16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015)

Speech Beyond Speech Towards a Better Understanding of the Most Important Biosignal

Dresden, Germany
6 – 10 September 2015

Volume 1 of 5

ISBN: 978-1-5108-1790-6
Oral Session 1: Feature Extraction and Modeling with Neural Networks

Chairs: Peter Bell, Bhuvana Ramabhadran
Room: Large Hall, Time 11:00 – 13:00, Monday, September 7, 2015

Learning the Speech Front-End with Raw Waveform CLDNNs
Tara N. Sainath, Ron J. Weiss, Andrew Senior, Kevin W. Wilson, Oriol Vinyals, Google, USA

Architectures for Deep Neural Network Based Acoustic Models Defined Over Windowed Speech Waveforms
Mayank Bhargava, Richard Rose, McGill University, Canada

Analysis of CNN-Based Speech Recognition System Using Raw Speech as Input
Dimitri Palaz 1, Mathew Magimai-Doss 1, Ronan Collobert 2
1Idiap Research Institute, Switzerland; 2Facebook, USA

Bilinear Map of Filter-Bank Outputs for DNN-Based Speech Recognition
Tetsuji Ogawa 1, Kenshiro Ueda 1, Kouichi Katsurada 2, Tetsunori Kobayashi 1, Tsuneo Nitta 1
1Waseda University, Japan; 2Toyoashi University of Technology, Japan

Speech Recognition with Temporal Neural Networks
Payton Lin 1, Dau-Cheng Lyu 2, Yun-Fan Chang 1, Yu Tsao 1
1Academia Sinica, Taiwan; 2ASUS, Taiwan

Convolutional Neural Networks for Acoustic Modeling of Raw Time Signal in LVCSR
Pavel Golk, Zoltán Tüske, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

Oral Session 2: Prosody

Chairs: Štefan Beňuš, Katalin Mády
Room: Conference 1, Time 11:00 – 13:00, Monday, September 7, 2015

Stable and Unstable Intervals as a Basic Segmentation Procedure of the Speech Signal
Ulrike Glavitsch 1, Lei He 2, Volker Dellwo 2
1EMPA, Switzerland; 2Universität Zürich, Switzerland

Polysyllabic Shortening and Word-Final Lengthening in English
Andreas Windmann 1, Juraj Simko 2, Petra Wagner 1
1Universität Bielefeld, Germany; 2University of Helsinki, Finland

The Acoustics of Word Stress in English as a Function of Stress Level and Speaking Style
Anders Eriksson, Mattias Heldner, Stockholm University, Sweden

Pitch Accent Distribution in German Infant-Directed Speech
Katharina Zahner, Muna Pohl, Bettina Braun, Universität Konstanz, Germany

Acoustic Correlates of Perceived Syllable Prominence in German
Hansjörg Mixdorff 1, Christian Cossio-Mercado 2, Angelika Hönemann 3, Jorge Gurlekian 2, Diego Evin 2, Humberto Torres 2
1BHT Berlin, Germany; 2Universidad de Buenos Aires, Argentina; 3Universität Bielefeld, Germany

Cross-Modality Matching of Linguistic and Emotional Prosody
Simone Simonetti, Jeesun Kim, Chris Davis, University of Western Sydney, Australia
Oral Session 3: Speech Intelligibility Enhancement

**Chairs:** Paavo Alku, Jon Barker  
**Room:** Conference 2+3, Time 11:00 – 13:00, Monday, September 7, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-1</th>
<th>11:00 – 11:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Fast Algorithm for Improved Intelligibility of Speech-in-Noise Based on Frequency and Time Domain Energy Reallocation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tudor-Cătălin Zorilă, Yannis Stylianou, Toshiba Research Europe, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-2</th>
<th>11:20 – 11:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intelligibility Enhancement of Casual Speech for Reverberant Environments Inspired by Clear Speech Properties</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maria Koutsogiannaki¹, Petko N. Petkov², Yannis Stylianou¹</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹University of Crete, Greece; ²Toshiba Research Europe, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-3</th>
<th>11:40 – 12:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intelligibility Enhancement of Vocal Announcements for Public Address Systems: A Design for All Through a Presbycusis Pre-Compensation Filter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A. Ben Jemaa¹, N. Mechergui¹, G. Courtois², A. Mudry³, S. Djaziri-Larbi⁴, M. Turki⁴, H. Lissek⁵, M. Jaidane⁴</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Université de Tunis El Manar, Tunisia; ²EPFL, Switzerland; ³Stanford University, USA; ⁴Telnet, Tunisia</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-4</th>
<th>12:00 – 12:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model-Based Integration of Reverberation for Noise-Adaptive Near-End Listening Enhancement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Henning Schepker¹, David Hülsmeyer², Jan Rennies², Simon Doclo¹</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Carl von Ossietzky Universität Oldenburg, Germany; ²Fraunhofer IDMT, Germany</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-5</th>
<th>12:20 – 12:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Online Lombard-Adaptation in Incremental Speech Synthesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sebastian Rottschäfer, Hendrik Buschmeier, Herwin van Welbergen, Stefan Kopp, Universität Bielefeld, Germany</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-3-6</th>
<th>12:40 – 13:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comparison of Gaussian Process Regression and Gaussian Mixture Models in Spectral Tilt Modelling for Intelligibility Enhancement of Telephone Speech</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emma Jokinen, Ulpu Remes, Paavo Alku, Aalto University, Finland</td>
<td></td>
</tr>
</tbody>
</table>

INTERSPEECH 2015 Table of Contents

Oral Session 4: Detecting and Predicting Mental and Social Disorders

**Chairs:** Julien Epps, Elizabeth Shriberg  
**Room:** Conference 4+5, Time 11:00 – 13:00, Monday, September 7, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-4-1</th>
<th>11:00 – 11:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Discriminative Reliability-Aware Classification Model with Applications to Intelligibility Classification in Pathological Speech</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naveen Kumar, Shrikanth S. Narayanan, University of Southern California, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-4-2</th>
<th>11:20 – 11:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voiced/Unvoiced Transitions in Speech as a Potential Bio-Marker to Detect Parkinson’s Disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td>J.R. Orozco-Arroyave¹, Florian Höning², J.D. Arias-Londoño¹, J.F. Vargas-Bonilla¹, Sabine Skodda³, J. Rusz⁴, Elmar Nöth²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Universidad de Antioquia, Colombia; ²FAU Erlangen-Nürnberg, Germany; ³Ruhr-Universität Bochum, Germany; ⁴Czech Technical University in Prague, Czech Republic</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-4-3</th>
<th>11:40 – 12:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-Frequency Components Analysis in Running Speech for the Automatic Detection of Parkinson’s Disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T. Villa-Cañas¹, J.D. Arias-Londoño¹, J.R. Orozco-Arroyave¹, J.F. Vargas-Bonilla¹, Elmar Nöth²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Universidad de Antioquia, Colombia; ²FAU Erlangen-Nürnberg, Germany</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-4-4</th>
<th>12:00 – 12:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automatic Detection of Parkinson’s Disease from Continuous Speech Recorded in Non-Controlled Noise Conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>J.C. Vásquez-Correa¹, T. Arias-Vergara¹, J.R. Orozco-Arroyave¹, J.F. Vargas-Bonilla¹, J.D. Arias-Londoño¹, Elmar Nöth²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Universidad de Antioquia, Colombia; ²FAU Erlangen-Nürnberg, Germany</td>
<td></td>
</tr>
<tr>
<td>Page 110</td>
<td>Mon-O-4-5</td>
<td>12:20 – 12:40</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Relevance Vector Machine for Depression Prediction</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Nicholas Cummins</em>¹, <em>Vidhyasaharan Sethu</em>¹, <em>Julien Epps</em>¹, <em>Jarek Krajewski</em>²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹ University of New South Wales, Australia; ² Universität Wuppertal, Germany</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 115</th>
<th>Mon-O-4-6</th>
<th>12:40 – 13:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Typicality and Emotion in the Voice of Children with Autism Spectrum Condition: Evidence Across Three Languages</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹ Technische Universität München, Germany; ² Imperial College London, UK; ³ University of Cambridge, UK; ⁴ Bar-Ilan University, Israel; ⁵ Karolinska Institute, Sweden; ⁶ Universität Paderborn, Germany</td>
</tr>
</tbody>
</table>

---

**Oral Session 5: Spoken Language Understanding**

*Chairs: David Suendermann-Oeft, Hugo Van hamme*

*Room: Conference 6, Time 11:00 – 13:00, Monday, September 7, 2015*

<table>
<thead>
<tr>
<th>Page 120</th>
<th>Mon-O-5-1</th>
<th>11:00 – 11:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Deep Contextual Language Understanding in Spoken Dialogue Systems</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Chunxi Liu</em>¹, <em>Puyang Xu</em>², <em>Ruhi Sarikaya</em>²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹ Johns Hopkins University, USA; ² Microsoft, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 125</th>
<th>Mon-O-5-2</th>
<th>11:20 – 11:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>RNN-Based Labeled Data Generation for Spoken Language Understanding</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Yik-Cheung Tam</em>, <em>Yangyang Shi</em>, <em>Hank Chen</em>, <em>Mei-Yuh Hwang</em>, Microsoft, China</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 130</th>
<th>Mon-O-5-3</th>
<th>11:40 – 12:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Is it Time to Switch to Word Embedding and Recurrent Neural Networks for Spoken Language Understanding?</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Vedran Vukotic</em>¹, <em>Christian Raymond</em>¹, <em>Guillaume Gravier</em>²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹ INSA de Rennes, France; ² INRIA, France</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 135</th>
<th>Mon-O-5-4</th>
<th>12:00 – 12:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Recurrent Neural Network and LSTM Models for Lexical Utterance Classification</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Suman Ravuri</em>¹, <em>Andreas Stolcke</em>²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹ ICSI, USA; ² Microsoft, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 140</th>
<th>Mon-O-5-5</th>
<th>12:20 – 12:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Semantic Retrieval of Personal Photos Using a Deep Autoencoder Fusing Visual Features with Speech Annotations Represented as Word/Paragraph Vectors</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Hung-tsung Lu</em>, <em>Yuan-ming Liou</em>, <em>Hung-yi Lee</em>, <em>Lin-shan Lee</em>, National Taiwan University, Taiwan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page 145</th>
<th>Mon-O-5-6</th>
<th>12:40 – 13:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>A Comparison of Normalization Techniques Applied to Latent Space Representations for Speech Analytics</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Mohamed Morchid</em>, <em>Richard Dufour</em>, <em>Driss Matrouf</em>, LIA, France</td>
</tr>
</tbody>
</table>
**Special Session 1: Active Perception in Human and Machine Speech Communication**  
*Chairs: Martin Heckmann, Dorothea Kolossa*  
*Room: Hall 1, Time 11:00 - 13:00, Monday, September 7, 2015*  

<table>
<thead>
<tr>
<th>Page</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>11:00 - 11:17</td>
<td>The Effect of Soft, Modal and Loud Voice Levels on Entrainment in Noisy Conditions</td>
<td>Éva Székely, Mark T. Keane, Julie Carson-Berndsen, University College Dublin, Ireland</td>
</tr>
<tr>
<td>160</td>
<td>11:34 - 11:51</td>
<td>Exploiting Top-Down Source Models to Improve Binaural Localisation of Multiple Sources in Reverberant Environments</td>
<td>Ning Ma, Guy J. Brown, Jose A. Gonzalez, University of Sheffield, UK</td>
</tr>
</tbody>
</table>
| 165  | 11:51 - 12:08 | Binaural Sound Source Localisation and Tracking Using a Dynamic Spherical Head Model | Christopher Schymura\(^1\), Fiete Winter\(^2\), Dorothea Kolossa\(^1\), Sascha Spors\(^2\)  
\(^1\)Ruhr-Universität Bochum, Germany; \(^2\)Universität Rostock, Germany |
| 170  | 12:08 - 12:25 | The Role of Temporal Resolution in Modulation-Based Speech Segregation | Tobias May, Thomas Bentsen, Torsten Dau, Technical University of Denmark, Denmark            |
| 175  | 12:25 - 12:42 | Improving Automatic Speech Recognition in Spatially-Aware Hearing Aids | Hendrik Kayser, Constantin Spille, Daniel Marquardt, Bernd T. Meyer, Carl von Ossietzky Universität Oldenburg, Germany |

---

**Mon-SP1 continued…**  

<table>
<thead>
<tr>
<th>Page</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>12:42 - 12:59</td>
<td>Dereverberation for Active Human-Robot Communication Robust to Speaker's Face Orientation</td>
<td>Randy Gomez, Levko Ivanchuk, Keisuke Nakamura, Takeshi Mizumoto, Kazuhiro Nakadai, Honda Research Institute Japan, Japan</td>
</tr>
</tbody>
</table>
# Poster Session 1: Speaker Recognition

*Chair: Sandro Cumani*

**Room:** Hall 2, Time 11:00 - 13:00, Monday, September 7, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Poster Session 1-1</th>
</tr>
</thead>
</table>
| 185  | Multi-Task Learning for Text-Dependent Speaker Verification  
Nanxin Chen, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China |
| 190  | JFA for Speaker Recognition with Random Digit Strings  
Themos Stafylakis¹, Patrick Kenny¹, Md. Jahangir Alam¹, Marcel Kockmann²  
¹CRIM, Canada; ²VoiceTrust, Canada |
| 195  | Structured Prediction for Speaker Identification in TV Series  
Elena Knyazeva, Guillaume Wisniewski, Hervé Bredin, François Yvon, LIMSI, France |
| 200  | Speaker Recognition by Means of Acoustic and Phonetically Informed GMMs  
Sandro Cumani, Pietro Laface, Farzana Kulsoom, Politecnico di Torino, Italy |
| 205  | A Fast Approach to Psychoacoustic Model Compensation for Robust Speaker Recognition in Additive Noise  
Ashish Panda, TCS Innovation Labs Mumbai, India |
| 210  | Blind Score Normalization Method for PLDA Based Speaker Recognition  
Danila Doroshin, Nikolay Lubimov, Marina Nastasenko, Mikhail Kotov, Stel Computer Systems, Russia |
| 214  | Non-Linear PLDA for i-Vector Speaker Verification  
Sergey Novoselov, Timur Pekhovsky, Oleg Kudashev, Valentin S. Mendelev, Alexey Prudnikov, Speech Technology Center, Russia |
| 219  | On the Need of Template Protection for Voice Authentication  
Carlos Vaquero, Patricia Rodriguez, AGNITiO, Spain |

*INTERSPEECH 2015 Table of Contents*

---

<table>
<thead>
<tr>
<th>Page</th>
<th>Poster Session 1-9</th>
</tr>
</thead>
</table>
| 224  | Evaluation and Calibration of Short-Term Aging Effects in Speaker Verification  
Finnian Kelly, John H.L. Hansen, University of Texas at Dallas, USA |
| 229  | Phone-Centric Local Variability Vector for Text-Constrained Speaker Verification  
Liping Chen¹, Kong Aik Lee², Bin Ma², Wu Guo¹, Haizhou Li², Li-Rong Dai¹  
¹USTC, China; ²A*STAR, Singapore |
| 234  | Cosine Distance Features for Robust Speaker Verification  
Kuruvachan K. George¹, C. Santhosh Kumar¹, Ramachandran K.I.¹, Ashish Panda²  
¹Amrita University, India; ²TCS Innovation Labs Mumbai, India |
| 239  | Voice Liveness Detection Algorithms Based on Pop Noise Caused by Human Breath for Automatic Speaker Verification  
Sayaka Shiota¹, Fernando Villavicencio², Junichi Yamagishi², Nobutaka Ono², Isao Echizen², Tomoko Matsui³  
¹Tokyo Metropolitan University, Japan; ²NII, Japan; ³ISM, Japan |
| 244  | Noise Robust Speaker Recognition with Convolutive Sparse Coding  
Antti Hurmalainen¹, Rahim Saedi², Tuomas Virtanen¹  
¹Tampere University of Technology, Finland; ²Aalto University, Finland |
| 249  | Combining Amplitude and Phase-Based Features for Speaker Verification with Short Duration Utterances  
Md. Jahangir Alam, Patrick Kenny, Themos Stafylakis, CRIM, Canada |

*INTERSPEECH 2015 Table of Contents*
Poster Session 2: Speech Synthesis 1

Chair: Inma Hernaez
Room: Hall 2, Time 11:00 - 13:00, Monday, September 7, 2015

Page 254
Mon-P-2-1
Phase Perception of the Glottal Excitation of Vocoded Speech
Tuomo Raitio¹, Lauri Juvela¹, Antti Suni², Martti Vainio², Paavo Alku¹
¹Aalto University, Finland; ²University of Helsinki, Finland

Page 259
Mon-P-2-2
Using Acoustics to Improve Pronunciation for Synthesis of Low Resource Languages
Sunayana Sitaram, Serena Jeblee, Alan W. Black, Carnegie Mellon University, USA

Page 264
Mon-P-2-3
Sub-Band Text-to-Speech Combining Sample-Based Spectrum with Statistically Generated Spectrum
Tadashi Inai¹, Sunao Hara¹, Masanobu Abe¹, Yusuke Ijima², Noboru Miyazaki², Hideyuki Mizuno²
¹Okayama University, Japan; ²NTT Corporation, Japan

Page 269
Mon-P-2-4
Pruning Redundant Synthesis Units Based on Static and Delta Unit Appearance Frequency
Heng Lu¹, Wei Zhang¹, Xu Shao¹, Quan Zhou², Wenhui Lei², Hongbin Zhou², Andrew Breen³
¹Nuance Communications, USA; ²Nuance Communications, China; ³Nuance Communications, UK

Page 274
Mon-P-2-5
Emotional Transplant in Statistical Speech Synthesis Based on Emotion Additive Model
Yamato Ohtani, Yu Nasu, Masahiro Morita, Masami Akamine, Toshiba Corporation, Japan

Page 279
Mon-P-2-6
Generalized Variable Parameter HMMs Based Acoustic-to-Articulatory Inversion
Xurong Xie, Xunying Liu, Lan Wang, Rongfeng Su, Chinese Academy of Sciences, China

Page 284
Mon-P-2-7
Semi-Supervised Training of a Voice Conversion Mapping Function Using a Joint-Autoencoder
Seyed Hamidreza Mohammadi, Alexander Kain, Oregon Health & Science University, USA

Page 289
Mon-P-2-8
On Glottal Source Shape Parameter Transformation Using a Novel Deterministic and Stochastic Speech Analysis and Synthesis System
Stefan Huber, Axel Roebel, IRCAM, France

Page 294
Mon-P-2-9
Fluent Personalized Speech Synthesis with Prosodic Word-Level Spontaneous Speech Generation
Yi-Chin Huang, Chung-Hsien Wu, Ming-Ge Shie, National Cheng-Kung University, Taiwan

Page 299
Mon-P-2-10
Non-Native Speech Synthesis Preserving Speaker Individuality Based on Partial Correction of Prosodic and Phonetic Characteristics
Yuji Oshima, Shinnosuke Takamichi, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan

Page 304
Mon-P-2-11
Evaluation of State Mapping Based Foreign Accent Conversion
Markus Toman, Michael Pucher, FTW, Austria

Page 309
Mon-P-2-12
Minimum Trajectory Error Training for Deep Neural Networks, Combined with Stacked Bottleneck Features
Zhizheng Wu, Simon King, University of Edinburgh, UK
Poster Session 3: Mining and Annotation of Spoken and Multimodal Resources

Chair: Murat Saraçlar
Room: Foyer, Time 11:00 – 13:00, Monday, September 7, 2015

Anomaly-Based Annotation Errors Detection in TTS Corpora
Jindřich Matoušek, Daniel Tihelka, University of West Bohemia, Czech Republic

Analysing Automatic Descriptions of Intonation with ICARUS
Katrin Schweitzer, Markus Gärtner, Arndt Riester, Ina Rösiger, Kerstin Eckart, Jonas Kuhn, Grzegorz Dogil, Universität Stuttgart, Germany

iCALL Corpus: Mandarin Chinese Spoken by Non-Native Speakers of European Descent
Nancy F. Chen, Rong Tong, Darren Wee, Peixuan Lee, Bin Ma, Haizhou Li, A*STAR, Singapore

Development of a Cantonese Dysarthric Speech Corpus
Ka Ho Wong¹, Yu Ting Yeung¹, Edwin H.Y. Chan¹, Patrick C.M. Wong¹, Gina-Anne Levow², Helen Meng¹
¹Chinese University of Hong Kong, China; ²University of Washington, USA

StyleX: A Corpus of Educational Videos for Research on Speaking Styles and Their Impact on Engagement and Learning
Harish Arsikere, Sonal Patil, Ranjeet Kumar, Kundan Shrivastava, Om Deshmukh, Xerox Research Center India, India

A Dialog Act Tagging Approach to Behavioral Coding: A Case Study of Addiction Counseling Conversations
Doğan Can¹, David C. Atkins², Shrikanth S. Narayanan¹
¹University of Southern California, USA; ²University of Washington, USA

Analysing Rhythm in Ritual Discourse in Yucatec Maya Using Automatic Speech Alignment
Valentina Vapnarsky¹, Claude Barras², Cédric Becquey¹, David Doukhan³, Martine Adda-Decker², Lori Lamel²
¹LESC (UMR 7186), France; ²LIMSI, France; ³IRCAM, France

Noise-Matched Training of CRF Based Sentence End Detection Models
Madina Hasan, Rama Duddipatla, Thomas Hain, University of Sheffield, UK

The Effect of Spectral Slope on Pitch Perception
Jianjing Kuang, Mark Liberman, University of Pennsylvania, USA

Special Event: Speaker Comparison for Forensic and Investigative Applications
Chairs: Jean-François Bonastre, Joseph P. Campbell
Room: Large Hall, Time 14:30 – 16:30, Monday, September 7, 2015
Combined Cine- and Tagged-MRI for Tracking Landmarks on the Tongue Surface
Honghao Bao¹, Wenhuan Lu¹, Kiyoshi Honda¹, Jianguo Wei¹, Qiang Fang², Jianwu Dang¹
¹Tianjin University, China; ²Chinese Academy of Social Sciences, China

Human Vocal Tract Growth: A Longitudinal Study of the Development of Various Anatomical Structures
Guillaume Barbier¹, Louis-Jean Boë¹, Guillaume Captier², Rafael Laboissière³
¹GIPSA, France; ²Université de Montpellier, France; ³LPNC, France

Analysis of Coarticulated Speech Using Estimated Articulatory Trajectories
Ganesh Sivaraman¹, Vikramjit Mitra², Mark K. Tiede³, Elliot Saltzman⁴, Louis Goldstein⁵, Carol Espy-Wilson¹
¹University of Maryland, USA; ²SRI International, USA; ³Haskins Laboratories, USA; ⁴Boston University, USA; ⁵University of Southern California, USA

Speech Planning in 4-Year-Old Children versus Adults: Acoustic and Articulatory Analyses
Guillaume Barbier¹, Pascal Perrier¹, Lucie Ménard², Yohan Payan³, Mark K. Tiede⁴, Joseph S. Perkell⁵
¹GIPSA, France; ²Université du Québec à Montréal, Canada; ³TIMC-IMAG, France; ⁴Haskins Laboratories, USA; ⁵MIT, USA

Morphological and Acoustic Analysis of the Vocal Tract Using a Multi-Speaker Volumetric MRI Dataset
Tokihiko Kaburagi, Kyushu University, Japan

Experimental Assessment of the Tongue Incompressibility Hypothesis During Speech Production
Zisis Iason Skordilis¹, Vikram Ramanarayanan², Louis Goldstein¹, Shrikanth S. Narayanan¹
¹University of Southern California, USA; ²Educational Testing Service, USA
Oral Session 7: Deep Neural Networks in Language and Accent Recognition

**Chairs:** Ville Hautamäki, Michael Wagner

**Room:** Conference 2+3, Time 14:30 – 16:30, Monday, September 7, 2015

**Page 389**

Multilingual Bottleneck Features for Language Recognition

Radek Fér, Pavel Matějka, František Grézl, Oldřich Plchot, Jan Černocký, Brno University of Technology, Czech Republic

**Page 394**

DNN Senone MAP Multinomial i-Vectors for Phonotactic Language Recognition

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

**Page 398**

Deep Bottleneck Network Based I-Vector Representation for Language Identification

Yan Song, Xinhai Hong, Bing Jiang, Ruilian Cui, Ian McLoughlin, Li-Rong Dai, USTC, China

**Page 403**

An End-to-End Approach to Language Identification in Short Utterances Using Convolutional Neural Networks

Alicia Lozano-Diez, Ruben Zazo-Candil, Javier Gonzalez-Domínguez, Doroteo T. Toledano, Joaquin Gonzalez-Rodriguez, Universidad Autónoma de Madrid, Spain

**Page 408**

Boosting Universal Speech Attributes Classification with Deep Neural Network for Foreign Accent Characterization

Ville Hautamäki ¹, Sabato Marco Siniscalchi ², Hamid Behravan ¹, Valerio Mario Salerno ², Ivan Kukanov ¹

¹ University of Eastern Finland, Finland; ² Università di Enna Kore, Italy

**Page 413**

Multilingual Tandem Bottleneck Feature for Language Identification

Wang Geng, Jie Li, Shanshan Zhang, Xinyuan Cai, Bo Xu, Chinese Academy of Sciences, China

Oral Session 8: Speech Transmission

**Chairs:** Tom Bäckström, Gernot Kubin

**Room:** Conference 4+5, Time 14:30 – 16:30, Monday, September 7, 2015

**Page 418**

On Compressibility of Neural Network Phonological Features for Low Bit Rate Speech Coding

Afsaneh Asaei, Milos Cernak, Hervé Bourlard, Idiap Research Institute, Switzerland

**Page 423**

Robust and Accurate LSF Location with Laguerre Method

Michał Lenarczyk, Polish Academy of Sciences, Poland

**Page 428**

Interactivity-Aware Playout Adaptation

Jochen Issing ¹, Nikolaus Färber ², Reinhard German ¹

¹ FAU Erlangen-Nürnberg, Germany; ² Fraunhofer IIS, Germany

**Page 433**

Advanced Time Shrinking Using a Drop Classifier Based on Codec Features

Jochen Issing ¹, Nikolaus Färber ², Reinhard German ¹

¹ FAU Erlangen-Nürnberg, Germany; ² Fraunhofer IIS, Germany

**Page 438**

Measuring and Monitoring Speech Quality for Voice Over IP with POLQA, ViSQOL and P.563

Andrew Hines ¹, Eoin Gillen ², Naomi Harte ²

¹ Dublin Institute of Technology, Ireland; ² Trinity College Dublin, Ireland

**Page 443**

Towards the Prediction of Human Speaker Identification Performance from Measured Speech Quality

Laura Fernández Gallardo ¹, Sebastian Möller ²

¹ University of Canberra, Australia; ² T-Labs, Germany
Oral Session 9: Language Modeling for Conversational Speech

Chairs: Michael Levit, Hsin Min Wang
Room: Conference 6, Time 14:30 – 16:30, Monday, September 7, 2015

- **Personalization of Word-Phrase-Entity Language Models**
  M. Levit, Andreas Stolcke, R. Subba, S. Parthasarathy, S. Chang, S. Xie, T. Anastasakis, Benoit Dumoulin, Microsoft, USA

- **Discriminative Bilinear Language Modeling for Broadcast Transcriptions**
  Akio Kobayashi, Manon Ichiki, Takahiro Oku, Kazuo Onoe, Shoel Sato, NHK, Japan

- **Recognize Foreign Low-Frequency Words with Similar Pairs**
  Xi Ma, Xiaoxi Wang, Dong Wang, Zhiyong Zhang, Tsinghua University, China

- **Combinations of Various Language Model Technologies Including Data Expansion and Adaptation in Spontaneous Speech Recognition**
  Ryo Masumura 1, Taichi Asami 1, Takanobu Oba 1, Hirokazu Masataki 1, Sumitaka Sakauchi 1, Akinori Ito 2
  1 NTT Corporation, Japan; 2 Tohoku University, Japan

- **Bringing Contextual Information to Google Speech Recognition**
  Petar Aleksic, Mohammadreza Ghodsi, Assaf Michaely, Cyril Allauzen, Keith Hall, Brian Roark, David Rybach, Pedro Moreno, Google, USA

Special Session 2a: Interspeech 2015 Computational Paralinguistics Challenge (ComParE): Degree of Nativeness, Parkinson’s & Eating Condition

Chairs: Anton Batliner, Stefan Steidl
Room: Hall 1, Time 14:30 – 16:30, Monday, September 7, 2015

- **The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson’s & Eating Condition**
  Björn Schuller 1, Stefan Steidl 2, Anton Batliner 2, Simone Hanke 3, Florian Hönig 2, J.R. Orozco-Arroyave 2, Elmar Nöth 2, Yue Zhang 3, Felix Weninger 3
  1 Imperial College London, UK; 2 FAU Erlangen-Nürnberg, Germany; 3 Technische Universität München, Germany

- **The Degree of Nativeness Sub-Challenge: The Data**
  Florian Hönig, FAU Erlangen-Nürnberg, Germany

- **Phrase Accentuation Verification and Phonetic Variation Measurement for the Degree of Nativeness Sub-Challenge**
  Claude Montacié, Marie-José Caraty, STIH (EA 4509), France

- **Combining Multiple Approaches to Predict the Degree of Nativeness**
  Eugénio Ribeiro 1, Jaime Ferreira 1, Julia Olcoz 2, Alberto Abad 1, Helena Moniz 1, Fernando Batista 1, Isabel Trancoso 1
  1 INESC-ID Lisboa, Portugal; 2 Universidad de Zaragoza, Spain
<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
<th>Authors</th>
<th>Institution(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:10 – 15:20</td>
<td>Automated Evaluation of Non-Native English Pronunciation Quality: Combining Knowledge- and Data-Driven Features at Multiple Time Scales</td>
<td>Matthew P. Black, Daniel Bone, Zisis Iason Skordilis, Rahul Gupta, Wei Xia, Pavlos Papadopoulos, Sanjeev Nallan Chakravarthula, Bo Xia, Maarten Van Segbroeck, Jangwon Kim, Panayiotis G. Georgiou, Shrikanth S. Narayanan</td>
<td>University of Southern California, USA</td>
</tr>
<tr>
<td>15:20 – 15:30</td>
<td>The Parkinson’s Condition Sub-Challenge: The Data</td>
<td>J.R. Orozco-Arroyave</td>
<td>Universidad de Antioquia, Colombia</td>
</tr>
<tr>
<td>15:30 – 15:40</td>
<td>Estimating the Severity of Parkinson’s Disease from Speech Using Linear Regression and Database Partitioning</td>
<td>Dávid Sztahó, Gábor Kiss, Klára Vicsi</td>
<td>BME, Hungary</td>
</tr>
<tr>
<td>15:40 – 15:50</td>
<td>Random Forest-Based Prediction of Parkinson’s Disease Progression Using Acoustic, ASR and Intelligibility Features</td>
<td>Alexander Zlotnik¹, Juan M. Montero¹, Rubén San-Segundo¹, Ascensión Gallardo-Antolín²</td>
<td>Universidade Politécnica de Madrid, Spain; Universidad Carlos III de Madrid, Spain</td>
</tr>
<tr>
<td>15:50 – 16:00</td>
<td>Automatic Recognition of Unified Parkinson’s Disease Rating from Speech with Acoustic, i-Vector and Phonotactic Features</td>
<td>Guozhen An¹, David Guy Brizan¹, Min Ma¹, Michelle Morales¹, Ali Raza Syed¹, Andrew Rosenberg²</td>
<td>CUNY Graduate Center, USA; CUNY Queens College, USA</td>
</tr>
<tr>
<td>16:00 – 16:10</td>
<td>Parkinson’s Condition Estimation Using Speech Acoustic and Inversely Mapped Articulatory Data</td>
<td>Seongjun Hahm, Jun Wang</td>
<td>University of Texas at Dallas, USA</td>
</tr>
<tr>
<td>16:10 – 16:20</td>
<td>Segment-Dependent Dynamics in Predicting Parkinson’s Disease</td>
<td>James R. Williamson¹, Thomas F. Quatieri¹, Brian S. Helfer¹, Joseph Perricone¹, Satrajit S. Ghosh², Gregory Ciccarelli¹, Daryush D. Mehta¹</td>
<td>MIT Lincoln Laboratory, USA; MIT, USA</td>
</tr>
<tr>
<td>16:20 – 16:30</td>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Poster Session 4: Pronunciation, Prosody and Audiovisual Features and Models

Chair: Anton Ragni
Room: Hall 2, Time 14:30 - 16:30, Monday, September 7, 2015

PAGE 523
Mon-P-4-1
Recognition of Voiced Sounds with a Continuous State HMM
S.M. Houghton, Colin J. Champion, Philip Weber, University of Birmingham, UK

PAGE 528
Mon-P-4-2
Learning Speech Rate in Speech Recognition
Xiangyu Zeng, Shi Yin, Dong Wang, Tsinghua University, China

PAGE 533
Mon-P-4-3
Pronunciation and Silence Probability Modeling for ASR
Guoguo Chen, Hainan Xu, Minhua Wu, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

PAGE 538
Mon-P-4-4
Exploring Minimal Pronunciation Modeling for Low Resource Languages
Marelie Davel¹, Etienne Barnard¹, Charl van Heerden¹, William Hartmann², Damianos Karakos², Richard Schwartz², Stavros Tsakalidis²
¹North-West University, South Africa; ²Raytheon BBN Technologies, USA

PAGE 543
Mon-P-4-5
Attribute Knowledge Integration for Speech Recognition Based on Multi-Task Learning Neural Networks
Hao Zheng¹, Zhanlei Yang¹, Liwei Qiao², Jianping Li², Wenju Liu¹
¹Chinese Academy of Sciences, China; ²SGCC, China

PAGE 548
Mon-P-4-6
Detecting Audio-Visual Synchrony Using Deep Neural Networks
Etienne Marcheret¹, Gerasimos Potamianos², Josef Vopicka³, Vaibhava Goel¹
¹IBM T.J. Watson Research Center, USA; ²University of Thessaly, Greece; ³IBM, Czech Republic

PAGE 553
Mon-P-4-7
Cross Database Training of Audio-Visual Hidden Markov Models for Phone Recognition
Shahram Kalantari, David Dean, Houman Ghaemmaghami, Sridha Sridharan, Clinton Fookes, Queensland University of Technology, Australia

PAGE 558
Mon-P-4-8
Incorporating Visual Information for Spoken Term Detection
Shahram Kalantari, David Dean, Sridha Sridharan, Queensland University of Technology, Australia

PAGE 563
Mon-P-4-9
Integration of Deep Bottleneck Features for Audio-Visual Speech Recognition
Hiroshi Ninomiya¹, Norihide Kitaoka², Satoshi Tamura³, Yurie Iribe⁴, Kazuya Takeda¹
¹Nagoya University, Japan; ²University of Tokushima, Japan; ³Gifu University, Japan; ⁴Aichi Prefectural University, Japan

PAGE 568
Mon-P-4-10
Automatic Detection of Sentence Prominence in Speech Using Predictability of Word-Level Acoustic Features
Sofoklis Kakouros, Okko Räsänen, Aalto University, Finland

PAGE 573
Mon-P-4-11
An Empirical Model of Emphatic Word Detection
Milos Cernak, Pierre-Edouard Honnet, Idiap Research Institute, Switzerland

PAGE 578
Mon-P-4-12
Using Tilt for Automatic Emphasis Detection with Bayesian Networks
Yishuang Ning¹, Zhiyong Wu¹, Xiaoyan Lou², Helen Meng¹, Jia Jia¹, Lianhong Cai¹
¹Tsinghua University, China; ²Beijing Samsung Telecom R&D Center, China
Poster Session 5: Speech Analysis and Representation
Chair: Jen-Tzung Chien
Room: Hall 2, Time 14:30 - 16:30, Monday, September 7, 2015

PAGE 583
Mon-P-5-1
Analysis of a Low-Dimensional Bottleneck Neural Network Representation of Speech for Modelling Speech Dynamics
Linxue Bai, Peter Jančovič, Martin Russell, Philip Weber, University of Birmingham, UK

PAGE 588
Mon-P-5-2
Statistical Acoustic-to-Articulatory Mapping Unified with Speaker Normalization Based on Voice Conversion
Hidetsugu Uchida, Daisuke Saito, Nobuaki Minematsu, Keikichi Hirose, University of Tokyo, Japan

PAGE 593
Mon-P-5-3
Analysis of Features from Analytic Representation of Speech Using MP-ABX Measures
Raghavendra Reddy Pappagari, Karthika Vijayan, K. Sri Rama Murty, IIT Hyderabad, India

PAGE 598
Mon-P-5-4
Source-Filter Separation of Speech Signal in the Phase Domain
Erfan Loweimi, Jon Barker, Thomas Hain, University of Sheffield, UK

PAGE 603
Mon-P-5-5
A Maximum Likelihood Approach to the Detection of Moments of Maximum Excitation and Its Application to High-Quality Speech Parameterization
Ranniery Maia¹, Yannis Stylianou¹, Masami Akamine²
¹Toshiba Research Europe, UK; ²Toshiba Corporation, Japan

PAGE 608
Mon-P-5-6
SABR: Sparse, Anchor-Based Representation of the Speech Signal
Christopher Liberatore, Sandesh Aryal, Zelun Wang, Seth Polsley, Ricardo Gutierrez-Osuna, Texas A&M University, USA

INTERSPEECH 2015 Table of Contents

Mon-P-5 continued…

PAGE 613
Mon-P-5-7
Automatic Transformation of Irregular to Regular Voice by Residual Analysis and Synthesis
Tamás Gábor Csapó, Géza Németh, BME, Hungary

PAGE 618
Mon-P-5-8
Optical Sensor Calibration for Electro-Optical Stomatography
Simon Preuß, Peter Birkholz, Technische Universität Dresden, Germany

PAGE 623
Mon-P-5-9
From Text to Formants — Indirect Model for Trajectory Prediction Based on a Multi-Speaker Parallel Speech Database
Kálmán Abari¹, Tamás Gábor Csapó², Bálint Pál Tóth², Gábor Olaszy²
¹University of Debrecen, Hungary; ²BME, Hungary

PAGE 628
Mon-P-5-10
Layered Nonnegative Matrix Factorization for Speech Separation
Chung-Chien Hsu, Jen-Tzung Chien, Tai-Shih Chi, National Chiao Tung University, Taiwan

PAGE 633
Mon-P-5-11
Robust Tongue Tracking in Ultrasound Images: A Multi-Hypothesis Approach
Catherine Laporte¹, Lucie Ménard²
¹École de Technologie Supérieure, Canada; ²Université du Québec à Montréal, Canada

PAGE 638
Mon-P-5-12
Objective Measures for Predicting the Intelligibility of Spectrally Smoothed Speech with Artificial Excitation
Danny Websdale, Thomas Le Cornu, Ben Milner, University of East Anglia, UK
Poster Session 6: Speech Recognition — Technologies and Systems for New Applications

Chair: Daniel Willett
Room: Foyer, Time 14:30 – 16:30, Monday, September 7, 2015

Page 643
Mon-P-6-1
Mispronunciation Detection Without Nonnative Training Data
Ann Lee, James Glass, MIT, USA

Page 648
Mon-P-6-2
Automatic Accentedness Evaluation of Non-Native Speech Using Phonetic and Sub-Phonetic Posterior Probabilities
Ramya Rasipuram, Milos Cernak, Alexandre Nachen, Mathew Magimai-Doss, Idiap Research Institute, Switzerland

Page 653
Mon-P-6-3
Using F0 Contours to Assess Nativeness in a Sentence Repeat Task
Min Ma1, Keelan Evanini2, Anastassia Loukina2, Xinhao Wang2, Klaus Zechn2
1CUNY Graduate Center, USA; 2Educational Testing Service, USA

Page 658
Mon-P-6-4
Using Linguistic Indicators of Difficulty to Identify Mild Cognitive Impairment
Rebecca Lunsford, Peter A. Heeman, Oregon Health & Science University, USA

Page 663
Mon-P-6-5
Automatic Intelligibility Measures Applied to Speech Signals Simulating Age-Related Hearing Loss
Lionel Fontan1, Jérôme Farinas1, Isabelle Ferrané1, Julien Pinquier1, Xavier Aumont2
1IRIT, France; 2LETRA (ANR-14-LAB4-0003-01), France

Page 668
Mon-P-6-6
Assessing Empathy Using Static and Dynamic Behavior Models Based on Therapist’s Language in Addiction Counseling
Sandeep Nallan Chakravarthula1, Bo Xiao1, Zac E. Imel2, David C. Atkins3, Panayiotis G. Georgiou1
1University of Southern California, USA; 2University of Utah, USA; 3University of Washington, USA

Page 673
Mon-P-6-7
SVitchboard II and FiSVer I: High-Quality Limited-Complexity Corpora of Conversational English Speech
Yuzong Liu, Rishabh Iyer, Katrin Kirchhoff, Jeff Bilmes, University of Washington, USA

Page 678
Mon-P-6-8
Fully Unsupervised Small-Vocabulary Speech Recognition Using a Segmental Bayesian Model
Herman Kamper1, Aren Jansen2, Sharon Goldwater1
1University of Edinburgh, UK; 2Johns Hopkins University, USA

Page 683
Mon-P-6-9
LSTM for Punctuation Restoration in Speech Transcripts
Ottokar Tilk, Tanel Alumäe, Tallinn University of Technology, Estonia

Page 688
Mon-P-6-10
Noise Robust Exemplar Matching for Speech Enhancement: Applications to Automatic Speech Recognition
Emre Yılmaz, Deepak Baby, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium

Page 693
Mon-P-6-11
A Study on Robust Detection of Pronunciation Erroneous Tendency Based on Deep Neural Network
Yingming Gao, Yanlu Xie, Wen Cao, Jinsong Zhang, BLCU, China

Page 697
Mon-P-6-12
Vowel Mispronunciation Detection Using DNN Acoustic Models with Cross-Lingual Training
Shrikant Joshi, Nachiket Deo, Preeti Rao, IIT Bombay, India

Page 702
Mon-P-6-13
Confidence-Features and Confidence-Scores for ASR Applications in Arbitration and DNN Speaker Adaptation
Kshitiz Kumar, Ziad Al Bawah, Yong Zhao, Chaojun Liu, Benoit Dumoulin, Yifan Gong, Microsoft, USA
**Topic Modeling for Conference Analytics**  
Pengfei Liu\(^1\), Shaob Jameel\(^1\), Wai Lam\(^1\), Bin Ma\(^2\), Helen Meng\(^1\)  
\(^1\)Chinese University of Hong Kong, China; \(^2\)A*STAR, Singapore

**Sparse Coding Based Features for Speech Units Classification**  
Pulkit Sharma, Vinayak Abrol, A.D. Dileep, Anil Kumar Sao, IIT Mandi, India

**Show and Tell Session 1**  
*Chair: Siegfried Kunzmann*  
*Room: Seminar 1, Time 14:30 – 16:30, Monday, September 7, 2015*

**Smarter Driving with IDA, the Intelligent Driving Assistant for Singapore**  
Andreea I. Niculescu, Ngoc Thuy Huong Thai, Chongjia Ni, Boon Pang Lim, Kheng Huei Yeow, Rafael E. Banchs, A*STAR, Singapore

**Talk it Out: Adding Speech Interaction to Support Informational and Transactional Applications on Public Touch-Screen Kiosks**  
Kheng Huei Yeow, Rafael E. Banchs, A*STAR, Singapore

**Conversational Agent and Management Tools for Conference and Tourism Domain**  
Luis Fernando D’Haro, Seokhwan Kim, Rafael E. Banchs, A*STAR, Singapore

**Latvian Speech-to-Text Transcription Service**  
Askars Salimbajevs, Jevgenijs Strigins, Tilde, Latvia

**System Supporting Speaker Identification in Emergency Call Center**  
Jakub Galka, Joanna Grzybowska, Magdalena Igras, Pawel Jaciw, Kamil Wajda, Marcin Witkowski, Mariusz Ziolk, AGH UST, Poland

**QAT\(^2\) — The QCRI Advanced Transcription and Translation System**  
Ahmed Abdelali, Ahmed Ali, Francisco Guzman, Felix Stahlberg, Stephan Vogel, Yifan Zhang, Hamad Bin Khalifa University, Qatar

**Implementation of a Live Dialectal Media Subtitling System**  
Michael Stadtchnitzer, Christoph Schmidt, Fraunhofer IAIS, Germany

**A System for Automatic Broadcast News Summarisation, Geolocation and Translation**  
Peter Bell, Catherine Lai, Clare Llewellyn, Alexandra Birch, Mark Sinclair, University of Edinburgh, UK

**Media Monitoring System for Latvian Radio and TV Broadcasts**  
Arturs Znotins\(^1\), Kaspars Polis\(^2\), Roberts Dargis\(^1\)  
\(^1\)University of Latvia, Latvia; \(^2\)LETA, Latvia

**Meeting Assistant Application**  
Michael Assayaq\(^1\), Jonathan Huang\(^2\), Jonathan Mamou\(^1\), Oren Pereg\(^1\), Saurav Sahay\(^2\), Oren Shamir\(^3\), Georg Stemmer\(^3\), Moshe Wasserblat\(^1\)  
\(^1\)Intel, Israel; \(^2\)Intel, USA; \(^3\)Intel, Germany
## Oral Session 10: Distant and Reverberant Speech Recognition

*Chairs: John H. L. Hansen, Rita Singh*

*Room: Large Hall, Time 17:00 – 19:00, Monday, September 7, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-1</th>
<th>17:00 – 17:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bayesian Integration of Sound Source Separation and Speech Recognition</td>
<td>Kousuke Itakura, Izaya Nishimita, Yoshiaki Bando, Katsutoshi Itoyama, Kazuyoshi Yoshi, Kyoto University, Japan</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-2</th>
<th>17:20 – 17:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Channel Selection in the Short-Time Modulation Domain for Distant Speech Recognition</td>
<td>Ivan Himawan¹, Petr Motlicek¹, Sridha Sridharan², David Dean², Dian Tjondronegoro²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Idiap Research Institute, Switzerland; ²Queensland University of Technology, Australia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-3</th>
<th>17:40 – 18:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Multi-Channel Speech Enhancement Framework for Robust NMF-Based Speech Recognition for Speech-Impaired Users</td>
<td>Gert Dekkers, Toon van Waterschoot, Bart Vanrumste, Bert Van Den Broeck, Jort F. Gemmeke, Hugo Van hamme, Peter Karsmakers, Katholieke Universiteit Leuven, Belgium</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-4</th>
<th>18:00 – 18:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sound Source Separation Algorithm Using Phase Difference and Angle Distribution Modeling Near the Target</td>
<td>Chanwoo Kim, Kean K. Chin, Google, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-5</th>
<th>18:20 – 18:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contaminated Speech Training Methods for Robust DNN-HMM Distant Speech Recognition</td>
<td>Mirco Ravanelli, Maurizio Omologo, FBK, Italy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-10-6</th>
<th>18:40 – 19:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distance-Aware DNNs for Robust Speech Recognition</td>
<td>Yajie Miao, Florian Metze, Carnegie Mellon University, USA</td>
</tr>
</tbody>
</table>

## Oral Session 11: Speech Analysis and Representation 1

*Chairs: Thomas Drugman, Panayiotis Georgiou*

*Room: Conference 1, Time 17:00 – 19:00, Monday, September 7, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-1</th>
<th>17:00 – 17:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vocal Tremor Analysis via AM-FM Decomposition of Empirical Modes of the Glottal Cycle Length Time Series</td>
<td>Christophe Mertens¹, Francis Grenez¹, François Viallet², Alain Ghio³, Sabine Skodda⁴, Jean Schoentgen⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Université Libre de Bruxelles, Belgium; ²Centre Hospitalier du Pays d’Aix, France; ³LPL, France; ⁴Ruhr-Universität Bochum, Germany</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-2</th>
<th>17:20 – 17:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimating Lower Vocal Tract Features with Closed-Open Phase Spectral Analyses</td>
<td>Elizabeth Godoy, Nicolas Malyska, Thomas F. Quatieri, MIT Lincoln Laboratory, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-3</th>
<th>17:40 – 18:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inductive Implementation of Segmental HMMs as CS-HMMs</td>
<td>S.M. Houghton, Colin J. Champion, University of Birmingham, UK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-4</th>
<th>18:00 – 18:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Discriminative Analysis Within and Across Voiced and Unvoiced Consonants in Neutral and Whispered Speech in Multiple Indian Languages</td>
<td>G. Nisha Meenakshi, Prasanta Kumar Ghosh, Indian Institute of Science, India</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-5</th>
<th>18:20 – 18:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aligning Meeting Recordings via Adaptive Fingerprinting</td>
<td>T.J. Tsai¹, Andreas Stolcke²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹University of California at Berkeley, USA; ²Microsoft, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-O-11-6</th>
<th>18:40 – 19:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>On Representation Learning for Artificial Bandwidth Extension</td>
<td>Matthias Zöhrer¹, Robert Peharz², Franz Pernkopf¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Technische Universität Graz, Austria; ²Medizinische Universität Graz, Austria</td>
</tr>
</tbody>
</table>
Oral Session 12: L2 Speech Perception and Production

Chairs: Tarayuki Arai, Catherine Best
Room: Conference 2+3, Time 17:00 – 19:00, Monday, September 7, 2015

PAGE 796
Mon-O-12-1
17:00 – 17:20
Perception and Production of Vowel Contrasts in German Learners of English
Helena Levy, Universität Freiburg, Germany

PAGE 801
Mon-O-12-2
17:20 – 17:40
Goodness of Tone (GOT) for Non-Native Mandarin Tone Recognition
Rong Tong, Nancy F. Chen, Bin Ma, Haizhou Li, A*STAR, Singapore

PAGE 806
Mon-O-12-3
17:40 – 18:00
The Effect of High-Variability Training on the Perception and Production of French Stops by German Native Speakers
Jeanin Jügler¹, Frank Zimmerer¹, Bernd Möbius¹, Christoph Draxler²
¹Universität des Saarlandes, Germany; ²LMU München, Germany

PAGE 811
Mon-O-12-4
18:00 – 18:20
Perception of Mandarin Tones by Native Tibetan Speakers
Wenfu Bao, Hui Feng, Jianwu Dang, Zhilei Liu, Yang Yu, Siyu Wang, Tianjin University, China

PAGE 815
Mon-O-12-5
18:20 – 18:40
Study of Acoustic Correlates of English Lexical Stress Produced by Native (L1) Bengali Speakers Compared to Native (L1) English Speakers
Shambhu Nath Saha, Shyamal Kr. Das Mandal, IIT Kharagpur, India

PAGE 820
Mon-O-12-6
18:40 – 19:00
Prosodic Phrasing Unique to the Acquisition of L2 Intonation — An Analysis of L2 Japanese Intonation by L1 Swedish Learners
Yasuko Nagano-Madsen, University of Gothenburg, Sweden

Oral Session 13: Information and Metadata Extraction from Speech

Chairs: Pedro Moreno, Olivier Siohan
Room: Conference 4+5, Time 17:00 – 19:00, Monday, September 7, 2015

PAGE 824
Mon-O-13-1
17:00 – 17:20
Fusion of LVCSR and Posteriorgram Based Keyword Search
Leda Sarı, Batuhan Gündoğdu, Murat Saraçlar, Boğaziçi Üniversitesi, Turkey

PAGE 829
Mon-O-13-2
17:20 – 17:40
Improving Speech Recognition and Keyword Search for Low Resource Languages Using Web Data
Gideon Mendels¹, Erica Cooper¹, Victor Soto¹, Julia Hirschberg¹, Mark J.F. Gales², Kate M. Knill²
¹Columbia University, USA; ²University of Cambridge, UK

PAGE 834
Mon-O-13-3
17:40 – 18:00
Two-Step Spoken Term Detection Using SVM Classifier Trained with Pre-Indexed Keywords Based on ASR Result
Kentaro Domoto¹, Takehito Utsuro¹, Naoki Sawada², Hiromitsu Nishizaki²
¹University of Tsukuba, Japan; ²University of Yamanashi, Japan

PAGE 839
Mon-O-13-4
18:00 – 18:20
Enhancing Low Resource Keyword Spotting with Automatically Retrieved Web Documents
Le Zhang, Damianos Karakos, William Hartmann, Roger Hsiao, Richard Schwartz, Stavros Tsakalidis, Raytheon BBN Technologies, USA

PAGE 844
Mon-O-13-5
18:20 – 18:40
A Comparison Between a DNN and a CRF Disfluency Detection and Reconstruction System
Dario Bertero, Linlin Wang, Ho Yin Chan, Pascale Fung, HKUST, China

PAGE 849
Mon-O-13-6
18:40 – 19:00
Recurrent Neural Networks for Incremental Disfluency Detection
Julian Hough, David Schlangen, Universität Bielefeld, Germany
Oral Session 14: Deep Neural Networks for Speech Synthesis

Chairs: Raul Fernandez, Yoshihiko Nankaku
Room: Conference 6, Time 17:00 – 19:00, Monday, September 7, 2015

Fusion of Multiple Parameterisations for DNN-Based Sinusoidal Speech Synthesis with Multi-Task Learning
Qiong Hu\textsuperscript{1}, Zhizheng Wu\textsuperscript{1}, Korin Richmond\textsuperscript{1}, Junichi Yamagishi\textsuperscript{1}, Yannis Stylianou\textsuperscript{2}, Ranniery Maia\textsuperscript{2}
\textsuperscript{1}University of Edinburgh, UK; \textsuperscript{2}Toshiba Research Europe, UK

An Investigation of Recurrent Neural Network Architectures for Statistical Parametric Speech Synthesis
Sivanand Achanta, Tejas Godambe, Suryakanth V. Gangashetty, IIT Hyderabad, India

Sequence Generation Error (SGE) Minimization Based Deep Neural Networks Training for Text-to-Speech Synthesis
Yuchen Fan, Yao Qian, Frank K. Soong, Lei He, Microsoft, China

Towards Minimum Perceptual Error Training for DNN-Based Speech Synthesis
Cassia Valentini-Botinhao, Zhizheng Wu, Simon King, University of Edinburgh, UK

Deep Neural Network-Based Statistical Parametric Speech Synthesis System Using Improved Time-Frequency Trajectory Excitation Model
Eunwoo Song, Hong-Goo Kang, Yonsei University, Korea

A Study of Speaker Adaptation for DNN-Based Speech Synthesis
Zhizheng Wu, Pawel Swietojanski, Christophe Veaux, Steve Renals, Simon King, University of Edinburgh, UK

Special Session 2b: Interspeech 2015 Computational Paralinguistics Challenge\textsuperscript{E} (ComParE): Degree of Nativeness, Parkinson’s & Eating Condition

Chairs: Anton Batliner, Stefan Steidl
Room: Hall 1, Time 17:00 – 19:00, Monday, September 7, 2015

The Eating Condition Sub-Challenge: The Data
Anton Batliner, FAU Erlangen-Nürnberg, Germany

Automatic Classification of Eating Conditions from Speech Using Acoustic Feature Selection and a Set of Hierarchical Support Vector Machine Classifiers
Abhay Prasad, Prasanta Kumar Ghosh, Indian Institute of Science, India

Combining Hierarchical Classification with Frequency Weighting for the Recognition of Eating Conditions
Johannes Wagner, Andreas Seiderer, Florian Lingenfelser, Elisabeth André, Universität Augsburg, Germany

Acoustic Group Feature Selection Using Wrapper Method for Automatic Eating Condition Recognition
Dara Pir\textsuperscript{1}, Theodore Brown\textsuperscript{2}
\textsuperscript{1}CUNY Graduate Center, USA; \textsuperscript{2}CUNY Queens College, USA

Comparing SVM, Softmax, and Shallow Neural Networks for Eating Condition Classification
Thomas Pellegrini, IRIT, France

Using Representation Learning and Out-of-Domain Data for a Paralinguistic Speech Task
Benjamin Milde, Chris Biemann, Technische Universität Darmstadt, Germany
Fisher Vectors with Cascaded Normalization for Paralinguistic Analysis
Heysem Kaya\textsuperscript{1}, Alexey A. Karpov\textsuperscript{2}, Albert Ali Salah\textsuperscript{1}
\textsuperscript{1}Boğaziçi Üniversitesi, Turkey; \textsuperscript{2}Russian Academy of Sciences, Russia

Automatic Estimation of Parkinson’s Disease Severity from Diverse Speech Tasks
Jangwon Kim, Md. Nasir, Rahul Gupta, Maarten Van Segbroeck, Daniel Bone, Matthew P. Black, Zisis Iason Skordilis, Zhaojun Yang, Panayiotis G. Georgiou, Shrikanth S. Narayanan, University of Southern California, USA

Assessing the Degree of Nativeness and Parkinson’s Condition Using Gaussian Processes and Deep Rectifier Neural Networks
Tamás Grósz\textsuperscript{1}, Róbert Busa-Fekete\textsuperscript{2}, Gábor Gosztolya\textsuperscript{3}, László Tóth\textsuperscript{3}
\textsuperscript{1}University of Szeged, Hungary; \textsuperscript{2}Universität Paderborn, Germany; \textsuperscript{3}MTA-SZTE RGAI, Hungary

The INTERSPEECH 2015 Computational Paralinguistics Challenge: A Summary of Results
Stefan Steidl, FAU Erlangen-Nürnberg, Germany

Wrapping Up: The Story of the Compare Challenges, What We Learned and Where to Go
Anton Batliner, FAU Erlangen-Nürnberg, Germany

Poster Session 7: Prosody 1
Chair: Hongwei Ding
Room: Hall 2, Time 17:00 – 19:00, Monday, September 7, 2015

Pitch Scaling as a Perceptual Cue for Questions in German
Jan Michalsky, Carl von Ossietzky Universität Oldenburg, Germany

Parameterization of Prosodic Headedness
Uwe D. Reichel\textsuperscript{1}, Katalin Mády\textsuperscript{2}, Štefan Beňuš\textsuperscript{3}
\textsuperscript{1}LMU München, Germany; \textsuperscript{2}Hungarian Academy of Sciences, Hungary; \textsuperscript{3}UKF, Slovak Republic

Detection of Mizo Tones
Biswajit Dev Sarma, Priyankoo Sarmah, Wendy Lalhminghlui, S.R. Mahadeva Prasanna, IIT Guwahati, India

The Intonation of Echo \textit{wh}-Questions
Sophie Repp, Lena Rosin, Humboldt-Universität zu Berlin, Germany

Immediately Postverbal Questions in Urdu
Farhat Jabeen, Tina Bögel, Miriam Butt, Universität Konstanz, Germany

Prosodic (Non-)Realisation of Broad, Narrow and Contrastive Focus in Hungarian: A Production and a Perception Study
Katalin Mády, Hungarian Academy of Sciences, Hungary

F0 Discontinuity as a Marker of Prosodic Boundary Strength in Lombard Speech
Štefan Beňuš\textsuperscript{1}, Uwe D. Reichel\textsuperscript{2}, Juraj Šimko\textsuperscript{3}
\textsuperscript{1}UKF, Slovak Republic; \textsuperscript{2}LMU München, Germany; \textsuperscript{3}University of Helsinki, Finland
Poster Session 8: Prosody 2

Chair: Hongwei Ding
Room: Hall 2, Time 17:00 – 19:00, Monday, September 7, 2015

Comparing Journalistic and Spontaneous Speech: Prosodic and Spectral Analysis
Cédric Gendrot, Martine Adda-Decker, Yaru Wu, LPP (UMR 7018), France

Rhythm Influences the Tonal Realisation of Focus
Nadja Schaufler, Katrin Schweitzer, Universität Stuttgart, Germany

Linguistic Measures of Pitch Range in Slavic and Germanic Languages
Bistra Andreeva¹, Bernd Möbius¹, Grazyna Demenko², Frank Zimmerer¹, Jeanin Jügler¹
¹Universtität des Saarlandes, Germany; ²Adam Mickiewicz University in Poznań, Poland

The Effect of Stress on Vowel Space in Daxi Hakka Chinese
Chunan Qiu, Jie Liang, Tongji University, China

Declination, Peak Height and Pitch Level in Declaratives and Questions of South Connaught Irish
Maria O’Reilly, Ailbhe Ni Chasaide, Trinity College Dublin, Ireland

Contextual Variation of Tones in Mizo
Priyankoo Sarmah, Leena Dihingia, Wendy Lahthinghlu, IIT Guwahati, India

The Prosodic Marking of Rhetorical Questions in German
Daniela Wochner, Jana Schlegel, Nicole Dehé, Bettina Braun, Universität Konstanz, Germany

Poster Session 9: Speaker and Language Recognition

Chair: Kong Aik Lee
Room: Foyer, Time 17:00 – 19:00, Monday, September 7, 2015

High-Resolution Acoustic Modeling and Compact Language Modeling of Language-Universal Speech Attributes for Spoken Language Identification
Yannan Wang¹, Jun Du¹, Li-Rong Dai¹, Chin-Hui Lee²
¹USTC, China; ²Georgia Institute of Technology, USA

Phonemes Frequency Based PLLR Dimensionality Reduction for Language Recognition
Saad Irzsa¹, Vidhyasaharan Sethu¹, Phu Ngoc Le¹, Eliathamby Ambikairajah¹, Haizhou Li²
¹University of New South Wales, Australia; ²A*STAR, Singapore

Exploiting i-Vector Posterior Covariances for Short-Duration Language Recognition
Sandro Cumani¹, Oldřich Plchot², Radek Féř²
¹Politecnico di Torino, Italy; ²Brno University of Technology, Czech Republic

Using the Beat Histogram for Speech Rhythm Description and Language Identification
Athanasios Lykartsis, Stefan Weinzierl, Technische Universität Berlin, Germany

Speaker Recognition for Speech Under Face Cover
Rahim Saeidi¹, Tuija Niemi², Hanna Karppelin³, Juoni Pohjalainen¹, Tomi Kinnunen⁴, Paavo Alku¹
¹Aalto University, Finland; ²National Bureau of Investigation, Finland; ³University of Helsinki, Finland; ⁴University of Eastern Finland, Finland
<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-P-9-6</th>
<th>Dataset-Invariant Covariance Normalization for Out-Domain PLDA Speaker Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Md. Hafizur Rahman, Ahilan Kanagasundaram, David Dean, Sridha Sridharan, Queensland University of Technology, Australia</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-7</td>
<td>Sparse Coding of Total Variability Matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longting Xu¹, Kong Aik Lee², Haizhou Li², Zhen Yang¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹NJUPT, China; ²A*STAR, Singapore</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-8</td>
<td>Duration Dependent Covariance Regularization in PLDA Modeling for Speaker Verification</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weicheng Cai¹, Ming Li², Lin Li³, QingYang Hong³</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹JRI, China; ²Sun Yat-sen University, China; ³Xiamen University, China</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-9</td>
<td>Exploiting Supervector Structure for Speaker Recognition Trained on a Small Development Set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hagai Aronowitz, IBM Research Haifa, Israel</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-10</td>
<td>Modified-Prior PLDA and Score Calibration for Duration Mismatch Compensation in Speaker Recognition System</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QingYang Hong¹, Lin Li¹, Ming Li², Ling Huang¹, Lihong Wan¹, Jun Zhang¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Xiamen University, China; ²Sun Yat-sen University, China</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-11</td>
<td>Speaker Verification Using Gaussian Posteriorgrams on Fixed Phrase Short Utterances</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sarfaraz Jelil¹, Rohan Kumar Das², Rohit Sinha², S.R. Mahadeva Prasanna²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹North-Eastern Hill University, India; ²IIT Guwahati, India</td>
</tr>
</tbody>
</table>

**INTERSPEECH 2015 Table of Contents**

<table>
<thead>
<tr>
<th>Page</th>
<th>Mon-P-9-12</th>
<th>Importance of Intelligible Phonemes for Human Speaker Recognition in Different Channel Bandwidths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Laura Fernández Gallardo¹, Sebastian Möller², Michael Wagner¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹University of Canberra, Australia; ²T-Labs, Germany</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-13</td>
<td>Denoising Autoencoder-Based Speaker Feature Restoration for Utterances of Short Duration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hitoshi Yamamoto, Takafumi Koshinaka, NEC Corporation, Japan</td>
</tr>
<tr>
<td>Page</td>
<td>Mon-P-9-14</td>
<td>Full Multicondition Training for Robust i-Vector Based Speaker Recognition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dayana Ribas³, Emmanuel Vincent², José Ramón Calvo¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹CENATAV, Cuba; ²INRIA, France</td>
</tr>
</tbody>
</table>
### Show and Tell Session 2

**Chair:** Volker Fischer  
**Room:** Seminar 1, Time 17:00 – 19:00, Monday, September 7, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1062</td>
<td>SARMATA 2.0 Automatic Polish Language Speech Recognition System</td>
<td>Bartosz Ziółko, Tomasz Jadczyk, Dawid Skurzok, Piotr Żelasko, Jakub Gałka, Tomasz Pędzimaz, Ireneusz Gawlik, Szymon Pałka, AGH UST, Poland</td>
</tr>
<tr>
<td>1064</td>
<td>Remeeting — Get More Out of Meetings</td>
<td>Arlo Faria, Korbinian Riedhammer, Remeeting, USA</td>
</tr>
<tr>
<td>1066</td>
<td>Web Application System for Pronunciation Practice by Children with Disabilities and to Support Cooperation of Teachers and Medical Workers</td>
<td>Ikuyo Masuda-Katsuse, Kindai University, Japan</td>
</tr>
<tr>
<td>1068</td>
<td>PATSY — It's All About Pronunciation!</td>
<td>Caroline Kaufhold¹, Vadim Gamidov², Andreas Kiessling², Klaus Reinhard², Elmar Noth³</td>
</tr>
<tr>
<td>1070</td>
<td>Real-Time Pitch Modification System for Speech and Singing Voice</td>
<td>Elias Aزارov, Maxim Vashkevich, Denis Likhachov, Alexander Petrovsky, BSUIR, Belarus</td>
</tr>
<tr>
<td>1072</td>
<td>Nao is Doing Humour in the CHIST-ERA JOKER Project</td>
<td>Guillaume Dubuisson Duplessis¹, Lucile Béchade², Mohamed A. Sehili¹, Agnès Delaborde¹, Vincent Letard¹, Anne-Laure Ligozat¹, Paul Deléglise³, Yannick Estève³, Sophie Rosset¹, Laurence Devillers¹</td>
</tr>
<tr>
<td>1074</td>
<td>ABIMS — Auditory Bewildered Interaction Measurement System</td>
<td>Lisa Lange, Bartholomäus Pfeiffer, Daniel Duran, Universität Stuttgart, Germany</td>
</tr>
</tbody>
</table>

### Oral Session 15: Neural Networks and Speaker Adaptation

**Chairs:** Penny Karanasou, Daniel Povey  
**Room:** Large Hall, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1076</td>
<td>Maximum a posteriori Adaptation of Network Parameters in Deep Models</td>
<td>Zhen Huang¹, Sabato Marco Siniscalchi¹, I-Fan Chen¹, Jinyu Li², Jiadong Wu¹, Chin-Hui Lee¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Georgia Institute of Technology, USA; ²Microsoft, USA</td>
</tr>
<tr>
<td>1081</td>
<td>Regularized Sequence-Level Deep Neural Network Model Adaptation</td>
<td>Yan Huang, Yifan Gong, Microsoft, USA</td>
</tr>
<tr>
<td>1086</td>
<td>Modeling Speaker Variability Using Long Short-Term Memory Networks for Speech Recognition</td>
<td>Xiangang Li, Xihong Wu, Peking University, China</td>
</tr>
<tr>
<td>1091</td>
<td>Intermediate-Layer DNN Adaptation for Offline and Session-Based Iterative Speaker Adaptation</td>
<td>Kshitiz Kumar, Chaojun Liu, Kaisheng Yao, Yifan Gong, Microsoft, USA</td>
</tr>
<tr>
<td>1096</td>
<td>Speaker Adaptation of Convolutional Neural Network Using Speaker Specific Subspace Vectors of SGMM</td>
<td>Murali Karthick B, Prateek Kolhar, S. Umesh, IIT Madras, India</td>
</tr>
<tr>
<td>1101</td>
<td>On Speaker Adaptation of Long Short-Term Memory Recurrent Neural Networks</td>
<td>Yajie Miao, Florian Metze, Carnegie Mellon University, USA</td>
</tr>
</tbody>
</table>
## Oral Session 16: Brain- and Other Biosignal-based Spoken Communication
**Chairs:** Thomas F. Quatieri, Tanja Schultz  
**Room:** Conference 1, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Session</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 1106 | Tue-O-16-1 | 09:00 – 09:20 | Automatic Identification of Received Language in MEG | Emilio Parisotto¹, Youness A. Ghassabeh², Matt J. MacDonald³, Adelina Cozma⁴, Elizabeth W. Pan⁵, Frank Rudzicz¹  
¹University of Toronto, Canada; ²UHN Toronto Rehabilitation Institute, Canada; ³SickKids Research Institute, Canada; ⁴University of Calgary, Canada; ⁵Hospital for Sick Children, Canada |
| 1111 | Tue-O-16-2 | 09:20 – 09:40 | Detection of Cardiovascular Reactivity in Speech | Laurens van der Werff, Jón Guðnason, Kamilla Run Jóhannsdóttir, Reykjavík University, Iceland |
| 1116 | Tue-O-16-3 | 09:40 – 10:00 | Lateralization in Emotional Speech Perception Following Transcranial Direct Current Stimulation | Alex Francois-Nienaber¹, Jed A. Meltzer², Frank Rudzicz¹  
¹University of Toronto, Canada; ²Rotman Research Institute, Canada |
| 1121 | Tue-O-16-4 | 10:00 – 10:20 | Speech Reconstruction from Human Auditory Cortex with Deep Neural Networks | Minda Yang¹, Sameer A. Sheth², Catherine A. Schevon², Guy M. McKhann II², Nima Mesgarani¹  
¹Columbia University, USA; ²Columbia University Medical Center, USA |
| 1126 | Tue-O-16-5 | 10:20 – 10:40 | Temporal Dynamics of the Speech Readiness Potential, and Its Use in a Neural Decoder of Speech-Motor Intention | Jonathan S. Brumberg, Nichol Castro, Akshatha Rao, University of Kansas, USA |
| 1131 | Tue-O-16-6 | 10:40 – 11:00 | Continuous Speech Recognition from ECoG | Dominic Heger¹, Christian Herff¹, Adriana de Pesters², Dominic Telaar¹, Peter Brunner², Gerwin Schalk², Tanja Schultz¹  
¹KIT, Germany; ²New York State Department of Health, USA |

## Oral Session 17: Deep Neural Networks in Speaker Recognition
**Chairs:** Patrick Kenny, Alan McCree  
**Room:** Conference 2+3, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Session</th>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1136</td>
<td>Tue-O-17-1</td>
<td>09:00 – 09:20</td>
<td>Locally-Connected and Convolutional Neural Networks for Small Footprint Speaker Recognition</td>
<td>Yu-hsin Chen, Ignacio Lopez-Moreno, Tiara N. Sainath, Mirko Visontai, Raziel Alvarez, Carolina Parada, Google, USA</td>
</tr>
<tr>
<td>1141</td>
<td>Tue-O-17-2</td>
<td>09:20 – 09:40</td>
<td>Insights into Deep Neural Networks for Speaker Recognition</td>
<td>Daniel Garcia-Romero, Alan McCree, Johns Hopkins University, USA</td>
</tr>
</tbody>
</table>
| 1146 | Tue-O-17-3 | 09:40 – 10:00 | A Unified Deep Neural Network for Speaker and Language Recognition | Fred Richardson¹, Douglas A. Reynolds¹, Najim Dehak²  
¹MIT Lincoln Laboratory, USA; ²MIT, USA |
| 1151 | Tue-O-17-4 | 10:00 – 10:20 | Investigation of Bottleneck Features and Multilingual Deep Neural Networks for Speaker Verification | Yao Tian, Meng Cai, Liang He, Jia Liu, Tsinghua University, China |
| 1156 | Tue-O-17-5 | 10:20 – 10:40 | Frequency Offset Correction in Single Sideband (SSB) Speech by Deep Neural Network for Speaker Verification | Hua Xing, Gang Liu, John I.L. Hansen, University of Texas at Dallas, USA |
| 1161 | Tue-O-17-6 | 10:40 – 11:00 | Exploring Robustness of DNN/RNN for Extracting Speaker Baum-Welch Statistics in Mismatched Conditions | Hao Zheng, Shanshan Zhang, Wenju Liu, Chinese Academy of Sciences, China |
### Oral Session 18: Speech Analysis and Representation 2

**Chairs:** Miloš Cerňák, Shinji Watanabe  
**Room:** Conference 4+5, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Presentation</th>
</tr>
</thead>
</table>
| 1166 | Tue-O-18-1  
09:00 – 09:20  
AM-FM Based Filter Bank Analysis for Estimation of Spectro-Temporal Envelopes and Its Application for Speaker Recognition in Noisy Reverberant Environments  
*Dhananjaya Gowda, Rahim Saeidi, Paavo Alku, Aalto University, Finland* |
| 1171 | Tue-O-18-2  
09:20 – 09:40  
Fast and Accurate Phase Unwrapping  
*Thomas Drugman, Yannis Stylianou, Toshiba Research Europe, UK* |
| 1176 | Tue-O-18-3  
09:40 – 10:00  
Sparse Representation with Temporal Max-Smoothing for Acoustic Event Detection  
*Xugang Lu¹, Peng Shen¹, Yu Tsao², Chiori Hori¹, Hisashi Kawai¹  
¹NICT, Japan; ²Academia Sinica, Taiwan* |
| 1181 | Tue-O-18-4  
10:00 – 10:20  
Estimation of Glottal Closure Instants from Telephone Speech Using a Group Delay-Based Approach That Considers Speech Signal as a Spectrum  
*Anushiya Rachel G., Vijayalakshmi P., Nagarajan T., SSN, India* |
| 1186 | Tue-O-18-5  
10:20 – 10:40  
The Role of Prosody and Voice Quality in Text-Dependent Categories of Storytelling Across Languages  
*Raul Montano, Francesc Alias, Universitat Ramon Llull, Spain* |
| 1191 | Tue-O-18-6  
10:40 – 11:00  
Neuromorphic Based Oscillatory Device for Incremental Syllable Boundary Detection  
*Alexandre Hyafil¹, Milos Cerňak²  
¹Universitat Pompeu Fabra, Spain; ²Idiap Research Institute, Switzerland* |

### Oral Session 19: Statistical Parametric Speech Synthesis

**Chairs:** Michael Pucher, Junichi Yamagishi  
**Room:** Conference 6, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Presentation</th>
</tr>
</thead>
</table>
| 1196 | Tue-O-19-1  
09:00 – 09:20  
Simultaneous Optimization of Multiple Tree Structures for Factor Analyzed HMM-Based Speech Synthesis  
*Takenori Yoshimura, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda, Nagoya Institute of Technology, Japan* |
| 1201 | Tue-O-19-2  
09:20 – 09:40  
HMM Training Strategy for Incremental Speech Synthesis  
*Mael Pouget¹, Thomas Hueber¹, Gerard Bailly¹, Timo Baumann²  
¹GIPSA, France; ²Universität Hamburg, Germany* |
| 1206 | Tue-O-19-3  
09:40 – 10:00  
Modulation Spectrum-Constrained Trajectory Training Algorithm for HMM-Based Speech Synthesis  
*Shinnosuke Takamichi¹, Tomoki Toda¹, Alan W. Black², Satoshi Nakamura¹  
¹NAIST, Japan; ²Carnegie Mellon University, USA* |
| 1211 | Tue-O-19-4  
10:00 – 10:20  
Random Forests for Statistical Speech Synthesis  
*Alan W. Black, Prasanna Kumar Muthukumar, Carnegie Mellon University, USA* |
| 1216 | Tue-O-19-5  
10:20 – 10:40  
Speaker Adaptation Using Relevance Vector Regression for HMM-Based Expressive TTS  
*Doo Hwa Hong, Joun Yeop Lee, Se Young Jang, Nam Soo Kim, Seoul National University, Korea* |
| 1221 | Tue-O-19-6  
10:40 – 11:00  
Towards a Linear Dynamical Model Based Speech Synthesizer  
*Vassilis Tsiaras¹, Ranniery Maia², Vassilis Diakoloukas¹, Yannis Stylianou², Vassilis Digalakis¹  
¹Technical University of Crete, Greece; ²Toshiba Research Europe, UK* |
Providing Objective Metrics of Team Communication Skills via Interpersonal Coordination Mechanisms
Céline De Looze\textsuperscript{1}, Brian Vaughan\textsuperscript{2}, Finnian Kelly\textsuperscript{3}, Alison Kay\textsuperscript{1}
\textsuperscript{1}Trinity College Dublin, Ireland; \textsuperscript{2}Dublin Institute of Technology, Ireland; \textsuperscript{3}University of Texas at Dallas, USA

Dialog Act Modeling for Virtual Personal Assistant Applications Using a Small Volume of Labeled Data and Domain Knowledge
Donghyeon Lee, Jinsik Lee, Eun-Kyoung Kim, Jaewon Lee, Samsung Electronics, Korea

A Polyglot Domain Optimised Text-to-Speech System for Railway Station Announcements
Csaba Zainkó, Mátyás Bartalis, Géza Németh, Gábor Olaszy, BME, Hungary

Development of Hindi Speech Recognition System of Agricultural Commodities Using Deep Neural Network
Partho Mandal, Shalini Jain, Gaurav Ojha, Anupam Shukla, ABV-IIITM, India

Real-Time Audio Signal Enhancement for Hands-Free Speech Applications
Thomas Fehér, Michael Freitag, Christian Gruber, voice INTER connect, Germany

Personalized Synthetic Voices for Speaking Impaired: Website and App
D. Erro\textsuperscript{1}, Inma Hernaez\textsuperscript{1}, Agustin Alonso\textsuperscript{1}, D. Garcia-Lorenzo\textsuperscript{1}, Eva Navas\textsuperscript{1}, J. Ye\textsuperscript{1}, H. Arzelus\textsuperscript{2}, Igor Jauk\textsuperscript{3}, N.Q. Hy\textsuperscript{4}, C. Magariños\textsuperscript{5}, R. Pérez-Ramón\textsuperscript{1}, M. Sulír\textsuperscript{6}, Xiaohai Tian\textsuperscript{4}, X. Wang\textsuperscript{7}
\textsuperscript{1}Universidad del País Vasco, Spain; \textsuperscript{2}VicomTech-IK4, Spain; \textsuperscript{3}Universitat Politècnica de Catalunya, Spain; \textsuperscript{4}NTU, Singapore; \textsuperscript{5}Universidade de Vigo, Spain; \textsuperscript{6}Technical University of Košice, Slovak Republic; \textsuperscript{7}USTC, China
Poster Session 10: Speech Recognition: Evaluation and Low-resource Languages

Chair: Jia Cui
Room: Hall 2, Time 09:00 – 11:00, Tuesday, September 8, 2015

Tue-P-10-1
Under-Resourced Speech Recognition Based on the Speech Manifold
Reza Sahraeian¹, Dirk Van Compernolle¹, Febe de Wet²
¹ Katholieke Universiteit Leuven, Belgium; ² CSIR, South Africa

Tue-P-10-2
Multilingual Features Based Keyword Search for Very Low-Resource Languages
Pavel Golik, Zoltán Tüske, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

Tue-P-10-3
Second Language Speech Recognition Using Multiple-Pass Decoding with Lexicon Represented by Multiple Reduced Phoneme Sets
Xiaoyun Wang, Seiichi Yamamoto, Doshisha University, Japan

Tue-P-10-4
Using Resources from a Closely-Related Language to Develop ASR for a Very Under-Resourced Language: A Case Study for Iban
Sarah Samson Juan¹, Laurent Besacier², Benjamin Lecouteux², Mohamed Dyab²
¹ UNIMAS, Malaysia; ² LIG (UMR 5217), France

Tue-P-10-5
Prediction of Speech Recognition Accuracy for Utterance Classification
Maxim L. Korenevsky, Andrey B. Smirnov, Valentin S. Mendelev, ITMO University, Russia

Tue-P-10-6
Error Bounds for Context Reduction and Feature Omission
Eugen Beck, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

Tue-P-10 continued...

Tue-P-10-7
A Metric for Evaluating Speech Recognizer Output Based on Human-Perception Model
Nobuyasu Itoh¹, Gakuto Kurata¹, Ryuki Tachibana¹, Masafumi Nishimura²
¹ IBM Research Tokyo, Japan; ² Shizuoka University, Japan

Tue-P-10-8
How to Evaluate ASR Output for Named Entity Recognition?
Mohamed Ameur Ben Jannet¹, Olivier Galibert¹, Martine Adda-Decker², Sophie Rosset³
¹ LNE, France; ² LPP (UMR 7018), France; ³ LIMSI, France
Poster Session 11: Emotion

Chair: Anton Batliner
Room: Hall 2, Time 09:00 - 11:00, Tuesday, September 8, 2015

Tue-P-11-1
Page 1294
Acoustic-Prosodic Analysis of Attitudinal Expressions in German
Hansjörg Mixdorff¹, Angelika Höнемann², Albert Rilliard³
¹BHT Berlin, Germany; ²Universität Bielefeld, Germany; ³LIMSI, France

Tue-P-11-2
Page 1299
Continuous Emotion Tracking Using Total Variability Space
Hossein Khaki, Engin Erzin, Koç Üniversitesi, Turkey

Tue-P-11-3
Page 1304
An Analysis of the Relationship Between Signal-Derived Vocal Arousal Score and Human Emotion Production and Perception
Chi-Chun Lee¹, Daniel Bone², Shrikanth S. Narayan²
¹National Tsing Hua University, Taiwan; ²University of Southern California, USA

Tue-P-11-4
Page 1309
Morphology of Vocal Affect Bursts: Exploring Expressive Interjections in Japanese Conversation
Hiroki Mori, Utsunomiya University, Japan

Tue-P-11-5
Page 1314
Emotion Clustering Based on Probabilistic Linear Discriminant Analysis
Mahnoosh Mehrabani, Ozlem Kalinli, Ruxin Chen, Sony, USA

Tue-P-11-6
Page 1319
Objective Study of the Performance Degradation in Emotion Recognition Through the AMR-WB+ Codec
Aaron Albin, Elliot Moore, Georgia Institute of Technology, USA

Tue-P-11-7
Page 1324
Analysis of Excitation Source Features of Speech for Emotion Recognition
Sudarsana Reddy Kadiri, P. Gangamohan, Suryakanth V. Gangashetty, B. Yegnanarayana, IIIT Hyderabad, India

INTERSPEECH 2015 Table of Contents

Tue-P-11 continued...

Tue-P-11-8
Page 1329
An Investigation of Emotion Change Detection from Speech
Zhaocheng Huang, Julien Epps, Eliathamby Ambikairajah, University of New South Wales, Australia

Tue-P-11-9
Page 1334
Crosslinguistic Comparison on the Perception of Mandarin Attitudinal Speech
Wentao Gu¹, Ping Tang¹, Keikichi Hirose², Véronique Aubergé³
¹Nanjing Normal University, China; ²University of Tokyo, Japan; ³CNRS, France

Tue-P-11-10
Page 1339
Conflict Intensity Estimation from Speech Using Greedy Forward-Backward Feature Selection
Gábor Gosztolya, MTA-SZTE RGAI, Hungary
### Poster Session 12: Spoken Language Understanding

**Chair:** Ruhi Sarikaya  
**Room:** Foyer, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page 1344</th>
<th>Tue-P-12-1</th>
</tr>
</thead>
</table>
| **Study of Entity-Topic Models for OOV Proper Name Retrieval**  
*Imran Sheikh, Irina Illina, Dominique Fohr, LORIA, France* |

<table>
<thead>
<tr>
<th>Page 1349</th>
<th>Tue-P-12-2</th>
</tr>
</thead>
</table>
| **Audio Quotation Marks for Natural Language Understanding**  
*Simon Boutin¹, Réal Tremblay², Patrick Cardinal¹, Doug Peters², Pierre Dumouchel¹*  
¹École de Technologie Supérieure, Canada; ²Nuance Communications, Canada |

<table>
<thead>
<tr>
<th>Page 1353</th>
<th>Tue-P-12-3</th>
</tr>
</thead>
</table>
| **Using Word Confusion Networks for Slot Filling in Spoken Language Understanding**  
*Xiaohao Yang, Jia Liu, Tsinghua University, China* |

<table>
<thead>
<tr>
<th>Page 1358</th>
<th>Tue-P-12-4</th>
</tr>
</thead>
</table>
| **Distributed Representation-Based Spoken Word Sense Induction**  
*Justin Chiu, Yajie Miao, Alan W. Black, Alexander I. Rudnicky, Carnegie Mellon University, USA* |

<table>
<thead>
<tr>
<th>Page 1363</th>
<th>Tue-P-12-5</th>
</tr>
</thead>
</table>
| **Structuring Lectures in Massive Open Online Courses (MOOCs) for Efficient Learning by Linking Similar Sections and Predicting Prerequisites**  
*Sheng-syan Shen¹, Hung-yi Lee¹, Shang-wen Li², Victor Zue², Lin-shan Lee¹*  
¹National Taiwan University, Taiwan; ²MIT, USA |

<table>
<thead>
<tr>
<th>Page 1368</th>
<th>Tue-P-12-6</th>
</tr>
</thead>
</table>
| **News Talk-Show Chaptering with Journalistic Genres**  
*Delphine Charlet, Géraldine Damnati, Jérémy Trione, Orange Labs, France* |

<table>
<thead>
<tr>
<th>Page 1373</th>
<th>Tue-P-12-7</th>
</tr>
</thead>
</table>
| **An Analysis of Time-Aggregated and Time-Series Features for Scoring Different Aspects of Multimodal Presentation Data**  
*Vikram Ramanarayanan, Lei Chen, Chee Wee Leong, Gary Feng, David Suendermann-Oeft, Educational Testing Service, USA* |

---

**Tue-P-12 continued…**

<table>
<thead>
<tr>
<th>Page 1378</th>
<th>Tue-P-12-8</th>
</tr>
</thead>
</table>
| **Incorporating Prosodic Prominence Evidence into Term Weights for Spoken Content Retrieval**  
*David N. Racca, Gareth J.F. Jones, Dublin City University, Ireland* |

<table>
<thead>
<tr>
<th>Page 1383</th>
<th>Tue-P-12-9</th>
</tr>
</thead>
</table>
| **Leveraging Word Embeddings for Spoken Document Summarization**  
*Kuan-Yu Chen¹, Shih-Hung Liu¹, Hsin-Min Wang¹, Berlin Chen², Hsin-Hsi Chen³*  
¹Academia Sinica, Taiwan; ²National Taiwan Normal University, Taiwan; ³National Taiwan University, Taiwan |

<table>
<thead>
<tr>
<th>Page 1388</th>
<th>Tue-P-12-10</th>
</tr>
</thead>
</table>
| **Mutually Exclusive Grounding for Weakly Supervised Non-Negative Matrix Factorisation**  
*Vincent Renkens, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium* |

<table>
<thead>
<tr>
<th>Page 1393</th>
<th>Tue-P-12-11</th>
</tr>
</thead>
</table>
| **Using Semantic Maps for Robust Natural Language Interaction with Robots**  
*Emanuele Bastianelli¹, Danilo Croce¹, Roberto Basili¹, Daniele Nardi²*  
¹Università di Roma “Tor Vergata”, Italy; ²Università di Roma “La Sapienza”, Italy |

<table>
<thead>
<tr>
<th>Page 1398</th>
<th>Tue-P-12-12</th>
</tr>
</thead>
</table>
| **Efficient Learning for Spoken Language Understanding Tasks with Word Embedding Based Pre-Training**  
*Yi Luan¹, Shinji Watanabe², Bret Harsham²*  
¹University of Washington, USA; ²MERL, USA |

<table>
<thead>
<tr>
<th>Page 1403</th>
<th>Tue-P-12-13</th>
</tr>
</thead>
</table>
| **Zero-Shot Semantic Parser for Spoken Language Understanding**  
*Emmanuel Ferreira, Bassam Jabaian, Fabrice Lefèvre, LIA, France* |

<table>
<thead>
<tr>
<th>Page 1408</th>
<th>Tue-P-12-14</th>
</tr>
</thead>
</table>
| **Adapting Lexical Representation and OOV Handling from Written to Spoken Language with Word Embedding**  
*Jeremie Tafforeau, Thierry Artieres, Benoît Favre, Frédéric Bechet, LIF (UMR 7279), France* |
### Poster Session 13: Language Modeling for Speech Recognition

**Chair:** Xunying Liu  
**Room:** Seminar 5+6, Time 09:00 – 11:00, Tuesday, September 8, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
<th>Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1413</td>
<td>Multi-Stream Long Short-Term Memory Neural Network Language Model</td>
<td>Ebru Arisoy(^1), Murat Saraçlar(^2)</td>
<td>MEF Üniversitesi, Turkey; Boğaziçi Üniversitesi, Turkey</td>
</tr>
<tr>
<td>1418</td>
<td>Composition-Based On-the-Fly Rescoring for Salient n-Gram Biasing</td>
<td>Keith Hall, Eunjoon Cho, Cyril Allauzen, Françoise Beaufays, Noah Coccaro, Kaisuke Nakajima, Michael Riley, Brian Roark, David Rybach, Linda Zhang</td>
<td>Google, USA</td>
</tr>
<tr>
<td>1423</td>
<td>Learning Phrase Patterns for ASR Name Error Detection Using Semantic Similarity</td>
<td>Alex Marin(^1), Mari Ostendorf(^2), Ji He(^2)</td>
<td>Microsoft, USA; University of Washington, USA</td>
</tr>
<tr>
<td>1428</td>
<td>Sparse Non-Negative Matrix Language Modeling for Skip-Grams</td>
<td>Noam Shazeer, Joris Pelemans, Ciprian Chelba</td>
<td>Google, USA</td>
</tr>
<tr>
<td>1433</td>
<td>Pruning Sparse Non-Negative Matrix N-Gram Language Models</td>
<td>Joris Pelemans, Noam Shazeer, Ciprian Chelba</td>
<td>Google, USA</td>
</tr>
<tr>
<td>1438</td>
<td>Geo-Location for Voice Search Language Modeling</td>
<td>Ciprian Chelba, Xuedong Zhang, Keith Hall</td>
<td>Google, USA</td>
</tr>
<tr>
<td>1443</td>
<td>On Efficient Training of Word Classes and Their Application to Recurrent Neural Network Language Models</td>
<td>Rami Botros, Kazuki Irie, Martin Sundermeyer, Hermann Ney</td>
<td>RWTH Aachen University, Germany</td>
</tr>
</tbody>
</table>

**Tue-P-13 continued…**

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
<th>Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1448</td>
<td>Deep Semantic Encodings for Language Modeling</td>
<td>Ali Orkan Bayer, Giuseppe Riccardi, Università di Trento, Italy</td>
<td></td>
</tr>
<tr>
<td>1453</td>
<td>Learning OOV Through Semantic Relatedness in Spoken Dialog Systems</td>
<td>Ming Sun, Yun-Nung Chen, Alexander I. Rudnicky, Carnegie Mellon University, USA</td>
<td></td>
</tr>
<tr>
<td>1458</td>
<td>TDTSO Language Modeling with Feedforward Neural Networks</td>
<td>Tze Yang Chong(^1), Rafael E. Banchs(^2), Eng Siong Chng(^1), Haizhou Li(^1)</td>
<td>TL@NTU, Singapore; A*STAR, Singapore</td>
</tr>
</tbody>
</table>
**Oral Session 20: Fast Efficient and Scalable Computing for Neural Nets**

*Chairs: Tara Sainath, Frank Seide*

*Room: Large Hall, Time 14:00 – 16:00, Tuesday, September 8, 2015*

---

**Page 1463**

**Tue-O-20-1**

14:00 – 14:20

*Improvements to the Pruning Behavior of DNN Acoustic Models*

*Matthias Paulik, Apple, USA*

---

**Page 1468**

**Tue-O-20-2**

14:20 – 14:40

*Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition*

*Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays, Google, USA*

---

**Page 1473**

**Tue-O-20-3**

14:40 – 15:00

*Compressing Deep Neural Networks Using a Rank-Constrained Topology*

*Preetum Nakkiran¹, Raziel Alvarez², Rohit Prabhavalkar², Carolina Parada²*

¹University of California at Berkeley, USA; ²Google, USA

---

**Page 1478**

**Tue-O-20-4**

15:00 – 15:20

*Convolutional Neural Networks for Small-Footprint Keyword Spotting*

*Tara N. Sainath, Carolina Parada, Google, USA*

---

**Page 1483**

**Tue-O-20-5**

15:20 – 15:40

*Efficient GPU Implementation of Convolutional Neural Networks for Speech Recognition*

*Ewout van den Berg¹, Daniel Brand², Rajesh Bordawekar², Leonid Rachevsky¹, Bhuvana Ramabhadran¹*

¹IBM Watson, USA; ²IBM T.J. Watson Research Center, USA

---

**Page 1488**

**Tue-O-20-6**

15:40 – 16:00

*Scalable Distributed DNN Training Using Commodity GPU Cloud Computing*

*Nikko Strom, Amazon.com, USA*

---

**Oral Session 21: Source Separation and Computational Auditory Scene Analysis**

*Chairs: Rajesh M. Hegde, Chin-Hui Lee*

*Room: Conference 1, Time 14:00 – 16:00, Tuesday, September 8, 2015*

---

**Page 1493**

**Tue-O-21-1**

14:00 – 14:20

*Joint Source Localization and Separation in Spherical Harmonic Domain Using a Sparsity Based Method*

*Sachin N. Kalkur, Sandeep Reddy C., Rajesh M. Hegde, IIT Kanpur, India*

---

**Page 1498**

**Tue-O-21-2**

14:20 – 14:40

*Regularized Non-Negative Matrix Factorization Using Alternating Direction Method of Multipliers and Its Application to Source Separation*

*Shaofei Zhang¹, Dong-Yan Huang², Lei Xie¹, Eng Siong Chng³, Haizhou Li², Minghui Dong²*

¹Northwestern Polytechnical University, China; ²A*STAR, Singapore; ³NTU, Singapore

---

**Page 1503**

**Tue-O-21-3**

14:40 – 15:00

*Two-Stage Multi-Target Joint Learning for Monaural Speech Separation*

*Shuai Nie¹, Shan Liang¹, Wei Xue¹, Xueliang Zhang², Wenju Liu¹, Like Dong³, Hong Yang³*

¹Chinese Academy of Sciences, China; ²Inner Mongolia University, China; ³SGCC, China

---

**Page 1508**

**Tue-O-21-4**

15:00 – 15:20

*Multi-Objective Learning and Mask-Based Post-Processing for Deep Neural Network Based Speech Enhancement*

*Yong Xu¹, Jun Du¹, Zhen Huang², Li-Rong Dai¹, Chin-Hui Lee²*

¹USTC, China; ²Georgia Institute of Technology, USA

---

**Page 1513**

**Tue-O-21-5**

15:20 – 15:40

*Discriminative Nonnegative Matrix Factorization Using Cross-Reconstruction Error for Source Separation*

*Kisoo Kwon¹, Jong Won Shin², Hyung Yong Kim¹, Nam Soo Kim¹*

¹Seoul National University, Korea; ²GIST, Korea
Tue-O-21 continued…

Using Audio and Visual Information for Single Channel Speaker Separation

Faheem Khan, Ben Milner, University of East Anglia, UK

---

**Oral Session 22: Emotion**

*Chairs: Carlos Busso Recabarren, Nick Campbell*
*Room: Conference 2+3, Time 14:00 – 16:00, Tuesday, September 8, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 22: Emotion</th>
</tr>
</thead>
</table>
| 1522  | Exploring Acoustic Differences Between Cantonese (Tonal) and English (Non-Tonal) Spoken Expressions of Emotions  
*Chee Seng Chong, Jeesun Kim, Chris Davis, University of Western Sydney, Australia* |
| 1527  | Valence, Arousal and Dominance Estimation for English, German, Greek, Portuguese and Spanish Lexica Using Semantic Models  
*Elisavet Palogiannidi¹, Elias Iosif², Polychronis Koutsakis¹, Alexandros Potamianos²*  
¹Technical University of Crete, Greece; ²Athena RC, Greece |
| 1532  | Dimensionality Reduction for Speech Emotion Features by Multiscale Kernels  
*Xinzhou Xu¹, Jun Deng¹, Wenming Zheng², Li Zhao², Björn Schuller³*  
¹Technische Universität München, Germany; ²Southeast University, China; ³Imperial College London, UK |
| 1537  | High-Level Feature Representation Using Recurrent Neural Network for Speech Emotion Recognition  
*Jinkyu Lee¹, Ivan Tashev²*  
¹Yonsei University, Korea; ²Microsoft, USA |
| 1541  | Speech Emotion Classification Using Tree-Structured Sparse Logistic Regression  
*Myung Jong Kim, Joohong Yoo, Younggwan Kim, Hoirin Kim, KAIST, Korea* |
| 1546  | Annotators’ Agreement and Spontaneous Emotion Classification Performance  
*Bogdan Vlasenko¹, Andreas Wendemuth²*  
¹Idiap Research Institute, Switzerland; ²Otto-von-Guericke-Universität Magdeburg, Germany |
Oral Session 23: Computational Models of Human Speech Perception

Chairs: Martin Cooke, Okko Räsänen
Room: Conference 4+5, Time 14:00 – 16:00, Tuesday, September 8, 2015

On the Nature of the Features Generated in the Human Auditory Pathway for Phone Recognition
Harald Höge, Universität der Bundeswehr München, Germany

How the Slope of the Speech Spectrum Affects the Perception of Speaker Size
Kodai Yamamoto¹, Toshio Irino¹, Ryuichi Nisimura¹, Hideki Kawahara¹, Roy D. Patterson²
¹Wakayama University, Japan; ²University of Cambridge, UK

Weakly-Supervised Word Learning is Improved by an Active Online Algorithm
Heikki Rasilo, Okko Räsänen, Aalto University, Finland

The Effect of Cochlear Implant Processing on Speaker Intelligibility: A Perceptual Study and Computer Model
Lin Lin¹, Jon Barker², Guy J. Brown²
¹Jilin University, China; ²University of Sheffield, UK

Phonetic-Phonological Feature Emerges by Associating Phonetic with Semantic Information — A GSOM-Based Modeling Study
Mengxue Cao¹, Aijun Li¹, Qiang Fang¹, Bernd J. Kröger²
¹Chinese Academy of Social Sciences, China; ²RWTH Aachen University, Germany

DIANA: Towards Computational Modeling Reaction Times in Lexical Decision in North American English
L. ten Bosch¹, L. Bowes¹, B. Tucker², M. Ernestus¹
¹Radboud Universiteit Nijmegen, The Netherlands; ²University of Alberta, Canada

Oral Session 24: Prosody Modeling for Speech Synthesis

Chairs: Katrin Schweitzer, Ingmar Steiner
Room: Conference 6, Time 14:00 – 16:00, Tuesday, September 8, 2015

Automatic Phrase Boundary Labeling of Speech Synthesis Database Using Context-Dependent HMMs and N-Gram Prior Distributions
Qian Chen¹, Zhen-Hua Ling¹, Chen-Yu Yang², Li-Rong Dai¹
¹USTC, China; ²A*STAR, Singapore

A Perceptual Investigation of Wavelet-Based Decomposition of $f_0$ for Text-to-Speech Synthesis
Manuel Sam Ribeiro, Junichi Yamagishi, Robert A.J. Clark, University of Edinburgh, UK

Duration Prediction Using Multi-Level Model for GPR-Based Speech Synthesis
Decha Moungsri, Tomoki Koriyama, Takao Kobayashi, Tokyo Institute of Technology, Japan

Data-Driven Foot-Based Intonation Generator for Text-to-Speech Synthesis
Mahsa Sadat Elyasi Langarani, Jan van Santen, Seyed Hamidreza Mohammadi, Alexander Kain, Oregon Health & Science University, USA

Weighted Correlation Based Atom Decomposition Intonation Modelling
Branislav Gerazov¹, Pierre-Edouard Honnet², Aleksandar Gjoreski¹, Philip N. Garner²
¹UKiM, Macedonia; ²Idiap Research Institute, Switzerland

Using Deep Bidirectional Recurrent Neural Networks for Prosodic-Target Prediction in a Unit-Selection Text-to-Speech System
Raul Fernandez¹, Asaf Rendel², Bhuvana Ramabhadran¹, Ron Hoory²
¹IBM T.J. Watson Research Center, USA; ²IBM Research Haifa, Israel
Special Session 4: Speech and Language Processing of Children’s Speech

Chairs: Kay Berkling, Diego Giuliani
Room: Hall 1, Time 14:00 – 16:00, Tuesday, September 8, 2015

Oral Presentations

**Large Vocabulary Automatic Speech Recognition for Children**
Hank Liao, Golan Pundak, Olivier Siohan, Melissa K. Carroll, Noah Coccaro, Qi-Ming Jiang, Tara N. Sainath, Andrew Senior, Françoise Beaufays, Michiel Bacchiani, Google, USA

**Acoustic-Prosodic Correlates of ‘Awkward’ Prosody in Story Retellings from Adolescents with Autism**
Daniel Bone¹, Matthew P. Black¹, Anil Ramakrishna¹, Ruth Grossman², Shrikanth S. Narayanan¹
¹University of Southern California, USA; ²Emerson College, USA

**Evidence of Phonological Processes in Automatic Recognition of Children’s Speech**
Eva Fringi¹, Jill Fain Lehman², Martin Russell¹
¹University of Birmingham, UK; ²Disney Research, USA

Poster Presentations

**Influence of Speaker Familiarity on Blind and Visually Impaired Children’s Perception of Synthetic Voices in Audio Games (Poster)**
Michael Pucher¹, Markus Toman¹, Dietmar Schabus¹, Cassia Valentini-Botinhao², Junichi Yamagishi², Bettina Zillinger³, Erich Schmid⁴
¹FTW, Austria; ²University of Edinburgh, UK; ³Fachhochschule Wiener Neustadt, Austria; ⁴Bundes-Blindenerziehungsinstitut, Austria

**Low-Memory Fast On-Line Adaptation for Acoustically Mismatched Children’s Speech Recognition (Poster)**
S. Shahnawazuddin, Rohit Sinha, IIT Guwahati, India

**Large Vocabulary Children’s Speech Recognition with DNN-HMM and SGMM Acoustic Modeling (Poster)**
Diego Giuliani¹, Bagher BabaAli²
¹FBK, Italy; ²University of Tehran, Iran

**HMM Adaptation for Child Speech Synthesis (Poster)**
Avashna Govender¹, Febe de Wet¹, Jules-Raymond Tapamo²
¹CSIR, South Africa; ²University of KwaZulu-Natal, South Africa

**Vocal Turn-Taking Patterns in Groups of Children Performing Collaborative Tasks: An Exploratory Study (Poster)**
Jaebok Kim, Khiet P. Truong, Vicky Charisi, Cristina Zaga, Manja Lohse, Dirk Heylen, Vanessa Evers, University of Twente, The Netherlands

**Towards an Automated Screening Tool for Pediatric Speech Delay (Poster)**
Roozbeh Sadeghian, Stephen A. Zahorian, Binghamton University, USA

**Children’s Reading Aloud Performance: A Database and Automatic Detection of Disfluencies (Poster)**
Jorge Proença¹, Dirce Celorico¹, Sara Candeias², Carla Lopes¹, Fernando Perdigão¹
¹Instituto de Telecomunicações, Portugal; ²Microsoft, Portugal

**Keyword Spotting in Multi-Player Voice Driven Games for Children (Poster)**
Harshavardhan Sundar¹, Jill Fain Lehman², Rita Singh²
¹Carnegie Mellon University, USA; ²Disney Research, USA
Tue-SP4 continued…

Age-Dependent Height Estimation and Speaker Normalization for Children’s Speech Using the First Three Subglottal Resonances (Poster)
Jinxi Guo¹, Rohit Paturi¹, Gary Yeung¹, Steven M. Lulich², Harish Arsikere³, Abeer Alwan¹
¹University of California at Los Angeles, USA; ²Indiana University, USA; ³Xerox Research Center India, India

Poster Session 14: Syllables and Segments

The Effect of Speakers’ Regional Varieties on Listeners’ Decision-Making
Adrian Leemann¹, Camilla Bernardasci², Francis Nolan¹
¹University of Cambridge, UK; ²Universität Zürich, Switzerland

Word-Initial Glottal Stop Insertion, Hiatus Resolution and Linking in British English
Robert Fuchs, Westfälische Wilhelms-Universität Münster, Germany

Acoustic Analysis of Mandarin Affricates
Shanpeng Li, Wentao Gu, Nanjing Normal University, China

Homophonous Phonotactic and Morphonotactic Consonant Clusters in Word-Final Position
Hannah Leykum, Sylvia Moosmüller, Wolfgang U. Dressler, Austrian Academy of Sciences, Austria

Consonant Duration and VOT as a Function of Syllable Complexity and Voicing in a Sub-Set of Spanish Clusters
Mark Gibson¹, Ana Maria Fernández Planas², Adamantios Gafos³, Emily Remirez¹
¹Rice University, USA; ²Universitat de Barcelona, Spain; ³Universität Potsdam, Germany

Hands-On Tool Producing Front Vowels for Phonetic Education: Aiming for Pronunciation Training with Tactile Sensation
Takayuki Arai, Sophia University, Japan

Acoustics of Articulatory Constraints: Vowel Classification and Nasalization
Indranil Dutta, Ayushi Pandey, EFL University, India

Voice-Conditioned Allophones of MOUTH and PRICE in Bahamian Creole
Janina Kraus, LMU München, Germany

Analysis of Spatial Variation with App-Based Crowdsourced Audio Data
Marie-José Kolly¹, Adrian Leemann², Florian Matter³
¹LIMSI, France; ²University of Cambridge, UK; ³University of Bern, Switzerland

Confusability in L2 Vowels: Analyzing the Role of Different Features
Mátyás Jani¹, Catia Cucchiari², Roeland van Hout², Helmer Strik²
¹Pázmány Péter Catholic University, Hungary; ²Radboud Universiteit Nijmegen, The Netherlands

Perception of French Speakers’ German Vowels
Frank Zimmerer, Jürgen Trouvain, Universität des Saarlandes, Germany

Unintuitive Phonetic Behavior in Tswana Post-Nasal Stops
Jagoda Bruni, Daniel Duran, Grzegorz Dogil, Universität Stuttgart, Germany
Poster Session 15: Speech Enhancement

Chair: Rainer Martin
Room: Hall 2, Time 14:00 - 16:00, Tuesday, September 8, 2015

Modeling Temporal Dependency for Robust Estimation of LP Model Parameters in Speech Enhancement
Chun Hoy Wong, Tan Lee, Yu Ting Yeung, P.C. Ching, Chinese University of Hong Kong, China

Learning a Speech Manifold for Signal Subspace Speech Denoising
Colin Vaz, Shrikanth S. Narayanan, University of Southern California, USA

An Iterative Speech Model-Based a priori SNR Estimator
Samy Elshamy1, Nilesh Madhu2, Wouter Tirry2, Tim Fingscheidt1
1 Technische Universität Braunschweig, Germany; 2 NXP Software, Belgium

Multi-Resolution Stacking for Speech Separation Based on Boosted DNN
Xiao-Lei Zhang, DeLiang Wang, Ohio State University, USA

Least Squares Estimate of the Initial Phases in STFT Based Speech Enhancement
Sidse Marie Norholm1, Martin Krawczyk-Becker2, Timo Gerkmann2, Steven van de Par2, Jesper Rindom Jensen1, Mads Grønbøll Christensen1
1 Aalborg University, Denmark; 2 Carl von Ossietzky Universität Oldenburg, Germany

Enhancement of Non-Stationary Speech Using Harmonic Chirp Filters
Sidse Marie Norholm, Jesper Rindom Jensen, Mads Grønbøll Christensen, Aalborg University, Denmark

Text-Informed Speech Enhancement with Deep Neural Networks
Keisuke Kinoshita, Marc Delcroix, Atsunori Ogawa, Tomohiro Nakatani, NTT Corporation, Japan

Complex Tensor Factorization in Modulation Frequency Domain for Single-Channel Speech Enhancement
Shogo Masaya, Masashi Unoki, JAIST, Japan

Systematic Integration of Acoustic Echo Canceller and Noise Reduction Modules for Voice Communication Systems
Hyeonjoo Kang1, JeeSok Lee1, Soonho Baek2, Hong-Goo Kang1
1 Yonsei University, Korea; 2 Samsung Electronics, Korea

DNN-Based Residual Echo Suppression
Chul Min Lee1, Jong Won Shin2, Nam Soo Kim1
1 Seoul National University, Korea; 2 GIST, Korea

Codebook-Based Speech Enhancement Using Markov Process and Speech-Presence Probability
Qi He, Changchun Bao, Feng Bao, Beijing University of Technology, China

On Optimal Smoothing in Minimum Statistics Based Noise Tracking
Aleksej Chinaev, Reinhold Haeb-Umbach, Universität Paderborn, Germany

A Data-Driven Speech Enhancement Method Based on Modeled Long-Range Temporal Dynamics
Yue Hao, Changchun Bao, Feng Bao, Feng Deng, Beijing University of Technology, China

Improved Phase Reconstruction in Single-Channel Speech Separation
Florian Mayer, Pejman Mowlaee, Technische Universität Graz, Austria
Poster Session 16: Spoken Language Understanding

Chair: Roger K. Moore
Room: Foyer, Time 14:00 - 16:00, Tuesday, September 8, 2015

Page 1800
Tue-P-16-1
Dialog State Tracking Using Long Short-Term Memory Neural Networks
Xiaohao Yang, Jia Liu, Tsinghua University, China

Page 1805
Tue-P-16-2
Detecting Repetitions in Spoken Dialogue Systems Using Phonetic Distances
José Lopes 1, Giampiero Salvi 1, Gabriel Skantze 1, Alberto Abad 2, Joakim Gustafson 1, Fernando Batista 2, Raveesh Meena 1, Isabel Trancoso 2
1 KTH, Sweden; 2 INESC-ID Lisboa, Portugal

Page 1810
Tue-P-16-3
Multi-Language Hypotheses Ranking and Domain Tracking for Open Domain Dialogue Systems
Paul A. Crook, Jean-Philippe Robichaud, Ruhi Sarikaya, Microsoft, USA

Page 1815
Tue-P-16-4
Measuring Mimicry in Task-Oriented Conversations: Degree of Mimicry is Related to Task Difficulty
Vijay Solanki, Alessandro Vinciarelli, Jane Stuart-Smith, Rachel Smith, University of Glasgow, UK

Page 1820
Tue-P-16-5
Auto-Imputing Radial Basis Functions for Neural-Network Turn-Taking Models
Kornel Laskowski, Carnegie Mellon University, USA

Page 1825
Tue-P-16-6
Effect of Gender and Call Duration on Customer Satisfaction in Call Center Big Data
Quim Llimona, Jordi Luque, Xavier Anguera, Zoraida Hidalgo, Souenil Park, Nuria Oliver, Telefónica I+D, Spain

Page 1830
Tue-P-16-7
Using Profile Similarity to Measure Agreement in Personality Perception
Zoraida Callejas 1, David Griol 2
1 Universidad de Granada, Spain; 2 Universidad Carlos III de Madrid, Spain

Tue-P-16 continued…

Page 1835
Tue-P-16-8
Relieving Mental Stress of Speakers Using a Tele-Operated Robot in Foreign Language Speech Education
Shizuka Nakamura, Miki Watanabe, Yuichiro Yoshikawa, Kohei Ogawa, Hiroshi Ishiguro, Osaka University, Japan

Page 1839
Tue-P-16-9
Backward Mimicry and Forward Influence in Prosodic Contour Choice in Standard American English
Agustín Gravano 1, Štefan Beňuš 2, Rivka Levitan 3, Julia Hirschberg 4
1 Universidad de Buenos Aires, Argentina; 2 UKF, Slovak Republic; 3 CUNY Brooklyn College, USA; 4 Columbia University, USA

Page 1844
Tue-P-16-10
The Role of Speakers and Context in Classifying Competition in Overlapping Speech
Shammur Absar Chowdhury, Morena Danielli, Giuseppe Riccardi, Università di Trento, Italy

Page 1849
Tue-P-16-11
Automatic Detection and Annotation of Disfluencies in Spoken French Corpora
George Christodoulides 1, Mathieu Avanzi 2
1 Université catholique de Louvain, Belgium; 2 University of Cambridge, UK

Page 1854
Tue-P-16-12
Clustering Novel Intents in a Conversational Interaction System with Semantic Parsing
Dilek Hakkani-Tür, Yun-Cheng Ju, Geoffroy Zweig, Gokhan Tur, Microsoft, USA

Page 1859
Tue-P-16-13
Semantic Analysis of Spoken Input Using Markov Logic Networks
Vladimir Despotovic 1, Oliver Walter 2, Reinhold Haeb-Umbach 2
1 University of Belgrade, Serbia; 2 Universität Paderborn, Germany
**Tue-P-16 continued…**

**Hierarchical Discriminative Model for Spoken Language Understanding Based on Convolutional Neural Network**  
*Jan Švec, Adam Chýlek, Luboš Smidl*, University of West Bohemia, Czech Republic

**Learning Semantic Hierarchy with Distributed Representations for Unsupervised Spoken Language Understanding**  
*Yun-Nung Chen, William Yang Wang, Alexander I. Rudnicky*, Carnegie Mellon University, USA

---

**Show and Tell Session 3**  
*Chair: Hemant Patil*  
*Room: Seminar 1, Time 14:00 – 16:00, Tuesday, September 8, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Show &amp; Tell</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 1874  | Show&Tell-3-1| Phontasia — A Game for Training German Orthography                   | Kay Berkling¹, Nadine Pflaumer², Alexei Coyplove³  
¹DHBW, Germany; ²Logopraxen, Germany; ³Inline Internet Online, Germany |
| 1876  | Show&Tell-3-2| E-Commu-Book: An Assistive Technology for Users with Speech Impairments| Ka Ho Wong, Wai Kim Leung, Helen Meng, Chinese University of Hong Kong, China                                          |
| 1878  | Show&Tell-3-3| Swiss GraphoGame: Concept and Design Presentation of a Computerised Reading Intervention for Children with High Risk for Poor Reading Outcomes| Martina Röthlisberger¹, Ilkana I. Karipidis¹, Georgette Pleisch¹, Volker Dellwo¹, Ulla Richardson², Silvia Brem¹  
¹Universität Zürich, Switzerland; ²University of Jyväskylä, Finland |
| 1880  | Show&Tell-3-4| Neolexon — A Therapy App for Patients with Aphasia                   | Jakob Pfab¹, Hanna Jakob², Mona Späth², Christoph Draxler¹  
¹LMU München, Germany; ²Städtisches Klinikum München, Germany |
| 1882  | Show&Tell-3-5| Acoustic Stress Detection for Improved Navigation of Educational Videos| Sonal Patil, Harish Arsikere, Om Deshmukh, Xerox Research Center India, India                                        |
| 1884  | Show&Tell-3-6| Multimodal Read-Aloud eBooks for Language Learning                  | Xavier Anguera, Sinkronigo, Spain                                                                                      |
| 1886  | Show&Tell-3-7| Speech Technologies for African Languages: Example of a Multilingual Calculator for Education| Laurent Besacier¹, Elodie Gauthier¹, Mathieu Mangeot¹, Philippe Bretier², Paul Bagshaw², Olivier Rosec², Thierry Moudenc², François Pellegrino³, Sylvie Voisin³, Egidio Marsico³, Pascal Nocera⁴  
¹LIG (UMR 5217), France; ²Voxygen, France; ³DDL (UMR 5596), France; ⁴LIA, France |

---

**Special Event: Privacy Issues in Speech Data Collection and Usage**  
*Chair: Michael Wagner*  
*Room: Large Hall, Time 16:30 – 17:30, Tuesday, September 8, 2015*
### Oral Session 25: Phonetic Recognition: Novel Approaches and Understanding

**Chairs:** Lukáš Burget, Yifan Gong  
**Room:** Large Hall, Time 09:00 – 11:00, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Paper Title</th>
<th>Authors</th>
<th>Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1888</td>
<td>Time-Frequency Kernel-Based CNN for Speech Recognition</td>
<td>Tuo Zhao$^1$, Yunxin Zhao$^1$, Xin Chen$^2$</td>
<td>$^1$University of Missouri, USA; $^2$Pearson Knowledge Technologies, USA</td>
</tr>
<tr>
<td>1898</td>
<td>Investigating Factor Analysis Features for Deep Neural Networks in Noisy Speech Recognition</td>
<td>Sriram Ganapathy, Samuel Thomas, Dimitrios Dimitriadis, Steven Rennie, IBM T.J. Watson Research Center, USA</td>
<td></td>
</tr>
<tr>
<td>1903</td>
<td>Ensemble of Gaussian Mixture Localized Neural Networks with Application to Phone Recognition</td>
<td>Ruchir Travadi, Shrikanth S. Narayanan, University of Southern California, USA</td>
<td></td>
</tr>
<tr>
<td>1908</td>
<td>DNN Derived Filters for Processing of Modulation Spectrum of Speech</td>
<td>Jan Pešán, Lukáš Burget, Hynek Hermansky, Karel Vesely, Brno University of Technology, Czech Republic</td>
<td></td>
</tr>
<tr>
<td>1912</td>
<td>Exploring How Deep Neural Networks Form Phonemic Categories</td>
<td>Tasha Nagamine$^1$, Michael L. Seltzer$^2$, Nima Mesgarani$^1$</td>
<td>$^1$Columbia University, USA; $^2$Microsoft, USA</td>
</tr>
</tbody>
</table>

---

### Oral Session 26: Varieties of Speech

**Chairs:** Mattias Heldner, Petra Wagner  
**Room:** Conference 1, Time 09:00 – 11:00, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Paper Title</th>
<th>Authors</th>
<th>Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1917</td>
<td>Pronunciation Accuracy and Intelligibility of Non-Native Speech</td>
<td>Anastassia Loukina, Melissa Lopez, Keelan Evanini, David Suendermann-Oeft, Alexei V. Ivanov, Klaus Zechnier, Educational Testing Service, USA</td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td>Productions of /h/ in German: French vs. German Speakers</td>
<td>Frank Zimmerer, Jürgen Trouvain, Universität des Saarlandes, Germany</td>
<td></td>
</tr>
<tr>
<td>1927</td>
<td>German Non-Native Realizations of French Voiced Fricatives in Final Position of a Group of Words</td>
<td>Anne Bonneau, Martine Cadot, LORIA, France</td>
<td></td>
</tr>
<tr>
<td>1932</td>
<td>From Newcastle MOUTH to Aussie Ears: Australians’ Perceptual Assimilation and Adaptation for Newcastle UK Vowels</td>
<td>Catherine T. Best$^1$, Jason A. Shaw$^1$, Gerard Docherty$^2$, Bronwen G. Evans$^3$, Paul Foulkes$^4$, Jennifer Hay$^5$, Jalal Al-Tamimi$^6$, Katharine Mair$^3$, Karen E. Mulak$^1$, Sophie Wood$^4$</td>
<td>$^1$University of Western Sydney, Australia; $^2$Griffith University, Australia; $^3$University College London, UK; $^4$University of York, UK; $^5$University of Canterbury, New Zealand; $^6$Newcastle University, UK</td>
</tr>
<tr>
<td>1937</td>
<td>Wubuy Coronal Stop Perception by Speakers of Three Dialects of Bangla</td>
<td>Rikke Louise Bundgaard-Nielsen$^1$, Brett Baker$^2$, Olga Maxwell$^2$, Janet Fletcher$^2$</td>
<td>$^1$University of Western Sydney, Australia; $^2$University of Melbourne, Australia</td>
</tr>
<tr>
<td>1942</td>
<td>Using Melody Metrics to Compare English Speech Read by Native Speakers and by L2 Chinese Speakers from Shanghai</td>
<td>Daniel Hirst$^1$, Hongwei Ding$^2$</td>
<td>$^1$LPL, France; $^2$Shanghai Jiao Tong University, China</td>
</tr>
</tbody>
</table>

---
Oral Session 27: Conversational Interaction

Chairs: Alan W. Black, Nigel G. Ward
Room: Conference 2+3, Time 09:00 – 11:00, Wednesday, September 9, 2015

PAGE 1947  Wed-O-27-1  09:00 – 09:20
Predicting Therapist Empathy in Motivational Interviews Using Language Features Inspired by Psycholinguistic Norms
James Gibson1, Nikolaos Malandrakis1, Francisco Romero1, David C. Atkins2, Shrikanth S. Narayanan1
1University of Southern California, USA; 2University of Washington, USA

PAGE 1952  Wed-O-27-2  09:20 – 09:40
Therapy Language Analysis Using Automatically Generated Psycholinguistic Norms
Nikolaos Malandrakis, Shrikanth S. Narayanan, University of Southern California, USA

PAGE 1957  Wed-O-27-3  09:40 – 10:00
A Dynamic Model for Behavioral Analysis of Couple Interactions Using Acoustic Features
Wei Xia1, James Gibson1, Bo Xiao1, Brian Baucom2, Panayiotis G. Georgiou1
1University of Southern California, USA; 2University of Utah, USA

PAGE 1962  Wed-O-27-4  10:00 – 10:20
Analysis and Modeling of the Role of Laughter in Motivational Interviewing Based Psychotherapy Conversations
Rahul Gupta1, Theodora Chaspari1, Panayiotis G. Georgiou1, David C. Atkins2, Shrikanth S. Narayanan1
1University of Southern California, USA; 2University of Washington, USA

The Discourse Value of Social Signals at Topic Change Moments
Francesca Bonin, Nick Campbell, Carl Vogel, Trinity College Dublin, Ireland

PAGE 1972  Wed-O-27-6  10:40 – 11:00
Automatic Detection of Uncertainty in Spontaneous German Dialogue
Tobias Schrank, Barbara Schuppler, Technische Universität Graz, Austria

Oral Session 28: Speech and Audio Segmentation and Classification

Chairs: Haizhou Li, Koichi Shinoda
Room: Conference 4+5, Time 09:00 – 11:00, Wednesday, September 9, 2015

PAGE 1977  Wed-O-28-1  09:00 – 09:20
Face Reading from Speech — Predicting Facial Action Units from Audio Cues
Fabien Ringeval1, Erik Marchi1, Marc Mehu2, Klaus Scherer3, Björn Schuller3
1Technische Universität München, Germany; 2Webster Vienna Private University, Austria; 3Université de Genève, Switzerland

PAGE 1982  Wed-O-28-2  09:20 – 09:40
A New Front-End for Classification of Non-Speech Sounds: A Study on Human Whistle
Mahesh Kumar Nandwana, Hynek Bořil, John H.L. Hansen, University of Texas at Dallas, USA

PAGE 1987  Wed-O-28-3  09:40 – 10:00
Robust Features for Sonorant Segmentation in Continuous Speech
Sri Harsha Dumpala, Bhanu Teja Nellore, Raghu Ram Nevali, Suryakanth V. Gangashetty, B. Yegnanarayana, IIIT Hyderabad, India

PAGE 1992  Wed-O-28-4  10:00 – 10:20
Reduction of Reverberation Effects in the MFCC Modulation Spectrum for Improved Classification of Acoustic Signals
Sebastian Gergen, Anil Nagathil, Rainer Martin, Ruhr-Universität Bochum, Germany

Spiking Neural Networks and the Generalised Hough Transform for Speech Pattern Detection
Jonathan Dennis, Huy Dat Tran, Haizhou Li, A*STAR, Singapore

PAGE 2002  Wed-O-28-6  10:40 – 11:00
Acoustic Event Recognition Using Dominant Spectral Basis Vectors
Woohyun Choi1, Sangwook Park1, David K. Han2, Hanseok Ko1
1Korea University, Korea; 2ONR, USA
Oral Session 29: Spoken Dialogue Systems

Chairs: Satoshi Nakamura, Alexander Rudnicky
Room: Conference 6, Time 09:00 – 11:00, Wednesday, September 9, 2015

Learning from Real Users: Rating Dialogue Success with Neural Networks for Reinforcement Learning in Spoken Dialogue Systems
Pei-Hao Su, David Vandyke, Milica Gašić, Dongho Kim, Nikola Mrkšić, Tsung-Hsien Wen, Steve Young, University of Cambridge, UK

A Framework to Develop Context-Aware Adaptive Dialogue System
David Griol 1, Zoraida Callejas 2, Ramón López-Cózar 2
1 Universidad Carlos III de Madrid, Spain; 2 Universidad de Granada, Spain

A Proposal to Develop Domain and Subtask-Adaptive Dialog Management Models
David Griol 1, Zoraida Callejas 2
1 Universidad Carlos III de Madrid, Spain; 2 Universidad de Granada, Spain

Hypotheses Ranking and State Tracking for a Multi-Domain Dialog System Using Multiple ASR Alternates
Omar Zia Khan, Jean-Philippe Robichaud, Paul A. Crook, Ruhi Sarikaya, Microsoft, USA

An Entropy Minimization Framework for Goal-Driven Dialogue Management
Ji Wu 1, Miao Li 1, Chin-Hui Lee 2
1 Tsinghua University, China; 2 Georgia Institute of Technology, USA

Context-Dependent Error Correction of Spoken Referring Expressions
Ingrid Zukerman, Andisheh Partovi, Su Nam Kim, Monash University, Australia

Special Session 5: Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof 2015)

Chairs: Tomi Kinnunen, Zhizheng Wu
Room: Hall 1, Time 09:00 – 11:00, Wednesday, September 9, 2015

Introductory Talk by the Organizers
Zhizheng Wu 1, Tomi Kinnunen 2
1 University of Edinburgh, UK; 2 University of Eastern Finland, Finland

ASVspoof 2015: The First Automatic Speaker Verification Spoofing and Countermeasures Challenge (Poster)
Zhizheng Wu 1, Tomi Kinnunen 2, Nicholas Evans 3, Junich Yamagishi 1, Cemal Hanifçi 2, Md. Sahidullah 2, Aleksandr Sizov 2
1 University of Edinburgh, UK; 2 University of Eastern Finland, Finland; 3 EURECOM, France

The AHOLAB RPS SSD Spoofing Challenge 2015 Submission (Poster)
Jon Sanchez, Ibon Saratxaga, Inma Hernaez, Eva Navas, D. Erro, Universidad del País Vasco, Spain

Human vs Machine Spoofing Detection on Wideband and Narrowband Data (Poster)
Mirjam Wester, Zhizheng Wu, Junich Yamagishi, University of Edinburgh, UK

Spoofing Speech Detection Using High Dimensional Magnitude and Phase Features: The NTU Approach for ASVspoof 2015 Challenge (Poster)
Xiong Xiao 1, Xiaohai Tian 2, Steven Du 1, Haihua Xu 1, Eng Siong Chng 1, Haizhou Li 1
1 TL@NTU, Singapore; 2 NTU, Singapore
| Page  | Wed-SP5-5 | Classifiers for Synthetic Speech Detection: A Comparison (Poster)  
Cemal Hanilçi, Tomi Kinnunen, Md. Sahidullah, Aleksandr Sizov, University of Eastern Finland, Finland |
|-------|-----------|------------------------------------------------------------------|
| Page  | Wed-SP5-6 | Combining Evidences from Mel Cepstral, Cochlear Filter Cepstral and Instantaneous Frequency Features for Detection of Natural vs. Spoofed Speech (Poster)  
Tanvina B. Patel, Hemant A. Patil, DA-IICT, India |
| Page  | Wed-SP5-7 | Spoofing Detection with DNN and One-Class SVM for the ASVspoof 2015 Challenge (Poster)  
Jesús Villalba, Antonio Miguel, Alfonso Ortega, Eduardo Lleida, Universidad de Zaragoza, Spain |
| Page  | Wed-SP5-8 | Development of CRIM System for the Automatic Speaker Verification Spoofing and Countermeasures Challenge 2015 (Poster)  
Md. Jahangir Alam, Patrick Kenny, Gautam Bhattacharya, Themis Stafylakis, CRIM, Canada |
| Page  | Wed-SP5-9 | Spoofing Countermeasure Based on Analysis of Linear Prediction Error (Poster)  
Artur Janicki, Warsaw University of Technology, Poland |
| Page  | Wed-SP5-10 | Simultaneous Utilization of Spectral Magnitude and Phase Information to Extract Supervectors for Speaker Verification Anti-Spoofing (Poster)  
Yi Liu¹, Yao Tian¹, Liang He¹, Jia Liu¹, Michael T. Johnson²  
¹Tsinghua University, China; ²Marquette University, USA |
| Page  | Wed-SP5-11 | A Comparison of Features for Synthetic Speech Detection (Poster)  
Md. Sahidullah, Tomi Kinnunen, Cemal Hanilçi, University of Eastern Finland, Finland |
|-------|-----------|------------------------------------------------------------------|
| Page  | Wed-SP5-12 | Relative Phase Information for Detecting Human Speech and Spoofed Speech (Poster)  
Longbiao Wang¹, Yohei Yoshida¹, Yuta Kawakami¹, Seiichi Nakagawa²  
¹Nagaoka University of Technology, Japan; ²Toyohashi University of Technology, Japan |
| Page  | Wed-SP5-13 | Robust Deep Feature for Spoofing Detection — The SJTU System for ASVspoof 2015 Challenge (Poster)  
Nanxin Chen, Yanmin Qian, Heinrich Dinkel, Bo Chen, Kai Yu, Shanghai Jiao Tong University, China |
| Wed-SP5-D | 10:30 – 11:00 | Open Discussion and Future Plans  
Junichi Yamagishi¹, Nicholas Evans²  
¹University of Edinburgh, UK; ²EURECOM, France |
Poster Session 17: Acoustic Modeling and Decoding Methods for Speech Recognition

Chair: Thomas Hain
Room: Hall 2, Time 09:00 – 11:00, Wednesday, September 9, 2015

**PAGE 2102**
**Wed-P-17-1**

Applying GPGPU to Recurrent Neural Network Language Model Based Fast Network Search in the Real-Time LVCSR

Kyungmin Lee, Chiyoun Park, Ilhwan Kim, Namhoon Kim, Jaewon Lee, Samsung Electronics, Korea

**PAGE 2107**
**Wed-P-17-2**

Real-Time Integration of Dynamic Context Information for Improving Automatic Speech Recognition

Youssef Oualil¹, Marc Schulder¹, Hartmut Helmke², Anna Schmidt¹,
Dietrich Klakow¹
¹Universität des Saarlandes, Germany; ²DLR, Germany

**PAGE 2112**
**Wed-P-17-3**

Rapid Vocabulary Addition to Context-Dependent Decoder Graphs

Cyril Allauzen, Michael Riley, Google, USA

**PAGE 2117**
**Wed-P-17-4**

Modeling Phonetic Context with Non-Random Forests for Speech Recognition

Hainan Xu, Guoguo Chen, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

**PAGE 2122**
**Wed-P-17-5**

Ant Colony Algorithm Applied to Automatic Speech Recognition Graph Decoding

Benjamin Lecouteux, Didier Schwab, LIG (UMR 5217), France

**PAGE 2127**
**Wed-P-17-6**

Garbage Modeling for On-Device Speech Recognition

Christophe Van Gysel¹, Leonid Velikovich², Ian McGraw², Françoise Beaufays²
¹University of Amsterdam, The Netherlands; ²Google, USA

INTERSECH 2015 Table of Contents

**Wed-P-17 continued…**

**PAGE 2132**
**Wed-P-17-7**

A Comparative Study of BNF and DNN Multilingual Training on Cross-Lingual Low-Resource Speech Recognition

Haihua Xu, Van Hai Do, Xiong Xiao, Eng Siong Chng, TL@NTU, Singapore

**PAGE 2137**
**Wed-P-17-8**

Neural Higher-Order Factors in Conditional Random Fields for Phoneme Classification

Martin Ratajczak¹, Sebastian Tschatschek², Franz Pernkopf¹
¹Technische Universität Graz, Austria; ²ETH Zürich, Switzerland

**PAGE 2142**
**Wed-P-17-9**

Stacked Auto-Encoder for ASR Error Detection and Word Error Rate Prediction

Shahab Jalalvand, Daniele Falavigna, FBK, Italy
Poster Session 18: Speech Production Measurements and Analyses
Chairs: Prasanta Ghosh, Frank Rudzicz
Room: Hall 2, Time 09:00 - 11:00, Wednesday, September 9, 2015

---

**Estimation of the Air-Tissue Boundaries of the Vocal Tract in the Mid-Sagittal Plane from Electromagnetic Articulograph Data**
Satyabrata Parida\(^1\), Ashok Kumar Pattem\(^2\), Prasanta Kumar Ghosh\(^2\)
\(^1\)IIT Kharagpur, India; \(^2\)Indian Institute of Science, India

**A New Italian Dataset of Parallel Acoustic and Articulatory Data**
Claudia Canevari, Leonardo Badino, Luciano Fadiga, Istituto Italiano di Tecnologia, Italy

**Error Analysis of Extracted Tongue Contours from 2D Ultrasound Images**
Tamás Gábor Csapó\(^1\), Steven M. Lulich\(^2\)
\(^1\)BME, Hungary; \(^2\)Indiana University, USA

**Accuracy of a Markerless Acquisition Technique for Studying Speech Articulators**
Andrea Bandini\(^1\), Slim Ouni\(^2\), Piero Cosi\(^3\), Silvia Orlandi\(^1\), Claudia Manfredi\(^1\)
\(^1\)Università di Firenze, Italy; \(^2\)LORIA, France; \(^3\)CNR-ISTC, Italy

**Measuring Oral and Nasal Airflow in Production of Chinese Plosive**
Yujie Chi, Kiyoshi Honda, Jianguo Wei, Hui Feng, Jianwu Dang, Tianjin University, China

**Enhanced Videokymographic Data Analysis Based on Vocal Folds Dynamics Modeling**
Carlo Drioli, Gian Luca Foresti, Università di Udine, Italy

**Interpolation of Tongue Fleshpoint Kinematics from Combined EMA Position and Orientation Data**
Andrew J. Kolb, Michael T. Johnson, Jeffrey Berry, Marquette University, USA

---

**A New Technique for Assessing Glottal Dynamics in Speech and Singing by Means of Optical-Flow Computation**
Gustavo Andrade-Miranda\(^1\), Nathalie Henrich Bernardoni\(^2\),
Juan Ignacio Godino-Llorente\(^1\)
\(^1\)Universidad Politécnica de Madrid, Spain; \(^2\)GIPSA, France

**On the Incompatibility of Trilling and Palatalization: A Single-Subject Study of Sustained Apical and Uvular Trills**
Alexei Kochetov, Phil Howson, University of Toronto, Canada

**Articulatory Movement Prediction Using Deep Bidirectional Long Short-Term Memory Based Recurrent Neural Networks and Word/Phone Embeddings**
Pengcheng Zhu, Lei Xie, Yunlin Chen, Northwestern Polytechnical University, China
Poster Session 19: Speech Synthesis 2
Chair: Cassia Valentini-Botinhao
Room: Foyer, Time 09:00 – 11:00, Wednesday, September 9, 2015

Combining Extreme Learning Machine and Decision Tree for Duration Prediction in HMM Based Speech Synthesis
Yang Wang, Minghao Yang, Zhengqi Wen, Jianhua Tao, Chinese Academy of Sciences, China

F0 Parameterization of Glottalized Tones for HMM-Based Vietnamese TTS
Duy Khanh Ninh, Yoichi Yamashita, Ritsumeikan University, Japan

Deep Neural Network Context Embeddings for Model Selection in Rich-Context HMM Synthesis
Thomas Merritt, Junichi Yamagishi, Zhizheng Wu, Oliver Watts, Simon King, University of Edinburgh, UK

An Investigation of Context Clustering for Statistical Speech Synthesis with Deep Neural Network
Bo Chen, Zhehuai Chen, Jiachen Xu, Kai Yu, Shanghai Jiao Tong University, China

Sentence-Level Control Vectors for Deep Neural Network Speech Synthesis
Oliver Watts, Zhizheng Wu, Simon King, University of Edinburgh, UK

Micro-Structure of Disfluencies: Basics for Conversational Speech Synthesis
Simon Betz, Petra Wagner, David Schlangen, Universität Bielefeld, Germany

Using Automatic Stress Extraction from Audio for Improved Prosody Modelling in Speech Synthesis
György Szaszák1, András Beke2, Gábor Olaszy1, Bálint Pál Tóth1
1BME, Hungary; 2Hungarian Academy of Sciences, Hungary

Reconstructing Voices Within the Multiple-Average-Voice-Model Framework
Pierre Lanchantin1, Christophe Veaux2, Mark J.F. Gales1, Simon King2, Junichi Yamagishi1
1University of Cambridge, UK; 2University of Edinburgh, UK

HMM Based Myanmar Text to Speech System
Ye Kyaw Thu1, Win Pa Pa2, Jinfu Ni1, Yoshinori Shiga1, Andrew Finch1, Chiori Hor1, Hisashi Kawai1, Eiichiro Sumita1
1NICT, Japan; 2UCSY, Myanmar

Multiple Feed-Forward Deep Neural Networks for Statistical Parametric Speech Synthesis
Shinji Takaki1, SangJin Kim2, Junichi Yamagishi1, JongJin Kim2
1NII, Japan; 2Naver Corporation, Korea
Special Event: Standardization Brown-Bag  
Chair: David Suendermann-Oeft  
Room: Large Hall, Time 12:30 – 14:00, Wednesday, September 9, 2015

INTER SPEECH 2015 Table of Contents

Oral Session 30: Spoken Translation & Speech-to-speech  
Chairs: Roland Kuhn, Frank Seide  
Room: Large Hall, Time 14:00 – 16:00, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Session</th>
<th>Chair</th>
<th>Title</th>
<th>Authors</th>
<th>Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>2247</td>
<td>Wed-O-30-1</td>
<td>14:00 – 14:20</td>
<td>Adapting Machine Translation Models Toward Misrecognized Speech with Text-to-Speech Pronunciation Rules and Acoustic Confusability</td>
<td>Nicholas Ruiz¹, Qin Gao², William Lewis², Marcello Federico¹</td>
<td>FBK, Italy; Microsoft, USA</td>
</tr>
<tr>
<td>2252</td>
<td>Wed-O-30-2</td>
<td>14:20 – 14:40</td>
<td>“Speech is Silver, but Silence is Golden”: Improving Speech-to-Speech Translation Performance by Slashing User’s Input</td>
<td>Frederic Bechet, Benoit Favre, Mickael Rouvier, LIF (UMR 7279), France</td>
<td></td>
</tr>
<tr>
<td>2257</td>
<td>Wed-O-30-3</td>
<td>14:40 – 15:00</td>
<td>A Study on the Stability and Effectiveness of Features in Quality Estimation for Spoken Language Translation</td>
<td>Raymond W.M. Ng, Kashif Shah, Lucia Specia, Thomas Hain, University of Sheffield, UK</td>
<td></td>
</tr>
<tr>
<td>2262</td>
<td>Wed-O-30-4</td>
<td>15:00 – 15:20</td>
<td>Efficient Language Model Adaptation for Automatic Speech Recognition of Spoken Translations</td>
<td>Joris Pelemans¹, Tom Vanallemeersch¹, Kris Demuynck², Hugo Van hamme¹, Patrick Wambacq¹</td>
<td>Katholieke Universiteit Leuven, Belgium; Ghent University, Belgium</td>
</tr>
<tr>
<td>2267</td>
<td>Wed-O-30-5</td>
<td>15:20 – 15:40</td>
<td>Speed or Accuracy? A Study in Evaluation of Simultaneous Speech Translation</td>
<td>Takashi Mieno, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura, NAIIST, Japan</td>
<td></td>
</tr>
<tr>
<td>2272</td>
<td>Wed-O-30-6</td>
<td>15:40 – 16:00</td>
<td>Large Scale Speech-to-Text Translation with Out-of-Domain Corpora Using Better Context-Based Models and Domain Adaptation</td>
<td>Marcin Junczys-Dowmunt¹, Pawel Przybylsz¹, Arleta Staszuk¹, Eun-Kyoung Kim², Jaewon Lee²</td>
<td>Samsung Poland R&amp;D Center, Poland; Samsung Electronics, Korea</td>
</tr>
</tbody>
</table>
Oral Session 31: Voice Activity Detection

Chairs: Ma Bin, Nam Soo Kim
Room: Conference 1, Time 14:00 – 16:00, Wednesday, September 9, 2015

A Statistical Model-Based Voice Activity Detection Using Multiple DNNs and Noise Awareness
Inyoung Hwang, Jaeseong Sim, Sang-Hyeon Kim, Kwang-Sub Song, Joon-Hyuk Chang, Hanyang University, Korea

A Universal VAD Based on Jointly Trained Deep Neural Networks
Qing Wang¹, Jun Du¹, Xiao Bao¹, Zi-Rui Wang¹, Li-Rong Dai¹, Chin-Hui Lee²
¹USTC, China; ²Georgia Institute of Technology, USA

Spectrographic Speech Mask Estimation Using the Time-Frequency Correlation of Speech Presence
Ge Zhan, Zhaoqiong Huang, Dongwen Ying, Jielin Pan, Yonghong Yan, Chinese Academy of Sciences, China

Complete-Linkage Clustering for Voice Activity Detection in Audio and Visual Speech
Houman Ghaemmaghami, David Dean, Shahram Kalantari, Sridha Sridharan, Clinton Fookes, Queensland University of Technology, Australia

A Model Based Voice Activity Detector for Noisy Environments
Kaavya Sriskandaraja, Vidhyasaharan Sethu, Phu Ngoc Le, Eliathamby Ambikairajah, University of New South Wales, Australia

An Unsupervised Visual-Only Voice Activity Detection Approach Using Temporal Orofacial Features
Fei Tao, John H.L. Hansen, Carlos Busso, University of Texas at Dallas, USA

Oral Session 32: Advances in iVector-based Speaker Verification

Chairs: Hagai Aronowitz, Douglas A. Reynolds
Room: Conference 2+3, Time 14:00 – 16:00, Wednesday, September 9, 2015

An I-Vector Backend for Speaker Verification
Patrick Kenny¹, Themos Stafylakis¹, Md. Jahangir Alam¹, Marcel Kockmann²
¹CRIM, Canada; ²VoiceTrust, Canada

Multi-Channel Speaker Verification Based on Total Variability Modelling
Joana Correia¹, Alessio Bruttì², Alberto Abad¹
¹INESC-ID Lisboa, Portugal; ²FBK, Italy

SNR-Invariant PLDA Modeling for Robust Speaker Verification
Na Li, Man-Wai Mak, Hong Kong Polytechnic University, China

Investigating In-Domain Data Requirements for PLDA Training
Md. Hafizur Rahman, David Dean, Ahilan Kanagasundaram, Sridha Sridharan, Queensland University of Technology, Australia

Migrating i-Vectors Between Speaker Recognition Systems Using Regression Neural Networks
Ondřej Glembek, Pavel Matějka, Oldřich Plchot, Jan Pešán, Lukáš Burget, Petr Schwarz, Brno University of Technology, Czech Republic

Improving PLDA Speaker Verification Using WMFD and Linear-Weighted Approaches in Limited Microphone Data Conditions
Ahilan Kanagasundaram, David Dean, Sridha Sridharan, Queensland University of Technology, Australia
### Oral Session 33: Voice Quality

**Chairs:** Paavo Alku, Carlos Ishi  
**Room:** Conference 4+5, Time 14:00 – 16:00, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Time</th>
<th>Title</th>
<th>Authors/Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2337</td>
<td>14:00</td>
<td>The Relationship Between Voice Source Parameters and the Maxima Dispersion Quotient (MDQ)</td>
<td>Christer Gobl, Irena Yanushevskaya, Ailbhe Ni Chasaide, Trinity College Dublin, Ireland</td>
</tr>
<tr>
<td>2342</td>
<td>14:20</td>
<td>Glottal Inverse Filtering Based on Quadratic Programming</td>
<td>Manu Airaksinen¹, Tom Bäckström², Paavo Alku¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹Aalto University, Finland; ²FAU Erlangen-Nürnberg, Germany</td>
</tr>
<tr>
<td>2347</td>
<td>14:40</td>
<td>Automatic Detection of Creaky Voice Using Epoch Parameters</td>
<td>N.P. Narendra, K. Sreenivasa Rao, IIT Kharagpur, India</td>
</tr>
<tr>
<td>2352</td>
<td>15:00</td>
<td>Perception of Voicing in the Absence of Native Voicing Experience</td>
<td>Rikke Louise Bundgaard-Nielsen¹, Brett Baker²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹University of Western Sydney, Australia; ²University of Melbourne, Australia</td>
</tr>
<tr>
<td>2357</td>
<td>15:20</td>
<td>The Relationship Between Acoustic and Perceived Intraspeaker Variability in Voice Quality</td>
<td>Jody Kreiman, Soo Jin Park, Patricia A. Keating, Abeer Alwan, University of California at Los Angeles, USA</td>
</tr>
<tr>
<td>2361</td>
<td>15:40</td>
<td>Perceptual Cues of Whispered Tones: Are They Really Special?</td>
<td>Li Jiao¹, Qiuwu Ma¹, Ting Wang¹, Yi Xu²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹Tongji University, China; ²University College London, UK</td>
</tr>
</tbody>
</table>

### Oral Session 34: Neural Networks for Language Modeling

**Chairs:** Ebru Arisoy, Tomas Mikolov  
**Room:** Conference 6, Time 14:00 – 16:00, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Time</th>
<th>Title</th>
<th>Authors/Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2366</td>
<td>14:00</td>
<td>Multiscale Recurrent Neural Network Based Language Model</td>
<td>Tsuyoshi Morioka¹, Tomoharu Iwata², Takaaki Hori², Tetsunori Kobayashi¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹Waseda University, Japan; ²NTT Corporation, Japan</td>
</tr>
<tr>
<td>2371</td>
<td>14:20</td>
<td>Bag-of-Words Input for Long History Representation in Neural Network-Based Language Models for Speech Recognition</td>
<td>Kazuki Irie, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany</td>
</tr>
<tr>
<td>2376</td>
<td>14:40</td>
<td>Efficient Machine Translation Decoding with Slow Language Models</td>
<td>Ahmad Emami, IBM T.J. Watson Research Center, USA</td>
</tr>
<tr>
<td>2380</td>
<td>15:00</td>
<td>Latent Words Recurrent Neural Network Language Models</td>
<td>Ryo Masumura¹, Taiichi Asami¹, Takanobu Oba¹, Hirokazu Masataki¹, Sumitaka Sakauchi¹, Akinori Ito²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹NTT Corporation, Japan; ²Tohoku University, Japan</td>
</tr>
<tr>
<td>2385</td>
<td>15:20</td>
<td>Combining Multiple-Type Input Units Using Recurrent Neural Network for LVCSR Language Modeling</td>
<td>Vataya Chunwijitra, Ananlada Chotimongkol, Chai Wutiwiwatchai, NECTEC, Thailand</td>
</tr>
<tr>
<td>2390</td>
<td>15:40</td>
<td>Prosodically-Enhanced Recurrent Neural Network Language Models</td>
<td>Siva Reddy Gangireddy¹, Steve Renals¹, Yoshihiko Nankaku², Akinobu Lee²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹University of Edinburgh, UK; ²Nagoya Institute of Technology, Japan</td>
</tr>
</tbody>
</table>
**Special Session 6: Biosignal-based Spoken Communication**

*Chairs: Matthias Janke, Michael Wand*

**Room: Hall 1, Time 14:00 – 16:00, Wednesday, September 9, 2015**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:05</td>
<td>Welcome and Introduction</td>
<td>Matthias Janke(^1), Michael Wand(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)Universität Bremen, Germany; (^2)IDSIA, Switzerland</td>
</tr>
<tr>
<td>14:05 – 14:20</td>
<td>A Comprehensive 3D Biomechanically-Driven Vocal Tract Model Including</td>
<td>Peter Anderson(^1), Negar M. Harandi(^1), Scott Moisik(^2), Ian Stavness(^3), Sidney Fels(^1)</td>
</tr>
<tr>
<td></td>
<td>Inverse Dynamics for Speech Research</td>
<td>(^1)University of British Columbia, Canada; (^2)MPI for Psycholinguistics, The Netherlands; (^3)University of Saskatchewan, Canada</td>
</tr>
<tr>
<td>14:20 – 14:35</td>
<td>Low Frequency Ultrasonic Voice Activity Detection Using Convolutional</td>
<td>Ian McLoughlin(^1), Yan Song(^2)</td>
</tr>
<tr>
<td></td>
<td>Neural Networks</td>
<td>(^1)University of Kent, UK; (^2)USTC, China</td>
</tr>
<tr>
<td>14:35 – 14:50</td>
<td>Real-Time Control of a DNN-Based Articulatory Synthesizer for Silent</td>
<td>Florent Bocquelet(^1), Thomas Hueber(^2), Laurent Girin(^2), Christophe Savariaux(^2),</td>
</tr>
<tr>
<td></td>
<td>Speech Conversion: A Pilot Study</td>
<td>Blaise Yvert(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)Clinatec, France; (^2)GIPSA, France</td>
</tr>
<tr>
<td>14:50 – 15:05</td>
<td>Tongue Tracking in Ultrasound Images Using EigenTongue Decomposition</td>
<td>Diandra Fabre, Thomas Hueber, Florent Bocquelet, Pierre Badin</td>
</tr>
<tr>
<td></td>
<td>and Artificial Neural Networks</td>
<td>(^1)GIPSA, France</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>INTERSPEECH 2015 Table of Contents</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Wed-SP6 continued...</strong></td>
</tr>
<tr>
<td>15:05 – 15:20</td>
<td>Speaker-Independent Silent Speech Recognition with Across-Speaker</td>
<td>Jun Wang, Seongjun Hahm, University of Texas at Dallas, USA</td>
</tr>
<tr>
<td></td>
<td>Articulatory Normalization and Speaker Adaptive Training</td>
<td><strong>INTERSPEECH 2015 Table of Contents</strong></td>
</tr>
<tr>
<td>15:20 – 15:35</td>
<td>Codebook Clustering for Unit Selection Based EMG-to-Speech Conversion</td>
<td>Lorenz Diener, Matthias Janke, Tanja Schultz, KIT, Germany</td>
</tr>
<tr>
<td>15:50 – 16:00</td>
<td>Panel and Discussion</td>
<td>Matthias Janke(^1), Michael Wand(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)Universität Bremen, Germany; (^2)IDSIA, Switzerland</td>
</tr>
</tbody>
</table>
Poster Session 20: Robust Speech Recognition: Features, Far-field and Reverberation

Chair: Andrew Senior
Room: Hall 2, Time 14:00 – 16:00, Wednesday, September 9, 2015

A Study on Deep Neural Network Acoustic Model Adaptation for Robust Far-Field Speech Recognition
Seyedmahdad Mirmamadi, John H.L. Hansen, University of Texas at Dallas, USA

Speech Dereverberation Using Long Short-Term Memory
Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara, Kyoto University, Japan

Reverberation Robust Acoustic Modeling Using i-Vectors with Time Delay Neural Networks
Vijayaditya Peddinti, Guoguo Chen, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

Delta-MelSpectra Features for Noise Robustness to DNN-Based ASR Systems
Kshitiz Kumar, Chaojun Liu, Yifan Gong, Microsoft, USA

Combating Reverberation in Large Vocabulary Continuous Speech Recognition
Vikramjit Mitra, Julien Van Hout, Mitchell McLaren, Wen Wang, Martin Graciarena, Dimitra Vergyri, Horacio Franco, SRI International, USA

Three Ways to Adapt a CTS Recognizer to Unseen Reverberated Speech in BUT System for the ASPIRE Challenge
Martin Karafiát, František Grézl, Lukáš Burget, Igor Szöke, Jan Černocký, Brno University of Technology, Czech Republic

Robust Parameter Estimation for Audio Declipping in Noise
Mark J. Harvilla, Richard M. Stern, Carnegie Mellon University, USA

Multi-Task Learning Deep Neural Networks for Speech Feature Denoising
Bin Huang¹, Dengfeng Ke², Hao Zheng², Bo Xu², Yanyan Xu¹, Kaile Su³
¹Beijing Forestry University, China; ²Chinese Academy of Sciences, China; ³Griffith University, Australia

Time-Frequency Masking for Large Scale Robust Speech Recognition
Yuxuan Wang¹, Ananya Misra², Kean K. Chin²
¹Ohio State University, USA; ²Google, USA

Efficient Use of DNN Bottleneck Features in Generalized Variable Parameter HMMs for Noise Robust Speech Recognition
Rongfeng Su, Xurong Xie, Xanying Liu, Lan Wang, Chinese Academy of Sciences, China

Investigating Modulation Spectrogram Features for Deep Neural Network-Based Automatic Speech Recognition
Deepak Baby, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium

Deep Neural Network Based Spectral Feature Mapping for Robust Speech Recognition
Kun Han, Yanzhang He, Deblin Bagchi, Eric Fosler-Lussier, DeLiang Wang, Ohio State University, USA

INTERSPEECH 2015 Table of Contents
### Poster Session 21: Social Signals, Assessment and Paralinguistics

*Chairs: Laurence Devillers, Julia Hirschberg*

*Room: Hall 2, Time 14:00 - 16:00, Wednesday, September 9, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2489</td>
<td>Analyzing Speech Rate Entrainment and Its Relation to Therapist Empathy in Drug Addiction Counseling</td>
<td>Bo Xiao¹, Zac E. Imel², David C. Atkins³, Panayiotis G. Georgiou¹, Shrikanth S. Narayanan¹</td>
</tr>
<tr>
<td>2494</td>
<td>Agreement and Disagreement Utterance Detection in Conversational Speech by Extracting and Integrating Local Features</td>
<td>Atsushi Ando, Taichi Asami, Manabu Okamoto, Hirokazu Masataki, Sumitaka Sakauchi, NTT Corporation, Japan</td>
</tr>
<tr>
<td>2499</td>
<td>Still Together?: The Role of Acoustic Features in Predicting Marital Outcome</td>
<td>Md. Nasir¹, Wei Xia¹, Bo Xiao¹, Brian Baucom², Shrikanth S. Narayanan¹, Panayiotis G. Georgiou¹</td>
</tr>
<tr>
<td>2504</td>
<td>On Evaluation Metrics for Social Signal Detection</td>
<td>Gábor Gosztolya, MTA-SZTE RGAI, Hungary</td>
</tr>
<tr>
<td>2509</td>
<td>Laughter and Filler Detection in Naturalistic Audio</td>
<td>Lakshmish Kaushik, Abhijeet Sangwan, John H.L. Hansen, University of Texas at Dallas, USA</td>
</tr>
<tr>
<td>2514</td>
<td>Automatic Formatted Transcripts for Videos</td>
<td>Aasish Pappu, Amanda Stent, Yahoo!, USA</td>
</tr>
</tbody>
</table>

---

**INTERSPEECH 2015 Table of Contents continued...**

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2519</td>
<td>Does My Speech Rock? Automatic Assessment of Public Speaking Skills</td>
<td>Lucas Azaïs¹, Adrien Payan¹, Tianjiao Sun¹, Guillaume Vidal¹, Tina Zhang¹, Eduardo Coutinho¹, Florian Eyben², Björn Schuller¹</td>
</tr>
<tr>
<td>2524</td>
<td>Verbal Intelligence Identification Based on Text Classification</td>
<td>Roman Sergienko, Alexander Schmitt, Universität Ulm, Germany</td>
</tr>
<tr>
<td>2529</td>
<td>A Multimodal Approach for Automatic Assessment of School Principals’ Oral Presentation During Pre-Service Training Program</td>
<td>Shan-Wen Hsiao¹, Hung-Ching Sun¹, Ming-Chuan Hsieh², Ming-Hsueh Tsai², Hsin-Chih Lin², Chi-Chun Lee¹</td>
</tr>
<tr>
<td>2534</td>
<td>Are You TED Talk Material? Comparing Prosody in Professors and TED Speakers</td>
<td>T.J. Tsai, University of California at Berkeley, USA</td>
</tr>
<tr>
<td>2539</td>
<td>Detection of Cognitive States and Their Correlation to Speech Recognition Performance in Speech-to-Speech Machine Translation Systems</td>
<td>Hayakawa Akira, Fasih Haider, Loredana Cerrato, Nick Campbell, Saturnino Luz, Trinity College Dublin, Ireland</td>
</tr>
</tbody>
</table>
Poster Session 22: Bandwidth Extension, Quality and Intelligibility Measures

Chair: Tim Fingscheidt
Room: Foyer, Time 14:00 – 16:00, Wednesday, September 9, 2015

PAGE 2544
Wed-P-22-1
Perceptual Speech Quality Dimensions in a Conversational Situation
Friedemann Köster, Sebastian Möller, T-Labs, Germany

PAGE 2549
Wed-P-22-2
Multidimensional Evaluation and Predicting Overall Speech Quality
Jens Berger, Anna Llagostera, SwissQual, Switzerland

PAGE 2553
Wed-P-22-3
On Speech Intelligibility Estimation of Phase-Aware Single-Channel Speech Enhancement
Andreas Gaich, Pejman Mowlaee, Technische Universität Graz, Austria

PAGE 2558
Wed-P-22-4
A Framework for the Evaluation of Microscopic Intelligibility Models
Ricard Marxer¹, Martin Cooke², Jon Barker¹
¹University of Sheffield, UK; ²Ikerbasque, Spain

PAGE 2563
Wed-P-22-5
A Binaural Short Time Objective Intelligibility Measure for Noisy and Enhanced Speech
Asgar Heidemann Andersen¹, Jan Mark de Haan², Zheng-Hua Tan¹, Jesper Jensen¹
¹Aalborg University, Denmark; ²Oticon, Denmark

PAGE 2568
Wed-P-22-6
A Glimpse-Based Approach for Predicting Binaural Intelligibility with Single and Multiple Maskers in Anechoic Conditions
Yan Tang¹, Martin Cooke², Bruno M. Fazenda¹, Trevor J. Cox¹
¹University of Salford, UK; ²Ikerbasque, Spain

PAGE 2573
Wed-P-22-7
Improving the Prediction Power of the Speech Transmission Index to Account for Non-Linear Distortions Introduced by Noise-Reduction Algorithms
Fei Chen, SUSTC, China

PAGE 2578
Wed-P-22-8
DNN-Based Speech Bandwidth Expansion and Its Application to Adding High-Frequency Missing Features for Automatic Speech Recognition of Narrowband Speech
Kehuang Li¹, Zhen Huang¹, Yong Xu², Chin-Hui Lee¹
¹Georgia Institute of Technology, USA; ²USTC, China

PAGE 2583
Wed-P-22-9
Speech Quality Evaluation of Artificial Bandwidth Extension: Comparing Subjective Judgments and Instrumental Predictions
Hannu Palakka¹, Ville Myllylä¹, Anssi Rämö², Paavo Alku³
¹Microsoft, Finland; ²Nokia Research Center, Finland; ³Aalto University, Finland

PAGE 2588
Wed-P-22-10
Synchronous Overlap and Add of Spectra for Enhancement of Excitation in Artificial Bandwidth Extension of Speech
M.A. Tuğtekin Turan, Engin Erzin, Koç Universitesi, Turkey

PAGE 2593
Wed-P-22-11
Speech Bandwidth Expansion Based on Deep Neural Networks
Yingxue Wang¹, Shenghui Zhao¹, Wenbo Liu², Ming Li², Jingming Kuang¹
¹Beijing University of Technology, China; ²Sun Yat-sen University, China

PAGE 2598
Wed-P-22-12
A Novel Method of Artificial Bandwidth Extension Using Deep Architecture
Bin Liu, Jianhua Tao, Zhengqi Wen, Ya Li, Danish Bukhari, Chinese Academy of Sciences, China
Show and Tell Session 4

Chair: Björn Hoffmeister

Room: Seminar 1, Time 14:00 – 16:00, Wednesday, September 9, 2015

PAGE 2603
Show&Tell-4-1

The RedDots Platform for Mobile Crowd-Sourcing of Speech Data
Kong Aik Lee, Guangsen Wang, Kam Pheng Ng, Hanwu Sun, Trung Hieu Nguyen, Ngoc Thuy Huong Thai, Bin Ma, Haizhou Li, A*STAR, Singapore

PAGE 2605
Show&Tell-4-2

Two Extensions of Umeda and Teranishi’s Physical Models of the Human Vocal Tract
Takayuki Arai, Sophia University, Japan

PAGE 2607
Show&Tell-4-3

Collaborative Annotation for Person Identification in TV Shows
Matheuz Budnik¹, Laurent Besacier¹, Johann Poignant², Hervé Bredin², Claude Barras², Mickael Stefas³, Pierrick Bruneau³, Thomas Tamisier³
¹LIG (UMR 5217), France; ²LIMSI, France; ³LIST, Luxembourg

PAGE 2609
Show&Tell-4-4

Phonetic/Linguistic Web Services at BAS
Thomas Kisler, Florian Schiel, Uwe D. Reichel, Christoph Draxler, LMU München, Germany

PAGE 2611
Show&Tell-4-5

Managing Speech Databases with emuR and the EMU-webApp
Raphael Winkelmann, LMU München, Germany

PAGE 2613
Show&Tell-4-6

Visual Comparison of Speaker Groups
Sebastian Wankerl, Florian Hönig, Anton Batliner, J.R. Orozco-Arroyave, Elmar Nöth, FAU Erlangen-Nürnberg, Germany

PAGE 2615
Show&Tell-4-7

Tools for Rapid Customization of S2S Systems for Emergent Domains
Rohit Kumar, Matthew E. Roy, Sanjika Hewavitharana, Dennis N. Mehay, Nina Zinovieva, Raytheon BBN Technologies, USA

PAGE 2617
Show&Tell-4-8

The Speech Recognition Virtual Kitchen Turns One
Florian Metze¹, Eric Riebling¹, Eric Fosler-Lussier², Andrew Plummer², Rebecca Bates³
¹Carnegie Mellon University, USA; ²Ohio State University, USA; ³Minnesota State University, USA

PAGE 2619
Show&Tell-4-9

Model-Based Adaptive Pre-Processing of Speech for Enhanced Intelligibility in Noise and Reverberation
Jan Rennies¹, Andreas Volgenandt¹, Henning Schepker², Simon Doclo¹
¹Fraunhofer IDMT, Germany; ²Carl von Ossietzky Universität Oldenburg, Germany

PAGE 2621
Show&Tell-4-10

Experiences with and New Application Ideas for the Interspeech App
Sebastian Möller, Tilo Westermann, T-Labs, Germany

PAGE 2623
Show&Tell-4-11

Traditional IVR and Visual IVR — Killing Two Birds with One Stone
Dmitry Sityaev, Praphul Kumar, Rajesh Ramchander, Genesys, USA
## Oral Session 35: Discriminative Acoustic Training Methods for ASR

*Chairs: Steve Renals, George Saon*

*Room: Large Hall, Time 16:30 – 18:30, Wednesday, September 9, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Wed-O-35-1</th>
<th>16:30 – 16:50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annotating Large Lattices with the Exact Word Error</td>
<td>Rogier C. van Dalen, Mark J.F. Gales, University of Cambridge, UK</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-35-2</td>
<td>16:50 – 17:10</td>
</tr>
<tr>
<td></td>
<td>Semi-Supervised Maximum Mutual Information Training of Deep Neural Network Acoustic Models</td>
<td>Vimal Manohar, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-35-3</td>
<td>17:10 – 17:30</td>
</tr>
<tr>
<td></td>
<td>Rectified Linear Neural Networks with Tied-Scalar Regularization for LVCSR</td>
<td>Shiliang Zhang¹, Hui Jiang², Si Wei¹, Li-Rong Dai¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹USTC, China; ²York University, Canada</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-35-4</td>
<td>17:30 – 17:50</td>
</tr>
<tr>
<td></td>
<td>Segmental Conditional Random Fields with Deep Neural Networks as Acoustic Models for First-Pass Word Recognition</td>
<td>Yanzhang He, Eric Fosler-Lussier, Ohio State University, USA</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-35-5</td>
<td>17:50 – 18:10</td>
</tr>
<tr>
<td></td>
<td>Distinct Triphone Acoustic Modeling Using Deep Neural Networks</td>
<td>Dongpeng Chen, Brian Mak, HKUST, China</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-35-6</td>
<td>18:10 – 18:30</td>
</tr>
<tr>
<td></td>
<td>Minimum Word Error Training of RNN-Based Voice Activity Detection</td>
<td>Gregory Gelly, Jean-Luc Gauvain, LIMSI, France</td>
</tr>
</tbody>
</table>

## Oral Session 36: Syllables and Segments

*Chairs: Anders Eriksson, Frank Zimmerer*

*Room: Conference 1, Time 16:30 – 18:30, Wednesday, September 9, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Wed-O-36-1</th>
<th>16:30 – 16:50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classification of Place-of-Articulation of Stop Consonants Using Temporal Analysis</td>
<td>A.P. Prathosh¹, A.G. Ramakrishnan², T.V. Ananthapadmanabha³</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Xerox Research Center India, India; ²Indian Institute of Science, India; ³Voice and Speech Systems, India</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-36-2</td>
<td>16:50 – 17:10</td>
</tr>
<tr>
<td></td>
<td>The Emergence of Nasal Velar Codas in Brazilian Portuguese: An rt-MRI Study</td>
<td>Marissa Barlaz, Maojing Fu, Zhi-Pei Liang, Ryan Shosted, Brad Sutton, University of Illinois at Urbana-Champaign, USA</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-36-3</td>
<td>17:10 – 17:30</td>
</tr>
<tr>
<td></td>
<td>Salient Dimensions in Implicit Phonotactic Learning</td>
<td>Elise Michon¹, Emmanuel Dupoux², Alejandra Cristina²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Boğaziçi Üniversitesi, Turkey; ²LSCP, France</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-36-4</td>
<td>17:30 – 17:50</td>
</tr>
<tr>
<td></td>
<td>An Acoustic Examination of the Three-Way Sibilant Contrast in Lower Sorbian</td>
<td>Phil Howson, University of Toronto, Canada</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-36-5</td>
<td>17:50 – 18:10</td>
</tr>
<tr>
<td></td>
<td>Investigating Consonant Reduction in Mandarin Chinese with Improved Forced Alignment</td>
<td>Jiahong Yuan, Mark Liberman, University of Pennsylvania, USA</td>
</tr>
<tr>
<td>Page</td>
<td>Wed-O-36-6</td>
<td>18:10 – 18:30</td>
</tr>
<tr>
<td></td>
<td>Durational Characteristics and Timing Patterns of Russian Onset Clusters at Two Speaking Rates</td>
<td>Marianne Pouplier¹, Stefania Marin¹, Alexei Kochetov²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹LMU München, Germany; ²University of Toronto, Canada</td>
</tr>
</tbody>
</table>
Oral Session 37: Topics in Paralinguistics

Chairs: Shrikanth Narayanan, Khiet Truong
Room: Conference 2+3, Time 16:30 – 18:30, Wednesday, September 9, 2015

---

**Vocal Biomarkers to Discriminate Cognitive Load in a Working Memory Task**
Thomas F. Quatieri¹, James R. Williamson¹, Christopher J. Small¹, Tejash Patel¹, Joseph Perricone¹, Daryush D. Mehta¹, Brian S. Helfer¹, Gregory Ciccarelli¹, Darrell Ricke¹, Nicolas Malyska¹, Jeff Palmer¹, Kristin Heaton², Marianna Eddy³, Joseph Moran³
¹MIT Lincoln Laboratory, USA; ²USARIEM, USA; ³NSRDEC, USA

---

**I-Vector Based Physical Task Stress Detection with Different Fusion Strategies**
Chunlei Zhang, Gang Liu, Chengzhu Yu, John H.L. Hansen, University of Texas at Dallas, USA

---

**Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech Using ASR**
László Tóth¹, Gábor Gosztolya¹, Veronika Vince¹, Ildikó Hoffmann², Gréta Szatlóczki², Edit Bíró², Fruzsina Zsura², Magdolna Pákáski², János Kálmán²
¹MTA-SZTE RGAI, Hungary; ²University of Szeged, Hungary

---

**Contemporary Stochastic Feature Selection Algorithms for Speech-Based Emotion Recognition**
Maxim Sidorov¹, Christina Brester², Alexander Schmitt¹
¹Universität Ulm, Germany; ²Siberian State Aerospace University, Russia

---

**Effect of Different Jitter-Induced Glottal Pulse Shape Changes in Periodicity Perturbation Measures**
Carlos A. Ferrer¹, Diana Torres¹, Eduardo González¹, José Ramón Calvo², Eduardo Castillo³
¹Universidad Central de Las Villas, Cuba; ²CENATAV, Cuba; ³University of New Brunswick, Canada

---

**Automatic Audio Sentiment Extraction Using Keyword Spotting**
Lakshmish Kaushik, Abhijeet Sangwan, John H.L. Hansen, University of Texas at Dallas, USA
## Oral Session 38: Spoken Language Processing

**Chairs:** Giuseppe Riccardi, Ingrid Zukerman  
**Room:** Conference 4+5, Time 16:30 – 18:30, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-1</th>
<th>16:30 – 16:50</th>
</tr>
</thead>
</table>
| 2714  | Unsupervised Relation Detection Using Automatic Alignment of Query Patterns Extracted from Knowledge Graphs and Query Click Logs  
Panupong Pasupat¹, Dilek Hakkani-Tür²  
¹Stanford University, USA; ²Microsoft, USA |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-2</th>
<th>16:50 – 17:10</th>
</tr>
</thead>
</table>
| 2719  | A Latent Variable Model for Joint Pause Prediction and Dependency Parsing  
The Tung Nguyen, Graham Neubig, Hiroyuki Shindo, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura, NAIST, Japan |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-3</th>
<th>17:10 – 17:30</th>
</tr>
</thead>
</table>
| 2724  | Extractive Meeting Summarization Through Speaker Zone Detection  
Mohammad Hadi Bokaei¹, Hossein Sameti¹, Yang Liu²  
¹Sharif University of Technology, Iran; ²University of Texas at Dallas, USA |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-4</th>
<th>17:30 – 17:50</th>
</tr>
</thead>
</table>
| 2729  | Positional Language Modeling for Extractive Broadcast News Speech Summarization  
Shih-Hung Liu¹, Kuan-Yu Chen¹, Berlin Chen², Hsin-Min Wang¹, Hsu-Chun Yen³, Wen-Lian Hsu¹  
¹Academia Sinica, Taiwan; ²National Taiwan Normal University, Taiwan; ³National Taiwan University, Taiwan |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-5</th>
<th>17:50 – 18:10</th>
</tr>
</thead>
</table>
| 2734  | Speech-Based Location Estimation of First Responders in a Simulated Search and Rescue Scenario  
Saeid Mokaram, Roger K. Moore, University of Sheffield, UK |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 38-6</th>
<th>18:10 – 18:30</th>
</tr>
</thead>
</table>
| 2739  | Constructive Feedback, Thinking Process and Cooperation: Assessing the Quality of Classroom Interaction  
Tahir Sousa¹, Lucie Flekova², Margot Mieskes³, Iryna Gurevych²  
¹University of Minnesota, USA; ²Technische Universität Darmstadt, Germany; ³Hochschule Darmstadt, Germany |

## Oral Session 39: Voice Conversion

**Chairs:** Masami Akamine, Yannis Stylianou  
**Room:** Conference 6, Time 16:30 – 18:30, Wednesday, September 9, 2015

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-1</th>
<th>16:30 – 16:50</th>
</tr>
</thead>
</table>
| 2744  | A Real-Time Variable-Q Non-Stationary Gabor Transform for Pitch Shifting  
Dong-Yan Huang, Minghui Dong, Haizhou Li, A*STAR, Singapore |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-2</th>
<th>16:50 – 17:10</th>
</tr>
</thead>
</table>
| 2749  | Many-to-Many Voice Conversion Based on Multiple Non-Negative Matrix Factorization  
Ryo Aihara, Testuya Takiguchi, Yasuo Arikii, Kobe University, Japan |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-3</th>
<th>17:10 – 17:30</th>
</tr>
</thead>
</table>
| 2754  | Statistical Singing Voice Conversion Based on Direct Waveform Modification with Global Variance  
Kazuhiro Kobayashi, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-4</th>
<th>17:30 – 17:50</th>
</tr>
</thead>
</table>
| 2759  | System Fusion for High-Performance Voice Conversion  
Xiaoai Tian¹, Zhizheng Wu², Siu Wa Lee³, Quy Hy Nguyen¹, Minghui Dong³, Eng Siong Chng¹  
¹NTU, Singapore; ²University of Edinburgh, UK; ³A*STAR, Singapore |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-5</th>
<th>17:50 – 18:10</th>
</tr>
</thead>
</table>
| 2764  | Speaker Adaptation Using Only Vocalic Segments via Frequency Warping  
Agustin Alonso, D. Erro, Eva Navas, Inma Hernaez, Universidad del País Vasco, Spain |

<table>
<thead>
<tr>
<th>Page</th>
<th>Oral Session 39-6</th>
<th>18:10 – 18:30</th>
</tr>
</thead>
</table>
| 2769  | Non-Audible Murmur Enhancement Based on Statistical Conversion Using Air- and Body-Conductive Microphones in Noisy Environments  
Yusuke Tajiiri, Kou Tanaka, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan |
### Special Session 7: Advanced Crowdsourcing for Speech and Beyond

**Chairs:** Gina-Anne Levow, Tim Polzehl  
**Room:** Hall 1, Time 16:30 – 18:30, Wednesday, September 9, 2015

**Introduction by the Organizers**  
**Tim Polzehl**¹, Gina-Anne Levow²  
¹T-Labs, Germany; ²University of Washington, USA

<table>
<thead>
<tr>
<th>Oral Presentations</th>
</tr>
</thead>
</table>
| **Transcribing Continuous Speech Using Mismatched Crowdsourcing**  
Preethi Jyothi, Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA |
| **Selection and Aggregation Techniques for Crowdsourced Semantic Annotation Task**  
Shammur Absar Chowdhury¹, Marcos Calvo², Arindam Ghosh¹, Evgeny A. Stepanov¹, Ali Orkan Bayer¹, Giuseppe Riccardi¹, Fernando García², Emilio Sanchis²  
¹Università di Trento, Italy; ²Universidad Politécnica de Valencia, Spain |
| **Controlling Quality and Handling Fraud in Large Scale Crowdsourcing Speech Data Collections**  
Spencer Rothwell, Ahmad Elshenawy, Steele Carter, Daniela Braga, Faraz Romani, Michael Kennewick, Bob Kennewick, VoiceBox Technologies, USA |

### Poster Presentations

**Robustness in Speech Quality Assessment and Temporal Training Expiry in Mobile Crowdsourcing Environments**  
Tim Polzehl, Babak Naderi, Friedemann Köster, Sebastian Möller, T-Labs, Germany

**Effect of Trapping Questions on the Reliability of Speech Quality Judgments in a Crowdsourcing Paradigm (Poster)**  
Babak Naderi, Tim Polzehl, Ina Wechsung, Friedemann Köster, Sebastian Möller, T-Labs, Germany

**Voice Äpp: A Mobile App for Crowdsourcing Swiss German Dialect Data (Poster)**  
Adrian Leemann¹, Marie-José Kolly², Jean-Philippe Goldman³, Volker Dellwo⁴, Ingrid Hove⁴, Ibrahim Almajai⁵, Sarah Grimm⁵, Sylvain Robert⁵, Daniel Wanitsch⁶  
¹University of Cambridge, UK; ²LIMSI, France; ³Université de Genève, Switzerland; ⁴Universität Zürich, Switzerland; ⁵ETH Zürich, Switzerland; ⁶iBros.ch, Switzerland

**Expert and Crowdsourced Annotation of Pronunciation Errors for Automatic Scoring Systems (Poster)**  
Anastassia Loukina, Melissa Lopez, Keelan Evanini, David Suendermann-Oeft, Klaus Zechner, Educational Testing Service, USA

**CapCap: An Output-Agreement Game for Video Captioning (Poster)**  
Hernisa Kacorri¹, Kaoru Shinkawa², Shin Saito²  
¹CUNY Graduate Center, USA; ²IBM Research Tokyo, Japan

**Auris populi: Crowdsourced Native Transcriptions of Dutch Vowels Spoken by Adult Spanish Learners (Poster)**  
Pepi Burgos, Eric Sanders, Catia Cucchiarini, Roeland van Hout, Helmer Strik, Radboud Universiteit Nijmegen, The Netherlands
Crowdsource a Little to Label a Lot: Labeling a Speech Corpus of Dialectal Arabic (Poster)
Samantha Wray, Ahmed Ali, Hamad Bin Khalifa University, Qatar

Using Keyword Spotting to Help Humans Correct Captioning Faster (Poster)
Yashesh Gaur, Florian Metze, Yajie Miao, Jeffrey P. Bigham, Carnegie Mellon University, USA

Validating and Optimizing a Crowdsourced Method for Gradient Measures of Child Speech (Poster)
Tara McAllister Byun¹, Elaine Hitchcock², Daphna Harel¹
¹New York University, USA; ²Montclair State University, USA

Poster Session 23: Robust Speech Recognition: Adaptation
Chair: Najim Dehak
Room: Hall 2, Time 16:30 – 18:30, Wednesday, September 9, 2015

Joint Training of Speech Separation, Filterbank and Acoustic Model for Robust Automatic Speech Recognition
Zhong-Qiu Wang, DeLiang Wang, Ohio State University, USA

Joint Environment and Speaker Normalization Using Factored Front-End CMLLR
Shakti Rath, Sunil Sivadas, Bin Ma, A*STAR, Singapore

Robust Speech Recognition Using DNN-HMM Acoustic Model Combining Noise-Aware Training with Spectral Subtraction
Akihiro Abe, Kazumasa Yamamoto, Seiichi Nakagawa, Toyohashi University of Technology, Japan

Robust i-Vector Extraction for Neural Network Adaptation in Noisy Environment
Chengzhu Yu¹, Atsunori Ogawa², Marc Delcroix², Takuya Yoshioka², Tomohiro Nakatani², John H.L. Hansen¹
¹University of Texas at Dallas, USA; ²NTT Corporation, Japan

Spectrally Selective Dithering for Distorted Speech Recognition
Michal Borsky, Petr Mizera, Petr Pollak, Czech Technical University in Prague, Czech Republic

Feature-Space Speaker Adaptation for Probabilistic Linear Discriminant Analysis Acoustic Models
Liang Lu, Steve Renals, University of Edinburgh, UK

Speaker Adaptation Using the I-Vector Technique for Bottleneck Features
Patrick Cardinal, Najim Dehak, Yu Zhang, James Glass, MIT, USA
I-Vector Estimation Using Informative Priors for Adaptation of Deep Neural Networks
Penny Karanasou, Mark J.F. Gales, Philip C. Woodland, University of Cambridge, UK

Robust i-Vector Based Adaptation of DNN Acoustic Model for Speech Recognition
Sri Garimella¹, Arindam Mandal², Nikko Strom², Bjorn Hoffmeister², Spyros Matsoukas², Sree Hari Krishnan Parthasarathi²
¹Amazon.com, India; ²Amazon.com, USA

GMM-Derived Features for Effective Unsupervised Adaptation of Deep Neural Network Acoustic Models
Natalia Tomashenko¹, Yuri Khokhlov²
¹Speech Technology Center, Russia; ²STC-innovations, Russia

Unsupervised Adaptation for Deep Neural Network Using Linear Least Square Method
Roger Hsiao, Tim Ng, Stavros Tsakalidis, Long Nguyen, Richard Schwartz, Raytheon BBN Technologies, USA

Ensemble Speaker Modeling Using Speaker Adaptive Training Deep Neural Network for Speaker Adaptation
Sheng Li¹, Xugang Lu², Yuya Akita¹, Tatsuya Kawahara¹
¹Kyoto University, Japan; ²NICT, Japan

Data-Selective Transfer Learning for Multi-Domain Speech Recognition
Mortaza Doulaty, Oscar Saz, Thomas Hain, University of Sheffield, UK

Automatic Detection of Equipment Alarms in a Neonatal Intensive Care Unit Environment: A Knowledge-Based Approach
Ganna Raboshchuk¹, Peter Jančovič², Climent Nadeu¹, Alex Peiró Lilja¹, Műnöver Kőkäer², Blanca Muñoz Mahamud³, Ana Riverola de Veciana¹
¹Universitat Politècnica de Catalunya, Spain; ²University of Birmingham, UK; ³Hospital Sant Joan de Déu Barcelona, Spain

"Multilingual" Deep Neural Network for Music Genre Classification
Jia Dai¹, Wenju Liu¹, Chongjia Ni², Like Dong³, Hong Yang³
¹Chinese Academy of Sciences, China; ²SDUFE, China; ³SGCC, China

Accurate Endpointing with Expected Pause Duration
Baiyang Liu, Bjorn Hoffmeister, Ariya Rastrow, Amazon.com, USA

Locality Constrained Transitive Distance Clustering on Speech Data
Wenbo Liu¹, Zhiding Yu², Bhiksha Raj², Ming Li¹
¹Sun Yat-sen University, China; ²Carnegie Mellon University, USA

Feature Extraction Strategies in Deep Learning Based Acoustic Event Detection
Miquel Espi, Masakiyo Fujimoto, Keisuke Kinoshita, Tomohiro Nakatani, NTT Corporation, Japan

An Acoustic Event Detection Framework and Evaluation Metric for Surveillance in Cars
Peter Transfeld, Simon Receveur, Tim Fingscheidt, Technische Universität Braunschweig, Germany
Diachronic Semantic Cohesion for Topic Segmentation of TV Broadcast News  
Abdessalam Bouchekif¹, Géraldine Damnati¹, Yannick Estève², Delphine Charlet¹, Nathalie Camelin²  
¹Orange Labs, France; ²LIUM, France

Comparison of Forced-Alignment Speech Recognition and Humans for Generating Reference VAD  
Ivan Kraljevski¹, Zheng-Hua Tan², Maria Paola Bissiri³  
¹voice INTER connect, Germany; ²Aalborg University, Denmark; ³Queen Margaret University, UK

Improving Voice Activity Detection in Movies  
Bernhard Lehner, Gerhard Widmer, Reinhard Sonnleitner, Johannes Kepler  
Universität Linz, Austria

Language-Independent Method for Analysis of German Stuttering Recordings  
Tomas Lustyk¹, Petr Bergl¹, Tino Haderlein², Elmar Nöth², Roman Cmejla¹  
¹Czech Technical University in Prague, Czech Republic; ²FAU Erlangen-Nürnberg, Germany

An Investigation of MDVP Parameters for Voice Pathology Detection on Three Different Databases  
Ahmed Al-nasheri, Zulfiquar Ali, Ghulam Muhammad, Mansour Alsulaiman, King Saud University, Saudi Arabia

Energy Distribution Analysis and Nonlinear Dynamical Analysis of Adductor Spasmodic Dysphonia  
Jiantao Wu¹, Ping Yu², Nan Yan¹, Lan Wang¹, Xiaohui Yang³, Manwa L. Ng⁴  
¹Chinese Academy of Sciences, China; ²PLA General Hospital, China; ³CUST, China; ⁴University of Hong Kong, China

Auditory-Visual Tone Perception in Hearing Impaired Thai Listeners  
Benjawan Kasisopa¹, Nittayapa Klangpornkun², Denis Burnham¹  
¹University of Western Sydney, Australia; ²Thammasat University, Thailand

Speech Intelligibility Decline in Individuals with Fast and Slow Rates of ALS Progression  
Panying Rong¹, Yana Yunusova², Jordan R. Green¹  
¹MGH Institute of Health Professions, USA; ²University of Toronto, Canada
Latency Analysis of Speech Shadowing Reveals Processing Differences in Japanese Adults Who Do and Do Not Stutter
Rong Na A, Koichi Mori, Naomi Sakai, NRCD Research Institute, Japan

A Syllable-Based Analysis of Speech Temporal Organization: A Comparison Between Speaking Styles in Dysarthric and Healthy Populations
Brigitte Bigi1, Katarzyna Klessa2, Laurianne Georgeton1, Christine Meunier1
1LPL, France; 2Adam Mickiewicz University in Poznań, Poland

Autonomous Measurement of Speech Intelligibility Utilizing Automatic Speech Recognition
Bernd T. Meyer, Birger Kollmeier, Jasper Ooster, Carl von Ossietzky Universität Oldenburg, Germany

Can You Hear Me? Acoustic Modifications in Speech Directed to Foreigners and Hearing-Impaired People
Monja Angelika Knoll, Melissa Johnstone, Charlene Blakely, University of the West of Scotland, UK

Improving Automatic Forced Alignment for Dysarthric Speech Transcription
Yu Ting Yeung, Ka Ho Wong, Helen Meng, Chinese University of Hong Kong, China

The RedDots Data Collection for Speaker Recognition
Kong Aik Lee1, Anthony Larcher2, Guangsen Wang1, Patrick Kenny3, Niko Brümmer4, David van Leeuwen5, Hagai Aronowitz6, Marcel Kockmann7, Carlos Vaquero8, Bin Ma1, Haizhou Li1, Themos Stafylakis3, Md. Jahangir Alam3, Albert Swart4, Javier Perez6
1A*STAR, Singapore; 2LIUM, France; 3CRIM, Canada; 4AGNITIO, South Africa; 5NovoLanguage, The Netherlands; 6IBM Research Haifa, Israel; 7VoiceTrust, Canada; 8AGNITIO, Spain

Noise-Robust Speaker Recognition Based on Morphological Component Analysis
Yongjun He1, Chen Chen2, Jiqing Han2
1HUST, China; 2HIT, China

Analysis of Mutual Duration and Noise Effects in Speaker Recognition: Benefits of Condition-Matched Cohort Selection in Score Normalization
Andreas Nautsch1, Rahim Saeidi2, Christian Rathgeb1, Christoph Busch1
1Hochschule Darmstadt, Germany; 2Aalto University, Finland

Robustness to Additive Noise of Locally-Normalized Cepstral Coefficients in Speaker Verification
Josué Fredes1, José Novoa1, Víctor Poblete1, Simon King2, Richard M. Stern3, Néstor Becerra Yoma1
1Universidad de Chile, Chile; 2University of Edinburgh, UK; 3Carnegie Mellon University, USA

Probabilistic Linear Discriminant Analysis for Robust Speaker Identification in Co-Channel Speech
Navid Shokouhi, John H.L. Hansen, University of Texas at Dallas, USA
Community Detection with Manifold Learning on Speaker i-Vector Space for Chinese
Hongcui Wang¹, Di Jin¹, Lantian Li², Jianwu Dang¹
¹Tianjin University, China; ²Tsinghua University, China

A Comparison of Neural Network Feature Transforms for Speaker Diarization
Sree Harsha Yella¹, Andreas Stolcke²
¹Idiap Research Institute, Switzerland; ²Microsoft, USA

Clustering Short Push-to-Talk Segments
Ilya Shapiro, Neta Rabin, Irit Opher, Itshak Lapidot, Afeka, Israel

Exploring ANN Back-Ends for i-Vector Based Speaker Age Estimation
Anna Fedorova¹, Ondřej Glembek², Tomi Kinnunen¹, Pavel Matějka²
¹University of Eastern Finland, Finland; ²Brno University of Technology, Czech Republic

Analysis of the Second Phase of the 2013–2014 i-Vector Machine Learning Challenge
Désiré Bansé¹, George R. Doddington¹, Daniel Garcia-Romero², John J. Godfrey², Craig S. Greenberg¹, Jaime Hernández-Cordero³, John M. Howard¹, Alvin F. Martin¹, Lisa P. Mason¹, Alan McCree², Douglas A. Reynolds⁴
¹NIST, USA; ²Johns Hopkins University, USA; ³DOD, USA; ⁴MIT Lincoln Laboratory, USA

NIST Language Recognition Evaluation — Plans for 2015
Alvin F. Martin¹, Craig S. Greenberg¹, John M. Howard¹, Désiré Bansé¹, George R. Doddington¹, Jaime Hernández-Cordero², Lisa P. Mason²
¹NIST, USA; ²DOD, USA

Oral Session 40: Dialogue and Discourse
Chairs: Jens Edlund, Kornel Laskowski
Room: Conference 1, Time 09:00 – 11:00, Thursday, September 10, 2015

Communicative Needs and Respiratory Constraints
Marcin Włodarczak¹, Mattias Heldner¹, Jens Edlund²
¹Stockholm University, Sweden; ²KTH, Sweden

Analysis and Classification of Cooperative and Competitive Dialogs
Uwe D. Reichel¹, Nina Pörner¹, Dianne Nowack¹, Jennifer Cole²
¹LMU München, Germany; ²University of Illinois at Urbana-Champaign, USA

Towards Automatic Detection of Reported Speech in Dialogue Using Prosodic Cues
Alessandra Cervone, Catherine Lai, Silvia Pareti, Peter Bell, University of Edinburgh, UK

Modeling Phrasing and Prominence Using Deep Recurrent Learning
Andrew Rosenberg¹, Raul Fernandez², Bhuvana Ramabhadran²
¹CUNY Queens College, USA; ²IBM T.J. Watson Research Center, USA

Pitch Declination and Reset as a Function of Utterance Duration in Conversational Speech Data
Céline De Looze, Irena Yanushevskaia, Andy Murphy, Eoghan O’Connor, Christer Gobl, Trinity College Dublin, Ireland

Investigating the Role of ‘yeah’ in Stance-Dense Conversation
Valerie Freeman, Gina-Anne Levow, Richard Wright, Mari Ostendorf, University of Washington, USA
Oral Session 41: Speaker Diarization

Chairs: Jean-François Bonastre, Itshack Lapidot
Room: Conference 2+3, Time 09:00 – 11:00, Thursday, September 10, 2015

Page 3081
Thu-O-41-1
09:00 – 09:20
Factor Analysis for Speaker Segmentation and Improved Speaker Diarization
Brecht Desplanques, Kris Demuynck, Jean-Pierre Martens, Ghent University, Belgium

Page 3086
Thu-O-41-2
09:20 – 09:40
Enhanced Speaker Diarization with Detection of Backchannels Using Eye-Gaze Information in Poster Conversations
Koji Inoue¹, Yukoh Wakabayashi², Hiromasa Yoshimoto¹, Katsuya Takanashi¹, Tatsuya Kawahara¹
¹Kyoto University, Japan; ²Ritsumeikan University, Japan

Page 3091
Thu-O-41-3
09:40 – 10:00
Novel Clustering Selection Criterion for Fast Binary Key Speaker Diarization
Héctor Delgado¹, Xavier Anguera², Corinne Fredouille³, Javier Serrano¹
¹Universidad Autónoma de Barcelona, Spain; ²Sinkronigo, Spain; ³LIA, France

Page 3096
Thu-O-41-4
10:00 – 10:20
Speaker Diarization with I-Vectors from DNN Senone Posteriors
Gregory Sell, Daniel Garcia-Romero, Alan McCree, Johns Hopkins University, USA

Page 3098
Thu-O-41-5
10:20 – 10:40
Using Voice-Quality Measurements with Prosodic and Spectral Features for Speaker Diarization
Abraham Woubie¹, Jordi Luque², Javier Hernando¹
¹Universitat Politècnica de Catalunya, Spain; ²Telefónica I+D, Spain

Page 3103
Thu-O-41-6
10:40 – 11:00
Integrating Online I-Vector Extractor with Information Bottleneck Based Speaker Diarization System
Srikanth Madikeri, Ivan Himawan, Petr Motlicek, Marc Ferras, Idiap Research Institute, Switzerland

Oral Session 42: L1/L2 Speech Perception and Acquisition

Chairs: Gábor Pintér, Jürgen Trouvain
Room: Conference 4+5, Time 09:00 – 11:00, Thursday, September 10, 2015

Page 3110
Thu-O-42-1
09:00 – 09:20
Enhanced Processing of a Lost Language: Linguistic Knowledge or Linguistic Skill?
Jiyoun Choi¹, Mirjam Broersma², Anne Cutler²
¹Hanyang University, Korea; ²MPI for Psycholinguistics, The Netherlands

Page 3115
Thu-O-42-2
09:20 – 09:40
Production Inconsistencies Delay Adaptation to Foreign Accents
Ann-Kathrin Grohe¹, Gregory J. Poarch², Adriana Hanulíková³, Andrea Weber¹
¹Eberhard Karls Universität Tübingen, Germany; ²Westfälische Wilhelms-Universität Münster, Germany; ³Universität Freiburg, Germany

Page 3120
Thu-O-42-3
09:40 – 10:00
Acquisition of English Speech Rhythm by Monolingual Children
Mikhail Ordin, Leona Polyanskaya, Universität Bielefeld, Germany

Page 3125
Thu-O-42-4
10:00 – 10:20
Durational Information in Word-Initial Lexical Embeddings in Spoken Dutch
Odette Scharenborg, Radboud Universiteit Nijmegen, The Netherlands

Page 3130
Thu-O-42-5
10:20 – 10:40
The Development of Categorical Perception of Lexical Tones in Mandarin-Speaking Preschoolers
Fei Chen, Nan Yan, Lan Wang, Tao Yang, Jian Tao Wu, Han Zhao, Gang Peng, Chinese Academy of Sciences, China

Page 3135
Thu-O-42-6
10:40 – 11:00
Perception of Italian Liquids by Japanese Listeners: Comparisons to Spanish Liquids
Tomohiko Ooigawa, Sophia University, Japan
Oral Session 43: LVCSR Systems and Applications

Chairs: Michel Bacchiani, Florian Metze
Room: Conference 6, Time 09:00 - 11:00, Thursday, September 10, 2015

The IBM 2015 English Conversational Telephone Speech Recognition System
George Saon, Hong-Kwang J. Kuo, Steven Rennie, Michael Picheny, IBM T.J. Watson Research Center, USA

The Cambridge University 2014 BOLT Conversational Telephone Mandarin Chinese LVCSR System for Speech Translation
Xunying Liu, Federico Flego, Linlin Wang, C. Zhang, Mark J.F. Gales, Philip C. Woodland, University of Cambridge, UK

The IBM BOLT Speech Transcription System
Samuel Thomas, George Saon, Hong-Kwang J. Kuo, Lidia Mangu, IBM T.J. Watson Research Center, USA

Improvements in RWTH LVCSR Evaluation Systems for Polish, Portuguese, English, Urdu, and Arabic
M. Ali Basha Shaik, Zoltán Tüske, M. Ali Tahir, Markus Nußbaum-Thom, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

Active Learning Based Data Selection for Limited Resource STT and KWS
Thiago Fraga-Silva1, Jean-Luc Gauvain 2, Lori Lamel2, Antoine Laurent1, Viet-Bac Le1, Abdel Messaoudi1
1Vocapia Research, France; 2LIMSI, France

Improved Hindi Broadcast ASR by Adapting the Language Model and Pronunciation Model Using a priori Syntactic and Morphophonemic Knowledge
Preethi Jyothi, Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA

Special Session 8: Zero Resource Speech Technologies: Unsupervised Discovery of Linguistic Units

Chairs: Xavier Anguera, Emmanuel Dupoux, Maarten Versteegh
Room: Hall 1, Time 09:00 - 11:00, Thursday, September 10, 2015

The Zero Resource Speech Challenge 2015
Maarten Versteegh1, Roland Thiollière1, Thomas Schatz1, Xuan Nga Cao1, Xavier Anguera2, Aren Jansen3, Emmanuel Dupoux1
1ENS, France; 2Telefónica I+D, Spain; 3Johns Hopkins University, USA

Poster Flash Presentation
Xavier Anguera1, Emmanuel Dupoux2, Maarten Versteegh3
1Telefónica I+D, Spain; 2HESS, France; 3ENS, France

Panel Discussion: Summary of Results, Discussion of Evaluation Metrics, and the Next Challenge
Xavier Anguera1, Emmanuel Dupoux2, Maarten Versteegh3
1Telefónica I+D, Spain; 2HESS, France; 3ENS, France

Poster Presentations

Discovering Discrete Subword Units with Binarized Autoencoders and Hidden-Markov-Model Encoders (Poster)
Leonardo Badino1, Alessio Mereta1, Lorenzo Rosasco2
1Istituto Italiano di Tecnologia, Italy; 2Università di Genova, Italy

A Hybrid Dynamic Time Warping-Deep Neural Network Architecture for Unsupervised Acoustic Modeling (Poster)
Roland Thiollière, Ewan Dunbar, Gabriel Synnaeve, Maarten Versteegh, Emmanuel Dupoux, ENS, France
Automatic Segmentation and Clustering of Speech Using Sparse Coding and Metaheuristic Search (Poster)
Wiehan Agenbag, Thomas Niesler, Stellenbosch University, South Africa

Parallel Inference of Dirichlet Process Gaussian Mixture Models for Unsupervised Acoustic Modeling: A Feasibility Study (Poster)
Hongjie Chen¹, Cheung-Chi Leung², Lei Xie¹, Bin Ma², Haizhou Li²
¹Northwestern Polytechnical University, China; ²A*STAR, Singapore

Using Articulatory Features and Inferred Phonological Segments in Zero Resource Speech Processing (Poster)
Pallavi Baljekar, Sunayana Sitaram, Prasanna Kumar Muthukumar, Alan W. Black, Carnegie Mellon University, USA

A Comparison of Neural Network Methods for Unsupervised Representation Learning on the Zero Resource Speech Challenge (Poster)
Daniel Renshaw¹, Herman Kamper¹, Aren Jansen², Sharon Goldwater¹
¹University of Edinburgh, UK; ²Johns Hopkins University, USA

Unsupervised Word Discovery from Speech Using Automatic Segmentation into Syllable-Like Units (Poster)
Okko Räsänen¹, Gabriel Doyle², Michael C. Frank²
¹Aalto University, Finland; ²Stanford University, USA

An Evaluation of Graph Clustering Methods for Unsupervised Term Discovery (Poster)
Vince Lyzinski, Gregory Sell, Aren Jansen, Johns Hopkins University, USA

Poster Session 27: Neural Networks: Novel Architectures for LVCSR
Chair: Rogier Van Dalen
Room: Hall 2, Time 09:00 - 11:00, Thursday, September 10, 2015

A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts
Vijayaditya Peddinti, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA

Long Short-Term Memory Based Convolutional Recurrent Neural Networks for Large Vocabulary Speech Recognition
Xiangang Li, Xihong Wu, Peking University, China

Parameterised Sigmoid and ReLU Hidden Activation Functions for DNN Acoustic Modelling
C. Zhang, Philip C. Woodland, University of Cambridge, UK

Discriminative Template Learning in Group-Convolutional Networks for Invariant Speech Representations
Chiyuan Zhang, Stephen Voinea, Georgios Evangelopoulos, Lorenzo Rosasco, Tomaso Poggio, MIT, USA

Investigation of Parametric Rectified Linear Units for Noise Robust Speech Recognition
Sunil Sivadas¹, Zhenzhou Wu², Ma Bin¹
¹A*STAR, Singapore; ²McGill University, Canada

Multi-Softmax Deep Neural Network for Semi-Supervised Training
Hang Su¹, Haihua Xu²
¹ICSI, USA; ²NTU, Singapore
Thu-P-27 continued…

**Thu-P-27-7**

**A Multi-Region Deep Neural Network Model in Speech Recognition**  
Jia Cui, George Saon, Bhuvana Ramabhadran, Brian Kingsbury, IBM T.J. Watson Research Center, USA

**Thu-P-27-8**

**A Study of the Recurrent Neural Network Encoder-Decoder for Large Vocabulary Speech Recognition**  
Liang Lu¹, Xingxing Zhang¹, Kyunghyun Cho², Steve Renals¹  
¹University of Edinburgh, UK; ²Université de Montréal, Canada

**Thu-P-27-9**

**Gaussian Free Cluster Tree Construction Using Deep Neural Network**  
Linchen Zhu, Kevin Kilgour, Sebastian Stüker, Alex Waibel, KIT, Germany

**Thu-P-27-10**

**Very Deep Convolutional Neural Networks for LVCSR**  
Mengxià Bi, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China

**Thu-P-27-11**

**Transferring Knowledge from a RNN to a DNN**  
William Chan, Nan Rosemary Ke, Ian Lane, Carnegie Mellon University, USA

**Thu-P-27-12**

**SVD-Based Universal DNN Modeling for Multiple Scenarios**  
Changliang Liu¹, Jinyu Li², Yifan Gong²  
¹Microsoft, China; ²Microsoft, USA

**Thu-P-27-13**

**Speech Enhancement and Recognition Using Multi-Task Learning of Long Short-Term Memory Recurrent Neural Networks**  
Zhuo Chen, Shinji Watanabe, Hakan Erdogan, John R. Hershey, MERL, USA

---

**Poster Session 28: Speech and Music Analysis**

*Chair: Hideki Kawahara*  
*Room: Hall 2, Time 09:00 – 11:00, Thursday, September 10, 2015*

**Thu-P-28-1**

**Speaker-Dependent Multipitch Tracking Using Deep Neural Networks**  
Yuzhou Liu, DeLiang Wang, Ohio State University, USA

**Thu-P-28-2**

**An Error Correction Scheme for GCI Detection Algorithms Using Pitch Smoothness Criterion**  
Sujith P.¹, A.P. Prathosh², A.G. Ramakrishnan³, Prasanta Kumar Ghosh³  
¹Ittiam Systems, India; ²Xerox Research Center India, India; ³Indian Institute of Science, India

**Thu-P-28-3**

**Robust Pitch Estimation in Noisy Speech Using ZTW and Group Delay Function**  
RaviShankar Prasad, B. Yegnanarayana, IIIT Hyderabad, India

**Thu-P-28-4**

**Robust Localization of Single Sound Source Based on Phase Difference Regression**  
Zhaoqiong Huang, Ge Zhan, Dongwen Ying, Yonghong Yan, Chinese Academy of Sciences, China

**Thu-P-28-5**

**Frequency Map Selection Using a RBFN-Based Classifier in the MVDR Beamformer for Speaker Localization in Reverberant Rooms**  
Daniele Salvati, Carlo Drioli, Gian Luca Foresti, Università di Udine, Italy

**Thu-P-28-6**

**Exploiting Deep Neural Networks and Head Movements for Binaural Localisation of Multiple Speakers in Reverberant Conditions**  
Ning Ma¹, Guy J. Brown¹, Tobias May²  
¹University of Sheffield, UK; ²Technical University of Denmark, Denmark
Thu-P-28 continued…

Joint Optimization of Recurrent Networks Exploiting Source Auto-Regression for Source Separation
Shuai Nie\textsuperscript{1}, Wei Xue\textsuperscript{1}, Shan Liang\textsuperscript{1}, Xueliang Zhang\textsuperscript{2}, Wenju Liu\textsuperscript{1}, Liwei Qiao\textsuperscript{3}, Jianping Li\textsuperscript{3}
\textsuperscript{1}Chinese Academy of Sciences, China; \textsuperscript{2}Inner Mongolia University, China; \textsuperscript{3}SGCC, China

Real-Time Audio-to-Score Alignment of Singing Voice Based on Melody and Lyric Information
Rong Gong, Philippe Cuvillier, Nicolas Obin, Arshia Cont, IRCAM, France

Vocal Separation from Monaural Music Using Adaptive Auditory Filtering Based on Kernel Back-Fitting
Jun-Yong Lee, Hye-Seung Cho, Hyoung-Gook Kim, Kwangwoon University, Korea

A Two-Stage Singing Voice Separation Algorithm Using Spectro-Temporal Modulation Features
Frederick Z. Yen, Mao-Chang Huang, Tai-Shih Chi, National Chiao Tung University, Taiwan

Robust Sound Event Classification Using LBP-HOG Based Bag-of-Audio-Words Feature Representation
Hyungjun Lim, Myung Jong Kim, Hoirin Kim, KAIST, Korea

Poster Session 29: Speech Synthesis 3

Sequence-to-Sequence Neural Net Models for Grapheme-to-Phoneme Conversion
Kaisheng Yao, Geoffrey Zweig, Microsoft, USA

Knowledge versus Data in TTS: Evaluation of a Continuum of Synthesis Systems
Rosie Kay\textsuperscript{1}, Oliver Watts\textsuperscript{1}, Roberto Barra Chicote\textsuperscript{2}, Cassie Mayo\textsuperscript{1}
\textsuperscript{1}University of Edinburgh, UK; \textsuperscript{2}Universidad Politécnica de Madrid, Spain

Improving G2P from Wiktionary and Other (Web) Resources
Steffen Eger, Goethe-Universität Frankfurt am Main, Germany

BLSTM Neural Networks for Speech Driven Head Motion Synthesis
Chuang Ding, Pengcheng Zhu, Lei Xie, Northwestern Polytechnical University, China

Articulatory Controllable Speech Modification Based on Gaussian Mixture Models with Direct Waveform Modification Using Spectrum Differential
Patrick Lumban Tobing, Kazuhiro Kobayashi, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, NAIST, Japan

Reconstructing Intelligible Audio Speech from Visual Speech Features
Thomas Le Cornu, Ben Milner, University of East Anglia, UK

Universal Grapheme-Based Speech Synthesis
Sunayana Sitaram\textsuperscript{1}, Alok Parlikar\textsuperscript{1}, Gopala Krishna Anumanchipalli\textsuperscript{2}, Alan W. Black\textsuperscript{1}
\textsuperscript{1}Carnegie Mellon University, USA; \textsuperscript{2}University of California at San Francisco, USA
Thu-P-29 continued…

Artificial Personality and Disfluency
Mirjam Wester¹, Matthew Aylett¹, Marcus Tomalin², Rasmus Dall¹
¹University of Edinburgh, UK; ²University of Cambridge, UK

Comparison of Chironomic Stylization versus Statistical Modeling of Prosody for Expressive Speech Synthesis
Marc Evrard, Samuel Delalez, Christophe d’Alessandro, Albert Rilliard, LIMSI, France

A Multi-Layer F0 Model for Singing Voice Synthesis Using a B-Spline Representation with Intuitive Controls
Luc Ardaillon, Gilles Degottex, Axel Roebel, IRCAM, France

Creating Expressive Synthetic Voices by Unsupervised Clustering of Audiobooks
Igor Jauk¹, Antonio Bonafonte¹, Paula Lopez-Otero², Laura Docio-Fernandez²
¹Universitat Politècnica de Catalunya, Spain; ²Universidade de Vigo, Spain

Articulatory-Based Conversion of Foreign Accents with Deep Neural Networks
Sandesh Aryal, Ricardo Gutierrez-Osuna, Texas A&M University, USA

Oral Session 44: Speech and Cognition in Adverse Conditions
Chairs: Louis ten Bosch, Bernd Meyer
Room: Large Hall, Time 14:00 – 16:00, Thursday, September 10, 2015

Action Planning and Congruency Effect Between Articulation and Grasping
Mikko Tiainen, Lari Vainio, Kaisa Tiippana, Naeem Komeilipoor, Martti Vainio, University of Helsinki, Finland

Cognitive Workload and Vocabulary Sparseness: Theory and Practice
Ron M. Hecht¹, Aharon Bar-Hillel², Stas Tiomkin¹, Hadar Levi¹, Omer Tsimhoni³, Naftali Tishby¹
¹Hebrew University of Jerusalem, Israel; ²Microsoft, Israel; ³General Motors, Israel

Counting Competing Speakers in a Timeframe — Human versus Computer
Valentin Andrei, Horia Cucu, Andi Buzo, Corneliu Burileanu, Universitatea Politehnica din București, Romania

Segmental Contribution to the Intelligibility of Ideal Binary-Masked Sentences
Fei Chen¹, Alexander Siu Tai Kwok²
¹SUSTC, China; ²University of Hong Kong, China

Perception of an Existing and Non-Existing L2 English Phoneme Behind Noise by Japanese Native Speakers
Mako Ishida, Takayuki Arai, Sophia University, Japan

Viseme Comparison Based on Phonetic Cues for Varying Speech Accents
Chitralekha Bhat, Sunil Kopparapu, TCS Innovation Labs Mumbai, India
**Oral Session 45: Audio Signal Analysis and Representation**

*Chairs: Keisuke Kinoshita, Pejman Mowlaee*

*Room: Conference 1, Time 14:00 - 16:00, Thursday, September 10, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-1</th>
<th>14:00 – 14:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quantifying Difference in Vocalizations of Bird Populations</td>
<td>Colm O'Reilly, Nicola M. Marples, David J. Kelly, Naomi Harte, Trinity College Dublin, Ireland</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-2</th>
<th>14:20 – 14:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reverberation-Robust Acoustic Indoor Localization</td>
<td>Jae Choi, Jeunghun Kim, Shin Jae Kang, Nam Soo Kim, Seoul National University, Korea</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-3</th>
<th>14:40 – 15:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>An Alternating Optimization Approach for Phase Retrieval</td>
<td>Huaiqing Ming¹, Dong-Yan Huang², Lei Xie¹, Haizhou Li¹, Minghui Dong²</td>
</tr>
<tr>
<td></td>
<td>¹Northwestern Polytechnical University, China; ²A*STAR, Singapore</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-4</th>
<th>15:00 – 15:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learning to Estimate Reverberation Time in Noisy and Reverberant Rooms</td>
<td>Xiong Xiao¹, Shengkui Zhao², Xionghu Zhong³, Douglas L. Jones², Eng Siong Chng¹, Haizhou Li¹</td>
</tr>
<tr>
<td></td>
<td>¹TL@NTU, Singapore; ²Advanced Digital Sciences Center, Singapore; ³NTU, Singapore</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-5</th>
<th>15:20 – 15:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direction of Arrival Estimation Based on Reverberation Weighting and Noise Error Estimator</td>
<td>Cheng Pang, Jie Zhang, Hong Liu, Peking University, China</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-45-6</th>
<th>15:40 – 16:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Representing Nonspeech Audio Signals Through Speech Classification Models</td>
<td>Huy Phan, Lars Hertel, Marco Maass, Radoslaw Mazur, Alfred Mertins, Universität zu Lübeck, Germany</td>
</tr>
</tbody>
</table>

**Oral Session 46: Robustness in Speaker Recognition**

*Chairs: Joseph P. Campbell, Pietro Laface*

*Room: Conference 2+3, Time 14:00 - 16:00, Thursday, September 10, 2015*

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-1</th>
<th>14:00 – 14:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mitigating the Effects of Non-Stationary Unseen Noises on Language Recognition Performance</td>
<td>Luciana Ferrer¹, Mitchell McLaren², Aaron Lawson², Martin Graciarena²</td>
</tr>
<tr>
<td></td>
<td>¹Universidad de Buenos Aires, Argentina; ²SRI International, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-2</th>
<th>14:20 – 14:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>An Information Theory Based Data-Homogeneity Measure for Voice Comparison</td>
<td>Moez Ajili¹, Jean-François Bonastre¹, Solange Rossato², Juliette Kahn³, Ishak Lapidot⁴</td>
</tr>
<tr>
<td></td>
<td>¹LIA, France; ²LIG (UMR 5217), France; ³LNE, France; ⁴Afeka, Israel</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-3</th>
<th>14:40 – 15:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The QUT-NOISE-SRE Protocol for the Evaluation of Noisy Speaker Recognition</td>
<td>David Dean, Ahilan Kanagasundaram, Houman Ghaemmaghami, Md. Hafizur Rahman, Sridha Sridharan, Queensland University of Technology, Australia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-4</th>
<th>15:00 – 15:20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score Stabilization for Speaker Recognition Trained on a Small Development Set</td>
<td>Hagai Aronowitz, IBM Research Haifa, Israel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-5</th>
<th>15:20 – 15:40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anti-Spoofing System: An Investigation of Measures to Detect Synthetic and Human Speech</td>
<td>Abhinav Misra, Shivesh Ranjan, Chunlei Zhang, John H.L. Hansen, University of Texas at Dallas, USA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-O-46-6</th>
<th>15:40 – 16:00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Likelihood Ratio-Based Forensic Voice Comparison in Microphone vs. Mobile Mismatched Conditions Using Japanese /ai/</td>
<td>Michael J. Carne, Australian National University, Australia</td>
</tr>
</tbody>
</table>
### Oral Session 47: Evaluation of Speech Synthesis

**Chairs:** Alan W. Black, Rob Clark  
**Room:** Conference 4+5, Time 14:00 – 16:00, Thursday, September 10, 2015

#### Thu-O-47-1

**Are We Using Enough Listeners? No! — An Empirically-Supported Critique of Interspeech 2014 TTS Evaluations**  
Mirjam Wester, Cassia Valentini-Botinhao, Gustav Eje Henter, University of Edinburgh, UK

#### Thu-O-47-2

**How to Compare TTS Systems: A New Subjective Evaluation Methodology Focused on Differences**  
Jonathan Chevelu¹, Damien Lolive¹, Sèbastien Le Maquè², David Guennec¹  
¹IRISA, France; ²Universität des Saarlandes, Germany

#### Thu-O-47-3

**Double-Ended Prediction of the Naturalness Ratings of the Blizzard Challenge 2008–2013**  
Lukas Latacz, Werner Verhelst, Vrije Universiteit Brussel, Belgium

#### Thu-O-47-4

**Entropy-Based Sentence Selection for Speech Synthesis Using Phonetic and Prosodic Contexts**  
Takashi Nose¹, Yusuke Arao², Takao Kobayashi², Komei Sugiura³, Yoshinori Shiga³, Akinori Ito¹  
¹Tohoku University, Japan; ²Tokyo Institute of Technology, Japan; ³NICT, Japan

#### Thu-O-47-5

**A Comparison of Speech Synthesis Systems Based on GPR, HMM, and DNN with a Small Amount of Training Data**  
Tomoki Koriyama, Takao Kobayashi, Tokyo Institute of Technology, Japan

#### Thu-O-47-6

**Objective Intelligibility Assessment of Text-to-Speech Systems Through Utterance Verification**  
Raphael Ullmann, Ramya Rasipuram, Mathew Magimai-Doss, Hervé Bourlard, Idiap Research Institute, Switzerland

### Oral Session 48: Adaptive Methods for LVCSR

**Chairs:** Mark Gales, Brian Mak  
**Room:** Conference 6, Time 14:00 – 16:00, Thursday, September 10, 2015

#### Thu-O-48-1

**Continuous Word Representation Using Neural Networks for Proper Name Retrieval from Diachronic Documents**  
Dominique Fohr, Irina Illina, LORIA, France

#### Thu-O-48-2

**Recurrent Neural Network Language Model Adaptation for Multi-Genre Broadcast Speech Recognition**  
X. Chen, T. Tan, Xunying Liu, Pierre Lanchantin, M. Wan, Mark J.F. Gales, Philip C. Woodland, University of Cambridge, UK

#### Thu-O-48-3

**Paragraph Vector Based Topic Model for Language Model Adaptation**  
Wengong Jin, Tianxing He, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China

#### Thu-O-48-4

**Personalized Speech Recognizer with Keyword-Based Personalized Lexicon and Language Model Using Word Vector Representations**  
Ching-Feng Yeh, Yuan-ming Liou, Hung-yi Lee, Lin-shan Lee, National Taiwan University, Taiwan

#### Thu-O-48-5

**Discriminative Data Selection for Lightly Supervised Training of Acoustic Model Using Closed Caption Texts**  
Sheng Li, Yuuya Akita, Tatsuya Kawahara, Kyoto University, Japan

#### Thu-O-48-6

**Cross-Lingual Transfer Learning During Supervised Training in Low Resource Scenarios**  
Amit Das, Mark Hasegawa-Johnson, University of Illinois at Urbana-Champaign, USA
Special Session 9: Robust Speech Processing Using Observation Uncertainty and Uncertainty Propagation

Chairs: Ramón F. Astudillo, Shinji Watanabe
Room: Hall 1, Time 14:00 – 16:00, Thursday, September 10, 2015

Thu-SP9-1 14:00 – 14:15
Session and Paper Overview
Ramón F. Astudillo1, Shinji Watanabe2, Ahmed Hussen Abdelaziz3,
Dorothea Kolossa3
1INESC-ID Lisboa, Portugal; 2MERL, USA; 3Ruhr-Universität Bochum, Germany
Poster Presentations

Uncertainty Propagation for Noise Robust Speaker Recognition: The Case of NIST-SRE (Poster)
Dayana Ribas1, Emmanuel Vincent2, José Ramón Calvo1
1CENATAV, Cuba; 2INRIA, France

Uncertainty Training and Decoding Methods of Deep Neural Networks Based on Stochastic Representation of Enhanced Features (Poster)
Yuuki Tachioka1, Shinji Watanabe2
1Mitsubishi Electric, Japan; 2MERL, USA

Accounting for Uncertainty of i-Vectors in Speaker Recognition Using Uncertainty Propagation and Modified Imputation (Poster)
Rahim Saeidi, Paavo Alku, Aalto University, Finland

Thu-SP9 continued...

Autoencoder Based Multi-Stream Combination for Noise Robust Speech Recognition (Poster)
Sri Harish Mallidi1, Tetsuji Ogawa2, Karel Veselý3, Phani S. Nidadavolu1,
Hynek Hermansky1
1Johns Hopkins University, USA; 2Waseda University, Japan; 3Brno University of Technology, Czech Republic

Uncertainty Decoding for DNN-HMM Hybrid Systems Based on Numerical Sampling (Poster)
Christian Huemmer1, Roland Maas1, Andreas Schwarz1, Ramón F. Astudillo2,
Walter Kellermann1
1FAU Erlangen-Nürnberg, Germany; 2INESC-ID Lisboa, Portugal

Uncertainty Propagation Through Deep Neural Networks (Poster)
Ahmed Hussen Abdelaziz1, Shinji Watanabe2, John R. Hershey2, Emmanuel Vincent3,
Dorothea Kolossa1
1Ruhr-Universität Bochum, Germany; 2MERL, USA; 3INRIA, France

Handling Derivative Filterbank Features in Bounded-Marginalization-Based Missing Data Automatic Speech Recognition (Poster)
Marco Kühne, VoiceBox Technologies, Germany

Large-Scale, Sequence-Discriminative, Joint Adaptive Training for Masking-Based Robust ASR (Poster)
Arun Narayanan, Ananya Misra, Kean K. Chin, Google, USA

Integration of DNN Based Speech Enhancement and ASR (Poster)
Ramón F. Astudillo, Joana Correia, Isabel Trancoso, INESC-ID Lisboa, Portugal
<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-1</th>
<th>A General Artificial Neural Network Extension for HTK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C. Zhang, Philip C. Woodland, University of Cambridge, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-2</th>
<th>Audio Augmentation for Speech Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tom Ko¹, Vijayaditya Peddinti², Daniel Povey², Sanjeev Khudanpur²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Huawei Technologies, China; ²Johns Hopkins University, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-3</th>
<th>A Diversity-Penalizing Ensemble Training Method for Deep Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Xiaoqiang Zhang, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-4</th>
<th>Deep Neural Network Training Emphasizing Central Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gakuto Kurata¹, Daniel Willett²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹IBM Research Tokyo, Japan; ²Nuance Communications, Germany</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-5</th>
<th>Training Deep Bidirectional LSTM Acoustic Model for LVCSR by a Context-Sensitive-Chunk BPTT Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kai Chen¹, Zhi-Jie Yan², Qiang Huo²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹USTC, China; ²Microsoft, China</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-6</th>
<th>Structured Output Layer with Auxiliary Targets for Context-Dependent Acoustic Modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pawel Swietojanski, Peter Bell, Steve Renals, University of Edinburgh, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-7</th>
<th>Complementary Tasks for Context-Dependent Deep Neural Network Acoustic Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peter Bell, Steve Renals, University of Edinburgh, UK</td>
<td></td>
</tr>
</tbody>
</table>

**Thu-P-30 continued…**

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-8</th>
<th>Towards End-to-End Speech Recognition for Chinese Mandarin Using Long Short-Term Memory Recurrent Neural Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jie Li, Heng Zhang, Xinyuan Cai, Bo Xu, Chinese Academy of Sciences, China</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-9</th>
<th>Improving Deep Neural Networks Based Multi-Accent Mandarin Speech Recognition Using I-Vectors and Accent-Specific Top Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mingming Chen¹, Zhanlei Yang¹, Jizhong Liang², Yanpeng Li², Wenju Liu¹</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Chinese Academy of Sciences, China; ²SGCC, China</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-10</th>
<th>Rapid Adaptation for Deep Neural Networks Through Multi-Task Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhen Huang¹, Jinyu Li², Sabato Marco Siniscalchi¹, I-Fan Chen¹, Ji Wu¹, Chin-Hui Lee¹</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Georgia Institute of Technology, USA; ²Microsoft, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-11</th>
<th>fMLLR Based Feature-Space Speaker Adaptation of DNN Acoustic Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sree Hari Krishnan Parthasarathi¹, Bjorn Hoffmeister¹, Spyros Matsoukas¹, Arindam Mandal¹, Nikko Strom¹, Sri Garimella²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>¹Amazon.com, USA; ²Amazon.com, India</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-12</th>
<th>I-Vector Dependent Feature Space Transformations for Adaptive Speech Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Xiangang Li, Xihong Wu, Peking University, China</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-13</th>
<th>Unsupervised Domain Discovery Using Latent Dirichlet Allocation for Acoustic Modelling in Speech Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mortaza Doulaty, Oscar Saz, Thomas Hain, University of Sheffield, UK</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-30-14</th>
<th>Training Data Selection for Acoustic Modeling via Submodular Optimization of Joint Kullback-Leibler Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Taichi Asami, Ryo Masumura, Hirokazu Masataki, Manabu Okamoto, Sumitaka Sakauchi, NTT Corporation, Japan</td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-1</td>
<td>Combination of NN and CRF Models for Joint Detection of Punctuation and Disfluencies</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eunah Cho, Kevin Kilgour, Jan Niehues, Alex Waibel, KIT, Germany</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-2</td>
<td>Tunable Keyword-Aware Language Modeling and Context Dependent Fillers for LVCSR-Based Spoken Keyword Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tze Siong Lau(^1), I-Fan Chen(^2), Chin-Hui Lee(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1) TL@NTU, Singapore; (^2) Georgia Institute of Technology, USA</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-3</td>
<td>Joint Decoding of Tandem and Hybrid Systems for Improved Keyword Spotting on Low Resource Languages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Haipeng Wang, Anton Ragni, Mark J.F. Gales, Kate M. Knill, Philip C. Woodland, C. Zhang, University of Cambridge, UK</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-4</td>
<td>Preserving Word-Level Emphasis in Speech-to-Speech Translation Using Linear Regression HSMMs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quoc Truong Do, Shinnosuke Takamichi, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura, NAIST, Japan</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-5</td>
<td>Phonology-Augmented Statistical Transliteration for Low-Resource Languages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hoang Gia Ngo(^1), Nancy F. Chen(^2), Binh Minh Nguyen(^1), Bin Ma(^2), Haizhou Li(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1) NUS, Singapore; (^2) A*STAR, Singapore</td>
</tr>
</tbody>
</table>

**Thu-P-31 continued...**

<table>
<thead>
<tr>
<th>Page</th>
<th>Thu-P-31-6</th>
<th>Evaluation of Re-Ranking by Prioritizing Highly Ranked Documents in Spoken Term Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Kazuki Oouchi(^1), Ryota Konno(^1), Takahiro Akyu(^1), Kazuma Konno(^1), Kazunori Kajima(^1), Kazuyo Tanaka(^2), Shi-wook Lee(^3), Yoshiaki Itoh(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1) Iwate Prefectural University, Japan; (^2) Tsukuba University, Japan; (^3) AIST, Japan</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-7</td>
<td>Distinctive Feature Based Representation of Speech for Query-by-Example Spoken Term Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abhijeet Saxena, B. Yegnanarayana, IIIT Hyderabad, India</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-8</td>
<td>Combination of Diverse Subword Units in Spoken Term Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shi-wook Lee(^1), Kazuyo Tanaka(^2), Yoshiaki Itoh(^3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1) AIST, Japan; (^2) Tsukuba University, Japan; (^3) Iwate Prefectural University, Japan</td>
</tr>
<tr>
<td>Page</td>
<td>Thu-P-31-9</td>
<td>Sparse Modeling of Posterior Exemplars for Keyword Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dhnananjay Ram, Afsaneh Asaei, Pranay Dighe, Hervé Bourlard, Idiap Research Institute, Switzerland</td>
</tr>
</tbody>
</table>
Poster Session 32: Stress, Load, and Pathologies

Chairs: John H. L. Hansen, Thomas F. Quatieri
Room: Foyer, Time 14:00 – 16:00, Thursday, September 10, 2015

---

**Thu-P-32-1**

Stress Level Detection Using Double-Layer Subband Filter

*Tin Lay Nwe, Qianli Xu, Cuntai Guan, Bin Ma, A*STAR, Singapore*

---

**Thu-P-32-2**

Prosodic Characteristics of Read Speech Before and After Treadmill Running

*Jürgen Trouvain*, *Khiet P. Truong*

1 Universität des Saarlandes, Germany; 2 University of Twente, The Netherlands

---

**Thu-P-32-3**

A Database for Analysis of Speech Under Physical Stress: Detection of Exercise Intensity While Running and Talking

*Khiet P. Truong*, *Arne Nieuwenhuys*, *Peter Beek*, *Vanessa Evers*

1 University of Twente, The Netherlands; 2 Radboud Universiteit Nijmegen, The Netherlands

---

**Thu-P-32-4**

Stressed Out: What Speech Tells Us About Stress

*Will Paul, Cecilia Ovesdotter Alm, Reynold Bailey, Joe Geigel, Linwei Wang*, Rochester Institute of Technology, USA

---

**Thu-P-32-5**

Prediction of Heart Rate Changes from Speech Features During Interaction with a Misbehaving Dialog System

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

---

**Thu-P-32-6**

Acoustic Correlates for Perceived Effort Levels in Expressive Speech

*Mary Pietrowicz, Mark Hasegawa-Johnson, Karrie Karahalios*, University of Illinois at Urbana-Champaign, USA

---

**Thu-P-32-7**

Pitch-Based Speech Perturbation Measures Using a Novel GCI Detection Algorithm: Application to Pathological Voice Classification

*Khalid Daoudi, Ashwini Jaya Kumar*, INRIA, France

---

**Thu-P-32-8**

Speech-Based Assessment of PTSD in a Military Population Using Diverse Feature Classes

*Dimitra Vergyri*, *Bruce Knoth*, *Elizabeth Shriberg*, *Vikramjit Mitra*,

*Mitchell McLaren*, *Luciana Ferrer*, *Pablo Garcia*, *Charles Marmar*

1 SRI International, USA; 2 NYU Langone Medical Center, USA

---

**Thu-P-32-9**

Cognitive Impairment Prediction in the Elderly Based on Vocal Biomarkers

*Bea Yu*, *Thomas F. Quatieri*, *James R. Williamson*, *James C. Mandt*

1 MIT Lincoln Laboratory, USA; 2 Center for Psych Consulting, USA

---

**Thu-P-32-10**

Automatic Age Detection in Normal and Pathological Voice

*J.-A. Gómez-García*, *L. Moro-Velázquez*, *Juan Ignacio Godino-Llorente*,

*G. Castellanos-Dominguez*

1 Universidad Politécnica de Madrid, Spain; 2 Universidad Nacional de Colombia, Colombia