IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

Effect of Intensity Resolution on Wiener Filtering and Spectral Subtraction Method for Noise Suppression in Cochlear Implants

Kalyan S. Kasturi[1], K. Dinesh[1] and K. A. Dwarakanadh[1]

Department of Electronics and Communication Engineering, RMD Engineering College, Chennai, India[1]

Abstract: Cochlear implants are biomedical devices that serve as advanced hearing aids for profoundly deaf people. In quiet listening conditions and when no background noise is present, cochlear implant devices produce high speech recognition. However, speech recognition using cochlear implant devices significantly drops in the presence of ambient noise. Since speech recognition using cochlear implants drops due to noise, incorporation of noise suppression methods to suppress noise can improve speech recognition using cochlear implants. In this research paper, the speech recognition improvement using two such noise suppression methods namely, spectral subtraction and Wiener filtering was evaluated using cochlear implant simulations. The performance evaluation of the two noise suppression methods was done in a systematic manner by varying the number of intensity steps. In this experiment the intensity resolution for cochlear implant simulation was varied in four different ways, and the performance of Wiener filtering and spectral subtraction was evaluated for the four different values of intensity resolution. Results indicated that the speech recognition using Wiener filtering was significantly higher than the speech recognition obtained by spectral subtraction method for the four values of intensity resolution.

Keywords: Noise Suppression, Cochlear Implants, Intensity Resolution, Spectral Subtraction, Wiener Filter, Spectral Subtraction Method.

I. INTRODUCTION

Cochlear implants are prosthetic devices developed to aid the profoundly deaf people to obtain partial hearing. A cochlear implant contains an electrode array is which inserted into the inner ear of the profoundly deaf patient. The electrodes of the cochlear implant are stimulated using electrical pulses and hence hearing by cochlear implants is referred to as electric hearing. A signal processor is used to design the electrical pulse stimuli to stimulate the electrodes of the cochlear implant.

A research study that investigated the speech recognition using cochlear implants was reported in [1]. More than 30 cochlear implant patients were tested on speech recognition in quiet listening conditions. The material used for testing was everyday sentences used in normal conversations. The mean sentences recognition of all the cochlear implant patients was found to be around 90%. Hence the speech recognition using cochlear implants is high in the absence of ambient noise.

Researchers also investigated the performance of cochlear implants in noisy listening conditions where the input speech is corrupted by unwanted noise. A research experiment that studied the effect of noise on speech recognition using cochlear implants was reported in [2]. More than 90 cochlear implant patients were tested on sentence recognition in the presence of noise. The noise was added to the sentences at two levels of signal to noise ratio at 10 dB and 5 dB levels. The sentence recognition in the presence of noise at 10 dB signal to noise ratio dropped significantly to around 70%. For the more severe case of 5 dB signal to noise ratio, the sentence recognition significantly dropped to around 45%.
Hence the addition of noise significantly reduces the speech recognition obtained by the cochlear implant devices. Since the presence of noise is the main reason for the drop in speech recognition using cochlear implants, a method that reduces the amplitude or level of noise can help to improve the performance of cochlear implants in noise.

II. RELATED WORK

The speech signal which is corrupted by added noise is referred to as noisy speech. To perform noise suppression, some form of filtering or weighting function is applied on the noisy speech signal to reduce the noise level and increase the level of speech signal. Such algorithms are called speech enhancement algorithms since they suppress the noise level and enhance the level of speech signal. In this research paper we investigated the use of two speech enhancement methods namely spectral subtraction and Wiener filtering to improve speech recognition in noise with cochlear implants.

A method that performs spectral subtraction by subtracting the noise power spectrum from the power spectrum of the input signal to cancel the noise was presented in [3]. The

noise power spectrum was estimated from the silence portions of the speech signal. The noise suppression was performed by subtracting the estimated noise power spectrum from the power spectrum of noisy speech.
Another enhancement method called Wiener filtering that improves the speech signal quality by using a weighting function to filter out noise portion was developed in [4]. In this method, the transfer function of the Wiener filter minimizes the mean square error between the desired speech signal and the estimated speech signal.

In this research paper we used cochlear implant simulations to evaluate the performance of Wiener filtering and spectral subtraction. Cochlear implant simulations provide a practically feasible method to test various algorithms for improving cochlear implant signal processing instead of conducting experiments with cochlear implant patients. The availability of cochlear implant patients for participation in research experiments is very low. Also testing algorithms with cochlear implant users is a very tedious process and a time consuming task. Hence several researchers have used sinusoidal synthesis based cochlear implant simulations to evaluate the performance of various processing methods for improving cochlear implant technology [5, 6].

One of the key factors in conducting the cochlear implant simulation is the number of intensity steps that are used to code the speech information. If the number of intensity steps in cochlear implant simulation is low the speech recognition score will be less, whereas high number of intensity steps produce high speech recognition score [6]. When using noise suppression methods for cochlear implants, an important question is how many intensity steps are required to achieve asymptotic level of speech recognition. In the current work we conducted cochlear implant simulations by varying the number of intensity steps to evaluate the performance of Wiener filtering and spectral subtraction enhancement methods in improving speech recognition score in presence of ambient noise.

III. RESEARCH METHODOLOGY

A. Subjects

Five normal hearing listeners participated in the listening experiment. All the research subjects were given training in listening to synthetic speech prior to the start of the experiment.

B. Speech Material

Lists of sentences from the HINT database [7] were used as the speech material. Speech-shaped noise from the HINT database was added to the sentences at 0 dB signal to noise ratio.

C. Signal Processing

Signal processing consisted of two steps including noise suppression and sinusoidal synthesis. In the noise suppression, we implement the noise suppression algorithms to produce enhanced speech stimuli. In the sinusoidal synthesis, we perform the cochlear implant simulation.

Noise Suppression

In the first step of signal processing, two sets of enhanced speech stimuli were generated by implementing the two noise suppression methods namely spectral subtraction and Wiener filtering.

If speech signal $x(t)$ is corrupted by uncorrelated noise $n(t)$ then the resultant noisy speech can be given as:

$$y(t) = x(t) + n(t) \qquad (1)$$

The frequency domain representation of the noisy speech is given as follows:

$$Y(\omega) = X(\omega) + N(\omega) \qquad (2)$$

Since noise is additive and uncorrelated to the speech, the corresponding spectral representation can be formulated as:

$$P_Y(\omega) = P_X(\omega) + P_N(\omega) \qquad (3)$$

Where $P_Y(\omega) = |Y(\omega)|^2$ and $P_X(\omega) = |X(\omega)|^2$
The 1st set of stimuli consisted of the speech stimuli that were enhanced using spectral subtraction according to Berouti et al. [3]. 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 given by:

$$P_X(\omega) = \begin{cases} P_Y(\omega) - \alpha.P_N(\omega), \text{if } P_X(\omega) > \beta.P_N(\omega) \\ \beta.P_N(\omega), \ \text{else} \end{cases} \qquad (4)$$

In the above equation, α denotes the over subtraction factor and b denotes the spectral floor.
The estimate of noise power spectrum is obtained by averaging the power spectrum during the silence portion of noisy speech signal. The time frame of initial 120 milliseconds of the noisy speech was used to capture the noise signal.
Choosing the over subtraction is critical to the success of spectral subtraction. A high value of over subtraction factor can distort the speech signal where as a low value can result in high residual noise.
The over subtraction factor α is computed by the following equation:

$$\alpha = \alpha_0 - \frac{SNR}{s} \qquad (5)$$

SNR is the segmental signal to noise ratio computed for each time frame. The values for various parameters are:

$$\alpha_0 = 4, \quad s = \frac{20}{3} \quad \text{and} \quad \beta = 0.01 \qquad (6)$$

The noise power spectrum can be estimated as the average of the power spectrum of the noisy signal over several frames during silence period. Finally, the inverse Fourier transform of the square root of the obtained power spectrum is calculated to obtain the enhanced signal.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

The 2nd stimuli set consisted of speech material that were enhanced using Wiener filtering [4]. The desired signal is obtained by filtering noisy signal with the Wiener filter whose frequency response is given by:

$$H(\omega) = \frac{P_X(\omega)}{P_X(\omega) + P_N(\omega)} \quad (7)$$

where $P_X(\omega) = |X(\omega)|^2$ and $P_N(\omega) = |N(\omega)|^2$ are the power spectrums of the clean signal and noise respectively.

The Wiener filter can be expressed in a more generalized form as given below:

$$H(\omega) = \left(\frac{P_X(\omega)}{P_X(\omega) + \alpha . P_N(\omega)}\right)^{\beta} \quad (8)$$

The enhanced spectral signal estimate is obtained by filtering the noisy signal using the MMSE Wiener filter as:

$$\hat{X}(\omega) = H(\omega).Y(\omega) \quad (9)$$

By taking inverse Fourier transform of the enhanced spectral signal estimate, the speech enhanced signal is generated.

The 3rd speech stimuli consisted of the sentence material to which speech-shaped noise was added at 0 dB SNR. These noisy sentences were used to compare "without-enhancement" condition with the two enhancement conditions.

The following labels are used to identify the three stimuli sets as given by:

- specsub – Processed speech using spectral subtraction.
- wiener – Processed speech using Wiener filter.
- noisy – noisy speech for comparison.

Sinusoidal Synthesis

The next step in signal processing consisted of subjecting the three sets of speech enhanced stimuli to sinusoidal synthesis as mentioned in Loizou et al. [5]. Eight frequency channels were used to conduct cochlear implant simulation.

TABLE I FREQUENCY LIMITS FOR 8 CHANNELS

Channel number	Frequency Limits: Lower Cut-off Frequency (Hz)	Center Frequency (Hz)	Upper Cut-off Frequency (Hz)
1	300	366	432
2	432	526	621
3	621	757	893
4	893	1089	1285
5	1285	1566	1848
6	1848	2253	2658
7	2658	3241	3823
8	3823	4662	5500

Test material was passed through to a low-pass filter with 6000 Hz cut-off frequency and then passed through a pre-emphasis filter with 2000 Hz frequency limit. The low-pass filtered and pre-emphasized speech material was then subjected to filtering. This was done by using 8 logarithmically spaced band-pass filters in the frequency range from 300 Hz to 5500 Hz using sixth-order Butterworth filters. The cut-off frequencies for the 8-channel case are shown in the Table 1.

The output of each channel was passed through a full-wave rectifier followed by a second order Butterworth low-pass filter with a center frequency of 400 Hz to obtain the envelope of each channel output. By computing the root mean square energy of the envelopes using a frame length of 4 milliseconds, the envelope amplitudes were calculated for each of the channels used in the cochlear implant simulation.

The envelope amplitudes were then uniformly quantized to 'Q' discrete levels (Q = 4, 8, 16, Unquantized) to generate quantized envelope amplitudes. The 'Unquantized' condition corresponded to the normal amplitude values without performing any quantization.
A sine wave signal with amplitude equal to quantized envelope amplitude and with frequency value equal to the center frequency was generated in all the channels. The output signal was generated by adding the sine wave signals of all the channels.

D. Procedure

The subjects were instructed to hear the sentence via the headset connected to a computer. They were asked to write down the words in the sentence they heard. A high quality headset - Sennheiser HD circumaural headset was used in the experiment.
A practice session with sentences processed in quiet was conducted. This was followed by another practice session with processed sentences in noise at 0 dB signal to noise ratio. The sentences used in practice sessions were not used in the experiment.
After the conclusion of practice, subjects were tested with the sentences processed through the spectral subtraction and Wiener filter enhancement methods as well as noisy speech for different number of intensity steps.
All the 5 subjects were tested using total of 12 conditions corresponding to the 3 stimuli sets and the 4 intensity steps used. The order of the different processing conditions was partially counterbalanced between the different subjects.

IV. RESULTS

The sentence recognition scores for the five subjects were averaged to obtain the mean correct recognition score for each of the 12 conditions of processing.
For the 4 values of intensity resolution (4, 8, 16 and Unquantized), the mean sentence recognition values for the noisy speech, processed speech using spectral subtraction and processed speech using Wiener filter are

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

given in the Table II. It can be noted from the Table II that mean sentence recognition scores gradually increase as the intensity resolution increases.

It can be observed that both Wiener filtering and spectral subtraction method produced higher sentence recognition than the noisy speech for the 4 variations of intensity resolution.

Statistical analysis showed that mean sentence recognition using spectral subtraction was significantly higher than the mean sentence recognition using noisy speech ($p<0.005$).

TABLE II THE PERFORMANCE OF SPECTRAL SUBTRACTION AND WIENER FILTER FOR DIFFERENT INTENSITY STEPS

Intensity Resolution	noisy	specsub	wiener
4	8	21	30
8	15	30	40
16	26	46	59
Unquantized	29	49	69

A pictorial representation of mean correct recognition as a function of intensity resolution is shown in Figure 1. It can be observed that the performance of the Wiener filter was higher than that of spectral subtraction for all the 4 values of intensity resolution. Statistical analysis using paired T-test showed that mean correct recognition using Wiener filter was significantly higher ($p<0.005$) than that of spectral subtraction.

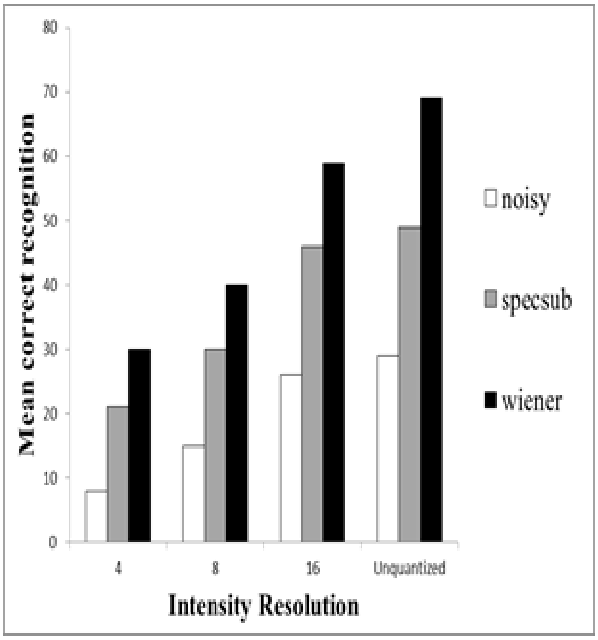

Fig. 1 Mean Correct Recognition as a Function of Intensity Resolution.

V. CONCLUSION

The present study evaluated the effect of intensity resolution on speech recognition in noisy listening conditions using Wiener filter and spectral subtraction enhancement from the cochlear implant perspective. A detailed comparison of the above two methods was performed using sinusoidal synthesis for 4 levels of intensity resolution. The results of the cochlear implant simulations indicated that the performance of Wiener filter was significantly higher than that of spectral subtraction enhancement.

ACKNOWLEDGMENT

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). The author also appreciates the efforts of the subjects who participated in the listening experiment.

REFERENCES

[1] A. J. Spahr, and M. F. Dorman, “Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices,” Arch. Otolaryngol. Head Neck Surg., vol. 130, pp. 624-628, 2004.
[2] B. Fetterman and E. Domico, “Speech recognition in background noise of cochlear implant patients,” Otolaryngol. Head Neck Surg., vol. 126, pp. 257-263, 2002.
[3] M. Berouti, R. Schwart and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” Proc. Int. Conf. Acoust., Speech, Signal Process., pp. 208-211, 1979.
[4] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” Proc. of the IEEE, vol. 67, no. 12, pp. 221-239, 1979.
[5] P. Loizou, M. Dorman and Z. Tu, “On the number of channels needed to understand speech,” J. Acoust. Soc. Am., vol. 106, no.4, pp. 2097-2103, 1999.
[6] P. Loizou, M. Dorman, O. Poroy and T. Spahr, “Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution,” J. Acoust. Soc. Am., vol. 108, no.5, pp. 2377-2387, 2000.
[7] M. Nilsson, S. Soli and J. Sullivan, “Development of Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise,” J. Acoust. Soc. Am., vol. 95, pp. 1085-1099, 1994.

BIOGRAPHY

Dr. Kalyan S Kasturi is an Associate Professor in the Department of Electronics and Communication Engineering at R.M.D Engineering College where he has been a faculty member since May 2010. He obtained B.Tech. Degree in Electronics and Communication Engineering in the year 1999 at VR Siddhartha Engineering College affiliated to Nagarjuna University, M.S. Degree in DSP and Communication at University of Texas at Dallas, USA in the year 2002 and Ph.D. In DSP and Communication at University of Texas at Dallas, USA in the year 2006. His area of interest is DSP and Communication. He is a senior member of IACSIT.