Investigation of a Method of Speech Signal Analysis Using Empirical Mode Decomposition and Its Applications

Tomoki Sawaguchi and Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923–1292 Japan
Phone/FAX:+81-761-51-1391/+81-761-51-1149
E-mail: {tomoki-s, unoki}@jaist.ac.jp

Abstract

In recent years, a number of noise reduction methods based on empirical mode decomposition (EMD) have been proposed in the field of speech signal processing. However, these methods cannot effectively reduce noise components from noisy speech they lack useful prior knowledge related to the noise characteristics. Moreover, because they reduce only the higher frequency components of noise, the overall effect on noise reduction seems to be insufficient. Our aim was to develope a speech signal analysis method that can adequately analyze non-stationary speech signals in time-frequency domains. We investigated the properties of an analysis method for non-stationary signals based on EMD and the characteristics of AM-FM representations in the intrinsic mode function (IMF) and have subsequently developed a method of noise reduction based on our investigations. Simulations were conducted to determine whether or not the proposed method can effectively reduce noise components from noisy speech. Results demonstrate that it can do so adequately.

1. Introduction

Currently, the Fourier transform and the wavelet transform are the standard techniques used to analyze signals in time-frequency analysis. These methods can analyze the temporal spectral fluctuations of the signal in the time-frequency domains, but only if the analytical signal is assumed to be stationary. Realistic signals (i.e., electroencephalogram (EEG) signals, seismic waves, speech signals, etc.) are non-stationary signals so these methods cannot precisely analyze the non-stationary fluctuations of the instantaneous amplitude and the instantaneous phase of the signal.

In recent years, the empirical mode decomposition (EMD) technique [1], originally proposed by Huang *et al.*, has been used for analyzing non-stationary signals. This technique can analyze EEG signals and explore the source of seismic waves, and it is currently being applied to speech signal processing. In particular, EMD-based noise reduction methods have been proposed to reduce musical noise [2] from restored speech and to classify robust voiced/unvoiced signals in noisy environments [3].

Because speech signals are generally non-stationary, speech representation based on EMD seems to be more suitable that conventional methods in terms of representing speech features such as non-stationary fluctuations. However, it is unclear how or if noisy speech can be represented as suitable forms (separately speech and noise), and it is also unclear whether these speech and noise components can be completely separated in the representations. Because these previously proposed methods [2, 3] use particular IMFs corresponding to noise components in speech in which the noise is to be reduced, they can remove IMFs of non-stationary speech by reducing the noise components on these representations.

We investigated the properties of the analysis method for non-stationary signals using EMD and the characteristics of the decomposed intrinsic mode function (IMF). We then examined the possibility of using EMD to reduce the noise in noisy speech signals.

2. Empirical Mode Decomposition (EMD)

2.1. Signal representation using EMD

EMD decomposes signal $x(t)$ into IMFs, $c_k(t)$, and negligible residue $r(t)$. $x(t)$ is represented as follows.

$$x(t) = \sum_{k=1}^{K} c_k(t) + r(t) \tag{1}$$

where k represents the channel number and K represents the number of IMFs. Here, the IMF must satisfy two conditions: (1) in the entire data set, the number of extrema and the number of zero crossings must either equal or differ at most by one, and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. K depends on the characteristics of the analytical signal so all signals may not have the same K.

Figure 1 shows the problem analysis diagram (PAD) of the EMD algorithm. In this algorithm, upper envelope $u(t)$ and lower envelope $l(t)$ are obtained from local maxima and local minima, respectively, by using cubic spline interpolation. Next, the mean value between $u(t)$ and $l(t)$ is subtracted from the original signal while mean value is not zero. Finally, while the IMF satisfies the two constraints, the original signal

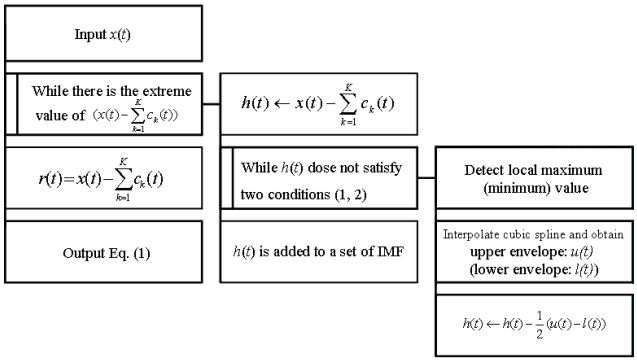

Figure 1: PAD of the empirical mode decomposition

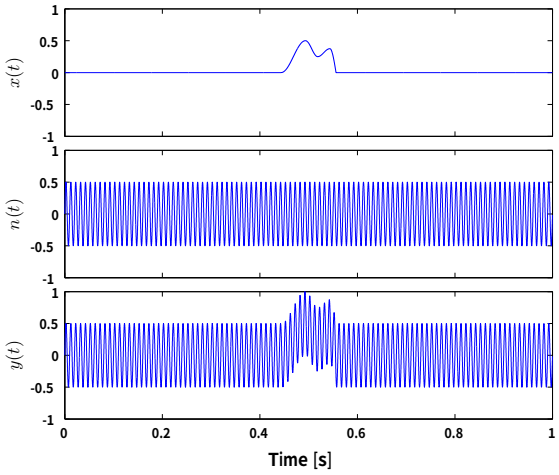

Figure 2: Mixed signal $y(t)$ composed of non-stationary signal $x(t)$ and stationary signal $n(t)$

is decomposed into IMFs by repeating these subtractions, as shown in Fig. 1.

2.2. Properties of EMD

In this section, we investigate the properties of EMD and the characteristics of the decomposed IMFs. First, we study how the IMFs are derived by the EMD algorithm. The mean value between the upper and lower envelopes is subtracted from the signal to derive the first IMF. This step is repeated while the mean value is not zero to derive the kth IMF. It can thus be understood that the decomposed IMFs are obtained with the intent of grouping them in a common envelope. In this case, the IMF can be matched to slow or fast fluctuations in the envelope.

Next, we investigate the characteristics of the decomposed IMFs. The first constraint indicated that IMFs must be a signal that alternates the extreme value and zero crossing in turn. This suggests that the IMFs are represented as an FM-signals without any band-limitation because there is no limitation of the pair-frequency of the extreme value and zero crossing. The second constraint indicated that the IMFs must have the same upper and lower envelopes. This suggests that the IMFs are common AM-signals. In summary, the decomposed

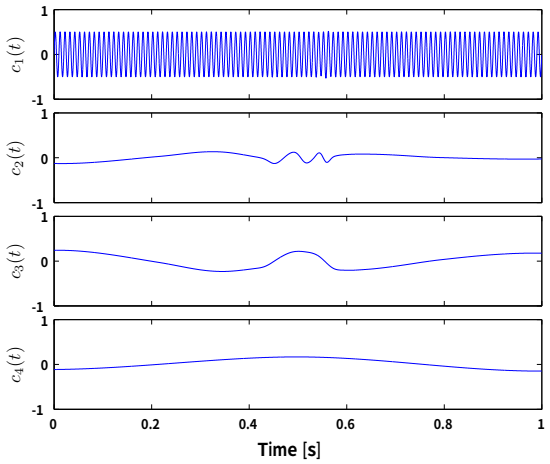

Figure 3: Decomposition of $y(t)$ (IMFs, $c_k(t)$)

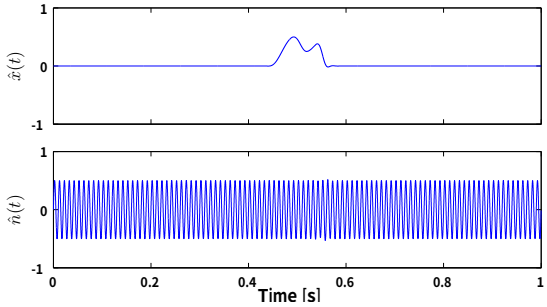

Figure 4: Resynthesized signals $\hat{x}(t)$ and $\hat{n}(t)$

IMFs can be regarded as AM-FM signals based on common-envelope decomposition.

2.3. Example of signal analysis using the EMD

We examined an example of signal analysis using EMD for the following mixture: $y(t)$ is composed of non-stationary signal $x(t)$ and stationary signal $n(t)$, as shown in Fig. 2. The decomposed IMFs of $y(t)$, $c_1(t)$, $c_2(t)$, $\cdots$, and $c_4(t)$, are shown in Fig. 3. Based on our investigations in Sec. 2.2, $c_1(t)$ can be regarded as a stationary signal with a constant envelope while the other IMFs $c_2(t), c_3(t)$, and $c_4(t)$ can be regarded as non-stationary signals. This means that the first IMF, $c_1(t)$, seems to be $\hat{n}(t)$ and the summed IMF, $c_2(t) + c_3(t) + c_4(t)$, seems to be $\hat{x}(t)$, as shown in Fig. 4. This result demonstrates that the essence of signal analysis based on EMD is to separate non-stationary signals from stationary signals during the signal representation procedure by a common-envelope-based decomposition.

3. EMD-Based Noise Reduction Method

We next consider the applicability of sound analysis based on EMD. In the previous section, we showed that EMD can easily separate stationary and non-stationary signals on the decomposed IMFs. With this advantage, we consider a sepa-

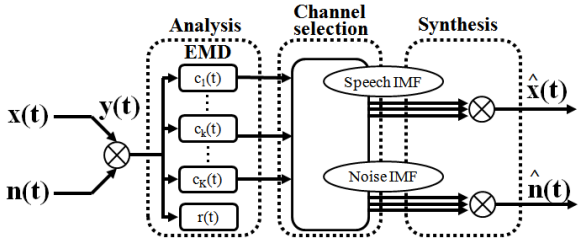

Figure 5: Proposed method

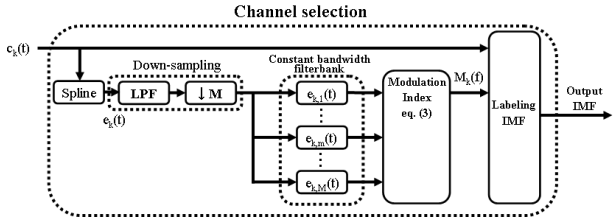

Figure 6: Channel selection of IMFs

ration of non-stationary speech signals and stationary white noise as an application. An EMD-based noise reduction method has already proposed by Molla & Hirose [3]. In this method, they focus on energy distribution of noise in the decomposed IMFs as prior knowledge to remove noise IMFs, and they then mandatorily remove the first two IMFs ($c_1(t)$ and $c_2(t)$) to reduce the noise components. This method results in the reduction of only higher frequency components of the noise and therefore seems to be insufficient.

We propose another approach to noise reduction based on EMD, as shown in Fig. 5. In our method, the channel selection of speech-IMFs and noise-IMFs, as shown in Fig. 6, is combined with the conventional method to separate noise IMFs from the decomposed IMFs. First, noisy speech $y(t)$ is decomposed by EMD. Next, temporal envelope $e_k(t)$ is extracted from decomposed IMF $c_k(t)$, by the Hilbert transform and low-pass filtering where the cut-off frequency is 20 Hz because modulation index (MI) of lower than 20 Hz is important for speech perception. Next, a modulation filterbank (with a constant bandwidth filterbank) is used to analyze the modulation characteristics of the IMF's temporal envelope $e_{k,m}(t)$. The MI of decomposed IMF $M_{k,m}$ is determined as

$$M_{k,m} = \frac{\max(e_{k,m}(t)) - \min(e_{k,m}(t))}{\max(e_{k,m}(t)) + \overline{e_{k,m}(t)}} \quad (2)$$

Channel selection (Fig. 6) is used to clarify the decomposed IMFs into speech IMFs and noise IMFs based on the modulation characteristics in Eq. (2). Finally, the proposed method resynthesizes restored speech $\hat{x}(t)$ as follows.

$$\hat{x}(t) = \sum_{k \in \mathbf{S}} c_k(t) \quad (3)$$

where $\mathbf{S}$ is a set of speech IMFs.

We examine the differences between the characteristics of the speech and noise IMFs. Here, we focus on the difference

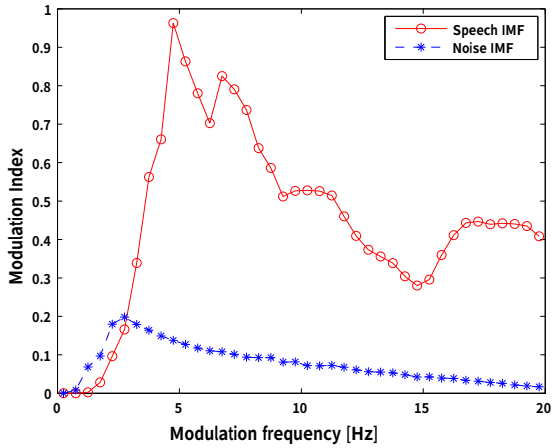

Figure 7: Modulation index of speech and noise IMFs

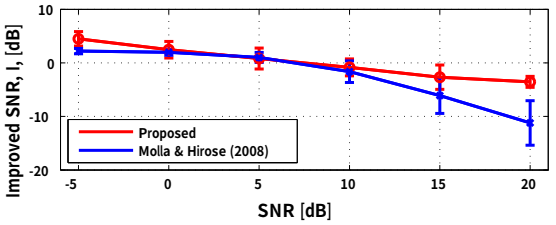

Figure 8: Evaluation result: improved SNR [dB]

between the modulation spectrum of speech and that of noise. It is well known that the dominant modulation frequency on the speech MI is roughly between 2 and 8 [Hz] [4], and we classify the decomposed IMFs of noisy speech in relation to these characteristics. The MIs of the speech IMF and noise IMF are shown in Fig. 7. In the proposed method, speech IMFs are defined as if the MI peak position is in a region between 2 and 8 Hz and the MI peak value is over 0.25.

4. Evaluation

We conducted simulations to determine the effectiveness of the proposed method compared with Molla & Hirose's method [3]. Thirty speech signals (each comprised of three words from five males and five females) from the ATR database a-set [5] were used in these simulations. White noise was added to original speech signals to obtain noisy speech signals with SNRs of -5, 0, 5, 10, 15, and 20 [dB]. An improved SNR, I, was used to evaluate the amplitude information as well as the signal's phase information. Here, I is defined as

$$I = 10\log_{10}\frac{\int \hat{x}^2(t)dt}{\int \hat{n}^2(t)dt} - 10\log_{10}\frac{\int x^2(t)dt}{\int n^2(t)dt} \quad (4)$$

The proposed method and Molla & Hirose's method were applied to 30 noisy speech signals. Improved SNRs for both methods were calculated by using Eq. (4). The results are shown in Fig. 8. In a low SNR condition, the noise reduction was more effective in the proposed method than in Molla

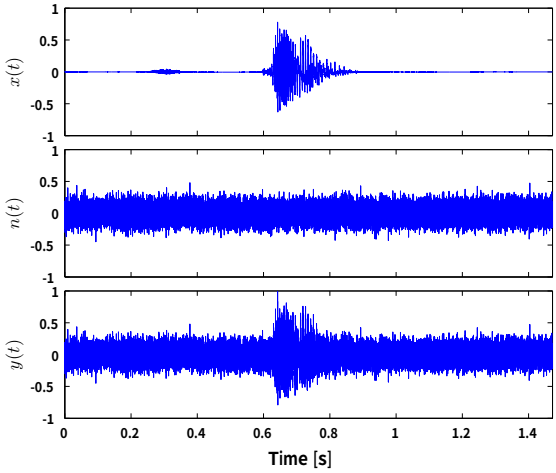

Figure 9: Noisy speech $y(t)$ composed of original speech $x(t)$ and Gaussian noise $n(t)$ (SNR $= 0$ [dB])

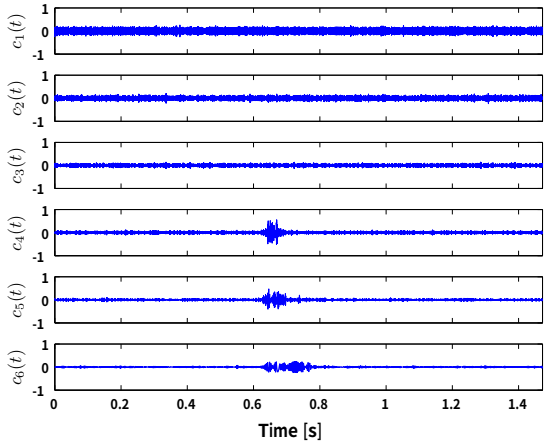

Figure 10: Decomposition of noisy speech $y(t)$ (the first six IMFs: $c_1(t)$, $c_2(t)$, $\cdots$, and $c_6(t)$)

& Hirose's method. In a high SNR condition, Molla & Hirose's method exhibited a speech signal that was over-filtered, which resulted in a restored signal that was corrupted due to over-subtraction. In the same condition, the proposed method reduced the noise components from noisy speech without distortion, because it uses channel selection to reduce only noise-IMFs in the decomposed IMFs.

We illustrate an example of noise reduction using the proposed method. A noisy speech $y(t)$ with a SNR of 0 dB is shown in Fig. 9. The noisy speech is decomposed by EMD and the decomposed IMFs (the first six) are then obtained, as shown in Fig. 10. These results demonstrated how EMD decomposes noisy speech into stationary noise IMFs and non-stationary speech IMFs. The proposed method reduced noise components in the decomposed IMFs and then restored the signal, as shown in Fig. 11.

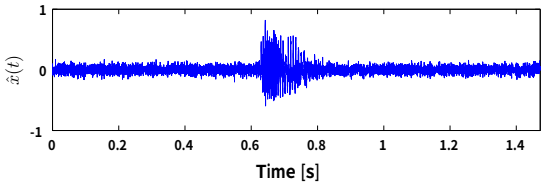

Figure 11: Restored signal $\hat{x}(t)$

5. Conclusion

We investigated the properties of an analysis method based on EMD for non-stationary signals and found that the essence of this method is to represent non-stationary signals as an AM-FM decomposition based on a common temporal envelope. We then investigated the characteristics of the decomposed IMFs from the noisy speech signal and found that EMD decomposes noisy speech into two separate IMFs, speech and noise. We used these findings to develope a noise reduction method based on EMD and then conducted simulations to evaluate its effectiveness in reducing noise components in noisy speech. Results demonstrate that non-stationary speech and stationary noise can effectively be separated by using our proposed method.

Acknowledgments

This work was supported by the Strategic Information and COmmunications R&D Promotion ProgrammE (SCOPE) (071705001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] N. E. Huang et al.: The Empirical Mode Decomposition and the Hilbert Spectrum for nonlinear and non-stationary time series analysis, Proc. the Royal Society: Math., Physi. & Eng. Sci., Vol. A454, pp. 903–995, 1998.

[2] T. Hasan and M. K. Hasan: Suppression of Residual Noise From Speech Signals Using Empirical Mode Decomposition, IEEE Signal Process. Letters., Vol. 16, No. 1, pp. 2–5, 2008.

[3] M. K. I. Molla and K. Hirose: Robust Voiced/Unvoiced Classification of Speech Signal Using Hilbert-Huang Transformation, Signal Process., Vol. 12, No. 6, pp. 473–482, 2008.

[4] S. Greenberg, H. Carvey, L. Hitchcock and S. Chang: Temporal properties of spontaneous speech-a syllable-centric perspective, Phonetics, Vol. 31 No. 3-4, pp. 465–485, 2003.

[5] K. Takeda, Y. Sagisaka, S. Katagiri and H. Kuwabara: A Japanese speech database for various kinds of research purposes, J. Acoust. Soc. Jpn., Vol. 44, No. 10, pp. 747–754, 1988.