2013 IEEE Workshop on Automatic Speech Recognition and Understanding

(ASRU 2013)

Olomouc, Czech Republic
8-12 December 2013
# TABLE OF CONTENTS

**LM: LANGUAGE MODELING**

**LM.1: K-COMPONENT RECURRENT NEURAL NETWORK LANGUAGE MODELS** .................1
**USING CURRICULUM LEARNING**
Yangyang Shi, Martha Larson, Catholijn M Jonker, Delft University of Technology, Netherlands

**LM.2: LEARNING A SUBWORD VOCABULARY BASED ON UNIGRAM LIKELIHOOD** ..........7
Matti Varjokallio, Mikko Kurimo, Sami Virpioja, Aalto University, Finland

**LM.3: EFFECTIVE PSEUDO-RELEVANCE FEEDBACK FOR LANGUAGE MODELING** ........13
**IN SPEECH RECOGNITION**
Berlin Chen, Yi-Wen Chen, National Taiwan Normal University, Taiwan; Kuan-Yu Chen, Institute of Information Science, Academia Sinica, Taiwan; Ea-Ee Jan, IBM Thomas J. Watson Research Center, United States

**LM.4: LEARNING BETTER LEXICAL PROPERTIES FOR RECURRENT OOV WORDS** ..........19
Long Qin, M*Modal Inc, United States; Alexander I. Rudnicky, Carnegie Mellon University, United States

**LM.5: JOINT TRAINING OF INTERPOLATED EXPONENTIAL N-GRAM MODELS** ..........25
Abhinav Sethy, Stanley Chen, Ebru Arisoy, Bhuvana Ramabhadran, IBM, United States; Kartik Audkhasi, Shrikanth Narayanan, University of Southern California, United States; Paul Vozila, Nuance Communications, United States

**LM.6: MIXTURE OF MIXTURE N-GRAM LANGUAGE MODELS** ........................................31
Hasim Sak, Cyril Allauzen, Kaisuke Nakajima, Francoise Beaufays, Google, United States

**AM: ACOUSTIC MODELING**

**AM.1: COMPACT ACOUSTIC MODELING BASED ON ACOUSTIC MANIFOLD USING A** ..........37
**MIXTURE OF FACTOR ANALYZERS**
Wen-Lin Zhang, Zhengzhou Information Science and Technology Institute, China; Wei-Qiang Zhang, Tsinghua University, China; Bi-Cheng Li, Zhengzhou Information Science and Technology Institute, China

**AM.2: A GENERALIZED DISCRIMINATIVE TRAINING FRAMEWORK FOR SYSTEM** ..........43
**COMBINATION**
Yuuki Tachioka, Mitsubishi Electric, Japan; Shinji Watanabe, Jonathan Le Roux, John Hershey, Mitsubishi Electric Research Laboratories, United States

**AM.3: ACOUSTIC MODELING USING TRANSFORM-BASED PHONE-CLUSTER** .................49
**ADAPTIVE TRAINING**
Vimal Manohar, Bhargav Srinivas Ch., Umesh Srinivasan, Indian Institute of Technology Madras, India

**AM.4: SPEAKER ADAPTATION OF NEURAL NETWORK ACOUSTIC MODELS USING** ..........55
**I-VECTORS**
George Saon, Hagen Soltau, David Nahamoo, Michael Picheny, IBM, United States
AM.5: NEIGHBOUR SELECTION AND ADAPTATION FOR RAPID .....................................................60

SPEAKER-DEPENDENT ASR
Udhyakumar Nallasamy, Carnegie Mellon University, United States; Mark Fuhs, Monika Woszczyna, M*Modal Inc, United States; Florian Metze, Carnegie Mellon University, United States; Tanja Schultz, Karlsruhe Institute of Technology, Germany

Dec: DECODER SEARCH

Dec.1: EFFICIENT NEARLY ERROR-LESS LVCSR DECODING BASED ON .................................................................66
INCREMENTAL FORWARD AND BACKWARD PASSES
David Nolden, Ralf Schlüter, Hermann Ney, RWTH Aachen University, Germany

SLU: SPOKEN LANGUAGE UNDERSTANDING

SLU.1: QUERY UNDERSTANDING ENHANCED BY HIERARCHICAL PARSING ...........................................72
STRUCTURES
Jingjing Liu, Panupong Pasupat, Yining Wang, Scott Cyphers, Jim Glass, Massachusetts Institute of Technology, United States

SLU.2: CONVOLUTIONAL NEURAL NETWORK BASED TRIANGULAR CRF FOR ...........................................78
JOINT INTENT DETECTION AND SLOT FILLING
Puyang Xu, Ruhi Sarikaya, Microsoft, United States

SLU.3: SEMANTIC ENTITY DETECTION FROM MULTIPLE ASR HYPOTHESES ...........................................84
WITHIN THE WFST FRAMEWORK
Jan Svec, Pavel Ircing, Lubos Smidl, University of West Bohemia, Czech Republic

SLU.4: ON-LINE ADAPTATION OF SEMANTIC MODELS FOR SPOKEN LANGUAGE ..................................90
UNDERSTANDING
Ali Orkan Bayer, Giuseppe Riccardi, University of Trento, Italy

SLU.5: DYSFLUENT SPEECH DETECTION BY IMAGE FORENSICS TECHNIQUES ...........................................96
Juraj Palfy, Sakhia Darjaa, Slovak Academy of Sciences, Slovakia; Jiri Pospichal, Slovak University of Technology, Slovakia

SLU.6: BARGE-IN EFFECTS IN BAYESIAN DIALOGUE ACT RECOGNITION AND SIMULATION .........................102
Heriberto Cuayahuitl, Nina Dethlefs, Helen Hastie, Oliver Lemon, Heriot-Watt University, United Kingdom

Dial: SPOKEN DIALOG SYSTEMS

Dial.1: EXPERT-BASED REWARD SHAPING AND EXPLORATION SCHEME FOR .........................................108
BOOSTING POLICY LEARNING OF DIALOGUE MANAGEMENT
Emmanuel Ferreira, Fabrice Lefèvre, Laboratoire Informatique d’Avignon, France

Dial.2: DIALOGUE MANAGEMENT FOR LEADING THE CONVERSATION IN ...........................................114
PERSUASIVE DIALOGUE SYSTEMS
Takuya Hiraoka, Yuki Yamauchi, Graham Neubig, Sakrini Sakti, Tomoki Toda, Satoshi Nakamura, Nara Institute of Science and Technology, Japan
Dial.3: UNSUPERVISED INDUCTION AND FILLING OF SEMANTIC SLOTS FOR SPOKEN DIALOGUE SYSTEMS USING FRAME-SEMANTIC PARSING
Yun-Nung Chen, William Yang Wang, Alexander I. Rudnicky, Carnegie Mellon University, United States

Multi: MULTILINGUAL SPEECH & LANGUAGE PROCESSING

Multi.1: CROSS-LINGUAL CONTEXT SHARING AND PARAMETER-TYING FOR MULTI-LINGUAL SPEECH RECOGNITION
Aanchan Mohan, Richard Rose, McGill University, Canada

Multi.2: IMPROVED PUNCTUATION RECOVERY THROUGH COMBINATION OF MULTIPLE SPEECH STREAMS
João Miranda, Instituto Superior Técnico / Carnegie Mellon University, Portugal; João Neto, Instituto Superior Técnico, Portugal; Alan Black, Carnegie Mellon University, United States

Multi.3: INVESTIGATION OF MULTILINGUAL DEEP NEURAL NETWORKS FOR SPOKEN TERM DETECTION
Kate Knill, Mark Gales, Shakti Rath, Phil Woodland, Chao Zhang, Shi-Xiong Zhang, University of Cambridge,

Multi.4: LANGUAGE STYLE AND DOMAIN ADAPTATION FOR CROSS-LANGUAGE SLU PORTING
Evgeny Stepanov, Ilya Kashkarev, Orkan Bayer, Giuseppe Riccardi, Arindam Ghosh, University of Trento, Italy

Robust: ROBUSTNESS IN ASR

Robust.1: AUTOMATIC MODEL COMPLEXITY CONTROL FOR GENERALIZED VARIABLE PARAMETER HMMS
Rongfeng Su, Shenzhen Institutes of Advanced Technology, China; Xunying Liu, Cambridge University, United Kingdom; Lan Wang, Shenzhen Institutes of Advanced Technology, China

Robust.2: IMPROVED CEPSTRAL MEAN AND VARIANCE NORMALIZATION USING BAYESIAN FRAMEWORK
Vishnu Prasad N, Umesh S, Indian Institute of Technology Madras, India

Robust.3: THE SECOND ‘CHIME’ SPEECH SEPARATION AND RECOGNITION CHALLENGE: AN OVERVIEW OF CHALLENGE SYSTEMS AND OUTCOMES
Emmanuel Vincent, Inria, France; Jon Barker, University of Sheffield, United Kingdom; Shinji Watanabe, Jonathan Le Roux, Mitsubishi Electric Research Laboratories, United States; Francesco Nesta, Conexant Systems, United States; Marco Matassoni, FBK-Irst, Italy

Robust.4: LEARNING STATE LABELS FOR SPARSE CLASSIFICATION OF SPEECH WITH MATRIX DECONVOLUTION
Antti Hurmalainen, Tuomas Virtanen, Tampere University of Technology, Finland

Robust.5: MODIFIED SPLICE AND ITS EXTENSION TO NON-Stereo DATA FOR NOISE ROBUST SPEECH RECOGNITION
Pavan Kumar D S, Vishnu Prasad N, Indian Institute of Technology Madras, India; Vikas Joshi, IBM India Research Labs, India; Umesh S, Indian Institute of Technology Madras, India

Robust.6: A PROPAGATION APPROACH TO MODELLING THE JOINT DISTRIBUTIONS OF CLEAN AND CORRUPTED SPEECH IN THE MEL-CEPSTRAL DOMAIN
Ramón Astudillo, INESC-ID Lisboa, Portugal
Robust.7: VECTOR TAYLOR SERIES BASED HMM ADAPTATION FOR GENERALIZED CEPSTRUM IN NOISY ENVIRONMENT
Soonho Baek, Hong-Goo Kang, Yonsei University, Republic of Korea

SDRKWS: SPOKEN DOCUMENT RETRIEVAL AND KEYWORD SPOTTING

SDRKWS.1: THE TAO OF ATWV: PROBING THE MYSTERIES OF KEYWORD SEARCH PERFORMANCE
Steven Wegmann, Arlo Faria, Adam Janin, Korbinian Riedhammer, Nelson Morgan, ICSI, United States

SDRKWS.2: TOWARDS UNSUPERVIZED SEMANTIC RETRIEVAL OF SPOKEN CONTENT WITH QUERY EXPANSION BASED ON AUTOMATICALLY DISCOVERED ACOUSTIC PATTERNS
Yun-Chiao Li, National Taiwan University, Taiwan; Hung-yi Lee, Academia Sinica, Taiwan; Cheng-Tao Chung, Chun-an Chan, Lin-shan Lee, National Taiwan University, Taiwan

SDRKWS.3: THE IBM KEYWORD SEARCH SYSTEM FOR THE DARPA RATS PROGRAM
Lidia Mangu, Hagen Soltau, Hong-Kwang Kuo, George Saon, IBM, United States

SDRKWS.4: SCORE NORMALIZATION AND SYSTEM COMBINATION FOR IMPROVED KEYWORD SPOTTING
Damianos Karakos, Richard Schwartz, Stavros Tsakalidis, Le Zhang, Shivesh Ranjan, Tim Ng, Roger Hsiao, Gururprasad Saikumar, Ivan Bulyko, Long Nguyen, John Makhoul, Raytheon BBN Technologies, United States; Frantisek Grezl, Mirko Hannemann, Martin Karafiat, Igor Szoke, Karel Vesely, Brno University of Technology, Czech Republic; Lori Lamel, CNRS-LIMSI, France; Viet-Bac Le, Vocapia Research, France

NewApp: NEW APPLICATIONS OF ASR

NewApp.1: EMOTION RECOGNITION FROM SPONTANEOUS SPEECH USING HIDDEN MARKOV MODELS WITH DEEP BELIEF NETWORKS
Duc Le, Emily Mower Provost, University of Michigan, United States

NewApp.2: AUTOMATIC PRONUNCIATION CLUSTERING USING A WORLD ENGLISH ARCHIVE AND PRONUNCIATION STRUCTURE ANALYSIS
Han-Ping Shen, National Cheng Kung University, Taiwan; Nobuaki Minematsu, The University of Tokyo, Japan; Takehiko Makino, Chuo University, Japan; Steven H. Weinberger, George Mason University, United States; Teeraphon Pongkittiphan, The University of Tokyo, Japan; Chung-Hsien Wu, National Cheng Kung University, Taiwan

NewApp.3: PHONETIC AND ANTHROPOMETRIC CONDITIONING OF MSA-KST COGNITIVE IMPAIRMENT CHARACTERIZATION SYSTEM
Alexei Ivanov, Shahab Jalalvand, Roberto Gretter, Daniele Falavigna, Fondazione Bruno Kessler, Italy

NewApp.4: ASR FOR ELECTRO-LARYNGEAL SPEECH
Anna Katharina Fuchs, Juan Andres Morales-Cordovilla, Martin Hagmüller, Graz University of Technology, Austria

NewApp.5: AUTOMATIC SENTIMENT EXTRACTION FROM YOUTUBE VIDEOS
Lakshmish Kaushik, Abhijeet Sangwan, John H. L. Hansen, University of Texas at Dallas, United States
SPFea: SPEECH SIGNAL PROCESSING AND FEATURE EXTRACTION

SPFea.1: ACOUSTIC CHARACTERISTICS RELATED TO THE PERCEPTUAL PITCH IN WHISPERED VOWELS
Hideaki Konno, Hideo Kanemitsu, Nobuyuki Takahashi, Hokkaido University of Education, Japan; Mineichi Kudo, Hokkaido University, Japan

SPFea.2: AN SVD-BASED SCHEME FOR MFCC COMPRESSION IN DISTRIBUTED SPEECH RECOGNITION SYSTEM
Azzedine Touazi, Mohamed Debyche, University of Science and Technology Houari Boumediene, Algeria

SPFea.3: A STUDY OF SUPERVISED INTRINSIC SPECTRAL ANALYSIS FOR TIMIT PHONE CLASSIFICATION
Reza Sahraeian, Dirk Van Compernolle, Katholieke Universiteit Leuven, Belgium

SPFea.4: MODELS OF TONE FOR TONAL AND NON-TONAL LANGUAGES
Florian Metze, Zaid A. W. Sheikh, Carnegie Mellon University, United States; Alex Waibel, Karlsruhe Institute of Technology / Carnegie Mellon University, Germany; Jonas Gehring, Kevin Kilgour, Quoc Bao Nguyen, Van Huy Nguyen, Karlsruhe Institute of Technology, Germany

NN: NEURAL NETWORKS IN ASR

NN.1: SEMI-SUPERVISED TRAINING OF DEEP NEURAL NETWORKS
Karel Vesely, Mirko Hannemann, Lukas Burget, Brno University of Technology, Czech Republic

NN.2: HYBRID SPEECH RECOGNITION WITH DEEP BIDIRECTIONAL LSTM
Alex Graves, Navdeep Jaitly, Abdel-rahman Mohamed, University of Toronto, Canada

NN.3: IMPROVING ROBUSTNESS OF DEEP NEURAL NETWORKS VIA SPECTRAL MASKING FOR AUTOMATIC SPEECH RECOGNITION
Bo Li, Khe Chai Sim, National University of Singapore, Singapore

NN.4: HYBRID ACOUSTIC MODELS FOR DISTANT AND MULTICHANNEL LARGE VOCABULARY SPEECH RECOGNITION
Pawel Swietojanski, Arnab Ghoshal, Steve Renals, University of Edinburgh, United Kingdom

NN.5: DEEP MAXOUT NEURAL NETWORKS FOR SPEECH RECOGNITION
Meng Cai, Yongzhe Shi, Jia Liu, Tsinghua University, China

NN.6: LEARNING FILTER BANKS WITHIN A DEEP NEURAL NETWORK FRAMEWORK
Tara Sainath, Brian Kingsbury, IBM, United States; Abdel-Rahman Mohamed, University of Toronto, Canada; Bhuvana Ramabhadran, IBM, United States

NN.7: ACCELERATING HESSIAN-FREE OPTIMIZATION FOR DEEP NEURAL NETWORKS BY IMPLICIT PRECONDITIONING AND SAMPLING
Tara Sainath, Lior Horesh, Brian Kingsbury, Aleksandr Aravkin, Bhuvana Ramabhadran, IBM, United States

NN.8: ELASTIC SPECTRAL DISTORTION FOR LOW RESOURCE SPEECH RECOGNITION WITH DEEP NEURAL NETWORKS
Naoyuki Kanda, Ryu Takeda, Yasunari Obuchi, Hitachi Ltd., Japan
NN.9: IMPROVEMENTS TO DEEP CONVOLUTIONAL NEURAL NETWORKS FOR LVCSR
Tara Sainath, Brian Kingsbury, IBM, United States; Abdel-Rahman Mohamed, George Dahl, University of Toronto, United States; George Saon, Hagen Soltau, Tomas Beran, Aleksandr Aravkin, Bhuvana Ramabhadran, IBM, United States

NN.10: COMBINING STOCHASTIC AVERAGE GRADIENT AND HESSIAN-FREE OPTIMIZATION FOR SEQUENCE TRAINING OF DEEP NEURAL NETWORKS
Pierre Dognin, Vaibhava Goel, IBM Research, United States

NN.11: ACCELERATING RECURRENT NEURAL NETWORK TRAINING VIA TWO STAGE CLASSES AND PARALLELIZATION
Zhiheng Huang, Geoffrey Zweig, Michael Levit, Benoit Dumoulin, Barlas Oguz, Shawn Chang, Microsoft, United States

NN.12: IMPACT OF DEEP MLP ARCHITECTURE ON DIFFERENT ACOUSTIC MODELING TECHNIQUES FOR UNDER-RESOURCED SPEECH RECOGNITION
David Imseng, Petr Motlicek, Philip N. Garner, Hervé Bourlard, Idiap Research Institute, Switzerland

NN.13: CONTEXT-DEPENDENT MODELLING OF DEEP NEURAL NETWORK USING LOGISTIC REGRESSION
Guangsen Wang, Khe Chai Sim, National University of Singapore, Singapore

NN.14: DNN ACOUSTIC MODELING WITH MODULAR MULTI-LINGUAL FEATURE EXTRACTION NETWORKS
Jonas Gehring, Quoc Bao Nguyen, Karlsruhe Institute of Technology, Germany; Florian Metze, Carnegie Mellon University, United States; Alex Waibel, Karlsruhe Institute of Technology, Germany

NN.15: DISCRIMINATIVE PIECEWISE LINEAR TRANSFORMATION BASED ON DEEP LEARNING FOR NOISE ROBUST AUTOMATIC SPEECH RECOGNITION
Yosuke Kashiwagi, Daisuke Saito, Nobuaki Minematsu, Keikichi Hirose, The University of Tokyo, Japan

NN.16: PORTING CONCEPTS FROM DNNS BACK TO GMMS
Kris Demuynck, Fabian Trifrenbach, Ghent University, Belgium

NN.17: HIERARCHICAL NEURAL NETWORKS AND ENHANCED CLASS POSTERIORS FOR SOCIAL SIGNAL CLASSIFICATION
Raymond Brueckner, Technische Universität München, Germany; Björn Schuller, Imperial College London, United Kingdom

NN.18: LARGE SCALE DEEP NEURAL NETWORK ACOUSTIC MODELING WITH SEMI-SUPERVISED TRAINING DATA FOR YOUTUBE VIDEO TRANSCRIPTION
Hank Liao, Erik McDermott, Andrew Senior, Google, United States

LowZero: ASR/SPEECH SEARCH WITH LOW OR ZERO RESOURCES

LowZero.1: ACOUSTIC DATA-DRIVEN PRONUNCIATION LEXICON FOR LARGE VOCABULARY SPEECH RECOGNITION
Liang Lu, Arnab Ghoshal, Steve Renals, University of Edinburgh, United Kingdom

LowZero.2: ACOUSTIC UNIT DISCOVERY AND PRONUNCIATION GENERATION FROM A GRAPHEME-BASED LEXICON
William Hartmann, Anindya Roy, Lori Lamel, Jean-Luc Gauvain, LIMSI-CNRS, France
LowZero.3: A HIERARCHICAL SYSTEM FOR WORD DISCOVERY EXPLOITING DTW-BASED INITIALIZATION
Oliver Walter, Timo Korthals, Reinhold Haeb-Umbach, University of Paderborn, Germany; Bhiksha Raj, Carnegie Mellon University, United States

LowZero.4: NMF-BASED KEYWORD LEARNING FROM SCARCE DATA
Bart Ons, Jort F. Gemmeke, Hugo Van hamme, Katholieke Universiteit Leuven, Belgium

LowZero.5: DEEP MAXOUT NETWORKS FOR LOW-RESOURCE SPEECH RECOGNITION
Yajie Miao, Florian Metze, Shourabh Rawat, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, United States

LowZero.6: COMBINATION OF DATA BORROWING STRATEGIES FOR LOW-RESOURCE LVCSR
Yanmin Qian, Kai Yu, Shanghai Jiao Tong University, China; Jia Liu, Tsinghua University, China

LowZero.7: FIXED-DIMENSIONAL ACOUSTIC EMBEDDINGS OF VARIABLE-LENGTH SECTIONS IN LOW-RESOURCE SETTINGS
Keith Levin, Johns Hopkins University, United States; Katharine Henry, University of Chicago, United States; Aren Jansen, Johns Hopkins University, United States; Karen Livescu, Toyota Technological Institute at Chicago, United States

LowZero.8: USING PROXIES FOR OOV KEYWORDS IN THE KEYWORD SEARCH TASK
Guoguo Chen, Oguz Yilmaz, Jan Trmal, Daniel Povey, Sanjeev Khudanpur, Johns Hopkins University, United States

LowZero.9: SEARCH RESULTS BASED N-BEST HYPOTHESIS RESCORING WITH MAXIMUM ENTROPY CLASSIFICATION
Fuchun Peng, Scott Roy, Ben Shahshahani, Francoise Beaufays, Google, United States

LowZero.10: USING WEB TEXT TO IMPROVE KEYWORD SPOTTING IN SPEECH
Ankur Gandhe, Long Qin, Florian Metze, Alexander I. Rudnicky, Ian Lane, Carnegie Mellon University, United States; Matthias Eck, Mobile Technologies, United States

LowZero.11: MULTI-STREAM TEMPORALLY VARYING WEIGHT REGRESSION FOR CROSS-LINGUAL SPEECH RECOGNITION
Shilin Liu, Khe Chai Sim, National University of Singapore, Singapore

LowZero.12: DISCRIMINATIVE SEMI-SUPERVISED TRAINING FOR KEYWORD SEARCH IN LOW RESOURCE LANGUAGES
Roger Hsiao, Tim Ng, Raytheon BBN Technologies, United States; Frantisek Grezl, Brno University of Technology, Czech Republic; Damianos Karakos, Stavros Tsakalidis, Long Nguyen, Richard Schwartz, Raytheon BBN Technologies, United States

LowZero.13: PROBABILISTIC LEXICAL MODELING AND UNSUPERVISED TRAINING FOR ZERO-RESERVED ASR
Ramya Rapisidou, Marzieh Razavi, Idiap Research Institute, École polytechnique fédérale de Lausanne, Switzerland; Mathew Magimai Doss, Idiap Research Institute, Switzerland
LowZero.14: LIGHTLY SUPERVISED AUTOMATIC SUBTITLING OF WEATHER ..........................452
FORECASTS
Joris Driesen, Steve Renals, University of Edinburgh, United Kingdom

LowZero.15: UNSUPERVISED WORD SEGMENTATION FROM NOISY INPUT ...............................458
Jahn Heymann, Oliver Walter, Reinhold Haeb-Umbach, University of Paderborn, Germany; Bhiksha Raj,
Carnegie Mellon University, United States

LowZero.16: AN EMPIRICAL STUDY OF CONFUSION MODELING IN KEYWORD ......................464
SEARCH FOR LOW RESOURCE LANGUAGES
Murat Saraclar, IBM / Bogazici University, United States; Abhinav Sethy, Bhuvana Ramabhadran, Lidia
Mangu, Jia Cui, Xiaodong Cui, Brian Kingsbury, IBM, United States; Jonathan Mamou, IBM Haifa Research
Labs, Israel

LowZero.17: SEMI-SUPERVISED BOOTSTRAPPING APPROACH FOR NEURAL ..........................470
NETWORK FEATURE EXTRACTOR TRAINING
Frantisek Grezl, Martin Karafiat, Brno University of Technology, Czech Republic