



.....

**A NEW SINUSOIDAL SPEECH CODING TECHNIQUE WITH SPEECH
ENHANCER AT LOW BIT RATES**

Samer J. Alabed

Darmstadt University of Technology
Darmstadt, Germany

Eyad A. Ibrahim

Zarqa University of Technology
Zarqa, Jordan

ABSTRACT

Speech coding deals with the problem of reducing the bit rate required for representing speech signals while preserving the quality of the speech reconstructed from that representation. In this paper, we propose a novel speech coding technique, not only to compress speech signal at low bit rate, but also to maintain its quality even if the received signal is corrupted by noise. The encoder of the proposed technique is based on speech analysis/synthesis model using a sinusoidal representation where the sinusoidal components are involved to form a nearly resemblance of the original speech waveform. In the proposed technique, the original frame is divided to voiced or unvoiced sub-frames based on their energies. The aim of the division and classification is to choose the best parameters that reduce the total bit rate and enable the receiver to recover the speech signal with a good quality. The parameters involved in the analysis stage are extracted from the short-time Fourier transform where the original speech signal is converted into frequency domain. Making use of the peak-picking technique, amplitudes of the selected peaks with their associated frequencies and phases of the original speech signal are extracted. In the next stage, novel parameter reduction and quantization techniques are performed to reduce the bit rate while preserving the quality of the recovered signal.

Keywords: Speech Coding, Speech Enhancement, Speech Compression, Waveform Speech Coder, Sinusoidal Model, Source Coding.

1. INTRODUCTION

Due to the redundancy in speech signals, speech coding used to compress speech is one of the most important speech processing steps. Speech coding or compression deals with the problem of obtaining compact representation of speech signals for efficient digital storage or transmission and in reducing the bit rate required for a speech representation while preserving the quality of speech reconstructed from that representation. Hence, the main objective of speech coding techniques is to

represent speech signal with minimum number of bits while maintaining its quality. Furthermore, speech coding techniques are used for improving bandwidth utilization and power efficiency in several applications such as digital telephony, multimedia applications, and security of digital communications which require the speech signal to be in digital format to facilitate its process, storage, and transmission.

Although digital speech brings flexibility and opportunities for encryption, it is also associated, when uncompressed, with a high data rate and, hence, high requirements of transmission bandwidth and storage. In wired communications, very large transmission bandwidths are now available, however, in wireless and satellite communications, transmission bandwidth is limited. Therefore, reducing the bit rate is necessary to reduce the required transmission bandwidth and memory storage.

In order to reduce the bit rate of speech signal while preserving its quality, speech coding provides sophisticated techniques to remove the redundancies and the irrelevant information from the speech signal. There are two categories of speech coding techniques: i-) techniques based on linear prediction [1] and ii-) techniques based on orthogonal transforms [1-19]. The techniques belonging to the first category are very well known [13-19] where one of them, called regular pulse excitation (RPE), is now used for the GSM standard [1]. The proposed technique described in details in this paper belongs to the second category.

The encoder, analysis stage, and the decoder, synthesis stage, are the two main components of any speech coding technique. In the analysis stage, the encoder encodes the speech signal in a compact form using a few parameters where the analog speech signal $s(t)$ is first sampled at rate $f_s \geq 2f_{max}$, where f_{max} is the maximum frequency content of $s(t)$ and the sampled discrete time signal is denoted by $s(n)$. Afterwards, one of the coding techniques such as pulse code modulation (PCM), differential PCM, predictive coding, ... , etc is used to encode the signal $s(n)$. In PCM coding technique, the discrete time signal $s(n)$ is quantized to one of the 2^R levels where each sample $s(n)$ is represented by R bits. In sinusoidal speech coding [2-9],[12], the encoder takes a group of samples at a time, extracts some parameters from them, and then converts the extracted parameters to binary bits. After that, the binary signal is transmitted to decoder. In the synthesis stage, the decoder reconstructs the parameters from the received binary bits. Making use the reconstructed parameters, the can recover the original speech signal.

In the proposed technique, sinusoidal speech coding is used to reduce the required bit rate of a speech signal while maintaining its quality. We first divided the speech signal to sub-frames and made voiced/ unvoiced classifications based on their energies. In the analysis stage and after converting the speech frame into frequency domain using the short-time Fourier transform, all peaks with their associated frequencies and phases are extracted using the peak-picking strategy. In the next stage, novel parameter reduction and quantization techniques as well as the concept of birth and death tracking of the involved frequencies are performed to reduce the required bit rate and enhance the quality of the recovered signal.

The layout of this paper is organized as follows: In section two, the implementation of the sinusoidal coder is introduced; this is followed by discussion of the proposed technique in section three. In the last section, the authors present the experimental results and conclusions.

2. IMPLEMENTATION OF THE SINUSOIDAL CODER

2.1. Analysis-synthesis model

A sinusoidal speech model is a vocoding strategy proposed in [1] to develop a new analysis/synthesis technique characterized by the amplitudes, frequencies and phases of the speech sine waves. This model has been shown to produce a high quality recovered speech at low data rates [1]-[12] where the k^{th} segment (frame) of the input speech is represented as a sum of finite number of sinusoidal waves with different amplitudes, frequencies, and phases, such that

$$s(n) = \sum_{k=1}^P A_k \cdot \sin(\omega_k n + \theta_k) \quad (1)$$

where A_k , ω_k , θ_k , and P represent the amplitude, frequency, phase of the k th sinusoidal wave, and the number of possible peaks, respectively. It has also been shown that the sinusoidal encoder is capable of representing both voiced and unvoiced speech frames [1]. In the analysis/synthesis model and after dividing the original speech signal into small frames, the analysis stage is used to extract parameters from each speech frame which represent it. The extracted parameters are used at the synthesis stage to reconstruct the speech frames which should be as close as possible to the original ones.

2.2. Encoder stage

The encoder processes the speech signal and converts it to a set of parameters, before quantizing them in order to transmit the resulting binary bits along the digital channel. In the proposed technique, we focus on minimizing the overall bit rate required to represent the speech signal while maintaining the perceptual quality of the reconstructed speech. First, the speech is sampled at 8 kHz and divided into main frames. Afterward, the main frames are categorized based on their energies into voiced and unvoiced frames, so that the unvoiced frame has less peaks as compared to the voiced frames.

In addition to that, each of the voiced main frames is further divided into N sub-frames which are also classified according to their energies, so that the sub-frame with higher energy gets more peaks than that with lower energy. The purpose of these classifications is to extract the best parameters which represent speech frames to achieve low bit rate and good quality for the reconstructed speech. The two parts of the proposed encoder stage are explained in the following subsections.

2.2.1 Peak-picking strategy

In order to make the speech signal wide sense stationary, the length of each main frame should be small enough. In the proposed technique, the encoder divides the speech signal into (20 to 40 ms) main frames and then transforms them into the frequency domain using the fast Fourier transform (FFT) technique. A crucial part in a sinusoidal modeling system is peak detection since the speech is reconstructed at the decoder using the detected peaks only. There are fundamental problems in the estimation of the meaningful peaks and their corresponding parameters. Most of these problems are related to the length of the analysis window where a short window is required to follow rapid changes in the input signal and a long window is needed to estimate accurate frequencies of the sinusoidal waves or to distinguish spectrally close sinusoids from each other. It is worth to mention that a Hanning window is used in the analysis stage, since it has very good side lobe structure which improves the speech quality.

In almost all the sinusoidal analysis systems, the peak detection and parameter estimation is performed in the frequency domain. This is natural, since each stable sinusoid corresponds to an impulse in the frequency domain. However, natural sounds are infinite-duration stable sinusoids.

The simplest technique for extracting sinusoidal waves of a speech signal is to choose a large number of local maximums in the magnitude of the STFT where a peak or a local maximum in the magnitude of the STFT indicates the presence of a sinusoidal wave. This method, often used in audio coding applications, is very fast and produces a fixed bit rate. However, to achieve a low bit rate, a small number of sinusoids should be chosen. A natural improvement of this technique is to use a threshold for peak detection where all local maximums of the STFT amplitudes above the threshold are interpreted as sinusoidal peaks.

In the proposed technique, the original speech is divided into main frames where each of them is also divided into 6 sub-frames. The peaks are selected by finding the location of change in spectral slope from positive to negative. A more accurate technique using a parabola that is fitted to peak and the location of its vertex is encoded as the peak frequency. Usually after performing this step, around eighty peaks are obtained. The obtained peaks are further reduced by the proposed reduction techniques, described latter, without significant loss of perceptual information. The amplitude spectrum is illustrated in Fig. 1.

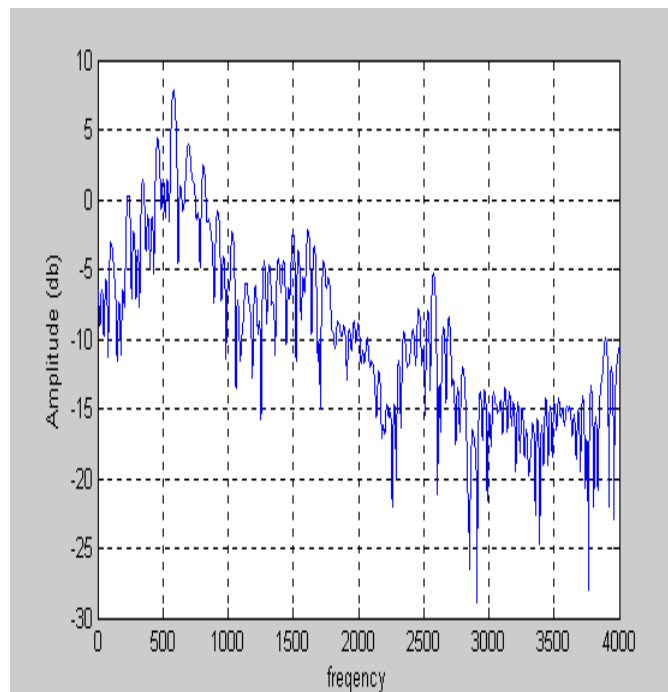


Fig. 1: Amplitude Spectral Domain of a Voiced Frame

After performing the proposed reduction techniques, we extract the frequency locations corresponding to the detected peaks as well as the significant phases. The last step is to quantize them before transmitting them to the receiver.

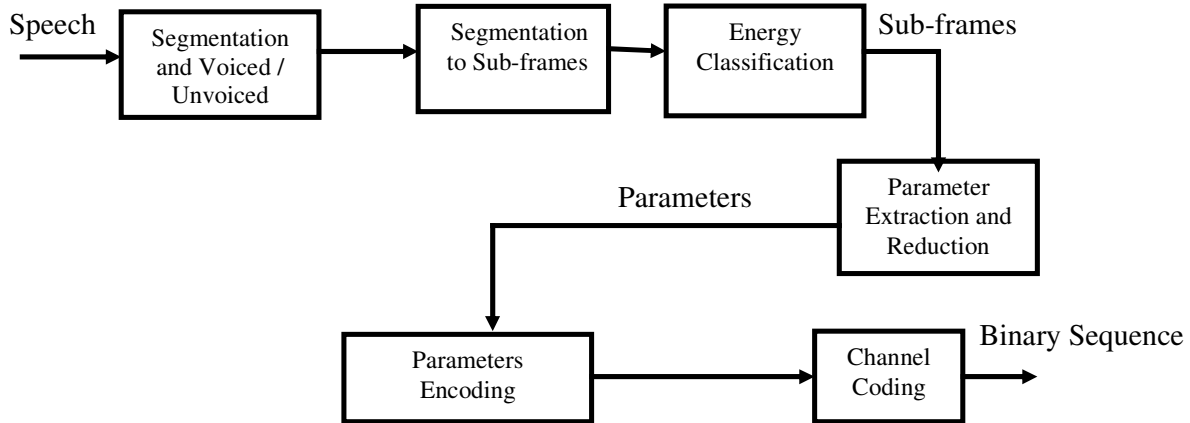
2.2.2. Parameters optimization

In our proposed technique, the encoding of the speech frames is based on selecting the most important peaks rather than encoding all peaks by dividing the frame to sub-frames and making proper classifications. The block diagram of our new encoder model is shown in Fig. 2 (a and b). In this model, the original speech is divided in time domain to main frames. After that, we classify these main frames to voiced and unvoiced frames using energy threshold where the energy of voiced frames should be higher than this threshold value while the energy of unvoiced frames is below it. If the main frame is voiced, it will be divided to N sub-frames. Afterward, we make energy classification to the sub-frames, so that the sub-frame with higher energy gets more peaks than that with lower energy. If the main frame is unvoiced, the same procedure is applied but there is no energy classification and all sub-frames have the same number of peaks which is the number chosen for the lowest energy sub-frame in the voiced frame. The purpose of dividing the main frames to N sub-frames and making the voiced and unvoiced classification is to choose the best peaks in these sub-frames that enable us to achieve a low bit rate and a good quality for the reconstructed speech.

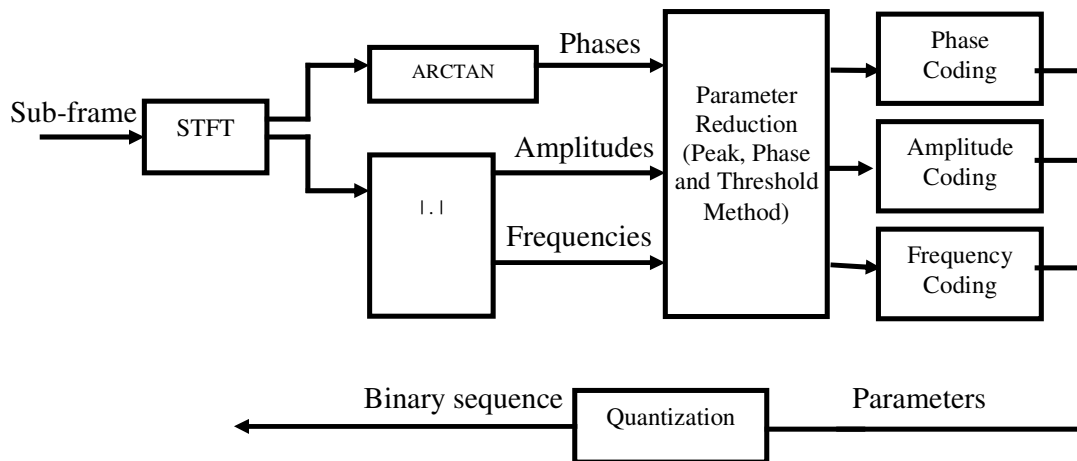
The parameter reduction is one of the most important parts in this model, since most errors occur in this stage. The aim of this part is to reduce the number of parameters described each main-

frame to (15-30) parameters. In addition to the proceeding reduction technique, another reduction of information can come out from quantization process. This justifies our main concern about this topic. Hence, after classifying the frames and dividing them to sub-frames, the following three encoding techniques are proposed to reduce the number of parameters.

- A. Peak reduction,
- B. Phase reduction,
- C. Threshold reduction.



(a)



(b)

Fig. 2: (a) The Encoder Stage, (b) Parameter Extraction and Reduction stage

A. Peak reduction technique

This technique is based on selecting the best N sinusoidal waves in each speech frame. The value of N depends on the required data rate. The following encoding procedure summarizes this technique:

1. Selecting the largest peaks for each sub-frame, after converting it to frequency domain.
2. If a group of peaks are close enough to each other, then choose the largest peak to represent them.

It should be noted that by doing this step, the speech signal is still having a very good quality which encouraging to go forward to the second reduction technique.

B. Phase reduction technique

This type of reduction aims to reduce the phase parameters which can be performed after determining whether the sub-frame is voiced or unvoiced where the voiced frame has the following characteristics:

- Its energy is greater than a preset threshold.
- Its zero crossing is less than that of the unvoiced (also less than a preset threshold value).
- It has a specific pitch value.

Note that the first item of the previous criteria is the most sufficient one to reduce the overall complexity; therefore, we depend on it in the binary decision process. If the frame is voiced, i.e., it has a large embedded energy, the encoder extracts its phases. Otherwise, the frame is considered as unvoiced and, in this case, its phases are estimated using the phase extraction equations proposed by McAulay and Quatieri in [2], [7] or Ahmadi and Spanias in [3], [4]. Once this procedure is performed, the number of phases is reduced with humble effect on speech quality, since human ear is less sensitive to phase distortion, so the elimination is justified.

C. Threshold reduction technique

This technique considered as the most efficient one among all reduction techniques described previously, in the sense that it reduces the number of peaks without affecting the voice perceptual sense. This technique chooses a threshold value that is very small, so that all the peaks below this value are eliminated. By doing this, not only the number of amplitudes, but also their corresponding number of frequency locations and corresponding phases are also reduced. Thus, this reduction technique reduces the total data rate required for transmission and enhances the recovered speech frames by filtering the peaks of the noise signal whose amplitudes are less than the threshold value. At the end of the day, this filtration is an advantageous technique. On the other hand, the increase of the threshold above a certain value produces a corrupted speech frame because of filtering important informational peaks. Therefore, the threshold value should be chosen based on exhaustive statistical study to confirm the optimal value.

After performing these reduction techniques, we end up having S amplitudes and frequencies plus $(0.5 S)$ phases. In other words, we have: S peaks plus S frequency locations plus $(0.5 S)$ phases for each main frame. In this paper, we use 6 bits for each amplitude and frequency location and 4 bits for each phase.

Thus, the required data rate for each frame = $(6 S + 6 S + 4 (0.5 S)) = 14 S$ bits/frame. The total data rate R can be computed as:

$$R = 14 S \text{ (bits/frame)} * N \text{ (frame/s)} = 14 N S \text{ bps.}$$

Some extra bits can also be used for control and error detection and correction. At this point, we turn to the quantization process which has same degree of importance.

2.2.3. Modeling and encoding technique

The quantization technique is defined as the process in which the dynamic range of the signal is divided into number of levels. The number of levels is determined from the formula $L = 2^k$ where k is the word size and L is number of distinct words. We assigned each level to a specific word after rounding the sample to the nearest level. This kind of quantization is called PCM.

In our model, the different techniques described in the next subsections are used to encode phase, frequency location, and amplitude of each sinusoid.

A. Sinusoidal phase modeling and encoding

The bits used to quantize the phases can be reduced by minimizing their entropy. In order to minimize the entropy of the phases, the encoder predicts differentially a phase from its past value and encodes the phase difference rather than the phase itself which has less entropy than the actual phase [3]. The differentially predicted phase is given by

$$\hat{\theta}_l^k = \theta_l^{k-1} + \omega_l^{k-1} \cdot T \quad l = 1, 2, \dots, L, \quad (2)$$

where the superscript k denotes the frame number, ω_l^k is the (l) sinusoid, T is the time interval between frames, and L is the number of sinusoidal components. The phase differences or residues are expressed as

$$\Delta\theta_l^k = \theta_l^k - \hat{\theta}_l^k \quad l = 1, 2, \dots, L \quad (3)$$

where the actual phase is used to estimate the phase difference $\Delta\theta_l^k$.

B. Sinusoidal frequency encoding

After transforming the speech frames into the frequency domain using the STFT strategy, the frequency location indices are integer values, i.e., in Matlab, the spectrum has 512 points in both sides. By taking one side (256 points) that represents the frequencies contained within one frame, frequency locations are from 1 to 256 which corresponding to the frequency range from 0 to 4000 Hz where (4) is used to get the frequency rang:

$$frequency = (location - 1) \cdot \frac{4000}{framesize} \quad (4)$$

The minimum number of bits required to encode each frequency location is 8 bits which is the normal case, however, in this model, the situation looks different, where only 6 bits per location are used and the results are almost similar. In the proposed model, the first frequency location represents low frequency components and the last frequency location represents high frequency component. Hence, we do not need to spend the same number of bits for each frequency location where higher frequency locations correspond to the high frequency components which have less effect in speech perception. Therefore, higher frequency locations can be quantized using fewer bits than lower frequency locations. This reduces the bit rate while keeping the speech quality almost the same. Hence, to implement this idea, we developed the following procedure:

1. Dividing the frequency locations by the STFT size to normalize the frequency location vector, and then we obtain (f_n) .

2. The normalized frequency location vector is transformed to another domain (u_n) to reduce the number of bits used to encode each frequency location where (u_n) is given by

$$u_n = \frac{(\log_e (1 + 4.f_n) - 0.072)}{1.528}.64 \quad (5)$$

After calculating U_n , we obtain values within the range of (1-64). Note that equation (5) is similar to μ -Law used in digital signal processing to compress the speech signal.

3. Round the result and then convert resulting value to binary.

C. Sinusoidal amplitude encoding

This technique is also important since the amplitude is susceptible to any change in it due to the quantization process. Therefore, we proposed an encoding technique that increases the resolution in order of (6-12) times than the resolution of the PCM. Let us assume that we have amplitudes: $x(n) = [\text{amp}_1, \text{amp}_2, \dots, \text{amp}_N]$ where N is the number of the considered peaks, then the proposed encoding technique is summarized as follows:

1. Take $\text{Log}_2(x_n)$ of the amplitude, in order to reduce the dynamic range.
2. The results of the first step are all negative, since all amplitude involved are less than unity.
3. The resulted dynamic range from the previous two steps is (-1,-20), because the lowest amplitude is 10^{-6} which is our predetermined threshold.
4. Take the absolute value of the results and then multiply them by (β) where the value of (β) is chosen to be 3 to make the dynamic range (1-64). Then, extract the values of a_n using (6).
5. Sort the amplitudes (a_n) in ascending order together with the associated phases and frequencies as a bundle. This step is justified because we note that there is a small difference between successive amplitudes in the same frame.
6. Take the integer part in the first amplitude a_0 (floor) and convert it to binary (q_0).
7. Subtract the value found in the step 6 above (q_0) from all other amplitudes (a_n).
8. Multiply the next amplitude (a_n) by a number (α) in the range (6-12).
9. Floor the value found in the previous step (q_i where $i=1, \dots, N-1$)
10. Convert the result to binary.
11. Subtract all remaining a_n 's by the output of the step 9 divided by α .
12. Repeat steps (8-11) until you finish a_n 's.

The general equations that represent the amplitude quantization are given by

$$a_n = \beta \cdot \text{abs}(\log_2(x(n))) \quad (6)$$

$$q_0 = \text{Floor}[(a_0)] \quad (7)$$

$$q_n = \text{Floor}\left\{\alpha \cdot a_n - [\alpha \cdot q_0 + \sum_{i=1}^{N-1} q_i]\right\} \quad (8)$$

2.3. Decoder stage

The decoder is used to reconstruct the original signal by decoding the parameters extracted in the encoder stage as shown in Fig. 3. These parameters are then used to reconstruct the speech frames by linearly summing the sine waves of different amplitudes, frequencies, and phases.

2.3.1. Decoding strategy

This strategy converts the received binary representation of the parameters to a decimal form. Three decoding techniques for the amplitudes, frequencies, and phases are required to recover them. The reconstructed parameters should be as similar as possible to the original ones.

A. Phase decoding technique

This process can be summarized as follows:

1. Dequantize the received binary bits corresponding to the phase differences.
2. Predict the phases from their past values using equation (2).
3. Add the estimated phase found in the previous step to the phase difference found in step 1.

B. Frequency decoding technique

1. Convert the received binary bits to decimal form $\hat{u}_n$.
2. The estimated normalized frequency location vector (z_n) is reconstructed from $\hat{u}_n$ using equation (9) which is the inverse of equation 5, given by:

$$z_n = \frac{(\exp(0.072 + 0.023875 \cdot \hat{u}_n) - 1)}{4} \quad (9)$$

3. Round z_n .

C. Amplitude decoding technique

Convert the binary signal to $[q_0, q_1, \dots, q_N]$, then we find

$$\begin{aligned} d_0 &= q_0 \\ d_1 &= q_0 + \frac{q_1}{\alpha} \\ d_2 &= q_0 + \frac{q_1}{\alpha} + \frac{q_2}{\alpha} \\ &\vdots \\ d_n &= q_0 + \sum_{i=1}^n \frac{q_i}{\alpha} \end{aligned} \quad (10)$$

where $n+1$ is the number of considered peaks. Note that after performing this step, the maximum error occur at $n = 0$. However, this error is very small. To reconstruct the signal parameter amplitudes (y_n), we use the following equation

$$y_n = -2^{\left(\frac{d_n}{\beta}\right)} \quad (11)$$

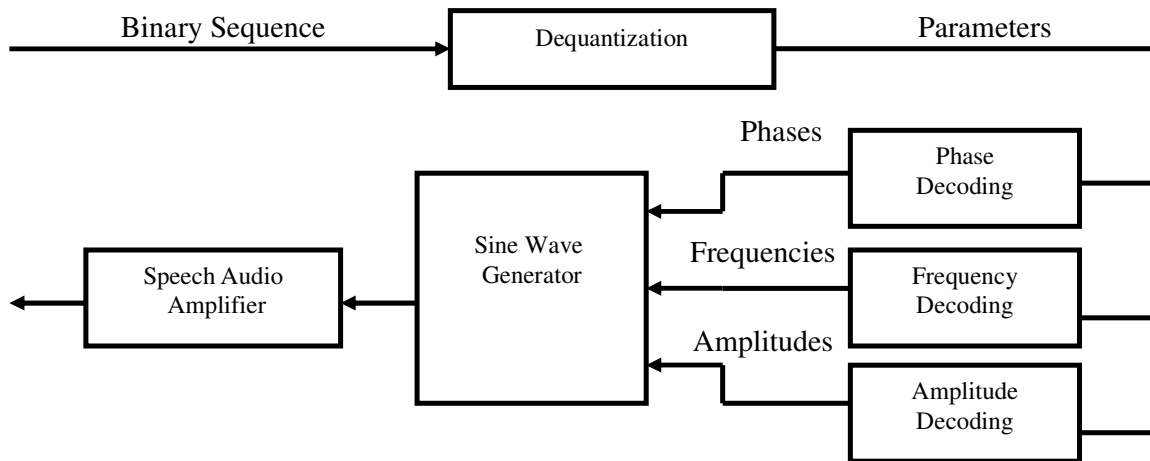


Figure (3): The Decoder Stage

3. ADVANTAGES OF THE PROPOSED SPEECH CODING TECHNIQUE

From the previous described section, we can conclude that the proposed speech coding technique

- Enjoys a very efficient and effective encoding and decoding procedure.
- Gives a reconstructed speech signal with high quality.
- Reduces the data rate to (3.6-8) kbps.
- Enhances the original signal when the received speech signal is corrupted by additive noise.
- Does not depend on a pitch (the fundamental frequency).
- Can be considered as a noise immune.
- Reduces the total required transmitted power due to minimizing the required bit rate.
- Allows error detection and correction procedures.

4. EXPERIMENTAL RESULTS

1. From literature it is advised to use a window size equal to 2.5 times the average pitch., therefore, the size of the main frame is between (20-40) ms. This means that the overlap and add percentage is 33.3% at the transmitter, and the FFT size is equal to 512 points.
2. After an exhaustive statistical study, the threshold value used in Sec. 2.2.2-C is selected to be less than (10^{-6}). As explained in Sec. 2.2.2-C, this step is to reduce the total number of peaks.
3. Hamming window is employed.
4. The data rate of the proposed technique is between 3.6 kbps to slightly less than 8 kbps. We remark that for high quality speech, the data rate is less than 8 kbps where the remainder of bits can be used for controlling and error detection and correction.
5. At the decoder, we perform an overlap and add with percentage equal to 50% to eliminate discontinuity of the received speech.

5. CONCLUSIONS

In this research, we propose a computationally efficient low bit rate speech coding technique based on the sinusoidal model with efficient speech enhancer. The proposed technique can reconstruct the transmitted speech signal at the decoder with good quality and intelligibility, even if it is corrupted by a thermal noise, at bit rate from 3.6 to 8 kbps. In our speech coding technique, we propose novel encoding techniques to minimize the total number of parameters extracted from frequency domain, i.e., amplitudes, frequency locations, and phases. The significant one is the threshold technique which not only reduces the number of parameters but also enhances the recovered speech signal. After that, we introduced new techniques, i.e., phase coding, amplitude coding, and frequency coding, to model and encode these parameters efficiently.

REFERENCES

1. Spanias, "Speech Coding: A Tutorial Review," Proc. of the IEEE, Vol. 82, No. 10, pp. 1541 - 1582, Oct. 94.
2. R.J. McAulay and T.F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Trans. On ASSP, Vol. ASSP-34, No. 4, pp. 744-754, August 1986.
3. Sassan Ahmadi & Andreas .S. Spanias, "New Techniques For Sinusoidal Coding of Speech at 2400 bps", Arizona State University. Proc. Asilomar-96, Nov 3-6, Pacific Grove, CA1.
4. Sassan Ahmadi and Anderias Spanias, "Low-bit rate speech coding based on harmonic sinusoidal models", In Proc. International Symposium on Digital Signal Processing (ISDSP), pp. 165-170, July 1996.
5. Remy Boyer and Julie Rosier, "Iterative Method for Harmonic and Exponentially Sinusoidal Models", Proc. Of the 5th Int. Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, September 26-28, 2002.
6. E. B. George and M. J. T. Smith. Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. IEEE Trans. Speech and Audio Proc., Vol.5, Number 5, pp.389–406, September 1997.
7. Robert J. McAulay and Thomas F. Quatieri, "Processing of Acoustic Waveforms," United States Patent, Dec. 28, 1999, Patent No.:Re.36, 478, Assignee: Massachusetts Institute of Technology, Cambridge, Mass.
8. K. Vos, R. Vafin, R. Heusdens, and W. B. Kleijn, "High quality consistent analysis-synthesis in sinusoidal coding", in Proc. AES 17th Int. Conf., 'High-Quality Audio Coding', pp. 244 – 250, 1999.
9. Izmirli, O., "Non-harmonic Sinusoidal Modeling Synthesis Using Short-time High-resolution Parameter Analysis" Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.
10. Harald Pobloth, Renat Vafin, and W. Bastiaan Kleijn, " Polar Quantization of Sinusoids from Speech Signal Blocks", EUROSPEECH 2003 – Geneva.
11. Mathieu Lagrange, Sylvain Marchand and Jean Bernard Rault, "Sinusoidal Parameter extraction and Component selection in a Non Stationary Model", Proc. Of the 5th Int. Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, September 26-28, 2002.
12. Ibrahim Mansour and Samer J. Alabed. "Using Sinusoidal Model to Implement Sinusoidal Speech Coder with Speech Enhancer". The 6th International Electrical and Electronics Engineering Conference (JIEEEEC), Volume 1, page 1-8, march 2006.
13. Kang Sangwon, Shin Yongwon, and Fischer Thomas. (2004). "Low-Complexity Predictive Trellis-Coded Quantization of Speech Line Spectral Frequencies". IEEE Transactions on Signal Processing, Vol. 52, No. 7.

14. Alku Paavo, and Bäckström Tom. (2004). "Linear Predictive Method for Improved Spectral Modeling of Lower Frequencies of Speech With Small Prediction Orders". IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 2.
15. Atal Bishnu. (1982). "Predictive coding of speech at low bit rates". IEEE Transactions on Communications, COM-30(4):600-614.
16. Brinker Albertus C. den, Voitishchuk, V., and Eijndhoven Stephanus J. L. van. (2004). "IIR-Based Pure Linear Prediction". IEEE Transactions on Speech and Audio Processing, Vol. 12, No. 1.
17. Papamichalis Panos. (1987). "Practical Approaches to Speech Coding", Prentice Hall, Inc. Texas Instruments, Inc. Rice University.
18. Härmä Aki. (2001). "Linear Predictive Coding With Modified Filter Structures". IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 8.
19. Hu Hwai-Tsu, and Wu Hsi-Tsung. (2000). "A Glottal-Excited Linear Prediction (GELP) Model for Low-Bit-Rate Speech Coding", Proc. Natl. Sci. Coun. ROC(A) Vol. 24. pp. 134-142.
20. Sudha.P.N and Dr U.Eranna, "Source and Adaptive Channel Coding Techniques for Wireless Communication", International Journal of Electronics and Communication Engineering & Technology (IJCET), Volume 3, Issue 3, 2012, pp. 314 - 323, ISSN Print: 0976-6464, ISSN Online: 0976-6472,.
21. P Mahalakshmi and M R Reddy, "Speech Processing Strategies for Cochlear Prostheses-The Past, Present and Future: A Tutorial Review", International Journal of Advanced Research in Engineering & Technology (IJARET), Volume 3, Issue 2, 2012, pp. 197 - 206, ISSN Print: 0976-6480, ISSN Online: 0976-6499.