

Bibliography

- [1] A. K. Syrdal et al., "Corpus-based Techniques in the AT&T NEXTGEN Synthesis System," in *Proc. ICSLP*, pp. 410–415, 2000.
- [2] M. Chu, H. Peng, H. Y. Yang and E. Chang, "Selecting Non-uniform Units from a Very Large Corpus for Concatenative Speech Synthesizer," in *Proc. ICASSP*, pp. 785–788, 2001.
- [3] F. C. Chou, C. Y. Tseng and L. S. Lee, "A Set of Corpus-based Text-to-speech Synthesis Technologies for Mandarin Chinese," *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 7, pp. 481–494, 2002.
- [4] H. Meng, C. K. Keung, S. K. Chung, T. Y. Fung and P. C. Ching, "CU Vocal: Corpus-based Syllable Concatenation for Chinese Speech Synthesis across Domains and Dialects," in *Proc. ICSLP*, pp. 2373–2376, 2002.
- [5] J. R. W. Yi, "Corpus-Based Unit Selection for Natural-Sounding Speech Synthesis," *Ph.D. thesis*, Massachusetts Institute of Technology, 2003.
- [6] Y. Meron, "High Quality Singing Synthesis Using the Selection-base Synthesis Scheme," *Ph.D. thesis*, University of Tokyo, 1999.
- [7] C. Y. Lin, T. Y. Lin and J. S. Jang, "A Corpus-based Singing Voice Synthesis System for Mandarin Chinese," in *Proc. ACM Multimedia*, pp. 359–362, 2005.
- [8] L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 25, No. 1, pp. 24–33, 1977.
- [9] J. C. Brown, "Musical Fundamental Frequency Tracking Using a Pattern Recognition Method," *Journal of the Acoustical Society of America*, Vol. 92, No. 3, pp. 1394–1402, 1992.
- [10] A. Choi, "Real-Time Fundamental Frequency Estimation by Least-Square Fitting," *IEEE Trans. on Speech and Audio Processing*, Vol. 5, No. 2, pp. 201–205, 1997.
- [11] F. Charpentier and Moulines, "Pitch-synchronous Waveform Processing Technique for Text-to-Speech Synthesis Using Diphones," in *Proc. Eurospeech*, pp. 13–19, 1989.
- [12] M. W. Macon, "Speech Synthesis Based on Sinusoidal Modeling," *PhD thesis*, Georgia Institute of Technology, October 1996.

- [13] Y. Stylianou, "Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis," *IEEE Trans. on Speech and Audio Processing*, Vol. 9, No.1, pp. 21–29, 2001.
- [14] E. Moulines et al. "A Real-Time French Text-to-Speech System Generating High-Quality Synthetic Speech," in *Proc. ICASSP*, pp. 309–312, 1990.
- [15] C. Y. Lin and J. S. Jang, "A Two-Phase Pitch Marking Method for TD-PSOLA Synthesis," *GESTS International Transaction on Speech Science and Engineering*, Vol. 1, No. 2, pp. 211–221, 2004. (Invited paper)
- [16] X. Huang, A. Acero and H. W. Hon, *Spoken Language Processing*, New Jersey, Prentice Hall, 2001.
- [17] F. Malfrère, O. Deroo and T. Dutoit, "Phonetic Alignment: Speech Synthesis based vs. Hybrid HMM/ANN," in *Proc. ICSLP*, pp. 1571–1574, 1998.
- [18] C. W. Wightman and D. T. Talkin, "The Aligner: Text-to-Speech Alignment Using Markov Models," in *Progress in Speech Synthesis*, pp. 313–323, 1997.
- [19] M. J. Makashay, C. W. Wightman, A. K. Syrdal and A. Conkie, "Perceptual Evaluation of Automatic Segmentation in Text-to-Speech Synthesis," in *Proc. ICSLP*, pp. 431–434, 2000.
- [20] J. P. H. van Santen, and R. Sproat, "High-Accuracy Automatic Segmentation," in *Proc. Eurospeech*, pp. 2809–2812, 1990.
- [21] A. Bonafonte, A. Nogueiras and A. Rodriguez-Garrido, "Explicit Segmentation of Speech Using Gaussian Models," in *Proc. ICSLP*, pp. 1269–1272, 1996.
- [22] T. D. Toledano, M. A. Rodriguez Crespo and J. G. EscaladaSardina, "Trying to Mimic Human Segmentation of Speech Using HMM and Fuzzy Logic Post-correction Rules," in *Proc. Third ESCA/COCOSDA Workshop on speech synthesis*, pp. 207–212, 1998.
- [23] F. C. Chou, C. Y. Tseng and L. S. Lee, "Automatic Segmental and Prosodic Labeling of Mandarin Speech," in *Proc. ICSLP*, pp. 1263–1266, 1998.
- [24] L. Wang et al. "Refining Segmental Boundaries for TTS Database Using Fine Contextual-Dependent Boundary Models", in *Proc. ICASSP*, pp. 641–644, 2004.
- [25] D. T. Toledano, "Neural Network Boundary Refining for Automatic Speech Segmentation," in *Proc. ICASSP*, pp. 3438–3441, 2000.

- [26] M. Sharma and R. Mammone, "Automatic Speech Segmentation Using Neural Tree Networks," in *Proc. IEEE Workshop, Neural Networks for Signal Processing*, pp. 282–290, 1995.
- [27] E. Y. Park, S. H. Kim and J. H. Chung, "Automatic Speech Synthesis Unit Generation with MLP Based Postprocessor Against Auto-segmented Phoneme Errors," in *Proc. Int. Joint Conf. Neural Networks*, pp. 2985–2990, 1999.
- [28] H. Kawai and T. Toda, "An Evaluation of Automatic Phone Segmentation for Concatenative Speech Synthesis," in *Proc. ICASSP*, pp. 677–680, 2004.
- [29] K. S. Lee, "MLP-based Phone Boundary Refining for A TTS Database," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 3, pp. 981–989, 2005.
- [30] J. S. Jang and M. Y. Gao, "A Query-by-singing System Based on Dynamic Programming," in *Proc. International Workshop on Intelligent Systems Resolutions*, pp. 85–89, 2000.
- [31] J. S. Jang and H. R. Lee, "Hierarchical Filtering Method for Content-based Music Retrieval via Acoustic Input," in *Proc. ACM Multimedia*, pp. 401–410, 2001.
- [32] C. Y. Lin, J. S. Jang and M. Y. Hsu, "An Automatic Singing Voice Rectifier Design," in *Proc. ACM Multimedia*, pp. 267–270, 2003.
- [33] T. Svendsen and F. Soong, "On The Automatic Segmentation of Speech Signals," in *Proc. ICASSP*, pp. 77–80, 1987.
- [34] H. Hon, A. Acero, X. Huang, J. Liu and M. Plumpe, "Automatic Generation of Synthesis Units for Trainable Text-to-speech System," in *Proc. ICASSP*, 1998, pp. 293–296, 1998.
- [35] A. Ljolje and M. D. Riley, "Automatic Segmentation and Labeling of Speech," in *Proc. ICASSP*, pp. 473–476, 1991.
- [36] A. Ljolje, J. Hirschberg and J. P. H. Van Santen, "Automatic Speech Segmentation for Concatenative Inventory Selection," in *Progress in Speech Synthesis*, pp. 305–311, 1997.
- [37] D. T. Toledano, L. A. H. Gomez and L. V. Grande, "Automatic Phonetic Segmentation," *IEEE Trans. Speech Audio Process.*, Vol. 11, No. 6, pp. 617–625, 2003.

- [38] C. G. J. Houben, "Automatic Labeling of Speech Using An Acoustic-phonetic Knowledge Base," in *Proc. EuroSpeech*, vol. 2, pp. 104–107, 1989.
- [39] K. Hatazaki, Y. Komor, T. Kawabata and K. Shikano, "Phoneme Segmentation Using Spectrogram Reading Knowledge," in *Proc. ICASPP*, pp. 393–396, 1989.
- [40] J. P. van Hermert, "Automatic Segmentation of Speech," *IEEE Trans. Speech Audio Process.*, Vol. 39, No. 4, pp. 1008–1012, 1991.
- [41] C. Y. Lin, J. S. Jang and P. C. Jao, "An Emotional Transformation Model for Chinese Expressive Speech Synthesis: from Neutral to Anger," *Speech Communication* (submitted after the first revision), 2007.
- [42] K. J. Chen and S. H. Liu, "Word Identification for Mandarin Chinese Sentences," *Proceedings of the Fifteenth International Conference on Computational Linguistics*, pp. 101–107, 1992.
- [43] C. L. Yeh and H. J. Lee, "Rule-based Word Identification for Mandarin Chinese Sentences - A Unification Approach," *International Journal of Computer Processing of Chinese and Oriental Languages*, Vol. 5, No. 2, pp. 97–118, 1991.
- [44] R. Sproat and C. Shih, "A Statistical Method for Finding Word Boundaries in Chinese Text," *International Journal of Computer Processing of Chinese and Oriental Languages*, Vol. 3, No. 4, pp. 336–351, 1990.
- [45] C. Y. Lin, J. S. Roger Jang and K. T. Chen, "Automatic Segmentation and Labeling for Mandarin Chinese Speech Corpora for Concatenation-based TTS," *International Journal of Computational Linguistics and Chinese Language Processing*, Vol. 10, No. 2, pp. 145–166, 2005.
- [46] D. Talkin, "A Robust Algorithm for Pitch Tracking," *Speech Coding and Synthesis*. Amsterdam, NL: Elsevier Science, 1995.
- [47] Speech Filing System (SFS) software, [Online]. Available: <http://www.phon.ucl.ac.uk/resource/sfs/>
- [48] Y. M. Cheng and D. O'Shaughnessy, "Automatic and Reliable Estimation of Glottal Closure Instant and Period," *IEEE Trans. Acoustic, Speech and Signal Processing*, Vol. 37, No. 11, pp. 1805–1815, 1989.
- [49] M. Sakamoto and T. Saito, "An Automatic Pitch-Marking Method using Wavelet Transform," in *Proc. ICSLP*, pp. 650–653, 2000.

- [50] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, pp. 339–340, Prentice Hall, 1993.
- [51] C. Y. Lin and J. S. Jang, “New Refinement Schemes for Voice Conversion,” in *Proc. ICME*, pp. 725–728, 2003.
- [52] C. Y. Lin, J. S. Jang and S. H. Hwang, “An On-The-Fly Mandarin Singing Voice Synthesis System,” in *Proc. IPCM*, pp. 631–638, 2002.
- [53] TCC-300, [Online]. Available: http://www.aclclp.org.tw/use_mat.php#tcc300edu
- [54] M. J. Makashay, C. W. Wightman, A. K. Syrdal and A. Conkie, “Perceptual Evaluation of Automatic Segmentation in Text-to-speech Synthesis,” in *Proc. ICSLP*, pp. 431–434, 2000.
- [55] HMM Toolkits (HTK), [Online]. Available: <http://htk.eng.cam.ac.uk/>
- [56] C. Chesta, O. Siohan, and C.-H. Lee, “Maximum a Posteriori Linear Regression for Hidden Markov Model Adaptation,” in *Proc. EuroSpeech*, pp. 211–214, 1999.
- [57] C. J. Leggetter and P. C. Woodland, “Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models,” *Comput. Speech Lang.*, Vol. 9, No. 2, pp. 171–185, 1995.
- [58] F. Zheng, Z. Song, L. Li, W. Yu, F. Zheng and W. Wu, “The Distance Measure for Line Spectrum Pairs Applied to Speech Recognition,” in *Proc. ICSLP*, pp. 1123–1126, 1998.
- [59] J. L. Shen, J. W. Hung and L. S. Lee, “Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments,” in *Proc. ICSLP*, pp. 232–235, 1998.
- [60] A. Whitney, “A Direct Method of Nonparametric Measurement Selection,” *IEEE Trans. on Comput.*, Vol. 20 No. 9, pp. 1100–1103, 1971.
- [61] P. Pudil, J. Novovicová and J. Kittler, “Floating Search Methods in Feature Selection,” *Pattern Recognition Letters*, Vol. 15, pp. 1119–1125, 1994
- [62] F. Fukunaga and P. M. Narendra, “A Branch and Bound Algorithm for Computing K-nearest Neighbors,” *IEEE Trans. Comput.*, Vol. 24, No. 7, pp. 750–753, 1975.
- [63] R. A. Redner and H. F. Walker, “Mixture Densities, Maximum Likelihood and the EM Algorithm,” *SIAM Review*, Vol. 26, No. 2, pp. 195–239, 1984.
- [64] S. Haykin, *Neural Networks: A Comprehensive Foundation*. 2nd ed., Prentice Hall, 1999.

- [65] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [66] R. D. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley, New York, 2001.
- [67] C. C. Chang and C. J. Lin, LIBSVM: A Library for Support Vector Machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [68] V. Vapnik, *Statistical Learning Theory*. Wiley, New York, 1998.
- [69] B. Schölkopf, A. Smola, R. Williamson and P. L. Bartlett, "New Support Vector Algorithms," In *Proc. Neural Computation*, pp. 1207–1245, 2000.
- [70] W. H. Press, "Least Squares Linear Regression," *Numerical Recipes in C*. Cambridge Univ. Press, 1992.
- [71] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with The Marquardt Algorithm," *IEEE Trans. Neural Network*, Vol. 5, No. 6, pp. 989–993, 1994.
- [72] J. Bi and K. P. Bennett, "A Geometric Approach to Support Vector Regression," *Neurocomputing*, Vol. 55, No. 1-2, pp. 79–108, 2003.

