

2016 IEEE Spoken Language Technology Workshop (SLT 2016)

**San Diego, California, USA
13-16 December 2016**



**IEEE Catalog Number: CFP16SLT-POD
ISBN: 978-1-5090-4904-2**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP16SLT-POD
ISBN (Print-On-Demand):	978-1-5090-4904-2
ISBN (Online):	978-1-5090-4903-5

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

ASR-1: SPEECH RECOGNITION 1

ASR-1.1: I-VECTOR ESTIMATION AS AUXILIARY TASK FOR MULTI-TASK LEARNING1 BASED ACOUSTIC MODELING FOR AUTOMATIC SPEECH RECOGNITION

Gueorgui Pironkov, Stephane Dupont, Thierry Dutoit, University of Mons, Belgium

ASR-1.2: BBN TECHNOLOGIES' OPENSAD SYSTEM8

Scott Novotney, Damianos Karakos, Jan Silovsky, Rich Schwartz, Raytheon BBN Technologies, United States

ASR-1.3: A STUDY OF SPEECH DISTORTION CONDITIONS IN REAL SCENARIOS13 FOR SPEECH PROCESSING APPLICATIONS

Dayana Ribas, Advanced Technologies Application Center (CENATAV), Cuba; Emmanuel Vincent, Institute for Research in Computer Science and Automation (INRIA), France; Jose R. Calvo, Advanced Technologies Application Center (CENATAV), Cuba

ASR-1.4: AUTOMATIC OPTIMIZATION OF DATA PERTURBATION DISTRIBUTIONS21 FOR MULTI-STYLE TRAINING IN SPEECH RECOGNITION

Mortaza Doulaty, The University of Sheffield, United Kingdom; Richard Rose, Olivier Siohan, Google Inc., United States

ASR-1.5: BATCH-NORMALIZED JOINT TRAINING FOR DNN-BASED DISTANT28 SPEECH RECOGNITION

Mirco Ravanelli, Fondazione Bruno Kessler, Italy; Philemon Brakel, University of Montreal, Canada; Maurizio Omologo, Fondazione Bruno Kessler, Italy; Yoshua Bengio, University of Montreal, Canada

ASR-1.6: DEEP BOTTLENECK FEATURES AND SOUND-DEPENDENT I-VECTORS35 FOR SIMULTANEOUS RECOGNITION OF SPEECH AND ENVIRONMENTAL SOUNDS

Sakriani Sakti, Seiji Kawanishi, Graham Neubig, Koichiro Yoshino, Satoshi Nakamura, Nara Institute of Science and Technology, Japan

ASR-1.7: LEARNING UTTERANCE-LEVEL NORMALISATION USING VARIATIONAL43 AUTOENCODERS FOR ROBUST AUTOMATIC SPEECH RECOGNITION

Shawn Tan, National University of Singapore, Singapore; Khe Chai Sim, Google Inc., Singapore

ASR-1.8: PERFORMANCE MONITORING FOR AUTOMATIC SPEECH RECOGNITION50 IN NOISY MULTI-CHANNEL ENVIRONMENTS

Bernd T. Meyer, Sri Harish Mallidi, Johns Hopkins University, United States; Angel Mario Castro Martínez, Carl von Ossietzky Universität Oldenburg, Germany; Guillermo Paya-Vaya, Leibniz Universität Hannover, Germany; Hendrik Kayser, Carl von Ossietzky Universität Oldenburg, Germany; Hynek Hermansky, Johns Hopkins University, United States

ASR-1.9: ITERATIVE TRAINING OF A DPGMM-HMM ACOUSTIC UNIT RECOGNIZER57 IN A ZERO RESOURCE SCENARIO

Michael Heck, Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology, Japan

ASR-1.10: TOWARD HUMAN-ASSISTED LEXICAL UNIT DISCOVERY WITHOUT64 TEXT RESOURCES

Chris Bartels, Wen Wang, Vikramjit Mitra, Colleen Richey, Andreas Kathol, Dimitra Vergyri, Harry Bratt, Chiachi Hung, SRI International, United States

ASR-1.11: A NONPARAMETRIC BAYESIAN APPROACH FOR AUTOMATIC DISCOVERY	71
OF A LEXICON AND ACOUSTIC UNITS	
<i>Amir Hossein Harati Nejad Torbati, Jibo Inc, United States; Joseph Picone, Temple University, United States</i>	
ASR-1.12: JOINTLY LEARNING TO ALIGN AND CONVERT GRAPHEMES TO	76
PHONEMES WITH NEURAL ATTENTION MODELS	
<i>Shubham Toshniwal, Karen Livescu, Toyota Technological Institute at Chicago, United States</i>	
 IVR: SPOKEN DIALOG SYSTEMS AND SPEECH APPLICATIONS	
IVR.1: DIALPORT: CONNECTING THE SPOKEN DIALOG RESEARCH COMMUNITY	83
TO REAL USER DATA	
<i>Tiancheng Zhao, Carnegie Mellon University, United States; Kyusong Lee, Pohang University of Science and Technology, United States; Maxine Eskenazi, Carnegie Mellon University, United States</i>	
IVR.2: WEAKLY SUPERVISED USER INTENT DETECTION FOR MULTI-DOMAIN	91
DIALOGUES	
<i>Ming Sun, Carnegie Mellon University, United States; Aasish Pappu, Yahoo Research, United States; Yun-Nung Chen, National Taiwan University, United States; Alexander Rudnicky, Carnegie Mellon University, United States</i>	
IVR.3: LEARNING DIALOGUE DYNAMICS WITH THE METHOD OF MOMENTS	98
<i>Merwan Barlier, Romain Laroche, NaDia Team, Orange Labs, France; Olivier Pietquin, Université Lille 1, France</i>	
IVR.4: TOWARDS A VIRTUAL PERSONAL ASSISTANT BASED ON A USER-DEFINED	106
PORTFOLIO OF MULTI-DOMAIN VOCAL APPLICATIONS	
<i>Tatiana Ekeinhorn-Komi, Jean Léon Bouraoui, Romain Laroche, Orange Labs, France; Fabrice Lefèvre, Université d'Avignon, France</i>	
IVR.5: SPEAKER INDEPENDENT DIARIZATION FOR CHILD LANGUAGE	114
ENVIRONMENT ANALYSIS USING DEEP NEURAL NETWORKS	
<i>Maryam Najafian, John H. L. Hansen, The University of Texas at Dallas, United States</i>	
IVR.6: AUTOMATIC PLAGIARISM DETECTION FOR SPOKEN RESPONSES IN AN	121
ASSESSMENT OF ENGLISH LANGUAGE PROFICIENCY	
<i>Xinhao Wang, Keelan Evanini, James Bruno, Matthew Mulholland, Educational Testing Service, United States</i>	
IVR.7: IMPROVED PREDICTION OF THE ACCENT GAP BETWEEN SPEAKERS OF	129
ENGLISH FOR INDIVIDUAL-BASED CLUSTERING OF WORLD ENGLISHES	
<i>Fumiya Shiozawa, Daisuke Saito, Nobuaki Minematsu, The University of Tokyo, Japan</i>	
IVR.8: SPEECH VS. TEXT: A COMPARATIVE ANALYSIS OF FEATURES FOR	136
DEPRESSION DETECTION SYSTEMS	
<i>Michelle Morales, The Graduate Center, CUNY, United States; Rivka Levitan, Brooklyn College, CUNY, United States</i>	
IVR.9: INCREMENTALLY LEARN THE RELEVANCE OF WORDS IN A DICTIONARY	144
FOR SPOKEN LANGUAGE ACQUISITION	
<i>Vincent Renkens, KULeuven, Belgium; Vikrant Tomar, Fluent.ai, Canada; Hugo Van hamme, KULeuven, Belgium</i>	

IVR.10: ABSTRACTIVE HEADLINE GENERATION FOR SPOKEN CONTENT BY	151
ATTENTIVE RECURRENT NEURAL NETWORKS WITH ASR ERROR MODELING	
<i>Lang-Chi Yu, Hung-yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan</i>	
IVR.11: EXTRACTIVE SPEECH SUMMARIZATION LEVERAGING CONVOLUTIONAL	158
NEURAL NETWORK TECHNIQUES	
<i>Chun-I Tsai, Hsiao-Tsung Hung, National Taiwan Normal University, Taiwan; Kuan-Yu Chen, Academia Sinica, Taiwan; Berlin Chen, National Taiwan Normal University, Taiwan</i>	
 MGB: MGB CHALLENGE	
MGB.1: THE NDSC TRANSCRIPTION SYSTEM FOR THE 2016 MULTI-GENRE	273
BROADCAST CHALLENGE	
<i>Xu-Kui Yang, Dan Qu, Wen-Lin Zhang, National Digital Switching System Engineering and Technological R&D Center, China; Wei-Qiang Zhang, Tsinghua University, China</i>	
MGB.2: THE MGB-2 CHALLENGE: ARABIC MULTI-DIALECT BROADCAST MEDIA	279
RECOGNITION	
<i>Ahmed Ali, Qatar Computing Research Institute, Qatar; Peter Bell, University of Edinburgh, United Kingdom; James Glass, Massachusetts Institute of Technology, United States; Yacine Messaoui, Aljazeera Media Networks, Qatar; Hamdy Mubarak, Qatar Computing Research Institute, Qatar; Steve Renals, University of Edinburgh, United Kingdom; Yifan Zhang, Qatar Computing Research Institute, Qatar</i>	
MGB.3: LIUM ASR SYSTEMS FOR THE 2016 MULTI-GENRE BROADCAST ARABIC	285
CHALLENGE	
<i>Natalia Tomashenko, University of Le Mans, France; Kévin Vythelingum, Voxygen, France; Anthony Rousseau, Yannick Estève, University of Le Mans, France</i>	
MGB.4: QCRI ADVANCED TRANSCRIPTION SYSTEM (QATS) FOR THE ARABIC	292
MULTI-DIALECT BROADCAST MEDIA RECOGNITION: MGB-2 CHALLENGE	
<i>Sameer Khurana, Ahmed Ali, Qatar Computing Research Institute, Qatar</i>	
MGB.5: DEVELOPMENT OF THE MIT ASR SYSTEM FOR THE 2016 ARABIC	299
MULTI-GENRE BROADCAST CHALLENGE	
<i>Tuka AlHanai, Wei-Ning Hsu, James Glass, Massachusetts Institute of Technology, United States</i>	
 NLP: NATURAL LANGUAGE PROCESSING	
NLP.1: PARALLEL LONG SHORT-TERM MEMORY FOR MULTI-STREAM	218
CLASSIFICATION	
<i>Mohamed Bouaziz, Mohamed Morchid, Richard Dufour, Georges Linarès, Renato De Mori, LIA - University of Avignon, France</i>	
NLP.2: IMPROVING MULTI-STREAM CLASSIFICATION BY MAPPING	224
SEQUENCE-EMBEDDING IN A HIGH DIMENSIONAL SPACE	
<i>Mohamed Bouaziz, Mohamed Morchid, Richard Dufour, Georges Linarès, LIA - University of Avignon, France</i>	
NLP.3: HIERARCHICAL ATTENTION MODEL FOR IMPROVED MACHINE	232
COMPREHENSION OF SPOKEN CONTENT	
<i>Wei Fang, Juei-Yang Hsu, Hung-yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan</i>	

NLP.4: COMPARING SPEAKER INDEPENDENT AND SPEAKER ADAPTED	239
CLASSIFICATION FOR WORD PROMINENCE DETECTION	
<i>Andrea Schnall, TU Darmstadt, Germany; Martin Heckmann, Honda Research Institut Europe GmbH, Germany</i>	
NLP.5: AUTOMATIC TURN SEGMENTATION FOR MOVIE & TV SUBTITLES	245
<i>Pierre Lison, Norwegian Computing Centre, Norway; Raveesh Meena, KTH Royal Institute of Technology, Sweden</i>	
NLP.6: VOICE SEARCH LANGUAGE MODEL ADAPTATION USING CONTEXTUAL	253
INFORMATION	
<i>Justin Scheiner, Ian Williams, Petar Aleksic, Google Inc., United States</i>	
NLP.7: ADAPTATION OF SVM FOR MIL FOR INFERRING THE POLARITY OF	258
MOVIES AND MOVIE REVIEWS	
<i>Joana Correia, Carnegie Mellon University, Instituto Superior Tecnico, and INESC-ID, United States; Isabel Trancoso, Instituto Superior Tecnico, and INESC-ID, Portugal; Bhiksha Raj, Carnegie Mellon University, United States</i>	
NLP.8: SEMANTICALLY DRIVEN INVERSION TRANSDUCTION GRAMMAR	265
INDUCTION FOR EARLY STAGE TRAINING OF SPOKEN LANGUAGE TRANSLATION	
<i>Meriem Beloucif, Dekai Wu, HKUST, Hong Kong SAR of China</i>	
 SLR-1: SPEAKER AND LANGUAGE RECOGNITION 1	
SLR-1.1: DEEP NEURAL NETWORK-BASED SPEAKER EMBEDDINGS FOR	165
END-TO-END SPEAKER VERIFICATION	
<i>David Snyder, Pegah Ghahremani, Daniel Povey, Daniel Garcia-Romero, Johns Hopkins University, United States; Yishay Carmiel, Spoken Communications, United States; Sanjeev Khudanpur, Johns Hopkins University, United States</i>	
SLR-1.2: END-TO-END ATTENTION BASED TEXT-DEPENDENT SPEAKER	171
VERIFICATION	
<i>Shixiong Zhang, Microsoft, United States; Zhuo Chen, Columbia University, United States; Yong Zhao, Jinyu Li, Yifan Gong, Microsoft, United States</i>	
SLR-1.3: FURTHER OPTIMISATIONS OF CONSTANT Q CEPSTRAL PROCESSING	179
FOR INTEGRATED UTTERANCE AND TEXT-DEPENDENT SPEAKER VERIFICATION	
<i>Héctor Delgado, Massimiliano Todisco, EURECOM, France; Md Sahidullah, University of Eastern Finland, Finland; Achintya K Sarkar, Aalborg University, Denmark; Nicholas Evans, EURECOM, France; Tomi Kinnunen, University of Eastern Finland, Finland; Zheng-Hua Tan, Aalborg University, Denmark</i>	
SLR-1.4: DEEP NEURAL NETWORK DRIVEN MIXTURE OF PLDA FOR ROBUST	186
I-VECTOR SPEAKER VERIFICATION	
<i>Na Li, Man-Wai Mak, The Hong Kong Polytechnic University, Hong Kong SAR of China; Jen-Tzung Chien, National Chiao Tung University, Taiwan</i>	
SLR-1.5: MODELLING SPEAKER AND CHANNEL VARIABILITY USING DEEP	192
NEURAL NETWORKS FOR ROBUST SPEAKER VERIFICATION	
<i>Gautam Bhattacharya, Jahangir Alam, Patrick Kenny, Vishwa Gupta, CRIM, Canada</i>	

SLR-1.6: ANALYSIS OF THE DNN-BASED SRE SYSTEMS IN MULTI-LANGUAGE CONDITIONS	199
<i>Ondrej Novotný, Pavel Matejka, Ondrej Glembek, Oldrich Plchot, František Grézl, Lukáš Burget, Jan Cernocký, Brno University of Technology, Czech Republic</i>	
SLR-1.7: EVALUATION AND CALIBRATION OF LOMBARD EFFECTS IN SPEAKER VERIFICATION	205
<i>Finnian Kelly, John H. L. Hansen, The University of Texas at Dallas, United States</i>	
SLR-1.8: PHONETIC CONTENT IMPACT ON FORENSIC VOICE COMPARISON	210
<i>Moez Ajili, Jean-françois Bonastre, Waad Ben Kheder, University of Avignon, France; Solange Rossato, University of Grenoble, France; Juliette Kahn, Laboratoire national de métrologie et d'essais, France</i>	
SLR-2: SPEAKER AND LANGUAGE RECOGNITION 2	
SLR-2.1: SPEECH ENHANCEMENT USING LONG SHORT-TERM MEMORY BASED RECURRENT NEURAL NETWORKS FOR NOISE ROBUST SPEAKER VERIFICATION	305
<i>Morten Kolbæk, Zheng-Hua Tan, Jesper Jensen, Aalborg University, Denmark</i>	
SLR-2.2: ENVIRONMENTALLY ROBUST AUDIO-VISUAL SPEAKER IDENTIFICATION	312
<i>Lea Schönherr, Dennis Orth, Ruhr-Universität Bochum, Germany; Martin Heckmann, Honda Research Institute Europe GmbH, Germany; Dorothea Kolossa, Ruhr-Universität Bochum, Germany</i>	
SLR-2.3: A ROBUST DIARIZATION SYSTEM FOR MEASURING DOMINANCE IN PEER-LED TEAM LEARNING GROUPS	319
<i>Harishchandra Dubey, Abhijeet Sangwan, John H. L. Hansen, The University of Texas at Dallas, United States</i>	
SLR-2.4: UNSUPERVISED K-MEANS CLUSTERING BASED OUT-OF-SET CANDIDATE SELECTION FOR ROBUST OPEN-SET LANGUAGE RECOGNITION	324
<i>Qian Zhang, John H. L. Hansen, The University of Texas at Dallas, United States</i>	
SLR-2.5: MULTI-LINGUAL DEEP NEURAL NETWORKS FOR LANGUAGE RECOGNITION	330
<i>Luis Murphy Marcos, University of Puerto Rico, Puerto Rico; Fred Richardson, MIT Lincoln Laboratory, United States</i>	
SLR-2.6: APPROACHES FOR LANGUAGE IDENTIFICATION IN MISMATCHED ENVIRONMENTS	335
<i>Shahan Nercessian, Pedro Torres-Carrasquillo, Gabriel Martinez-Montes, MIT Lincoln Laboratory, United States</i>	
SLR-2.7: A FACTOR ANALYSIS MODEL OF SEQUENCES FOR LANGUAGE RECOGNITION	341
<i>Mohamed Omar, IBM T. J. Watson Research Center, United States</i>	
SLU: SPOKEN LANGUAGE UNDERSTANDING	
SLU.1: SYNTAX OR SEMANTICS? KNOWLEDGE-GUIDED JOINT SEMANTIC FRAME PARSING	348
<i>Yun-Nung Chen, National Taiwan University, Taiwan; Dilek Hakkani-Tur, Gokhan Tur, Google Research, United States; Asli Celikyilmaz, Jianfeng Guo, Li Deng, Microsoft Research, United States</i>	

SLU.2: A LOG-LINEAR WEIGHTING APPROACH IN THE WORD2VEC SPACE FOR SPOKEN LANGUAGE UNDERSTANDING	356
<i>Killian Janod, University of Avignon - Orkis, France; Mohamed Morchid, Richard Dufour, Georges Linarès, University of Avignon, France</i>	
SLU.3: QUATERNION NEURAL NETWORKS FOR SPOKEN LANGUAGE UNDERSTANDING	362
<i>Titouan Parcollet, Mohamed Morchid, Pierre-Michel Bousquet, Richard Dufour, Georges Linarès, University of Avignon, France; Renato De mori, McGill University, Canada</i>	
SLU.4: ROBUST UTTERANCE CLASSIFICATION USING MULTIPLE CLASSIFIERS IN THE PRESENCE OF SPEECH RECOGNITION ERRORS	369
<i>Takeshi Homma, Hitachi Ltd., Japan; Kazuaki Shima, Takuya Matsumoto, Clarion Co Ltd., Japan</i>	
SLU.5: PRE-FILTERED DYNAMIC TIME WARPING FOR POSTERIORGRAM BASED KEYWORD SEARCH	376
<i>Gözde Çetinkaya, Batuhan Gündogdu, Murat Saraçlar, Bogazici University, Turkey</i>	
SLU.6: MULTIMODAL DEEP NEURAL NETS FOR DETECTING HUMOR IN TV SITCOMS	383
<i>Dario Bertero, Pascale Fung, The Hong Kong University of Science and Technology, Hong Kong SAR of China</i>	
SLU.7: AN OVERVIEW OF END-TO-END LANGUAGE UNDERSTANDING AND DIALOG MANAGEMENT FOR PERSONAL DIGITAL ASSISTANTS	391
<i>Ruhi Sarikaya, Paul Crook, Alex Marin, Minwoo Jeong, Jean-Philippe Robichaud, Asli Celikyilmaz, Young-Bum Kim, Alexandre Rochette, Omar Zia Khan, Xiuahu Liu, Daniel Boies, Tasos Anastasakos, Zhalleh Feizollahi, Nikhil Ramesh, Hisami Suzuki, Roman Holenstein, Elizabeth Krawczyk, Vasiliy Radostev, Microsoft, United States</i>	
SLU.8: SEMANTIC MODEL FOR FAST TAGGING OF WORD LATTICES	398
<i>Leonid Velikovich, Google Inc., United States</i>	
SLU.9: OPTIMIZING NEURAL NETWORK HYPERPARAMETERS WITH GAUSSIAN PROCESSES FOR DIALOG ACT CLASSIFICATION	406
<i>Franck DERNONCOURT, Ji Young Lee, Massachusetts Institute of Technology, United States</i>	
SLU.10: INTENT DETECTION USING SEMANTICALLY ENRICHED WORD EMBEDDINGS	414
<i>Joo-Kyung Kim, The Ohio State University, United States; Gokhan Tur, Google Research, United States; Asli Celikyilmaz, Microsoft Research, United States; Bin Cao, Ye-Yi Wang, Microsoft, United States</i>	
 ASR-2: SPEECH RECOGNITION 2	
ASR-2.1: AN UNSUPERVISED VOCABULARY SELECTION TECHNIQUE FOR CHINESE AUTOMATIC SPEECH RECOGNITION	420
<i>Yike Zhang, Pengyuan Zhang, Ta Li, Yonghong Yan, Institute of Acoustics, Chinese Academy of Sciences, China</i>	
ASR-2.2: DYNAMIC ADJUSTMENT OF LANGUAGE MODELS FOR AUTOMATIC SPEECH RECOGNITION USING WORD SIMILARITY	426
<i>Anna Currey, Irina Illina, Dominique Fohr, LORIA-INRIA, France</i>	

ASR-2.3: PUNCTUATED TRANSCRIPTION OF MULTI-GENRE BROADCASTS USING ACOUSTIC AND LEXICAL APPROACHES	433
<i>Ondrej Klejch, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
ASR-2.4: CONTEXTUAL LANGUAGE MODEL ADAPTATION USING DYNAMIC CLASSES	441
<i>Lucy Vasserman, Ben Haynor, Petar Aleksic, Google Inc., United States</i>	
ASR-2.5: UNSUPERVISED CONTEXT LEARNING FOR SPEECH RECOGNITION	447
<i>Assaf Hurwitz Michaely, Mohammadreza Ghodsi, Zelin Wu, Justin Scheiner, Petar Aleksic, Google Inc., United States</i>	
ASR-2.6: AUTOMATED OPTIMIZATION OF DECODER HYPER-PARAMETERS FOR ONLINE LVCSR	454
<i>Akshay Chandrashekar, Ian Lane, Carnegie Mellon University, United States</i>	
ASR-2.7: SEQUENCE TRAINING AND ADAPTATION OF HIGHWAY DEEP NEURAL NETWORKS	461
<i>Liang Lu, Toyota Technological Institute at Chicago, United States</i>	
ASR-2.8: A PRIORITIZED GRID LONG SHORT-TERM MEMORY RNN FOR SPEECH RECOGNITION	467
<i>Wei-Ning Hsu, Yu Zhang, James Glass, Massachusetts Institute of Technology, United States</i>	
ASR-2.9: MAX-POOLING LOSS TRAINING OF LONG SHORT-TERM MEMORY NETWORKS FOR SMALL-FOOTPRINT KEYWORD SPOTTING	474
<i>Ming Sun, Anirudh Raju, Amazon.com, United States; George Tucker, Google Inc., United States; Sankaran Panchapagesan, Gengshen Fu, Arindam Mandal, Spyros Matsoukas, Nikko Strom, Shiv Vitaladevuni, Amazon.com, United States</i>	
ASR-2.10: VERY DEEP CONVOLUTIONAL NEURAL NETWORKS FOR ROBUST SPEECH RECOGNITION	481
<i>Yanmin Qian, Shanghai Jiao Tong University, China; Philip C Woodland, Cambridge University, United Kingdom</i>	
ASR-2.11: DEEP LEARNING WITH MAXIMAL FIGURE-OF-MERIT COST TO ADVANCE MULTI-LABEL SPEECH ATTRIBUTE DETECTION	489
<i>Ivan Kukanov, Ville Hautamäki, University of Eastern Finland, Finland; Marco Siniscalchi, The Kore University of Enna, Italy; Kehuang Li, Georgia Institute of Technology, United States</i>	
ASR-2.12: END-TO-END TRAINING APPROACHES FOR DISCRIMINATIVE SEGMENTAL MODELS	496
<i>Hao Tang, Weiran Wang, Kevin Gimpel, Karen Livescu, Toyota Technological Institute at Chicago, United States</i>	
ASR-2.13: DISCRIMINATIVE ACOUSTIC WORD EMBEDDINGS: RECURRENT NEURAL NETWORK-BASED APPROACHES	503
<i>Shane Settle, Karen Livescu, Toyota Technological Institute at Chicago, United States</i>	

DSTC: FIFTH DIALOG STATE TRACKING CHALLENGE

DSTC.1: THE FIFTH DIALOG STATE TRACKING CHALLENGE.....511

Seokhwan Kim, Luis Fernando D'Haro, Rafael Enrique Banchs, Institute for Infocomm Research, Singapore; Jason D. Williams, Microsoft Research, United States; Matthew Henderson, Google Inc., United States; Koichiro Yoshino, Nara Institute of Science and Technology, Japan

DSTC.2: RECURRENT CONVOLUTIONAL NEURAL NETWORKS FOR518 **STRUCTURED SPEECH ACT TAGGING**

Takashi Ushio, Hongjie Shi, Mitsuru Endo, Katsuyoshi Yamagami, Noriaki Horii, Panasonic Corporation, Japan

DSTC.3: THE MSIIP SYSTEM FOR DIALOG STATE TRACKING CHALLENGE 5.....525

Ying Su, Miao Li, Ji Wu, Tsinghua University, China

DSTC.4: NEURAL DIALOG STATE TRACKER FOR LARGE ONTOLOGIES BY531 **ATTENTION MECHANISM**

Youngsoo Jang, Jiyeon Ham, Byung-Jun Lee, Youngjae Chang, Kee-Eung Kim, KAIST, Republic of Korea

DSTC.5: LSTM ENCODER-DECODER FOR DIALOGUE RESPONSE GENERATIONN/A

Zhenlong Yu, Caixia Yuan, Xiaojie Wang, Guohua Yang, Beijing University of Posts and Telecommunications, China

DSTC.6: TRACKING DIALOG STATES USING AN AUTHOR-TOPIC BASED544 **REPRESENTATION**

Richard Dufour, Mohamed Morchid, Titouan Parcollet, LIA - University of Avignon, France

DSTC.7: DIALOG STATE TRACKING WITH ATTENTION-BASED552 **SEQUENCE-TO-SEQUENCE LEARNING**

Takaaki Hori, Mitsubishi Electric Research Laboratories, United States; Hai Wang, Toyota Technological Institute at Chicago, United States; Chiori Hori, Shinji Watanabe, Bret Harsham, Jonathan Le Roux, John Hershey, Mitsubishi Electric Research Laboratories, United States; Yusuke Koji, Yi Jing, Mitsubishi Electric Corporation, Japan; Zhaocheng Zhu, Peking University, China; Takeyuki Aikawa, Mitsubishi Electric Corporation, Japan

DSTC.8: A MULTICHANNEL CONVOLUTIONAL NEURAL NETWORK FOR559 **CROSS-LANGUAGE DIALOG STATE TRACKING**

Hongjie Shi, Takashi Ushio, Mitsuru Endo, Katsuyoshi Yamagami, Noriaki Horii, Panasonic Corporation, Japan

ASR-3: SPEECH RECOGNITION 3 AND SYNTHESIS

ASR-3.1: CODE-SWITCHING DETECTION USING MULTILINGUAL DNNS.....610

Emre Yilmaz, Henk Van den Heuvel, David Van Leeuwen, Radboud University Nijmegen, Netherlands

ASR-3.2: ATTRIBUTE BASED SHARED HIDDEN LAYERS FOR CROSS-LANGUAGE617 **KNOWLEDGE TRANSFER**

Vipul Arora, Aditi Lahiri, University of Oxford, United Kingdom; Henning Reetz, Goethe University, Germany

ASR-3.3: TOWARDS ACOUSTIC MODEL UNIFICATION ACROSS DIALECTS.....624

Mohamed Elfeky, Meysam Bastani, Xavier Velez, Pedro Moreno, Austin Waters, Google Inc., United States

ASR-3.4: BOOSTING PERFORMANCE ON LOW-RESOURCE LANGUAGES BY	629
STANDARD CORPORA: AN ANALYSIS	
<i>František Grézl, Martin Karafiat, Brno University of Technology, Czech Republic</i>	
ASR-3.5: MULTILINGUAL BLSTM AND SPEAKER-SPECIFIC VECTOR ADAPTATION IN	637
2016 BUT BABEL SYSTEM	
<i>Martin Karafiat, Murali Karthick Baskar, Pavel Matejka, Karel Vesely, František Grézl, Jan Cernocký, Speech@FIT VUT, Czech Republic</i>	
ASR-3.6: DNN ADAPTATION FOR RECOGNITION OF CHILDREN SPEECH	644
THROUGH AUTOMATIC UTTERANCE SELECTION	
<i>Marco Matassoni, Daniele Falavigna, Diego Giuliani, Fondazione Bruno Kessler, Italy</i>	
ASR-3.7: LOW-RANK BASES FOR FACTORIZED HIDDEN LAYER ADAPTATION OF	652
DNN ACOUSTIC MODELS	
<i>Lahiru Samarakoon, National University of Singapore, Singapore; Khe Chai Sim, Google Inc., United States</i>	
ASR-3.8: DEEP NEURAL NETWORK BASED ACOUSTIC MODEL PARAMETER	659
REDUCTION USING MANIFOLD REGULARIZED LOW RANK MATRIX FACTORIZATION	
<i>Hoon Chung, Jeom Ja Kang, Ki Young Park, Sung Joo Lee, Jeon Gue Park, Electronics and Telecommunications Research Institute, Republic of Korea</i>	
ASR-3.9: AUTOMATED STRUCTURE DISCOVERY AND PARAMETER TUNING OF	665
NEURAL NETWORK LANGUAGE MODEL BASED ON EVOLUTION STRATEGY	
<i>Tomohiro Tanaka, Tokyo Institute of Technology, Japan; Takafumi Moriya, NTT Corporation, Japan; Takahiro Shinozaki, Tokyo Institute of Technology, Japan; Shinji Watanabe, Takaaki Hori, Mitsubishi Electric Research Laboratories, United States; Kevin Duh, Johns Hopkins University, United States</i>	
ASR-3.10: ENTROPY-BASED PRUNING OF HIDDEN UNITS TO REDUCE DNN	672
PARAMETERS	
<i>Gautam Mantena, National University of Singapore, Singapore; Khe Chai Sim, Google Inc., United States</i>	
ASR-3.11: INFLUENCE OF CORPUS SIZE AND CONTENT ON THE PERCEPTUAL	680
QUALITY OF A UNIT SELECTION MARYTTS VOICE	
<i>Florian Hinterleitner, Benjamin Weiss, Sebastian Möller, TU Berlin, Germany</i>	
ASR-3.12: MEDIAN-BASED GENERATION OF SYNTHETIC SPEECH DURATIONS	686
USING A NON-PARAMETRIC APPROACH	
<i>Srikanth Ronanki, Oliver Watts, Simon King, Gustav Henter, University of Edinburgh, United Kingdom</i>	
ASR-3.13: F0 TRANSFORMATION TECHNIQUES FOR STATISTICAL VOICE	693
CONVERSION WITH DIRECT WAVEFORM MODIFICATION WITH SPECTRAL DIFFERENTIAL	
<i>Kazuhiro Kobayashi, Nara Institute of Science and Technology, Japan; Tomoki Toda, Nagoya University, Japan; Satoshi Nakamura, Nara Institute of Science and Technology, Japan</i>	
 MP: MULTIMODAL AND SPEECH PROCESSING	
MP.1: RECOGNIZING EMOTIONS IN SPOKEN DIALOGUE WITH	565
HIERARCHICALLY FUSED ACOUSTIC AND LEXICAL FEATURES	
<i>Leimin Tian, Johanna Moore, Catherine Lai, The University of Edinburgh, United Kingdom</i>	

MP.2: LOOK, LISTEN, AND DECODE: MULTIMODAL SPEECH RECOGNITION WITH IMAGES	573
<i>Felix Sun, David Harwath, James Glass, Massachusetts Institute of Technology, United States</i>	
MP.3: AUDIO-VISUAL SPEECH ACTIVITY DETECTION IN A TWO-SPEAKER SCENARIO INCORPORATING DEPTH INFORMATION FROM A PROFILE OR FRONTAL VIEW	579
<i>Spyridon Thermos, Gerasimos Potamianos, University of Thessaly, Greece</i>	
MP.4: ANALYSIS OF USER BEHAVIOR WITH MULTIMODAL VIRTUAL CUSTOMER SERVICE AGENTS	585
<i>Ian Beaver, Cynthia Freeman, NextIT Corporation, United States</i>	
MP.5: HIGH QUALITY AGREEMENT-BASED SEMI-SUPERVISED TRAINING DATA FOR ACOUSTIC MODELING	592
<i>Félix de Chaumont Quitry, Asa Oines, Pedro Moreno, Eugene Weinstein, Google Inc., United States</i>	
MP.6: BLIND SPEECH SEGMENTATION USING SPECTROGRAM IMAGE-BASED FEATURES AND MEL CEPSTRAL COEFFICIENTS	597
<i>Adriana Stan, Technical University of Cluj-Napoca, Romania; Cassia Valentini-Botinhao, University of Edinburgh, United Kingdom; Bogdan Orza, Mircea Giurgiu, Technical University of Cluj-Napoca, Romania</i>	
MP.7: DISCRIMINATIVE MULTIPLE SOUND SOURCE LOCALIZATION BASED ON DEEP NEURAL NETWORKS USING INDEPENDENT LOCATION MODEL	603
<i>Ryu Takeda, Kazunori Komatani, Osaka University, Japan</i>	