ICSV14
Cairns • Australia
9-12 July, 2007

IN-CAR SPEECH ENHANCEMENT USING ENSEMBLE EMPIRICAL MODE DECOMPOSITION

Jiafang Zhang, Hai Huang and Xiangxian Chen

Department of Instrument Engineering, Zhejiang University
Hangzhou 310027, P.R.China
jiafang322@gmail.com

Abstract
The performance of the human-machine dialogue at in-car environment is considerably deteriorated by background noises and other disturbances. In this paper, the authors present an in-car speech enhancement (ICSE) method to improve quality of speech signals suffering the in-car noises. The method is based on a novel signal processing technology called the ensemble empirical mode decomposition (EEMD). By using EEMD, the noisy speech signals are decomposed as a set of intrinsic mode functions (IMF). Then, the nonlinear least-square estimation and signal-to-noise ratio (SNR) are employed to find out the optimal weighting coefficients of those IMFs dominating the speech signals. Finally, the enhanced speech signals, where in-car noises are suppressed, are obtained by the reconstruction technique based on the weighting IMFs. Results of the work show that, EEMD is an effective technology of separating pure speech from in-car noises, and the weighting coefficients proposed in this study are also effective for noises reduction as considering the similarity of the signal waveform.

Keywords: Speech enhancement; Ensemble empirical mode decomposition (EEMD); In-car noise

1. INTRODUCTION

In-car human-machine speech-based interactions are becoming increasingly important, and finding more and more applications. For example, for safety sake it is desirable for car-drivers to use voice to setup and locate destinations or implement other control orders in GPS navigation systems rather than use their hands and eyes. However, the performance of in-car human-machine dialogue is considerably deteriorated by background noises and other disturbances [1]. Therefore, to develop effective speech front-end processing techniques used for suppressing the background noises are the key step in order to improve the speech intelligibility at the in-car noisy environment.

The in-car noises originate from engines, pumps, audio equipments, road, wind and air-conditioning, radio and communication, and so on. They are usually time-varying and non-stationary [1, 2]. In order to extract the original voice from the noisy speech, some of in-car speech enhancement techniques were developed in recent years, where the most

common approaches for one-channel speech enhancement are spectral subtraction method, adaptive filter and signal subspace method. These methods produce certain effects in practical uses, but also have some inherent limitations. Some authors have made improvements to these methods since then. Ercelebi proposed a new method of speech enhancement based on time-frequency analysis and adaptive digital filtering [3]. Huang *et al*. suggested an energy-constrained signal subspace (ECSS) method of automatic speech enhancement and recognition under additive noise condition [4]. Hu *et al*. extended three complementary spectral subtraction techniques: the spectral smoothing, the formant intensification and the comb filtering [5]. Westerlund *et al*. suggested a sub-band signal decomposition and adaptive gain equalizer based on short-term SNR estimation [6]. These methods make certain improvements, but for effectively recognizing the non-stationary and nonlinear speech signals [7] in bad (sometimes harsh) in-car noisy environment, further investigations are still needed.

Empirical mode decomposition (EMD) is a new signal processing method used mainly to analyse non-stationary signals [8]. It is an intuitive, direct, a posteriori and adaptive signal processing method. The basic component functions of EMD are derived from the analysed signals. These basic functions well describe the inherent characteristics of the analysed non-stationary signal. To overcome so-called mode mixing problem, a noise-assisted data analysis method, named as ensemble empirical mode decomposition (EEMD), was developed [9]. EEMD employs statistical characteristics of the white noise and the scale separation principle of EMD, eliminating the mode mixing automatically. Hence it makes the results more accurate and more explicit in physical sense.

In this paper, authors introduce an in-car speech enhancement (ICSE) method based on EEMD. A satisfying performance of the speech enhancement is achieved by this technique. This paper is organized as follows. In Section 2, basic principles of EMD and EEMD are briefly introduced. In Section 3, in-car speech enhancement method developed by the authors based on EEMD is presented. In Section 4 the experiment setup and simulation results to prove validity of the ICSE method are shown. Finally, Section 5 provides conclusions.

2. BASIC PRINCIPLES OF EMD AND EEMD

Hilbert-Huang transform is an adaptive signal processing method used to analyse the non-stationary signals [8]. Analysed signals can be decomposed into a set of intrinsic components named intrinsic mode functions (IMFs) by a shifting process, which is described as follows. Firstly, the upper and lower envelopes of the signals x(t), as well as their mean value $m_1(t)$, are calculated respectively. Then the difference $h_1(t) = x(t) - m_1(t)$ is obtained. However, $h_1(t)$ rarely satisfies the IMF properties. Therefore, the sifting has to be implemented for another time, until obtained difference, say $h_{1k}(t)$, satisfies the IMF properties, then it can be taken as the first IMF component, denoted by $c_1(t) = h_{1k}(t)$. Next, taking rest data $r_1(t) = x(t) - c_1(t)$ as 'new' signals and implementing the same step on it, we can obtain the second IMF $c_2(t)$. This procedure should be repeatedly used. Finally, $c_1(t)$, $c_2(t)$, …, $c_n(t)$ and the residue $r_n(t)$ are all defined. When the decomposition procedure finished, the analysed signals can be expressed as

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{1}$$

where $r_n(t)$ is a monotonic function which usually can be negligible in signal analysis, and n is the numbers of IMFs.

EEMD is an improvement of EMD. It can eliminate the mode mixing automatically, and guarantee the accuracy of the follow-up data treatments.

The principle of EEMD can be briefly presented as follows. By adding white noise, the filter bank can separate the whole time-frequency space uniformly with the constituting components of different scales. When the signal is added to this uniformly distributed white background, the signal in different scales is automatically projected onto proper scales of reference founded by the white noise in the background. Because each of the noise-added decompositions includes the signal and the added white noise, each individual trial may produce very noisy results. For the noise in each trial is different in separate trials, it can be cancelled out in the ensemble mean of enough trails. At the end, the ensemble mean is considered as the true answer, for the only persistent part is the signal as more and more trials are added in the ensemble to cancel the added noise.

The procedure of EEMD consists of 4 steps as follows. Step 1: adding a white noise series to the analysed data. Step 2: decomposing the data with added white noise into IMFs. Step 3: using another white noise and repeating step 1 and 2. Step 4: when the employed white noise series enough, calculating ensemble means of the obtained IMFs and taking them as final IMF results.

3. IN-CAR SPEECH ENHANCEMENT METHOD

In this section, an ICSE method based on EEMD is introduced. According to the method, by using nonlinear least-square estimation and signal-to-noise ratio (SNR) to find out some optimal weighting coefficients of those IMFs dominating the speech signals, one can obtain enhanced speech signals, where in-car noises are suppressed, by means of the reconstruction technique based on the weighting IMFs.

3.1 Optimal Number of IMFs

EEMD extracts a set of IMFs from the analysed speech. Due to the adaptability of the constant ratio band-pass filters (that is, the centre frequency, bandwidth, and the groups of the filters are determined by the signal itself) [10, 11], and different characteristics between speech and noise, they can be well decomposed in IMFs. Speech signals concentrate in lower IMFs, while the in-car noises concentrate in higher IMFs. According to experiment results, the first three IMFs are considered as the speech signals and are weighted and added to suppress in-car noises, and then the pure speech is obtained.

3.2 Optimal Weighting Coefficients

How to use nonlinear least-square estimation and signal-to-noise ratio (SNR) to find out the optimal weighting coefficients of the IMFs? It is as follows: 1) by using nonlinear least square estimation in a lot of experiment data, the weighting coefficient of the third IMF was determined as a_3=1.2; 2) the weighting coefficients of the first and second IMF, a_1 and a_2 are determined in accordance with the ratio of SNR, that is, $a_1 = a_3 \times \mathrm{SNR\_IMF}_1 / \mathrm{SNR\_IMF}_3$ and $a_2 = a_3 \times \mathrm{SNR\_IMF}_2 / \mathrm{SNR\_IMF}_3$, where $\mathrm{SNR\_IMF}_1$, $\mathrm{SNR\_IMF}_2$ and $\mathrm{SNR\_IMF}_3$ are the SNR of the first, second and third IMF respectively. Hence, the processed speech signal can be expressed as:

$$x_e(t) = \sum_{i=1}^{3} a_i c_i(t) \tag{2}$$

3.3 Post-treatment

Furthermore, before use of the weighted first three IMFs, small energy parts of every IMF are removed to completely eliminate the noises in quiet signal segments. Because in a certain time interval there exist speech segments in some of the IMFs and quiet segments in other IMFs, with the treatment the noises in speech segments can be suppressed and the quality of speech improved further.

The signals through the above processing are finally filtered by a low-pass FIR filter to remove the high-frequency noises in the first three IMFs.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Conditions and Signal Acquisition

For the study a small speech database including noises in the automobile environment and pure speech signals in quiet background was set up in the authors' lab. Ten types of in-car environmental noises were recorded, which are classified based on different car states, such as the car stopping or running, the car speeding up, different traffic conditions, the air-conditioning on or off, the windows open or closed. In quiet background, 30 voice words read by six individuals (5 males and 1 female) were recorded in twice. The car used in the experiment is a MAZIDA 323 sedan. The microphone and acoustic scales used are B&K 4155 and B&K 2230. The recording software is Goldwave. Each speech signal was sampled at a frequency of 20 kHz and memorized in 16 bits.

To implement the study of in-car speech with different SNR, noisy speech can be obtained by adding the noise to the pure speech signal in different ratio:

$$Y_s = a \times X_p + b \times X_n \tag{3}$$

where Y_s is noisy speech, X_p is pure speech, X_n is in-car noise, a and b are coefficients with $a+b=1$.

SNR is employed to measure the quantity of added noise in noisy speech. SNR is defined as follows:

$$SNR = 10 \times \lg \frac{E_p}{E_n} = 10 \times \lg \frac{(E_s - E_n)}{E_n} \tag{4}$$

where, E_p is pure speech energy, E_n is noise energy, and E_s is noisy speech energy.

4.2 Analysis of Speech without In-car Noise

As a basic function of the EEMD, each IMF factually and completely expresses an independent component of the speech because of its adaptability, and materializes the sound waveform and voice. In the frequency domain, these IMFs give the speech signal components in different frequency bands. As the order of IMF increases, the corresponding frequency interval decreases. Table 1 shows the energy (mean square value) distribution over the IMFs extracted from various speeches by EEMD. It can be seen that the first several IMFs contain the most energy. The energy sum of the first three IMFs even is 60% ~ 80% of the total

energy, where energy of IMF_3 accounts for almost half of the total energy. It means that the first three IMFs together basically constitute the original pure speech.

Table 1. Energy distribution over the IMFs extracted by EEMD from various speeches

		IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
sousuo	energy(10^{-3})	0.1653	0.0638	3.8033	1.2703	0.2894	0.0148	0.0034	0.0016	0.0031	0.0015
	percentage	2.94%	1.14%	67.71%	22.62%	5.15%	0.26%	0.06%	0.03%	0.06%	0.03%
xiaoheshan	energy(10^{-3})	0.0416	0.2017	1.2797	0.5701	0.1881	0.1405	0.0026	0.0004	0.0005	0.0007
	percentage	1.71%	8.31%	52.75%	23.50%	7.75%	5.79%	1.07%	0.02%	0.02%	0.03%
yiyuan	energy(10^{-3})	0.1789	0.4067	2.849	0.9044	0.5199	0.0122	0.0019	0.0010	0.0026	0.0020
	percentage	3.66%	8.34%	58.40%	18.54%	10.66%	0.25%	0.04%	0.02%	0.05%	0.04%
youyi	energy(10^{-3})	0.0117	0.1603	3.2193	0.9449	0.6676	0.0892	0.0031	0.0045	0.0030	0.0019
	percentage	0.23%	3.13%	63.06%	18.51%	13.08%	1.75%	0.06%	0.09%	0.06%	0.04%
liuxia	energy(10^{-3})	0.0235	0.0058	0.1223	0.0759	0.0137	0.0011	0.0003	0.0001	0.0001	0.0001
	percentage	9.68%	2.39%	50.34%	31.24%	5.63%	0.46%	0.13%	0.04%	0.04%	0.06%

4.3 Analysis of In-car Noise

Table 2. Energy distribution over the IMFs extracted by EEMD from typical in-car environments

		IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
stopped with the engine running, windows closed, and air-conditioning on	energy(10^{-5})	0.0066	0.0057	0.0272	0.8721	0.8170	0.7227	0.2179	0.0899	0.0062	0.0017
	percentage	0.24%	0.21%	0.68%	31.52%	29.53%	26.12%	7.87%	3.25%	0.55%	0.06%
running in the town traffic at the speed of 60 km/h, windows closed , and air-conditioning off	energy(10^{-3})	0.0014	0.0008	0.0125	0.0585	0.0587	0.1548	0.3355	0.0074	0.0011	0.0016
	percentage	0.23%	0.12%	1.98%	9.24%	9.29%	24.48%	53.06%	1.17%	0.17%	0.26%
running in town traffic at the speed of 70 km/h, windows closed , and air-conditioning on	energy(10^{-3})	0.0007	0.0003	0.0012	0.0594	0.1145	0.1456	0.0380	0.0034	0.0009	0.0004
	percentage	0.20%	0.09%	0.32%	16.30%	31.43%	39.96%	10.43%	0.93%	0.25 %	0.10%
running in the freeway at the speed of 100 km/h, windows open, and air-conditioning off	energy(10^{-3})	0.0069	0.0039	0.0074	0.1460	0.4133	0.5455	1.1023	0.1394	0.0170	0.0008
	percentage	0.29%	0.17%	0.31%	6.13%	17.36%	22.86%	46.28%	5.85%	0.71%	0.03%
running in the freeway at the speed of 100 km/h, windows closed , and air-conditioning off	energy(10^{-3})	0.0010	0.0005	0.0096	0.0972	0.1808	0.1287	0.0505	0.0094	0.0007	0.0004
	percentage	0.20%	0.09%	2.01%	20.31%	37.76%	26.88%	10.55%	1.97%	0.15%	0.08%

Table 2 shows the energy distribution over the IMFs for the typical in-car noisy environments. The energy distribution and dominated order number for different in-car noises are not the same. The energy mainly concentrates in the IMFs with order 4 to 7, and this four IMFs account for almost 90% of the total energy.

4.4 Analysis of the Speech with In-car Noise

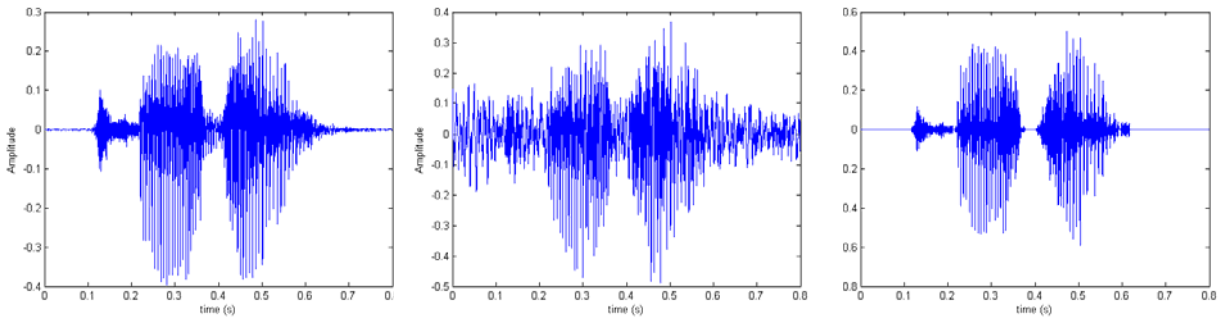
Table 3 shows the energy distribution over the IMFs extracted from various noisy speeches by EEMD, where the SNR is 0, the car runs on the highway with the speed of 100 km/h, and the air-conditioning and the windows are both closed. Comparing with Table 1, it can be seen that IMF_1~IMF_3 are very close in energy, and from IMF_4, especially in IMF_5~IMF_7, the energy of noisy speech is much larger than the energy of pure speech. In contrast with the decomposition results from different noisy environments shown in Table 2, although energy distribution over IMF_1~IMF_3 is basically the same, the energy distribution over IMF_4~IMF_7 is vary different.

Table 3. Energy distribution over the IMFs extracted by EEMD from various noisy speeches

SNR=0		IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
sousuo	energy(10^{-3})	0.1697	0.0539	3.4043	1.7078	2.1885	1.7293	0.3168	0.0642	0.0171	0.0058
	percentage	1.76%	0.56%	35.25%	17.68%	22.66%	17.91%	3.28%	0.67%	0.18%	0.06%
xiaoheshan	energy(10^{-3})	0.0443	0.1430	1.1527	0.9533	1.2165	0.9684	0.1522	0.0350	0.0059	0.0020
	percentage	0.95%	3.06%	24.67%	20.40%	26.03%	20.72%	3.26%	0.75%	0.13%	0.04%
yiyuan	energy(10^{-3})	0.1318	0.4602	2.7305	1.4137	2.904	1.917	0.2233	0.0337	0.0093	0.0028
	percentage	1.34%	4.68%	27.79%	14.39%	29.55%	19.51%	2.27%	0.341%	0.10%	0.03%
youyi	energy(10^{-3})	0.0195	0.1202	2.8885	1.5339	2.9207	2.0448	0.2681	0.0673	0.0281	0.0018
	percentage	0.20%	1.21%	29.20%	15.51%	29.52%	20.67%	2.71%	0.68%	0.28%	0.02%
liuxia	energy(10^{-3})	0.0267	0.0051	0.1135	0.1192	0.1053	0.0935	0.0145	0.0030	0.0010	0.0005
	percentage	5.54%	1.06%	23.53%	24.71%	21.83%	19.39%	3.01%	0.62%	0.21%	0.10%

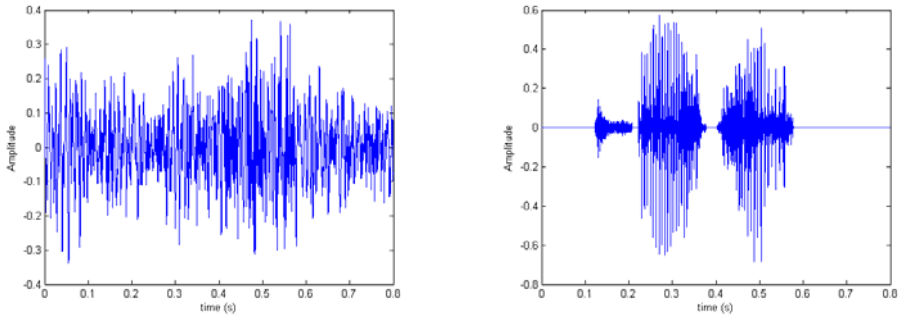
These results obtained by using the EEMD on the in-car noises and speeches indicate that, almost half of the speech energy concentrates on IMF_3, and the energy of the first three IMFs accounts for 60%~80% of the total speech energy. Noises mainly distribute in IMF_4~IMF_7, and their energy sum accounts for 90% of the total noise energy. It shows that EEMD method can effectively separate pure speech from in-car noisy speech.

4.5 Experimental Results



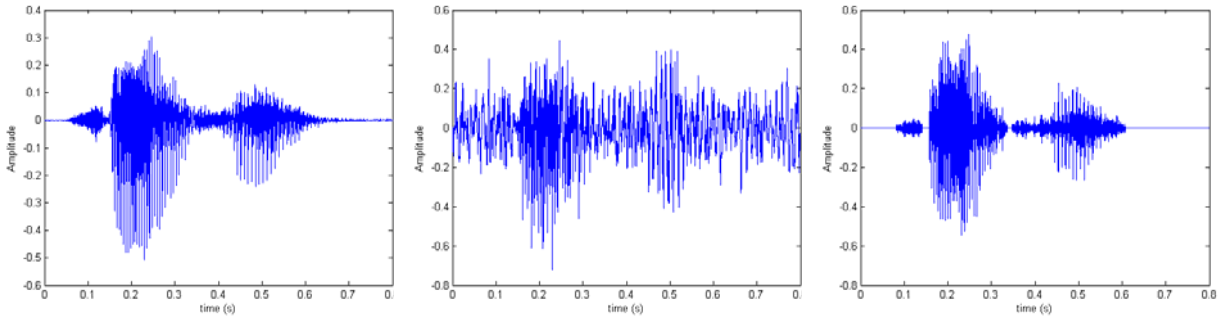
(a) (b) (c)

Figure 1. Speech enhancement process of speech signal "chazhao"(SNR＝0）: (a) the waveform of pure speech, (b) the waveform of noisy speech, (c) enhancement speech.



(a) (b)

Figure 2. Speech enhancement process of speech signal "chazhao"(SNR＝-5）: (a) the waveform of noisy speech, (b) enhancement speech



(a) (b) (c)

Figure 3. Speech enhancement process of speech signal "shanghai"(SNR＝-5): (a) the waveform of pure speech, (b) the waveform of noisy speech, (c) enhancement speech.

Figure 1 shows the speech enhancement process of speech signal "chazhao", where SNR is 0, car runs on the highway with the speed of 100 km/h, and air-conditioning and windows are both closed. It can be seen that from the similarity of signal waveforms shown in Figure 1, this process can effectively suppresses the noises.

Figures 2(a) and 2(b) show respectively the waveforms of noisy and enhanced speech “chazhao”, where the SNR is -5. Figures 3(a) - 3(c) show respectively the pure, noisy and enhanced speech “shanghai”, where the noise environment is that car runs in the town traffic with the speed of 70 km/h, windows closed, air-conditioning on, and SNR=-5. They all give an effective noise reduction.

5. CONCLUSION

In this study, based on the ensemble empirical mode decomposition (EEMD) technique, the authors develop an improved ICSE method used to enhance the speech seriously polluted by the in-car noises. EEMD is a very effective technology for separating pure speech from the in-car noises. The weighting coefficients determined by EEMD from experimental results are efficient for noises reduction, as shown in the similarity of the extracted signal waveforms. The experimental results demonstrated that this method can be applied in practice to effectively suppress (even eliminate) in-car noises, enhance voice, and keep the reality and natural characteristics of the non-stationary, non-linear speech.

ACKNOWLEDGMENT

This work was supported by the Y. C. Tang disciplinary development fund from Zhejiang University.

REFERENCES

[1] H. Abut, J. H. L. Hansen and K. Takeda, DSP for in-vehicle and mobile systems, *Springer*, 2005.

[2] L. D. Poulat, Robust speech recognition techniques evaluation for telephony server based in-car applications. *ICASSP* 2004, I-65-I-68.

[3] E. Ercelebi, Speech enhancement based on the discrete Gabor transform and multi-notch

adaptive digital filters. *Applied Acoustics* **65,** 739-762 (2004).

[4] J. Huang and Y. X. Zhao, An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises. *Speech Communication* **26,** 165-181(1998).

[5] H. T. Hu, F. J. Kuo and H. J. Wang, Supplementary schemes to spectral subtraction for speech enhancement. *Speech Communication* **36,** 205-218 (2002).

[6] N. Westerlund, M. Dahl and I. Claesson, Speech enhancement for personal communication using an adaptive gain equalizer. *Signal Processing* **85,** 1089-1101 (2005).

[7] A. Kumar and S. K. Mullick, Nonlinear Dynamic Analysis of Speech, *J Acoust Soc Am*, 100:615, 1996.

[8] N. E. Huang, *et al*, 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.

[9] Z. Wu and N. E. Huang, Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method, *Proceedings of the first international conference on the advance of Hilbert-Huang Transform and it's applications*, National Central University, Taiwan, 2006.

[10] P. Flandrin, G. Rilling and P. Goncalves, Empirical mode decomposition as a filter bank. *IEEE signal processing letters*, 2004, 11(2): 112-114.

[11] Z. Wu and N. E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society*. London: The Royal Society, 460: 1597-1611, 2004