

Research Article

Cochlear Implant Speech Processing Using Wavelet Transform

M. Mehrzad,^{1,2} M. D. Abolhassani,^{1,2} A. H. Jafari,^{1,2} J. Alirezaie,^{3,4} and M. Sangargir^{1,2}

¹ Medical Physics & Biomedical Engineering Department, Tehran University of Medical Sciences, Tehran 1417613151, Iran

² Biomedical Group, Research Centre for Science and Technology in Medicine, Tehran 14185615, Iran

³ Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3

⁴ Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Correspondence should be addressed to A. H. Jafari, h.jafari@tums.ac.ir

Received 9 April 2012; Accepted 29 May 2012

Academic Editors: W.-L. Hwang, S. Kwong, and A. Rubio Ayuso

Copyright © 2012 M. Mehrzad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a method for coding speech signals for the simulation of a cochlear implant. The method is based on a wavelet packet decomposition strategy. We used wavelet packet db4 for 7 levels, generated a series of channels with bandwidths exactly the same as nucleus device, and applied an input stimulus to each channel. The processed signal was then reconstructed and compared to the original signal, which preserved the contents to a high percentage. Finally, performance of the wavelet packet decomposition in terms of computational complexity was compared to other commonly used strategies in cochlear implants. The results showed the power of this method in processing of the input signal for implant users with less complexity than other methods, while maintaining the contents of the input signal to a very good extent.

1. Introduction

Human cochlea is a coiled duct, filled with fluid, and divided into three sections: the scala vestibuli, scala media, and scala tympani. The scala media is separated from the scala tympani by basilar membrane. Input sound vibrations create a pressure difference between both sides of the basilar membrane which results in a corresponding movement of basilar membrane. Above the basilar membrane lies the tectorial membrane, and between the two membranes, lie the hair cells.

There are two types of the hair cells, the outer hair cells (OHCs) and the inner hair cells (IHCs). Three outer hair cells deliver information to one inner hair cell.

Mechanical movement of OHCs in response to the input sound vibrations is transformed by IHCs into neural activity.

Any damage to the cochlea is mainly associated with damage or loss of the outer hair cells, while the inner hair cells are mainly remained intact [1]. In case of missing of OHCs in any part of the cochlea, the frequencies associated with this specific location along the basilar membrane will not be sensed any more.

Cochlear implants are being widely used for the patients with severe damages in the cochlea. The main parts of

a cochlear implant from inside to the outside of the ear are an array electrode (surgically placed into the inner ear), a receiver (intracranial), a transmitter (placed on the skull using a magnet, fitted on the exact location of the intracranial receiver), speech processor (placed behind the ear), and a microphone which collects the sound [2].

A damaged cochlea is not able to analyze the input signal into proper frequency bands. A speech processor is designed to overcome this disability and simulate the function of a healthy cochlea.

Many studies have been conducted to develop speech processing techniques. The previous techniques were not successful in proving time and frequency resolutions at the same time. Filter banks, for instance, despite of providing excellent temporal resolution, are not successful in providing adequate frequency resolution. On the other hand, FFT is strong in frequency resolution while limited in temporal resolution [3].

Wavelet has overcome the limitations of the previous methods by providing both time and frequency resolutions [4]. The DWT is defined by the following equation:

$$W(j, k) = \sum_j \sum_k x(k) 2^{-j/2} \psi(2^{-j}n - k), \quad (1)$$

where ψ is the mother wavelet, $x(k)$ is the input signal, the coefficient 2^{-j} represents the scale, and the constant k is used to generate different positions in the wavelet. As a result, both temporal and frequency resolutions can be obtained without any restriction.

This paper is organized as follows. Section 2 describes the preprocessing method (Section 2.1), a justification of why wavelet has been chosen as the leading strategy (Section 2.2), steps taken towards the selection of the wavelet-packet-decomposition tree, and signal analysis (Section 2.3). Section 3 covers the results. Finally, the conclusions are given in Section 4.

2. Material and Methods

2.1. Preprocessing. The electrical format of the input acoustic signal is generated by a microphone. A gain controller is used to enhance quality of this signal. In order to preemphasize the high frequency components corresponding to consonant parts of the speech, a first order high-pass Butterworth filter with cut-off frequency of 1.2 KHz is applied [5].

The signal is the digitalized using an A/D and further filtering is performed to remove its DC component (Figure 1(a)).

2.2. Selection of the Wavelet-Packet-Decomposition Tree. The next stage is signal analysis. As described in the introduction, we chose wavelet transform for this purpose. MATLAB provides a family of wavelet packets including Haar, Daubechies, Symlets, Coiflets, and Meyer. The main challenge is now to select the best wavelet-packet-decomposition tree (WPD) which provides the best speech recognition after the reconstruction, while each family has its own advantages. Daubechies provides a maximally flat response, while the Symmlet is more symmetrical [6].

In 2006, Nogueira et al. applied a db3 and Haar wavelets to decompose the input signal. The results were tested on a group of 7 patients which showed the significant efficiency of db3 [6]. Studies of Gopalakrishna et al. have shown that Daubechies (db4, for six levels) and Symlets have similar results [3]. Comparing the mentioned wavelets, we chose wavelet packet db4 for 7 levels.

There are several reasons for choosing wavelet packet db4. Firstly, db is compact support which results in finite number of filter parameters and faster implementation. We demand a filter of shorter length which results in better frequency resolution and noise reduction. The length of the filter in db wavelet ($2N$, where N is the order of the filter) is less than others (e.g., $6N$ for coiflet).

Secondly, db wavelet has maximum number of vanishing moments amongst others which results in a better regularity of the signal. Therefore, signal components with fine amplitudes will not be missed.

Finally, db has identical forward/backward filter parameters which enables us with a fast, exact reconstruction of the speech signal.

Thus, db4 meets our expectations for speech signal processing.

TABLE 1: List of frequency bands of different channels.

Channel number	Cut-off frequencies in Hz	Node's number
1	7000–8000	[7,56]–[7,63]
2	6000–7000	[7,48]–[7,55]
3	5250–6000	[7,42]–[7,47]
4	4625–5250	[7,37]–[7,41]
5	4000–4625	[7,32]–[7,36]
6	3500–4000	[7,28]–[7,31]
7	3000–3500	[7,24]–[7,27]
8	2625–3000	[7,21]–[7,23]
9	2250–2625	[7,18]–[7,20]
10	2000–2250	[7,16]–[7,17]
11	1750–2000	[7,14]–[7,15]
12	1500–1750	[7,12]–[7,13]
13	1250–1500	[7,10]–[7,11]
14	1125–1250	[7,9]
15	1000–1125	[7,8]
16	875–1000	[7,7]
17	750–875	[7,6]
18	625–750	[7,5]
19	500–625	[7,4]
20	375–500	[7,3]
21	250–375	[7,2]
22	125–250	[7,1]

After choosing the suitable wavelet packet, we decomposed the signal in seven levels to be of the same frequency range as the nucleus device for each channel.

The next stage is signal analysis as described in the next section.

2.3. Signal Analysis. Figure 1 shows a block diagram of the decomposition strategy. We applied wavelet packet db4 for 7 levels which will generate a series of channels with bandwidths exactly the same as nucleus device [7]. The correspondent phi (W_0) and psi ($W_1 - W_6$) functions are shown in Figure 2.

Figure 3 shows the related wavelet-packet tree. In this figure, the numbers next to the tree branches indicate the channel numbers. By adding some nodes together, the correspondent frequency range is generated and specified to the demanded channel. Node [7,0] includes dc information which is not used in the calculations in this paper. Table 1 presents a list of frequency bands of different channels.

We will assume a sampling frequency of 16 KHz, simulation rate of 250–2400 pulse per second, and a number of 22 channels corresponding to the 22 electrodes in the cochlear implant. Decomposition of the signal in each level causes further down sampling and decrease in the simulation rate. Depending on the patient's MAP, in order to increase the simulation rate, the output needs to be repeated or linearly interpolated [5].

After decomposition of the frequency band of the input signal into the mentioned 22 channels, the envelope of the signal is detected for each band. A second-order low-pass

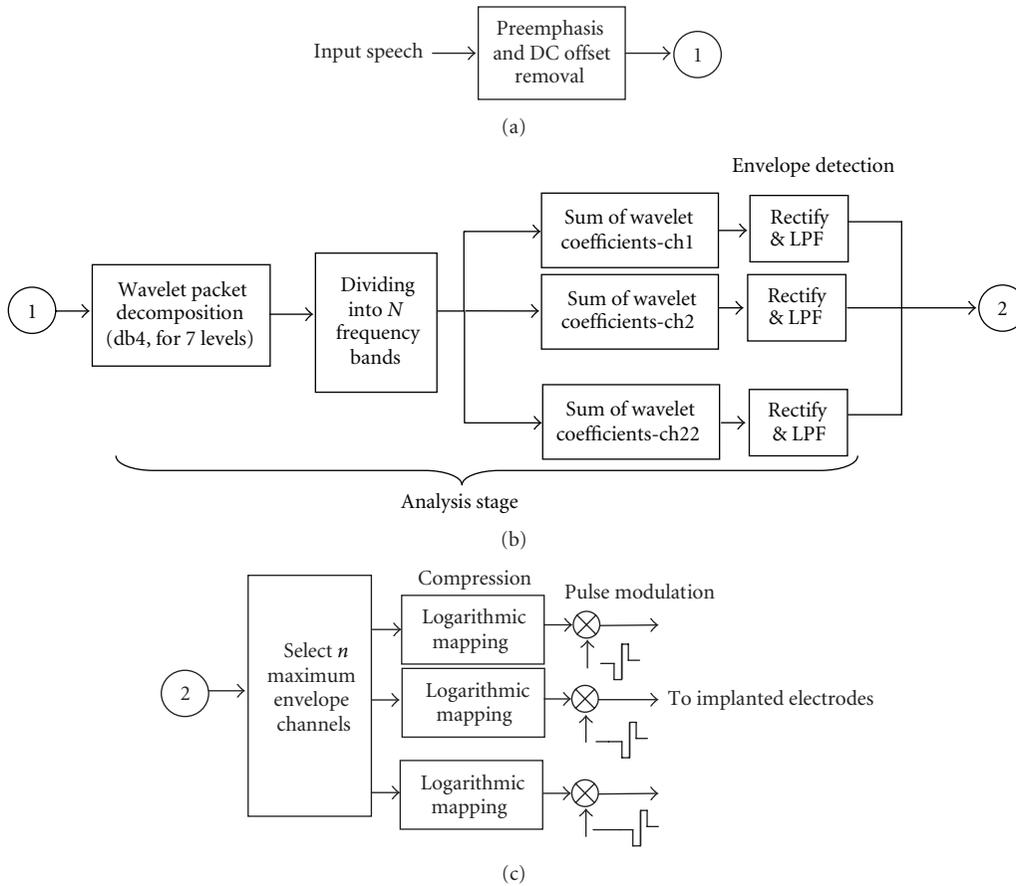


FIGURE 1: Block diagram of the decomposition strategy.

Butterworth filter with cut-off frequency of 200–400 Hz is applied for smoothing [8, 9], as shown in Figure 1(b).

Then, n -of- m strategy is used to select 10 channels with maximum amplitude out of 22. The 10 selected channels contain the maximum share in energy of the signal and not all of the other channels are needed to provide the proper hearing sensation [10].

Finally, amplitude matching is needed to map the decomposed signal to the dynamic range of the human ear. For this purpose, a nonlinear logarithmic function is used [11], shown in Figure 1(c).

3. Results

After preprocessing of an input signal and implementation of the WPD, the signal was analyzed as described in Section 2.3. The decomposed signal was then reconstructed to test the accuracy of the method.

The cross-correlation between the reconstructed and original signals was over 90% (dependent on the type of the input signal), showing the power of this technique and its capability in representation of a high percentage of the original signal for the implant user, despite eliminating

information of some nodes (with less share in the energy of the signal).

Furthermore, the complexity of the algorithm needs to be low, so as to minimize the power requirements. The latter criterion is important as it can prolong the battery life of the speech processor. Table 2 includes a comparison list of approximate number of multiplications in various speech processing techniques. The result shows that wavelet's complexity applied in this study is significantly lower than other methods.

4. Conclusion

In this paper, we presented a wavelet-based technique to decompose the input signal into different frequency bands. For the decomposition strategy, we chose WPD (wavelet packet db4 for 7 levels) and generated a series of channels with bandwidths exactly the same as nucleus device. Reconstruction of the decomposed signal showed that our technique can generate the processing with less complexity than other methods, while maintaining over 90% of the original signal. This makes the strategy a good choice with respect to accuracy and power saving for cochlear implant users.

TABLE 2: Comparison of approximate number of multiplications in various speech processing technique.

Filter bank	Approximate number of real multiplications for $m = 256$ sample frame, 128-point FFT, 22 channels, and $l = 12, 24,$ and 30 length wavelet packet filter		
	FFT (N -point)	Recursive DFT update	Wavelet
Channels * (6th order BPF)	$2N(\log_2 N) + (N/2) * 2$	2nd order all-pole window $(N/2) * 4 + (N/2) * 2$	$m * l * 7 + (m/8) * \text{channels} * 5$
112 k	519 k	100 k	25 k
		167 k	47 k
			57 k

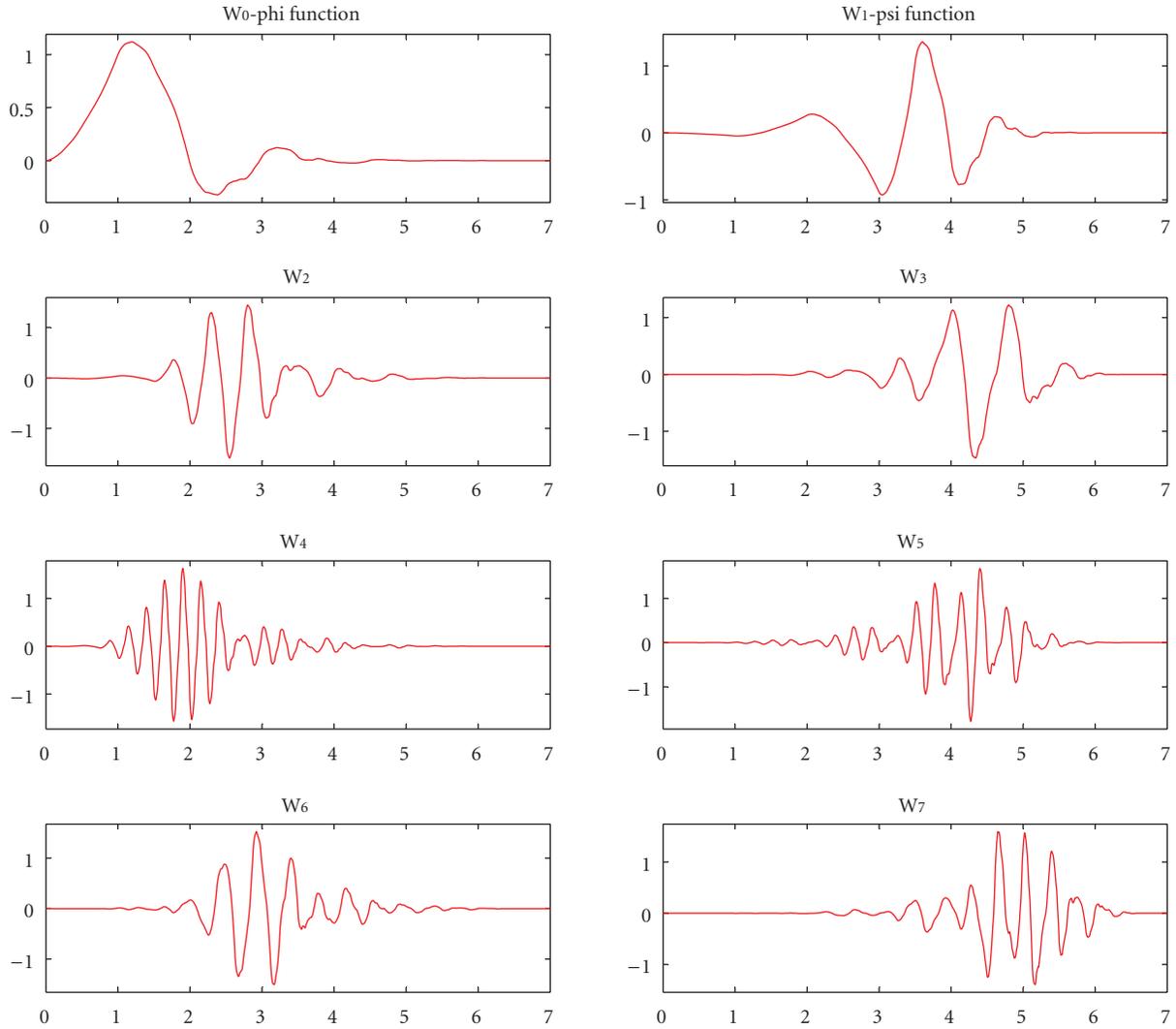


FIGURE 2: Phi (W_0) and psi ($W_1 - W_6$) functions correspondent to db4 for 7 levels.

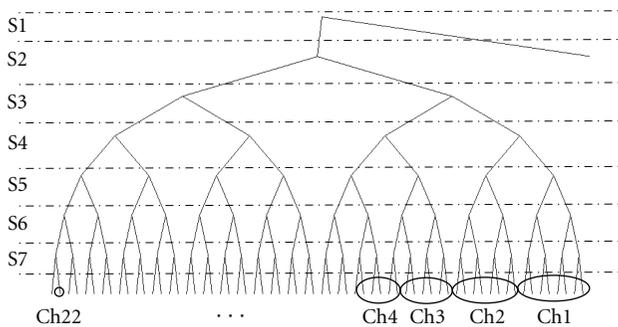


FIGURE 3: Related wavelet-packet tree of db4 for 7 levels.

References

[1] M. C. Liberman and L. W. Dodds, "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and

alterations of threshold tuning curves," *Hearing Research*, vol. 16, no. 1, pp. 55–74, 1984.
 [2] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine*, vol. 15, no. 5, pp. 101–130, 1998.
 [3] V. Gopalakrishna, N. Kehtarnavaz, and P. C. Loizou, "A recursive wavelet-based strategy for real-time cochlear implant speech processing on PDA platforms," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 2053–2063, 2010.
 [4] S. Sunny, D. Peter, and K. P. Jacob, "Recognition of speech signals: an experimental comparison of linear predictive coding and discrete wavelet transforms," *International Journal of Engineering Science*, vol. 4, no. 4, pp. 1594–1601, 2012.
 [5] 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 *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 3763–3766, September 2006.
 [6] W. Nogueira, A. Giese, B. Edler, and A. Büchner, "Wavelet packet filterbank for speech processing strategies in cochlear

- implants,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, pp. V121–V124, May 2006.
- [7] Software User Manual, N95246F Issue 1, Nucleus MATLAB Toolbox 2.11, Cochlear Corporation, Lane Cove, NSW, 2002.
 - [8] 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, no. 6332, pp. 236–238, 1991.
 - [9] P. C. Loizou, “Speech processing in vocoder-centric cochlear implants,” *Advances in Oto-Rhino-Laryngology*, vol. 64, pp. 109–143, 2006.
 - [10] 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, no. 12, pp. 868–875, 2009.
 - [11] 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,” *Journal of the Acoustical Society of America*, vol. 106, no. 2, pp. L18–L23, 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

