

# Lossy Coding of Speech Signals using Subband Coding

Dr. B. Kirubagari  
Department of Computer Science and  
Engineering  
Annamalai University

T. Akilan  
Department of Computer Science and  
Engineering  
Annamalai University

**Abstract**—Speech coding is a methodology of representing a digitized speech signal using as few bits as could reasonably be expected, keeping up the quality at the same time. In Ubiquitous environments, analysis and encryption of speech plays a discriminating part in different acoustic-based coding systems. In this work, a new speech coding technique using subband coding is proposed for reducing the memory occupied of the speech signals. The amplitude values of the input is extracted after pre-processing, decomposing and windowing, the values are transformed into frequency domain by applying discrete cosine transform (DCT). The initial 20 coefficients, which holds the maximum content of speech features are separated and coded using subband coding. To reconstruct the speech signal the signal is retransformed into time domain by applying inverse discrete cosine transform (IDCT). The experiments are conducted using a speech signal at 8 KHz with 16 bits per sample. The signal to noise ratio (SNR) demonstrates the effectiveness of the model used.

**Keywords**— DCT, SNR, Subband, Quantization, Windowing.

## I. INTRODUCTION

Speech coding is the art of creating a minimally redundant representation of the speech signal that can be efficiently transmitted or stored in digital media, and decoding the signal with the best possible perceptual quality [1]. Today, speech coders have been taken, the essential components are used in telecommunications and in multimedia infrastructures. In the same way as other different signals, however, a sampled speech signal contains a lot of information that is either redundant (nonzero mutual information between successive samples in the signal) or perceptually irrelevant (information that is not perceived by human listeners). Most telecommunications coders are lossy, implying that the synthesized speech is perceptually similar to the original but may be physically dissimilar.

A speech decoder gets coded frames and synthesizes reconstructed speech. Standards typically dictate the input–output relationships of both coder and decoder. Speech coders differ fundamentally in bit rate (measured in bits per sample or bits per second), complexity (measured in operations per second), delay (measured in

milliseconds in the middle of recording and playback), and perceptual quality of the synthesized speech. Narrowband (NB) coding alludes to coding of speech signals whose bandwidth is less than 4 kHz (8 kHz sampling rate), while wideband (WB) coding refers to coding of 7-kHz-bandwidth signals (14–16 kHz sampling rate). Subband coders are broadly utilized for high-quality audio coding. The advantage of subband coding is that each band can be coded distinctively and that the coding error in each band can be controlled in relation to human perceptual characteristics.

## II. PROPOSED METHODOLOGY

### A. Speech compression

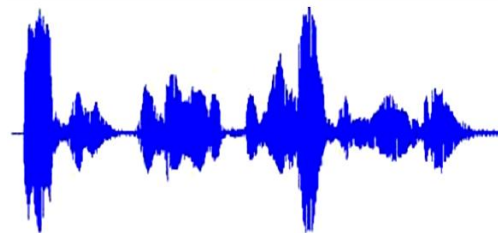


Figure 1: The Uncoded Speech Signal

Speech compression may be varying the amounts of compression in data according to the sampling rate utilized. This gives distinctive levels of system complexity and compressed quality of speech data. The recorded waveform which is compressed can be transmitted with or without loss. The digital audio data is handled through mixing, filtering and equalization. The speech signal is fed into an encoder that uses fewer bits than original audio data bit rate [2]. This results in reducing the transmission bandwidth of digital audio streams and also reduces storage size of audio files. Compression can be classified into lossy and lossless. Lossy compression is transparent to human perceptibility yet lossless being have a compressing factor from 6 to 1. The uncoded speech signal is shown in the figure1 and figure 2 shows block diagram of Speech Coding using Subband Coding.

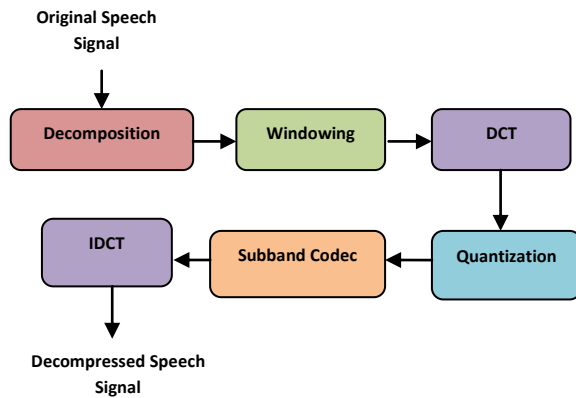


Figure 2: Block diagram of Speech coding using Subband Coding

### B. Subband Codec

The procedure of breaking the input speech signals into sub signals using band pass filters and coding each signals independently is called subband coding. To keep the number of samples to be coded at the very least, the sampling rate for the signals in each band is reduced by decimation. Since the band pass filters are not ideal, there is some overlap between nearby bands and aliasing occurs during decimation. Ignoring the distortion or noise due to compression, Quadrature mirror filter (QMF) banks permit the associating that happens filtering and sub sampling at the encoder to be cancelled at the decoder. The codecs used in each band can be PCM, ADPCM, or even an analysis-by-synthesis method. The advantage of subband coding is that each band can be coded differently and that the coding error in each band can be controlled in relation to human perceptual characteristics. Transform coding methods were initially applied to still images however later explored for speech [3,4]. The essential principle is that a piece of speech samples is worked on by a discrete unitary transform and the resulting transform coefficients are quantized and coded for transmission to the recipient. Low bit rates and good performance can be acquired because more bits can be allotted to the perceptually important coefficients. For well-designed transforms, many coefficients require not be coded at all, but are simply discarded, and acceptable performance is still attained to. The distinction between transforms and filter bank methods is somewhat blurred, and the choice between a filter bank implementation and a transform technique may simply be a design choice. Subband encoder and decoder is shown in the figure 3 and 4.

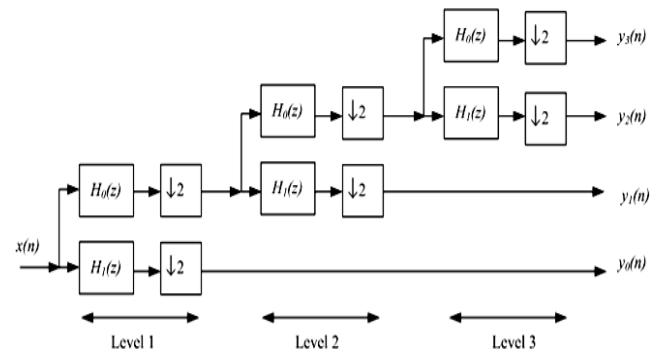


Figure 3: Subband Encoder

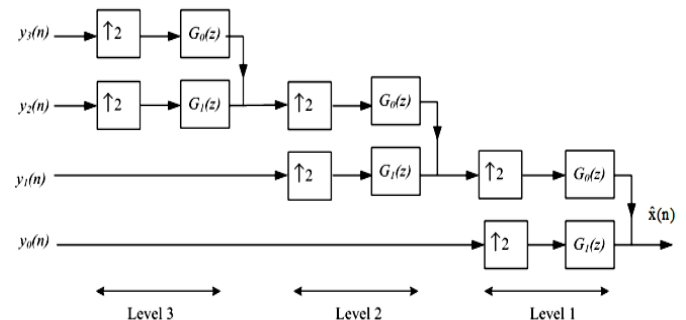


Figure 4: Subband Decoder

### C. Decomposition

Wavelets decompose a signal into different resolutions or frequency bands. Signal compression is focused around the concept that selecting small number of approximation coefficients and some of the detail coefficients can represent the signal components accurately. Speech is initially investigated to focus the voiced and unvoiced parts of the signal [5]. Decomposition of the voiced part into periodic and aperiodic components is then accomplished by first recognizing the frequency regions of harmonic and noise components in the spectral domain. The signal corresponding to the noise regions is used as a first approximation to the aperiodic component.

### D. Windowing

The windows are applied to raw speech frames in order to reduce the spectral leakages effect. For most phonemes the properties of the speech signals remain invariant for a short period of time (5-100 ms). Hence for a short window of time, traditional signal processing methods can be applied relatively successful. A large portion of speech processing, in fact it is carried out in this way by taking short windows (overlapping possibly) and processing them [6]. The short window of signal is called frame. A long signal (of speech for instance or ideal impulse response) is increased with a window function of finite length,

giving limited length weighted (normally) form of the original signal. In speech processing, the shape of the window function is not that crucial but usually some soft window like Hanning, Hamming, triangle, half parallelogram, not with right angles. In this proposed work we use Hamming-window, the window is enhanced to minimize the maximum (closest) side lobe, providing for a height of about one-fifth that of the Hanning window. This window is zero at the edges and rises gradually to be 1 in the middle. When this window is used the edges of the signal are de-emphasised and the edge effects are reduced. It is important to use a Hamming (or the similar *Hann* window) in some kinds of analysis, especially the frequency domain methods. The strategy for using overlapping short-time signals and forming the reconstruction by summing partially overlapping frames is called overlap-add-method.

#### E. DCT/IDCT

The vector value of input speech signal is isolated into smaller frames and arranged into matrix form. DCT operation is performed on the matrix. DCT operation is performed and the elements are sorted in their matrix form to find components and their indices [7,8]. Here 90 values of speech signals are taken and they fed into for further processing to acquire the good perception. The elements are arranged in descending order. After the arrangement has been carried out, the higher values are chosen, then the threshold values are obtained. The coefficients below the threshold values are discarded. Subsequently decreasing the size of the signal which results in compression.

$$X_c(k) = (1/N) \sum_{n=0}^{N-1} x_n \cos(k2\pi n/N)$$

Where,

Where  $k=0,1,2,\dots,N-1$

The original form of data is attained back by using reconstruction process. Then, we perform IDCT operation on the signal. In this manner the signal is reconstructed.

$$x_c(k) = \sum_{n=0}^{N-1} c[u] X_n \cos(k2\pi n/N)$$

$X_n$  is the DCT result, and

$c[u] = 1$  for  $u=0$ , and  $c[u] = 2$  for  $u=1,2,3,\dots,N-1$

Where  $k=0,1,2,\dots,N-1$

#### F. Quantization

The sampled analog signal must be converted from a voltage value to a binary number that the computer can read. The conversion from infinitely precise amplitude to a binary number is called quantization. During quantization, the A/D converter uses a finite number of evenly spaced values to represent the analog signal. The number of distinctive values is determined by the number of bits used for the transformation. Typically, the converter chooses the digital value that is closest to the actual sampled value. A device or algorithmic function that performs quantization is known as a quantizer. The round-off error acquainted by quantization is alluded with as quantization error. In analog-to-digital conversion, the contrast between the actual analog value and quantized digital value is called quantization error or quantization distortion [9]. This error is either due to rounding or truncation. The error signal is sometimes modelled as an additional random signal called quantization noise on account of its stochastic behaviour. Quantization is included to some degree in nearly all digital signal processing, as the methodology of representing a signal in digital form ordinarily involves rounding. Quantization also forms the center of basically all lossy compression algorithms. The first 90 values of speech signals can be taken which holds the maximum content of speech signals

### III. EXPERIMENTAL RESULTS AND DISCUSSIONS

#### 3.1 Screen Shots

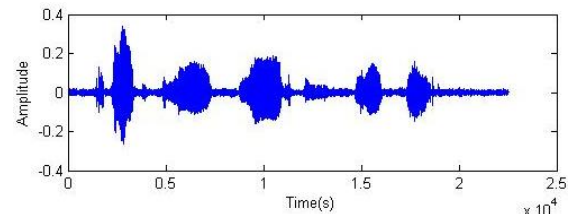


Figure 5: Noisy Speech Signal

The noisy signals are taken from the various areas like airport, babble, car, exhibition, restaurant, station, street, and train. And these noise are further coded and compressed by using DCT with subband coding technique. Figure 5 shows the noisy speech signal.

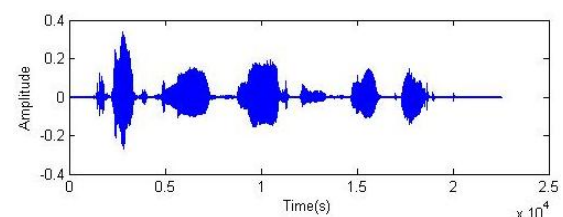


Figure 6: Clean signal

The finest hearing speech signal is called clean signal. This clean signals having original information of the user voice (figure 6). Some additional noises, which are listed above is added to the clean signal to get noisy speech signal.

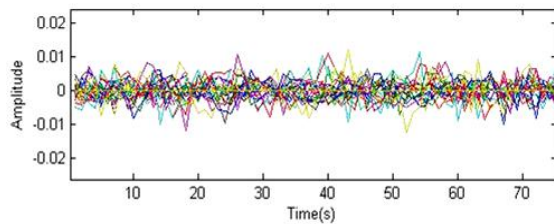


Figure 7: Applying DCT for Noisy Speech Signal

By applying DCT, the elements are sorted in their matrix form to find components and their indices. The arrangement has been done, a Threshold value is decided. The coefficients below the threshold values are discarded. Hence reducing the size of the signal which results in compression. Applying DCT for noisy speech signal is shown in the Figure 7.

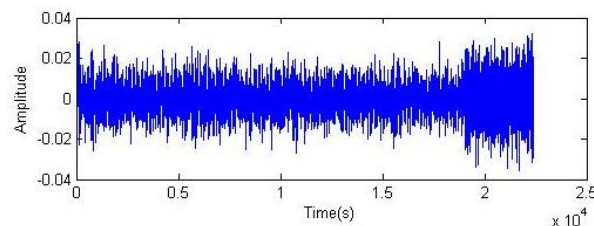


Figure 8: Reconstructed Noisy Signal

The Threshold value is then converted back into the original form by using reconstruction process [figure 8]. In this process, the result of the threshold value is regains the original frequency. This frequency is more or less equal to the original signal in 85% accuracy.

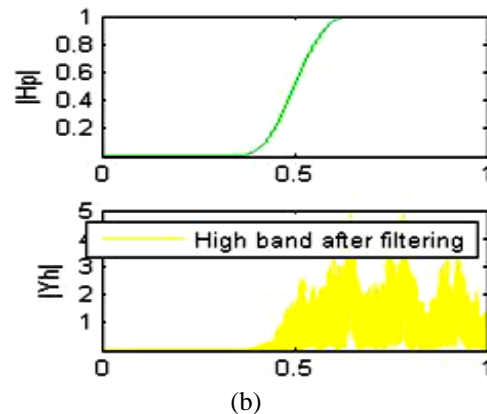
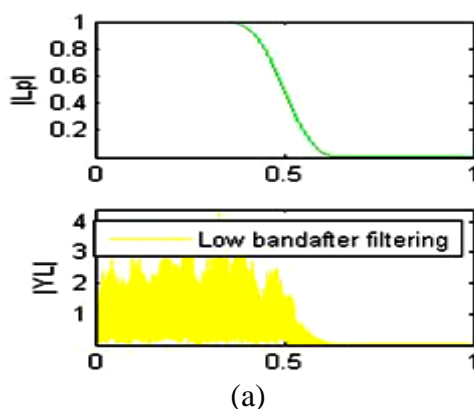


Figure 9 (a): Applying Low Pass Filter,  
(b): Applying High Pass Filter

The Low-Pass Filter, the speech is muted and we can only hear the low frequencies in the wave file. But with the High-Pass Filter, the speech signal is barely audible such that we can only hear the high frequencies that are spoken in the speech signal. the Low-Pass waveform is shown to display only the low frequencies [figure 9 (a)] while the High-Pass filter is only displaying the high frequencies from the sound wave [figure 9 (b)]. And obtained compressed speech signal is shown in the figure 10.

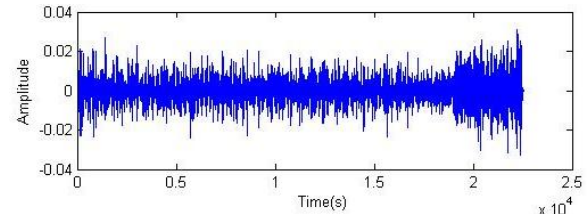


Figure 10: Compressed Speech Signal

### 3.2 Database

Experiments are done on noisy speech signal database (NOIZEUS), which contains noisy signals. The noisy signals are taken in different areas like airport, babble, street, restaurant, exhibition, train, car, station. Noisy is add to the original clean signals, it will processed and finally SNR value will be taken to compares the level of a clean signal to the level of noisy signal [10].

Table 1: SNR values for 0 decibel speech signal

0 dB	Airport	Babble	Exhibition	Restaurant
SNR	2.2829	2.3382	0.8531	2.3078
Values	0.8162	1.7363	0.7987	1.3133

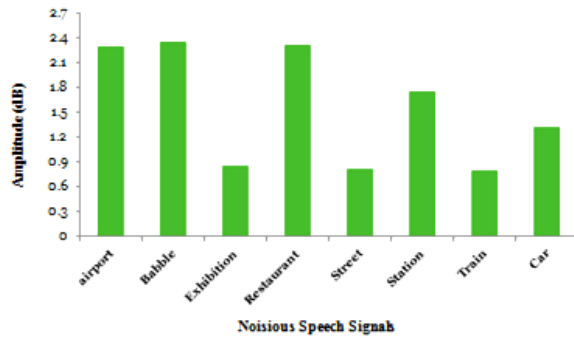


Figure 11: SNR values for 0 dB Speech signals

The SNR is applied to the 0 dB of noisy speech signal, better results will be obtained for Airport, Babble and Restaurant.

Table 2: SNR values for 5 decibel speech signal:

5dB	Airport	Babble	Exhibition	Restaurant
SNR Values	4.7607	4.6134	1.964	4.3242
	Street	Station	Train	Car
	3.3806	3.1754	2.0885	4.0706

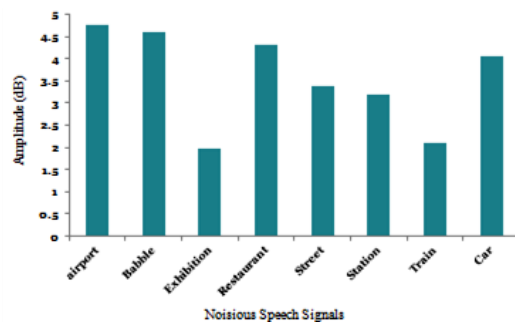


Figure 12: SNR values for 5 dB Speech signals

The SNR is applied to the 5 dB of noisy speech signal, better results will be obtained for Airport, Babble, Restaurant and Car.

Table 3: SNR values for 10 decibel speech signal:

10 dB	Airport	Babble	Exhibition	Restaurant
SNR Value	8.8565	8.602	5.4182	9.0188
	Street	Station	Train	Car
	9.119	7.7203	4.445	5.9361

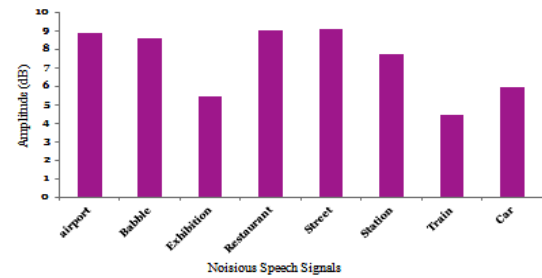


Figure 13: SNR values for 10 dB Speech signals

The SNR is applied to the 10 dB of noisy speech signal, better results will be obtained for Airport, Babble, Restaurant, Street.

Table 4: SNR values for 15 decibel speech signal:

15 dB	Airport	Babble	Exhibition	Restaurant
SNR Values	15.1591	12.8307	9.5267	13.754
	Street	Station	Train	Car
	9.7573	12.0957	8.262	12.2973

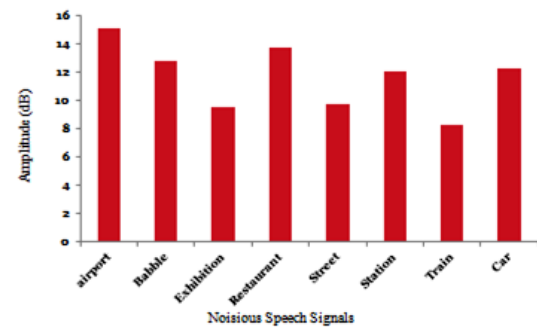


Figure 14: SNR values for 15 dB Speech signals

The SNR is applied to the 15 dB noisy speech signal. Except airport noise, the remaining results are in lesser quality while compared with the airport noisy speech signal.

#### IV. CONCLUSION

Speech coding is an emerging research area and speech compression is a standard for designing and compressing audio and speech signals which are transmitted to the recipient end. This work focuses on developing an efficient speech coding techniques using subband coding. DCT based speech compression approaches are used to produce the better results. This experiment was conducted with the noiseus database. Speech signal is reconstructed from the coded features. We tried out playing the reconstructed speech signal after processing the noisy speech. The subband coding technique is worked and produces an efficient result in these harsh conditions. Few



of them (the more experienced) could understand each word from the corrupted utterance. While hearing the speech, that the clear speech can be achieved by applying subband coding technique.

#### V. REFERENCES

- [1]. Ulrich Benzler, Student Member, IEEE, "Spatial Scalable Video Coding Using a Combined Subband-DCT Approach", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 7, pp. 1080-1087, October 2000.
- [2]. Huijun Ding, Ing Yann Soon, and Chai Kiat Yeo, "A DCT-Based Speech Enhancement System With Pitch Synchronous Analysis", IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 8, pp. 2614-2623, November 2011.
- [3]. K. Satyapriya, Yugandhar Dasari, "Performance Analysis Of Speech Coding Techniques", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 11, pp. 5725-5732, November 2013.
- [4]. Yang-Jeng Chen, Robert C. Maher, "Sub-Band Coding Of Audio Using Recursively Indexed Quantization", pp. 1-4.
- [5]. Sheetal D. Gunjal, Dr. Rajeshree D. Raut, "Advance Source Coding Techniques for Audio/Speech Signal: A Survey", Int.J.Computer Technology & Applications, Vol 3 (4), 1335-1342, August 2012.
- [6]. Sangita Roy, Dola B. Gupta, Sheli Sinha Chaudhuri and P. K. Banerjee, "Studies and Implementation of Subband Coder and Decoder of Speech Signal Using Rayleigh Distribution", Emerging Trends in Computing and Communication, copyright in Springer India, pp. 11-25, 2014.
- [7]. Sorin Dusan, James L. Flanagan, Amod Karve, and Mridul Balaraman, "Speech Compression by Polynomial Approximation", IEEE Transactions On Audio, Speech, And Language Processing, VOL. 15, NO. 2, pp. 387- 395, February 2007.
- [8]. Chandra R. Murthy, Ethan R. Duni, and Bhaskar D. Rao, "High-Rate Vector Quantization for Noisy Channels With Applications to Wideband Speech Spectrum Compression", IEEE Transactions on Signal Processing, vol. 59, no. 11, pp. 5390- 5403, November 2011.
- [9]. Serajul Haque, Roberto Togneri, and Anthony Zaknich, "An Auditory Motivated Asymmetric Compression Technique for Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2111- 2124, September 2011.
- [10]. Hu, Y. And Loizou, P. "Subjective evaluation and compression of speech enhancement algorithms", speech communication, vol.49, no. 7, pp. 588-601, 2007.