2014 9th International Symposium on Chinese Spoken Language Processing

(ISCSLP 2014)

Singapore
12-14 September 2014
ORAL SESSIONS
<table>
<thead>
<tr>
<th>Oral Session A1</th>
<th>Deep Neural Networks in Speech Recognition - I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 10:30 - 12:10 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Lin-shan Lee</td>
</tr>
<tr>
<td>Venue</td>
<td>Room A</td>
</tr>
</tbody>
</table>

**A1-1: Speaker Adaptation of Hybrid NN/HMM Model for Speech Recognition Based on Singular Value Decomposition**  
Shaofei Xue, Hui Jiang and Lirong Dai

**A1-2: Deep Neural Network Acoustic Modeling For Native and Non-Native Mandarin Speech Recognition**  
Xin Chen and Jian Cheng

**A1-3: Labeling Unsegmented Sequence Data with DNN-HMM and Its Application for Speech Recognition**  
Xiangang Li and Xihong Wu

**A1-4: Mandarin Speech Recognition Using Convolution Neural Network with Augmented Tone Features**  
Xinhui Hu, Xugang Lu and Chiori Hori

**A1-5: Research on Deep Neural Network’s Hidden Layers in Phoneme Recognition**  
Yuan Ma, Jianwu Dang and Weifeng Li

<table>
<thead>
<tr>
<th>Oral Session A2</th>
<th>Language Modelling and Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 13:30 - 15:30 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Chung Hsien Wu</td>
</tr>
<tr>
<td>Venue</td>
<td>Room A</td>
</tr>
</tbody>
</table>

**A2-1: Joint-Character-PON N-Gram Language Modeling for Chinese Speech Recognition**  
Bin Wang, Zhijian Ou, Jian Li and Akinori Kawamura

**A2-2: Linear Model Incorporating Feature Ranking for Chinese Documents Readability**  
Gang Sun, Zhiwei Jiang, Qing Gu and Daoxu Chen

**A2-3: Rapid Bayesian Learning for Recurrent Neural Network Language Model**  
Jen-Tzung Chien, Yuan-Chu Ku and Mou-Yue Huang

**A2-4: Minimum Classification Error Rate Training of Supervised Topic Mixture Model for Multi-label Text Categorization**  
Zhiyang He, Ping Lv and Ji Wu
A2-5: Investigation of Using Different Chinese Word Segmentation Standards and Algorithms for Automatic Speech Recognition  
Chongjia Ni and Cheung-Chi Leung

A2-6: Deep Belief Network based CRF for Spoken Language Understanding  
Xiaohao Yang and Jia Liu

Oral Session A3  
Speaker Recognition

Date/Time  
Friday, 12 September 2014 / 16:00 - 18:00 Hrs

Chair  
Thomas Zheng

Venue  
Room A

A3-1: Local Variability Vector for Text-Independent Speaker Verification  
Liping Chen, Kong Aik Lee, Bin Ma, Wu Guo, Haizhou Li and Li Rong Dai

A3-2: Single-sided Approach to Discriminative PLDA Training for Text-Independent Speaker Verification without Using Expanded i-vector  
Ikuya Hirano, Kong Aik Lee, Zhaofeng Zhang, Longbiao Wang and Atsuhiko Kai

A3-3: Relevance Vector Machines with Empirical Likelihood-Ratio Kernels for PLDA Speaker Verification  
Wei Rao and Man-Wai Mak

A3-4: Data-driven Tree Structure Based UBM Reconstruction for Speaker Verification  
Rong Zheng and Bo Xu

A3-5: An iVector Extractor Using Pre-trained Neural Networks for Speaker Verification  
Shanshan Zhang, Rong Zheng and Bo Xu

A3-6: An Iterative Framework for Unsupervised Learning in the PLDA based Speaker Verification  
Wenbo Liu, Zhiding Yu and Ming Li
### Oral Session A4: Speech Recognition - I

**Date/Time**
Saturday, 13 September 2014 / 10:00 - 12:00 Hrs

**Chair**
Chin-Hui Lee

**Venue**
Room A

1. **A4-1: Speaker Adaptive Bottleneck Features Extraction for LVCSR Based on Discriminative Learning of Speaker Codes**  
   Changqing Kong, Shaofei Xue, Jianqing Gao, Wu Guo, Lirong Dai and Hui Jiang

2. **A4-2: Error-Driven Pronunciation Dictionary Construction for Mandarin Speech Recognition**  
   Yi Liu, Xiangang Li and Xihong Wu

3. **A4-3: Multilevel Sampling and Aggregation for Discriminative Training**  
   Yunxin Zhao, Tuo Zhao and Xin Chen

   Tuo Zhao, Yunxin Zhao and Xin Chen

5. **A4-5: Modeling Inter-cluster and Intra-cluster Discrimination among Triphones**  
   Tom Ko, Brian Mak and Dongpeng Chen

6. **A4-6: Robust Voice Activity Detection Based on Concept of Modulation Transfer Function in Noisy Reverberant Environments**  
   Shota Morita, Masashi Unoki, Xugang Lu and Masato Akagi

---

### Oral Session A5: Deep Neural Networks in Speech Recognition - II

**Date/Time**
Sunday, 14 September 2014 / 10:00 - 12:00 Hrs

**Chair**
Li Deng

**Venue**
Room A

1. **A5-1: TANDEM-Bottleneck Feature Combination using Hierarchical Deep Neural Networks**  
   Mirco Ravanelli, Van Hai Do and Adam Janin

2. **A5-2: A General Framework for Multi-Accent Mandarin Speech Recognition Using Adaptive Neural Networks**  
   Xiang Sui, Huiyong Wang and Lan Wang

3. **A5-3: Decision Tree based State Tying for Speech Recognition using DNN Derived Embeddings**  
   Xiangang Li and Xihong Wu
A5-4: Acoustic Emotion Recognition using Deep Neural Network  
Jianwei Niu, Yanmin Qian and Kai Yu

A5-5: Convolutional Maxout Neural Networks for Low-Resource Speech Recognition  
Meng Cai, Yongzhe Shi, Jian Kang, Jia Liu and Tengrong Su

A5-6: Multiple Time-Span Feature Fusion for Deep Neural Network Modeling  
Chongjia Ni, Nancy Chen and Bin Ma

Oral Session A6  
Speaker and Language Recognition

Date/Time  
Sunday, 14 September 2014 / 13:30 - 15:30 Hrs

Chair  
Man Wai Mak

Venue  
Room A

A6-1: Performance Evaluation of Deep Bottleneck Features for Spoken Language Identification  
Bing Jiang, Yan Song, Si Wei, Meng-Ge Wang, Ian McLoughlin and Li-Rong Dai

A6-2: Discriminative Boosting Regression Backend for Phonotactic Language Recognition  
Wei-wei Liu, Wei-Qiang Zhang and Jia Liu

A6-3: Phonotactic Language Recognition Based on DNN-HMM Acoustic Model  
Wei-wei Liu, Meng Cai, Hua Yuan, Xiao-Bei Shi, Wei-Qiang Zhang and Jia Liu

A6-4: A Fusion Approach to Spoken Language Identification Based on Combining Multiple Phone Recognizers and Speech Attribute Detectors  
Yannan Wang, Jun Du, Lirong Dai and Chin-Hui Lee

A6-5: A New Fast and Memory Effective I-Vector Extraction Based on Factor Analysis of KLD Derived GMM Supervector  
Zhi-Yi Li, Wei-Qiang Zhang, Yao Tian and Jia Liu

A6-6: Improved Multitaper PNCC Feature for Robust Speaker Verification  
Yi Liu, Liang He and Jia Liu
### Oral Session B1

**Keyword Search and Spoken Language Application**

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Friday, 12 September 2014 / 10:30 - 12:10 Hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Hsin-Min Wang</td>
</tr>
<tr>
<td>Venue</td>
<td>Room B</td>
</tr>
</tbody>
</table>

#### B1-1: Personalized Video Summarization Based on Multi-layered Probabilistic Latent Semantic Analysis with shared Topics
Cheng-Tao Chung, Hsin-Kuan Hsiung, Cheng-Kuang Wei and Lin-shan Lee

#### B1-2: Interlocutor Personality Perception based on BFI Profiles and Coupled HMMs in a Dyadic Conversation
Ming-Hsiang Su, Yu-Ting Zheng and Chung-Hsien Wu

#### B1-3: The Vietnamese Speech Recognition Based on Rectified Linear Units Deep Neural Network and Spoken Term Detection System Combination
Shifu Xiong, Wu Guo and Diyuan Liu

#### B1-4: Improving Keyword Search by Query Expansion in a Probabilistic Framework
Zhipeng Chen, Zhiyang He, Ping Lv and Ji Wu

#### B1-5: A Novel Keyword+LVCSR-Filler Based Grammar Network Representation for Spoken Keyword Search
I-Fan Chen, Chongjia Ni, Boon Pang Lim, Nancy F. Chen and Chin-Hui Lee

### Oral Session B2

**Speech Synthesis and Voice Conversion**

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Friday, 12 September 2014 / 13:30 -15:30 Hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Minghui Dong</td>
</tr>
<tr>
<td>Venue</td>
<td>Room B</td>
</tr>
</tbody>
</table>

#### B2-1: Pitch Transformation in Neural Network based Voice Conversion
Feng-Long Xie, Yao Qian, Frank K. Soong and Haifeng Li

#### B2-2: Integrating Global Variance of Log Power Spectrum Derived from LSPs into MGE Training for HMM-Based Parametric Speech Synthesis
Yu-Sheng Sun, Zhen-Hua Ling, Xiang Yin and Li-Rong Dai

Jingjie Li, Lan Vince McLoughlin and Yan Song

#### B2-4: Correlation-based Frequency Warping for Voice Conversion
Xiaohai Tian, Zhizheng Wu, S. W. Lee and Eng Siong Chng
B2-5: Automatic Speech Data Clustering with Human Perception based Weighted Distance  
Xixin Wu, Zhiyong Wu, Jia Jia, Helen Meng, Lianhong Cai and Weifeng Li

B2-6: Frame Correlation Based Autoregressive GMM Method for Voice Conversion  
Xian Li and Zeng-fu Wang

<table>
<thead>
<tr>
<th>Oral Session B3</th>
<th>Dialogue System and Language Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 16:00 - 18:00 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Kai Yu</td>
</tr>
<tr>
<td>Venue</td>
<td>Room B</td>
</tr>
</tbody>
</table>

B3-1: An Ontology Semantic Tree based Natural Language Interface  
Shusen Li, Zhiyang He and Jia Wu

Xiaohao Yang, Zhenfeng Chen and Jia Liu

B3-3: An Experimental Comparative Study on Prosodic Features between Ningbo EFL Learners and American Native Speakers-in the Case of Production of Yes-No Questions  
Dongxia Qian, Yuan Jia, Aijun Li and Liang Xu

B3-4: Cross-language Comparison of F0 Range in Speakers of Native Chinese, Native Japanese and Chinese L2 of Japanese: Preliminary Results of a Corpus-based Analysis  
Shuju Shi, Jinsong Zhang and Yanlu Xie

B3-5: A New Neural Network Based Logistic Regression Classifier For Improving Mispronunciation Detection of L2 Language Learners  
Wenping Hu, Yao Qian and Frank Soong

B3-6: Mispronunciation Detection and Diagnosis in L2 English Speech Using Multi-Distribution Deep Neural Networks  
Kun Li and Helen Meng
<table>
<thead>
<tr>
<th>Oral session B4</th>
<th>Speech Prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date/Time</strong></td>
<td>Saturday, 13 September 2014 / 10:00 - 12:00 Hrs</td>
</tr>
<tr>
<td><strong>Chair</strong></td>
<td>Chiu-yu Tseng</td>
</tr>
<tr>
<td><strong>Venue</strong></td>
<td>Room B</td>
</tr>
</tbody>
</table>

**B4-1: The Perception of Mandarin Tones by Learners from Heritage and Non-heritage Backgrounds**
Kimiko Tsukada, Hui Ling Xu and Nan Xu Rattanasone

**B4-2: A Preliminary Research on Rhetorical Structural and Prosodic Features in Chinese Reading Texts**
Liang Zhang, Yuan Jia and Aijun Li

**B4-3: Superpositional HMM-based Intonation Synthesis using a Functional F0 Model**
Jinfu Ni, Yoshinori Shiga and Chiori Hori

**B4-4: Improving F0 Prediction Using Bidirectional Associative Memories and Syllable-Level F0 Features for HMM-based Mandarin Speech Synthesis**
Li Gao, Zhen-Hua Ling, Ling-Hui Chen and Li-Rong Dai

**B4-5: The Power of Special Characters in Prosodic Word Prediction for Chinese TTS**
Zhengchen Zhang and Minghui Dong

**B4-6: Learning Model-based F0 Production through Goal-directed Babbling**
Hao Liu and Yi Xu

<table>
<thead>
<tr>
<th>Oral Session B5</th>
<th>Speech Perception and Production</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date/Time</strong></td>
<td>Sunday, 14 September 2014 / 10:00 - 12:00 Hrs</td>
</tr>
<tr>
<td><strong>Chair</strong></td>
<td>Aijun Li</td>
</tr>
<tr>
<td><strong>Venue</strong></td>
<td>Room B</td>
</tr>
</tbody>
</table>

**B5-1: Effects of Preceding Contexts on the Categorical Perception of Mandarin Tones**
Fei Chen
Kunyu Xu and Gang Peng

**B5-2: A New Framework of Neurocomputational Model for Speech Production**
Han Yan
Jianwu Dang, Mengxue Cao and Bernd J. Kröger

**B5-3: A Multi-channel/Multi-speaker Articulatory Database in Chinese Mandarin for Speech Visualization and Acoustic-to-articulatory Mapping**
Dan Zhang Xiaqian Liu Nan Yan Lan Wang Yun Zhu and Hui Chen
B5-4: Surface Electromyographic Activity of Non-Laryngeal Neck Muscles in Cantonese Tone Production
Shing Yu
Tan Lee and Manwa L. Ng

B5-5: Novel Approach for Estimating Length of the Vocal Folds using Fujisaki Model
Tanvina Patel and Hemant Arjun Patil

B5-6: Tone confusion in spoken and whispered Mandarin Chinese
Ian McLoughlin
Yan Xu and Yan Song

<table>
<thead>
<tr>
<th>Oral Session B6</th>
<th>Speech Analysis and Enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Sunday, 14 September 2014 / 13:30 - 15:30 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Tan Lee</td>
</tr>
<tr>
<td>Venue</td>
<td>Room B</td>
</tr>
</tbody>
</table>

B6-1: Spectral Patch Based Sparse Coding for Acoustic Event Detection
Xugang Lu, Yu Tsao, Peng Shen and Chiori Hori

B6-2: Multidimensional Acoustic Analysis for Voice Quality Assessment based on the GRBAS Scale
Ping Yu, Zhijian Wang, Shanshan Liu, Nan Yan, Lan Wang and Manwa Ng

B6-3: Investigation on Articulatory and Acoustic Characteristics of Dysarthria
Chengran Zhang, Jianwu Dang, Jianan Zhang and Jianguo Wei

B6-4: Multipitch Tracking Based on Linear Programming Relaxation and Parsity-Based Pitch Candidate Estimation
Feng Huang and Tan Lee

B6-5: Cross-language Transfer Learning for Deep Neural Network Based Speech Enhancement
Yong Xu, Jun Du, Lirong Dai and Chin-Hui Lee

B6-6: Speech Separation Based on Improved Deep Neural Networks with Dual Outputs of Speech Features for Both Target and Interfering Speakers
Yanhui Tu, Jun Du, Yong Xu, Lirong Dai and Chin-Hui Lee
NCMMSC Session | Emotional Speech Processing
---|---
C1 | 
**Date/Time**: Friday, 12 September 2014 / 10:30 - 12:10 Hrs

**Chair**: Jianhua Tao

**Venue**: Room C

**C1-1**: Improving Generation Performance of Speech Emotion Recognition by Denoising Autoencoders  
Linlin Chao, Jianhua Tao, Minghao Yang and Ya Li

**C1-2**: Survey on Discriminative Feature Selection For Speech Emotion Recognition  
Xin Xu, Ya Li, Xiaoying Xu, Zhengqi Wen, Hao Che, Shanfeng Liu and Jianhua Tao

**C1-3**: The Emotion Recognition from Uyghur Sentences Based on Combination of Class Discriminating Words and Sentiment Dictionary  
Abdusalam Dawut, Hussein Yusuf and Askar Hamdulla

**C1-4**: Performance Analysis of Different Keyword Extraction Algorithms for Emotion Recognition from Uyghur Text  
Seyyare Imam, Rayilam Parhat, Askar Hamdulla and Zhijun Li

**C1-5**: Study of Pitch of “Dearing” as Emotional Speech  
Youran Lin and Jiangping Kong

**C1-6**: The Expression of Emotions by Text and Speech  
Xiaoying Xu, Huimin Wang, Ya Li, Wei Lai and Jianhua Tao

NCMMSC Session | Multimodal Observation and Analysis for Speech Production
---|---
C2 | 
**Date/Time**: Friday, 12 September 2014 / 13:30 - 15:30 Hrs

**Chair**: Jianwu Dang

**Venue**: Room C

**C2-1**: A Mass-spring Tongue Model with Efficient Collision Detection and Response During Speech  
Rui Li, Jun Yu, Chen Jiang, Changwei Luo and Zengfu Wang

**C2-2**: The Modeling of Tongue Tip in Standard Chinese using MRI  
Gaowu Wang, Jianwu Dang and Jiangping Kong

**C2-3**: The Chest and Abdomen Breathing in Reading literature in Mandarin  
Feng Yang and Jiangping Kong
C2-4: The Effects of Focal Stress on the Articulatory and Acoustic Properties of Segments in Standard Chinese
Yinghao Li and Jiangping Kong

C2-5: Definition and Extraction of Lip Protrusion Based on the Facial Skeleton Data
Xiaosheng Pan, Menghan Zhang and Wee Chung Liew

C2-6: Visualization of Mandarin Articulation Driven by Ultrasound Data
Jianan Zhang, Jianguo Wei, Chengran Zhang, Dian Huang and Jianwu Dang

C2-7: Correlations between Vocal Tract Parameters and Body Heights in Adult Humans
Honglin Cao and Jiangping Kong

C2-8: A Novel 3D Geometric Articulatory Model
Qiang Fang, Jianguo Wei, Wenhuan Lu, Jie Liu and Chan Song

<table>
<thead>
<tr>
<th>NCMMSC Session</th>
<th>(A) Multimodal Observation and Analysis for Speech Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>(B) Front-end Processing for Distant-talking Speech Recognition</td>
</tr>
<tr>
<td>Chair</td>
<td>Jianwu Dang</td>
</tr>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 16:00 - 18:00 Hrs</td>
</tr>
<tr>
<td>Venue</td>
<td>Room C</td>
</tr>
</tbody>
</table>

C3A-1: Mapping Between Ultrasound and Vowel Speech using DNN Framework
Xinyuan Zheng, Jianguo Wei, Wenhuan Lu, Qiang Fang and Jianwu Dang

C3A-2: Automatic Speech Recognition under Robot Ego Noises
Jianrong Wang, Ju Zhang, Jianguo Wei, Wenhuan Lu and Jianwu Dang

C3B-3: Realizing Speech Enhancement by Combining EEMD and K-SVD Dictionary Training Algorithm
Hao Chen, Zhenye Gan and Hongwu Yang

C3B-4: Single-channel Dereverberation for Distant-talking Speech Recognition by Combining Denoising Autoencoder and Temporal Structure Normalization
Yuma Ueda, Longbiao Wang, Atsuhioko Kai, Xiong Xiao, Eng Siong Chng and Haizhou Li

C3B-5: Distant-Talking Speech Recognition using Multi-Channel LMS and Multiple-Step Linear Prediction
Satoshi Shiota, Longbiao Wang, Kyohei Odani, Atsuhioko Kai and Weifeng Li
C3B-6: Speech Enhancement via Low-rank Matrix Decomposition and Image Based Masking
Liyang Liu, Zhaogui Ding, Weifeng Li, Longbiao Wang and Qingmin Liao

C3B-7: Experimental Study on Dereverberation and Noise Reduction for Distant Speech Recognition
Zhong-Hua Fu, Lei Xie and Hang Lv

<table>
<thead>
<tr>
<th>NCMMSC Session</th>
<th>(A) Computational Audio/Speech Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4</td>
<td></td>
</tr>
<tr>
<td>Chair</td>
<td>Lei Xie</td>
</tr>
<tr>
<td>(B) Speech Prosody and Language Modeling for Agglutinative Languages</td>
<td></td>
</tr>
<tr>
<td>Chair</td>
<td>Askar</td>
</tr>
<tr>
<td>Date/Time</td>
<td>Saturday, 13 September 2014 / 10:00 - 12:00 Hrs</td>
</tr>
<tr>
<td>Venue</td>
<td>Room C</td>
</tr>
</tbody>
</table>

C4A-1: Algorithm of Pure Tone Audiometry based on Multiple Judgment
Yuhao Wu, Jia Jia, Xiulong Zhang and Lianhong Cai

C4A-2: An Improved Pitch Extraction Algorithm for Speech Processing
Xiao Chen and Bo Xu

C4A-3: A Robust High Resolution Speaker DOA Estimation under Reverberant Environment
Yifan Guo, Y.X. Zou and Yongqing Wang

C4A-4: A Hybrid Virtual Bass System with Improved Phase Vocoder and High Efficiency
Shaofei Zhang, Lei Xie, Zhong-Hua Fu and Yougen Yuan

C4B-5: An Electropalatographic and Electroglottographic Study on Domain-Initial Strengthening in Korean
Yinghao Li and Jinghua Zhang

C4B-6: Multilayer Structure based Lexicon Optimization for Agglutinative Languages
Ablimit Mijit, Akbar Pattar and Askar Hamdulla

C4B-7: Prosody Modeling for Uyghur TTS
Gulmire Imam, Guljamal Mamateli, Maynur Ablitip and Askar Hamdulla

C4B-8: Document Classification based on Word Vectors
Rong Liu, Dong Wang and Chao Xing
<table>
<thead>
<tr>
<th>NCMMMSC Session</th>
<th>Robust Speaker Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5</td>
<td>Date/Time</td>
</tr>
<tr>
<td></td>
<td>Sunday, 14 September 2014 / 10:00 - 12:00 Hrs</td>
</tr>
<tr>
<td></td>
<td>Chair</td>
</tr>
<tr>
<td></td>
<td>Thomas Zheng</td>
</tr>
<tr>
<td></td>
<td>Venue</td>
</tr>
<tr>
<td></td>
<td>Room C</td>
</tr>
</tbody>
</table>

- **C5-1: Multi-Scale Kernels for Short Utterance Speaker Recognition**
  Wei-Qiang Zhang, Junhong Zhao, Wen-Lin Zhang and Jia Liu

- **C5-2: Speaker Verification Based on SVM and Total Variability**
  Sheng Zhang, Jie Xu, Guoping Hu, Wu Guo and Xiaokong Ma

- **C5-3: Speaker Verification using Fisher Vector**
  Yao Tian, Liang He, Zhi-Yi Li, Wei-Lan Wu, Wei-Qiang Zhang and Jia Liu

- **C5-4: Research on Generalization Property of Time-Varying Fbank-Weighted MFCC for i-Vector based Speaker Verification**
  Jun Wang, Lantian Li, Dong Wang and Thomas Fang Zheng

- **C5-5: Score Regulation based on GMM Token Ratio Similarity for Speaker Recognition**
  Yingchun Yang and Licai Deng

- **C5-6: Research on Truncated Speech in Speaker Verification**
  Fanhu Bie, Dong Wang and Thomas Fang Zheng

---

<table>
<thead>
<tr>
<th>NCMMMSC Session</th>
<th>Speech and Language Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6</td>
<td>Date/Time</td>
</tr>
<tr>
<td></td>
<td>Sunday, 14 September 2014 / 13:30 - 15:30 Hrs</td>
</tr>
<tr>
<td></td>
<td>Chair</td>
</tr>
<tr>
<td></td>
<td>Aijun Li</td>
</tr>
<tr>
<td></td>
<td>Venue</td>
</tr>
<tr>
<td></td>
<td>Room C</td>
</tr>
</tbody>
</table>

- **C6-1: The Undulating Scale of Intonations of Exclamatory Sentences in Uyghur from the view of Experimental Phonetics**
  Hankiz Yilahun, Gulmire Imam, Maynur Ablitip, Guljamal Mamateli and Askar Hamdulla

- **C6-2: A Proficient Trilingual’s Production of Sibilant Fricatives of Mandarin Chinese, Korean and English**
  Jinghua Zhang and Yinghao Li

- **C6-3: GSOM-based Modeling Study on Phoneme Acquisition**
  Mengxue Cao, Aijun Li and Qiang Fang
C6-4: Influences of Vowels on Perception of Nasal Codas in Mandarin for Japanese Learners and Chinese
Zuyan Wang and Jinsong Zhang

C6-5: Automatic Mispronunciation Detection for Mandarin Chinese based on Articulation Place and Articulation Manner
Richeng Duan, Jinsong Zhang, Yanlu Xie and Wen Cao

C6-6: The Training of The Tone of Mandarin Two-Syllable Words based on Pitch Projection Synthesis Speech
Yanlu Xie, Bei Zhang and Jinsong Zhang

C6-7: The Text Analysis and Processing of Thai Language Text to Speech Conversion System
Xuee Lin, Jian Yang and Juan Zhao
POSTER SESSIONS
<table>
<thead>
<tr>
<th>Poster Session</th>
<th>Speech Recognition - II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 13:30 - 15:30 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Dong Wang</td>
</tr>
<tr>
<td>Venue</td>
<td>Topaz Foyer</td>
</tr>
</tbody>
</table>

**P2-1: A Low Complexity Cluster Model Interpolation based On-Line Adaptation Technique for Spoken Query Systems**  
S Shahnawazuddin and Rohit Sinha

**P2-2: Corpus and Transcription System of Chinese Lecture Room**  
Sheng Li, Yuya Akita and Tatsuya Kawahara

**P2-3: Improving Training Time of Deep Neural Network With Asynchronous Averaged Stochastic Gradient Descent**  
Zhao You and Bo Xu

**P2-4: Investigation of Stochastic Hessian-Free Optimization In Deep Neural Networks For Speech Recognition**  
Zhao You and Bo Xu

**P2-5: Acoustic Feature Conversion using a Polynomial Based Feature Transferring Algorithm**  
Syu-Siang Wang, Payton Lin, Dau-Cheng Lyu, Yu Tsao, Hsin-Te Hwang and Borching Su

**P2-6: Recurrent Neural Network Language Model with Part-of-speech for Mandarin Speech Recognition**  
Caixia Gong, Xiangang Li and Xihong Wu

**P2-7: Effectiveness of Fractal Dimension for ASR in Low Resource Language**  
Mohammadi Zaki, Nirmesh J. Shah and Hemant A. Patil

**P2-8: Unsupervised Acoustic Model Training for the Korean Language**  
Antoine Laurent, William Hartmann and Lori Lamel

**P2-9: Cross-language Speech Attribute Detection and Phone Recognition for Tibetan Using Deep Learning**  
Hui Wang, Yue Zhao, Yanmin Xu, Xiaona Xu, Xingmei Suo and Qiang Ji

**P2-10 Speech Emotion Recognition Based on Wavelet Packet Coefficient Model**  
Kunxia Wang, Ning An and Lian Li
<table>
<thead>
<tr>
<th>Poster Session</th>
<th>Prosody and Speech Synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date/Time</td>
<td>Friday, 12 September 2014 / 16:00 - 18:00 Hrs</td>
</tr>
<tr>
<td>Chair</td>
<td>Jinsong Zhang</td>
</tr>
<tr>
<td>Venue</td>
<td>Topaz Foyer</td>
</tr>
</tbody>
</table>

**P3-1: Improving Segmental GMM Based Voice Conversion Method with Target Frame Selection**  
Hung-Yan Gu and Sung-Fung Tsai

**P3-2: Investigation of Social Media on Depression**  
Wei Tong Mok, Rachael Sing, Xituting Jiang and Swee Lan See

**P3-3: The Typology of Focus Realization of Northern Mandarin**  
Wenjun Duan and Yuan Jia

**P3-4: Combining Prosodic and Spectral Features for Mandarin Intonation Recognition**  
Wei Bao, Ya Li, Mingliang Gu, Jianhua Tao, Linlin Chao and Shanfeng Liu

**P3-5: Investigating Effect of Rich Syntactic Features on Mandarin Prosodic Phrase Boundaries Prediction**  
Hao Che, Zhengqi Wen, Ya Li and Jianhua Tao

**P3-6: Context Features Based Unit Selection and Weight Prediction In Concatenation Speech Synthesis System**  
Shanfeng Liu, Zhengqi Wen, Ya Li, Jianhua Tao and Bin Liu

**P3-7: A Speaker Adaptation of Speaking Rate-dependent Hierarchical Prosodic Model for Mandarin TTS**  
Po-Chun Wang, I-Bin Liao, Chen-Yu Chiang, Yih-Ru Wang and Sin-Horng Chen

**P3-8: Evaluation of Parameter Generation Using High Order Dynamic Features and Long Span Windows for HMM based Speech Synthesis**  
Yang Wang and Jianhua Tao

**P3-9: Fusion of Magnitude and Phase-based Features for Objective Evaluation of TTS Voice**  
Hardik Sailor and Hemant Arjun Patil

**P3-10: Deterministic Annealing EM Algorithm for Developing TTS System in Gujarati**  
Nirmesh J. Shah, Hemant A. Patil, Maulik C. Madhavi, Hardik B. Sailor and Tanvina B. Patel
<table>
<thead>
<tr>
<th>Poster Session</th>
<th>Speech and Audio Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td></td>
</tr>
</tbody>
</table>

**Date/Time**  
Saturday, 13 September 2014 / 10:00 - 12:00 Hrs

**Chair**  
Zhonghua Fu

**Venue**  
Topaz Foyer

- **P4-1: Efficient Voice Activity Detection Algorithm based On Sub-Band Temporal Envelope and Sub-Band Long-Term Signal Variability**  
  Bin Liu, Jianhua Tao, Fuyuan Mo, Ya Li, Zhengqi Wen and Liu Shanfeng

- **P4-2: Correlations Between Body Heights And Formant Frequencies in Young Male Speakers: A Pilot Study**  
  Honglin Cao, Yingli Wang and Jiangping Kong

- **P4-3: Classification of Pathological Infant Cries using Modulation Spectrogram Features**  
  Anshu Chittora and Hemant Arjun Patil

- **P4-4: Exploiting Speech Source Information for Vowel Landmark Detection for Low Resource Language**  
  Ankur Undhad, Hemant Arjun Patil and Maulik Madhavi

- **P4-5: Effect of Vocoder Type to Mandarin Speech Recognition in Cochlear Implant Simulation**  
  Fei Chen and Ada H.Y. Lau

- **P4-6: Speech Analysis Method Based on Source-Filter Model Using Multivariate Empirical Mode Decomposition in Log-Spectrum Domain**  
  Surasak Boonkla, Masashi Unoki, Stanislav S. Makhanov and Chai Wutiwiwatchai

- **P4-7: Signal to noise ratio estimation based on an optimal design of subband voice activity detection**  
  Shota Morita, Xugang Lu and Masashi Unoki

- **P4-8: Soft Constrained Leading Voice Separation with Music Score Guidance**  
  Renbo Zhao, Siu-Wa Lee, Dongyan Huang and Minghui Dong

- **P4-9: Using hierarchical method to improve real time for audio-based surveillance system**  
  Aiying Zhang

- **P4-10: A Non-uniformly Distributed Three-microphone Array for Speech Enhancement in Directional and Diffuse Noise Field**  
  Chung-Chien Hsu, Kah-Meng Cheong and Tai-Shih Chi
Poster Session | Language, Pronunciation, Speaker and Emotion Processing
---|---
Date/Time | Sunday, 14 September 2014 / 10:00 - 12:00 Hrs
Chair | Dongyan Huang
Venue | Topaz Foyer

**P5-1: Speech Emotion Classification using Acoustic Features**
Shizhe Chen, Qin Jin, Xirong Li, Jieping Xu and Gang Yang

**P5-2: Acoustic Emotion Recognition based on Fusion of Multiple Features-Dependent Deep Boltzmann Machines**
Kelvin Poon Feng, Dongyan Huang, Minghui Dong and Haizhou Li

**P5-3: Speech Based Emotion Recognition Using Spectral Feature Extraction and an Ensemble of kNN Classifiers**
Steven Rieger, Rajani Muralleedharan and Ravi Ramachandran

**P5-4: The Discrimination of z-zh c-ch s-sh by Proficient Speakers of Chinese as Second Language**
Li Mei and Jing Zhu

**P5-5: A Study on the Long-term Retention Effects of Japanese C2L Learners to Distinguish Chinese Tone 2 and Tone 3 After Perceptual Training**
Xiaoli Feng, Yue Sun and Jiinsong Zhang

**P5-6: An Effective and Robust Approach to Mandarin Spoken Language Understanding in Specific Domain**
Zhiyang He, Ping Lv and Ji Wu

**P5-7: A Bottom-Up Kernel of Pattern Learning for Relation Extraction**
Chunyun Zhang, Weiran Xu, Jun Guo and Sheng Gao

**P5-8: Global Discriminative Model for Dependency Parsing in NLP Pipeline**
Miao Li, Hongyi Ding and Ji Wu

**P5-9: Fusion of SNR-Dependent PLDA Models for Noise Robust Speaker Verification**
Xiaomin Pang and Manwai Mak

**P5-10 Exploiting Variable Length Teager Energy Operator in Melcepstral Features for Person Recognition from Humming**
Maulik Madhavi and Hemant Arjun Patil

**P5-11 Psychoacoustic Model Compensation with Robust Feature Set For Speaker Verification in Additive Noise**
Ashish Panda

**P5-12 Where and How to Make an Emphasis? - L2 Distinct Prosody and Why**
Chiu-yu Tseng and Chao-yu Su