Original Article

www.jmss.mui.ac.ir

An Undecimated Wavelet-based Method for Cochlear Implant Speech Processing

Fatemeh Hajiaghababa, Saeed Kermani[1], Hamid R. Marateb[2]

Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran, [1]Department of Medical Physics and Medical Engineering,Isfahan University of Medical Sciences,Isfahan,Iran, [2]Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Submission: 11-08-2014 Accepted: 17-09-2014

ABSTRACT

A cochlear implant is an implanted electronic device used to provide a sensation of hearing to a person who is hard of hearing. The cochlear implant is often referred to as a bionic ear. This paper presents an undecimated wavelet-based speech coding strategy for cochlear implants, which gives a novel speech processing strategy. The undecimated wavelet packet transform (UWPT) is computed like the wavelet packet transform except that it does not down-sample the output at each level. The speech data used for the current study consists of 30 consonants, sampled at 16 kbps. The performance of our proposed UWPT method was compared to that of infinite impulse response (IIR) filter in terms of mean opinion score (MOS), short-time objective intelligibility (STOI) measure and segmental signal-to-noise ratio (SNR). Undecimated wavelet had better segmental SNR in about 96% of the input speech data. The MOS of the proposed method was twice in comparison with that of the IIR filter-bank. The statistical analysis revealed that the UWT-based N-of-M strategy significantly improved the MOS, STOI and segmental SNR (P < 0.001) compared with what obtained with the IIR filter-bank based strategies. The advantage of UWPT is that it is shift-invariant which gives a dense approximation to continuous wavelet transform. Thus, the information loss is minimal and that is why the UWPT performance was better than that of traditional filter-bank strategies in speech recognition tests. Results showed that the UWPT could be a promising method for speech coding in cochlear implants, although its computational complexity is higher than that of traditional filter-banks.

Key words: *Cochlear implant, mean opinion score, undecimated wavelet transforms*

INTRODUCTION

The human hearing system consists of external, middle and inner ear. Sound undergoes a series of transformations as it travels through auditory nerve and into the brain afterwards. The function of the external ear is to collect the sound waves and focusing them on the eardrum, separating the external ear from the middle ear, and to convert the sound waves into mechanical vibrations. In the inner ear, the cochlea, which resembles a snail shell is filled with fluid. It transforms the mechanical vibrations to vibrations in fluid. Pressure variations within the fluid of the cochlea displace the basilar membrane.[1] The displacements of this flexible membrane have information about the frequency of the acoustic signal.

The hair cells are attached to the basilar membrane and bend according to their displacements. These two parts translate mechanical vibrations into neural information. Thus, due to damaged hair cells, the auditory system is not able to transform mechanical sound signal to electrical nerve impulses, resulting in hearing impairment.[1] Researches have shown that the most common cause of deafness is the loss of hair cells, rather than the loss of auditory neurons.[2] The basis of the cochlear implant approach is that the neurons could be directly excited through electrical stimulation.

In the last decades, cochlear implant system has been improved profoundly.[2] It is a prosthetic device that could be implanted in the inner ear thus providing partial hearing. The cochlear implant system consists of an external processor, which selects and arranges sounds picked up by the microphone and an internal element that is implanted inside the body by means of a surgical operation.[3]

The main part of a cochlear implant system is the signal processor, which converts the signal into electrical pulses based on the speech processing strategy. The processing

Address for correspondence:
Fatemeh Hajiaghababa, Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.
E-mail: hajaghababa@yahoo.com

in the speech processor can aim to either preserve either waveform or envelope information.[1]

There are several speech processing strategies to drive electrical pulses. Most of which use a linear filter-bank for spectral analysis performed in the human cochlea. Since the model used for the cochlea is a set of nonlinear overlapping band-pass filters, one possibility is to use nonlinear strategies.[4,5] For example, Kim *et al.* proposed an active nonlinear model of the basilar membrane in the cochlea, called the dual resonance nonlinear (DRNL) model.[6] They have also simplified the DRNL to a model called simple dual path nonlinear.[7,8]

Albalate *et al.* investigated the influence of speech intelligibility in cochlear implants users when filter-banks are used with different time-frequency resolutions.[9] Gopalakrishna *et al*. have presented the real-time implementation of wavelet-based advanced combination encoder on PDA platforms for cochlear implant.[10,11] A new cochlear implant acoustic simulation model was proposed by Mahalakshmi and Reddy based on a critical band finite-impulse response (FIR) filter-bank.[12]

A number of factors must be considered when implementing speech processing algorithms in cochlear implant applications: First, the complexity of the algorithm must be low, so as to minimize the power requirements; second, time-frequency resolution is required to be somehow similar to that of healthy ear[10] and third, the implementation must provide acceptable temporal resolution, which in turn will allow for the possibility of high-simulation rates.

The use of Fourier transform provides an excellent frequency resolution, but at the cost of limited temporal resolution. This is partially solved through the short-time Fourier transform (STFT) by using sliding analysis windows. However, the STFT uses a fixed window length and still cannot always simultaneously resolve short events and closely spaced long-duration tones in speech. Gopalakrishna *et al*. presented a real-time, and interactive implementation of the recursive Fourier transform approach on personal digital assistant (PDA) platforms for cochlear implant signal processing applications.[13]

The wavelet transform minimizes the limitation of the uncertainty principle by varying the length of the moving window with variant scaling factor. Wavelet transform is a time-frequency analysis for nonstationary signals, such as speech, electroencephalography, electrocardiography and so on.[14] The wavelet transform can be regarded as a bank of band-pass filters with constant Q-factor (the ratio of the bandwidth and the central frequency). The wavelet analysis has a distinct ability to detect local features of the signal in both time and frequency, such as the plosive fine structures of the speech and other transients. The speech processing property of cochlea is similar to that of wavelet transform; Since the cochlea is composed of a number of band-pass filters with constant Q-factors.[15] A damaged cochlea is not able to analyze the input speech into proper frequency bands. A speech processor is designed to overcome this defect and simulate the function of a healthy cochlea. The speech processor decomposes the input signal into different frequency bands,[2] and creates appropriate signals for application in the electrode array.

In the present study, we proposed the use of a speech processing strategy based on undecimated wavelet transform for frequency decomposition. To provide a denser approximation and to preserve the translation invariance, the undecimated wavelet packet transform (UWPT) has been introduced and was invented several times with different names as shift-invariant discrete wavelet transform (DWT),[16,17] algorithm *à trous* (with holes) and redundant discrete wavelet transform.[18] The UWPT is computed in a similar manner as the wavelet packet transform except that it does not down-sample the output at each level.[19] In Starck *et al.*,[20] it was shown that thresholding using an undecimated transform rather than a decimated one can improve the result in de-noising applications.

This paper is organized as follows. In the next section, information about speech processing strategies in cochlear implants is provided, and an undecimated wavelet-based strategy is described. The results are presented in Section III, where the performance of the method is assessed in terms of mean opinion score (MOS), short-time objective intelligibility (STOI) and segmental signal-to-noise ratio (SNR). Finally in Section IV the discussion and conclusion are given respectively.

MATERIALS AND METHODS

Speech Processing Strategies in Cochlear Implants

Processing strategies are used to translate incoming acoustic stimuli into electrical pulses that stimulate auditory nerve fibers. The various speech processing strategies developed for cochlear implants can be divided into three categories: Waveform strategies (e.g. compressed analog and continuous interleaved sampling (CIS), feature-extraction strategies (e.g. F0/F2, F0/F1/F2 and MPEAK) and "N-of-M" strategies.[21]

Continuous Interleaved Sampling

Researchers at the Research Triangle Institute developed the CIS approach to avoid the deformity of speech caused due to channel interaction by the summation of the current fields. It is referred to the channel interaction issue by using nonsimultaneous, interleaved pulses. In the CIS strategy, the acoustic signal passes through a set of band-pass filters that divide the waveform into four channels. Then, the envelopes of the band-passed waveforms are extracted by rectification and low-pass filtering.[21]

Some devices for instance use the fast Fourier transform (FFT) for spectral analysis while others use the Hilbert transform to extract the envelope instead of full-wave rectification and low-pass filtering. The envelope outputs are finally compressed and then used to modulate biphasic pulses. The compression is done by using a logarithmic function to fit the patient's dynamic range of electrically evoked hearing. The channel interaction problem is minimized by using nonsimultaneous, interleaved pulses. The CIS strategy is implemented in several implants: Clarion, Nucleus and Med-EL. The difference between these implants using CIS is mainly the number of channels (8 for Clarion, 22 for Nucleus and 12 for Med-EL).

N-of-M Strategy

N-of-M strategy divides the speech signal into M sub-bands and extracts the envelope information from each band of the signal. N bands that have the largest amplitude are then selected for stimulation (N out of M).[3] Only the electrodes corresponding to the N selected outputs are stimulated at each cycle. Thus, the bandwidth of a cochlear implant is limited by the number of channels (electrodes) and the overall stimulation rate. The channel stimulation rate represents the temporal resolution of the implant, while the total number of electrodes M represents the frequency resolution.

The basic aim here is to increase the temporal resolution by neglecting the least important spectral components and to concentrate on the more important features. Advanced combinational encoder (ACE) and SPEAK strategies, both of which are N-of-M type.[22] The SPEAK strategy uses a 20-channel band-pass filter-bank to perform a spectral analysis. The ACE strategy is similar to the SPEAK strategy but uses 22 channels and has the capability to provide stimulation at higher pulse rates of up to 2400 pps per channel.

Undecimated Wavelet-based Method

The UWPT is a translation invariant and redundant transform, where no decimation is done after the filtering. The key advantage of UWPT is that it is redundant and shift-invariant and gives a dense approximation to the continuous wavelet transform (CWT) than that provided by the orthonormal discrete wavelet transform.[23] Undecimated DWT (UDWT) coefficients are a collection of all DWT coefficients of different shifts of the signal. There is no down-sampling at all in the multi-resolution algorithm.[20] We can also consider UDWT coefficients as a collection of coefficients of DWTs with different down-sampling schemes. In the filter-bank implementation, this means both even samples and odd samples of the filtering output are kept and separately filtered at the next stage of iteration. The UWT W using the filter-bank *(h, g)* of a 1-D signal c_0 leads to a set $W = \{w_1, w_2, \ldots, w_j, c_j\}$ where w_j are the wavelet coefficients at scale j and c_j are the coefficients at coarsest resolution. Each new resolution is iteratively calculated using the Eqs. (1) and (2):[20]

$$c_{j+1}[l] = \sum_k h[k] c_j[l + 2^j k] \quad (1)$$

$$w_{j+1}[l] = \sum_k g[k] c_j[l + 2^j k] \quad (2)$$

Note that in the UWPT, the coarsest resolution is also iteratively decomposed like the fine resolution. The undecimated wavelet approach can be used to decompose the input speech signal into a number of frequency bands. Similar to the FFT-based N-of-M strategy, the number of maximum amplitude channel output, can be selected using a logarithmic compression map and stimulation. A second-order Butterworth low-pass filter (cut-off frequency 400 Hz) was used to obtain smooth envelopes of speech signals. The block diagram of the undecimated wavelet-based N-of-M strategy is shown in Figure 1. In this strategy, input speech signals are passed through a 6-stage wavelet packet decomposition yielding a 64-band output. A channel output is computed by summing up all the frequency-band output falling within the frequency range of the channel and is passed through a rectifier and

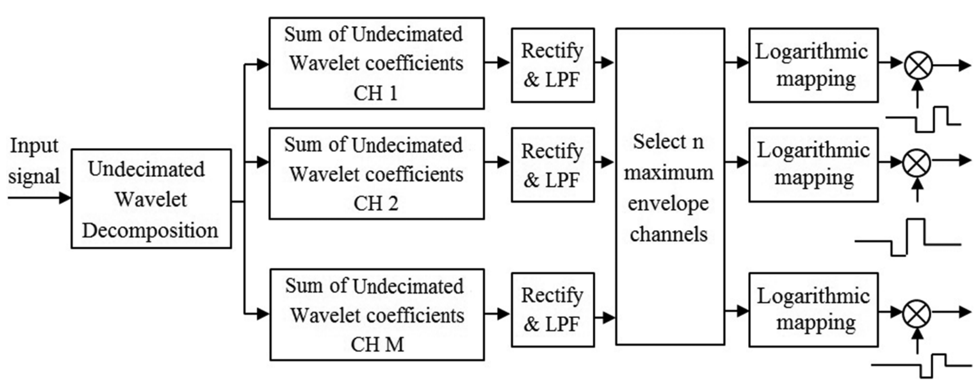

Figure 1: Block diagram of the N-of-M strategy using undecimated wavelet transform

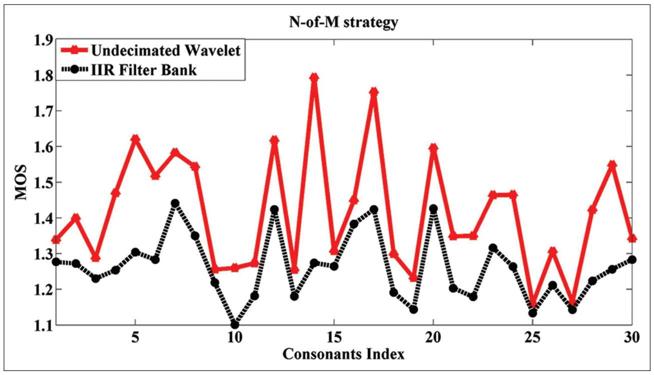

Figure 2: Comparison of mean opinion score for undecimated wavelet, and infinite impulse response filter-bank, both with N-of-M, implementations

then low-pass filtered to extract the channel envelope. The number of channels can be varied. The block diagrams of the traditional base CIS and N-of-M strategies were shown in Figures 1 and 2 of the manuscript written by Gopalakrishna *et al.*[10] Comparing with our proposed N-of-M structure, the FFT block was replaced with the undecimated wavelet and the rectifier and LPF was taken from the CIS strategy.[3]

With this undecimated wavelet-decomposition tree, the bandwidths of the channels become exactly the same as the frequency bands used in the Nucleus device.[10] The same approach can be applied to any other implant device with different frequency spacing and different number of channels.[24] Table 1 gives the cutoff frequencies of the channels corresponding to the binary tree structures for the input sampling rate of 16 kbps.

In our implementation, 512 sample windows were used to compute the undecimated wavelet-decomposition coefficient for six-stage decomposition. Both the Symmlet and Daubechies wavelet basis functions produced similar outputs.

Validation

The function of the proposed speech processing in cochlear implant devices was primarily to decompose the input speech signal into a number of frequency bands to extract 8 bands which have the largest amplitude for stimulation. The input speech was analyzed using undecimated wavelet-based on the specifications discussed in Section II. The envelope of the signal was derived by obtaining the absolute value of the signal at each time instant, that is, performing full-wave rectification. A second order infinite impulse response (IIR) low-pass filter with the cut-off frequency of 400 Hz was used to obtain smooth envelopes of the speech signals. To verify the function of the proposed method in the speech processor in cochlear implant, three validation criteria (MOS, STOI and segmental SNR) were used. The speech data used for the current study consisted of 30 consonants,[25] sampled at 16 kbps.

Table 1: Lower and upper cutoff frequencies of the channels (input sampling rate is 16 kbps)

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

Mean opinion score

The MOS test is widely known as an index for speech quality rating.[26] In recent years, some objectives MOS assessment methods were developed, such as perceptual evaluation of speech quality (PESQ). It evaluates the audible distortions based on the perceptual domain representation of two signals, namely, an original signal and a reduced signal which is the output of the system under test. On the other hand, ITU-T G.107 defines the E-model, a computational model combining all the impairment parameters into a total value. The principle of the E-model is based on the suppositions that transmission impairments can be transformed into psychological factors. The fundamental output of the E-model is a transmission rating factor R-value which is directly converted to a MOS estimate.[27] It is given by the Eq. (3):

$$R = R0 - Ie - Id - Is + A \tag{3}$$

where *Ro* depicts the basic SNR, '*Is*' represents the impairments occurring simultaneously with the voice signal, '*Id*' represents the impairments caused by delay, and '*Ie*' represents the impairments caused by low bit rate codecs.[28] The advantage factor *A* can be used for compensation when there are other advantages of access to the user. *R* can be transformed into a MOS scale by the Eq. (4):[29]

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + R(R-60)(100-R).7.10^{-6} & 0 < R < 100 \\ 4.5 & R > 100 \end{cases} \tag{4}$$

A version of PESQ known as P. 862.1 MOS-listening quality objective (MOS-LQO) optimized on a large corpus of subjective data representing different applications and languages, performs better than the original PESQ.

Thus, P. 862.1 MOS-LQO could be used as the estimate of the subjective MOS. It is obtained by first running the PESQ algorithm via a hardware toolbox called digital speech level analyzer (DSLA) and then mapping the measured PESQ result by:

$$y = 0.999 + \frac{4}{1 + e^{-1.4945x + 4.6607}} \quad (5)$$

Where x and y represent the raw PESQ score and the mapped P. 862.1 MOS-LQO score, respectively.[26] Also, DSLA is a measurement tool manufactured by Malden Electronics Ltd., Surrey, U.K. to perform MOS measurement.

Short-time objective intelligibility

In the development process of noise-reduction algorithms, objective measures are an essential tool for predicting quality and intelligibility of degraded speech signals. Otherwise, its quality or intelligibility would have been predicted using subjective listening that is costly and time consuming.

Some objective measures showed promising results for noisy speech subjected to reverberation and spectral subtraction, but has only been evaluated for stationary speech-shaped noise. They are less suitable for speech signals distorted by nonstationary noise sources and processed by time-varying and nonlinear filtering systems. To better take this type of distortions into account, STOI measure[30] by Taal *et al.* has proposed. This measure is the average linear correlation coefficient between a time-frequency representation of clean and noisy speech over time frames.

Among all objective measures, the STOI measure has the highest ability in predicting speech intelligibility because it provides highest correlation between objective prediction and subjective listening scores. This is different from other measures, which typically consider the complete signal at once, or use a very short analysis length. In general, STOI showed better correlation with speech intelligibility compared with other reference objective intelligibility models. STOI is the method that works well in most conditions.[31]

Time-domain signal-to-noise ratio

The time domain measures are usually applicable to analog or waveform coding systems. Their target is to reproduce the waveform itself. Acknowledge of SNR have an important role for system optimization. SNR and segmental SNR (SNRseg) are the usual performance measures used.[32,33] However, SNR is a poor assessor of subjective voice quality for a large range of speech distortion and therefore is of little interest as a general objective measure of voice quality. On the other hand, SNRseg represents one of the most popular classes of the time domain measures.

Segmental SNR calculates the average of the SNR values of short segments (15-20 ms). It is given as the following:

$$SNRseg = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10} \sum_{i=N_m}^{N_m+N-1}\left(\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right) \quad (6)$$

where x *(i)* and y *(i)* are the original and processed speech samples indexed by i is the number of samples, N and M are the segment length and the number of segments, respectively. Only frames with SNRseg in the range of − 10 to 35 dB were considered in the average.

RESULTS

The validation of the proposed method in terms of MOS, STOI and SNRseg quality measures were presented in the [Figures 2-4]. Also, [Figure 5] showed the comparison of MOS for CIS and N-of-M undecimated wavelet implementations.

Figure 2 shows the MOS scores obtained by each input speech for undecimated wavelet and IIR filter-bank base N-of-M strategy. In the N-of-M strategy, eight maximum amplitude analysis channels were selected out of 22. Figure 2 represented the oscillatory behavior of MOS according to the N-of-M strategy. Our proposed method had MOS values about two times than those of the IIR filter-bank indicating good performance score for the undecimated wavelet compared with IIR filter-bank. The average MOS values for the undecimated wavelet and the IIR filter-bank N-of-M implementations were 1.42 ± 0.16 and 1.26 ± 0.09, respectively.

The other objective measure of speech quality, the STOI, was used for comparing both methods implementations. Figure 3 shows the results in terms of the STOI for undecimated

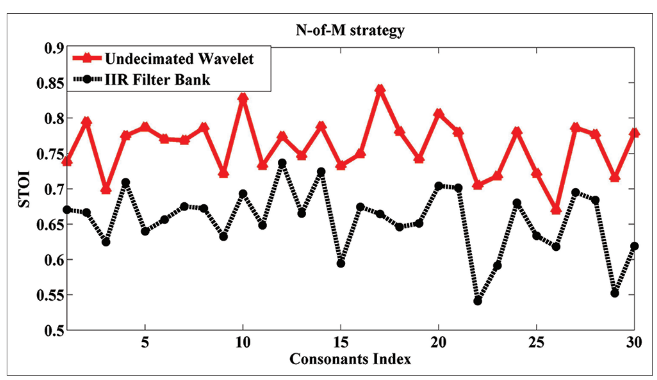

Figure 3: Comparison of short-time objective intelligibility for undecimated wavelet, and infinite impulse response filter-bank, both with N-of-M, implementations

wavelet and IIR filter-bank based N-of-M strategy. The STOI values for the undecimated wavelet and the IIR filter-bank N-of-M implementations were 0.76 ± 0.03 and 0.65 ± 0.04, respectively.

Figure 4 shows the SNRseg for undecimated wavelet as another validation index and compared it with that of IIR filter-bank. Although the IIR filter-bank is a conventional and commercial method, undecimated wavelet has better scores in about 96% of the input speech data. The average SNRseg values for the undecimated wavelet and the IIR filter-bank N-of-M implementations were 7.47 ± 5.09 and −0.26 ± 3.33, respectively.

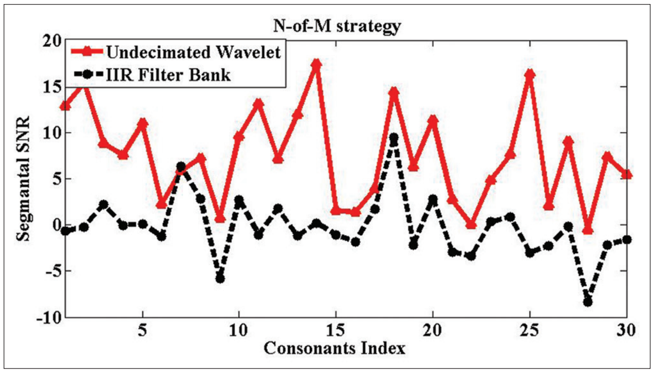

Figure 4: Comparison of segmental signal to noise ratio for undecimated wavelet and infinite impulse response filter-bank, both with N-of-M, implementations

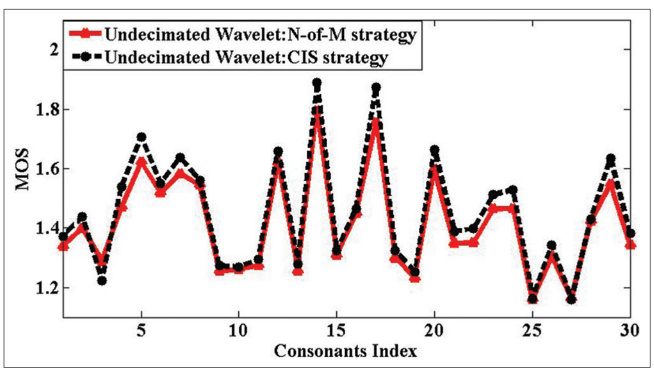

Figure 5: Comparison of mean opinion score for continuous interleaved sampling and N-of-M undecimated wavelet implementations

The MOS of 30 input speech data for the undecimated wavelet were showed in Figure 5 based on N-of-M and CIS strategies. The number of frequency bands was taken to be 22 for both strategies to ensure a fair comparison.[34] Eight frequency bands with the largest amplitude were extracted for stimulation in the N-of-M strategy. The average MOS values for the undecimated wavelet with N-of-M and CIS implementations were 1.42 ± 0.16 and 1.45 ± 0.19, respectively.

The electrode stimulation patterns (electrodograms) represent the activity of the electrode array for a given input signal. Figure 6 demonstrates the spectrogram of the input word "test" and the corresponding electrodogram. The spectrogram shows the amount of energy in a frequency versus time. Time is represented on the X-axis, and frequency on the Y-axis. In electrodogram, the X-axis represents time and the Y-axis is the exciting electrodes of the CI, and the colors indicate the level of energy for each electrode. The comparable color mapping was used for both spectrogram and electrodogram. White and black colors indicate the maximum and minimum energy intensities, respectively. As the bandwidths are not the same for all channels, the comparison between the spectrogram and electrodogram must be made with caution. For example, the bandwidth frequency of channel 1 is 7000-8000 Hz, while it is 125-250 Hz for channel 22.

DISCUSSIONS AND CONCLUSION

In this article, we presented an undecimated wavelet-based strategy to decompose the input speech signal into different frequency bands. The speech data used in our method consisted of 30 consonants that could be increased

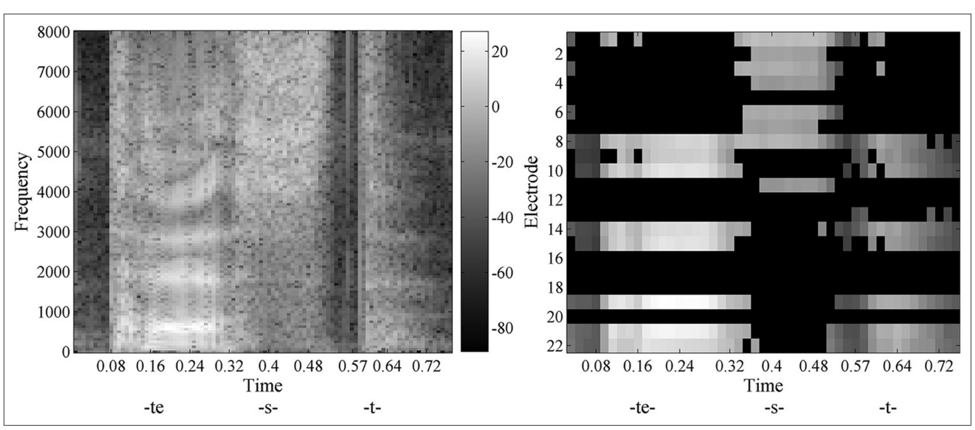

Figure 6: The spectrogram of the original acoustic signal (the word "test") at the microphone input of the sound processor (left). And the corresponding electrodogram using results obtained from undecimated wavelet strategy (right). The decomposition was done by N-of-M strategy (8 channels out of 22). The colormap shows the intensity

to achieve more generalized results. In the undecimated wavelet transform, *Sym2* wavelet was selected since it is suited for speech analysis. Also we compared the performance of the proposed undecimated wavelet-based N-of-M strategy with that of IIR filter-bank based N-of-M strategy, in terms of MOS, STOI and SNRseg.

The discrete wavelet transform is very efficient from the computational point of view.[24] The computational complexities of UDWT, WT and FFT are O ($N\log_2 N$), O (N) and O ($N\log_2 N$), respectively for a signal of length N.[16] The only drawback of WT is that it is not translation invariant. Translations of the original signal lead to different wavelet coefficients. In order to overcome this and to get more complete characteristic of the analyzed signal the undecimated wavelet transform was proposed. The UDWT has been independently discovered several times, for different purposes and under different names, e.g. shift/translation invariant wavelet transforms, redundant wavelet transform, or stationary wavelet transform. To grain noise reduction in ultrasonic nondestructive testing of materials, redundant wavelet processing was applied.[35] For various test signals and SNRs undecimated wavelet de-noising (UWD) performed considerably better than CWT. In contrast to CWT, UWD is shifted-invariant. Also, in contrast to continuous wavelet de-noising, smooth and accurate estimates can be computed simultaneously.[16]

The paired-samples *t*-test showed that the MOS, STOI and SNRseg scores obtained by the input speech data for undecimated wavelet-based N-of-M strategy yielded to a performance significantly higher that what obtained with filter-bank (t = 7.68, 15.88, 8.97 respectively; df = 29; $P < 0.001$). This finding showed that the proposed method outperformed the classical filter-bank implementation in terms of all of the performance criteria considered in this study.

A similar analysis showed that most of the performance indices used in this study for undecimated wavelet with N-of-M implantation were statistically different from those of CIS (t = −5.74, −10.60, −1.52 respectively; df = 29; P = 0, 0, 0.138). Thus, the results obtained based on CIS strategy in terms of MOS and STOI are significantly higher than N-of-M implantation.

Since the performance indices followed the normal distribution (one-sample Kolmogorov–Smirnov test; $P > 0.05$), parametric test *t*-test was applied for the inferential statistical analysis. The high power of the parametric tests in addition with the controlled Type-I error (α =0.05), could provide the fact that the results of this study could be generalized to any *similar* speech dataset. Thus, it could be deduced that the cochlear implant speech processing strategies using undecimated wavelet achieve a good performance in terms of MOS, STOI and SNRseg when compared with strategies using an IIR filter-bank. Although, our results have only been compared with the filter-bank, it is a conventional method commonly used in commercial strategies. Also, the computational complexity in the filter-bank is less than the wavelet method.

The main advantage of this type of decomposition of the input speech signal into frequency components compared with that of the IIR filter-bank is improving the deaf patients hearing ability. The basic advantage of IIR or FIR band-pass filters will lead to a simple design in filter configuration. Figure 5 illustrates the comparison of MOS for CIS and N-of-M, undecimated wavelet, implementations. The number of analysis channels is taken to be 22 for both strategies to ensure a reasonable comparison. When 8 channels or less were selected, significant differences were found between the N-of-M and CIS strategies.

In Figure 6 the areas with a white color, having the highest energies, are formants. In our example, they are near 625, 1900 and 3000 Hz. The white area on the spectrogram for 625 Hz formant is distributed in 0.16-0.29 s. This is in consistent with the strongest stimulation in electrodes 19 and 21. The next formant occurred in 0.15-0.28 s in the spectrogram, which is in consistent with the stimulation of electrode 10. Finally, the third 3000 Hz formant was provided by the electrode 8. Meanwhile, the main distinguished features, formants and variety of intensities of the speech signal were transferred and presented by using the proposed sound coding and speech processing.

To summarize, the implementation of filter-bank using undecimated wavelet transform presented a novel method to analyze speech signals in cochlear implant. Simulation results indicated that applying undecimated wavelet transform on speech processor for cochlear implant is feasible.

The UWT has the advantages of fast calculation, programmable filter parameters, and the same filter structures. The property of WT is in good agreement with the function of cochlea, so the method discussed in this paper might give a novel speech processing strategy for cochlear implants based on wavelet analysis. Furthermore, the implementation of speech processing in cochlear implant with wavelet transform might provide a new method for researches on hearing restoration for totally deaf people. Further studies on the application of wavelet transform to practical cochlear implant should be investigated in the future works.

REFERENCES

1. Loizou PC. Mimicking the human ear. IEEE Signal Process Mag 1998;15:101-30.

2. Loizou PC. Introduction to cochlear implants. IEEE Eng Med Biol Mag 1999;18:32-42.
3. Zeng FG, Rebscher S, Harrison W, Sun X, Feng H. Cochlear implants: System design, integration, and evaluation. IEEE Rev Biomed Eng 2008;1:115-42.
4. Lopez-Poveda EA, Meddis R. A human nonlinear cochlear filterbank. J Acoust Soc Am 2001;110:3107-18.
5. Lopez-Najera A, Lopez-Poveda EA, Meddis R. Further studies on the dual-resonance nonlinear filter model of cochlear frequency selectivity: Responses to tones. J Acoust Soc Am 2007;122:2124-34.
6. Kim KH, Choi SJ, Kim JH, Kim DH. An improved speech processing strategy for cochlear implants based on an active nonlinear filterbank model of the biological cochlea. IEEE Trans Biomed Eng 2009;56:828-36.
7. Kim KH, Choi SJ, Kim JH. A Speech Processor for Cochlear Implant using a Simple Dual Path Nonlinear Model of Basilar Membrane. ICBME; 2009. p. 257-60.
8. Kim KH, Choi SJ, Kim JH. A sound processor for cochlear implant using a simple dual path nonlinear model of basilar membrane. Comput Math Methods Med 2013;2013:153039.
9. Albalate A, Nogueira W, Edler B, Buchner A. Signal analysis by using adaptive filterbanks in cochlear implants. Biomedical Circuits and Systems Conference, 2006 BioCAS 2006 IEEE; 2006 Nov. 29 2006-Dec. 1 2006.
10. Gopalakrishna V, Kehtarnavaz N, Loizou PC. A recursive wavelet-based strategy for real-time cochlear implant speech processing on PDA platforms. IEEE Trans Biomed Eng 2010;57:2053-63.
11. Gopalakrishna V, Kehtarnavaz N, Loizou PC. Real-time implementation of wavelet-based Advanced Combination Encoder on PDA platforms for cochlear implant studies. Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on; 2010, 14-19 March; 2010.
12. Mahalakshmi P, Reddy MR. Cochlear implant acoustic simulation model based on critical band filters. Int J Electron Commun Eng Technol 2012;3:116-29.
13. Gopalakrishna V, Kehtarnavaz N, Loizou PC. Real-time PDA-based recursive Fourier transform Implementation for cochlear implant applications. Acoustics, Speech and Signal Processing, 2009 ICASSP 2009 IEEE International Conference on; 2009 19-24 April, 2009.
14. Kaibao N, Ning L, Shangkai G. Implementation of CIS speech processing strategy for cochlear implants by using wavelet transform. Signal Processing Proceedings, 1998 ICSP '98 1998 Fourth International Conference on 1998; 1998.
15. Tian G, Shuli Y, Datian Y. Application of wavelet in speech processing of cochlear implant. Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference; 2005 1-4, September.
16. Lang M, Guo H, Odegard JE, Burrus CS, Wells RO Jr. Noise reduction using an undecimated discrete wavelet transform. IEEE Signal Process Lett 1996;3:10-2.
17. Enders J, Geng W, Li P, Frazier MW, Scholl DJ. The shift-invariant discrete wavelet transform and application to speech waveform analysis. J Acoust Soc Am 2005;117:2122-33.
18. Fowler JE. The redundant discrete wavelet transform and additive noise. IEEE Signal Process Lett 2005;12:629-32.
19. Shensa M. The discrete wavelet transform: Wedding the a trous and Mallat algorithms. Signal Process IEEE Trans 1992;40:2464-82.
20. Starck JL, Fadili J, Murtagh F. The undecimated wavelet decomposition and its reconstruction. IEEE Trans Image Process 2007;16:297-309.
21. Loizou PC. Signal processing for cochlear prosthesis: A tutorial review. Circuits and Systems, 1997 Proceedings of the 40th Midwest Symposium on; 1997 3-6 Aug, 1997.
22. Nogueira W, Büchner A, Lenarz T, Edler B. A psychoacoustic NofM-type speech coding strategy for cochlear implants. EURASIP Journal on Applied Signal Processing. 2005;2005:3044-59.
23. Guo H. Theory and applications of the shift-invariant, time-varying and undecimated wavelet transforms [Masters Thesis]: Rice University; 1995.
24. Mehrzad M, Abolhassani M, Jafari A, Alirezaie J, Sangargir M. Cochlear Implant Speech Processing Using Wavelet Transform. ISRN Signal Processing, 2012; 2012.
25. Database. Available from: http://www.mpi.nl/world/dcsp/diphones/material/download.html. [Last accesed on 2014 Aug 11].
26. Lijing D, El-Hennawey MS, Goubran RA. Measurement of the Effects of Temporal clipping on Speech Quality. Instrumentation and Measurement Technology Conference, 2005 IMTC 2005 Proceedings of the IEEE; 2005 16-19 May, 2005.
27. Bergstra JA, Middelburg C. ITU-T Recommendation G. 107: The E-Model, a Computational Model for Use in Transmission Planning; 2003.
28. Lijing D, Goubran RA. Speech quality prediction in VoIP using the extended E-model. Global Telecommunications Conference, 2003 GLOBECOM '03 IEEE; 2003 1-5 Dec, 2003.
29. Lijing D, Goubran RA. Assessment of effects of packet loss on speech quality in VoIP. Haptic, Audio and Visual Environments and Their Applications, 2003 HAVE 2003 Proceedings The 2nd IEEE Internatioal Workshop on; 2003 20-21 Sept, 2003.
30. Taal CH, Hendriks RC, Heusdens R, Jensen J. An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. Audio, Speech, and Language Processing. IEEE Trans 2011;19:2125-36.
31. Jensen J, Taal CH. Speech Intelligibility Prediction Based on Mutual Information. Audio, Speech, and Language Processing. IEEE/ACM Trans 2014;22:430-40.
32. Mermelstein P. Evaluation of a segmental SNR measure as an indicator of the quality of ADPCM coded speech. J Acoust Soc Am 1979;66:1664-7.
33. Cohen I. Speech enhancement using a noncausal *a priori* SNR estimator. IEEE Signal Process Lett 2004;11:725-8.
34. Patrick JF, Clark GM. The Nucleus 22-channel cochlear implant system. Ear Hear 1991;12 4 Suppl: 3S-9.
35. Pardo E, San Emeterio JL, Rodriguez MA, Ramos A. Noise reduction in ultrasonic NDT using undecimated wavelet transforms. Ultrasonics 2006;44 Suppl 1:e1063-7.

How to cite this article: Hajiaghababa F, Kermani S, Marateb HR. An Undecimated Wavelet-based Method for Cochlear Implant Speech Processing. J Med Sign Sence 2014;4:247-55.

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES

Fatemeh hajiaghababa received the B.S. and M.S. degrees in communication engineering from Islamic Azad University of Najafabad, Iran, in 2009 and 2014, respectively. Her research interests include digital signal processing and speech processing for cochlear implants.

E-mail: hajaghababa@yahoo.com

Saeed Kermani obtained his B.S. from the Department of Electrical Engineering of Isfahan University of Technology in Isfahan, Iran, 1987, and he received the M.S. in Bioelectric Engineering from Sharif University of Technology, in 1992 and his Ph.D. 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 and image processing techniques

E-mail: kermani@med.mui.ac.ir

Hamid Reza Marateb received the B.S. and M.S. degrees from Shahid Beheshti University of Medical Science and Amirkabir University of Technology, Tehran, Iran, in 2000 and 2003, respectively. He received his Ph.D. and post-doctoral fellowship from the Laboratory of Engineering of Neuromuscular Systems, Politecnico di Torino, Turin, Italy in 2011 and 2012, respectively. He was a visiting researcher at Stanford University in 2009 and at Aalborg University in 2010. He was a visiting professor in UPC, Barcelona, in 2012. He is currently with the biomedical engineering department, faculty of engineering, the University of Isfahan, IRAN. His research interests include intra-muscular and surface electromyography and expert-based systems in bioinformatics.

E-mail: h.marateb@eng.ui.ac.ir