DEVELOPMENT OF SPEECH BANDWIDTH EXTENSION TECHNIQUE USING DISCRETE COSINE TRANSFORM

Dr. M. Ranga Rao[1], N.Parvathi[2] , A.Lavanya[3], P.B.N.S.Krishna Vamsi[4], A.Subrahmanyam[5]
[1] Professor, [2,3,4,5] Students
Sri Vasavi Institute of Engineering & Technology, Nandamuru

ABSTRACT

Public telephone systems transmit speech across a limited frequency range, about 300–3400 Hz, called narrowband (NB) which results in a significant reduction of quality and intelligibility of speech. This project proposes a new transform-domain speech bandwidth extension algorithm to transmit the information about the missing speech frequencies over a hidden channel, i.e., the related spectral envelope and gain parameters are hidden within the narrowband speech signal using discrete cosine transform (DCT)-based data hiding technique. The hidden information is recovered reliably at the receiver to produce a wideband signal of much higher quality. Obtained results will confirm that the excellent reconstructed wideband signal quality of the proposed algorithm.

1 INTRODUCTION

The human speech contains frequencies beyond the bandwidth of the existing telephone network i.e. (0-4)kHz. Apparently, the transmission of speech through the telephone network results in a loss of portions of speech spectrum causing a significant reduction in speech quality and intelligibility. The transmission of wideband (WB) speech which lies in the range of (0-8)kHz. And this WB of telephone networks would increase the quality of speech signal compared to narrow band speech. But this requires building of new telephone networks that support larger bandwidths which are expensive and will likely take time to be established. It is, therefore, desirable to enhance the bandwidth at the receiver using speech bandwidth extension (BWE) using DCT based data hiding techniques without modifying infrastructure of the existing telephone networks.

The input speech signal is analyzed and its high frequencies are embedded in the low frequencies. Then the low frequencies are transferred to the receiver and at the receiver end the wideband speech is reconstructed using a concise representation of the high frequencies and the low band signal. The concise representation of the high frequencies is based on a frequency selective spectral linear prediction technique.

But this requires building of new telephone networks that support larger bandwidths which are expensive and will likely take time to be established. It is, therefore, desirable to enhance the bandwidth at the receiver using speech bandwidth extension (BWE) using DCT based data hiding techniques without modifying infrastructure of the existing telephone networks.

1.1 Speech Processing

Speech processing is the study of speech signals and the processing methods ofsignals. The signals are usually processed in digital representation, so speech processing can be regarded as a special case of digital signal processing applied to speech signals. Aspects of speech processing include acquisition, manipulation, storage, transfer and output of speech signals. Fig1.1 shows speech signal.

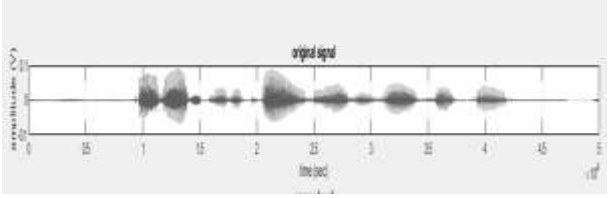
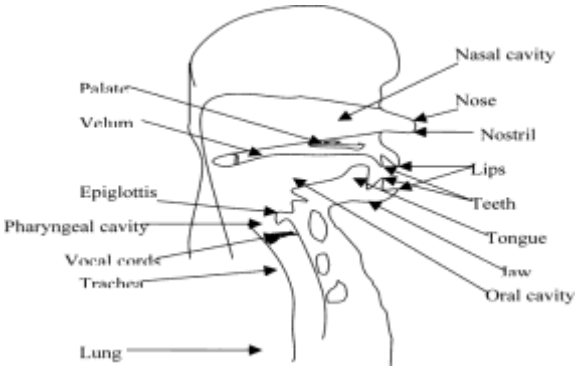


Fig 1.1 SPEECH SIGNAL

1.2 Production of speech signal

Speech processing is the study of speech signals and the processing methods of signals. The signals are usually processed in digital representation, so speech processing can be regarded as a special case of digital signal processing applied to speech signals. Aspects of speech processing include acquisition, manipulation, storage, transfer and output of speech signals. Fig1.1 shows speech signal.



The vocal cord is situated in larynx called the adams apple as shown in fig 1.2. The vocal cord is the source for speech production in humans. It generates two kinds of speech sounds. They are voiced and unvoiced. The vibration of vocal cords produces the sound called the voiced and the unvoiced sound due to turbulence of flow of air at a construction at all possible sites in the vocal tract. The frequency of vibration of the cord is determined by several factors like the tension exerted by the muscle, it's mass and it's length. These factors vary according to sexes and age. The vocal tract is divided into two parts. First one is called the oral cavity which is highly mobile and consists of the tongue, pharynx, plate, lips, and jaw etc., as shown in fig 1.2. The position of these organs is varied to produce different speech sounds, which we hear as the radiation from the lips or nostrils. The second one is the nasal tract where is immobile but is coupled with oral tract by changing the position of the velum. The shape of the vocal tract responds better for some basic frequency produced by vocal cord than others. This is the essential mechanism for the production of different speech sounds. The lowest resonance frequency for a 3 particular shape of the vocal tract is called the first formant (f1) and next the second formant frequency (f2) and so on.

1.3.1 VOICED SPEECH

If the input excitation is nearly periodic impulse sequence, then the corresponding speech looks visually nearly periodic and is termed as voiced speech. The speech production process for the voiced speech can be pictorially represented as shown in fig 1.3.1. During the production of voiced speech, the air exhaling out of lungs through the trachea is interrupted periodically by the vibrating vocal folds. Due to this, the glottal wave is generated that excites the speech production system resulting in the voiced speech. A typical glottal waveform and the corresponding voiced speech are shown in Figure 1.3.1. Thus grossly, when we look at the speech signal waveform, if it looks nearly periodic in nature, then it can be marked as voiced speech.

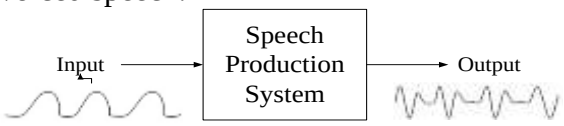


Fig 1.3.1: BLOCK DIAGRAM REPRESENTATION OF VOICED SPEECH PRODUCTION

1.3.2 UNVOICED SPEECH

If the excitation is random noise-like, then the resulting speech will also be random noise-like without any periodic nature and is termed as unvoiced speech. The speech production process for the unvoiced speech can be pictorially represented as in fig 1.3.2.

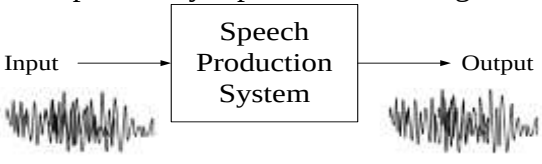


Fig 1.3.2: BLOCK DIAGRAM REPRESENTATION OF UNVOICED SPEECH PRODUCTION

During the production of unvoiced speech, the air exhaling out of lungs through the trachea is not interrupted by the vibrating vocal folds. However, starting from glottis, somewhere along the length of vocal tract, total or partial closure occurs which results in

obstructing air flow completely or narrowly. This modification of airflow results in stop or frication excitation and excites the vocal tract system to produce unvoiced speech. The typical nature of excitation and resulting unvoiced speech are shown in figure 1.3.2 itself. As it can be seen, the unvoiced speech will not have any periodic nature. This will be the main distinction between voiced and unvoiced speech.

1.4 PHONETIC CLASSIFICATION

The phonetic alphabet is usually divided into two categories, vowels and consonants. Vowels are always voiced sounds and they are produced with the vocal cords in vibration, while consonants may be either voiced or unvoiced. Vowels have considerably higher amplitudes than consonants and they are also more stable and easier to analyze and describe acoustically. As consonants involve very rapid changes they are more difficult to synthesize properly.

English consonants may be classified by the manner of articulation as plosives, fricatives, nasals, liquids, and semivowels. Further classification may be made by the place of articulation as labials(lips), dentals (teeth), alveolars (gums), palatals(palate), velars (soft palate) , glottal (glottis), and labiodentals (lips and teeth).

1.4.1 PLOSIVES OR STOP CONSONANTS

The vocal tract is closed causing stop or attenuated sound. When the tract reopens, it noise-like, impulse-like, or burst sound. /k/, /p/, /t/, /g/, /b/, /d/ are some of the examples of plosives. The frequency range of plosives is around from 0 to 8kHz. The vocal tract structure of some of the plosives is shown in fig 1.4.1.

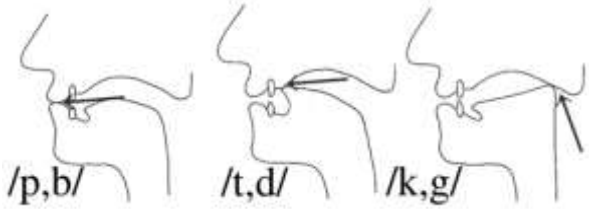


Fig 1.4.1: PLOSIVES

1.4.2 FRICATIVES

The vocal tract is constricted in some place so the turbulent air flow causes noise which is modified by the vocal tract resonances. /f/, /h/, /s/ are some of the examples of fricatives. The frequency range of fricatives is around from 0 to 8 kHz. The vocal tract structure of some of the fricatives are shown in Fig 1.4.2.

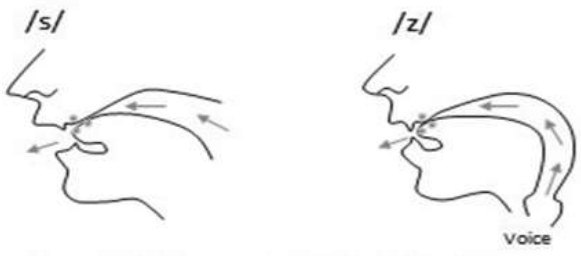


Fig 1.4.2: FRICATIVES

1.4.3 NASALS

The vocal tract is closed but the velum opens a route to the nasal cavity. The generated voiced sound is affected by both vocal and nasal tract. Some of the examples of nasals are /n/, /m/, /ng/. Frequency range of nasal sounds is around 0 to 4KHz.

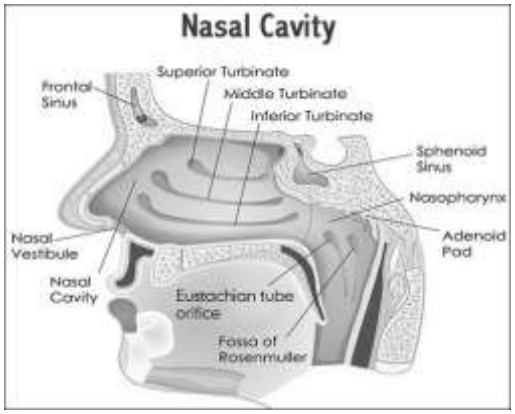


Fig 1.4.3: NASAL CAVITY

1.4.4 LATERALS

The top of the tongue closes the vocal tract leaving a side route for the air flow. The vocal tract structure of sound /l/ is shown in fig 1.4.4.

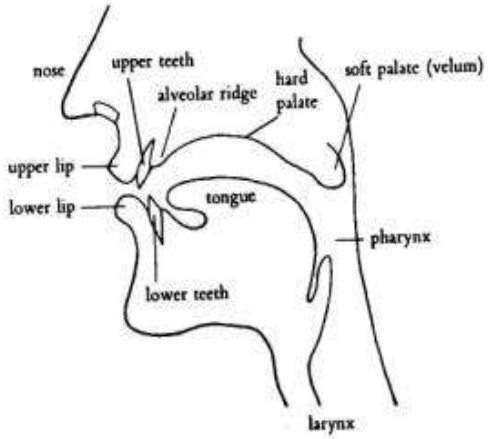


Fig 1.4.4: LATERALS

The transmission of speech through the telephone network results in a loss of portions of speech spectrum i.e. from (4-8) kHz causing a significant reduction in speech quality and intelligibility. Due to limited frequency bandwidth, it is difficult to distinguish plosive and fricative utterances from each other on the phone. Sometimes, it is difficult to distinguish a mother from her daughter and a father from his son on the phone. To overcome the problems associated with limited frequency bandwidth, wideband speech which lies in the range of (0-8) kHz can be transmitted through telephone networks which would increase the quality, intelligibility and perceived naturalness of speech signal compared to narrow band speech. But this requires building of new telephone networks that support larger bandwidths which are expensive and will likely take time to be established. It is, therefore, desirable to enhance the bandwidth at the receiver using speech bandwidth extension (BWE) techniques without modifying infrastructure of the existing telephone networks.

1.5 METHODS OF SPEECH BANDWIDTH EXTENSION

There are two methods for speech bandwidth extension. They are:

1. Artificial Speech Bandwidth Extension
2. Speech Bandwidth extension using data hiding

1.5.1 ARTIFICIAL SPEECH BAND WIDTH EXTENSION

In artificial speech bandwidth extension (ABE) method, WB signal is reconstructed by estimating the missing speech frequencies from the NB signal alone. ABE techniques are motivated from the fact that mutual dependence exists between the NB and the high frequencies of speech usually ranging from 4kHz to 8kHz which is illuminated from the speech production model. The dependency between the frequency bands justifies the estimation of the (4-8) kHz extended band from the narrow band. According to the ABE techniques based on all source-filter model, bandwidth extension is a matter of estimating an excitation signal and a vocal tract filter that modifies the spectral envelope.

1.5.2 SPEECH BANDWIDTH EXTENSION USING DATA HIDING

The quality of telephone speech is improved by hiding the perceptually important components outside the PSTN bandwidth into the telephone bandwidth. Linear prediction parameters are extracted from the down sampled version of the high frequency components of speech signal existing above NB and are embedded in the NB speech signal. The embedded information is extracted at the receiving end to reconstruct the wideband speech signal.

1.6 MOTIVATION

The transmission of human speech through telephone networks results in a loss of portions of speech signal spectrum causing a significant reduction in speech quality. So, our motivation is to improve the quality of the reconstructed wideband signal at the receiver side.

1.7 OBJECTIVE

The objective of our project is to develop speech bandwidth extension technique using transform-domain data hiding which will improve the quality of the reconstructed WB signal.

2 LITERATURE SURVEY

2.1 Telephony Speech Enhancement by Data Hiding

In, the proposed method using data hiding technique is used to extend the public switched telephone network (PSTN) channel bandwidth. Based on the perceptual masking principle, the inaudible spectrum components within the telephone bandwidth can be removed without degrading the speech quality, providing a hidden channel to transmit extra information. The audible components outside the PSTN bandwidth, which are spread out by using orthogonal pseudo-noise codes, are embedded into this hidden channel and then transmitted through the PSTN channel bandwidth. While this hidden signal is not audible to the human ear, it can be extracted at the receiver end. It results in a final speech signal with a wider bandwidth than the normal PSTN channel. Using both theoretical and simulation analysis, it is shown that the proposed approach is robust to quantization errors and channel noises with better quality.

2.3 Speech Bandwidth Extension Based on In-Band Transmission of Higher Frequencies

In, a method for wideband speech transmission is proposed which is fully backwards compatible with narrowband telephone systems. Here, a pitch-scaled version of higher speech frequencies (4 – 6.4) kHz is inserted into the previously "unused" 3.4 – 4 kHz frequency range of standard telephone speech. This operation is reverted at the decoder side. A consistently good wideband speech quality can be achieved, even after transmission over common codecs. The proposed system facilitates fully backwards compatible transmission of higher speech frequencies over various speech codecs.

2.4 Simulation and Overall Comparative Evaluation of Performance Between Different Techniques for High Band Feature Extraction Based on Artificial Bandwidth Extension of Speech over proposed Global System for Mobile Full Rate Narrow Band Coder.

In, computation of high band (HB) feature extraction based on LPC (Linear Prediction Coding) and MFCC (Mel Frequency Cepstral Co-efficient) techniques has been addressed. Further, HB features are embedded into encoded bit stream of global system for mobile (GSM) full rate (FR) 06.10 coder using joint source coding and data hiding before being transmitted to receiving terminal. At receiver, HB features are extracted to reproduce HB portion of speech and their results evaluated in terms of quality.

2.5 Speech Bandwidth Extension Aided by Magnitude Spectrum Data Hiding

In, a NB speech BWE using magnitude spectrum data hiding technique has been proposed. The down-sampled frequency-shifted version of missing speech signal is analysed using a Code excited linear prediction (CELP) coder, and the extracted CELP parameters are embedded in the low – amplitude high frequency part of the magnitude spectrum of NB speech without alerting the low frequency part. Spread spectrum (SS) technique is employed in this paper for successful extraction of the embedded information. The proposed method is demonstrated to produce a much better quality speech signal than the conventional speech bandwidth extension techniques. Hence it is suitable for extending the bandwidth of the existing telephone networks without making changes to the existing telephone networks.

3 PAPER DESCRIPTION

3.1 SPEECH BANDWIDTH EXTENSION USING TRANSFORM DOMAIN DATA HIDING

This chapter gives the description of transmitter and receiver of speech bandwidth extension using transform domain data hiding. A narrowband speech bandwidth extension (BWE) method using Discrete Domain Transform (DCT) based data hiding method is proposed. Initially, a wide band speech (0-8)

kHz, with sampling rate of 16khz is band-splitted using low pass filter (LPF) and a high pass filter (HPF) respectively. The LPF output (0-4) kHz is decimated to provide the narrow band signal. The HPF output (4-8) kHz is decimated to provide the extended band signal. The extended band signal parameters are embedded into the narrow band signal that can be transmitted through the telephone network channel. The embedded information is extracted at the receiving end to reconstruct the wideband speech signal of high quality.

3.2 TRANSMITTER

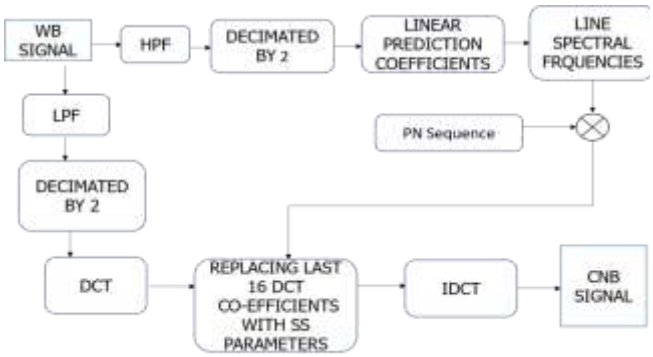


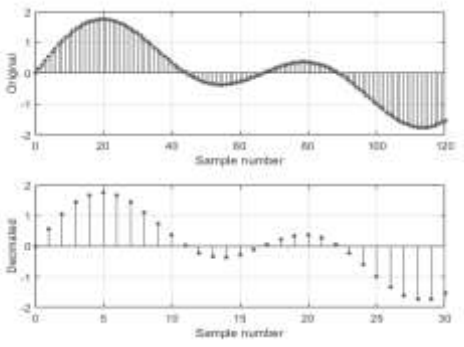
Fig: 4.2 BLOCK DIAGRAM OF TRANSMITTER

3.2.1 DESCRIPTION

The speech bandwidth extension transmitter is shown in fig 4.2. In that initially, a wide band speech (0-8) kH z, with sampling rate of 16khz is band- splitted using low pass filter(LPF) and a high pass filter(HPF) respectively. The LPF output (0-4) kHz is decimated to provide the narrow band signal. The HPF output (4-8) kHz is decimated to provide the extended band signal. Both the narrow band and extended band are partitioned into frames of duration 20msec. DCT is then applied on each frame of narrow band signal. Linear predictive coding is then performed on each frame of extended band signal to generate 12 linear predictive coefficients. These LPCs are converted into Line Spectral Frequencies (LSFs) and Spread spectrum technique is applied on LSFs values, which are then embedded in the last 16 DCT coefficients. Then inverse DCT is applied on the resultant signal, which is termed as composite narrow band signal and is transmitted over the channel.

3.2.2 DECIMATION

In digital signal processing, decimation is the process of reducing the sampling rate of a signal that is already sampled, by an integer factor M. The decimation factor is usually an integer. This reduces the data rate or size of data. In practice, this usually implies low pass filtering a signal, then throwing away some of its samples. In decimation, the sampling rate is reduced from Fs to Fs/M by discarding M – 1 samples for every M samples in the original sequence.



3.2.3 DISCRETE COSINE TRANSFORM

A discrete cosine transform (DCT) converts a signal from its original domain (often time) to a representation in the frequency domain and vice versa. DCT expresses a finite sequence of data points in

terms of a sum of cosine functions oscillating at different frequencies. In particular, a DCT is a Fourier-related transform similar to the discrete Fourier transform (DFT), but using only real numbers. The DCTs are generally related to Fourier Series coefficients of a periodically and symmetrically extended sequence whereas DFTs are related to Fourier Series coefficients of a periodically extended sequence.

3.2.4 LINEAR PREDICTIVE ANALYSIS:

Speech signal is produced as the convolution of excitation source and time varying vocal tract system components. To study these components separately some homomorphic analysis like cepstral analysis are developed to separate the excitation source and vocal tract system components. This deconvolution in frequency domain is a computational intensive process. To reduce such complexity and finding the components separately from time domain itself, LPC analysis is developed.

LPC stands for linear predictive coding, it is one of the most powerful signal analysis technique and one of the most useful methods for encoding good quality speech at a low bit rate and provides highly accurate estimates of speech parameters. This method is used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form. LPC method is used to compress the spectral information for its efficient storage and transmission. The basic idea behind linear predictive analysis is that specific speech sample at the current time can be approximated as linear combination of past speech samples.

The prediction of current sample as a linear combination of past p samples form the basis of linear prediction analysis where p is the order of prediction. The predicted sample $\hat{s}(n)$ can be represented as follows,

$$\hat{s}(n) = \sum_{k=1}^{p} a_k . s(n-k) \qquad (4.6)$$

where $a_k s$ are the linear prediction coefficients and s(n) is the windowed speech sequence obtained by multiplying short time speech frame with a hamming or similar type of window which is given by,

$$s(n) = x(n).w(n) \qquad (4.7)$$

where ω(n) is the windowing sequence. The prediction error e(n) can be computed by the difference between actual sample s(n)and the predicted sample $\hat{s}(n)$ which is given by,

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= s(n) \\ &\quad - \sum_{k=1}^{p} a_k . s(n-k) \end{aligned} \qquad (4.8)$$

$$e(n) = s(n) - \hat{s}(n) \qquad (4.9)$$

$$e(n) = s(n) - \sum_{k=1}^{p} a_k . s(n-k) \qquad (4.10)$$

The primary objective of LP analysis is to compute the LP coefficients which minimized the prediction error e(n). The popular method for computing the LP coefficients by least squares auto correlation method. This is achieved by minimizing the total prediction error. The total prediction error can be represented as follows,

$$E = \sum_{n=-\infty}^{\infty} e^2(n) \qquad (4.11)$$

This can be expanded using the equation (4.10) as follows,

$$E = \sum_{n=-\infty}^{\infty} \left[s(n) - \sum_{k=1}^{p} a_k . s(n-k) \right]^2 \qquad (4.12)$$

The values of aks which minimize the total prediction error E can be computed by finding $\frac{\partial E}{\partial a_k}$ and equating to zero for k=0,1,2,...p.

$$\frac{\partial E}{\partial a_k} = 0 \qquad (4.13)$$

for each a_k give p linear equations with p unknowns, the solution of which gives the LP coefficients. This can be represented as follows,

$$\frac{\partial E}{\partial a_k} = \frac{\partial}{\partial a_k} \cdot \sum_{n=-\infty}^{\infty} \left[s(n) - \sum_{k=1}^{p} a_k . s(n-k) \right]^2 = 0 \quad (4.14)$$

The differentiated expression can be written as,

$$\sum_{n=-\infty}^{\infty} s(n-i).s(n) = \sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} s(n-i).s(n-k) \quad (4.15)$$

where i=1, 2, 3...p. The equation (4.15) can be written in terms of autocorrelation sequence R(i) as follows,

$$\sum_{k=1}^{p} a_k . R(i-k) = R(i) \quad (4.16)$$

for i=1,2,3...p.

Where the autocorrelation sequence used in equation (4.16) can be written as follows,

$$R(i) = \sum_{n=i}^{N-1} s(n)s(n-i) \quad (4.17)$$

for i= 1,2,3...p and N is the length of the sequence.

This can be represented in the matrix form as follows,

$$R.A = -r \quad (4.18)$$

where R is the pXp symmetric matrix of elements R(i, k) = R(|i-k|), (1<=i, k<=p), r is a column vector with elements (R(1),R(2), ...R(P)) and finally A is the column vector of LPC coefficients (a(1), a(2), ...a(p)). It can be shown that R is toeplitz matrix which can be represented as,

$$R = \begin{bmatrix} R(1), R(2), R(3), \ldots \ldots R(P) \\ R(2), R(1), R(2) \ldots \ldots R(P-1) \\ R(3), R(2), R(1) \ldots \ldots R(P-2) \\ \vdots \\ R(P), R(P-1), R(P-2) \ldots . R(1) \end{bmatrix} \quad (4.19)$$

The LP coefficients can be computed as shown,

$$A = -R^{-1}.r \quad (4.20)$$

where R^{-1} is the inverse of the matrix R.

3.2.4.1 LEVINSON-DURBIN ALGORITHM

The solution for finding the optimal prediction coefficients a= $[a1, a2 \ldots, ap]^T$ of a predictor of order P is given by

$$a = R_{ss}^{-1} r_{ss} \quad (4.21)$$

where, $r_{ss} = \left[r_{ss}^{(1)}, r_{ss}^{(2)}, \ldots \ldots r_{ss}^{(p)} \right]^T$ (4.22)

and the matrix given by

$$R_{ss} = \begin{bmatrix} r_{ss}^0 & r_{ss}^1 \ldots \ldots & r_{ss}^{p-1} \\ r_{ss}^1 & r_{ss}^0 \ldots \ldots & r_{ss}^{p-2} \\ r_{ss}^{p-1} & r_{ss}^{p-2} \ldots \ldots & r_{ss}^{01} \end{bmatrix} \quad (4.23)$$

Involving elements given by the eliminated short-term autocorrelation function

$$R_{ss}(m) = \sum_{k=0}^{N-1-m} s(k)s(k+m) \quad (4.24)$$

of a speech segment[s(0),s(1),..,s(N-1)] of length N. Computing the solution of (4.21) by inverting the matrix Rss without exploiting the special properties of this matrix results in P3 operations. In addition the limited precision of the elements of Rss and the limited accuracy of the processing unit may cause numerical problems when inverting the matrix. Therefore methods have been developed that take advantage of the special properties of this matrix. The method presented here is called Levinson-Durbin recursion. The Levinson-Durbin recursion is a recursive-in-model-order solution for solving (4.21).This means that the coefficients of an order P predictor are calculated out of the solution for an order P-1 predictor. By exploiting the Toeplitz structure of Rss the Levinson-Durbin recursion reduces the computational complexity to P2 operations.

Let us recall (4.21) and denote the order by adding a superscript in brackets

$$R_{ss}{}^p a^p = r_{ss}{}^p \quad (4.25)$$

We can denote the system of equations defined in (4.25) as

$$r_0 a_1^{(p)} + r_0 a_2^{(p)} + \ldots\ldots\ldots$$
$$+ r_{p-2} a_{p-1}^{(p)} = r_1$$
$$r_{p-2} a_1^{(p)} + r_{p-3} a_2^{(p)} + \ldots\ldots\ldots \quad (4.26)$$

$+r_1 a_p^{(p)} = r_{p-1}$
$r_{p-1}a_1^{(p)} + r_{p-2}a_2^{(p)}+\ldots\ldots..$
$+r_0 a_p^{(p)} = r_p$

By subtracting the last expressions the system of equations can now be split into a system of equations containing the first p-1 rows

$r_0 a_1^{(p)} + r_0 a_2^{(p)}+\ldots\ldots.. \quad r_{p-2}a_{p-1}^{(p)}=$
$r_{1-}r_{p-1}a_p^{(p)}$
$r_{p-2}a_1^{(p)} + r_{p-3}a_2^{(p)}+\ldots\ldots..$
$r_1 a_p^{(p)}= r_{p-1-}r_1 a_p^{(p)}$ (4.27)
$r_{p-1}a_1^{(p)} + r_{p-2}a_2^{(p)}+\ldots\ldots..$
$r_0 a_p^{(p)}= r_{p-}r_0 a_p^{(p)}$

and an equation containing the last row

$r_{p-1}a_1^{(p)} + r_{p-2}a_2^{(p)}+\ldots\ldots..$
$r_1 a_{p-1}^{(p)}= r_{p-}r_0 a_p^{(p)}$ (4.28)

Now the system of equations in 4.27 contains on the left side the matrixR^{p-1}. This matrix is obtained skipping the last row and the last column of R^p . By defining

$$r^{(P)} = \left[r_0^{(P)} r_0^{(p-1\ldots\ldots r_{0)(1)}}\right]^T \quad (4.29)$$

The system of equations 4.27 can be written as

$$R^{(p-1)}\begin{bmatrix} a_1^{(p)} \\ a_2^{(P)} \\ \vdots \\ a_{(P-1)}^{(P)} \end{bmatrix} = r^{(p-1)} - a_p^{(P)}\bar{r}^{(p-1)} \quad (4.30)$$

In the same manner we can rewrite (4.28)

$$\left(\tilde{r}^{(p-1)}\right)^T R^{(p-1)}\begin{bmatrix} a_1^{(P)} \\ a_2^{(P)} \\ \vdots \\ a_{p-1}^{(P)} \end{bmatrix} = r_p - a_p^{(P)} r_0 \quad (4.31)$$

Now we multiply (4.30) with $R^{(P-1)\wedge -1}$, which is the inverse of R^{p-1}

$$\begin{bmatrix} a_1^{(p)} \\ a_2^{(P)} \\ \vdots \\ a_{p-1}^{(p)} \end{bmatrix} = R^{(P-1)^{-1}} r^{(P-1)} - a_p^{(p)} R^{(p-1)^{-1}} \tilde{r}^{p-1} \quad (4.32)$$

If we now define in parallel to $\bar{r}^p$,

$$\bar{a}^p = \left[a_p^{(P)}, a_{p-1}^{(p)}, , , , a_1^{(P)}\right]^r \quad (4.33)$$

Exploiting the symmetry of (4.21)

$$R^p \tilde{a}^{(p)} = \tilde{r}^p \quad (4.34)$$

Now we can write (4.32) as

$$\begin{bmatrix} a_1^{(p)} \\ a_2^{(p)} \\ a_3^{(p)} \\ \vdots \\ a_{p-1}^{(p)} \end{bmatrix} = a^{(p-1)-}a_p^{(P)} a^{(p-1)} \quad (4.35)$$

This finally is the instruction how to calculate the coefficients $\square_{\square}^{(\square)}$, i=1….P-1, of a predictor of order P out of coefficients of an order P-1 predictor and $\square_{\square}^{(\square)}$. For calculating the single coefficients (4.35) can be written as:

$\square_{\square}^{(\square)} = \square_{\square}^{(\square-1)}\square_{\square}^{(\square)}\square_{\square-1}^{(\square-1)}$, for i = 1 …. P-1 (4.36)

This is known as the Levinson-recursion. The coefficient that is missing up to now $\square_{\square}^{(\square)}$ is called reflection coefficient or PARCOR coefficient (partial correlation coefficient).

Using (4.35) in (4.31) results in

$$\square_{\square}^{(\square)}(\square_0 - \left(\square^{(\square-1)}\right)^{\square}\square^{(\square-1)}) = \square_{\square} - \left(\square^{(\square-1)}\right)^{\square}\square^{(\square-1)} \quad (4.37)$$

Now we can isolate $\square_{\square}^{(\square)}$ is

$$\square_{\square}^{(\square)} = \frac{(\square_{\square} - \left(\square^{(\square-1)}\right)^{\square}\square^{(p-1)}}{(\square_0 - \left(\square^{(\square-1)}\right)^{\square}\square^{(\square-1)}} \quad (4.38)$$

Using (4.36) and (4.38) all coefficients of an order P predictor can be calculated out of an order P-1 predictor with the autocorrelation values of the input process or the estimated autocorrelation values.

3.2.4.2 LINE SPECTRAL FREQUENCIES

Line spectral pairs (LSP) or **line spectral frequencies (LSF)** are used to represent linear prediction coefficients (LPC) for transmission over a channel. LSPs have several properties (e.g. smaller sensitivity to quantization noise) that make them superior to direct quantization of LPCs. The degradation of the signal will be less when we embed the LSFs rather than LPCs.

3.3 RECEIVER

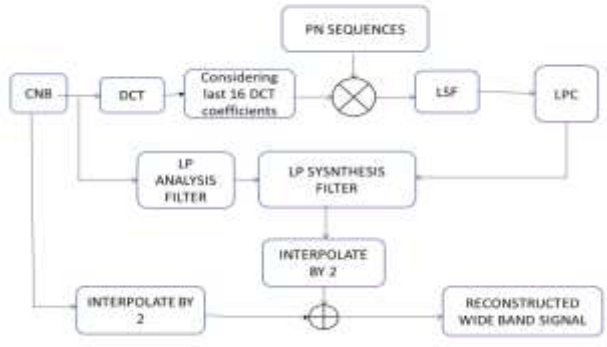


Fig 4.3: BLOCK DIAGRAM OF RECEIVER

3.3.1 DESCRIPTION

The block diagram of receiver is depicted in fig 4.3. At receiver side, firstly the spread spectrum parameters are despreaded using the PN sequences and Hadamard codes and the composite narrowband signal is divided into frames and then DCT is applied on each frame of composite narrow band signal. DCT coefficients are generated from it. 10 LSF coefficients which are embedded in narrowband at the transmitter section are extracted. These are converted to LPC coefficients. To get the excitation of the composite narrowband signal, Linear Predictive Coding (LPC) analysis is applied on the composite narrowband signal. Excitation and extracted extended band LPCs are given as inputs to this synthesis filter from which extended band is reconstructed. Reconstructed signal is then interpolated by factor 2. This signal is then combined with the composite narrowband signal which is also interpolated by factor 2 to get reconstructed wide band signal with high quality.

3.3.2 INTERPOLATION

Interpolation means increasing sampling rate by an integer factor. It increases the original sampling rate for a sequence to a higher rate. It performs interpolation by inserting zeros into the original sequence and then by applying a special low pass filter.

For example, if compact disc audio at 8000 samples/sec is interpolated by a factor of 2, the resulting sampling rate is 16000 samples/sec.

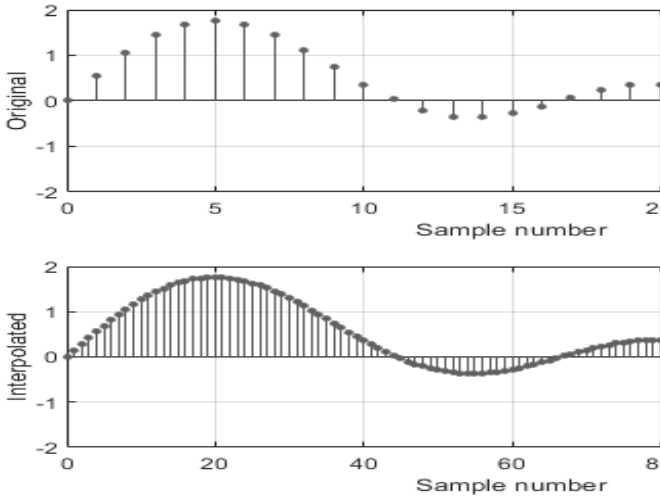


Fig 3.3.3: RECONSTRUCTION OF THE SPEECH SIGNALS USING RESIDUAL BY LP FILTERING

RESULTS

A speech bandwidth extension method using data hiding technique based on discrete cosine transform (DCT) for extending the bandwidth of the existing narrowband in telephone network is proposed here. Public Switched Telephone Networks (PSTN) are bandlimited and transmits speech only in the range of about (0-4) kHz. Due to this the information present in higher frequencies (4-8) kHz components are missed. Hence, the quality and intelligibility of speech is reduced.

To overcome this problem, speech bandwidth extension technique using data hiding based on discrete cosine transform is used. At the transmitter, some parameters corresponding to the extended band signal are embedded into the low-amplitude high-frequency positions of magnitude spectrum of narrow band signal. At the receiver, the out-of-band information is extracted from the narrowband which can be used to combine with the narrowband signal, providing a signal with wider bandwidth. The quality of the reconstructed wideband signal is improved using this method.

Hence it is possible for extending the bandwidth of the telephone networks without making changes to the existing telephone networks. The data hiding

technique has an advantage of using real missing speech band information instead of the estimated missing speech signal as that in the artificial speech bandwidth extension. Thus this technique is more accurate in reconstructing the wideband speech.

4.1 TRANSMITTER WAVEFORMS

The original wideband signal is shown in the fig 4.1.1.

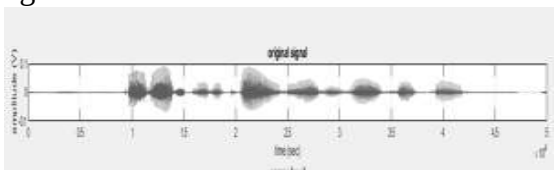


Fig 4.1.1: ORIGINAL SIGNAL

Narrowband signal is shown in the fig 4.1.2.

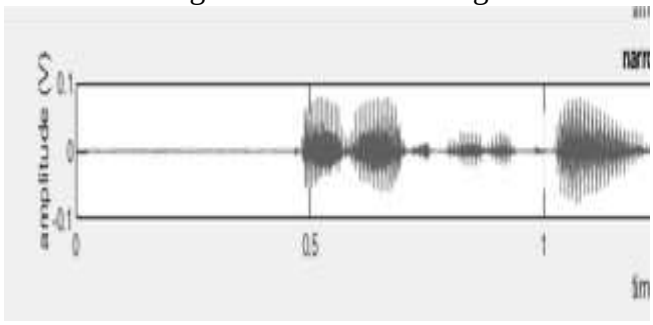


Fig 4.1.2: NARROW BAND SIGNAL

Fig 4.1.3 shows the extended band signal.

Fig 4.1.3: EXTENDED BAND SIGNAL

Linear prediction coefficients of extended band signal are shown in fig 4.1.4.

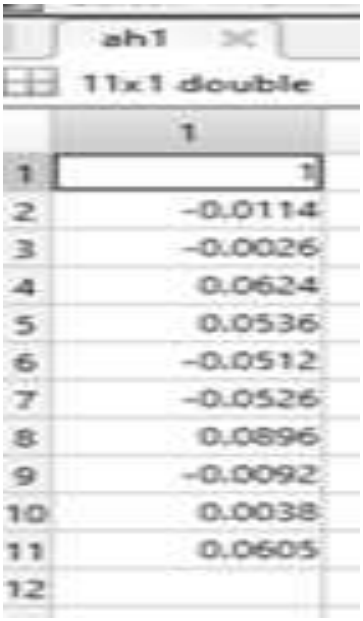


Fig 4.1.4: LINEAR PREDICTION COEFFICIENTS OF EXTENDED BAND SIGNAL

Line spectral frequencies of extended band signal are shown in fig 4.1.5.

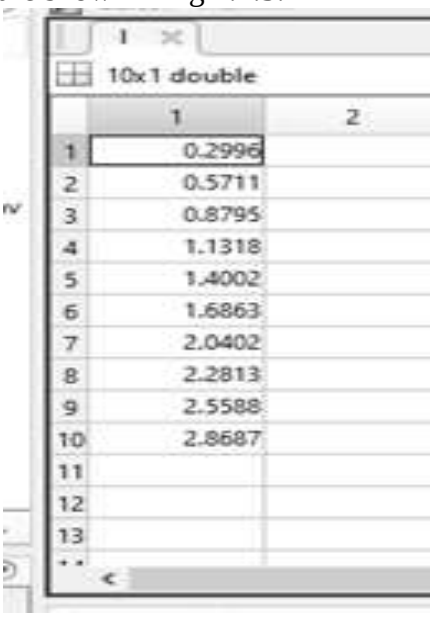


Fig 4.1.5: LINE SPECTRAL FREQUENCIES OF EXTENDED BAND SIGNAL

Composite narrowband signal is shown in the fig 4.1.6.

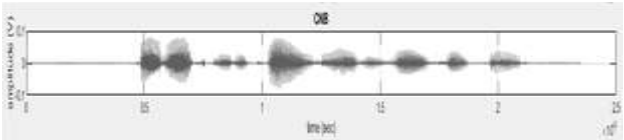


Fig 4.1.6: COMPOSITE NARROW BAND

4.2 RECEIVER WAVEFORMS

Line spectral frequencies of the extended band signal that are extracted from composite narrow band signal are shown in the fig 4.2.1

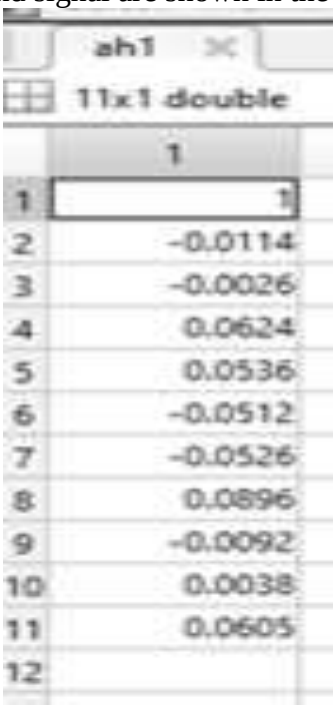


Fig 4.2.1: EXTENDED BAND SPECTRAL FREQUENCIES EXTRACTED FROM COMPOSITE NARROW BAND SIGNAL

Fig 4.2.2 shows linear prediction coefficients of extended band signal that are retrieved from line spectral frequencies.

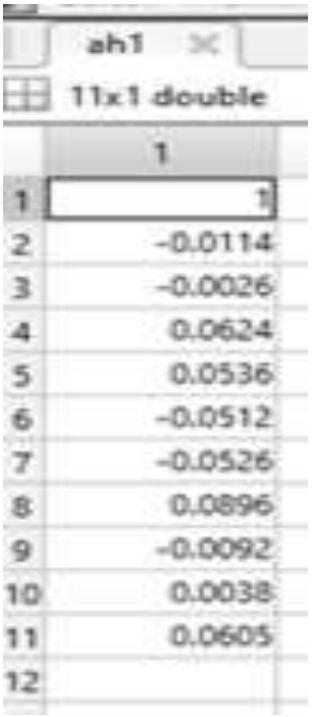


Fig 4.2.2: EXTENDED BAND LINEAR PREDICTION COEFFICIENTS RETRIEVED FROM LINE SPECTRAL FREQUENCIES

Fig 4.2.3 represents the linear prediction coefficients of composite narrow band signal.

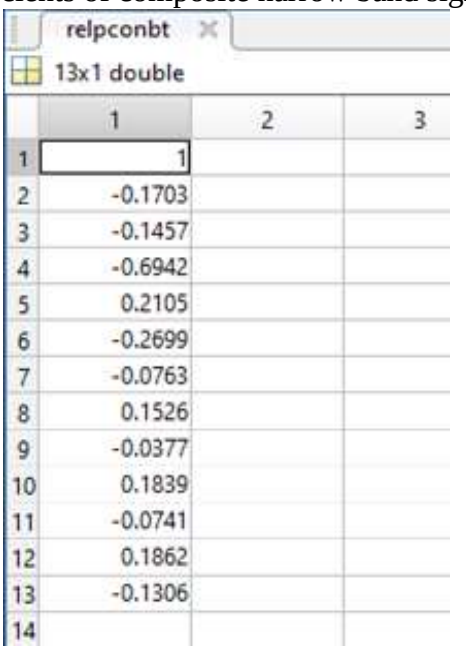


Fig 4.2.3: LINEAR PREDICTION COEFFICIENTS OF COMPOSITE NARROW BAND SIGNAL

Reconstructed wideband signal which is the output of the receiver is depicted in fig 4.2.4.

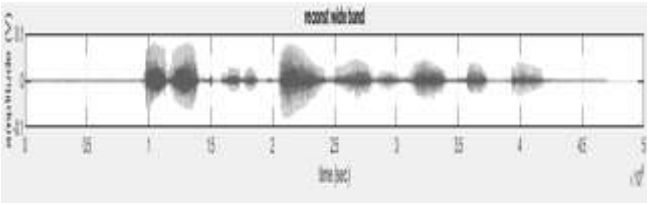


Fig 4.2.4: RECONSTRUCTED WIDE BAND SIGNAL

CONCLUSION

So, in public switching telephone network (PSTN), the speech transmitted is band limited to certain frequency range (0-4) kHz. Due to this, information present in the higher frequencies are missing. So, the quality and intelligibility of speech is reduced. To overcome this problem without disturbing present telephone networks our method uses data hiding technique using discrete cosine transform (DCT) to provide a better-quality wideband speech signal. The missing components are embedded into magnitude spectrum of narrow band. The embedded information is extracted at the receiving end to reconstruct the wideband speech signal thus by increasing the quality of speech.

REFERENCES

[1] Siyue Chen and Henry Leungu, **"*Telephony Speech Enhancement by Data Hiding*",** IEEE Transactions on Instrumentation and Measurement, Vol.56, No.1, February 2007.

[2] Laura Laaksonen, Hannu Pulakka, Ville Myllyla, and Paavo Alku, **"*Development, Evaluation and Implementation of an Artificial Bandwidth Extension Method of Telephone Speech in Mobile Terminal*",** IEEE Transactions on Consumer Electronics, Vol.55, No.2, May 2009.

[3] Bernd Geiser and Peter, **"*Speech bandwidth extension based on in-band transmission of higher frequencies*",** IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013.

[4] Ninad S. Bhatt, **"*Simulation and overall comparative evaluation of performance between different techniques for high band feature extraction based on artificial bandwidth extension of speech over proposed global system for mobile full rate narrow band coder*",** Int J Speech Technol, 2016.

[5] T. Kishore Kumar, **"*Speech bandwidth extension aided by Magnitude Spectrum Data Hiding*",** Circuits, Syst and Signal Processing, Vol.36, No.11, 2017.