International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015
ISSN 2229-5518

Evaluation of Levinson-Durbin Recursion Method for Source-Filter Model based Artificial Bandwidth Extension Systems

G.Gandhimathi[1] , Dr. A. Hemalatha [2] Dr.S.P.K.Babu[3]

[1,2,3] School of Architecture Engineering and Technology. Periyar Maniammai University Thanjavur, India

gandhimathiece@pmu.edu

Abstract - Artificial Band Extension (ABE) techniques are used to generate a wideband signal from the narrowband signal. Since most of the high frequency components and the fricative consonants were absent even in the narrowband representation of the sound, it is a challenging task to create those missing components in the wideband equivalent signal. Generally, there are different techniques of autoregressive (AR) modeling to find the filter coefficients which is important in source-filter based ABE. In this paper, the performance of two of autoregressive (AR) modeling methods namely 1. Autocorrelation Method (LPC), 2.Levinson-Durbin Recursion Method, were evaluated. Generally, these estimation methods lead to approximately the same results (same coefficients) for a particular autoregressive parameters. But, the small differences in such estimations will have a great impact on the quality of reproduced sound . In this work, we implemented a source-filter model based artificial speech bandwidth extension systems with the above two AR modeling methods and validated their performance with suitable metrics.

Keywords: *Artificial bandwidth Extension , AR models , Levinson-Durbin Recursion*

1. INTRODUCTION

Best feature with high quality speech is needed in all miniature hands-held digital communication devices . Wideband (WB) codecs can resolve those problems exploring more bandwidth with high transmission bit rates . Presently due to the availability of Narrow band (NB) codec in all hand held devices can produce only low quality sound . We can achieve a good quality speech with less transmission bit rates by introducing Artificial Bandwidth Extension (ABWE) Technique. Using ABWE the missing low frequency component of NB (> 3.4 KHz to 4 KHz) can be predicted , and the high frequency component (4KHz to 7KHz) of input speech signals are created artificially at the receiving end .

1.1 Problem Specification

WB speech signal is shown in Fig 1 (top) and the corresponding NB representation is shown in Fig 1 (bottom) . In the NB , high frequency components of above 3.4kHz is missing. In the recording , even the WB signal above 7 khz is almost absent. It shows that , in the WB spectrum. the first harmonic components of the band 3.4kHz. to 4kHz is also missing (above 7kHz). Here the challenging task will be the Creation of missing components between 3.4kHz. to 4.6kHz

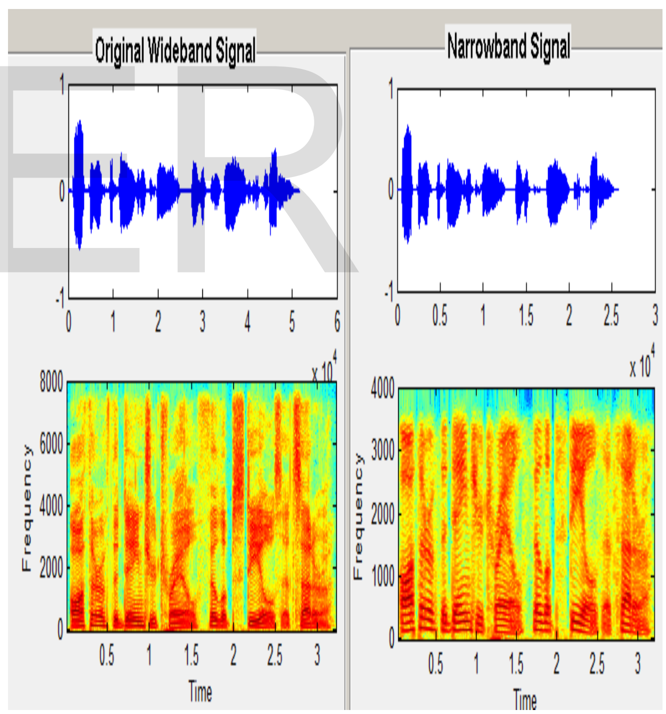


Fig 1. WB and NB Signal (Time domain-Top) & (Frequency domain -Bottom)

In this work , by developing a source-filter model based ABE system and evaluate its performance by introducing autoregressive (AR) modeling methods namely Autocorrelation Method (LPC), Levinson-Durbin Recursion Method using ARCTIC sound database [9] of Carnegie Mellon University. Compare the proposed ABE systems its limitations also

1.2 Previous works

Schitzier (1998) gives a solution to reduce the bitrates for coding wideband speech, is to code the parameters of wider bandwidth speech with significant increase in bitrates relative to NB coders. Makhoul and Benarti (1979), Carl and Heute (1994), Yoshida and Abe (1994), Jax and Vary (2000) were discussed another approach that employ the WB enhancement by analysis and synthesis model. This technique synthesis the WB speech from the Pitch, Voicing , and spectral envelope information of NB speech. Many ABE methods, e.g. codebooks [2,3] , linear mapping [4] , Neural Networks etc is used to estimate the missing components. Again Jax and Vary [6, 7] found the potential features of speech and evaluate their performance for BWE application.

This paper organized as; in section II discussed the proposed methodology. The design of the proposed source filter modeled Levinson-Durbin Recursion Method ABE is to be discussed in Section III, Section IV convolutes the results, Section V wrap up its performance and the future work.

2. SOURCE-FILTER MODEL (VOCAL TRACT) OF SPEECH PRODUCTION

The Vocal tract assumed as all pole filter. Speech generation can be modeled by source-filter model of VT and this is shown in Fig 2.

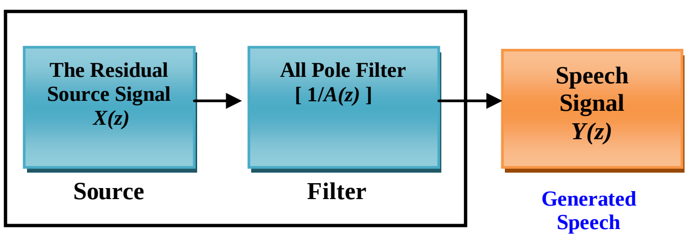


Fig 2. Source-Filter Model of Vocal Tract

Let VT be a all-pole filter with p^{th} order

$$\frac{1}{A(z)} = \frac{1}{1+a_1 z^{-1} + \dots\dots\dots + a_p z^{-p}} \dots\dots\dots\dots(1)$$

Here $a_1 \dots a_p$ are the filter coefficients and there by estimated using linear prediction method

Output Speech signal *Y(z)* in frequency domain is formed by multiplication of residual/source signal *X(z)* by the VT all-pole filter 1/*A(z)* or by filtering *X(z)* by 1/*A(z)* in Time domain .

$$Y(z) = X(z) / A(z) \quad \dots\dots\dots\dots\dots\dots\dots(2)$$

A(z) can be obtained by using the filter coefficients $a_1 \dots a_p$

$$X(z) = Y(z)\,A(z) \quad \dots\dots\dots\dots\dots\dots\dots(3)$$

2.1 AUTOREGRESSIVE MODELING

2.1.1 Autoregressive Modeling

The successive samples y of an autoregressive process linearly depend on their predecessors:

$$y_t + a_1 y_{t-1} + .a_2 y_{t-2} + \dots\dots + a_p y_{t-p}. = \eta_t \dots\dots(4)$$

in which a_i are the autoregressive parameters and the innovations η_t are a stationary purely random process with zero mean. It can be shown that the auto covariance function R_τ for delays 0 to p is related to the autoregressive parameters a_i through the Yule-Walker equation for the autoregressive process (Priestley, 1994):

$$\begin{pmatrix} R_0 & R_1 & \cdots & R_{p-1} \\ R_1 & R_0 & \cdots & R_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p-1} & R_{p-2} & \cdots & R_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{pmatrix} \dots\dots\dots\dots(5)$$

An estimated autoregressive model of the same order p can be written as

$$y_t + \hat{a}_1 y_{t-1} + .\hat{a}_2 y_{t-2} + \dots\dots + \hat{a}_p y_{t-p}. = \hat{\eta}_t \dots\dots\dots\dots(6)$$

in which

$\hat{a}_i$ are the autoregressive-parameter estimates and

$\hat{\eta}_i$ are the estimated innovations.

A clear distinction should be made between the autoregressive process (Eq. (4)) and the corresponding autoregressive model (Eq. (6) (Broersen and Wensink, 1993). Using Eq. (6), each data sample can be predicted from its predecessors:

$$\hat{y}_i = -\sum_{i=1}^{p} \hat{a}_i y_{t-1} \dots\dots\dots\dots\dots(7)$$

As the samples y_t cannot be predicted exactly, a residue is introduced, which is defined as the difference between the measured value and the estimated value:

$$residue \equiv y_t - \hat{y}_t = \hat{\eta}_t \ldots\ldots\ldots\ldots\ldots(8)$$

which means that the residue is equal to the estimated innovation, as introduced in Eq. (6).

It is assumed in these equations that the autoregressive model order p is known. In practice, the model order has to be estimated as well, which is usually done using Akaike's criterion (Priestley, 1994). Suppose that the estimation realization y consists of N data points (an estimation realization contains those data points that are used for parameter estimation). Two methods of autoregressive-parameter estimation from these data samples shall be considered here

1. LPC (Autocorrelation Method)
2. Levinson-Durbin Recursion Method

2.2. Different Methods of Spectrum Estimation

In many applications one is interested in evaluating the power spectrum of a random signal rather then the Fourier transform of a particular waveform. The power spectrum tells us about the expected or average power of a signal at each frequency in the spectrum.The various methods of spectrum estimation are available. They can be categorized as follows:

- Nonparametric methods
- Parametric methods
- Subspace methods

Nonparametric methods are those in which the estimate of the PSD is made directly from the signal itself. The nonparametric approaches do not assume any specific parametric model for power spectral density (PSD). They are based solely on the estimation of the autocorrelation sequence of the random process from the observed data. The simplest such method is the periodogram. An improved version of the periodogram is Welch's method. A more modern nonparametric technique is the multitaper method (MTM).

For the parametric approaches we first postulate a model for the process of interest, where the model is described by a small number of parameters. Based on the model, the PSD of the process can be expressed in terms of the model parameters. In other words, parametric methods are those in which the signal whose PSD we want to estimate is assumed to be output of a linear system driven by white noise. Examples are the Yule-Walker autoregressive (AR) method and the Burg method. These methods estimate the PSD by first estimating the parameters (coefficients) of the linear system that hypothetically "generates" the signal. They tend to produce better results than classical nonparametric methods when the data length of the available signal is relatively short.

Subspace methods, also known as high-resolution methods or super-resolution methods, generate frequency component estimates for a signal based on an eigen analysis or eigen decomposition of the correlation matrix. Examples are the multiple signal classification (MUSIC) method or the eigenvector (EV) method. These methods are best suited for line spectra - that is, spectra of sinusoidal signals - and are effective in the detection of sinusoids buried in noise, especially when the signal to noise ratios are low.

2.2.1. LPC (Autocorrelation Method)

LPC uses the autocorrelation method of AR modeling to find the filter coefficients. The generated filter might not model the process exactly even if the data sequence is truly an AR process of the correct order.

2.2.2 Levinson-Durbin Recursion Method

The Levinson-Durbin recursion is an algorithm for finding an all-pole IIR filter with a prescribed deterministic autocorrelation sequence. It has applications in filter design, coding, and spectral estimation.

3. ABE WITH DIFFERENT AR MODELING METHODS

The main scope of this work is to evaluate different AR modeling methods under a source filter based ABE system. There are a number of alternate useful representations of the predictor coefficients. The most important are the line spectrum pairs, reflection coefficients, log-area ratios and the roots of the predictor polynomial. For speech coding purposes, LP polynomial A(z) can be decomposed into LSF. LSFs have good quantization and interpolation properties and are thus widely used in speech coding.

In this work, we assumed an simple and ideal code book. That is, for each narrowband LSF/AR coefficients, directly the corresponding wideband LSF/AR coefficients were mapped. If we design a ABE system without code book, (one such as a steganography based ABE system where the wideband information itself directly embedded inside the narrowband signal itself) the simple model explained in the following algorithm will be sufficient to design such system. (We can design it even without AR-LSF conversion, but, we used AR-LSF conversion in this model so that this method can be used for the system which may use a codebook also)

3.1 The outline of Artificial Bandwidth Extention

Each narrowband signal frame is decomposed into a source part and a filter part and the parts are extended separately. The vocal tract is modeled as an all-pole filter and the filter coefficients are estimated using AR modeling. The model residual is used as a source signal. The vocal tract model is extended using the most suitable wideband model taken from a codebook or directly decoded from the file and the residual signal by time domain zero-insertion. The created signal is added to a resampled and delayed version of the original narrowband signal to form an artificial wideband signal.

3.2. The Simplified ABE Algorithm

The basic idea is to create a signal that contains the frequencies that are missing from the original narrowband signal. (here we assume that the wideband LSF/LR coefficients were resolved from the encoded narrowband signal itself)

In the following algorithm in Fig 3 , for AR modeling, we will use any one of the selected algorithm at a time.

1. The narrowband signals which is to be extended is opened and a suitable narrowband pre-emphasis filter was applied on the narrowband signals.
2. Increase the sampling rate of the narrowband signal using time domain zero-insertion. This will create a signal whose spectrum on the band 4-8kHz is a mirrored copy of the spectrum on the band 0-4kHz (check the spectrogram).
3. The n^{th}-order AR-coefficients of each narrowband frame is calculated and converted in to into LSF coefficients.
4. The wideband LSF coefficients were resolved and converted into AR-coefficients for waveform synthesis (Filter part Extension).
5. Using the narrowband signal frames from (1) and its AR-coefficients coefficients calculate the source signal.
6. Extend the narrowband source signal using zero insertion technique. (source extension)
7. Using the extended source signal from(6), and the AR-coefficients from (4) calculate the output signal.
8. The extended frames from(7) are concatenated using overlap-add technique. For overlap-add, a 25ms analysis window and a 10ms synthesis window are used. The time difference between adjacent frames is 5ms in both analysis and synthesis.
9. To get the final synthetic wideband signal, the signal from (9) and the delayed version of resampled narrowband signal were added/mixed. Since the extension causes a delay in the signal so the resampled narrowband signal must be delayed to synchronize the signals.

During AR Modeling, the order of 10 is used in the case of narrowband signal and the order of 18 is used for wideband signal.

4. RESULTS AND DISCUSSION

4.1 The Speech Database

Carnegie Mellon University ARCTIC database by SLT (CMU ARCTIC SLT 0.95) contains (1132 utterances) a recording of the phonetically balanced US English speech by a female US English speaker. The speaker is experienced in building synthetic voices.

4.2 Performance Evaluation

We used three methods for evaluating the performance of the ABE techniques. First comparing the spectrograms of extended signal with the spectrograms original wideband signal , then second Mean Opinion Score (Listening test) and third Log Spectral Distance .

4.2.1 Visualization using Spectrogram

Spectrogram, is a visual demonstration of sound. Since it is based on real measurements of the varying frequency component of a sound with time, spectrogram provides more absolute and precise information.

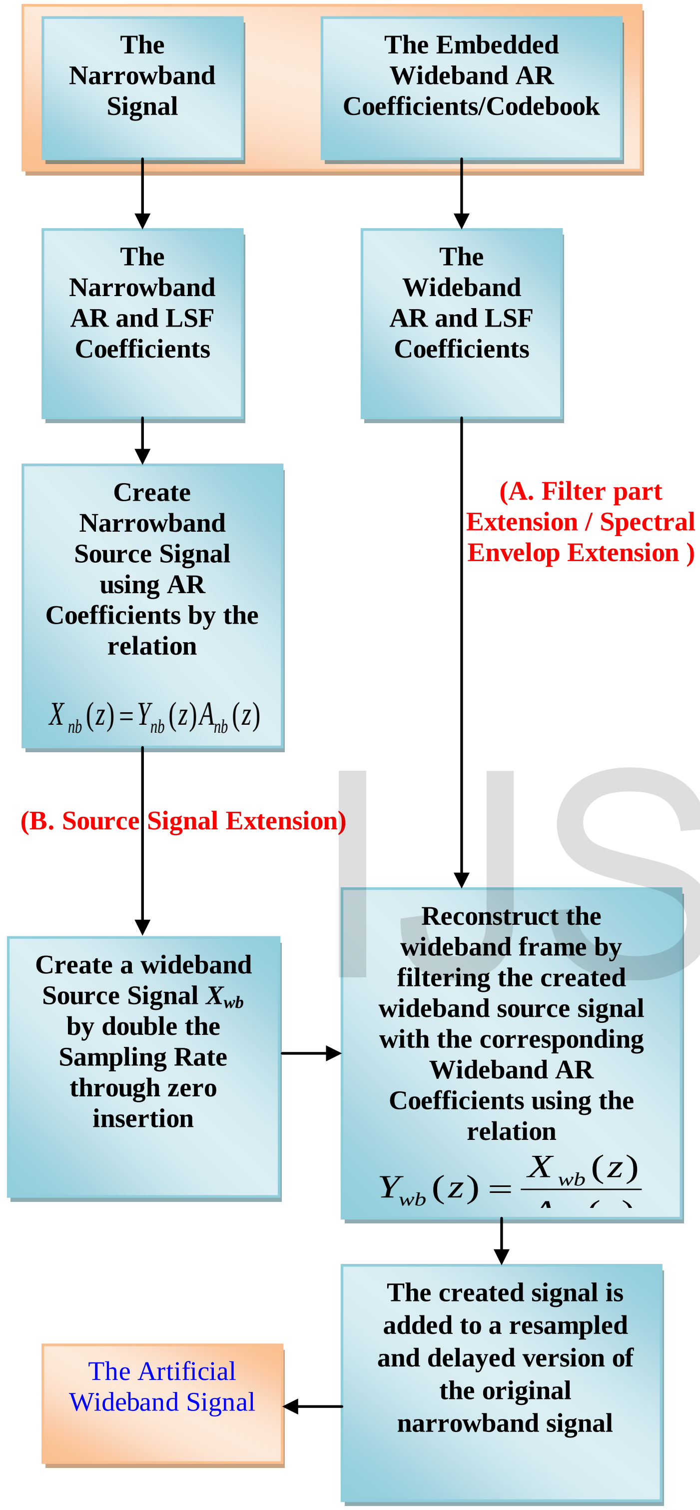


Fig 3 Analysis and Synthesis Model

4.2.2 Mean Opinion Score (MOS)

For subjective measure, MOS preference test is conducted between the Original NB Signal, the interpolated version of NB signal and the ABW Extended version of signal. The objective of this test is to find which signal has the more hearable or understandable high frequency component sounds such as "sh". We selected 10 NB files in which the high frequency components are almost absent but present in the corresponding original WB signals.

The 10 audio files of NB signal, TDI version , and the ABWE signals were presented to 10 listeners (5 females and 5 males) between 21 and 25 years of age having no auditory disorders were involved in the test. All the listeners had adequate knowledge on English and Phonetics for understanding all the given English speech signals/files.

For each listener the test was arranged individually in a quiet room using a simple graphical user interface (GUI) on a computer screen. Through high quality headphones, test sample files were played on both ears of each listener. Before itself , the listeners had a little practice session.. During the practice session the listeners can be allowed to adjust the volume setting to a suitable level. The arithmetic mean of all the individual scores is MOS, and generally its range is from 1 for worst case to 5 for best. The test results are tabulated in Table 2. The average comparison of bandwidth extension using TDI and ABWE are shown in Fig 4 .Spectrogram comparison and Log Spectral Distance (LSD) of ABE of narrow band *arctic a0001.wav* using Auto Correlation and Levinson – Durbin method are shown in Fig 5

4.2.3 *Log Spectral Distance (LSD).*

It is based upon difference between the logarithmic distance of the original spectra **P (ω)** and the recreated wideband spectra **P($\hat{\omega}$).**

$$D_{LSD} = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[10\log_{10}\frac{P(\omega)}{\hat{P}(\omega)}\right.d\omega}$$

Table 1. MOS Scores and their Meaning

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

ISSN 2229-5518

The values do not need to be whole numbers. Certain thresholds and limits are often expressed in decimal values from this MOS spectrum. For instance, a value of 4.0 to 4.5 is referred to as toll-quality and causes complete satisfaction. generally, the values dropping below 3.5 are termed not acceptable by many users. The Spectrogram comparison of LPC method and LD Method is tabulated in Fig 5

Table 2 MOS of arctic_a0001 –a0010.wav

Audio File	Mean Opinion Score		
	Original NB Signal	Interpolated WB Signal	ABE WB Signal
1	3.00	3.05	4.00
2	3.00	3.40	4.00
3	3.25	3.50	4.50
4	3.00	3.75	4.00
5	3.00	3.50	4.25
6	3.25	3.00	4.75
7	3.00	3.25	4.50
8	3.00	3.50	4.50
9	3.00	3.73	4.25
10	3.00	3.50	4.00
Avg.	3.05	3.463	4.275

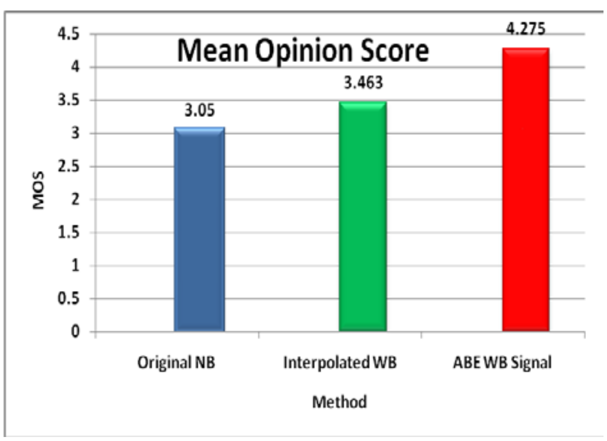
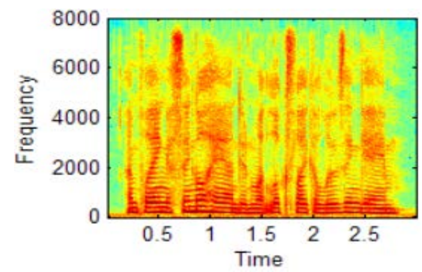
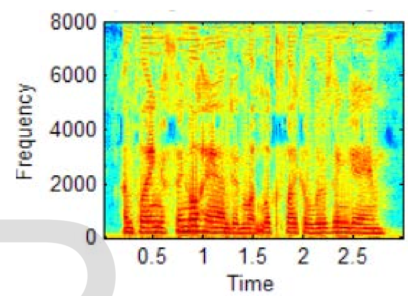
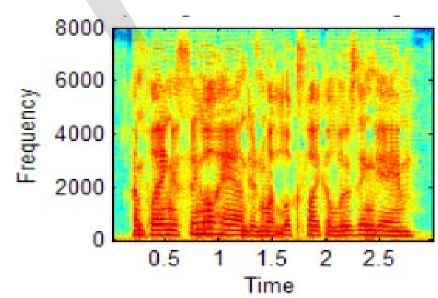


Fig 4 . Comparison of Bandwidth extension using TDI and ABWE

<table>
<tr><th>Method</th><th>Spectrogram</th></tr>
<tr><td>Original Wideband Signal</td><td>
</td></tr>
<tr><td>Autocorrelation Method (lpc)</td><td>

Log Spectral Distance : 1.441486</td></tr>
<tr><td>Levinson-Durbin Recursion Method</td><td>

Log Spectral Distance : 1.270368</td></tr>
</table>

Fig 5 . Spectrogram Comparison of Arctic a0010.wav

4.3 Observations :

ISSN 2229-5518

- Based on our observations, in Table 3 , the High value of MOS indicaates the ABWE using analysis and synthesis method is superior than TDI
- In Fig 5 the less LSD indicates the LD method is superior than Auto correlation method for calculating the LSP parameter and this less LSD provides better speech quality.

5. CONCLUSION

In this paper, we designed the Source-filter model based on vocal tract and implemented ABWE algorithm using LPC coefficients. From the Spectrogram comparison, our proposed algorithm can able to represent consonant sounds, mostly fricative consonants (th /, sh/, / Isl /, xl/ ch, /, etc). Artificial Band Extension (ABE) techniques are used to generate a wideband signal from the narrowband signal. The spectrogram of extended signals shows the obvious creation of missing bands. The performance MOS preference test proved that from the NB signal the ABE system created almost the original WB signal. In most of the practical system, will be available at the transmitting end and before transmission, it will be converted in to NB signal to minimize transmission cost (narrowband channel is only available). In such cases, the wideband LPC coefficients can be directly transmitted along with the NB signal. So that, at the receiving end, instead of maintaining a codebook, the bandwidth extension can be done by directly using the WB LPC coefficients. So to model such fast and efficient system without codebook , we may use a custom audio encoding and decoding technique to incorporate the transfer of WB LPC code at both transmitting and receiving end.

6. REFERENCES

[1] Carl,H., and Heute,U., 1994 “Bandwidth Enhancement of narrow band speech signals”, Signal Processing VII, Theories and applications, EUSIPCO, Vol 2, pp. 1178-1181.

[2] CCITT, 1988 ,“7 kHz Audio Coding Within 64 kBit/s” Recommendation G.722, Vol. Fascile III.4 of Blue Book, Melbourne.

[3] Cheng,Y.M., ’Shaughnessy,D.O, Mermelstein,P., 1994 “Statistical Recovery of Wideband Speech from Narrowband Speech”. IEEE Transactions on Speech and Audio Processing, vol.2, no4,pp.544– 548.

[4] Chennoukh,S., Gerrits, A., Miet, G.,and Sluitjer, R., May 2001 Speech enhancement via frequency bandwidth extension using line spectral frequencies," Pro. IEEE Int. Conf. On Acoustics, Speech, Signal Processing, vol. 1, pp. 665- 668.

[5] 3GPP TS 26.171, March 2001 “Speech Codec speech processing functions; AMR WB Speech Codec; General Description”. Version 1.

[6] Gandhimathi,G., Narmadh,C., and Lakshmi,C. August 2010 ,” Simulation of Narrow Band Speech Signal using BPN Networks, International Journal of Computer Applications (0975 - 8887) Volume 5- No.9.

[7] Gandhimathi.G., Jayakumar,S., 2013” Speech enhancement Using Artificial Bandwidth Extension Algorithm in Multicast conferencing through Cloud services “, Information Technology Journal , ISSN 1812- 5638 pp.1-8.

[8] Heide, D.A. , Kang, G.S., May 1998, "Speech enhancement for band limited speech", in Proc of the ICASSP, Vol. 1, Seattle, WA, USA, pp. 393-396.

[9] http://www.festvox.org/cmu_arctic/ , 2003 Carnegie Mellon University, Copyright (c)

[10] Jax P., and Vary, P., 2000 “ Wide band Extension of speech using Hidden Markov Model” in Proc. IEEE workshop on speech coding .

[11] Jax P., and Vary, P., May 2002. "An upper bound on the quality of artificial bandwidth extension of narrowband speech signals", Pro. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) vol. 1, pp 237- 240, Orlando, FL, USA.

[12] Jax P., and Vary, P., May 2004., "Feature Selection for Improved Bandwidth Extension of Speech Signals", Pro.IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) vol. 1, pp. 697-700, Montreal, Canada.