

# **17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)**

Understanding Speech Processing  
in Humans and Machines

San Francisco, California, USA  
8-12 September 2016

Volume 2 of 5

ISBN: 978-1-5108-3313-5

**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© (2016) by International Speech Communications Association  
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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

International Speech Communications 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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Keynote 1: ISCA Medalist: John Makhoul

Chair: Haizhou Li

Room: Grand Ballroom ABC, Time 09:30-10:30, Friday, 9 Sept. 2016

---

PAGE 1  
Fri-Keynote-1  
09:30

**A 50-Year Retrospective on Speech and Language Processing**  
*John Makhoul, BBN Technologies, USA*

## Fri-O-1-1 : Neural Networks in Speech Recognition

Chairs: Penny Karanasou, Bhuvana Ramabhadran

Room: Grand Ballroom A, Time 11:00-13:00, Friday, 9 Sept. 2016

---

PAGE 2  
Fri-O-1-1-1  
11:00

**Improving English Conversational Telephone Speech Recognition**  
*Ivan Medennikov<sup>1</sup>, Alexey Prudnikov<sup>2</sup>, Alexander Zlatovitskiy<sup>1</sup>*  
*<sup>1</sup>STC-innovations, Russia; <sup>2</sup>ITMO University, Russia*

PAGE 7  
Fri-O-1-1-2  
11:20

**The IBM 2016 English Conversational Telephone Speech Recognition System**  
*George Saon, Tom Sercu, Steven Rennie, Hong-Kwang J. Kuo, IBM, USA*

PAGE 12  
Fri-O-1-1-3  
11:40

**Small-Footprint Deep Neural Networks with Highway Connections for Speech Recognition**  
*Liang Lu, Steve Renals, University of Edinburgh, UK*

PAGE 17  
Fri-O-1-1-4  
12:00

**Deep Convolutional Neural Networks with Layer-Wise Context Expansion and Attention**  
*Dong Yu<sup>1</sup>, Wayne Xiong<sup>1</sup>, Jasha Droppo<sup>1</sup>, Andreas Stolcke<sup>1</sup>, Guoli Ye<sup>2</sup>, Jinyu Li<sup>1</sup>, Geoffrey Zweig<sup>1</sup>*  
*<sup>1</sup>Microsoft, USA; <sup>2</sup>Microsoft, China*

PAGE 22  
Fri-O-1-1-5  
12:20

**Lower Frame Rate Neural Network Acoustic Models**  
*Golan Pundak, Tara N. Sainath, Google, USA*

PAGE 27  
Fri-O-1-1-6  
12:40

**Improved Neural Network Initialization by Grouping Context-Dependent Targets for Acoustic Modeling**  
*Gakuto Kurata<sup>1</sup>, Brian Kingsbury<sup>2</sup>*  
*<sup>1</sup>IBM, Japan; <sup>2</sup>IBM, USA*

## Fri-O-1-2 : Special Session: Auditory-Visual Expressive Speech and Gesture in Humans and Machines

Chairs: *Jeesun Kim, Gérard Bailly*

Room: *Grand Ballroom BC, Time 11:00-13:00, Friday, 9 Sept. 2016*

---

PAGE 32 Fri-O-1-2-1 11:00	<b>Automatic Scoring of Monologue Video Interviews Using Multimodal Cues</b> <i>Lei Chen, Gary Feng, Michelle Martin-Raugh, Chee Wee Leong, Christopher Kitchen, Su-Youn Yoon, Blair Lehman, Harrison Kell, Chong Min Lee, Educational Testing Service, USA</i>
PAGE 37 Fri-O-1-2-2 11:15	<b>The Sound of Disgust: How Facial Expression May Influence Speech Production</b> <i>Chee Seng Chong, Jeesun Kim, Chris Davis, Western Sydney University, Australia</i>
PAGE 42 Fri-O-1-2-3 11:30	<b>Analyzing Temporal Dynamics of Dyadic Synchrony in Affective Interactions</b> <i>Zhaojun Yang, Shrikanth S. Narayanan, University of Southern California, USA</i>
PAGE 47 Fri-O-1-2-4 11:45	<b>Audiovisual Speech Scene Analysis in the Context of Competing Sources</b> <i>Attigodu C. Ganesh, Frédéric Berthommier, Jean-Luc Schwartz, GIPSA, France</i>
PAGE 52 Fri-O-1-2-5 12:00	<b>Head Motion Generation with Synthetic Speech: A Data Driven Approach</b> <i>Najmeh Sadoughi, Carlos Busso, University of Texas at Dallas, USA</i>
PAGE 57 Fri-O-1-2-6 12:15	<b>The Consistency and Stability of Acoustic and Visual Cues for Different Prosodic Attitudes</b> <i>Jeesun Kim, Chris Davis, Western Sydney University, Australia</i>
N/A Fri-O-1-2-7 12:30	<b>Introduction to Poster Presentation of Part II</b> <i>Jeesun Kim<sup>1</sup>, Gérard Bailly<sup>2</sup></i> <i><sup>1</sup>Western Sydney University, Australia; <sup>2</sup>GIPSA, France</i>

## Fri-O-1-3 : Prosody

Chairs: *Chiu-Yu Tseng, Štefan Beňuš*

Room: *Bayview A, Time 11:00-13:00, Friday, 9 Sept. 2016*

---

PAGE 62 Fri-O-1-3-1 11:00	<b>The Unit of Speech Encoding: The Case of Romanian</b> <i>Irene Vogel<sup>1</sup>, Laura Spinu<sup>2</sup></i> <i><sup>1</sup>University of Delaware, USA; <sup>2</sup>University of Western Ontario, Canada</i>
PAGE 67 Fri-O-1-3-2 11:20	<b>The Perceptual Effect of L1 Prosody Transplantation on L2 Speech: The Case of French Accented German</b> <i>Jeanin Jügler, Frank Zimmerer, Jürgen Trouvain, Bernd Möbius, Universität des Saarlandes, Germany</i>
PAGE 72 Fri-O-1-3-3 11:40	<b>Organizing Syllables into Sandhi Domains — Evidence from F0 and Duration Patterns in Shanghai Chinese</b> <i>Bijun Ling, Jie Liang, Tongji University, China</i>
PAGE 77 Fri-O-1-3-4 12:00	<b>Automatic Analysis of Phonetic Speech Style Dimensions</b> <i>Neville Ryant, Mark Liberman, Linguistic Data Consortium, USA</i>
PAGE 82 Fri-O-1-3-5 12:20	<b>The Acoustic Manifestation of Prominence in Stressless Languages</b> <i>Angeliki Athanasopoulou, Irene Vogel, University of Delaware, USA</i>
PAGE 87 Fri-O-1-3-6 12:40	<b>The Rhythmic Constraint on Prosodic Boundaries in Mandarin Chinese Based on Corpora of Silent Reading and Speech Perception</b> <i>Wei Lai<sup>1</sup>, Jiahong Yuan<sup>1</sup>, Ya Li<sup>2</sup>, Xiaoying Xu<sup>3</sup>, Mark Liberman<sup>1</sup></i> <i><sup>1</sup>University of Pennsylvania, USA; <sup>2</sup>Chinese Academy of Sciences, China; <sup>3</sup>Beijing Normal University, China</i>

## Fri-O-1-4: Speech and Language Processing for Clinical Health Applications

Chairs: Julien Epps, Elmar Nöth

Room: Bayview B, Time 11:00-13:00, Friday, 9 Sept. 2016

---

- PAGE 92  
Fri-O-1-4-1  
11:00
- Toward Development and Evaluation of Pain Level-Rating Scale for Emergency Triage based on Vocal Characteristics and Facial Expressions**  
*Fu-Sheng Tsai<sup>1</sup>, Ya-Ling Hsu<sup>1</sup>, Wei-Chen Chen<sup>1</sup>, Yi-Ming Weng<sup>2</sup>, Chip-Jin Ng<sup>2</sup>, Chi-Chun Lee<sup>1</sup>*  
<sup>1</sup>National Tsing Hua University, Taiwan; <sup>2</sup>Chang Gung Memorial Hospital, Taiwan
- PAGE 97  
Fri-O-1-4-2  
11:20
- Predicting Severity of Voice Disorder from DNN-HMM Acoustic Posteriors**  
*Tan Lee, Yuanyuan Liu, Yu Ting Yeung, Thomas K.T. Law, Kathy Y.S. Lee, Chinese University of Hong Kong, China*
- PAGE 102  
Fri-O-1-4-3  
11:40
- Long-Term Stability of Tracheoesophageal Voices**  
*Klaske E. van Sluis, Michiel W.M. van den Brekel, Frans J.M. Hilgers, Rob J.J.H. van Son, Netherlands Cancer Institute, The Netherlands*
- PAGE 107  
Fri-O-1-4-4  
12:00
- Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection**  
*Gábor Gosztolya<sup>1</sup>, László Tóth<sup>1</sup>, Tamás Grósz<sup>2</sup>, Veronika Vincze<sup>1</sup>, Ildikó Hoffmann<sup>2</sup>, Gréta Szatlóczi<sup>2</sup>, Magdolna Pákáski<sup>2</sup>, János Kálmán<sup>2</sup>*  
<sup>1</sup>MTA-SZTE RGAI, Hungary; <sup>2</sup>University of Szeged, Hungary
- PAGE 112  
Fri-O-1-4-5  
12:20
- Towards an Automated Screening Tool for Developmental Speech and Language Impairments**  
*Jen J. Gong<sup>1</sup>, Maryann Gong<sup>1</sup>, Dina Levy-Lambert<sup>1</sup>, Jordan R. Green<sup>2</sup>, Tiffany P. Hogan<sup>2</sup>, John V. Guttag<sup>1</sup>*  
<sup>1</sup>MIT, USA; <sup>2</sup>MGH Institute of Health Professions, USA
- PAGE 117  
Fri-O-1-4-6  
12:40
- Spectral Enhancement of Cleft Lip and Palate Speech**  
*Vikram C.M., Nagaraj Adiga, S.R. Mahadeva Prasanna, IIT Guwahati, India*

## Fri-O-1-5: Speech Coding and Audio Processing for Noise Reduction

Chairs: Tom Bäckström, Dayana Ribas

Room: Seacliff BCD, Time 11:00-13:00, Friday, 9 Sept. 2016

---

- PAGE 122  
Fri-O-1-5-1  
11:00
- Assessing Level-Dependent Segmental Contribution to the Intelligibility of Speech Processed by Single-Channel Noise-Suppression Algorithms**  
*Tian Guan<sup>1</sup>, Guangxing Chu<sup>1</sup>, Fei Chen<sup>2</sup>, Feng Yang<sup>3</sup>*  
<sup>1</sup>Tsinghua University, China; <sup>2</sup>SUSTC, China; <sup>3</sup>Shenzhen Children's Hospital, China
- PAGE 126  
Fri-O-1-5-2  
11:20
- Effectiveness of Near-End Speech Enhancement Under Equal-Loudness and Equal-Level Constraints**  
*Tudor-Cătălin Zorilă<sup>1</sup>, Sheila Flanagan<sup>2</sup>, Brian C.J. Moore<sup>2</sup>, Yannis Stylianou<sup>1</sup>*  
<sup>1</sup>Toshiba Research Europe, UK; <sup>2</sup>University of Cambridge, UK
- PAGE 131  
Fri-O-1-5-3  
11:40
- Speech Synthesis in Noisy Environment by Enhancing Strength of Excitation and Formant Prominence**  
*Bidisha Sharma, S.R. Mahadeva Prasanna, IIT Guwahati, India*
- PAGE 136  
Fri-O-1-5-4  
12:00
- Relative Contributions of Amplitude and Phase to the Intelligibility Advantage of Ideal Binary Masked Sentences**  
*Lei Wang, Shufeng Zhu, Diliang Chen, Yong Feng, Fei Chen, SUSTC, China*
- PAGE 140  
Fri-O-1-5-5  
12:20
- Predicting Binaural Speech Intelligibility from Signals Estimated by a Blind Source Separation Algorithm**  
*Qingju Liu<sup>1</sup>, Yan Tang<sup>2</sup>, Philip J.B. Jackson<sup>1</sup>, Wenwu Wang<sup>1</sup>*  
<sup>1</sup>University of Surrey, UK; <sup>2</sup>University of Salford, UK
- PAGE 145  
Fri-O-1-5-6  
12:40
- Automated Pause Insertion for Improved Intelligibility Under Reverberation**  
*Petko N. Petkov, Norbert Braunschweiler, Yannis Stylianou, Toshiba Research Europe, UK*

## Fri-O-1-6 : Speech Analysis

Chair: Hakan Erdogan

Room: Seacliff A, Time 11:00-13:00, Friday, 9 Sept. 2016

---

- PAGE 150  
Fri-O-1-6-1  
11:00      **Automatic Classification of Phonation Modes in Singing Voice: Towards Singing Style Characterisation and Application to Ethnomusicological Recordings**  
*Jean-Luc Rouas, Leonidas Ioannidis, LaBRI (UMR 5800), France*
- PAGE 155  
Fri-O-1-6-2  
11:20      **Novel Nonlinear Prediction Based Features for Spoofed Speech Detection**  
*Himanshu N. Bhavsar, Tanvina B. Patel, Hemant A. Patil, DA-IICT, India*
- PAGE 160  
Fri-O-1-6-3  
11:40      **Robust Vowel Landmark Detection Using Epoch-Based Features**  
*Sri Harsha Dumpala, Bhanu Teja Nellore, Raghu Ram Nevali, Suryakanth V. Gangashetty, B. Yegnanarayana, IIIT Hyderabad, India*
- PAGE 165  
Fri-O-1-6-4  
12:00      **Sensitivity of Quantitative RT-MRI Metrics of Vocal Tract Dynamics to Image Reconstruction Settings**  
*Johannes Töger, Yongwan Lim, Sajjan Goud Lingala, Shrikanth S. Narayanan, Krishna S. Nayak, University of Southern California, USA*
- PAGE 170  
Fri-O-1-6-5  
12:20      **Sound Pattern Matching for Automatic Prosodic Event Detection**  
*Milos Cernak, Afsaneh Asaei, Pierre-Edouard Honnet, Philip N. Garner, Hervé Broulard, Idiap Research Institute, Switzerland*
- PAGE 175  
Fri-O-1-6-6  
12:40      **Automatic Classification of Lexical Stress in English and Arabic Languages Using Deep Learning**  
*Mostafa Shahin<sup>1</sup>, Julien Epps<sup>2</sup>, Beena Ahmed<sup>1</sup>*  
*<sup>1</sup>Texas A&M University, Qatar; <sup>2</sup>University of New South Wales, Australia*

## Fri-P-1-1 : First and Second Language Acquisition

Chair: Maria Luisa Garcia Lecumberri

Room: Pacific Concourse - Poster A, Time 11:00-13:00, Friday, 9 Sept. 2016

---

- PAGE 180  
Fri-P-1-1-1  
11:00      **Development of Mandarin Onset-Rime Detection in Relation to Age and Pinyin Instruction**  
*Fei Chen, Nan Yan, Xunan Huang, Hao Zhang, Lan Wang, Gang Peng, Chinese Academy of Sciences, China*
- PAGE 185  
Fri-P-1-1-2  
11:00      **Joint Effect of Dialect and Mandarin on English Vowel Production: A Case Study in Changsha EFL Learners**  
*Xinyi Wen, Yuan Jia, Chinese Academy of Social Sciences, China*
- PAGE 190  
Fri-P-1-1-3  
11:00      **Effects of L1 Phonotactic Constraints on L2 Word Segmentation Strategies**  
*Tamami Katayama, Prefectural University of Hiroshima, Japan*
- PAGE 195  
Fri-P-1-1-4  
11:00      **Putting German [ʃ] and [ç] in Two Different Boxes: Native German vs L2 German of French Learners**  
*Jane Wottawa<sup>1</sup>, Martine Adda-Decker<sup>1</sup>, Frédéric Isel<sup>2</sup>*  
*<sup>1</sup>LPP (UMR 7018), France; <sup>2</sup>MoDyCo, France*
- PAGE 200  
Fri-P-1-1-5  
11:00      **Naturalness Judgement of L2 English Through Dubbing Practice**  
*Dean Luo<sup>1</sup>, Ruxin Luo<sup>2</sup>, Lixin Wang<sup>3</sup>*  
*<sup>1</sup>Shenzhen Institute of Information Technology, China; <sup>2</sup>Shenzhen Polytechnic, China; <sup>3</sup>Shenzhen Seaskyland Technologies, China*
- PAGE 204  
Fri-P-1-1-6  
11:00      **Audiovisual Training Effects for Japanese Children Learning English /r/-/l/**  
*Yasuaki Shinohara, Waseda University, Japan*
- PAGE 208  
Fri-P-1-1-7  
11:00      **L2 Acquisition and Production of the English Rhotic Pharyngeal Gesture**  
*Sarah Harper, Louis Goldstein, Shrikanth S. Narayanan, University of Southern California, USA*

## Fri-P-1-2 : Speech and Hearing Disorders & Perception

Chair: Frank Rudzicz

Room: Pacific Concourse - Poster B, Time 11:00-13:00, Friday, 9 Sept. 2016

---

- PAGE 213  
Fri-P-1-2-1  
11:00      **Auditory-Visual Perception of VCVs Produced by People with Down Syndrome: Preliminary Results**  
*Alexandre Hennequin, Amélie Rochet-Capellan, Marion Dohen, GIPSA, France*
- PAGE 218  
Fri-P-1-2-2  
11:00      **Combining Non-Pathological Data of Different Language Varieties to Improve DNN-HMM Performance on Pathological Speech**  
*Emre Yilmaz, Mario Ganzeboom, Catia Cucchiarini, Helmer Strik, Radboud Universiteit Nijmegen, The Netherlands*
- PAGE 223  
Fri-P-1-2-3  
11:00      **Evaluation of a Phone-Based Anomaly Detection Approach for Dysarthric Speech**  
*Imed Laaridh<sup>1</sup>, Corinne Fredouille<sup>1</sup>, Christine Meunier<sup>2</sup>*  
*<sup>1</sup>LLA, France; <sup>2</sup>LPL, France*
- PAGE 228  
Fri-P-1-2-4  
11:00      **Recognition of Dysarthric Speech Using Voice Parameters for Speaker Adaptation and Multi-Taper Spectral Estimation**  
*Chitralekha Bhat, Bhavik Vachhani, Sunil Kopperapu, Tata Consultancy Services, India*
- PAGE 233  
Fri-P-1-2-5  
11:00      **Impaired Categorical Perception of Mandarin Tones and its Relationship to Language Ability in Autism Spectrum Disorders**  
*Fei Chen<sup>1</sup>, Nan Yan<sup>1</sup>, Xiaojie Pan<sup>2</sup>, Feng Yang<sup>3</sup>, Zhuanzhuan Ji<sup>1</sup>, Lan Wang<sup>1</sup>, Gang Peng<sup>1</sup>*  
*<sup>1</sup>Chinese Academy of Sciences, China; <sup>2</sup>Shenzhen Love Wisdom Special Children Rehabilitation Center, China; <sup>3</sup>Shenzhen Children's Hospital, China*

- PAGE 238  
Fri-P-1-2-6  
11:00      **Perceived Naturalness of Electrolaryngeal Speech Produced Using sEMG-Controlled vs. Manual Pitch Modulation**  
*K.F. Nagle<sup>1</sup>, J.T. Heaton<sup>2</sup>*  
*<sup>1</sup>Seton Hall University, USA; <sup>2</sup>Massachusetts General Hospital, USA*
- PAGE 243  
Fri-P-1-2-7  
11:00      **Identifying Hearing Loss from Learned Speech Kernels**  
*Shamima Najnin, Bonny Banerjee, Lisa Lucks Mendel, Masoumeh Heidari Kapourchali, Jayanta Kumar Dutta, Sungmin Lee, Chhayakanta Patro, Monique Pousson, University of Memphis, USA*
- PAGE 248  
Fri-P-1-2-8  
11:00      **Differential Effects of Velopharyngeal Dysfunction on Speech Intelligibility During Early and Late Stages of Amyotrophic Lateral Sclerosis**  
*Panying Rong<sup>1</sup>, Yana Yunusova<sup>2</sup>, Jordan R. Green<sup>1</sup>*  
*<sup>1</sup>MGH Institute of Health Professions, USA; <sup>2</sup>University of Toronto, Canada*
- PAGE 253  
Fri-P-1-2-9  
11:00      **The Production of Intervocalic Glides in Non Dysarthric Parkinsonian Speech**  
*V. Delvaux, V. Roland, K. Huet, M. Piccaluga, M.C. Haelewyck, B. Harmegnies, Université de Mons, Belgium*
- PAGE 257  
Fri-P-1-2-10  
11:00      **Auditory Processing Impairments Under Background Noise in Children with Non-Syndromic Cleft Lip and/or Palate**  
*Yang Feng, Zhang Lu, Shenzhen Children's Hospital, China*
- PAGE 262  
Fri-P-1-2-11  
11:00      **Modulation Spectral Features for Predicting Vocal Emotion Recognition by Simulated Cochlear Implants**  
*Zhi Zhu<sup>1</sup>, Ryota Miyauchi<sup>1</sup>, Yukiko Araki<sup>2</sup>, Masashi Unoki<sup>1</sup>*  
*<sup>1</sup>JAIST, Japan; <sup>2</sup>Kanazawa University, Japan*

INTERSPEECH 2016 Table of Contents

---

PAGE 267 Fri-P-1-2-12 11:00	<b>Automatic Discrimination of Soft Voice Onset Using Acoustic Features of Breathy Voicing</b> <i>Keiko Ochi<sup>1</sup>, Koichi Mori<sup>2</sup>, Naomi Sakai<sup>2</sup>, Nobutaka Ono<sup>1</sup></i> <sup>1</sup> NII, Japan; <sup>2</sup> National Rehabilitation Center for Persons with Disabilities, Japan
PAGE 272 Fri-P-1-2-13 11:00	<b>Effect of Noise on Lexical Tone Perception in Cantonese-Speaking Amusics</b> <i>Jing Shao, Caicai Zhang, Gang Peng, Yike Yang, William S.-Y. Wang, Hong Kong Polytechnic University, China</i>
PAGE 277 Fri-P-1-2-14 11:00	<b>Audio-Visual Speech Recognition Using Bimodal-Trained Bottleneck Features for a Person with Severe Hearing Loss</b> <i>Yuki Takashima<sup>1</sup>, Ryo Aihara<sup>1</sup>, Tetsuya Takiguchi<sup>1</sup>, Yasuo Arika<sup>1</sup>, Nobuyuki Mitani<sup>2</sup>, Kiyohiro Omori<sup>2</sup>, Kaoru Nakazono<sup>2</sup></i> <sup>1</sup> Kobe University, Japan; <sup>2</sup> Hyogo Institute of Assistive Technology, Japan
PAGE 282 Fri-P-1-2-15 11:00	<b>Perception of Tone in Whispered Mandarin Sentences: The Case for Singapore Mandarin</b> <i>Yuling Gu, Boon Pang Lim, Nancy F. Chen, A*STAR, Singapore</i>

INTERSPEECH 2016 Table of Contents

---

**Fri-P-1-3: Speech Synthesis Poster**

*Chair: Alan Black*

*Room: Pacific Concourse - Poster C, Time 11:00-13:00, Friday, 9 Sept. 2016*

---

PAGE 287 Fri-P-1-3-1 11:00	<b>A KL Divergence and DNN-Based Approach to Voice Conversion without Parallel Training Sentences</b> <i>Feng-Long Xie<sup>1</sup>, Frank K. Soong<sup>2</sup>, Haifeng Li<sup>1</sup></i> <sup>1</sup> Harbin Institute of Technology, China; <sup>2</sup> Microsoft, China
PAGE 292 Fri-P-1-3-2 11:00	<b>Parallel Dictionary Learning for Voice Conversion Using Discriminative Graph-Embedded Non-Negative Matrix Factorization</b> <i>Ryo Aihara, Tetsuya Takiguchi, Yasuo Arika, Kobe University, Japan</i>
PAGE 297 Fri-P-1-3-3 11:00	<b>Speech Bandwidth Extension Using Bottleneck Features and Deep Recurrent Neural Networks</b> <i>Yu Gu, Zhen-Hua Ling, Li-Rong Dai, USTC, China</i>
PAGE 302 Fri-P-1-3-4 11:00	<b>Voice Conversion Based on Matrix Variate Gaussian Mixture Model Using Multiple Frame Features</b> <i>Yi Yang, Hidetsugu Uchida, Daisuke Saito, Nobuaki Minematsu, University of Tokyo, Japan</i>
PAGE 307 Fri-P-1-3-5 11:00	<b>Voice Conversion Based on Trajectory Model Training of Neural Networks Considering Global Variance</b> <i>Naoki Hosaka, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, Keiichi Tokuda, Nagoya Institute of Technology, Japan</i>
PAGE 312 Fri-P-1-3-6 11:00	<b>Comparing Articulatory and Acoustic Strategies for Reducing Non-Native Accents</b> <i>Sandesh Aryal, Ricardo Gutierrez-Osuna, Texas A&amp;M University, USA</i>



INTERSPEECH 2016 Table of Contents

---

PAGE 317 Fri-P-1-3-7 11:00	<b>Cross-Lingual Speaker Adaptation for Statistical Speech Synthesis Using Limited Data</b> <i>Seyyed Saeed Sarfjoo, Cenk Demiroglu, Özyeğin Üniversitesi, Turkey</i>
PAGE 322 Fri-P-1-3-8 11:00	<b>Personalized, Cross-Lingual TTS Using Phonetic Posteriorgrams</b> <i>Lifa Sun, Hao Wang, Shiyin Kang, Kun Li, Helen Meng, Chinese University of Hong Kong, China</i>
PAGE 327 Fri-P-1-3-9 11:00	<b>Acoustic Analysis of Syllables Across Indian Languages</b> <i>Anusha Prakash, Jeena J. Prakash, Hema A. Murthy, IIT Madras, India</i>
PAGE 332 Fri-P-1-3-10 11:00	<b>Objective Evaluation Methods for Chinese Text-To-Speech Systems</b> <i>Teng Zhang<sup>1</sup>, Zhipeng Chen<sup>1</sup>, Ji Wu<sup>1</sup>, Sam Lai<sup>2</sup>, Wenhui Lei<sup>2</sup>, Carsten Isert<sup>2</sup></i> <i><sup>1</sup>Tsinghua University, China; <sup>2</sup>BMW Group Technology Office China, China</i>
PAGE 337 Fri-P-1-3-11 11:00	<b>Objective Evaluation Using Association Between Dimensions Within Spectral Features for Statistical Parametric Speech Synthesis</b> <i>Yusuke Ijima<sup>1</sup>, Taichi Asami<sup>1</sup>, Hideyuki Mizuno<sup>2</sup></i> <i><sup>1</sup>NTT, Japan; <sup>2</sup>Tokyo University of Science, Japan</i>
PAGE 342 Fri-P-1-3-12 11:00	<b>A Hierarchical Predictor of Synthetic Speech Naturalness Using Neural Networks</b> <i>Takenori Yoshimura<sup>1</sup>, Gustav Eje Henter<sup>2</sup>, Oliver Watts<sup>2</sup>, Mirjam Wester<sup>2</sup>, Junichi Yamagishi<sup>2</sup>, Keiichi Tokuda<sup>1</sup></i> <i><sup>1</sup>Nagoya Institute of Technology, Japan; <sup>2</sup>University of Edinburgh, UK</i>
PAGE 347 Fri-P-1-3-13 11:00	<b>Text-to-Speech for Individuals with Vision Loss: A User Study</b> <i>Monika Podsiadło<sup>1</sup>, Shweta Chahar<sup>2</sup></i> <i><sup>1</sup>Google, USA; <sup>2</sup>Google, UK</i>

INTERSPEECH 2016 Table of Contents

---

PAGE 352 Fri-P-1-3-14 11:00	<b>Speech Enhancement for a Noise-Robust Text-to-Speech Synthesis System Using Deep Recurrent Neural Networks</b> <i>Cassia Valentini-Botinhao<sup>1</sup>, Xin Wang<sup>2</sup>, Shinji Takaki<sup>2</sup>, Junichi Yamagishi<sup>1</sup></i> <i><sup>1</sup>University of Edinburgh, UK; <sup>2</sup>NII, Japan</i>
PAGE 357 Fri-P-1-3-15 11:00	<b>Data Selection and Adaptation for Naturalness in HMM-Based Speech Synthesis</b> <i>Erica Cooper, Alison Chang, Yocheved Levitan, Julia Hirschberg, Columbia University, USA</i>

## Fri-P-1-4: Topics in Speech Processing

Chair: Javier Hernando

Room: Pacific Concourse - Poster D, Time 11:00-13:00, Friday, 9 Sept. 2016

---

PAGE 362  
Fri-P-1-4-1  
11:00

### **A Portable Automatic PA-TA-KA Syllable Detection System to Derive Biomarkers for Neurological Disorders**

*Fei Tao<sup>1</sup>, Louis Daudet<sup>2</sup>, Christian Poellabauer<sup>2</sup>, Sandra L. Schneider<sup>3</sup>, Carlos Busso<sup>1</sup>*  
<sup>1</sup>University of Texas at Dallas, USA; <sup>2</sup>University of Notre Dame, USA; <sup>3</sup>Saint Mary's College, USA

PAGE 367  
Fri-P-1-4-2  
11:00

### **Deep Neural Networks for i-Vector Language Identification of Short Utterances in Cars**

*Omid Ghahabi, Antonio Bonafonte, Javier Hernando, Asunción Moreno, Universitat Politècnica de Catalunya, Spain*

PAGE 372  
Fri-P-1-4-3  
11:00

### **Improving i-Vector and PLDA Based Speaker Clustering with Long-Term Features**

*Abraham Woubie<sup>1</sup>, Jordi Luque<sup>2</sup>, Javier Hernando<sup>1</sup>*

<sup>1</sup>Universitat Politècnica de Catalunya, Spain; <sup>2</sup>Telefónica I+D, Spain

## Fri-S&T-1: Show & Tell Session 1

Chairs: Shiva Sundaram, Nicolas Scheffer

Room: Market Street Foyer, Time 11:00-13:00, Friday, 9 Sept. 2016

---

PAGE 377  
Fri-S&T-1-1  
11:00

### **Open Language Interface for Voice Exploitation (OLIVE)**

*Aaron Lawson, Mitchell McLaren, Harry Bratt, Martin Graciarena, Horacio Franco, Christopher George, Allen Stauffer, Chris Bartels, Julien VanHout, SRI International, USA*

PAGE 379  
Fri-S&T-1-2  
11:00

### **A Multimodal Dialogue System for Air Traffic Control Trainees Based on Discrete-Event Simulation**

*Luboš Šmíd, Adam Chýlek, Jan Švec, University of West Bohemia, Czech Republic*

PAGE 381  
Fri-S&T-1-3  
11:00

### **LIG-AIKUMA: A Mobile App to Collect Parallel Speech for Under-Resourced Language Studies**

*Elodie Gauthier<sup>1</sup>, David Blachon<sup>1</sup>, Laurent Besacier<sup>1</sup>, Guy-Noël Kouarata<sup>2</sup>, Martine Adda-Decker<sup>2</sup>, Annie Rialland<sup>2</sup>, Gilles Adda<sup>3</sup>, Grégoire Bachman<sup>2</sup>*  
<sup>1</sup>LIG (UMR 5217), France; <sup>2</sup>LPP (UMR 7018), France; <sup>3</sup>LIMSI, France

PAGE 383  
Fri-S&T-1-4  
11:00

### **ARET — Automatic Reading of Educational Texts for Visually Impaired Students**

*Martin Grüber, Jindřich Matoušek, Zdeněk Hanzlíček, Zdeněk Krňoul, Zbyněk Zajíc, University of West Bohemia, Czech Republic*

## Fri-O-2-1 : New Trends in Neural Networks for Speech Recognition

Chairs: Arnab Ghoshal, Rohit Prabhavalkar

Room: Grand Ballroom A, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 385  
Fri-O-2-1-1  
14:30      **Segmental Recurrent Neural Networks for End-to-End Speech Recognition**  
*Liang Lu<sup>1</sup>, Lingpeng Kong<sup>2</sup>, Chris Dyer<sup>2</sup>, Noah A. Smith<sup>3</sup>, Steve Renals<sup>1</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>Carnegie Mellon University, USA; <sup>3</sup>University of Washington, USA*
- PAGE 390  
Fri-O-2-1-2  
14:50      **Acoustic Modeling Using Bidirectional Gated Recurrent Convolutional Units**  
*Markus Nussbaum-Thom, Jia Cui, Bhuvana Ramabhadran, Vaibhava Goel, IBM, USA*
- PAGE 395  
Fri-O-2-1-3  
15:10      **Exploiting Depth and Highway Connections in Convolutional Recurrent Deep Neural Networks for Speech Recognition**  
*Wei-Ning Hsu, Yu Zhang, Ann Lee, James Glass, MIT, USA*
- PAGE 400  
Fri-O-2-1-4  
15:30      **Stimulated Deep Neural Network for Speech Recognition**  
*Chunyang Wu<sup>1</sup>, Penny Karanasou<sup>1</sup>, Mark J.F. Gales<sup>1</sup>, Khe Chai Sim<sup>2</sup>*  
*<sup>1</sup>University of Cambridge, UK; <sup>2</sup>NUS, Singapore*
- PAGE 405  
Fri-O-2-1-5  
15:50      **Phonetic Context Embeddings for DNN-HMM Phone Recognition**  
*Leonardo Badino, Istituto Italiano di Tecnologia, Italy*
- PAGE 410  
Fri-O-2-1-6  
16:10      **Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks**  
*Ying Zhang, Mohammad Pezeshki, Philémon Brakel, Saizheng Zhang, César Laurent, Yoshua Bengio, Aaron Courville, Université de Montréal, Canada*

## Fri-O-2-2 : Special Session: The RedDots Challenge: Towards Characterizing Speakers from Short Utterances

Chairs: Kong Aik Lee, Anthony Larcher, Hagai Aronowitz, Guangsen Wang, Patrick Kenny

Room: Grand Ballroom BC, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 415  
Fri-O-2-2-1  
14:30      **Joint Speaker and Lexical Modeling for Short-Term Characterization of Speaker**  
*Guangsen Wang, Kong Aik Lee, Trung Hieu Nguyen, Hanwu Sun, Bin Ma, A\*STAR, Singapore*
- PAGE 420  
Fri-O-2-2-2  
14:45      **Tandem Features for Text-Dependent Speaker Verification on the RedDots Corpus**  
*Md Jahangir Alam, Patrick Kenny, Vishwa Gupta, CRIM, Canada*
- PAGE 425  
Fri-O-2-2-3  
15:00      **Text Dependent Speaker Verification Using Un-Supervised HMM-UBM and Temporal GMM-UBM**  
*Achintya Kr. Sarkar, Zheng-Hua Tan, Aalborg University, Denmark*
- PAGE 430  
Fri-O-2-2-4  
15:15      **Utterance Verification for Text-Dependent Speaker Recognition: A Comparative Assessment Using the RedDots Corpus**  
*Tomi Kinnunen<sup>1</sup>, Md. Sahidullah<sup>1</sup>, Ivan Kukanov<sup>1</sup>, Héctor Delgado<sup>2</sup>, Massimiliano Todisco<sup>2</sup>, Achintya Kr. Sarkar<sup>3</sup>, Nicolai Bæk Thomsen<sup>3</sup>, Ville Hautamäki<sup>1</sup>, Nicholas Evans<sup>2</sup>, Zheng-Hua Tan<sup>3</sup>*  
*<sup>1</sup>University of Eastern Finland, Finland; <sup>2</sup>EURECOM, France; <sup>3</sup>Aalborg University, Denmark*
- PAGE 435  
Fri-O-2-2-5  
15:30      **Parallel Speaker and Content Modelling for Text-Dependent Speaker Verification**  
*Jianbo Ma, Saad Irtza, Kaavya Sriskandaraja, Vidhyasaharan Sethu, Eliathamby Ambikairajah, University of New South Wales, Australia*

PAGE 440  
Fri-O-2-2-6  
15:45

**i-Vector/HMM Based Text-Dependent Speaker Verification System for RedDots Challenge**

*Hossein Zeinali<sup>1</sup>, Hossein Sameti<sup>1</sup>, Lukáš Burget<sup>2</sup>, Jan Černocký<sup>2</sup>,  
Nooshin Maghsoodi<sup>1</sup>, Pavel Matějka<sup>2</sup>*

*<sup>1</sup>Sharif University of Technology, Iran; <sup>2</sup>Brno University of Technology, Czech Republic*

PAGE 445  
Fri-O-2-2-7  
16:00

**Exploring Session Variability and Template Aging in Speaker Verification for Fixed Phrase Short Utterances**

*Rohan Kumar Das, Sarfaraz Jelil, S.R. Mahadeva Prasanna, IIT Guwahati, India*

**Fri-O-2-3: Articulatory Measurements and Analysis**

*Chairs: Jianwu Dang, Adam Lammert*

*Room: Bayview A, Time 14:30-16:30, Friday, 9 Sept. 2016*

---

PAGE 450  
Fri-O-2-3-1  
14:30

**Prediction of the Articulatory Movements of Unseen Phonemes of a Speaker Using the Speech Structure of Another Speaker**

*Hidetsugu Uchida, Daisuke Saito, Nobuaki Minematsu, University of Tokyo, Japan*

PAGE 455  
Fri-O-2-3-2  
14:50

**Vocal Tract Length Normalization for Speaker Independent Acoustic-to-Articulatory Speech Inversion**

*Ganesh Sivaraman<sup>1</sup>, Vikramjit Mitra<sup>2</sup>, Hosung Nam<sup>3</sup>, Mark Tiede<sup>4</sup>,  
Carol Espy-Wilson<sup>1</sup>*

*<sup>1</sup>University of Maryland, USA; <sup>2</sup>SRI International, USA; <sup>3</sup>Korea University, Korea; <sup>4</sup>Haskins Laboratories, USA*

PAGE 460  
Fri-O-2-3-3  
15:10

**Investigation of Speed-Accuracy Tradeoffs in Speech Production Using Real-Time Magnetic Resonance Imaging**

*Adam C. Lammert<sup>1</sup>, Christine H. Shadle<sup>2</sup>, Shrikanth S. Narayanan<sup>3</sup>,  
Thomas F. Quatieri<sup>1</sup>*

*<sup>1</sup>MIT Lincoln Laboratory, USA; <sup>2</sup>Haskins Laboratories, USA; <sup>3</sup>University of Southern California, USA*

PAGE 465  
Fri-O-2-3-4  
15:30

**Characterizing Vocal Tract Dynamics Across Speakers Using Real-Time MRI**

*Tanner Sorensen, Asterios Toutios, Louis Goldstein, Shrikanth S. Narayanan,  
University of Southern California, USA*

PAGE 470  
Fri-O-2-3-5  
15:50

**Tracking Contours of Orofacial Articulators from Real-Time MRI of Speech**

*Mathieu Labrunie<sup>1</sup>, Pierre Badin<sup>1</sup>, Dirk Voit<sup>2</sup>, Arun A. Joseph<sup>2</sup>, Laurent Lamalle<sup>3</sup>,  
Coriandre Vilain<sup>1</sup>, Louis-Jean Boë<sup>1</sup>, Jens Frahm<sup>2</sup>*

*<sup>1</sup>GIPSA, France; <sup>2</sup>BiomedNMR, Germany; <sup>3</sup>IRMaGe, France*

PAGE 475  
Fri-O-2-3-6  
16:10

**State-of-the-Art MRI Protocol for Comprehensive Assessment of Vocal Tract Structure and Function**

*Sajan Goud Lingala<sup>1</sup>, Asterios Toutios<sup>1</sup>, Johannes Töger<sup>1</sup>, Yongwan Lim<sup>1</sup>, Yinghua Zhu<sup>1</sup>, Yoon-Chul Kim<sup>2</sup>, Colin Vaz<sup>1</sup>, Shrikanth S. Narayanan<sup>1</sup>, Krishna S. Nayak<sup>1</sup>*

<sup>1</sup>University of Southern California, USA; <sup>2</sup>Samsung Medical Center, Korea

**Fri-O-2-4: Automatic Assessment of Emotions**

*Chairs: Carlos Busso, Elizabeth Shriberg*

*Room: Bayview B, Time 14:30-16:30, Friday, 9 Sept. 2016*

---

PAGE 480  
Fri-O-2-4-1  
14:30

**DBN-ivector Framework for Acoustic Emotion Recognition**

*Rui Xia, Yang Liu, University of Texas at Dallas, USA*

PAGE 485  
Fri-O-2-4-2  
14:50

**An Investigation of Emotional Speech in Depression Classification**

*Brian Stasak<sup>1</sup>, Julien Epps<sup>1</sup>, Nicholas Cummins<sup>1</sup>, Roland Goecke<sup>2</sup>*

<sup>1</sup>University of New South Wales, Australia; <sup>2</sup>University of Canberra, Australia

PAGE 490  
Fri-O-2-4-3  
15:10

**Retrieving Categorical Emotions Using a Probabilistic Framework to Define Preference Learning Samples**

*Reza Lotfian, Carlos Busso, University of Texas at Dallas, USA*

PAGE 495  
Fri-O-2-4-4  
15:30

**At the Border of Acoustics and Linguistics: Bag-of-Audio-Words for the Recognition of Emotions in Speech**

*Maximilian Schmitt, Fabien Ringeval, Björn Schuller, Universität Passau, Germany*

PAGE 500  
Fri-O-2-4-5  
15:50

**Speech Emotion Recognition Using Affective Saliency**

*Arodami Chorianopoulou<sup>1</sup>, Polychronis Koutsakis<sup>2</sup>, Alexandros Potamianos<sup>3</sup>*

<sup>1</sup>Technical University of Crete, Greece; <sup>2</sup>Murdoch University, Australia; <sup>3</sup>NTUA, Greece

PAGE 505  
Fri-O-2-4-6  
16:10

**Laughter Valence Prediction in Motivational Interviewing Based on Lexical and Acoustic Cues**

*Rahul Gupta<sup>1</sup>, Nishant Nath<sup>1</sup>, Taruna Agrawal<sup>1</sup>, Panayiotis Georgiou<sup>1</sup>, David C. Atkins<sup>2</sup>, Shrikanth S. Narayanan<sup>1</sup>*

<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Washington, USA

## Fri-O-2-5: Acoustic and Articulatory Phonetics

Chairs: Ailbhe Ni Chasaide, Mattias Heldner

Room: Seacliff BCD, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 510  
Fri-O-2-5-1  
14:30      **Respiratory Belts and Whistles: A Preliminary Study of Breathing Acoustics for Turn-Taking**  
*Marcin Włodarczak, Mattias Heldner, Stockholm University, Sweden*
- PAGE 515  
Fri-O-2-5-2  
14:50      **/r/ as Language Marker in Bilingual Speech Production and Perception**  
*Constantijn Kaland, Vincenzo Galatà, Lorenzo Spreafico, Alessandro Vietti, Libera Università di Bolzano, Italy*
- PAGE 520  
Fri-O-2-5-3  
15:10      **Evaluation of Phonatory Behavior of German and French Speakers in Native and Non-Native Speech**  
*Manfred Pützer<sup>1</sup>, Frank Zimmerer<sup>1</sup>, Wolfgang Wokurek<sup>2</sup>, Jeanin Jügler<sup>1</sup>*  
*<sup>1</sup>Universität des Saarlandes, Germany; <sup>2</sup>Universität Stuttgart, Germany*
- PAGE 525  
Fri-O-2-5-4  
15:30      **Today's Most Frequently Used  $F_0$  Estimation Methods, and Their Accuracy in Estimating Male and Female Pitch in Clean Speech**  
*Sofia Strömbergsson, Karolinska Institute, Sweden*
- PAGE 530  
Fri-O-2-5-5  
15:50      **A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform**  
*Lei He, Volker Dellwo, Universität Zürich, Switzerland*
- PAGE 535  
Fri-O-2-5-6  
16:10      **Likelihood Ratio Calculation in Acoustic-Phonetic Forensic Voice Comparison: Comparison of Three Statistical Modelling Approaches**  
*Ewald Enzinger, University of New South Wales, Australia*

## Fri-O-2-6: Source Separation and Spatial Audio

Chair: Emmanuel Vincent

Room: Seacliff A, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 540  
Fri-O-2-6-1  
14:30      **A Sparse Spherical Harmonic-Based Model in Subbands for Head-Related Transfer Functions**  
*Xiaoke Qi, Jianhua Tao, Chinese Academy of Sciences, China*
- PAGE 545  
Fri-O-2-6-2  
14:50      **Single-Channel Multi-Speaker Separation Using Deep Clustering**  
*Yusuf Isik, Jonathan Le Roux, Zhuo Chen, Shinji Watanabe, John R. Hershey, MERL, USA*
- PAGE 550  
Fri-O-2-6-3  
15:10      **Jointly Optimizing Activation Coefficients of Convolutional NMF Using DNN for Speech Separation**  
*Hao Li<sup>1</sup>, Shuai Nie<sup>2</sup>, Xueliang Zhang<sup>1</sup>, Hui Zhang<sup>1</sup>*  
*<sup>1</sup>Inner Mongolia University, China; <sup>2</sup>Chinese Academy of Sciences, China*
- PAGE 555  
Fri-O-2-6-4  
15:30      **A Feature Study for Masking-Based Reverberant Speech Separation**  
*Masood Delfarah, DeLiang Wang, Ohio State University, USA*
- PAGE 560  
Fri-O-2-6-5  
15:50      **Discriminative Layered Nonnegative Matrix Factorization for Speech Separation**  
*Chung-Chien Hsu, Tai-Shih Chi, Jen-Tzung Chien, National Chiao Tung University, Taiwan*
- PAGE 565  
Fri-O-2-6-6  
16:10      **On Discriminative Framework for Single Channel Audio Source Separation**  
*Arpita Gang, Pravesh Biyani, IIT Delhi, India*

## Fri-P-2-1 : Special Session: Auditory-Visual Expressive Speech and Gesture in Humans and Machines

Chairs: Jeusun Kim, Gérard Bailly

Room: Pacific Concourse - Poster A, Time 14:30-16:30, Friday, 9 Sept. 2016

---

PAGE 570  
Fri-P-2-1-1  
14:30

### Generating Natural Video Descriptions via Multimodal Processing

*Qin Jin<sup>1</sup>, Junwei Liang<sup>2</sup>, Xiaozhu Lin<sup>1</sup>*

<sup>1</sup>Renmin University of China, China; <sup>2</sup>Carnegie Mellon University, USA

PAGE 575  
Fri-P-2-1-2  
14:30

### Feature-Level Decision Fusion for Audio-Visual Word Prominence Detection

*Martin Heckmann, Honda Research Institute Europe, Germany*

PAGE 580  
Fri-P-2-1-3  
14:30

### Acoustic and Visual Analysis of Expressive Speech: A Case Study of French Acted Speech

*Slim Ouni<sup>1</sup>, Vincent Colotte<sup>1</sup>, Sara Dahmani<sup>1</sup>, Soumaya Azzi<sup>2</sup>*

<sup>1</sup>LORIA, France; <sup>2</sup>Polytech Clermont-Ferrand, France

PAGE 585  
Fri-P-2-1-4  
14:30

### Characterization of Audiovisual Dramatic Attitudes

*Adela Barbulescu<sup>1</sup>, Rémi Ronfard<sup>1</sup>, Gérard Bailly<sup>2</sup>*

<sup>1</sup>Inria, France; <sup>2</sup>GIPSA, France

PAGE 590  
Fri-P-2-1-5  
14:30

### Conversational Engagement Recognition Using Auditory and Visual Cues

*Yuyun Huang, Emer Gilmartin, Nick Campbell, Trinity College Dublin, Ireland*

PAGE 595  
Fri-P-2-1-6  
14:30

### An Acoustic Analysis of Child-Child and Child-Robot Interactions for Understanding Engagement during Speech-Controlled Computer Games

*Theodora Chaspari, Jill Fain Lehman, Disney Research, USA*

PAGE 600  
Fri-P-2-1-7  
14:30

### Auditory-Visual Lexical Tone Perception in Thai Elderly Listeners with and without Hearing Impairment

*Benjawan Kasisopa<sup>1</sup>, Chutamane Onsuwan<sup>2</sup>, Charturong Tantibundhit<sup>2</sup>, Nittayapa Klangpornkun<sup>2</sup>, Suparak Techacharoenrungrueang<sup>3</sup>, Sudaporn Luksaneeyanawin<sup>3</sup>, Denis Burnham<sup>1</sup>*

<sup>1</sup>Western Sydney University, Australia; <sup>2</sup>Thammasat University, Thailand; <sup>3</sup>Chulalongkorn University, Thailand

PAGE 605  
Fri-P-2-1-8  
14:30

### Use of Agreement/Disagreement Classification in Dyadic Interactions for Continuous Emotion Recognition

*Hossein Khaki, Engin Erzin, Koç Üniversitesi, Turkey*

## Fri-P-2-2 : Special Session: Intelligibility Under the Microscope

Chairs: Ricard Marxer, Martin Cooke, Jon Barker

Room: Pacific Concourse - Poster B, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 610  
Fri-P-2-2-1  
14:30  
**Microscopic Multilingual Matrix Test Predictions Using an ASR-Based Speech Recognition Model**  
*Marc René Schädler, David Hülsmeier, Anna Warzybok, Sabine Hochmuth, Birger Kollmeier, Carl von Ossietzky Universität Oldenburg, Germany*
- PAGE 615  
Fri-P-2-2-2  
14:30  
**DNN-Based Automatic Speech Recognition as a Model for Human Phoneme Perception**  
*Mats Exter<sup>1</sup>, Bernd T. Meyer<sup>2</sup>*  
*<sup>1</sup>Carl von Ossietzky Universität Oldenburg, Germany; <sup>2</sup>Johns Hopkins University, USA*
- PAGE 620  
Fri-P-2-2-3  
14:30  
**Undoing Misperceptions: A Microscopic Analysis of Consistent Confusions Through Signal Modifications**  
*Attila Máté Tóth<sup>1</sup>, Martin Cooke<sup>2</sup>*  
*<sup>1</sup>Universidad del País Vasco, Spain; <sup>2</sup>Ikerbasque, Spain*
- PAGE 625  
Fri-P-2-2-4  
14:30  
**Blind Non-Intrusive Speech Intelligibility Prediction Using Twin-HMMs**  
*Mahdie Karbasi<sup>1</sup>, Ahmed Hussien Abdelaziz<sup>2</sup>, Hendrik Meutzner<sup>1</sup>, Dorothea Kolossa<sup>1</sup>*  
*<sup>1</sup>Ruhr-Universität Bochum, Germany; <sup>2</sup>ICSI, USA*
- PAGE 630  
Fri-P-2-2-5  
14:30  
**Misperceptions Arising from Speech-in-Babble Interactions**  
*Attila Máté Tóth<sup>1</sup>, Martin Cooke<sup>2</sup>, Jon Barker<sup>3</sup>*  
*<sup>1</sup>Universidad del País Vasco, Spain; <sup>2</sup>Ikerbasque, Spain; <sup>3</sup>University of Sheffield, UK*
- PAGE 635  
Fri-P-2-2-6  
14:30  
**Introducing Temporal Rate Coding for Speech in Cochlear Implants: A Microscopic Evaluation in Humans and Models**  
*Anja Eichenauer<sup>1</sup>, Mathias Dietz<sup>1</sup>, Bernd T. Meyer<sup>2</sup>, Tim Jürgens<sup>1</sup>*  
*<sup>1</sup>Carl von Ossietzky Universität Oldenburg, Germany; <sup>2</sup>Johns Hopkins University, USA*

- PAGE 640  
Fri-P-2-2-7  
14:30  
**Language Effects in Noise-Induced Word Misperceptions**  
*Maria Luisa Garcia Lecumberri<sup>1</sup>, Jon Barker<sup>2</sup>, Ricard Marxer<sup>2</sup>, Martin Cooke<sup>3</sup>*  
*<sup>1</sup>Universidad del País Vasco, Spain; <sup>2</sup>University of Sheffield, UK; <sup>3</sup>Ikerbasque, Spain*
- PAGE 645  
Fri-P-2-2-8  
14:30  
**Speech Reductions Cause a De-Weighting of Secondary Acoustic Cues**  
*Léo Varnet<sup>1</sup>, Fanny Meunier<sup>2</sup>, Michel Hoen<sup>3</sup>*  
*<sup>1</sup>LSP (UMR 8248), France; <sup>2</sup>BCL (UMR 7320), France; <sup>3</sup>Oticon Medical, France*
- PAGE 650  
Fri-P-2-2-9  
14:30  
**Using Phonologically Weighted Levenshtein Distances for the Prediction of Microscopic Intelligibility**  
*Lionel Fontan<sup>1</sup>, Isabelle Ferrané<sup>2</sup>, Jérôme Farinas<sup>2</sup>, Julien Pinquier<sup>2</sup>, Xavier Aumont<sup>1</sup>*  
*<sup>1</sup>Archean Technologies, France; <sup>2</sup>IRIT, France*
- PAGE 655  
Fri-P-2-2-10  
14:30  
**The Impact of Manner of Articulation on the Intelligibility of Voicing Contrast in Noise: Cross-Linguistic Implications**  
*Mayuki Matsui, NINJAL, Japan*
- PAGE 660  
Fri-P-2-2-11  
14:30  
**Directly Comparing the Listening Strategies of Humans and Machines**  
*Michael I. Mandel, CUNY Brooklyn College, USA*



**Fri-P-2-3 : Spoken Documents, Spoken Understanding and Semantic Analysis**

Chair: Helen Meng

Room: Pacific Concourse - Poster C, Time 14:30-16:30, Friday, 9 Sept. 2016

---

- PAGE 665  
Fri-P-2-3-1  
14:30
- LSTM-Based NeuroCRFs for Named Entity Recognition**  
*Marc-Antoine Rondeau<sup>1</sup>, Yi Su<sup>2</sup>*  
<sup>1</sup>McGill University, Canada; <sup>2</sup>Nuance Communications, Canada
- PAGE 670  
Fri-P-2-3-2  
14:30
- Exploring Word Mover's Distance and Semantic-Aware Embedding Techniques for Extractive Broadcast News Summarization**  
*Shih-Hung Liu<sup>1</sup>, Kuan-Yu Chen<sup>1</sup>, Yu-Lun Hsieh<sup>1</sup>, Berlin Chen<sup>2</sup>, Hsin-Min Wang<sup>1</sup>, Hsu-Chun Yen<sup>3</sup>, Wen-Lian Hsu<sup>1</sup>*  
<sup>1</sup>Academia Sinica, Taiwan; <sup>2</sup>National Taiwan Normal University, Taiwan; <sup>3</sup>National Taiwan University, Taiwan
- PAGE 675  
Fri-P-2-3-3  
14:30
- Improved Neural Bag-of-Words Model to Retrieve Out-of-Vocabulary Words in Speech Recognition**  
*Imran Sheikh<sup>1</sup>, Irina Illina<sup>1</sup>, Dominique Fohr<sup>1</sup>, Georges Linarès<sup>2</sup>*  
<sup>1</sup>LORIA, France; <sup>2</sup>LIA, France
- PAGE 680  
Fri-P-2-3-4  
14:30
- Beyond Utterance Extraction: Summary Recombination for Speech Summarization**  
*Jérémy Trione, Benoit Favre, Frederic Bechet, LIF (UMR 7279), France*
- PAGE 685  
Fri-P-2-3-5  
14:30
- Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling**  
*Bing Liu, Ian Lane, Carnegie Mellon University, USA*

- PAGE 690  
Fri-P-2-3-6  
14:30
- Domain Adaptation of Recurrent Neural Networks for Natural Language Understanding**  
*Aaron Jaech<sup>1</sup>, Larry Heck<sup>2</sup>, Mari Ostendorf<sup>1</sup>*  
<sup>1</sup>University of Washington, USA; <sup>2</sup>Google, USA
- PAGE 695  
Fri-P-2-3-7  
14:30
- LATTICERNN: Recurrent Neural Networks Over Lattices**  
*Faisal Ladhak, Ankur Gandhe, Markus Dreyer, Lambert Mathias, Ariya Rastrow, Björn Hoffmeister, Amazon.com, USA*
- PAGE 700  
Fri-P-2-3-8  
14:30
- Learning Document Representations Using Subspace Multinomial Model**  
*Santosh Kesiraju, Lukáš Burget, Igor Szóke, Jan Černocký, Brno University of Technology, Czech Republic*
- PAGE 705  
Fri-P-2-3-9  
14:30
- Attention-Based Convolutional Neural Networks for Sentence Classification**  
*Zhiwei Zhao, Youzheng Wu, Sony, China*
- PAGE 710  
Fri-P-2-3-10  
14:30
- Spoken Language Understanding in a Latent Topic-Based Subspace**  
*Mohamed Morchid, Mohamed Bouaziz, Waad Ben Kheder, Killian Janod, Pierre-Michel Bousquet, Richard Dufour, Georges Linarès, LIA, France*
- PAGE 715  
Fri-P-2-3-11  
14:30
- Multi-Domain Joint Semantic Frame Parsing Using Bi-Directional RNN-LSTM**  
*Dilek Hakkani-Tür<sup>1</sup>, Gokhan Tur<sup>1</sup>, Asli Celikyilmaz<sup>1</sup>, Yun-Nung Chen<sup>2</sup>, Jianfeng Gao<sup>1</sup>, Li Deng<sup>1</sup>, Ye-Yi Wang<sup>1</sup>*  
<sup>1</sup>Microsoft, USA; <sup>2</sup>National Taiwan University, Taiwan
- PAGE 720  
Fri-P-2-3-12  
14:30
- Deep Stacked Autoencoders for Spoken Language Understanding**  
*Killian Janod, Mohamed Morchid, Richard Dufour, Georges Linarès, Renato De Mori, LIA, France*

PAGE 725  
Fri-P-2-3-13  
14:30

**Labeled Data Generation with Encoder-Decoder LSTM for Semantic Slot Filling**  
*Gakuto Kurata<sup>1</sup>, Bing Xiang<sup>2</sup>, Bowen Zhou<sup>2</sup>*  
<sup>1</sup>IBM, Japan; <sup>2</sup>IBM, USA

PAGE 730  
Fri-P-2-3-14  
14:30

**Exploring the Correlation of Pitch Accents and Semantic Slots for Spoken Language Understanding**  
*Sabrina Stehwien, Ngoc Thang Vu, Universität Stuttgart, Germany*

PAGE 735  
Fri-P-2-3-15  
14:30

**Analysis on Gated Recurrent Unit Based Question Detection Approach**  
*Yaodong Tang, Zhiyong Wu, Helen Meng, Mingxing Xu, Lianhong Cai, Tsinghua University, China*

## Fri-P-2-4: Spoken Term Detection

*Chairs: Laura June Mayfield Tomokiyo*

*Room: Pacific Concourse - Poster D, Time 14:30-16:30, Friday, 9 Sept. 2016*

---

PAGE 740  
Fri-P-2-4-1  
14:30

**Combining State-Level Spotting and Posterior-Based Acoustic Match for Improved Query-by-Example Spoken Term Detection**  
*Shuji Oishi, Tatsuya Matsuba, Mitsuaki Makino, Atsuhiko Kai, Shizuoka University, Japan*

PAGE 745  
Fri-P-2-4-2  
14:30

**A Novel Discriminative Score Calibration Method for Keyword Search**  
*Zhiqiang Lv, Meng Cai, Wei-Qiang Zhang, Jia Liu, Tsinghua University, China*

PAGE 750  
Fri-P-2-4-3  
14:30

**Segmented Dynamic Time Warping for Spoken Query-by-Example Search**  
*Jorge Proença, Fernando Perdigão, Instituto de Telecomunicações, Portugal*

PAGE 755  
Fri-P-2-4-4  
14:30

**Generating Complementary Acoustic Model Spaces in DNN-Based Sequence-to-Frame DTW Scheme for Out-of-Vocabulary Spoken Term Detection**  
*Shi-wook Lee<sup>1</sup>, Kazuyo Tanaka<sup>2</sup>, Yoshiaki Itoh<sup>3</sup>*  
<sup>1</sup>AIST, Japan; <sup>2</sup>University of Tsukuba, Japan; <sup>3</sup>Iwate Prefectural University, Japan

PAGE 760  
Fri-P-2-4-5  
14:30

**Multi-Task Learning and Weighted Cross-Entropy for DNN-Based Keyword Spotting**  
*Sankaran Panchapagesan, Ming Sun, Aparna Khare, Spyros Matsoukas, Arindam Mandal, Björn Hoffmeister, Shiv Vitaladevuni, Amazon.com, USA*

PAGE 765  
Fri-P-2-4-6  
14:30

**Audio Word2Vec: Unsupervised Learning of Audio Segment Representations Using Sequence-to-Sequence Autoencoder**  
*Yu-An Chung, Chao-Chung Wu, Chia-Hao Shen, Hung-Yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan*

PAGE 770  
Fri-P-2-4-7  
14:30

**Non-Uniform Boosted MCE Training of Deep Neural Networks for Keyword Spotting**

*Zhong Meng, Biing-Hwang Juang, Georgia Institute of Technology, USA*

PAGE 775  
Fri-P-2-4-8  
14:30

**Language Model Data Augmentation for Keyword Spotting in Low-Resourced Training Conditions**

*Arseniy Gorin<sup>1</sup>, Rasa Lileikytė<sup>1</sup>, Guangpu Huang<sup>1</sup>, Lori Lamel<sup>1</sup>, Jean-Luc Gauvain<sup>1</sup>, Antoine Laurent<sup>2</sup>*

*<sup>1</sup>LIMSI, France; <sup>2</sup>Vocapia Research, France*

**Fri-S&T-2: Show & Tell Session 2**

*Chairs: Nicolas Scheffer, Shiva Sundaram*

*Room: Market Street Foyer, Time 14:30–16:30, Friday, 9 Sept. 2016*

---

PAGE 780  
Fri-S&T-2-1  
14:30

**STON: Efficient Subtitling in Dutch Using State-of-the-Art Tools**

*Lyan Verwimp<sup>1</sup>, Brecht Desplanques<sup>2</sup>, Kris Demuyne<sup>2</sup>, Joris Pelemans<sup>1</sup>, Marieke Lycke<sup>3</sup>, Patrick Wambacq<sup>1</sup>*

*<sup>1</sup>Katholieke Universiteit Leuven, Belgium; <sup>2</sup>Ghent University, Belgium; <sup>3</sup>VRT, Belgium*

PAGE 782  
Fri-S&T-2-2  
14:30

**An Automatic Training Tool for Air Traffic Control Training**

*Petr Stanislav, Luboš Šmíd, Jan Švec, University of West Bohemia, Czech Republic*

PAGE 784  
Fri-S&T-2-3  
14:30

**Digitala: An Augmented Test and Review Process Prototype for High-Stakes Spoken Foreign Language Examination**

*Reima Karhila<sup>1</sup>, Aku Rouhe<sup>1</sup>, Peter Smit<sup>1</sup>, André Mansikkaniemi<sup>1</sup>, Heini Kallio<sup>2</sup>, Erik Lindroos<sup>2</sup>, Raili Hildén<sup>2</sup>, Martti Vainio<sup>2</sup>, Mikko Kurimo<sup>1</sup>*

*<sup>1</sup>Aalto University, Finland; <sup>2</sup>University of Helsinki, Finland*

PAGE 786  
Fri-S&T-2-4  
14:30

**Exploring Collections of Multimedia Archives Through Innovative Interfaces in the Context of Digital Humanities**

*Géraldine Damnati, Delphine Charlet, Marc Denjean, Orange Labs, France*

## Fri-O-3-1 : Feature Extraction and Acoustic Modeling Using Neural Networks for ASR

Chairs: Hui Jiang, Kate Krill

Room: Grand Ballroom A, Time 17:00-19:00, Friday, 9 Sept. 2016

---

- PAGE 788  
Fri-O-3-1-1  
17:00      **Learning Neural Network Representations Using Cross-Lingual Bottleneck Features with Word-Pair Information**  
*Yougen Yuan<sup>1</sup>, Cheung-Chi Leung<sup>2</sup>, Lei Xie<sup>1</sup>, Bin Ma<sup>2</sup>, Haizhou Li<sup>2</sup>*  
*<sup>1</sup>Northwestern Polytechnical University, China; <sup>2</sup>A\*STAR, Singapore*
- PAGE 793  
Fri-O-3-1-2  
17:20      **Novel Front-End Features Based on Neural Graph Embeddings for DNN-HMM and LSTM-CTC Acoustic Modeling**  
*Yuzong Liu<sup>1</sup>, Katrin Kirchhoff<sup>2</sup>*  
*<sup>1</sup>Amazon.com, USA; <sup>2</sup>University of Washington, USA*
- PAGE 798  
Fri-O-3-1-3  
17:40      **Articulatory Feature Extraction Using CTC to Build Articulatory Classifiers Without Forced Frame Alignments for Speech Recognition**  
*Basil Abraham, S. Umesh, Neethu Mariam Joy, IIT Madras, India*
- PAGE 803  
Fri-O-3-1-4  
18:00      **On the Role of Nonlinear Transformations in Deep Neural Network Acoustic Models**  
*Tasha Nagamine<sup>1</sup>, Michael L. Seltzer<sup>2</sup>, Nima Mesgarani<sup>1</sup>*  
*<sup>1</sup>Columbia University, USA; <sup>2</sup>Microsoft, USA*
- PAGE 808  
Fri-O-3-1-5  
18:20      **Complex Linear Projection (CLP): A Discriminative Approach to Joint Feature Extraction and Acoustic Modeling**  
*Ehsan Variani, Tara N. Sainath, Izhak Shafran, Michiel Bacchiani, Google, USA*
- PAGE 813  
Fri-O-3-1-6  
18:40      **Modeling Time-Frequency Patterns with LSTM vs. Convolutional Architectures for LVCSR Tasks**  
*Tara N. Sainath, Bo Li, Google, USA*

## Fri-O-3-2 : Special Session: The Speakers in the Wild (SITW) Speaker Recognition Challenge

Chairs: Mitchell McLaren, Aaron Lawson, Luciana Ferrer, Diego Castan

Room: Grand Ballroom BC, Time 17:00-19:00, Friday, 9 Sept. 2016

---

- PAGE 818  
Fri-O-3-2-1  
17:00      **The Speakers in the Wild (SITW) Speaker Recognition Database**  
*Mitchell McLaren<sup>1</sup>, Luciana Ferrer<sup>2</sup>, Diego Castan<sup>1</sup>, Aaron Lawson<sup>1</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>Universidad de Buenos Aires, Argentina*
- PAGE 823  
Fri-O-3-2-2  
17:15      **The 2016 Speakers in the Wild Speaker Recognition Evaluation**  
*Mitchell McLaren<sup>1</sup>, Luciana Ferrer<sup>2</sup>, Diego Castan<sup>1</sup>, Aaron Lawson<sup>1</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>Universidad de Buenos Aires, Argentina*
- PAGE 828  
Fri-O-3-2-3  
17:30      **Analysis of Speaker Recognition Systems in Realistic Scenarios of the SITW 2016 Challenge**  
*Ondřej Novotný, Pavel Matějka, Oldřich Plchot, Ondřej Glembek, Lukáš Burget, Jan Černocký, Brno University of Technology, Czech Republic*
- PAGE 833  
Fri-O-3-2-4  
17:45      **A Speaker Recognition System for the SITW Challenge**  
*Oleg Kudashev<sup>1</sup>, Sergey Novoselov<sup>1</sup>, Konstantin Simonchik<sup>1</sup>, Alexandr Kozlov<sup>2</sup>*  
*<sup>1</sup>ITMO University, Russia; <sup>2</sup>Speech Technology Center, Russia*
- PAGE 838  
Fri-O-3-2-5  
18:00      **Speakers In The Wild (SITW): The QUT Speaker Recognition System**  
*H. Ghaemmaghami, M.H. Rahman, Ivan Himawan, David Dean, Ahilan Kanagasundaram, Sridha Sridharan, Clinton Fookes, Queensland University of Technology, Australia*
- PAGE 843  
Fri-O-3-2-6  
18:15      **AUT System for SITW Speaker Recognition Challenge**  
*Abbas Khosravani, Mohammad Mehdi Homayounpour, Amirkabir University of Technology, Iran*

PAGE 848  
Fri-O-3-2-7  
18:30

**LIA System for the SITW Speaker Recognition Challenge**

*Waad Ben Kheder, Moez Ajili, Pierre-Michel Bousquet, Driss Matrouf, Jean-François Bonastre, LIA, France*

PAGE 853  
Fri-O-3-2-8  
18:45

**Investigating Various Diarization Algorithms for Speaker in the Wild (SITW) Speaker Recognition Challenge**

*Yi Liu, Yao Tian, Liang He, Jia Liu, Tsinghua University, China*

**Fri-O-3-3 : Non-Native Speech Perception**

*Chairs: Outi Tuomainen, Florian Hintz*

*Room: Bayview A, Time 17:00-19:00, Friday, 9 Sept. 2016*

---

PAGE 858  
Fri-O-3-3-1  
17:00

**Does the Importance of Word-Initial and Word-Final Information Differ in Native versus Non-Native Spoken-Word Recognition?**

*Odette Scharenborg<sup>1</sup>, Juul Coumans<sup>1</sup>, Sofoklis Kakouros<sup>2</sup>, Roeland van Hout<sup>1</sup>  
<sup>1</sup>Radboud Universiteit Nijmegen, The Netherlands; <sup>2</sup>Aalto University, Finland*

PAGE 863  
Fri-O-3-3-2  
17:20

**The Effect of Sentence Accent on Non-Native Speech Perception in Noise**

*Odette Scharenborg<sup>1</sup>, Elea Kolkman<sup>1</sup>, Sofoklis Kakouros<sup>2</sup>, Brechtje Post<sup>3</sup>  
<sup>1</sup>Radboud Universiteit Nijmegen, The Netherlands; <sup>2</sup>Aalto University, Finland;  
<sup>3</sup>University of Cambridge, UK*

PAGE 868  
Fri-O-3-3-3  
17:40

**The Effects of Modified Speech Styles on Intelligibility for Non-Native Listeners**

*Martin Cooke<sup>1</sup>, Maria Luisa Garcia Lecumberri<sup>2</sup>  
<sup>1</sup>Ikerbasque, Spain; <sup>2</sup>Universidad del Pais Vasco, Spain*

PAGE 873  
Fri-O-3-3-4  
18:00

**The Influence of Language Experience on the Categorical Perception of Vowels: Evidence from Mandarin and Korean**

*Hao Zhang<sup>1</sup>, Fei Chen<sup>1</sup>, Nan Yan<sup>1</sup>, Lan Wang<sup>1</sup>, Feng Shi<sup>2</sup>, Manwa L. Ng<sup>3</sup>  
<sup>1</sup>Chinese Academy of Sciences, China; <sup>2</sup>Nankai University, China; <sup>3</sup>University of Hong Kong, China*

PAGE 878  
Fri-O-3-3-5  
18:20

**Multiple Influences on Vocabulary Acquisition: Parental Input Dominates**

*Dominic W. Massaro, University of California at Santa Cruz, USA*

PAGE 883  
Fri-O-3-3-6  
18:40

**Can Intensive Exposure to Foreign Language Sounds Affect the Perception of Native Sounds?**

*Jian Gong<sup>1</sup>, Maria Luisa Garcia Lecumberri<sup>2</sup>, Martin Cooke<sup>3</sup>  
<sup>1</sup>JUST, China; <sup>2</sup>Universidad del País Vasco, Spain; <sup>3</sup>Ikerbasque, Spain*

**Fri-O-3-4: Behavioral Signal Processing and Speaker State and Traits Analytics**

*Chairs: Joseph Tepperman, Chi-Chun (Jeremy) Lee*

*Room: Bayview B, Time 17:00–19:00, Friday, 9 Sept. 2016*

---

PAGE 888  
Fri-O-3-4-1  
17:00

**Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration**

*Nikoletta Bassiou, Andreas Tsiartas, Jennifer Smith, Harry Bratt, Colleen Richey, Elizabeth Shriberg, Cynthia D'Angelo, Nonye Alozie, SRI International, USA*

PAGE 893  
Fri-O-3-4-2  
17:20

**Complexity in Prosody: A Nonlinear Dynamical Systems Approach for Dyadic Conversations; Behavior and Outcomes in Couples Therapy**

*Md. Nasir<sup>1</sup>, Brian Baucom<sup>2</sup>, Shrikanth S. Narayanan<sup>1</sup>, Panayiotis Georgiou<sup>1</sup>  
<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Utah, USA*

PAGE 898  
Fri-O-3-4-3  
17:40

**Couples Behavior Modeling and Annotation Using Low-Resource LSTM Language Models**

*Shao-Yen Tseng<sup>1</sup>, Sandeep Nallan Chakravarthula<sup>1</sup>, Brian Baucom<sup>2</sup>, Panayiotis Georgiou<sup>1</sup>  
<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Utah, USA*

PAGE 903  
Fri-O-3-4-4  
18:00

**Speech Likability and Personality-Based Social Relations: A Round-Robin Analysis over Communication Channels**

*Laura Fernández Gallardo, Benjamin Weiss, T-Labs, Germany*

PAGE 908  
Fri-O-3-4-5  
18:20

**Behavioral Coding of Therapist Language in Addiction Counseling Using Recurrent Neural Networks**

*Bo Xiao<sup>1</sup>, Doğan Can<sup>1</sup>, James Gibson<sup>1</sup>, Zac E. Imel<sup>2</sup>, David C. Atkins<sup>3</sup>, Panayiotis Georgiou<sup>1</sup>, Shrikanth S. Narayanan<sup>1</sup>  
<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Utah, USA; <sup>3</sup>University of Washington, USA*

PAGE 913  
Fri-O-3-4-6  
18:40

**Factor Analysis Based Speaker Normalisation for Continuous Emotion Prediction**

*Ting Dang, Vidhyasaharan Sethu, Eliathamby Ambikairajah, University of New South Wales, Australia*

## Fri-O-3-5: Spoken Term Detection

*Chairs: Isabel Trancoso, Roland Kuhn*

*Room: Seacliff BCD, Time 17:00-19:00, Friday, 9 Sept. 2016*

---

- PAGE 918  
Fri-O-3-5-1  
17:00      **Subspace Detection of DNN Posterior Probabilities via Sparse Representation for Query by Example Spoken Term Detection**  
*Dhananjay Ram, Afsaneh Asaei, Hervé Bourlard, Idiap Research Institute, Switzerland*
- PAGE 923  
Fri-O-3-5-2  
17:20      **Unsupervised Bottleneck Features for Low-Resource Query-by-Example Spoken Term Detection**  
*Hongjie Chen<sup>1</sup>, Cheung-Chi Leung<sup>2</sup>, Lei Xie<sup>1</sup>, Bin Ma<sup>2</sup>, Haizhou Li<sup>2</sup>*  
*<sup>1</sup>Northwestern Polytechnical University, China; <sup>2</sup>A\*STAR, Singapore*
- PAGE 928  
Fri-O-3-5-3  
17:40      **A Nonparametric Bayesian Approach for Spoken Term Detection by Example Query**  
*Amir Hossein Harati Nejad Torbati, Joseph Picone, Temple University, USA*
- PAGE 933  
Fri-O-3-5-4  
18:00      **Rescoring Hypothesized Detections of Out-of-Vocabulary Keywords Using Subword Samples**  
*Van Tung Pham<sup>1</sup>, Haihua Xu<sup>2</sup>, Xiong Xiao<sup>2</sup>, Nancy F. Chen<sup>3</sup>, Eng Siong Chng<sup>1</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>NTU, Singapore; <sup>2</sup>TL@NTU, Singapore; <sup>3</sup>A\*STAR, Singapore*
- PAGE 938  
Fri-O-3-5-5  
18:20      **Unrestricted Vocabulary Keyword Spotting Using LSTM-CTC**  
*Yimeng Zhuang, Xuankai Chang, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China*
- PAGE 943  
Fri-O-3-5-6  
18:40      **Interactive Spoken Content Retrieval by Deep Reinforcement Learning**  
*Yen-Chen Wu, Tzu-Hsiang Lin, Yang-De Chen, Hung-Yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan*

## Fri-O-3-6: Co-Inference of Production and Acoustics

*Chair: Carol Epsy-Wilson*

*Room: Seacliff A, Time 17:00-19:00, Friday, 9 Sept. 2016*

---

- PAGE 948  
Fri-O-3-6-1  
17:00      **Relating Estimated Cyclic Spectral Peak Frequency to Measured Epilarynx Length Using Magnetic Resonance Imaging**  
*Elizabeth Godoy, Andrew Dumas, Jennifer Melot, Nicolas Malyska, Thomas F. Quatieri, MIT Lincoln Laboratory, USA*
- PAGE 953  
Fri-O-3-6-2  
17:20      **Acoustic-to-Articulatory Inversion Mapping Based on Latent Trajectory Gaussian Mixture Model**  
*Patrick Lumban Tobing<sup>1</sup>, Tomoki Toda<sup>2</sup>, Hirokazu Kameoka<sup>3</sup>, Satoshi Nakamura<sup>1</sup>*  
*<sup>1</sup>NAIST, Japan; <sup>2</sup>Nagoya University, Japan; <sup>3</sup>NTT, Japan*
- PAGE 958  
Fri-O-3-6-3  
17:40      **Formant Estimation and Tracking Using Deep Learning**  
*Yehoshua Dissen, Joseph Keshet, Bar-Ilan University, Israel*
- PAGE 963  
Fri-O-3-6-4  
18:00      **Convex Hull Convolutional Non-Negative Matrix Factorization for Uncovering Temporal Patterns in Multivariate Time-Series Data**  
*Colin Vaz, Asterios Toutios, Shrikanth S. Narayanan, University of Southern California, USA*
- PAGE 968  
Fri-O-3-6-5  
18:20      **Majorisation-Minimisation Based Optimisation of the Composite Autoregressive System with Application to Glottal Inverse Filtering**  
*Lauri Juvela<sup>1</sup>, Hirokazu Kameoka<sup>2</sup>, Manu Airaksinen<sup>1</sup>, Junichi Yamagishi<sup>3</sup>, Paavo Alku<sup>1</sup>*  
*<sup>1</sup>Aalto University, Finland; <sup>2</sup>University of Tokyo, Japan; <sup>3</sup>NII, Japan*
- PAGE 973  
Fri-O-3-6-6  
18:40      **F<sub>0</sub> Contour Analysis Based on Empirical Mode Decomposition for DNN Acoustic Modeling in Mandarin Speech Recognition**  
*Xiaoyun Wang<sup>1</sup>, Xugang Lu<sup>1</sup>, Hisashi Kawai<sup>1</sup>, Seiichi Yamamoto<sup>2</sup>*  
*<sup>1</sup>NICT, Japan; <sup>2</sup>Doshisha University, Japan*

## Fri-P-3-1 : Acoustic and Articulatory Phonetics

Chair: Marija Tabain

Room: Pacific Concourse - Poster A, Time 17:00-19:00, Friday, 9 Sept. 2016

---

- PAGE 978  
Fri-P-3-1-1  
17:00      **Vowels and Diphthongs in Cangnan Southern Min Chinese Dialect**  
*Fang Hu, Chunyu Ge, Chinese Academy of Social Sciences, China*
- PAGE 983  
Fri-P-3-1-2  
17:00      **Diphthongization of Nuclear Vowels and the Emergence of a Tetraphthong in Hetang Cantonese**  
*Wenqi Hu<sup>1</sup>, Fang Hu<sup>2</sup>, Jian Jin<sup>1</sup>*  
*<sup>1</sup>Sun Yat-Sen University, China; <sup>2</sup>Chinese Academy of Social Sciences, China*
- PAGE 988  
Fri-P-3-1-3  
17:00      **PhonVoc: A Phonetic and Phonological Vocoding Toolkit**  
*Milos Cernak, Philip N. Garner, Idiap Research Institute, Switzerland*
- PAGE 993  
Fri-P-3-1-4  
17:00      **Vowels and Diphthongs in the Taiyuan Jin Chinese Dialect**  
*Liping Xia, Fang Hu, Chinese Academy of Social Sciences, China*
- PAGE 998  
Fri-P-3-1-5  
17:00      **The Effects of Prosody on French V-to-V Coarticulation: A Corpus-Based Study**  
*Giuseppina Turco, Cécile Fougeron, Nicolas Audibert, LPP (UMR 7018), France*
- PAGE 1002  
Fri-P-3-1-6  
17:00      **An Acoustic Analysis of /r/ in Tyrolean**  
*Vincenzo Galatà, Lorenzo Spreafico, Alessandro Vietti, Constantijn Kaland, Libera Università di Bolzano, Italy*
- PAGE 1007  
Fri-P-3-1-7  
17:00      **Hyperarticulated Production of Korean Glides by Age Group**  
*Seung-Eun Chang, Minsook Kim, University of California at Berkeley, USA*
- PAGE 1011  
Fri-P-3-1-8  
17:00      **Coda Stop and Taiwan Min Checked Tone Sound Changes**  
*Ho-hsien Pan, Hsiao-tung Huang, Shao-ren Lyu, National Chiao Tung University, Taiwan*

## Fri-P-3-2 : Prosody, Phonation and Voice Quality

Chair: Irene Vogel

Room: Pacific Concourse - Poster B, Time 17:00-19:00, Friday, 9 Sept. 2016

---

- PAGE 1016  
Fri-P-3-2-1  
17:00      **The Influence of Modality and Speaking Style on the Assimilation Type and Categorization Consistency of Non-Native Speech**  
*Sarah E. Fenwick, Catherine T. Best, Chris Davis, Michael D. Tyler, Western Sydney University, Australia*
- PAGE 1021  
Fri-P-3-2-2  
17:00      **Prosodic Convergence with Spoken Stimuli in Laboratory Data**  
*Margaret Zellers, Universität Stuttgart, Germany*
- PAGE 1026  
Fri-P-3-2-3  
17:00      **Effects of Stress on Fricatives: Evidence from Standard Modern Greek**  
*Charalambos Themistocleous<sup>1</sup>, Angelandria Savva<sup>2</sup>, Andrie Aristodemou<sup>2</sup>*  
*<sup>1</sup>University of Gothenburg, Sweden; <sup>2</sup>University of Cyprus, Cyprus*
- PAGE 1030  
Fri-P-3-2-4  
17:00      **Analysis of Chinese Syllable Durations in Running Speech of Japanese L2 Learners**  
*Yue Sun<sup>1</sup>, Shudon Hsiao<sup>1</sup>, Yoshinori Sagisaka<sup>1</sup>, Jinsong Zhang<sup>2</sup>*  
*<sup>1</sup>Waseda University, Japan; <sup>2</sup>BLCU, China*
- PAGE 1034  
Fri-P-3-2-5  
17:00      **Automatic Paragraph Segmentation with Lexical and Prosodic Features**  
*Catherine Lai<sup>1</sup>, Mireia Farrús<sup>2</sup>, Johanna D. Moore<sup>1</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>Universitat Pompeu Fabra, Spain*
- PAGE 1039  
Fri-P-3-2-6  
17:00      **Automatic Glottal Inverse Filtering with Non-Negative Matrix Factorization**  
*Manu Airaksinen<sup>1</sup>, Lauri Juvela<sup>1</sup>, Tom Bäckström<sup>2</sup>, Paavo Alku<sup>1</sup>*  
*<sup>1</sup>Aalto University, Finland; <sup>2</sup>FAU Erlangen-Nürnberg, Germany*
- PAGE 1044  
Fri-P-3-2-7  
17:00      **Speaker Identity and Voice Quality: Modeling Human Responses and Automatic Speaker Recognition**  
*Soo Jin Park<sup>1</sup>, Caroline Sigouin<sup>2</sup>, Jody Kreiman<sup>1</sup>, Patricia Keating<sup>1</sup>, Jinxi Guo<sup>1</sup>, Gary Yeung<sup>1</sup>, Fang-Yu Kuo<sup>1</sup>, Abeer Alwan<sup>1</sup>*  
*<sup>1</sup>University of California at Los Angeles, USA; <sup>2</sup>Université Laval, Canada*



PAGE 1049 Fri-P-3-2-8 17:00	<b>Analysis of Glottal Stop in Assam Sora Language</b> <i>Sishir Kalita, Luke Horo, Priyankoo Sarmah, S.R. Mahadeva Prasanna, S. Dandapat, IIT Guwahati, India</i>
PAGE 1054 Fri-P-3-2-9 17:00	<b>Acoustic Differences Between English /t/ Glottalization and Phrasal Creak</b> <i>Marc Garellek, Scott Seyfarth, University of California at San Diego, USA</i>
PAGE 1059 Fri-P-3-2-10 17:00	<b>The Acoustics of Lexical Stress in Italian as a Function of Stress Level and Speaking Style</b> <i>Anders Eriksson<sup>1</sup>, Pier Marco Bertinetto<sup>2</sup>, Mattias Heldner<sup>1</sup>, Rosalba Nodari<sup>2</sup>, Giovanna Lenoci<sup>2</sup></i> <i><sup>1</sup>Stockholm University, Sweden; <sup>2</sup>Scuola Normale Superiore, Italy</i>
PAGE 1064 Fri-P-3-2-11 17:00	<b>Cross-Gender and Cross-Dialect Tone Recognition for Vietnamese</b> <i>Antje Schweitzer, Ngoc Thang Vu, Universität Stuttgart, Germany</i>
PAGE 1069 Fri-P-3-2-12 17:00	<b>Prosody Modification Using Allpass Residual of Speech Signals</b> <i>Karthika Vijayan, K. Sri Rama Murty, IIT Hyderabad, India</i>
PAGE 1074 Fri-P-3-2-13 17:00	<b>Analyzing the Contribution of Top-Down Lexical and Bottom-Up Acoustic Cues in the Detection of Sentence Prominence</b> <i>Sofoklis Kakouros<sup>1</sup>, Joris Pelemans<sup>2</sup>, Lyan Verwimp<sup>2</sup>, Patrick Wambacq<sup>2</sup>, Okko Räsänen<sup>1</sup></i> <i><sup>1</sup>Aalto University, Finland; <sup>2</sup>Katholieke Universiteit Leuven, Belgium</i>
PAGE 1079 Fri-P-3-2-14 17:00	<b>A Longitudinal Study of Children's Intonation in Narrative Speech</b> <i>Jeffrey Kallay, Melissa A. Redford, University of Oregon, USA</i>

### Fri-P-3-3: Speech Production Analysis and Modeling

Chair: Gayeon Son

Room: Pacific Concourse - Poster C, Time 17:00-19:00, Friday, 9 Sept. 2016

---

PAGE 1084 Fri-P-3-3-1 17:00	<b>Velum Control for Oral Sounds</b> <i>Reed Blaylock, Louis Goldstein, Shrikanth S. Narayanan, University of Southern California, USA</i>
PAGE 1089 Fri-P-3-3-2 17:00	<b>F0 Development in Acquiring Korean Stop Distinction</b> <i>Gayeon Son, University of Pennsylvania, USA</i>
PAGE 1094 Fri-P-3-3-3 17:00	<b>Phonetic Reduction Can Lead to Lengthening, and Enhancement Can Lead to Shortening</b> <i>Clara Cohen, Matt Carlson, Pennsylvania State University, USA</i>
PAGE 1099 Fri-P-3-3-4 17:00	<b>Mechanical Production of [b], [m] and [w] Using Controlled Labial and Velopharyngeal Gestures</b> <i>Takayuki Arai, Sophia University, Japan</i>
PAGE 1104 Fri-P-3-3-5 17:00	<b>An Improved 3D Geometric Tongue Model</b> <i>Qiang Fang<sup>1</sup>, Yun Chen<sup>2</sup>, Haibo Wang<sup>1</sup>, Jianguo Wei<sup>2</sup>, Jianrong Wang<sup>2</sup>, Xiyu Wu<sup>3</sup>, Aijun Li<sup>1</sup></i> <i><sup>1</sup>Chinese Academy of Social Sciences, China; <sup>2</sup>Tianjin University, China; <sup>3</sup>Peking University, China</i>
PAGE 1108 Fri-P-3-3-6 17:00	<b>Congruency Effect Between Articulation and Grasping in Native English Speakers</b> <i>Mikko Tiainen<sup>1</sup>, Fatima M. Felisberti<sup>2</sup>, Kaisa Tiippana<sup>1</sup>, Martti Vainio<sup>1</sup>, Juraj Simko<sup>1</sup>, Jiri Lukavsky<sup>3</sup>, Lari Vainio<sup>1</sup></i> <i><sup>1</sup>University of Helsinki, Finland; <sup>2</sup>Kingston University London, UK; <sup>3</sup>Czech Academy of Sciences, Czech Republic</i>

PAGE 1113 Fri-P-3-3-7 17:00	<b>Emergence of Vocal Developmental Sequences in a Predictive Coding Model of Speech Acquisition</b> <i>Shamima Najnin, Bonny Banerjee, University of Memphis, USA</i>
PAGE 1118 Fri-P-3-3-8 17:00	<b>Categorization of Natural Spanish Whistled Vowels by Naïve Spanish Listeners</b> <i>Julien Meyer<sup>1</sup>, Laure Dentel<sup>2</sup>, Fanny Meunier<sup>3</sup></i> <i><sup>1</sup>GIPSA, France; <sup>2</sup>World Whistles Research Association, France; <sup>3</sup>L2C2, France</i>
PAGE 1122 Fri-P-3-3-9 17:00	<b>Between- and Within-Speaker Effects of Bilingualism on F0 Variation</b> <i>Rob Voigt, Dan Jurafsky, Meghan Sumner, Stanford University, USA</i>
PAGE 1127 Fri-P-3-3-10 17:00	<b>Vowel Characteristics in the Assessment of L2 English Pronunciation</b> <i>Calbert Graham, Paula Buttery, Francis Nolan, University of Cambridge, UK</i>
PAGE 1132 Fri-P-3-3-11 17:00	<b>Kulning (Swedish Cattle Calls): Acoustic, EGG, Stroboscopic and High-Speed Video Analyses of an Unusual Singing Style</b> <i>Ahmed Geneid<sup>1</sup>, Anne-Maria Laukkanen<sup>2</sup>, Anita McAllister<sup>3</sup>, Robert Eklund<sup>4</sup></i> <i><sup>1</sup>Helsinki University Hospital, Finland; <sup>2</sup>University of Tampere, Finland; <sup>3</sup>Karolinska Institute, Sweden; <sup>4</sup>Linköping University, Sweden</i>
PAGE 1136 Fri-P-3-3-12 17:00	<b>Glottal Squeaks in VC Sequences</b> <i>Miša Hejná<sup>1</sup>, Pertti Palo<sup>2</sup>, Scott Moisik<sup>3</sup></i> <i><sup>1</sup>Newcastle University, UK; <sup>2</sup>Queen Margaret University, UK; <sup>3</sup>NTU, Singapore</i>
PAGE 1141 Fri-P-3-3-13 17:00	<b>Automatic Pronunciation Generation by Utilizing a Semi-Supervised Deep Neural Networks</b> <i>Naoya Takahashi<sup>1</sup>, Tofiqh Naghibi<sup>2</sup>, Beat Pfister<sup>2</sup></i> <i><sup>1</sup>Sony, Japan; <sup>2</sup>ETH Zürich, Switzerland</i>

## Fri-P-3-4: Spoken Dialogue Systems

Chair: Ruhi Sarikaya

Room: Pacific Concourse - Poster D, Time 17:00-19:00, Friday, 9 Sept. 2016

---

PAGE 1146 Fri-P-3-4-1 17:00	<b>Personalized Natural Language Understanding</b> <i>Xiaohu Liu, Ruhi Sarikaya, Liang Zhao, Yong Ni, Yi-Cheng Pan, Microsoft, USA</i>
PAGE 1151 Fri-P-3-4-2 17:00	<b>A Sequence-to-Sequence Model for User Simulation in Spoken Dialogue Systems</b> <i>Layla El Asri, Jing He, Kaheer Suleman, Maluuba Research, Canada</i>
PAGE 1156 Fri-P-3-4-3 17:00	<b>Root Cause Analysis of Miscommunication Hotspots in Spoken Dialogue Systems</b> <i>Spiros Georgiladakis<sup>1</sup>, Georgia Athanasopoulou<sup>1</sup>, Raveesh Meena<sup>2</sup>, José Lopes<sup>2</sup>, Arodami Chorianopoulou<sup>3</sup>, Elisavet Palogiannidi<sup>3</sup>, Elias Iosif<sup>1</sup>, Gabriel Skantze<sup>2</sup>, Alexandros Potamianos<sup>1</sup></i> <i><sup>1</sup>NTUA, Greece; <sup>2</sup>KTH, Sweden; <sup>3</sup>Technical University of Crete, Greece</i>
PAGE 1161 Fri-P-3-4-4 17:00	<b>Making Personal Digital Assistants Aware of What They Do Not Know</b> <i>Omar Zia Khan, Ruhi Sarikaya, Microsoft, USA</i>
PAGE 1166 Fri-P-3-4-5 17:00	<b>Implementing Acoustic-Prosodic Entrainment in a Conversational Avatar</b> <i>Rivka Levitan<sup>1</sup>, Štefan Beňuš<sup>2</sup>, Ramiro H. Gálvez<sup>3</sup>, Agustín Gravano<sup>3</sup>, Florencia Savoretti<sup>3</sup>, Marian Trnka<sup>4</sup>, Andreas Weise<sup>1</sup>, Julia Hirschberg<sup>5</sup></i> <i><sup>1</sup>CUNY Brooklyn College, USA; <sup>2</sup>UKF, Slovak Republic; <sup>3</sup>Universidad de Buenos Aires, Argentina; <sup>4</sup>Slovak Academy of Sciences, Slovak Republic; <sup>5</sup>Columbia University, USA</i>
PAGE 1171 Fri-P-3-4-6 17:00	<b>Perceived Usability and Cognitive Demand of Secondary Tasks in Spoken Versus Visual-Manual Automotive Interaction</b> <i>Annika Silvervarg<sup>1</sup>, Sofia Lindvall<sup>1</sup>, Jonatan Andersson<sup>1</sup>, Ida Esberg<sup>2</sup>, Christian Jernberg<sup>2</sup>, Filip Frumerie<sup>2</sup>, Arne Jönsson<sup>1</sup></i> <i><sup>1</sup>Linköping University, Sweden; <sup>2</sup>Volvo, Sweden</i>

INTERSPEECH 2016 Table of Contents

---

**Fri-S&T-3 : Show & Tell Session 3**

*Chairs: Shiva Sundaram, Nicolas Scheffer*

*Room: Market Street Foyer, Time 17:00–19:00, Friday, 9 Sept. 2016*

---

PAGE 1176  
Fri-S&T-3-1  
17:00

**Zara: An Empathetic Interactive Virtual Agent**

*Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Wan Yan, Ricky Ho Yin Chan, HKUST, China*

PAGE 1178  
Fri-S&T-3-2  
17:00

**Measuring Pronunciation Improvement in Users of CAPT Tool TipTopTalk!**

*Cristian Tejedor-García, David Escudero-Mancebo, Enrique Cámara-Arenas, César González-Ferreras, Valentín Cardeñoso-Payo, Universidad de Valladolid, Spain*

PAGE 1180  
Fri-S&T-3-3  
17:00

**SparkNG: Interactive MATLAB Tools for Introduction to Speech Production, Perception and Processing Fundamentals and Application of the Aliasing-Free L-F Model Component**

*Hideki Kawahara, Wakayama University, Japan*

PAGE 1182  
Fri-S&T-3-4  
17:00

**Real-Time Tracking of Speakers' Emotions, States, and Traits on Mobile Platforms**

*Erik Marchi, Florian Eyben, Gerhard Hagerer, Björn Schuller, audeERING, Germany*

INTERSPEECH 2016 Table of Contents

---

**Sat-SE-1 : Special Event: Mindfulness**

*Chair: David Suendermann-Oeft*

*Room: Grand Ballroom ABC, Time 08:00–08:25, Saturday, 10 Sept. 2016*

---

N/A  
Sat-SE-1  
08:00

**Mindfulness Special Event**

*Nikki Mirghafori, ICSI, USA*

**Keynote 2: Edward Chang**

*Chairs: Andreas Stolcke*

*Room: Grand Ballroom ABC, Time 08:30–09:30, Saturday, 10 Sept. 2016*

---

PAGE 1184  
Sat-Keynote-2  
08:30

**The Human Speech Cortex**

*Edward Chang, University of California at San Francisco, USA*

**Sat-SE-2 : Special Event: Speaker Comparison for Forensic and Investigative Applications II**

*Chair: Joe Campbell*

*Room: Grand Ballroom A, Time 10:00–12:00, Saturday, 10 Sept. 2016*

---

N/A  
Sat-SE-2  
10:00

**Speaker Comparison for Forensic and Investigative Applications II**

*Jean-François Bonastre<sup>1</sup>, Joseph P. Campbell<sup>2</sup>, Anders Eriksson<sup>3</sup>, Hiro Nakasone<sup>4</sup>, Reva Schwartz<sup>5</sup>*

*<sup>1</sup>LIA, France; <sup>2</sup>MIT Lincoln Laboratory, USA; <sup>3</sup>Stockholm University, Sweden; <sup>4</sup>FBI, USA; <sup>5</sup>NIST, USA*

## Sat-O-4-2 : Special Session: Clinical and Neuroscience-Inspired Vocal Biomarkers of Neurological and Psychiatric Disorders

*Chairs: Nicholas Cummins, Julien Epps, Emily Mower Provost, Thomas Quatieri, Stefan Scherer*

*Room: Grand Ballroom BC, Time 10:00-12:00, Saturday, 10 Sept. 2016*

---

PAGE 1185  
Sat-O-4-2-1  
10:00

### **Acoustic-Prosodic and Turn-Taking Features in Interactions with Children with Neurodevelopmental Disorders**

*Daniel Bone<sup>1</sup>, Somer Bishop<sup>2</sup>, Rahul Gupta<sup>1</sup>, Sungbok Lee<sup>1</sup>, Shrikanth S. Narayanan<sup>1</sup>*  
*<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of California at San Francisco, USA*

PAGE 1190  
Sat-O-4-2-2  
10:15

### **Automatic Detection of Parkinson's Disease Based on Modulated Vowels**

*Daria Hemmerling<sup>1</sup>, Juan Rafael Orozco-Arroyave<sup>2</sup>, Andrzej Skalski<sup>1</sup>, Janusz Gajda<sup>1</sup>, Elmar Nöth<sup>3</sup>*  
*<sup>1</sup>AGH UST, Poland; <sup>2</sup>Universidad de Antioquia, Colombia; <sup>3</sup>FAU Erlangen-Nürnberg, Germany*

PAGE 1195  
Sat-O-4-2-3  
10:30

### **Towards Automatic Detection of Amyotrophic Lateral Sclerosis from Speech Acoustic and Articulatory Samples**

*Jun Wang<sup>1</sup>, Prasanna V. Kothalkar<sup>1</sup>, Beiming Cao<sup>1</sup>, Daragh Heitzman<sup>2</sup>*  
*<sup>1</sup>University of Texas at Dallas, USA; <sup>2</sup>Texas Neurology, USA*

PAGE 1200  
Sat-O-4-2-4  
10:45

### **Neurophysiological Vocal Source Modeling for Biomarkers of Disease**

*Gregory Ciccarelli<sup>1</sup>, Thomas F. Quatieri<sup>1</sup>, Satrajit S. Ghosh<sup>2</sup>*  
*<sup>1</sup>MIT Lincoln Laboratory, USA; <sup>2</sup>MIT, USA*

PAGE 1205  
Sat-O-4-2-5  
11:00

### **Relation of Automatically Extracted Formant Trajectories with Intelligibility Loss and Speaking Rate Decline in Amyotrophic Lateral Sclerosis**

*Rachelle L. Horwitz-Martin<sup>1</sup>, Thomas F. Quatieri<sup>1</sup>, Adam C. Lammert<sup>1</sup>, James R. Williamson<sup>1</sup>, Yana Yunusova<sup>2</sup>, Elizabeth Godoy<sup>1</sup>, Daryush D. Mehta<sup>1</sup>, Jordan R. Green<sup>3</sup>*  
*<sup>1</sup>MIT Lincoln Laboratory, USA; <sup>2</sup>University of Toronto, Canada; <sup>3</sup>Harvard-MIT SHBT, USA*

PAGE 1210  
Sat-O-4-2-6  
11:15

### **Automatic Analysis of Typical and Atypical Encoding of Spontaneous Emotion in the Voice of Children**

*Fabien Ringeval<sup>1</sup>, Erik Marchi<sup>1</sup>, Charline Grossard<sup>2</sup>, Jean Xavier<sup>2</sup>, Mohamed Chetouani<sup>2</sup>, David Cohen<sup>2</sup>, Björn Schuller<sup>1</sup>*  
*<sup>1</sup>audEERING, Germany; <sup>2</sup>UPMC, France*

PAGE 1215  
Sat-O-4-2-7  
11:30

### **Recognition of Depression in Bipolar Disorder: Leveraging Cohort and Person-Specific Knowledge**

*Soheil Khorram, John Gideon, Melvin McInnis, Emily Mower Provost, University of Michigan, USA*

PAGE 1220  
Sat-O-4-2-8  
11:45

### **Diagnosing People with Dementia Using Automatic Conversation Analysis**

*Bahman Mirheidari<sup>1</sup>, Daniel Blackburn<sup>1</sup>, Markus Reuber<sup>2</sup>, Traci Walker<sup>1</sup>, Heidi Christensen<sup>1</sup>*  
*<sup>1</sup>University of Sheffield, UK; <sup>2</sup>Royal Hallamshire Hospital, UK*

## Sat-O-4-3: Special Session: Singing Synthesis Challenge: Fill-In the Gap

Chairs: *Christophe d'Alessandro, Axel Roebel, Olivier Deroo*

Room: *Bayview A, Time 10:00-12:00, Saturday, 10 Sept. 2016*

---

- PAGE 1225  
Sat-O-4-3-1  
10:00      **SERAPHIM: A Wavetable Synthesis System with 3D Lip Animation for Real-Time Speech and Singing Applications on Mobile Platforms**  
*Paul Yaozhu Chan<sup>1</sup>, Minghui Dong<sup>1</sup>, Grace Xue Hui Ho<sup>2</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>NTU, Singapore*
- PAGE 1230  
Sat-O-4-3-2  
10:15      **Expressive Singing Synthesis Based on Unit Selection for the Singing Synthesis Challenge 2016**  
*Jordi Bonada, Martí Umbert, Merlijn Blaauw, Universitat Pompeu Fabra, Spain*
- PAGE 1235  
Sat-O-4-3-3  
10:30      **Vocal Effort Modification for Singing Synthesis**  
*Olivier Perrotin, Christophe d'Alessandro, LIMSI, France*
- PAGE 1240  
Sat-O-4-3-4  
10:45      **Bertsokantari: a TTS Based Singing Synthesis System**  
*Eder del Blanco, Inma Hernaez, Eva Navas, Xabier Sarasola, D. Erro, Universidad del País Vasco, Spain*
- PAGE 1245  
Sat-O-4-3-5  
11:00      **Evaluation of Singing Synthesis: Methodology and Case Study with Concatenative and Performative Systems**  
*Lionel Feugère<sup>1</sup>, Christophe d'Alessandro<sup>1</sup>, Samuel Delalez<sup>1</sup>, Luc Ardaillon<sup>2</sup>, Axel Roebel<sup>2</sup>*  
*<sup>1</sup>LIMSI, France; <sup>2</sup>IRCAM, France*
- PAGE 1250  
Sat-O-4-3-6  
11:15      **Expressive Control of Singing Voice Synthesis Using Musical Contexts and a Parametric F0 Model**  
*Luc Ardaillon, Celine Chabot-Canet, Axel Roebel, IRCAM, France*
- PAGE 1255  
Sat-O-4-3-7  
11:30      **Optimal Unit Stitching in a Unit Selection Singing Synthesis System**  
*Marius Cotescu, Acapela Group, Belgium*

## Sat-O-4-4: Conversation and Interaction

Chairs: *Margaret Zellers, Julia Hirschberg*

Room: *Bayview B, Time 10:00-12:00, Saturday, 10 Sept. 2016*

---

- PAGE 1260  
Sat-O-4-4-1  
10:00      **The Perception of Overlapping Speech: Effects of Speaker Prosody and Listener Attitudes**  
*Katherine Hilton, Stanford University, USA*
- PAGE 1265  
Sat-O-4-4-2  
10:20      **Who Do You Think Will Speak Next? Perception of Turn-Taking Cues in Slovak and Argentine Spanish**  
*Agustín Gravano<sup>1</sup>, Pablo Brusco<sup>1</sup>, Štefan Beňuš<sup>2</sup>*  
*<sup>1</sup>Universidad de Buenos Aires, Argentina; <sup>2</sup>UKF, Slovak Republic*
- PAGE 1270  
Sat-O-4-4-3  
10:40      **Disentrainment may be a Positive Thing: A Novel Measure of Unsigned Acoustic-Prosodic Synchrony, and its Relation to Speaker Engagement**  
*Juan M. Pérez, Ramiro H. Gálvez, Agustín Gravano, Universidad de Buenos Aires, Argentina*
- PAGE 1275  
Sat-O-4-4-4  
11:00      **Respiratory Turn-Taking Cues**  
*Marcin Włodarczak, Mattias Heldner, Stockholm University, Sweden*
- PAGE 1280  
Sat-O-4-4-5  
11:20      **The Discourse Marker “so” in Turn-Taking and Turn-Releasing Behavior**  
*Emma Rennie<sup>1</sup>, Rebecca Lunsford<sup>2</sup>, Peter A. Heeman<sup>2</sup>*  
*<sup>1</sup>Reed College, USA; <sup>2</sup>Oregon Health & Science University, USA*
- PAGE 1285  
Sat-O-4-4-6  
11:40      **Acoustic Properties of Formality in Conversational Japanese**  
*Ethan Sherr-Ziarko, University of Oxford, UK*

## Sat-O-4-5: Automatic Learning of Representations

Chair: Carolina Parada

Room: Seacliff BCD, Time 10:00-12:00, Saturday, 10 Sept. 2016

---

- PAGE 1290  
Sat-O-4-5-1  
10:00      **Inferring Phonemic Classes from CNN Activation Maps Using Clustering Techniques**  
*Thomas Pellegrini, Sandrine Mouysset, IRIT, France*
- PAGE 1295  
Sat-O-4-5-2  
10:20      **Joint Learning of Speaker and Phonetic Similarities with Siamese Networks**  
*Neil Zeghidour<sup>1</sup>, Gabriel Synnaeve<sup>1</sup>, Nicolas Usunier<sup>1</sup>, Emmanuel Dupoux<sup>2</sup>*  
*<sup>1</sup>Facebook, France; <sup>2</sup>ENS, France*
- PAGE 1300  
Sat-O-4-5-3  
10:40      **Unsupervised Learning of Acoustic Units Using Autoencoders and Kohonen Nets**  
*Vikramjit Mitra, Dimitra Vergyri, Horacio Franco, SRI International, USA*
- PAGE 1305  
Sat-O-4-5-4  
11:00      **Learning Multiscale Features Directly from Waveforms**  
*Zhenyao Zhu, Jesse H. Engel, Awni Hannun, Baidu Research, USA*
- PAGE 1310  
Sat-O-4-5-5  
11:20      **Supervised Learning of Acoustic Models in a Zero Resource Setting to Improve DPGMM Clustering**  
*Michael Heck, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan*
- PAGE 1315  
Sat-O-4-5-6  
11:40      **Semi-Supervised and Cross-Lingual Knowledge Transfer Learnings for DNN Hybrid Acoustic Models Under Low-Resource Conditions**  
*Haihua Xu<sup>1</sup>, Hang Su<sup>2</sup>, Chongjia Ni<sup>3</sup>, Xiong Xiao<sup>1</sup>, Hao Huang<sup>4</sup>, Eng Siong Chng<sup>1</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>TL@NTU, Singapore; <sup>2</sup>University of California at Berkeley, USA; <sup>3</sup>A\*STAR, Singapore; <sup>4</sup>Xinjiang University, China*

## Sat-O-4-6: Language Modeling for Conversational Speech and Confidence Measures

Chairs: Renato de Mori, Dimitra Vergyri

Room: Seacliff A, Time 10:00-12:00, Saturday, 10 Sept. 2016

---

- PAGE 1320  
Sat-O-4-6-1  
10:00      **Recurrent Out-of-Vocabulary Word Detection Using Distribution of Features**  
*Taichi Asami<sup>1</sup>, Ryo Masumura<sup>1</sup>, Yushi Aono<sup>1</sup>, Koichi Shinoda<sup>2</sup>*  
*<sup>1</sup>NTT, Japan; <sup>2</sup>Tokyo Institute of Technology, Japan*
- PAGE 1325  
Sat-O-4-6-2  
10:20      **Investigation of Semi-Supervised Acoustic Model Training Based on the Committee of Heterogeneous Neural Networks**  
*Naoyuki Kanda, Shoji Harada, Xugang Lu, Hisashi Kawai, NICT, Japan*
- PAGE 1330  
Sat-O-4-6-3  
10:40      **Acoustic Word Embeddings for ASR Error Detection**  
*Sahar Ghannay, Yannick Estève, Nathalie Camelin, Paul deléglise, LIUM, France*
- PAGE 1335  
Sat-O-4-6-4  
11:00      **Combining Semantic Word Classes and Sub-Word Unit Speech Recognition for Robust OOV Detection**  
*Axel Horndasch, Anton Batliner, Caroline Kaufhold, Elmar Nöth, FAU Erlangen-Nürnberg, Germany*
- PAGE 1340  
Sat-O-4-6-5  
11:20      **Web Data Selection Based on Word Embedding for Low-Resource Speech Recognition**  
*Chuangdong Xie<sup>1</sup>, Wu Guo<sup>1</sup>, Guoping Hu<sup>2</sup>, Junhua Liu<sup>3</sup>*  
*<sup>1</sup>USTC, China; <sup>2</sup>Ministry of Public Security, China; <sup>3</sup>iFLYTEK, China*
- PAGE 1345  
Sat-O-4-6-6  
11:40      **Colloquialising Modern Standard Arabic Text for Improved Speech Recognition**  
*Sarah Al-Shareef, Thomas Hain, University of Sheffield, UK*

## Sat-P-4-1 : Topics in Speech Perception

Chair: Bernd Meyer

Room: Pacific Concourse - Poster A, Time 10:00-12:00, Saturday, 10 Sept. 2016

---

PAGE 1350  
Sat-P-4-1-1  
10:00

### **Pitch-Range Perception: The Dynamic Interaction Between Voice Quality and Fundamental Frequency**

*Jianjing Kuang, Mark Liberman, University of Pennsylvania, USA*

PAGE 1355  
Sat-P-4-1-2  
10:00

### **Comparing the Contributions of Amplitude and Phase to Speech Intelligibility in a Vocoder-Based Speech Synthesis Model**

*Fei Chen<sup>1</sup>, Benson C.L. Chiao<sup>2</sup>*

*<sup>1</sup>SUSTC, China; <sup>2</sup>University of Hong Kong, China*

PAGE 1359  
Sat-P-4-1-3  
10:00

### **Modeling Noise Influence to Speech Intelligibility Non-Intrusively by Reduced Speech Dynamic Range**

*Fei Chen, SUSTC, China*

PAGE 1363  
Sat-P-4-1-4  
10:00

### **Do GMM Phoneme Classifiers Perceive Synthetic Sibilants as Humans Do?**

*Gábor Pintér, Hiroki Watanabe, Kobe University, Japan*

PAGE 1368  
Sat-P-4-1-5  
10:00

### **Neural Responses to Speech-Specific Modulations Derived from a Spectro-Temporal Filter Bank**

*Marina Frye<sup>1</sup>, Cristiano Micheli<sup>1</sup>, Inga M. Schepers<sup>1</sup>, Gerwin Schalk<sup>2</sup>,*

*Jochem W. Rieger<sup>1</sup>, Bernd T. Meyer<sup>3</sup>*

*<sup>1</sup>Carl von Ossietzky Universität Oldenburg, Germany; <sup>2</sup>NCAN, USA; <sup>3</sup>Johns Hopkins University, USA*

PAGE 1373  
Sat-P-4-1-6  
10:00

### **Comparing Different Methods for Analyzing ERP Signals**

*Kimberley Mulder, Louis ten Bosch, Lou Boves, Radboud Universiteit Nijmegen, The Netherlands*

PAGE 1378  
Sat-P-4-1-7  
10:00

### **Supplementary Motor Area Activation in Disfluency Perception: An fMRI Study of Listener Neural Responses to Spontaneously Produced Unfilled and Filled Pauses**

*Robert Eklund<sup>1</sup>, Martin Ingvar<sup>2</sup>*

*<sup>1</sup>Linköping University, Sweden; <sup>2</sup>Karolinska Institute, Sweden*

PAGE 1382  
Sat-P-4-1-8  
10:00

### **Vowel Fundamental and Formant Frequency Contributions to English and Mandarin Sentence Intelligibility**

*Daniel Fogerty<sup>1</sup>, Fei Chen<sup>2</sup>*

*<sup>1</sup>University of South Carolina, USA; <sup>2</sup>SUSTC, China*



## Sat-P-4-2 : Behavioral Signal Processing and Speaker State and Traits Analytics

Chair: *Chung-Hsien Wu*

Room: *Pacific Concourse - Poster B, Time 10:00-12:00, Saturday, 10 Sept. 2016*

---

- PAGE 1387  
Sat-P-4-2-1  
10:00      **Attention Assisted Discovery of Sub-Utterance Structure in Speech Emotion Recognition**  
*Che-Wei Huang, Shrikanth S. Narayanan, University of Southern California, USA*
- PAGE 1392  
Sat-P-4-2-2  
10:00      **Combining CNN and BLSTM to Extract Textual and Acoustic Features for Recognizing Stances in Mandarin Ideological Debate Competition**  
*Linchuan Li, Zhiyong Wu, Mingxing Xu, Helen Meng, Lianhong Cai, Tsinghua University, China*
- PAGE 1397  
Sat-P-4-2-3  
10:00      **Inter-Speech Clicks in an Interspeech Keynote**  
*Jürgen Trouvain, Zofia Malisz, Universität des Saarlandes, Germany*
- PAGE 1402  
Sat-P-4-2-4  
10:00      **Speaker Age Classification and Regression Using i-Vectors**  
*Joanna Grzybowska, Stanisław Kacprzak, AGH UST, Poland*
- PAGE 1407  
Sat-P-4-2-5  
10:00      **Sparsely Connected and Disjointly Trained Deep Neural Networks for Low Resource Behavioral Annotation: Acoustic Classification in Couples' Therapy**  
*Haoqi Li<sup>1</sup>, Brian Baucom<sup>2</sup>, Panayiotis Georgiou<sup>1</sup>*  
*<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Utah, USA*
- PAGE 1412  
Sat-P-4-2-6  
10:00      **Automatically Classifying Self-Rated Personality Scores from Speech**  
*Guozhen An<sup>1</sup>, Sarah Ita Levitan<sup>2</sup>, Rivka Levitan<sup>3</sup>, Andrew Rosenberg<sup>4</sup>, Michelle Levine<sup>2</sup>, Julia Hirschberg<sup>2</sup>*  
*<sup>1</sup>CUNY Graduate Center, USA; <sup>2</sup>Columbia University, USA; <sup>3</sup>CUNY Brooklyn College, USA; <sup>4</sup>CUNY Queens College, USA*
- PAGE 1417  
Sat-P-4-2-7  
10:00      **Estimation of Children's Physical Characteristics from Their Voices**  
*Jill Fain Lehman, Rita Singh, Disney Research, USA*

- PAGE 1422  
Sat-P-4-2-8  
10:00      **Talking to a System and Talking to a Human: A Study from a Speech-to-Speech, Machine Translation Mediated Map Task**  
*Hayakawa Akira<sup>1</sup>, Saturnino Luz<sup>2</sup>, Nick Campbell<sup>1</sup>*  
*<sup>1</sup>Trinity College Dublin, Ireland; <sup>2</sup>University of Edinburgh, UK*
- PAGE 1427  
Sat-P-4-2-9  
10:00      **Predicting Affective Dimensions Based on Self Assessed Depression Severity**  
*Rahul Gupta, Shrikanth S. Narayanan, University of Southern California, USA*
- PAGE 1432  
Sat-P-4-2-10  
10:00      **Enhancement of Automatic Oral Presentation Assessment System Using Latent N-Grams Word Representation and Part-of-Speech Information**  
*Wen-Yu Huang<sup>1</sup>, Shan-Wen Hsiao<sup>1</sup>, Hung-Ching Sun<sup>1</sup>, Ming-Chuan Hsieh<sup>2</sup>, Ming-Hsueh Tsai<sup>2</sup>, Chi-Chun Lee<sup>1</sup>*  
*<sup>1</sup>National Tsing Hua University, Taiwan; <sup>2</sup>NAER, Taiwan*
- PAGE 1437  
Sat-P-4-2-11  
10:00      **Use of Vowels in Discriminating Speech-Laugh from Laughter and Neutral Speech**  
*Sri Harsha Dumpala, P. Gangamohan, Suryakanth V. Gangashetty, B. Yegnanarayana, IIIT Hyderabad, India*
- PAGE 1442  
Sat-P-4-2-12  
10:00      **A Convex Model for Linguistic Influence in Group Conversations**  
*Kan Kawabata, Visar Berisha, Anna Scaglione, Amy LaCross, Arizona State University, USA*
- PAGE 1447  
Sat-P-4-2-13  
10:00      **A Deep Learning Approach to Modeling Empathy in Addiction Counseling**  
*James Gibson<sup>1</sup>, Doğan Can<sup>1</sup>, Bo Xiao<sup>1</sup>, Zac E. Imel<sup>2</sup>, David C. Atkins<sup>3</sup>, Panayiotis Georgiou<sup>1</sup>, Shrikanth S. Narayanan<sup>1</sup>*  
*<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Utah, USA; <sup>3</sup>University of Washington, USA*
- PAGE 1452  
Sat-P-4-2-14  
10:00      **Unipolar Depression vs. Bipolar Disorder: An Elicitation-Based Approach to Short-Term Detection of Mood Disorder**  
*Kun-Yi Huang<sup>1</sup>, Chung-Hsien Wu<sup>1</sup>, Yu-Ting Kuo<sup>1</sup>, Fong-Lin Jang<sup>2</sup>*  
*<sup>1</sup>National Cheng Kung University, Taiwan; <sup>2</sup>Chi Mei Medical Center, Taiwan*



## Sat-P-4-3 : Speech Synthesis Poster

Chair: Ingmar Steiner

Room: Pacific Concourse - Poster C, Time 10:00-12:00, Saturday, 10 Sept. 2016

---

- PAGE 1457  
Sat-P-4-3-1  
10:00  
**Conditional Random Fields for the Tunisian Dialect Grapheme-to-Phoneme Conversion**  
*Abir Masmoudi<sup>1</sup>, Mariem Ellouze<sup>2</sup>, Fethi Bougares<sup>1</sup>, Yannick Esève<sup>1</sup>, Lamia Belguith<sup>2</sup>*  
<sup>1</sup>LIUM, France; <sup>2</sup>University of Sfax, Tunisia
- PAGE 1462  
Sat-P-4-3-2  
10:00  
**Efficient Thai Grapheme-to-Phoneme Conversion Using CRF-Based Joint Sequence Modeling**  
*Sittipong Saychum, Sarawoot Kongyoung, Anocha Rugchatjaroen, Patcharika Chootrakool, Sawit Kasuriya, Chai Wutiwiwatchai, NECTEC, Thailand*
- PAGE 1467  
Sat-P-4-3-3  
10:00  
**An Articulatory-Based Singing Voice Synthesis Using Tongue and Lips Imaging**  
*Aurore Jaumard-Hakoun<sup>1</sup>, Kele Xu<sup>1</sup>, Clémence Leboullenger<sup>1</sup>, Pierre Roussel-Ragot<sup>2</sup>, Bruce Denby<sup>3</sup>*  
<sup>1</sup>UPMC, France; <sup>2</sup>Institut Langevin, France; <sup>3</sup>Tianjin University, China
- PAGE 1472  
Sat-P-4-3-4  
10:00  
**Phoneme Embedding and its Application to Speech Driven Talking Avatar Synthesis**  
*Xu Li<sup>1</sup>, Zhiyong Wu<sup>1</sup>, Helen Meng<sup>1</sup>, Jia Jia<sup>1</sup>, Xiaoyan Lou<sup>2</sup>, Lianhong Cai<sup>1</sup>*  
<sup>1</sup>Tsinghua University, China; <sup>2</sup>Beijing Samsung Telecom R&D Center, China
- PAGE 1477  
Sat-P-4-3-5  
10:00  
**Expressive Speech Driven Talking Avatar Synthesis with DBLSTM Using Limited Amount of Emotional Bimodal Data**  
*Xu Li<sup>1</sup>, Zhiyong Wu<sup>1</sup>, Helen Meng<sup>1</sup>, Jia Jia<sup>1</sup>, Xiaoyan Lou<sup>2</sup>, Lianhong Cai<sup>1</sup>*  
<sup>1</sup>Tsinghua University, China; <sup>2</sup>Beijing Samsung Telecom R&D Center, China
- PAGE 1482  
Sat-P-4-3-6  
10:00  
**Audio-to-Visual Speech Conversion Using Deep Neural Networks**  
*Sarah Taylor<sup>1</sup>, Akihiro Kato<sup>1</sup>, Iain Matthews<sup>2</sup>, Ben Milner<sup>1</sup>*  
<sup>1</sup>University of East Anglia, UK; <sup>2</sup>Disney Research, USA

- PAGE 1487  
Sat-P-4-3-7  
10:00  
**Generative Acoustic-Phonemic-Speaker Model Based on Three-Way Restricted Boltzmann Machine**  
*Toru Nakashika, Yasuhiro Minami, University of Electro-Communications, Japan*
- PAGE 1492  
Sat-P-4-3-8  
10:00  
**Articulatory Synthesis Based on Real-Time Magnetic Resonance Imaging Data**  
*Asterios Toutios, Tanner Sorensen, Krishna Somandepalli, Rachel Alexander, Shrikanth S. Narayanan, University of Southern California, USA*
- PAGE 1497  
Sat-P-4-3-9  
10:00  
**Deep Neural Network Based Acoustic-to-Articulatory Inversion Using Phone Sequence Information**  
*Xurong Xie, Xunying Liu, Lan Wang, Chinese Academy of Sciences, China*
- PAGE 1502  
Sat-P-4-3-10  
10:00  
**Articulatory-to-Acoustic Conversion with Cascaded Prediction of Spectral and Excitation Features Using Neural Networks**  
*Zheng-Chen Liu, Zhen-Hua Ling, Li-Rong Dai, USTC, China*
- PAGE 1507  
Sat-P-4-3-11  
10:00  
**Generating Gestural Scores from Acoustics Through a Sparse Anchor-Based Representation of Speech**  
*Christopher Liberatore, Ricardo Gutierrez-Osuna, Texas A&M University, USA*
- PAGE 1512  
Sat-P-4-3-12  
10:00  
**On the Suitability of Vocalic Sandwiches in a Corpus-Based TTS Engine**  
*David Guennec, Damien Lolive, IRISA, France*
- PAGE 1517  
Sat-P-4-3-13  
10:00  
**Unsupervised Stress Information Labeling Using Gaussian Process Latent Variable Model for Statistical Speech Synthesis**  
*Decha Moungsri, Tomoki Koriyama, Takao Kobayashi, Tokyo Institute of Technology, Japan*
- PAGE 1522  
Sat-P-4-3-14  
10:00  
**Using Zero-Frequency Resonator to Extract Multilingual Intonation Structure**  
*Jinfu Ni, Yoshinori Shiga, Hisashi Kawai, NICT, Japan*

## Sat-P-4-4: Resources and Annotation of Resources

Chair: Dilek Hakkani-Tur

Room: Pacific Concourse - Poster D, Time 10:00-12:00, Saturday, 10 Sept. 2016

---

PAGE 1527  
Sat-P-4-4-1  
10:00

### **A DNN-HMM Approach to Story Segmentation**

*Jia Yu<sup>1</sup>, Xiong Xiao<sup>2</sup>, Lei Xie<sup>1</sup>, Eng Siong Chng<sup>2</sup>, Haizhou Li<sup>2</sup>*

<sup>1</sup>Northwestern Polytechnical University, China; <sup>2</sup>TL@NTU, Singapore

PAGE 1532  
Sat-P-4-4-2  
10:00

### **The SIWIS Database: A Multilingual Speech Database with Acted Emphasis**

*Jean-Philippe Goldman<sup>1</sup>, Pierre-Edouard Honnet<sup>2</sup>, Rob Clark<sup>3</sup>, Philip N. Garner<sup>2</sup>, Maria Ivanova<sup>1</sup>, Alexandros Lazaridis<sup>2</sup>, Hui Liang<sup>4</sup>, Tiago Macedo<sup>1</sup>, Beat Pfister<sup>4</sup>, Manuel Sam Ribeiro<sup>3</sup>, Eric Wehrli<sup>1</sup>, Junichi Yamagishi<sup>3</sup>*

<sup>1</sup>Université de Genève, Switzerland; <sup>2</sup>Idiap Research Institute, Switzerland; <sup>3</sup>University of Edinburgh, UK; <sup>4</sup>ETH Zürich, Switzerland

PAGE 1536  
Sat-P-4-4-3  
10:00

### **Open Source Speech and Language Resources for Frisian**

*Emre Yilmaz<sup>1</sup>, Henk van den Heuvel<sup>1</sup>, Jelske Dijkstra<sup>2</sup>, Hans Van de Velde<sup>2</sup>, Frederik Kampstra<sup>3</sup>, Jouke Algra<sup>3</sup>, David Van Leeuwen<sup>1</sup>*

<sup>1</sup>Radboud Universiteit Nijmegen, The Netherlands; <sup>2</sup>Fryske Akademy, The Netherlands; <sup>3</sup>Omrop Fryslân, The Netherlands

PAGE 1541  
Sat-P-4-4-4  
10:00

### **The SRI CLEO Speaker-State Corpus**

*Andreas Kathol, Elizabeth Shriberg, Massimiliano de Zambotti, SRI International, USA*

PAGE 1545  
Sat-P-4-4-5  
10:00

### **SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese**

*Nancy F. Chen<sup>1</sup>, Rong Tong<sup>1</sup>, Darren Wee<sup>2</sup>, Peixuan Lee<sup>2</sup>, Bin Ma<sup>1</sup>, Haizhou Li<sup>1</sup>*

<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>NUS, Singapore

PAGE 1550  
Sat-P-4-4-6  
10:00

### **The SRI Speech-Based Collaborative Learning Corpus**

*Colleen Richey, Cynthia D'Angelo, Nonye Alozie, Harry Bratt, Elizabeth Shriberg, SRI International, USA*

PAGE 1555  
Sat-P-4-4-7  
10:00

### **An Expectation Maximization Approach to Joint Modeling of Multidimensional Ratings Derived from Multiple Annotators**

*Anil Ramakrishna<sup>1</sup>, Rahul Gupta<sup>1</sup>, Ruth B. Grossman<sup>2</sup>, Shrikanth S. Narayanan<sup>1</sup>*

<sup>1</sup>University of Southern California, USA; <sup>2</sup>Emerson College, USA

PAGE 1560  
Sat-P-4-4-8  
10:00

### **Voting Detector: A Combination of Anomaly Detectors to Reveal Annotation Errors in TTS Corpora**

*Jindřich Matoušek, Daniel Tihelka, University of West Bohemia, Czech Republic*

## Sat-S&T-4: Show & Tell Session 4

Chairs: Nicolas Scheffer, Shiva Sundaram

Room: Market Street Foyer, Time 10:00–12:00, Saturday, 10 Sept. 2016

---

- PAGE 1565  
Sat-S&T-4-1  
10:00
- The Magic Stone: A Video Game to Improve Communication Skills of People with Intellectual Disabilities**  
*Mario Corrales-Astorgano<sup>1</sup>, David Escudero-Mancebo<sup>1</sup>, César González-Ferreras<sup>1</sup>, Yurena Gutiérrez-González<sup>2</sup>, Valle Flores-Lucas<sup>1</sup>, Valentín Cardeñoso-Payo<sup>1</sup>, Lourdes Aguilar-Cuevas<sup>2</sup>*  
<sup>1</sup>Universidad de Valladolid, Spain; <sup>2</sup>Universidad Autónoma de Barcelona, Spain
- PAGE 1567  
Sat-S&T-4-2  
10:00
- Identifying Perceptually Similar Voices with a Speaker Recognition System Using Auto-Phonetic Features**  
*Finnian Kelly<sup>1</sup>, Anil Alexander<sup>1</sup>, Oscar Forth<sup>1</sup>, Samuel Kent<sup>1</sup>, Jonas Lindh<sup>2</sup>, Joel Åkesson<sup>2</sup>*  
<sup>1</sup>Oxford Wave Research, UK; <sup>2</sup>Voxalys, Sweden
- PAGE 1569  
Sat-S&T-4-3  
10:00
- A Real-Time Framework for Visual Feedback of Articulatory Data Using Statistical Shape Models**  
*Kristy James<sup>1</sup>, Alexander Hewer<sup>1</sup>, Ingmar Steiner<sup>1</sup>, Stefanie Wuhler<sup>2</sup>*  
<sup>1</sup>Universität des Saarlandes, Germany; <sup>2</sup>Inria, France
- PAGE 1571  
Sat-S&T-4-4  
10:00
- Flexible, Rapid Authoring of Goal-Orientated, Multi-Turn Dialogues Using the Task Completion Platform**  
*Alex Marin, Paul Crook, Omar Zia Khan, Vasily Radostev, Khushboo Aggarwal, Ruhi Sarikaya, Microsoft, USA*

## Sat-O-5-1: Acoustic Model Adaptation

Chairs: Brian Mak, Kai Yu

Room: Grand Ballroom A, Time 13:30–15:30, Saturday, 10 Sept. 2016

---

- PAGE 1573  
Sat-O-5-1-1  
13:30
- Context Adaptive Neural Network for Rapid Adaptation of Deep CNN Based Acoustic Models**  
*Marc Delcroix, Keisuke Kinoshita, Atsunori Ogawa, Takuya Yoshioka, Dung T. Tran, Tomohiro Nakatani, NTT, Japan*
- PAGE 1578  
Sat-O-5-1-2  
13:50
- Transfer Learning with Bottleneck Feature Networks for Whispered Speech Recognition**  
*Boon Pang Lim<sup>1</sup>, Faith Wong<sup>2</sup>, Yuyao Li<sup>2</sup>, Jia Wei Bay<sup>2</sup>*  
<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>Nanyang Girls' High School, Singapore
- PAGE 1583  
Sat-O-5-1-3  
14:10
- Adaptation of Neural Networks Constrained by Prior Statistics of Node Co-Activations**  
*Tasha Nagamine, Zhuo Chen, Nima Mesgarani, Columbia University, USA*
- PAGE 1588  
Sat-O-5-1-4  
14:30
- Domain Adaptation of CNN Based Acoustic Models Under Limited Resource Settings**  
*Masayuki Suzuki<sup>1</sup>, Ryuki Tachibana<sup>1</sup>, Samuel Thomas<sup>2</sup>, Bhuvana Ramabhadran<sup>2</sup>, George Saon<sup>2</sup>*  
<sup>1</sup>IBM, Japan; <sup>2</sup>IBM, USA
- PAGE 1593  
Sat-O-5-1-5  
14:50
- Subspace LHUC for Fast Adaptation of Deep Neural Network Acoustic Models**  
*Lahiru Samarakoon, Khe Chai Sim, NUS, Singapore*
- PAGE 1598  
Sat-O-5-1-6  
15:10
- Improving Children's Speech Recognition Through Out-of-Domain Data Augmentation**  
*Joachim Fainberg<sup>1</sup>, Peter Bell<sup>1</sup>, Mike Lincoln<sup>2</sup>, Steve Renals<sup>1</sup>*  
<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>Quorate Technology, UK

## Sat-O-5-2 : Special Session: Sharing Research and Education Resources for Understanding Speech Processing

Chairs: Eric Fosler-Lussier, Rebecca Bates, Florian Metze

Room: Grand Ballroom BC, Time 13:30-15:30, Saturday, 10 Sept. 2016

---

PAGE 1603  
Sat-O-5-2-1  
13:30

### Virtual Machines and Containers as a Platform for Experimentation

*Florian Metze<sup>1</sup>, Eric Riebling<sup>1</sup>, Anne S. Warlaumont<sup>2</sup>, Erika Bergelson<sup>3</sup>*

<sup>1</sup>Carnegie Mellon University, USA; <sup>2</sup>University of California at Merced, USA;

<sup>3</sup>University of Rochester, USA

PAGE 1608  
Sat-O-5-2-2  
13:45

### CloudCAST — Remote Speech Technology for Speech Professionals

*Phil Green<sup>1</sup>, Ricard Marxer<sup>1</sup>, Stuart Cunningham<sup>1</sup>, Heidi Christensen<sup>1</sup>, Frank Rudzicz<sup>2</sup>, Maria Yancheva<sup>3</sup>, André Coy<sup>4</sup>, Massimiliano Malavasi<sup>5</sup>, Lorenzo Desideri<sup>5</sup>, Fabio Tamburini<sup>6</sup>*

<sup>1</sup>University of Sheffield, UK; <sup>2</sup>Toronto Rehabilitation Institute, Canada; <sup>3</sup>University of Toronto, Canada; <sup>4</sup>University of West Indies, Jamaica; <sup>5</sup>AIAS Onlus Bologna, Italy;

<sup>6</sup>Università di Bologna, Italy

PAGE 1613  
Sat-O-5-2-3  
14:00

### webASR 2 — Improved Cloud Based Speech Technology

*Thomas Hain, Jeremy Christian, Oscar Saz, Salil Deena, Madina Hasan,*

*Raymond W.M. Ng, Rosanna Milner, Mortaza Doulaty, Yulan Liu, University of*

*Sheffield, UK*

PAGE 1618  
Sat-O-5-2-4  
14:15

### Sharing Speech Synthesis Software for Research and Education Within Low-Tech and Low-Resource Communities

*Andrew R. Plummer, Mary E. Beckman, Ohio State University, USA*

PAGE 1623  
Sat-O-5-2-5  
14:30

### The Berkeley Phonetics Machine

*Ronald L. Sprouse, Keith Johnson, University of California at Berkeley, USA*

PAGE 1627  
Sat-O-5-2-6  
14:45

### Experiences with Shared Resources for Research and Education in Speech and Language Processing

*Rebecca Bates<sup>1</sup>, Eric Fosler-Lussier<sup>2</sup>, Florian Metze<sup>3</sup>, Martha Larson<sup>4</sup>, Gina-Anne Levow<sup>5</sup>, Emily Mower Provost<sup>6</sup>*

<sup>1</sup>Minnesota State University, USA; <sup>2</sup>Ohio State University, USA; <sup>3</sup>Carnegie Mellon University, USA; <sup>4</sup>Technische Universiteit Delft, The Netherlands; <sup>5</sup>University of Washington, USA; <sup>6</sup>University of Michigan, USA

N/A  
Sat-O-5-2-7  
15:00

### Panel and Audience Discussion: How do we Develop, Disseminate, and Sustain Shared Resources from User and Developer Perspectives?

## Sat-O-5-3: Special Session: Voice Conversion Challenge

Chairs: Tomoki Toda, Junichi Yamagishi, Fernando Villavicencio, Zhizheng Wu, Ling-Hui Chen, Daisuke Saito, Mirjam Wester  
Room: Bayview A, Time 13:30-15:30, Saturday, 10 Sept. 2016

---

PAGE 1632  
Sat-O-5-3-1  
13:30

### The Voice Conversion Challenge 2016

*Tomoki Toda*<sup>1</sup>, *Ling-Hui Chen*<sup>2</sup>, *Daisuke Saito*<sup>3</sup>, *Fernando Villavicencio*<sup>4</sup>,  
*Mirjam Wester*<sup>5</sup>, *Zhizheng Wu*<sup>5</sup>, *Junichi Yamagishi*<sup>4</sup>  
<sup>1</sup>Nagoya University, Japan; <sup>2</sup>USTC, China; <sup>3</sup>University of Tokyo, Japan; <sup>4</sup>NII, Japan;  
<sup>5</sup>University of Edinburgh, UK

PAGE 1637  
Sat-O-5-3-2  
13:45

### Analysis of the Voice Conversion Challenge 2016 Evaluation Results

*Mirjam Wester*, *Zhizheng Wu*, *Junichi Yamagishi*, University of Edinburgh, UK

PAGE 1642  
Sat-O-5-3-3  
14:00

### The USTC System for Voice Conversion Challenge 2016: Neural Network Based Approaches for Spectrum, Aperiodicity and $F_0$ Conversion

*Ling-Hui Chen*<sup>1</sup>, *Li-Juan Liu*<sup>2</sup>, *Zhen-Hua Ling*<sup>1</sup>, *Yuan Jiang*<sup>2</sup>, *Li-Rong Dai*<sup>1</sup>  
<sup>1</sup>USTC, China; <sup>2</sup>iFLYTEK, China

PAGE 1647  
Sat-O-5-3-4  
14:15

### A Voice Conversion Mapping Function Based on a Stacked Joint-Autoencoder

*Seyed Hamidreza Mohammadi*, *Alexander Kain*, Oregon Health & Science University, USA

PAGE 1652  
Sat-O-5-3-5  
14:30

### Locally Linear Embedding for Exemplar-Based Spectral Conversion

*Yi-Chiao Wu*, *Hsin-Te Hwang*, *Chin-Cheng Hsu*, *Yu Tsao*, *Hsin-Min Wang*, Academia Sinica, Taiwan

PAGE 1657  
Sat-O-5-3-6  
14:45

### Applying Spectral Normalisation and Efficient Envelope Estimation and Statistical Transformation for the Voice Conversion Challenge 2016

*Fernando Villavicencio*<sup>1</sup>, *Junichi Yamagishi*<sup>1</sup>, *Jordi Bonada*<sup>2</sup>, *Felipe Espic*<sup>3</sup>  
<sup>1</sup>NII, Japan; <sup>2</sup>Universitat Pompeu Fabra, Spain; <sup>3</sup>University of Edinburgh, UK

PAGE 1662  
Sat-O-5-3-7  
15:00

### ML Parameter Generation with a Reformulated MGE Training Criterion — Participation in the Voice Conversion Challenge 2016

*D. Erro*, *A. Alonso*, *L. Serrano*, *D. Tavarez*, *I. Odriozola*, *Xabier Sarasola*,  
*Eder del Blanco*, *J. Sanchez*, *I. Saratxaga*, *Eva Navas*, *Inma Hernaez*, Universidad del País Vasco, Spain

PAGE 1667  
Sat-O-5-3-8  
15:15

### The NU-NAIST Voice Conversion System for the Voice Conversion Challenge 2016

*Kazuhiro Kobayashi*<sup>1</sup>, *Shinnosuke Takamichi*<sup>1</sup>, *Satoshi Nakamura*<sup>1</sup>, *Tomoki Toda*<sup>2</sup>  
<sup>1</sup>NAIST, Japan; <sup>2</sup>Nagoya University, Japan

## Sat-O-5-4: Intelligibility and Masking

*Chairs: Fei (Felix) Chen, Tan Lee*

*Room: Bayview B, Time 13:30-15:30, Saturday, 10 Sept. 2016*

---

- PAGE 1672  
Sat-O-5-4-1  
13:30      **Release from Energetic Masking Caused by Repeated Patterns of Glimpsing Windows**  
*Maury Lander-Portnoy, University of Southern California, USA*
- PAGE 1677  
Sat-O-5-4-2  
13:50      **Glimpsing Predictions for Natural and Vocoded Sentence Intelligibility During Modulation Masking: Effect of the Glimpse Cutoff Criterion**  
*Bobby Gibbs II, Daniel Fogerty, University of South Carolina, USA*
- PAGE 1682  
Sat-O-5-4-3  
14:10      **Temporal Envelopes in Sine-Wave Speech Recognition**  
*Li Xu, Ohio University, USA*
- PAGE 1687  
Sat-O-5-4-4  
14:30      **Understanding Periodically Interrupted Mandarin Speech**  
*Jing Liu<sup>1</sup>, Rosanna H.N. Tong<sup>2</sup>, Fei Chen<sup>1</sup>*  
*<sup>1</sup>SUSTC, China; <sup>2</sup>University of Hong Kong, China*
- PAGE 1692  
Sat-O-5-4-5  
14:50      **Factors Affecting the Intelligibility of Sine-Wave Speech**  
*Fei Chen<sup>1</sup>, Daniel Fogerty<sup>2</sup>*  
*<sup>1</sup>SUSTC, China; <sup>2</sup>University of South Carolina, USA*
- PAGE 1696  
Sat-O-5-4-6  
15:10      **Effects of Urgent Speech and Preceding Sounds on Speech Intelligibility in Noisy and Reverberant Environments**  
*Nao Hodoshima, Tokai University, Japan*

## Sat-O-5-5: Robust Speaker Recognition and Anti-Spoofing

*Chairs: Tomi Kinnunen, Nicholas Evans*

*Room: Seacliff BCD, Time 13:30-15:30, Saturday, 10 Sept. 2016*

---

- PAGE 1700  
Sat-O-5-5-1  
13:30      **Integrated Spoofing Countermeasures and Automatic Speaker Verification: An Evaluation on ASVspoof 2015**  
*Md. Sahidullah<sup>1</sup>, Héctor Delgado<sup>2</sup>, Massimiliano Todisco<sup>2</sup>, Hong Yu<sup>3</sup>, Tomi Kinnunen<sup>1</sup>, Nicholas Evans<sup>2</sup>, Zheng-Hua Tan<sup>3</sup>*  
*<sup>1</sup>University of Eastern Finland, Finland; <sup>2</sup>EURECOM, France; <sup>3</sup>Aalborg University, Denmark*
- PAGE 1705  
Sat-O-5-5-2  
13:50      **Cross-Database Evaluation of Audio-Based Spoofing Detection Systems**  
*Pavel Korshunov, Sébastien Marcel, Idiap Research Institute, Switzerland*
- PAGE 1710  
Sat-O-5-5-3  
14:10      **Investigation of Sub-Band Discriminative Information Between Spoofed and Genuine Speech**  
*Kaavya Sriskandaraja, Vidhyasaharan Sethu, Phu Ngoc Le, Eliathamby Ambikairajah, University of New South Wales, Australia*
- PAGE 1715  
Sat-O-5-5-4  
14:30      **An Investigation of Spoofing Speech Detection Under Additive Noise and Reverberant Conditions**  
*Xiaohai Tian<sup>1</sup>, Zhizheng Wu<sup>2</sup>, Xiong Xiao<sup>3</sup>, Eng Siong Chng<sup>1</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>NTU, Singapore; <sup>2</sup>University of Edinburgh, UK; <sup>3</sup>TL@NTU, Singapore*
- PAGE 1720  
Sat-O-5-5-5  
14:50      **Robust Speaker Recognition with Combined Use of Acoustic and Throat Microphone Speech**  
*Md. Sahidullah<sup>1</sup>, Rosa Gonzalez Hautamäki<sup>1</sup>, Dennis Alexander Lehmann Thomsen<sup>2</sup>, Tomi Kinnunen<sup>1</sup>, Zheng-Hua Tan<sup>2</sup>, Ville Hautamäki<sup>1</sup>, Robert Parts<sup>3</sup>, Martti Pitkänen<sup>3</sup>*  
*<sup>1</sup>University of Eastern Finland, Finland; <sup>2</sup>Aalborg University, Denmark; <sup>3</sup>Aplcomp, Finland*

PAGE 1725  
Sat-O-5-5-6  
15:10

**Statistical Modeling of Speaker's Voice with Temporal Co-Location for Active Voice Authentication**

*Zhong Meng, Biing-Hwang Juang, Georgia Institute of Technology, USA*

**Sat-O-5-6: Speech Enhancement and Applications**

*Chairs: Horacio Franco, Omid Sadjadi*

*Room: Seacliff A, Time 13:30-15:30, Saturday, 10 Sept. 2016*

---

PAGE 1730  
Sat-O-5-6-1  
13:30

**Joint Enhancement and Coding of Speech by Incorporating Wiener Filtering in a CELP Codec**

*Johannes Fischer, Tom Bäckström, FAU Erlangen-Nürnberg, Germany*

PAGE 1735  
Sat-O-5-6-2  
13:50

**Multi-Channel Linear Prediction Based on Binaural Coherence for Speech Dereverberation**

*Hong Liu, Xiuling Wang, Miao Sun, Cheng Pang, Peking University, China*

PAGE 1740  
Sat-O-5-6-3  
14:10

**Single-Channel Speech Enhancement Using Double Spectrum**

*Martin Blass<sup>1</sup>, Pejman Mowlae<sup>1</sup>, W. Bastiaan Kleijn<sup>2</sup>*

*<sup>1</sup>Technische Universität Graz, Austria; <sup>2</sup>Victoria University of Wellington, New Zealand*

PAGE 1745  
Sat-O-5-6-4  
14:30

**On the Appropriateness of Complex-Valued Neural Networks for Speech Enhancement**

*Lukas Drude<sup>1</sup>, Bhiksha Raj<sup>2</sup>, Reinhold Haeb-Umbach<sup>1</sup>*

*<sup>1</sup>Universität Paderborn, Germany; <sup>2</sup>Carnegie Mellon University, USA*

PAGE 1750  
Sat-O-5-6-5  
14:50

**Introducing the Turbo-Twin-HMM for Audio-Visual Speech Enhancement**

*Steffen Zeiler<sup>1</sup>, Hendrik Meutzner<sup>1</sup>, Ahmed Hussien Abdelaziz<sup>2</sup>, Dorothea Kolossa<sup>1</sup>*

*<sup>1</sup>Ruhr-Universität Bochum, Germany; <sup>2</sup>ICSI, USA*

PAGE 1755  
Sat-O-5-6-6  
15:10

**Assessing Speech Quality in Speech-Aware Hearing Aids Based on Phoneme Posteriorgrams**

*Constantin Spille<sup>1</sup>, Hendrik Kayser<sup>1</sup>, Hynek Hermansky<sup>2</sup>, Bernd T. Meyer<sup>2</sup>*

*<sup>1</sup>Carl von Ossietzky Universität Oldenburg, Germany; <sup>2</sup>Johns Hopkins University, USA*



## Sat-P-5-1 : Speech Analysis

Chair: Asterios Toutios

Room: Pacific Concourse - Poster A, Time 13:30-15:30, Saturday, 10 Sept. 2016

---

- PAGE 1760  
Sat-P-5-1-1  
13:30  
**Time-Varying Quasi-Closed-Phase Weighted Linear Prediction Analysis of Speech for Accurate Formant Detection and Tracking**  
*Dhananjaya Gowda, Paavo Alku, Aalto University, Finland*
- PAGE 1765  
Sat-P-5-1-2  
13:30  
**Improved Depiction of Tissue Boundaries in Vocal Tract Real-Time MRI Using Automatic Off-Resonance Correction**  
*Yongwan Lim, Sajjan Goud Lingala, Asterios Toutios, Shrikanth S. Narayanan, Krishna S. Nayak, University of Southern California, USA*
- PAGE 1770  
Sat-P-5-1-3  
13:30  
**Modeling and Transforming Speech Using Variational Autoencoders**  
*Merlijn Blaauw, Jordi Bonada, Universitat Pompeu Fabra, Spain*
- PAGE 1775  
Sat-P-5-1-4  
13:30  
**Phase-Encoded Speech Spectrograms**  
*Chandra Sekhar Seelamantula, Indian Institute of Science, India*
- PAGE 1780  
Sat-P-5-1-5  
13:30  
**Towards Minimally Invasive Velar State Detection in Normal and Silent Speech**  
*Peter Birkholz<sup>1</sup>, Petko Bakardjiev<sup>1</sup>, Steffen Kürbis<sup>1</sup>, Rico Petrick<sup>2</sup>*  
*<sup>1</sup>Technische Universität Dresden, Germany; <sup>2</sup>Linguwerk, Germany*
- PAGE 1785  
Sat-P-5-1-6  
13:30  
**RNN-BLSTM Based Multi-Pitch Estimation**  
*Jianshu Zhang, Jian Tang, Li-Rong Dai, USTC, China*
- PAGE 1790  
Sat-P-5-1-7  
13:30  
**TUSK: A Framework for Overviewing the Performance of F0 Estimators**  
*Masanori Morise<sup>1</sup>, Hideki Kawahara<sup>2</sup>*  
*<sup>1</sup>University of Yamanashi, Japan; <sup>2</sup>Wakayama University, Japan*
- PAGE 1795  
Sat-P-5-1-8  
13:30  
**A Robust Non-Parametric and Filtering Based Approach for Glottal Closure Instant Detection**  
*Pradeep Rengaswamy, Gurunath Reddy M., K. Sreenivasa Rao, Pallab Dasgupta, IIT Kharagpur, India*

## Sat-P-5-2 : Speaker Recognition

Chairs: Eliathamby Ambikairajah, Rahim Saeidi

Room: Pacific Concourse - Poster B, Time 13:30-15:30, Saturday, 10 Sept. 2016

---

- PAGE 1800  
Sat-P-5-2-1  
13:30  
**Analysis of Face Mask Effect on Speaker Recognition**  
*Rahim Saeidi, Ilkka Huhtakallio, Paavo Alku, Aalto University, Finland*
- PAGE 1805  
Sat-P-5-2-2  
13:30  
**Data Selection for Within-Class Covariance Estimation**  
*Elliot Singer<sup>1</sup>, Tyler Campbell<sup>2</sup>, Douglas Reynolds<sup>1</sup>*  
*<sup>1</sup>MIT Lincoln Laboratory, USA; <sup>2</sup>Rensselaer Polytechnic Institute, USA*
- PAGE 1810  
Sat-P-5-2-3  
13:30  
**Inter-Task System Fusion for Speaker Recognition**  
*M. Ferras, Srikanth Madikeri, S. Dey, Petr Motlicek, Hervé Boudlard, Idiap Research Institute, Switzerland*
- PAGE 1815  
Sat-P-5-2-4  
13:30  
**Mahalanobis Metric Scoring Learned from Weighted Pairwise Constraints in I-Vector Speaker Recognition System**  
*Zhenchun Lei, Yanhong Wan, Jian Luo, Yingen Yang, Jiangxi Normal University, China*
- PAGE 1820  
Sat-P-5-2-5  
13:30  
**Novel Subband Autoencoder Features for Detection of Spoofed Speech**  
*Meet H. Soni, Tanvina B. Patel, Hemant A. Patil, DA-IICT, India*
- PAGE 1825  
Sat-P-5-2-6  
13:30  
**On the Issue of Calibration in DNN-Based Speaker Recognition Systems**  
*Mitchell McLaren<sup>1</sup>, Diego Castan<sup>1</sup>, Luciana Ferrer<sup>2</sup>, Aaron Lawson<sup>1</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>Universidad de Buenos Aires, Argentina*
- PAGE 1830  
Sat-P-5-2-7  
13:30  
**Probabilistic Approach Using Joint Long and Short Session i-Vectors Modeling to Deal with Short Utterances for Speaker Recognition**  
*Waad Ben Kheder, Driss Matrouf, Moez Ajili, Jean-François Bonastre, LIA, France*



PAGE 1835 Sat-P-5-2-8 13:30	<b>Short Utterance Variance Modelling and Utterance Partitioning for PLDA Speaker Verification</b> <i>Ahilan Kanagasundaram, David Dean, Sridha Sridharan, Clinton Fookes, Ivan Himawan, Queensland University of Technology, Australia</i>
PAGE 1839 Sat-P-5-2-9 13:30	<b>Speaker-Dependent Dictionary-Based Speech Enhancement for Text-Dependent Speaker Verification</b> <i>Nicolai Bæk Thomsen, Dennis Alexander Lehmann Thomsen, Zheng-Hua Tan, Børge Lindberg, Søren Holdt Jensen, Aalborg University, Denmark</i>
PAGE 1844 Sat-P-5-2-10 13:30	<b>Text-Available Speaker Recognition System for Forensic Applications</b> <i>Chengzhu Yu, Chunlei Zhang, Finnian Kelly, Abhijeet Sangwan, John H.L. Hansen, University of Texas at Dallas, USA</i>
PAGE 1848 Sat-P-5-2-11 13:30	<b>Transfer Learning for Speaker Verification on Short Utterances</b> <i>Qingyang Hong, Lin Li, Lihong Wan, Jun Zhang, Feng Tong, Xiamen University, China</i>
PAGE 1853 Sat-P-5-2-12 13:30	<b>Twin Model G-PLDA for Duration Mismatch Compensation in Text-Independent Speaker Verification</b> <i>Jianbo Ma<sup>1</sup>, Vidhyasaharan Sethu<sup>1</sup>, Eliathamby Ambikairajah<sup>1</sup>, Kong Aik Lee<sup>2</sup></i> <i><sup>1</sup>University of New South Wales, Australia; <sup>2</sup>A*STAR, Singapore</i>
PAGE 1858 Sat-P-5-2-13 13:30	<b>Universal Background Sparse Coding and Multilayer Bootstrap Network for Speaker Clustering</b> <i>Xiao-Lei Zhang, Northwestern Polytechnical University, China</i>
PAGE 1863 Sat-P-5-2-14 13:30	<b>Improving Deep Neural Networks Based Speaker Verification Using Unlabeled Data</b> <i>Yao Tian<sup>1</sup>, Meng Cai<sup>2</sup>, Liang He<sup>1</sup>, Wei-Qiang Zhang<sup>1</sup>, Jia Liu<sup>1</sup></i> <i><sup>1</sup>Tsinghua University, China; <sup>2</sup>Microsoft, China</i>

## Sat-P-5-3: Decoding, System Combination

Chair: Alexey Karpov

Room: Pacific Concourse - Poster C, Time 13:30-15:30, Saturday, 10 Sept. 2016

---

PAGE 1868 Sat-P-5-3-1 13:30	<b>Maximum a posteriori Based Decoding for CTC Acoustic Models</b> <i>Naoyuki Kanda, Xugang Lu, Hisashi Kawai, NICT, Japan</i>
PAGE 1873 Sat-P-5-3-2 13:30	<b>Phonetic and Phonological Posterior Search Space Hashing Exploiting Class-Specific Sparsity Structures</b> <i>Afsaneh Asaei, Gil Luyet, Milos Cernak, Hervé Bourlard, Idiap Research Institute, Switzerland</i>
PAGE 1878 Sat-P-5-3-3 13:30	<b>Model Compression Applied to Small-Footprint Keyword Spotting</b> <i>George Tucker<sup>1</sup>, Minhua Wu<sup>2</sup>, Ming Sun<sup>2</sup>, Sankaran Panchapagesan<sup>2</sup>, Gengshen Fu<sup>2</sup>, Shiv Vitaladevuni<sup>2</sup></i> <i><sup>1</sup>Google, USA; <sup>2</sup>Amazon.com, USA</i>
PAGE 1883 Sat-P-5-3-4 13:30	<b>Why do ASR Systems Despite Neural Nets Still Depend on Robust Features</b> <i>Angel Mario Castro Martinez, Marc René Schädler, Carl von Ossietzky Universität Oldenburg, Germany</i>
PAGE 1888 Sat-P-5-3-5 13:30	<b>An Adaptive Multi-Band System for Low Power Voice Command Recognition</b> <i>Qing He<sup>1</sup>, Gregory W. Wornell<sup>1</sup>, Wei Ma<sup>2</sup></i> <i><sup>1</sup>MIT, USA; <sup>2</sup>Texas Instruments, USA</i>
PAGE 1893 Sat-P-5-3-6 13:30	<b>Memory-Efficient Modeling and Search Techniques for Hardware ASR Decoders</b> <i>Michael Price, Anantha Chandrakasan, James Glass, MIT, USA</i>
PAGE 1898 Sat-P-5-3-7 13:30	<b>Log-Linear System Combination Using Structured Support Vector Machines</b> <i>J. Yang, Anton Ragni, Mark J.F. Gales, Kate M. Knill, University of Cambridge, UK</i>

PAGE 1903 Sat-P-5-3-8 13:30	<b>Efficient Segmental Cascades for Speech Recognition</b> <i>Hao Tang, Weiran Wang, Kevin Gimpel, Karen Livescu, TTIC, USA</i>
PAGE 1908 Sat-P-5-3-9 13:30	<b>A WFST Framework for Single-Pass Multi-Stream Decoding</b> <i>Sirui Xu, Eric Fosler-Lussier, Ohio State University, USA</i>
PAGE 1913 Sat-P-5-3-10 13:30	<b>Comparison of Multiple System Combination Techniques for Keyword Spotting</b> <i>William Hartmann, Le Zhang, Kerri Barnes, Roger Hsiao, Stavros Tsakalidis, Richard Schwartz, Raytheon BBN Technologies, USA</i>
PAGE 1918 Sat-P-5-3-11 13:30	<b>Rescoring by Combination of Posteriorgram Score and Subword-Matching Score for Use in Query-by-Example</b> <i>Masato Obara<sup>1</sup>, Kazunori Kojima<sup>1</sup>, Kazuyo Tanaka<sup>2</sup>, Shi-wook Lee<sup>3</sup>, Yoshiaki Itoh<sup>1</sup></i> <i><sup>1</sup>Iwate Prefectural University, Japan; <sup>2</sup>University of Tsukuba, Japan; <sup>3</sup>AIST, Japan</i>
PAGE 1923 Sat-P-5-3-12 13:30	<b>Phone Synchronous Decoding with CTC Lattice</b> <i>Zhehuai Chen<sup>1</sup>, Wei Deng<sup>2</sup>, Tao Xu<sup>2</sup>, Kai Yu<sup>1</sup></i> <i><sup>1</sup>Shanghai Jiao Tong University, China; <sup>2</sup>AlSpeech, China</i>

## Sat-P-5-4: Special Session: Clinical and Neuroscience-Inspired Vocal Biomarkers of Neurological and Psychiatric Disorders

*Chairs: Nicholas Cummins, Julien Epps, Emily Mower Provost, Thomas Quatieri, Stefan Scherer*

*Room: Pacific Concourse - Poster D, Time 13:30-15:30, Saturday, 10 Sept. 2016*

---

PAGE 1928 Sat-P-5-4-1 13:30	<b>Speech Features for Depression Detection</b> <i>Saurabh Sahu, Carol Espy-Wilson, University of Maryland, USA</i>
PAGE 1933 Sat-P-5-4-2 13:30	<b>Parkinson's Disease Progression Assessment from Speech Using GMM-UBM</b> <i>T. Arias-Vergara<sup>1</sup>, J.C. Vasquez-Correa<sup>1</sup>, Juan Rafael Orozco-Arroyave<sup>1</sup>, J.F. Vargas-Bonilla<sup>1</sup>, Elmar Nöth<sup>2</sup></i> <i><sup>1</sup>Universidad de Antioquia, Colombia; <sup>2</sup>FAU Erlangen-Nürnberg, Germany</i>
PAGE 1938 Sat-P-5-4-3 13:30	<b>Speech-Based Detection of Alzheimer's Disease in Conversational German</b> <i>Jochen Weiner, Christian Herff, Tanja Schultz, Universität Bremen, Germany</i>
PAGE 1943 Sat-P-5-4-4 13:30	<b>Cross-Cultural Depression Recognition from Vocal Biomarkers</b> <i>Sharifa Alghowinem<sup>1</sup>, Roland Goecke<sup>2</sup>, Julien Epps<sup>3</sup>, Michael Wagner<sup>2</sup>, Jeffrey Cohn<sup>4</sup></i> <i><sup>1</sup>Australian National University, Australia; <sup>2</sup>University of Canberra, Australia; <sup>3</sup>University of New South Wales, Australia; <sup>4</sup>University of Pittsburgh, USA</i>
PAGE 1948 Sat-P-5-4-5 13:30	<b>Speech Recognition in Alzheimer's Disease and in its Assessment</b> <i>Luke Zhou<sup>1</sup>, Kathleen C. Fraser<sup>1</sup>, Frank Rudzicz<sup>2</sup></i> <i><sup>1</sup>University of Toronto, Canada; <sup>2</sup>Toronto Rehabilitation Institute, Canada</i>
PAGE 1953 Sat-P-5-4-6 13:30	<b>Does She Speak RTT? Towards an Earlier Identification of Rett Syndrome Through Intelligent Pre-Linguistic Vocalisation Analysis</b> <i>Florian B. Pokorny<sup>1</sup>, Peter B. Marschik<sup>1</sup>, Christa Einspieler<sup>1</sup>, Björn Schuller<sup>2</sup></i> <i><sup>1</sup>Medizinische Universität Graz, Austria; <sup>2</sup>Universität Passau, Germany</i>
PAGE 1958 Sat-P-5-4-7 13:30	<b>Speech Rhythm in Parkinson's Disease: A Study on Italian</b> <i>Massimo Pettorino<sup>1</sup>, Maria Grazia Busà<sup>2</sup>, Elisa Pellegrino<sup>1</sup></i> <i><sup>1</sup>Università di Napoli "L'Orientale", Italy; <sup>2</sup>Università di Padova, Italy</i>

## Sat-S&T-5: Show & Tell Session 5

*Chairs: Shiva Sundaram, Nicolas Scheffer*

*Room: Market Street Foyer, Time 13:30–15:30, Saturday, 10 Sept. 2016*

---

PAGE 1962  
Sat-S&T-5-1  
13:30

### **English Language Speech Assistant**

*Xavier Anguera<sup>1</sup>, Vu Van<sup>2</sup>*

*<sup>1</sup>ELSA, Portugal; <sup>2</sup>ELSA, USA*

PAGE 1964  
Sat-S&T-5-2  
13:30

### **Remeeting — Deep Insights to Conversations**

*Allen Guo, Arlo Faria, Korbinian Riedhammer, Remeeting, USA*

PAGE 1966  
Sat-S&T-5-3  
13:30

### **SERAPHIM Live! — Singing Synthesis for the Performer, the Composer, and the 3D Game Developer**

*Paul Yaozhu Chan<sup>1</sup>, Minghui Dong<sup>1</sup>, Grace Xue Hui Ho<sup>2</sup>, Haizhou Li<sup>1</sup>*

*<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>NTU, Singapore*

PAGE 1968  
Sat-S&T-5-4  
13:30

### **My-Own-Voice: A Web Service That Allows You to Create a Text-to-Speech Voice From Your Own Voice**

*Fabrice Malfre, Olivier Deroo, Emmanuelle Franques, Jonathan Hourez, Nicolas Mazars, Vincent Pagel, Geoffrey Wilfart, Acapela Group, Belgium*

## Keynote 3: Anne Fernald

*Chair: Shri Narayanan*

*Room: Grand Ballroom ABC, Time 08:30–09:30, Sunday, 11 Sept. 2016*

---

PAGE 1970  
Sun-Keynote-3  
08:30

### **Talking with Kids Really Matters: Early Language Experience Shapes Later Life Chances**

*Anne Fernald, Stanford University, USA*

## Sun-O-6-1 : Far-Field Speech Processing

Chairs: Jinyu Li, Tomohiro Nakatani

Room: Grand Ballroom A, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 1971  
Sun-O-6-1-1  
10:00      **Reducing the Computational Complexity of Multimicrophone Acoustic Models with Integrated Feature Extraction**  
*Tara N. Sainath, Arun Narayanan, Ron J. Weiss, Ehsan Variani, Kevin W. Wilson, Michiel Bacchiani, Izhak Shafran, Google, USA*
- PAGE 1976  
Sun-O-6-1-2  
10:20      **Neural Network Adaptive Beamforming for Robust Multichannel Speech Recognition**  
*Bo Li, Tara N. Sainath, Ron J. Weiss, Kevin W. Wilson, Michiel Bacchiani, Google, USA*
- PAGE 1981  
Sun-O-6-1-3  
10:40      **Improved MVDR Beamforming Using Single-Channel Mask Prediction Networks**  
*Hakan Erdogan<sup>1</sup>, John R. Hershey<sup>2</sup>, Shinji Watanabe<sup>2</sup>, Michael I. Mandel<sup>3</sup>, Jonathan Le Roux<sup>2</sup>*  
*<sup>1</sup>Sabancı Üniversitesi, Turkey; <sup>2</sup>MERL, USA; <sup>3</sup>CUNY Brooklyn College, USA*
- PAGE 1986  
Sun-O-6-1-4  
11:00      **Channel Selection for Distant Speech Recognition Exploiting Cepstral Distance**  
*Cristina Guerrero<sup>1</sup>, Georgina Tryfou<sup>1</sup>, Maurizio Omologo<sup>2</sup>*  
*<sup>1</sup>Università di Trento, Italy; <sup>2</sup>FBK, Italy*
- PAGE 1991  
Sun-O-6-1-5  
11:20      **Multichannel Spatial Clustering for Robust Far-Field Automatic Speech Recognition in Mismatched Conditions**  
*Michael I. Mandel<sup>1</sup>, Jon Barker<sup>2</sup>*  
*<sup>1</sup>CUNY Brooklyn College, USA; <sup>2</sup>University of Sheffield, UK*
- PAGE 1996  
Sun-O-6-1-6  
11:40      **Far-Field ASR Without Parallel Data**  
*Vijayaditya Peddinti, Vimal Manohar, Yiming Wang, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA*

## Sun-O-6-2 : Special Session: Interspeech 2016 Computational Paralinguistics Challenge (ComParE): Deception, Sincerity & Native Language

Chairs: Björn Schuller, Stefan Steidl, Anton Batliner, Julia Hirschberg, Judee Burgoon, Eduardo Coutinho

Room: Grand Ballroom BC, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2001  
Sun-O-6-2-1  
10:00      **The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language**  
*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA*
- N/A  
Sun-O-6-2-2  
10:10      **The Deception Sub-Challenge: The Data**  
*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA*
- PAGE 2006  
Sun-O-6-2-3  
10:20      **Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection**  
*Sarah Ita Levitan<sup>1</sup>, Guozhen An<sup>2</sup>, Min Ma<sup>2</sup>, Rivka Levitan<sup>3</sup>, Andrew Rosenberg<sup>4</sup>, Julia Hirschberg<sup>2</sup>*  
*<sup>1</sup>Columbia University, USA; <sup>2</sup>CUNY Graduate Center, USA; <sup>3</sup>CUNY Brooklyn College, USA; <sup>4</sup>CUNY Queens College, USA*

PAGE 2011  
Sun-O-6-2-4  
10:30

**Is Deception Emotional? An Emotion-Driven Predictive Approach**  
*Shahin Amiriparian, Jouni Pohjalainen, Erik Marchi, Sergey Pugachevskiy, Björn Schuller, Universität Passau, Germany*

PAGE 2016  
Sun-O-6-2-5  
10:40

**Prosodic Cues and Answer Type Detection for the Deception Sub-Challenge**  
*Claude Montacié, Marie-José Caraty, STIH (EA 4509), France*

N/A  
Sun-O-6-2-6  
10:50

**The Sincerity Sub-Challenge: The Data**  
*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA*

PAGE 2021  
Sun-O-6-2-7  
11:00

**Automatic Estimation of Perceived Sincerity from Spoken Language**  
*Brandon M. Booth, Rahul Gupta, Pavlos Papadopoulos, Ruchir Travadi, Shrikanth S. Narayanan, University of Southern California, USA*

PAGE 2026  
Sun-O-6-2-8  
11:10

**Estimating the Sincerity of Apologies in Speech by DNN Rank Learning and Prosodic Analysis**  
*Gábor Gosztolya<sup>1</sup>, Tamás Grósz<sup>1</sup>, György Szaszák<sup>2</sup>, László Tóth<sup>3</sup>*  
*<sup>1</sup>University of Szeged, Hungary; <sup>2</sup>BME, Hungary; <sup>3</sup>MTA-SZTE RGAI, Hungary*

PAGE 2031  
Sun-O-6-2-9  
11:20

**Minimization of Regression and Ranking Losses with Shallow Neural Networks on Automatic Sincerity Evaluation**  
*Hung-Shin Lee<sup>1</sup>, Yu Tsao<sup>2</sup>, Chi-Chun Lee<sup>3</sup>, Hsin-Min Wang<sup>2</sup>, Wei-Cheng Lin<sup>3</sup>, Wei-Chen Chen<sup>3</sup>, Shan-Wen Hsiao<sup>3</sup>, Shyh-Kang Jeng<sup>1</sup>*  
*<sup>1</sup>National Taiwan University, Taiwan; <sup>2</sup>Academia Sinica, Taiwan; <sup>3</sup>National Tsing Hua University, Taiwan*

[Search]

*Sun-O-6-2 continues next page...*

PAGE 2036  
Sun-O-6-2-10  
11:30

**Prediction of Deception and Sincerity from Speech Using Automatic Phone Recognition-Based Features**  
*Robert Herms, Technische Universität Chemnitz, Germany*

PAGE 2041  
Sun-O-6-2-11  
11:40

**Sincerity and Deception in Speech: Two Sides of the Same Coin? A Transfer- and Multi-Task Learning Perspective**  
*Yue Zhang<sup>1</sup>, Felix Weninger<sup>2</sup>, Zhao Ren<sup>3</sup>, Björn Schuller<sup>1</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>Nuance Communications, Germany; <sup>3</sup>Northwestern Polytechnical University, China*

PAGE 2046  
Sun-O-6-2-12  
11:50

**Fusing Acoustic Feature Representations for Computational Paralinguistics Tasks**  
*Heysem Kaya<sup>1</sup>, Alexey A. Karpov<sup>2</sup>*  
*<sup>1</sup>Namik Kemal Üniversitesi, Turkey; <sup>2</sup>Russian Academy of Sciences, Russia*

## Sun-O-6-3: Special Session: Speech, Audio, and Language Processing Techniques Applied to Bird and Animal Vocalizations

Chairs: Naomi Harte, Peter Jančovič, Karl-L. Schuchmann

Room: Bayview A, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

N/A Sun-O-6-3-1 10:00	<b>Introduction</b> <i>Naomi Harte</i> <sup>1</sup> , <i>Peter Jančovič</i> <sup>2</sup> , <i>Karl-L. Schuchmann</i> <sup>3</sup> <sup>1</sup> Trinity College Dublin, Ireland; <sup>2</sup> University of Birmingham, UK; <sup>3</sup> ZFMK, Germany
N/A Sun-O-6-3-2 10:05	<b>Poster Overview Presentations</b> <i>Naomi Harte</i> <sup>1</sup> , <i>Peter Jančovič</i> <sup>2</sup> , <i>Karl-L. Schuchmann</i> <sup>3</sup> <sup>1</sup> Trinity College Dublin, Ireland; <sup>2</sup> University of Birmingham, UK; <sup>3</sup> ZFMK, Germany
N/A Sun-O-6-3-3 11:15	<b>Discussion</b> <i>Naomi Harte</i> <sup>1</sup> , <i>Peter Jančovič</i> <sup>2</sup> , <i>Karl-L. Schuchmann</i> <sup>3</sup> <sup>1</sup> Trinity College Dublin, Ireland; <sup>2</sup> University of Birmingham, UK; <sup>3</sup> ZFMK, Germany
N/A Sun-O-6-3-4 11:55	<b>Closing Remarks</b> <i>Naomi Harte</i> <sup>1</sup> , <i>Peter Jančovič</i> <sup>2</sup> , <i>Karl-L. Schuchmann</i> <sup>3</sup> <sup>1</sup> Trinity College Dublin, Ireland; <sup>2</sup> University of Birmingham, UK; <sup>3</sup> ZFMK, Germany

## Sun-O-6-4: Dialogue Systems and Analysis of Dialogue

Chairs: Jens Edlund, Alexandros Potamianos

Room: Bayview B, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

PAGE 2051 Sun-O-6-4-1 10:00	<b>A Stochastic Model for Computer-Aided Human-Human Dialogue</b> <i>Merwan Barlier</i> <sup>1</sup> , <i>Romain Laroche</i> <sup>1</sup> , <i>Olivier Pietquin</i> <sup>2</sup> <sup>1</sup> Orange Labs, France; <sup>2</sup> CRISAL, France
PAGE 2056 Sun-O-6-4-2 10:20	<b>Highlighting Psychological Features for Predicting Child Interjections During Story Telling</b> <i>Gaël Lejeune, François Rioult, Bruno Crémilleux, GREYC, France</i>
PAGE 2060 Sun-O-6-4-3 10:40	<b>Hybrid Dialogue State Tracking for Real World Human-to-Human Dialogues</b> <i>Kai Sun, Su Zhu, Lu Chen, Siqui Yao, Xueyang Wu, Kai Yu, Shanghai Jiao Tong University, China</i>
PAGE 2065 Sun-O-6-4-4 11:00	<b>Automatic Recognition of Social Roles Using Long Term Role Transitions in Small Group Interactions</b> <i>Gaurav Fotedar</i> <sup>1</sup> , <i>Aditya Gaonkar P.</i> <sup>1</sup> , <i>Saikat Chatterjee</i> <sup>2</sup> , <i>Prasanta Kumar Ghosh</i> <sup>1</sup> <sup>1</sup> Indian Institute of Science, India; <sup>2</sup> KTH, Sweden
PAGE 2070 Sun-O-6-4-5 11:20	<b>On the Influence of Gender on Interruptions in Multiparty Dialogue</b> <i>Paul Van Eecke</i> <sup>1</sup> , <i>Raquel Fernández</i> <sup>2</sup> <sup>1</sup> Sony, France; <sup>2</sup> Universiteit van Amsterdam, The Netherlands
PAGE 2075 Sun-O-6-4-6 11:40	<b>Detection of User Escalation in Human-Computer Interactions</b> <i>Ian Beaver, Cynthia Freeman, NextIT, USA</i>

## Sun-O-6-5: Interaction between Speech Production and Perception

Chairs: Outi Tuomainen, Christopher Davis

Room: Seacliff BCD, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2080  
Sun-O-6-5-1  
10:00      **Assessing Idiosyncrasies in a Bayesian Model of Speech Communication**  
*Marie-Lou Barnaud<sup>1</sup>, Julien Diard<sup>2</sup>, Pierre Bessière<sup>3</sup>, Jean-Luc Schwartz<sup>1</sup>*  
*<sup>1</sup>GIPSA, France; <sup>2</sup>LPNC, France; <sup>3</sup>ISIR, France*
- PAGE 2085  
Sun-O-6-5-2  
10:20      **Prosodic and Linguistic Analysis of Semantic Fluency Data: A Window into Speech Production and Cognition**  
*Maria K. Wolters<sup>1</sup>, Najoung Kim<sup>2</sup>, Jung-Ho Kim<sup>2</sup>, Sarah E. MacPherson<sup>1</sup>, Jong C. Park<sup>2</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>KAIST, Korea*
- PAGE 2090  
Sun-O-6-5-3  
10:40      **Sensorimotor Response to Visual Imagery of Tongue Displacement**  
*William F. Katz<sup>1</sup>, Divya Prabhakaran<sup>2</sup>*  
*<sup>1</sup>University of Texas at Dallas, USA; <sup>2</sup>Plano East Senior High School, USA*
- PAGE 2095  
Sun-O-6-5-4  
11:00      **Does Auditory-Motor Learning of Speech Transfer from the CV Syllable to the CVCV Word?**  
*Tiphaine Caudrelier, Pascal Perrier, Jean-Luc Schwartz, Amélie Rochet-Capellan, GIPSA, France*
- PAGE 2100  
Sun-O-6-5-5  
11:20      **Exemplar Dynamics in Phonetic Convergence of Speech Rate**  
*Antje Schweitzer, Michael Walsh, Universität Stuttgart, Germany*
- PAGE 2105  
Sun-O-6-5-6  
11:40      **Articulation Rate in Adverse Listening Conditions in Younger and Older Adults**  
*Outi Tuomainen, Valerie Hazan, University College London, UK*

## Sun-O-6-6: Multimodal Processing

Chairs: Florian Metze, Angeliki Metallinou

Room: Seacliff A, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2110  
Sun-O-6-6-1  
10:00      **Error Correction in Lightly Supervised Alignment of Broadcast Subtitles**  
*Julia Olcoz<sup>1</sup>, Oscar Saz<sup>2</sup>, Thomas Hain<sup>2</sup>*  
*<sup>1</sup>Universidad de Zaragoza, Spain; <sup>2</sup>University of Sheffield, UK*
- PAGE 2115  
Sun-O-6-6-2  
10:20      **Automatic Genre and Show Identification of Broadcast Media**  
*Mortaza Doulaty, Oscar Saz, Raymond W.M. Ng, Thomas Hain, University of Sheffield, UK*
- PAGE 2120  
Sun-O-6-6-3  
10:40      **Speaker-Targeted Audio-Visual Models for Speech Recognition in Cocktail-Party Environments**  
*Guan-Lin Chao, William Chan, Ian Lane, Carnegie Mellon University, USA*
- PAGE 2125  
Sun-O-6-6-4  
11:00      **Text-Dependent Audiovisual Synchrony Detection for Spoofing Detection in Mobile Person Recognition**  
*Amit Aides, Hagai Aronowitz, IBM, Israel*
- PAGE 2130  
Sun-O-6-6-5  
11:20      **Improving Boundary Estimation in Audiovisual Speech Activity Detection Using Bayesian Information Criterion**  
*Fei Tao, John H.L. Hansen, Carlos Busso, University of Texas at Dallas, USA*
- PAGE 2135  
Sun-O-6-6-6  
11:40      **Dynamic Stream Weighting for Turbo-Decoding-Based Audiovisual ASR**  
*Sebastian Gergen<sup>1</sup>, Steffen Zeiler<sup>1</sup>, Ahmed Hussien Abdelaziz<sup>2</sup>, Robert Nickel<sup>3</sup>, Dorothea Kolossa<sup>1</sup>*  
*<sup>1</sup>Ruhr-Universität Bochum, Germany; <sup>2</sup>ICSI, USA; <sup>3</sup>Bucknell University, USA*



## Sun-P-6-1 : Pitch, Tone, and Music

Chair: Vikramjit Mitra

Room: Pacific Concourse - Poster A, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2140  
Sun-P-6-1-1  
10:00      **Retrieval of Textual Song Lyrics from Sung Inputs**  
*Anna M. Kruspe, Fraunhofer IDMT, Germany*
- PAGE 2145  
Sun-P-6-1-2  
10:00      **Phoneme, Phone Boundary, and Tone in Automatic Scoring of Mandarin Proficiency**  
*Jiahong Yuan, Mark Liberman, University of Pennsylvania, USA*
- PAGE 2150  
Sun-P-6-1-3  
10:00      **Tone Classification in Mandarin Chinese Using Convolutional Neural Networks**  
*Charles Chen, Razvan Bunescu, Li Xu, Chang Liu, Ohio University, USA*
- PAGE 2155  
Sun-P-6-1-4  
10:00      **Robust Estimation of Fundamental Frequency Using Single Frequency Filtering Approach**  
*Vishala Pannala, G. Aneesa, Sudarsana Reddy Kadiri, B. Yegnanarayana, IIIT Hyderabad, India*
- PAGE 2160  
Sun-P-6-1-5  
10:00      **A Fast and Accurate Fundamental Frequency Estimator Using Recursive Moving Average Filters**  
*Ryunosuke Daido, Yuji Hisaminato, Yamaha, Japan*
- PAGE 2165  
Sun-P-6-1-6  
10:00      **Frequency Estimation from Waveforms Using Multi-Layered Neural Networks**  
*Prateek Verma, Ronald W. Schafer, Stanford University, USA*

## Sun-P-6-2 : Speaker Diarization and Recognition

Chairs: Thomas Hain, Itshak Lapidot

Room: Pacific Concourse - Poster B, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2170  
Sun-P-6-2-1  
10:00      **Speaker Linking and Applications Using Non-Parametric Hashing Methods**  
*Douglas E. Sturim, William M. Campbell, MIT Lincoln Laboratory, USA*
- PAGE 2175  
Sun-P-6-2-2  
10:00      **Iterative PLDA Adaptation for Speaker Diarization**  
*Gaël Le Lan<sup>1</sup>, Delphine Charlet<sup>1</sup>, Anthony Larcher<sup>2</sup>, Sylvain Meignier<sup>2</sup>*  
*<sup>1</sup>Orange Labs, France; <sup>2</sup>LIUM, France*
- PAGE 2180  
Sun-P-6-2-3  
10:00      **A Speaker Diarization System for Studying Peer-Led Team Learning Groups**  
*Harishchandra Dubey, Lakshminish Kaushik, Abhijeet Sangwan, John H.L. Hansen, University of Texas at Dallas, USA*
- PAGE 2185  
Sun-P-6-2-4  
10:00      **DNN-Based Speaker Clustering for Speaker Diarisation**  
*Rosanna Milner, Thomas Hain, University of Sheffield, UK*
- PAGE 2190  
Sun-P-6-2-5  
10:00      **On the Importance of Efficient Transition Modeling for Speaker Diarization**  
*Itshak Lapidot<sup>1</sup>, Jean-François Bonastre<sup>2</sup>*  
*<sup>1</sup>Afeka Tel Aviv Academic College of Engineering, Israel; <sup>2</sup>LIA, France*
- PAGE 2194  
Sun-P-6-2-6  
10:00      **Priors for Speaker Counting and Diarization with AHC**  
*Gregory Sell, Alan McCree, Daniel Garcia-Romero, Johns Hopkins University, USA*
- PAGE 2199  
Sun-P-6-2-7  
10:00      **Two-Pass IB Based Speaker Diarization System Using Meeting-Specific ANN Based Features**  
*Nauman Dawalatabad<sup>1</sup>, Srikanth Madikeri<sup>2</sup>, Chandra Sekhar C.<sup>1</sup>, Hema A. Murthy<sup>1</sup>*  
*<sup>1</sup>IIT Madras, India; <sup>2</sup>Idiap Research Institute, Switzerland*



PAGE 2204 Sun-P-6-2-8 10:00	<b>DNN-Based Amplitude and Phase Feature Enhancement for Noise Robust Speaker Identification</b> <i>Zeyan Oo<sup>1</sup>, Yuta Kawakami<sup>1</sup>, Longbiao Wang<sup>1</sup>, Seiichi Nakagawa<sup>2</sup>, Xiong Xiao<sup>3</sup>, Masahiro Iwahashi<sup>1</sup></i> <sup>1</sup> Nagaoka University of Technology, Japan; <sup>2</sup> Toyohashi University of Technology, Japan; <sup>3</sup> NTU, Singapore
PAGE 2209 Sun-P-6-2-9 10:00	<b>Unit-Selection Attack Detection Based on Unfiltered Frequency-Domain Features</b> <i>Ulrich Scherhag, Andreas Nautsch, Christian Rathgeb, Christoph Busch, Hochschule Darmstadt, Germany</i>
PAGE 2214 Sun-P-6-2-10 10:00	<b>Investigating the Impact of Dialect Prestige on Lexical Decision</b> <i>Mairym Lloréns Monteserín, Jason Zevin, University of Southern California, USA</i>
PAGE 2219 Sun-P-6-2-11 10:00	<b>Speaker Verification Using Short Utterances with DNN-Based Estimation of Subglottal Acoustic Features</b> <i>Jinxi Guo<sup>1</sup>, Gary Yeung<sup>1</sup>, Deepak Muralidharan<sup>1</sup>, Harish Arsikere<sup>2</sup>, Amber Afshan<sup>1</sup>, Abeer Alwan<sup>1</sup></i> <sup>1</sup> University of California at Los Angeles, USA; <sup>2</sup> Xerox Research Center India, India
PAGE 2223 Sun-P-6-2-12 10:00	<b>Factor Analysis Based Speaker Verification Using ASR</b> <i>Hang Su, Steven Wegmann, ICSI, USA</i>
PAGE 2228 Sun-P-6-2-13 10:00	<b>Joint Sound Source Separation and Speaker Recognition</b> <i>Jeroen Zegers, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium</i>
PAGE 2233 Sun-P-6-2-14 10:00	<b>Robust Multichannel Gender Classification from Speech in Movie Audio</b> <i>Naveen Kumar, Md. Nasir, Panayiotis Georgiou, Shrikanth S. Narayanan, University of Southern California, USA</i>

### Sun-P-6-3: Speech Synthesis Poster

Chair: Keiichi Tokuda

Room: Pacific Concourse - Poster C, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

PAGE 2238 Sun-P-6-3-1 10:00	<b>Recent Advances in Google Real-Time HMM-Driven Unit Selection Synthesizer</b> <i>Xavi Gonzalvo, Siamak Tazari, Chun-an Chan, Markus Becker, Alexander Gutkin, Hanna Silen, Google, USA</i>
PAGE 2243 Sun-P-6-3-2 10:00	<b>First Step Towards End-to-End Parametric TTS Synthesis: Generating Spectral Parameters with Neural Attention</b> <i>Wenfu Wang, Shuang Xu, Bo Xu, Chinese Academy of Sciences, China</i>
PAGE 2248 Sun-P-6-3-3 10:00	<b>The Parameterized Phoneme Identity Feature as a Continuous Real-Valued Vector for Neural Network Based Speech Synthesis</b> <i>Zhengqi Wen, Ya Li, Jianhua Tao, Chinese Academy of Sciences, China</i>
PAGE 2253 Sun-P-6-3-4 10:00	<b>Improved Time-Frequency Trajectory Excitation Vocoder for DNN-Based Speech Synthesis</b> <i>Eunwoo Song<sup>1</sup>, Frank K. Soong<sup>1</sup>, Hong-Goo Kang<sup>2</sup></i> <sup>1</sup> Microsoft, China; <sup>2</sup> Yonsei University, Korea
PAGE 2258 Sun-P-6-3-5 10:00	<b>Voice Quality Control Using Perceptual Expressions for Statistical Parametric Speech Synthesis Based on Cluster Adaptive Training</b> <i>Yamato Ohtani, Koichiro Mori, Masahiro Morita, Toshiba, Japan</i>
PAGE 2263 Sun-P-6-3-6 10:00	<b>Waveform Generation Based on Signal Reshaping for Statistical Parametric Speech Synthesis</b> <i>Felipe Espic, Cassia Valentini-Botinhao, Zhizheng Wu, Simon King, University of Edinburgh, UK</i>
PAGE 2268 Sun-P-6-3-7 10:00	<b>Speaker Representations for Speaker Adaptation in Multiple Speakers' BLSTM-RNN-Based Speech Synthesis</b> <i>Yi Zhao, Daisuke Saito, Nobuaki Minematsu, University of Tokyo, Japan</i>

PAGE 2273 Sun-P-6-3-8 10:00	<b>Fast, Compact, and High Quality LSTM-RNN Based Statistical Parametric Speech Synthesizers for Mobile Devices</b> <i>Heiga Zen, Yannis Agiomyrgiannakis, Niels Egberts, Fergus Henderson, Przemysław Szczepaniak, Google, UK</i>
PAGE 2278 Sun-P-6-3-9 10:00	<b>An Investigation of DNN-Based Speech Synthesis Using Speaker Codes</b> <i>Nobukatsu Hojo<sup>1</sup>, Yusuke Ijima<sup>1</sup>, Hideyuki Mizuno<sup>2</sup></i> <i><sup>1</sup>NTT, Japan; <sup>2</sup>Tokyo University of Science, Japan</i>
PAGE 2283 Sun-P-6-3-10 10:00	<b>Using Text and Acoustic Features in Predicting Glottal Excitation Waveforms for Parametric Speech Synthesis with Recurrent Neural Networks</b> <i>Lauri Juvela<sup>1</sup>, Xin Wang<sup>2</sup>, Shinji Takaki<sup>2</sup>, Manu Airaksinen<sup>1</sup>, Junichi Yamagishi<sup>2</sup>, Paavo Alku<sup>1</sup></i> <i><sup>1</sup>Aalto University, Finland; <sup>2</sup>NII, Japan</i>
PAGE 2288 Sun-P-6-3-11 10:00	<b>Model Integration for HMM- and DNN-Based Speech Synthesis Using Product-of-Experts Framework</b> <i>Kentaro Tachibana<sup>1</sup>, Tomoki Toda<sup>2</sup>, Yoshinori Shiga<sup>1</sup>, Hisashi Kawai<sup>1</sup></i> <i><sup>1</sup>NICT, Japan; <sup>2</sup>Nagoya University, Japan</i>
PAGE 2293 Sun-P-6-3-12 10:00	<b>Idlak Tangle: An Open Source Kaldi Based Parametric Speech Synthesiser Based on DNN</b> <i>Blaise Potard<sup>1</sup>, Matthew P. Aylett<sup>1</sup>, David A. Baude<sup>1</sup>, Petr Motlicek<sup>2</sup></i> <i><sup>1</sup>CereProc, UK; <sup>2</sup>Idiap Research Institute, Switzerland</i>
PAGE 2298 Sun-P-6-3-13 10:00	<b>Probabilistic Amplitude Demodulation Features in Speech Synthesis for Improving Prosody</b> <i>Alexandros Lazaridis, Milos Cernak, Philip N. Garner, Idiap Research Institute, Switzerland</i>

PAGE 2303 Sun-P-6-3-14 10:00	<b>On Smoothing and Enhancing Dynamics of Pitch Contours Represented by Discrete Orthogonal Polynomials for Prosody Generation</b> <i>Chen-Yu Chiang, National Taipei University, Taiwan</i>
PAGE 2308 Sun-P-6-3-15 10:00	<b>An Investigation of Recurrent Neural Network Architectures Using Word Embeddings for Phrase Break Prediction</b> <i>Anandaswarup Vadapalli, Suryakanth V. Gangashetty, IIIT Hyderabad, India</i>
PAGE 2313 Sun-P-6-3-16 10:00	<b>Model-Based Parametric Prosody Synthesis with Deep Neural Network</b> <i>Hao Liu<sup>1</sup>, Heng Lu<sup>2</sup>, Xu Shao<sup>2</sup>, Yi Xu<sup>1</sup></i> <i><sup>1</sup>University College London, UK; <sup>2</sup>Nuance Communications, USA</i>

## Sun-P-6-4: Language Model Adaptation

Chair: Ilya Oparin

Room: Pacific Concourse - Poster D, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2318  
Sun-P-6-4-1  
10:00
- Active and Semi-Supervised Learning in ASR: Benefits on the Acoustic and Language Models**  
*Thomas Drugman<sup>1</sup>, Janne Pytköinen<sup>2</sup>, Reinhard Kneser<sup>1</sup>*  
*<sup>1</sup>Amazon.com, Germany; <sup>2</sup>Amazon.com, Finland*
- PAGE 2323  
Sun-P-6-4-2  
10:00
- Learning N-Gram Language Models from Uncertain Data**  
*Vitaly Kuznetsov<sup>1</sup>, Hank Liao<sup>2</sup>, Mehryar Mohri<sup>1</sup>, Michael Riley<sup>2</sup>, Brian Roark<sup>2</sup>*  
*<sup>1</sup>New York University, USA; <sup>2</sup>Google, USA*
- PAGE 2328  
Sun-P-6-4-3  
10:00
- Entropy Based Pruning for Non-Negative Matrix Based Language Models with Contextual Features**  
*Barlas Oğuz, Issac Alphonso, Shuangyu Chang, Microsoft, USA*
- PAGE 2333  
Sun-P-6-4-4  
10:00
- Unsupervised Adaptation of Recurrent Neural Network Language Models**  
*Siva Reddy Gangireddy, Pawel Swietojanski, Peter Bell, Steve Renals, University of Edinburgh, UK*
- PAGE 2338  
Sun-P-6-4-5  
10:00
- Contextual Prediction Models for Speech Recognition**  
*Yoni Halpern<sup>1</sup>, Keith Hall<sup>1</sup>, Vlad Schogol<sup>1</sup>, Michael Riley<sup>1</sup>, Brian Roark<sup>1</sup>, Gleb Skobeltsyn<sup>2</sup>, Martin Bäuml<sup>2</sup>*  
*<sup>1</sup>Google, USA; <sup>2</sup>Google, Switzerland*
- PAGE 2343  
Sun-P-6-4-6  
10:00
- Combining Feature and Model-Based Adaptation of RNNLMs for Multi-Genre Broadcast Speech Recognition**  
*Salil Deena, Madina Hasan, Mortaza Doulaty, Oscar Saz, Thomas Hain, University of Sheffield, UK*

## Sun-S&T-6: Show & Tell Session 6

Chairs: Nicolas Scheffer, Shiva Sundaram

Room: Market Street Foyer, Time 10:00-12:00, Sunday, 11 Sept. 2016

---

- PAGE 2348  
Sun-S&T-6-1  
10:00
- A Low Cost Desktop Robot and Tele-Presence Device for Interactive Speech Research**  
*Michael C. Brady, Tufts University, USA*
- PAGE 2350  
Sun-S&T-6-2  
10:00
- Silent-Speech Command Word Recognition Using Electro-Optical Stomatography**  
*Simon Stone, Peter Birkholz, Technische Universität Dresden, Germany*
- PAGE 2352  
Sun-S&T-6-3  
10:00
- An Engine for Online Video Search in Large Archives of the Holocaust Testimonies**  
*Petr Stanislav, Jan Švec, Pavel Ircing, University of West Bohemia, Czech Republic*

## Sun-O-7-1 : Robustness in Speech Processing

Chairs: Michael Seltzer, Ozlem Kalinli

Room: Grand Ballroom A, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2354  
Sun-O-7-1-1  
13:30
- Data Selection by Sequence Summarizing Neural Network in Mismatch Condition Training**  
*Kateřina Źmoliková<sup>1</sup>, Martin Karafiat<sup>1</sup>, Karel Vesely<sup>1</sup>, Marc Delcroix<sup>2</sup>, Shinji Watanabe<sup>3</sup>, Lukáš Burget<sup>1</sup>, Jan Černocký<sup>1</sup>*  
<sup>1</sup>Brno University of Technology, Czech Republic; <sup>2</sup>NTT, Japan; <sup>3</sup>MERL, USA
- PAGE 2359  
Sun-O-7-1-2  
13:50
- Incorporating a Generative Front-End Layer to Deep Neural Network for Noise Robust Automatic Speech Recognition**  
*Souvik Kundu<sup>1</sup>, Khe Chai Sim<sup>1</sup>, Mark J.F. Gales<sup>2</sup>*  
<sup>1</sup>NUS, Singapore; <sup>2</sup>University of Cambridge, UK
- PAGE 2364  
Sun-O-7-1-3  
14:10
- Robust Speech Recognition Using Generalized Distillation Framework**  
*Konstantin Markov<sup>1</sup>, Tomoko Matsui<sup>2</sup>*  
<sup>1</sup>University of Aizu, Japan; <sup>2</sup>ISM, Japan
- PAGE 2369  
Sun-O-7-1-4  
14:30
- Adversarial Multi-Task Learning of Deep Neural Networks for Robust Speech Recognition**  
*Yusuke Shinohara, Toshiba, Japan*
- PAGE 2373  
Sun-O-7-1-5  
14:50
- The Use of Locally Normalized Cepstral Coefficients (LNCC) to Improve Speaker Recognition Accuracy in Highly Reverberant Rooms**  
*Víctor Poblete<sup>1</sup>, Juan Pablo Escudero<sup>1</sup>, Josué Fredes<sup>2</sup>, José Novoa<sup>2</sup>, Richard M. Stern<sup>3</sup>, Simon King<sup>4</sup>, Néstor Becerra Yoma<sup>2</sup>*  
<sup>1</sup>Universidad Austral de Chile, Chile; <sup>2</sup>Universidad de Chile, Chile; <sup>3</sup>Carnegie Mellon University, USA; <sup>4</sup>University of Edinburgh, UK
- PAGE 2378  
Sun-O-7-1-6  
15:10
- Two-Stage Data Augmentation for Low-Resourced Speech Recognition**  
*William Hartmann, Tim Ng, Roger Hsiao, Stavros Tsakalidis, Richard Schwartz, Raytheon BBN Technologies, USA*

## Sun-O-7-2 : Special Session: Interspeech 2016 Computational Paralinguistics Challenge (ComParE): Deception, Sincerity & Native Language

Chairs: Björn Schuller, Stefan Steidl, Anton Batliner, Julia Hirschberg, Judee Burgoon, Eduardo Coutinho

Room: Grand Ballroom BC, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- N/A  
Sun-O-7-2-1  
13:30
- The Native Language Sub-Challenge: The Data**  
*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA
- PAGE 2383  
Sun-O-7-2-2  
13:40
- Native Language Identification Using Spectral and Source-Based Features**  
*Avni Rajpal<sup>1</sup>, Tanvina B. Patel<sup>1</sup>, Hardik B. Sailor<sup>1</sup>, Maulik C. Madhavi<sup>1</sup>, Hemant A. Patil<sup>1</sup>, Hiroya Fujisaki<sup>2</sup>*  
<sup>1</sup>DA-IICT, India; <sup>2</sup>University of Tokyo, Japan
- PAGE 2388  
Sun-O-7-2-3  
13:50
- Accent Identification by Combining Deep Neural Networks and Recurrent Neural Networks Trained on Long and Short Term Features**  
*Yishan Jiao, Ming Tu, Visar Berisha, Julie Liss, Arizona State University, USA*
- PAGE 2393  
Sun-O-7-2-4  
14:00
- Convolutional Neural Networks with Data Augmentation for Classifying Speakers' Native Language**  
*Gil Keren, Jun Deng, Jouni Pohjalainen, Björn Schuller, Universität Passau, Germany*
- PAGE 2398  
Sun-O-7-2-5  
14:10
- Native Language Detection Using the I-Vector Framework**  
*Mohammed Senoussaoui<sup>1</sup>, Patrick Cardinal<sup>1</sup>, Najim Dehak<sup>2</sup>, Alessandro L. Koerich<sup>1</sup>*  
<sup>1</sup>École de Technologie Supérieure, Canada; <sup>2</sup>Johns Hopkins University, USA

PAGE 2403  
Sun-O-7-2-6  
14:20

**Within-Speaker Features for Native Language Recognition in the Interspeech 2016 Computational Paralinguistics Challenge**

*Mark Huckvale, University College London, UK*

PAGE 2408  
Sun-O-7-2-7  
14:30

**Multimodal Fusion of Multirate Acoustic, Prosodic, and Lexical Speaker Characteristics for Native Language Identification**

*Prashanth Gurunath Shivakumar, Sandeep Nallan Chakravarthula, Panayiotis Georgiou, University of Southern California, USA*

PAGE 2413  
Sun-O-7-2-8  
14:40

**Exploiting Phone Log-Likelihood Ratio Features for the Detection of the Native Language of Non-Native English Speakers**

*Alberto Abad, Eugénio Ribeiro, Fábio Kepler, Ramon Astudillo, Isabel Trancoso, INESC-ID Lisboa, Portugal*

PAGE 2418  
Sun-O-7-2-9  
14:50

**Determining Native Language and Deception Using Phonetic Features and Classifier Combination**

*Gábor Gosztolya<sup>1</sup>, Tamás Grósz<sup>1</sup>, Róbert Busa-Fekete<sup>2</sup>, László Tóth<sup>3</sup>*  
*<sup>1</sup>University of Szeged, Hungary; <sup>2</sup>Universität Paderborn, Germany; <sup>3</sup>MTA-SZTE RGAI, Hungary*

N/A  
Sun-O-7-2-10  
15:00

**The INTERSPEECH 2016 Computational Paralinguistics Challenge: A Summary of Results**

*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA*

[Search]

*Sun-O-7-2 continues next page...*

N/A  
Sun-O-7-2-11  
15:10

**Discussion**

*Björn Schuller<sup>1</sup>, Stefan Steidl<sup>2</sup>, Anton Batliner<sup>3</sup>, Julia Hirschberg<sup>4</sup>, Judee K. Burgoon<sup>5</sup>, Alice Baird<sup>4</sup>, Aaron Elkins<sup>5</sup>, Yue Zhang<sup>1</sup>, Eduardo Coutinho<sup>1</sup>, Keelan Evanini<sup>6</sup>*  
*<sup>1</sup>Imperial College London, UK; <sup>2</sup>FAU Erlangen-Nürnberg, Germany; <sup>3</sup>Universität Passau, Germany; <sup>4</sup>Columbia University, USA; <sup>5</sup>University of Arizona, USA; <sup>6</sup>Educational Testing Service, USA*

[Search]

## Sun-O-7-3: Acoustic and Articulatory Phonetics

Chairs: Francisco Lacerda, Mary Beckman

Room: Bayview A, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2423  
Sun-O-7-3-1  
13:30      **A Preliminary Ultrasound Study of Nasal and Lateral Coronals in Arrernte**  
*Marija Tabain<sup>1</sup>, Richard Beare<sup>2</sup>*  
<sup>1</sup>La Trobe University, Australia; <sup>2</sup>Monash University, Australia
- PAGE 2428  
Sun-O-7-3-2  
13:50      **Illustrating the Production of the International Phonetic Alphabet Sounds Using Fast Real-Time Magnetic Resonance Imaging**  
*Asterios Toutios<sup>1</sup>, Sajan Goud Lingala<sup>1</sup>, Colin Vaz<sup>1</sup>, Jangwon Kim<sup>1</sup>, John Esling<sup>2</sup>, Patricia Keating<sup>3</sup>, Matthew Gordon<sup>4</sup>, Dani Byrd<sup>1</sup>, Louis Goldstein<sup>1</sup>, Krishna S. Nayak<sup>1</sup>, Shrikanth S. Narayanan<sup>1</sup>*  
<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of Victoria, Canada; <sup>3</sup>University of California at Los Angeles, USA; <sup>4</sup>University of California at Santa Barbara, USA
- PAGE 2433  
Sun-O-7-3-3  
14:10      **Marginal Contrast Among Romanian Vowels: Evidence from ASR and Functional Load**  
*Margaret E.L. Renwick<sup>1</sup>, Ioana Vasilescu<sup>2</sup>, Camille Dutrey<sup>3</sup>, Lori Lamel<sup>2</sup>, Bianca Vieru<sup>4</sup>*  
<sup>1</sup>University of Georgia, USA; <sup>2</sup>LIMSI, France; <sup>3</sup>LNE, France; <sup>4</sup>Vocapia Research, France
- PAGE 2438  
Sun-O-7-3-4  
14:30      **Effects of Subglottal-Coupling and Interdental-Space on Formant Trajectories During Front-to-Back Vowel Transitions in Chinese**  
*Shuanglin Fan, Kiyoshi Honda, Jianwu Dang, Hui Feng, Tianjin University, China*
- PAGE 2443  
Sun-O-7-3-5  
14:50      **Perceptual Lateralization of Coda Rhotic Production in Puerto Rican Spanish**  
*Mairym Lloréns Monteserín, Shrikanth S. Narayanan, Louis Goldstein, University of Southern California, USA*
- PAGE 2448  
Sun-O-7-3-6  
15:10      **Interaction Between Lexical Tone and Intonation: An EMA Study**  
*Hao Yi, Sam Tilsen, Cornell University, USA*

## Sun-O-7-4: Speech Synthesis Oral I: Neural Networks

Chairs: Heiga Zen, Zhen-Hua Ling

Room: Bayview B, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2453  
Sun-O-7-4-1  
13:30      **Deep Bidirectional LSTM Modeling of Timbre and Prosody for Emotional Voice Conversion**  
*Huaiping Ming<sup>1</sup>, Dongyan Huang<sup>1</sup>, Lei Xie<sup>2</sup>, Jie Wu<sup>2</sup>, Minghui Dong<sup>1</sup>, Haizhou Li<sup>1</sup>*  
<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>Northwestern Polytechnical University, China
- PAGE 2458  
Sun-O-7-4-2  
13:50      **Visual Speech Synthesis Using Dynamic Visemes, Contextual Features and DNNs**  
*Ausdang Thangthai, Ben Milner, Sarah Taylor, University of East Anglia, UK*
- PAGE 2463  
Sun-O-7-4-3  
14:10      **A Template-Based Approach for Speech Synthesis Intonation Generation Using LSTMs**  
*Srikanth Ronanki, Gustav Eje Henter, Zhizheng Wu, Simon King, University of Edinburgh, UK*
- PAGE 2468  
Sun-O-7-4-4  
14:30      **Multi-Language Multi-Speaker Acoustic Modeling for LSTM-RNN Based Statistical Parametric Speech Synthesis**  
*Bo Li<sup>1</sup>, Heiga Zen<sup>2</sup>*  
<sup>1</sup>Google, USA; <sup>2</sup>Google, UK
- PAGE 2473  
Sun-O-7-4-5  
14:50      **GlottDNN — A Full-Band Glottal Vocoder for Statistical Parametric Speech Synthesis**  
*Manu Airaksinen<sup>1</sup>, Bajibabu Bollepalli<sup>1</sup>, Lauri Juvela<sup>1</sup>, Zhizheng Wu<sup>2</sup>, Simon King<sup>2</sup>, Paavo Alku<sup>1</sup>*  
<sup>1</sup>Aalto University, Finland; <sup>2</sup>University of Edinburgh, UK
- PAGE 2478  
Sun-O-7-4-6  
15:10      **Singing Voice Synthesis Based on Deep Neural Networks**  
*Masanari Nishimura, Kei Hashimoto, Keiichiro Oura, Yoshihiko Nankaku, Keiichi Tokuda, Nagoya Institute of Technology, Japan*

## Sun-O-7-5: Speech Quality & Intelligibility

Chairs: Yannis Stylianou, Sebastian Möller

Room: Seacliff BCD, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2483  
Sun-O-7-5-1  
13:30      **Blind Recovery of Perceptual Models in Distributed Speech and Audio Coding**  
*Tom Bäckström, Florin Ghido, Johannes Fischer, FAU Erlangen-Nürnberg, Germany*
- PAGE 2488  
Sun-O-7-5-2  
13:50      **Glimpse-Based Metrics for Predicting Speech Intelligibility in Additive Noise Conditions**  
*Yan Tang<sup>1</sup>, Martin Cooke<sup>2</sup>*  
*<sup>1</sup>University of Salford, UK; <sup>2</sup>Ikerbasque, Spain*
- PAGE 2493  
Sun-O-7-5-3  
14:10      **Analyzing the Relation Between Overall Quality and the Quality of Individual Phases in a Telephone Conversation**  
*Friedemann Köster, Sebastian Möller, T-Labs, Germany*
- PAGE 2498  
Sun-O-7-5-4  
14:30      **Intelligibility Enhancement at the Receiving End of the Speech Transmission System — Effects of Far-End Noise Reduction**  
*Emma Jokinen, Paavo Alku, Aalto University, Finland*
- PAGE 2503  
Sun-O-7-5-5  
14:50      **Intelligibility of Disordered Speech: Global and Detailed Scores**  
*Mario Ganzeboom, Marjoke Bakker, Catia Cucchiari, Helmer Strik, Radboud Universiteit Nijmegen, The Netherlands*
- PAGE 2508  
Sun-O-7-5-6  
15:10      **Modulation Enhancement of Temporal Envelopes for Increasing Speech Intelligibility in Noise**  
*Maria Koutsogiannaki, Yannis Stylianou, University of Crete, Greece*

## Sun-O-7-6: Speech Translation and Metadata for Linguistic/Discourse Structure

Chairs: Tanja Schultz, Laurent Besacier

Room: Seacliff A, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2513  
Sun-O-7-6-1  
13:30      **Dynamic Transcription for Low-Latency Speech Translation**  
*Jan Niehues, Thai Son Nguyen, Eunah Cho, Thanh-Le Ha, Kevin Kilgour, Markus Müller, Matthias Sperber, Sebastian Stüker, Alex Waibel, KIT, Germany*
- PAGE 2518  
Sun-O-7-6-2  
13:50      **Learning a Translation Model from Word Lattices**  
*Oliver Adams<sup>1</sup>, Graham Neubig<sup>2</sup>, Trevor Cohn<sup>1</sup>, Steven Bird<sup>1</sup>*  
*<sup>1</sup>University of Melbourne, Australia; <sup>2</sup>NAIST, Japan*
- PAGE 2523  
Sun-O-7-6-3  
14:10      **Disfluency Detection Using a Bidirectional LSTM**  
*Vicky Zayats, Mari Ostendorf, Hannaneh Hajishirzi, University of Washington, USA*
- PAGE 2528  
Sun-O-7-6-4  
14:30      **Sentence Boundary Detection Based on Parallel Lexical and Acoustic Models**  
*Xiaoyin Che, Sheng Luo, Haojin Yang, Christoph Meinel, Universität Potsdam, Germany*
- PAGE 2533  
Sun-O-7-6-5  
14:50      **Transferring Emphasis in Speech Translation Using Hard-Attentional Neural Network Models**  
*Quoc Truong Do, Sakriani Sakti, Graham Neubig, Satoshi Nakamura, NAIST, Japan*
- PAGE 2538  
Sun-O-7-6-6  
15:10      **Better Evaluation of ASR in Speech Translation Context Using Word Embeddings**  
*Ngoc-Tien Le, Christophe Servan, Benjamin Lecouteux, Laurent Besacier, LIG (UMR 5217), France*



## Sun-P-7-1 : Speech Coding and Audio Processing for Noise Reduction

Chairs: Tom Bäckström, Dayana Ribas

Room: Pacific Concourse - Poster A, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

PAGE 2543  
Sun-P-7-1-1  
13:30

### **Entropy Coding of Spectral Envelopes for Speech and Audio Coding Using Distribution Quantization**

*Srikanth Korse<sup>1</sup>, Tobias Jähnel<sup>2</sup>, Tom Bäckström<sup>1</sup>*  
*<sup>1</sup>Fraunhofer IIS, Germany; <sup>2</sup>FAU Erlangen-Nürnberg, Germany*

PAGE 2548  
Sun-P-7-1-2  
13:30

### **An Objective Evaluation Methodology for Blind Bandwidth Extension**

*Stéphane Villette, Sen Li, Pravin Ramadas, Daniel J. Sinder, Qualcomm Technologies, USA*

PAGE 2553  
Sun-P-7-1-3  
13:30

### **EVS Channel Aware Mode Robustness to Frame Erasures**

*Anssi Rämö, Antti Kurittu, Henri Toukoma, Nokia, Finland*

PAGE 2558  
Sun-P-7-1-4  
13:30

### **An Interaural Magnification Algorithm for Enhancement of Naturally-Occurring Level Differences**

*Shadi Pirhosseinloo, Kostas Kokkinakis, University of Kansas, USA*

PAGE 2562  
Sun-P-7-1-5  
13:30

### **Probabilistic Spatial Filter Estimation for Signal Enhancement in Multi-Channel Automatic Speech Recognition**

*Hendrik Kayser<sup>1</sup>, Niko Moritz<sup>2</sup>, Jörn Anemüller<sup>1</sup>*  
*<sup>1</sup>Carl von Ossietzky Universität Oldenburg, Germany; <sup>2</sup>Fraunhofer IDMT, Germany*

PAGE 2567  
Sun-P-7-1-6  
13:30

### **Improved *a priori* SAP Estimator in Complex Noisy Environment for Dual Channel Microphone System**

*Youna Ji, Young-cheol Park, Yonsei University, Korea*

PAGE 2572  
Sun-P-7-1-7  
13:30

### **A Spectral Modulation Sensitivity Weighted Pre-Emphasis Filter for Active Noise Control System**

*Kah-Meng Cheong<sup>1</sup>, Yuh-Yuan Wang<sup>2</sup>, Tai-Shih Chi<sup>1</sup>*  
*<sup>1</sup>National Chiao Tung University, Taiwan; <sup>2</sup>TANGENT Microelectromechanics, Taiwan*

PAGE 2577  
Sun-P-7-1-8  
13:30

### **Semi-Coupled Dictionary Based Automatic Bandwidth Extension Approach for Enhancing Children's ASR**

*Ganji Sreeram, Rohit Sinha, IIT Guwahati, India*



## Sun-P-7-2 : Special Session: Speech, Audio, and Language Processing Techniques Applied to Bird and Animal Vocalizations

Chairs: Naomi Harte, Peter Jančovič, Karl-L. Schuchmann

Room: Pacific Concourse - Poster B, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2582  
Sun-P-7-2-1  
13:30
- Bird Song Synthesis Based on Hidden Markov Models**  
*Jordi Bonada<sup>1</sup>, Robert Lachlan<sup>2</sup>, Merlijn Blaauw<sup>1</sup>*  
<sup>1</sup>Universitat Pompeu Fabra, Spain; <sup>2</sup>Queen Mary University of London, UK
- PAGE 2587  
Sun-P-7-2-2  
13:30
- Noise-Robust Hidden Markov Models for Limited Training Data for Within-Species Bird Phrase Classification**  
*Kantapon Kaewtip, Charles Taylor, Abeer Alwan, University of California at Los Angeles, USA*
- PAGE 2592  
Sun-P-7-2-3  
13:30
- A Framework for Automated Marmoset Vocalization Detection and Classification**  
*Alan Wisler<sup>1</sup>, Laura J. Brattain<sup>2</sup>, Rogier Landman<sup>3</sup>, Thomas F. Quatieri<sup>2</sup>*  
<sup>1</sup>Arizona State University, USA; <sup>2</sup>MIT Lincoln Laboratory, USA; <sup>3</sup>Broad Institute of MIT and Harvard, USA
- PAGE 2597  
Sun-P-7-2-4  
13:30
- Call Alternation Between Specific Pairs of Male Frogs Revealed by a Sound-Imaging Method in Their Natural Habitat**  
*Ikkyu Aihara<sup>1</sup>, Takeshi Mizumoto<sup>2</sup>, Hiromitsu Awano<sup>3</sup>, Hiroshi G. Okuno<sup>4</sup>*  
<sup>1</sup>University of Tsukuba, Japan; <sup>2</sup>Honda Research Institute Japan, Japan; <sup>3</sup>Kyoto University, Japan; <sup>4</sup>Waseda University, Japan
- PAGE 2602  
Sun-P-7-2-5  
13:30
- Sinusoidal Modelling for Ecoacoustics**  
*Patrice Guyot<sup>1</sup>, Alice Eldridge<sup>1</sup>, Ying Chen Eyre-Walker<sup>1</sup>, Alison Johnston<sup>2</sup>, Thomas Pellegrini<sup>3</sup>, Mika Peck<sup>1</sup>*  
<sup>1</sup>University of Sussex, UK; <sup>2</sup>British Trust for Ornithology, UK; <sup>3</sup>IRIT, France
- 

- PAGE 2607  
Sun-P-7-2-6  
13:30
- Individual Identity in Songbirds: Signal Representations and Metric Learning for Locating the Information in Complex Corvid Calls**  
*Dan Stowell<sup>1</sup>, Veronica Morfi<sup>1</sup>, Lisa F. Gill<sup>2</sup>*  
<sup>1</sup>Queen Mary University of London, UK; <sup>2</sup>MPI for Ornithology, Germany
- PAGE 2612  
Sun-P-7-2-7  
13:30
- Recognition of Multiple Bird Species Based on Penalised Maximum Likelihood and HMM-Based Modelling of Individual Vocalisation Elements**  
*Peter Jančovič, Münevver Köküer, University of Birmingham, UK*
- PAGE 2617  
Sun-P-7-2-8  
13:30
- Cost Effective Acoustic Monitoring of Bird Species**  
*Ciira wa Maina, DeKUT, Kenya*
- PAGE 2621  
Sun-P-7-2-9  
13:30
- Feature Learning and Automatic Segmentation for Dolphin Communication Analysis**  
*Daniel Kohlsdorf<sup>1</sup>, Denise Herzing<sup>2</sup>, Thad Starner<sup>1</sup>*  
<sup>1</sup>Georgia Institute of Technology, USA; <sup>2</sup>Wild Dolphin Project, USA
- PAGE 2626  
Sun-P-7-2-10  
13:30
- Localizing Bird Songs Using an Open Source Robot Audition System with a Microphone Array**  
*Reiji Suzuki<sup>1</sup>, Shiho Matsubayashi<sup>1</sup>, Kazuhiro Nakadai<sup>2</sup>, Hiroshi G. Okuno<sup>3</sup>*  
<sup>1</sup>Nagoya University, Japan; <sup>2</sup>Honda Research Institute Japan, Japan; <sup>3</sup>Waseda University, Japan
- PAGE 2631  
Sun-P-7-2-11  
13:30
- Robust Detection of Multiple Bioacoustic Events with Repetitive Structures**  
*Frank Kurth, Fraunhofer FKIE, Germany*
- PAGE 2636  
Sun-P-7-2-12  
13:30
- A Real-Time Parametric General-Purpose Mammalian Vocal Synthesiser**  
*Roger K. Moore, University of Sheffield, UK*
- PAGE 2641  
Sun-P-7-2-13  
13:30
- YIN-Bird: Improved Pitch Tracking for Bird Vocalisations**  
*Colm O'Reilly, Nicola M. Marples, David J. Kelly, Naomi Harte, Trinity College Dublin, Ireland*

## Sun-P-7-3: Learning, Education and Different Speech

Chair: Thomas Pellegrini

Room: Pacific Concourse - Poster C, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2646  
Sun-P-7-3-1  
13:30      **Mispronunciation Detection Leveraging Maximum Performance Criterion Training of Acoustic Models and Decision Functions**  
*Yao-Chi Hsu, Ming-Han Yang, Hsiao-Tsung Hung, Berlin Chen, National Taiwan Normal University, Taiwan*
- PAGE 2651  
Sun-P-7-3-2  
13:30      **Using Clinician Annotations to Improve Automatic Speech Recognition of Stuttered Speech**  
*Peter A. Heeman<sup>1</sup>, Rebecca Lunsford<sup>1</sup>, Andy McMillin<sup>2</sup>, J. Scott Yaruss<sup>3</sup>*  
*<sup>1</sup>BioSpeech, USA; <sup>2</sup>Oregon Health & Science University, USA; <sup>3</sup>University of Pittsburgh, USA*
- PAGE 2656  
Sun-P-7-3-3  
13:30      **Deep Neural Networks for Voice Quality Assessment Based on the GRBAS Scale**  
*Simin Xie<sup>1</sup>, Nan Yan<sup>2</sup>, Ping Yu<sup>3</sup>, Manwa L. Ng<sup>4</sup>, Lan Wang<sup>2</sup>, Zhuanzhuan Ji<sup>2</sup>*  
*<sup>1</sup>Wuhan University of Technology, China; <sup>2</sup>Chinese Academy of Sciences, China; <sup>3</sup>PLA General Hospital, China; <sup>4</sup>University of Hong Kong, China*
- PAGE 2661  
Sun-P-7-3-4  
13:30      **Automated Screening of Speech Development Issues in Children by Identifying Phonological Error Patterns**  
*Lauren Ward<sup>1</sup>, Alessandro Stefani<sup>1</sup>, Daniel Smith<sup>1</sup>, Andreas Duenser<sup>1</sup>, Jill Freyne<sup>1</sup>, Barbara Dodd<sup>2</sup>, Angela Morgan<sup>3</sup>*  
*<sup>1</sup>CSIRO, Australia; <sup>2</sup>University of Melbourne, Australia; <sup>3</sup>MCRI, Australia*
- PAGE 2666  
Sun-P-7-3-5  
13:30      **Automatic Pronunciation Evaluation of Non-Native Mandarin Tone by Using Multi-Level Confidence Measures**  
*Ju Lin, Yanlu Xie, Jinsong Zhang, BLCU, China*

- PAGE 2671  
Sun-P-7-3-6  
13:30      **Dysarthric Speech Recognition Using Kullback-Leibler Divergence-Based Hidden Markov Model**  
*Myungjong Kim<sup>1</sup>, Jun Wang<sup>1</sup>, Hoirin Kim<sup>2</sup>*  
*<sup>1</sup>University of Texas at Dallas, USA; <sup>2</sup>KAIST, Korea*
- PAGE 2676  
Sun-P-7-3-7  
13:30      **Detection of Total Syllables and Canonical Syllables in Infant Vocalizations**  
*Anne S. Warlaumont<sup>1</sup>, Heather L. Ramsdell-Hudock<sup>2</sup>*  
*<sup>1</sup>University of California at Merced, USA; <sup>2</sup>Idaho State University, USA*
- PAGE 2681  
Sun-P-7-3-8  
13:30      **Improving Automatic Recognition of Aphasic Speech with AphasiaBank**  
*Duc Le, Emily Mower Provost, University of Michigan, USA*
- PAGE 2686  
Sun-P-7-3-9  
13:30      **Pronunciation Assessment of Japanese Learners of French with GOP Scores and Phonetic Information**  
*Vincent Laborde<sup>1</sup>, Thomas Pellegrini<sup>1</sup>, Lionel Fontan<sup>1</sup>, Julie Mauclair<sup>1</sup>, Halima Sahraoui<sup>2</sup>, Jérôme Farinas<sup>1</sup>*  
*<sup>1</sup>IRIT, France; <sup>2</sup>Octogone-Lordat (EA4156), France*
- PAGE 2691  
Sun-P-7-3-10  
13:30      **Pronunciation Error Detection for New Language Learners**  
*Sean Robertson, Cosmin Munteanu, Gerald Penn, University of Toronto, Canada*
- PAGE 2696  
Sun-P-7-3-11  
13:30      **L2 English Rhythm in Read Speech by Chinese Students**  
*Hongwei Ding<sup>1</sup>, Xinpeng Xu<sup>2</sup>*  
*<sup>1</sup>Shanghai Jiao Tong University, China; <sup>2</sup>Shanghai East High School, China*

## Sun-P-7-4: Dialogue Systems and Analysis of Dialogue

Chair: Joakim Gustafson

Room: Pacific Concourse - Poster D, Time 13:30-15:30, Sunday, 11 Sept. 2016

---

- PAGE 2701  
Sun-P-7-4-1  
13:30      **Improving the Probabilistic Framework for Representing Dialogue Systems with User Response Model**  
*Miao Li, Zhipeng Chen, Ji Wu, Tsinghua University, China*
- PAGE 2706  
Sun-P-7-4-2  
13:30      **Dialogue Session Segmentation by Embedding-Enhanced TextTiling**  
*Yiping Song<sup>1</sup>, Lili Mou<sup>1</sup>, Rui Yan<sup>1</sup>, Li Yi<sup>2</sup>, Zinan Zhu<sup>2</sup>, Xiaohua Hu<sup>2</sup>, Ming Zhang<sup>1</sup>*  
*<sup>1</sup>Peking University, China; <sup>2</sup>Central China Normal University, China*
- PAGE 2711  
Sun-P-7-4-3  
13:30      **Target-Based State and Tracking Algorithm for Spoken Dialogue System**  
*Miao Li, Zhiyang He, Ji Wu, Tsinghua University, China*
- PAGE 2716  
Sun-P-7-4-4  
13:30      **Neural Attention Models for Sequence Classification: Analysis and Application to Key Term Extraction and Dialogue Act Detection**  
*Sheng-syun Shen, Hung-Yi Lee, National Taiwan University, Taiwan*
- PAGE 2721  
Sun-P-7-4-5  
13:30      **Objective Language Feature Analysis in Children with Neurodevelopmental Disorders During Autism Assessment**  
*Manoj Kumar<sup>1</sup>, Rahul Gupta<sup>1</sup>, Daniel Bone<sup>1</sup>, Nikolaos Malandrakis<sup>1</sup>, Somer Bishop<sup>2</sup>, Shrikanth S. Narayanan<sup>1</sup>*  
*<sup>1</sup>University of Southern California, USA; <sup>2</sup>University of California at San Francisco, USA*
- PAGE 2726  
Sun-P-7-4-6  
13:30      **Improving Generalisation to New Speakers in Spoken Dialogue State Tracking**  
*Iñigo Casanueva, Thomas Hain, Phil Green, University of Sheffield, UK*
- PAGE 2731  
Sun-P-7-4-7  
13:30      **Towards Machine Comprehension of Spoken Content: Initial TOEFL Listening Comprehension Test by Machine**  
*Bo-Hsiang Tseng, Sheng-syun Shen, Hung-Yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan*

## Sun-O-8-1: Topics in Speech Recognition

Chairs: Karen Livescu, Eric McDermott

Room: Grand Ballroom A, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2736  
Sun-O-8-1-1  
16:00      **How Neural Network Depth Compensates for HMM Conditional Independence Assumptions in DNN-HMM Acoustic Models**  
*Suman Ravuri<sup>1</sup>, Steven Wegmann<sup>2</sup>*  
*<sup>1</sup>ICSI, USA; <sup>2</sup>Semantic Machines, USA*
- PAGE 2741  
Sun-O-8-1-2  
16:20      **Jointly Learning to Locate and Classify Words Using Convolutional Networks**  
*Dimitri Palaz<sup>1</sup>, Gabriel Synnaeve<sup>2</sup>, Ronan Collobert<sup>1</sup>*  
*<sup>1</sup>Facebook, USA; <sup>2</sup>Facebook, France*
- PAGE 2746  
Sun-O-8-1-3  
16:40      **On the Efficient Representation and Execution of Deep Acoustic Models**  
*Raziel Alvarez, Rohit Prabhavalkar, Anton Bakhtin, Google, USA*
- PAGE 2751  
Sun-O-8-1-4  
17:00      **Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI**  
*Daniel Povey<sup>1</sup>, Vijayaditya Peddinti<sup>1</sup>, Daniel Galvez<sup>2</sup>, Pegah Ghahremani<sup>1</sup>, Vimal Manohar<sup>1</sup>, Xingyu Na<sup>3</sup>, Yiming Wang<sup>1</sup>, Sanjeev Khudanpur<sup>1</sup>*  
*<sup>1</sup>Johns Hopkins University, USA; <sup>2</sup>Cornell University, USA; <sup>3</sup>Lele Innovation & Intelligence Technology, China*
- PAGE 2756  
Sun-O-8-1-5  
17:20      **Virtual Adversarial Training Applied to Neural Higher-Order Factors for Phone Classification**  
*Martin Ratajczak<sup>1</sup>, Sebastian Tschiatschek<sup>2</sup>, Franz Pernkopf<sup>1</sup>*  
*<sup>1</sup>Technische Universität Graz, Austria; <sup>2</sup>ETH Zürich, Switzerland*
- PAGE 2761  
Sun-O-8-1-6  
17:40      **Sequence Student-Teacher Training of Deep Neural Networks**  
*Jeremy H.M. Wong, Mark J.F. Gales, University of Cambridge, UK*

## Sun-O-8-2 : Special Session: Realism in Robust Speech Processing

Chairs: Dayana Ribas, Emmanuel Vincent, John Hansen

Room: Grand Ballroom BC, Time 16:00–18:00, Sunday, 11 Sept. 2016

---

PAGE 2766 Sun-O-8-2-1 16:00	<b>Robustness in Speech, Speaker, and Language Recognition: “You’ve Got to Know Your Limitations”</b> <i>John H.L. Hansen, Hynek Bořil, University of Texas at Dallas, USA</i>
PAGE 2771 Sun-O-8-2-2 16:15	<b>The Use of Read versus Conversational Lombard Speech in Spectral Tilt Modeling for Intelligibility Enhancement in Near-End Noise Conditions</b> <i>Emma Jokinen, Ulpu Remes, Paavo Alku, Aalto University, Finland</i>
PAGE 2776 Sun-O-8-2-3 16:30	<b>Corpora for the Evaluation of Robust Speaker Recognition Systems</b> <i>Douglas E. Sturim, Pedro A. Torres-Carrasquillo, Joseph P. Campbell, MIT Lincoln Laboratory, USA</i>
PAGE 2781 Sun-O-8-2-4 16:45	<b>A French Corpus for Distant-Microphone Speech Processing in Real Homes</b> <i>Nancy Bertin<sup>1</sup>, Ewen Camberlein<sup>1</sup>, Emmanuel Vincent<sup>2</sup>, Romain Lebarbenchon<sup>1</sup>, Stéphane Peillon<sup>3</sup>, Éric Lamande<sup>3</sup>, Sunit Sivasankaran<sup>2</sup>, Frédéric Bimbot<sup>1</sup>, Irina Illina<sup>4</sup>, Ariane Tom<sup>5</sup>, Sylvain Fleury<sup>5</sup>, Éric Jamet<sup>5</sup></i> <sup>1</sup> IRISA, France; <sup>2</sup> Inria, France; <sup>3</sup> VoiceBox Technologies, France; <sup>4</sup> LORIA, France; <sup>5</sup> CRPCC (EA 1285), France
PAGE 2786 Sun-O-8-2-5 17:00	<b>Realistic Multi-Microphone Data Simulation for Distant Speech Recognition</b> <i>Mirco Ravanelli, Piergiorgio Svaizer, Maurizio Omologo, FBK, Italy</i>
PAGE 2791 Sun-O-8-2-6 17:15	<b>Synthesis of Device-Independent Noise Corpora for Realistic ASR Evaluation</b> <i>Hannes Gamper, Mark R.P. Thomas, Lyle Corbin, Ivan Tashev, Microsoft, USA</i>

PAGE 2796 Sun-O-8-2-7 17:30	<b>Speaker Recognition Using Real vs Synthetic Parallel Data for DNN Channel Compensation</b> <i>Fred Richardson, Michael Brandstein, Jennifer Melot, Douglas Reynolds, MIT Lincoln Laboratory, USA</i>
N/A Sun-O-8-2-8 17:45	<b>Discussion</b> <i>Dayana Ribas<sup>1</sup>, Emmanuel Vincent<sup>2</sup>, John H.L. Hansen<sup>3</sup>, Emma Jokinen<sup>4</sup>, Mirco Ravanelli<sup>5</sup>, Hannes Gamper<sup>6</sup>, Fred Richardson<sup>7</sup></i> <sup>1</sup> CENATAV, Cuba; <sup>2</sup> Inria, France; <sup>3</sup> University of Texas at Dallas, USA; <sup>4</sup> Aalto University, Finland; <sup>5</sup> FBK, Italy; <sup>6</sup> Microsoft, USA; <sup>7</sup> MIT Lincoln Laboratory, USA

## Sun-O-8-3: Spoken Word Recognition

Chairs: Michael McAuliffe, Odette Scharenborg

Room: Bayview A, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2801  
Sun-O-8-3-1  
16:00      **Combining Data-Oriented and Process-Oriented Approaches to Modeling Reaction Time Data**  
*Louis ten Bosch, Lou Boves, M. Ernestus, Radboud Universiteit Nijmegen, The Netherlands*
- PAGE 2806  
Sun-O-8-3-2  
16:20      **Do Listeners Learn Better from Natural Speech?**  
*Michael McAuliffe<sup>1</sup>, Molly Babel<sup>2</sup>, Charlotte Vaughn<sup>3</sup>*  
*<sup>1</sup>McGill University, Canada; <sup>2</sup>University of British Columbia, Canada; <sup>3</sup>University of Oregon, USA*
- PAGE 2811  
Sun-O-8-3-3  
16:40      **Processing and Adaptation to Ambiguous Sounds during the Course of Perceptual Learning**  
*Polina Drozdova, Roeland van Hout, Odette Scharenborg, Radboud Universiteit Nijmegen, The Netherlands*
- PAGE 2816  
Sun-O-8-3-4  
17:00      **The Effect of Background Noise on the Activation of Phonological and Semantic Information During Spoken-Word Recognition**  
*Florian Hintz, Odette Scharenborg, Radboud Universiteit Nijmegen, The Netherlands*
- PAGE 2821  
Sun-O-8-3-5  
17:20      **Relationships Between Functional Load and Auditory Confusability Under Different Speech Environments**  
*Shinae Kang<sup>1</sup>, Clara Cohen<sup>2</sup>*  
*<sup>1</sup>Georgetown University, USA; <sup>2</sup>Pennsylvania State University, USA*
- PAGE 2826  
Sun-O-8-3-6  
17:40      **The Role of Pitch in Punjabi Word Identification**  
*Jasmeen Kanwal<sup>1</sup>, Amanda Ritchart<sup>2</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>University of California at San Diego, USA*

## Sun-O-8-4: Speech Synthesis Oral: High Level Linguistic Features

Chairs: Kishore Prahallad, Rob Clark

Room: Bayview B, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2831  
Sun-O-8-4-1  
16:00      **Improving TTS with Corpus-Specific Pronunciation Adaptation**  
*Marie Tahon, Raheel Qader, Gwénolé Lecorvé, Damien Lolive, IRISA, France*
- PAGE 2836  
Sun-O-8-4-2  
16:20      **Deep Bidirectional Long Short-Term Memory Recurrent Neural Networks for Grapheme-to-Phoneme Conversion Utilizing Complex Many-to-Many Alignments**  
*Amr El-Desoky Mousa, Björn Schuller, Universität Passau, Germany*
- PAGE 2841  
Sun-O-8-4-3  
16:40      **Predicting Pronunciations with Syllabification and Stress with Recurrent Neural Networks**  
*Daan van Esch, Mason Chua, Kanishka Rao, Google, USA*
- PAGE 2846  
Sun-O-8-4-4  
17:00      **Adaptive Latency for Part-of-Speech Tagging in Incremental Text-to-Speech Synthesis**  
*Maël Pouget, Olha Nahorna, Thomas Hueber, Gérard Bailly, GIPSA, France*
- PAGE 2851  
Sun-O-8-4-5  
17:20      **Redefining the Linguistic Context Feature Set for HMM and DNN TTS Through Position and Parsing**  
*Rasmus Dall<sup>1</sup>, Kei Hashimoto<sup>2</sup>, Keiichiro Oura<sup>2</sup>, Yoshihiko Nankaku<sup>2</sup>, Keiichi Tokuda<sup>2</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>Nagoya Institute of Technology, Japan*
- PAGE 2856  
Sun-O-8-4-6  
17:40      **Enhance the Word Vector with Prosodic Information for the Recurrent Neural Network Based TTS System**  
*Xin Wang, Shinji Takaki, Junichi Yamagishi, NII, Japan*

## Sun-O-8-5: Speech Enhancement

Chairs: Dimitrios Dimitriadis, Martin Cooke

Room: Seacliff BCD, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2861  
Sun-O-8-5-1  
16:00      **Local Sparsity Based Online Dictionary Learning for Environment-Adaptive Speech Enhancement with Nonnegative Matrix Factorization**  
*Kwang Myung Jeon, Hong Kook Kim, GIST, Korea*
- PAGE 2866  
Sun-O-8-5-2  
16:20      **Noise Aware and Combined Noise Models for Speech Denoising in Unknown Noise Conditions**  
*Pavlos Papadopoulos, Colin Vaz, Shrikanth S. Narayanan, University of Southern California, USA*
- PAGE 2870  
Sun-O-8-5-3  
16:40      **Causal Speech Enhancement Combining Data-Driven Learning and Suppression Rule Estimation**  
*Syedmahdad Mirsamadi<sup>1</sup>, Ivan Tashev<sup>2</sup>*  
*<sup>1</sup>University of Texas at Dallas, USA; <sup>2</sup>Microsoft, USA*
- PAGE 2875  
Sun-O-8-5-4  
17:00      **A Phase-Based Time-Frequency Masking for Multi-Channel Speech Enhancement in Domestic Environments**  
*Alessio Brutti<sup>1</sup>, Antigoni Tsiami<sup>2</sup>, Athanasios Katsamanis<sup>3</sup>, Petros Maragos<sup>2</sup>*  
*<sup>1</sup>FBK, Italy; <sup>2</sup>NTUA, Greece; <sup>3</sup>Athena RIC, Greece*
- PAGE 2880  
Sun-O-8-5-5  
17:20      **Generalizing Steady State Suppression for Enhanced Intelligibility Under Reverberation**  
*Petko N. Petkov, Yannis Stylianou, Toshiba Research Europe, UK*
- PAGE 2885  
Sun-O-8-5-6  
17:40      **Speech Intelligibility Prediction Based on the Envelope Power Spectrum Model with the Dynamic Compressive Gammachirp Auditory Filterbank**  
*Katsuhiko Yamamoto<sup>1</sup>, Toshio Irino<sup>1</sup>, Toshie Matsui<sup>1</sup>, Shoko Araki<sup>2</sup>, Keisuke Kinoshita<sup>2</sup>, Tomohiro Nakatani<sup>2</sup>*  
*<sup>1</sup>Wakayama University, Japan; <sup>2</sup>NTT, Japan*

## Sun-O-8-6: Dialogue: Backchannels and Turntaking

Chairs: Zofia Malisz, David Traum

Room: Seacliff A, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2890  
Sun-O-8-6-1  
16:00      **Prediction and Generation of Backchannel Form for Attentive Listening Systems**  
*Tatsuya Kawahara<sup>1</sup>, Takashi Yamaguchi<sup>1</sup>, Koji Inoue<sup>1</sup>, Katsuya Takanashi<sup>1</sup>, Nigel Ward<sup>2</sup>*  
*<sup>1</sup>Kyoto University, Japan; <sup>2</sup>University of Texas at El Paso, USA*
- PAGE 2895  
Sun-O-8-6-2  
16:20      **Measuring Turn-Taking Offsets in Human-Human Dialogues**  
*Rebecca Lunsford, Peter A. Heeman, Emma Rennie, Oregon Health & Science University, USA*
- PAGE 2900  
Sun-O-8-6-3  
16:40      **Using Past Speaker Behavior to Better Predict Turn Transitions**  
*Tomer Meshorer, Peter A. Heeman, Oregon Health & Science University, USA*
- PAGE 2905  
Sun-O-8-6-4  
17:00      **Quantitative Analysis of Backchannels Uttered by an Interviewer During Neuropsychological Tests**  
*Gérard Bailly<sup>1</sup>, Frédéric Elisei<sup>1</sup>, Alexandra Juphard<sup>2</sup>, Olivier Moreaud<sup>2</sup>*  
*<sup>1</sup>GIPSA, France; <sup>2</sup>CHU de Grenoble, France*
- PAGE 2910  
Sun-O-8-6-5  
17:20      **Predicting User Satisfaction from Turn-Taking in Spoken Conversations**  
*Shammur Absar Chowdhury, Evgeny A. Stepanov, Giuseppe Riccardi, Università di Trento, Italy*
- PAGE 2915  
Sun-O-8-6-6  
17:40      **Towards Building an Attentive Artificial Listener: On the Perception of Attentiveness in Feedback Utterances**  
*Catharine Oertel<sup>1</sup>, Joakim Gustafson<sup>1</sup>, Alan W. Black<sup>2</sup>*  
*<sup>1</sup>KTH, Sweden; <sup>2</sup>Carnegie Mellon University, USA*

## Sun-P-8-1 : Language Recognition

Chair: Honza Cernocky

Room: Pacific Concourse - Poster A, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

PAGE 2920  
Sun-P-8-1-1  
16:00

### Language Recognition via Sparse Coding

*Youngjune L. Gwon<sup>1</sup>, William M. Campbell<sup>1</sup>, Douglas E. Sturim<sup>1</sup>, H.T. Kung<sup>2</sup>*

<sup>1</sup>MIT Lincoln Laboratory, USA; <sup>2</sup>Harvard University, USA

PAGE 2925  
Sun-P-8-1-2  
16:00

### A Feature Normalisation Technique for PLLR Based Language Identification Systems

*Sarith Fernando, Vidhyasaharan Sethu, Eliathamby Ambikairajah, University of New South Wales, Australia*

PAGE 2930  
Sun-P-8-1-3  
16:00

### An Investigation of Deep Neural Network Architectures for Language Recognition in Indian Languages

*Mounika K.V., Sivanand Achanta, Lakshmi H. R., Suryakanth V. Gangashetty,*

*Anil Kumar Vuppala, IIT Hyderabad, India*

PAGE 2934  
Sun-P-8-1-4  
16:00

### Automatic Dialect Detection in Arabic Broadcast Speech

*Ahmed Ali<sup>1</sup>, Najim Dehak<sup>2</sup>, Patrick Cardinal<sup>3</sup>, Sameer Khurana<sup>1</sup>,*

*Sree Harsha Yella<sup>2</sup>, James Glass<sup>2</sup>, Peter Bell<sup>4</sup>, Steve Renals<sup>4</sup>*

<sup>1</sup>Hamad Bin Khalifa University, Qatar; <sup>2</sup>MIT, USA; <sup>3</sup>École de Technologie Supérieure, Canada; <sup>4</sup>University of Edinburgh, UK

PAGE 2939  
Sun-P-8-1-5  
16:00

### Combining Weak Tokenisers for Phonotactic Language Recognition in a Resource-Constrained Setting

*Raymond W.M. Ng, Bhusan Chettri, Thomas Hain, University of Sheffield, UK*

PAGE 2944  
Sun-P-8-1-6  
16:00

### End-to-End Language Identification Using Attention-Based Recurrent Neural Networks

*Wang Geng, Wenfu Wang, Yuanyuan Zhao, Xinyuan Cai, Bo Xu, Chinese Academy of Sciences, China*

PAGE 2949  
Sun-P-8-1-7  
16:00

### Enhancing Multilingual Recognition of Emotion in Speech by Language Identification

*Hesam Sagha<sup>1</sup>, Pavel Matějka<sup>2</sup>, Maryna Gavryukova<sup>1</sup>, Filip Povolny<sup>2</sup>, Erik Marchi<sup>1</sup>, Björn Schuller<sup>1</sup>*

<sup>1</sup>Universität Passau, Germany; <sup>2</sup>Phonexia Brno, Czech Republic



## Sun-P-8-2: Speech and Audio Segmentation and Classification

Chair: Martine Adda-Decker

Room: Pacific Concourse - Poster B, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 2954  
Sun-P-8-2-1  
16:00      **Deep Neural Network Bottleneck Features for Acoustic Event Recognition**  
*Seongkyu Mun<sup>1</sup>, Suwon Shon<sup>1</sup>, Wooil Kim<sup>2</sup>, Hanseok Ko<sup>1</sup>*  
*<sup>1</sup>Korea University, Korea; <sup>2</sup>Incheon National University, Korea*
- PAGE 2958  
Sun-P-8-2-2  
16:00      **Combining Energy and Cross-Entropy Analysis for Nuclear Segments Detection**  
*Antonio Origlia, Francesco Cutugno, Università di Napoli Federico II, Italy*
- PAGE 2963  
Sun-P-8-2-3  
16:00      **Anchored Speech Detection**  
*Roland Maas, Sree Hari Krishnan Parthasarathi, Brian King, Ruitong Huang, Björn Hoffmeister, Amazon.com, USA*
- PAGE 2968  
Sun-P-8-2-4  
16:00      **Towards Smart-Cars That Can Listen: Abnormal Acoustic Event Detection on the Road**  
*Mahesh Kumar Nandwana, Taufiq Hasan, Robert Bosch, USA*
- PAGE 2972  
Sun-P-8-2-5  
16:00      **Hierarchical Classification of Speaker and Background Noise and Estimation of SNR Using Sparse Representation**  
*K.V. Vijay Girish<sup>1</sup>, A.G. Ramakrishnan<sup>1</sup>, T.V. Ananthapadmanabha<sup>2</sup>*  
*<sup>1</sup>Indian Institute of Science, India; <sup>2</sup>Voice and Speech Systems, India*
- PAGE 2977  
Sun-P-8-2-6  
16:00      **Robust Sound Event Detection in Continuous Audio Environments**  
*Haomin Zhang<sup>1</sup>, Ian McLoughlin<sup>2</sup>, Yan Song<sup>1</sup>*  
*<sup>1</sup>USTC, China; <sup>2</sup>University of Kent, UK*
- PAGE 2982  
Sun-P-8-2-7  
16:00      **Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Recognition**  
*Naoya Takahashi<sup>1</sup>, Michael Gygli<sup>2</sup>, Beat Pfister<sup>2</sup>, Luc Van Gool<sup>2</sup>*  
*<sup>1</sup>Sony, Japan; <sup>2</sup>ETH Zürich, Switzerland*

- PAGE 2987  
Sun-P-8-2-8  
16:00      **Artificial Neural Network-Based Feature Combination for Spatial Voice Activity Detection**  
*Stefan Meier, Walter Kellermann, FAU Erlangen-Nürnberg, Germany*
- PAGE 2992  
Sun-P-8-2-9  
16:00      **HAPPY Team Entry to NIST OpenSAD Challenge: A Fusion of Short-Term Unsupervised and Segment i-Vector Based Speech Activity Detectors**  
*Tomi Kinnunen<sup>1</sup>, Alexey Sholokhov<sup>1</sup>, Elie Khoury<sup>2</sup>, Dennis Alexander Lehmann Thomsen<sup>3</sup>, Md. Sahidullah<sup>1</sup>, Zheng-Hua Tan<sup>3</sup>*  
*<sup>1</sup>University of Eastern Finland, Finland; <sup>2</sup>Pindrop, USA; <sup>3</sup>Aalborg University, Denmark*
- PAGE 2997  
Sun-P-8-2-10  
16:00      **Manual versus Automated: The Challenging Routine of Infant Vocalisation Segmentation in Home Videos to Study Neuro(mal)development**  
*Florian B. Pokorny<sup>1</sup>, Robert Peharz<sup>1</sup>, Wolfgang Roth<sup>2</sup>, Matthias Zöhrer<sup>2</sup>, Franz Pernkopf<sup>3</sup>, Peter B. Marschik<sup>1</sup>, Björn Schuller<sup>4</sup>*  
*<sup>1</sup>Medizinische Universität Graz, Austria; <sup>2</sup>Technische Universität Graz, Austria; <sup>3</sup>BioTechMed-Graz, Austria; <sup>4</sup>Universität Passau, Germany*
- PAGE 3002  
Sun-P-8-2-11  
16:00      **Minimizing Annotation Effort for Adaptation of Speech-Activity Detection Systems**  
*Luciana Ferrer<sup>1</sup>, Martin Graciarena<sup>2</sup>*  
*<sup>1</sup>Universidad de Buenos Aires, Argentina; <sup>2</sup>SRI International, USA*



## Sun-P-8-3: New Products and Services

Chair: Chuck Wooters

Room: Pacific Concourse - Poster C, Time 16:00-18:00, Sunday, 11 Sept. 2016

---

- PAGE 3007  
Sun-P-8-3-1  
16:00      **Progress and Prospects for Spoken Language Technology: What Ordinary People Think**  
*Roger K. Moore, Hui Li, Shih-Hao Liao, University of Sheffield, UK*
- PAGE 3012  
Sun-P-8-3-2  
16:00      **Progress and Prospects for Spoken Language Technology: Results from Four Sexennial Surveys**  
*Roger K. Moore, Ricard Marxer, University of Sheffield, UK*
- PAGE 3017  
Sun-P-8-3-3  
16:00      **On Employing a Highly Mismatched Crowd for Speech Transcription**  
*Purushotam Radadia, Rahul Kumar, Kanika Kalra, Shirish Karande, Sachin Lodha, Tata Consultancy Services, India*
- PAGE 3022  
Sun-P-8-3-4  
16:00      **Sage: The New BBN Speech Processing Platform**  
*Roger Hsiao, Ralf Meermeier, Tim Ng, Zhongqiang Huang, Maxwell Jordan, Enoch Kan, Tanel Alumäe, Jan Silovsky, William Hartmann, Francis Keith, Omer Lang, Manhung Siu, Owen Kimball, Raytheon BBN Technologies, USA*
- PAGE 3027  
Sun-P-8-3-5  
16:00      **DNN-Based Feature Enhancement Using Joint Training Framework for Robust Multichannel Speech Recognition**  
*Kang Hyun Lee, Tae Gyoon Kang, Woo Hyun Kang, Nam Soo Kim, Seoul National University, Korea*
- PAGE 3032  
Sun-P-8-3-6  
16:00      **Deep Neural Network Frontend for Continuous EMG-Based Speech Recognition**  
*Michael Wand, Jürgen Schmidhuber, IDSIA, Switzerland*
- PAGE 3037  
Sun-P-8-3-7  
16:00      **Overcoming Data Sparsity in Acoustic Modeling of Low-Resource Language by Borrowing Data and Model Parameters from High-Resource Languages**  
*Basil Abraham, S. Umesh, Neethu Mariam Joy, IIT Madras, India*

- PAGE 3042  
Sun-P-8-3-8  
16:00      **Multi-Language Neural Network Language Models**  
*Anton Ragni, Edgar Dakin, Xie Chen, Mark J.F. Gales, Kate M. Knill, University of Cambridge, UK*
- PAGE 3047  
Sun-P-8-3-9  
16:00      **Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration**  
*Ottokar Tilk<sup>1</sup>, Tanel Alumäe<sup>2</sup>*  
*<sup>1</sup>Tallinn University of Technology, Estonia; <sup>2</sup>Raytheon BBN Technologies, USA*
- PAGE 3052  
Sun-P-8-3-10  
16:00      **TheanoLM — An Extensible Toolkit for Neural Network Language Modeling**  
*Seppo Enarvi, Mikko Kurimo, Aalto University, Finland*
- PAGE 3057  
Sun-P-8-3-11  
16:00      **Selection of Multi-Genre Broadcast Data for the Training of Automatic Speech Recognition Systems**  
*P. Lanchantin, Mark J.F. Gales, Penny Karanasou, X. Liu, Y. Qian, L. Wang, P.C. Woodland, C. Zhang, University of Cambridge, UK*
- PAGE 3062  
Sun-P-8-3-12  
16:00      **Manipulating Word Lattices to Incorporate Human Corrections**  
*Yashesh Gaur, Florian Metze, Jeffrey P. Bigham, Carnegie Mellon University, USA*
- PAGE 3066  
Sun-P-8-3-13  
16:00      **Context-Aware Restaurant Recommendation for Natural Language Queries: A Formative User Study in the Automotive Domain**  
*Philipp Fischer<sup>1</sup>, Cornelius Styp von Rekowski<sup>2</sup>, Andreas Nürnberger<sup>2</sup>*  
*<sup>1</sup>Mercedes-Benz R&D North America, USA; <sup>2</sup>Otto-von-Guericke-Universität Magdeburg, Germany*
- PAGE 3071  
Sun-P-8-3-14  
16:00      **Teaming Up: Making the Most of Diverse Representations for a Novel Personalized Speech Retrieval Application**  
*Stephanie Pancoast<sup>1</sup>, Murat Akbacak<sup>2</sup>*  
*<sup>1</sup>Airbnb, USA; <sup>2</sup>Apple, USA*

PAGE 3076  
Sun-P-8-3-15  
16:00

**Automatic Speech Transcription for Low-Resource Languages — The Case of YoloXóchitl Mixtec (Mexico)**

*Vikramjit Mitra<sup>1</sup>, Andreas Kathol<sup>1</sup>, Jonathan D. Amith<sup>2</sup>, Rey Castillo García<sup>3</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>Gettysburg College, USA; <sup>3</sup>Secretaría de Educación Pública, Mexico*

PAGE 3081  
Sun-P-8-3-16  
16:00

**Real-Time Presentation Tracking Using Semantic Keyword Spotting**

*Reza Asadi, Harriet J. Fell, Timothy Bickmore, Ha Trinh, Northeastern University, USA*

**Sun-P-8-4: Low Resource Speech Recognition**

*Chair: Kartik Audhkhasi*

*Room: Pacific Concourse - Poster D, Time 16:00-18:00, Sunday, 11 Sept. 2016*

---

PAGE 3086  
Sun-P-8-4-1  
16:00

**Deriving Phonetic Transcriptions and Discovering Word Segmentations for Speech-to-Speech Translation in Low-Resource Settings**

*Andrew Wilkinson, Tiancheng Zhao, Alan W. Black, Carnegie Mellon University, USA*

PAGE 3091  
Sun-P-8-4-2  
16:00

**Unsupervised Joint Estimation of Grapheme-to-Phoneme Conversion Systems and Acoustic Model Adaptation for Non-Native Speech Recognition**

*Satoshi Tsujioka, Sakriani Sakti, Koichiro Yoshino, Graham Neubig, Satoshi Nakamura, NAIST, Japan*

PAGE 3096  
Sun-P-8-4-3  
16:00

**Learning Personalized Pronunciations for Contact Name Recognition**

*Antoine Bruguier, Fuchun Peng, Françoise Beaufays, Google, USA*

PAGE 3101  
Sun-P-8-4-4  
16:00

**Generation and Pruning of Pronunciation Variants to Improve ASR Accuracy**

*Zhenhao Ge, Aravind Ganapathiraju, Ananth N. Iyer, Scott A. Randal, Felix I. Wyss, Interactive Intelligence, USA*

PAGE 3106  
Sun-P-8-4-5  
16:00

**Optimizing Speech Recognition Evaluation Using Stratified Sampling**

*Janne Pytkönen<sup>1</sup>, Thomas Drugman<sup>2</sup>, Max Bisani<sup>2</sup>*  
*<sup>1</sup>Amazon.com, Finland; <sup>2</sup>Amazon.com, Germany*

## Keynote 4: Dan Jurafsky

Chair: Panos Georgiou

Room: Grand Ballroom ABC, Time 08:30-09:30, Monday, 12 Sept. 2016

---

PAGE 3111  
Mon-Keynote-4  
08:30

### **Ketchup, Interdisciplinarity, and the Spread of Innovation in Speech and Language Processing**

*Dan Jurafsky, Stanford University, USA*

---

## Mon-SE-3: Special Event: Speech Ventures

Chairs: Nicolas Scheffer, Korbinian Riedhammer, Alex Lebrun, David Suendermann-Oeft

Room: Grand Ballroom A, Time 10:00-12:00, Monday, 12 Sept. 2016

---

N/A  
Mon-SE-3  
10:00

### **Speech Ventures**

*Nicolas Scheffer<sup>1</sup>, Korbinian Riedhammer<sup>2</sup>, Alexandre Lebrun<sup>1</sup>, David Suendermann-Oeft<sup>3</sup>*

<sup>1</sup>Facebook, USA; <sup>2</sup>Remeeting, USA; <sup>3</sup>Educational Testing Service, USA

---

[Search]

## Mon-O-9-2: Special Session: Speech and Language Technologies for Human-Machine Conversation-Based Language Education

Chairs: Yao Qian, Helen Meng, Frank Soong

Room: Grand Ballroom BC, Time 10:00-12:00, Monday, 12 Sept. 2016

---

PAGE 3112  
Mon-O-9-2-1  
10:00

### **Context Aware Mispronunciation Detection for Mandarin Pronunciation Training**

*Rong Tong, Nancy F. Chen, Bin Ma, Haizhou Li, A\*STAR, Singapore*

PAGE 3117  
Mon-O-9-2-2  
10:15

### **DNN Online with iVectors Acoustic Modeling and Doc2Vec Distributed Representations for Improving Automated Speech Scoring**

*Jidong Tao, Lei Chen, Chong Min Lee, Educational Testing Service, USA*

PAGE 3122  
Mon-O-9-2-3  
10:30

### **Self-Adaptive DNN for Improving Spoken Language Proficiency Assessment**

*Yao Qian, Xinhao Wang, Keelan Evanini, David Suendermann-Oeft, Educational Testing Service, USA*

PAGE 3127  
Mon-O-9-2-4  
10:45

### **Detecting Mispronunciations of L2 Learners and Providing Corrective Feedback Using Knowledge-Guided and Data-Driven Decision Trees**

*Wei Li<sup>1</sup>, Kehuang Li<sup>1</sup>, Sabato Marco Siniscalchi<sup>1</sup>, Nancy F. Chen<sup>2</sup>, Chin-Hui Lee<sup>1</sup>*

<sup>1</sup>Georgia Institute of Technology, USA; <sup>2</sup>A\*STAR, Singapore

PAGE 3132  
Mon-O-9-2-5  
11:00

### **Phoneme Set Design Considering Integrated Acoustic and Linguistic Features of Second Language Speech**

*Xiaoyun Wang, Tsuneo Kato, Seiichi Yamamoto, Doshisha University, Japan*

PAGE 3137  
Mon-O-9-2-6  
11:15

### **HMM-Based Non-Native Accent Assessment Using Posterior Features**

*Ramya Rasipuram, Milos Cernak, Mathew Magimai-Doss, Idiap Research Institute, Switzerland*

PAGE 3142  
Mon-O-9-2-7  
11:30

### **Automatic Assessment and Error Detection of Shadowing Speech: Case of English Spoken by Japanese Learners**

*Shuju Shi<sup>1</sup>, Yosuke Kashiwagi<sup>1</sup>, Shohei Toyama<sup>1</sup>, Junwei Yue<sup>1</sup>, Yutaka Yamauchi<sup>2</sup>, Daisuke Saito<sup>1</sup>, Nobuaki Minematsu<sup>1</sup>*

<sup>1</sup>University of Tokyo, Japan; <sup>2</sup>Tokyo International University, Japan

## Mon-O-9-3: Phonation and Voice Quality

*Chairs: Paavo Alku, Marc Garellek*

*Room: Bayview A, Time 10:00-12:00, Monday, 12 Sept. 2016*

---

- PAGE 3147  
Mon-O-9-3-1  
10:00      **Multiplicity of the Acoustic Correlates of the Fortis-Lenis Contrast: Plosives in Aberystwyth English**  
*Miša Hejrná, Newcastle University, UK*
- PAGE 3152  
Mon-O-9-3-2  
10:20      **Automatic Measurement of Voice Onset Time and Prevoicing Using Recurrent Neural Networks**  
*Yossi Adi<sup>1</sup>, Joseph Keshet<sup>1</sup>, Olga Dmitrieva<sup>2</sup>, Matt Goldrick<sup>3</sup>*  
*<sup>1</sup>Bar-Ilan University, Israel; <sup>2</sup>Purdue University, USA; <sup>3</sup>Northwestern University, USA*
- PAGE 3156  
Mon-O-9-3-3  
10:40      **L1-L2 Interference: The Case of Final Devoicing of French Voiced Fricatives in Final Position by German Learners**  
*Sucheta Ghosh<sup>1</sup>, Camille Fauth<sup>2</sup>, Aghilas Sini<sup>1</sup>, Yves Laprie<sup>1</sup>*  
*<sup>1</sup>LORIA, France; <sup>2</sup>LiLpa, France*
- PAGE 3161  
Mon-O-9-3-4  
11:00      **Perceptual Salience of Voice Source Parameters in Signaling Focal Prominence**  
*Irena Yanushevskaya, Andy Murphy, Christer Gobl, Ailbhe Ní Chasaide, Trinity College Dublin, Ireland*
- PAGE 3166  
Mon-O-9-3-5  
11:20      **Classification of Voice Modality Using Electroglottogram Waveforms**  
*Michal Borsky<sup>1</sup>, Daryush D. Mehta<sup>2</sup>, Julius P. Gudjohnsen<sup>1</sup>, Jon Gudnason<sup>1</sup>*  
*<sup>1</sup>Reykjavik University, Iceland; <sup>2</sup>Massachusetts General Hospital, USA*
- PAGE 3171  
Mon-O-9-3-6  
11:40      **Voice-Quality Difference Between the Vowels in Filled Pauses and Ordinary Lexical Items**  
*Kikuo Maekawa<sup>1</sup>, Hiroki Mori<sup>2</sup>*  
*<sup>1</sup>NINJAL, Japan; <sup>2</sup>Utsunomiya University, Japan*

## Mon-O-9-4: Speech Synthesis Oral: Prosody and Expressive Speech

*Chairs: Junichi Yamagishi, Maria Wolters*

*Room: Bayview B, Time 10:00-12:00, Monday, 12 Sept. 2016*

---

- PAGE 3176  
Mon-O-9-4-1  
10:00      **Generation of Emotion Control Vector Using MDS-Based Space Transformation for Expressive Speech Synthesis**  
*Yan-You Chen, Chung-Hsien Wu, Yu-Fong Huang, National Cheng Kung University, Taiwan*
- PAGE 3181  
Mon-O-9-4-2  
10:20      **Direct Expressive Voice Training Based on Semantic Selection**  
*Igor Jauk, Antonio Bonafonte, Universitat Politècnica de Catalunya, Spain*
- PAGE 3186  
Mon-O-9-4-3  
10:40      **Syllable-Level Representations of Suprasegmental Features for DNN-Based Text-to-Speech Synthesis**  
*Manuel Sam Ribeiro, Oliver Watts, Junichi Yamagishi, University of Edinburgh, UK*
- PAGE 3191  
Mon-O-9-4-4  
11:00      **Pause Prediction from Text for Speech Synthesis with User-Definable Pause Insertion Likelihood Threshold**  
*Norbert Braunschweiler, Ranniery Maia, Toshiba Research Europe, UK*
- PAGE 3196  
Mon-O-9-4-5  
11:20      **A Hybrid System for Continuous Word-Level Emphasis Modeling Based on HMM State Clustering and Adaptive Training**  
*Quoc Truong Do<sup>1</sup>, Tomoki Toda<sup>2</sup>, Graham Neubig<sup>1</sup>, Sakriani Sakti<sup>1</sup>, Satoshi Nakamura<sup>1</sup>*  
*<sup>1</sup>NAIST, Japan; <sup>2</sup>Nagoya University, Japan*
- PAGE 3201  
Mon-O-9-4-6  
11:40      **Improving Prosodic Boundaries Prediction for Mandarin Speech Synthesis by Using Enhanced Embedding Feature and Model Fusion Approach**  
*Yibin Zheng, Ya Li, Zhengqi Wen, Xingguang Ding, Jianhua Tao, Chinese Academy of Sciences, China*

## Mon-O-9-5: Language Recognition

Chairs: Jean-Francois Bonastre, Haizhou Li

Room: Seacliff BCD, Time 10:00-12:00, Monday, 12 Sept. 2016

---

PAGE 3206  
Mon-O-9-5-1  
10:00

### Results of The 2015 NIST Language Recognition Evaluation

Hui Zhao<sup>1</sup>, Désiré Bansé<sup>1</sup>, George Doddington<sup>1</sup>, Craig Greenberg<sup>1</sup>,  
Jaime Hernández-Cordero<sup>2</sup>, John Howard<sup>1</sup>, Lisa Mason<sup>2</sup>, Alvin Martin<sup>1</sup>,  
Douglas Reynolds<sup>3</sup>, Elliot Singer<sup>3</sup>, Audrey Tong<sup>1</sup>  
<sup>1</sup>NIST, USA; <sup>2</sup>DoD, USA; <sup>3</sup>MIT Lincoln Laboratory, USA

PAGE 3211  
Mon-O-9-5-2  
10:20

### The 2015 NIST Language Recognition Evaluation: The Shared View of I2R, Fantastic4 and SingaMS

Kong Aik Lee<sup>1</sup>, Haizhou Li<sup>1</sup>, Li Deng<sup>2</sup>, Ville Hautamäki<sup>3</sup>, Wei Rao<sup>4</sup>, Xiong Xiao<sup>4</sup>,  
Anthony Larcher<sup>5</sup>, Hanwu Sun<sup>1</sup>, Trung Hieu Nguyen<sup>1</sup>, Guangsen Wang<sup>1</sup>,  
Aleksandr Sizov<sup>1</sup>, Jianshu Chen<sup>2</sup>, Ivan Kukanov<sup>3</sup>, Amir Hossein Poorjam<sup>3</sup>,  
Trung Ngo Trong<sup>3</sup>, Cheng-Lin Xu<sup>4</sup>, Haihua Xu<sup>4</sup>, Bin Ma<sup>1</sup>, Eng Siong Chng<sup>4</sup>,  
Sylvain Meignier<sup>5</sup>  
<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>Microsoft, USA; <sup>3</sup>University of Eastern Finland, Finland; <sup>4</sup>NTU, Singapore; <sup>5</sup>LIUM, France

PAGE 3216  
Mon-O-9-5-3  
10:40

### Pair-Wise Distance Metric Learning of Neural Network Model for Spoken Language Identification

Xugang Lu<sup>1</sup>, Peng Shen<sup>1</sup>, Yu Tsao<sup>2</sup>, Hisashi Kawai<sup>1</sup>  
<sup>1</sup>NICT, Japan; <sup>2</sup>Academia Sinica, Taiwan

PAGE 3221  
Mon-O-9-5-4  
11:00

### Non-Iterative Parameter Estimation for Total Variability Model Using Randomized Singular Value Decomposition

Ruchir Travadi, Shrikanth S. Narayanan, University of Southern California, USA

PAGE 3226  
Mon-O-9-5-5  
11:20

### Stacked Long-Term TDNN for Spoken Language Recognition

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

Mon-O-9-5 continued...

PAGE 3231  
Mon-O-9-5-6  
11:40

### A Divide-and-Conquer Approach for Language Identification Based on Recurrent Neural Networks

G. Gelly<sup>1</sup>, Jean-Luc Gauvain<sup>1</sup>, V.B. Le<sup>2</sup>, A. Messaoudi<sup>2</sup>  
<sup>1</sup>LIMSI, France; <sup>2</sup>Vocapia Research, France

## Mon-O-9-6: Spoken Language Understanding Systems

Chairs: Helen Meng, Gokhan Tur

Room: Seacliff A, Time 10:00-12:00, Monday, 12 Sept. 2016

---

- PAGE 3236  
Mon-O-9-6-1  
10:00      **Context-Sensitive and Role-Dependent Spoken Language Understanding Using Bidirectional and Attention LSTMs**  
*Chiori Hori, Takaaki Hori, Shinji Watanabe, John R. Hershey, MERL, USA*
- PAGE 3241  
Mon-O-9-6-2  
10:20      **A Step Beyond Local Observations with a Dialog Aware Bidirectional GRU Network for Spoken Language Understanding**  
*Vedran Vukotić<sup>1</sup>, Christian Raymond<sup>1</sup>, Guillaume Gravier<sup>2</sup>*  
*<sup>1</sup>INSA de Rennes, France; <sup>2</sup>Inria, France*
- PAGE 3245  
Mon-O-9-6-3  
10:40      **End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding**  
*Yun-Nung Chen<sup>1</sup>, Dilek Hakkani-Tür<sup>2</sup>, Gokhan Tur<sup>2</sup>, Jianfeng Gao<sup>2</sup>, Li Deng<sup>2</sup>*  
*<sup>1</sup>National Taiwan University, Taiwan; <sup>2</sup>Microsoft, USA*
- PAGE 3250  
Mon-O-9-6-4  
11:00      **Sequential Convolutional Neural Networks for Slot Filling in Spoken Language Understanding**  
*Ngoc Thang Vu, Universität Stuttgart, Germany*
- PAGE 3255  
Mon-O-9-6-5  
11:20      **A New Pre-Training Method for Training Deep Learning Models with Application to Spoken Language Understanding**  
*Asli Celikyilmaz, Ruhi Sarikaya, Dilek Hakkani-Tür, Xiaohu Liu, Nikhil Ramesh, Gokhan Tur, Microsoft, USA*
- PAGE 3260  
Mon-O-9-6-6  
11:40      **Joint Syntactic and Semantic Analysis with a Multitask Deep Learning Framework for Spoken Language Understanding**  
*Jeremie Tafforeau, Frederic Bechet, Thierry Artiere, Benoit Favre, LIF (UMR 7279), France*

## Mon-P-9-1: Language Recognition

Chair: Xiong Xiao

Room: Pacific Concourse - Poster A, Time 10:00-12:00, Monday, 12 Sept. 2016

---

- PAGE 3265  
Mon-P-9-1-1  
10:00      **Exploiting Hidden-Layer Responses of Deep Neural Networks for Language Recognition**  
*Ruizhi Li<sup>1</sup>, Sri Harish Mallidi<sup>1</sup>, Lukáš Burget<sup>2</sup>, Oldřich Plchot<sup>2</sup>, Najim Dehak<sup>1</sup>*  
*<sup>1</sup>Johns Hopkins University, USA; <sup>2</sup>Brno University of Technology, Czech Republic*
- PAGE 3270  
Mon-P-9-1-2  
10:00      **Out of Set Language Modelling in Hierarchical Language Identification**  
*Saad Irtza<sup>1</sup>, Vidhyasaharan Sethu<sup>1</sup>, Sarith Fernando<sup>1</sup>, Eliathamby Ambikairajah<sup>1</sup>, Haizhou Li<sup>2</sup>*  
*<sup>1</sup>University of New South Wales, Australia; <sup>2</sup>A\*STAR, Singapore*
- PAGE 3275  
Mon-P-9-1-3  
10:00      **Language Identification Based on Generative Modeling of Posteriorgram Sequences Extracted from Frame-by-Frame DNNs and LSTM-RNNs**  
*Ryo Masumura, Taichi Asami, Hirokazu Masataki, Yushi Aono, Sumitaka Sakauchi, NTT, Japan*
- PAGE 3280  
Mon-P-9-1-4  
10:00      **Gating Recurrent Enhanced Memory Neural Networks on Language Identification**  
*Wang Geng, Yuanyuan Zhao, Wenfu Wang, Xinyuan Cai, Bo Xu, Chinese Academy of Sciences, China*
- PAGE 3285  
Mon-P-9-1-5  
10:00      **Sequence Summarizing Neural Networks for Spoken Language Recognition**  
*Jan Pešán, Lukáš Burget, Jan Černocký, Brno University of Technology, Czech Republic*
- PAGE 3289  
Mon-P-9-1-6  
10:00      **The Role of Spectral Resolution in Foreign-Accented Speech Perception**  
*Michelle R. Kapolowicz, Vahid Montazeri, Peter F. Assmann, University of Texas at Dallas, USA*

PAGE 3294  
Mon-P-9-1-7  
10:00

**THU-EE System Description for NIST LRE 2015**

*Liang He, Yao Tian, Yi Liu, Jiaming Xu, Weiwei Liu, Cai Meng, Jia Liu, Tsinghua University, China*

PAGE 3299  
Mon-P-9-1-8  
10:00

**Variation in Spoken North Sami Language**

*Kristiina Jokinen<sup>1</sup>, Trung Ngo Trong<sup>1</sup>, Ville Hautamäki<sup>2</sup>*

*<sup>1</sup>University of Helsinki, Finland; <sup>2</sup>University of Eastern Finland, Finland*

**Mon-P-9-2: Music, Audio, and Source Separation**

*Chair: Yang Liu*

*Room: Pacific Concourse - Poster B, Time 10:00-12:00, Monday, 12 Sept. 2016*

---

PAGE 3304  
Mon-P-9-2-1  
10:00

**Improved Music Genre Classification with Convolutional Neural Networks**

*Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing, SCUT, China*

PAGE 3309  
Mon-P-9-2-2  
10:00

**Enhanced Harmonic Content and Vocal Note Based Predominant Melody Extraction from Vocal Polyphonic Music Signals**

*Gurunath Reddy M., K. Sreenivasa Rao, IIT Kharagpur, India*

PAGE 3314  
Mon-P-9-2-3  
10:00

**Long Short-Term Memory for Speaker Generalization in Supervised Speech Separation**

*Jitong Chen, DeLiang Wang, Ohio State University, USA*

PAGE 3319  
Mon-P-9-2-4  
10:00

**Phonotactic Language Identification for Singing**

*Anna M. Kruspe, Fraunhofer IDMT, Germany*

PAGE 3324  
Mon-P-9-2-5  
10:00

**Comparing the Influence of Spectro-Temporal Integration in Computational Speech Segregation**

*Thomas Bentsen, Tobias May, Abigail A. Kressner, Torsten Dau, Technical University of Denmark, Denmark*

PAGE 3329  
Mon-P-9-2-6  
10:00

**Blind Speech Separation with GCC-NMF**

*Sean U.N. Wood, Jean Rouat, Université de Sherbrooke, Canada*

PAGE 3334  
Mon-P-9-2-7  
10:00

**Effects of Cochlear Hearing Loss on the Benefits of Ideal Binary Masking**

*Vahid Montazeri, Shaikat Hossain, Peter F. Assmann, University of Texas at Dallas, USA*

PAGE 3339  
Mon-P-9-2-8  
10:00

**Combining Mask Estimates for Single Channel Audio Source Separation Using Deep Neural Networks**

*Emad M. Grais, Gerard Roma, Andrew J.R. Simpson, Mark D. Plumbley, University of Surrey, UK*



PAGE 3344 Mon-P-9-2-9 10:00	<b>Monaural Source Separation Using a Random Forest Classifier</b> <i>Cosimo Riday, Saurabh Bhargava, Richard H.R. Hahnloser, Shih-Chii Liu, Universität Zürich, Switzerland</i>
PAGE 3349 Mon-P-9-2-10 10:00	<b>Adaptive Group Sparsity for Non-Negative Matrix Factorization with Application to Unsupervised Source Separation</b> <i>Xu Li, Ziteng Wang, Xiaofei Wang, Qiang Fu, Yonghong Yan, Chinese Academy of Sciences, China</i>
PAGE 3354 Mon-P-9-2-11 10:00	<b>A Robust Dual-Microphone Speech Source Localization Algorithm for Reverberant Environments</b> <i>Yanmeng Guo<sup>1</sup>, Xiaofei Wang<sup>1</sup>, Chao Wu<sup>1</sup>, Qiang Fu<sup>1</sup>, Ning Ma<sup>2</sup>, Guy J. Brown<sup>2</sup></i> <i><sup>1</sup>Chinese Academy of Sciences, China; <sup>2</sup>University of Sheffield, UK</i>
PAGE 3359 Mon-P-9-2-12 10:00	<b>Speech Localisation in a Multitalker Mixture by Humans and Machines</b> <i>Ning Ma, Guy J. Brown, University of Sheffield, UK</i>
PAGE 3364 Mon-P-9-2-13 10:00	<b>Reverberation-Robust One-Bit TDOA Based Moving Source Localization for Automatic Camera Steering</b> <i>Harshavardhan Sundar, Gokul Deepak Manavalan, T.V. Sreenivas, Chandra Sekhar Seelamantula, Indian Institute of Science, India</i>
PAGE 3369 Mon-P-9-2-14 10:00	<b>Multi-Talker Speech Recognition Based on Blind Source Separation with ad hoc Microphone Array Using Smartphones and Cloud Storage</b> <i>Keiko Ochi<sup>1</sup>, Nobutaka Ono<sup>1</sup>, Shigeki Miyabe<sup>2</sup>, Shoji Makino<sup>2</sup></i> <i><sup>1</sup>NII, Japan; <sup>2</sup>University of Tsukuba, Japan</i>

### Mon-P-9-3: Acoustic Modeling with Neural Networks

Chair: Samuel Thomas

Room: Pacific Concourse - Poster C, Time 10:00-12:00, Monday, 12 Sept. 2016

---

PAGE 3374 Mon-P-9-3-1 10:00	<b>Phase-Aware Signal Processing for Automatic Speech Recognition</b> <i>Johannes Fahringer, Tobias Schrank, Johannes Stahl, Pejman Mowlae, Franz Pernkopf, Technische Universität Graz, Austria</i>
PAGE 3379 Mon-P-9-3-2 10:00	<b>Unsupervised Deep Auditory Model Using Stack of Convolutional RBMs for Speech Recognition</b> <i>Hardik B. Sailor, Hemant A. Patil, DA-IICT, India</i>
PAGE 3384 Mon-P-9-3-3 10:00	<b>Interpretation of Low Dimensional Neural Network Bottleneck Features in Terms of Human Perception and Production</b> <i>Philip Weber, Linxue Bai, Martin Russell, Peter Jančovič, Stephen Houghton, University of Birmingham, UK</i>
PAGE 3389 Mon-P-9-3-4 10:00	<b>Compact Feedforward Sequential Memory Networks for Large Vocabulary Continuous Speech Recognition</b> <i>Shiliang Zhang<sup>1</sup>, Hui Jiang<sup>2</sup>, Shifu Xiong<sup>1</sup>, Si Wei<sup>1</sup>, Li-Rong Dai<sup>1</sup></i> <i><sup>1</sup>USTC, China; <sup>2</sup>York University, Canada</i>
PAGE 3394 Mon-P-9-3-5 10:00	<b>Future Context Attention for Unidirectional LSTM Based Acoustic Model</b> <i>Jian Tang<sup>1</sup>, Shiliang Zhang<sup>1</sup>, Si Wei<sup>2</sup>, Li-Rong Dai<sup>1</sup></i> <i><sup>1</sup>USTC, China; <sup>2</sup>iFLYTEK, China</i>
PAGE 3399 Mon-P-9-3-6 10:00	<b>Hybrid Accelerated Optimization for Speech Recognition</b> <i>Jen-Tzung Chien<sup>1</sup>, Pei-Wen Huang<sup>1</sup>, Tan Lee<sup>2</sup></i> <i><sup>1</sup>National Chiao Tung University, Taiwan; <sup>2</sup>Chinese University of Hong Kong, China</i>
PAGE 3404 Mon-P-9-3-7 10:00	<b>On Online Attention-Based Speech Recognition and Joint Mandarin Character-Pinyin Training</b> <i>William Chan, Ian Lane, Carnegie Mellon University, USA</i>



PAGE 3409 Mon-P-9-3-8 10:00	<b>GMM-Free Flat Start Sequence-Discriminative DNN Training</b> <i>Gábor Gosztolya<sup>1</sup>, Tamás Grósz<sup>2</sup>, László Tóth<sup>1</sup></i> <i><sup>1</sup>MTA-SZTE RGAI, Hungary; <sup>2</sup>University of Szeged, Hungary</i>
PAGE 3414 Mon-P-9-3-9 10:00	<b>Open-Domain Audio-Visual Speech Recognition: A Deep Learning Approach</b> <i>Yajie Miao, Florian Metze, Carnegie Mellon University, USA</i>
PAGE 3419 Mon-P-9-3-10 10:00	<b>Multidimensional Residual Learning Based on Recurrent Neural Networks for Acoustic Modeling</b> <i>Yuanyuan Zhao, Shuang Xu, Bo Xu, Chinese Academy of Sciences, China</i>
PAGE 3424 Mon-P-9-3-11 10:00	<b>Towards Online-Recognition with Deep Bidirectional LSTM Acoustic Models</b> <i>Albert Zeyer, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</i>
PAGE 3429 Mon-P-9-3-12 10:00	<b>Advances in Very Deep Convolutional Neural Networks for LVCSR</b> <i>Tom Sercu, Vaibhava Goel, IBM, USA</i>
PAGE 3434 Mon-P-9-3-13 10:00	<b>Acoustic Modelling from the Signal Domain Using CNNs</b> <i>Pegah Ghahremani, Vimal Manohar, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA</i>
PAGE 3439 Mon-P-9-3-14 10:00	<b>Distilling Knowledge from Ensembles of Neural Networks for Speech Recognition</b> <i>Yevgen Chebotar<sup>1</sup>, Austin Waters<sup>2</sup></i> <i><sup>1</sup>University of Southern California, USA; <sup>2</sup>Google, USA</i>
PAGE 3444 Mon-P-9-3-15 10:00	<b>Triphone State-Tying via Deep Canonical Correlation Analysis</b> <i>Weiran Wang, Hao Tang, Karen Livescu, TTIC, USA</i>
PAGE 3449 Mon-P-9-3-16 10:00	<b>Low-Rank Representation of Nearest Neighbor Posterior Probabilities to Enhance DNN Based Acoustic Modeling</b> <i>Gil Luyet, Pranay Dighe, Afsaneh Asaei, Hervé Bourlard, Idiap Research Institute, Switzerland</i>

## Mon-P-9-4: Robustness and Adaptation

Chair: Kyu Jeong Han

Room: Pacific Concourse - Poster D, Time 10:00-12:00, Monday, 12 Sept. 2016

---

PAGE 3454 Mon-P-9-4-1 10:00	<b>Improving Large Vocabulary Accented Mandarin Speech Recognition with Attribute-Based I-Vectors</b> <i>Hao Zheng<sup>1</sup>, Shanshan Zhang<sup>1</sup>, Liwei Qiao<sup>2</sup>, Jianping Li<sup>2</sup>, Wenju Liu<sup>1</sup></i> <i><sup>1</sup>Chinese Academy of Sciences, China; <sup>2</sup>Shanxi Electric Power, China</i>
PAGE 3459 Mon-P-9-4-2 10:00	<b>Pitch-Adaptive Front-End Features for Robust Children's ASR</b> <i>S. Shahnawazuddin, Abhishek Dey, Rohit Sinha, IIT Guwahati, India</i>
PAGE 3464 Mon-P-9-4-3 10:00	<b>ASR Confidence Estimation with Speaker-Adapted Recurrent Neural Networks</b> <i>Miguel Ángel del-Agua, Santiago Piqueras, Adrià Giménez, Alberto Sanchis, Jorge Civera, Alfons Juan, Universidad Politécnica de Valencia, Spain</i>
PAGE 3469 Mon-P-9-4-4 10:00	<b>Automatic Correction of ASR Outputs by Using Machine Translation</b> <i>Luis Fernando D'Haro, Rafael E. Banchs, A*STAR, Singapore</i>
PAGE 3474 Mon-P-9-4-5 10:00	<b>A Framework for Practical Multistream ASR</b> <i>Sri Harish Mallidi, Hynek Hermansky, Johns Hopkins University, USA</i>
PAGE 3479 Mon-P-9-4-6 10:00	<b>DNNs for Unsupervised Extraction of Pseudo FMLLR Features Without Explicit Adaptation Data</b> <i>Neethu Mariam Joy, Murali Karthick Baskar, S. Umesh, Basil Abraham, IIT Madras, India</i>
PAGE 3484 Mon-P-9-4-7 10:00	<b>Multi-Attribute Factorized Hidden Layer Adaptation for DNN Acoustic Models</b> <i>Lahiru Samarakoon, Khe Chai Sim, NUS, Singapore</i>
PAGE 3489 Mon-P-9-4-8 10:00	<b>Speaker Normalization Through Feature Shifting of Linearly Transformed i-Vector</b> <i>Jahyun Goo, Younggwon Kim, Hyungjun Lim, Hoirin Kim, KAIST, Korea</i>

## Mon-SE-4 : Special Event: Computational Approaches to Linguistic Code Switching

*Chairs: Julia Hirschberg, Abeer Alwan, Mona Diab, Thamar Solorio*  
*Room: Grand Ballroom A, Time 12:15-13:00, Monday, 12 Sept. 2016*

---

N/A  
Mon-SE-4  
12:15

### **Computational Approaches to Linguistic Code Switching**

*Mona Diab<sup>1</sup>, Pascale Fung<sup>2</sup>, Julia Hirschberg<sup>3</sup>, Thamar Solorio<sup>4</sup>*

*<sup>1</sup>George Washington University, USA; <sup>2</sup>HKUST, China; <sup>3</sup>Columbia University, USA;*

*<sup>4</sup>University of Houston, USA*

[\[Search\]](#)

## Mon-O-10-1 : Neural Networks for Language Modeling

*Chairs: Steve Renals, Ebru Arisoy*  
*Room: Grand Ballroom A, Time 13:30-15:30, Monday, 12 Sept. 2016*

---

PAGE 3494  
Mon-O-10-1-1  
13:30

### **Compositional Neural Network Language Models for Agglutinative Languages**

*Ebru Arisoy<sup>1</sup>, Murat Saraclar<sup>2</sup>*

*<sup>1</sup>MEF Üniversitesi, Turkey; <sup>2</sup>Boğaziçi Üniversitesi, Turkey*

PAGE 3499  
Mon-O-10-1-2  
13:50

### **NN-Grams: Unifying Neural Network and n-Gram Language Models for Speech Recognition**

*Babak Damavandi, Shankar Kumar, Noam Shazeer, Antoine Bruguier, Google, USA*

PAGE 3504  
Mon-O-10-1-3  
14:10

### **Recurrent Neural Network Language Model with Incremental Updated Context Information Generated Using Bag-of-Words Representation**

*Md. Akmal Haidar, Mikko Kurimo, Aalto University, Finland*

PAGE 3509  
Mon-O-10-1-4  
14:30

### **Sequential Recurrent Neural Networks for Language Modeling**

*Youssef Oualil, Clayton Greenberg, Mittul Singh, Dietrich Klakow, Universität des Saarlandes, Germany*

PAGE 3514  
Mon-O-10-1-5  
14:50

### **Word-Phrase-Entity Recurrent Neural Networks for Language Modeling**

*Michael Levit, Sarangarajan Parthasarathy, Shuangyu Chang, Microsoft, USA*

PAGE 3519  
Mon-O-10-1-6  
15:10

### **LSTM, GRU, Highway and a Bit of Attention: An Empirical Overview for Language Modeling in Speech Recognition**

*Kazuki Irie, Zoltán Tüske, Tamer Alkhouli, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany*

## Mon-O-10-2: Special Session: Sub-Saharan African Languages: From Speech Fundamentals to Applications

*Chairs: Martine Adda-Decker, Laurent Besacier, Marelie Davel, Larry Hyman, Martin Jansche, Francois Pellegrino, Olivier Rosec, Sebastian Stüker, Martha Tachbelie Yifiru*

*Room: Grand Ballroom BC, Time 13:30-15:30, Monday, 12 Sept. 2016*

---

PAGE 3524  
Mon-O-10-2-1  
13:30

### **Automatic Speech Recognition Using Probabilistic Transcriptions in Swahili, Amharic, and Dinka**

*Amit Das, Preethi Jyothi, Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA*

PAGE 3529  
Mon-O-10-2-2  
13:45

### **Speed Perturbation and Vowel Duration Modeling for ASR in Hausa and Wolof Languages**

*Elodie Gauthier<sup>1</sup>, Laurent Besacier<sup>1</sup>, Sylvie Voisin<sup>2</sup>  
<sup>1</sup>LIG (UMR 5217), France; <sup>2</sup>DDL (UMR 5596), France*

PAGE 3534  
Mon-O-10-2-3  
14:00

### **Improving the Lwazi ASR Baseline**

*Charl van Heerden, Neil Kleynhans, Marelie Davel, North-West University, South Africa*

PAGE 3539  
Mon-O-10-2-4  
14:15

### **Preliminary Experiments on Unsupervised Word Discovery in Mboshi**

*Pierre Godard<sup>1</sup>, Gilles Adda<sup>1</sup>, Martine Adda-Decker<sup>1</sup>, Alexandre Allauzen<sup>1</sup>, Laurent Besacier<sup>2</sup>, H el ene Bonneau-Maynard<sup>1</sup>, Guy-No el Kouarata<sup>3</sup>, Kevin L oser<sup>1</sup>, Annie Rialland<sup>3</sup>, Fran ois Yvon<sup>1</sup>  
<sup>1</sup>LMSI, France; <sup>2</sup>LIG (UMR 5217), France; <sup>3</sup>LPP (UMR 7018), France*

PAGE 3544  
Mon-O-10-2-5  
14:30

### **Unsupervised Phoneme Segmentation of Previously Unseen Languages**

*Marco Vetter<sup>1</sup>, Markus M uller<sup>1</sup>, Fatima Hamlaoui<sup>2</sup>, Graham Neubig<sup>3</sup>, Satoshi Nakamura<sup>3</sup>, Sebastian St uker<sup>1</sup>, Alex Waibel<sup>1</sup>  
<sup>1</sup>KIT, Germany; <sup>2</sup>ZAS Berlin, Germany; <sup>3</sup>NAIST, Japan*

PAGE 3549  
Mon-O-10-2-6  
14:45

### **CNN-Based Phone Segmentation Experiments in a Less-Represented Language**

*C eline Manenti, Thomas Pellegrini, Julien Pinquier, IRIT, France*

PAGE 3554  
Mon-O-10-2-7  
15:00

### **Part-of-Speech Tagging and Chunking in Text-to-Speech Synthesis for South African Languages**

*Georg I. Schl unz, Nkosikhona Dlamini, Rynhardt P. Kruger, CSIR, South Africa*

PAGE 3559  
Mon-O-10-2-8  
15:15

### **The Effect of Postlexical Deletion on Automatic Speech Recognition in Fast Spontaneously Spoken Zulu**

*Ewald van der Westhuizen, Thomas Niesler, Stellenbosch University, South Africa*

## Mon-O-10-3: Speech Production Models

Chairs: Sofia Strömbergsson, Takayuki Arai

Room: Bayview A, Time 13:30-15:30, Monday, 12 Sept. 2016

---

PAGE 3564  
Mon-O-10-3-1  
13:30

### **A New Model of Speech Motor Control Based on Task Dynamics and State Feedback**

*Vikram Ramanarayanan<sup>1</sup>, Benjamin Parrell<sup>2</sup>, Louis Goldstein<sup>3</sup>,  
Srikantan Nagarajan<sup>4</sup>, John Houde<sup>4</sup>*

<sup>1</sup>Educational Testing Service, USA; <sup>2</sup>University of Delaware, USA; <sup>3</sup>University of Southern California, USA; <sup>4</sup>University of California at San Francisco, USA

PAGE 3569  
Mon-O-10-3-2  
13:50

### **Using a Biomechanical Model and Articulatory Data for the Numerical Production of Vowels**

*Saeed Dabbaghchian<sup>1</sup>, Marc Arnela<sup>2</sup>, Olov Engwall<sup>1</sup>, Oriol Guasch<sup>2</sup>, Ian Stavness<sup>3</sup>,  
Pierre Badin<sup>4</sup>*

<sup>1</sup>KTH, Sweden; <sup>2</sup>Universitat Ramon Llull, Spain; <sup>3</sup>University of Saskatchewan, Canada; <sup>4</sup>GIPSA, France

PAGE 3574  
Mon-O-10-3-3  
14:10

### **A New Model for Acoustic Wave Propagation and Scattering in the Vocal Tract**

*Jianguo Wei<sup>1</sup>, Wendan Guan<sup>1</sup>, Darcy Q. Hou<sup>1</sup>, Dingyi Pan<sup>2</sup>, Wenhuan Lu<sup>1</sup>,  
Jianwu Dang<sup>1</sup>*

<sup>1</sup>Tianjin University, China; <sup>2</sup>Zhejiang University, China

PAGE 3579  
Mon-O-10-3-4  
14:30

### **Uncontrolled Manifolds in Vowel Production: Assessment with a Biomechanical Model of the Tongue**

*Andrew Szabados, Pascal Perrier, GIPSA, France*

PAGE 3584  
Mon-O-10-3-5  
14:50

### **Experimental Validation of Sound Generated from Flow in Simplified Vocal Tract Model of Sibilant /s/**

*Tsukasa Yoshinaga<sup>1</sup>, Kazunori Nozaki<sup>2</sup>, Shigeo Wada<sup>1</sup>*

<sup>1</sup>Osaka University, Japan; <sup>2</sup>Osaka University Dental Hospital, Japan

PAGE 3588

### **Bayesian Modeling in Speech Motor Control: A Principled Structure for the Integration of Various Constraints**

*Jean-François Patri<sup>1</sup>, Pascal Perrier<sup>1</sup>, Julien Diard<sup>2</sup>*

<sup>1</sup>GIPSA, France; <sup>2</sup>LPNC, France

## Mon-O-10-4: Speaker States and Traits

Chairs: *Agustin Gravano, Mark Huckvale*

Room: *Bayview B, Time 13:30-15:30, Monday, 12 Sept. 2016*

---

- PAGE 3593  
Mon-O-10-4-1  
13:30
- Facing Realism in Spontaneous Emotion Recognition from Speech: Feature Enhancement by Autoencoder with LSTM Neural Networks**  
*Zixing Zhang<sup>1</sup>, Fabien Ringeval<sup>2</sup>, Jing Han<sup>1</sup>, Jun Deng<sup>1</sup>, Erik Marchi<sup>1</sup>, Björn Schuller<sup>1</sup>*  
<sup>1</sup>Universität Passau, Germany; <sup>2</sup>LIG (UMR 5217), France
- PAGE 3598  
Mon-O-10-4-2  
13:50
- Defining Emotionally Salient Regions Using Qualitative Agreement Method**  
*Srinivas Parthasarathy, Carlos Busso, University of Texas at Dallas, USA*
- PAGE 3603  
Mon-O-10-4-3  
14:10
- Representation Learning for Speech Emotion Recognition**  
*Sayan Ghosh<sup>1</sup>, Eugene Laksana<sup>1</sup>, Louis-Philippe Morency<sup>2</sup>, Stefan Scherer<sup>1</sup>*  
<sup>1</sup>University of Southern California, USA; <sup>2</sup>Carnegie Mellon University, USA
- PAGE 3608  
Mon-O-10-4-4  
14:30
- Multilingual Speech Emotion Recognition System Based on a Three-Layer Model**  
*Xingfeng Li, Masato Akagi, JAIST, Japan*
- PAGE 3613  
Mon-O-10-4-5  
14:50
- Analysis of Multi-Lingual Emotion Recognition Using Auditory Attention Features**  
*Ozlem Kalinli, Sony, USA*
- PAGE 3618  
Mon-O-10-4-6  
15:10
- On the Correlation and Transferability of Features Between Automatic Speech Recognition and Speech Emotion Recognition**  
*Haytham M. Fayek, Margaret Lech, Lawrence Cavedon, RMIT University, Australia*

## Mon-O-10-5: Speaker Recognition

Chairs: *John Hansen, Koichi Shinoda*

Room: *Seacliff BCD, Time 13:30-15:30, Monday, 12 Sept. 2016*

---

- PAGE 3623  
Mon-O-10-5-1  
13:30
- On the Influence of Text Content on Pass-Phrase Strength for Short-Duration Text-Dependent Automatic Speaker Authentication**  
*Giacomo Valenti<sup>1</sup>, Adrien Daniel<sup>1</sup>, Nicholas Evans<sup>2</sup>*  
<sup>1</sup>NXP Software, France; <sup>2</sup>EURECOM, France
- PAGE 3628  
Mon-O-10-5-2  
13:50
- Articulation Rate Filtering of CQCC Features for Automatic Speaker Verification**  
*Massimiliano Todisco, Héctor Delgado, Nicholas Evans, EURECOM, France*
- PAGE 3633  
Mon-O-10-5-3  
14:10
- The IBM Speaker Recognition System: Recent Advances and Error Analysis**  
*Seyed Omid Sadjadi<sup>1</sup>, Jason W. Pelecanos<sup>1</sup>, Sriram Ganapathy<sup>2</sup>*  
<sup>1</sup>IBM, USA; <sup>2</sup>Indian Institute of Science, India
- PAGE 3638  
Mon-O-10-5-4  
14:30
- Probabilistic Approach Using Joint Clean and Noisy i-Vectors Modeling for Speaker Recognition**  
*Waad Ben Kheder, Driss Matrouf, Moez Ajili, Jean-François Bonastre, LIA, France*
- PAGE 3643  
Mon-O-10-5-5  
14:50
- Generalized Discriminant Analysis (GDA) for Improved i-Vector Based Speaker Recognition**  
*Fahimeh Bahmaninezhad, John H.L. Hansen, University of Texas at Dallas, USA*
- PAGE 3648  
Mon-O-10-5-6  
15:10
- Noise and Metadata Sensitive Bottleneck Features for Improving Speaker Recognition with Non-Native Speech Input**  
*Yao Qian, Jidong Tao, David Suendermann-Oeft, Keelan Evanini, Alexei V. Ivanov, Vikram Ramanarayanan, Educational Testing Service, USA*

## Mon-O-10-6: VAD and Audio Events

Chair: Malcolm Slaney

Room: Seacliff A, Time 13:30-15:30, Monday, 12 Sept. 2016

---

- PAGE 3653  
Mon-O-10-6-1  
13:30      **Robust Audio Event Recognition with 1-Max Pooling Convolutional Neural Networks**  
*Huy Phan, Lars Hertel, Marco Maass, Alfred Mertins, Universität zu Lübeck, Germany*
- PAGE 3658  
Mon-O-10-6-2  
13:50      **Audio-Based Distributional Representations of Meaning Using a Fusion of Feature Encodings**  
*Giannis Karamanolakis<sup>1</sup>, Elias Iosif<sup>1</sup>, Athanasia Zlatintsi<sup>1</sup>, Aggelos Pikrakis<sup>2</sup>, Alexandros Potamianos<sup>1</sup>*  
*<sup>1</sup>NTUA, Greece; <sup>2</sup>University of Piraeus, Greece*
- PAGE 3663  
Mon-O-10-6-3  
14:10      **Robust DNN-Based VAD Augmented with Phone Entropy Based Rejection of Background Speech**  
*Yuya Fujita, Ken-ichi Iso, Yahoo! JAPAN, Japan*
- PAGE 3668  
Mon-O-10-6-4  
14:30      **Feature Learning with Raw-Waveform CLDNNs for Voice Activity Detection**  
*Ruben Zazo<sup>1</sup>, Tara N. Sainath<sup>2</sup>, Gabor Simko<sup>2</sup>, Carolina Parada<sup>2</sup>*  
*<sup>1</sup>Universidad Autónoma de Madrid, Spain; <sup>2</sup>Google, USA*
- PAGE 3673  
Mon-O-10-6-5  
14:50      **The SRI System for the NIST OpenSAD 2015 Speech Activity Detection Evaluation**  
*Martin Graciararena<sup>1</sup>, Luciana Ferrer<sup>2</sup>, Vikramjit Mitra<sup>1</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>Universidad de Buenos Aires, Argentina*
- PAGE 3678  
Mon-O-10-6-6  
15:10      **Model Adaptation and Active Learning in the BBN Speech Activity Detection System for the DARPA RATS Program**  
*Damianos Karakos, Scott Novotney, Le Zhang, Richard Schwartz, Raytheon BBN Technologies, USA*

## Mon-P-10-1: Spoken Term Detection

Chair: Nancy Chen

Room: Pacific Concourse - Poster A, Time 13:30-15:30, Monday, 12 Sept. 2016

---

- PAGE 3683  
Mon-P-10-1-1  
13:30      **Fusion Strategies for Robust Speech Recognition and Keyword Spotting for Channel- and Noise-Degraded Speech**  
*Vikramjit Mitra<sup>1</sup>, Julien VanHout<sup>1</sup>, Wen Wang<sup>1</sup>, Chris Bartels<sup>1</sup>, Horacio Franco<sup>1</sup>, Dimitra Vergyri<sup>1</sup>, Abeer Alwan<sup>2</sup>, Adam Janin<sup>3</sup>, John H.L. Hansen<sup>4</sup>, Richard M. Stern<sup>5</sup>, Abhijeet Sangwan<sup>4</sup>, Nelson Morgan<sup>3</sup>*  
*<sup>1</sup>SRI International, USA; <sup>2</sup>University of California at Los Angeles, USA; <sup>3</sup>ICSI, USA; <sup>4</sup>University of Texas at Dallas, USA; <sup>5</sup>Carnegie Mellon University, USA*
- PAGE 3688  
Mon-P-10-1-2  
13:30      **Recurrent Neural Network-Based Phoneme Sequence Estimation Using Multiple ASR Systems' Outputs for Spoken Term Detection**  
*Naoki Sawada, Hiromitsu Nishizaki, University of Yamanashi, Japan*
- PAGE 3693  
Mon-P-10-1-3  
13:30      **Enhancing Data-Driven Phone Confusions Using Restricted Recognition**  
*Mark Kane<sup>1</sup>, Julie Carson-Berndsen<sup>2</sup>*  
*<sup>1</sup>Daon, Ireland; <sup>2</sup>University College Dublin, Ireland*
- PAGE 3698  
Mon-P-10-1-4  
13:30      **Rapid Update of Multilingual Deep Neural Network for Low-Resource Keyword Search**  
*Chongjia Ni<sup>1</sup>, Lei Wang<sup>1</sup>, Cheung-Chi Leung<sup>1</sup>, Feng Rao<sup>2</sup>, Li Lu<sup>2</sup>, Bin Ma<sup>1</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>Tencent, China*
- PAGE 3703  
Mon-P-10-1-5  
13:30      **Toward High-Performance Language-Independent Query-by-Example Spoken Term Detection for MediaEval 2015: Post-Evaluation Analysis**  
*Cheung-Chi Leung<sup>1</sup>, Lei Wang<sup>1</sup>, Haihua Xu<sup>2</sup>, Jingyong Hou<sup>3</sup>, Van Tung Pham<sup>2</sup>, Hang Lv<sup>3</sup>, Lei Xie<sup>3</sup>, Xiong Xiao<sup>2</sup>, Chongjia Ni<sup>1</sup>, Bin Ma<sup>1</sup>, Eng Siong Chng<sup>2</sup>, Haizhou Li<sup>1</sup>*  
*<sup>1</sup>A\*STAR, Singapore; <sup>2</sup>NTU, Singapore; <sup>3</sup>Northwestern Polytechnical University, China*

## Mon-P-10-2 : Speech Enhancement and Noise Reduction

Chairs: Hynek Hermansky, Hemant Patil

Room: Pacific Concourse - Poster B, Time 13:30-15:30, Monday, 12 Sept. 2016

---

- PAGE 3708  
Mon-P-10-2-1  
13:30  
**Novel Subband Autoencoder Features for Non-Intrusive Quality Assessment of Noise Suppressed Speech**  
*Meet H. Soni, Hemant A. Patil, DA-IICT, India*
- PAGE 3713  
Mon-P-10-2-2  
13:30  
**SNR-Based Progressive Learning of Deep Neural Network for Speech Enhancement**  
*Tian Gao<sup>1</sup>, Jun Du<sup>1</sup>, Li-Rong Dai<sup>1</sup>, Chin-Hui Lee<sup>2</sup>*  
*<sup>1</sup>USTC, China; <sup>2</sup>Georgia Institute of Technology, USA*
- PAGE 3718  
Mon-P-10-2-3  
13:30  
**A Novel Risk-Estimation-Theoretic Framework for Speech Enhancement in Nonstationary and Non-Gaussian Noise Conditions**  
*Jishnu Sadasivan, Chandra Sekhar Seelamantula, Indian Institute of Science, India*
- PAGE 3723  
Mon-P-10-2-4  
13:30  
**Two-Stage Temporal Processing for Single-Channel Speech Enhancement**  
*Suman Samui, Indrajit Chakrabarti, Soumya Kanti Ghosh, IIT Kharagpur, India*
- PAGE 3728  
Mon-P-10-2-5  
13:30  
**A Class-Specific Speech Enhancement for Phoneme Recognition: A Dictionary Learning Approach**  
*Nazreen P.M., A.G. Ramakrishnan, Prasanta Kumar Ghosh, Indian Institute of Science, India*
- PAGE 3733  
Mon-P-10-2-6  
13:30  
**Robust Example Search Using Bottleneck Features for Example-Based Speech Enhancement**  
*Atsunori Ogawa<sup>1</sup>, Shogo Seki<sup>1</sup>, Keisuke Kinoshita<sup>1</sup>, Marc Delcroix<sup>1</sup>, Takuya Yoshioka<sup>1</sup>, Tomohiro Nakatani<sup>1</sup>, Kazuya Takeda<sup>2</sup>*  
*<sup>1</sup>NTT, Japan; <sup>2</sup>Nagoya University, Japan*
- PAGE 3738  
Mon-P-10-2-7  
13:30  
**Speech Enhancement in Multiple-Noise Conditions Using Deep Neural Networks**  
*Anurag Kumar<sup>1</sup>, Dinei Florencio<sup>2</sup>*  
*<sup>1</sup>Carnegie Mellon University, USA; <sup>2</sup>Microsoft, USA*

- PAGE 3743  
Mon-P-10-2-8  
13:30  
**Perception Optimized Deep Denoising AutoEncoders for Speech Enhancement**  
*Prashanth Gurunath Shivakumar, Panayiotis Georgiou, University of Southern California, USA*
- PAGE 3748  
Mon-P-10-2-9  
13:30  
**HMM-Based Speech Enhancement Using Sub-Word Models and Noise Adaptation**  
*Akihiro Kato, Ben Milner, University of East Anglia, UK*
- PAGE 3753  
Mon-P-10-2-10  
13:30  
**Semi-Supervised Joint Enhancement of Spectral and Cepstral Sequences of Noisy Speech**  
*Li Li<sup>1</sup>, Hirokazu Kameoka<sup>1</sup>, Takuya Higuchi<sup>2</sup>, Hiroshi Saruwatari<sup>1</sup>*  
*<sup>1</sup>University of Tokyo, Japan; <sup>2</sup>NTT, Japan*
- PAGE 3758  
Mon-P-10-2-11  
13:30  
**A priori SNR Estimation Using a Generalized Decision Directed Approach**  
*Aleksej Chinaev, Reinhold Haeb-Umbach, Universität Paderborn, Germany*
- PAGE 3763  
Mon-P-10-2-12  
13:30  
**A DNN-HMM Approach to Non-Negative Matrix Factorization Based Speech Enhancement**  
*Ziteng Wang, Xu Li, Xiaofei Wang, Qiang Fu, Yonghong Yan, Chinese Academy of Sciences, China*
- PAGE 3768  
Mon-P-10-2-13  
13:30  
**SNR-Aware Convolutional Neural Network Modeling for Speech Enhancement**  
*Szu-Wei Fu<sup>1</sup>, Yu Tsao<sup>1</sup>, Xugang Lu<sup>2</sup>*  
*<sup>1</sup>Academia Sinica, Taiwan; <sup>2</sup>NICT, Japan*
- PAGE 3773  
Mon-P-10-2-14  
13:30  
**An Iterative Phase Recovery Framework with Phase Mask for Spectral Mapping with an Application to Speech Enhancement**  
*Kehuang Li<sup>1</sup>, Bo Wu<sup>2</sup>, Chin-Hui Lee<sup>1</sup>*  
*<sup>1</sup>Georgia Institute of Technology, USA; <sup>2</sup>Xidian University, China*
- PAGE 3778  
Mon-P-10-2-15  
13:30  
**A Novel Research to Artificial Bandwidth Extension Based on Deep BLSTM Recurrent Neural Networks and Exemplar-Based Sparse Representation**  
*Bin Liu, Jianhua Tao, Chinese Academy of Sciences, China*



## Mon-P-10-3 : Far-Field, Robustness and Adaptation

Chair: Shinji Watanabe

Room: Pacific Concourse - Poster C, Time 13:30-15:30, Monday, 12 Sept. 2016

---

- PAGE 3783  
Mon-P-10-3-1  
13:30      **Coping with Unseen Data Conditions: Investigating Neural Net Architectures, Robust Features, and Information Fusion for Robust Speech Recognition**  
*Vikramjit Mitra, Horacio Franco, SRI International, USA*
- PAGE 3788  
Mon-P-10-3-2  
13:30      **On the Use of Gaussian Mixture Model Framework to Improve Speaker Adaptation of Deep Neural Network Acoustic Models**  
*Natalia Tomashenko<sup>1</sup>, Yuri Khokhlov<sup>2</sup>, Yannick Estève<sup>1</sup>*  
*<sup>1</sup>LIUM, France; <sup>2</sup>STC-innovations, Russia*
- PAGE 3793  
Mon-P-10-3-3  
13:30      **Analytical Assessment of Dual-Stream Merging for Noise-Robust ASR**  
*Louis ten Bosch<sup>1</sup>, Bert Cranen<sup>1</sup>, Yang Sun<sup>2</sup>*  
*<sup>1</sup>Radboud Universiteit Nijmegen, The Netherlands; <sup>2</sup>Nuance Communications, Germany*
- PAGE 3798  
Mon-P-10-3-4  
13:30      **Use of Generalised Nonlinearity in Vector Taylor Series Noise Compensation for Robust Speech Recognition**  
*Erfan Loweimi, Jon Barker, Thomas Hain, University of Sheffield, UK*
- PAGE 3803  
Mon-P-10-3-5  
13:30      **Joint Optimization of Denoising Autoencoder and DNN Acoustic Model Based on Multi-Target Learning for Noisy Speech Recognition**  
*Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara, Kyoto University, Japan*
- PAGE 3808  
Mon-P-10-3-6  
13:30      **Optimization of Speech Enhancement Front-End with Speech Recognition-Level Criterion**  
*Takuya Higuchi, Takuya Yoshioka, Tomohiro Nakatani, NTT, Japan*

- PAGE 3813  
Mon-P-10-3-7  
13:30      **Factorized Linear Input Network for Acoustic Model Adaptation in Noisy Conditions**  
*Dung T. Tran, Marc Delroix, Atsunori Ogawa, Tomohiro Nakatani, NTT, Japan*
- PAGE 3818  
Mon-P-10-3-8  
13:30      **Data Augmentation Using Multi-Input Multi-Output Source Separation for Deep Neural Network Based Acoustic Modeling**  
*Yusuke Fujita, Ryoich Takashima, Takeshi Homma, Masahito Togami, Hitachi, Japan*
- PAGE 3823  
Mon-P-10-3-9  
13:30      **Microphone Distance Adaptation Using Cluster Adaptive Training for Robust Far Field Speech Recognition**  
*Animesh Prasad, Khe Chai Sim, NUS, Singapore*
- PAGE 3828  
Mon-P-10-3-10  
13:30      **An Investigation on the Use of i-Vectors for Robust ASR**  
*Dimitrios Dimitriadis<sup>1</sup>, Samuel Thomas<sup>1</sup>, Sriram Ganapathy<sup>2</sup>*  
*<sup>1</sup>IBM, USA; <sup>2</sup>Indian Institute of Science, India*
- PAGE 3833  
Mon-P-10-3-11  
13:30      **The Sheffield Wargame Corpus — Day Two and Day Three**  
*Yulan Liu<sup>1</sup>, Charles Fox<sup>2</sup>, Madina Hasan<sup>1</sup>, Thomas Hain<sup>1</sup>*  
*<sup>1</sup>University of Sheffield, UK; <sup>2</sup>University of Leeds, UK*
- PAGE 3838  
Mon-P-10-3-12  
13:30      **Recurrent Models for Auditory Attention in Multi-Microphone Distant Speech Recognition**  
*Suyoun Kim, Ian Lane, Carnegie Mellon University, USA*
- PAGE 3843  
Mon-P-10-3-13  
13:30      **Semi-Supervised Speaker Adaptation for In-Vehicle Speech Recognition with Deep Neural Networks**  
*Wonkyum Lee<sup>1</sup>, Kyu J. Han<sup>2</sup>, Ian Lane<sup>1</sup>*  
*<sup>1</sup>Carnegie Mellon University, USA; <sup>2</sup>Ford, USA*



## Mon-P-10-4: Low Resource Speech Recognition

Chair: Hung-Yi Lee

Room: Pacific Concourse - Poster D, Time 13:30-15:30, Monday, 12 Sept. 2016

---

- PAGE 3848  
Mon-P-10-4-1  
13:30      **Semi-Supervised Training in Deep Learning Acoustic Model**  
*Yan Huang, Yongqiang Wang, Yifan Gong, Microsoft, USA*
- PAGE 3853  
Mon-P-10-4-2  
13:30      **Multilingual Data Selection for Low Resource Speech Recognition**  
*Samuel Thomas, Kartik Audhkhasi, Jia Cui, Brian Kingsbury, Bhuvana Ramabhadran, IBM, USA*
- PAGE 3858  
Mon-P-10-4-3  
13:30      **An Investigation on Training Deep Neural Networks Using Probabilistic Transcriptions**  
*Amit Das, Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA*
- PAGE 3863  
Mon-P-10-4-4  
13:30      **Analysis of Mismatched Transcriptions Generated by Humans and Machines for Under-Resourced Languages**  
*Van Hai Do<sup>1</sup>, Nancy F. Chen<sup>2</sup>, Boon Pang Lim<sup>2</sup>, Mark Hasegawa-Johnson<sup>1</sup>*  
*<sup>1</sup>ADSC, Singapore; <sup>2</sup>A\*STAR, Singapore*
- PAGE 3868  
Mon-P-10-4-5  
13:30      **ASR for South Slavic Languages Developed in Almost Automated Way**  
*Jan Nouza, Radek Safarik, Petr Cervá, Technical University of Liberec, Czech Republic*
- PAGE 3873  
Mon-P-10-4-6  
13:30      **Improving Under-Resourced Language ASR Through Latent Subword Unit Space Discovery**  
*Marzieh Razavi, Mathew Magimai-Doss, Idiap Research Institute, Switzerland*
- PAGE 3878  
Mon-P-10-4-7  
13:30      **Language Adaptive DNNs for Improved Low Resource Speech Recognition**  
*Markus Müller, Sebastian Stüker, Alex Waibel, KIT, Germany*
- PAGE 3883  
Mon-P-10-4-8  
13:30      **Improved Multilingual Training of Stacked Neural Network Acoustic Models for Low Resource Languages**  
*Tanel Alumäe, Stavros Tsakalidis, Richard Schwartz, Raytheon BBN Technologies, USA*