Improvements in Speech Recognition in Noise using Speech Enhancement Methods as a Function of Tonotopic Frequency Shift in Cochlear Implants

Kalyan S. Kasturi, P. Vishal Reddy, Vijayasimha Reddy Bonagiry and S. Vinod

Department of Electronics and Communication Engineering RMD Engineering College Chennai – 601206, India.

Received 2 September 2016; Accepted 2 December 2016; Published 31 December 2016

Address For Correspondence:
Kalyan S Kasturi, Department of Electronics and Communication Engineering RMD Engineering College Chennai – 601206, India
E-mail: kalyankasturi@hotmail.com

Copyright © 2016 by authors and American-Eurasian Network for Scientific Information (AENSI Publication).
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

ABSTRACT

Profound hearing loss severely debilitates a person’s life. Cochlear implant is a biomedical device used to restore hearing to a profoundly deaf person. Early cochlear implant devices consisted of a single electrode of stimulation. Current cochlear implant devices use anywhere from 8-22 electrodes of stimulation. In quiet conditions these devices produce high speech recognition. However speech recognition using cochlear implants drops in presence of noise. Researchers have used few speech enhancement algorithms to improve speech recognition in presence of noise. A systematic comparison of various speech enhancement methods in the perspective of cochlear implants has not been performed and such comparison could help the device manufacturers in choosing an appropriate speech enhancement algorithm. The current research work compares three noise reduction methods namely Wiener filtering method, Subspace method and spectral subtraction method in terms of frequency upshifting. Cochlear implant simulations were conducted to evaluate the effect of tonotopic frequency shift on the performance of the three speech enhancement algorithms mentioned above by varying frequency shift in three different ways. The cochlear implant simulations indicated that both Wiener filtering method and subspace method performed better than spectral subtraction method for all three conditions of tonotopic frequency shift.

KEYWORDS: Speech Enhancement, Cochlear Implants, Tonotopic Frequency Shift, Wiener Filtering.

INTRODUCTION

According to the “place theory”, the coding of frequency is performed by the auditory system using the vibrations of the basilar membrane in the cochlea. The vibrations of the basilar membrane in the cochlea cause the corresponding auditory nerves to be stimulated via “hair cells” that are attached to the basilar membrane. Most of the profoundly deaf people experience hearing loss due to the impairment of hair cells that connect the cochlea to the auditory nerve [1].

Cochlear implants are bio-medical devices consisting of an electrode array implanted into the inner ear of a profoundly deaf person to directly stimulate the auditory neurons in order to restore partial hearing [1]. A cochlear implant contains a signal processor and implanted electronics. The number of electrodes is referred to as the number of channels used in the cochlear implant. A speech processor controls the signals that are delivered to the electrode array. The current cochlear implant devices can produce 90% percent correct sentence recognition using 16 or more channels in several users [2].

The presence of background noise significantly degrades speech recognition with cochlear implant devices. Fetterman and Domico [3] studied the effect of addition of multi-talker babble noise on speech recognition using cochlear implant devices. The mean sentence recognition in quiet of 96 cochlear implant patients was around 85%. Addition of 10 dB babble noise caused sentence recognition to drop to around 70%. Further addition of 5 dB babble noise caused the sentence recognition to drop to around 45%.

Since speech recognition with cochlear implants is significantly decreased by the presence of background noise, a way to improve cochlear implant performance in noisy listening conditions is to use noise reduction techniques that suppress background noise.

The research study by Mauger et al. [4], studied cochlear implant user’s speech perception and listening preference of noise reduction with a range of gain functions. They used an advantageous gain function which is optimal for cochlear implant recipients. Using the optimised gain function, they reported a 27% improvement in speech weighted noise.

In summary, some noise reduction methods have been developed to improve speech recognition with cochlear implants but with varying degrees of success.

In the general area of speech enhancement, several noise reduction methods have been developed by various researchers [5-7]. However, not much research has been done to systematically compare these several noise reduction methods in the context of cochlear implant devices. In addition, investigation of the effect of various significant parameters of interest, such as spectral shifting, on the several noise reduction methods has not been conducted.

In cochlear implants, the various electrodes are implanted at various tonotopic locations in the inner ear. These various tonotopic locations correspond to different frequency bands. Since the electrode array insertion is performed using a surgical procedure, the spectral information delivered by the electrodes may not necessarily be matched to the actual tonotopic locations of the implanted electrodes. Due to this frequency mismatch, the speech recognition with cochlear implants decreases. This frequency mismatch commonly causes a spectral shifting effect, also called tontopic frequency shift in the cochlear implants [8]. Studying the effect of the frequency upshifting on the performance of various noise reduction methods is important since all cochlear implant users are affected by the spectral shift.

In the current work, three speech enhancement methods namely Wiener filter, spectral subtraction and subspace method are systematically compared using cochlear implant simulations to fully assess the effect of spectral shift on these three speech enhancement methods.

**Methodology:**

**A. Subjects:**

Five normal hearing listeners served as the subjects for the listening experiment. Sufficient training was provided to the subjects in listening to synthetic speech stimuli before the experiment.

**B. Speech Material:**

Sentences from the HINT database [9] were used as test material for sentence recognition. Noisy speech stimuli were created by adding speech-shaped noise from the HINT database to the sentences at 0 dB signal to noise ratio.

**C. Signal Processing:**

The sentence material was processed in two steps. In the first step of speech enhancement, the various noise reduction methods were implemented to produce the enhanced speech stimuli by using the three noise reduction methods namely spectral subtraction, wiener filtering and subspace method. In the second step of tonotopic shift simulation, a noise-band synthesis was performed using the various enhanced speech stimuli processed through eight channels with three degrees of spectral shift.

**Speech Enhancement:**

In this stage, three sets of enhanced speech stimuli were generated by implementing three noise reduction methods namely spectral subtraction, wiener filtering and MMSE spectral amplitude estimator.

The 1st set of noisy speech material was enhanced using Wiener filtering [5]. If speech signal \(x(t)\) is corrupted by uncorrelated noise \(n(t)\) then the resultant noisy speech be represented as \(y(t)\). We can calculate the power spectrum \(P_X(\omega)\) of the signal \(x(t)\) by taking the square of the modulus of the Fourier transform. The enhanced spectral signal is obtained by filtering noisy signal with the Wiener filter whose frequency response is given as per the below equation:

\[
H(\omega) = \frac{P_X(\omega)}{P_X(\omega) + P_N(\omega)}
\]  

(1)

By taking inverse Fourier transform of the enhanced spectral signal estimate, the enhanced signal is generated.
The 2nd set of noisy speech files were enhanced according to Berouti et al. [6] using spectral subtraction. The noise power spectrum is subtracted from the power spectrum of noisy speech to obtain an estimate of the power spectrum of the desired speech as follows:

\[ P_N(\omega) = \begin{cases} 
P_Y(\omega) - \alpha, & \text{if } P_Y(\omega) > \beta 
\end{cases} P_N(\omega), \text{ otherwise} \]  

(2)

The noise power spectrum is obtained by taking the average of the power spectrum of the noisy signal over several frames during silence period. \( \alpha \) is the over subtraction factor and \( \beta \) is the spectral floor. The inverse Fourier transform of the square root of the obtained power spectrum is calculated to obtain the desired signal.

The 3rd set of noisy speech material was enhanced using subspace estimator given in [7]. The subspace method divides the noisy speech vector as ‘desired signal’ subspace and ‘noise’ subspace. The noisy speech is represented by \( \tilde{y} = \tilde{x} + \tilde{n} \) using vector notation. To reduce the speech distortion with a limit on the noise distortion the estimator \( \tilde{H} \) can be obtained as:

\[ \tilde{H} = (\tilde{\Delta})^{-T} \tilde{\Lambda}^{-1} (\tilde{\Lambda} + \mu \tilde{I})^{-1} \]  

(3)

In the above equation, \( \mu \) is the Lagrange multiplier, \( \tilde{\Delta} \) is Eigen vector matrix and \( \tilde{\Lambda} \) is the diagonal Eigen value matrix.

**Simulation of Tonotopic Shift:**

In the second step of processing, the three speech enhanced sets along with a noisy speech set were subjected to cochlear implant simulation.

The inner ear has a snail-shaped structure with the outer end called as base and the inner end called as apex. The low frequency sound signals cause the basilar membrane to vibrate with highest oscillation occurring at the apex of the cochlea. The high frequency sound signals cause the basilar membrane to vibrate with highest peak occurring at the base of the cochlea.

When the electrode array of the cochlear implant is inserted into the inner ear, the low frequency electrodes cannot reach all the way up to the apex. Hence those low frequency electrodes will be located at slightly higher frequency locations in the inner ear. This creates a frequency mismatch between the low frequency stimuli and the actual location of the electrodes providing electrical stimulation. This is called frequency upshifting or tonotopic shift and causes reduction in the performance of cochlear implant.

In the processing condition where no spectral shift is present, the synthesis filters are exactly identical to the analysis filters. In order to simulate the spectral shift and the frequency mismatch in cochlear implants, the synthesis filters are shifted by a particular amount of frequency from the analysis filters.

Test material was subjected to a low-pass filter with cut-off frequency of 6000 Hz and then pre-emphasized using a pre-emphasis filter with a frequency boundary of 2000 Hz.

Speech material was then band-pass filtered into 8 frequency bands ranging in frequency from 300 Hz to 5500 Hz using sixth-order Butterworth filters. These filters are also called analysis filters.

A noise-band synthesis as described by Shannon et al. [10] was implemented to perform cochlear implant simulation. The output of each frequency channel was passed through a rectifier followed by a second order Butterworth low-pass filter with a center frequency of 160 Hz to obtain the envelope of each frequency channel. In order to remove the fine structure of the speech signal, the envelope of each channel was modulated with white noise to generate the various channel outputs. The outputs of the various frequency channels are again passed through band-pass filters which are called synthesis filters. Finally the processed speech was synthesized by summing up the outputs of all the frequency channels.

The relationship between the frequency and distance along the cochlea is given by the Greenwood’s equation [11] as follows:

\[ f = 165.4 \times \left(10^{(35-d)/0.06} - 0.88\right) \]  

(4)

In the above equation ‘\( f \)’ is frequency in Hertz and ‘\( d \)’ is distance in millimeters (mm).

Spectral shift was performed in 3 different ways by shifting the synthesis filters from the analysis filters by a frequency separation that corresponded to a shift of 0, 3 and 4.5 mm along the cochlear region. A frequency shift of 0 mm corresponds to the un-shifted condition where analysis and synthesis filter edge frequencies are the same and are given as per Equation (4).

The basal upward shifts of 3 mm was simulated by decreasing the distance by 3 and using the resulting shifted frequencies to generate the altered synthesis filter edges as given by the following equations:

\[ d_M = d - 3 \]  

(5)

\[ f = 165.4 \times \left(10^{(35-d_M)/0.06} - 0.88\right) \]  

(6)

The basal upward shifts of 4.5 mm was simulated by decreasing the distance by 4.5 and using the resulting shifted frequencies to generate the altered synthesis filter edges as given by the following equations:

\[ d_M = d - 4.5 \]  

(6)
\[ f = 165.4 \times \left(10^{(35-dN)+0.06} - 0.88\right) \quad (8) \]

Corresponding to the 4 sets of stimuli and 3 spectral shift values (0, 3 and 4.5 mm), a total of 12 test conditions were generated.

D. Procedure

The speech recognition experiments were performed using high quality Sennheiser HD circumaural headphones connected to a computer. A graphical user interface was developed to play the test material via the headphones. The subjects were instructed that they will hear a sentence via the headphones and must write down the all the words they understood.

To make the subjects familiar with synthesized speech, a practice session with 4 lists of ten sentences in quiet processed using 16 channels was conducted.

Following this, a pilot test session with ten sentences in quiet, processed using 16 channels was performed. Only the subjects who scored above 90 percent in the pilot test were allowed to participate in the experiment.

Following the completion of pilot test, the subjects were tested with the sentences processed through the various spectral shift conditions. The order of the various test conditions was randomized.

RESULTS AND DISCUSSION

The sentence recognition scores of the five subjects who participated in the experiment were averaged to generate the mean percent correct score for each condition. Thus the mean percent correct scores were obtained for all the 12 test conditions as described above.

The Table 1 shows the percent correct scores using the three noise reduction methods as well as the noisy condition for various values of spectral shift.

The condition where the spectral shift was 4.5mm resulted in low speech recognition due to greater frequency mismatch. On the other hand spectral shift value of 3 mm, resulted in moderately higher speech recognition due to smaller amount of frequency mismatch.

The spectral shift value of 0 mm, corresponds to the condition of no frequency mismatch and hence produced highest speech recognition across all the 4 different processing methods compared to the other two conditions of spectral up shifting.

<table>
<thead>
<tr>
<th>Spectral shift</th>
<th>noisy</th>
<th>specsub</th>
<th>wiener</th>
<th>subspace</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 mm</td>
<td>6</td>
<td>10</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>3 mm</td>
<td>18</td>
<td>28</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>0 mm</td>
<td>29</td>
<td>49</td>
<td>69</td>
<td>70</td>
</tr>
</tbody>
</table>

The spectral subtraction algorithm produced the lowest mean percent correct score among the three speech enhancement methods that were tested in this speech recognition experiment. This result is in agreement with the fact that spectral subtraction algorithm leads to speech distortion and introduces musical noise as a byproduct [12]. Statistical analysis using paired T-test also indicated that the Wiener estimator produced higher speech recognition than the spectral subtraction method (p<0.005). Statistical analysis indicated that the performance of subspace method was also significantly higher than the performance of spectral subtraction method (p<0.005). The performance of subspace method was similar to that of Wiener filter for all conditions of spectral shift.

The mean percent correct scores for the three noise reduction methods are presented in the pictorial form in the Figure 1. It can be observed that the speech recognition scores using both Wiener filtering method and subspace method were higher than traditional spectral subtraction method for all the conditions of spectral shift.
Conclusion:

Due to the variations in electrode-array insertion depths, tonotopic frequency shift is inherent to cochlear implants and limits the speech recognition with cochlear implants. The results of the present research study indicated that the performance of the Wiener filtering and subspace method was better than spectral subtraction method for the various conditions of spectral shift. Hence the use of Wiener filtering and subspace method to perform noise reduction in cochlear implants can benefit cochlear implant patients.

ACKNOWLEDGEMENT

This research was supported by Fast Track Young Scientist Research Grant via (DST No: SR/FTP/ETA-26/2011) given by the Department of Science and Technology (DST) and Science and Engineering Research Board (SERB), India. The author also appreciates the efforts of the subjects who participated in the listening experiment.

REFERENCES
