Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Signal Processing in Acoustics
Session 2pSP: Acoustic Signal Processing for Various Applications

2pSP5. Wind noise reduction using empirical mode decomposition

Kohei Yatabe* and Yasuhiro Oikawa

***Corresponding author's address: The Department of Intermedia Art and Science, Waseda University, Shinjuku-ku, 169-8555, Tokyo, Japan, k.yatabe@asagi.waseda.jp**

One common problem of outdoor recordings is a contamination of wind noise which has highly non-stationary characteristics. Although there are a lot of noise reduction methods which work well for general kinds of noises, most methods perform worse for wind noise due to its non-stationary nature. Therefore, wind noise reduction need specific technique to overcome this non-stationarity. Empirical mode decomposition (EMD) is a relatively new method to decompose a signal into several data-driven bases which are modeled as amplitude and frequency modulated sinusoids that represent wind noise better than quasi-stationary analysis methods such as short-time Fourier transform since it assumes an analyzing signal as non-stationary. Thus, EMD has a potential to reduce wind noise from recorded sounds in an entirely different way from ordinary methods. In this paper, the method to apply EMD as a wind noise suppressor is presented. The experiment is performed on a female speech superimposed with wind noise, and the results showed its possibility.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800880]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

A wind noise contamination is one of the most common problems in outdoor recordings, where portable devices such as a handy video camera and a smart phone are used without wind shields. However, such small and light-weighted portable devices can hardly mount wind shields which usually degrade their usability. Likewise, it is hard to mount a microphone array afterward even that is a well-known highly effective way to reduce wind noise. Therefore, a signal processing technique to suppress wind noise after a recording is demanded. There exist a number of noise reduction methods working effectively when a signal and noise can be assumed stationary in a short-time period. However, wind noise is a highly non-stationary signal that causes difficulty on a noise estimation and attenuation. Thus, a specific technique for a wind noise reduction is needed.

There are several methods approaching wind noise reduction using spectral subtraction [1], non-negative sparse coding [2], adaptive postfiltering [3] and morphological technique [4]. However, most of those methods use quasi-stationary analysis methods such as short-time Fourier spectra. Hence, there might be a possibility to improve the wind noise reduction methods if an analysis method suitable for a non-stationary signal is available.

Empirical mode decomposition (EMD) is an algorithm for instantaneous frequency analyses working together with Hilbert transform by decomposing a signal into several amplitude and frequency modulated (AM-FM) sinusoids termed intrinsic mode function (IMF) [5]. Since EMD does not need to assume an analyzing signal as stationary, it might have a potential to handle wind noise better than traditional frequency analysis methods. In this paper, the method to apply EMD as a wind noise suppressor is proposed. EMD is combined with a support vector machine (SVM) to discriminate noise-dominant IMFs from speech-dominant IMFs, and each noise-dominant IMF are subtracted from the noisy mixture to obtain a cleaner speech signal. This simple noise components subtraction showed the potential of EMD to handle wind noise as in the experimental results.

EMPIRICAL MODE DECOMPOSITION

EMD decomposes a signal $x(t)$ into IMFs, which are represented as AM-FM sinusoids, and a final residual trend $r(t)$, which is a non zero-mean low-order polynomial component:

$$x(t) = \sum_k IMF_k(t) + r(t). \tag{1}$$

Each IMF is obtained by subtraction of the local mean from a signal. In the original EMD [5], local mean values are estimated from mean envelopes with the following iterative procedure.

1. Identify successive extrema of an input signal $s(t)$.
2. Estimate an upper envelope $e_U(t)$ and a lower envelope $e_L(t)$ by interpolating upper and lower extrema respectively.
3. Calculate a mean envelope $m(t)$ by averaging the upper and lower envelope: $m(t) = (e_U(t) + e_L(t))/2$.
4. Subtract the mean envelope $m(t)$ from $s(t)$ to obtain a proto-IMF $p(t) = s(t) - m(t)$.
5. Assume the proto-IMF $p(t)$ as an input signal and repeat 1-4. If $p(t)$ satisfies certain criteria, treat $p(t)$ as an IMF and end the loop.
6. Treat the residual $r(t) = x(t) - \sum IMF$ as a new input signal and repeat 1-5. When $r(t)$ becomes negligibly small, or when $r(t)$ becomes a monotonic function from which no more IMF can be extracted, end the loop.

TABLE 1: Five possible features for a discrimination between speech-dominant IMFs and wind noise-dominant IMFs where $A(t)$ denotes instantaneous amplitude, $\omega(t)$ denotes instantaneous frequency, and $X(n)$ denotes a magnitude spectrum. Each feature is tested on SVM to reduce the dimension of the feature space.

weighted mean of instantaneous frequency	$\overline{\omega} = \frac{\sum A(t)\omega(t)}{\sum A(t)}$
weighted standard deviation of instantaneous frequency	$\sigma = \sqrt{\frac{1}{T}\sum\left(\frac{A(t)(\omega(t)-\overline{\omega})}{\sum A(t)}\right)^2}$
spectral flatness measure	$SFM = \frac{\sqrt[n]{\prod X^2(n)}}{\frac{1}{N}\sum X^2(n)} = \frac{\exp\left(\frac{1}{N}\sum \ln X^2(n)\right)}{\frac{1}{N}\sum X^2(n)}$
normalized spectral peak	$NSP = \max \frac{\lvert X(n)\rvert}{\sum \lvert X(n)\rvert}$
kurtosis of spectrum	$kurt = \frac{\frac{1}{N}\sum (X(n)-\overline{X})^4}{\left(\frac{1}{N}\sum (X(n)-\overline{X})^2\right)^2}$

Steps 1-5 in the above procedure are so-called sifting process. The standard deviation computed from the two consecutive sifting results $p(t)$ is a widely used criterion for step 5. Moreover, third-order spline is the standard interpolation method for EMD.

There are many variation of the original EMD in order to achieve better separation of each intrinsic mode. In this paper, doubly-iterative EMD [6] which estimates interpolating points as zero-crossings of the first IMF of a first-order derivative of a signal is used in order to obtain a better frequency separation.

CLASSIFICATION WITH SUPPORT VECTOR MACHINE

SVM is a widely used binary linear classifier which learns a maximum-margin hyperplane to classify two classes of data. It is usually combined with kernel method which allows nonlinear classification in a higher dimension feature space. In this paper, Gaussian radial basis function

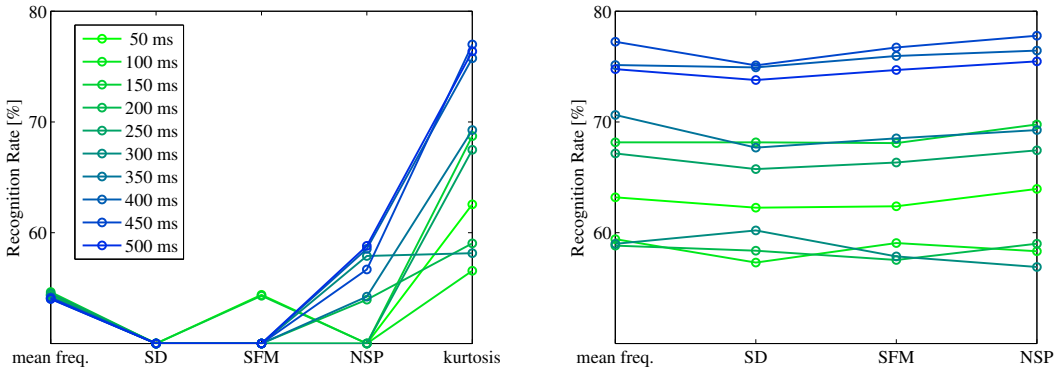
$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{2}$$

is used as the kernel function.

In order to achieve a better classifier, the selection of the features based on recognition rate is performed. Possible features for a discrimination between speech-dominant IMFs and wind noise-dominant IMFs are listed in Table 1.

Since the components of a speech and wind noise should be in different frequency range, it is reasonable to handle IMFs with frequency information. One way to think about frequency information is an arithmetic mean of instantaneous frequency which is suitable frequency representation of an IMF. However, instantaneous frequency is often fluctuate greatly, which is the strong cause of an interpretation error, at the time where instantaneous amplitude is small. Hence, instantaneous frequency is weighted by instantaneous amplitude to calculate the mean frequency. Kurtosis of a spectrum is also an option to deal with frequency information especially in the lower frequency range where wind noise is dominative.

Another criterion for the speech- and wind noise-dominant IMFs discrimination is the degree of frequency modulation which usually lowers tonality. A simple way to handle it is to take a standard deviation of instantaneous frequency which is weighted by instantaneous amplitude to



(a) The recognition rate of the one dimensional feature spaces.

(b) The recognition rate of the two dimensional feature spaces who have kurtosis of a spectrum as the common dimension.

FIGURE 1: The recognition rate of each criteria. A recorded speech and wind noise, whose duration are about three seconds, were divided into two groups: one for the learning and the other for the recognition test. Each group of the signals are segmented into short-time frames to extract the features. The results are shown in the percentage that the constructed classifier correctly recognized the IMFs of the testing frames.

decrease interpretation error. Spectral flatness measure and normalized spectral peak are also possible criteria since they can be used as measures of the degree of frequency modulation.

These criteria are tested and only the effective criteria are used for the recognition. Figure 1(a) illustrates the recognition rate of each criterion for several frame length. From this result, kurtosis of a spectrum is selected as the first criterion. Figure 1(b) illustrates the two dimensional recognition rate where kurtosis of a spectrum is common for every additional criterion. Yet there are a clear dependency on the frame length, normalized spectral peak which produced averagely better recognition rates for each frame length is selected as the second criterion. For more than three features, there were no significant difference of recognition rate compare to the two dimensional feature space. Thus, these two criteria are used as features for the rest of the paper.

WIND NOISE REDUCTION METHOD

EMD decomposes a signal into several IMFs, which are narrower-band components of the signal, from a locally higher frequency component to lower ones. Since IMFs are represented as AM-FM sinusoids, a simple subtraction of noise-dominant IMFs from a noisy speech has an enough potential to suppress wind noise.

Proposed Method

The proposed wind noise reduction method is described as the following procedure:

1. Segment a noisy signal into short-time frames. Each frame may overlap to an another frame.
2. Apply EMD on each frame.
3. Classify each IMF by SVM to discriminate noise-dominant IMFs from speech-dominant IMFs.
4. Subtract noise-dominant IMFs from the noisy speech to obtain a cleaner speech signal.

Figure 2 shows the block diagram of the proposed method.

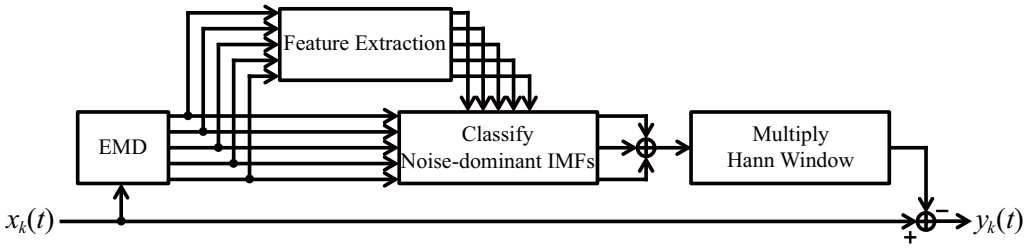
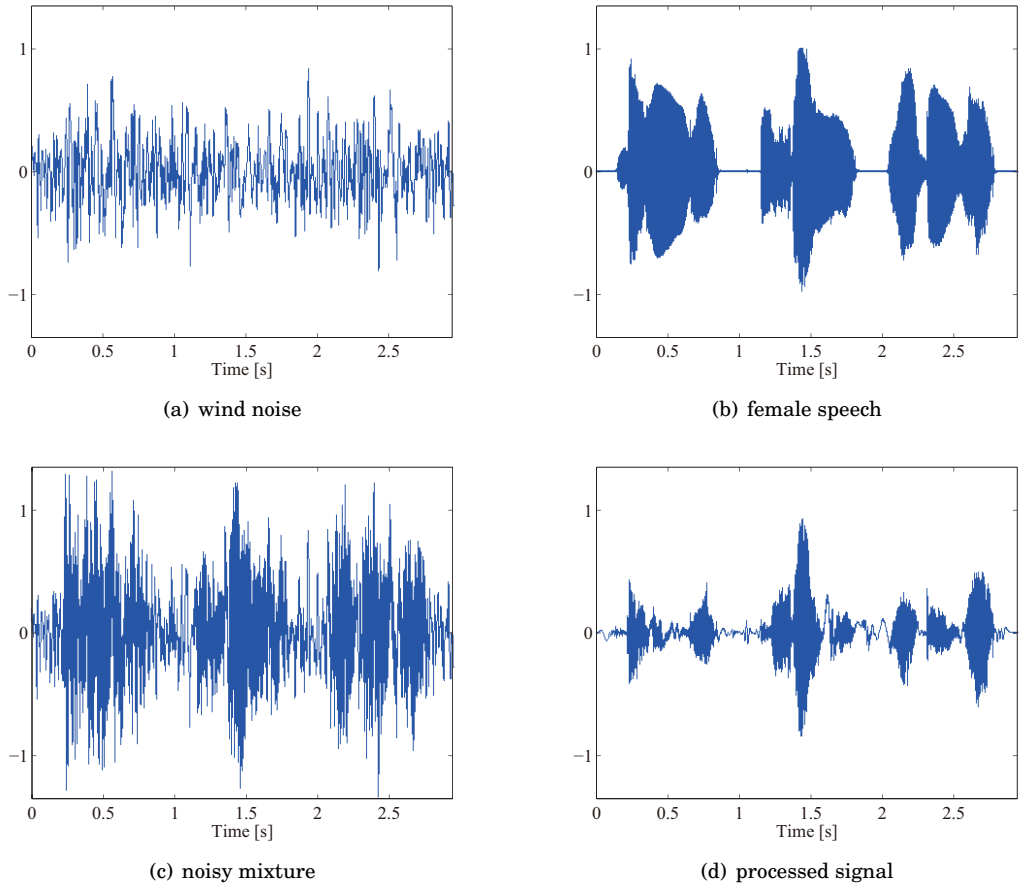


FIGURE 2: The block diagram of the proposed method where $x_k(t)$ denotes k th frame of the input noisy mixture and $y_k(t)$ denotes the processed signal. Each frame is decomposed into IMFs by EMD. Each IMF is classified based on the features explained in the third section. The noise-dominant IMFs are summed up and subtracted from the input signal after the multiplication of Hann window in order to obtain smooth connection between consecutive frames.



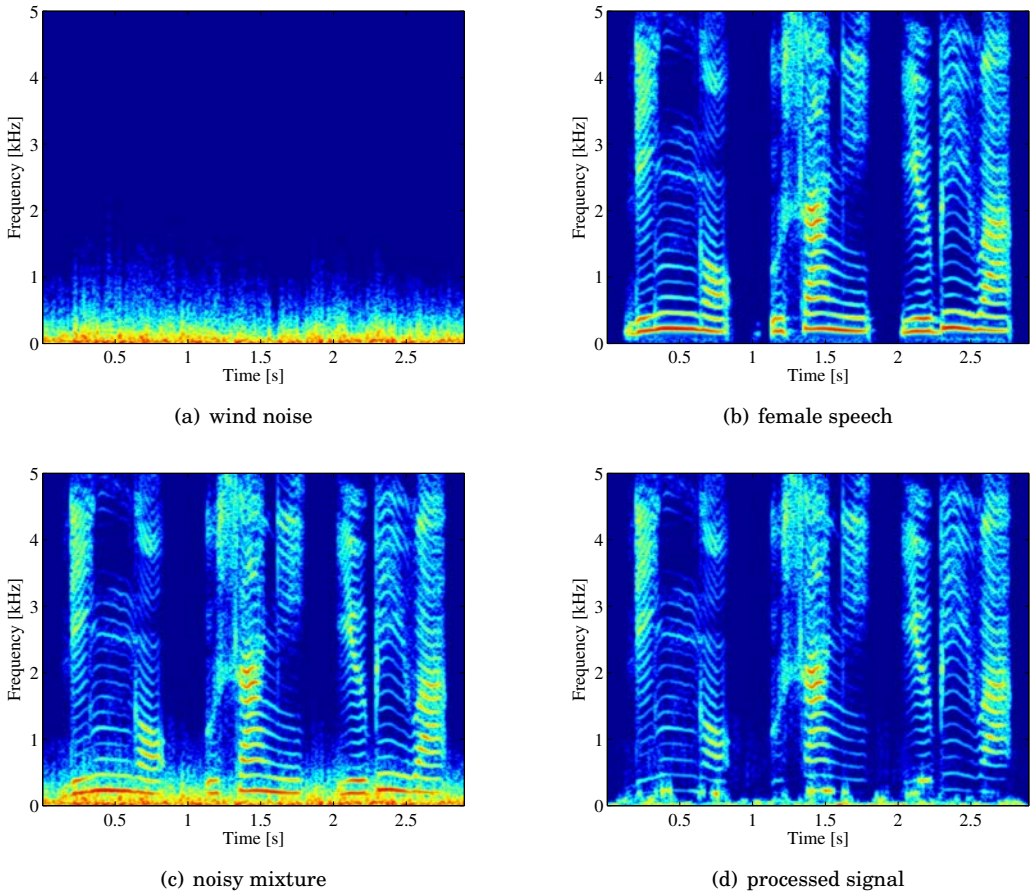
(a) wind noise (b) female speech

(c) noisy mixture (d) processed signal

FIGURE 3: The wave form of the tested signals. (a) Wind noise recorded by a consumer portable recorder at outside without a wind shield. (b) Clean female speech recorded in a sound booth. (c) Noisy speech signal obtained from the superposition of (a) and (b) whose SNR is 0 dB. (d) Processed speech signal obtained by the proposed method whose frame length was set to 150 ms.

Experiments

The proposed method was tested on a noisy speech signal which was artificially composed from a speech only and a wind only signals. Wind noise was recorded by a consumer portable recorder (SONY PCM-D50) at outside without a wind shield, and superimposed on a female speech whose duration is about 3 s. The frame length was set to 150 ms, and consecutive frames were overlapped 75%. Each noise-dominant IMF was multiplied by Hann window before the



(a) wind noise (b) female speech

(c) noisy mixture (d) processed signal

FIGURE 4: The spectrograms of the same signals as in Fig. 3.

subtraction for a smooth connection between the consecutive frames. SVM was trained at the beginning to construct a classifier using another data, which were recorded at the same time as the testing signal, whose duration is about 1 s.

Figure 3 shows the wave form of the testing signals and the processed signal. The noisy mixture was composed by the superposition of the speech and wind noise so that its signal to noise ratio became 0 dB. From Fig. 3(d), it can be seen that each spoken word has a cleaner wave form than the noisy mixture. Figure 4 shows the spectrograms of the same signals as in Fig. 3. It can be confirm that the proposed method reduced the wind noise evidently as in Fig. 4. However, it is also depicted that lower frequency components of the speech signal were subtracted with wind noise.

In order to verify the effect of the method, the signal to noise ratio

$$SNR = 10\log_{10}\left(\frac{\sum s(t)^2}{\sum\left(s(t)-\hat{s}(t)\right)^2}\right) \tag{3}$$

of the various tested signals, where $s(t)$ denotes the original speech signal and $\hat{s}(t)$ denotes the obtained speech signal from a noisy mixture, are calculated as in Table 2.

The proposed method was compared to the spectral enhancement method [7], which uses OM-LSA speech estimator and MCRA noise estimator, whose MATLAB code can be found in [8]. It can be seen that the proposed method obtained competing results when the signal to noise ratio of the noisy mixture is extremely low. This result shows the possibility of EMD as a wind

TABLE 2: SNR of the noisy mixture and the processed signals. The proposed method is compared to the spectral enhancement method using OM-LSA speech estimator and MCRA noise estimator [7].

input [dB]	proposed method [dB]	spectral enhancement [dB]
0	1.51 (+1.51)	14.1 (+14.1)
-5	1.22 (+6.22)	9.46 (+14.5)
-10	1.57 (+11.6)	5.05 (+15.1)
-15	-1.25 (+13.8)	0.40 (+15.4)
-20	-3.78 (+16.2)	-4.37 (+15.6)

noise suppressor. However, its noise reduction effect decreases as the signal to noise ratio of the noisy mixture increases. This decay of the performance can be due to the low recognition capability of the classifier. Nevertheless, there are a trade-off between the frame length and the recognition rate. As in Fig. 1, the recognition rate generally becomes higher when the frame length becomes longer. On the other hand, the quality of the decomposition of EMD becomes higher when the frame length becomes shorter due to the so-called mode mixing phenomenon [9] which is a interfusion of speech components and wind noise components within an IMF in this case. There remains an extensive room to improve the proposed method by upgrading accuracy of both classification and decomposition.

CONCLUSIONS

In this paper, the method to apply EMD to a wind noise reduction was introduced. Although the basic idea of the proposed method which is subtractions of noise-dominant IMFs from a noisy mixture is quite simple, the experimental result showed the possibility of EMD as a wind noise suppressor especially in the situation that a speech signal is extremely contaminated by wind noise. In order to achieve a better performance, the accuracy of both decomposition and classification of speech- and wind-related components must be enhanced. This might be achieved by a modification of EMD specialized in a characteristic of wind noise, and construction of a better classifier based on more effective features or using another information such as statistics.

REFERENCES

[1] S. Kuroiwa, Y. Mori, S. Tsuge, M. Takashina, and F. Ren, "Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure," Int. Conf. Commun. Technology (ICCT), 1–5 (2006).

[2] M. N. Schmidt, J. Larsen, and F. T. Hsiao, "Wind noise reduction using non-negative sparse coding," Workshop Mach. Learning Signal Process., 431–436 (2007).

[3] E. Nemer, and W. Leblanc, "Single-microphone wind noise reduction by adaptive postfiltering," Workshop Applicat. Signal Process. Audio Acoust., 177–180 (2009).

[4] C. Hofmann, T. Wolff, M. Buck, T. Haulick, and W. Kellermann, "A morphological approach to single-channel wind-noise suppression," Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC), 1–4 (2012).

[5] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond. A **454**, 903–995 (1998).

[6] Y. Kopsinis, and S. McLaughlin, "Improved EMD using doubly-iterative sifting and high order spline interpolation," EURASIP J. Adv. Signal Process. **2008**, 1–8 (2008).

[7] I. Cohen, and B. Berdugo, "Speech Enhancement for Non-Stationary Noise Environments," Signal Process. **81**, 2403–2418 (2001).

[8] http://webee.technion.ac.il/people/IsraelCohen/
[9] N. E. Huang, M. C. Wu, S. R. Long, S. S. Shen, W. Qu, P. Gloersen, and K. L. Fan, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," Proc. R. Soc. Lond. A **459**, 2317–2345, (2003).