

2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2017)

**Okinawa, Japan
16 – 20 December 2017**



**IEEE Catalog Number: CFP17SRW-POD
ISBN: 978-1-5090-4789-5**

**Copyright © 2017 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:	CFP17SRW-POD
ISBN (Print-On-Demand):	978-1-5090-4789-5
ISBN (Online):	978-1-5090-4788-8

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

ADV.1: ASR IN ADVERSE ENVIRONMENTS

ADV.1.1: NOISE-ROBUST EXEMPLAR MATCHING FOR RESCORING1 QUERY-BY-EXAMPLE SEARCH

Emre Yilmaz, Radboud University, Netherlands; Julien van Hout, Horacio Franco, SRI International, United States

ADV.1.2: LEARNING SPEAKER REPRESENTATION FOR NEURAL NETWORK BASED8 MULTICHANNEL SPEAKER EXTRACTION

Katerina Zmolikova, Brno University of Technology, Czech Republic; Marc Delcroix, Keisuke Kinoshita, Takuya Higuchi, Atsunori Ogawa, Tomohiro Nakatani, NTT, Japan

ADV.1.3: UNSUPERVISED DOMAIN ADAPTATION FOR ROBUST SPEECH16 RECOGNITION VIA VARIATIONAL AUTOENCODER-BASED DATA AUGMENTATION

Wei-Ning Hsu, Yu Zhang, James Glass, Massachusetts Institute of Technology, United States

ADV.1.4: BINAURAL PROCESSING FOR ROBUST RECOGNITION OF DEGRADED24 SPEECH

Anjali Menon, Carnegie Mellon University, United States; Chanwoo Kim, Google Inc., United States; Umpei Kurokawa, Richard Stern, Carnegie Mellon University, United States

ADV.1.5: MEETING RECOGNITION WITH ASYNCHRONOUS DISTRIBUTED32 MICROPHONE ARRAY

Shoko Araki, NTT, Japan; Nobutaka Ono, National Institute of Informatics, Japan; Keisuke Kinoshita, Marc Delcroix, NTT, Japan

ADV.1.6: ADVERSARIAL TRAINING FOR DATA-DRIVEN SPEECH ENHANCEMENT40 WITHOUT PARALLEL CORPUS

Takuya Higuchi, Keisuke Kinoshita, Marc Delcroix, Tomohiro Nakatani, NTT, Japan

ADV.1.7: TACKLING UNSEEN ACOUSTIC CONDITIONS IN QUERY-BY-EXAMPLE48 SEARCH USING TIME AND FREQUENCY CONVOLUTION FOR MULTILINGUAL DEEP BOTTLENECK FEATURES

Julien van Hout, Vikramjit Mitra, Horacio Franco, Chris Bartels, Dimitra Vergyri, SRI International, United States

ADV.1.8: IMPROVING SEPARATION OF OVERLAPPED SPEECH FOR MEETING55 CONVERSATIONS USING UNCALIBRATED MICROPHONE ARRAY

Keisuke Nakamura, Randy Gomez, Honda Research Institute Japan Co., Ltd., Japan

ASR.1: AUTOMATIC SPEECH RECOGNITION I

ASR.1.1: REDUCING THE COMPUTATIONAL COMPLEXITY FOR WHOLE WORD63 MODELS

Hagen Soltau, Hank Liao, Hasim Sak, Google Inc., United States

ASR.1.2: INVESTIGATION OF LATTICE-FREE MAXIMUM MUTUAL INFORMATION-BASED ACOUSTIC MODELS WITH SEQUENCE-LEVEL KULLBACK-LEIBLER DIVERGENCE	69
<i>Naoyuki Kanda, Yusuke Fujita, Kenji Nagamatsu, Hitachi Ltd., Japan</i>	
ASR.1.3: SEMI-SUPERVISED TRAINING STRATEGIES FOR DEEP NEURAL NETWORKS	77
<i>Matthew Gibson, Gary Cook, Puming Zhan, Nuance Communications Inc., United Kingdom</i>	
ASR.1.4: MULTI-TASK ENSEMBLES WITH TEACHER-STUDENT TRAINING	84
<i>Jeremy Heng Meng Wong, Mark J. F. Gales, University of Cambridge, United Kingdom</i>	
ASR.1.5: LANGUAGE DIARIZATION FOR SEMI-SUPERVISED BILINGUAL ACOUSTIC MODEL TRAINING	91
<i>Emre Yilmaz, Radboud University, Netherlands; Mitchell McLaren, SRI International, United States; Henk van den Heuvel, David A. van Leeuwen, Radboud University, Netherlands</i>	
ASR.1.6: FUTURE WORD CONTEXTS IN NEURAL NETWORK LANGUAGE MODELS	97
<i>Xie Chen, University of Cambridge, United Kingdom; Xunying Liu, Chinese University of Hong Kong, China; Anton Ragni, Yu Wang, Mark J. F. Gales, University of Cambridge, United Kingdom</i>	
ASR.1.7: FUTURE VECTOR ENHANCED LSTM LANGUAGE MODEL FOR LVCSR	104
<i>Qi Liu, Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China</i>	
ASR.1.8: ACOUSTIC-TO-WORD MODEL WITHOUT OOV	111
<i>Jinyu Li, Guoli Ye, Rui Zhao, Jasha Droppo, Yifan Gong, Microsoft, United States</i>	
ASR.1.9: TURBO FUSION OF MAGNITUDE AND PHASE INFORMATION FOR DNN-BASED PHONEME RECOGNITION	118
<i>Timo Lohrenz, Tim Fingscheidt, TU Braunschweig, Germany</i>	
ASR.1.10: COMPUTATIONAL COST REDUCTION OF LONG SHORT-TERM MEMORY BASED ON SIMULTANEOUS COMPRESSION OF INPUT AND HIDDEN STATE	126
<i>Takashi Masuko, Toshiba Corporation, Japan</i>	
ASR.1.11: CROSS-DOMAIN SPEECH RECOGNITION USING NONPARALLEL CORPORA WITH CYCLE-CONSISTENT ADVERSARIAL NETWORKS	134
<i>Masato Mimura, Shinsuke Sakai, Tatsuya Kawahara, Kyoto University, Japan</i>	
ASR.1.12: WERD: USING SOCIAL TEXT SPELLING VARIANTS FOR EVALUATING DIALECTAL SPEECH RECOGNITION	141
<i>Ahmed Ali, Preslav Nakov, Qatar Computing Research Institute, Qatar; Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
ASR.1.13: CHARACTER-BASED UNITS FOR UNLIMITED VOCABULARY CONTINUOUS SPEECH RECOGNITION	149
<i>Peter Smit, Siva Reddy Gangireddy, Seppo Enarvi, Sami Virpioja, Mikko Kurimo, Aalto University, Finland</i>	
ASR.1.14: GATED CONVOLUTIONAL NETWORKS BASED HYBRID ACOUSTIC MODELS FOR LOW RESOURCE SPEECH RECOGNITION	157
<i>Jian Kang, Wei-Qiang Zhang, Jia Liu, Tsinghua university, China</i>	

ASR.1.15: LATTICE RESCORING STRATEGIES FOR LONG SHORT TERM MEMORY	165
LANGUAGE MODELS IN SPEECH RECOGNITION	
<i>Shankar Kumar, Michael Nirschl, Daniel Holtmann-Rice, Hank Liao, Ananda Theertha Suresh, Felix Yu, Google Inc., United States</i>	
ASR.1.16: SYLLABLE-BASED ACOUSTIC MODELING WITH CTC-SMBR-LSTM	173
<i>Zhongdi Qu, Parisa Haghani, Eugene Weinstein, Pedro Moreno, Google Inc., United States</i>	
ASR.1.17: SEQUENCE TRAINING OF DNN ACOUSTIC MODELS WITH NATURAL	178
GRADIENT	
<i>Adnan Haider, Philip Woodland, University of Cambridge, United Kingdom</i>	
ASR.1.18: CONSISTENT DNN UNCERTAINTY TRAINING AND DECODING FOR	185
ROBUST ASR	
<i>Karan Nathwani, Emmanuel Vincent, INRIA, Nancy, France; Irina Illina, INRIA-LORIA, Nancy, France</i>	
 ASR.2: AUTOMATIC SPEECH RECOGNITION II	
ASR.2.1: EXPLORING ARCHITECTURES, DATA AND UNITS FOR STREAMING	193
END-TO-END SPEECH RECOGNITION WITH RNN-TRANSDUCER	
<i>Kanishka Rao, Hasim Sak, Rohit Prabhavalkar, Google Inc., United States</i>	
ASR.2.2: UNSUPERVISED ADAPTATION OF STUDENT DNNs LEARNED FROM	200
TEACHER RNNs FOR IMPROVED ASR PERFORMANCE	
<i>Lahiru Samarakoon, Brian Mak, Hong Kong University of Science and Technology, China</i>	
ASR.2.3: EXPLORING NEURAL TRANSDUCERS FOR END-TO-END SPEECH	206
RECOGNITION	
<i>Eric Battenberg, Jitong Chen, Rewon Child, Adam Coates, Yashesh Gaur, Yi Li, Hairong Liu, Sanjeev Satheesh, Anuroop Sriram, Zhenyao Zhu, Baidu SVAIL, United States</i>	
ASR.2.4: UNSUPERVISED ADAPTATION WITH DOMAIN SEPARATION NETWORKS	214
FOR ROBUST SPEECH RECOGNITION	
<i>Zhong Meng, Georgia Institute of Technology, United States; Zhuo Chen, Vadim Mazalov, Jinyu Li, Yifan Gong, Microsoft Corporation, United States</i>	
ASR.2.5: INCREMENTAL TRAINING AND CONSTRUCTING THE VERY DEEP	222
CONVOLUTIONAL RESIDUAL NETWORK ACOUSTIC MODELS	
<i>Sheng Li, Xugang Lu, Peng Shen, Ryoichi Takashima, National Institute of Information and Communications Technology, Japan; Tatsuya Kawahara, Kyoto University, Japan; Hisashi Kawai, National Institute of Information and Communications Technology, Japan</i>	
ASR.2.6: ON LATTICE GENERATION FOR LARGE VOCABULARY SPEECH	228
RECOGNITION	
<i>David Rybach, Michael Riley, Johan Schalkwyk, Google Inc., Germany</i>	
ASR.2.7: SIMPLIFYING VERY DEEP CONVOLUTIONAL NEURAL NETWORK	236
ARCHITECTURES FOR ROBUST SPEECH RECOGNITION	
<i>Joanna Rownicka, Steve Renals, Peter Bell, University of Edinburgh, United Kingdom</i>	
ASR.2.8: LANGUAGE MODELING WITH HIGHWAY LSTM	244
<i>Gakuto Kurata, Bhuvana Ramabhadran, George Saon, Abhinav Sethy, IBM Research, Japan</i>	

ASR.2.9: DIRECT MODELING OF RAW AUDIO WITH DNNS FOR WAKE WORD DETECTION	252
<i>Kenichi Kumatani, Sankaran Panchapagesan, Minhua Wu, Minjae Kim, Nikko Strom, Gautam Tiwari, Arindam Mandal, Amazon Inc., United States</i>	
ASR.2.10: IMPROVING THE EFFICIENCY OF FORWARD-BACKWARD ALGORITHM USING BATCHED COMPUTATION IN TENSORFLOW	258
<i>Khe Chai Sim, Arun Narayanan, Tom Bagby, Tara Sainath, Michiel Bacchiani, Google Inc., United States</i>	
ASR.2.11: LANGUAGE INDEPENDENT END-TO-END ARCHITECTURE FOR JOINT LANGUAGE IDENTIFICATION AND SPEECH RECOGNITION	265
<i>Shinji Watanabe, Johns Hopkins University, United States; Takaaki Hori, John Hershey, Mitsubishi Electric Research Laboratories, United States</i>	
ASR.2.12: KEYWORD SPOTTING FOR GOOGLE ASSISTANT USING CONTEXTUAL SPEECH RECOGNITION	272
<i>Assaf Hurwitz Michaely, Xuedong Zhang, Gabor Simko, Carolina Parada, Petar Aleksic, Google Inc., United States</i>	
ASR.2.13: INVESTIGATION OF TRANSFER LEARNING FOR ASR USING LF-MMI TRAINED NEURAL NETWORKS	279
<i>Pegah Ghahremani, Vimal Manohar, Hossein Hadian, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, United States</i>	
ASR.2.14: MULTI-LEVEL LANGUAGE MODELING AND DECODING FOR OPEN VOCABULARY END-TO-END SPEECH RECOGNITION	287
<i>Takaaki Hori, Mitsubishi Electric Research Laboratories, United States; Shinji Watanabe, Johns Hopkins University, United States; John Hershey, Mitsubishi Electric Research Laboratories, United States</i>	
ASR.2.15: LANGUAGE MODELING WITH NEURAL TRANS-DIMENSIONAL RANDOM FIELDS	294
<i>Bin Wang, Zhijian Ou, Tsinghua university, China</i>	
ASR.2.16: LISTENING WHILE SPEAKING: SPEECH CHAIN BY DEEP LEARNING	301
<i>Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology, Japan</i>	
ASR.2.17: ATTENTION-BASED WAV2TEXT WITH FEATURE TRANSFER LEARNING	309
<i>Andros Tjandra, Sakriani Sakti, Satoshi Nakamura, Nara Institute of Science and Technology, Japan</i>	

CH.1: CHALLENGE OVERVIEW

CH.1.1: SPEECH RECOGNITION CHALLENGE IN THE WILD: ARABIC MGB-3	316
<i>Ahmed Ali, Stephan Vogel, Qatar Computing Research Institute, Qatar; Steve Renals, University of Edinburgh, United Kingdom</i>	
CH.1.2: THE ZERO RESOURCE SPEECH CHALLENGE 2017	323
<i>Ewan Dunbar, Xuan Nga Cao, Juan Benjumea, Julien Karadayi, Mathieu Bernard, ENS, EHESS, PSL Research University, CNRS, INRIA, France; Laurent Besacier, LIG, Université Grenoble Alpes, CNRS, Grenoble INP, France; Xavier Anguera, Elsa Corp., Portugal; Emmanuel Dupoux, ENS, EHESS, PSL Research University, CNRS, INRIA, France</i>	

CH.1.3: THE BLIZZARD MACHINE LEARNING CHALLENGE 2017331
Kei Sawada, Keiichi Tokuda, Nagoya Institute of Technology, Japan; Simon King, University of Edinburgh, United Kingdom; Alan W. Black, Carnegie Mellon University, United States

**MGB.1: MULTI-GENRE BROADCAST MEDIA TRANSCRIPTION CHALLENGE:
MGB-3**

MGB.1.1: AALTO SYSTEM FOR THE 2017 ARABIC MULTI-GENRE BROADCAST CHALLENGE338

Peter Smit, Siva Reddy Gangireddy, Seppo Enarvi, Sami Virpioja, Mikko Kurimo, Aalto University, Finland

MGB.1.2: JHU KALDI SYSTEM FOR ARABIC MGB-3 ASR CHALLENGE USING DIARIZATION, AUDIO-TRANSCRIPT ALIGNMENT AND TRANSFER LEARNING346

Vimal Manohar, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, United States

MGB.1.3: AUTOMATIC SPEECH RECOGNITION OF ARABIC MULTI-GENRE BROADCAST MEDIA353

Maryam Najafian, Wei-Ning Hsu, Massachusetts Institute of Technology, United States; Ahmed Ali, Qatar Computing Research Institute, Qatar; James Glass, Massachusetts Institute of Technology, United States

MGB.1.4: UTD-CRSS SUBMISSION FOR MGB-3 ARABIC DIALECT IDENTIFICATION: FRONT-END AND BACK-END ADVANCEMENTS ON BROADCAST SPEECH360

Ahmet E. Bulut, Qian Zhang, Chunlei Zhang, Fahimeh Bahmaninezhad, John H. L. Hansen, University of Texas at Dallas, United States

MGB.1.5: MGB-3 BUT SYSTEM: LOW-RESOURCE ASR ON EGYPTIAN YOUTUBE DATA368

Karel Veselý, Karthick Murali Baskar, Mireia Diez, Karel Beneš, Brno University of Technology, Czech Republic

MGB.1.6: MIT-QCRI ARABIC DIALECT IDENTIFICATION SYSTEM FOR THE 2017 MULTI-GENRE BROADCAST CHALLENGE374

Suwon Shon, MIT CSAIL, United States; Ahmed Ali, Qatar Computing Research Institute, Qatar; James Glass, MIT CSAIL, United States

NEW.1: NEW APPLICATIONS OF ASR

NEW.1.1: SEEING AND HEARING TOO: AUDIO REPRESENTATION FOR VIDEO CAPTIONING381

Shun Po Chuang, Chia-Hung Wan, Pang-Chi Huang, Chi-Yu Yang, Hung-Yi Lee, National Taiwan University, Taiwan

NEW.1.2: MULTITASK TRAINING WITH UNLABELED DATA FOR END-TO-END SIGN LANGUAGE FINGERSPELLING RECOGNITION389

Bowen Shi, Karen Livescu, Toyota Technological Institute at Chicago, United States

NEW.1.3: A HIERARCHICAL ATTENTION BASED MODEL FOR OFF-TOPIC SPONTANEOUS SPOKEN RESPONSE DETECTION397

Andrey Malinin, Kate Knill, Mark J. F. Gales, University of Cambridge, United Kingdom

NEW.1.4: A CONTEXT-AWARE SPEECH RECOGNITION AND UNDERSTANDING	404
SYSTEM FOR AIR TRAFFIC CONTROL DOMAIN	
<i>Youssef Oualil, Dietrich Klakow, György Szaszák, Saarland University, Germany; Ajay Srinivasamurthy, Idiap Research Institute, Switzerland; Hartmut Helmke, German Aerospace Center, Germany; Petr Motlicek, Idiap Research Institute, Switzerland</i>	
NEW.1.5: SPOKEN LANGUAGE BIOMARKERS FOR DETECTING COGNITIVE	409
IMPAIRMENT	
<i>Tuka Alhanai, Massachusetts Institute of Technology, United States; Rhoda Au, Boston University School of Medicine and Public Health, United States; James Glass, Massachusetts Institute of Technology, United States</i>	
NEW.1.6: DBLSTM BASED MULTILINGUAL ARTICULATORY FEATURE	417
EXTRACTION FOR LANGUAGE DOCUMENTATION	
<i>Markus Müller, Sebastian Stüker, Alex Waibel, Karlsruhe Institute of Technology, Germany</i>	
NEW.1.7: LEARNING MODALITY-INVARIANT REPRESENTATIONS FOR SPEECH AND	424
IMAGES	
<i>Kenneth Leidal, David Harwath, James Glass, Massachusetts Institute of Technology, United States</i>	
NEW.1.8: EARLY AND LATE INTEGRATION OF AUDIO FEATURES FOR AUTOMATIC	430
VIDEO DESCRIPTION	
<i>Chiori Hori, Takaaki Hori, Tim Marks, John Hershey, Mitsubishi Electric Research Laboratories, United States</i>	
NEW.1.9: CRACKING THE COCKTAIL PARTY PROBLEM BY MULTI-BEAM DEEP	437
ATTRACTOR NETWORK	
<i>Zhuo Chen, Jinyu Li, Xiong Xiao, Takuya Yoshioka, Huaming Wang, Zhenghao Wang, Yifan Gong, Microsoft Corporation, United States</i>	
NEW.1.10: GROUND TRUTH ESTIMATION OF SPOKEN ENGLISH FLUENCY	445
SCORE USING DECORRELATION PENALIZED LOW-RANK MATRIX FACTORIZATION	
<i>Hoon Chung, Yun Kyung Lee, Jeon Gue Park, Electronics and Telecommunications Research Institute, Korea (South)</i>	
 SLP.1: SPOKEN LANGUAGE PROCESSING, DIALOG, MACHINE TRANSLATION	
SLP.1.1: EXPLORING THE USE OF ACOUSTIC EMBEDDINGS IN NEURAL	450
MACHINE TRANSLATION	
<i>Salil Deena, Raymond Wai Man Ng, Pranava Madhyastha, Lucia Specia, Thomas Hain, University of Sheffield, United Kingdom</i>	
SLP.1.2: UNWRITTEN LANGUAGES DEMAND ATTENTION TOO! WORD	458
DISCOVERY WITH ENCODER-DECODER MODELS	
<i>Marcely Zanon Boito, Alexandre Bérard, Laboratoire d'Informatique de Grenoble, France; Aline Villavicencio, Institute of Informatics, Federal University of Rio Grande do Sul, Brazil; Laurent Besacier, Laboratoire d'Informatique de Grenoble, France</i>	
SLP.1.3: NEURAL RELEVANCE-AWARE QUERY MODELING FOR SPOKEN	466
DOCUMENT RETRIEVAL	
<i>Tien-Hong Lo, Ying-Wen Chen, National Taiwan Normal University, Taiwan; Kuan-Yu Chen, National Taiwan University of Science and Technology, Taiwan; Hsin-Min Wang, Academia Sinica, Taiwan; Berlin Chen, National Taiwan Normal University, Taiwan</i>	

SLP.1.4: STREAMING SMALL-FOOTPRINT KEYWORD SPOTTING USING SEQUENCE-TO-SEQUENCE MODELS	474
<i>Yanzhang He, Rohit Prabhavalkar, Kanishka Rao, Wei Li, Anton Bakhtin, Ian McGraw, Google Inc., United States</i>	
SLP.1.5: ITERATIVE POLICY LEARNING IN END-TO-END TRAINABLE TASK-ORIENTED NEURAL DIALOG MODELS	482
<i>Bing Liu, Ian Lane, Carnegie Mellon University, United States</i>	
SLP.1.6: DENOTATION EXTRACTION FOR INTERACTIVE LEARNING IN DIALOGUE SYSTEMS	490
<i>Miroslav Vodolán, Filip Jurcícek, Charles University in Prague, Czech Republic</i>	
SLP.1.7: MITIGATING THE IMPACT OF SPEECH RECOGNITION ERRORS ON CHATBOT USING SEQUENCE-TO-SEQUENCE MODEL	497
<i>Pin-Jung Chen, I-Hung Hsu, Yi Yao Huang, Hung-Yi Lee, National Taiwan University, Taiwan</i>	
SLP.1.8: DEEP QUATERNION NEURAL NETWORKS FOR SPOKEN LANGUAGE UNDERSTANDING	504
<i>Titouan Parcollet, Mohamed Morchid, Georges Linarès, University of Avignon, France</i>	
SLP.1.9: TOPIC SEGMENTATION IN ASR TRANSCRIPTS USING BIDIRECTIONAL RNNs FOR CHANGE DETECTION	512
<i>Imran Sheikh, TATA Consultancy Services, India; Dominique Fohr, CNRS, France; Irina Illina, Université de Lorraine, France</i>	
SLP.1.10: GROUNDED LANGUAGE UNDERSTANDING FOR MANIPULATION INSTRUCTIONS USING GAN-BASED CLASSIFICATION	519
<i>Komei Sugiura, Hisashi Kawai, National Institute of Information and Communications Technology, Japan</i>	
SLP.1.11: HIERARCHICAL RECURRENT NEURAL NETWORK FOR STORY SEGMENTATION USING FUSION OF LEXICAL AND ACOUSTIC FEATURES	525
<i>Emiru Tsunoo, Ondrej Klejch, Peter Bell, Steve Renals, University of Edinburgh, United Kingdom</i>	
SLP.1.12: PERSONALIZED WORD REPRESENTATIONS CARRYING PERSONALIZED SEMANTICS LEARNED FROM SOCIAL NETWORK POSTS	533
<i>Zih-Wei Lin, Tzu-Wei Sung, Hung-Yi Lee, Lin-Shan Lee, National Taiwan University, Taiwan</i>	
SLP.1.13: SPEAKER-SENSITIVE DUAL MEMORY NETWORKS FOR MULTI-TURN SLOT TAGGING	541
<i>Young-Bum Kim, Alexa Brain/Amazon, United States; Sungjin Lee, Microsoft Research, United States; Ruhi Sarikaya, Alexa Brain/Amazon, United States</i>	
SLP.1.14: ONENET: JOINT DOMAIN, INTENT, SLOT PREDICTION FOR SPOKEN LANGUAGE UNDERSTANDING	547
<i>Young-Bum Kim, Alexa Brain/Amazon, United States; Sungjin Lee, Microsoft Research, United States; Karl Stratos, Toyota Technological Institute, United States</i>	
SLP.1.15: DYNAMIC TIME-AWARE ATTENTION TO SPEAKER ROLES AND CONTEXTS FOR SPOKEN LANGUAGE UNDERSTANDING	554
<i>Po-Chun Chen, Ta-Chung Chi, Shang-Yu Su, Yun-Nung Chen, National Taiwan University, Taiwan</i>	
SLP.1.16: SCALABLE MULTI-DOMAIN DIALOGUE STATE TRACKING	561
<i>Abhinav Rastogi, Dilek Hakkani-Tur, Larry Heck, Google Inc., United States</i>	

SLP.1.17: EXPLORING ASR-FREE END-TO-END MODELING TO IMPROVE569
SPOKEN LANGUAGE UNDERSTANDING IN A CLOUD-BASED DIALOG SYSTEM
Yao Qian, Rutuja Ubale, Vikram Ramanaryanan, Patrick Lange, David Suendermann-Oeft, Keelan Evanini, Eugene Tsuprun, Educational Testing Service, United States

SLR.1: SPEAKER/LANGUAGE RECOGNITION

SLR.1.1: LEVERAGING SIDE INFORMATION FOR SPEAKER IDENTIFICATION WITH577
THE ENRON CONVERSATIONAL TELEPHONE SPEECH COLLECTION
Ning Gao, University of Maryland, College Park, United States; Gregory Sell, Human Language Technology Center of Excellence, The Johns Hopkins University, United States; Douglas Oard, University of Maryland, College Park, United States; Mark Dredze, Human Language Technology Center of Excellence, The Johns Hopkins University, United States

SLR.1.2: END-TO-END TEXT-INDEPENDENT SPEAKER VERIFICATION WITH584
FLEXIBILITY IN UTTERANCE DURATION
Chunlei Zhang, University of Texas at Dallas, United States; Kazuhito Koishida, Microsoft Corporation, United States

SLR.1.3: SPOOFING DETECTION VIA SIMULTANEOUS VERIFICATION OF591
AUDIO-VISUAL SYNCHRONICITY AND TRANSCRIPTION
Lea Schönherr, Steffen Zeiler, Dorothea Kolossa, Ruhr-Universität Bochum, Germany

SLR.1.4: ADVERSARIAL MANIFOLD LEARNING FOR SPEAKER RECOGNITION599
Jen-Tzung Chien, Kang-Ting Peng, National Chiao Tung University, Taiwan

SLR.1.5: IMPROVING NATIVE LANGUAGE (L1) IDENTIFICATION WITH BETTER VAD606
AND TDNN TRAINED SEPARATELY ON NATIVE AND NON-NATIVE ENGLISH
CORPORA
Yao Qian, Keelan Evanini, Patrick Lange, Robert A. Pugh, Rutuja Ubale, Educational Testing Service, United States; Frank K Soong, Microsoft Research Asia, China

SLR.1.6: MULTI-VIEW (JOINT) PROBABILITY LINEAR DISCRIMINATION ANALYSIS614
FOR J-VECTOR BASED TEXT DEPENDENT SPEAKER VERIFICATION
Ziqiang Shi, Liu Liu, Mengjiao Wang, Rujie Liu, Fujitsu Research and Development Center, China

SLR.1.7: LEVERAGING NATIVE LANGUAGE SPEECH FOR ACCENT IDENTIFICATION621
USING DEEP SIAMESE NETWORKS
Aditya Siddhant, Carnegie Mellon University, United States; Preethi Jyothi, Indian Institute of Technology Bombay, India; Sriram Ganapathy, Indian Institute of Science, India

SLR.1.8: COMPARISON OF MULTIPLE FEATURES AND MODELING METHODS629
FOR TEXT-DEPENDENT SPEAKER VERIFICATION
Yi Liu, Liang He, Yao Tian, Zhuzi Chen, Jia Liu, Tsinghua University, China; Michael T. Johnson, University of Kentucky, United States

SLR.1.9: INVESTIGATING NATIVE AND NON-NATIVE ENGLISH CLASSIFICATION637
AND TRANSFER EFFECTS USING LEGENDRE POLYNOMIAL COEFFICIENT
CLUSTERING
Rachel Rakov, The Graduate Center, CUNY, United States; Andrew Rosenberg, IBM Research, United States

SML.1: SPEECH SYNTHESIS AS A MACHINE LEARNING PROBLEM

SML.1.1: THE CMU ENTRY TO BLIZZARD MACHINE LEARNING CHALLENGE644

Pallavi Baljekar, Sai Krishna Rallabandi, Alan W. Black, Carnegie Mellon University, United States

SML.1.2: THE USTC SYSTEM FOR BLIZZARD MACHINE LEARNING CHALLENGE650 2017-ES2

Ya Jun Hu, University of Science and Technology of China, China; Li Juan Liu, Chuang Ding, IFLYTEK CO.,LTD., China; Zhen Hua Ling, Li Rong Dai, University of Science and Technology of China, China

SML.1.3: THE IFLYTEK SYSTEM FOR BLIZZARD MACHINE LEARNING657 CHALLENGE 2017-ES1

Li Juan Liu, Chuang Ding, IFLYTEK CO.,LTD., China; Ya Jun Hu, Zhen Hua Ling, University of Science and Technology of China, China; Yuan Jiang, Ming Zhou, Si Wei, IFLYTEK CO.,LTD., China

TTS.1: TEXT-TO-SPEECH SYSTEMS

TTS.1.1: MINIMALLY SUPERVISED WRITTEN-TO-SPOKEN TEXT665 NORMALIZATION

Axel H. Ng, Kyle Gorman, Richard Sproat, Google Inc., United States

TTS.1.2: PERCEPTUAL QUALITY AND MODELING ACCURACY OF EXCITATION671 PARAMETERS IN DLSTM-BASED SPEECH SYNTHESIS SYSTEMS

Eunwoo Song, Yonsei University, Korea (South); Frank K. Soong, Microsoft Research Asia, China; Hong-Goo Kang, Yonsei University, Korea (South)

TTS.1.3: SPARSE REPRESENTATION OF PHONETIC FEATURES FOR VOICE677 CONVERSION WITH AND WITHOUT PARALLEL DATA

Berrak Sisman, Haizhou Li, National University of Singapore, Singapore; Kay Chen Tan, City University of Hong Kong, China

TTS.1.4: STATISTICAL PARAMETRIC SPEECH SYNTHESIS USING GENERATIVE685 ADVERSARIAL NETWORKS UNDER A MULTI-TASK LEARNING FRAMEWORK

*Shan Yang, Lei Xie, Northwestern Polytechnical University, China; Xiao Chen, Xiaoyan Lou, Xuan Zhu, Samsung R&D Institute of China, China; Dongyan Huang, Institute for Infocomm Research, A*STAR, Singapore; Haizhou Li, National University of Singapore, Singapore*

TTS.1.5: ERROR DETECTION OF GRAPHEME-TO-PHONEME CONVERSION IN692 TEXT-TO-SPEECH SYNTHESIS USING SPEECH SIGNAL AND LEXICAL CONTEXT

Kévin Vythelingum, Yannick Estève, Le Mans University, France; Olivier Rosec, Voxygen, France

TTS.1.6: SUBBAND WAVENET WITH OVERLAPPED SINGLE-SIDEBAND698 FILTERBANKS

Takuma Okamoto, Kentaro Tachibana, National Institute of Information and Communications Technology, Japan; Tomoki Toda, Nagoya University, Japan; Yoshinori Shiga, Hisashi Kawai, National Institute of Information and Communications Technology, Japan

TTS.1.7: INTEGRATED SPEAKER-ADAPTIVE SPEECH SYNTHESIS705

Moquan Wan, Gilles Degottex, Mark J. F. Gales, University of Cambridge, United Kingdom

TTS.1.8: AN INVESTIGATION OF MULTI-SPEAKER TRAINING FOR WAVENET VOCODER712
Tomoki Hayashi, Akira Tamamori, Kazuhiro Kobayashi, Kazuya Takeda, Tomoki Toda, Nagoya University, Japan

ZRS.1: THE ZERO RESOURCE SPEECH CHALLENGE 2017

ZRS.1.1: AN EMBEDDED SEGMENTAL K-MEANS MODEL FOR UNSUPERVISED SEGMENTATION AND CLUSTERING OF SPEECH719

Herman Kamper, Stellenbosch University, South Africa; Karen Livescu, Toyota Technological Institute at Chicago, United States; Sharon Goldwater, University of Edinburgh, United Kingdom

ZRS.1.2: MULTILINGUAL BOTTLE-NECK FEATURE LEARNING FROM UNTRANSCRIBED SPEECH727

*Hongjie Chen, Northwestern Polytechnical University, China; Cheung-Chi Leung, Institute for Infocomm Research, A*STAR, Singapore; Lei Xie, Northwestern Polytechnical University, China; Bin Ma, Institute for Infocomm Research, A*STAR, Singapore; Haizhou Li, National University of Singapore, Singapore*

ZRS.1.3: EXTRACTING BOTTLENECK FEATURES AND WORD-LIKE PAIRS FROM UNTRANSCRIBED SPEECH FOR FEATURE REPRESENTATION734

*Yougen Yuan, Northwestern Polytechnical University, China; Cheung-Chi Leung, Institute for Infocomm Research, A*STAR, Singapore; Lei Xie, Hongjie Chen, Northwestern Polytechnical University, China; Bin Ma, Institute for Infocomm Research, A*STAR, Singapore; Haizhou Li, National University of Singapore, Singapore*

ZRS.1.4: FEATURE OPTIMIZED DPGMM CLUSTERING FOR UNSUPERVISED SUBWORD MODELING: A CONTRIBUTION TO ZEROSPEECH 2017740

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

ZRS.1.5: COMPOSITE EMBEDDING SYSTEMS FOR ZEROSPEECH2017 TRACK1747

Hayato Shibata, Taku Kato, Takahiro Shinozaki, Tokyo Institute of Technology, Japan; Shinji Watanabe, Johns Hopkins University, United States

ZRS.1.6: DEEP LEARNING METHODS FOR UNSUPERVISED ACOUSTIC MODELING - LEAP SUBMISSION TO ZEROSPEECH CHALLENGE 2017754

Ansari TK, Rajath Kumar, Sonali Singh, Sriram Ganapathy, Indian Institute of Science, India

ZRS.1.7: UNSUPERVISED HMM POSTERIOGRAMS FOR LANGUAGE INDEPENDENT ACOUSTIC MODELING IN ZERO RESOURCE CONDITIONS762

Ansari TK, Rajath Kumar, Sonali Singh, Sriram Ganapathy, Susheela Devi, Indian Institute of Science, India