AENSI Journals
Australian Journal of Basic and Applied Sciences
ISSN:1991-8178
Journal home page: www.ajbasweb.com

Tamil Words Speech Synthesis in Cochlear Implant Using Acoustic Model

[1]T. Jayasankar, [2]Dr.J.Arputha Vijayaselvi, [3]K. Rajasekaran, [4]J. Jayalakshmi

[1]*Assistant Professor, Department of Electronics and Communication Engineering, Anna University BIT Campus, Tiruchirappalli.*
[2]*Dean (R&D) Department of Electronics and Communication Engineering, Kings College of Engineering, Pudukkottai.*
[3]Assistant Professor, Department of Electronics and Communication Engineering, *M.A.R College of Engineering, Tiruchirappalli.*
[4]*P.G Scholar, Communication Systems, PABCET, Tiruchirappalli.*

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form
8 May 2014
Accepted 20 May 2014
Available online 17 June 2014

Keywords:
Cochlear Implant, Speech synthesis, Channel vocoder, Impulse.

ABSTRACT

A cochlear implant (CI) is a surgically implanted electronic device that provides a sense of sound to a person who is profoundly deaf or severely hard of hearing. In this paper we have to analysis and synthesis the Tamil words in cochlear implants using the acoustic model. The channel vocoder is developed to perform the analysis and synthesis of Tamil words. System specific parameters are analyzed by developing uniform bandwidth filter bank based acoustic CI model. Acoustic CI simulations are generated for isolated words of Tamil language. The effect of number of channels on the speech quality is analyzed and a 21-channel vocoder is found to yield the best response with minimum mean square error. The impulse train is used as an excitation source to produce the synthesized speech.

© 2014 AENSI Publisher All rights reserved.

To Cite This Article: T. Jayasankar, Dr.J.Arputha Vijayaselvi, K. Rajasekaran, J.Jayalakshmi Tamil Words Speech Synthesis In Cochlear Implant Using Acoustic Model. ***Aust. J. Basic & Appl. Sci.,*** *8(10): 349-356, 2014*

INTRODUCTION

The cochlea is a part of the inner ear. It is a spiral structure with each portion sensitive to a particular frequency Loizou, (1998). The cochlea is therefore responsible for the decomposition of the speech signal. The cochlear implant is a prosthetic device used to restore partial hearing to a person with profound and bilateral hearing loss. It is different from a hearing aid in that, the latter simply amplifies the sound while the cochlear implant decomposes the sound into a number of frequency components and relays the decomposed signals to the brain.

A microphone picks up the speech signal; the external speech processor processes the speech signal using various speech processing strategies, and generates an encoded speech data to be received by the internal receiver stimulator through RF (radio frequency) link. The receiver decodes the signal and transmits the specified stimulation waveform to the intra-cochlear electrodes Loizou, (1998), Zeng *et al* (2008). The receiver stimulator is used for stimulating the auditory nerve via electrode array that enables understanding of the human speech by brain.

Design of cochlear implants started with single channel implants. These implants provided electrical stimulation at a single site in the cochlea using single electrode and tested with patients in the early 1970s that were capable of conveying time/envelope information, limited spectral information and had insufficient speech recognition Loizou, (1998). In order to improve the spectral information, multi-channel cochlear implants are developed in 1980s, which has multiple sites in the cochlea to provide electrical stimulation using an array of electrodes.

Multi-channel cochlear implants are the widely used systems. The major speech processing strategies that are used in the multichannel cochlear implants are waveform and feature extraction strategies based on whether the speech waveform or the feature is processed Vijayalakshmi et.al(2011).

Speech Data Collection:

Vowel sounds and Tamil words are recorded from a male speaker. The data were recorded at a sampling frequency of 16 kHz with a head mounted carbon microphone of frequency range 20 Hz – 20 kHz using s PRAAT tool.

Corresponding Author: T. Jayasankar, Assistant Professor, Department of Electronics and Communication Engineering, Anna University BIT Campus, Tiruchirappalli,Tamilnadu,India
E-mail: jayasankar_t@rediffmail.com

Channel Vocoder Based Model for Analysis on Vowels:

Initially the formant vocoder was developed Gladston et.al (2010) for this vowel synthesis but the hardware implementation of a formant vocoder would be complex since it requires units to estimate the formant frequencies Therefore, instead of choosing cut off frequencies of filters based on the formant frequencies, a number of uniform and width filters can be used. Such a system is the channel vocoder.

Effect of Number of Channels in Speech Quality:

The effect of the number of channels on speech quality is analyzed. The number of channels is increased from 3 to 21 in steps of 3. The quality of the synthetic speech is found to improve with increasing number of channels.

Result:

A 21-channel vocoder, with bandpass filters of bandwidth 400Hz, yields a better quality of synthetic speech than others with minimum mean square error. The vowel /அ/ recorded from a female speaker and the synthetic vowel generated by a 21-channel vocoder, are shown in Fig 1.

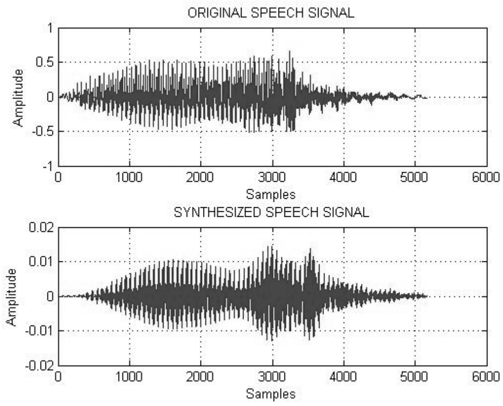


Fig. 1: Natural speech vowel /அ/ recorded from a female speaker and synthetic speech from 21-channel vocoder.

Channel Vocoder Based Model for Analysis on Words:

The acoustic CI models are developed based on channel vocoder which is an analysis-synthesis system The block diagram of an acoustic model based on Analysis-synthesis system as shown Fig2.. It consists of a speech analyzer and a speech synthesizer. The input speech signal is first filtered into a number of contiguous frequency channels using a bank of band-pass filters. The envelope of the signal in each channel is estimated by full-wave rectification and low-pass filtering. In addition to envelope estimation, the vocoder analyzer estimates the pitch period of the signal. This information is transmitted along with the envelope information.

The synthesizer modulates the received envelopes by voiced excitation source. The excitation signal is a train of impulses separated by pitch period as estimated in analyzer section. The modulated signals are subsequently band pass filtered by the same filters used in analyzer side and then added together to produce the synthesized speech waveform.

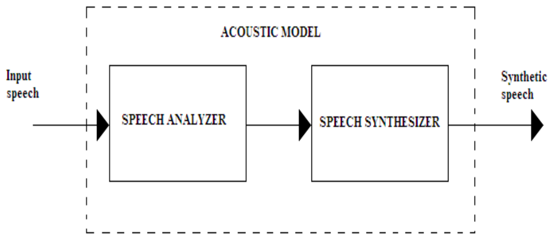


Fig. 2: General block diagram of an acoustic model based on Analysis-synthesis system.

Channel Vocoder Analyzer:

A block diagram of the channel vocoder analyzer is given in Fig 3. It consists of an analyzer filter bank for decomposing the incoming signal into its frequency components, a bank of envelope detector for extracting the envelopes of filtered signals and a pitch estimator for estimating the pitch period.

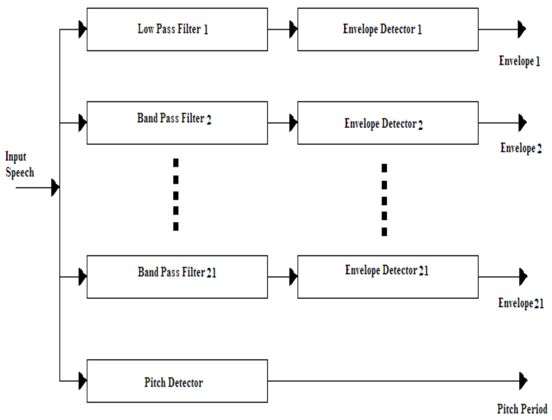


Fig. 3: Channel Vocoder Analyzer.

Filter design:

In channel vocoder, band pass filters are used for the frequency range of 200Hz to 8200Hz. A LPF having cut-off frequency of 200Hz is used for the first channel. The remaining channels are of uniform bandwidth range from 1200Hz to 400Hz. The numbers of channels are varied from 6-21 in steps of 3. The analyzer and synthesizer filters are of same type. The filter types used for analysis are IIR chebyshev type-1, chebyshev type-2 and their effects on speech signal are analyzed. The mean squared difference between the original speech and synthesized speech for various filter orders are obtained and it is shown in the Table 1. The number of channels required for this model is chosen as 21 with minimum mean square difference.

Table 1: Filter orders and corresponding mean squared difference.

Filter order	Mean squared difference
1	0.0033
2	0.0032
3	0.0031
4	0,0031
5	**0.0030**
6	1.4861e+238
7	4.3540e+300
8	1.6547e+299
9	2.7285e+296
10	3.3515e+299

Envelope detector:

This envelope detection method involves squaring the input signal and sending it through a low-order low pass filter. The squaring of signal results in demodulation of input by using the input as its own carrier wave. This means that half the energy of the signal is pushed up to higher frequencies and half is shifted down toward DC. Two operations are performed in order to preserve correct scaling. First the input signal is amplified by a factor of 2 so as to match final energy equal to its original energy. Then the square root of filtered signal is obtained with the purpose of reversing the scaling distortion due to squaring. The block diagram of envelope detector is shown in Fig 4.

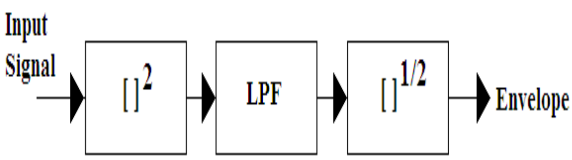


Fig. 4: Envelope Detector.

Pitch Estimator:

The pitch period of input speech signal is estimated using autocorrelation method. It is calculated by using a user defined function called 'pitchauto'. The computed pitch period is sent to synthesizer side for generation of excitation signal.

Australian Journal of Basic and Applied Sciences, 8(10) July 2014, Pages: 349-356

Channel Vocoder Synthesizer:

A block diagram of the channel vocoder synthesizer is given in figure 5. It consists of a bank of product modulator, a filter bank same as that of the analyzer filter bank, an excitation source for generating an excitation signal and an added that sum up all channel signals to produce the synthesized speech.

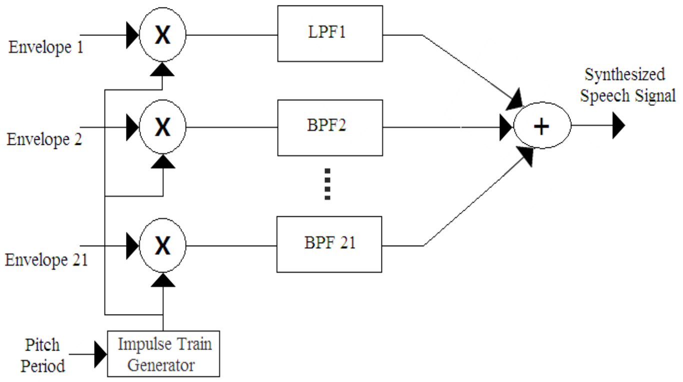


Fig. 5:.Channel Vocoder Synthesizer.

Excitation source:

The excitation source for voiced speech is an impulse train generator which generates train of impulses separated by pitch period. It is done by computing the number of samples by multiplying pitch period and sampling frequency. For the computed samples alone, the amplitude value is one.

Modulator and adder:

At receiver side, the envelope signals are product modulated by an excitation signal and passed through a synthesizer filter bank. The filtered signals are summed together by an adder to produce the synthesized speech signal.

Optimization:

In this model, some channel outputs are not having useful information. Hence such channel outputs are identified before summing the all channel synthesized filter outputs at the synthesizer section. It is done by taking autocorrelation of all channel synthesized filter outputs and finding the filtered outputs with complex autocorrelation values. Such channel filtered outputs are replaced with zeroes and then added all channel outputs to produce the synthesized words.

Input:

The input speech signals are words / அம்மா /,/ அன்பு /,/ மலர் /,/ மரம் /,/ முகம் /,/ நிறம் / and /பணம் / recorded from a single speaker using a microphone at the sampling frequency of 16000 Hz.

Acoustic model parameters:

Sampling Frequency	: 16000Hz
Frequency Range	: 0-8200Hz
Filter Type	: IIR Chebyshev type-2
No. of Channels	: 21 (1 LPF +20 BPF)
Bandwidth	: 400Hz
Order of filter	: 5

Results:

The optimum system specific parameters of acoustics CI model are chosen based on mean square difference between the original and synthesized vowels. The uniform bandwidth filter-bank CI model is designed with these optimum parameters has low mean square difference and improved speech quality. Synthesized speech is generated for isolated Tamil words input. The plot of signal at the output of each stage is displayed and played back the synthesized speech using ‘sound’ function.

The Waveform of input and synthesized speech for the Tamil word / அம்மா / is shown in Fig 6 and mean square difference of different Tamil words are shown in the table2.

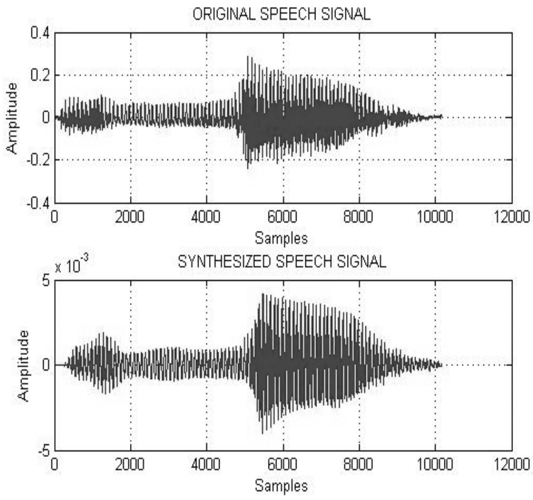


Fig. 6: Waveform of input and synthesized speech for the tamil word / அம்மா /.

Word Synthesis Using A Critical Bandwidth Filter Bank Based Acoustic CI Model:

Critical band is the smallest band of frequencies that activate the same part of basilar membrane and human ear can able to discriminate two tones that differ in critical bands. It is the frequency bandwidth of auditory filter created by cochlea.

The human auditory system is a very complex and highly non-linear one. The cochlea is often modeled as a bank of highly overlapping asymmetrical non-linear filters. The basilar membrane in the cochlea performs the frequency to place transformation. The bandwidth of each filter is often called as critical bandwidth which increases with frequency. Different critical bands correspond to different regions in the cochlea. It is defined as the bandwidth at which subjective responses of the hearing system change abruptly.

Table 2: Centre frequencies & corresponding ERBs.

Channel No.	fc in Hz	ERB in Hz	fcl = fc-ERB	fcu=fc+ERB
1	100	35.49	64.51	135.49
2	300	57.08	242.92	357.08
3	450	73.27	376.73	523.27
4	530	81.90	448.10	611.90
5	650	94.86	555.14	744.86
6	790	109.97	680.03	899.97
7	950	127.24	822.76	1077.24
8	1100	143.43	956.57	1243.43
9	1350	170.41	1179.59	1520.41
10	1600	197.40	1402.60	1797.40
11	1800	218.99	1581.01	2018.99
12	2100	251.37	1848.63	2351.37
13	2400	283.75	2116.25	2683.75
14	3200	370.10	2829.90	3570.10
15	3800	434.87	3365.13	4234.87
16	4600	521.21	4078.79	5121.21
17	5300	596.78	4703.22	5896.78
18	6400	715.51	5684.49	7115.51
19	6800	758.68	6041.32	7558.68
20	7400	823.45	6576.55	8223.45
21	7900	909.79	6990.21	8809.79

Design of CI model based on critical bands:

The CI model based on critical bands is also based on channel vocoder principle. The channel vocoder has two sections: vocoder analyzer and vocoder synthesizer. In the analyzer section, a filter bank is designed based on critical bands of the human auditory system. The critical band of each auditory band-pass filter is computed using equivalent rectangular bandwidth (ERB). If the center frequencies (fc) of filters are known, then the corresponding ERBs are calculated using the following formula,

$$ERB=24.7((0.00437*fc)+1) \qquad (1)$$

Table 2 lists the center frequencies, corresponding ERB values of filter bank. The lower and upper cut-off frequencies of each filter are calculated as (fc ± ERB). The input Tamil word is divided into frequency components by passing through the analyzer filter bank. The envelopes of 21 channel filtered signals are

Australian Journal of Basic and Applied Sciences, 8(10) July 2014, Pages: 349-356

obtained using an envelope detector. Then the pitch period of input vowel is estimated using a pitch detector by autocorrelation method.

The analyzer outputs such as envelopes and pitch period are passed through the synthesizer section. Since vowels are voiced sounds, the excitation signal is a train of impulses separated by pitch period. This excitation signal modulates the envelopes of all channels. The modulated signals of all channels are filtered through the synthesizer filter bank same as that of the analyzer filter bank. Then the filtered outputs are added to produce the synthesized word.

Results:

The waveform of input clean speech signal / அம்மா /, which is sampled at a frequency of 16,000Hz, is shown in Fig 7.

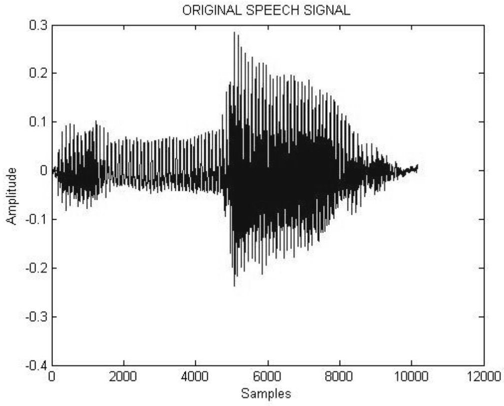


Fig. 7: Waveform of the tamil word / அம்மா /

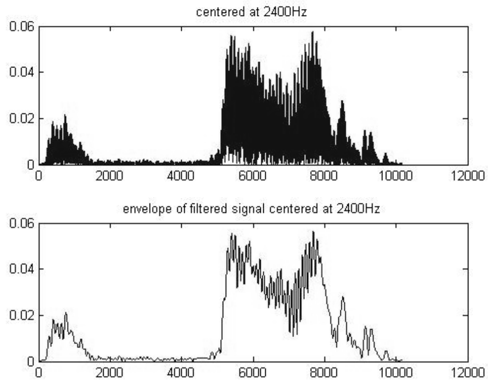
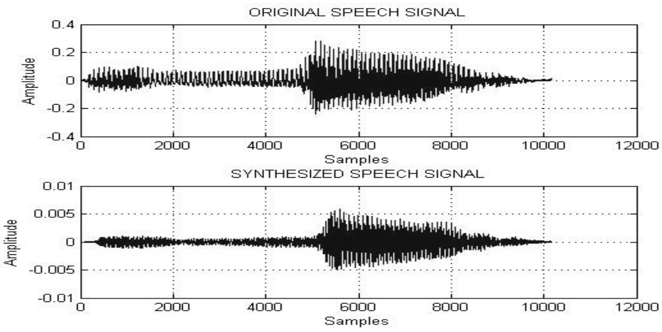


Fig. 8: (a)Waveform of filtered output of channel 13 (b)Waveform of the envelope channel 13 output.



Australian Journal of Basic and Applied Sciences, 8(10) July 2014, Pages: 349-356

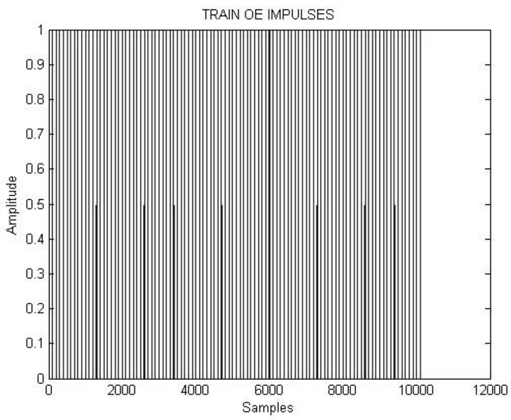


Fig. 9: Output of impulse train generator.

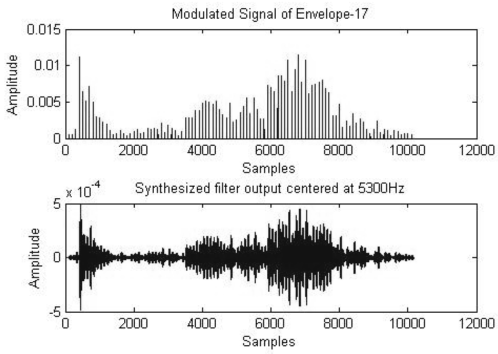


Fig. 10: (a) Waveform of modulated output of channel 16 (b) Waveform of the corresponding filtered output

Fig. 10: Waveform of input and synthesized speech for the Tamil word / அம்மா /

Discussion:

The optimum system specific parameters of acoustics CI model are chosen based on mean square difference between the original and synthesized vowels. The uniform bandwidth filter-bank CI model is designed with these optimum parameters has low mean square difference and improved speech intelligibility. Synthesized speech is generated for all Tamil words input. The plot of signal at the output of each stage is displayed and played back the synthesized speech using 'sound' function. The model performance is evaluated by conducting a closed-set listening test. In this test, a set of 15 listeners listened to these synthesized words and rated the intelligibility in 5 point grading system which varies from highly intelligible with a grade point of 5 to highly unintelligible with a rating of 1. The mean square opinion for this model is 4.2.

The noise signals are added to the adaptive filtered input speech in different magnitudes and mean squared difference between the original input and synthesized speech is used to analyze the model performance.

In the CI model based on critical bands, the filter bank has non-uniform bandwidth and it increases with frequency. This model is similar to that of uniform bandwidth filter bank based CI model except for center frequency and bandwidths. This model has a MOS of 4.5.

Table 3: Mean Square Difference Between Uniform Bandwidth Filter-Based CI Model And Auditory CI Model.

S.No	Tamil Words	Mean Square Difference	
		Uniform bandwidth filter-based CI model	Critical bandwidth filter-based CI model
1	அம்மா	0.0030	0.0029
2	அ ன்பு	0.0022	0.0022
3	ம ல ர்	0.0043	0.0042
4	ம ரம்	0.0031	0.0031
5	முக ம்	0.0010	0.0009
6	நிறம்	0.0030	0.0028
7	ப ணம்	0.0015	0.0015

Table 3 shows the mean square difference between the original and the synthesized Tamil words for the above two models. It is seen that the mean square difference of second model is lesser than the mean square difference of first model. Therefore, the performance of auditory CI model is better than the uniform bandwidth filter-based CI model.

Conclusion:

This paper proposes design of acoustic CI models based on waveform based strategy for synthesis of Tamil words. The uniform bandwidth filter based CI model & Critical bandwidth filter bank based CI model and their performance is evaluated based on mean square difference and speech quality. The filter parameters like filter type, filter bandwidth and numbers of channels which are system specific are chosen for this model design based on minimum mean square difference and speech quality. The envelope is detected using a simple method of squaring and low pass filtering. The pitch period of input speech is estimated accurately using autocorrelation method.

The model performance is evaluated using a closed-set listening test that showed a mean opinion score of 4.2 for uniform bandwidth filter-based CI model and 4.5 for auditory CI model based on critical bands. The auditory CI model outperforms the uniform bandwidth filter-based CI model in terms of intelligibility and mean square difference.

REFERENCES

An, S.K., S. Park, S.B. Jun, C.J. Lee, K.M. Byun, J.H. Sung, B.S. Wilson, S.J. Rebscher, S.H. Oh and S.J. Kim, 2007. “Design for a Simplified Cochlear Implant System”, *IEEE Trans. on Biomedical Engineering*, Vol. 54(6): 973-982.

Eddington, D.K., W.M. Rabinowitz and L. Dellzome, 2001. “Sound Processing for Cochlear Implants”, in *Proceedings of International IEEE EMBC*, pp: 3449-3452.

Gladston, A.R., P. Vijayalakshmi, N. Thangavelu, 2012. "Improving speech intelligibility in cochlear implants using vocoder-centric acoustic models," *Recent Trends In Information Technology (ICRTIT), 2012 International Conference on*, pp: 66-71.

Gold, B. and N. Morgan, 2000. “Speech and audio signal processing - processing and perception of speech and music”. John Wiley and Sons. Inc.

Jayanthan Ra, V., P. Vijayalakshmi and P.M. Kumar, 2009. “Auditory model based acoustic CI simulations for patients with profound hearing loss” – *in Proc. of Conference on Implantable Auditory Prosthesis (CIAP*), pp: 228.

Loizou, P.C., 2006. “Speech processing in vocoder-centric cochlear implants” *Cochlear and Brainstem Implants. Adv Otorhinolaryngol. Basel, Karger*, 64: 109-143.

Loizou, P.C., 1998. ”Mimicking the human ear” *IEEE Signal Processing magazine***,** 15(5): 101-130.

Liu, C. and Q. Fu, 2007. “Estimation of vowel recognition with cochlear implant simulations”, IEEE Trans. on Biomedical Engineering, 54(1): 74-81.

Loizou, P.C., 1999. “Signal-processing techniques for Cochlear implants” *IEEE Engineering in Medicine and Biology*, May/Jun, pp: 34-46.

McDermott, H.J., 1998. “How cochlear implants have expanded our understanding of speech perception” in *Proc. of the IEEE EMBC*, 20(5): 2251-2256.

Rabiner, L.R. and R.W. Schafer, 1978. Digital processing of speech signals. Prentice Hall.

Sit, J., A.M. Simonson, A.J. Oxenham, M. Faltys and R. Sarpeshkar, 2007. “A Low- Power Asynchronous Interleaved Sampling Algorithm forCochlear Implants That Encodes Envelope and Phase Information”, *IEEE Trans. on Biomedical Engineering*, 54(1): 138-149.

Vijayalakshmi, P., T. Nagarajan and Preethi Mahadevan, 2011. “Improving Speech Intelligibility in Cochlear Implants using Acoustic Models’’, WSEAS Transactions On Signal Processing, 4(7): 131-144.

Vijayalakshmi, P.P.M., Kumar, Ra. V. Jayanthan and T. Nagarajan, 2009. “Cochlear implant models based on critical band filters” – *in Proceedings of IEEE TENCON 2009*, Singapore, pp: 1-5.

Zeng, F.G., S. Rebscher, W. Harrison, X. Sun and H. Feng, 2008. *“*Cochlear Implants: System Design, Integration and Evaluation*” - IEEE Reviews in Biomedical Engineering*, 1: 115-142.