

LPC and Wavelet Techniques for Speech Compression

B.Ravi Kiran

Assistant Professor,
Department of ECE,
A.I.E.T, Thagarapavalasa,
Visakhapatnam.

D.Lakshmi Narayana

Associate Professor,
Department of ECE,
A.I.E.T, Thagarapavalasa,
Visakhapatnam.

B.Naveen

Assistant Professor,
Department of ECE,
A.I.E.T, Thagarapavalasa,
Visakhapatnam.

V.Ravi Tejesvi

Assistant Professor,
Department of ECE,
A.I.E.T, Thagarapavalasa,
Visakhapatnam.

Kiran.lionheart@gmail.com

Abstract:

The past decade has witnessed substantial progress towards the application of low-rate speech coders to civilian and military communications as well as computer-related voice applications. Central to this progress has been the development of new speech coders capable of producing high-quality speech at low data rates. Most of these coders incorporate mechanisms to represent the spectral properties of speech, provide for speech waveform matching, and optimize the coder's performance for the human ear. A number of these coders have already been adopted in national and international cellular telephony standards. In mobile communication systems, service providers are continuously met with the challenge of accommodating more users within a limited allocated bandwidth. For this reason, manufactures and service providers are continuously in search of low bit-rate speech coders that deliver toll-quality speech. In this paper the simulated low bit rate vocoder (LPC) using MATLAB was implemented. The result obtained from LPC was compared with other implemented voice compression using wavelet transform. From the results we see that the performance of wavelet transform was better than LPC.

Keywords: LPC.DWT.RELP.CELP.

1.INTRODUCTION :

Speech is a very basic way for humans to convey information to one another. With a bandwidth of only 4 kHz, speech can convey information with the emotion of a human voice. People want to be able to hear someone's voice from anywhere in the world, as if the person was in the same room. As a result a greater emphasis is being placed on the design of new and efficient speech coders for voice communication and transmission; today applications of speech coding and compression have become very numerous. Many applications involve the real time coding of speech signals, for use in mobile satellite communications, cellular telephony, and audio for videophones or

video teleconferencing systems. Other applications include the storage of speech for speech synthesis and playback, or for the transmission of voice at a later time. Some examples include voice mail systems, voice memo wrist-watches, voice logging recorders and interactive PC software. Traditionally speech coders can be classified into two categories: waveform coders and analysis/synthesis vocoders (from .voice coders.). Waveform coders attempt to copy the actual shape of the signal produced by the microphone and its associated analogue circuits. A popular waveform coding technique is pulse code modulation (PCM), which is used in telephony today. Vocoders use an entirely different approach to speech coding, known as parameter coding, or analysis/synthesis coding where no attempt is made at reproducing the exact speech waveform at the receiver, only a signal perceptually equivalent to it.

These systems provide much lower data rates by using a functional model of the human speaking mechanism at the receiver. One of the most popular techniques for analysis/synthesis coding of speech is called Linear Predictive Coding (LPC), which we will be introduce in the first part of our project. Some higher quality vocoders include RELP (Residual Excited Linear Prediction) and CELP (Code Excited Linear Prediction). In the second part of our paper, we look at another technique for analyzing and compressing speech signals using wavelets.

Very simply wavelets are mathematical functions of finite duration with an average value of zero that are useful in representing data or other functions. Any signal can be represented by a set of scaled and translated versions of a basic function called the mother wavelet. This set of wavelet functions forms the wavelet coefficients at different scales and positions and results from taking the wavelet transform of the original signal. The coefficients represent the signal in the wavelet domain and all data operations can be performed using just the corresponding wavelet coefficients.

IILPC SYSTEM IMPLEMENTATION:

Fig-1 shows the steps of linear predictor code (LPC).

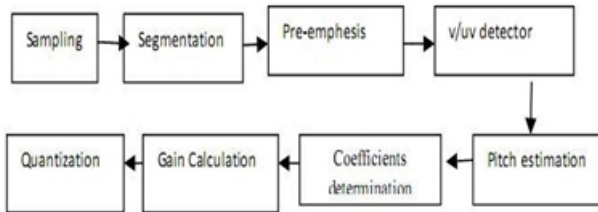


Fig (1) LPC encoder block diagram

1- Sampling:

First, the speech is sampled at a frequency appropriate to capture all of the necessary frequency components important for processing and recognition. According to the Nyquist theorem, the sampling frequency must be at least twice the bandwidth of the continuous-time signal in order to avoid aliasing. For voice transmission, 10 kHz is typically the sampling frequency of choice, though 8 kHz is not unusual. This is because, for almost all speakers, all significant speech energy is contained in those frequencies below 4 kHz (although some women and children violate this assumption).

2- Segmentation:

The speech is then segmented into blocks for processing. Properties of speech signals change with time. To process them effectively it is necessary to work on a frame-by-frame basis, where a frame consists of a certain number of samples. The actual duration of the frame is known as length. Typically, length is selected between 10 and 30 ms or 80 and 240 samples. Within this short interval, properties of the signal remain roughly constant. Simple LPC analysis uses equal length blocks of between 10 and 30ms. Less than 10ms does not encompass a full period of some low frequency voiced sounds for male speakers. For certain frames with male speech sounded synthetic at 10ms sample windows, pitch detection became impossible. More than 30ms violates the basic principle of stationarity.

3- Pre_emphasis:

The typical spectral envelope of the speech signal has a high frequency roll-off due to radiation effects of the sound from the lips.

Hence, high-frequency components have relatively low amplitude, which increases the dynamic range of the speech spectrum. As a result, LP analysis requires high computational precision to capture the features at the high end of the spectrum. One simple solution is to process the speech signal using the filter with system function

$$H(z) = 1 - \alpha z^{-1} \quad (1)$$

This is high pass in nature. The purpose is to augment the energy of the high frequency spectrum. The effect of the filter can also be thought of as a flattening process, where the spectrum is “whitened”. Denoting $x[n]$ as the input to the filter and $y[n]$ as the output, the following difference equation applies:

$$y[n] = x[n] - \alpha x[n - 1] \quad (2)$$

The filter described in (1) is known as the pre-emphasis filter. By pre-emphasizing, the dynamic range of the power spectrum is reduced. This process substantially reduces numerical problems during LP analysis, especially for low-precision devices. A value of α near 0.9 is usually selected. It is common to find in a typical speech coding scheme that the input speech is first pre-emphasized using (1). To keep a similar spectral shape for the synthetic speech, it is filtered by the de-emphasis filter with system function.

$$G(z) = \frac{1}{1 - \alpha z^{-1}} \quad (3)$$

At the decoder side, which is the inverse filter with respect to pre-emphasis.

4- Voicing detector:

The purpose of the voicing detector is to classify a given frame as voiced or unvoiced. In many instances, voiced/unvoiced classification can easily be accomplished by observing the waveform; a frame with clear periodicity is designated as voiced, and a frame with noise-like appearance is labeled as unvoiced. In other instances, however, the boundary between voiced and unvoiced is unclear; this happens for transition frames, where the signal goes from voiced to unvoiced or vice versa. The necessity to perform a strict voiced/unvoiced classification is indeed one of the fundamental limitations of the LPC model. In this section we discuss some measurements that a voicing detector relies on to accomplish its task. For reliable operation, the detector must take into account as many

parameters as possible so as to achieve a high degree of robustness. These parameters are input to a linear classifier having binary output. The voicing detector is one of the most critical components of the LPC coder, since misclassification of voicing states can have disastrous consequences on the quality of the synthetic speech. These parameters are discussed below.

Energy:

This is the most obvious and simple indicator of voicedness. Typically, voiced sounds are several orders of magnitude higher in energy than unvoiced signals. For the frame (of length N) ending at instant m, the energy is given by Serves a similar purpose.

$$Eng[m] = \sum_{n=m-N+1}^m |y[n]|^2 \dots (4)$$

For simplicity, the magnitude sum function defined by

$$MSF[m] = \sum_{n=m-N+1}^m |y[n]| \dots (5)$$

Since voiced speech has energy concentrated in the low-frequency region, due to the relatively low value of the pitch frequency, better discrimination can be obtained by low pass filtering the speech signal prior to energy calculation. That is, only energy of low-frequency components is taken into account. A bandwidth of 800 Hz is adequate for the purpose since the highest pitch frequency is around 500Hz.[4]

a- Zero Crossing Rate

The zero crossing rate of the frame ending at time instant m is defined by

$$ZC[m] = \frac{1}{2} \sum_{n=m-N+1}^m |sgn(y[n]) - sgn(y[n-1])| \dots (6)$$

With $sgn(\cdot)$ the sign function returning ± 1 depending on the sign of the operand. Equation (3.6) computes the zero crossing rates by checking the samples in pairs to determine where the zero crossings occur. Note that a zero crossing is said to occur if successive samples have different signs. For voiced speech, the zero crossing rate is relatively low due to the presence of the pitch frequency component (of low-frequency nature), whereas for unvoiced speech, the zero crossing rate is high due to the noise-like appearance of the signal with a large portion of energy located in the high-frequency region.

b- Pitch period

Since voiced speech concentrated in the low-frequency region, as a consequence, its pitch period has higher values than the unvoiced.

Voicing Detector Design:

A voicing detector can rely on the parameters discussed so far (energy, zero crossing rate, and pitch period) to make the proper decision. A simple detector can be implemented by using just one parameter as input. For instance, the zero crossing rate can be used for voicing detection in the following manner: if the rate is lower than a certain threshold, the frame is declared voiced; otherwise, it is unvoiced. The design problem is therefore to find the proper threshold so that a voicing decision can be accomplished reliably. By analyzing a large amount of speech signals, it is possible to come up with a reasonable value of a decision threshold so as to minimize the total classification error. Relying on just one parameter, however, limits the robustness of the system. For the voicing detector using the zero crossing rates alone, noise contamination can increase the rate in such a way that voiced frames are classified as unvoiced frames. Thus, using more parameters of the frame is necessary to improve the reliability in voicing detection.[4]

5- Pitch period estimation

One of the most important parameters in speech analysis, synthesis, and coding applications is the fundamental frequency, or pitch, of voiced speech. Pitch frequency is directly related to the speaker and sets the unique characteristic of a person. Voicing is generated when the airflow from the lungs is periodically interrupted by movements of the vocal cords. The time between successive vocal cord openings is called the fundamental period, or pitch period. For men, the possible pitch frequency range is usually found somewhere between 50 and 250 Hz, while for women the range usually falls between 120 and 500 Hz. In terms of period, the range for a male is 4 to 20 ms, while for a female it is 2 to 8ms. Pitch period must be estimated at every frame. By comparing a frame with past samples, it is possible to identify the period in which the signal repeats itself, resulting in an estimate of the actual pitch period. Note that the estimation procedure makes sense only for voiced frames. Meaningless results are obtained for unvoiced frames due to their random nature.

Design of a pitch period estimation algorithm is a complex undertaking due to lack of perfect periodicity, interference with formants of the vocal tract, uncertainty of the starting instance of a voiced segment, and other real-world elements such as noise and echo. In practice, pitch period estimation is implemented as a trade-off between computational complexity and performance. Many techniques have been proposed for the estimation of pitch period and only one is included here.[4]

a- The Autocorrelation Method

The pitch period could be estimated by taking the average separation between peaks. The overall peaks and troughs in the spectrum are referred to as the formant structure (where the formants are the frequencies where resonances occur). [5] The autocorrelation of a stationary sequence $x(n)$ is defined as

$$R_x(\tau) = x(n) * x(n + \tau) = \frac{1}{N} \sum_{n=0}^{N-\tau} x(n)x(n + \tau) \dots \dots \dots (7)$$

Where τ is termed the lag. Auto means self or from one signal, and correlation means relation between two samples. An autocorrelation is the average correlation between two samples from one signal that are separated by τ samples. It should be noted that the upper limit in the summation will be less than $N-1$ when τ is positive, and the lower limit will be greater than 0 when τ is negative. Thus, the autocorrelation can be rewritten as

$$R_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-\tau} x(n)x(n + \tau) \dots \dots \dots (8)$$

6-Coefficients determination

The coefficients of the difference equation (the prediction coefficients) characterize the formants, so the LPC system needs to estimate these coefficients. The estimate is done as mentioned above by minimizing the mean-square error between the predicted signal and the actual signal. This is a straight forward problem, in principle. In practice, it involves (1) the computation of a matrix of coefficient values, and (2) the solution of a set of linear equations. An efficient algorithm known as the Levinson-Durbin algorithm is used to estimate the linear prediction coefficients from a given speech waveform.

7- Gain Calculation:

Power of the prediction-error sequence is calculated next, which is different for voiced and unvoiced frames.

Denoting the prediction-error sequence as $e[n], n=0, N-1$, with N being the length of the frame, we have for the unvoiced case

$$p = \frac{1}{N} \sum_{n=0}^{N-1} e^2[n] \dots \dots \dots (9)$$

For the voiced case, power is calculated using an integer number of pitch periods:

$$p = \frac{1}{[N/T]T} \sum_{n=0}^{[N/T]T-1} e^2[n] \dots \dots \dots (10)$$

It is assumed that $N > T$, and hence use of the floor function ensures that the summation is always performed within the frame's boundaries. Gain computation is performed as follows. For the unvoiced case, denoting the gain by g , we have

$$g = \sqrt{p} \dots \dots \dots (11)$$

Since the white noise generator has unit-variance output.[4]

For the voiced case, the power of the impulse train having an amplitude of g and a period of T , measured over an interval of length $[N/T]T$, must equal p . Carrying out the operation yields

$$g = \sqrt{p} \dots \dots \dots (12)$$

8- Quantization

Usually direct quantization of the predictor coefficients is not considered. To ensure stability of the coefficients (the poles must lie within the unit circle in the z -plane) a relatively high accuracy (8-10 bits per coefficients) is required. This comes from the effect that small changes in the predictor coefficients lead to relatively large changes in the pole positions. Quantizing intermediate values is less problematic than quantifying the predictor coefficients directly. These intermediate values are called Line spectral frequency coefficients (LSFs). Line spectral frequency coefficients (LSFs) were first introduced by Itakura (1975) as an alternative representation of LPCs (LSFs are mathematically equivalent (one-to-one) to LPCs). Due to many desirable properties, the LSF has received widespread acceptance in speech coding applications. Line spectral frequency, possesses several desirable features that make it attractive as an alternative LPC representation. The values of the LSFs directly control the property of the signal in the frequency domain, and changes of one parameter have a local effect on the spectrum.

Also, the LSFs are bounded, they are located inside the $(0, \pi)$ interval and ordered $(0 < \omega_1 < \omega_2 < \omega_3 < \dots < \pi)$. LSF are more amenable to quantization. LSFs are more correlated from one frame to the next than LPCs. For frame size of 20 msec. There are 50 frames/sec. 2400 bps is equivalent to 48 bits/frame. These bits are allocated as in Table (1). The 34 bits for the LSFs are allocated as in Table (2).

TABLE I.

Parameter Name	Parameter Notation	Rate (bit/frame)
LSF		34
Gain	G	7
Voiced/Unvoiced Period	& V/UV,T	7
Total		48 bits

TABLE II.

LSF	# of bits
	3
	4
	4
	4
	4
	3
	3
	3
	3
	3
Total	34 Bits

The gain G is encoded using a 7-bit non-uniform scalar quantization.[3]

I.RESULTS:

In This paper the sampling frequency was 8 kHz and the bit rate was 2400 bits per second. A comparison between the original speech sentences and the LPC reconstructed speech has been studied. We have observed that the reconstructed speech has lower quality than the input speech sentences. The reconstructed signals sound mechanized and noisy being nearly unintelligible. The reconstructed speech sounds guttural with a lower pitch than the original Sound. The Sound seems to be whispered. The noisy feeling is very strong. We have presented an approach for classifying the v/uv part of the speech in an effective and simple way. The algorithm showed good results in classifying the speech as we segmented the speech into many frames. In table (3) a specification of each signal is tabulated.

TABLE III.

File name	Male speaker "s4.wav"	Female speaker "original.lau"
sampling rate	8000Hz	8000Hz
Number of bits per sample	16bits	8bits
file length	3 seconds	4.0960 seconds
sample/frame	160 samples	160 samples
length of the original signal	24000 samples	32768
length of the constructed signal	23840 samples	32640
original bit rate	8000*16= 128 kbps	8000*8= 64 kbps
bit rate after compression	2.4 kbps	2.4 kbps
compression ratio	128/2.4 = 53.33	64/2.4 = 26.6
thresh_msf	2.8828	1.7053
thresh_zc	62.3859	93.9044
thresh_pitch	29.0235	27.7059

TABLE IV. TABLE(4) SHOWS THE PERFORMANCE MUASERMENTS (SIGNAL TO NOISE RATIO AND COMPRESSION FACTOR) OF THE TWO SPEECH FILES:

Speech signal	Male	Female
SNR	-0.3975	-0.9825
CF	1.0067	1.0039

Looking at the SNR computed in table (4), it is obvious that both sounds are noisy as they have a negative SNR value. The noise in the synthesized files is more stronger than in the actual signals.

III.SPEECH COMPRESSION USING DISCRETE WAVE TRANSFORM (DWT):

Speech compression using DWT is shown in step below.

1-Choice of Wavelet

The choice of the mother-wavelet function used in designing high quality speech coders is of prime importance. Several different criteria can be used in selecting an optimal wavelet function. The objective is to minimize reconstructed error and maximize signal to noise ratio (SNR). In general optimum wavelets can be selected based on the energy conservation properties in the approximation part of the wavelet coefficients. A suitable criterion for selecting optimum mother wavelets is related to the amount of energy a wavelet basis function can concentrate into the level 1 approximation coefficients.

2-Wavelet Decomposition:

Wavelets work by decomposing a signal into different resolutions or frequency bands, and this task is carried out by choosing the wavelet function and computing the Discrete Wavelet Transform (DWT). Signal compression is based on the concept that selecting a small number of approximation coefficients (at a suitably chosen level) and some of the detail coefficients can accurately represent regular signal components. Choosing a decomposition level for the DWT usually depends on the type of signal being analyzed or some other suitable criterion such as entropy. For the processing of speech signals decomposition up to level 5 is adequate. This fact is derived from the experiments.

3-Truncation of Coefficients:

After calculating the wavelet transform of the speech signal, compression involves truncating wavelet coefficients below a threshold. From the experiments that we conducted, we found that most of the coefficients have small magnitudes. Speaking in general terms, more than 90% of the wavelet coefficients were found to be insignificant, and their truncation to zero made an imperceptible difference to the signal. This means that most of the speech energy is in the high-valued coefficients, which are few. Thus the small valued coefficients can be truncated or zeroed and then be used to reconstruct the signal. Two different approaches are available for calculating thresholds:

IV.GLOBAL THRESHOLD:

The aim of Global Thresholding is to retain the largest absolute value coefficients, regardless of the scale in the wavelet decomposition tree. Global thresholds are calculated by setting the % of coefficients to be truncated.

V.LEVEL DEPENDENT THRESHOLD:

This approach consists of applying visually determined level dependent thresholds to each decomposition level in the Wavelet Transform. The truncation of insignificant coefficients can be optimized when such a level dependent thresholding is used. Thus this approach to thresholding selects the highest absolute valued coefficients at each level.

4-Encoding:

Signal compression is achieved by first truncating small-valued coefficients and then efficiently encoding them. One way of representing the high-magnitude coefficients is to store the coefficients along with their respective positions in the wavelet transform vector. For a speech signal of frame size F, taking the DWT generates a frame of size T, slightly larger than F. If only the largest L coefficients are retained, then the compression ratio C is given by:

$$C = \frac{F}{2L} \tag{13}$$

Another approach to compression is to encode consecutive zero valued coefficients, with two bytes. One byte to indicate a sequence of zeros in the wavelet transforms vector and the second byte representing the number of consecutive zeros.

VIPERFORMANCE MEASURES:

a. Compression factor:

It is the ratio of the original signal to the compressed signal.

$$CF = \frac{\text{length of the original signal}}{\text{length of the compressed signal}} \tag{14}$$

b. Retained signal energy:

This indicates the amount of energy retained in the compressed signal as a percentage of the energy of original signal.

c. Signal to noise ratio (SNR):

This value gives the quality of reconstructed signal. Higher the value, the better. It is given by:

$$SNR = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_e^2} \right) \tag{15}$$

Where σ_x and σ_e are respectively the mean square of the Speech signal and the mean square difference between the original and reconstructed signals.

d- Percentage of zero coefficients:

It is given by the following relation:

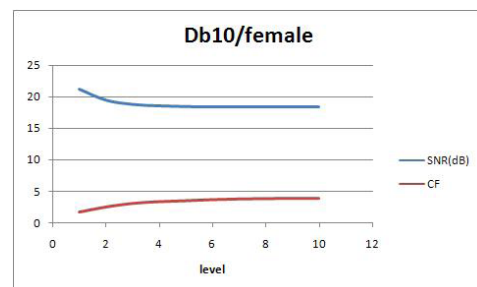
$$\frac{100 \times (\# \text{ of zeros of the current decomposition})}{\# \text{ of coefficients}} \tag{16}$$

- Signal energy in the first level approximation:

This quantity helps in the selection of appropriate mother wavelet for compression. The higher the amount of energy in the first level approximation better is the wavelet for compression of that signal.

VII.CHOOSING THE OPTIMUM LEVEL:

For choosing the optimum level, we have tested a number of mother wavelets at different levels for both speech signals, the results were recorded in the following figures (2):



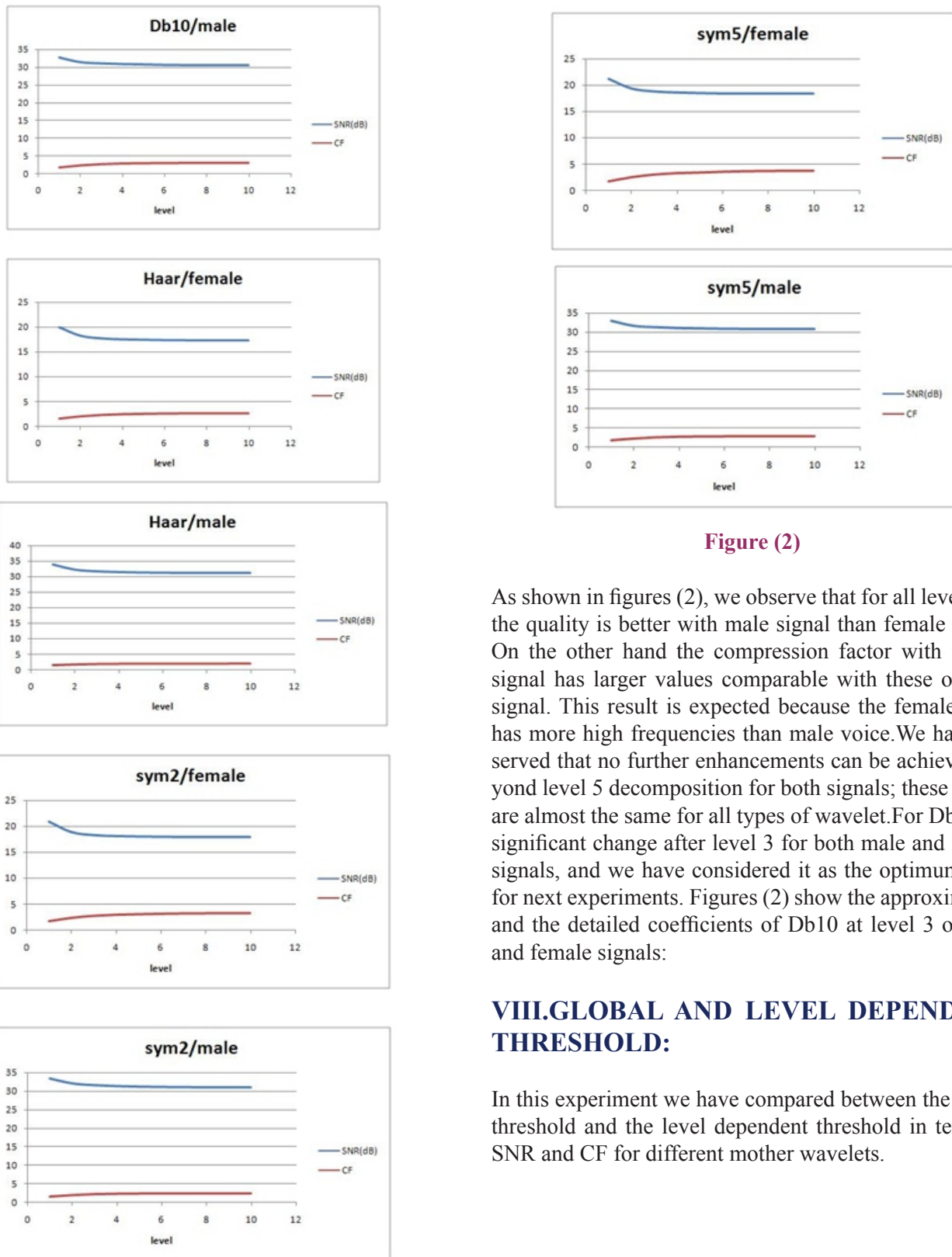


Figure (2)

As shown in figures (2), we observe that for all levels that the quality is better with male signal than female signal; On the other hand the compression factor with female signal has larger values comparable with these of male signal. This result is expected because the female voice has more high frequencies than male voice. We have observed that no further enhancements can be achieved beyond level 5 decomposition for both signals; these results are almost the same for all types of wavelet. For Db10, no significant change after level 3 for both male and female signals, and we have considered it as the optimum level for next experiments. Figures (2) show the approximation and the detailed coefficients of Db10 at level 3 of male and female signals:

VIII.GLOBAL AND LEVEL DEPENDENT THRESHOLD:

In this experiment we have compared between the global threshold and the level dependent threshold in terms of SNR and CF for different mother wavelets.

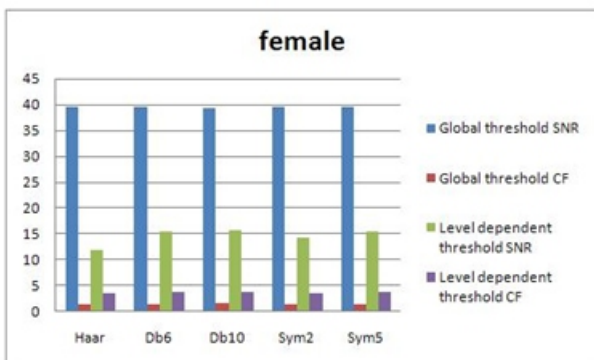
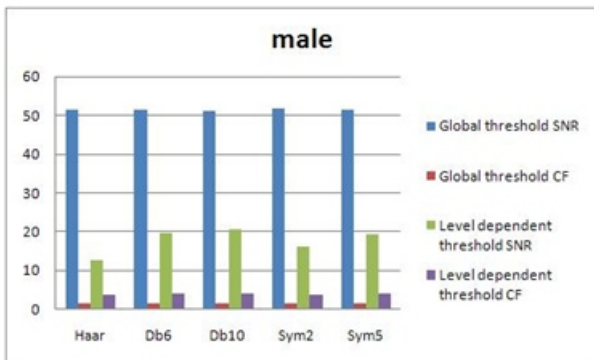


Figure (3)

Results of global and level dependent threshold For both speech signals, the results showed that the global threshold leads to a better quality (SNR) as in fig (3).

IX EFFECTS OF THE THRESHOLD:

In this experiment, we have studied the effects of varying threshold value on the speech signals in terms of SNR and CF. We have used db10 at level 3, the threshold value was slowly increased, and the corresponding values of the SNR and CF were recorded in the following figures:

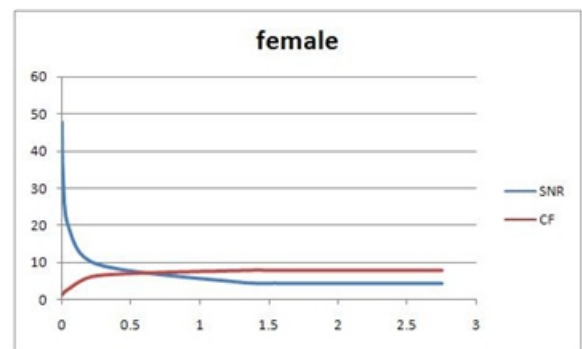
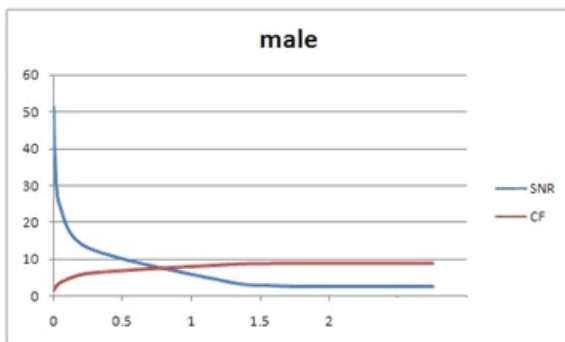


Figure (4)

From fig (4), we observed that after some particular threshold value, the SNR and CF do not change. This is because at this point all the detail coefficients are truncated to zero and only approximation coefficients was remain. Male signals have relatively more approximation coefficients than female voices. It needs a larger threshold value for truncation the details, because they have a relatively large magnitude.

IX-Performance of Recorded Speech Coding Level 3, global Threshold

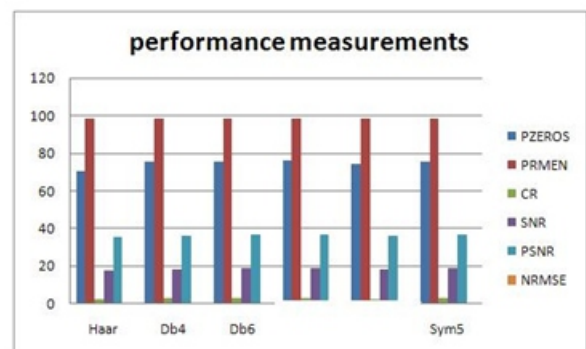
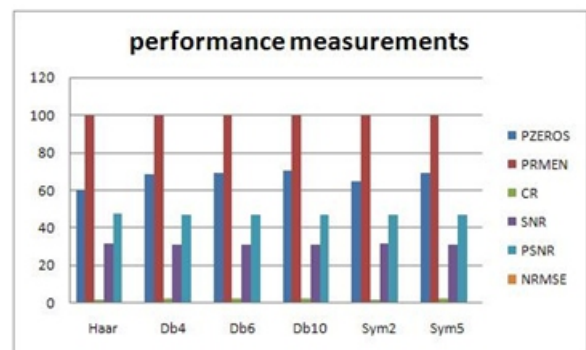


Figure (5)

These values were taken at constant global threshold and decomposition level 3 for all wavelet families.

From fig (5), the performance of db6 and db10 were the best for the female speech signal. For the male speech signal, db6, db10 and sym5 showed almost a good similar performance.

X.CONCLUSION :

The performance of the Discrete Wavelet Transform in compressing speech signals was tested and the following points were observed. High compression ratios were achieved with acceptable SNR. No further enhancements were achieved beyond level 5 decomposition. The NRMSE is remained almost constant for all experiments with negligible changes. Overall global threshold lead to better results than the level dependent threshold technique. After studying the two techniques, the LPC and wavelet we cannot say that one of them is the best. Each having some very good aspects, as well as some drawbacks. And finally using one of them is decided based on our application requirements (high quality, low bit rates....)The tradeoffs between quality on one side and bandwidth and complexity on the other side clearly appear here. If we want a better quality, the complexity of the system should be increased or a larger bandwidth has to be used.

Speech coding based on Linear Predictive Coding (LPC) is a successful and very commonly used method for years and has found many application fields from mobile phones to voice mails. The overly simplistic model that the LPC coder relies on has relatively low computational cost and makes the low bit-rate speech coder a practical reality. However, is also highly inaccurate in various circumstances, creating annoying artifacts in the synthetic speech. LPC10 and other vocoders have historically operated in the telephony bandwidth (0- 4kHz) as this bandwidth is thought to contain all the information necessary to make speech intelligible.

However we have found that the quality and intelligibility of speech coded at bit rates as low as 2.4Kbit/s in this way is not adequate for many current commercial applications. The common problem in the LPC based speech coding is to obtain more realistic speech synthesis at the receiver part. In many instances, a speech frame cannot be classified as strictly voiced or strictly unvoiced. Indeed, there are transition frames (voiced to unvoiced and unvoiced to voiced) that the LPC model fails to correctly sort. This inaccuracy of the model generates annoying artifacts such as buzzes and tonal noises.

The use of strictly random noise or a strictly periodic impulse train as excitation does not match practical observations using real speech signals. In general, the excitation for unvoiced frames can be reasonably approximated with white noise. For voiced frames, however, the excitation signal is a combination of a quasiperiodic component with noise. Thus, the use of an impulse train is a coarse approximation that degrades the naturalness of synthetic speech. For the LPC coder, the excitation pulses are obtained by exciting an allpass filter using an impulse train.

In our paper, we use wavelet decomposition to rearrange the energy of the high-frequency components to have more realistic synthesized speech. LPC technique is not very efficient for representing speech and does not compete with newly proposed technique because it assumes signal stationary within a given time frame and may therefore lack the ability to analyze the localized events accurately. Signals such as image and speech have different characteristics at different time or space, i.e., they are non stationary. To analyze these signals, both frequency and time information are needed simultaneously, i.e. a time-frequency representation of the signal is needed. The Wavelet Transform proved to be a useful tool for analysis of non-stationary signals like speech; it uses short windows at high frequencies and long windows at low frequencies. This results in multi-resolution analysis by which the signal is analyzed with different resolutions at different frequencies. The non uniform bandwidths corresponding to the wavelet transform is a good model for the human auditory system which has decreasing frequency resolution for increasing frequencies.

Furthermore, the LPC approach assumes a particular linear (all pole) model of speech production, which strictly speaking is not the case and not able to capture the nasalized and unvoiced sounds properly. As we have seen, no phase information is captured by the LPC model: neither voiced nor unvoiced frames have explicit parameters containing clues about the phase. The synthetic speech sounds like the original because the magnitude spectrum, or power spectral density, is similar to the original signal. Even though a human listener is relatively insensitive to the phase, retaining some phase information adds naturalness to the synthetic speech, leading to an improvement in quality. It is important to note that during speech synthesis of the LPC decoder, phase information for unvoiced frames can generally be ignored, since noise perceptions practically phase less.

LP analysis both in time domain and in frequency domain has restrictions of capturing peaks either in frequency domain or in time domain. Some of the peaks in time domain are more important as in the case of plosives sounds, which have high frequencies. Peaks at lower frequencies correspond to formants and are also very useful features especially for the identification of vowels. The poles in conventional LPC can describe sharp spectral resonances very precisely. Wavelet-based features can capture transients much better than LPC. The wavelet transform is able to model the details of a speech signal that correspond to the nasalized and unvoiced sounds portions of speech better than LPC. Since the most important advantage of wavelet over LPC scheme is that it can preserve both the harmonic structure and the formant structure of the speech signal.

Another important advantages of wavelet transform is that it concentrates speech formation into a few neighboring coefficients. Also when applying the DWT to a given speech signal many coefficients of small values (depending on level we choose) are thus considered insignificant. The retained coefficients will still have the larger percentage of energy in the signal. If we go back to the results of both LPC and Wavelet, we can say that the signal to noise ratio of the Wavelet (as one of the most important measurements of the performance) has high values comparable to LPC which produce a good quality by the Wavelet. Another good advantage of wavelet over LPC is that the compression factor is not constant and can be varied. In general, for the vocoders we can note that increasing the bit rate is not the best idea to improve the performance since the improvement in the quality is not linear.

REFERENCES:

[1]Robust spectral parameter coding in speech processing: nadim batry

[2]Wikipedia, the free encyclopedia.

[3]Data compression.com

[4]SPEECH CODING ALGORITHMS.Foundation and Evolution of Standardized Coders WAI C. CHU Mobile Media Laboratory DoCoMo USA Labs San Jose, California

[5]Final Year Electronic and Electrical Engineering “Applications of Digital Signal Processing”

[6]Speech compression and tone detection in a real time system. By Kristina burglund.

[7]Answers.com

[8]International Journal of Biomedical Sciences Volume 1 Number 4. “The Effect of Different Compression Schemes on Speech Signals”
.Jalal Karam, and Raed Saad

[9]ELEC 4801 THESIS PROJECT. Thesis “Speech Compression Using Wavelets” submitted by Nikhil Rao

[10]Ben Gold and Nelson Morgan. ‘Speech and Audio Signal Processing’. John Wiley and Sons, 2000.

[11]Speech coding. a tutorial review by Andres s. Spanias member IEEE.

[12]Speech coding fundamentals and applications. Mark Hasegawa Johnson.

[13]Audio compression using wavelet techniques. Matias Zanartu.

[14]On the use of zero crossing rate for an application of classification of percussive sounds. Fabian Gouyon and Francois Pachet