

# **2010 7th International Symposium on Chinese Spoken Language Processing**

**(ISCSLP 2010)**

**Tainan, Vaiwan  
29 November – 3 December 2010**



**IEEE Catalog Number: CFP10583-PRT  
ISBN: 978-1-4244-6244-5**

## Table of Contents

<b>Automatic Transcription of Parliamentary Meetings and Classroom Lectures – A Sustainable Approach and Real System Evaluations –</b> <i>Tatsuya KAWAHARA</i>	<b>P.1</b>
<b>Intelligibility Investigation of Single-Channel Noise Reduction Algorithms for Chinese and Japanese</b> <i>Junfeng LI, Lin YANG, Yonghong YAN, Chau Duc THANH and Masato AKAGI</i>	<b>P.7</b>
<b>DCT-based Processing of Dynamic Features for Robust Speech Recognition</b> <i>Wen-Chi LIN, Hao-Teng FAN and Jieh-Weih HUNG</i>	<b>P.12</b>
<b>Speech Enhancement as a Functional Approximation and Generalization</b> <i>Xugang LU, Masashi UNOKI, Ryosuke ISOTANI, Hisashi KAWAI and Satoshi NAKAMURA</i>	<b>P.18</b>
<b>Spectral Trajectory Estimation Using Nonnegative Matrix Factorization for Model-Based Monaural Speech Separation</b> <i>Chun-Man MAK, Tan LEE and S.W. LEE</i>	<b>P.23</b>
<b>An Environment Structuring Framework to Facilitating Suitable Prior Density Estimation for MAPLR on Robust Speech Recognition</b> <i>Yu TSAO, Ryosuke ISOTANI, Hisashi KAWAI and Satoshi NAKAMURA</i>	<b>P.29</b>
<b>Dual-microphone Noise Reduction Based on Semi-Blind DUET</b> <i>Zhong-Hua FU, Lei XIE and Dong-Mei JIANG</i>	<b>P.33</b>
<b>Acoustic and Articulatory Analysis on Mandarin Chinese Vowels in Emotional Speech</b> <i>Aijun LI, Qiang FANG, Fang HU, Lu ZHENG, Hong WANG and Jianwu DANG</i>	<b>P.38</b>
<b>Effect of Speech Rate on Inter-segmental Coarticulation in Standard Chinese</b> <i>Ying-Hao LI, Jiang-Ping KONG</i>	<b>P.44</b>
<b>Discrimination between Natural and Unnatural Articulations based on Articulatory Structure</b> <i>Akikazu NISHIKIDO, Shin-ichi KAWAMOTO and Jianwu DANG</i>	<b>P.50</b>
<b>An Initial Investigation of L1 and L2 Discourse Speech Planning in English</b> <i>Chiu-Yu TSENG, Zhao-Yu SU, Chi-Feng HUANG and Tanya VISCEGLIA</i>	<b>P.55</b>
<b>Effects of Syllable Positions on Taiwanese Mandarin Sibilant Perception</b> <i>Chenhao CHIU and Molly BABEL</i>	<b>P.60</b>
<b>Toward a Comprehensive Vowel Space for Whispered Speech</b> <i>Hamid R. SHARIFZADEH, Ian V. MCLOUGHLIN and Martin J. RUSSELL</i>	<b>P.65</b>
<b>Detection of Intonation in L2 English Speech of Native Mandarin Learners</b> <i>Kun LI, Shuang ZHANG, Mingxing LI, Wai-Kit LO and Helen MENG</i>	<b>P.69</b>
<b>Improving the Informativeness of Verbose Queries Using Summarization Techniques for Spoken Document Retrieval</b> <i>Shih-Hsiang LIN, Berlin CHEN and Ea-Ee JAN</i>	<b>P.75</b>
<b>Forward Optimal Measures for Automatic Mispronunciation Detection</b> <i>Changliang LIU, Fuping PAN, Fengpei GE, Bin DONG and Yonghong YAN</i>	<b>P.80</b>

<b>Capturing L2 Segmental Mispronunciations with Joint-sequence Models in Computer-Aided Pronunciation Training (CAPT)</b>	<b>P.84</b>
<i>Xiaojun QIAN, Helen MENG and Frank SOONG</i>	
<b>A Novel Approach for Proper Name Transliteration Verification</b>	<b>P.89</b>
<i>Ea-Ee JAN, Niyu GE, Shih-Hsiang LIN, Salim ROUKOS and Jeffrey SORENSEN</i>	
<b>Aligning Singing Voice with MIDI Melody Using Synthesized Audio Signal</b>	<b>P.95</b>
<i>Minghui DONG, Paul CHAN, Ling CEN and Haizhou LI</i>	
<b>Minimum Generation Error Training for HMM-based Prediction of Articulatory Movements</b>	<b>P.99</b>
<i>Tian-Yi ZHAO, Zhen-Hua LING, Ming LEI, Li-Rong DAI and Qing-Feng LIU</i>	
<b>Mandarin-English Bilingual Phone Modeling and Combining MPE Based Discriminative Training for Cross-Language Speech Recognition</b>	<b>P.103</b>
<i>Yanmin QIAN and Jia LIU</i>	
<b>Subvector-quantized High-density Discrete Hidden Markov Model and its Re-estimation</b>	<b>P.109</b>
<i>Guoli YE and Brian MAK</i>	
<b>Problems of Modeling Phone Deletion in Conversational Speech for Speech Recognition</b>	<b>P.114</b>
<i>Brian MAK and Tom KO</i>	
<b>Speaker Adaptation of Stochastic Segment Models Using Maximum Likelihood Linear Regression</b>	<b>P.119</b>
<i>Hao CHAO and Wen-Ju LIU</i>	
<b>A Study of Large Vocabulary Speech Recognition Decoding Using Finite-State Graphs</b>	<b>P.123</b>
<i>Zhijian OU and Ji XIAO</i>	
<b>Rendering a Personalized Photo-Real Talking Head from Short Video Footage</b>	<b>P.129</b>
<i>Lijuan WANG, Wei HAN, Xiaojun QIAN, Frank K. SOONG</i>	
<b>Automatic Prosody Prediction and Detection with Conditional Random Field (CRF) Models</b>	<b>P.135</b>
<i>Yao QIAN, Zhizheng WU, Xuezhe MA and Frank SOONG</i>	
<b>Development of an Articulatory Visual-Speech Synthesizer to Support Language Learning</b>	<b>P.139</b>
<i>Ka-Ho WONG, Wai-Kim LEUNG, Wai-Kit LO and Helen MENG</i>	
<b>Statistical Modeling of Syllable-Level F0 Features for HMM-based Unit Selection Speech Synthesis</b>	<b>P.144</b>
<i>Zhen-Hua LING, Zhi-Guo WANG and Li-Rong DAI</i>	
<b>Modeling Prosody Patterns for Chinese Expressive Text-to-Speech Synthesis</b>	<b>P.148</b>
<i>Zhiyong WU, Lianhong CAI and Helen M. MENG</i>	
<b>A Method for Modeling and Generating Mandarin Tone Contour with Phrase Intonation Based on the Generation Process Model</b>	<b>P.153</b>
<i>Miaomiao WANG, Miaomiao WEN, Keikichi HIROSE and Nobuaki MINEMATSU</i>	
<b>The Description of iFlyTek Speech Lab System for NIST2009 Language Recognition Evaluation</b>	<b>P.157</b>
<i>Ying XU, Yan SONG, Yan-Hua LONG, Hai-Bing ZHONG and Li-Rong DAI</i>	

<b>UBM Data Selection for Effective Speaker Modeling</b>	<b>P.162</b>
<i>Chien-Lin HUANG and Haizhou LI</i>	
<b>Factor Analysis Based Spatial Correlation Modeling for Speaker Verification</b>	<b>P.166</b>
<i>Er-Yu WANG, Wu GUO, Li-Rong DAI, Kong-Aik LEE, Bin MA and Hai-Zhou LI</i>	
<b>Dialect-Based Speaker Classification Using Speaker-Invariant Dialect Features</b>	<b>P.171</b>
<i>Xuebin MA, Ruiyuan XU, Nobuaki MINEMATSU, Yu QIAO, Keikichi HIROSE and Aijun LI</i>	
<b>Using Cepstral and Prosodic Features for Chinese Accent Identification</b>	<b>P.177</b>
<i>Jue HOU, Yi LIU, Thomas Fang ZHENG, Jesper OLSEN and Jilei TIAN</i>	
<b>Speaker Verification Using Support Vector Machine with LLR-based Sequence Kernels</b>	<b>P.182</b>
<i>Yi-Hsiang CHAO, Wei-Ho TSAI and Hsin-Min WANG</i>	
<b>Phonetic Clustering Based Confidence Measure for Embedded Speech Recognition</b>	<b>P.186</b>
<i>Zhi-Guo WANG, Cong LIU, Hai-Kun WANG, Yu HU and Li-Rong DAI</i>	
<b>Audio Visual Speech Recognition Based on Multi-Stream DBN Models with Articulatory Features</b>	<b>P.190</b>
<i>Dong-Mei JIANG, Peng WU, Feng-Na WANG, Hichem SAHLI, and Werner VERHELST</i>	
<b>A Study on Functional Loads of Phonetic Contrasts under Context Based On Mutual Information of Chinese Text and Phonemes</b>	<b>P.194</b>
<i>Jinsong ZHANG, Wei LI, Yuxia HOU, Wen CAO and Ziyu XIONG</i>	
<b>A Study on Hakka and Mixed Hakka-Mandarin Speech Recognition</b>	<b>P.199</b>
<i>Tsai-Lu TSAI, Chen-Yu CHIANG, Hsiu-Min YU, Lieh-Shih LO, Yih-Ru WANG and Sin-Horng CHEN</i>	
<b>Auditory Front-ends for Noise-Robust Automatic Speech Recognition</b>	<b>P.205</b>
<i>Ja-Zang YEH and Chia-Ping CHEN</i>	
<b>Robust Speaker Localization in a Disturbance Noise Environment Using a Distributed Microphone System</b>	<b>P.209</b>
<i>Kook CHO, Takanobu NISHIURA and Yoichi YAMASHITA</i>	
<b>An Integrated Framework for Transcribing Mandarin-English Code-mixed Lectures with Improved Acoustic and Language Modeling</b>	<b>P.214</b>
<i>Ching-Feng YEH, Chao-Yu HUANG, Liang-Che SUN and Lin-Shan LEE</i>	
<b>Large Vocabulary Uyghur Continuous Speech Recognition based on Stems and Suffixes</b>	<b>P.220</b>
<i>Xin LI, Shang CAI, Yafei YANG, Jielin PAN and Yonghong YAN</i>	
<b>Topic-weak-correlated Latent Dirichlet Allocation</b>	<b>P.224</b>
<i>Yimin TAN and Zhijian OU</i>	
<b>Building Topic Mixture Language Models using the Document Soft Classification Notion of Topic Models</b>	<b>P.229</b>
<i>Shuanhu BAI, Cheung-Chi LEUNG, Chien-Lin HUANG, Bin MA and Haizhou LI</i>	
<b>Data-Driven Lexicon Refinement using Local and Web Resources for Chinese Speech Recognition</b>	<b>P.233</b>
<i>Hua ZHANG, Xuan ZHU, Tengrong SU, Kiwan EOM and Jaewon LEE</i>	
<b>The Psychoacoustic Approach towards Enhancing Speech Intelligibility in Noise</b>	<b>P.238</b>
<i>Paul Yaozhu CHAN, Minghui DONG, Ling CEN and Haizhou LI</i>	

<b>Improving Mandarin Chinese STT System with Random Forests Language Models</b>	<b>P.242</b>
<i>Ilya OPARIN, Lori LAMEL and Jean-Luc GAUVAIN</i>	
<b>Semantics-Based Language Modeling for Cantonese-English Code-mixing Speech Recognition</b>	<b>P.246</b>
<i>Houwei CAO, P. C. CHING, Tan LEE and Yu Ting YEUNG</i>	
<b>Web-Based Keyword Adapted Language Modeling for Keyword Spotting</b>	<b>P.251</b>
<i>Wenzhu SHEN, Ji WU and Wei LI</i>	
<b>Spontaneous Mandarin Speech Understanding Using Utterance Classification: A Case Study</b>	<b>P.256</b>
<i>Yun-Cheng JU and Jasha DROPPA</i>	
<b>Adaptive Segment Model for Spoken Document Retrieval</b>	<b>P.261</b>
<i>Chuang-Hua CHUEH and Jen-Tzung CHIEN</i>	
<b>Sentence Decomplexification Using Holistic Aspect-Based Clause Detection for Long Sentence Understanding</b>	<b>P.265</b>
<i>Chao-Hong LIU and Chung-Hsien WU</i>	
<b>SURE-MSE speech enhancement for robust speech recognition</b>	<b>P.271</b>
<i>Nengheng ZHENG, Xia LI, Thierry BLU and Tan LEE</i>	
<b>A Novel Subspace Speech Enhancement Approach based on Test of Hypothesis and Masking Properties</b>	<b>P.275</b>
<i>Wenju LIU, Ning CHENG and Chao LI</i>	
<b>A Speedup Method for the Separation of Speech Signals in Frequency Domain</b>	<b>P.281</b>
<i>Shih-Hsun CHEN and Hsiao-Chuan WANG</i>	
<b>A Novel Algorithm of Seeking FrFT Order for Speech Enhancement</b>	<b>P.286</b>
<i>Duojia MA, Xiang XIE and JingMing KUANG</i>	
<b>Non-Negative Matrix Factorization Based Discriminative Features for Speaker Verification</b>	<b>P.291</b>
<i>Yanhua LONG, Li-Rong DAI, Er-Yu WANG, Bin MA and Wu GUO</i>	
<b>Multidimensional Scaling for Fast Speaker Clustering</b>	<b>P.296</b>
<i>Chi-Chun HSIA, Yu-Hsien CHIU, Kuo-Yuan LEE and Chih-Chieh CHUANG</i>	
<b>An Enhanced Fishervoice Subspace Framework for Text Independent Speaker Verification</b>	<b>P.300</b>
<i>Weiwu JIANG, Helen MENG and Zhifeng LI</i>	
<b>Frame Selection of Interview Channel for NIST Speaker Recognition Evaluation</b>	<b>P.305</b>
<i>Hanwu SUN, Bin MA and Haizhou LI</i>	
<b>Speaker Verification against Synthetic Speech</b>	<b>P.309</b>
<i>Lian-Wu CHEN, Wu GUO and Li-Rong DAI</i>	
<b>Spectro-temporal Smoothed Auditory Spectra for Robust Speaker Identification</b>	<b>P.313</b>
<i>Ting-Han LIN, Chung-Chien HSU and Tai-Shih CHI</i>	
<b>Multi-Feature Combination for Speaker Recognition</b>	<b>P.318</b>
<i>Zhi Yi LI, Liang HE, Wei Qiang ZHANG and Jia LIU</i>	

<b>Effects of F0 Dimensions in Perception of Mandarin Tones</b>	<b>P.322</b>
<i>Bin LI, Caicai ZHANG</i>	
<b>Investigation of the Relation between Acoustic Features and Articulation -- An Application to Emotional Speech Analysis</b>	<b>P.326</b>
<i>Yongxin WANG, Jianwu DANG and Lianhong CAI</i>	
<b>Investigation of Muscle Activation in Speech Production Based on an Articulatory Model</b>	<b>P.330</b>
<i>Xiyu WU, Qiang FANG and Jianwu DANG</i>	
<b>The Relation between Larynx Height and F0 during the Four Tones of Mandarin in X-ray Movie</b>	<b>P.335</b>
<i>Gaowu WANG and Jiangping KONG</i>	
<b>Does Semantic Stress Have Effect on Duration and Pitch Patterns of Prosodic Words in Presenters' Speech?</b>	<b>P.339</b>
<i>Yu ZOU, Wei HE, Min HOU and Yonglin TENG</i>	
<b>Downstep in High-Low Sequences in Chinese</b>	<b>P.344</b>
<i>Maolin WANG, Hua WU and Aijun LI</i>	
<b>Relation between Focus and Accent in Standard Chinese</b>	<b>P.348</b>
<i>Yuan JIA and Aijun LI</i>	
<b>Mandarin Prosodic Break Detection Based on Complementary Model</b>	<b>P.353</b>
<i>Chong-Jia NI, Wen-Ju LIU and Bo XU</i>	
<b>Acoustic Development of Vowels in Children's Speech</b>	<b>P.358</b>
<i>Wai-Sum LEE and Eric ZEE</i>	
<b>GMM-based Voice Conversion with Explicit Modelling on Feature Transform</b>	<b>P.364</b>
<i>Ling-Hui CHEN, Zhen-Hua LING, Wu GUO and Li-Rong DAI</i>	
<b>Study on Attenuated Tone for Mandarin Text-To-Speech</b>	<b>P.369</b>
<i>Xiaoyan LOU and Jian LI</i>	
<b>Automatic Phrase Boundary Labeling for Mandarin TTS Corpus Using Context-Dependent HMM</b>	<b>P.374</b>
<i>Chen-Yu YANG, Zhen-Hua LING, Heng LU, Wu GUO and Li-Rong DAI</i>	
<b>Hierarchical Pitch Target Model for Mandarin Speech</b>	<b>P.378</b>
<i>Zhiping ZHANG, Xinhao WANG, Yansuo YU and Xihong WU</i>	
<b>Generating Emotional Speech from Neutral Speech</b>	<b>P.383</b>
<i>Ling CEN, Paul CHAN, Minghui DONG and Haizhou LI</i>	
<b>Mandarin to Lanzhou Dialect Conversion based on Five Degree Tone Model</b>	<b>P.387</b>
<i>Hong-Wu YANG, Wei-Tong GUO, Dong PEI and Qing-Qing LIANG</i>	
<b>Improving GMM-Based Spectral Conversion with Optimal Conversion Function Selection</b>	<b>P.392</b>
<i>Hsin-Te HWANG, Weng-Liang WU and Sin-Horng CHEN</i>	
<b>Prosody Phrase Boundary Prediction with Ensemble Learning</b>	<b>P.397</b>
<i>Lifu YI, Jian LI, Lei HE, Jie HAO and Rui ZHAO</i>	

<b>Error Diagnosis Using Penalized Probabilistic FOIL for Chinese as a Second Language Learner</b>	<b>P.401</b>
<i>Ru-Yng CHANG, Chung-Hsien WU and Philips Kokoh PRASETYO</i>	
<b>Automatic Lexical Stress Detection for Chinese Learners' of English</b>	<b>P.407</b>
<i>Jinyu CHEN and Lan WANG</i>	
<b>Robust Pronunciation Evaluation in Adverse Environments</b>	<b>P.412</b>
<i>Si WEI, Qianyong GAO, Guoping HU and Yu HU</i>	
<b>A Distinctive Feature Based Method for Evaluating The Phonetic Transcription Of A Non-Native Speech Database</b>	<b>P.416</b>
<i>Jinsong ZHANG, Dongning WANG, Wen CAO and Ziyu XIONG</i>	
<b>Multi-Modal Feature Integration for Story Boundary Detection in Broadcast News</b>	<b>P.420</b>
<i>Mimi LU, Lei XIE, Zhong-Hua FU, Dong-Mei JIANG and Yan-Ning ZHANG</i>	
<b>Confidence Estimation for Spoken Language Translation based on Round Trip Translation</b>	<b>P.426</b>
<i>Dong YU, Wei WEI, Lei JIA and Bo XU</i>	
<b>High Performance Chinese Spoken Term Detection Based on Term Expansion</b>	<b>P.430</b>
<i>Wei LI, Ji WU and Ping LV</i>	
<b>Perception and Analysis of Linearly Approximated F0 Contours in Cantonese Speech</b>	<b>P.435</b>
<i>Yujia LI and Tan LEE</i>	
<b>The Duration Analysis of the Checked Tones in Cantonese Speech</b>	<b>P.440</b>
<i>Xiaoying XU, Jianhua TAO, Ling ZHANG and Yingchao LU</i>	
<b>Constructing Online Audio Dictionaries for Bilingual Mandarin-Taiwan Dialects Based on Web 2.0 Concept</b>	<b>P.446</b>
<i>Neng-Huang PAN, Feng-Long HUANG, Chun-Hsien HO, Xin-Wei LIN and Shu-Hau SHIU</i>	
<b>Combining HMM Spectrum Models and ANN Prosody Models for Speech Synthesis of Syllable Prominent Languages</b>	<b>P.451</b>
<i>Hung-Yan GU, Ming-Yen LAI and Sung-Feng TSAI</i>	
<b>A Combined Approach to the Polysemy Problems in a Chinese to Taiwanese TTS System</b>	<b>P.455</b>
<i>Yih-Jeng LIN, Ming-Shing YU and Chin-Yu LIN</i>	
<b>Language Identification In Code-Switching Speech Using Word-based Lexical Model</b>	<b>P.460</b>
<i>Dau-Cheng LYU, Ren-Yuan LYU, Cing-Lei ZHU and Ming-Tat KO</i>	
<b>A Survey on Recent Progress in the ASAT/SIRKUS Paradigm</b>	<b>P.465</b>
<i>Sabato Marco SINISCALCHI, Torbjørn SVENDSEN and Chin-Hui LEE</i>	
<b>Automatic Voice Onset Time Estimation of Stops in Continuous Speech</b>	<b>P.471</b>
<i>Chi-Yueh LIN and Hsiao-Chuan WANG</i>	
<b>Human Speech Model Based on Information Separation and Its Application to Speech Processing</b>	<b>P.477</b>
<i>Nobuaki MINEMATSU</i>	

**Robust Speaker Verification Using Phase Information of Speech**

*Ning WANG, P. C. CHING and Tan LEE*

**P.483**

**Phone Boundary Refinement Using Ranking Methods**

*Hung-Yi LO and Hsin-Min WANG*

**P.488**