Speech Analysis Method Based on Source-Filter Model Using Multivariate Empirical Mode Decomposition

Surasak BOONKLA[1,2,3], Masashi UNOKI[1], Stanislav S. MAKHANOV[2], Chai WUTIWIWATCHAI[3]

[1]School of Information Science, Japan Advanced Institute of Science and Technology, Japan.
[2]Sirindhorn International Institute of Technology, Thammasat University, Thailand.
[3]National Electronics and Computer Technology Center, Thailand.

Abstract : A new speech analysis method based on source-filter model using multivariate empirical mode decomposition (MEMD) is proposed. The proposed decomposes log spectra of adjacent frame of speech signals into intrinsic mode functions (IMFs). The IMFs are divided into two groups using common mode alignment property of MEMD. The first group characterized by spectral fine structure is used to estimate the fundamental frequency (F0) based on ACF whereas the second group characterized by the frequency response of the vocal-tract filter is used to evaluate formant frequencies by the peak picking technique. The proposed method automatically separates the glottal-source and the vocal-tract filter. The results show that the proposed method could provide high accuracy of F0 estimation and correct formant frequencies of vocal-tract.

Keywords : Multivariate empirical mode decomposition, speech analysis, fundamental frequency, formant frequency

1. Introduction & Motivation

Existing speech analysis methods still have limitations. Most of them can analyze two important information such as F0 and formant frequencies. Thus, development of new speech analysis method is important. We aim to proposed MEMD-based speech analysis method.

2. Multivariate signal and common mode

A multivariate signal can be formed by multi-channel or single-channel. Important information which is common to all sub signals, $x_j(t)$, as shown in Fig. 1 in the red box can be extracted by multivariate empirical mode decomposition (MEMD) into IMF, $z_i(t)$, such that $x_j(t) = \sum_{i=1}^{K} z_i(t)$.

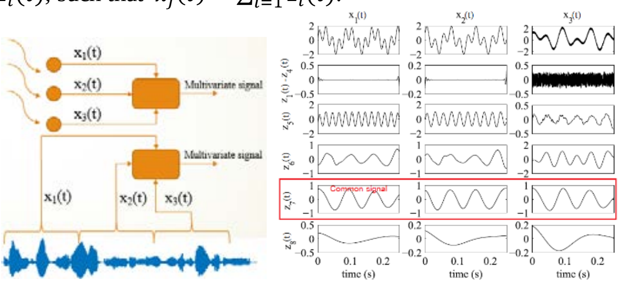

Figure 1. Multivariate signal and common mode extraction.

3. Proposed method

Log spectra of speech signal are decomposed into IMFs. They are divided into two groups corresponding to glottal-source, $\log|E(\omega)|$, and vocal-tract filter, $\log|V(\omega)|$. The value of M is of Eq. (3) is determined by the location of common mode of glottal-source using correlation coefficient as shown in Fig. 2.

$$s(t) = e(t) * v(t) \quad (1)$$

$$\log|S(\omega)| = \log|E(\omega)| + \log|V(\omega)| \quad (2)$$

$$\log|S(\omega)| = \underbrace{\sum_{i=1}^{M} z_i(\omega)}_{source} + \underbrace{\sum_{M+1}^{K} z_i(\omega)}_{filter} \quad (3)$$

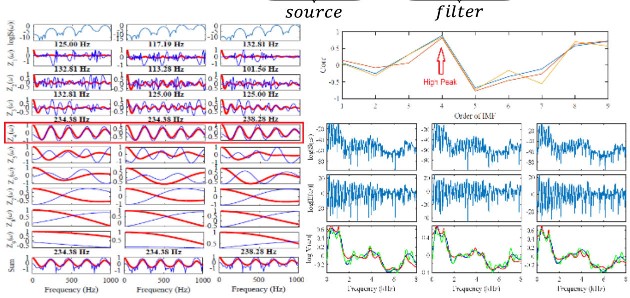

Figure 2. IMFs, common mode, and source-filter separation.

After IMFs separation, the first group is used for F0 estimation and the second group is used for formant estimation. The result are shown in Fig. 3.

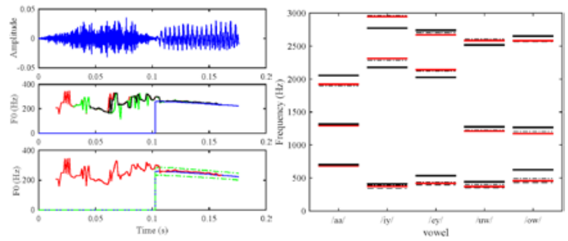

Figure 3. Results of F0 and formant frequencies estimations.

4. Conclusion

MEMD can be used for speech analysis method for voiced speech signal as well as LP and cepstrum techniques. The results indicate the efficiency of the proposed method.

References

[1] S. Boonkla, M. Unoki, S. S. Makhanov, and C. Wutiwiwatchai, "Speech analysis method based on source-filter model using multivariate empirical mode decomposition in log-spectrum domain," Proc. IEEE Int. Symposium on Chinese Spoken Language Processing (ISCSLP), pp. 555-559, Sept. 2014.