Chapter 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

This chapter deals with the study of literature in doing this work. The literature studied is explained in detail by highlighting the points concluded in each paper. The references included in this thesis report are all explained in detail.

Speech coding is a process of obtaining a compact representation for the speech signals by reducing the number of bits used to represent a speech sample without any reduction in the perceptual quality and is well defined by A.S.Spanias in 1994 [1], K.Sayoood in Introduction to Data Compression 1996 [2], Saeed V. Vaseghi in Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications in 1996[3], John C. Bellamy in Digital Telephony in 1999 [4] and Wai C.Chu in Speech coding algorithms [5]. The historical perspective of the speech coding methodologies, the properties of speech signal and the performance measures used in speech coding are best explicated by W.B.Kleijn and K.K.Paliwal in 1995 [6].

1998 [9] illustrated the quality measurements that are important in low bit-rate speech coding systems and various distance measures. Standardization of speech coders had become important in daily life applications and various speech coders for low bit-rate applications were standardized by ITU and the standardization process for each coder are placed in plain words by R.V.Cox and P.Kroon in 1996 [10].

There are different types of low bit-rate coders namely Linear Predictive (LP) Coder, Code Excited Linear Predictive (CELP) Coder and Mixed Excitation Linear Predictive (MELP) Coder. The theoretical foundations of these low bit-rate coders are better explicated by Ming Yang in 2004 [11]. In this work LPC coder is used as it is the widely used low bit-rate coder. There are a wide variety of applications involving low bit-rate coders which are very important in applications involving band limited channels. During the past decade considerable progress has been made in the advancement of low bit-rate speech coders for both civilian, military communications and computer based voice applications. Richharia in Satellite communication systems: design principles 1999 [12] and Gibson in 2005 [13] gave a first-rate explanation about various types of functions involving low bit-rate coders.

There exists a wide diversity of coders that are classified based on the bit-rate and the coding techniques used. They are: waveform coders, parametric coders and hybrid coders. Purnhagen in 1999 [14] and E.Shlomot, V.Cuperman, A.Gersho in 1998 and 2001 [15-16] conferred the aspects with these coders and concluded that 4Kbps
hybrid coding scheme has comparable performance to code excited linear predictive (CELP) Coder. Linear predictive coder is a form of parametric coder which is the widely used low bit-rate coder. It is developed by the Department of defense, USA, in Federal standard 1015 for encoding human speech. Thomas E.Tremain in 1982 [17] explained plainly the LPC-10 algorithm. Speech coding algorithms written by Wai C.Chu [5] explained clearly the highlights of FS1015 LPC coder, methods used for extracting the parameters required by the FS1015 LPC coder and the limitations of LPC coder. LPC coder is a standardized low bit-rate coder and its narration is best given by Atal in 2006 [18] and this coder is commonly used by many researchers for low bit-rate applications. The linear predictive algorithm is used for extracting the feature parameters at the front end of a speech recognition system and is used as an analyzer to speech coding system. This is preeminently expressed by C.H. Lee in 1988 [19].

Makhoul in 1975 [20] expounded linear predictive coding as a digital technique used to encode an analog signal, where a value of the current sample is likely to be a linear combination of the past sample values. This corresponds to modeling the signal spectrum as all pole zero spectrum in frequency domain. Gunnar Fant in Acoustic Theory of Speech Production 1970 [21], Rabiner and Schafer in Digital Processing of Speech Signals 1978 [22], Paget in Human Speech 1999 [23] and Al-Akaidi in Fractal Speech Processing 2004 [24] explained the method of human speech production, the depiction of it using a
linear predictive coder, the relation between the vocal tract and the LPC filter and the information regarding pitch.

Speech analysis is a process of analyzing the speech signal, the parameters like pitch, gain, and voiced & unvoiced decisions of a frame are obtained by analyzing the speech signal frame wise and the analysis process is good given by Pei Hongwen and Shen Fengji [25], R. McAulay and T. Quatieri [26], Chin-Hui Lee [27], F.F. Tzeng [28], S. Zahorian and P. Gordy [29] and B.Yegnanarayana and P. Satyanarayana Murthy [30].

The extraction of linear predictive coefficients, their use in speech analysis and synthesis as filter coefficients, and the conversion of linear predictive coefficients to line spectral frequencies are given with a good detailed explanation from [31-33].

Voiced and Unvoiced decisions of a frame, the parameters used for making the decisions, their way of extraction and the way the decisions are made is explained clearly with good figures and block diagrams in [34-37].

In [38-45] the extraction of pitch for voiced and unvoiced frames of speech signal using different methods, the importance of pitch of in the synthesis of a speech signal is best illustrated.

Speech synthesis is a process of reconstructing the speech signal from the speech parameters received at the receiving end. The parameters gain, linear predictive coefficients, voiced & unvoiced decisions of a frame and pitch plays an important role in speech
synthesis and the process involved in speech synthesis is best enlightened in [46-49].

The most important part of a linear predictive coder is the quantizer which is used to quantize high bit-rate vectors to low bit-rate vectors. According to C.E. Shannon 1959 [50] quantizing a set of values (vectors) is more resourceful than quantizing individual scalar values when the dimension of a vector is large the rate distortion function of the vector quantizer comes close to the Shannon limit. In 1998 Gray and Neuhoff [51] published a survey paper reporting the ancient times of quantization starting from the origin and discussed various quantization techniques in a concise manner. In later days vector quantizers that can produce finest results even for vectors of smaller dimensions has been developed and given in detail by Makhoul, Roucos and Gish in 1985 [52], T.F.Quatieri in Discrete-Time speech signal processing 2004 [53], So.Stephen and K.K.Paliwal in 2007 [54].

The filter parameters used for the analysis and synthesis of speech signals are the linear predictive coefficients (LPC). The estimation of LPC parameters from the speech signal is given with good explanation by Rabiner and Schafer in Digital Processing of Speech Signals 1978 [22] and A.M.Kondoz in Digital Speech: Coding for Low Bit-Rate Communication Systems 2005 [31]. The parameters used for vector quantization are the line spectral frequencies (LSFs) so as to make an easy check of filter stability after vector quantization. After vector quantization if the samples of an LSF vector obey the ascending order

With the development of non-variational design algorithms such as the Linde, Buzo, Gray (LBG) algorithm, Vector Quantization (VQ) has become the powerful tool and is best explained with a numerical example by Y. Linde, A. Buzo & R.M. Gray in 1980 [59], later by A. Buzo, A.H. Gray, R.M. Gray & J. Markel in 1980 [60], N. Sugamura & N. Farvardin in 1988 [61], A. Gersho & R.M. Gray in 1992 [62], and J. Pan and T.R. Fischer in 1994 [63].

In early 1975 the properties of quantization is clearly explained by R. Viswanathan and J. Makhoul [64]. The usefulness of line spectral frequencies for vector quantization over linear predictive coefficients and the stability achieved by them are best explained by J. Zhou, Y. Shoham and A. Akansu [65], A.D. Subramaniam and B.D. Rao [66], F.K. Soong and B.H. Juang [67] and W.R. Gardner and B.D. Rao [68].

Vector quantizers have many advantages over scalar quantizers in terms of space filling, shape and memory requirements. These advantages are clearly elucidated in 2007 by So. Stephen and K.K. Paliwal [54]. There exists an ample variety of vector quantization techniques, the primary of which is the Unconstrained Vector Quantization technique. So. Stephen and K.K. Paliwal in 2007 [54]
explained Unconstrained Vector Quantization as a quantization technique used for quantizing vectors of full length. It is not a widely used technique as its computational complexity and memory requirements are very high [69-71].

The practical limitations of Unconstrained Vector Quantization technique is solved to a great extent using the product code vector quantization techniques. The practical limitations of the vector quantizer such as the computational complexity and memory requirements required for the search and storage of the codebooks are best alleviated using product code vector quantizers, which divide the process of vector quantization into parts, each part is quantized independently, J.Sabin and R.M.Gray in 1984 [72] explained the theory regarding the reduction of computational complexity and memory requirements and portrayed them as the measures of quality for a speech coder and highlighted the advantages of product code vector quantizers. The first product code vector quantizer is the Split Vector Quantizer (SVQ) in which vectors of larger dimensions are reduced to vectors of smaller dimensions by dividing a vector into parts. Split vector quantization is first proposed by K.K.Paliwal & B.S.Atal in 1993 [73] and explained the way of reducing the complexity. C.S.Xydeas and C.Papanastasiou in 1995 [74] proposed an extension to split vector quantization and discussed in detail the way of obtaining an LSF matrix and the method to obtain an individual sub matrices. In 2004 F.Norden and T.Eriksson [75]
explained the losses that arise from the lower dimension of a split and suggested conditional quantization as a method to reduce the losses.

The disadvantages of Unconstrained Vector Quantization technique is rectified by the development of Multistage Vector Quantization technique where the complexity and memory requirements are reduced. In 1982 B.H.Juang and A.H.Gray Jr [76] proposed Multiple stage Vector Quantization as an extension to unconstrained vector quantizer design and explained the way to operate at higher bit-rates without any increase in distortion and also explained the way to reduce the computational complexity and storage requirements. W.Y.Chan, S.Gupta and A.Gersho in 1992 [77] proposed a method for the joint design of codebooks at each stage for optimizing the overall performance but the performance has been improved at a modest rate without any effect in the encoding and storage complexity. W.P.LeBlanc, B.Bhattacharya, S.A.Mahmoud and V.Cuperman in 1993 [78] proposed a tree searched Multistage Vector Quantization scheme and a new joint codebook design scheme. This scheme has obtained a spectral distortion less than 1 dB, improved the speed of convergence and discussed the performance measures of vector quantization. J.Pan in 1996 [79] proposed a two stage pyramidal lattice vector quantization by introducing the tree structure into the first stage of vector quantization so as to reduce the complexity. In 2004 V.Krishnan, D.V.Anderson and K.K.Truong [80] presented channel-optimized Multistage Vector Quantization (CO-MSVQ) codec, in which the stage codebooks are jointly designed.
Using this method the quality of the reconstructed speech signal has been improved and the mean of the spectral distortion is also reduced.

The spectral distortion, computational complexity and memory requirements of Multistage Vector Quantizer is further reduced using the Split-Multistage Vector Quantizer (S-MSVQ) which is a hybrid of Split Vector Quantizer and Multistage Vector Quantizer and is outlined by So.Stephen and K.K.Paliwal in 2007 [54].

Another product code vector quantizer used to decrease the computational complexity and memory requirements of Split and Multistage Vector Quantizers is the Switched Split Vector Quantizer (SSVQ) which is a hybrid of Switch Vector Quantizer and Split Vector Quantizer. It is implemented in two ways hard decision and soft decision schemes. It was first proposed by So.Stephen and K.K.Paliwal. In 2005 So.Stephen and K.K.Paliwal discussed the implementation aspects of SSVQ in wideband speech coding and proved that SSVQ has better trade-off between bit-rate and spectral distortion performance compared to SVQ [81]. Later in 2007 So.Stephen and K.K.Paliwal [54] published a paper explaining the concepts of vector quantization, and compared the results with the existing vector quantization techniques such as Split Vector Quantization, Multistage Vector Quantization and explained that SSVQ is better compared to SVQ and MSVQ in terms of spectral distortion and computational complexity. S.Chatterjee & T.V.Sreenivas in 2007 [82] proposed a formulae for optimum bit allocation to Split Vector Quantizer and switch Vector quantizer and illustrated
parametric method is advantageous compared to non parametric method of Switched Split Vector Quantization. S.Chatterjee and T.V.Sreenivas in 2008 [83] alleviated the coding loss by exploiting the correlation between sub-vectors and proposed a saving of two bits per vector with the parametric method.

In voice banking to have a good quality input to the speech recognizer the noise present in the speech signal is reduced before it is applied as an input to the recognizer. S.F.Boll in 1979 [91] clearly explained the method of reducing acoustic noise and the way to reduce narrowband noise components. Also explained that this algorithm will not work when the signal and noise energies become equal and when the noise becomes highly non stationary. Later in 1994 Flores and Young [92] adopted a continuous spectral subtraction scheme in speech recognition application and improved the recognition accuracy for much noisy environments. M.Safayani, H.Sameti, B.Babaali and M.T.Manzuri Shalmani in 2007 [93] presented a method of adjusting the multi band spectral subtraction filter parameters based on the results of a speech recognition system and showed an improvement in the recognition rate using this method. In this work the speech signal quality produced using various vector quantization techniques has been examined by applying the coded outputs to a speech recognizer using Hidden Markov Model (HMM). L.R.Rabiner in 1989 [94] was the first to describe the Hidden Markov Model technique and clearly explained the problems in it. Vicente Palazon and Andres Marzal in 2007 [95] explained the concept
of linear Hidden Markov Model with a neat diagram which is the left-right model or Bakis model. Later Mikael Nilsson & Marcus Ejnarsson in 2002 [96] and Nick Bardici & Bjorn Skarin in 2006 [97] provided a good review of speech recognition using Hidden Markov Model technique and clearly explained the steps involved in speech recognition. A.S. Spanias and F.H. Wu in 1991[98] made a good review of different types of coders and briefly described the relation between speech coding, voice recognition and speaker independent recognition. Mayukh Bhaowal and Kunal Chawla in 2004 [99] explained the concepts related to the use of vector quantized linear predictive coefficients in speech recognition using Hidden Markov Model technique and also discussed about the scaling to be used in various algorithms used in HMM. Information regarding voice banking is placed in plain words by Robert Dale in 2002 [102] and explained the steps involved in voice banking with good flow charts.