Hall 12

PAGE 176 Mon-O-1-5-1 11:00-11:20	High Quality, Lightweight and Adaptable TTS Using LPCNet <i>Zvi Kons, Slava Shechtman, Alex Sorin, Carmel Rabinovitz, Ron Hoory, IBM, Israel</i>
PAGE 181 Mon-O-1-5-2 11:20-11:40	Towards Achieving Robust Universal Neural Vocoding <i>Jaime Lorenzo-Trueba¹, Thomas Drugman², Javier Latorre³, Thomas Merritt¹, Bartosz Putrycz¹, Roberto Barra-Chicote¹, Alexis Moinet¹, Vatsal Aggarwal¹</i> <i>¹Amazon, UK; ²Amazon, Germany; ³Apple, UK</i>
PAGE 186 Mon-O-1-5-3 11:40-12:00	Expediting TTS Synthesis with Adversarial Vocoding <i>Paarth Neekhara, Chris Donahue, Miller Puckette, Shlomo Dubnov, Julian McAuley, University of California at San Diego, USA</i>
PAGE 191 Mon-O-1-5-4 12:00-12:20	Analysis by Adversarial Synthesis — A Novel Approach for Speech Vocoding <i>Ahmed Mustafa¹, Arijit Biswas², Christian Bergler³, Julia Schottenhamml³, Andreas Maier³</i> <i>¹Fraunhofer IIS, Germany; ²Dolby, Germany; ³FAU Erlangen-Nürnberg, Germany</i>
PAGE 196 Mon-O-1-5-5 12:20-12:40	Quasi-Periodic WaveNet Vocoder: A Pitch Dependent Dilated Convolution Model for Parametric Speech Generation <i>Yi-Chiao Wu, Tomoki Hayashi, Patrick Lumban Tobing, Kazuhiro Kobayashi, Tomoki Toda, Nagoya University, Japan</i>
PAGE 201 Mon-O-1-5-6 12:40-13:00	A Speaker-Dependent WaveNet for Voice Conversion with Non-Parallel Data <i>Xiaohai Tian¹, Eng Siong Chng², Haizhou Li¹</i> <i>¹NUS, Singapore; ²NTU, Singapore</i>

Mon-O-2-1 : Attention Mechanism for Speaker State Recognition

Chairs: Carlos Busso, Chi-Chun Lee

Time 14:30-16:30, Monday 16 Sept 2019, Main Hall

- Mon-O-2-1-1
14:30-15:10
N/A
- Survey Talk: When Attention Meets Speech Applications: Speech & Speaker Recognition Perspective**
Kyu J. Han, Ramon Prieto, Tao Ma, ASAPP, USA
- PAGE 206
Mon-O-2-1-2
15:10-15:30
- Attention-Enhanced Connectionist Temporal Classification for Discrete Speech Emotion Recognition**
Ziping Zhao¹, Zhongtian Bao¹, Zixing Zhang², Nicholas Cummins³, Haishuai Wang¹, Björn W. Schuller³
¹Tianjin Normal University, China; ²Imperial College London, UK; ³Universität Augsburg, Germany
- PAGE 211
Mon-O-2-1-3
15:30-15:50
- Attentive to Individual: A Multimodal Emotion Recognition Network with Personalized Attention Profile**
Jeng-Lin Li, Chi-Chun Lee, National Tsing Hua University, Taiwan
- PAGE 216
Mon-O-2-1-4
15:50-16:10
- A Saliency-Based Attention LSTM Model for Cognitive Load Classification from Speech**
Ascensión Gallardo-Antolín¹, Juan Manuel Montero²
¹Universidad Carlos III de Madrid, Spain; ²Universidad Politécnica de Madrid, Spain
- PAGE 221
Mon-O-2-1-5
16:10-16:30
- A Hierarchical Attention Network-Based Approach for Depression Detection from Transcribed Clinical Interviews**
Adria Mallol-Ragolta, Ziping Zhao, Lukas Stappen, Nicholas Cummins, Björn W. Schuller, Universität Augsburg, Germany

Mon-O-2-2 : ASR Neural Network Training — 1

Chair: Mark Gales

Time 14:30-16:30, Monday 16 Sept 2019, Hall 1

- PAGE 226
Mon-O-2-2-1
14:30-14:50
- Untranscribed Web Audio for Low Resource Speech Recognition**
Andrea Carmantini, Peter Bell, Steve Renals, University of Edinburgh, UK
- PAGE 231
Mon-O-2-2-2
14:50-15:10
- RWTH ASR Systems for LibriSpeech: Hybrid vs Attention**
Christoph Lüscher, Eugen Beck, Kazuki Irie, Markus Kitza, Wilfried Michel, Albert Zeyer, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany
- PAGE 236
Mon-O-2-2-3
15:10-15:30
- Auxiliary Interference Speaker Loss for Target-Speaker Speech Recognition**
Naoyuki Kanda¹, Shota Horiguchi¹, Ryoichi Takashima¹, Yusuke Fujita¹, Kenji Nagamatsu¹, Shinji Watanabe²
¹Hitachi, Japan; ²Johns Hopkins University, USA
- PAGE 241
Mon-O-2-2-4
15:30-15:50
- Speaker Adaptation for Attention-Based End-to-End Speech Recognition**
Zhong Meng, Yashesh Gaur, Jinyu Li, Yifan Gong, Microsoft, USA
- PAGE 246
Mon-O-2-2-5
15:50-16:10
- Large Margin Training for Attention Based End-to-End Speech Recognition**
Peidong Wang¹, Jia Cui², Chao Weng², Dong Yu²
¹Ohio State University, USA; ²Tencent, USA
- PAGE 251
Mon-O-2-2-6
16:10-16:30
- Large-Scale Mixed-Bandwidth Deep Neural Network Acoustic Modeling for Automatic Speech Recognition**
Khoi-Nguyen C. Mac¹, Xiaodong Cui², Wei Zhang², Michael Picheny²
¹University of Illinois at Urbana-Champaign, USA; ²IBM, USA

Mon-O-2-3: Zero-Resource ASR

Chairs: Jan Černocký, Marco Siniscalchi

Time 14:30-16:30, Monday 16 Sept 2019, Hall 2

- PAGE 256
Mon-O-2-3-1
14:30-14:50
- SparseSpeech: Unsupervised Acoustic Unit Discovery with Memory-Augmented Sequence Autoencoders**
Benjamin Milde, Chris Biemann, Universität Hamburg, Germany
- PAGE 261
Mon-O-2-3-2
14:50-15:10
- Bayesian Subspace Hidden Markov Model for Acoustic Unit Discovery**
Lucas Ondel, Hari Krishna Vydana, Lukáš Burget, Jan Černocký, Brno University of Technology, Czech Republic
- PAGE 266
Mon-O-2-3-3
15:10-15:30
- Speaker Adversarial Training of DPGMM-Based Feature Extractor for Zero-Resource Languages**
Yosuke Higuchi, Naohiro Tawara, Tetsunori Kobayashi, Tetsuji Ogawa, Waseda University, Japan
- PAGE 271
Mon-O-2-3-4
15:30-15:50
- Building Large-Vocabulary ASR Systems for Languages Without Any Audio Training Data**
Manasa Prasad¹, Daan van Esch¹, Sandy Ritchie², Jonas Fromseier Mortensen²
¹Google, USA; ²Google, UK
- PAGE 276
Mon-O-2-3-5
15:50-16:10
- Towards Bilingual Lexicon Discovery From Visually Grounded Speech Audio**
Emmanuel Azuh, David Harwath, James Glass, MIT, USA
- PAGE 281
Mon-O-2-3-6
16:10-16:30
- Improving Unsupervised Subword Modeling via Disentangled Speech Representation Learning and Transformation**
Siyuan Feng, Tan Lee, CUHK, China

Mon-O-2-4: Sociophonetics

Chairs: Barbara Schuppler, Marcin Włodarczyk

Time 14:30-16:30, Monday 16 Sept 2019, Hall 11

- PAGE 286
Mon-O-2-4-1
14:30-14:50
- Listeners' Ability to Identify the Gender of Preadolescent Children in Different Linguistic Contexts**
Shawn Nissen, Sharalee Blunck, Anita Dromey, Christopher Dromey, Brigham Young University, USA
- PAGE 291
Mon-O-2-4-2
14:50-15:10
- Sibilant Variation in New Englishes: A Comparative Sociophonetic Study of Trinidadian and American English /s(tr)-Retraction**
Wiebke Ahlers¹, Philipp Meer²
¹Universität Osnabrück, Germany; ²Universität Münster, Germany
- PAGE 296
Mon-O-2-4-3
15:10-15:30
- Tracking the New Zealand English NEAR/SQUARE Merger Using Functional Principal Components Analysis**
Michele Gubian¹, Jonathan Harrington¹, Mary Stevens¹, Florian Schiel¹, Paul Warren²
¹LMU München, Germany; ²Victoria University of Wellington, New Zealand
- PAGE 301
Mon-O-2-4-4
15:30-15:50
- Phonetic Accommodation in a Wizard-of-Oz Experiment: Intonation and Segments**
Iona Gessinger¹, Bernd Möbius¹, Bistra Andreeva¹, Eran Raveh¹, Ingmar Steiner²
¹Universität des Saarlandes, Germany; ²audEERING, Germany
- PAGE 306
Mon-O-2-4-5
15:50-16:10
- PASCAL and DPA: A Pilot Study on Using Prosodic Competence Scores to Predict Communicative Skills for Team Working and Public Speaking**
Oliver Niebuhr¹, Jan Michalsky²
¹University of Southern Denmark, Denmark; ²FAU Erlangen-Nürnberg, Germany
- PAGE 311
Mon-O-2-4-6
16:10-16:30
- Towards the Prosody of Persuasion in Competitive Negotiation. The Relationship Between f0 and Negotiation Success in Same Sex Sales Tasks**
Jan Michalsky¹, Heike Schoormann², Thomas Schultze³
¹FAU Erlangen-Nürnberg, Germany; ²Carl von Ossietzky Universität Oldenburg, Germany; ³GAU Göttingen, Germany

Mon-O-2-5: Resources – Annotation – Evaluation

Chairs: Martine Adda-Decker, Michael Picheny

Time 14:30-16:30, Monday 16 Sept 2019, Hall 12

- PAGE 316
Mon-O-2-5-1
14:30-14:50
- VESUS: A Crowd-Annotated Database to Study Emotion Production and Perception in Spoken English**
Jacob Sager, Ravi Shankar, Jacob Reinhold, Archana Venkataraman, Johns Hopkins University, USA
- PAGE 321
Mon-O-2-5-2
14:50-15:10
- Building the Singapore English National Speech Corpus**
Jia Xin Koh¹, Aqilah Mislán², Kevin Khoo², Brian Ang², Wilson Ang², Charmaine Ng¹, Ying-Ying Tan¹
¹NTU, Singapore; ²IMDA, Singapore
- PAGE 326
Mon-O-2-5-3
15:10-15:30
- Challenging the Boundaries of Speech Recognition: The MALACH Corpus**
Michael Picheny, Zoltán Tüske, Brian Kingsbury, Kartik Audhkhasi, Xiaodong Cui, George Saon, IBM, USA
- PAGE 331
Mon-O-2-5-4
15:30-15:50
- NITK Kids' Speech Corpus**
Pravin Bhaskar Ramteke¹, Sujata Supanekar¹, Pradyoth Hegde¹, Hanna Nelson², Venkataraja Aithal², Shashidhar G. Koolagudi¹
¹NITK Surathkal, India; ²SOAHS Manipal, India
- PAGE 336
Mon-O-2-5-5
15:50-16:10
- Towards Variability Resistant Dialectal Speech Evaluation**
Ahmed Ali¹, Salam Khalifa², Nizar Habash²
¹HBKU, Qatar; ²New York University Abu Dhabi, UAE
- PAGE 341
Mon-O-2-5-6
16:10-16:30
- How to Annotate 100 Hours in 45 Minutes**
Per Fallgren, Zofia Malisz, Jens Edlund, KTH, Sweden

Mon-P-1-A: Speaker Recognition and Diarization

Chair: Suwon Shon

Time 11:00-13:00, Monday 16 Sept 2019, Gallery A

- PAGE 346
Mon-P-1-A-1
11:00-13:00
- Bayesian HMM Based x-Vector Clustering for Speaker Diarization**
Mireia Diez, Lukáš Burget, Shuai Wang, Johan Rohdin, Jan Černocký, Brno University of Technology, Czech Republic
- PAGE 351
Mon-P-1-A-2
11:00-13:00
- Unleashing the Unused Potential of i-Vectors Enabled by GPU Acceleration**
Ville Vestman¹, Kong Aik Lee¹, Tomi H. Kinnunen², Takafumi Koshinaka¹
¹NEC, Japan; ²University of Eastern Finland, Finland
- PAGE 356
Mon-P-1-A-3
11:00-13:00
- MCE 2018: The 1st Multi-Target Speaker Detection and Identification Challenge Evaluation**
Suwon Shon¹, Najim Dehak², Douglas Reynolds³, James Glass¹
¹MIT, USA; ²Johns Hopkins University, USA; ³MIT Lincoln Laboratory, USA
- PAGE 361
Mon-P-1-A-4
11:00-13:00
- Improving Aggregation and Loss Function for Better Embedding Learning in End-to-End Speaker Verification System**
Zhifu Gao¹, Yan Song¹, Ian McLoughlin², Pengcheng Li¹, Yiheng Jiang¹, Li-Rong Dai¹
¹USTC, China; ²University of Kent, UK
- PAGE 366
Mon-P-1-A-5
11:00-13:00
- LSTM Based Similarity Measurement with Spectral Clustering for Speaker Diarization**
Qingjian Lin¹, Ruiqing Yin², Ming Li¹, Hervé Bredin², Claude Barras²
¹Duke Kunshan University, China; ²LIMSI (UPR 3251), France
- PAGE 371
Mon-P-1-A-6
11:00-13:00
- Who Said That?: Audio-Visual Speaker Diarisation of Real-World Meetings**
Joon Son Chung, Bong-jin Lee, Icksang Han, Naver, Korea

INTERSPEECH 2019 Table of Contents

Mon-P-1-A continued...

PAGE 376
Mon-P-1-A-7
11:00-13:00

Multi-PLDA Diarization on Children's Speech

Jiamin Xie, Leibny Paola García-Perera, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

PAGE 381
Mon-P-1-A-8
11:00-13:00

Speaker Diarization Using Leave-One-Out Gaussian PLDA Clustering of DNN Embeddings

Alan McCree, Gregory Sell, Daniel Garcia-Romero, Johns Hopkins University, USA

PAGE 386
Mon-P-1-A-9
11:00-13:00

Speaker-Corrupted Embeddings for Online Speaker Diarization

Omid Ghahabi, Volker Fischer, EML European Media Laboratory, Germany

PAGE 391
Mon-P-1-A-10
11:00-13:00

Speaker Diarization with Lexical Information

Tae Jin Park¹, Kyu J. Han², Jing Huang², Xiaodong He³, Bowen Zhou³, Panayiotis Georgiou¹, Shrikanth Narayanan¹

¹University of Southern California, USA; ²JD.com, USA; ³JD.com, China

PAGE 396
Mon-P-1-A-11
11:00-13:00

Joint Speech Recognition and Speaker Diarization via Sequence Transduction

Laurent El Shafey, Hagen Soltau, Izhak Shafran, Google, USA

PAGE 401
Mon-P-1-A-12
11:00-13:00

Normal Variance-Mean Mixtures for Unsupervised Score Calibration

Sandro Cumani, Politecnico di Torino, Italy

PAGE 406
Mon-P-1-A-13
11:00-13:00

Speaker Augmentation and Bandwidth Extension for Deep Speaker Embedding

Hitoshi Yamamoto, Kong Aik Lee, Koji Okabe, Takafumi Koshinaka, NEC, Japan

PAGE 411
Mon-P-1-A-14
11:00-13:00

Large-Scale Speaker Diarization of Radio Broadcast Archives

Emre Yilmaz¹, Adem Derinel¹, Kun Zhou¹, Henk van den Heuvel², Niko Brummer³, Haizhou Li¹, David A. van Leeuwen²

¹NUS, Singapore; ²Radboud Universiteit Nijmegen, The Netherlands; ³Cyberupt, South Africa

INTERSPEECH 2019 Table of Contents

Mon-P-1-A continued...

PAGE 416
Mon-P-1-A-15
11:00-13:00

Toeplitz Inverse Covariance Based Robust Speaker Clustering for Naturalistic Audio Streams

Harishchandra Dubey, Abhijeet Sangwan, John H.L. Hansen, University of Texas at Dallas, USA

Mon-P-1-B: ASR for Noisy and Far-Field Speech

Chair: Ozlem Kalinli

Time 11:00-13:00, Monday 16 Sept 2019, Gallery B

- PAGE 421
Mon-P-1-B-1
11:00-13:00
- Examining the Combination of Multi-Band Processing and Channel Dropout for Robust Speech Recognition**
György Kovács¹, László Tóth², Dirk Van Compernelle³, Marcus Liwicki¹
¹Luleå University of Technology, Sweden; ²University of Szeged, Hungary; ³Katholieke Universiteit Leuven, Belgium
- PAGE 426
Mon-P-1-B-2
11:00-13:00
- Label Driven Time-Frequency Masking for Robust Continuous Speech Recognition**
Meet Soni, Ashish Panda, TCS Innovation Labs Mumbai, India
- PAGE 431
Mon-P-1-B-3
11:00-13:00
- Speaker-Invariant Feature-Mapping for Distant Speech Recognition via Adversarial Teacher-Student Learning**
Long Wu, Hangting Chen, Li Wang, Pengyuan Zhang, Yonghong Yan, Chinese Academy of Sciences, China
- PAGE 436
Mon-P-1-B-4
11:00-13:00
- Full-Sentence Correlation: A Method to Handle Unpredictable Noise for Robust Speech Recognition**
Ji Ming, Danny Crookes, Queen's University Belfast, UK
- PAGE 441
Mon-P-1-B-5
11:00-13:00
- Generative Noise Modeling and Channel Simulation for Robust Speech Recognition in Unseen Conditions**
Meet Soni, Sonal Joshi, Ashish Panda, TCS Innovation Labs Mumbai, India
- PAGE 446
Mon-P-1-B-6
11:00-13:00
- Far-Field Speech Enhancement Using Heteroscedastic Autoencoder for Improved Speech Recognition**
Shashi Kumar, Shakti P. Rath, Samsung, India

Mon-P-1-B continued...

- PAGE 451
Mon-P-1-B-7
11:00-13:00
- End-to-End SpeakerBeam for Single Channel Target Speech Recognition**
Marc Delcroix¹, Shinji Watanabe², Tsubasa Ochiai¹, Keisuke Kinoshita¹, Shigeki Karita¹, Atsunori Ogawa¹, Tomohiro Nakatani¹
¹NTT, Japan; ²Johns Hopkins University, USA
- PAGE 456
Mon-P-1-B-8
11:00-13:00
- NIESR: Nuisance Invariant End-to-End Speech Recognition**
I-Hung Hsu, Ayush Jaiswal, Premkumar Natarajan, University of Southern California, USA
- PAGE 461
Mon-P-1-B-9
11:00-13:00
- Knowledge Distillation for Throat Microphone Speech Recognition**
Takahito Suzuki¹, Jun Ogata², Takashi Tsunakawa¹, Masafumi Nishida¹, Masafumi Nishimura¹
¹Shizuoka University, Japan; ²AIST, Japan
- PAGE 466
Mon-P-1-B-10
11:00-13:00
- Improved Speaker-Dependent Separation for CHiME-5 Challenge**
Jian Wu¹, Yong Xu², Shi-Xiong Zhang², Lianwu Chen³, Meng Yu², Lei Xie¹, Dong Yu²
¹Northwestern Polytechnical University, China; ²Tencent, USA; ³Tencent, China
- PAGE 471
Mon-P-1-B-11
11:00-13:00
- Bridging the Gap Between Monaural Speech Enhancement and Recognition with Distortion-Independent Acoustic Modeling**
Peidong Wang, Ke Tan, DeLiang Wang, Ohio State University, USA
- PAGE 476
Mon-P-1-B-12
11:00-13:00
- Enhanced Spectral Features for Distortion-Independent Acoustic Modeling**
Peidong Wang, DeLiang Wang, Ohio State University, USA
- PAGE 481
Mon-P-1-B-13
11:00-13:00
- Universal Adversarial Perturbations for Speech Recognition Systems**
Paarth Neekhara, Shehzeen Hussain, Prakhar Pandey, Shlomo Dubnov, Julian McAuley, Farinaz Koushanfar, University of California at San Diego, USA

Mon-P-1-B continued...

PAGE 486
Mon-P-1-B-14
11:00-13:00

One-Pass Single-Channel Noisy Speech Recognition Using a Combination of Noisy and Enhanced Features

Masakiyo Fujimoto, Hisashi Kawai, NICT, Japan

PAGE 491
Mon-P-1-B-15
11:00-13:00

Jointly Adversarial Enhancement Training for Robust End-to-End Speech Recognition

Bin Liu¹, Shuai Nie¹, Shan Liang¹, Wenju Liu¹, Meng Yu², Lianwu Chen³, Shouye Peng⁴, Changliang Li⁵

¹Chinese Academy of Sciences, China; ²Tencent, USA; ³Tencent, China; ⁴Xueersi Online School, China; ⁵Kingsoft, China

Mon-P-1-C : Social Signals Detection and Speaker Traits Analysis

Chair: Elizabeth Shriberg

Time 11:00-13:00, Monday 16 Sept 2019, Gallery C

PAGE 496
Mon-P-1-C-1
11:00-13:00

Predicting Humor by Learning from Time-Aligned Comments

Zixiaofan Yang, Bingyan Hu, Julia Hirschberg, Columbia University, USA

PAGE 501
Mon-P-1-C-2
11:00-13:00

Predicting the Leading Political Ideology of YouTube Channels Using Acoustic, Textual, and Metadata Information

Yoan Dinkov¹, Ahmed Ali², Ivan Koychev¹, Preslav Nakov²

¹Sofia University, Bulgaria; ²HBKU, Qatar

PAGE 506
Mon-P-1-C-3
11:00-13:00

Mitigating Gender and L1 Differences to Improve State and Trait Recognition

Guozhen An¹, Rivka Levitan²

¹CUNY Queensborough Community College, USA; ²CUNY Graduate Center, USA

PAGE 510
Mon-P-1-C-4
11:00-13:00

Deep Learning Based Mandarin Accent Identification for Accent Robust ASR

Felix Weninger¹, Yang Sun², Junho Park¹, Daniel Willett², Puming Zhan¹

¹Nuance Communications, USA; ²Nuance Communications, Germany

PAGE 515
Mon-P-1-C-5
11:00-13:00

Calibrating DNN Posterior Probability Estimates of HMM/DNN Models to Improve Social Signal Detection from Audio Data

Gábor Gosztolya¹, László Tóth²

¹MTA-SZTE RGAI, Hungary; ²University of Szeged, Hungary

PAGE 520
Mon-P-1-C-6
11:00-13:00

Conversational and Social Laughter Synthesis with WaveNet

Hiroki Mori¹, Tomohiro Nagata¹, Yoshiko Arimoto²

¹Utsunomiya University, Japan; ²Chiba Institute of Technology, Japan

PAGE 524
Mon-P-1-C-7
11:00-13:00

Laughter Dynamics in Dyadic Conversations

Bogdan Ludusan, Petra Wagner, Universität Bielefeld, Germany

Mon-P-1-C continued...

PAGE 529
Mon-P-1-C-8
11:00-13:00

Towards an Annotation Scheme for Complex Laughter in Speech Corpora

Khiet P. Truong¹, Jürgen Trouvain², Michel-Pierre Jansen¹

¹Universiteit Twente, The Netherlands; ²Universität des Saarlandes, Germany

PAGE 534
Mon-P-1-C-9
11:00-13:00

Using Speech to Predict Sequentially Measured Cortisol Levels During a Trier Social Stress Test

Alice Baird¹, Shahin Amiriparian¹, Nicholas Cummins¹, Sarah Sturmhuber², Johanna Janson², Eva-Maria Messner³, Harald Baumeister³, Nicolas Rohleder², Björn W. Schuller¹

¹Universität Augsburg, Germany; ²FAU Erlangen-Nürnberg, Germany; ³Universität Ulm, Germany

PAGE 539
Mon-P-1-C-10
11:00-13:00

Sincerity in Acted Speech: Presenting the Sincere Apology Corpus and Results

Alice Baird¹, Eduardo Coutinho², Julia Hirschberg³, Björn W. Schuller¹

¹Universität Augsburg, Germany; ²University of Liverpool, UK; ³Columbia University, USA

PAGE 544
Mon-P-1-C-11
11:00-13:00

Do not Hesitate! — Unless You Do it Shortly or Nasally: How the Phonetics of Filled Pauses Determine Their Subjective Frequency and Perceived Speaker Performance

Oliver Niebuhr, Kerstin Fischer, University of Southern Denmark, Denmark

PAGE 549
Mon-P-1-C-12
11:00-13:00

Phonet: A Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech

J.C. Vásquez-Correa, Philipp Klumpp, Juan Rafael Orozco-Arroyave, Elmar Nöth, FAU Erlangen-Nürnberg, Germany

Mon-P-1-D: Applications of Language Technologies

Chair: Ngoc Thang Vu

Time 11:00-13:00, Monday 16 Sept 2019, Hall 10/D

PAGE 554
Mon-P-1-D-1
11:00-13:00

Code-Switching Sentence Generation by Generative Adversarial Networks and its Application to Data Augmentation

Ching-Ting Chang, Shun-Po Chuang, Hung-Yi Lee, National Taiwan University, Taiwan

PAGE 559
Mon-P-1-D-2
11:00-13:00

Comparative Analysis of Think-Aloud Methods for Everyday Activities in the Context of Cognitive Robotics

Moritz Meier, Celeste Mason, Felix Putze, Tanja Schultz, Universität Bremen, Germany

PAGE 564
Mon-P-1-D-3
11:00-13:00

RadioTalk: A Large-Scale Corpus of Talk Radio Transcripts

Doug Beeferman, William Brannon, Deb Roy, MIT, USA

PAGE 569
Mon-P-1-D-4
11:00-13:00

Qualitative Evaluation of ASR Adaptation in a Lecture Context: Application to the PASTEL Corpus

Salima Mdhaffar¹, Yannick Estève², Nicolas Hernandez³, Antoine Laurent¹, Richard Dufour², Solen Quiniou³

¹LIUM (EA 4023), France; ²LIA (EA 4128), France; ³LS2N (UMR 6004), France

PAGE 574
Mon-P-1-D-5
11:00-13:00

Active Annotation: Bootstrapping Annotation Lexicon and Guidelines for Supervised NLU Learning

Federico Marinelli¹, Alessandra Cervone¹, Giuliano Tortoreto², Evgeny A. Stepanov², Giuseppe Di Fabbrizio³, Giuseppe Riccardi¹

¹Università di Trento, Italy; ²VUI, Italy; ³VUI, USA

PAGE 579
Mon-P-1-D-6
11:00-13:00

Automatic Lyric Transcription from Karaoke Vocal Tracks: Resources and a Baseline System

Gerardo Roa Dabike, Jon Barker, University of Sheffield, UK

Mon-P-1-D continued...

PAGE 584
Mon-P-1-D-7
11:00-13:00

Detecting Mismatch Between Speech and Transcription Using Cross-Modal Attention

Qiang Huang, Thomas Hain, University of Sheffield, UK

PAGE 589
Mon-P-1-D-8
11:00-13:00

EpaDB: A Database for Development of Pronunciation Assessment Systems

Jazmín Vidal, Luciana Ferrer, Leonardo Brambilla, UBA, Argentina

PAGE 594
Mon-P-1-D-9
11:00-13:00

Automatic Compression of Subtitles with Neural Networks and its Effect on User Experience

Katrin Angerbauer, Heike Adel, Ngoc Thang Vu, Universität Stuttgart, Germany

PAGE 599
Mon-P-1-D-10
11:00-13:00

Integrating Video Retrieval and Moment Detection in a Unified Corpus for Video Question Answering

Hongyin Luo¹, Mitra Mohtarami¹, James Glass¹, Karthik Krishnamurthy², Brigitte Richardson²

¹MIT, USA; ²Ford, USA

Mon-P-1-E: Speech and Audio Characterization and Segmentation

Chair: Korbinian Riedhammer

Time 11:00-13:00, Monday 16 Sept 2019, Hall 10/E

PAGE 604
Mon-P-1-E-1
11:00-13:00

Early Identification of Speech Changes Due to Amyotrophic Lateral Sclerosis Using Machine Classification

Sarah E. Gutz¹, Jun Wang², Yana Yunusova³, Jordan R. Green¹

¹MGH Institute of Health Professions, USA; ²University of Texas at Austin, USA;

³University of Toronto, Canada

PAGE 609
Mon-P-1-E-2
11:00-13:00

Automatic Detection of Breath Using Voice Activity Detection and SVM Classifier with Application on News Reports

Mohamed Ismail Yasar Arafath K., Aurobinda Routray, IIT Kharagpur, India

PAGE 614
Mon-P-1-E-3
11:00-13:00

Acoustic Scene Classification Using Teacher-Student Learning with Soft-Labels

Hee-Soo Heo, Jee-weon Jung, Hye-jin Shim, Ha-Jin Yu, University of Seoul, Korea

PAGE 619
Mon-P-1-E-4
11:00-13:00

Rare Sound Event Detection Using Deep Learning and Data Augmentation

Yanping Chen, Hongxia Jin, Samsung, USA

PAGE 624
Mon-P-1-E-5
11:00-13:00

A Combination of Model-Based and Feature-Based Strategy for Speech-to-Singing Alignment

Bidisha Sharma, Haizhou Li, NUS, Singapore

PAGE 629
Mon-P-1-E-6
11:00-13:00

Dr.VOT: Measuring Positive and Negative Voice Onset Time in the Wild

Yosi Shrem¹, Matthew Goldrick², Joseph Keshet¹

¹Bar-Ilan University, Israel; ²Northwestern University, USA

PAGE 634
Mon-P-1-E-7
11:00-13:00

Effects of Base-Frequency and Spectral Envelope on Deep-Learning Speech Separation and Recognition Models

J. Hui, Y. Wei, S.T. Chen, R.H.Y. So, HKUST, China

INTERSPEECH 2019 Table of Contents

Mon-P-1-E continued...

- PAGE 639
Mon-P-1-E-8
11:00-13:00 **Phone Aware Nearest Neighbor Technique Using Spectral Transition Measure for Non-Parallel Voice Conversion**
Nirmesh J. Shah, Hemant A. Patil, DA-IICT, India
- PAGE 644
Mon-P-1-E-9
11:00-13:00 **Weakly Supervised Syllable Segmentation by Vowel-Consonant Peak Classification**
Ravi Shankar, Archana Venkataraman, Johns Hopkins University, USA
- PAGE 649
Mon-P-1-E-10
11:00-13:00 **An Approach to Online Speaker Change Point Detection Using DNNs and WFSTs**
Lukas Mateju, Petr Cerva, Jindrich Zdansky, Technical University of Liberec, Czech Republic
- PAGE 654
Mon-P-1-E-11
11:00-13:00 **Regression and Classification for Direction-of-Arrival Estimation with Convolutional Recurrent Neural Networks**
Zhenyu Tang, John D. Kanu, Kevin Hogan, Dinesh Manocha, University of Maryland at College Park, USA

INTERSPEECH 2019 Table of Contents

Mon-P-2-A: Neural Techniques for Voice Conversion and Waveform Generation

Chair: Gustav Eje Henter

Time 14:30-16:30, Monday 16 Sept 2019, Gallery A

- PAGE 659
Mon-P-2-A-1
14:30-16:30 **Non-Parallel Voice Conversion Using Weighted Generative Adversarial Networks**
Dipjyoti Paul¹, Yannis Pantazis², Yannis Stylianou¹
¹University of Crete, Greece; ²FORTH, Greece
- PAGE 664
Mon-P-2-A-2
14:30-16:30 **One-Shot Voice Conversion by Separating Speaker and Content Representations with Instance Normalization**
Ju-chieh Chou, Hung-Yi Lee, National Taiwan University, Taiwan
- PAGE 669
Mon-P-2-A-3
14:30-16:30 **One-Shot Voice Conversion with Global Speaker Embeddings**
Hui Lu¹, Zhiyong Wu¹, Dongyang Dai¹, Runnan Li¹, Shiyin Kang², Jia Jia¹, Helen Meng¹
¹Tsinghua University, China; ²Tencent, China
- PAGE 674
Mon-P-2-A-4
14:30-16:30 **Non-Parallel Voice Conversion with Cyclic Variational Autoencoder**
Patrick Lumban Tobing, Yi-Chiao Wu, Tomoki Hayashi, Kazuhiro Kobayashi, Tomoki Toda, Nagoya University, Japan
- PAGE 679
Mon-P-2-A-5
14:30-16:30 **StarGAN-VC2: Rethinking Conditional Methods for StarGAN-Based Voice Conversion**
Takuhiro Kaneko, Hirokazu Kameoka, Kou Tanaka, Nobukatsu Hojo, NTT, Japan
- PAGE 684
Mon-P-2-A-6
14:30-16:30 **Robustness of Statistical Voice Conversion Based on Direct Waveform Modification Against Background Sounds**
Yusuke Kurita, Kazuhiro Kobayashi, Kazuya Takeda, Tomoki Toda, Nagoya University, Japan

INTERSPEECH 2019 Table of Contents

Mon-P-2-A continued...

PAGE 689
Mon-P-2-A-7
14:30-16:30

Fast Learning for Non-Parallel Many-to-Many Voice Conversion with Residual Star Generative Adversarial Networks

Shengkui Zhao, Trung Hieu Nguyen, Hao Wang, Bin Ma, Alibaba Group, Singapore

PAGE 694
Mon-P-2-A-8
14:30-16:30

GELP: GAN-Excited Linear Prediction for Speech Synthesis from Mel-Spectrogram

Lauri Juvela¹, Bajibabu Bollepalli¹, Junichi Yamagishi², Paavo Alku¹
¹Aalto University, Finland; ²NII, Japan

PAGE 699
Mon-P-2-A-9
14:30-16:30

Probability Density Distillation with Generative Adversarial Networks for High-Quality Parallel Waveform Generation

Ryuichi Yamamoto¹, Eunwoo Song², Jae-Min Kim²
¹LINE, Japan; ²NAVER, Korea

PAGE 704
Mon-P-2-A-10
14:30-16:30

One-Shot Voice Conversion with Disentangled Representations by Leveraging Phonetic Posteriorgrams

Seyed Hamidreza Mohammadi, Taehwan Kim, ObEN, USA

PAGE 709
Mon-P-2-A-11
14:30-16:30

Investigation of F0 Conditioning and Fully Convolutional Networks in Variational Autoencoder Based Voice Conversion

Wen-Chin Huang¹, Yi-Chiao Wu², Chen-Chou Lo¹, Patrick Lumban Tobing², Tomoki Hayashi², Kazuhiro Kobayashi², Tomoki Toda², Yu Tsao¹, Hsin-Min Wang¹
¹Academia Sinica, Taiwan; ²Nagoya University, Japan

PAGE 714
Mon-P-2-A-12
14:30-16:30

Jointly Trained Conversion Model and WaveNet Vocoder for Non-Parallel Voice Conversion Using Mel-Spectrograms and Phonetic Posteriorgrams

Songxiang Liu¹, Yuewen Cao¹, Xixin Wu¹, Lifa Sun², Xunying Liu¹, Helen Meng¹
¹CUHK, China; ²SpeechX, China

INTERSPEECH 2019 Table of Contents

Mon-P-2-A continued...

PAGE 719
Mon-P-2-A-13
14:30-16:30

Generative Adversarial Networks for Unpaired Voice Transformation on Impaired Speech

Li-Wei Chen¹, Hung-Yi Lee¹, Yu Tsao²
¹National Taiwan University, Taiwan; ²Academia Sinica, Taiwan

PAGE 724
Mon-P-2-A-14
14:30-16:30

Group Latent Embedding for Vector Quantized Variational Autoencoder in Non-Parallel Voice Conversion

Shaojin Ding, Ricardo Gutierrez-Osuna, Texas A&M University, USA

PAGE 729
Mon-P-2-A-15
14:30-16:30

Semi-Supervised Voice Conversion with Amortized Variational Inference

Cory Stephenson, Gokce Keskin, Anil Thomas, Oguz H. Elibol, Intel, USA

Mon-P-2-B: Model Adaptation for ASR

Chair: Penny Karanasou

Time 14:30-16:30, Monday 16 Sept 2019, Gallery B

PAGE 734
Mon-P-2-B-1
14:30-16:30

Exploiting Semi-Supervised Training Through a Dropout Regularization in End-to-End Speech Recognition

Subhadeep Dey¹, Petr Motlicek¹, Trung Bui², Franck Dernoncourt²
¹Idiap Research Institute, Switzerland; ²Adobe, USA

PAGE 739
Mon-P-2-B-2
14:30-16:30

Improved Vocal Tract Length Perturbation for a State-of-the-Art End-to-End Speech Recognition System

Chanwoo Kim, Minkyu Shin, Abhinav Garg, Dhananjaya Gowda, Samsung, Korea

PAGE 744
Mon-P-2-B-3
14:30-16:30

Multi-Accent Adaptation Based on Gate Mechanism

Han Zhu, Li Wang, Pengyuan Zhang, Yonghong Yan, Chinese Academy of Sciences, China

PAGE 749
Mon-P-2-B-4
14:30-16:30

Unsupervised Adaptation with Adversarial Dropout Regularization for Robust Speech Recognition

Pengcheng Guo, Sining Sun, Lei Xie, Northwestern Polytechnical University, China

PAGE 754
Mon-P-2-B-5
14:30-16:30

Cumulative Adaptation for BLSTM Acoustic Models

Markus Kitzka, Pavel Golik, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

PAGE 759
Mon-P-2-B-6
14:30-16:30

Fast DNN Acoustic Model Speaker Adaptation by Learning Hidden Unit Contribution Features

Xurong Xie¹, Xunying Liu¹, Tan Lee¹, Lan Wang²
¹CUHK, China; ²Chinese Academy of Sciences, China

Mon-P-2-B continued...

PAGE 764
Mon-P-2-B-7
14:30-16:30

End-to-End Adaptation with Backpropagation Through WFST for On-Device Speech Recognition System

Emiru Tsunoo, Yosuke Kashiwagi, Satoshi Asakawa, Toshiyuki Kumakura, Sony, Japan

PAGE 769
Mon-P-2-B-8
14:30-16:30

Learning Speaker Aware Offsets for Speaker Adaptation of Neural Networks

Leda Sari¹, Samuel Thomas², Mark A. Hasegawa-Johnson¹
¹University of Illinois at Urbana-Champaign, USA; ²IBM, USA

PAGE 774
Mon-P-2-B-9
14:30-16:30

An Investigation into On-Device Personalization of End-to-End Automatic Speech Recognition Models

Khe Chai Sim, Petr Zadrzil, Françoise Beaufays, Google, USA

PAGE 779
Mon-P-2-B-10
14:30-16:30

A Multi-Accent Acoustic Model Using Mixture of Experts for Speech Recognition

Abhinav Jain, Vishwanath P. Singh, Shakti P. Rath, Samsung, India

PAGE 784
Mon-P-2-B-11
14:30-16:30

Personalizing ASR for Dysarthric and Accented Speech with Limited Data

Joel Shor¹, Dotan Emanuel¹, Oran Lang¹, Omry Tuval¹, Michael Brenner¹, Julie Cattiau¹, Fernando Vieira², Maevie McNally², Taylor Charbonneau², Melissa Nollstadt², Avinatan Hassidim¹, Yossi Matias¹
¹Google, USA; ²ALS TDI, USA

Mon-P-2-C : Dialogue Speech Understanding

Chair: Ian Lane

Time 14:30-16:30, Monday 16 Sept 2019, Gallery C

- PAGE 789
Mon-P-2-C-1
14:30-16:30
- Mitigating Noisy Inputs for Question Answering**
Denis Peskov¹, Joe Barrow¹, Pedro Rodriguez¹, Graham Neubig², Jordan Boyd-Graber¹
¹University of Maryland at College Park, USA; ²Carnegie Mellon University, USA
- PAGE 794
Mon-P-2-C-2
14:30-16:30
- One-vs-All Models for Asynchronous Training: An Empirical Analysis**
Rahul Gupta, Aman Alok, Shankar Ananthakrishnan, Amazon, USA
- PAGE 799
Mon-P-2-C-3
14:30-16:30
- Adapting a FrameNet Semantic Parser for Spoken Language Understanding Using Adversarial Learning**
Gabriel Marzinotto¹, Géraldine Damnati¹, Frédéric Béchet²
¹Orange Labs, France; ²LIS (UMR 7020), France
- PAGE 804
Mon-P-2-C-4
14:30-16:30
- M2H-GAN: A GAN-Based Mapping from Machine to Human Transcripts for Speech Understanding**
Titouan Parcollet¹, Mohamed Morchid¹, Xavier Bost², Georges Linarès¹
¹LIA (EA 4128), France; ²Orkis, France
- PAGE 809
Mon-P-2-C-5
14:30-16:30
- Ultra-Compact NLU: Neuronal Network Binarization as Regularization**
Munir Georges¹, Krzysztof Czarnowski², Tobias Bocklet¹
¹Intel, Germany; ²Intel, Poland
- PAGE 814
Mon-P-2-C-6
14:30-16:30
- Speech Model Pre-Training for End-to-End Spoken Language Understanding**
Loren Lugosch¹, Mirco Ravanelli¹, Patrick Ignoto², Vikrant Singh Tomar², Yoshua Bengio¹
¹Université de Montréal, Canada; ²Fluent.ai, Canada

Mon-P-2-C continued...

- PAGE 819
Mon-P-2-C-7
14:30-16:30
- Spoken Language Intent Detection Using Confusion2Vec**
Prashanth Gurunath Shivakumar, Mu Yang, Panayiotis Georgiou, University of Southern California, USA
- PAGE 824
Mon-P-2-C-8
14:30-16:30
- Investigating Adaptation and Transfer Learning for End-to-End Spoken Language Understanding from Speech**
Natalia Tomashenko¹, Antoine Caubrière², Yannick Estève¹
¹LIA (EA 4128), France; ²LIUM (EA 4023), France
- PAGE 829
Mon-P-2-C-9
14:30-16:30
- Topic-Aware Dialogue Speech Recognition with Transfer Learning**
Yuanfeng Song¹, Di Jiang², Xueyang Wu¹, Qian Xu², Raymond Chi-Wing Wong¹, Qiang Yang¹
¹HKUST, China; ²WeBank, China
- PAGE 834
Mon-P-2-C-10
14:30-16:30
- Improving Conversation-Context Language Models with Multiple Spoken Language Understanding Models**
Ryo Masumura, Tomohiro Tanaka, Atsushi Ando, Hosana Kamiyama, Takanobu Oba, Satoshi Kobashikawa, Yushi Aono, NTT, Japan
- PAGE 839
Mon-P-2-C-11
14:30-16:30
- Meta Learning for Hyperparameter Optimization in Dialogue System**
Jen-Tzung Chien, Wei Xiang Lieow, National Chiao Tung University, Taiwan
- PAGE 844
Mon-P-2-C-12
14:30-16:30
- Zero Shot Intent Classification Using Long-Short Term Memory Networks**
Kyle Williams, Microsoft, USA
- PAGE 849
Mon-P-2-C-13
14:30-16:30
- A Comparison of Deep Learning Methods for Language Understanding**
Mandy Korpusik¹, Zoe Liu², James Glass²
¹Loyola Marymount University, USA; ²MIT, USA
- PAGE 854
Mon-P-2-C-14
14:30-16:30
- Slot Filling with Weighted Multi-Encoders for Out-of-Domain Values**
Yuka Kobayashi, Takami Yoshida, Kenji Iwata, Hiroshi Fujimura, Toshiba, Japan

Mon-P-2-D: Speech Production and Silent Interfaces

Chair: Jana Brunner

Time 14:30-16:30, Monday 16 Sept 2019, Hall 10/D

- PAGE 859
Mon-P-2-D-1
14:30-16:30
- Multi-Corpus Acoustic-to-Articulatory Speech Inversion**
Nadee Seneviratne¹, Ganesh Sivaraman², Carol Espy-Wilson¹
¹University of Maryland at College Park, USA; ²Pindrop, USA
- PAGE 864
Mon-P-2-D-2
14:30-16:30
- Towards a Speaker Independent Speech-BCI Using Speaker Adaptation**
Debadatta Dash¹, Alan Wisler¹, Paul Ferrari², Jun Wang¹
¹University of Texas at Dallas, USA; ²University of Texas at Austin, USA
- PAGE 869
Mon-P-2-D-3
14:30-16:30
- Identifying Input Features for Development of Real-Time Translation of Neural Signals to Text**
Janaki Sheth¹, Ariel Tankus², Michelle Tran¹, Lindy Comstock¹, Itzhak Fried¹, William Speier¹
¹University of California at Los Angeles, USA; ²Tel Aviv University, Israel
- PAGE 874
Mon-P-2-D-4
14:30-16:30
- Exploring Critical Articulator Identification from 50Hz RT-MRI Data of the Vocal Tract**
Samuel Silva¹, António Teixeira¹, Conceição Cunha², Nuno Almeida¹, Arun A. Joseph³, Jens Frahm³
¹Universidade de Aveiro, Portugal; ²LMU München, Germany; ³MPI for Biophysical Chemistry, Germany
- PAGE 879
Mon-P-2-D-5
14:30-16:30
- Towards a Method of Dynamic Vocal Tract Shapes Generation by Combining Static 3D and Dynamic 2D MRI Speech Data**
Ioannis K. Douros¹, Anastasiia Tsukanova¹, Karyna Isaieva², Pierre-André Vuissoz², Yves Laprie¹
¹Loria (UMR 7503), France; ²IADI (Inserm U1254), France

Mon-P-2-D continued...

- PAGE 884
Mon-P-2-D-6
14:30-16:30
- Temporal Coordination of Articulatory and Respiratory Events Prior to Speech Initiation**
Oksana Rasskazova¹, Christine Mooshammer¹, Susanne Fuchs²
¹Humboldt-Universität zu Berlin, Germany; ²ZAS, Germany
- PAGE 889
Mon-P-2-D-7
14:30-16:30
- Zooming in on Spatiotemporal V-to-C Coarticulation with Functional PCA**
Michele Gubian, Manfred Pastötter, Marianne Pouplier, LMU München, Germany
- PAGE 894
Mon-P-2-D-8
14:30-16:30
- Ultrasound-Based Silent Speech Interface Built on a Continuous Vocoder**
Tamás Gábor Csapó¹, Mohammed Salah Al-Radhi¹, Géza Németh¹, Gábor Gosztolya², Tamás Grósz³, László Tóth³, Alexandra Markó⁴
¹BME, Hungary; ²MTA-SZTE RGAI, Hungary; ³University of Szeged, Hungary; ⁴MTA-ELTE LingArt, Hungary
- PAGE 899
Mon-P-2-D-9
14:30-16:30
- Assessing Acoustic and Articulatory Dimensions of Speech Motor Adaptation with Random Forests**
Eugen Klein¹, Jana Brunner¹, Phil Hoole²
¹Humboldt-Universität zu Berlin, Germany; ²LMU München, Germany
- PAGE 904
Mon-P-2-D-11
14:30-16:30
- Speech Organ Contour Extraction Using Real-Time MRI and Machine Learning Method**
Hironori Takemoto¹, Tsubasa Goto¹, Yuya Hagihara¹, Sayaka Hamanaka¹, Tatsuya Kitamura², Yukiko Nota³, Kikuo Maekawa³
¹Chiba Institute of Technology, Japan; ²Konan University, Japan; ³NINJAL, Japan
- PAGE 909
Mon-P-2-D-12
14:30-16:30
- CNN-Based Phoneme Classifier from Vocal Tract MRI Learns Embedding Consistent with Articulatory Topology**
K.G. van Leeuwen, P. Bos, S. Trebeschi, M.J.A. van Alphen, L. Voskuilen, L.E. Smeele, F. van der Heijden, R.J.J.H. van Son, NKI, The Netherlands

Mon-P-2-D continued...

PAGE 914
Mon-P-2-D-13
14:30-16:30

Strength and Structure: Coupling Tones with Oral Constriction Gestures

Doris Mücke¹, Anne Hermes¹, Sam Tilsen²

¹Universität zu Köln, Germany; ²Cornell University, USA

Mon-P-2-E: Speech Signal Characterization 2

Chair: Jean-François Bonastre

Time 14:30-16:30, Monday 16 Sept 2019, Hall 10/E

PAGE 919
Mon-P-2-E-1
14:30-16:30

Salient Speech Representations Based on Cloned Networks

W. Bastiaan Kleijn, Felicia S.C. Lim, Michael Chinen, Jan Skoglund, Google, USA

PAGE 924
Mon-P-2-E-2
14:30-16:30

ASR Inspired Syllable Stress Detection for Pronunciation Evaluation Without Using a Supervised Classifier and Syllable Level Features

Manoj Kumar Ramanathi, Chiranjeevi Yarra, Prasanta Kumar Ghosh, Indian Institute of Science, India

PAGE 929
Mon-P-2-E-3
14:30-16:30

Acoustic and Articulatory Feature Based Speech Rate Estimation Using a Convolutional Dense Neural Network

Renuka Mannem¹, Jhansi Mallela², Aravind Illa¹, Prasanta Kumar Ghosh¹
¹Indian Institute of Science, India; ²RGUKT, India

PAGE 934
Mon-P-2-E-5
14:30-16:30

Predictive Auxiliary Variational Autoencoder for Representation Learning of Global Speech Characteristics

Sebastian Springenberg, Egor Lakomkin, Cornelius Weber, Stefan Wermter, Universität Hamburg, Germany

PAGE 939
Mon-P-2-E-6
14:30-16:30

Unsupervised Low-Rank Representations for Speech Emotion Recognition

Georgios Paraskevopoulos¹, Efthymios Tzinis², Nikolaos Ellinas¹, Theodoros Giannakopoulos³, Alexandros Potamianos¹

¹NTUA, Greece; ²University of Illinois at Urbana-Champaign, USA; ³Behavioral Signal Technologies, USA

PAGE 944
Mon-P-2-E-7
14:30-16:30

On the Suitability of the Riesz Spectro-Temporal Envelope for WaveNet Based Speech Synthesis

Jitendra Kumar Dhiman¹, Nagaraj Adiga², Chandra Sekhar Seelamantula¹

¹Indian Institute of Science, India; ²University of Crete, Greece

Mon-P-2-E continued...

PAGE 949
Mon-P-2-E-8
14:30-16:30

Autonomous Emotion Learning in Speech: A View of Zero-Shot Speech Emotion Recognition

Xinzhou Xu¹, Jun Deng², Nicholas Cummins³, Zixing Zhang⁴, Li Zhao⁵, Björn W. Schuller³

¹NJUPT, China; ²Agile Robots, Germany; ³Universität Augsburg, Germany; ⁴Imperial College London, UK; ⁵Southeast University, China

PAGE 954
Mon-P-2-E-9
14:30-16:30

An Improved Goodness of Pronunciation (GoP) Measure for Pronunciation Evaluation with DNN-HMM System Considering HMM Transition Probabilities

Sweekar Sudhakara, Manoj Kumar Ramanathi, Chiranjeevi Yarra, Prasanta Kumar Ghosh, Indian Institute of Science, India

PAGE 959
Mon-P-2-E-10
14:30-16:30

Low Resource Automatic Intonation Classification Using Gated Recurrent Unit (GRU) Networks Pre-Trained with Synthesized Pitch Patterns

Atreyee Saha¹, Chiranjeevi Yarra², Prasanta Kumar Ghosh²

¹Jadavpur University, India; ²Indian Institute of Science, India

Mon-S&T-1: Applications in Language Learning and Healthcare

Time 14:30-16:30, Monday 16 Sept 2019, Hall 4

PAGE 964
Mon-S&T-1-1
14:30-16:30

Apkinson: A Mobile Solution for Multimodal Assessment of Patients with Parkinson's Disease

J.C. Vásquez-Correa¹, T. Arias-Vergara¹, Philipp Klumpp¹, M. Strauss¹, A. Küderle¹, N. Roth¹, S. Bayerl², N. García-Ospina³, P.A. Perez-Toro³, L.F. Parra-Gallego³, Cristian David Rios-Urrego³, D. Escobar-Grisales³, Juan Rafael Orozco-Arroyave¹, B. Eskofier¹, Elmar Nöth¹

¹FAU Erlangen-Nürnberg, Germany; ²Technische Hochschule Rosenheim, Germany;

³Universidad de Antioquia, Colombia

PAGE 966
Mon-S&T-1-2
14:30-16:30

Depression State Assessment: Application for Detection of Depression by Speech

Gábor Kiss, Dávid Sztahó, Klára Vicsi, BME, Hungary

PAGE 968
Mon-S&T-1-3
14:30-16:30

SPIRE-fluent: A Self-Learning App for Tutoring Oral Fluency to Second Language English Learners

Chiranjeevi Yarra¹, Aparna Srinivasan¹, Sravani Gottimukkala², Prasanta Kumar Ghosh¹

¹Indian Institute of Science, India; ²RGUKT, India

PAGE 970
Mon-S&T-1-4
14:30-16:30

Using Real-Time Visual Biofeedback for Second Language Instruction

Shawn Nissen, Rebecca Nissen, Brigham Young University, USA

PAGE 972
Mon-S&T-1-5
14:30-16:30

Splash: Speech and Language Assessment in Schools and Homes

A. Miwardelli¹, I. Gallagher², J. Gibson¹, N. Katsos¹, Kate M. Knill¹, H. Wood¹

¹University of Cambridge, UK; ²ig Projects, UK

INTERSPEECH 2019 Table of Contents

Mon-S&T-1 continued...

PAGE 974
Mon-S&T-1-6
14:30-16:30

Using Ultrasound Imaging to Create Augmented Visual Biofeedback for Articulatory Practice

Colin T. Annand, Maurice Lamb, Sarah Dugan, Sarah R. Li, Hannah M. Woeste, T. Douglas Mast, Michael A. Riley, Jack A. Masterson, Neeraja Mahalingam, Kathryn J. Eary, Caroline Spencer, Suzanne Boyce, Stephanie Jackson, Anoosha Baxi, Renée Seward, University of Cincinnati, USA

PAGE 976
Mon-S&T-1-7
14:30-16:30

Speech-Based Web Navigation for Limited Mobility Users

Vasily Radostev, Serge Berger, Justin Tabrizi, Pasha Kamyshev, Hisami Suzuki, Microsoft, USA

INTERSPEECH 2019 Table of Contents

Tue-K-2 : Keynote 2: Tanja Schultz

Chair: To be confirmed

Time 08:30-09:30, Tuesday 17 Sept 2019, Main Hall

Tue-K-2
08:30-09:30
N/A

Biosignal Processing for Human-Machine Interaction

Tanja Schultz, Universität Bremen, Germany

Tue-SS-3-6: The Second DIHARD Speech Diarization Challenge (DIHARD II)

Chairs: Neville Ryant, Alejandrina Cristia

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 3

PAGE 978
Tue-SS-3-6-1
10:00-10:20

The Second DIHARD Diarization Challenge: Dataset, Task, and Baselines

Neville Ryant¹, Kenneth Church², Christopher Cieri¹, Alejandrina Cristia³, Jun Du⁴,
Sriram Ganapathy⁵, Mark Liberman¹

¹University of Pennsylvania, USA; ²Baidu Research, USA; ³LSCP (UMR 8554), France;
⁴USTC, China; ⁵Indian Institute of Science, India

PAGE 983
Tue-SS-3-6-2
10:20-10:40

LEAP Diarization System for the Second DIHARD Challenge

Prachi Singh¹, Harsha Vardhan M.A.¹, Sriram Ganapathy¹, A. Kanagasundaram²

¹Indian Institute of Science, India; ²University of Jaffna, Sri Lanka

PAGE 988
Tue-SS-3-6-3
10:40-11:00

ViVoLAB Speaker Diarization System for the DIHARD 2019 Challenge

Ignacio Viñals, Pablo Gimeno, Alfonso Ortega, Antonio Miguel, Eduardo Lleida,
Universidad de Zaragoza, Spain

PAGE 993
Tue-SS-3-6-4
11:00-11:20

UWB-NTIS Speaker Diarization System for the DIHARD II 2019 Challenge

Zbyněk Zajíc, Marie Kunešová, Marek Hruží, Jan Vaněk, University of West Bohemia,
Czech Republic

PAGE 998
Tue-SS-3-6-5
11:20-11:40

The Second DIHARD Challenge: System Description for USC-SAIL Team

Tae Jin Park, Manoj Kumar, Nikolaos Flemotomos, Monisankha Pal, Raghuvveer Peri,
Rimita Lahiri, Panayiotis Georgiou, Shrikanth Narayanan, University of Southern
California, USA

PAGE 1003
Tue-SS-3-6-6
11:40-12:00

Speaker Diarization with Deep Speaker Embeddings for DIHARD Challenge II

Sergey Novoselov¹, Aleksei Gusev¹, Artem Ivanov², Timur Pekhovsky²,
Andrey Shulipa¹, Anastasia Avdeeva¹, Artem Gorlanov², Alexandr Kozlov²

¹ITMO University, Russia; ²STC-innovations, Russia

Tue-SS-4-4: The 2019 Automatic Speaker Verification Spoofing and Countermeasures Challenge: ASVspoof Challenge — O

Chairs: Junichi Yamagishi, Nicholas Evans

Time 14:30-15:30, Tuesday 17 Sept 2019, Hall 11

PAGE 1008
Tue-SS-4-4-16
14:30-15:00

ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection

Massimiliano Todisco¹, Xin Wang², Ville Vestman³, Md. Sahidullah⁴,
Héctor Delgado¹, Andreas Nautsch¹, Junichi Yamagishi², Nicholas Evans¹,
Tomi H. Kinnunen³, Kong Aik Lee⁵

¹EURECOM, France; ²NII, Japan; ³University of Eastern Finland, Finland; ⁴Loria (UMR
7503), France; ⁵NEC, Japan

Tue-SS-4-4-17
15:00-15:30

Discussion

Tue-SS-4-A: The 2019 Automatic Speaker Verification Spoofing and Countermeasures Challenge: ASVspoof Challenge — P

Chairs: Junichi Yamagishi, Nicholas Evans

Time 13:30-14:30, Tuesday 17 Sept 2019, Gallery A

- PAGE 1013
Tue-SS-4-4-1
13:30-14:30 **ASSERT: Anti-Spoofing with Squeeze-Excitation and Residual Networks**
Cheng-I Lai, Nanxin Chen, Jesús Villalba, Najim Dehak, Johns Hopkins University, USA
- PAGE 1018
Tue-SS-4-4-2
13:30-14:30 **Ensemble Models for Spoofing Detection in Automatic Speaker Verification**
Bhusan Chettri¹, Daniel Stoller¹, Veronica Morfi¹, Marco A. Martínez Ramírez¹, Emmanouil Benetos¹, Bob L. Sturm²
¹Queen Mary University of London, UK; ²KTH, Sweden
- PAGE 1023
Tue-SS-4-4-3
13:30-14:30 **The DKU Replay Detection System for the ASVspoof 2019 Challenge: On Data Augmentation, Feature Representation, Classification, and Fusion**
Weicheng Cai, Haiwei Wu, Danwei Cai, Ming Li, Duke Kunshan University, China
- PAGE 1028
Tue-SS-4-4-4
13:30-14:30 **Robust Bayesian and Light Neural Networks for Voice Spoofing Detection**
Radosław Białobrzeski, Michał Kośmider, Mateusz Matuszewski, Marcin Plata, Alexander Rakowski, Samsung, Poland
- PAGE 1033
Tue-SS-4-4-5
13:30-14:30 **STC Antispoofing Systems for the ASVspoof2019 Challenge**
Galina Lavrentyeva¹, Sergey Novoselov², Andzhukaev Tseren¹, Marina Volkova¹, Artem Gorlanov¹, Alexandr Kozlov¹
¹STC-innovations, Russia; ²ITMO University, Russia
- PAGE 1038
Tue-SS-4-4-6
13:30-14:30 **The SJTU Robust Anti-Spoofing System for the ASVspoof 2019 Challenge**
Yexin Yang, Hongji Wang, Heinrich Dinkel, Zhengyang Chen, Shuai Wang, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China

Tue-SS-4-A continued...

- PAGE 1043
Tue-SS-4-4-7
13:30-14:30 **IIIT-H Spoofing Countermeasures for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2019**
K.N.R.K. Raju Alluri, Anil Kumar Vuppala, IIIT Hyderabad, India
- PAGE 1048
Tue-SS-4-4-8
13:30-14:30 **Anti-Spoofing Speaker Verification System with Multi-Feature Integration and Multi-Task Learning**
Rongjin Li, Miao Zhao, Zheng Li, Lin Li, Qingyang Hong, Xiamen University, China
- PAGE 1053
Tue-SS-4-4-9
13:30-14:30 **Speech Replay Detection with x-Vector Attack Embeddings and Spectral Features**
Jennifer Williams, Joanna Rownicka, University of Edinburgh, UK
- PAGE 1058
Tue-SS-4-4-10
13:30-14:30 **Long Range Acoustic Features for Spoofed Speech Detection**
Rohan Kumar Das, Jichen Yang, Haizhou Li, NUS, Singapore
- PAGE 1063
Tue-SS-4-4-11
13:30-14:30 **Transfer-Representation Learning for Detecting Spoofing Attacks with Converted and Synthesized Speech in Automatic Speaker Verification System**
Su-Yu Chang, Kai-Cheng Wu, Chia-Ping Chen, National Sun Yat-sen University, Taiwan
- PAGE 1068
Tue-SS-4-4-12
13:30-14:30 **A Light Convolutional GRU-RNN Deep Feature Extractor for ASV Spoofing Detection**
Alejandro Gomez-Alanis¹, Antonio M. Peinado¹, Jose A. Gonzalez², Angel M. Gomez¹
¹Universidad de Granada, Spain; ²Universidad de Málaga, Spain
- PAGE 1073
Tue-SS-4-4-13
13:30-14:30 **Detecting Spoofing Attacks Using VGG and SincNet: BUT-Omilia Submission to ASVspoof 2019 Challenge**
Hossein Zeinali¹, Themis Stafylakis², Georgia Athanasopoulou², Johan Rohdin¹, Ioannis Gkinis², Lukáš Burget¹, Jan Černocký¹
¹Brno University of Technology, Czech Republic; ²Omilia, Greece

Tue-SS-4-A continued...

PAGE 1078
Tue-SS-4-4-14
13:30-14:30

Deep Residual Neural Networks for Audio Spoofing Detection
Moustafa Alzantot, Ziqi Wang, Mani B. Srivastava, University of California at Los Angeles, USA

PAGE 1083
Tue-SS-4-4-15
13:30-14:30

Replay Attack Detection with Complementary High-Resolution Information Using End-to-End DNN for the ASVspoof 2019 Challenge
Jee-weon Jung, Hye-jin Shim, Hee-Soo Heo, Ha-Jin Yu, University of Seoul, Korea

Tue-SS-5-6: The Zero Resource Speech Challenge 2019: TTS Without T

Chairs: Ewan Dunbar, Laurent Besacier

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 3

PAGE 1088
Tue-SS-5-6-1
16:00-16:15

The Zero Resource Speech Challenge 2019: TTS Without T
Ewan Dunbar¹, Robin Algayres², Julien Karadayi², Mathieu Bernard², Juan Benjumea², Xuan-Nga Cao², Lucie Miskic¹, Charlotte Dugrain¹, Lucas Ondel³, Alan W. Black⁴, Laurent Besacier⁵, Sakriani Sakti⁶, Emmanuel Dupoux²
¹LLF (UMR 7110), France; ²LSCP (UMR 8554), France; ³Brno University of Technology, Czech Republic; ⁴Carnegie Mellon University, USA; ⁵LIG (UMR 5217), France; ⁶NAIST, Japan

PAGE 1093
Tue-SS-5-6-2
16:15-16:30

Combining Adversarial Training and Disentangled Speech Representation for Robust Zero-Resource Subword Modeling
Siyuan Feng, Tan Lee, Zhiyuan Peng, CUHK, China

PAGE 1098
Tue-SS-5-6-3
16:30-16:45

Temporally-Aware Acoustic Unit Discovery for Zerospeech 2019 Challenge
Bolaji Yusuf, Alican Gök, Batuhan Gundogdu, Oyku Deniz Kose, Murat Saraclar, Boğaziçi Üniversitesi, Turkey

PAGE 1103
Tue-SS-5-6-4
16:45-17:00

Unsupervised Acoustic Unit Discovery for Speech Synthesis Using Discrete Latent-Variable Neural Networks
Ryan Eloff¹, André Nortje¹, Benjamin van Niekerk¹, Avashna Govender², Leanne Nortje¹, Arnu Pretorius¹, Elan van Biljon¹, Ewald van der Westhuizen¹, Lisa van Staden¹, Herman Kamper¹
¹Stellenbosch University, South Africa; ²University of Edinburgh, UK

PAGE 1108
Tue-SS-5-6-5
17:05-17:20

Unsupervised End-to-End Learning of Discrete Linguistic Units for Voice Conversion
Andy T. Liu, Po-chun Hsu, Hung-Yi Lee, National Taiwan University, Taiwan

INTERSPEECH 2019 Table of Contents

Tue-SS-5-6 continued...

PAGE 1113
Tue-SS-5-6-6
17:20-17:35

Zero Resource Speech Synthesis Using Transcripts Derived from Perceptual Acoustic Units

Karthik Pandia D. S., Hema A. Murthy, IIT Madras, India

PAGE 1118
Tue-SS-5-6-7
17:35-17:50

VQVAE Unsupervised Unit Discovery and Multi-Scale Code2Spec Inverter for Zerospeech Challenge 2019

Andros Tjandra¹, Berrak Sisman², Mingyang Zhang², Sakriani Sakti¹, Haizhou Li², Satoshi Nakamura¹

¹NAIST, Japan; ²NUS, Singapore

Tue-SS-5-6-8
17:50-18:00

General Discussion

INTERSPEECH 2019 Table of Contents

Tue-O-3-1 : Speech Translation

Chairs: Chiori Hori, Hermann Ney

Time 10:00-12:00, Tuesday 17 Sept 2019, Main Hall

Tue-O-3-1-1
10:00-10:40
N/A

Survey Talk: A Survey on Speech Translation

Jan Niehues, Universiteit Maastricht, The Netherlands

PAGE 1123
Tue-O-3-1-2
10:40-11:00

Direct Speech-to-Speech Translation with a Sequence-to-Sequence Model

Ye Jia, Ron J. Weiss, Fadi Biadsy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu, Google, USA

PAGE 1128
Tue-O-3-1-3
11:00-11:20

End-to-End Speech Translation with Knowledge Distillation

Yuchen Liu¹, Hao Xiong², Jiajun Zhang¹, Zhongjun He², Hua Wu², Haifeng Wang², Chengqing Zong¹

¹Chinese Academy of Sciences, China; ²Baidu, China

PAGE 1133
Tue-O-3-1-4
11:20-11:40

Adapting Transformer to End-to-End Spoken Language Translation

Mattia A. Di Gangi¹, Matteo Negri², Marco Turchi²

¹Università di Trento, Italy; ²FBK, Italy

PAGE 1138
Tue-O-3-1-5
11:40-12:00

Unsupervised Phonetic and Word Level Discovery for Speech to Speech Translation for Unwritten Languages

Steven Hillis, Anushree Prasanna Kumar, Alan W. Black, Carnegie Mellon University, USA

Tue-O-3-2 : Speaker Recognition 1

Chairs: Kong Aik Lee, Oldřich Plchot

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 1

- PAGE 1143
Tue-O-3-2-1
10:00-10:20
- Deep Speaker Recognition: Modular or Monolithic?**
Gautam Bhattacharya¹, Jahangir Alam², Patrick Kenny²
¹McGill University, Canada; ²CRIM, Canada
- PAGE 1148
Tue-O-3-2-2
10:20-10:40
- On the Usage of Phonetic Information for Text-Independent Speaker Embedding Extraction**
Shuai Wang¹, Johan Rohdin², Lukáš Burget², Oldřich Plchot², Yanmin Qian¹, Kai Yu¹, Jan Černocký²
¹Shanghai Jiao Tong University, China; ²Brno University of Technology, Czech Republic
- PAGE 1153
Tue-O-3-2-3
10:40-11:00
- Learning Speaker Representations with Mutual Information**
Mirco Ravanelli, Yoshua Bengio, Université de Montréal, Canada
- PAGE 1158
Tue-O-3-2-4
11:00-11:20
- Multi-Task Learning with High-Order Statistics for x-Vector Based Text-Independent Speaker Verification**
Lanhua You, Wu Guo, Li-Rong Dai, Jun Du, USTC, China
- PAGE 1163
Tue-O-3-2-5
11:20-11:40
- Data Augmentation Using Variational Autoencoder for Embedding Based Speaker Verification**
Zhanghao Wu, Shuai Wang, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China
- PAGE 1168
Tue-O-3-2-6
11:40-12:00
- Deep Neural Network Embeddings with Gating Mechanisms for Text-Independent Speaker Verification**
Lanhua You, Wu Guo, Li-Rong Dai, Jun Du, USTC, China

Tue-O-3-3 : Dialogue Understanding

Chairs: Yannick Estève, David Griol

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 2

- PAGE 1173
Tue-O-3-3-1
10:00-10:20
- Neural Transition Systems for Modeling Hierarchical Semantic Representations**
Riyaz Bhat, John Chen, Rashmi Prasad, Srinivas Bangalore, Interactions, USA
- PAGE 1178
Tue-O-3-3-2
10:20-10:40
- Mining Polysemous Triplets with Recurrent Neural Networks for Spoken Language Understanding**
Vedran Vukotić¹, Christian Raymond²
¹LAMARK, France; ²IRISA (UMR 6074), France
- PAGE 1183
Tue-O-3-3-3
10:40-11:00
- Iterative Delexicalization for Improved Spoken Language Understanding**
Avik Ray, Yilin Shen, Hongxia Jin, Samsung, USA
- PAGE 1188
Tue-O-3-3-4
11:00-11:20
- End-to-End Spoken Language Understanding: Bootstrapping in Low Resource Scenarios**
Swapnil Bhosale, Imran Sheikh, Sri Harsha Dumpala, Sunil Kumar Kopparapu, TCS Innovation Labs Mumbai, India
- PAGE 1193
Tue-O-3-3-5
11:20-11:40
- Recognition of Intentions of Users' Short Responses for Conversational News Delivery System**
Hiroaki Takatsu¹, Katsuya Yokoyama¹, Yoichi Matsuyama¹, Hiroshi Honda², Shinya Fujie¹, Tetsunori Kobayashi¹
¹Waseda University, Japan; ²Honda, Japan
- PAGE 1198
Tue-O-3-3-6
11:40-12:00
- Curriculum-Based Transfer Learning for an Effective End-to-End Spoken Language Understanding and Domain Portability**
Antoine Caubrière¹, Natalia Tomashenko², Antoine Laurent¹, Emmanuel Morin³, Nathalie Camelin¹, Yannick Estève²
¹LIUM (EA 4023), France; ²LIA (EA 4128), France; ³LS2N (UMR 6004), France

Tue-O-3-4: Speech in the Brain

Chairs: Ann Bradlow, Okko Räsänen

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 11

- PAGE 1203
Tue-O-3-4-1
10:00-10:20
- Spatial and Spectral Fingerprint in the Brain: Speaker Identification from Single Trial MEG Signals**
Debadatta Dash¹, Paul Ferrari², Jun Wang¹
¹University of Texas at Dallas, USA; ²University of Texas at Austin, USA
- PAGE 1208
Tue-O-3-4-2
10:20-10:40
- ERP Signal Analysis with Temporal Resolution Using a Time Window Bank**
Annika Nijveld, L. ten Bosch, Mirjam Ernestus, Radboud Universiteit Nijmegen, The Netherlands
- PAGE 1213
Tue-O-3-4-3
10:40-11:00
- Phase Synchronization Between EEG Signals as a Function of Differences Between Stimuli Characteristics**
L. ten Bosch, K. Mulder, L. Boves, Radboud Universiteit Nijmegen, The Netherlands
- PAGE 1218
Tue-O-3-4-4
11:00-11:20
- The Processing of Prosodic Cues to Rhetorical Question Interpretation: Psycholinguistic and Neurolinguistics Evidence**
Mariya Kharaman, Manluolan Xu, Carsten Eulitz, Bettina Braun, Universität Konstanz, Germany
- PAGE 1223
Tue-O-3-4-5
11:20-11:40
- The Neural Correlates Underlying Lexically-Guided Perceptual Learning**
Odette Scharenborg¹, Jiska Koemans², Cybelle Smith³, Mark A. Hasegawa-Johnson³, Kara D. Federmeier³
¹Technische Universiteit Delft, The Netherlands; ²Radboud Universiteit Nijmegen, The Netherlands; ³University of Illinois at Urbana-Champaign, USA
- PAGE 1228
Tue-O-3-4-6
11:40-12:00
- Speech Quality Evaluation of Synthesized Japanese Speech Using EEG**
Ivan Halim Parmonangan¹, Hiroki Tanaka¹, Sakriani Sakti¹, Shinnosuke Takamichi², Satoshi Nakamura¹
¹NAIST, Japan; ²University of Tokyo, Japan

Tue-O-3-5: Far-Field Speech Recognition

Chairs: Sanjeev Khudanpur, Michael Seltzer

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 12

- PAGE 1233
Tue-O-3-5-1
10:00-10:20
- Multi-Microphone Adaptive Noise Cancellation for Robust Hotword Detection**
Yiteng Huang, Turaj Z. Shabestary, Alexander Gruenstein, Li Wan, Google, USA
- PAGE 1238
Tue-O-3-5-2
10:20-10:40
- Multi-Task Multi-Network Joint-Learning of Deep Residual Networks and Cycle-Consistency Generative Adversarial Networks for Robust Speech Recognition**
Shengkui Zhao, Chongjia Ni, Rong Tong, Bin Ma, Alibaba Group, Singapore
- PAGE 1243
Tue-O-3-5-3
10:40-11:00
- R-Vectors: New Technique for Adaptation to Room Acoustics**
Yuri Khokhlov¹, Alexander Zatvornitskiy², Ivan Medennikov¹, Ivan Sorokin¹, Tatiana Prisyach¹, Aleksei Romanenko¹, Anton Mitrofanov¹, Vladimir Bataev¹, Andrei Andrusenko¹, Mariya Korenevskaya¹, Oleg Petrov³
¹STC-innovations, Russia; ²Speech Technology Center, Russia; ³ITMO University, Russia
- PAGE 1248
Tue-O-3-5-4
11:00-11:20
- Guided Source Separation Meets a Strong ASR Backend: Hitachi/Paderborn University Joint Investigation for Dinner Party ASR**
Naoyuki Kanda¹, Christoph Boeddeker², Jens Heitkaemper², Yusuke Fujita¹, Shota Horiguchi¹, Kenji Nagamatsu¹, Reinhold Haeb-Umbach²
¹Hitachi, Japan; ²Universität Paderborn, Germany
- PAGE 1253
Tue-O-3-5-5
11:20-11:40
- Unsupervised Training of Neural Mask-Based Beamforming**
Lukas Drude, Jahn Heymann, Reinhold Haeb-Umbach, Universität Paderborn, Germany
- PAGE 1258
Tue-O-3-5-6
11:40-12:00
- Acoustic Model Ensembling Using Effective Data Augmentation for CHiME-5 Challenge**
Feng Ma¹, Li Chai¹, Jun Du¹, Diyuan Liu², Zhongfu Ye¹, Chin-Hui Lee³
¹USTC, China; ²iFlytek, China; ³Georgia Tech, USA

Tue-O-4-1 : Speaker and Language Recognition 1

Chairs: Haizhou Li, Md Jahangir Alam

Time 13:30-15:30, Tuesday 17 Sept 2019, Main Hall

- Tue-O-4-1-1
13:30-14:10
N/A
- Survey Talk: End-to-End Deep Neural Network Based Speaker and Language Recognition**
Ming Li, Weicheng Cai, Danwei Cai, Duke Kunshan University, China
- PAGE 1263
Tue-O-4-1-2
14:10-14:30
- Attention Based Hybrid i-Vector BLSTM Model for Language Recognition**
Bharat Padi¹, Anand Mohan², Sriram Ganapathy²
¹minds.ai, India; ²Indian Institute of Science, India
- PAGE 1268
Tue-O-4-1-3
14:30-14:50
- RawNet: Advanced End-to-End Deep Neural Network Using Raw Waveforms for Text-Independent Speaker Verification**
Jee-weon Jung, Hee-Soo Heo, Ju-ho Kim, Hye-jin Shim, Ha-Jin Yu, University of Seoul, Korea
- PAGE 1273
Tue-O-4-1-4
14:50-15:10
- Target Speaker Extraction for Multi-Talker Speaker Verification**
Wei Rao¹, Chenglin Xu², Eng Siong Chng², Haizhou Li¹
¹NUS, Singapore; ²NTU, Singapore
- PAGE 1278
Tue-O-4-1-5
15:10-15:30
- Improving Keyword Spotting and Language Identification via Neural Architecture Search at Scale**
Hanna Mazzawi, Xavi Gonzalvo, Aleks Kracun, Prashant Sridhar, Niranjan Subrahmanya, Ignacio Lopez Moreno, Hyun Jin Park, Patrick Violette, Google, USA

Tue-O-4-2 : Speech Synthesis: Towards End-to-End

Chairs: Esther Klabbbers, Rob Clark

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 1

- PAGE 1283
Tue-O-4-2-1
13:30-13:50
- Forward-Backward Decoding for Regularizing End-to-End TTS**
Yibin Zheng¹, Xi Wang², Lei He², Shifeng Pan², Frank K. Soong², Zhengqi Wen¹, Jianhua Tao¹
¹Chinese Academy of Sciences, China; ²Microsoft, China
- PAGE 1288
Tue-O-4-2-2
13:50-14:10
- A New GAN-Based End-to-End TTS Training Algorithm**
Haohan Guo¹, Frank K. Soong², Lei He², Lei Xie¹
¹Northwestern Polytechnical University, China; ²Microsoft, China
- PAGE 1293
Tue-O-4-2-3
14:10-14:30
- Robust Sequence-to-Sequence Acoustic Modeling with Stepwise Monotonic Attention for Neural TTS**
Mutian He¹, Yan Deng², Lei He²
¹Beihang University, China; ²Microsoft, China
- PAGE 1298
Tue-O-4-2-4
14:30-14:50
- Joint Training Framework for Text-to-Speech and Voice Conversion Using Multi-Source Tacotron and WaveNet**
Mingyang Zhang¹, Xin Wang², Fuming Fang², Haizhou Li¹, Junichi Yamagishi²
¹NUS, Singapore; ²NII, Japan
- PAGE 1303
Tue-O-4-2-5
14:50-15:10
- Training Multi-Speaker Neural Text-to-Speech Systems Using Speaker-Imbalanced Speech Corpora**
Hieu-Thi Luong¹, Xin Wang¹, Junichi Yamagishi¹, Nobuyuki Nishizawa²
¹NII, Japan; ²KDDI Research, Japan
- PAGE 1308
Tue-O-4-2-6
15:10-15:30
- Real-Time Neural Text-to-Speech with Sequence-to-Sequence Acoustic Model and WaveGlow or Single Gaussian WaveRNN Vocoders**
Takuma Okamoto¹, Tomoki Toda², Yoshinori Shiga¹, Hisashi Kawai¹
¹NICT, Japan; ²Nagoya University, Japan

Tue-O-4-3: Semantic Analysis and Classification

Chairs: Dilek Hakkani-Tür, Mari Ostendorf

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 2

- PAGE 1313
Tue-O-4-3-1
13:30-13:50 **Fusion Strategy for Prosodic and Lexical Representations of Word Importance**
Sushant Kafle, Cecilia Ovesdotter Alm, Matt Huenerfauth, RIT, USA
- PAGE 1318
Tue-O-4-3-2
13:50-14:10 **Self Attention in Variational Sequential Learning for Summarization**
Jen-Tzung Chien, Chun-Wei Wang, National Chiao Tung University, Taiwan
- PAGE 1323
Tue-O-4-3-3
14:10-14:30 **Multi-Modal Sentiment Analysis Using Deep Canonical Correlation Analysis**
Zhongkai Sun¹, Prathusha K. Sarma¹, William Sethares¹, Erik P. Bucy²
¹UW-Madison, USA; ²Texas Tech University, USA
- PAGE 1328
Tue-O-4-3-4
14:30-14:50 **Interpreting and Improving Deep Neural SLU Models via Vocabulary Importance**
Yilin Shen¹, Wenhu Chen², Hongxia Jin¹
¹Samsung, USA; ²University of California at Santa Barbara, USA
- PAGE 1333
Tue-O-4-3-5
14:50-15:10 **Assessing the Semantic Space Bias Caused by ASR Error Propagation and its Effect on Spoken Document Summarization**
Máté Ákos Tündik, Valér Kaszás, György Szaszák, BME, Hungary
- PAGE 1338
Tue-O-4-3-6
15:10-15:30 **Latent Topic Attention for Domain Classification**
Peisong Huang, Peijie Huang, Wencheng Ai, Jiande Ding, Jinchuan Zhang, SCAU, China

Tue-O-4-5: Speech and Audio Source Separation and Scene Analysis 1

Chairs: Elmar Nöth, Tomohiro Nakatani

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 12

- PAGE 1343
Tue-O-4-5-1
13:30-13:50 **A Unified Bayesian Source Modelling for Determined Blind Source Separation**
Chaitanya Narisetty, NEC, Japan
- PAGE 1348
Tue-O-4-5-2
13:50-14:10 **Recursive Speech Separation for Unknown Number of Speakers**
Naoya Takahashi¹, Sudarsanam Parthasaarathy², Nabarun Goswami², Yuki Mitsufuji¹
¹Sony, Japan; ²Sony, India
- PAGE 1353
Tue-O-4-5-3
14:10-14:30 **Practical Applicability of Deep Neural Networks for Overlapping Speaker Separation**
Pieter Appeltans, Jeroen Zegers, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium
- PAGE 1358
Tue-O-4-5-4
14:30-14:50 **Speech Separation Using Independent Vector Analysis with an Amplitude Variable Gaussian Mixture Model**
Zhaoyi Gu, Jing Lu, Kai Chen, Nanjing University, China
- PAGE 1363
Tue-O-4-5-5
14:50-15:10 **Improved Speech Separation with Time-and-Frequency Cross-Domain Joint Embedding and Clustering**
Gene-Ping Yang, Chao-I Tuan, Hung-Yi Lee, Lin-shan Lee, National Taiwan University, Taiwan
- PAGE 1368
Tue-O-4-5-6
15:10-15:30 **WHAM!: Extending Speech Separation to Noisy Environments**
Gordon Wichern¹, Joe Antognini², Michael Flynn², Licheng Richard Zhu², Emmett McQuinn², Dwight Crow², Ethan Manilow¹, Jonathan Le Roux¹
¹MERL, USA; ²Whisper.ai, USA

Tue-O-5-1 : Speech Intelligibility

Chairs: Martin Cooke, Sebastian Möller

Time 16:00-18:00, Tuesday 17 Sept 2019, Main Hall

- Tue-O-5-1-1
16:00-16:40
N/A
- Survey Talk: Preserving Privacy in Speaker and Speech Characterisation**
Andreas Nautsch, EURECOM, France
- PAGE 1373
Tue-O-5-1-2
16:40-17:00
- Evaluating Near End Listening Enhancement Algorithms in Realistic Environments**
Carol Chermaz¹, Cassia Valentini-Botinhao¹, Henning Schepker², Simon King¹
¹University of Edinburgh, UK; ²Carl von Ossietzky Universität Oldenburg, Germany
- PAGE 1378
Tue-O-5-1-3
17:00-17:20
- Improvement and Assessment of Spectro-Temporal Modulation Analysis for Speech Intelligibility Estimation**
Amin Edraki¹, Wai-Yip Chan¹, Jesper Jensen², Daniel Fogerty³
¹Queen's University, Canada; ²Aalborg University, Denmark; ³University of South Carolina, USA
- PAGE 1383
Tue-O-5-1-4
17:20-17:40
- Listener Preference on the Local Criterion for Ideal Binary-Masked Speech**
Zhuohuang Zhang, Yi Shen, Indiana University, USA
- PAGE 1388
Tue-O-5-1-5
17:40-18:00
- Using a Manifold Vocoder for Spectral Voice and Style Conversion**
Tuan Dinh¹, Alexander Kain¹, Kris Tjaden²
¹Oregon Health & Science University, USA; ²University at Buffalo, USA

Tue-O-5-2 : ASR Neural Network Architectures 1

Chairs: Brian Kingsbury, Renato DeMori

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 1

- PAGE 1393
Tue-O-5-2-1
16:00-16:20
- Multi-Span Acoustic Modelling Using Raw Waveform Signals**
P. von Platen, Chao Zhang, P.C. Woodland, University of Cambridge, UK
- PAGE 1398
Tue-O-5-2-2
16:20-16:40
- An Analysis of Local Monotonic Attention Variants**
André Merboldt, Albert Zeyer, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany
- PAGE 1403
Tue-O-5-2-3
16:40-17:00
- Layer Trajectory BLSTM**
Eric Sun, Jinyu Li, Yifan Gong, Microsoft, USA
- PAGE 1408
Tue-O-5-2-4
17:00-17:20
- Improving Transformer-Based End-to-End Speech Recognition with Connectionist Temporal Classification and Language Model Integration**
Shigeki Karita¹, Nelson Enrique Yalta Soplín², Shinji Watanabe³, Marc Delcroix¹, Atsunori Ogawa¹, Tomohiro Nakatani¹
¹NTT, Japan; ²Waseda University, Japan; ³Johns Hopkins University, USA
- PAGE 1413
Tue-O-5-2-5
17:20-17:40
- Trainable Dynamic Subsampling for End-to-End Speech Recognition**
Shucong Zhang, Erfan Loweimi, Yumo Xu, Peter Bell, Steve Renals, University of Edinburgh, UK
- PAGE 1418
Tue-O-5-2-6
17:40-18:00
- Shallow-Fusion End-to-End Contextual Biasing**
Ding Zhao, Tara N. Sainath, David Rybach, Pat Rondon, Deepti Bhatia, Bo Li, Ruoming Pang, Google, USA

Tue-O-5-3 : Speech and Language Analytics for Mental Health

Chairs: Carol Espy-Wilson, Khiet Truong

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 2

PAGE 1423
Tue-O-5-3-1
16:00-16:20

Modeling Interpersonal Linguistic Coordination in Conversations Using Word Mover's Distance

Md. Nasir¹, Sandeep Nallan Chakravarthula¹, Brian R.W. Baucom², David C. Atkins³, Panayiotis Georgiou¹, Shrikanth Narayanan¹

¹University of Southern California, USA; ²University of Utah, USA; ³University of Washington, USA

PAGE 1428
Tue-O-5-3-2
16:20-16:40

Bag-of-Acoustic-Words for Mental Health Assessment: A Deep Autoencoding Approach

Wenchao Du¹, Louis-Philippe Morency¹, Jeffrey Cohn², Alan W. Black¹

¹Carnegie Mellon University, USA; ²University of Pittsburgh, USA

PAGE 1433
Tue-O-5-3-3
16:40-17:00

Objective Assessment of Social Skills Using Automated Language Analysis for Identification of Schizophrenia and Bipolar Disorder

Rohit Voleti¹, Stephanie Woolridge², Julie M. Liss¹, Melissa Milanovic², Christopher R. Bowie², Visar Berisha¹

¹Arizona State University, USA; ²Queen's University, Canada

PAGE 1438
Tue-O-5-3-4
17:00-17:20

Into the Wild: Transitioning from Recognizing Mood in Clinical Interactions to Personal Conversations for Individuals with Bipolar Disorder

Katie Matton, Melvin G. McInnis, Emily Mower Provost, University of Michigan, USA

PAGE 1443
Tue-O-5-3-5
17:20-17:40

Detecting Depression with Word-Level Multimodal Fusion

Morteza Rohanian, Julian Hough, Matthew Purver, Queen Mary University of London, UK

Tue-O-5-3 continued...

PAGE 1448
Tue-O-5-3-6
17:40-18:00

Assessing Neuromotor Coordination in Depression Using Inverted Vocal Tract Variables

Carol Espy-Wilson¹, Adam C. Lammert², Nadee Seneviratne¹, Thomas F. Quatieri²

¹University of Maryland at College Park, USA; ²MIT Lincoln Laboratory, USA

Tue-O-5-4: Dialogue Modelling

Chairs: Frederic Bechet, Giuseppe Riccardi

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 11

- PAGE 1453
Tue-O-5-4-1
16:00-16:20 **Towards Universal Dialogue Act Tagging for Task-Oriented Dialogues**
Shachi Paul, Rahul Goel, Dilek Hakkani-Tür, Amazon, USA
- PAGE 1458
Tue-O-5-4-2
16:20-16:40 **HyST: A Hybrid Approach for Flexible and Accurate Dialogue State Tracking**
Rahul Goel, Shachi Paul, Dilek Hakkani-Tür, Amazon, USA
- PAGE 1463
Tue-O-5-4-3
16:40-17:00 **Multi-Lingual Dialogue Act Recognition with Deep Learning Methods**
Jiří Martínek¹, Pavel Král¹, Ladislav Lenc¹, Christophe Cerisara²
¹University of West Bohemia, Czech Republic; ²Loria (UMR 7503), France
- PAGE 1468
Tue-O-5-4-4
17:00-17:20 **BERT-DST: Scalable End-to-End Dialogue State Tracking with Bidirectional Encoder Representations from Transformer**
Guan-Lin Chao, Ian Lane, Carnegie Mellon University, USA
- PAGE 1473
Tue-O-5-4-5
17:20-17:40 **Discovering Dialog Rules by Means of an Evolutionary Approach**
David Griol¹, Zoraida Callejas²
¹Universidad Carlos III de Madrid, Spain; ²Universidad de Granada, Spain
- PAGE 1478
Tue-O-5-4-6
17:40-18:00 **Active Learning for Domain Classification in a Commercial Spoken Personal Assistant**
Xi C. Chen, Adithya Sagar, Justine T. Kao, Tony Y. Li, Christopher Klein, Stephen Pulman, Ashish Garg, Jason D. Williams, Apple, USA

Tue-O-5-5: Speaker Recognition Evaluation

Chairs: Tomi Kinnunen, Lukáš Burget

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 12

- PAGE 1483
Tue-O-5-5-1
17:40-18:00 **The 2018 NIST Speaker Recognition Evaluation**
Seyed Omid Sadjadi¹, Craig Greenberg¹, Elliot Singer², Douglas Reynolds², Lisa Mason³, Jaime Hernandez-Cordero³
¹NIST, USA; ²MIT Lincoln Laboratory, USA; ³DoD, USA
- PAGE 1488
Tue-O-5-5-2
16:20-16:40 **State-of-the-Art Speaker Recognition for Telephone and Video Speech: The JHU-MIT Submission for NIST SRE18**
Jesús Villalba¹, Nanxin Chen¹, David Snyder¹, Daniel Garcia-Romero¹, Alan McCree¹, Gregory Sell¹, Jonas Borgstrom², Fred Richardson², Suwon Shon³, François Grondin³, Réda Dehak⁴, Leibny Paola García-Perera¹, Daniel Povey¹, Pedro A. Torres-Carrasquillo², Sanjeev Khudanpur¹, Najim Dehak¹
¹Johns Hopkins University, USA; ²MIT Lincoln Laboratory, USA; ³MIT, USA; ⁴EPITA-LSE, France
- PAGE 1493
Tue-O-5-5-3
16:40-17:00 **x-Vector DNN Refinement with Full-Length Recordings for Speaker Recognition**
Daniel Garcia-Romero, David Snyder, Gregory Sell, Alan McCree, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

Tue-O-5-5 continued...

PAGE 1497
Tue-O-5-5-4
17:20-17:40

I4U Submission to NIST SRE 2018: Leveraging from a Decade of Shared Experiences

Kong Aik Lee¹, Ville Hautamäki², Tomi H. Kinnunen², Hitoshi Yamamoto¹, Koji Okabe¹, Ville Vestman¹, Jing Huang³, Guohong Ding⁴, Hanwu Sun⁵, Anthony Larcher⁶, Rohan Kumar Das⁷, Haizhou Li⁷, Mickael Rouvier⁸, Pierre-Michel Bousquet⁸, Wei Rao⁹, Qing Wang¹⁰, Chunlei Zhang¹¹, Fahimeh Bahmaninezhad¹¹, Héctor Delgado¹², Massimiliano Todisco¹²
*¹NEC, Japan; ²University of Eastern Finland, Finland; ³JD.com, USA; ⁴JD.com, China; ⁵A*STAR, Singapore; ⁶LIUM (EA 4023), France; ⁷NUS, Singapore; ⁸LIA (EA 4128), France; ⁹NTU, Singapore; ¹⁰Northwestern Polytechnical University, China; ¹¹University of Texas at Dallas, USA; ¹²EURECOM, France*

PAGE 1502
Tue-O-5-5-5
16:00-16:20

Pindrop Labs' Submission to the First Multi-Target Speaker Detection and Identification Challenge

Elie Houry, Khaled Lakhdar, Andrew Vaughan, Ganesh Sivaraman, Parav Nagarsheth, Pindrop, USA

PAGE 1506
Tue-O-5-5-6
17:00-17:20

Speaker Recognition Benchmark Using the CHiME-5 Corpus

Daniel Garcia-Romero, David Snyder, Shinji Watanabe, Gregory Sell, Alan McCree, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

Tue-P-3-A: Speech Synthesis: Data and Evaluation

Chair: Cassia Valentini-Botinhao

Time 10:00-12:00, Tuesday 17 Sept 2019, Gallery A

PAGE 1511
Tue-P-3-A-1
10:00-12:00

Investigating the Effects of Noisy and Reverberant Speech in Text-to-Speech Systems

David Ayllón, Héctor A. Sánchez-Hevia, Carol Figueroa, Pierre Lanchantin, ObEN, USA

PAGE 1516
Tue-P-3-A-2
10:00-12:00

Selection and Training Schemes for Improving TTS Voice Built on Found Data

F.-Y. Kuo, L.C. Ouyang, S. Aryal, Pierre Lanchantin, ObEN, USA

PAGE 1521
Tue-P-3-A-3
10:00-12:00

All Together Now: The Living Audio Dataset

David A. Braude¹, Matthew P. Aylett¹, Caoimhín Laoide-Kemp¹, Simone Ashby², Kristen M. Scott², Brian Ó Raghallaigh³, Anna Braudo⁴, Alex Brouwer⁴, Adriana Stan⁵
¹CereProc, UK; ²Universidade da Madeira, Portugal; ³Dublin City University, Ireland; ⁴University of Edinburgh, UK; ⁵Technical University of Cluj-Napoca, Romania

PAGE 1526
Tue-P-3-A-4
10:00-12:00

LibriTTS: A Corpus Derived from LibriSpeech for Text-to-Speech

Heiga Zen¹, Viet Dang², Rob Clark², Yu Zhang³, Ron J. Weiss³, Ye Jia³, Zhifeng Chen³, Yonghui Wu³
¹Google, Japan; ²Google, UK; ³Google, USA

PAGE 1531
Tue-P-3-A-5
10:00-12:00

Corpus Design Using Convolutional Auto-Encoder Embeddings for Audio-Book Synthesis

Meysam Shamsi, Damien Lolive, Nelly Barbot, Jonathan Chevelu, IRISA (UMR 6074), France

PAGE 1536
Tue-P-3-A-6
10:00-12:00

Evaluating Intention Communication by TTS Using Explicit Definitions of Illocutionary Act Performance

Nobukatsu Hojo, Noboru Miyazaki, NTT, Japan

Tue-P-3-A continued...

- PAGE 1541
Tue-P-3-A-7
10:00-12:00
MOSNet: Deep Learning-Based Objective Assessment for Voice Conversion
Chen-Chou Lo¹, Szu-Wei Fu¹, Wen-Chin Huang¹, Xin Wang², Junichi Yamagishi², Yu Tsao¹, Hsin-Min Wang¹
¹Academia Sinica, Taiwan; ²NII, Japan
- PAGE 1546
Tue-P-3-A-8
10:00-12:00
Investigating the Robustness of Sequence-to-Sequence Text-to-Speech Models to Imperfectly-Transcribed Training Data
Jason Fong, Pilar Oplustil Gallegos, Zack Hodari, Simon King, University of Edinburgh, UK
- PAGE 1551
Tue-P-3-A-9
10:00-12:00
Using Pupil Dilation to Measure Cognitive Load When Listening to Text-to-Speech in Quiet and in Noise
Avashna Govender¹, Anita E. Wagner², Simon King¹
¹University of Edinburgh, UK; ²Rijksuniversiteit Groningen, The Netherlands
- PAGE 1556
Tue-P-3-A-10
10:00-12:00
A Multimodal Real-Time MRI Articulatory Corpus of French for Speech Research
Ioannis K. Douros¹, Jacques Felblinger², Jens Frahm³, Karyna Isaieva², Arun A. Joseph³, Yves Laprie¹, Freddy Odille², Anastasiia Tsukanova¹, Dirk Voit³, Pierre-André Vuissoz²
¹Loria (UMR 7503), France; ²IADI (Inserm U1254), France; ³MPI for Biophysical Chemistry, Germany
- PAGE 1561
Tue-P-3-A-11
10:00-12:00
A Chinese Dataset for Identifying Speakers in Novels
Jia-Xiang Chen, Zhen-Hua Ling, Li-Rong Dai, USTC, China
- PAGE 1566
Tue-P-3-A-12
10:00-12:00
CSS10: A Collection of Single Speaker Speech Datasets for 10 Languages
Kyubyong Park¹, Thomas Mulc²
¹Kakao Brain, Korea; ²Expedia Group, USA

Tue-P-3-B: Model Training for ASR

Chair: Alberto Abad

Time 10:00-12:00, Tuesday 17 Sept 2019, Gallery B

- PAGE 1571
Tue-P-3-B-1
10:00-12:00
Attention Model for Articulatory Features Detection
Ievgen Karaulov, Dmytro Tkanov, SciForce, Ukraine
- PAGE 1576
Tue-P-3-B-2
10:00-12:00
Unbiased Semi-Supervised LF-MMI Training Using Dropout
Sibo Tong, Apoorv Vyas, Philip N. Garner, Hervé Broulard, Idiap Research Institute, Switzerland
- PAGE 1581
Tue-P-3-B-3
10:00-12:00
Acoustic Model Optimization Based on Evolutionary Stochastic Gradient Descent with Anchors for Automatic Speech Recognition
Xiaodong Cui, Michael Picheny, IBM, USA
- PAGE 1586
Tue-P-3-B-4
10:00-12:00
Whether to Pretrain DNN or not?: An Empirical Analysis for Voice Conversion
Nirmesh J. Shah¹, Hardik B. Sailor², Hemant A. Patil¹
¹DA-IICT, India; ²University of Sheffield, UK
- PAGE 1591
Tue-P-3-B-5
10:00-12:00
Detection of Glottal Closure Instants from Raw Speech Using Convolutional Neural Networks
Mohit Goyal, Varun Srivastava, Prathosh A. P., IIT Delhi, India
- PAGE 1596
Tue-P-3-B-6
10:00-12:00
Lattice-Based Lightly-Supervised Acoustic Model Training
Joachim Fainberg, Ondřej Klejch, Steve Renals, Peter Bell, University of Edinburgh, UK
- PAGE 1601
Tue-P-3-B-7
10:00-12:00
Comparison of Lattice-Free and Lattice-Based Sequence Discriminative Training Criteria for LVCSR
Wilfried Michel, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany
- PAGE 1606
Tue-P-3-B-8
10:00-12:00
End-to-End Automatic Speech Recognition with a Reconstruction Criterion Using Speech-to-Text and Text-to-Speech Encoder-Decoders
Ryo Masumura, Hiroshi Sato, Tomohiro Tanaka, Takafumi Moriya, Yusuke Ijima, Takano Oba, NTT, Japan

INTERSPEECH 2019 Table of Contents

Tue-P-3-B continued...

- PAGE 1611
Tue-P-3-B-9
10:00-12:00
Char+CV-CTC: Combining Graphemes and Consonant/Vowel Units for CTC-Based ASR Using Multitask Learning
Abdelwahab Heba¹, Thomas Pellegrini¹, Jean-Pierre Lorré², Régine Andre-Obrecht¹
¹IRIT (UMR 5505), France; ²LINAGORA, France
- PAGE 1616
Tue-P-3-B-10
10:00-12:00
Guiding CTC Posterior Spike Timings for Improved Posterior Fusion and Knowledge Distillation
Gakuto Kurata¹, Kartik Audhkhasi²
¹IBM, Japan; ²IBM, USA
- PAGE 1621
Tue-P-3-B-11
10:00-12:00
Direct Neuron-Wise Fusion of Cognate Neural Networks
Takashi Fukuda, Masayuki Suzuki, Gakuto Kurata, IBM, Japan
- PAGE 1626
Tue-P-3-B-12
10:00-12:00
Two Tiered Distributed Training Algorithm for Acoustic Modeling
Pranav Ladkat, Oleg Rybakov, Radhika Arava, Sree Hari Krishnan Parthasarathi, I-Fan Chen, Nikko Strom, Amazon, USA
- PAGE 1631
Tue-P-3-B-13
10:00-12:00
Exploring the Encoder Layers of Discriminative Autoencoders for LVCSR
Pin-Tuan Huang¹, Hung-Shin Lee¹, Syu-Siang Wang¹, Kuan-Yu Chen², Yu Tsao¹, Hsin-Min Wang¹
¹Academia Sinica, Taiwan; ²Taiwan Tech, Taiwan
- PAGE 1636
Tue-P-3-B-14
10:00-12:00
Multi-Task CTC Training with Auxiliary Feature Reconstruction for End-to-End Speech Recognition
Gakuto Kurata¹, Kartik Audhkhasi²
¹IBM, Japan; ²IBM, USA
- PAGE 1641
Tue-P-3-B-15
10:00-12:00
Framework Supervised Training Towards End-to-End Speech Recognition Models: First Results
Mohan Li, Yuanjiang Cao, Weicong Zhou, Min Liu, Toshiba China R&D Center, China

INTERSPEECH 2019 Table of Contents

Tue-P-3-C: Network Architectures for Emotion and Paralinguistics Recognition

Chair: Emily Mower Provost

Time 10:00-12:00, Tuesday 17 Sept 2019, Gallery C

- PAGE 1646
Tue-P-3-C-1
10:00-12:00
Deep Hierarchical Fusion with Application in Sentiment Analysis
Efthymios Georgiou, Charilaos Papaioannou, Alexandros Potamianos, NTUA, Greece
- PAGE 1651
Tue-P-3-C-2
10:00-12:00
Leveraging Acoustic Cues and Paralinguistic Embeddings to Detect Expression from Voice
Vikramjit Mitra, Sue Booker, Erik Marchi, David Scott Farrar, Ute Dorothea Peitz, Bridget Cheng, Ermine Teves, Anuj Mehta, Devang Naik, Apple, USA
- PAGE 1656
Tue-P-3-C-3
10:00-12:00
Analysis of Deep Learning Architectures for Cross-Corpus Speech Emotion Recognition
Jack Parry, Dimitri Palaz, Georgia Clarke, Pauline Lecomte, Rebecca Mead, Michael Berger, Gregor Hofer, Speech Graphics, UK
- PAGE 1661
Tue-P-3-C-4
10:00-12:00
A Path Signature Approach for Speech Emotion Recognition
Bo Wang¹, Maria Liakata², Hao Ni², Terry Lyons¹, Alejo J. Nevado-Holgado¹, Kate Saunders¹
¹University of Oxford, UK; ²Alan Turing Institute, UK
- PAGE 1666
Tue-P-3-C-5
10:00-12:00
Employing Bottleneck and Convolutional Features for Speech-Based Physical Load Detection on Limited Data Amounts
Olga Egorow¹, Tarik Mrech², Norman Weißkirchen¹, Andreas Wendemuth¹
¹Otto-von-Guericke-Universität Magdeburg, Germany; ²Fraunhofer IFF, Germany
- PAGE 1671
Tue-P-3-C-6
10:00-12:00
Speech Emotion Recognition in Dyadic Dialogues with Attentive Interaction Modeling
Jinming Zhao, Shizhe Chen, Jingjun Liang, Qin Jin, Renmin University of China, China

Tue-P-3-C continued...

PAGE 1676
Tue-P-3-C-7
10:00-12:00

Predicting Group Performances Using a Personality Composite-Network Architecture During Collaborative Task

Shun-Chang Zhong¹, Yun-Shao Lin¹, Chun-Min Chang¹, Yi-Ching Liu², Chi-Chun Lee¹

¹National Tsing Hua University, Taiwan; ²National Taiwan University, Taiwan

PAGE 1681
Tue-P-3-C-8
10:00-12:00

Enforcing Semantic Consistency for Cross Corpus Valence Regression from Speech Using Adversarial Discrepancy Learning

Gao-Yi Chao, Yun-Shao Lin, Chun-Min Chang, Chi-Chun Lee, National Tsing Hua University, Taiwan

PAGE 1686
Tue-P-3-C-9
10:00-12:00

Deep Learning of Segment-Level Feature Representation with Multiple Instance Learning for Utterance-Level Speech Emotion Recognition

Shuiyang Mao, P.C. Ching, Tan Lee, CUHK, China

PAGE 1691
Tue-P-3-C-10
10:00-12:00

Towards Robust Speech Emotion Recognition Using Deep Residual Networks for Speech Enhancement

Andreas Triantafyllopoulos¹, Gil Keren², Johannes Wagner¹, Ingmar Steiner¹, Björn W. Schuller¹

¹audEERING, Germany; ²Universität Augsburg, Germany

PAGE 1696
Tue-P-3-C-11
10:00-12:00

Towards Discriminative Representations and Unbiased Predictions: Class-Specific Angular Softmax for Speech Emotion Recognition

Zhixuan Li¹, Liang He¹, Jingyang Li², Li Wang², Wei-Qiang Zhang¹

¹Tsinghua University, China; ²Ministry of Public Security, China

PAGE 1701
Tue-P-3-C-12
10:00-12:00

Learning Temporal Clusters Using Capsule Routing for Speech Emotion Recognition

Md. Asif Jalal¹, Erfan Loweimi², Roger K. Moore¹, Thomas Hain¹

¹University of Sheffield, UK; ²University of Edinburgh, UK

Tue-P-3-D: Acoustic Phonetics

Chair: Michele Gubian

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 10/D

PAGE 1706
Tue-P-3-D-1
10:00-12:00

L2 Pronunciation Accuracy and Context: A Pilot Study on the Realization of Geminate in Italian as L2 by French Learners

Sonia d'Apolito, Barbara Gili Fivela, Università del Salento, Italy

PAGE 1711
Tue-P-3-D-2
10:00-12:00

The Monophthongs of Formal Nigerian English: An Acoustic Analysis

Nisad Jamakovic¹, Robert Fuchs²

¹WWU Münster, Germany; ²Universität Hamburg, Germany

PAGE 1716
Tue-P-3-D-3
10:00-12:00

Quantifying Fundamental Frequency Modulation as a Function of Language, Speaking Style and Speaker

Pablo Arantes¹, Anders Eriksson²

¹Universidade Federal de São Paulo, Brazil; ²Stockholm University, Sweden

PAGE 1721
Tue-P-3-D-4
10:00-12:00

The Voicing Contrast in Stops and Affricates in the Western Armenian of Lebanon

Niamh E. Kelly, Lara Keshishian, AUB, Lebanon

PAGE 1726
Tue-P-3-D-5
10:00-12:00

“Gra[f]e!” Word-Final Devoicing of Obstruents in Standard French: An Acoustic Study Based on Large Corpora

Adèle Jatteau¹, Ioana Vasilescu¹, Lori Lamel¹, Martine Adda-Decker¹, Nicolas Audibert²

¹LMSI (UPR 3251), France; ²LPP (UMR 7018), France

PAGE 1731
Tue-P-3-D-6
10:00-12:00

Acoustic Indicators of Deception in Mandarin Daily Conversations Recorded from an Interactive Game

Chih-Hsiang Huang, Huang-Cheng Chou, Yi-Tong Wu, Chi-Chun Lee, Yi-Wen Liu, National Tsing Hua University, Taiwan

INTERSPEECH 2019 Table of Contents

Tue-P-3-D continued...

PAGE 1736
Tue-P-3-D-7
10:00-12:00

Prosodic Effects on Plosive Duration in German and Austrian German

Barbara Schuppler¹, Margaret Zellers²

¹Technische Universität Graz, Austria; ²Christian-Albrechts-Universität zu Kiel, Germany

PAGE 1741
Tue-P-3-D-8
10:00-12:00

Cross-Lingual Consistency of Phonological Features: An Empirical Study

Cibu Johnny, Alexander Gutkin, Martin Jansche, Google, UK

PAGE 1746
Tue-P-3-D-9
10:00-12:00

Are IP Initial Vowels Acoustically More Distinct? Results from LDA and CNN Classifications

Fanny Guitard-Ivent¹, Gabriele Chignoli¹, Cécile Fougeron¹, Laurianne Georgeton²

¹LPP (UMR 7018), France; ²SCPTS, France

PAGE 1751
Tue-P-3-D-10
10:00-12:00

Neural Network-Based Modeling of Phonetic Durations

Xizi Wei¹, Meilyn Hunt², Adrian Skilling²

¹University of Birmingham, UK; ²Apple, UK

PAGE 1756
Tue-P-3-D-11
10:00-12:00

An Acoustic Study of Vowel Undershoot in a System with Several Degrees of Prominence

Janina Mołczanow, Beata Łukaszewicz, Anna Łukaszewicz, University of Warsaw, Poland

PAGE 1761
Tue-P-3-D-12
10:00-12:00

A Preliminary Study of Charismatic Speech on YouTube: Correlating Prosodic Variation with Counts of Subscribers, Views and Likes

Stephanie Berger¹, Oliver Niebuhr², Margaret Zellers¹

¹Christian-Albrechts-Universität zu Kiel, Germany; ²University of Southern Denmark, Denmark

PAGE 1766
Tue-P-3-D-13
10:00-12:00

Phonetic Detail Encoding in Explaining the Size of Speech Planning Window

Shan Luo, Yangzhou University, China

INTERSPEECH 2019 Table of Contents

Tue-P-3-D continued...

PAGE 1771
Tue-P-3-D-14
10:00-12:00

Acoustic Cues to Topic and Narrow Focus in Egyptian Arabic

Dina El Zarka¹, Barbara Schuppler², Francesco Cangemi³

¹Universität Graz, Austria; ²Technische Universität Graz, Austria; ³Universität zu Köln, Germany

PAGE 1776
Tue-P-3-D-15
10:00-12:00

Acoustic and Articulatory Study of Ewe Vowels: A Comparative Study of Male and Female

Kowovi Comivi Alowonou¹, Jianguo Wei¹, Wenhuan Lu¹, Zhicheng Liu¹, Kiyoshi Honda¹, Jianwu Dang²

¹Tianjin University, China; ²JAIST, Japan

Tue-P-3-E: Speech Enhancement: Noise Attenuation

Chair: Mahadeva Prasanna

Time 10:00-12:00, Tuesday 17 Sept 2019, Hall 10/E

- PAGE 1781
Tue-P-3-E-1
10:00-12:00
- Speech Augmentation via Speaker-Specific Noise in Unseen Environment**
Ya'nan Guo¹, Ziping Zhao², Yide Ma³, Björn W. Schuller¹
¹Universität Augsburg, Germany; ²Tianjin Normal University, China; ³Lanzhou University, China
- PAGE 1786
Tue-P-3-E-2
10:00-12:00
- UNetGAN: A Robust Speech Enhancement Approach in Time Domain for Extremely Low Signal-to-Noise Ratio Condition**
Xiang Hao, Xiangdong Su, Zhiyu Wang, Hui Zhang, Batushiren, Inner Mongolia University, China
- PAGE 1791
Tue-P-3-E-3
10:00-12:00
- Towards Generalized Speech Enhancement with Generative Adversarial Networks**
Santiago Pascual¹, Joan Serra², Antonio Bonafonte¹
¹Universitat Politècnica de Catalunya, Spain; ²Telefónica I+D, Spain
- PAGE 1796
Tue-P-3-E-4
10:00-12:00
- A Convolutional Neural Network with Non-Local Module for Speech Enhancement**
Xiaoqi Li, Yaxing Li, Meng Li, Shan Xu, Yuanjie Dong, Xinrong Sun, Shengwu Xiong, Wuhan University of Technology, China
- PAGE 1801
Tue-P-3-E-5
10:00-12:00
- IA-NET: Acceleration and Compression of Speech Enhancement Using Integer-Adder Deep Neural Network**
Yu-Chen Lin¹, Yi-Te Hsu², Szu-Wei Fu¹, Yu Tsao², Tei-Wei Kuo¹
¹National Taiwan University, Taiwan; ²Academia Sinica, Taiwan
- PAGE 1806
Tue-P-3-E-6
10:00-12:00
- KL-Divergence Regularized Deep Neural Network Adaptation for Low-Resource Speaker-Dependent Speech Enhancement**
Li Chai¹, Jun Du¹, Chin-Hui Lee²
¹USTC, China; ²Georgia Tech, USA

Tue-P-3-E continued...

- PAGE 1811
Tue-P-3-E-7
10:00-12:00
- Speech Enhancement with Wide Residual Networks in Reverberant Environments**
Jorge Llombart, Dayana Ribas, Antonio Miguel, Luis Vicente, Alfonso Ortega, Eduardo Lleida, Universidad de Zaragoza, Spain
- PAGE 1816
Tue-P-3-E-8
10:00-12:00
- A Scalable Noisy Speech Dataset and Online Subjective Test Framework**
Chandan K.A. Reddy, Ebrahim Beyrami, Jamie Pool, Ross Cutler, Sriram Srinivasan, Johannes Gehrke, Microsoft, USA
- PAGE 1821
Tue-P-3-E-9
10:00-12:00
- Speech Enhancement for Noise-Robust Speech Synthesis Using Wasserstein GAN**
Nagaraj Adiga¹, Yannis Pantazis², Vassilis Tsiaras¹, Yannis Stylianou¹
¹University of Crete, Greece; ²FORTH, Greece
- PAGE 1826
Tue-P-3-E-10
10:00-12:00
- A Non-Causal FFTNet Architecture for Speech Enhancement**
Muhammed Shifas P.V., Nagaraj Adiga, Vassilis Tsiaras, Yannis Stylianou, University of Crete, Greece
- PAGE 1831
Tue-P-3-E-11
10:00-12:00
- Speech Enhancement with Variance Constrained Autoencoders**
D.T. Braithwaite, W. Bastiaan Kleijn, Victoria University of Wellington, New Zealand

Tue-P-4-B: Language Learning and Databases

Chair: Febe De Wet

Time 13:30-15:30, Tuesday 17 Sept 2019, Gallery B

- PAGE 1836
Tue-P-4-B-1
13:30-15:30
- A Deep Learning Approach to Automatic Characterisation of Rhythm in Non-Native English Speech**
Konstantinos Kyriakopoulos, Kate M. Knill, Mark J.F. Gales, University of Cambridge, UK
- PAGE 1841
Tue-P-4-B-2
13:30-15:30
- Language Learning Using Speech to Image Retrieval**
Danny Merks, Stefan L. Frank, Mirjam Ernestus, Radboud Universiteit Nijmegen, The Netherlands
- PAGE 1846
Tue-P-4-B-3
13:30-15:30
- Using Alexa for Flashcard-Based Learning**
Lucy Skidmore, Roger K. Moore, University of Sheffield, UK
- PAGE 1851
Tue-P-4-B-4
13:30-15:30
- The 2019 Inaugural Fearless Steps Challenge: A Giant Leap for Naturalistic Audio**
John H.L. Hansen, Aditya Joglekar, Meena Chandra Shekhar, Vinay Kothapally, Chengzhu Yu, Lakshmish Kaushik, Abhijeet Sangwan, University of Texas at Dallas, USA
- PAGE 1856
Tue-P-4-B-5
13:30-15:30
- Completely Unsupervised Phoneme Recognition by a Generative Adversarial Network Harmonized with Iteratively Refined Hidden Markov Models**
Kuan-Yu Chen, Che-Ping Tsai, Da-Rong Liu, Hung-Yi Lee, Lin-shan Lee, National Taiwan University, Taiwan
- PAGE 1861
Tue-P-4-B-6
13:30-15:30
- Analysis of Native Listeners' Facial Microexpressions While Shadowing Non-Native Speech — Potential of Shadowers' Facial Expressions for Comprehensibility Prediction**
Tasavat Trisitichoke, Shintaro Ando, Daisuke Saito, Nobuaki Minematsu, University of Tokyo, Japan

Tue-P-4-B continued...

- PAGE 1866
Tue-P-4-B-7
13:30-15:30
- Transparent Pronunciation Scoring Using Articulatorily Weighted Phoneme Edit Distance**
Reima Karhila¹, Anna-Riikka Smolander², Sari Ylinen², Mikko Kurimo¹
¹Aalto University, Finland; ²University of Helsinki, Finland
- PAGE 1871
Tue-P-4-B-8
13:30-15:30
- Development of Robust Automated Scoring Models Using Adversarial Input for Oral Proficiency Assessment**
Su-Youn Yoon, Chong Min Lee, Klaus Zechner, Keelan Evanini, Educational Testing Service, USA
- PAGE 1876
Tue-P-4-B-9
13:30-15:30
- Impact of ASR Performance on Spoken Grammatical Error Detection**
Y. Lu, Mark J.F. Gales, Kate M. Knill, P. Manakul, L. Wang, Y. Wang, University of Cambridge, UK
- PAGE 1881
Tue-P-4-B-10
13:30-15:30
- Self-Imitating Feedback Generation Using GAN for Computer-Assisted Pronunciation Training**
Seung Hee Yang, Minhwa Chung, Seoul National University, Korea

Tue-P-4-C: Emotion and Personality in Conversation

Chair: To be confirmed

Time 13:30-15:30, Tuesday 17 Sept 2019, Gallery C

- PAGE 1886
Tue-P-4-C-1
13:30-15:30 **Joint Student-Teacher Learning for Audio-Visual Scene-Aware Dialog**
Chiori Hori, Anoop Cherian, Tim K. Marks, Takaaki Hori, MERL, USA
- PAGE 1891
Tue-P-4-C-2
13:30-15:30 **Topical-Chat: Towards Knowledge-Grounded Open-Domain Conversations**
Karthik Gopalakrishnan, Behnam Hedayatnia, Qinlang Chen, Anna Gottardi, Sanjeev Kwatra, Anu Venkatesh, Raefer Gabriel, Dilek Hakkani-Tür, Amazon, USA
- PAGE 1896
Tue-P-4-C-3
13:30-15:30 **Analyzing Verbal and Nonverbal Features for Predicting Group Performance**
Uliyana Kubasova, Gabriel Murray, McKenzie Braley, University of the Fraser Valley, Canada
- PAGE 1901
Tue-P-4-C-4
13:30-15:30 **Identifying Therapist and Client Personae for Therapeutic Alliance Estimation**
Victor R. Martinez¹, Nikolaos Flemotomos¹, Victor Ardulov¹, Krishna Somandepalli¹, Simon B. Goldberg², Zac E. Imel³, David C. Atkins⁴, Shrikanth Narayanan¹
¹University of Southern California, USA; ²UW-Madison, USA; ³University of Utah, USA; ⁴University of Washington, USA
- PAGE 1906
Tue-P-4-C-5
13:30-15:30 **Do Hesitations Facilitate Processing of Partially Defective System Utterances? An Exploratory Eye Tracking Study**
Kristin Haake¹, Sarah Schimke¹, Simon Betz², Sina Zarrieß²
¹Technische Universität Dortmund, Germany; ²Universität Bielefeld, Germany
- PAGE 1911
Tue-P-4-C-6
13:30-15:30 **Influence of Contextuality on Prosodic Realization of Information Structure in Chinese Dialogues**
Bin Li¹, Yuan Jia²
¹UCASS, China; ²CASS, China

Tue-P-4-C continued...

- PAGE 1916
Tue-P-4-C-7
13:30-15:30 **Cross-Lingual Transfer Learning for Affective Spoken Dialogue Systems**
Kristijan Gjoreski, Aleksandar Gjoreski, Ivan Kraljevski, Diane Hirschfeld, voice INTER connect, Germany
- PAGE 1921
Tue-P-4-C-8
13:30-15:30 **Identifying Personality Traits Using Overlap Dynamics in Multiparty Dialogue**
Mingzhi Yu¹, Emer Gilmartin², Diane Litman¹
¹University of Pittsburgh, USA; ²Trinity College Dublin, Ireland
- PAGE 1926
Tue-P-4-C-9
13:30-15:30 **Identifying Mood Episodes Using Dialogue Features from Clinical Interviews**
Zakaria Aldeneh¹, Mimansa Jaiswal¹, Michael Picheny², Melvin G. McInnis¹, Emily Mower Provost¹
¹University of Michigan, USA; ²IBM, USA
- PAGE 1931
Tue-P-4-C-10
13:30-15:30 **Do Conversational Partners Entrain on Articulatory Precision?**
Nichola Lubold¹, Stephanie A. Borrie², Tyson S. Barrett², Megan Willi³, Visar Berisha¹
¹Arizona State University, USA; ²Utah State University, USA; ³California State University Chico, USA
- PAGE 1936
Tue-P-4-C-11
13:30-15:30 **Conversational Emotion Analysis via Attention Mechanisms**
Zheng Lian, Jianhua Tao, Bin Liu, Jian Huang, Chinese Academy of Sciences, China

Tue-P-4-D: Voice Quality, Speech Perception, and Prosody

Chair: Immacolata Sonia d'Apolito

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 10/D

PAGE 1941 Tue-P-4-D-1 13:30-15:30	The Effect of Phoneme Distribution on Perceptual Similarity in English <i>Emma O'Neill, Julie Carson-Berndsen, University College Dublin, Ireland</i>
PAGE 1946 Tue-P-4-D-2 13:30-15:30	Prosodic Representations of Prominence Classification Neural Networks and Autoencoders Using Bottleneck Features <i>Sofoklis Kakouros, Antti Suni, Juraj Šimko, Martti Vainio, University of Helsinki, Finland</i>
PAGE 1951 Tue-P-4-D-3 13:30-15:30	Compensation for French Liquid Deletion During Auditory Sentence Processing <i>Sharon Peperkamp, Alvaro Martin Iturralde Zurita, LSCP (UMR 8554), France</i>
PAGE 1956 Tue-P-4-D-4 13:30-15:30	Prosodic Factors Influencing Vowel Reduction in Russian <i>Daniil Kocharov, Tatiana Kachkovskaia, Pavel Skrelin, Saint Petersburg State University, Russia</i>
PAGE 1961 Tue-P-4-D-5 13:30-15:30	Time to Frequency Domain Mapping of the Voice Source: The Influence of Open Quotient and Glottal Skew on the Low End of the Source Spectrum <i>Christer Gobl, Ailbhe Ni Chasaide, Trinity College Dublin, Ireland</i>
PAGE 1966 Tue-P-4-D-6 13:30-15:30	Testing the Distinctiveness of Intonational Tunes: Evidence from Imitative Productions in American English <i>Eleanor Chodroff¹, Jennifer S. Cole²</i> <i>¹University of York, UK; ²Northwestern University, USA</i>
PAGE 1971 Tue-P-4-D-7 13:30-15:30	A Study of a Cross-Language Perception Based on Cortical Analysis Using Biomimetic STRFs <i>Sangwook Park¹, David K. Han², Mounya Elhilali¹</i> <i>¹Johns Hopkins University, USA; ²U.S. Army Research Laboratory, USA</i>

Tue-P-4-D continued...

PAGE 1976 Tue-P-4-D-8 13:30-15:30	Perceptual Evaluation of Early versus Late F0 Peaks in the Intonation Structure of Czech Question-Word Questions <i>Pavel Šturm, Jan Volín, Charles University, Czech Republic</i>
PAGE 1981 Tue-P-4-D-9 13:30-15:30	Acoustic Correlates of Phonation Type in Chichimec <i>Anneliese Kelterer¹, Barbara Schuppler²</i> <i>¹Universität Graz, Austria; ²Technische Universität Graz, Austria</i>
PAGE 1986 Tue-P-4-D-10 13:30-15:30	F0 Variability Measures Based on Glottal Closure Instants <i>Yu-Ren Chien, Michal Borský, Jón Guðnason, Reykjavik University, Iceland</i>
PAGE 1990 Tue-P-4-D-11 13:30-15:30	Recognition of Creaky Voice from Emergency Calls <i>Lauri Tavi¹, Tanel Alumäe², Stefan Werner¹</i> <i>¹University of Eastern Finland, Finland; ²TalTech, Estonia</i>

Tue-P-4-E : Speech Signal Characterization 3

Chair: Hiroshi Shimodaira

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 10/E

PAGE 1995 Tue-P-4-E-1 13:30-15:30	Direct F0 Estimation with Neural-Network-Based Regression <i>Shuzhuang Xu, Hiroshi Shimodaira, University of Edinburgh, UK</i>
PAGE 2000 Tue-P-4-E-2 13:30-15:30	Real Time Online Visual End Point Detection Using Unidirectional LSTM <i>Tanay Sharma, Rohith Chandrashekar Aralikatti, Dilip Kumar Margam, Abhinav Thanda, Sharad Roy, Pujitha Appan Kandala, Shankar M. Venkatesan, Samsung, India</i>
PAGE 2005 Tue-P-4-E-3 13:30-15:30	Fully-Convolutional Network for Pitch Estimation of Speech Signals <i>Luc Ardaillon, Axel Roebel, STMS (UMR 9912), France</i>
PAGE 2010 Tue-P-4-E-4 13:30-15:30	Vocal Pitch Extraction in Polyphonic Music Using Convolutional Residual Network <i>Mingye Dong¹, Jie Wu², Jian Luan²</i> <i>¹USTC, China; ²Microsoft, China</i>
PAGE 2015 Tue-P-4-E-5 13:30-15:30	Multi-Level Adaptive Speech Activity Detector for Speech in Naturalistic Environments <i>Bidisha Sharma, Rohan Kumar Das, Haizhou Li, NUS, Singapore</i>
PAGE 2020 Tue-P-4-E-6 13:30-15:30	On the Importance of Audio-Source Separation for Singer Identification in Polyphonic Music <i>Bidisha Sharma, Rohan Kumar Das, Haizhou Li, NUS, Singapore</i>
PAGE 2025 Tue-P-4-E-7 13:30-15:30	Investigating the Physiological and Acoustic Contrasts Between Choral and Operatic Singing <i>Hiroko Terasawa¹, Kenta Wakasa¹, Hideki Kawahara², Ken-Ichi Sakakibara³</i> <i>¹University of Tsukuba, Japan; ²Wakayama University, Japan; ³Health Sciences University of Hokkaido, Japan</i>

Tue-P-4-E continued...

PAGE 2030 Tue-P-4-E-8 13:30-15:30	Optimizing Voice Activity Detection for Noisy Conditions <i>Ruixi Lin¹, Charles Costello², Charles Jankowski³, Vishwas Mruthyunjaya⁴</i> <i>¹NUS, Singapore; ²Plus One Robotics, USA; ³CloudMinds Technology, USA; ⁴AI SERA, USA</i>
PAGE 2035 Tue-P-4-E-9 13:30-15:30	Small-Footprint Magic Word Detection Method Using Convolutional LSTM Neural Network <i>Taiki Yamamoto¹, Ryota Nishimura¹, Masayuki Misaki², Norihide Kitaoka¹</i> <i>¹Tokushima University, Japan; ²Panasonic, Japan</i>
PAGE 2040 Tue-P-4-E-10 13:30-15:30	Acoustic Modeling for Automatic Lyrics-to-Audio Alignment <i>Chitralekha Gupta, Emre Yilmaz, Haizhou Li, NUS, Singapore</i>
PAGE 2045 Tue-P-4-E-11 13:30-15:30	Two-Dimensional Convolutional Recurrent Neural Networks for Speech Activity Detection <i>Anastasios Vafeiadis¹, Eleftherios Fanioudakis², Ilyas Potamitis², Konstantinos Votis¹, Dimitrios Giakoumis¹, Dimitrios Tzovaras¹, Liming Chen³, Raouf Hamzaoui³</i> <i>¹CERTH, Greece; ²TEI Crete, Greece; ³De Montfort University, UK</i>
PAGE 2050 Tue-P-4-E-12 13:30-15:30	A Study of Soprano Singing in Light of the Source-Filter Interaction <i>Tokihiko Kaburagi, Kyushu University, Japan</i>

Tue-P-5-A: Speech Synthesis: Pronunciation, Multilingual, and Low Resource

Chair: Erica Cooper

Time 16:00-18:00, Tuesday 17 Sept 2019, Gallery A

PAGE 2055 Tue-P-5-A-1 16:00-18:00	Boosting Character-Based Chinese Speech Synthesis via Multi-Task Learning and Dictionary Tutoring <i>Yuxiang Zou, Linhao Dong, Bo Xu, Chinese Academy of Sciences, China</i>
PAGE 2060 Tue-P-5-A-2 16:00-18:00	Building a Mixed-Lingual Neural TTS System with Only Monolingual Data <i>Liumeng Xue¹, Wei Song², Guanghui Xu², Lei Xie¹, Zhizheng Wu³</i> <i>¹Northwestern Polytechnical University, China; ²JD.com, China; ³JD.com, USA</i>
PAGE 2065 Tue-P-5-A-3 16:00-18:00	Neural Machine Translation for Multilingual Grapheme-to-Phoneme Conversion <i>Alex Sokolov, Tracy Rohlin, Ariya Rastrow, Amazon, USA</i>
PAGE 2070 Tue-P-5-A-4 16:00-18:00	Analysis of Pronunciation Learning in End-to-End Speech Synthesis <i>Jason Taylor, Korin Richmond, University of Edinburgh, UK</i>
PAGE 2075 Tue-P-5-A-5 16:00-18:00	End-to-End Text-to-Speech for Low-Resource Languages by Cross-Lingual Transfer Learning <i>Yuan-Jui Chen, Tao Tu, Cheng-chieh Yeh, Hung-Yi Lee, National Taiwan University, Taiwan</i>
PAGE 2080 Tue-P-5-A-6 16:00-18:00	Learning to Speak Fluently in a Foreign Language: Multilingual Speech Synthesis and Cross-Language Voice Cloning <i>Yu Zhang¹, Ron J. Weiss¹, Heiga Zen², Yonghui Wu¹, Zhifeng Chen¹, R.J. Skerry-Ryan¹, Ye Jia¹, Andrew Rosenberg¹, Bhuvana Ramabhadran¹</i> <i>¹Google, USA; ²Google, Japan</i>
PAGE 2085 Tue-P-5-A-7 16:00-18:00	Unified Language-Independent DNN-Based G2P Converter <i>Markéta Jůzová, Daniel Tihelka, Jakub Vít, University of West Bohemia, Czech Republic</i>

Tue-P-5-A continued...

PAGE 2090 Tue-P-5-A-8 16:00-18:00	Disambiguation of Chinese Polyphones in an End-to-End Framework with Semantic Features Extracted by Pre-Trained BERT <i>Dongyang Dai¹, Zhiyong Wu¹, Shiyin Kang², Xixin Wu³, Jia Jia¹, Dan Su², Dong Yu⁴, Helen Meng¹</i> <i>¹Tsinghua University, China; ²Tencent, China; ³CUHK, China; ⁴Tencent, USA</i>
PAGE 2095 Tue-P-5-A-9 16:00-18:00	Transformer Based Grapheme-to-Phoneme Conversion <i>Sevinj Yolchuyeva, Géza Németh, Bálint Gyires-Tóth, BME, Hungary</i>
PAGE 2100 Tue-P-5-A-10 16:00-18:00	Developing Pronunciation Models in New Languages Faster by Exploiting Common Grapheme-to-Phoneme Correspondences Across Languages <i>Harry Bleyan¹, Sandy Ritchie², Jonas Fromseier Mortensen², Daan van Esch¹</i> <i>¹Google, USA; ²Google, UK</i>
PAGE 2105 Tue-P-5-A-11 16:00-18:00	Cross-Lingual, Multi-Speaker Text-To-Speech Synthesis Using Neural Speaker Embedding <i>Mengnan Chen¹, Minchuan Chen², Shuang Liang², Jun Ma², Lei Chen¹, Shaojun Wang², Jing Xiao²</i> <i>¹East China Normal University, China; ²Ping An Technology, China</i>
PAGE 2110 Tue-P-5-A-12 16:00-18:00	Polyphone Disambiguation for Mandarin Chinese Using Conditional Neural Network with Multi-Level Embedding Features <i>Zexin Cai, Yaogen Yang, Chuxiong Zhang, Xiaoyi Qin, Ming Li, Duke Kunshan University, China</i>
PAGE 2115 Tue-P-5-A-13 16:00-18:00	Token-Level Ensemble Distillation for Grapheme-to-Phoneme Conversion <i>Hao Sun¹, Xu Tan², Jun-Wei Gan², Hongzhi Liu¹, Sheng Zhao², Tao Qin², Tie-Yan Liu²</i> <i>¹Peking University, China; ²Microsoft, China</i>

Tue-P-5-B: Cross-Lingual and Multilingual ASR

Chair: Sakriani Sakti

Time 16:00-18:00, Tuesday 17 Sept 2019, Gallery B

PAGE 2120 Tue-P-5-B-1 16:00-18:00	Multilingual Speech Recognition with Corpus Relatedness Sampling <i>Xinjian Li, Siddharth Dalmia, Alan W. Black, Florian Metze, Carnegie Mellon University, USA</i>
PAGE 2125 Tue-P-5-B-2 16:00-18:00	Multi-Dialect Acoustic Modeling Using Phone Mapping and Online i-Vectors <i>Harish Arsikere, Ashtosh Sapru, Sri Garimella, Amazon, India</i>
PAGE 2130 Tue-P-5-B-3 16:00-18:00	Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model <i>Anjuli Kannan, Arindrima Datta, Tara N. Sainath, Eugene Weinstein, Bhuvana Ramabhadran, Yonghui Wu, Ankur Bapna, Zhifeng Chen, Seungji Lee, Google, USA</i>
PAGE 2135 Tue-P-5-B-4 16:00-18:00	Recognition of Latin American Spanish Using Multi-Task Learning <i>Carlos Mendes¹, Alberto Abad², João Paulo Neto¹, Isabel Trancoso²</i> <i>¹VoiceInteraction, Portugal; ²INESC-ID, Portugal</i>
PAGE 2140 Tue-P-5-B-5 16:00-18:00	End-to-End Accented Speech Recognition <i>Thibault Viglino¹, Petr Motlicek², Milos Cernak³</i> <i>¹EPFL, Switzerland; ²Idiap Research Institute, Switzerland; ³Logitech, Switzerland</i>
PAGE 2145 Tue-P-5-B-6 16:00-18:00	End-to-End Articulatory Attribute Modeling for Low-Resource Multilingual Speech Recognition <i>Sheng Li, Chenchen Ding, Xugang Lu, Peng Shen, Tatsuya Kawahara, Hisashi Kawai, NICT, Japan</i>
PAGE 2150 Tue-P-5-B-7 16:00-18:00	Exploiting Monolingual Speech Corpora for Code-Mixed Speech Recognition <i>Karan Taneja¹, Satarupa Guha², Preethi Jyothi¹, Basil Abraham²</i> <i>¹IIT Bombay, India; ²Microsoft, India</i>

Tue-P-5-B continued...

PAGE 2155 Tue-P-5-B-8 16:00-18:00	Phoneme-Based Contextualization for Cross-Lingual Speech Recognition in End-to-End Models <i>Ke Hu, Antoine Bruguier, Tara N. Sainath, Rohit Prabhavalkar, Golan Pundak, Google, USA</i>
PAGE 2160 Tue-P-5-B-9 16:00-18:00	Constrained Output Embeddings for End-to-End Code-Switching Speech Recognition with Only Monolingual Data <i>Yerbolat Khassanov¹, Haihua Xu², Van Tung Pham¹, Zhiping Zeng², Eng Siong Chng¹, Chongjia Ni³, Bin Ma³</i> <i>¹NTU, Singapore; ²Temasek Laboratories @ NTU, Singapore; ³Alibaba Group, Singapore</i>
PAGE 2165 Tue-P-5-B-10 16:00-18:00	On the End-to-End Solution to Mandarin-English Code-Switching Speech Recognition <i>Zhiping Zeng¹, Yerbolat Khassanov², Van Tung Pham¹, Haihua Xu¹, Eng Siong Chng¹, Haizhou Li³</i> <i>¹Temasek Laboratories @ NTU, Singapore; ²NTU, Singapore; ³NUS, Singapore</i>
PAGE 2170 Tue-P-5-B-11 16:00-18:00	Towards Language-Universal Mandarin-English Speech Recognition <i>Shiliang Zhang¹, Yuan Liu¹, Ming Lei¹, Bin Ma², Lei Xie³</i> <i>¹Alibaba Group, China; ²Alibaba Group, Singapore; ³Northwestern Polytechnical University, China</i>

Tue-P-5-C: Spoken Term Detection, Confidence Measure, and End-to-End Speech Recognition

Chair: Yu Tsao

Time 16:00-18:00, Tuesday 17 Sept 2019, Gallery C

PAGE 2175 Tue-P-5-C-1 16:00-18:00	Improving ASR Confidence Scores for Alexa Using Acoustic and Hypothesis Embeddings <i>Prakhar Swarup¹, Roland Maas², Sri Garimella¹, Sri Harish Mallidi², Björn Hoffmeister²</i> ¹ Amazon, India; ² Amazon, USA
PAGE 2180 Tue-P-5-C-2 16:00-18:00	Investigation of Transformer Based Spelling Correction Model for CTC-Based End-to-End Mandarin Speech Recognition <i>Shiliang Zhang, Ming Lei, Zhijie Yan, Alibaba Group, China</i>
PAGE 2185 Tue-P-5-C-3 16:00-18:00	Improving Performance of End-to-End ASR on Numeric Sequences <i>Cal Peyser, Hao Zhang, Tara N. Sainath, Zelin Wu, Google, USA</i>
PAGE 2190 Tue-P-5-C-4 16:00-18:00	A Time Delay Neural Network with Shared Weight Self-Attention for Small-Footprint Keyword Spotting <i>Ye Bai¹, Jiangyan Yi¹, Jianhua Tao¹, Zhengqi Wen¹, Zhengkun Tian¹, Chenghao Zhao², Cunhang Fan¹</i> ¹ Chinese Academy of Sciences, China; ² Jiangsu Normal University, China
PAGE 2195 Tue-P-5-C-5 16:00-18:00	Sub-Band Convolutional Neural Networks for Small-Footprint Spoken Term Classification <i>Chieh-Chi Kao, Ming Sun, Yixin Gao, Shiv Vitaladevuni, Chao Wang, Amazon, USA</i>
PAGE 2200 Tue-P-5-C-6 16:00-18:00	Investigating Radical-Based End-to-End Speech Recognition Systems for Chinese Dialects and Japanese <i>Sheng Li, Xugang Lu, Chenchen Ding, Peng Shen, Tatsuya Kawahara, Hisashi Kawai, NICT, Japan</i>

Tue-P-5-C continued...

PAGE 2205 Tue-P-5-C-7 16:00-18:00	Joint Decoding of CTC Based Systems for Speech Recognition <i>Jiaqi Guo¹, Yongbin You², Yanmin Qian¹, Kai Yu¹</i> ¹ Shanghai Jiao Tong University, China; ² AISpeech, China
PAGE 2210 Tue-P-5-C-8 16:00-18:00	A Joint End-to-End and DNN-HMM Hybrid Automatic Speech Recognition System with Transferring Sharable Knowledge <i>Tomohiro Tanaka, Ryo Masumura, Takafumi Moriya, Takanobu Oba, Yushi Aono, NTT, Japan</i>
PAGE 2215 Tue-P-5-C-9 16:00-18:00	Active Learning Methods for Low Resource End-to-End Speech Recognition <i>Karan Malhotra, Shubham Bansal, Sriram Ganapathy, Indian Institute of Science, India</i>
PAGE 2220 Tue-P-5-C-10 16:00-18:00	Analysis of Multilingual Sequence-to-Sequence Speech Recognition Systems <i>Martin Karafiát¹, Murali Karthick Baskar¹, Shinji Watanabe², Takaaki Hori³, Matthew Wiesner², Jan Černocký¹</i> ¹ Brno University of Technology, Czech Republic; ² Johns Hopkins University, USA; ³ MERL, USA
PAGE 2225 Tue-P-5-C-11 16:00-18:00	Lattice Generation in Attention-Based Speech Recognition Models <i>Michał Zapotoczny, Piotr Pietrzak, Adrian Lańcucki, Jan Chorowski, University of Wrocław, Poland</i>
PAGE 2230 Tue-P-5-C-12 16:00-18:00	Sampling from Stochastic Finite Automata with Applications to CTC Decoding <i>Martin Jansche, Alexander Gutkin, Google, UK</i>
PAGE 2235 Tue-P-5-C-13 16:00-18:00	ShrinkML: End-to-End ASR Model Compression Using Reinforcement Learning <i>Lukasz Dudziak, Mohamed S. Abdelfattah, Ravichander Vipperla, Stefanos Laskaridis, Nicholas D. Lane, Samsung, UK</i>

Tue-P-5-C continued...

PAGE 2240
Tue-P-5-C-14
16:00-18:00

Acoustic-to-Phrase Models for Speech Recognition
Yashesh Gaur, Jinyu Li, Zhong Meng, Yifan Gong, Microsoft, USA

PAGE 2245
Tue-P-5-C-15
16:00-18:00

Performance Monitoring for End-to-End Speech Recognition
Ruizhi Li, Gregory Sell, Hynek Hermansky, Johns Hopkins University, USA

Tue-P-5-D: Speech Perception

Chair: Ellen Marklund

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 10/D

PAGE 2250
Tue-P-5-D-1
16:00-18:00

The Role of Musical Experience in the Perceptual Weighting of Acoustic Cues for the Obstruent Coda Voicing Contrast in American English
Michelle Cohn, Georgia Zellou, Santiago Barreda, University of California at Davis, USA

PAGE 2255
Tue-P-5-D-2
16:00-18:00

Individual Differences in Implicit Attention to Phonetic Detail in Speech Perception
Natalie Lewandowski¹, Daniel Duran²
¹Independent Scientist, Germany; ²Albert-Ludwigs-Universität Freiburg, Germany

PAGE 2260
Tue-P-5-D-3
16:00-18:00

Effects of Natural Variability in Cross-Modal Temporal Correlations on Audiovisual Speech Recognition Benefit
Kaylah Lalonde, Boys Town National Research Hospital, USA

PAGE 2265
Tue-P-5-D-4
16:00-18:00

Listening with Great Expectations: An Investigation of Word Form Anticipations in Naturalistic Speech
M. Bentum, L. ten Bosch, A. van den Bosch, Mirjam Ernestus, Radboud Universiteit Nijmegen, The Netherlands

PAGE 2270
Tue-P-5-D-5
16:00-18:00

Quantifying Expectation Modulation in Human Speech Processing
M. Bentum, L. ten Bosch, A. van den Bosch, Mirjam Ernestus, Radboud Universiteit Nijmegen, The Netherlands

PAGE 2275
Tue-P-5-D-6
16:00-18:00

Perception of Pitch Contours in Speech and Nonspeech
Daniel R. Turner, Ann R. Bradlow, Jennifer S. Cole, Northwestern University, USA

PAGE 2280
Tue-P-5-D-7
16:00-18:00

Analyzing Reaction Time and Error Sequences in Lexical Decision Experiments
L. ten Bosch, L. Boves, K. Mulder, Radboud Universiteit Nijmegen, The Netherlands

INTERSPEECH 2019 Table of Contents

Tue-P-5-D continued...

- PAGE 2285
Tue-P-5-D-8
16:00-18:00
Automatic Detection of the Temporal Segmentation of Hand Movements in British English Cued Speech
Li Liu¹, Jianze Li², Gang Feng³, Xiao-Ping Zhang¹
¹Ryerson University, Canada; ²CUHK, China; ³GIPSA-lab (UMR 5216), France
- PAGE 2290
Tue-P-5-D-9
16:00-18:00
Place Shift as an Autonomous Process: Evidence from Japanese Listeners
Yuriko Yokoe, Sophia University, Japan
- PAGE 2295
Tue-P-5-D-10
16:00-18:00
A Perceptual Study of CV Syllables in Both Spoken and Whistled Speech: A Tashlhiyt Berber Perspective
Julien Meyer¹, Laure Dentel², Silvain Gerber¹, Rachid Ridouane³
¹GIPSA-lab (UMR 5216), France; ²World Whistles Research Association, France; ³LPP (UMR 7018), France
- PAGE 2300
Tue-P-5-D-11
16:00-18:00
Consonant Classification in Mandarin Based on the Depth Image Feature: A Pilot Study
Han-Chi Hsieh, Wei-Zhong Zheng, Ko-Chiang Chen, Ying-Hui Lai, National Yang-Ming University, Taiwan
- PAGE 2305
Tue-P-5-D-12
16:00-18:00
The Different Roles of Expectations in Phonetic and Lexical Processing
Shiri Lev-Ari¹, Robin Dodsworth², Jeff Mielke², Sharon Peperkamp³
¹RHUL, UK; ²North Carolina State University, USA; ³LSCP (UMR 8554), France
- PAGE 2310
Tue-P-5-D-13
16:00-18:00
Perceptual Adaptation to Device and Human Voices: Learning and Generalization of a Phonetic Shift Across Real and Voice-AI Talkers
Bruno Ferenc Segedin, Michelle Cohn, Georgia Zellou, University of California at Davis, USA
- PAGE 2315
Tue-P-5-D-14
16:00-18:00
End-to-End Convolutional Sequence Learning for ASL Fingerspelling Recognition
Katerina Papadimitriou, Gerasimos Potamianos, University of Thessaly, Greece

INTERSPEECH 2019 Table of Contents

Tue-P-5-E: Topics in Speech and Audio Signal Processing

Chair: To be confirmed

Time 16:00-18:00, Tuesday 17 Sept 2019, Hall 10/E

- PAGE 2320
Tue-P-5-E-1
16:00-18:00
Multiview Shared Subspace Learning Across Speakers and Speech Commands
Krishna Somandepalli, Naveen Kumar, Arindam Jati, Panayiotis Georgiou, Shrikanth Narayanan, University of Southern California, USA
- PAGE 2325
Tue-P-5-E-2
16:00-18:00
A Machine Learning Based Clustering Protocol for Determining Hearing Aid Initial Configurations from Pure-Tone Audiograms
Celzy Belitz, Hussnain Ali, John H.L. Hansen, University of Texas at Dallas, USA
- PAGE 2330
Tue-P-5-E-3
16:00-18:00
Acoustic Scene Classification with Mismatched Devices Using CliqueNets and Mixup Data Augmentation
Truc Nguyen, Franz Pernkopf, Technische Universität Graz, Austria
- PAGE 2335
Tue-P-5-E-4
16:00-18:00
DeepLung: Smartphone Convolutional Neural Network-Based Inference of Lung Anomalies for Pulmonary Patients
Mohsin Y. Ahmed¹, Md. Mahbubur Rahman², Jilong Kuang²
¹University of Virginia, USA; ²Samsung, USA
- PAGE 2340
Tue-P-5-E-5
16:00-18:00
On the Use/Misuse of the Term 'Phoneme'
Roger K. Moore, Lucy Skidmore, University of Sheffield, UK
- PAGE 2345
Tue-P-5-E-6
16:00-18:00
Understanding and Visualizing Raw Waveform-Based CNNs
Hannah Muckenhirn¹, Vinayak Abrol², Mathew Magimai-Doss¹, Sébastien Marcel¹
¹Idiap Research Institute, Switzerland; ²University of Oxford, UK
- PAGE 2350
Tue-P-5-E-7
16:00-18:00
Fréchet Audio Distance: A Reference-Free Metric for Evaluating Music Enhancement Algorithms
Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, Matthew Sharifi, Google, Switzerland

Tue-P-5-E continued...

PAGE 2355
Tue-P-5-E-8
16:00-18:00

ReMASC: Realistic Replay Attack Corpus for Voice Controlled Systems
Yuan Gong, Jian Yang, Jacob Huber, Mitchell MacKnight, Christian Poellabauer,
University of Notre Dame, USA

PAGE 2360
Tue-P-5-E-9
16:00-18:00

Analyzing Intra-Speaker and Inter-Speaker Vocal Tract Impedance Characteristics in a Low-Dimensional Feature Space Using t-SNE
Balamurali B.T., Jer-Ming Chen, SUTD, Singapore

Tue-S&T-2 : Speech Processing and Analysis

Time 13:30-15:30, Tuesday 17 Sept 2019, Hall 4

PAGE 2364
Tue-S&T-2-1
13:30-15:30

Directional Audio Rendering Using a Neural Network Based Personalized HRTF
Geon Woo Lee, Jung Hyuk Lee, Seong Ju Kim, Hong Kook Kim, GIST, Korea

PAGE 2366
Tue-S&T-2-2
13:30-15:30

Online Speech Processing and Analysis Suite
Wikus Pienaar, Daan Wissing, North-West University, South Africa

PAGE 2368
Tue-S&T-2-3
13:30-15:30

Formant Pattern and Spectral Shape Ambiguity of Vowel Sounds, and Related Phenomena of Vowel Acoustics — Exemplary Evidence
Dieter Maurer¹, Heidy Suter¹, Christian d'Hereuse¹, Volker Dellwo²
¹Zürcher Hochschule der Künste, Switzerland; ²Universität Zürich, Switzerland

PAGE 2370
Tue-S&T-2-4
13:30-15:30

Sound Tools eXtended (STx) 5.0 — A Powerful Sound Analysis Tool Optimized for Speech
Anton Noll, Jonathan Stuefer, Nicola Klingler, Hannah Leykum, Carina Lozo,
Jan Luttenberger, Michael Pucher, Carolin Schmid, Austrian Academy of Sciences,
Austria

PAGE 2372
Tue-S&T-2-5
13:30-15:30

FarSpeech: Arabic Natural Language Processing for Live Arabic Speech
Mohamed Eldesouki, Naassih Gopee, Ahmed Ali, Kareem Darwish, HBKU, Qatar

PAGE 2374
Tue-S&T-2-6
13:30-15:30

A System for Real-Time Privacy Preserving Data Collection for Ambient Assisted Living
Fasih Haider, Saturnino Luz, University of Edinburgh, UK

PAGE 2376
Tue-S&T-2-7
13:30-15:30

NUS Speak-to-Sing: A Web Platform for Personalized Speech-to-Singing Conversion
Chitralekha Gupta, Karthika Vijayan, Bidisha Sharma, Xiaoxue Gao, Haizhou Li, NUS,
Singapore

Wed-K-3: Keynote 3: Manfred Kaltenbacher

Chair: *To be confirmed*

Time 08:30-09:30, Wednesday 18 Sept 2019, Main Hall

Wed-K-3
08:30-09:30
N/A

Physiology and Physics of Voice Production
Manfred Kaltenbacher, Technische Universität Wien, Austria

Wed-SS-6-4: The Interspeech 2019 Computational Paralinguistics Challenge (ComParE)

Chair: *Björn Schuller*

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 11

PAGE 2378
Wed-SS-6-4-1
10:00-10:15

The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity

Björn W. Schuller¹, Anton Batliner², Christian Bergler³, Florian B. Pokorny⁴, Jarek Krajewski⁵, Margaret Cychosz⁶, Ralf Vollmann⁷, Sonja-Dana Roelen⁸, Sebastian Schnieder⁸, Erika Bergelson⁹, Alejandrina Cristia¹⁰, Amanda Seidl¹¹, Anne S. Warlaumont¹², Lisa Yankowitz¹³, Elmar Nöth³, Shahin Amiriparian², Simone Hantke², Maximilian Schmitt²

¹Imperial College London, UK; ²Universität Augsburg, Germany; ³FAU Erlangen-Nürnberg, Germany; ⁴Medizinische Universität Graz, Austria; ⁵Bergische Universität Wuppertal, Germany; ⁶University of California at Berkeley, USA; ⁷Universität Graz, Austria; ⁸Rheinische Fachhochschule Köln, Germany; ⁹Duke University, USA; ¹⁰LSCP (UMR 8554), France; ¹¹Purdue University, USA; ¹²University of California at Los Angeles, USA; ¹³University of Pennsylvania, USA

PAGE 2383
Wed-SS-6-4-2
10:15-10:23

Using Speech Production Knowledge for Raw Waveform Modelling Based Styrian Dialect Identification

S. Pavankumar Dubagunta, Mathew Magimai-Doss, Idiap Research Institute, Switzerland

PAGE 2388
Wed-SS-6-4-3
10:23-10:31

Deep Neural Baselines for Computational Paralinguistics

Daniel Elsner, Stefan Langer, Fabian Ritz, Robert Mueller, Steffen Illium, LMU München, Germany

INTERSPEECH 2019 Table of Contents

Wed-SS-6-4 continued...

PAGE 2393
Wed-SS-6-4-4
10:31-10:39

Styrian Dialect Classification: Comparing and Fusing Classifiers Based on a Feature Selection Using a Genetic Algorithm

Thomas Kisler, Raphael Winkelmann, Florian Schiel, LMU München, Germany

PAGE 2398
Wed-SS-6-4-5
10:39-10:47

Using Attention Networks and Adversarial Augmentation for Styrian Dialect Continuous Sleepiness and Baby Sound Recognition

Sung-Lin Yeh¹, Gao-Yi Chao¹, Bo-Hao Su¹, Yu-Lin Huang¹, Meng-Han Lin¹, Yin-Chun Tsai¹, Yu-Wen Tai¹, Zheng-Chi Lu¹, Chieh-Yu Chen², Tsung-Ming Tai², Chiu-Wang Tseng², Cheng-Kuang Lee², Chi-Chun Lee¹
¹National Tsing Hua University, Taiwan; ²NVIDIA, Taiwan

PAGE 2403
Wed-SS-6-4-6
10:47-10:55

Ordinal Triplet Loss: Investigating Sleepiness Detection from Speech

Peter Wu, SaiKrishna Rallabandi, Alan W. Black, Eric Nyberg, Carnegie Mellon University, USA

PAGE 2408
Wed-SS-6-4-7
10:55-11:03

Voice Quality and Between-Frame Entropy for Sleepiness Estimation

Vijay Ravi, Soo Jin Park, Amber Afshan, Abeer Alwan, University of California at Los Angeles, USA

PAGE 2413
Wed-SS-6-4-8
11:03-11:11

Using Fisher Vector and Bag-of-Audio-Words Representations to Identify Styrian Dialects, Sleepiness, Baby & Orca Sounds

Gábor Gosztolya, MTA-SZTE RGAI, Hungary

PAGE 2418
Wed-SS-6-4-9
11:11-11:19

Instantaneous Phase and Long-Term Acoustic Cues for Orca Activity Detection

Rohan Kumar Das, Haizhou Li, NUS, Singapore

PAGE 2423
Wed-SS-6-4-10
11:19-11:27

Relevance-Based Feature Masking: Improving Neural Network Based Whale Classification Through Explainable Artificial Intelligence

Dominik Schiller, Tobias Huber, Florian Lingenfels, Michael Dietz, Andreas Seiderer, Elisabeth André, Universität Augsburg, Germany

INTERSPEECH 2019 Table of Contents

Wed-SS-6-4 continued...

PAGE 2428
Wed-SS-6-4-11
11:27-11:35

Spatial, Temporal and Spectral Multiresolution Analysis for the INTERSPEECH 2019 ComParE Challenge

Marie-José Caraty, Claude Montacé, STIH (EA 4509), France

PAGE 2433
Wed-SS-6-4-12
11:35-11:43

The DKU-LENOVO Systems for the INTERSPEECH 2019 Computational Paralinguistic Challenge

Haiwei Wu, Weiqing Wang, Ming Li, Duke Kunshan University, China

Wed-SS-6-4-13
11:43-12:00
N/A

Overview on Approaches and Results

Wed-SS-7-3: The VOiCES from a Distance Challenge — O

Chairs: Mahesh Kumar Nandwana, Mitchell McLaren

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 2

- PAGE 2438
Wed-O-7-3-1
13:30-13:40
- The VOiCES from a Distance Challenge 2019**
Mahesh Kumar Nandwana¹, Julien van Hout¹, Colleen Richey¹, Mitchell McLaren¹, Maria A. Barrios², Aaron Lawson¹
¹SRI International, USA; ²Lab41, USA
- PAGE 2443
Wed-O-7-3-2
13:40-13:45
- STC Speaker Recognition Systems for the VOiCES from a Distance Challenge**
Sergey Novoselov¹, Aleksei Gusev¹, Artem Ivanov¹, Timur Pekhovsky¹, Andrey Shulipa², Galina Lavrentyeva¹, Vladimir Volokhov¹, Alexandr Kozlov¹
¹STC-innovations, Russia; ²ITMO University, Russia
- PAGE 2448
Wed-O-7-3-3
13:45-13:50
- Analysis of BUT Submission in Far-Field Scenarios of VOiCES 2019 Challenge**
Pavel Matějka, Oldřich Plchot, Hossein Zeinali, Ladislav Mošner, Anna Silnova, Lukáš Burget, Ondřej Novotný, Ondřej Glembek, Brno University of Technology, Czech Republic
- PAGE 2453
Wed-O-7-3-4
13:50-13:55
- The STC ASR System for the VOiCES from a Distance Challenge 2019**
Ivan Medennikov¹, Yuri Khokhlov¹, Aleksei Romanenko², Ivan Sorokin¹, Anton Mitrofanov¹, Vladimir Bataev¹, Andrei Andrusenko¹, Tatiana Prisyach¹, Mariya Korenevskaya¹, Oleg Petrov², Alexander Zatzvornitskiy³
¹STC-innovations, Russia; ²ITMO University, Russia; ³Speech Technology Center, Russia
- PAGE 2458
Wed-O-7-3-5
13:55-14:00
- The I2R's ASR System for the VOiCES from a Distance Challenge 2019**
*Tze Yuang Chong, Kye Min Tan, Kah Kuan Teh, Chang Huai You, Hanwu Sun, Huy Dat Tran, A*STAR, Singapore*

Wed-SS-7-A: The VOiCES from a Distance Challenge — P

Chairs: Mahesh Kumar Nandwana, Mitchell McLaren

Time 13:30-15:30, Wednesday 18 Sept 2019, Gallery A

- Wed-SS-7-A-1
14:00-15:30
N/A
- The VOiCES from a Distance Challenge 2019**
Mahesh Kumar Nandwana¹, Julien van Hout¹, Colleen Richey¹, Mitchell McLaren¹, Maria A. Barrios², Aaron Lawson¹
¹SRI International, USA; ²Lab41, USA
- Wed-SS-7-A-2
14:00-15:30
N/A
- STC Speaker Recognition Systems for the VOiCES from a Distance Challenge**
Sergey Novoselov¹, Aleksei Gusev¹, Artem Ivanov¹, Timur Pekhovsky¹, Andrey Shulipa², Galina Lavrentyeva¹, Vladimir Volokhov¹, Alexandr Kozlov¹
¹STC-innovations, Russia; ²ITMO University, Russia
- Wed-SS-7-A-3
14:00-15:30
N/A
- Analysis of BUT Submission in Far-Field Scenarios of VOiCES 2019 Challenge**
Pavel Matějka, Oldřich Plchot, Hossein Zeinali, Ladislav Mošner, Anna Silnova, Lukáš Burget, Ondřej Novotný, Ondřej Glembek, Brno University of Technology, Czech Republic
- Wed-SS-7-A-4
14:00-15:30
N/A
- The STC ASR System for the VOiCES from a Distance Challenge 2019**
Ivan Medennikov¹, Yuri Khokhlov¹, Aleksei Romanenko², Ivan Sorokin¹, Anton Mitrofanov¹, Vladimir Bataev¹, Andrei Andrusenko¹, Tatiana Prisyach¹, Mariya Korenevskaya¹, Oleg Petrov², Alexander Zatzvornitskiy³
¹STC-innovations, Russia; ²ITMO University, Russia; ³Speech Technology Center, Russia
- Wed-SS-7-A-5
14:00-15:30
N/A
- The I2R's ASR System for the VOiCES from a Distance Challenge 2019**
*Tze Yuang Chong, Kye Min Tan, Kah Kuan Teh, Chang Huai You, Hanwu Sun, Huy Dat Tran, A*STAR, Singapore*

Wed-SS-7-A continued...

PAGE 2463
Wed-SS-7-A-6
14:00-15:30

Multi-Task Discriminative Training of Hybrid DNN-TVM Model for Speaker Verification with Noisy and Far-Field Speech

Arindam Jati, Raghuvveer Peri, Monisankha Pal, Tae Jin Park, Naveen Kumar, Ruchir Travadi, Panayiotis Georgiou, Shrikanth Narayanan, University of Southern California, USA

PAGE 2468
Wed-SS-7-A-7
14:00-15:30

The JHU Speaker Recognition System for the VOICES 2019 Challenge

David Snyder, Jesús Villalba, Nanxin Chen, Daniel Povey, Gregory Sell, Najim Dehak, Sanjeev Khudanpur, Johns Hopkins University, USA

PAGE 2473
Wed-SS-7-A-8
14:00-15:30

Intel Far-Field Speaker Recognition System for VOICES Challenge 2019

Jonathan Huang¹, Tobias Bocklet²
¹Intel, USA; ²Intel, Germany

PAGE 2478
Wed-SS-7-A-9
14:00-15:30

The I2R's Submission to VOICES Distance Speaker Recognition Challenge 2019

*Hanwu Sun, Kah Kuan Teh, Ivan Kukanov, Huy Dat Tran, A*STAR, Singapore*

PAGE 2483
Wed-SS-7-A-10
14:00-15:30

The LeVoice Far-Field Speech Recognition System for VOICES from a Distance Challenge 2019

Yulong Liang, Lin Yang, Xuyang Wang, Yingjie Li, Chen Jia, Junjie Wang, Lenovo, China

PAGE 2488
Wed-SS-7-A-11
14:00-15:30

The JHU ASR System for VOICES from a Distance Challenge 2019

Yiming Wang, David Snyder, Hainan Xu, Vimal Manohar, Phani Sankar Nidadavolu, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

PAGE 2493
Wed-SS-7-A-12
14:00-15:30

The DKU System for the Speaker Recognition Task of the 2019 VOICES from a Distance Challenge

Danwei Cai, Xiaoyi Qin, Weicheng Cai, Ming Li, Duke Kunshan University, China

Wed-SS-8-6: Voice Quality Characterization for Clinical Voice Assessment: Voice Production, Acoustics, and Auditory Perception

Chairs: Philipp Aichinger, Carlo Drioli

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 3

PAGE 2498
Wed-SS-8-6-1
16:00-16:15

Identifying Distinctive Acoustic and Spectral Features in Parkinson's Disease

Yermiyahu Hauptman¹, Ruth Aloni-Lavi¹, Itshak Lapidot¹, Tanya Gurevich², Yael Manor², Stav Naor², Noa Diamant², Irit Opher¹
¹Afeka College, Israel; ²Ichilov, Israel

PAGE 2503
Wed-SS-8-6-2
16:15-16:30

Aerodynamics and Lumped-Masses Combined with Delay Lines for Modeling Vertical and Anterior-Posterior Phase Differences in Pathological Vocal Fold Vibration

Carlo Drioli¹, Philipp Aichinger²
¹Università di Udine, Italy; ²Medizinische Universität Wien, Austria

PAGE 2508
Wed-SS-8-6-3
16:30-16:45

Mel-Frequency Cepstral Coefficients of Voice Source Waveforms for Classification of Phonation Types in Speech

Sudarsana Reddy Kadiri, Paavo Alku, Aalto University, Finland

PAGE 2513
Wed-SS-8-6-4
16:45-17:00

Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations

Sunghye Cho¹, Mark Liberman¹, Neville Ryant¹, Meredith Cola², Robert T. Schultz², Julia Parish-Morris²
¹University of Pennsylvania, USA; ²Children's Hospital of Philadelphia, USA

PAGE 2518
Wed-SS-8-6-5
17:00-17:15

Analysis and Synthesis of Vocal Flutter and Vocal Jitter

Jean Schoentgen¹, Philipp Aichinger²
¹Université libre de Bruxelles, Belgium; ²Medizinische Universität Wien, Austria

INTERSPEECH 2019 Table of Contents

Wed-SS-8-6 continued...

PAGE 2523
Wed-SS-8-6-6
17:15-17:30

Reliability of Clinical Voice Parameters Captured with Smartphones — Measurements of Added Noise and Spectral Tilt

Felix Schaeffler, Stephen Jannetts, Janet Beck, Queen Margaret University, UK

PAGE 2528
Wed-SS-8-6-7
17:30-17:45

Say What? A Dataset for Exploring the Error Patterns That Two ASR Engines Make

Meredith Moore, Michael Saxon, Hemanth Venkateswara, Visar Berisha, Sethuraman Panchanathan, Arizona State University, USA

Wed-SS-8-6-8
17:45-18:00

Discussion

INTERSPEECH 2019 Table of Contents

Wed-O-6-1 : Prosody

Chairs: Margaret Zellers, Jan Michalsky

Time 10:00-12:00, Wednesday 18 Sept 2019, Main Hall

Wed-O-6-1-1
10:00-10:40
N/A

Survey Talk: Prosody Research and Applications: The State of the Art

Nigel G. Ward, University of Texas at El Paso, USA

PAGE 2533
Wed-O-6-1-2
10:40-11:00

Dimensions of Prosodic Prominence in an Attractor Model

Simon Roessig, Doris Mücke, Lena Pagel, Universität zu Köln, Germany

PAGE 2538
Wed-O-6-1-3
11:00-11:20

Comparative Analysis of Prosodic Characteristics Using WaveNet Embeddings

Antti Suni¹, Marcin Włodarczak², Martti Vainio¹, Juraj Šimko¹

¹University of Helsinki, Finland; ²Stockholm University, Sweden

PAGE 2543
Wed-O-6-1-4
11:20-11:40

The Role of Voice Quality in the Perception of Prominence in Synthetic Speech

Andy Murphy, Irena Yanushevskaya, Ailbhe Ní Chasaide, Christer Gobl, Trinity College Dublin, Ireland

PAGE 2548
Wed-O-6-1-5
11:40-12:00

Phonological Awareness of French Rising Contours in Japanese Learners

Rachel Albar, Hiyon Yoo, LLF (UMR 7110), France

Wed-O-6-2 : Speech and Audio Classification 1

Chairs: Katrin Kirchhoff, Roger Moore

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 1

PAGE 2553 Wed-O-6-2-1 10:00-10:20	Audio Classification of Bit-Representation Waveform <i>Masaki Okawa, Takuya Saito, Naoki Sawada, Hiromitsu Nishizaki, University of Yamanashi, Japan</i>
PAGE 2558 Wed-O-6-2-2 10:20-10:40	Locality-Constrained Linear Coding Based Fused Visual Features for Robust Acoustic Event Classification <i>Manjunath Mulimani, Shashidhar G. Koolagudi, NITK Surathkal, India</i>
PAGE 2563 Wed-O-6-2-3 10:40-11:00	Learning How to Listen: A Temporal-Frequential Attention Model for Sound Event Detection <i>Yu-Han Shen, Ke-Xin He, Wei-Qiang Zhang, Tsinghua University, China</i>
PAGE 2568 Wed-O-6-2-4 11:00-11:20	A Deep Residual Network for Large-Scale Acoustic Scene Analysis <i>Logan Ford, Hao Tang, François Grondin, James Glass, MIT, USA</i>
PAGE 2573 Wed-O-6-2-5 11:20-11:40	Supervised Classifiers for Audio Impairments with Noisy Labels <i>Chandan K.A. Reddy, Ross Cutler, Johannes Gehrke, Microsoft, USA</i>
PAGE 2578 Wed-O-6-2-6 11:40-12:00	Self-Attention for Speech Emotion Recognition <i>Lorenzo Tarantino¹, Philip N. Garner², Alexandros Lazaridis³</i> <i>¹EPFL, Switzerland; ²Idiap Research Institute, Switzerland; ³Swisscom, Switzerland</i>

Wed-O-6-3 : Singing and Multimodal Synthesis

Chairs: Julia Hirschberg, Sunayana Sitaram

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 2

PAGE 2583 Wed-O-6-3-1 10:00-10:20	Unsupervised Singing Voice Conversion <i>Eliya Nachmani, Lior Wolf, Facebook, Israel</i>
PAGE 2588 Wed-O-6-3-2 10:20-10:40	Adversarially Trained End-to-End Korean Singing Voice Synthesis System <i>Juheon Lee, Hyeong-Seok Choi, Chang-Bin Jeon, Junghyun Koo, Kyogu Lee, Seoul National University, Korea</i>
PAGE 2593 Wed-O-6-3-3 10:40-11:00	Singing Voice Synthesis Using Deep Autoregressive Neural Networks for Acoustic Modeling <i>Yuan-Hao Yi, Yang Ai, Zhen-Hua Ling, Li-Rong Dai, USTC, China</i>
PAGE 2598 Wed-O-6-3-4 11:00-11:20	Conditional Variational Auto-Encoder for Text-Driven Expressive AudioVisual Speech Synthesis <i>Sara Dahmani, Vincent Colotte, Valérian Girard, Slim Ouni, Loria (UMR 7503), France</i>
PAGE 2603 Wed-O-6-3-5 11:20-11:40	A Strategy for Improved Phone-Level Lyrics-to-Audio Alignment for Speech-to-Singing Synthesis <i>David Ayllón, Fernando Villavicencio, Pierre Lanchantin, ObEN, USA</i>
PAGE 2608 Wed-O-6-3-6 11:40-12:00	Modeling Labial Coarticulation with Bidirectional Gated Recurrent Networks and Transfer Learning <i>Théo Biasutto-Lervat, Sara Dahmani, Slim Ouni, Loria (UMR 7503), France</i>

Wed-O-6-5 : ASR Neural Network Training — 2

Chairs: Rohit Prabhavalkar, Jia Cui

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 12

- PAGE 2613
Wed-O-6-5-1
10:00-10:20
- SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition**
Daniel S. Park¹, William Chan², Yu Zhang¹, Chung-Cheng Chiu¹, Barret Zoph¹, Ekin D. Cubuk¹, Quoc V. Le¹
¹Google, USA; ²Google, Canada
- PAGE 2618
Wed-O-6-5-2
10:20-10:40
- Forget a Bit to Learn Better: Soft Forgetting for CTC-Based Automatic Speech Recognition**
Kartik Audhkhasi, George Saon, Zoltán Tüske, Brian Kingsbury, Michael Picheny, IBM, USA
- PAGE 2623
Wed-O-6-5-3
10:40-11:00
- Online Hybrid CTC/Attention Architecture for End-to-End Speech Recognition**
Haoran Miao, Gaofeng Cheng, Pengyuan Zhang, Ta Li, Yonghong Yan, Chinese Academy of Sciences, China
- PAGE 2628
Wed-O-6-5-4
11:00-11:20
- A Highly Efficient Distributed Deep Learning System for Automatic Speech Recognition**
Wei Zhang, Xiaodong Cui, Ulrich Finkler, George Saon, Abdullah Kayi, Alper Buyuktosunoglu, Brian Kingsbury, David Kung, Michael Picheny, IBM, USA
- PAGE 2633
Wed-O-6-5-5
11:20-11:40
- Knowledge Distillation for End-to-End Monaural Multi-Talker ASR System**
Wangyou Zhang, Xuankai Chang, Yanmin Qian, Shanghai Jiao Tong University, China
- PAGE 2638
Wed-O-6-5-6
11:40-12:00
- Analysis of Deep Clustering as Preprocessing for Automatic Speech Recognition of Sparsely Overlapping Speech**
Tobias Menne, Ilya Sklyar, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

Wed-O-7-1 : Bilingualism, L2, and Non-Nativeness

Chairs: Mirjam Broersma, Sharon Peperkamp

Time 13:30-15:30, Wednesday 18 Sept 2019, Main Hall

- Wed-O-7-1-1
13:30-14:10
N/A
- Survey Talk: Recognition of Foreign-Accented Speech: Challenges and Opportunities for Human and Computer Speech Communication**
Ann R. Bradlow, Northwestern University, USA
- PAGE 2643
Wed-O-7-1-2
14:10-14:30
- The Effects of Time Expansion on English as a Second Language Individuals**
John S. Novak III, Daniel Bunn, Robert V. Kenyon, University of Illinois at Chicago, USA
- PAGE 2648
Wed-O-7-1-3
14:30-14:50
- Capturing L1 Influence on L2 Pronunciation by Simulating Perceptual Space Using Acoustic Features**
Shuju Shi¹, Chilin Shih¹, Jinsong Zhang²
¹University of Illinois at Urbana-Champaign, USA; ²BLCU, China
- PAGE 2653
Wed-O-7-1-4
14:50-15:10
- Cognitive Factors in Thai-Naïve Mandarin Speakers' Imitation of Thai Lexical Tones**
Juqiang Chen, Catherine T. Best, Mark Antoniou, Western Sydney University, Australia
- PAGE 2658
Wed-O-7-1-5
15:10-15:30
- Foreign-Language Knowledge Enhances Artificial-Language Segmentation**
Annie Tremblay¹, Mirjam Broersma²
¹University of Kansas, USA; ²Radboud Universiteit Nijmegen, The Netherlands

Wed-O-7-2: Spoken Term Detection

Chairs: Stavros Tsakalidis, Xavier Anguera

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 1

- PAGE 2663
Wed-O-7-2-1
13:30-13:50
- Neural Named Entity Recognition from Subword Units**
Abdalghani Abujabal¹, Judith Gaspers²
¹MPI for Informatics, Germany; ²Amazon, Germany
- PAGE 2668
Wed-O-7-2-2
13:50-14:10
- Unsupervised Acoustic Segmentation and Clustering Using Siamese Network Embeddings**
Saurabhchand Bhati¹, Shekhar Nayak², K. Sri Rama Murty², Najim Dehak¹
¹Johns Hopkins University, USA; ²IIT Hyderabad, India
- PAGE 2673
Wed-O-7-2-3
14:10-14:30
- An Empirical Evaluation of DTW Subsampling Methods for Keyword Search**
Bolaji Yusuf, Murat Saraclar, Boğaziçi Üniversitesi, Turkey
- PAGE 2678
Wed-O-7-2-4
14:30-14:50
- Linguistically-Informed Training of Acoustic Word Embeddings for Low-Resource Languages**
Zixiaofan Yang, Julia Hirschberg, Columbia University, USA
- PAGE 2683
Wed-O-7-2-5
14:50-15:10
- Multimodal Word Discovery and Retrieval with Phone Sequence and Image Concepts**
Liming Wang, Mark A. Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA
- PAGE 2688
Wed-O-7-2-6
15:10-15:30
- Empirical Evaluation of Sequence-to-Sequence Models for Word Discovery in Low-Resource Settings**
Marcely Zanon Boito¹, Aline Villavicencio², Laurent Besacier¹
¹LIG (UMR 5217), France; ²University of Essex, UK

Wed-O-7-4: Speech and Audio Source Separation and Scene Analysis 2

Chairs: Hugo Van hamme, Masahito Togami

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 11

- PAGE 2693
Wed-O-7-4-1
13:30-13:50
- Direct-Path Signal Cross-Correlation Estimation for Sound Source Localization in Reverberation**
Wei Xue¹, Ying Tong¹, Guohong Ding¹, Chao Zhang², Tao Ma³, Xiaodong He¹, Bowen Zhou¹
¹JD.com, China; ²JD.com, UK; ³JD.com, USA
- PAGE 2698
Wed-O-7-4-2
13:50-14:10
- Multiple Sound Source Localization with SVD-PHAT**
François Grondin, James Glass, MIT, USA
- PAGE 2703
Wed-O-7-4-3
14:10-14:30
- Robust DOA Estimation Based on Convolutional Neural Network and Time-Frequency Masking**
Wangyou Zhang, Ying Zhou, Yanmin Qian, Shanghai Jiao Tong University, China
- PAGE 2708
Wed-O-7-4-4
14:30-14:50
- Multichannel Loss Function for Supervised Speech Source Separation by Mask-Based Beamforming**
Yoshiki Masuyama¹, Masahito Togami², Tatsuya Komatsu²
¹Waseda University, Japan; ²LINE, Japan
- PAGE 2713
Wed-O-7-4-5
14:50-15:10
- Direction-Aware Speaker Beam for Multi-Channel Speaker Extraction**
Guanjun Li¹, Shan Liang¹, Shuai Nie¹, Wenju Liu¹, Meng Yu², Lianwu Chen³, Shouye Peng⁴, Changliang Li⁵
¹Chinese Academy of Sciences, China; ²Tencent, USA; ³Tencent, China; ⁴Xueersi Online School, China; ⁵Kingsoft, China
- PAGE 2718
Wed-O-7-4-6
15:10-15:30
- Multimodal SpeakerBeam: Single Channel Target Speech Extraction with Audio-Visual Speaker Clues**
Tsubasa Ochiai, Marc Delcroix, Keisuke Kinoshita, Atsunori Ogawa, Tomohiro Nakatani, NTT, Japan

Wed-O-7-5 : Speech Enhancement: Single Channel 2

Chairs: John H.L. Hansen, Timo Gerkmann

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 12

PAGE 2723
Wed-O-7-5-1
13:30-13:50

Speech Denoising with Deep Feature Losses

François G. Germain¹, Qifeng Chen², Vladlen Koltun³

¹Stanford University, USA; ²HKUST, China; ³Intel, USA

PAGE 2728
Wed-O-7-5-2
13:50-14:10

VoiceFilter: Targeted Voice Separation by Speaker-Conditioned Spectrogram Masking

Quan Wang¹, Hannah Muckenhirn², Kevin Wilson¹, Prashant Sridhar¹, Zelin Wu¹, John R. Hershey¹, Rif A. Saurous¹, Ron J. Weiss¹, Ye Jia¹, Ignacio Lopez Moreno¹

¹Google, USA; ²Idiap Research Institute, Switzerland

PAGE 2733
Wed-O-7-5-3
14:10-14:30

Incorporating Symbolic Sequential Modeling for Speech Enhancement

Chien-Feng Liao¹, Yu Tsao¹, Xugang Lu², Hisashi Kawai²

¹Academia Sinica, Taiwan; ²NICT, Japan

PAGE 2738
Wed-O-7-5-4
14:30-14:50

Maximum a posteriori Speech Enhancement Based on Double Spectrum

Pejman Mowlae¹, Daniel Scheran², Johannes Stahl², Sean U.N. Wood², W. Bastiaan Kleijn³

¹Widex, Denmark; ²Technische Universität Graz, Austria; ³Victoria University of Wellington, New Zealand

PAGE 2743
Wed-O-7-5-5
14:50-15:10

Coarse-to-Fine Optimization for Speech Enhancement

Jian Yao, Ahmad Al-Dahle, Apple, USA

PAGE 2748
Wed-O-7-5-6
15:10-15:30

Kernel Machines Beat Deep Neural Networks on Mask-Based Single-Channel Speech Enhancement

Like Hui, Siyuan Ma, Mikhail Belkin, Ohio State University, USA

Wed-O-8-1 : Multimodal ASR

Chair: To be confirmed

Time 16:00-18:00, Wednesday 18 Sept 2019, Main Hall

Wed-O-8-1-1
16:00-16:40

Survey Talk: Multimodal Processing of Speech and Language

Florian Metze, Carnegie Mellon University, USA

PAGE 2753
Wed-O-8-1-2
16:40-17:00

MobiVSR : Efficient and Light-Weight Neural Network for Visual Speech Recognition on Mobile Devices

Nilay Shrivastava¹, Astitwa Saxena¹, Yaman Kumar², Rajiv Ratn Shah³, Amanda Stent⁴, Debanjan Mahata⁴, Preeti Kaur¹, Roger Zimmermann⁵

¹NSUT, India; ²Adobe, India; ³IIT Delhi, India; ⁴Bloomberg, USA; ⁵NUS, Singapore

PAGE 2758
Wed-O-8-1-3
17:00-17:20

Speaker Adaptation for Lip-Reading Using Visual Identity Vectors

Pujitha Appan Kandala, Abhinav Thanda, Dilip Kumar Margam, Rohith Chandrashekar Aralikatti, Tanay Sharma, Sharad Roy, Shankar M. Venkatesan, Samsung, India

PAGE 2763
Wed-O-8-1-4
17:20-17:40

MobiLipNet: Resource-Efficient Deep Learning Based Lipreading

Alexandros Koumparoulis, Gerasimos Potamianos, University of Thessaly, Greece

PAGE 2768
Wed-O-8-1-5
17:40-18:00

LipSound: Neural Mel-Spectrogram Reconstruction for Lip Reading

Leyuan Qu, Cornelius Weber, Stefan Wermter, Universität Hamburg, Germany

Wed-O-8-2 : ASR Neural Network Architectures 2

Chairs: Bhuvana Ramabhadran, Peter Bell

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 1

- PAGE 2773
Wed-O-8-2-1
16:00-16:20
- Two-Pass End-to-End Speech Recognition**
Tara N. Sainath, Ruoming Pang, David Rybach, Yanzhang He, Rohit Prabhavalkar, Wei Li, Mirkó Visontai, Qiao Liang, Trevor Strohman, Yonghui Wu, Ian McGraw, Chung-Cheng Chiu, Google, USA
- PAGE 2778
Wed-O-8-2-2
16:20-16:40
- Extract, Adapt and Recognize: An End-to-End Neural Network for Corrupted Monaural Speech Recognition**
Max W.Y. Lam¹, Jun Wang¹, Xunying Liu², Helen Meng², Dan Su¹, Dong Yu³
¹Tencent, China; ²CUHK, China; ³Tencent, USA
- PAGE 2783
Wed-O-8-2-3
16:40-17:00
- Multi-Task Multi-Resolution Char-to-BPE Cross-Attention Decoder for End-to-End Speech Recognition**
Dhananjaya Gowda, Abhinav Garg, Kwangyoung Kim, Mehul Kumar, Chanwoo Kim, Samsung, Korea
- PAGE 2788
Wed-O-8-2-4
17:00-17:20
- Multi-Stride Self-Attention for Speech Recognition**
Kyu J. Han¹, Jing Huang¹, Yun Tang¹, Xiaodong He², Bowen Zhou²
¹JD.com, USA; ²JD.com, China
- PAGE 2793
Wed-O-8-2-5
17:20-17:40
- LF-MMI Training of Bayesian and Gaussian Process Time Delay Neural Networks for Speech Recognition**
Shoukang Hu, Xurong Xie, Shansong Liu, Max W.Y. Lam, Jianwei Yu, Xixin Wu, Xunying Liu, Helen Meng, CUHK, China
- PAGE 2798
Wed-O-8-2-6
17:40-18:00
- Self-Teaching Networks**
Liang Lu, Eric Sun, Yifan Gong, Microsoft, USA

Wed-O-8-3 : Training Strategy for Speech Emotion Recognition

Chairs: Björn Schuller, Felix Burkhardt

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 2

- PAGE 2803
Wed-O-8-3-1
16:00-16:20
- Improved End-to-End Speech Emotion Recognition Using Self Attention Mechanism and Multitask Learning**
Yuanchao Li¹, Tianyu Zhao², Tatsuya Kawahara²
¹Honda, Japan; ²Kyoto University, Japan
- PAGE 2808
Wed-O-8-3-2
16:20-16:40
- Continuous Emotion Recognition in Speech — Do We Need Recurrence?**
Maximilian Schmitt, Nicholas Cummins, Björn W. Schuller, Universität Augsburg, Germany
- PAGE 2813
Wed-O-8-3-3
16:40-17:00
- Speech Based Emotion Prediction: Can a Linear Model Work?**
Anda Ouyang, Ting Dang, Vidhyasaharan Sethu, Eliathamby Ambikairajah, UNSW, Australia
- PAGE 2818
Wed-O-8-3-4
17:00-17:20
- Speech Emotion Recognition Based on Multi-Label Emotion Existence Model**
Atsushi Ando, Ryo Masumura, Hosana Kamiyama, Satoshi Kobashikawa, Yushi Aono, NTT, Japan
- PAGE 2823
Wed-O-8-3-5
17:20-17:40
- Gender De-Biasing in Speech Emotion Recognition**
Cristina Gorrostieta¹, Reza Lotfian¹, Kye Taylor¹, Richard Brutti¹, John Kane²
¹Cogito, USA; ²Cogito, Ireland
- PAGE 2828
Wed-O-8-3-6
17:40-18:00
- CycleGAN-Based Emotion Style Transfer as Data Augmentation for Speech Emotion Recognition**
Fang Bao, Michael Neumann, Ngoc Thang Vu, Universität Stuttgart, Germany

Wed-O-8-4: Voice Conversion for Style, Accent, and Emotion

Chairs: Helen Meng, Tomoki Toda

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 11

- PAGE 2833
Wed-O-8-4-1
16:00-16:20
- Lombard Speech Synthesis Using Transfer Learning in a Tacotron Text-to-Speech System**
Bajibabu Bollepalli, Lauri Juvela, Paavo Alku, Aalto University, Finland
- PAGE 2838
Wed-O-8-4-2
16:20-16:40
- Augmented CycleGANs for Continuous Scale Normal-to-Lombard Speaking Style Conversion**
Shreyas Seshadri¹, Lauri Juvela¹, Paavo Alku¹, Okko Räsänen²
¹Aalto University, Finland; ²Tampere University, Finland
- PAGE 2843
Wed-O-8-4-3
16:40-17:00
- Foreign Accent Conversion by Synthesizing Speech from Phonetic Posteriorgrams**
Guanlong Zhao, Shaojin Ding, Ricardo Gutierrez-Osuna, Texas A&M University, USA
- PAGE 2848
Wed-O-8-4-4
17:00-17:20
- A Multi-Speaker Emotion Morphing Model Using Highway Networks and Maximum Likelihood Objective**
Ravi Shankar, Jacob Sager, Archana Venkataraman, Johns Hopkins University, USA
- PAGE 2853
Wed-O-8-4-5
17:20-17:40
- Effects of Waveform PMF on Anti-Spoofing Detection**
Itshak Lapidot¹, Jean-François Bonastre²
¹Afeka College, Israel; ²LIA (EA 4128), France
- PAGE 2858
Wed-O-8-4-6
17:40-18:00
- Nonparallel Emotional Speech Conversion**
Jian Gao¹, Deep Chakraborty², Hamidou Tembine¹, Olaitan Olaleye³
¹New York University, USA; ²UMass Amherst, USA; ³Signify, USA

Wed-O-8-5: Speaker Recognition 2

Chairs: Nicholas Evans, Seyed Omid Sadjadi

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 12

- PAGE 2863
Wed-O-8-5-1
16:00-16:20
- Self-Supervised Speaker Embeddings**
Themos Stafylakis¹, Johan Rohdin², Oldřich Plchot², Petr Mizera¹, Lukáš Burget²
¹Omilia, Greece; ²Brno University of Technology, Czech Republic
- PAGE 2868
Wed-O-8-5-2
16:20-16:40
- Privacy-Preserving Speaker Recognition with Cohort Score Normalisation**
*Andreas Nautsch¹, Jose Patino¹, Amos Treiber², Themis Stafylakis³, Petr Mizera³,
Massimiliano Todisco¹, Thomas Schneider², Nicholas Evans¹*
¹EURECOM, France; ²Technische Universität Darmstadt, Germany; ³Omilia, Greece
- PAGE 2873
Wed-O-8-5-3
16:40-17:00
- Large Margin Softmax Loss for Speaker Verification**
Yi Liu, Liang He, Jia Liu, Tsinghua University, China
- PAGE 2878
Wed-O-8-5-4
17:00-17:20
- A Deep Neural Network for Short-Segment Speaker Recognition**
Amirhossein Hajavi, Ali Etemad, Queen's University, Canada
- PAGE 2883
Wed-O-8-5-5
17:20-17:40
- Deep Speaker Embedding Extraction with Channel-Wise Feature Responses and Additive Supervision Softmax Loss Function**
Jianfeng Zhou, Tao Jiang, Zheng Li, Lin Li, Qingyang Hong, Xiamen University, China
- PAGE 2888
Wed-O-8-5-6
17:40-18:00
- VoiceID Loss: Speech Enhancement for Speaker Verification**
Suwon Shon, Hao Tang, James Glass, MIT, USA

Wed-P-6-A : Speaker Recognition and Anti-Spoofing

Chair: Jesus Villalba Lopez

Time 10:00-12:00, Wednesday 18 Sept 2019, Gallery A

PAGE 2893 Wed-P-6-A-1 10:00-12:00	Blind Channel Response Estimation for Replay Attack Detection <i>Anderson R. Avila¹, Jahangir Alam², Douglas O'Shaughnessy¹, Tiago H. Falk¹</i> ¹ INRS-EMT, Canada; ² CRIM, Canada
PAGE 2898 Wed-P-6-A-2 10:00-12:00	Energy Separation-Based Instantaneous Frequency Estimation for Cochlear Cepstral Feature for Replay Spoof Detection <i>Ankur T. Patil¹, Rajul Acharya¹, Pulikonda Aditya Sai², Hemant A. Patil¹</i> ¹ DA-IICT, India; ² IIT Vadodara, India
PAGE 2903 Wed-P-6-A-3 10:00-12:00	Optimization of False Acceptance/Rejection Rates and Decision Threshold for End-to-End Text-Dependent Speaker Verification Systems <i>Victoria Mingote, Antonio Miguel, Dayana Ribas, Alfonso Ortega, Eduardo Lleida,</i> <i>Universidad de Zaragoza, Spain</i>
PAGE 2908 Wed-P-6-A-4 10:00-12:00	Deep Hashing for Speaker Identification and Retrieval <i>Lei Fan, Qing-Yuan Jiang, Ya-Qi Yu, Wu-Jun Li, Nanjing University, China</i>
PAGE 2913 Wed-P-6-A-5 10:00-12:00	Adversarial Optimization for Dictionary Attacks on Speaker Verification <i>Mirko Marras¹, Pawel Korus², Nasir Memon², Gianni Fenu¹</i> ¹ Università di Cagliari, Italy; ² New York University, USA
PAGE 2918 Wed-P-6-A-6 10:00-12:00	An Adaptive-Q Cochlear Model for Replay Spoofing Detection <i>Tharshini Gunendradasan¹, Eliathamby Ambikairajah¹, Julien Epps¹, Haizhou Li²</i> ¹ UNSW, Australia; ² NUS, Singapore
PAGE 2923 Wed-P-6-A-7 10:00-12:00	An End-to-End Text-Independent Speaker Verification Framework with a Keyword Adversarial Network <i>Sungrack Yun, Janghoon Cho, Jungyun Eum, Wonil Chang, Kyuwoong Hwang,</i> <i>Qualcomm, Korea</i>

Wed-P-6-A continued...

PAGE 2928 Wed-P-6-A-8 10:00-12:00	Shortcut Connections Based Deep Speaker Embeddings for End-to-End Speaker Verification System <i>Soonshin Seo, Daniel Jun Rim, Minkyu Lim, Donghyun Lee, Hosung Park, Junseok Oh,</i> <i>Changmin Kim, Ji-Hwan Kim, Sogang University, Korea</i>
PAGE 2933 Wed-P-6-A-9 10:00-12:00	Device Feature Extractor for Replay Spoofing Detection <i>Chang Huai You¹, Jichen Yang², Huy Dat Tran¹</i> ¹ A*STAR, Singapore; ² NUS, Singapore
PAGE 2938 Wed-P-6-A-10 10:00-12:00	Cross-Domain Replay Spoofing Attack Detection Using Domain Adversarial Training <i>Hongji Wang, Heinrich Dinkel, Shuai Wang, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China</i>
PAGE 2943 Wed-P-6-A-11 10:00-12:00	A Study of x-Vector Based Speaker Recognition on Short Utterances <i>A. Kanagasundaram¹, S. Sridharan², G. Sriram³, S. Prachi³, C. Fookes²</i> ¹ University of Jaffna, Sri Lanka; ² QUT, Australia; ³ Indian Institute of Science, India
PAGE 2948 Wed-P-6-A-12 10:00-12:00	Tied Mixture of Factor Analyzers Layer to Combine Frame Level Representations in Neural Speaker Embeddings <i>Nanxin Chen, Jesús Villalba, Najim Dehak, Johns Hopkins University, USA</i>
PAGE 2953 Wed-P-6-A-13 10:00-12:00	Biologically Inspired Adaptive-Q Filterbanks for Replay Spoofing Attack Detection <i>Buddhi Wickramasinghe, Eliathamby Ambikairajah, Julien Epps, UNSW, Australia</i>
PAGE 2958 Wed-P-6-A-14 10:00-12:00	On Robustness of Unsupervised Domain Adaptation for Speaker Recognition <i>Pierre-Michel Bousquet, Mickael Rouvier, LIA (EA 4128), France</i>
PAGE 2963 Wed-P-6-A-15 10:00-12:00	Large-Scale Speaker Retrieval on Random Speaker Variability Subspace <i>Suwon Shon¹, Younggun Lee², Taesu Kim²</i> ¹ MIT, USA; ² Nexasience, Korea

Wed-P-6-B: Rich Transcription and ASR Systems

Chair: Xunying Liu

Time 10:00-12:00, Wednesday 18 Sept 2019, Gallery B

PAGE 2968 Wed-P-6-B-1 10:00-12:00	Meeting Transcription Using Asynchronous Distant Microphones <i>Takuya Yoshioka, Dimitrios Dimitriadis, Andreas Stolcke, William Hinthorn, Zhuo Chen, Michael Zeng, Xuedong Huang, Microsoft, USA</i>
PAGE 2973 Wed-P-6-B-2 10:00-12:00	Detection and Recovery of OOVs for Improved English Broadcast News Captioning <i>Samuel Thomas, Kartik Audhkhasi, Zoltán Tüske, Yinghui Huang, Michael Picheny, IBM, USA</i>
PAGE 2978 Wed-P-6-B-3 10:00-12:00	Improving Large Vocabulary Urdu Speech Recognition System Using Deep Neural Networks <i>Muhammad Umar Farooq, Farah Adeeba, Sahar Rauf, Sarmad Hussain, UET Lahore, Pakistan</i>
PAGE 2983 Wed-P-6-B-4 10:00-12:00	Hybrid Arbitration Using Raw ASR String and NLU Information — Taking the Best of Both Embedded World and Cloud World <i>Min Tang, Nuance Communications, USA</i>
PAGE 2988 Wed-P-6-B-5 10:00-12:00	Leveraging a Character, Word and Prosody Triplet for an ASR Error Robust and Agglutination Friendly Punctuation Approach <i>György Szaszák¹, Máté Ákos Tündik²</i> <i>¹Telepathy Labs, Switzerland; ²BME, Hungary</i>
PAGE 2993 Wed-P-6-B-6 10:00-12:00	The Airbus Air Traffic Control Speech Recognition 2018 Challenge: Towards ATC Automatic Transcription and Call Sign Detection <i>Thomas Pellegrini¹, Jérôme Farinas¹, Estelle Delpech², François Lancelot²</i> <i>¹IRIT (UMR 5505), France; ²Airbus, France</i>

Wed-P-6-B continued...

PAGE 2998 Wed-P-6-B-7 10:00-12:00	KITE: Automatic Speech Recognition for Unmanned Aerial Vehicles <i>Dan Oneață, Horia Cucu, UPB, Romania</i>
PAGE 3003 Wed-P-6-B-8 10:00-12:00	Exploring Methods for the Automatic Detection of Errors in Manual Transcription <i>Xiaofei Wang, Jinyi Yang, Ruizhi Li, Samik Sadhu, Hynek Hermansky, Johns Hopkins University, USA</i>
PAGE 3008 Wed-P-6-B-9 10:00-12:00	Improved Low-Resource Somali Speech Recognition by Semi-Supervised Acoustic and Language Model Training <i>Astik Biswas, Raghav Menon, Ewald van der Westhuizen, Thomas Niesler, Stellenbosch University, South Africa</i>
PAGE 3013 Wed-P-6-B-10 10:00-12:00	The Althingi ASR System <i>Inga R. Helgadóttir, Anna Björk Nikulásdóttir, Michal Borský, Judy Y. Fong, Róbert Kjaran, Jón Guðnason, Reykjavik University, Iceland</i>
PAGE 3018 Wed-P-6-B-11 10:00-12:00	CRIM's Speech Transcription and Call Sign Detection System for the ATC Airbus Challenge Task <i>Vishwa Gupta, Lise Rebout, Gilles Boulianne, Pierre-André Ménard, Jahangir Alam, CRIM, Canada</i>

Wed-P-6-C : Speech and Language Analytics for Medical Applications

Chair: Daniel Bone

Time 10:00-12:00, Wednesday 18 Sept 2019, Gallery C

- PAGE 3023
Wed-P-6-C-1
10:00-12:00
- Optimizing Speech-Input Length for Speaker-Independent Depression Classification**
Tomasz Rutowski¹, Amir Harati², Yang Lu², Elizabeth Shriberg³
¹Ellipsis Health, Poland; ²Ellipsis Health, Canada; ³Ellipsis Health, USA
- PAGE 3028
Wed-P-6-C-2
10:00-12:00
- A New Approach for Automating Analysis of Responses on Verbal Fluency Tests from Subjects At-Risk for Schizophrenia**
Mary Pietrowicz¹, Carla Agurto¹, Raquel Norel¹, Elif Eyigoz¹, Guillermo Cecchi¹, Zarina R. Bilgrami², Cheryl Corcoran²
¹IBM, USA; ²Icahn School of Medicine at Mount Sinai, USA
- PAGE 3033
Wed-P-6-C-3
10:00-12:00
- Comparison of Telephone Recordings and Professional Microphone Recordings for Early Detection of Parkinson's Disease, Using Mel-Frequency Cepstral Coefficients with Gaussian Mixture Models**
Laetitia Jeancolas¹, Graziella Mangone², Jean-Christophe Corvol², Marie Vidailhet², Stéphane Lehericy², Badr-Eddine Benkelfat¹, Habib Benali³, Dijana Petrovska-Delacrétaz¹
¹SAMOVAR (UMR 5157), France; ²ICM (UMRS 1127 UMR 7225), France; ³Concordia University, Canada
- PAGE 3038
Wed-P-6-C-4
10:00-12:00
- Spectral Subspace Analysis for Automatic Assessment of Pathological Speech Intelligibility**
Parvaneh Janbakhshi, Ina Kodrasi, Hervé Bourlard, Idiap Research Institute, Switzerland

Wed-P-6-C continued...

- PAGE 3043
Wed-P-6-C-5
10:00-12:00
- An Investigation of Therapeutic Rapport Through Prosody in Brief Psychodynamic Psychotherapy**
Carolina De Pasquale¹, Charlie Cullen², Brian Vaughan¹
¹Technological University Dublin, Ireland; ²University of the West of Scotland, UK
- PAGE 3048
Wed-P-6-C-6
10:00-12:00
- Feature Representation of Pathophysiology of Parkinsonian Dysarthria**
Alice Rueda¹, J.C. Vásquez-Correa², Cristian David Rios-Urrego², Juan Rafael Orozco-Arroyave², Sridhar Krishnan¹, Elmar Nöth³
¹Ryerson University, Canada; ²Universidad de Antioquia, Colombia; ³FAU Erlangen-Nürnberg, Germany
- PAGE 3053
Wed-P-6-C-7
10:00-12:00
- Neural Transfer Learning for Cry-Based Diagnosis of Perinatal Asphyxia**
Charles C. Onu, Jonathan Lebensold, William L. Hamilton, Doina Precup, McGill University, Canada
- PAGE 3058
Wed-P-6-C-8
10:00-12:00
- Investigating the Variability of Voice Quality and Pain Levels as a Function of Multiple Clinical Parameters**
Hui-Ting Hong¹, Jeng-Lin Li¹, Yi-Ming Weng², Chip-Jin Ng², Chi-Chun Lee¹
¹National Tsing Hua University, Taiwan; ²Chang Gung Memorial Hospital, Taiwan
- PAGE 3063
Wed-P-6-C-9
10:00-12:00
- Assessing Parkinson's Disease from Speech Using Fisher Vectors**
José Vicente Egas López¹, Juan Rafael Orozco-Arroyave², Gábor Gosztolya¹
¹University of Szeged, Hungary; ²Universidad de Antioquia, Colombia
- PAGE 3068
Wed-P-6-C-10
10:00-12:00
- Feature Space Visualization with Spatial Similarity Maps for Pathological Speech Data**
Philipp Klumpp, J.C. Vásquez-Correa, Tino Haderlein, Elmar Nöth, FAU Erlangen-Nürnberg, Germany

Wed-P-6-C continued...

PAGE 3073
Wed-P-6-C-11
10:00-12:00

Predicting Behavior in Cancer-Afflicted Patient and Spouse Interactions Using Speech and Language

Sandeep Nallan Chakravarthula¹, Haoqi Li¹, Shao-Yen Tseng¹, Maija Reblin², Panayiotis Georgiou¹

¹University of Southern California, USA; ²Moffitt Cancer Center, USA

PAGE 3078
Wed-P-6-C-12
10:00-12:00

Automatic Assessment of Language Impairment Based on Raw ASR Output

Ying Qin¹, Tan Lee¹, Anthony Pak Hin Kong²

¹CUHK, China; ²University of Central Florida, USA

Wed-P-6-D: Speech Perception in Adverse Listening Conditions

Chair: Nao Hodoshima

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 10/D

PAGE 3083
Wed-P-6-D-1
10:00-12:00

Effects of Spectral and Temporal Cues to Mandarin Concurrent-Vowels Identification for Normal-Hearing and Hearing-Impaired Listeners

Zhen Fu, Xihong Wu, Jing Chen, Peking University, China

PAGE 3088
Wed-P-6-D-2
10:00-12:00

Disfluencies and Human Speech Transcription Errors

Vicky Zayats, Trang Tran, Richard Wright, Courtney Mansfield, Mari Ostendorf, University of Washington, USA

PAGE 3093
Wed-P-6-D-3
10:00-12:00

The Influence of Distraction on Speech Processing: How Selective is Selective Attention?

Sandra I. Parhammer¹, Miriam Ebersberg¹, Jenny Tippmann², Katja Stärk³, Andreas Opitz⁴, Barbara Hinger⁵, Sonja Rossi¹

¹Medizinische Universität Innsbruck, Austria; ²Technische Universität Dresden, Germany; ³MPI for Psycholinguistics, The Netherlands; ⁴Universität Leipzig, Germany; ⁵Universität Innsbruck, Austria

PAGE 3098
Wed-P-6-D-4
10:00-12:00

Subjective Evaluation of Communicative Effort for Younger and Older Adults in Interactive Tasks with Energetic and Informational Masking

Valerie Hazan, Outi Tuomainen, Linda Taschenberger, University College London, UK

PAGE 3103
Wed-P-6-D-5
10:00-12:00

Perceiving Older Adults Producing Clear and Lombard Speech

Chris Davis, Jeeseun Kim, Western Sydney University, Australia

PAGE 3108
Wed-P-6-D-6
10:00-12:00

Phone-Attribute Posteriors to Evaluate the Speech of Cochlear Implant Users

T. Arias-Vergara¹, Juan Rafael Orozco-Arroyave¹, Milos Cernak², S. Gollwitzer³, M. Schuster³, Elmar Nöth⁴

¹Universidad de Antioquia, Colombia; ²Logitech, Switzerland; ³LMU München, Germany; ⁴FAU Erlangen-Nürnberg, Germany

INTERSPEECH 2019 Table of Contents

Wed-P-6-D continued...

PAGE 3113 Wed-P-6-D-7 10:00-12:00	Effects of Urgent Speech and Congruent/Incongruent Text on Speech Intelligibility in Noise and Reverberation <i>Nao Hodoshima, Tokai University, Japan</i>
PAGE 3118 Wed-P-6-D-8 10:00-12:00	Quantifying Cochlear Implant Users' Ability for Speaker Identification Using CI Auditory Stimuli <i>Nursadul Mamun, Ria Ghosh, John H.L. Hansen, University of Texas at Dallas, USA</i>
PAGE 3123 Wed-P-6-D-9 10:00-12:00	Lexically Guided Perceptual Learning of a Vowel Shift in an Interactive L2 Listening Context <i>E. Felker, Mirjam Ernestus, Mirjam Broersma, Radboud Universiteit Nijmegen, The Netherlands</i>
PAGE 3128 Wed-P-6-D-10 10:00-12:00	Talker Intelligibility and Listening Effort with Temporally Modified Speech <i>Maximilian Paulus, Valerie Hazan, Patti Adank, University College London, UK</i>
PAGE 3133 Wed-P-6-D-11 10:00-12:00	R²SPIN: Re-Recording the Revised Speech Perception in Noise Test <i>Lauren Ward¹, Catherine Robinson², Matthew Paradis³, Katherine M. Tucker², Ben Shirley¹</i> <i>¹University of Salford, UK; ²BBC, UK; ³American Academy of Dramatic Arts, USA</i>
PAGE 3138 Wed-P-6-D-12 10:00-12:00	Contributions of Consonant-Vowel Transitions to Mandarin Tone Identification in Simulated Electric-Acoustic Hearing <i>Fei Chen, SUSTech, China</i>

INTERSPEECH 2019 Table of Contents

Wed-P-6-E: Speech Enhancement: Single Channel 1

Chair: Rainer Martin

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 10/E

PAGE 3143 Wed-P-6-E-1 10:00-12:00	Monaural Speech Enhancement with Dilated Convolutions <i>Shadi Pirhosseinloo, Jonathan S. Brumberg, University of Kansas, USA</i>
PAGE 3148 Wed-P-6-E-2 10:00-12:00	Noise Adaptive Speech Enhancement Using Domain Adversarial Training <i>Chien-Feng Liao¹, Yu Tsao¹, Hung-Yi Lee², Hsin-Min Wang¹</i> <i>¹Academia Sinica, Taiwan; ²National Taiwan University, Taiwan</i>
PAGE 3153 Wed-P-6-E-3 10:00-12:00	Environment-Dependent Attention-Driven Recurrent Convolutional Neural Network for Robust Speech Enhancement <i>Meng Ge¹, Longbiao Wang¹, Nan Li¹, Hao Shi¹, Jianwu Dang¹, Xiangang Li²</i> <i>¹Tianjin University, China; ²DiDi Chuxing, China</i>
PAGE 3158 Wed-P-6-E-4 10:00-12:00	A Statistically Principled and Computationally Efficient Approach to Speech Enhancement Using Variational Autoencoders <i>Manuel Pariente, Antoine Deleforge, Emmanuel Vincent, Loria (UMR 7503), France</i>
PAGE 3163 Wed-P-6-E-5 10:00-12:00	Speech Enhancement Using Forked Generative Adversarial Networks with Spectral Subtraction <i>Ju Lin¹, Sufeng Niu², Zice Wei¹, Xiang Lan¹, Adriaan J. van Wijngaarden³, Melissa C. Smith¹, Kuang-Ching Wang¹</i> <i>¹Clemson University, USA; ²LinkedIn, USA; ³Nokia Bell Labs, USA</i>
PAGE 3168 Wed-P-6-E-6 10:00-12:00	Specialized Speech Enhancement Model Selection Based on Learned Non-Intrusive Quality Assessment Metric <i>Ryandhimas E. Zezario¹, Szu-Wei Fu¹, Xugang Lu², Hsin-Min Wang¹, Yu Tsao¹</i> <i>¹Academia Sinica, Taiwan; ²NICT, Japan</i>

INTERSPEECH 2019 Table of Contents

Wed-P-6-E continued...

- PAGE 3173
Wed-P-6-E-7
10:00-12:00
- Speaker-Aware Deep Denoising Autoencoder with Embedded Speaker Identity for Speech Enhancement**
Fu-Kai Chuang¹, Syu-Siang Wang¹, Jieh-weih Hung², Yu Tsao³, Shih-Hau Fang¹
¹Yuan Ze University, Taiwan; ²National Chi Nan University, Taiwan; ³Academia Sinica, Taiwan
- PAGE 3178
Wed-P-6-E-8
10:00-12:00
- Investigation of Cost Function for Supervised Monaural Speech Separation**
Yun Liu¹, Hui Zhang¹, Xueliang Zhang¹, Yuhang Cao²
¹Inner Mongolia University, China; ²Unisound, China
- PAGE 3183
Wed-P-6-E-9
10:00-12:00
- Deep Attention Gated Dilated Temporal Convolutional Networks with Intra-Parallel Convolutional Modules for End-to-End Monaural Speech Separation**
Ziqiang Shi¹, Huibin Lin¹, Liu Liu¹, Rujie Liu¹, Jiqing Han², Anyan Shi³
¹Fujitsu, China; ²Harbin Institute of Technology, China; ³ShuangFeng First, China
- PAGE 3188
Wed-P-6-E-10
10:00-12:00
- Masking Estimation with Phase Restoration of Clean Speech for Monaural Speech Enhancement**
Xianyun Wang, Changchun Bao, Beijing University of Technology, China
- PAGE 3193
Wed-P-6-E-11
10:00-12:00
- Progressive Speech Enhancement with Residual Connections**
Jorge Llombart, Dayana Ribas, Antonio Miguel, Luis Vicente, Alfonso Ortega, Eduardo Lleida, Universidad de Zaragoza, Spain

INTERSPEECH 2019 Table of Contents

Wed-P-7-B: Speech Recognition and Beyond

Chair: Langzhou Chen

Time 13:30-15:30, Wednesday 18 Sept 2019, Gallery B

- PAGE 3198
Wed-P-7-B-1
13:30-15:30
- Acoustic Model Bootstrapping Using Semi-Supervised Learning**
Langzhou Chen¹, Volker Leutnant²
¹Amazon, UK; ²Amazon, Germany
- PAGE 3203
Wed-P-7-B-2
13:30-15:30
- Bandwidth Embeddings for Mixed-Bandwidth Speech Recognition**
Gautam Mantena, Ozlem Kalinli, Ossama Abdel-Hamid, Don McAllaster, Apple, USA
- PAGE 3208
Wed-P-7-B-3
13:30-15:30
- Adversarial Black-Box Attacks on Automatic Speech Recognition Systems Using Multi-Objective Evolutionary Optimization**
Shreya Khare¹, Rahul Aralikkatte², Senthil Mani¹
¹IBM, India; ²University of Copenhagen, Denmark
- PAGE 3213
Wed-P-7-B-4
13:30-15:30
- Towards Debugging Deep Neural Networks by Generating Speech Utterances**
Bilal Soomro, Anssi Kanervisto, Trung Ngo Trong, Ville Hautamäki, University of Eastern Finland, Finland
- PAGE 3218
Wed-P-7-B-5
13:30-15:30
- Compression of CTC-Trained Acoustic Models by Dynamic Frame-Wise Distillation or Segment-Wise N-Best Hypotheses Imitation**
Haisong Ding¹, Kai Chen², Qiang Huo²
¹USTC, China; ²Microsoft, China
- PAGE 3223
Wed-P-7-B-6
13:30-15:30
- Keyword Spotting for Hearing Assistive Devices Robust to External Speakers**
Iván López-Espejo, Zheng-Hua Tan, Jesper Jensen, Aalborg University, Denmark
- PAGE 3228
Wed-P-7-B-7
13:30-15:30
- Latent Dirichlet Allocation Based Acoustic Data Selection for Automatic Speech Recognition**
Mortaza Doulaty¹, Thomas Hain²
¹Microsoft, Germany; ²University of Sheffield, UK

Wed-P-7-B continued...

PAGE 3233
Wed-P-7-B-8
13:30-15:30

Target Speaker Recovery and Recognition Network with Average x-Vector and Global Training

Wenjie Li, Pengyuan Zhang, Yonghong Yan, Chinese Academy of Sciences, China

PAGE 3238
Wed-P-7-B-9
13:30-15:30

Lyrics Recognition from Singing Voice Focused on Correspondence Between Voice and Notes

Motoyuki Suzuki, Sho Tomita, Tomoki Morita, Osaka Institute of Technology, Japan

PAGE 3242
Wed-P-7-B-10
13:30-15:30

Transfer Learning from Audio-Visual Grounding to Speech Recognition

Wei-Ning Hsu, David Harwath, James Glass, MIT, USA

Wed-P-7-C : Emotion Modeling and Analysis

Chair: Tan Lee

Time 13:30-15:30, Wednesday 18 Sept 2019, Gallery C

PAGE 3247
Wed-P-7-C-1
13:30-15:30

Cross-Corpus Speech Emotion Recognition Using Semi-Supervised Transfer Non-Negative Matrix Factorization with Adaptation Regularization

Hui Luo, Jiqing Han, Harbin Institute of Technology, China

PAGE 3252
Wed-P-7-C-2
13:30-15:30

Modeling User Context for Valence Prediction from Narratives

Aniruddha Tammewar¹, Alessandra Cervone¹, Eva-Maria Messner², Giuseppe Riccardi¹

¹Università di Trento, Italy; ²Universität Ulm, Germany

PAGE 3257
Wed-P-7-C-3
13:30-15:30

Front-End Feature Compensation and Denoising for Noise Robust Speech Emotion Recognition

Rupayan Chakraborty, Ashish Panda, Meghna Pandharipande, Sonal Joshi, Sunil Kumar Kopparapu, TCS Innovation Labs Mumbai, India

PAGE 3262
Wed-P-7-C-4
13:30-15:30

The Contribution of Acoustic Features Analysis to Model Emotion Perceptual Process for Language Diversity

Xingfeng Li, Masato Akagi, JAIST, Japan

PAGE 3267
Wed-P-7-C-5
13:30-15:30

Design and Development of a Multi-Lingual Speech Corpora (TaMaR-EmoDB) for Emotion Analysis

Rajeev Rajan, Haritha U. G., Sujitha A. C., Rejisha T. M., College of Engineering Trivandrum, India

PAGE 3272
Wed-P-7-C-6
13:30-15:30

Speech Emotion Recognition with a Reject Option

Kusha Sridhar, Carlos Busso, University of Texas at Dallas, USA

PAGE 3277
Wed-P-7-C-7
13:30-15:30

Development of Emotion Rankers Based on Intended and Perceived Emotion Labels

Zhenghao Jin, Houwei Cao, New York Institute of Technology, USA

Wed-P-7-C continued...

PAGE 3282
Wed-P-7-C-8
13:30-15:30

Emotion Recognition from Natural Phone Conversations in Individuals with and without Recent Suicidal Ideation

John Gideon¹, Heather T. Schatten², Melvin G. McInnis¹, Emily Mower Provost¹
¹University of Michigan, USA; ²Brown University, USA

PAGE 3287
Wed-P-7-C-9
13:30-15:30

An Acoustic and Lexical Analysis of Emotional Valence in Spontaneous Speech: Autobiographical Memory Recall in Older Adults

Deniece S. Nazareth¹, Ellen Tournier², Sarah Leimkötter¹, Esther Janse², Dirk Heylen¹, Gerben J. Westerhof¹, Khiet P. Truong¹
¹Universiteit Twente, The Netherlands; ²Radboud Universiteit Nijmegen, The Netherlands

PAGE 3292
Wed-P-7-C-10
13:30-15:30

Does the Lombard Effect Improve Emotional Communication in Noise? — Analysis of Emotional Speech Acted in Noise

Yi Zhao¹, Atsushi Ando², Shinji Takaki¹, Junichi Yamagishi¹, Satoshi Kobashikawa²
¹NII, Japan; ²NTT, Japan

PAGE 3297
Wed-P-7-C-11
13:30-15:30

Linear Discriminant Differential Evolution for Feature Selection in Emotional Speech Recognition

Soumaya Gharsellaoui, Sid Ahmed Selouani, Mohammed Sidi Yakoub, Université de Moncton, Canada

PAGE 3302
Wed-P-7-C-12
13:30-15:30

Multi-Modal Learning for Speech Emotion Recognition: An Analysis and Comparison of ASR Outputs with Ground Truth Transcription

Saurabh Sahu, Vikramjit Mitra, Nadee Seneviratne, Carol Espy-Wilson, University of Maryland at College Park, USA

Wed-P-7-D: Articulatory Phonetics

Chair: To be confirmed

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 10/D

PAGE 3307
Wed-P-7-D-1
13:30-15:30

Articulatory Characteristics of Secondary Palatalization in Romanian Fricatives

Laura Spinu¹, Maida Percival², Alexei Kochetov²
¹CUNY Kingsborough Community College, USA; ²University of Toronto, Canada

PAGE 3312
Wed-P-7-D-2
13:30-15:30

Articulation of Vowel Length Contrasts in Australian English

Louise Ratko, Michael Proctor, Felicity Cox, Macquarie University, Australia

PAGE 3317
Wed-P-7-D-3
13:30-15:30

V-to-V Coarticulation Induced Acoustic and Articulatory Variability of Vowels: The Effect of Pitch-Accent

Andrea Deme¹, Márton Bartók¹, Tekla Etelka Grácsi², Tamás Gábor Csapó³, Alexandra Markó¹
¹ELTE, Hungary; ²MTA-ELTE LingArt, Hungary; ³BME, Hungary

PAGE 3322
Wed-P-7-D-4
13:30-15:30

The Contribution of Lip Protrusion to Anglo-English /r/: Evidence from Hyper- and Non-Hyperarticulated Speech

Hannah King¹, Emmanuel Ferragne²
¹CLILLAC-ARP (EA 3967), France; ²LPP (UMR 7018), France

PAGE 3327
Wed-P-7-D-5
13:30-15:30

Articulatory Analysis of Transparent Vowel /i:/ in Harmonic and Antiharmonic Hungarian Stems: Is There a Difference?

Alexandra Markó¹, Márton Bartók¹, Tamás Gábor Csapó², Tekla Etelka Grácsi², Andrea Deme¹
¹ELTE, Hungary; ²MTA-ELTE LingArt, Hungary

PAGE 3332
Wed-P-7-D-6
13:30-15:30

On the Role of Oral Configurations in European Portuguese Nasal Vowels

Conceição Cunha¹, Samuel Silva², António Teixeira², Catarina Oliveira², Paula Martins², Arun A. Joseph³, Jens Frahm³
¹LMU München, Germany; ²Universidade de Aveiro, Portugal; ³MPI for Biophysical Chemistry, Germany

Wed-P-7-E: Speech and Audio Classification 2

Chair: *To be confirmed*

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 10/E

- PAGE 3337
Wed-P-7-E-1
13:30-15:30
- Residual + Capsule Networks (ResCap) for Simultaneous Single-Channel Overlapped Keyword Recognition**
Yan Xiong, Visar Berisha, Chaitali Chakrabarti, Arizona State University, USA
- PAGE 3342
Wed-P-7-E-2
13:30-15:30
- A Study for Improving Device-Directed Speech Detection Toward Frictionless Human-Machine Interaction**
Che-Wei Huang, Roland Maas, Sri Harish Mallidi, Björn Hoffmeister, Amazon, USA
- PAGE 3347
Wed-P-7-E-3
13:30-15:30
- Unsupervised Methods for Audio Classification from Lecture Discussion Recordings**
Hang Su, Borislav Dzozdo, Xixin Wu, Xunying Liu, Helen Meng, CUHK, China
- PAGE 3352
Wed-P-7-E-4
13:30-15:30
- Neural Whispered Speech Detection with Imbalanced Learning**
Takanori Ashihara, Yusuke Shinohara, Hiroshi Sato, Takafumi Moriya, Kiyooki Matsui, Takaaki Fukutomi, Yoshikazu Yamaguchi, Yushi Aono, NTT, Japan
- PAGE 3357
Wed-P-7-E-5
13:30-15:30
- Deep Learning for Orca Call Type Identification — A Fully Unsupervised Approach**
Christian Bergler¹, Manuel Schmitt¹, Rachael Xi Cheng², Andreas Maier¹, Volker Barth³, Elmar Nöth¹
¹FAU Erlangen-Nürnberg, Germany; ²IZW in FVB, Germany; ³Anthro-Media, Germany
- PAGE 3362
Wed-P-7-E-6
13:30-15:30
- Open-Vocabulary Keyword Spotting with Audio and Text Embeddings**
Niccolò Sacchi¹, Alexandre Nanchen², Martin Jaggi¹, Milos Cernak³
¹EPFL, Switzerland; ²Idiap Research Institute, Switzerland; ³Logitech, Switzerland
- PAGE 3367
Wed-P-7-E-7
13:30-15:30
- ToneNet: A CNN Model of Tone Classification of Mandarin Chinese**
Qiang Gao, Shutao Sun, Yaping Yang, CUC, China

Wed-P-7-E continued...

- PAGE 3372
Wed-P-7-E-8
13:30-15:30
- Temporal Convolution for Real-Time Keyword Spotting on Mobile Devices**
Seungwoo Choi, Seokjun Seo, Beomjun Shin, Hyeongmin Byun, Martin Kersner, Beomsu Kim, Dongyoung Kim, Sungjoo Ha, Hyperconnect, Korea
- PAGE 3377
Wed-P-7-E-9
13:30-15:30
- Audio Tagging with Compact Feedforward Sequential Memory Network and Audio-to-Audio Ratio Based Data Augmentation**
Zhiying Huang, Shiliang Zhang, Ming Lei, Alibaba Group, China
- PAGE 3382
Wed-P-7-E-10
13:30-15:30
- Music Genre Classification Using Duplicated Convolutional Layers in Neural Networks**
Hansi Yang, Wei-Qiang Zhang, Tsinghua University, China
- PAGE 3387
Wed-P-7-E-11
13:30-15:30
- A Storyteller's Tale: Literature Audiobooks Genre Classification Using CNN and RNN Architectures**
Nehory Carmi, Azaria Cohen, Mireille Avigal, Anat Lerner, OUI, Israel

Wed-P-8-A: Speech Coding and Evaluation

Chair: Tom Bäckström

Time 16:00-18:00, Wednesday 18 Sept 2019, Gallery A

PAGE 3391 Wed-P-8-A-1 16:00-18:00	Parameter Enhancement for MELP Speech Codec in Noisy Communication Environment <i>Min-Jae Hwang¹, Hong-Goo Kang²</i> ¹ Search Solutions, Korea; ² Yonsei University, Korea
PAGE 3396 Wed-P-8-A-2 16:00-18:00	Cascaded Cross-Module Residual Learning Towards Lightweight End-to-End Speech Coding <i>Kai Zhen¹, Jongmo Sung², Mi Suk Lee², Seungkwon Beack², Minje Kim¹</i> ¹ Indiana University Bloomington, USA; ² ETRI, Korea
PAGE 3401 Wed-P-8-A-3 16:00-18:00	End-to-End Optimization of Source Models for Speech and Audio Coding Using a Machine Learning Framework <i>Tom Bäckström, Aalto University, Finland</i>
PAGE 3406 Wed-P-8-A-4 16:00-18:00	A Real-Time Wideband Neural Vocoder at 1.6kb/s Using LPCNet <i>Jean-Marc Valin¹, Jan Skoglund²</i> ¹ Amazon, USA; ² Google, USA
PAGE 3411 Wed-P-8-A-5 16:00-18:00	Super-Wideband Spectral Envelope Modeling for Speech Coding <i>Guillaume Fuchs¹, Chamran Ashour², Tom Bäckström³</i> ¹ Fraunhofer IIS, Germany; ² Ericsson, Sweden; ³ Aalto University, Finland
PAGE 3416 Wed-P-8-A-6 16:00-18:00	Speech Audio Super-Resolution for Speech Recognition <i>Xinyu Li, Venkata Chebiyyam, Katrin Kirchhoff, Amazon, USA</i>
PAGE 3421 Wed-P-8-A-7 16:00-18:00	Artificial Bandwidth Extension Using H^∞ Optimization <i>Deepika Gupta, Hanumant Singh Shekhawat, IIT Guwahati, India</i>

Wed-P-8-A continued...

PAGE 3426 Wed-P-8-A-8 16:00-18:00	Quality Degradation Diagnosis for Voice Networks — Estimating the Perceived Noisiness, Coloration, and Discontinuity of Transmitted Speech <i>Gabriel Mittag, Sebastian Möller, Technische Universität Berlin, Germany</i>
PAGE 3431 Wed-P-8-A-9 16:00-18:00	A Cross-Entropy-Guided (CEG) Measure for Speech Enhancement Front-End Assessing Performances of Back-End Automatic Speech Recognition <i>Li Chai¹, Jun Du¹, Chin-Hui Lee²</i> ¹ USTC, China; ² Georgia Tech, USA
PAGE 3436 Wed-P-8-A-10 16:00-18:00	Extending the E-Model Towards Super-Wideband and Fullband Speech Communication Scenarios <i>Sebastian Möller¹, Gabriel Mittag¹, Thilo Michael¹, Vincent Barriac², Hitoshi Aoki³</i> ¹ Technische Universität Berlin, Germany; ² Orange Labs, France; ³ NTT, Japan

Wed-P-8-B : Feature Extraction for ASR

Chair: Sriram Ganapathy

Time 16:00-18:00, Wednesday 18 Sept 2019, Gallery B

PAGE 3441 Wed-P-8-B-1 16:00-18:00	Modulation Vectors as Robust Feature Representation for ASR in Domain Mismatched Conditions <i>Samik Sadhu, Hynek Hermansky, Johns Hopkins University, USA</i>
PAGE 3446 Wed-P-8-B-2 16:00-18:00	Prosody Usage Optimization for Children Speech Recognition with Zero Resource Children Speech <i>Chenda Li, Yanmin Qian, Shanghai Jiao Tong University, China</i>
PAGE 3451 Wed-P-8-B-3 16:00-18:00	Unsupervised Raw Waveform Representation Learning for ASR <i>Purvi Agrawal, Sriram Ganapathy, Indian Institute of Science, India</i>
PAGE 3456 Wed-P-8-B-4 16:00-18:00	Low-Dimensional Bottleneck Features for On-Device Continuous Speech Recognition <i>David B. Ramsay¹, Kevin Kilgour², Dominik Roblek², Matthew Sharifi²</i> <i>¹MIT, USA; ²Google, Switzerland</i>
PAGE 3460 Wed-P-8-B-5 16:00-18:00	Binary Speech Features for Keyword Spotting Tasks <i>Alexandre Riviello, Jean-Pierre David, Polytechnique Montréal, Canada</i>
PAGE 3465 Wed-P-8-B-6 16:00-18:00	wav2vec: Unsupervised Pre-Training for Speech Recognition <i>Steffen Schneider, Alexei Baevski, Ronan Collobert, Michael Auli, Facebook, USA</i>
PAGE 3470 Wed-P-8-B-7 16:00-18:00	Automatic Detection of Prosodic Focus in American English <i>Sunghye Cho¹, Mark Liberman¹, Yong-cheol Lee²</i> <i>¹University of Pennsylvania, USA; ²Cheongju University, Korea</i>

Wed-P-8-B continued...

PAGE 3475 Wed-P-8-B-8 16:00-18:00	Feature Exploration for Almost Zero-Resource ASR-Free Keyword Spotting Using a Multilingual Bottleneck Extractor and Correspondence Autoencoders <i>Raghav Menon¹, Herman Kamper¹, Ewald van der Westhuizen¹, John Quinn², Thomas Niesler¹</i> <i>¹Stellenbosch University, South Africa; ²United Nations Global Pulse, Uganda</i>
PAGE 3480 Wed-P-8-B-9 16:00-18:00	On Learning Interpretable CNNs with Parametric Modulated Kernel-Based Filters <i>Erfan Loweimi, Peter Bell, Steve Renals, University of Edinburgh, UK</i>

Wed-P-8-C : Lexicon and Language Model for Speech Recognition

Chair: Xunying Liu

Time 16:00-18:00, Wednesday 18 Sept 2019, Gallery C

- PAGE 3485
Wed-P-8-C-1
16:00-18:00
- Reverse Transfer Learning: Can Word Embeddings Trained for Different NLP Tasks Improve Neural Language Models?**
Lyan Verwimp¹, Jerome R. Bellegarda²
¹Katholieke Universiteit Leuven, Belgium; ²Apple, USA
- PAGE 3490
Wed-P-8-C-2
16:00-18:00
- Joint Grapheme and Phoneme Embeddings for Contextual End-to-End ASR**
Zhehuai Chen¹, Mahaveer Jain², Yongqiang Wang², Michael L. Seltzer², Christian Fuegen²
¹Shanghai Jiao Tong University, China; ²Facebook, USA
- PAGE 3495
Wed-P-8-C-3
16:00-18:00
- Character-Aware Sub-Word Level Language Modeling for Uyghur and Turkish ASR**
Chang Liu¹, Zhen Zhang², Pengyuan Zhang¹, Yonghong Yan¹
¹Chinese Academy of Sciences, China; ²CNCERT, China
- PAGE 3500
Wed-P-8-C-4
16:00-18:00
- Connecting and Comparing Language Model Interpolation Techniques**
Ernest Pusateri, Christophe Van Gysel, Rami Botros, Sameer Badaskar, Mirko Hannemann, Youssef Oualil, Ilya Oparin, Apple, USA
- PAGE 3505
Wed-P-8-C-5
16:00-18:00
- Enriching Rare Word Representations in Neural Language Models by Embedding Matrix Augmentation**
Yerbolat Khassanov¹, Zhiping Zeng², Van Tung Pham¹, Haihua Xu², Eng Siong Chng¹
¹NTU, Singapore; ²Temasek Laboratories @ NTU, Singapore
- PAGE 3510
Wed-P-8-C-6
16:00-18:00
- Comparative Study of Parametric and Representation Uncertainty Modeling for Recurrent Neural Network Language Models**
Jianwei Yu, Max W.Y. Lam, Shoukang Hu, Xixin Wu, Xu Li, Yewen Cao, Xunying Liu, Helen Meng, CUHK, China

Wed-P-8-C continued...

- PAGE 3515
Wed-P-8-C-7
16:00-18:00
- Improving Automatically Induced Lexicons for Highly Agglutinating Languages Using Data-Driven Morphological Segmentation**
Wiehan Aagenbag, Thomas Niesler, Stellenbosch University, South Africa
- PAGE 3520
Wed-P-8-C-8
16:00-18:00
- Attention-Based Word Vector Prediction with LSTMs and its Application to the OOV Problem in ASR**
Alejandro Coucheiro-Limeres, Fernando Fernández-Martínez, Rubén San-Segundo, Javier Ferreiros-López, Universidad Politécnica de Madrid, Spain
- PAGE 3525
Wed-P-8-C-9
16:00-18:00
- Code-Switching Sentence Generation by Bert and Generative Adversarial Networks**
Yingying Gao, Junlan Feng, Ying Liu, Leijing Hou, Xin Pan, Yong Ma, China Mobile, China
- PAGE 3530
Wed-P-8-C-10
16:00-18:00
- Unified Verbalization for Speech Recognition & Synthesis Across Languages**
Sandy Ritchie¹, Richard Sproat², Kyle Gorman², Daan van Esch², Christian Schallhart¹, Nikos Bampounis¹, Benoît Brard¹, Jonas Fromseier Mortensen¹, Millie Holt¹, Eoin Mahon¹
¹Google, UK; ²Google, USA
- PAGE 3535
Wed-P-8-C-11
16:00-18:00
- Better Morphology Prediction for Better Speech Systems**
Dravyansh Sharma¹, Melissa Wilson², Antoine Bruguier³
¹Carnegie Mellon University, USA; ²University of Oklahoma, USA; ³Google, USA

Wed-P-8-D: First and Second Language Acquisition

Chair: To be confirmed

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 10/D

- PAGE 3540
Wed-P-8-D-1
16:00-18:00
- Vietnamese Learners Tackling the German /ft/ in Perception**
Anke Sennema¹, Silke Hamann²
¹Universität Wien, Austria; ²Universiteit van Amsterdam, The Netherlands
- PAGE 3544
Wed-P-8-D-2
16:00-18:00
- An Articulatory-Acoustic Investigation into GOOSE-Fronting in German-English Bilinguals Residing in London, UK**
Scott Lewis, Adib Mehrabi, Esther de Leeuw, Queen Mary University of London, UK
- PAGE 3549
Wed-P-8-D-3
16:00-18:00
- Multimodal Articulation-Based Pronunciation Error Detection with Spectrogram and Acoustic Features**
Sabrina Jenne, Ngoc Thang Vu, Universität Stuttgart, Germany
- PAGE 3554
Wed-P-8-D-4
16:00-18:00
- Using Prosody to Discover Word Order Alternations in a Novel Language**
Anouschka Foltz¹, Sarah Cooper², Tamsin M. McKelvey²
¹Universität Graz, Austria; ²Bangor University, UK
- PAGE 3559
Wed-P-8-D-5
16:00-18:00
- Speaking Rate, Information Density, and Information Rate in First-Language and Second-Language Speech**
Ann R. Bradlow, Northwestern University, USA
- PAGE 3564
Wed-P-8-D-6
16:00-18:00
- Articulation Rate as a Metric in Spoken Language Assessment**
Calbert Graham, Francis Nolan, University of Cambridge, UK
- PAGE 3569
Wed-P-8-D-7
16:00-18:00
- Learning Alignment for Multimodal Emotion Recognition from Speech**
Haiyang Xu¹, Hui Zhang¹, Kun Han², Yun Wang³, Yiping Peng¹, Xiangang Li¹
¹DiDi Chuxing, China; ²DiDi Research America, USA; ³Peking University, China

Wed-P-8-D continued...

- PAGE 3574
Wed-P-8-D-8
16:00-18:00
- Liquid Deletion in French Child-Directed Speech**
Sharon Peperkamp, Monica Hegde, Maria Julia Carbajal, LSCP (UMR 8554), France
- PAGE 3579
Wed-P-8-D-9
16:00-18:00
- Towards Detection of Canonical Babbling by Citizen Scientists: Performance as a Function of Clip Length**
Amanda Seidl¹, Anne S. Warlaumont², Alejandrina Cristia³
¹Purdue University, USA; ²University of California at Los Angeles, USA; ³LSCP (UMR 8554), France
- PAGE 3584
Wed-P-8-D-10
16:00-18:00
- Nasal Consonant Discrimination in Infant- and Adult-Directed Speech**
Bogdan Ludusan¹, Annett Jorschick¹, Reiko Mazuka²
¹Universität Bielefeld, Germany; ²RIKEN, Japan
- PAGE 3589
Wed-P-8-D-11
16:00-18:00
- No Distributional Learning in Adults from Attended Listening to Non-Speech**
Ellen Marklund, Johan Sjons, Lisa Gustavsson, Elisabet Eir Cortes, Stockholm University, Sweden
- PAGE 3594
Wed-P-8-D-12
16:00-18:00
- A Computational Model of Early Language Acquisition from Audiovisual Experiences of Young Infants**
Okko Räsänen, Khazar Khorrami, Tampere University, Finland
- PAGE 3599
Wed-P-8-D-13
16:00-18:00
- The Production of Chinese Affricates /ts/ and /ts^h/ by Native Urdu Speakers**
Dan Du, Jinsong Zhang, BLCU, China

Wed-P-8-E: Speech and Audio Classification 3

Chair: Johannes Gehrke

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 10/E

PAGE 3604 Wed-P-8-E-1 16:00-18:00	Multi-Stream Network with Temporal Attention for Environmental Sound Classification <i>Xinyu Li, Venkata Chebiyyam, Katrin Kirchhoff, Amazon, USA</i>
PAGE 3609 Wed-P-8-E-2 16:00-18:00	Neural Network Distillation on IoT Platforms for Sound Event Detection <i>Gianmarco Cerutti, Rahul Prasad, Alessio Brutti, Elisabetta Farella, FBK, Italy</i>
PAGE 3614 Wed-P-8-E-3 16:00-18:00	Class-Wise Centroid Distance Metric Learning for Acoustic Event Detection <i>Xugang Lu¹, Peng Shen¹, Sheng Li¹, Yu Tsao², Hisashi Kawai¹</i> <i>¹NICT, Japan; ²Academia Sinica, Taiwan</i>
PAGE 3619 Wed-P-8-E-4 16:00-18:00	A Hybrid Approach to Acoustic Scene Classification Based on Universal Acoustic Models <i>Xue Bai¹, Jun Du¹, Zi-Rui Wang¹, Chin-Hui Lee²</i> <i>¹USTC, China; ²Georgia Tech, USA</i>
PAGE 3624 Wed-P-8-E-5 16:00-18:00	Hierarchical Pooling Structure for Weakly Labeled Sound Event Detection <i>Ke-Xin He, Yu-Han Shen, Wei-Qiang Zhang, Tsinghua University, China</i>
PAGE 3629 Wed-P-8-E-6 16:00-18:00	Sound Event Detection in Multichannel Audio Using Convolutional Time-Frequency-Channel Squeeze and Excitation <i>Wei Xia¹, Kazuhito Koishida²</i> <i>¹University of Texas at Dallas, USA; ²Microsoft, USA</i>
PAGE 3634 Wed-P-8-E-7 16:00-18:00	A Robust Framework for Acoustic Scene Classification <i>Lam Pham, Ian McLoughlin, Huy Phan, Ramaswamy Palaniappan, University of Kent, UK</i>

Wed-P-8-E continued...

PAGE 3639 Wed-P-8-E-8 16:00-18:00	Compression of Acoustic Event Detection Models with Quantized Distillation <i>Bowen Shi¹, Ming Sun², Chieh-Chi Kao², Viktor Rozgic², Spyros Matsoukas², Chao Wang²</i> <i>¹TTIC, USA; ²Amazon, USA</i>
PAGE 3644 Wed-P-8-E-9 16:00-18:00	An End-to-End Audio Classification System Based on Raw Waveforms and Mix-Training Strategy <i>Jiaxu Chen, Jing Hao, Kai Chen, Di Xie, Shicai Yang, Shiliang Pu, Hikvision, China</i>
PAGE 3649 Wed-P-8-E-10 16:00-18:00	Few-Shot Audio Classification with Attentional Graph Neural Networks <i>Shilei Zhang, Yong Qin, Kewei Sun, Yonghua Lin, IBM, China</i>
PAGE 3654 Wed-P-8-E-11 16:00-18:00	Semi-Supervised Audio Classification with Consistency-Based Regularization <i>Kangkang Lu, Chuan-Sheng Foo, Kah Kuan Teh, Huy Dat Tran, Vijay Ramaseshan Chandrasekhar, A*STAR, Singapore</i>

Wed-S&T-3: Speech and Speaker Recognition

Time 10:00-12:00, Wednesday 18 Sept 2019, Hall 4

- PAGE 3659
Wed-S&T-3-1
10:00-12:00
- Avaya Conversational Intelligence: A Real-Time System for Spoken Language Understanding in Human-Human Call Center Conversations**
Jan Mizgajski, Adrian Szymczak, Robert Głowski, Piotr Szymański, Piotr Żelasko, Lukasz Augustyniak, Mikołaj Morzy, Yishay Carmiel, Jeff Hodson, Łukasz Wójciak, Daniel Smoczyk, Adam Wróbel, Bartosz Borowik, Adam Artajew, Marcin Baran, Cezary Kwiatkowski, Marzena Żyła-Hoppe, Avaya, USA
- PAGE 3661
Wed-S&T-3-2
10:00-12:00
- Robust Keyword Spotting via Recycle-Pooling for Mobile Game**
Shounan An, Youngsoo Kim, Hu Xu, Jinwoo Lee, Myungwoo Lee, Insoo Oh, Netmarble, Korea
- PAGE 3663
Wed-S&T-3-3
10:00-12:00
- Multimodal Dialog with the MALACH Audiovisual Archive**
Adam Chýlek, Luboš Šmídl, Jan Švec, University of West Bohemia, Czech Republic
- PAGE 3665
Wed-S&T-3-4
10:00-12:00
- SpeechMarker: A Voice Based Multi-Level Attendance Application**
*Sarfaraz Jelil¹, Abhishek Shrivastava¹, Rohan Kumar Das², S.R. Mahadeva Prasanna¹, Rohit Sinha¹
¹IIT Guwahati, India; ²NUS, Singapore*
- PAGE 3667
Wed-S&T-3-5
10:00-12:00
- Robust Sound Recognition: A Neuromorphic Approach**
*Jibin Wu¹, Zihan Pan¹, Malu Zhang¹, Rohan Kumar Das¹, Yansong Chua², Haizhou Li¹
¹NUS, Singapore; ²A*STAR, Singapore*
- PAGE 3669
Wed-S&T-3-6
10:00-12:00
- The CUHK Dysarthric Speech Recognition Systems for English and Cantonese**
*Shoukang Hu¹, Shansong Liu¹, Heng Fai Chang², Mengzhe Geng¹, Jiani Chen¹, Lau Wing Chung¹, To Ka Hei¹, Jianwei Yu¹, Ka Ho Wong¹, Xunying Liu¹, Helen Meng¹
¹CUHK, China; ²University of Cambridge, UK*

Wed-S&T-4: Speech Annotation and Labelling

Time 13:30-15:30, Wednesday 18 Sept 2019, Hall 4

- PAGE 3671
Wed-S&T-4-1
13:30-15:30
- BAS Web Services for Automatic Subtitle Creation and Anonymization**
Florian Schiel, Thomas Kisler, LMU München, Germany
- PAGE 3673
Wed-S&T-4-2
13:30-15:30
- A User-Friendly and Adaptable Re-Implementation of an Acoustic Prominence Detection and Annotation Tool**
Jana Voße, Petra Wagner, Universität Bielefeld, Germany
- PAGE 3675
Wed-S&T-4-3
13:30-15:30
- PyToBI: A Toolkit for ToBI Labeling Under Python**
Mónica Domínguez, Patrick Louis Rohrer, Juan Soler-Company, Universitat Pompeu Fabra, Spain
- PAGE 3677
Wed-S&T-4-4
13:30-15:30
- GECKO — A Tool for Effective Annotation of Human Conversations**
Golan Levy, Raquel Sitman, Ido Amir, Eduard Golshtein, Ran Mochary, Eilon Reshef, Roi Reichart, Omri Allouche, Gong.io, Israel
- PAGE 3679
Wed-S&T-4-5
13:30-15:30
- SLP-AA: Tools for Sign Language Phonetic and Phonological Research**
Roger Yu-Hsiang Lo, Kathleen Currie Hall, University of British Columbia, Canada
- PAGE 3681
Wed-S&T-4-6
13:30-15:30
- SANTLR: Speech Annotation Toolkit for Low Resource Languages**
Xinjian Li, Zhong Zhou, Siddharth Dalmia, Alan W. Black, Florian Metzger, Carnegie Mellon University, USA

Wed-S&T-5: Speech Synthesis

Time 16:00-18:00, Wednesday 18 Sept 2019, Hall 4

PAGE 3683
Wed-S&T-5-1
16:00-18:00

Web-Based Speech Synthesis Editor

Martin Grüber, Jakub Vít, Jindřich Matoušek, University of West Bohemia, Czech Republic

PAGE 3685
Wed-S&T-5-2
16:00-18:00

GFM-Voc: A Real-Time Voice Quality Modification System

Olivier Perrotin¹, Ian McLoughlin²
¹GIPSA-lab (UMR 5216), France; ²University of Kent, UK

PAGE 3687
Wed-S&T-5-3
16:00-18:00

Off the Cuff: Exploring Extemporaneous Speech Delivery with TTS

Éva Székely, Gustav Eje Henter, Jonas Beskow, Joakim Gustafson, KTH, Sweden

PAGE 3689
Wed-S&T-5-4
16:00-18:00

Synthesized Spoken Names: Biases Impacting Perception

Lucas Kessler, Cecilia Ovesdotter Alm, Reynold Bailey, RIT, USA

PAGE 3691
Wed-S&T-5-5
16:00-18:00

Unbabel Talk — Human Verified Translations for Voice Instant Messaging

Luís Bernardo¹, Mathieu Giquel¹, Sebastião Quintas², Paulo Dimas¹, Helena Moniz¹, Isabel Trancoso²
¹Unbabel, Portugal; ²Universidade de Lisboa, Portugal

PAGE 3693
Wed-S&T-5-6
16:00-18:00

Adjusting Pleasure-Arousal-Dominance for Continuous Emotional Text-to-Speech Synthesizer

Azam Rabiee, Tae-Ho Kim, Soo-Young Lee, KAIST, Korea

Thu-K-4: Keynote 4: Mirella Lapata

Chair: To be confirmed

Time 08:30-09:30, Thursday 19 Sept 2019, Main Hall

Thu-K-4
08:30-09:30

Learning Natural Language Interfaces with Neural Models

Mirella Lapata, University of Edinburgh, UK

Thu-SS-9-6: Privacy in Speech and Audio Interfaces

Chairs: Tom Bäckström, Stephan Sigg, Rainer Martin

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 3

PAGE 3695
Thu-SS-9-6-1
10:00-10:20

The GDPR & Speech Data: Reflections of Legal and Technology Communities, First Steps Towards a Common Understanding

Andreas Nautsch¹, Catherine Jasserand², Els Kindt³, Massimiliano Todisco¹, Isabel Trancoso⁴, Nicholas Evans¹

¹EURECOM, France; ²Rijksuniversiteit Groningen, The Netherlands; ³Katholieke Universiteit Leuven, Belgium; ⁴INESC-ID, Portugal

PAGE 3700
Thu-SS-9-6-2
10:20-10:40

Privacy-Preserving Adversarial Representation Learning in ASR: Reality or Illusion?

Brij Mohan Lal Srivastava¹, Aurélien Bellet¹, Marc Tommasi¹, Emmanuel Vincent²

¹CRISTAL (UMR 9189), France; ²Loria (UMR 7503), France

PAGE 3705
Thu-SS-9-6-3
10:40-11:00

Privacy-Preserving Siamese Feature Extraction for Gender Recognition versus Speaker Identification

Alexandru Nelus, Silas Rech, Timm Koppelman, Henrik Biermann, Rainer Martin, Ruhr-Universität Bochum, Germany

PAGE 3710
Thu-SS-9-6-4
11:00-11:20

Privacy-Preserving Variational Information Feature Extraction for Domestic Activity Monitoring versus Speaker Identification

Alexandru Nelus¹, Janek Ebberts², Reinhold Haeb-Umbach², Rainer Martin¹

¹Ruhr-Universität Bochum, Germany; ²Universität Paderborn, Germany

PAGE 3715
Thu-SS-9-6-5
11:20-11:40

Extracting Mel-Frequency and Bark-Frequency Cepstral Coefficients from Encrypted Signals

Patricia Thaine, Gerald Penn, University of Toronto, Canada

Thu-SS-9-6 continued...

PAGE 3720
Thu-SS-9-6-6
11:40-12:00

Sound Privacy: A Conversational Speech Corpus for Quantifying the Experience of Privacy

Pablo Pérez Zarazaga¹, Sneha Das¹, Tom Bäckström¹, V. Vidyadhara Raju V.², Anil Kumar Vuppala²

¹Aalto University, Finland; ²IIT Hyderabad, India

Thu-SS-10-5 : Speech Technologies for Code-Switching in Multilingual Communities

Chairs: Sunayana Sitaram, Alan W. Black

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 12

- PAGE 3725
Thu-0-10-5-1
13:30-13:47
- Improving Code-Switched Language Modeling Performance Using Cognate Features**
Victor Soto, Julia Hirschberg, Columbia University, USA
- PAGE 3730
Thu-0-10-5-2
13:47-14:04
- Linguistically Motivated Parallel Data Augmentation for Code-Switch Language Modeling**
Grande Lee, Xianghu Yue, Haizhou Li, NUS, Singapore
- PAGE 3735
Thu-0-10-5-3
14:04-14:21
- Variational Attention Using Articulatory Priors for Generating Code Mixed Speech Using Monolingual Corpora**
SaiKrishna Rallabandi, Alan W. Black, Carnegie Mellon University, USA
- PAGE 3740
Thu-0-10-5-4
14:21-14:38
- Code-Switching Detection Using ASR-Generated Language Posteriors**
Qinyi Wang, Emre Yilmaz, Adem Derinel, Haizhou Li, NUS, Singapore
- PAGE 3745
Thu-0-10-5-5
14:38-14:55
- Semi-Supervised Acoustic Model Training for Five-Lingual Code-Switched ASR**
Astik Biswas¹, Emre Yilmaz², Febe de Wet¹, Ewald van der Westhuizen¹, Thomas Niesler¹
¹Stellenbosch University, South Africa; ²NUS, Singapore
- PAGE 3750
Thu-0-10-5-6
14:55-15:12
- Multi-Graph Decoding for Code-Switching ASR**
Emre Yilmaz¹, Samuel Cohen¹, Xianghu Yue¹, David A. van Leeuwen², Haizhou Li¹
¹NUS, Singapore; ²Radboud Universiteit Nijmegen, The Netherlands
- PAGE 3755
Thu-0-10-5-7
15:12-15:30
- End-to-End Multilingual Multi-Speaker Speech Recognition**
Hiroshi Seki¹, Takaaki Hori¹, Shinji Watanabe², Jonathan Le Roux¹, John R. Hershey¹
¹MERL, USA; ²Johns Hopkins University, USA

Thu-O-9-1 : Speech Synthesis: Articulatory and Physical Approaches

Chairs: Paavo Alku, Frank Soong

Time 10:00-12:00, Thursday 19 Sept 2019, Main Hall

- Thu-O-9-1-1
10:00-10:40
- Survey Talk: Realistic Physics-Based Computational Voice Production**
Oriol Guasch, Universitat Ramon Llull, Spain
- PAGE 3760
Thu-O-9-1-2
10:40-11:00
- An Extended Two-Dimensional Vocal Tract Model for Fast Acoustic Simulation of Single-Axis Symmetric Three-Dimensional Tubes**
Debasish Ray Mohapatra¹, Victor Zappi², Sidney Fels¹
¹University of British Columbia, Canada; ²Northeastern University, USA
- PAGE 3765
Thu-O-9-1-3
11:00-11:20
- Perceptual Optimization of an Enhanced Geometric Vocal Fold Model for Articulatory Speech Synthesis**
Peter Birkholz¹, Susanne Drechsel², Simon Stone¹
¹Technische Universität Dresden, Germany; ²MLU Halle-Wittenberg, Germany
- PAGE 3770
Thu-O-9-1-4
11:20-11:40
- Articulatory Copy Synthesis Based on a Genetic Algorithm**
Yingming Gao, Simon Stone, Peter Birkholz, Technische Universität Dresden, Germany
- PAGE 3775
Thu-O-9-1-5
11:40-12:00
- A Phonetic-Level Analysis of Different Input Features for Articulatory Inversion**
Abdolreza Sabzi Shahrehabaki¹, Negar Olfati¹, Ali Shariq Imran¹, Sabato Marco Siniscalchi², Torbjørn Svendsen¹
¹NTNU, Norway; ²Università di Enna "Kore", Italy

Thu-O-9-2 : Sequence-to-Sequence Speech Recognition

Chairs: Jan Chorowski, Shinji Watanabe

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 1

- PAGE 3780
Thu-O-9-2-1
10:00-10:20
- Advancing Sequence-to-Sequence Based Speech Recognition**
Zoltán Tüske, Kartik Audhkhasi, George Saon, IBM, USA
- PAGE 3785
Thu-O-9-2-2
10:20-10:40
- Sequence-to-Sequence Speech Recognition with Time-Depth Separable Convolutions**
Awni Hannun, Ann Lee, Qiantong Xu, Ronan Collobert, Facebook, USA
- PAGE 3790
Thu-O-9-2-3
10:40-11:00
- Semi-Supervised Sequence-to-Sequence ASR Using Unpaired Speech and Text**
Murali Karthick Baskar¹, Shinji Watanabe², Ramon Astudillo³, Takaaki Hori⁴, Lukáš Burget¹, Jan Černocký¹
¹Brno University of Technology, Czech Republic; ²Johns Hopkins University, USA; ³IBM, USA; ⁴MERL, USA
- PAGE 3795
Thu-O-9-2-4
11:00-11:20
- Learn Spelling from Teachers: Transferring Knowledge from Language Models to Sequence-to-Sequence Speech Recognition**
Ye Bai, Jiangyan Yi, Jianhua Tao, Zhengkun Tian, Zhengqi Wen, Chinese Academy of Sciences, China
- PAGE 3800
Thu-O-9-2-5
11:20-11:40
- On the Choice of Modeling Unit for Sequence-to-Sequence Speech Recognition**
Kazuki Irie¹, Rohit Prabhavalkar², Anjali Kannan², Antoine Bruguier², David Rybach², Patrick Nguyen²
¹RWTH Aachen University, Germany; ²Google, USA
- PAGE 3805
Thu-O-9-2-6
11:40-12:00
- Listen, Attend, Spell and Adapt: Speaker Adapted Sequence-to-Sequence ASR**
Felix Weninger¹, Jesús Andrés-Ferrer², Xinwei Li¹, Puming Zhan¹
¹Nuance Communications, USA; ²Nuance Communications, Spain

Thu-O-9-3 : Search Methods for Speech Recognition

Chairs: Takaaki Hori, Murat Saraçlar

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 2

- PAGE 3810
Thu-O-9-3-1
10:00-10:20
- Lattice Re-Scoring During Manual Editing for Automatic Error Correction of ASR Transcripts**
Anna V. Rúnarsdóttir, Inga R. Helgadóttir, Jón Guðnason, Reykjavik University, Iceland
- PAGE 3815
Thu-O-9-3-2
10:20-10:40
- GPU-Based WFST Decoding with Extra Large Language Model**
Daisuke Fukunaga, Yoshiki Tanaka, Yuichi Kageyama, Sony, Japan
- PAGE 3820
Thu-O-9-3-3
10:40-11:00
- Real-Time One-Pass Decoder for Speech Recognition Using LSTM Language Models**
Javier Jorge, Adrià Giménez, Javier Iranzo-Sánchez, Jorge Civera, Albert Sanchis, Alfons Juan, Universidad Politécnica de Valencia, Spain
- PAGE 3825
Thu-O-9-3-4
11:00-11:20
- Vectorized Beam Search for CTC-Attention-Based Speech Recognition**
Hiroshi Seki¹, Takaaki Hori², Shinji Watanabe³, Niko Moritz², Jonathan Le Roux²
¹Toyohashi University of Technology, Japan; ²MERL, USA; ³Johns Hopkins University, USA
- PAGE 3830
Thu-O-9-3-5
11:20-11:40
- Contextual Recovery of Out-of-Lattice Named Entities in Automatic Speech Recognition**
Jack Serrino¹, Leonid Velikovich², Petar Aleksic², Cyril Allauzen²
¹MIT, USA; ²Google, USA
- PAGE 3835
Thu-O-9-3-6
11:40-12:00
- Sequence-to-Sequence Learning via Attention Transfer for Incremental Speech Recognition**
Sashi Novitasari, Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan

Thu-O-9-4: Audio Signal Characterization

Chairs: Franz Pernkopf, Reinhold Häb-Umbach

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 11

- PAGE 3840
Thu-O-9-4-1
10:00-10:20
- Unsupervised Representation Learning with Future Observation Prediction for Speech Emotion Recognition**
Zheng Lian, Jianhua Tao, Bin Liu, Jian Huang, Chinese Academy of Sciences, China
- PAGE 3845
Thu-O-9-4-2
10:20-10:40
- Spatio-Temporal Attention Pooling for Audio Scene Classification**
Huy Phan¹, Oliver Y. Chén², Lam Pham¹, Philipp Koch³, Maarten De Vos², Ian McLoughlin¹, Alfred Mertins³
¹University of Kent, UK; ²University of Oxford, UK; ³Universität zu Lübeck, Germany
- PAGE 3850
Thu-O-9-4-3
10:40-11:00
- Subspace Pooling Based Temporal Features Extraction for Audio Event Recognition**
Qiyu Shi, Hui Luo, Jiqing Han, Harbin Institute of Technology, China
- PAGE 3855
Thu-O-9-4-4
11:00-11:20
- Multi-Scale Time-Frequency Attention for Acoustic Event Detection**
Jingyang Zhang¹, Wenhao Ding¹, Jintao Kang², Liang He¹
¹Tsinghua University, China; ²Ministry of Public Security, China
- PAGE 3860
Thu-O-9-4-5
11:20-11:40
- Acoustic Scene Classification by Implicitly Identifying Distinct Sound Events**
Hongwei Song¹, Jiqing Han¹, Shiwen Deng², Zhihao Du¹
¹Harbin Institute of Technology, China; ²Harbin Normal University, China
- PAGE 3865
Thu-O-9-4-6
11:40-12:00
- Parameter-Transfer Learning for Low-Resource Individualization of Head-Related Transfer Functions**
Xiaoke Qi¹, Lu Wang²
¹CUPL, China; ²Shenzhen University, China

Thu-O-9-5: Speech and Voice Disorders 1

Chairs: Heidi Christensen, Phil Green

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 12

- PAGE 3870
Thu-O-9-5-1
10:00-10:20
- Prosodic Characteristics of Mandarin Declarative and Interrogative Utterances in Parkinson's Disease**
Lei Liu, Meng Jian, Wentao Gu, Nanjing Normal University, China
- PAGE 3875
Thu-O-9-5-2
10:20-10:40
- Study of the Performance of Automatic Speech Recognition Systems in Speakers with Parkinson's Disease**
Laureano Moro-Velazquez¹, JaeJin Cho¹, Shinji Watanabe¹, Mark A. Hasegawa-Johnson², Odette Scharenborg³, Heejin Kim², Najim Dehak¹
¹Johns Hopkins University, USA; ²University of Illinois at Urbana-Champaign, USA; ³Technische Universiteit Delft, The Netherlands
- PAGE 3880
Thu-O-9-5-3
10:40-11:00
- Towards the Speech Features of Mild Cognitive Impairment: Universal Evidence from Structured and Unstructured Connected Speech of Chinese**
Tianqi Wang¹, Chongyuan Lian¹, Jingshen Pan¹, Quanlei Yan¹, Feiqi Zhu², Manwa L. Ng³, Lan Wang¹, Nan Yan¹
¹Chinese Academy of Sciences, China; ²Shenzhen Luohu People's Hospital, China; ³HKU, China
- PAGE 3885
Thu-O-9-5-4
11:00-11:20
- Child Speech Disorder Detection with Siamese Recurrent Network Using Speech Attribute Features**
Jiarui Wang, Ying Qin, Zhiyuan Peng, Tan Lee, CUHK, China
- PAGE 3890
Thu-O-9-5-5
11:20-11:40
- Interpretable Deep Learning Model for the Detection and Reconstruction of Dysarthric Speech**
Daniel Korzekwa¹, Roberto Barra-Chicote², Bozena Kostek³, Thomas Drugman⁴, Mateusz Lajszczak²
¹Amazon, Poland; ²Amazon, UK; ³Gdańsk University of Technology, Poland; ⁴Amazon, Germany

Thu-O-9-5 continued...

PAGE 3895
Thu-O-9-5-6
11:40-12:00

Vocal Biomarker Assessment Following Pediatric Traumatic Brain Injury: A Retrospective Cohort Study

Camille Noufi¹, Adam C. Lammert¹, Daryush D. Mehta¹, James R. Williamson¹, Gregory Ciccarelli¹, Douglas Sturim¹, Jordan R. Green², Thomas F. Campbell³, Thomas F. Quatieri¹

¹MIT Lincoln Laboratory, USA; ²MGH Institute of Health Professions, USA; ³University of Texas at Dallas, USA

Thu-O-10-1 : Neural Networks for Language Modeling

Chairs: Ariya Rastrow, Ralf Schlüter

Time 13:30-15:30, Thursday 19 Sept 2019, Main Hall

Thu-O-10-1-1
13:30-14:10

Survey Talk: Reaching Over the Gap: Cross- and Interdisciplinary Research on Human and Automatic Speech Processing

Odette Scharenborg, Technische Universiteit Delft, The Netherlands

PAGE 3900
Thu-O-10-1-2
14:10-14:30

Improved Deep Duel Model for Rescoring N-Best Speech Recognition List Using Backward LSTM and Ensemble Encoders

Atsunori Ogawa, Marc Delcroix, Shigeki Karita, Tomohiro Nakatani, NTT, Japan

PAGE 3905
Thu-O-10-1-3
14:30-14:50

Language Modeling with Deep Transformers

Kazuki Irie, Albert Zeyer, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

PAGE 3910
Thu-O-10-1-4
14:50-15:10

Scalable Multi Corpora Neural Language Models for ASR

Anirudh Raju, Denis Filimonov, Gautam Tiwari, Guitang Lan, Ariya Rastrow, Amazon, USA

PAGE 3915
Thu-O-10-1-5
15:10-15:30

Who Needs Words? Lexicon-Free Speech Recognition

Tatiana Likhomanenko, Gabriel Synnaeve, Ronan Collobert, Facebook, USA

Thu-O-10-2 : Representation Learning of Emotion and Paralinguistics

Chairs: Tatsuya Kawahara, Chung-Hsien Wu

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 1

- PAGE 3920
Thu-O-10-2-1
13:30-13:50
- Direct Modelling of Speech Emotion from Raw Speech**
Siddique Latif¹, Rajib Rana¹, Sara Khalifa², Raja Jurdak², Julien Epps³
¹University of Southern Queensland, Australia; ²CSIRO, Australia; ³UNSW, Australia
- PAGE 3925
Thu-O-10-2-2
13:50-14:10
- Improving Emotion Identification Using Phone Posteriors in Raw Speech Waveform Based DNN**
Mousmita Sarma¹, Pegah Ghahremani², Daniel Povey², Nagendra Kumar Goel³, Kandarpa Kumar Sarma¹, Najim Dehak²
¹Gauhati University, India; ²Johns Hopkins University, USA; ³GoVivace, USA
- PAGE 3930
Thu-O-10-2-3
14:10-14:30
- Pyramid Memory Block and Timestep Attention for Speech Emotion Recognition**
Miao Cao, Chun Yang, Fang Zhou, Xu-cheng Yin, USTB, China
- PAGE 3935
Thu-O-10-2-4
14:30-14:50
- Robust Speech Emotion Recognition Under Different Encoding Conditions**
Christopher Oates, Andreas Triantafyllopoulos, Ingmar Steiner, Björn W. Schuller, audEERING, Germany
- PAGE 3940
Thu-O-10-2-5
14:50-15:10
- Using the Bag-of-Audio-Word Feature Representation of ASR DNN Posteriors for Paralinguistic Classification**
Gábor Gosztolya, MTA-SZTE RGAI, Hungary
- PAGE 3945
Thu-O-10-2-6
15:10-15:30
- Disentangling Style Factors from Speaker Representations**
Jennifer Williams, Simon King, University of Edinburgh, UK

Thu-O-10-3 : World's Languages and Varieties

Chairs: Maria da Conceição Cunha, Pavel Šturm

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 2

- PAGE 3950
Thu-O-10-3-1
13:30-13:50
- Sentence Prosody and *Wh*-Indeterminates in Taiwan Mandarin**
Yu-Yin Hsu¹, Anqi Xu²
¹PolyU, China; ²University College London, UK
- PAGE 3955
Thu-O-10-3-2
13:50-14:10
- Frication as a Vowel Feature? — Evidence from the Rui'an Wu Chinese Dialect**
Fang Hu¹, Youjue He²
¹CASS, China; ²UCASS, China
- PAGE 3960
Thu-O-10-3-3
14:10-14:30
- Vowels and Diphthongs in the Xupu Xiang Chinese Dialect**
Zhenrui Zhang¹, Fang Hu²
¹UCASS, China; ²CASS, China
- PAGE 3965
Thu-O-10-3-4
14:30-14:50
- Age-Related Changes in European Portuguese Vowel Acoustics**
Luciana Albuquerque, Catarina Oliveira, António Teixeira, Pedro Sa-Couto, Daniela Figueiredo, Universidade de Aveiro, Portugal
- PAGE 3970
Thu-O-10-3-5
14:50-15:10
- Vowel-Tone Interaction in Two Tibeto-Burman Languages**
Wendy Lalhminghlu, Viyazonuo Terhijja, Priyankoo Sarmah, IIT Guwahati, India
- PAGE 3975
Thu-O-10-3-6
15:10-15:30
- The Vowel System of Korebaju**
Jenifer Vega Rodríguez, Université Sorbonne Nouvelle, France

Thu-O-10-4: Adaptation and Accommodation in Conversation

Chair: *To be confirmed*

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 11

- PAGE 3980
Thu-O-10-4-1
13:30-13:50
- Fundamental Frequency Accommodation in Multi-Party Human-Robot Game Interactions: The Effect of Winning or Losing**
Omnia Ibrahim¹, Gabriel Skantze², Sabine Stoll¹, Volker Dellwo¹
¹Universität Zürich, Switzerland; ²KTH, Sweden
- PAGE 3985
Thu-O-10-4-2
13:50-14:10
- Pitch Accent Trajectories Across Different Conditions of Visibility and Information Structure — Evidence from Spontaneous Dyadic Interaction**
Petra Wagner, Nataliya Bryhadyr, Marin Schröer, Universität Bielefeld, Germany
- PAGE 3990
Thu-O-10-4-3
14:10-14:30
- The Green Tree — Lengthening Position Influences Uncertainty Perception**
Simon Betz¹, Sina Zarrieß¹, Éva Székely², Petra Wagner¹
¹Universität Bielefeld, Germany; ²KTH, Sweden
- PAGE 3995
Thu-O-10-4-4
14:30-14:50
- CNN-BLSTM Based Question Detection from Dialogs Considering Phase and Context Information**
Yuke Si¹, Longbiao Wang¹, Jianwu Dang¹, Mengfei Wu¹, Aijun Li²
¹Tianjin University, China; ²CASS, China
- PAGE 4000
Thu-O-10-4-5
14:50-15:10
- Mirroring to Build Trust in Digital Assistants**
Katherine Metcalf, Barry-John Theobald, Garrett Weinberg, Robert Lee, Ing-Marie Jonsson, Russ Webb, Nicholas Apostoloff, Apple, USA
- PAGE 4005
Thu-O-10-4-6
15:10-15:30
- Three's a Crowd? Effects of a Second Human on Vocal Accommodation with a Voice Assistant**
Eran Raveh¹, Ingo Siegert², Ingmar Steiner³, Iona Gessinger¹, Bernd Möbius¹
¹Universität des Saarlandes, Germany; ²Otto-von-Guericke-Universität Magdeburg, Germany; ³audEERING, Germany

Thu-P-9-A: Speaker and Language Recognition 2

Chair: *Sandro Cumani*

Time 10:00-12:00, Thursday 19 Sept 2019, Gallery A

- PAGE 4010
Thu-P-9-A-1
10:00-12:00
- Adversarial Regularization for End-to-End Robust Speaker Verification**
Qing Wang¹, Pengcheng Guo¹, Sining Sun¹, Lei Xie¹, John H.L. Hansen²
¹Northwestern Polytechnical University, China; ²University of Texas at Dallas, USA
- PAGE 4015
Thu-P-9-A-2
10:00-12:00
- Combining Speaker Recognition and Metric Learning for Speaker-Dependent Representation Learning**
João Monteiro, Jahangir Alam, Tiago H. Falk, INRS-EMT, Canada
- PAGE 4020
Thu-P-9-A-3
10:00-12:00
- VAE-Based Regularization for Deep Speaker Embedding**
Yang Zhang, Lantian Li, Dong Wang, Tsinghua University, China
- PAGE 4025
Thu-P-9-A-4
10:00-12:00
- Language Recognition Using Triplet Neural Networks**
Victoria Mingote¹, Diego Castan², Mitchell McLaren², Mahesh Kumar Nandwana², Alfonso Ortega¹, Eduardo Lleida¹, Antonio Miguel¹
¹Universidad de Zaragoza, Spain; ²SRI International, USA
- PAGE 4030
Thu-P-9-A-5
10:00-12:00
- Spatial Pyramid Encoding with Convex Length Normalization for Text-Independent Speaker Verification**
Youngmoon Jung¹, Youngwan Kim², Hyungjun Lim¹, Yeunju Choi¹, Hoirin Kim¹
¹KAIST, Korea; ²LG Electronics, Korea
- PAGE 4035
Thu-P-9-A-6
10:00-12:00
- End-to-End Losses Based on Speaker Basis Vectors and All-Speaker Hard Negative Mining for Speaker Verification**
Hee-Soo Heo, Jee-weon Jung, Il-Ho Yang, Sung-Hyun Yoon, Hye-jin Shim, Ha-Jin Yu, University of Seoul, Korea
- PAGE 4040
Thu-P-9-A-7
10:00-12:00
- An Effective Deep Embedding Learning Architecture for Speaker Verification**
Yiheng Jiang¹, Yan Song¹, Ian McLoughlin², Zhifu Gao¹, Li-Rong Dai¹
¹USTC, China; ²University of Kent, UK

Thu-P-9-A continued...

PAGE 4045 Thu-P-9-A-8 10:00-12:00	Far-Field End-to-End Text-Dependent Speaker Verification Based on Mixed Training Data with Transfer Learning and Enrollment Data Augmentation <i>Xiaoyi Qin, Danwei Cai, Ming Li, Duke Kunshan University, China</i>
PAGE 4050 Thu-P-9-A-9 10:00-12:00	Two-Stage Training for Chinese Dialect Recognition <i>Zongze Ren, Guofu Yang, Shugong Xu, Shanghai University, China</i>
PAGE 4055 Thu-P-9-A-10 10:00-12:00	Investigation on Blind Bandwidth Extension with a Non-Linear Function and its Evaluation of x-Vector-Based Speaker Verification <i>Ryota Kaminishi, Haruna Miyamoto, Sayaka Shiota, Hitoshi Kiya, Tokyo Metropolitan University, Japan</i>
PAGE 4060 Thu-P-9-A-11 10:00-12:00	Auto-Encoding Nearest Neighbor i-Vectors for Speaker Verification <i>Umair Khan, Miquel India, Javier Hernando, Universitat Politècnica de Catalunya, Spain</i>
PAGE 4065 Thu-P-9-A-12 10:00-12:00	Towards a Fault-Tolerant Speaker Verification System: A Regularization Approach to Reduce the Condition Number <i>Siqi Zheng, Gang Liu, Hongbin Suo, Yun Lei, Alibaba Group, China</i>
PAGE 4070 Thu-P-9-A-13 10:00-12:00	Deep Learning Based Multi-Channel Speaker Recognition in Noisy and Reverberant Environments <i>Hassan Taherian, Zhong-Qiu Wang, DeLiang Wang, Ohio State University, USA</i>
PAGE 4075 Thu-P-9-A-14 10:00-12:00	Joint Optimization of Neural Acoustic Beamforming and Dereverberation with x-Vectors for Robust Speaker Verification <i>Joon-Young Yang, Joon-Hyuk Chang, Hanyang University, Korea</i>
PAGE 4080 Thu-P-9-A-15 10:00-12:00	A New Time-Frequency Attention Mechanism for TDNN and CNN-LSTM-TDNN, with Application to Language Identification <i>Xiaoxiao Miao¹, Ian McLoughlin¹, Yonghong Yan²</i> <i>¹University of Kent, UK; ²Chinese Academy of Sciences, China</i>

Thu-P-9-B: Medical Applications and Visual ASR

Chair: To be confirmed

Time 10:00-12:00, Thursday 19 Sept 2019, Gallery B

PAGE 4085 Thu-P-9-B-1 10:00-12:00	An Attention-Based Hybrid Network for Automatic Detection of Alzheimer's Disease from Narrative Speech <i>Jun Chen, Ji Zhu, Jieping Ye, University of Michigan, USA</i>
PAGE 4090 Thu-P-9-B-2 10:00-12:00	Investigating the Lombard Effect Influence on End-to-End Audio-Visual Speech Recognition <i>Pingchuan Ma, Stavros Petridis, Maja Pantic, Imperial College London, UK</i>
PAGE 4095 Thu-P-9-B-3 10:00-12:00	"Computer, Test My Hearing": Accurate Speech Audiometry with Smart Speakers <i>Jasper Ooster¹, Pia Nancy Porysek Moreta¹, Jörg-Hendrik Bach², Inga Holube³, Bernd T. Meyer¹</i> <i>¹Carl von Ossietzky Universität Oldenburg, Germany; ²HörTech, Germany; ³Jade Hochschule, Germany</i>
PAGE 4100 Thu-P-9-B-4 10:00-12:00	Synchronising Audio and Ultrasound by Learning Cross-Modal Embeddings <i>Aciel Eshky, Manuel Sam Ribeiro, Korin Richmond, Steve Renals, University of Edinburgh, UK</i>
PAGE 4105 Thu-P-9-B-5 10:00-12:00	Automatic Hierarchical Attention Neural Network for Detecting AD <i>Yilin Pan, Bahman Mirheidari, Markus Reuber, Annalena Venneri, Daniel Blackburn, Heidi Christensen, University of Sheffield, UK</i>
PAGE 4110 Thu-P-9-B-6 10:00-12:00	Deep Sensing of Breathing Signal During Conversational Speech <i>Venkata Srikanth Nallanthighal¹, Aki Härmä¹, Helmer Strik²</i> <i>¹Philips, The Netherlands; ²Radboud Universiteit Nijmegen, The Netherlands</i>

Thu-P-9-B continued...

PAGE 4115
Thu-P-9-B-7
10:00-12:00

Parrottron: An End-to-End Speech-to-Speech Conversion Model and its Applications to Hearing-Impaired Speech and Speech Separation

Fadi Biadisy, Ron J. Weiss, Pedro J. Moreno, Dimitri Kanvesky, Ye Jia, Google, USA

PAGE 4120
Thu-P-9-B-8
10:00-12:00

Exploiting Visual Features Using Bayesian Gated Neural Networks for Disordered Speech Recognition

Shansong Liu¹, Shoukang Hu¹, Yi Wang², Jianwei Yu¹, Rongfeng Su³, Xunying Liu¹, Helen Meng¹

¹CUHK, China; ²University of Cambridge, UK; ³Chinese Academy of Sciences, China

PAGE 4125
Thu-P-9-B-9
10:00-12:00

Video-Driven Speech Reconstruction Using Generative Adversarial Networks

Konstantinos Vougioukas, Pingchuan Ma, Stavros Petridis, Maja Pantic, Imperial College London, UK

PAGE 4130
Thu-P-9-B-10
10:00-12:00

On the Use of Pitch Features for Disordered Speech Recognition

Shansong Liu, Shoukang Hu, Xunying Liu, Helen Meng, CUHK, China

PAGE 4135
Thu-P-9-B-11
10:00-12:00

Large-Scale Visual Speech Recognition

Brendan Shillingford¹, Yannis Assael¹, Matthew W. Hoffman¹, Thomas Paine¹, Cían Hughes¹, Utsav Prabhu², Hank Liao², Hasim Sak², Kanishka Rao², Lorryne Bennett¹, Marie Mulville¹, Misha Denil¹, Ben Coppin¹, Ben Laurie¹, Andrew Senior¹, Nando de Freitas¹

¹DeepMind, UK; ²Google, USA

Thu-P-9-C: Turn Management in Dialogue

Chair: Helena Moniz

Time 10:00-12:00, Thursday 19 Sept 2019, Gallery C

PAGE 4140
Thu-P-9-C-1
10:00-12:00

Investigating Linguistic and Semantic Features for Turn-Taking Prediction in Open-Domain Human-Computer Conversation

S. Zahra Razavi, Benjamin Kane, Lenhart K. Schubert, University of Rochester, USA

PAGE 4145
Thu-P-9-C-2
10:00-12:00

Benchmarking Benchmarks: Introducing New Automatic Indicators for Benchmarking Spoken Language Understanding Corpora

Frédéric Béchet¹, Christian Raymond²

¹LIS (UMR 7020), France; ²IRISA (UMR 6074), France

PAGE 4150
Thu-P-9-C-3
10:00-12:00

A Neural Turn-Taking Model without RNN

Chaoran Liu, Carlos Ishi, Hiroshi Ishiguro, ATR HIL, Japan

PAGE 4155
Thu-P-9-C-4
10:00-12:00

An Incremental Turn-Taking Model for Task-Oriented Dialog Systems

Andrei C. Coman¹, Koichiro Yoshino², Yukitoshi Murase², Satoshi Nakamura², Giuseppe Riccardi¹

¹Università di Trento, Italy; ²NAIST, Japan

PAGE 4160
Thu-P-9-C-5
10:00-12:00

Personalized Dialogue Response Generation Learned from Monologues

Feng-Guang Su, Aliyah R. Hsu, Yi-Lin Tuan, Hung-Yi Lee, National Taiwan University, Taiwan

PAGE 4165
Thu-P-9-C-6
10:00-12:00

Voice Quality as a Turn-Taking Cue

Mattias Heldner¹, Marcin Włodarczak¹, Štefan Beňuš², Agustín Gravano³

¹Stockholm University, Sweden; ²UKF, Slovak Republic; ³UBA, Argentina

PAGE 4170
Thu-P-9-C-7
10:00-12:00

Turn-Taking Prediction Based on Detection of Transition Relevance Place

Kohei Hara, Koji Inoue, Katsuya Takanashi, Tatsuya Kawahara, Kyoto University, Japan

Thu-P-9-C continued...

PAGE 4175
Thu-P-9-C-8
10:00-12:00

Analysis of Effect and Timing of Fillers in Natural Turn-Taking
Divesh Lala, Shizuka Nakamura, Tatsuya Kawahara, Kyoto University, Japan

PAGE 4180
Thu-P-9-C-9
10:00-12:00

Multimodal Response Obligation Detection with Unsupervised Online Domain Adaptation
Shota Horiguchi, Naoyuki Kanda, Kenji Nagamatsu, Hitachi, Japan

PAGE 4185
Thu-P-9-C-10
10:00-12:00

Follow-Up Question Generation Using Neural Tensor Network-Based Domain Ontology Population in an Interview Coaching System
Ming-Hsiang Su, Chung-Hsien Wu, Yi Chang, National Cheng Kung University, Taiwan

Thu-P-9-D: Corpus Annotation and Evaluation

Chair: To be confirmed

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 10/D

PAGE 4190
Thu-P-9-D-1
10:00-12:00

On the Role of Style in Parsing Speech with Neural Models
Trang Tran¹, Jiahong Yuan², Yang Liu², Mari Ostendorf¹
¹University of Washington, USA; ²LAIX, USA

PAGE 4195
Thu-P-9-D-2
10:00-12:00

On the Contributions of Visual and Textual Supervision in Low-Resource Semantic Speech Retrieval
Ankita Pasad¹, Bowen Shi¹, Herman Kamper², Karen Livescu¹
¹TTIC, USA; ²Stellenbosch University, South Africa

PAGE 4200
Thu-P-9-D-3
10:00-12:00

Automatic Detection of Off-Topic Spoken Responses Using Very Deep Convolutional Neural Networks
Xinhao Wang, Su-Youn Yoon, Keelan Evanini, Klaus Zechner, Yao Qian, Educational Testing Service, USA

PAGE 4205
Thu-P-9-D-4
10:00-12:00

Rescoring Keyword Search Confidence Estimates with Graph-Based Re-Ranking Using Acoustic Word Embeddings
Anna Piunova, Eugen Beck, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

PAGE 4210
Thu-P-9-D-5
10:00-12:00

SpeechYOLO: Detection and Localization of Speech Objects
Yael Segal, Tzeviya Sylvia Fuchs, Joseph Keshet, Bar-Ilan University, Israel

PAGE 4215
Thu-P-9-D-6
10:00-12:00

Prosodic Phrase Alignment for Machine Dubbing
Alp Öktem¹, Mireia Farrús², Antonio Bonafonte³
¹Col-lectivaT, Spain; ²Universitat Pompeu Fabra, Spain; ³Universitat Politècnica de Catalunya, Spain

INTERSPEECH 2019 Table of Contents

Thu-P-9-D continued...

PAGE 4220
Thu-P-9-D-7
10:00-12:00

Spot the Pleasant People! Navigating the Cocktail Party Buzz
Christina Tännander, Per Fallgren, Jens Edlund, Joakim Gusafsson, KTH, Sweden

PAGE 4225
Thu-P-9-D-8
10:00-12:00

Neural Text Clustering with Document-Level Attention Based on Dynamic Soft Labels
Zhi Chen, Wu Guo, Li-Rong Dai, Zhen-Hua Ling, Jun Du, USTC, China

PAGE 4230
Thu-P-9-D-9
10:00-12:00

Noisy BiLSTM-Based Models for Disfluency Detection
Nguyen Bach, Fei Huang, Alibaba Group, USA

PAGE 4235
Thu-P-9-D-10
10:00-12:00

Subword RNNLM Approximations for Out-Of-Vocabulary Keyword Search
Mittul Singh¹, Sami Virpioja², Peter Smit¹, Mikko Kurimo¹
¹Aalto University, Finland; ²University of Helsinki, Finland

PAGE 4240
Thu-P-9-D-11
10:00-12:00

Simultaneous Detection and Localization of a Wake-Up Word Using Multi-Task Learning of the Duration and Endpoint
Takashi Maekaku, Yusuke Kida, Akihiko Sugiyama, Yahoo, Japan

INTERSPEECH 2019 Table of Contents

Thu-P-9-E: Speech Enhancement: Multi-Channel and Intelligibility

Chair: Antonio Peinado

Time 10:00-12:00, Thursday 19 Sept 2019, Hall 10/E

PAGE 4245
Thu-P-9-E-1
10:00-12:00

On Mitigating Acoustic Feedback in Hearing Aids with Frequency Warping by All-Pass Networks
Ching-Hua Lee, Kuan-Lin Chen, fred harris, Bhaskar D. Rao, Harinath Garudadri, University of California at San Diego, USA

PAGE 4250
Thu-P-9-E-2
10:00-12:00

Deep Multitask Acoustic Echo Cancellation
Amin Fazel, Mostafa El-Khamy, Jungwon Lee, Samsung, USA

PAGE 4255
Thu-P-9-E-3
10:00-12:00

Deep Learning for Joint Acoustic Echo and Noise Cancellation with Nonlinear Distortions
Hao Zhang, Ke Tan, DeLiang Wang, Ohio State University, USA

PAGE 4260
Thu-P-9-E-4
10:00-12:00

Harmonic Beamformers for Non-Intrusive Speech Intelligibility Prediction
Charlotte Sørensen¹, Jesper B. Boldt², Mads G. Christensen¹
¹Aalborg University, Denmark; ²GN Hearing, Denmark

PAGE 4265
Thu-P-9-E-5
10:00-12:00

Convolutional Neural Network-Based Speech Enhancement for Cochlear Implant Recipients
Nursadul Mamun, Soheil Khorram, John H.L. Hansen, University of Texas at Dallas, USA

PAGE 4270
Thu-P-9-E-6
10:00-12:00

Validation of the Non-Intrusive Codebook-Based Short Time Objective Intelligibility Metric for Processed Speech
Charlotte Sørensen¹, Jesper B. Boldt², Mads G. Christensen¹
¹Aalborg University, Denmark; ²GN Hearing, Denmark

INTERSPEECH 2019 Table of Contents

Thu-P-9-E continued...

PAGE 4275
Thu-P-9-E-7
10:00-12:00

Predicting Speech Intelligibility of Enhanced Speech Using Phone Accuracy of DNN-Based ASR System

Kenichi Arai¹, Shoko Araki¹, Atsunori Ogawa¹, Keisuke Kinoshita¹, Tomohiro Nakatani¹, Katsuhiko Yamamoto², Toshio Irino²
¹NTT, Japan; ²Wakayama University, Japan

PAGE 4280
Thu-P-9-E-8
10:00-12:00

A Novel Method to Correct Steering Vectors in MVDR Beamformer for Noise Robust ASR

Suliang Bu¹, Yunxin Zhao¹, Mei-Yuh Hwang²
¹University of Missouri, USA; ²Mobvoi, USA

PAGE 4285
Thu-P-9-E-9
10:00-12:00

End-to-End Multi-Channel Speech Enhancement Using Inter-Channel Time-Restricted Attention on Raw Waveform

Hyeonseung Lee, Hyung Yong Kim, Woo Hyun Kang, Jeunghun Kim, Nam Soo Kim,
Seoul National University, Korea

PAGE 4290
Thu-P-9-E-10
10:00-12:00

Neural Spatial Filter: Target Speaker Speech Separation Assisted with Directional Information

Rongzhi Gu¹, Lianwu Chen², Shi-Xiong Zhang³, Jimeng Zheng², Yong Xu³, Meng Yu³, Dan Su², Yuexian Zou¹, Dong Yu³
¹Peking University, China; ²Tencent, China; ³Tencent, USA

PAGE 4295
Thu-P-9-E-11
10:00-12:00

My Lips Are Concealed: Audio-Visual Speech Enhancement Through Obstructions

Triantafyllos Afouras, Joon Son Chung, Andrew Zisserman, University of Oxford, UK

INTERSPEECH 2019 Table of Contents

Thu-P-10-A: Speaker Recognition 3

Chair: Hynek Boril

Time 13:30-15:30, Thursday 19 Sept 2019, Gallery A

PAGE 4300
Thu-P-10-A-1
13:30-15:30

End-to-End Neural Speaker Diarization with Permutation-Free Objectives

Yusuke Fujita¹, Naoyuki Kanda¹, Shota Horiguchi¹, Kenji Nagamatsu¹, Shinji Watanabe²
¹Hitachi, Japan; ²Johns Hopkins University, USA

PAGE 4305
Thu-P-10-A-2
13:30-15:30

Self Multi-Head Attention for Speaker Recognition

Miquel India, Pooyan Safari, Javier Hernando, Universitat Politècnica de Catalunya, Spain

PAGE 4310
Thu-P-10-A-3
13:30-15:30

Phonetically-Aware Embeddings, Wide Residual Networks with Time-Delay Neural Networks and Self Attention Models for the 2018 NIST Speaker Recognition Evaluation

Ignacio Viñals, Dayana Ribas, Victoria Mingote, Jorge Llombart, Pablo Gimeno, Antonio Miguel, Alfonso Ortega, Eduardo Lleida, Universidad de Zaragoza, Spain

PAGE 4315
Thu-P-10-A-4
13:30-15:30

Variational Domain Adversarial Learning for Speaker Verification

Youzhi Tu¹, Man-Wai Mak¹, Jen-Tzung Chien²
¹PolyU, China; ²National Chiao Tung University, Taiwan

PAGE 4320
Thu-P-10-A-5
13:30-15:30

A Unified Framework for Speaker and Utterance Verification

Tianchi Liu, Maulik Madhavi, Rohan Kumar Das, Haizhou Li, NUS, Singapore

PAGE 4325
Thu-P-10-A-6
13:30-15:30

Analysis of Critical Metadata Factors for the Calibration of Speaker Recognition Systems

Mahesh Kumar Nandwana¹, Luciana Ferrer², Mitchell McLaren¹, Diego Castan¹, Aaron Lawson¹
¹SRI International, USA; ²UBA, Argentina

INTERSPEECH 2019 Table of Contents

Thu-P-10-A continued...

PAGE 4330
Thu-P-10-A-7
13:30-15:30

Factorization of Discriminatively Trained i-Vector Extractor for Speaker Recognition

Ondřej Novotný, Oldřich Plchot, Ondřej Glembek, Lukáš Burget, Brno University of Technology, Czech Republic

PAGE 4335
Thu-P-10-A-8
13:30-15:30

End-to-End Speaker Identification in Noisy and Reverberant Environments Using Raw Waveform Convolutional Neural Networks

Daniele Salvati, Carlo Drioli, Gian Luca Foresti, Università di Udine, Italy

PAGE 4340
Thu-P-10-A-9
13:30-15:30

Whisper to Neutral Mapping Using Cosine Similarity Maximization in i-Vector Space for Speaker Verification

Abinay Reddy Naini, Achuth Rao M.V., Prasanta Kumar Ghosh, Indian Institute of Science, India

PAGE 4345
Thu-P-10-A-10
13:30-15:30

Mixup Learning Strategies for Text-Independent Speaker Verification

*Yingke Zhu¹, Tom Ko², Brian Mak¹
¹HKUST, China; ²SUSTech, China*

PAGE 4350
Thu-P-10-A-11
13:30-15:30

Optimizing a Speaker Embedding Extractor Through Backend-Driven Regularization

*Luciana Ferrer¹, Mitchell McLaren²
¹UBA, Argentina; ²SRI International, USA*

PAGE 4355
Thu-P-10-A-12
13:30-15:30

The NEC-TT 2018 Speaker Verification System

*Kong Aik Lee¹, Hitoshi Yamamoto¹, Koji Okabe¹, Qionggiong Wang¹, Ling Guo¹, Takafumi Koshinaka¹, Jiacen Zhang², Koichi Shinoda²
¹NEC, Japan; ²Tokyo Tech, Japan*

INTERSPEECH 2019 Table of Contents

Thu-P-10-A continued...

PAGE 4360
Thu-P-10-A-13
13:30-15:30

Autoencoder-Based Semi-Supervised Curriculum Learning for Out-of-Domain Speaker Verification

Siqi Zheng, Gang Liu, Hongbin Suo, Yun Lei, Alibaba Group, China

PAGE 4365
Thu-P-10-A-14
13:30-15:30

Multi-Channel Training for End-to-End Speaker Recognition Under Reverberant and Noisy Environment

Danwei Cai, Xiaoyi Qin, Ming Li, Duke Kunshan University, China

PAGE 4370
Thu-P-10-A-15
13:30-15:30

The DKU-SMIIP System for NIST 2018 Speaker Recognition Evaluation

Danwei Cai, Weicheng Cai, Ming Li, Duke Kunshan University, China

Thu-P-10-B: NN Architectures for ASR

Chair: Martin Karafiát

Time 13:30-15:30, Thursday 19 Sept 2019, Gallery B

- PAGE 4375
Thu-P-10-B-1
13:30-15:30
- Pretraining by Backtranslation for End-to-End ASR in Low-Resource Settings**
Matthew Wiesner, Adithya Renduchintala, Shinji Watanabe, Chunxi Liu, Najim Dehak, Sanjeev Khudanpur, Johns Hopkins University, USA
- PAGE 4380
Thu-P-10-B-2
13:30-15:30
- Cross-Attention End-to-End ASR for Two-Party Conversations**
Suyoun Kim, Siddharth Dalmia, Florian Metzger, Carnegie Mellon University, USA
- PAGE 4385
Thu-P-10-B-3
13:30-15:30
- Towards Using Context-Dependent Symbols in CTC Without State-Tying Decision Trees**
Jan Chorowski, Adrian Łańcucki, Bartosz Kostka, Michał Zopotoczny, University of Wrocław, Poland
- PAGE 4390
Thu-P-10-B-4
13:30-15:30
- An Online Attention-Based Model for Speech Recognition**
Ruchao Fan¹, Pan Zhou², Wei Chen³, Jia Jia², Gang Liu¹
¹BUPT, China; ²Tsinghua University, China; ³Sogou, China
- PAGE 4395
Thu-P-10-B-5
13:30-15:30
- Self-Attention Transducers for End-to-End Speech Recognition**
Zhengkun Tian, Jiangyan Yi, Jianhua Tao, Ye Bai, Zhengqi Wen, Chinese Academy of Sciences, China
- PAGE 4400
Thu-P-10-B-6
13:30-15:30
- Improving Transformer-Based Speech Recognition Systems with Compressed Structure and Speech Attributes Augmentation**
Sheng Li, Dabre Raj, Xugang Lu, Peng Shen, Tatsuya Kawahara, Hisashi Kawai, NICT, Japan
- PAGE 4405
Thu-P-10-B-7
13:30-15:30
- Extending an Acoustic Data-Driven Phone Set for Spontaneous Speech Recognition**
Jeong-Uk Bang¹, Mu-Yeol Choi², Sang-Hun Kim², Oh-Wook Kwon¹
¹Chungbuk National University, Korea; ²ETRI, Korea

Thu-P-10-B continued...

- PAGE 4410
Thu-P-10-B-8
13:30-15:30
- Joint Maximization Decoder with Neural Converters for Fully Neural Network-Based Japanese Speech Recognition**
Takafumi Moriya, Jian Wang, Tomohiro Tanaka, Ryo Masumura, Yusuke Shinohara, Yoshikazu Yamaguchi, Yushi Aono, NTT, Japan
- PAGE 4415
Thu-P-10-B-9
13:30-15:30
- Real to H-Space Encoder for Speech Recognition**
Titouan Parcollet¹, Mohamed Morchid¹, Georges Linarès¹, Renato De Mori²
¹LIA (EA 4128), France; ²McGill University, Canada
- PAGE 4420
Thu-P-10-B-10
13:30-15:30
- ECTC-DOCD: An End-to-End Structure with CTC Encoder and OCD Decoder for Speech Recognition**
Cheng Yi, Feng Wang, Bo Xu, Chinese Academy of Sciences, China
- PAGE 4425
Thu-P-10-B-11
13:30-15:30
- End-to-End Multi-Speaker Speech Recognition Using Speaker Embeddings and Transfer Learning**
Pavel Denisov, Ngoc Thang Vu, Universität Stuttgart, Germany

Thu-P-10-C : Speech Synthesis: Text Processing, Prosody, and Emotion

Chair: Éva Székely

Time 13:30-15:30, Thursday 19 Sept 2019, Gallery C

- PAGE 4430
Thu-P-10-C-1
13:30-15:30
- Pre-Trained Text Embeddings for Enhanced Text-to-Speech Synthesis**
*Tomoki Hayashi¹, Shinji Watanabe², Tomoki Toda¹, Kazuya Takeda¹,
Shubham Toshniwal³, Karen Livescu³*
¹Nagoya University, Japan; ²Johns Hopkins University, USA; ³TTIC, USA
- PAGE 4435
Thu-P-10-C-2
13:30-15:30
- Spontaneous Conversational Speech Synthesis from Found Data**
Éva Székely, Gustav Eje Henter, Jonas Beskow, Joakim Gustafson, KTH, Sweden
- PAGE 4440
Thu-P-10-C-3
13:30-15:30
- Fine-Grained Robust Prosody Transfer for Single-Speaker Neural Text-To-Speech**
Viacheslav Klimkov¹, Srikanth Ronanki², Jonas Rohnke², Thomas Drugman¹
¹Amazon, Germany; ²Amazon, UK
- PAGE 4445
Thu-P-10-C-4
13:30-15:30
- Speech Driven Backchannel Generation Using Deep Q-Network for Enhancing Engagement in Human-Robot Interaction**
Nusrah Hussain, Ençin Erzin, T. Metin Sezgin, Yücel Yemez, Koç Üniversitesi, Turkey
- PAGE 4450
Thu-P-10-C-5
13:30-15:30
- Semi-Supervised Prosody Modeling Using Deep Gaussian Process Latent Variable Model**
Tomoki Koriyama¹, Takao Kobayashi²
¹University of Tokyo, Japan; ²Tokyo Tech, Japan
- PAGE 4455
Thu-P-10-C-6
13:30-15:30
- Bootstrapping a Text Normalization System for an Inflected Language. Numbers as a Test Case**
Anna Björk Nikulásdóttir, Jón Guðnason, Reykjavik University, Iceland

Thu-P-10-C continued...

- PAGE 4460
Thu-P-10-C-7
13:30-15:30
- Exploiting Syntactic Features in a Parsed Tree to Improve End-to-End TTS**
Haohan Guo¹, Frank K. Soong², Lei He², Lei Xie¹
¹Northwestern Polytechnical University, China; ²Microsoft, China
- PAGE 4465
Thu-P-10-C-8
13:30-15:30
- Duration Modeling with Global Phoneme-Duration Vectors**
Jinfu Ni, Yoshinori Shiga, Hisashi Kawai, NICT, Japan
- PAGE 4470
Thu-P-10-C-9
13:30-15:30
- Improving Speech Synthesis with Discourse Relations**
Adèle Aubin¹, Alessandra Cervone², Oliver Watts¹, Simon King¹
¹University of Edinburgh, UK; ²Università di Trento, Italy
- PAGE 4475
Thu-P-10-C-10
13:30-15:30
- Visualization and Interpretation of Latent Spaces for Controlling Expressive Speech Synthesis Through Audio Analysis**
Noé Tits¹, Fengna Wang², Kevin El Haddad¹, Vincent Pagel², Thierry Dutoit¹
¹Université de Mons, Belgium; ²Acapela Group, Belgium
- PAGE 4480
Thu-P-10-C-11
13:30-15:30
- Pre-Trained Text Representations for Improving Front-End Text Processing in Mandarin Text-to-Speech Synthesis**
Bing Yang, Jiaqi Zhong, Shan Liu, Tencent, China
- PAGE 4485
Thu-P-10-C-12
13:30-15:30
- A Mandarin Prosodic Boundary Prediction Model Based on Multi-Task Learning**
Huashan Pan, Xiulin Li, Zhiqiang Huang, Databaker Technology, China
- PAGE 4489
Thu-P-10-C-13
13:30-15:30
- Dual Encoder Classifier Models as Constraints in Neural Text Normalization**
Ajda Gokcen¹, Hao Zhang², Richard Sproat²
¹University of Washington, USA; ²Google, USA

Thu-P-10-C continued...

PAGE 4494
Thu-P-10-C-14
13:30-15:30

Knowledge-Based Linguistic Encoding for End-to-End Mandarin Text-to-Speech Synthesis

Jingbei Li¹, Zhiyong Wu¹, Runnan Li¹, Pengpeng Zhi², Song Yang², Helen Meng³
¹Tsinghua University, China; ²TAL, China; ³CUHK, China

PAGE 4499
Thu-P-10-C-15
13:30-15:30

Automated Emotion Morphing in Speech Based on Diffeomorphic Curve Registration and Highway Networks

Ravi Shankar, Hsi-Wei Hsieh, Nicolas Charon, Archana Venkataraman, Johns Hopkins University, USA

Thu-P-10-D: Speech and Voice Disorders 2

Chair: Laureano Moro-Velazquez

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 10/D

PAGE 4504
Thu-P-10-D-1
13:30-15:30

Use of Beiwe Smartphone App to Identify and Track Speech Decline in Amyotrophic Lateral Sclerosis (ALS)

Kathryn P. Connaghan¹, Jordan R. Green¹, Sabrina Paganoni², James Chan³, Harli Weber³, Ella Collins³, Brian Richburg¹, Marziye Eshghi¹, J.P. Onnela², James D. Berry²

¹MGH Institute of Health Professions, USA; ²Harvard University, USA; ³Massachusetts General Hospital, USA

PAGE 4509
Thu-P-10-D-2
13:30-15:30

Profiling Speech Motor Impairments in Persons with Amyotrophic Lateral Sclerosis: An Acoustic-Based Approach

Hannah P. Rowe, Jordan R. Green, MGH Institute of Health Professions, USA

PAGE 4514
Thu-P-10-D-3
13:30-15:30

Diagnosing Dysarthria with Long Short-Term Memory Networks

Alex Mayle¹, Zhiwei Mou², Razvan Bunescu¹, Sadegh Mirshekarian¹, Li Xu¹, Chang Liu¹

¹Ohio University, USA; ²Jinan University, China

PAGE 4519
Thu-P-10-D-4
13:30-15:30

Modification of Devoicing Error in Cleft Lip and Palate Speech

Protima Nomo Sudro, S.R. Mahadeva Prasanna, IIT Guwahati, India

PAGE 4524
Thu-P-10-D-5
13:30-15:30

Reduced Task Adaptation in Alternating Motion Rate Tasks as an Early Marker of Bulbar Involvement in Amyotrophic Lateral Sclerosis

Marziye Eshghi¹, Panying Rong², Antje S. Mefferd³, Kaila L. Stipancic¹, Yana Yunusova⁴, Jordan R. Green¹

¹MGH Institute of Health Professions, USA; ²University of Kansas, USA; ³Vanderbilt University, USA; ⁴University of Toronto, Canada

INTERSPEECH 2019 Table of Contents

Thu-P-10-D continued...

PAGE 4529
Thu-P-10-D-6
13:30-15:30

Towards the Speech Features of Early-Stage Dementia: Design and Application of the Mandarin Elderly Cognitive Speech Database

Tianqi Wang¹, Quanlei Yan¹, Jingshen Pan¹, Feiqi Zhu², Rongfeng Su¹, Yi Guo³, Lan Wang¹, Nan Yan¹

¹Chinese Academy of Sciences, China; ²Shenzhen Luohu People's Hospital, China; ³Shenzhen People's Hospital, China

PAGE 4534
Thu-P-10-D-7
13:30-15:30

Acoustic Characteristics of Lexical Tone Disruption in Mandarin Speakers After Brain Damage

Wenjun Chen¹, Jeroen van de Weijer², Shuangshuang Zhu³, Qian Qian³, Manna Wang³

¹SISU, China; ²Shenzhen University, China; ³Shanghai Sunshine Rehabilitation Center, China

PAGE 4539
Thu-P-10-D-8
13:30-15:30

Intragastric Variation in Natural Sentence Production: Essential Tremor Patients Treated with DBS

Anne Hermes¹, Doris Mücke², Tabea Thies², Michael T. Barbe³

¹LPP (UMR 7018), France; ²Universität zu Köln, Germany; ³Uniklinik Köln, Germany

PAGE 4544
Thu-P-10-D-9
13:30-15:30

Nasal Air Emission in Sibilant Fricatives of Cleft Lip and Palate Speech

Sishir Kalita, Protima Nomo Sudro, S.R. Mahadeva Prasanna, S. Dandapat, IIT Guwahati, India

PAGE 4549
Thu-P-10-D-10
13:30-15:30

Parallel vs. Non-Parallel Voice Conversion for Esophageal Speech

Luis Serrano, Sneha Raman, David Tavarez, Eva Navas, Inma Hernaez, Universidad del País Vasco, Spain

PAGE 4554
Thu-P-10-D-11
13:30-15:30

Hypernasality Severity Detection Using Constant Q Cepstral Coefficients

Akhilesh Kumar Dubey, S.R. Mahadeva Prasanna, S. Dandapat, IIT Guwahati, India

INTERSPEECH 2019 Table of Contents

Thu-P-10-D continued...

PAGE 4559
Thu-P-10-D-12
13:30-15:30

Automatic Depression Level Detection via ℓ_p -Norm Pooling

Mingyue Niu, Jianhua Tao, Bin Liu, Cunhang Fan, Chinese Academy of Sciences, China

PAGE 4564
Thu-P-10-D-13
13:30-15:30

Comparison of Speech Tasks and Recording Devices for Voice Based Automatic Classification of Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis

Suhas B.N.¹, Deep Patel¹, Nithin Rao², Yamini Belur³, Pradeep Reddy³, Nalini Atchayaram³, Ravi Yadav³, Dipanjan Gope¹, Prasanta Kumar Ghosh¹

¹Indian Institute of Science, India; ²University of Southern California, USA; ³NIMHANS, India

Thu-P-10-E: Speech and Audio Source Separation and Scene Analysis 3

Chair: To be confirmed

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 10/E

- PAGE 4569
Thu-P-10-E-1
13:30-15:30
- A Modified Algorithm for Multiple Input Spectrogram Inversion**
Dongxiao Wang¹, Hirokazu Kameoka², Koichi Shinoda¹
¹Tokyo Tech, Japan; ²NTT, Japan
- PAGE 4574
Thu-P-10-E-2
13:30-15:30
- A Comprehensive Study of Speech Separation: Spectrogram vs Waveform Separation**
Fahimeh Bahmaninezhad¹, Jian Wu², Rongzhi Gu³, Shi-Xiong Zhang⁴, Yong Xu⁴, Meng Yu⁴, Dong Yu⁴
¹University of Texas at Dallas, USA; ²Northwestern Polytechnical University, China; ³Peking University, China; ⁴Tencent, USA
- PAGE 4579
Thu-P-10-E-3
13:30-15:30
- Evaluating Audiovisual Source Separation in the Context of Video Conferencing**
Berkay İnan¹, Milos Cernak², Helmut Grabner², Helena Peic Tukuljac¹, Rodrigo C.G. Pena¹, Benjamin Ricaud¹
¹EPFL, Switzerland; ²Logitech, Switzerland
- PAGE 4584
Thu-P-10-E-4
13:30-15:30
- Influence of Speaker-Specific Parameters on Speech Separation Systems**
David Ditter, Timo Gerkmann, Universität Hamburg, Germany
- PAGE 4589
Thu-P-10-E-5
13:30-15:30
- CNN-LSTM Models for Multi-Speaker Source Separation Using Bayesian Hyper Parameter Optimization**
Jeroen Zegers, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium
- PAGE 4594
Thu-P-10-E-6
13:30-15:30
- Towards Joint Sound Scene and Polyphonic Sound Event Recognition**
Helen L. Bear, Inês Nolasco, Emmanouil Benetos, Queen Mary University of London, UK

Thu-P-10-E continued...

- PAGE 4599
Thu-P-10-E-7
13:30-15:30
- Discriminative Learning for Monaural Speech Separation Using Deep Embedding Features**
Cunhang Fan, Bin Liu, Jianhua Tao, Jiangyan Yi, Zhengqi Wen, Chinese Academy of Sciences, China
- PAGE 4604
Thu-P-10-E-8
13:30-15:30
- Probabilistic Permutation Invariant Training for Speech Separation**
Midia Yousefi, Soheil Khorram, John H.L. Hansen, University of Texas at Dallas, USA
- PAGE 4609
Thu-P-10-E-9
13:30-15:30
- Which Ones Are Speaking? Speaker-Inferred Model for Multi-Talker Speech Separation**
Jing Shi, Jiaming Xu, Bo Xu, Chinese Academy of Sciences, China
- PAGE 4614
Thu-P-10-E-10
13:30-15:30
- End-to-End Monaural Speech Separation with Multi-Scale Dynamic Weighted Gated Dilated Convolutional Pyramid Network**
Ziqiang Shi¹, Huibin Lin¹, Liu Liu¹, Rujie Liu¹, Shoji Hayakawa², Shouji Harada², Jiqing Han³
¹Fujitsu, China; ²Fujitsu, Japan; ³Harbin Institute of Technology, China
- PAGE 4619
Thu-P-10-E-11
13:30-15:30
- End-to-End Music Source Separation: Is it Possible in the Waveform Domain?**
Francesc Lluís, Jordi Pons, Xavier Serra, Universitat Pompeu Fabra, Spain

Thu-S&T-6: Speech-to-Text and Speech Assessment

Time 13:30-15:30, Thursday 19 Sept 2019, Hall 4

PAGE 4624
Thu-S&T-6-1
13:30-15:30

Elpis, an Accessible Speech-to-Text Tool

*Ben Foley, Alina Rakhi, Nicholas Lambourne, Nicholas Buckeridge, Janet Wiles,
University of Queensland, Australia*

PAGE 4626
Thu-S&T-6-2
13:30-15:30

Framework for Conducting Tasks Requiring Human Assessment

*Martin Grüber, Adam Chýlek, Jindřich Matoušek, University of West Bohemia, Czech
Republic*

PAGE 4628
Thu-S&T-6-3
13:30-15:30

Multimedia Simultaneous Translation System for Minority Language Communication with Mandarin

*Shen Huang¹, Bojie Hu¹, Shan Huang¹, Pengfei Hu¹, Jian Kang¹, Zhiqiang Lv¹,
Jinghao Yan¹, Qi Ju¹, Shiyin Kang¹, Deyi Tuo¹, Guangzhi Li¹, Nurmemet Yolwas²
¹Tencent, China; ²Xin Jiang University, China*

PAGE 4630
Thu-S&T-6-4
13:30-15:30

The SAIL LABS Media Mining Indexer and the CAVA Framework

Erinc Dikici, Gerhard Backfried, Jürgen Riedler, SAIL LABS Technology, Austria

PAGE 4632
Thu-S&T-6-5
13:30-15:30

CaptionAI: A Real-Time Multilingual Captioning Application

*Nagendra Kumar Goel, Mousmita Sarma, Saikiran Valluri, Dharmeshkumar Agrawal,
Steve Braich, Tejendra Singh Kuswah, Zikra Iqbal, Surbhi Chauhan, Raj Karbar,
GoVivace, USA*