An Improved Spectral Subtraction Algorithm for Noise Reduction in Cochlear Implants

Saeed Kermani[1], Marjan Mozaffari legha[2]

Abstract
Cochlear implants are widely known as the unique ways for persons with severe to profound hearing loss to restore some degree of hearing. Speech enhancement strategies play an important role in improving the cochlear implant. In this paper, a noise reduction algorithm is proposed that applies a spectral subtraction using the classifications between the speech dominant and the noise one in each channel. The proposed classifications use the spectrum entropy of observation signal in each channel. The performance of the proposed noise reduction algorithm is evaluated with segmental SNR using Noisy92 sentences embedded in babble and white Gaussian noise (WGN) at 0–20 dB SNR. On the basis of comparing segmental, and visually inspecting the enhanced spectrograms, the proposed method was found to effectively reduce noise.

Keywords: Cochlear implants, improved spectral subtraction, noise reduction

1. Introduction

Cochlear implant (CI) is an auditory prosthesis device for restoring hearing function of patients with sensory-neural hearing loss, using electrical stimulation of auditory nerve [1]. Most CI users reach 80% word recognition scores in quiet listening conditions[2]. However, speech recognition scores are corrupted in noisy environments[3]. Several studies have been proposed to develop speech processing techniques for CI in noisy conditions. Most of modern devices utilize a filter-bank for the frequency decomposition of speech. This is a simulation of frequency decomposition function of biological cochlea which is based on placement coding theory. Outputs from each channel of the filter-bank are used to modulate the amplitudes of electrical stimulation pulses. In most current multichannel devices, a simple linear band pass filter is used for the frequency decomposition[4]. Many studies described that CI users are more vulnerable to noise than normal-hearing listeners[5].The main reason of this abnormality returns to limitation of spectral resolution. Typically, there are maximally 22 channels in a CI system, which is much less than the number of frequency bands used in a normal-hearing person.

- ***Saeed Kermani***, Department of Medical Physics and Medical Engineering in the School of Medicine of Isfahan University of Medical Sciences, Iran. (E-mail: kermani@med.mui.ac.ir).
- ***Marjan Mozaffari legha***, Department of Medical Physics and Medical Engineering in the School of Medicine of Isfahan University of Medical Sciences, Iran. (E-mail: M_mozaffari@ymail.com).

The first solution is to offer more frequency bands (or electrodes) in a CI device. However, due to technological constrains, little progress has been reported in CIs. Therefore, speech enhancement will play an important role for CI's users. Several studies have described to improve speech intelligibility for the hearing impaired[6],[7]. Some of these algorithms were based on existence of two or more microphones. Studies showed that an adaptive beam-forming algorithm based on multiple-microphone could considerably improve the speech recognition of CI listeners when the speech and noise signals were from different directions [6]. However, due to the constrained dimension, it is not feasible to implement second microphone for unilateral CI recipients. Therefore, single-microphone noise reduction algorithms are more interesting and more practicable for implementation. Preferably, a noise reduction algorithm is need to be simple to implement and, most significantly, to be embedded in the existing coding strategies rather than being used as a pre-processor. Several single-microphone noise-reduction strategies have been proposed for cochlear implants which are based either on spectral subtraction [7], or on statistical-model-based methods[8], or on subspace method[9] . The above noise-reduction algorithms have provided little advantage for CI listeners[10]. For a multi-channel spectral subtraction, which is often used in the current CI system, the noise levels in different frequency bands are estimated and the Signal-to-Noise Ratio (SNR) in each band of the noisy speech is determined.

The speech signal is estimated by subtracting the estimated noise spectral magnitude from the noisy speech spectral magnitude. A gain function is used to determine a level of reduction to be applied to the signal to optimally remove the noise. The main disadvantage of this method is the "musical noise" artifact, which is due to the inaccuracy of the spectrum estimation.

In the present study, a modified spectral subtraction method is proposed using the classifications between the speech dominant and the noise one in each channel. The proposed classifications use spectrum entropy of observation signal in each channel. Objective evaluation of the proposed algorithm was done with segmental SNR using Noisy92 sentences embedded in babble and white Gaussian noise (WGN) at 0–20 dB SNR. This paper is organized as follows. Section 2 describes the proposed method. Section 3 covers the results. Finally, the conclusions are given in Section 4.

2. Materials and Methods

2.1. Preprocessing. In CI, input audio signal, which is captured by a microphone, is pre-emphasized to compensate the high-frequency components corresponding to consonant parts of the speech. A first order high-pass Butterworth filter with cut-off frequencyof1.2 KHz is applied [11]. The signal is the digitalized using an A/D (Figure 1(a)).

2.2. Improved Spectral Subtraction Algorithm

The next stage is signal analysis. The proposed speech enhancement technique is based on the spectral subtraction algorithm [12].

The observation signal $y(k)$ is windowed by hamming window, and then the windowed signal is transformed to the frequency domain by applying FFT. We consider a speech signal $s(k)$corrupted by an additive background noise$n(k)$. The observation signal $y(k)$ can be expressed by

$$y(k) = s(k) + n(k) \quad (1)$$

$$Y(w,r) = S(w,r) + N(w,r) \quad (2)$$

Where$Y(w,r)$, $S(w,r)$ and$N(w,r)$ denote the short-time Fourier transforms of$y(k)$, $s(k)$and $n(k)$ for frame$r$, respectively. Also, $s(k)$is assumed to be uncorrelated with$n(k)$. If the noise spectrum $|N(w,r)|$ is estimated as$|\widehat{N}(w,r)|$, the estimation of the short-time speech spectrum $|\hat{S}(w,r)|$is represented by

$$\left|\hat{S}(w,r)\right| = H(w,r)|Y(w,r)| \quad (3)$$

$$H(w,r) = \begin{cases} \sqrt{1-\alpha\ \mathrm{SNRpost}(w,r)^2} & \text{if } \mathrm{J} \geq 0 \\ \beta\sqrt{\mathrm{SNRpost}(w,r)^2} & \text{otherwise} \end{cases} \quad (4)$$

$$\mathrm{J} = \frac{1}{(\alpha+\beta)} - \mathrm{SNRpost}(w,r)^2 \quad (5)$$

$$\mathrm{SNRpost}(w,r) = \frac{\left|\widehat{\mathrm{N}}(w,r)\right|}{|\mathrm{Y}(w,r)|} \quad (6)$$

Where $H(w,r)$is gain function, (α ≥1) is the over subtraction factor and (β ≥0) is flooring level factor. When J is bigger than zero the spectral subtraction is carried out. On the other hand, spectral flooring is carried out when J < 0.We set two over-subtraction factors for speech dominant and noise one, namely $\alpha 1$ and $\alpha 2$. The $\alpha 1$ is set as the value to reduce the noise with little distortion to speech. The $\alpha 2$ is set $\alpha 1 < \alpha 2$ to fully reduce the noise. Previous study [13]showed that for 16 kHz sampling frequency with 256 samples per frame and a 50% overlap, the values of $\alpha 1$ is around 2.5, $\alpha 2$ is around 6 and β is 0.001. These values are adopted for this study.

The classifications of speech/noise-dominant using entropy of the spectrum of observation signal are carried out. This step is described in detail in the next subsection. Estimation of noise spectrum is carried out in the same manner as estimation of noise spectrum based on Quantile based noise estimation for spectral subtraction [14].Once the subtraction is calculated in the spectral domain with (3) and (4) the enhanced speech signal $\hat{s}(k)$ is obtained as

$$\hat{s}(k) = \mathrm{IFFT}[|\hat{s}(w,r)|.\, e^{j\arg(Y(w,r))}] \quad (7)$$

Where the phase of the observation signal is used for the enhanced speech signal. By combining the $|\hat{S}(w,r)|$ and the phase of the observation signal, the enhanced speech signal $\hat{s}(k)$ is obtained by applying the inverse fast Fourier transform.

ISSN 2229-5518

2.3. Speech/Noise-dominant classification

We propose a new classification scheme between the speech dominant and the noise Dominant signals. The available estimation of the stationary background noise spectrum can be used to locate regions of energy level higher than that of the background. The higher energy of these regions may be due to either speech or a high-energy nonstationary noise component. Since it is not possible to distinguish the two possibilities based on instantaneous energy alone, we turn to different features for speech–noise discrimination. The relatively flat spectral structure of noise-dominated regions can be captured by quantification of ''spectral flatness''. Entropy is a related measure which is used in the voice activity detection (VAD)algorithm[15].Based on the assumption that the signal spectrum is more organized during speech segments than during noise segments, the proposed classification use entropy of the spectrum of observation signal in each critical band. Entropy can be re defined for a frequency band as

$$H_i = -\frac{1}{\log(L)} \sum_{k=b_i}^{e_i} P(|X(k)|^2) \log(|X(k)|^2) \qquad (8)$$

Where$P(|X(k)|^2) = \frac{|X(k)|^2}{\sum_{k=b_i}^{e_i}(|X(k)|^2)}$is the ''probability'' ofthe frequency bin''k''.H takes maximum value of ''l''when the signal is a white noise, and minimum valueof 0 when it is a pure tone.

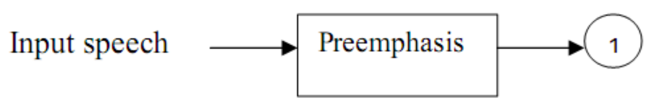
2.4. Signal Analysis. Figure 1 shows a block diagram of the CI. We applied filter bank with center frequency arranged from low to high to simulate frequency response of basilar membrane of cochlea[16]. Table 1 shows a list of frequency bands of different channels. Sampling frequency of 16 KHz, simulation rate of 250–2400 pulse per second, and a number of 22channels corresponding to the 22 electrodes in the cochlear implant are assumed.

In order to increase the simulation rate, the output needs to be interpolated depending on the patient's MAP[11]. After decomposition of the frequency band of the input signal into the 22 channels, the envelope of the signal is identified for each band. A second-order low-pass Butterworth filter with cut-off frequency of 200–400Hz is applied for smoothing[17].It is shown in Figure 1(c). The cut-off frequency of the low-pass filter determines the slowly varying rate information preserved in the envelope.

Then, n-of-m strategy is used to choice 10 channels with maximum amplitude out of 22. These selected channels cover the maximum energy of the signal [18].Lastly, amplitude matching with a nonlinear logarithmic function is done to map the decomposed signal to the dynamic range of the patient's hearing[19].

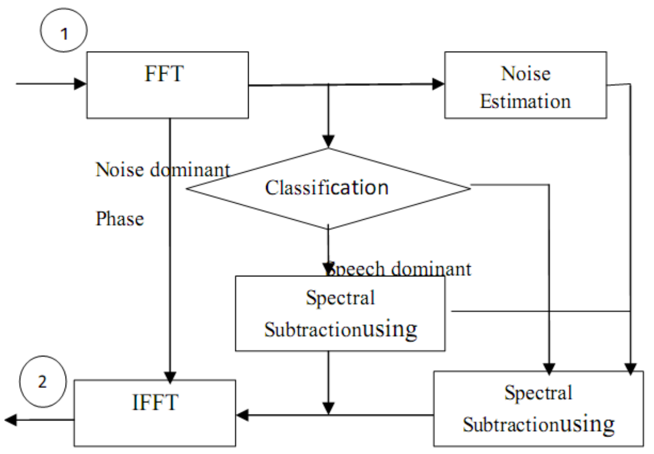
Table1- List of frequency bands of 22-channels

Channel number	Cut-off frequencies in Hz
1	7000–8000
2	6000–7000
3	5250–6000
4	4625–5250
5	4000–4625
6	3500–4000
7	3000–3500
8	2625–3000
9	2250–2625
10	2000–2250
11	1750–2000
12	1500–1750
13	1250–1500
14	1125–1250
15	1000–1125
16	875–1000
17	750–875
18	625–750
19	500–625
20	375–500
21	250–375
22	125–250



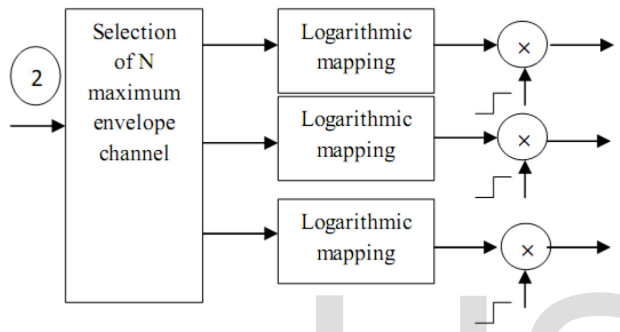
(a)

ISSN 2229-5518



(b)

Pulse Modulation



(c)

Fig1. Block diagram of the decomposition strategy in CI

3. Results

The performance of the proposed noise reduction algorithm is evaluated with segmental SNR using Noisy92 sentences embedded in multi-talker babble and white Gaussian noise (WGN) at 0–20 dB SNR. It illustrates the effectiveness of the proposed system for reducing background noise in cochlear implant. In table 2, the improvement of segmental SNR for multi-talker babbling and white Gaussian noise is shown. The segmental SNR is defined as

$$\mathrm{seg.SNR} = \frac{10}{M} . \sum_{m=0}^{M-1} \log . \left(\sum_{k=Nm}^{Nm+N-1} s\verb|^|2(k) \Big/ \sum_{k=Nm}^{Nm+N-1} \{\hat{s}(k) - s(k)\}\verb|^|2 \right) \qquad (9)$$

Where N is the segment length and M is the number of segments in the speech signal. From table 2, we find that the proposed method is suitable for speech processor of CI.

Table2-segmental SNR for multi-talker babbling and white Gaussian noise

	0	5	10	15	20
without enhanced algorithm	1.1	2.11	4.67	9.7	13.3
enhanced algorithmic in multi-talker babble	3.3	10	14.55	21.8	24.6
enhanced algorithm in WGN	3.4	10.21	12.87	20.3	26.1

For better comparison, spectrograms of synthesized signal without & with proposed noise removal strategy are shown. It illustrates the effectiveness of the proposed system for reducing background noise in cochlear implant. Fig.2 shows spectrogram of synthesized sound without using denoising strategy. From Fig.3, we see that the noise is reduced by improved spectral subtraction (ISS) method for multi-talker babbling noise. Fig.4and Fig5 shows spectrogram of synthesized sound without and with using denoising strategy for WGN noise.

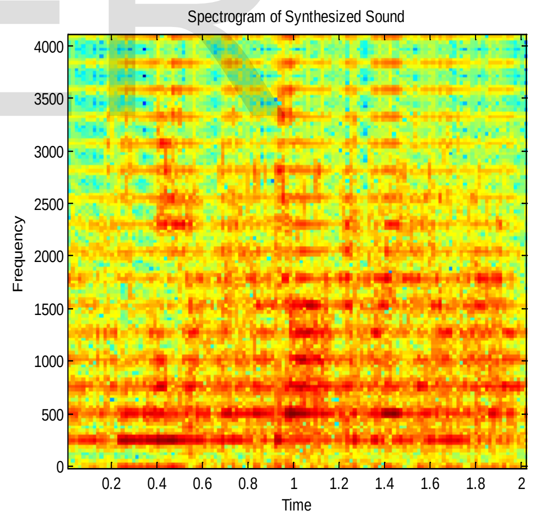


Fig2.spectrogram of synthesized sound for multi-talker babbling noise

ISSN 2229-5518

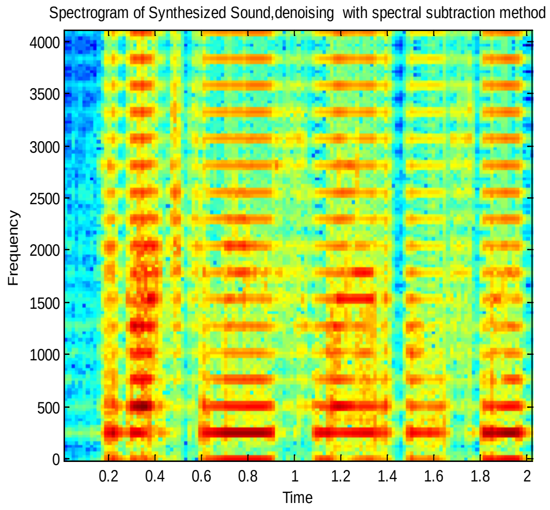


Fig3.spectrogram of synthesized sound, denoising with ISS method for multi-talker babbling noise

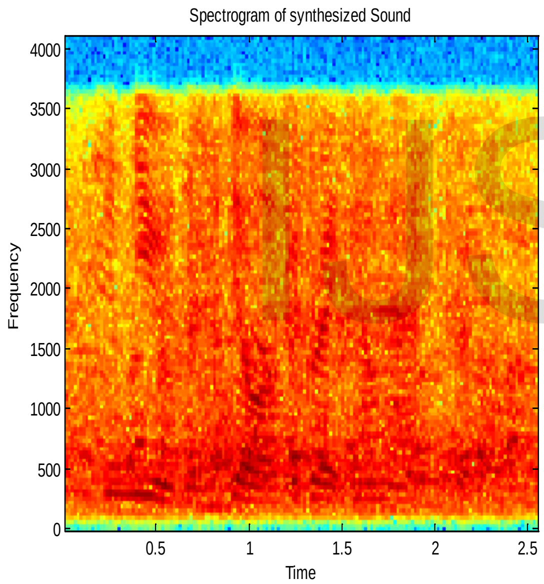


Fig4.spectrogram of synthesized sound for WGN noise

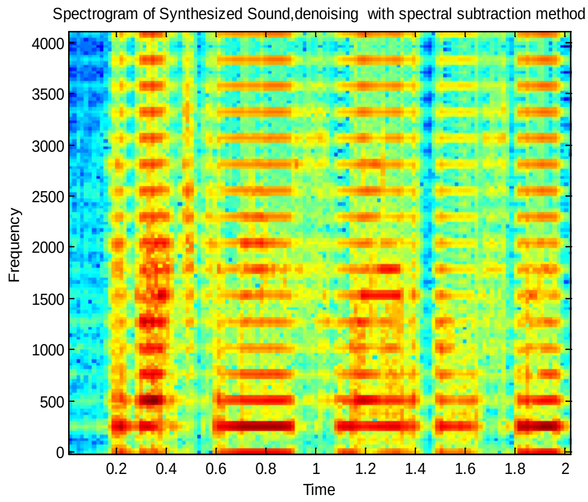


Fig5.spectrogram of synthesized sound, denoising with ISS method for WGN noise

4. Conclusion

In this paper, we introduced a novel enhancement method which is performed on a band-by-band basis for each time frame. Based on both the decision on whether a particular band in a frame is speech or noise dominant, an appropriate amount of noise is reduced using modified spectral subtraction. The performance of the proposed noise reduction algorithm is evaluated with segmental SNR using Noisy92 sentences embedded in babble and white Gaussian noise (WGN) at 0–20 dB SNR. On the basis of comparing segmental SNR, and visually inspecting the enhanced spectrograms, the proposed method was found to effectively reduce noise.

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] 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.

[3] C. C. Dunn, W. Noble, R. S. Tyler, M. Kordus, B. J. Gantz, and H. Ji, "Bilateral and unilateral cochlear implant users compared on speech perception in noise," *Ear and hearing,* vol. 31, p. 296, 2010.

[4] J. Yao and Y.-T. Zhang, "The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations," *Biomedical Engineering, IEEE Transactions on,* vol. 49, pp. 1299-1309, 2002.

[5] B. L. Fetterman and E. H. Domico, "Speech recognition in background noise of cochlear implant patients,"

Otolaryngology--Head and Neck Surgery, vol. 126, pp. 257-263, 2002.

[6] R. Van Hoesel and G. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *The Journal of the Acoustical Society of America,* vol. 97, pp. 2498-2503, 1995.

[7] L.-P. Yang and Q.-J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *The Journal of the Acoustical Society of America,* vol. 117, pp. 1001-1004, 2005.

[8] P. W. Dawson, S. J. Mauger, and A. A. Hersbach, "Clinical evaluation of signal-to-noise ratio–based noise reduction in Nucleus® cochlear implant recipients," *Ear and hearing,* vol. 32, pp. 382-390, 2011.

[9] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *Speech and Audio Processing, IEEE Transactions on,* vol. 11, pp. 334-341, 2003.

[10] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech communication,* vol. 49, pp. 588-601, 2007.

[11] A. Paglialonga, G. Tognola, G. Baselli, M. Parazzini, P. Ravazzani, and F. Grandori, "Speech processing for cochlear implants with the discrete wavelet transform: feasibility study and performance evaluation," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 3763-3766.

[12] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 27, pp. 113-120, 1979.

[13] A. Guérin, R. Le Bouquin-Jeannés, and G. Faucon, "A two-sensor noise reduction system: applications for hands-free car kit," *EURASIP Journal on Applied Signal Processing,* vol. 2003, pp. 1125-1134, 2003.

[14] M. F. Dorman and P. C. Loizou, "Mechanisms of vowel recognition for Ineraid patients fit with continuous interleaved sampling processors," *The Journal of the Acoustical Society of America,* vol. 102, pp. 581-587, 1997.

[15] P. Renevey and A. Drygajlo, "Entropy based voice activity detection in very noisy conditions," in *INTERSPEECH*, 2001, pp. 1887-1890.

[16] M. Mehrzad, M. Abolhassani, A. Jafari, J. Alirezaie, and M. Sangargir, "Introduction of a Novel Mapping for Cochlear Implant Speech Processing," in *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*, 2012, pp. 330-334.

[17] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature,* vol. 352, pp. 236-238, 1991.

[18] A. Buechner, C. Frohne-Buechner, P. Boyle, R.-D. Battmer, and T. Lenarz, "A high rate n-of-m speech processing strategy for the first generation Clarion cochlear implant," *International journal of audiology,* vol. 48, pp. 868-875, 2009.

[19] Q.-J. Fu and R. V. Shannon, "Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping," *The Journal of the Acoustical Society of America,* vol. 106, pp. L18-L23, 1999.

Saeed Kermani obtained his BS from the Department of Electrical Engineering of Isfahan University of Technology in Isfahan, Iran, 1987, and he received the MS in Bioelectric Engineering from Sharif University of Technology, in 1992 and his PhD in Bioelectric Engineering at AmirKabir University of Technology, Tehran, Iran, in 2008. He is Assistant Professor of Medical Engineering at the Department of Medical Physics and Medical Engineering in the School of Medicine of Isfahan University of Medical Sciences, Iran. His research interests are in biomedical signal processing techniques

Marjan Mozaffari Legha was born in Kerman, Iran, in 1988. She received her B.Sc. from Shahid Beheshti University in 2010, and the M.Sc. degrees from Medical Physics and Medical Engineering in the School of Medicine of Isfahan University of Medical Sciences in 2012. She's currently PhD student of Biomedical Engineering from Shahid Beheshti University. She is interested in digital signal processing and speech processing and Bioelectromagnetics.