Science Arena Publications
Specialty Journal of Electronic and Computer Sciences
Available online at www.sciarena.com
2016, Vol, 2 (1): 56-60

Noise Reduction in Cochlear Implant using Empirical Mode Decomposition

Marjan Mozaffari Legha[1], Akram Soltanipursardo[2], Abdollah Navardimard[3]
[1]*Medical Physics & Biomedical Engineering Department, Shahid Beheshti University of Medical Sciences, Tehran, Iran.Email:* Mozaffarilegha.m@gmail.com
[2]*Department of Power Engineering, Institute of Higher Education Javid, Jiroft, Iran.*
Email: S.soltani3397@gmail.com
[3]*Department of Power Engineering, Institute of Higher Education Javid, Jiroft, Iran.*
Email: Abdullh.navardi@gmail.com

Abstract: Cochlear implants are being widely used for the patients with severe to profound sensori-neural hearing loss. Speech coding algorithms play an important role in improving the performance of cochlear implant. At recent years, the performance of CI has been improved for most users under the silent environment. However, as the background noise level increases, speech recognition scores are degraded considerably. In this paper, the Empirical Mode Decomposition and a selected modes approaches are applied as a speech enhancement method for cochlear implants. This algorithm is developed to extract features, called intrinsic mode functions, by a sifting process. Then, IMFs are selected based on CMSE criteria to decrease the noise effect. Also, the Choi-Williams time-frequency technique is applied to extract different components of the resulting signal. Finally, performance of this algorithm in terms of correlation analysis was compared to continuous interleaved sampling (CIS), frequency amplitude modulation encoding (FAME) and Hilbert Huang Transform Stimulating (HHTS) strategies. The results showed the highest correlation coefficient between spectrum of synthesized signal and original speech with proposed method.
Keywords *: Noise, Cochlear, algorithm, implant*

INTRODUCTION

Cochlear implant (CI) is an electronic prosthetic device surgically implanted into the inner ear for restoring some degree of hearing of profoundly deaf patients with sensory-neural origin[1]. It includes internal and external components. The external part consists of a microphone that picks up sound, a signal processor that converts sound into electric impulses, and a transmitter that is magnetically attached to the internal device to which it transmits the electric impulses via radio waves. The impulses are sent to an array of electrodes, which are surgically inserted into the cochlea. The electrodes stimulate the auditory nerves, providing auditory information to the brain[2].Most CI users achieve 80% word recognition scores in quiet listening conditions[3]. However, speech recognition scores are degraded in noisy conditions[4]. Several studies have been proposed to develop speech processing techniques for CI. In CIS strategy, envelope characteristics of speech signal is extracted [5].It utilizes a filter-bank for the frequency decomposition of incoming speech which is a simplification of frequency decomposition function of biological cochlea. Outputs from each channel of the filter-bank are used to modulate the amplitudes of electrical stimulation pulses. In FAME strategy, envelope, frequency and phase information are extracted. This algorithm provides too much indiscriminate information. These techniques are not successful in providing time and frequency resolutions at the same time. Wavelet Transform (WT) overcome the limitations of the previous methods by providing both time and frequency resolutions [6]. However, it suffers to analyze non-stationary signals like speech and depends on the basis wavelet. In the last decade, a new nonlinear technique, termed empirical mode decomposition (EMD), has been introduced by N. E. Huang et al. [7]for adaptively representing non stationary signals. The most important characteristic of EMD is that the basis functions are directly derived from the speech signal itself. HHTS strategy[8] is used to encode both temporal envelope and instantaneous frequency (IF) of input speech signal for CI. While this strategy has allowed cochlear implant users to achieve good speech recognition in quiet, their performance in noise is severely compromised. Also, an IF has a true meaning only for mono component signals, where there is only one frequency or at least a narrow range of frequencies varying as a function of time. Speech signals do not show these necessary characteristics. In this paper, in the first time, we propose a new speech enhancement approach based on the EMD technique and a CMSE criteria applied to select modes. The basic idea is to reconstruct the signal with IMFs by selecting only IMFs that satisfy CMSE criteria. In the second time, the Choi-Williams time-frequency technique is applied to extract different components of the resulting signal. The proposed algorithm can effectively reduce noise in comparison with HHTS strategy. This paper is outlined as follows. Section 2 describes theoretical overview of EMD (Section 2.1), noise reduction with EMD and CMSE algorithm (Section 2.2), CHOI-WILLIAMS Time-Frequency Technique (Section 2.3) and stimulation of new denoising EMD Stimulating Algorithm for CIs (Section 2.4) are described. Section 3 covers the results. Section 4 devotes to the conclusion.

1. Material and Methods

2.1. Theoretical overview of EMD

The Empirical Mode Decomposition (EMD) has been proposed as an adaptive time-frequency data analysis method[7]. This adaptive technique is derived from the simple assumption that any signal consists of different intrinsic mode functions (IMF) each of them representing an embedded distinctive oscillation on a separated time-scale. An IMF is defined by two criteria: i) the number of extrema and of zero crossings must either equal or differ at most by one, and, ii) at any instant in time, the mean value of the envelope defined by the local maxima and the envelope of the local minima is zero. The following plan offers an idea about the principle algorithm of the EMD:

1. **Initialize**$r_0(t) = x(t); j = 1$
2. **Extract** the j_th IMF:
(a) **Initialize**$h_0(t) = r_j(t); k = 1$
(b) Locate local **maxima** and **minima** of $h_{k-1}(t)$
(c) Cubic **spline interpolation** to define upper and lower envelope of $h_{k-1}(t)$
(d) Calculate **mean**$m_{k-1}(t)$ from **upper and lower envelope** of $h_{k-1}(t)$
(e) **Define**$h_k(t) = h_{k-1}(t) - m_{k-1}(t)$
(f) If **stopping criteria** are satisfied then $h_j(t) = h_k(t)$ else goto 2. (b) with $k = k + 1$
3. **Define**$r_j(t) = r_{j-1}(t) - h_j(t)$
4. If $r_j(t)$ still has at least two extrema then goto 2. (a) with $j = j + 1$ else theEMD is finished
5. $r_j(t)$is the residue of $x(t)$

At the end of this numerical sifting process the signal $x(t)$ can be expressed:

$$x(t) = \sum_{j=1}^{n} h_j(t) + r_n(t)$$

Where$h_j(t)$ indicates the j_th IMF, n as the number of sifted IMF and $r_n(t)$denotes a residue which can be understood as the trend of the signal.

1.2. Noise reduction with EMD and CMSE algorithm

Consider a noise-contaminated speech model described by (1):

$x[n] = s[n] + t[n]$ (1)

Where$x[n]$is the noisy speech signal, $s[n]$is the original noise-free speech, and $t[n]$is the noise source. With calculating IMF of noisy speech signal and comparison them together, it is revealed that when noise is added to the clean speech, the first few IMFs contain most of the noise energy as well as some of the speech. However, it can also be seen that the EMD decomposition drives a substantial amount of the speech energy to latter IMFs along with some residual noise. The mode selection method is based on this assumption that the first IMFs (high-frequency modes) are mostly dominated by noise and are not representative for information specific to the original signal. Thus, the enhanced signal is reconstructed only by a few IMFs in which pure signal mostly predominates. In fact, there will be a mode, $IMF_{ks}(t)$ from which the energy distribution of the original signal is greater than the noise. The simple of this approach is to set to zero the first $Ks-1$IMFs [9].

Another method is to find an approximation of the original signal $x(t)$that minimizes the mean square error (MSE) defined by[10]:

$$MSE(x,\tilde{x}) \triangleq \frac{1}{N}\sum_{i=1}^{N}[X(t_i)-\tilde{x}(t_i)]^2 \qquad (2)$$

Where $=[x(t_1),x(t_2),\dots x(t_N)]^t$,$\tilde{x}=[\tilde{x}(t_1),\tilde{x}(t_2),\dots\tilde{x}(t_N)]^t$ and N is the signal length. After decomposing the signal $x(t)$through the EMD algorithm, $\tilde{x}(t)$is reconstructed as follows:

$$\tilde{x}_{ks}(t)=\sum_{k=ks}^{n} imf_k(t)+r_n(t), \quad ks\in\{2,3,\dots,n\} \quad (3)$$

Since, the original signal $x(t)$ is unknown; the MSE cannot obviously be calculated. Thus a distortion measure, termed consecutive MSE (CMSE) [9]is used. The CMSE is defined as:

$$CMSE(\tilde{X}_K,\tilde{X}_{K+1}) \triangleq \frac{1}{N}\sum_{i=1}^{N}[\tilde{X}_K(t_i)-\tilde{X}_{K+1}(t_i)]^2=\frac{1}{N}\sum_{i=1}^{N}[imf_k(t_i)]^2 \quad k\in\{1,2,..,n\}(4)$$

By using the CMSE criterion, the IMF order corresponding to the first significant change in the energy distribution is identified.

1.3. CHOI-WILLIAMS Time-Frequency Technique

Time-frequency representations have found extensive application in problems requiring time-varying spectral analysis [11]. The most significant class of time-frequency representations is known as the Cohen's Class [12]. Between the different time-frequency methods belonging to this class, the Choi–Williams distribution was chosen for its remarkable properties. The Choi–Williams distribution $CWD(t,f)$ was a significant step in the field of time-frequency analysis where it opened the way for enhancing resolution with cross-terms reduction [13]:

$$CWD_S=2\iint_{-\infty}^{\infty}\frac{\sqrt{\sigma}}{4|\tau|\sqrt{\pi}}e^{\frac{-\sigma x^2}{(16\tau^2)}}s(t+x+\frac{\tau}{2})s^*(t+x-\frac{\tau}{2})\,dxd\tau \qquad (5)$$

Where $\sigma$is a real parameter that can control the resolution and the cross-terms reduction. This can show excellent performance in reducing cross-terms while keeping high resolution, with a compromise between these two requirements decided by the parameter.

1.4. Denoising EMD Stimulating Algorithm

Figure 1 is a block diagram representing acoustic synthesis of proposed algorithm. One of the first processing steps in cochlear implants is to apply pre-emphasis to the signal. The pre-emphasis filter attenuates low frequencies and amplifies high frequencies, to compensate for the typical 6 dB/octave spectral roll-off of speech signals. It makes the low-energy, high-frequency consonants to stand out better against the high-energy, low-frequency vowels.

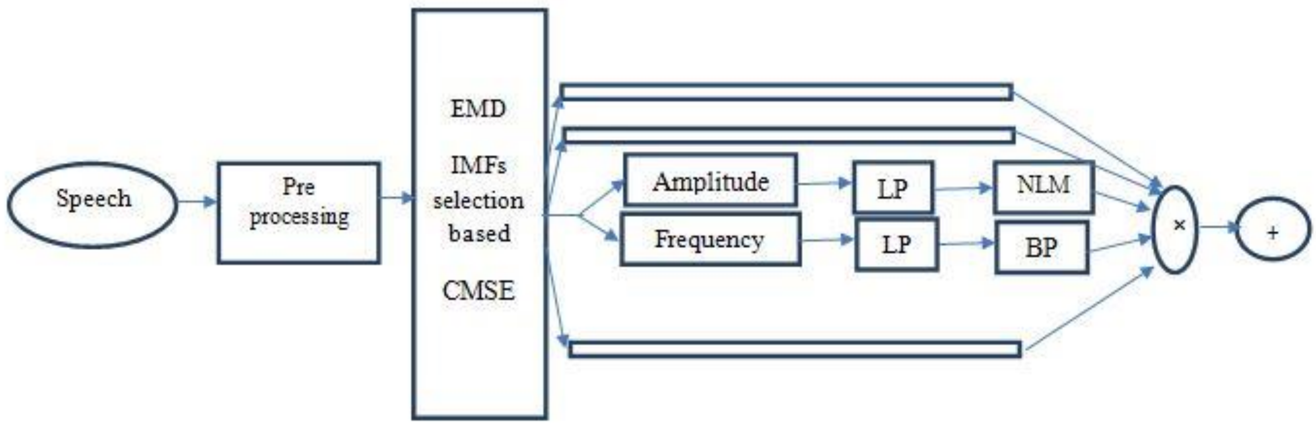

Figure 1- Block diagram of denoising EMD algorithm

Then the signal is processed through empirical mode composition, in each band, using the CMSE criterion, the IMF selection is done. The outputs of n channels are passed through two independent parallel paths to extract the amplitude envelope and frequency information in each band. The envelopes of the derived IMFs are then extracted by mode computation and low-pass filtering (LP).The cutoff frequency of LP is typically400 Hz. Finally, envelope matching is needed to map the decomposed signal to the dynamic range of the human ear. For this purpose, a nonlinear logarithmic function (NLM, as shown in figure 1) is used [14]. At the same time, in another path the frequency is derived from output speech signal of $CWD(t,f)$ in each band. After low pass filtering (LP), the frequency depth of processed signal in each band is limited at about 500Hz. At last, synthesized speech signal could obtained by summarizing each sub-band's stimuli.

2. Results

The cross-correlation between spectrums of synthesized and original signals was calculated, showing the power of this method and its capability in representation of a high percentage of the original signal for the implant user. Correlation coefficients were obtained in different environmental conditions (quiet, 5dB, 10dB, 15dB). In this computer simulation, Noisy92 sentences as database were processed by CIS, FAME, HHTS and proposed algorithm. Table 1 shows correlation coefficient between spectrum of reconstructed signal and original one which is deteriorating for each speech coding algorithm with noise increasing. This indicated that regardless of the type of algorithm, the efficiency of encoding can be reduced by increasing of noise. The performance of proposed algorithm is better than other three algorithms in four different environmental conditions.

Table1-Mean values of absolute correlation coefficients $\bar{r}$

Listening condition	CIS	FAME	HHT	Proposed algorithm
Quiet	0.0892	0.3174	0.4347	0.5320
5 dB	0.0041	0.3060	0.4122	0.4621
10 dB	0.0038	0.2974	0.4009	0.4103
15 dB	0.0036	0.2795	0.3793	0.4021

4. Conclusion

In this paper, we presented a denoising EMD-based technique to decompose the input signal into different frequency bands and a selected modes approaches are applied as a speech enhancement method for cochlear implants. This algorithm is developed to extract features, called intrinsic mode functions, by a sifting process. Then, IMFs are selected based on CMSE criteria to decrease the noise effect. Also, the Choi-Williams time-frequency technique is applied to extract different components of the resulting signal. Reconstruction of the decomposed signal showed that our technique can produce the processing with higher correlation than other methods.

REFERENCES

[1] K. Nie, G. Stickney, and F.-G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *Biomedical Engineering, IEEE Transactions on,* vol. 52, pp. 64-73, 2005.

[2] P. C. Loizou, "Mimicking the human ear," *Signal Processing Magazine, IEEE,* vol. 15, pp. 101-130, 1998.

[3] A. J. Spahr and M. F. Dorman, "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," *Archives of Otolaryngology–Head & Neck Surgery,* vol. 130, pp. 624-628, 2004.

[4] M. K. Qin and A. J. Oxenham, "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *The Journal of the Acoustical Society of America,* vol. 114, pp. 446-454, 2003.

[5] B. S. Wilson, D. T. Lawson, M. Zerbi, C. C. Finley, and R. D. Wolford, "New processing strategies in cochlear implantation," *Otology & Neurotology,* vol. 16, pp. 669-675, 1995.

[6] S. Sunny, S. David Peter, and K. Poulose Jacob, "Recognition of speech signals: an experimental comparison of linear predictive coding and discrete wavelet transforms," *International Journal of Engineering Science & Technology,* vol. 4, 2012.

[7] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng*, et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences,* vol. 454, pp. 903-995, 1998.

[8] H. Liu, W. Wang, K. Li, and Z. Zhang, "A novel speech coding algorithm for cochlear implants," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, 2012, pp. 403-406.

[9] J.-C. Cexus, "Analyse des signaux non-stationnaires par transformation de Huang, Opérateur deTeager-Kaiser, et Transformation de Huang-Teager (THT)," Université Rennes 1, 2005.

[10] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *Instrumentation and Measurement, IEEE Transactions on,* vol. 56, pp. 2196-2202, 2007.

[11] L. Cohen, *Time-frequency analysis: theory and applications*: Prentice-Hall, Inc., 1995.

[12] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE,* vol. 77, pp. 941-981, 1989.

[13] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 37, pp. 862-871, 1989.

[14] Q.-J. Fu and R. V. Shannon, "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," *The Journal of the Acoustical Society of America,* vol. 107, pp. 589-597, 2000.