Compression of Speech Signals using MSVQ and its Enhancement Using Spectral Subtraction & Kalman filter and its performance comparisions

M.Suman¹, Dr.Habibulla Khan², Dr.M. Madhavi Latha³ and D.Aruna Kumari⁴

¹,⁴ Department of Electronics and Computer Engineering, K.L.University, Vaddeswaram, Guntur
⁵ suman.maloji@kluniversity.in and aruna_D@kluniversity.in

²,⁶ Department of Electronics and Communication Engineering, K.L.University, Vaddeswaram, Guntur
² habibullaiah@rediffmail.com

³ Department of Electronics and Communication Engineering, JNTUCE, Hyderabad
³ mlmakkena@gmail.com

Abstract: Coding algorithms seek to minimize the bit rate in the digital representation of a signal without an objectionable loss of signal quality in the process. Speech enhancement means improvement in intelligibility and/or quality of a speech signal. This paper deals with multistage vector quantization technique used for coding of narrow band speech signals. The parameter used for coding of speech signals are the line spectral frequencies, so as to ensure filter stability after quantization. The code books used for quantization are generated by using Linde, Buzo and Gray(LBG) algorithm. The results of the multistage vector quantizer are compared with unconstrained vector quantization Technique. The performance of quantization is measured in terms of spectral distortion measured in dB, Computational complexity measured in KFlops and Memory Requirements measured in Floats. From the results it can be proved that multistage vector quantization is having better spectral distortion performance, less computational complexity and memory requirements when compared to unconstrained vector quantization. Speech enhancement is a special case of signal estimation as speech is non stationary and hence human ear is the final judge and we does not believe in mathematical error criterion. The compressed speech signals are enhanced using spectral subtraction and Kalman Filter the outputs are compared with original signals with 0dB & 10dB noise.

KEYWORDS: Linear predictive Coding, Multi stage vector quantization, Line Spectral Frequencies (LSF), Spectral Subtraction.

I. INTRODUCTION

Speech has arguably been most important form of human communication. Since languages were first conceived over the ages, many forms of communication have been developed to convey information across a distance, but the relatively recent invention of the telephone has revalorized this process. The demand for efficient communication & data storage is continuously increasing. The purpose of speech coding research is to address the problem of accommodating more users over such limited capacity by coding speech before transmitting it across a network. As suggested by the research scholar M. satya sai Ram [5][6][7]

The advantages with coded speech signals are:
- Lower sensitivity to channel noise
- Easier to error-protect, encrypt, multiplex and packetize.
- Efficient transmission over bandwidth constrained channels due to lower bit rate.

The quantization technique should have less computational and memory requirements and it should not result in suboptimal quantization performance intelligibility. Speech coders operating at low bit rates necessitate efficient encoding of linear predictive coding (LPC) coefficients. Line spectral frequencies (LSF) parameters are currently one of the most efficient choices of transmission parameters for the LPC coefficients.

Multi Stage Vector Quantization (MSVQ) can achieve very low encoding and storage complexity in comparison to unstructured vector quantization. However, the conventional MSVQ is suboptimal with respect to the overall performance measure. This paper proposes a new technology to design the decoder codebook, which is different from the encoder codebook to optimize the overall performance. The performance improvement is achieved with no effect on encoding complexity, both storage and time consuming, but a modest increase in storage complexity of decoder. Speech coding is the compression of speech (into a code) for transmission with speech codecs that use audio signal processing and speech processing techniques.[6]

The aim of this paper is to provide a general review of MSVQ, and to compare its performance with unconstrained vector quantization technique and Kalman filter. The practical limitations, Regarding computational complexity and memory requirements as a function of bit rate are discussed.
II. DIMENSIONS OF PERFORMANCE IN SPEECH COMPRESSION

Speech coders attempt to minimize the bit rate for transmission or storage of the signal while maintaining required levels of speech quality, communication delay, and complexity of implementation (power consumption). We will now provide brief descriptions of the above parameters of performance, with particular reference to speech. [4]

Speech Quality
Bit Rate
Communication Delay
Complexity

III. MULTISTAGE VECTOR QUANTIZATION

Several techniques can be employed in calculating the codebooks in MSVQ design. The simplest method is to train the codebooks sequentially. The codebook for the first stage is computed in a traditional manner using, e.g., GLA and the training data is quantized with the obtained one-stage vector quantizer. The resulting quantization error vectors are used as the training data for the second stage. This is repeated for all stages, with each new codebook trained using the error between the original and the reconstructed vectors including all the previous stages.

LPC quantization performance, ideally it should be used to design the vector quantizer. However, it is very difficult to design a vector quantizer using this distortion measure. Therefore, simpler distance measures (such as the Euclidean and the weighted Euclidean distance measures) between the original and quantized LPC parameter vectors (in some suitable representation such as the LSF representation) are used to design the LPC vector quantizer. To find the best LPC parametric representation for the Euclidean distance measure, the study of the 2-stage vector quantizer, the distance measure in the following three domains: the LSF domain, the arcsine reflection coefficient domain and the log-area ratio domain is done. The 2-stage vector quantizer performs better with the LSF representation than with the other two representations. [1][2]

The Euclidean distance measure used for vector quantization in the preceding section provides equal weights to individual components of the LSF vector, which obviously are not proportional to their spectral sensitivities. Paliwal and Atal [10] have proposed a weighted Euclidean distance measure in the LSF domain which tries to assign weights to individual LSFs according to their spectral sensitivities. The weighted Euclidean distance measure between the test LSF vector \( f \) and the reference LSF vector is given by

IV. SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION

Speech enhancement is an area of speech processing where the goal is to improve the intelligibility and/or pleasantness of a speech signal. The most common approach in speech enhancement is noise removal, where we, by estimation of noise characteristics, can cancel noise components and retain only the clean speech signal. The basic problem with this approach is that if we remove those parts of the signal that resemble noise, we are also bounded to remove those parts of the speech signal that resemble noise. In other words, speech enhancement procedures, often inadvertently, also corrupt the speech signal when attempting to remove noise.

Algorithms must therefore compromise between effectiveness of noise removal and level of distortion in the speech signal. Current speech processing algorithms can roughly be divided into three domains, spectral subtraction, sub-space analysis and filtering algorithms[1][2]

Figure 2 Block Diagram Spectral subtraction

Spectral subtraction algorithms operate in the spectral domain by removing, from each spectral band, that amount of energy which corresponds to the noise
contribution. While spectral subtraction is effective in estimating the spectral magnitude of the speech signal, the phase of the original signal is not retained, which produces a clearly audible distortion known as “ringing”.

One of the most popular methods of reducing the effect of background (additive) noise is Spectral Subtraction. Suppose your speech signal \( x(n) \) is corrupted by background noise \( n(n) \); that is:

\[
Y(n) = x(n) + N(n)
\]

Where \( y(n) \), \( x(n) \) and \( N(n) \) are the signal, the additive noise and the noisy signal respectively, and \( n \) is the discrete time index. In the frequency domain, it is expressed as In the frequency domain, it is expressed as

\[
Y(f) = X(f) + N(f)
\]  

A. Subtracting the Noise Spectrum

The effect of additive noise on the magnitude spectrum of a signal is to increase the mean and the variance of the spectrum. The increase in the variance of the signal spectrum results from the random fluctuations of the noise, and cannot be cancelled out. The increase in the mean of the signal spectrum can be removed by subtraction of an estimate of the mean of the noise spectrum from the noisy signal spectrum.[2] The noisy signal model in the time domain is given by

\[
y(n) = x(n) + n(n)
\]

Where \( y(n) \), \( x(n) \) and \( n(n) \) are the signal, the additive noise and the noisy signal respectively, and \( n \) is the discrete time index. In the frequency domain, it is expressed as

\[
Y(f) = X(f) + N(f)
\]

Where \( Y(f) \), \( X(f) \) and \( N(f) \) are the Fourier transforms of the noisy signal \( y(m) \), the original signal \( x(m) \) and the noise \( n(m) \) respectively, and \( f \) is the frequency variable. In spectral subtraction, the incoming signal \( x(m) \) is buffered and divided into segments of \( N \) samples length. Each segment is windowed, using a Hanning or a Hamming window, and then transformed via discrete Fourier transform (DFT) to \( N \) spectral samples.

\[
yw(n) = w(m)y(m)
\]

\[
= w(m) [ x(m) + n(m) ]
\]

\[
= xw(m) + nw(m)
\]

A. Wiener filter: This filter is the precursor of Kalman filter. The goal of Wiener filter is to remove the noise from a corrupted signal. The Wiener filters are characterized by the following concepts: Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation. Requirement: the filter must be physically realizable, i.e. causal (this requirement can be dropped, resulting in a non-causal solution). Performance criteria: minimum mean-square error.

B. Kalman filter

Kalman filter is simply an optimal recursive data processing algorithm. If we focus on the word optimal, its definition depends on the criteria chosen to evaluate. A feature is called optimum if the Kalman filter incorporates all the information provided. It processes all the measurements available, regardless the precision, to estimate the current value of the interest variables, using: The performance of quantization is measured in terms of Spectral distortion measured in dB, Computational complexity measured in Kflops and Memory Requirements measured in Floats.

VI. RESULTS

The performance of quantization is measured in terms of Spectral distortion measured in dB, Computational complexity measured in Kflops and Memory Requirements measured in Floats.

TABLE –I: SPECTRAL DISTORTION FOR MSVQ
Tables I, II and III shows the spectral distortion (dB), computational complexities (Kflops/frame) and memory requirements (floats) at various bit rates for unconstrained and three stage multistage vector quantizer. From Table-I it is observed that MSVQ has better spectral distortion performance. From Table-II &III it is observed that MSVQ has less computational complexities and memory requirements when compared to unconstrained vector quantization Technique. The code books used for quantization are generated by using Linde, Buzo and Gray (LBG) algorithm.

Experimental Outputs

Table-I: Spectral Distortion and Computational Complexities at Various Bit Rates

<table>
<thead>
<tr>
<th>Bits / frame</th>
<th>SD(dB)</th>
<th>2-4 dB</th>
<th>&gt;4 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>24(8+8+8)</td>
<td>0.984</td>
<td>1.38</td>
<td>0</td>
</tr>
<tr>
<td>23(7+8+8)</td>
<td>1.238</td>
<td>1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>22(7+7+8)</td>
<td>1.345</td>
<td>0.85</td>
<td>0.13</td>
</tr>
<tr>
<td>21(7+7+7)</td>
<td>1.4</td>
<td>1.08</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table-II: Complexity and Memory Requirements for Unconstrained Vector Quantization

<table>
<thead>
<tr>
<th>Bits / frame</th>
<th>Complexity (Kflops/frame)</th>
<th>ROM (floats)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24(8+8+8)</td>
<td>30.717</td>
<td>7680</td>
</tr>
<tr>
<td>23(7+8+8)</td>
<td>25.597</td>
<td>6400</td>
</tr>
<tr>
<td>22(7+7+8)</td>
<td>20.477</td>
<td>5120</td>
</tr>
<tr>
<td>21(7+7+7)</td>
<td>15.357</td>
<td>3840</td>
</tr>
<tr>
<td>20(6+7+7)</td>
<td>12.797</td>
<td>3200</td>
</tr>
<tr>
<td>19(6+6+7)</td>
<td>10.237</td>
<td>2560</td>
</tr>
<tr>
<td>18(6+6+6)</td>
<td>7.677</td>
<td>1920</td>
</tr>
</tbody>
</table>

From Table-I, II & III that as the number of bits/frame decreases, the complexity and memory requirements are also decreased but the spectral distortion has increased and transparency in quantization is achieved at 24 bits/frame.
VII. CONCLUSIONS AND FUTURE WORK

Speech coding is a method of reducing the amount of information required to represent a speech signal. In this paper two methods of speech coding techniques i.e. unconstrained vector quantization and multi stage vector quantization are analyzed. From results it can be concluded that Multi Stage Vector Quantization is having the less computational complexity & memory requirement when compared to unconstrained vector quantization. But the Spectral distortion performance of the Multi Stage Vector Quantizer is better when compared to unconstrained Vector Quantizer. The decreasing computational complexity & memory requirement with multi stage Vector Quantizer is due to less availability of bits at each stage of quantizer. Fig 5&6 represents that original speech signal at 0dB and 10dB noise is compressed using MSVQ . These compressed outputs are first enhanced using spectral subtraction. As per the fig 7&8 we can observe that spectral subtraction has better performance at 0dB . Secondly the same compressed signals at 0dB and 10dB are enhanced using kalman filter. It is observed from fig 9&10 that kalman filter has better performance at 10dB .

As per the experimental outputs [fig 7] spectral subtraction has the better performance at 0dB and kalman filter [fig 10] provides better performance at 10 dB noise.

These speech enhancement methods employ similar principles as speech recognition, speech coding and general compression techniques. Application of some filtering techniques enhances the signal further so that total signal can be restored back. An analysis has been done to verify the systematic approach of speech enhancement and hence it is proved that it is better way is to First Compress the signal and then Enhance the signal.

This paper can be further extended with different noise signals and the performance characteristics of spectral subtraction and Kalman filter can be compared after compression.

VIII. REFERENCES