Literature Review

Yonatan Vaizman

I. Audio processing

[1] for speech coding tutorial.....

[2] for joint spectral-envelope and f0 estimation....

[3] efficient solution to sparse linear prediction analysis for speech...

Books: [4]–[7]

A. LPC

Good tutorial for linear prediction framework [8]. [9], [10]... [11], [12] sparse linear prediction for speech....

B. Vocoders

[13]....

C. Frequency-warped LPC

D. Fourier space excitation-filter framework

E. Filterbanks (overcomplete codebooks?) for speech analysis

Speech coding with VQ [10]....

II. Optimization and learning

A. LPC optimization

Good tutorial for linear prediction framework [8]. Makhoul talks about the squared error criterion, its advantages and shortcomings and when it might not fit the best spectral envelop (when the excitation input to the system is a pitched periodic pulse train — in which case the peaks of the estimated filter coincide with the harmonics instead of the better fitting envelop).

B. L2, L1, Lp minimizations

L1 minimization (least absolute error criterion) for speech [14]....

Linear prediction with L1 norm [15]...

Adaptive Lp linear prediction [16]...

[11], [12] sparse linear prediction for speech....

Stable IIR design based on Lp error minimization [17]....

Giri and Rao - block sparse excitation criterion [18]...

C. Optimization of mixtures

D. Sparsecoding and compressed sensing

E. LTI filter clustering

Perceptually consistent measures of spectral distance [19]....

Speech coding with VQ [10]....

Spectral distance measures [19]....

VQ in speech coding [20]...

III. MIR applications

A. Source separation

[21]....

B. Melody extraction and automatic transcription

[22]–[24].... Guitar chords and fingering [25]....

C. Chord recognition

D. Instrument recognition

[26], [27].... Polyphonic and polyinstrument [28].... Instrument recognition (temporal and cepstral features) [29].... Identifying woodwind instruments [30]..... Multitrack mixing: [31], [32]....

IV. Datasets

References

[1] A. S. Spanias, "Speech coding: a tutorial review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541–1582, 1994.

[2] H. Kameoka, N. Ono, and S. Sagayama, "Speech spectrum modeling for joint estimation of spectral envelope and fundamental frequency," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1507–1516, 2010.

[3] V. Khanagha and K. Daoudi, "An efficient solution to sparse linear prediction analysis of speech," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 1, pp. 1–9, 2013.

[4] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice-hall Englewood Cliffs, 1978, vol. 100.

[5] W. C. Chu, *Speech coding algorithms: foundation and evolution of standardized coders*. John Wiley & Sons, 2004.

[6] W. B. Kleijn and K. K. Paliwal, *Speech coding and synthesis*. Elsevier Science Inc., 1995.

[7] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Upper Saddle River (NJ, USA): Prentice Hall, 1993.

[8] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[9] ——, "Stable and efficient lattice methods for linear prediction," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 5, pp. 423–428, 1977.

[10] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, vol. 73, no. 11, pp. 1551–1588, 1985.

[11] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen, and M. Moonen, "Sparse linear predictors for speech processing." in *INTERSPEECH*, 2008, pp. 1353–1356.

[12] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1644–1657, 2012.

[13] J. Makhoul, R. Viswanathan, and W. Russell, "A framework for the objective evaluation of vocoder speech quality," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'76.*, vol. 1. IEEE, 1976, pp. 103–106.

[14] E. Denoel and J.-P. Solvay, "Linear prediction of speech with a least absolute error criterion," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, no. 6, pp. 1397–1403, 1985.

[15] J. Schroeder and R. Yarlagadda, "Linear predictive spectral estimation via the¡ i¿ l¡/i¿¡ sub¿ 1¡/sub¿ norm," *Signal processing*, vol. 17, no. 1, pp. 19–29, 1989.

[16] J. Lansford and R. Yarlagadda, "Adaptive l p approach to speech coding," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*. IEEE, 1988, pp. 335–338.

[17] C.-C. Tseng, "Design of stable iir digital filter based on least p-power error criterion," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 51, no. 9, pp. 1879–1888, 2004.

[18] R. Giri and B. D. Rao, "Block sparse excitation based all-pole modeling with applications to speech," in *ICAASP*, 2014.

[19] R. Viswanathan, J. Makhoul, and W. Russell, "Towards perceptually consistent measures of spectral distance," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'76.*, vol. 1. IEEE, 1976, pp. 485–488.

[20] A. Gersho, S. Wang, and K. Zeger, "Vector quantization techniques in speech coding," *Advances in Speech Signal Processing, S. Furui and MM Sondhi, eds., Mar-cel Dekker, New York*, pp. 49–84, 1992.

[21] M. Slaney, D. Naar, and R. F. Lyon, "Auditory model inversion for sound separation," in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, vol. 2. IEEE, 1994, pp. II–77.

[22] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: challenges and future directions," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 407–434, 2013.

[23] K. O'Hanlon and M. D. Plumbley, "Automatic music transcription using row weighted decompositions," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 16–20.

[24] G. Peeters, "Music pitch representation by periodicity measures based on combined temporal and spectral representations," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5. IEEE, 2006, pp. V–V.

[25] A. M. Barbancho, A. Klapuri, L. J. Tardon, and I. Barbancho, "Automatic transcription of guitar chords and fingering from audio," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 915–921, 2012.

[26] K. D. Martin and Y. E. Kim, "Musical instrument identification: A pattern-recognition approach," *The Journal of the Acoustical Society of America*, vol. 104, no. 3, pp. 1768–1768, 1998.

[27] L.-F. Yu, L. Su, and Y.-H. Yang, "Sparse cepstral codes and power scale for instrument identification," in *Proc. ICASSP*, 2014.

[28] P. Hamel, S. Wood, and D. Eck, "Automatic identification of instrument classes in polyphonic and poly-instrument audio." in *ISMIR*. International Society for Music Information Retrieval conference (ISMIR), 2009, pp. 399–404.

[29] A. Eronen and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 2. IEEE, 2000, pp. II753–II756.

[30] J. C. Brown, O. Houix, and S. McAdams, "Feature dependence in the automatic identification of musical woodwind instruments," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1064–1072, 2001.

[31] J. Scott and Y. E. Kim, "Instrument identification informed multi-track mixing," in *International Society for Music Information Retrieval conference (ISMIR)*, 2013.

[32] J. Scott, M. Prockup, E. M. Schmidt, and Y. E. Kim, "Automatic multi-track mixing using linear dynamical systems," in *Proceedings of the 8th Sound and Music Computing Conference, Padova, Italy*, 2011.